Editorial

Mondo Digitale is a long-standing and well-acknowledged AICA-sponsored publication within the Italian ICT magazines scenario. The mission of Mondo Digitale is to be a reference publication for practitioners, enthusiasts and interested Readers wishing to keep the pace with the recent ICT advances. Within this framework, this special issue of Mondo Digitale opens a window on a selection of the scientific works presented at the last AICA Conference, held in Fisciano (Salerno, Italy) from September 18 to 20, 2013, on the Theme "Digital Frontiers, from Digital Divide to Smart Society" (Frontiere Digitali: dal Digital Divide alla Smart Society) http://www.aica2013.it/ . Therefore, besides their scientific results and advances, the papers that the Reader will find in this issue draw a faceted picture of activities currently in progress inside main national Universities as well as Research and Teaching Institutions.

The 2013-AICA-Conference attracted more than 150 contributions by 210 Authors from more than 45 different affiliations, and was organized in scientific sessions, spanning, among all, Cloud Computing, e-Government, e-Health, Smart-Cities, Biometrics, and e-Learning.

Fifteen papers, reshaped in their extended version, after a further peer-review, have been selected by the Editorial Board of Mondo Digitale in collaboration with the Program Committee of AICA, and listed below.

As the Reader will notice, this issue of Mondo Digitale has taken the chance to introduce an additional innovation. To meet the constant evolution of modern specialized publishing, realized in the context of an ever-growing editorial competition, which goes beyond national borders, involving common interests in many English language countries, Mondo Digitale comes out to-day with its first English version. This makes it possible to further enhance the visibility opportunities provided by Scopus indexing, which was granted to Mondo Digitale more than a decade ago, in 2003. This choice allows deeper insights and spreading of scientific achievements, making them immediately fully accessible to the most reputed world research communities.

As Readers in enterprises, schools and academies will immediately perceive, a new trend has started in this issue of Mondo Digitale. Still keeping in mind its leadership and service roles inside the Italian scenario, Mondo Digitale is now widening its horizons and perspectives, in an effort to cross the National borders and to reach also an international Readership.

The Editorial Board strongly believes that, as knowledge sharing and dissemination are border-free, Mondo Digitale could take this unique opportunity to present and promote the Italian research to the international communities, as well as to help stimulating the collaboration of Italian researchers with Colleagues from other Countries.

As Authors we thank you for your collaboration, and as Readers we thank you for your interest in Mondo Digitale. Please do not forget to send us your feedback, reviews and suggestions at mondodigitale@aicanet.it.

Claudio Giovanni Demartini
Full Professor at Politecnico di Torino

Network Access: Social Law Implied in the Constitution for the Use of an Active Citizenship

V. Amenta

Abstract. *This article raises the configuration of the right of access to the Internet as a new law, which, although does not have an explicit regulatory approval, has a strong tone constitutional. The objective will be just what the legal status of the right of access to the Internet, which is proposed in its intrinsic structure and biphasic. From one side it is configured as a fundamental right and thus comprise all activities relating to the right to be informed and to inform. On the other hand, the Internet presents itself as social right instrumental to the exercise of other rights. This right can be claimed by members of a given community against their political authorities. But in national experiences such as the Italian, the formal equality, it seems inadequate to implement the principle of equality for the presence of subjective situations and unequal starting positions that benefit certain subsidiaries against others.*

Keywords: access, breadband, network.

1. The right of access to the Internet

In a context where information, computerization and globalization of processes, rules and networks now seem irreplaceable keys of access to freedom, progress and democracy, the sensation is that a new category of rights has been born and is developing within our society [Bovero, 2004].

The new law resulting from the evolution of technology is the right to digital freedom, which manifests a new aspect of the ancient idea of personal freedom and reflects the advance of new frontiers of human freedom towards a future society, and which takes its place in the prism of contemporary constitutionalism.

In the primitive version which dates to 1981, digital freedom was configured with a dual meaning, positive and negative.

The negative digital freedom expresses the right not to make public certain information of a personal, private, confidential nature; however, positive digital freedom expresses the ability to exercise the right of control over personal data

which have escaped from the circle of privacy and become input elements of an electronic program; and thus positive freedom of information, or recognized subjective right, to know, to correct, to remove or add data to a personal e-file.

A careful doctrine states that the right to freedom of information takes on a new form of the traditional right of personal freedom, such as the right to check information about one's own person, as a right of *habeas data* [Frosini, 2011].

The goal of *constitutional habeas data* is to ensure freedom of information, as a personal guarantee to know and access personal information existing in databases, to control their content and then modify them in case of inaccuracy or improper storage or treatment and to decide on their circulation [Rozo Acuna, 2002].

With the advent of the Web, the right to digital information has become a claim to freedom in the active sense, not freedom *from* but freedom *of*, which is to be able to utilize computer tools of all kinds. It is the right to participate in a virtual society, created by the advent of computers in a technological society: it is a society of mobile components and dynamic relationships, in which each individual participant is sovereign of his own decisions. We are dealing with a new kind of freedom, that is to communicate with whoever you want, disseminating one's own ideas, thoughts and materials, and the freedom to receive. It thus delineates the freedom to communicate as the freedom to transmit and receive.

In this context, based on a technological concept of freedom of communication, the contents of traditional constitutional freedoms are struggling and slow to emerge, in particular those regarding communication and expression.

The matter of fundamental rights is presented as a meeting point between issues of great importance, including the definition of subjective rights, the concept of constitution and the meaning of democracy. Democracy in the 21st century takes a different form from that of previous centuries: the meanings of representation and sovereignty change, advancing a new mass democracy, which breaks the closed circles of the *élites* in power, forcing so to speak representatives of the will of the people to descend on the telematic marketplace and deal directly with representatives, in the new forms of technical policy. More than anything else, the aim of current democracy, in the field of fundamental rights remains that of undertaking a precise blend of external legal experience, typical of so-called normal citizen, and the experience of internal law, its classic operators of law. The aim is very ambitious; i.e., to return to the citizen an awareness of his primary role in this phase of changing the law.

When one uses the term "fundamental rights" it refers generally to the phrase "human rights." However, if on one hand human rights aspire to universalization, and are protected by international regulations, on the other hand fundamental rights are not simply individual rights, but subjective rights that perform a "functional" task of a specific rule of constitutional law.

Within the issue under consideration, this work explores the problem of protecting recent digital rights, belonging to the so-called fourth-generation rights, particularly with regard to one law, the so-called *right of access to the*

network, which is currently configured as the most innovative law, but is slow to emerge as a fundamental guaranteed right.
The content of the right to electronic access is not only broader than traditional forms of the freedom to communicate, but, on the contrary, seems to have a support function with respect to a long list of traditional rights. And it is in fact the new structure of telecommunications that has reinforced the idea and the need to think about the right to Internet access, understood as a sort of mother-right, compared to that which will be acknowledged in order to connect to the Internet.

It is correct to believe, as stated in an authoritative doctrine [Zeno-Zencovich, 2004], that the social role of electronic access cannot be underestimated since in modern times it is equivalent to not providing everyone with the necessary conditions to establish relationships with others and to ensure the full expression of one's own personality.

Among the umbras indicated by jurists in the emergence of the right to Internet access as law, is the need to qualify this right; that is, if it can be placed within the fundamental rights set forth in our Constitution, specifying whether it is the case of a fundamental right or social right, or if it should receive autonomous legal regulation.

1.1 Right of access to the network: social right or fundamental right implicit in the Costitution?

The interpreter who chooses to follow the first route cannot avoid starting from the consideration of Articles 15 and 21 of our Constitution.
The basis of these articles lies in the awareness that regardless of gender, race or religion, everyone has the right not only to the freedom to express their thoughts, but also the freedom to gather information. These elements are essential for the solid development of the personality of each individual, and therefore for genuine equality between citizens and the ability of each person to participate in social life. Article 15 of the Constitution recognizes the right of individuals and social groups to communicate their thoughts to one or more subjects and at the same time guarantees confidentiality.
Constitutional protection under art. 15 is lost in the broader protection of communication in general, as if to compose a perfect triptych with Articles 13 and 14 of the Constitution, expression of the three aspects (physical, spatial and spiritual) of the inviolability of the human person [Di Lello, 2007].

The objective scope of protection of Article 15 is precisely that of private communication, consisting of both correspondence in the strict sense or any other form of interpersonal communication.

Freedom of information is understood not so much as the freedom to seek and obtain certain news, being concerned with the apprehension of knowledge in general, as the use of information sources, regardless of the information contained in them.

In order to speak of communication, in accordance with our Constitution, we can detect the *animus* of the subject of confidentiality and the *determinability* of the persons to whom our communication is addressed.

However, we must understand whether the communication model created by the Internet can make it difficult to distinguish between private communication (within art. 15 of the Constitution) and public expression of thought, which is protected by art. 21 of the Constitution.

Most of the doctrine holds that the Internet medium will benefit, as appropriate, from the guarantees of the means of expression and *sic et simpliciter* or additional collateral guarantees granted to the media for interpersonal communication. Thus thought directed to specific recipients through software that ensures confidentiality will fall within the guarantees of Article 15; on the contrary, we would not be dealing with a confidential communication if the user is using forms aimed at the public demonstration of his own thought. And in fact, in a communication posted in a forum, chat or blog, the typical guarantees of freedom of expression or the regime of freedom of assembly could well apply [Pubusa, 2006].

If one prefers instead a merely literal interpretation of Article 21, the freedom of expression, its existence and its exercise would be remitted to the person who expresses their thoughts, and in particular, with their consent, to the acquisition of information related to the formation and expression of their own thoughts. The scope of freedom of information would then be determined by a person other than its owner, and the protection would be only occasional, applied only if the freedom of expression is violated. Asking whether it is possible to appeal to a right that appears to be only proclaimed but not also "protected", is to ask whether the right "exists", also in the case that the law had not given it a precise formation.

The provisions codifying fundamental rights are usually formulated in extremely general and indeterminate terms. Faced with lists of rights formulated thus, the question arises whether it is perhaps inevitable that fundamental rights have been proclaimed in provisions formulated in a generic and indeterminate way [Ferrajoli, 2010].

Thus, the identification of a new law, such as the right of access, may be achieved by simply using a non-restrictive reading of the basic types of fundamental rights. It might be postulated that the fundamental rights constitute a *genus* with an inherent capacity for expansion, which is duplicated within in accordance with the guidelines implied for that category. In fact, fundamental rights exist before the State and thus do not depend on the law, but constitute a limit to the free production of laws. In jurisdictions with long and pluralist constitutions, the provisions that codify fundamental rights are usually formulated in extremely general and indeterminate terms. A constitutional provision providing for fundamental rights in strictly circumstantial terms would be unreasonable. Thus, a fundamental right could be considered the implicit assumption of certain explicit fundamental rights: the latter cannot be explained if you do not affirm the first. However, in the Italian constitutional order, the creation of implicit fundamental rights is greatly facilitated by the presence of Article 2 of the Constitution, usually interpreted as a meta-rule that allows the

identification of all rights considered essential to the development of the individual [Pino, 2010].

From this arrangement there may derive a broad and systematic interpretation not only of art. 21, but of all constitutional precepts related to the problem of information, in order to make the related freedom autonomous and distinct from freedom of expression: however, this is not to deny the natural connections that unite the expression of thought to the acquisition of knowledge, but to find an approach that by distinguishing between legal regimes, enhances both of them, so that they can be mutually reinforcing.

In the light of this interpretation, Article 21 of the Constitution could be read as if it were written in the following way: everyone has the right to use any source of information (and the Internet is a great means of dissemination of information) available in order to spread one's own thinking.

What has been said so far is thus not opposed to extending the existing constitutional system to the Internet, given that the closing words of art. 21, any other means of communication, has allowed the constitutional provision to survive the advent of new means of mass communication.

The freedom of art. 21 of the Constitution is guaranteed, according to the wording of that article, to "anyone", and it follows that the access to the medium must be recognized for everyone in their geographic location, regardless of social status.

Thus, there may be no obstacle to the application of constitutional law to electronic communication.
Moreover, if one considers that the extension of freedom of expression is influenced not only by the content of the message, but also the means by which it is exercised, the Internet may adhere to the spirit of the constitutional provision better than any other means.

The perspective outlined up to now refers to the manifestation of thought. Instead, the perspective of an individual "surfing" the Web to search for news is different. In this case, the user's activity does not fall within the activity of expressing thought in the strictest sense, and yet it can equally fall under the regime of article 21 by virtue of the corollary of the principle of the right to information.

In fact, corollaries of the freedom of information are the right to inform oneself, understood as the right to seek information (that is, the right to access information through any means that contains it, electronic or not) and the right to be informed, that is the right have information sources available. Given their essence, these two corollaries cannot exist if not related in the sense that the former cannot exist without the second, and the latter has no reason to be without the other.

Thus, in reference to electronic communication, the right to inform oneself does not refer to a right attributable to the subject to receive information with a particular content, but to the absence of restrictions on Internet access to research information already present on the Web. Therefore, this is not a demand, but a simple freedom of the subject to not be denied access [Corte Costit., 1969].

From this point of view, the right to surf the Internet would equally be guaranteed as a right to seek information, as a necessary and indispensable corollary of freedom of information, but the guarantee does not cover the search for information that is not available.

The freedom of information is not limited however to communication and information. It also embraces political freedom and institutional organization, which also incorporates the use of services by users/citizens. It is clear that technological progress is increasingly destined to change the institutional apparatus experienced and that the democratic process is strongly influenced by the way information circulates, that is, where the possibility of its use of by all citizens is characterized by being a prerequisite of that process. Thus, electronic freedom provides impetus to the democratic principle and that of impartiality, which have their constitutional origin in Article 3 of the Constitution, which must be integrated with Articles 1 and 2 of the same constitutional text. By analyzing and following the circular path that starts from Article 1 and arrives at Article 3, three key concepts are deduced that address the entire question: inclusion, participation, and public policy. These expressions reflect the imperative – the binding principle -- inherent in art. 3 , paragraph 2 , which requires a constant connection between aims and organizations, the founding values of the constitution, and from which follows a structure of public authorities who implement them. Article 1, paragraph 2 establishes that sovereignty belongs to the people who exercise it in the manner prescribed by the constitution in various forms, to arrive at art. 3, paragraph 2 that provides for the removal of obstacles to allowing the full development of the individual for the purpose of effective participation in the political, economic and social development of the country. It is thus a full circle: the exercise of sovereignty, and in particular of individual rights, is in some way aimed at a person's relatedness to allow full participation, i.e., an active and responsible citizenship.

The first regulatory response and the first representation of the individual in this new role as digital citizen is to be found in the Digital Administration Code (CAD), approved by the d.lgs.82/2005. This legislation identifies and outlines a “species” of digital citizen, which adapts to both natural and legal persons and which rests on a double strand of rights and obligations, or the right to demand from public offices interaction in digital mode and the obligation of the administration to adopt adequate facilities to fulfilling the user’s request.

Among the various standards, emphasized here are administrative procedures, the right to make electronic payments with central government, the right to quality of service in terms of information and communication technology, and the right to communicate through email. This budding relationship between citizen and administration must be placed within the context of a broader social law for which the Italian State must be spokesman in order to ensure fruition by all users, assuming the right to education and the enhancement of computer culture. These last rights are considered by the Constitutional Court as "*corresponding to objectives of general interest, that is the development of culture, in this case through the use of the software tool, the pursuit of which is headed by the Republic in all its ramifications (art. 9 Cost.)* [Corte Costit., 2004]".

The right of access considered in this way therefore qualifies as an instrumental social right, which allows the use of other traditional social rights. This new form of participation, rather than appearing as a new phase in the development of rights, is emerging as an alternative mode of access to them. Technological innovation has opened the way for a new understanding of citizenship. In essence, the exercise of many basic rights has been facilitated by the direct relationship between the citizen and the institutional apparatus, mediated by tools that by their very nature fall outside the physical space in which the relationship between them usually takes place.

E-citizenship, by virtue of its intangible nature, would seem to offer unprecedented opportunities to exercise civil and political rights that until now remained only on paper and no less, to claim new types of rights. It thus offers a new possibility for the exercise of social, economic and political rights.

As has been noted, Citizenship in the electronic age not only presumes widespread digital literacy, but requires the concrete possibility of easy access to the Internet . Today, the new citizenship is the right to not be excluded from use and benefit of public structures and in particular the use of online resources that these structures can offer. Inclusion and access are essential aspects of the new citizenship [Masucci, 2003].

On one hand there are those who have extolled the benefits of the technology implications of the network, claiming that it will be able, by itself, to enrich their human capital and ultimately, to improve their lives; on the other, there are those who stress that the uneven diffusion of the Internet among the population will result in increasing inequality, improving the prospects of those who are already in privileged positions and at the same time denying opportunities for advancement to the non-privileged.

1.2 Freedom of information vs the right of access to the Internet in the guidelines of the Courts

The findings outlined above are further corroborated by the same positions taken by the Constitutional Court, when this expressed its own evaluations, creating the basis for an information law that has helped enrich the paradigm of constitutional freedom, especially in light of the views expressed in the European context.

The guidelines outlined by the Court on one hand emphasisized the fundamental nature of article 21 of the Constitution, intended as a "cornerstone of democracy," and on the other hand, refined the breakdown between content regulation and the the means of mass communication, assessed as " services that are objectively public or otherwise in the public interest. [Corte Cost., 1977]"

In the guidelinesof this jurisprudence, the assumption was created that the "*right to information*" has increasingly come to connect with a general theory of democracy. From this we can affirm the significance of co-essentiality that the Court, in numerous decisions [Corte Costit., 1972], has referred to the relationship between freedom of expression used for information purposes and the form of the democratic state, whose essence implies "*plurality of sources of*

information, free access to the same through all forms of communication, absence of unjustified legal obstacles to the circulation of news and ideas."

The doctrine did not fail to highlight the "limited and dated content of this discipline to the point of referring to it as a kind of institutional myopia, due to the fact that it had targeted the experiences of the past, far more than future prospects".

The Constituents, when they first confronted the issue of freedom of expression and the means of its exercise, were moved by the desire to remove or restrict the instruments of control over the press that had been disseminated by the Fascist regime, such as authorization, censorship or seizure of printed material, rather than outline a information system that was defined based on the possible development of the pluralist democracy they were building. And it is for this reason that Article 21 of the Constitution is not given to search any reference to the means of mass communication which were already established. These are discrepancies which -- as pointed out by an authoritative doctrine – become even more important if we are "to compare the content of art. 21 with the formulations expressed in other international constitutional documents contemporary to our republican charter", such as those contained in Article 19 of the Universal Declaration of Human Rights, adopted by the UN in 1948 (which recognizes the right of every individual to "*seek, receive and impart information and ideas through any media*") and Art. 10 of the European Convention for the Protection of Human Rights and Fundamental Freedoms, signed in 1950, which includes the freedom of expression, "*the freedom to receive and impart information and ideas without possibility of interference from the public authorities*."

The scheme, at the time set out for printing, could be extended by analogy to the Internet medium. Several court rulings can aid in this. For example, there is the well-known judgment of the Court of Milan [Court of Milan, 2010] pronounced in the case of Google vs Vivi Down, which recognized that the legal status of the service provider is equivalent to that of a publisher of mass-media. The decision originated with an incident in September 2006, in which the platform Google-Videos uploaded a video of a boy with Down's syndrome who was beaten up by some classmates. The footage was left on the platform for about 2 months and seen by about 5,000 people before being removed by the service provider, after a warning from the judicial police. The video involved the students who had abused the boy and a teacher from the school. The latter were accused of having insulted the reputation of the association "Vividown" and the student's disability, in accordance with Article 110, paragraphs 2 and 385, and paragraphs 1 and 3 of the Penal Code, and to have omitted the correct treatment of personal data pursuant to art. 167 of Legislative Decree no. 196/2003. The Court of Milan sentenced Google for unauthorized use for commercial purposes of sensitive data belonging to other people, in the same way as if it were a publisher of mass media.

In addition to the experience of our country's Courts we cannot neglect foreign verdicts in the U.S. and French contexts.

Among the former was a ruling by the Federal Court of the District of Pennsylvania in 1996 (and later the U.S. Supreme Court in 1997). The

antecedent of the decision is represented by the *Communications Decency Act*, an act contested in Federal Court in 1996, which provided penalties for users that introduced Web content considered morally inappropriate.
In its verdict the Court held that the free use of the Web is protected by the First Amendment regarding the freedom of religion, speech and press, and the interest to stimulate freedom of expression in a democratic society was superior to any alleged benefit of censorship. The matter was referred to the Supreme Court which upheld the decision of the federal courts and adopted the decision by which the judges expressed comments regarding the Internet and more specifically the relationship between the Internet and constitutional freedoms. Thus, after a well-articulated debate the judges of the Supreme Court confirmed the decision of the unconstitutional nature of the law since it conflicted with the First Amendment of the Constitution, thereby highlighting the aspect of the expression of thought instead of that concerning the confidentiality and privacy of communication. In this sense, the final impact of their motivation is important: *"The facts established show that the expansion of the Internet has been and continues to be phenomenal. It is the tradition of our constitutional jurisprudence to presume, in the absence of evidence to the contrary, that public regulation of the content of the manifestations of thought is more likely to interfere with the free exchange of ideas than to encourage it. The interest in fostering freedom of expression in a democratic society is superior to any alleged, but not demonstrated, benefit of censorship*."

From this ruling emerged the connection of network access to the exercise of fundamental freedoms. It should be noted that the Supreme Court, in using the First Amendment as a parameter of the unconstitutional nature of the law repressing freedom on the Internet, revived it, giving it a new meaning, which is not and cannot be the original one, because age of technology it not only protects the traditional right to freedom of thought, but also electronic freedom of speech, electronic freedom of the press, and freedom of assembly. Thus, the first amendment to the Constitution affirms and guarantees the right to freedom of information, as a new constitutional right of freedom obtainable from traditional rights and constitutional principles which should be read and interpreted in the context of a technological society.

Of course, in the light of the above, the role of judges will be increasingly reinforced and empowered, since it is for them to reinterpret the old constitutional tradition in the light of technological innovation. Thus, it is up to them to tackle the task of how to be jurists in a technological society, involved by dwelling now in the new world of the age of automation and to live with the legal problems that arise from it. This new condition of jurists, who now participate in the two forms of activity, the humanities and technology, reflects the general condition of man today.

Equally significant is the second case, of French jurisprudence and featuring the *Conseil Costitutionnel*. The French Constitutional Council was called upon to rule on the constitutionality of law n. 2009/669, known as the HADOPI Law, regarding the diffusion and protection of creative work on the Internet. The fragment of interest here is that in which the French constitutional court considered access to

the Internet in the same way as the principle of freedom of expression embodied in Article 19 of the Universal Declaration of Human Rights of 1948 and art. 11 of the Declaration of the Rights of Man and Citizen of 1789, still in force in French law. The judges of the Conseil Costitutionnel, based on those provisions of the Statutes, have been able to affirm that the right to communication also includes the freedom to access the services of on-line communication, tools for participation in democratic life and for the expression of ideas and opinions.

Thus, an individual's freedoms also include access to the Web. Therefore, according to the Council, the penalty of disconnection from the Internet for acts of piracy cannot be imposed by an administrative order, but requires a judicial decision, as is the case for the limitation of other personal freedoms [De Marco, 2005].

Moreover, the original text of the Hadopi law would have been inconsistent with the position of the European Parliament, which in its Lambridis Recommendatio invited Member States to exclude preventive and generalized measures to limit the rights of citizens on the Internet, including disconnection from the Internet. In compliance with the Lambridis recommendation, in May 2009 the European Parliament suggested the adoption of Amendment No 138/46 regarding the package of the five directives on the transformation of the European telecommunications sector (the *Telecom Package*) with the aim of identifying access to the Internet as a fundamental right of end users. This large European telecommunications reform came into force in our legal system with Law No. 337/2009.

1.2.1 Right of access to the Internet: a need for explicit rules?

In Italy, the issue of the protection of freedom of expression in relation to the establishment of a right of access to the Internet under the profile considered has recently been addressed in the legal literature by an authoritative jurist, Stefano Rodotà. At the Internet Governance Forum in Rome in November 2010, the eminent scholar developed a proposal for the adoption of Article 21-*bis*, to ensure that the Internet is recognized as a fundamental right of all Italian citizens.

The wording of this Article reads [Rodotà, 2010]: "*Everyone has an equal right to access the Internet network, on equal terms, in technologically appropriate ways that remove all economic and social obstacles.*"

Article 21-*bis*, as proposed, would complement art. 21, which already exists and which guarantees freedom of the press. The illustrious author pointed out that an extension of the Constitution, albeit only the first part, is now necessary. In fact, openness to a right to the Internet indirectly but clearly strengthens the principle of network neutrality and consideration of knowledge on the Web as a common good, to which access must be granted.

By analyzing the content proposed for the new standard, we can grasp the following key points:

- The new article sheds light on the problem of the *digital divide* when it states "in a technologically appropriate manner", and imposes on the State the commitment to overcome it.

- In addition, all citizens, "on equal terms", must have access to the Internet: and must be able to do so "in a technologically appropriate manner", i.e., by providing an ADSL line of at least a guaranteed minimum speed, without the cost of this falling on the citizens themselves ("that remove all economic and social obstacles").

The term *digital divide* refers to the exclusion of some parts of the country from the possibility/need to use the services offered by the new networks, due to lack of investment in connection infrastructure. In practice, it takes the form of unequal access of the citizen to the advantages of using the electronic system compared to others who benefit from it.

In addition to any potential barriers to access, the digital divide can also relate to other factors, including the availability of information, the quality of technical resources, and the personal ability to use technology. Thus, in addition to the infrastructural digital divide, we must also take into account the *social digital divide*, namely the lack of involvement of part of the population in the use of IT tools and the new services.

In this respect, rather than a binary structure, the relationship of the individuals with the computer reflects a multiple structure, the result of the combination of a plurality of variables. This combination has been defined, in figurative language, as *"a rainbow, the result of the presence of physical media, software, content, services, infrastructure and so on* [Clement-Shade, 2000]*."*

Recognizing digital inequality is important since it allows one to place the aspect of access in a broader context, characterized by a strong focus on the impact that technology has on social inequalities. These considerations allow division of the concept of digital divide into three aspects, namely access, use and skills.

Later we will enter into the merits of the subdivision of these classes, but for the moment it is important to emphasize that the quality of access, the availability of increasingly sophisticated technological equipment enabling a continuous connection, together with the possession of the skills that permit one to achieve more and better aims, are all elements pertaining to the sphere of Internet use.

The social problem of digital literacy, understood in its double meaning of the concrete possibility of an individual to own the physical support that conveys the service and an appropriate level of technical expertise, has been neglected until recently.

Nowadays we are witnessing a remarkable expansion of the telecommunications sector, such that it should also be supported by an increase in computerization of the public sector through an effective expansion in the use of new technology. This creates an obvious gap between those who have the ability to go online and those who cannot.

The state's duty is thus to eliminate the cognitive obstacles that hinder true digital equality.

Therefore, we could hypothesize that the right of access is connected, by a close relationship of cause and effect, to the condition of aid from the State

[Sadowsib, De Rooij, Smits, 2006], in order to ensure an infrastructure program that provides the right to access the services offered by broadband, regardless of the potential user's geographic location.
Moreover, it is necessary for the State to facilitate the purchase of computers for the sector of individuals who are unable to deal with a similar expense, employing the tool of financial incentives.

A question, then, arises: does lack of access to the Internet lead to discrimination for the deprived, creating an obstacle to the possibility of their full social inclusion? Does an affirmative answer to this question suggest that broadband should be included among essential services guaranteed and protected by the State, and thus be configured as an essential public service?

This might mean that the state should intervene to remedy any deficiencies in infrastructure, where if operators were to decide that it was not cost-effective to invest in, and that the State should decide in advance the limits of this service and who is involved in ensuring the provision of the same to end users.

If Article 21-*bis* becomes a reality, the Internet should be recognized as a constitutional and basic social right of all citizens.

This policy is not original, but has already been introduced in Finland since July 2010. This nation allowed broadband connections to become a right for all residents. To achieve this, the Government determined that all providers were obliged to provide each resident a broadband line with a minimum speed of 1 Mbps, which by 2015 will reach 100 Mbps.

Methods for the growth of broadband adopted by Finland have relied primarily on market forces, encouraging state intervention only where it was absolutely necessary. More specifically, even in the case when market failures were recorded, national subsidies could not in any case exceed one-third of the amount required in the project, with a maximum of a further one-third for European and local funds. The private forces would thus be obliged to cover at least one-third of the cost of the project.

In line with this, Finland is one of several European countries that have adopted a coherent and effective policy of development of broadband. The country falls among the top ten in Europe regarding broadband penetration rate, and occupies first place for use of broadband technology in business [I-com, 2010].

We believe that the explicit provision of the right to Internet access as a fundamental right by adding a new Article to the Constitution, would eliminate at the root every problem of the existence and legal status of the right in question; however, it can be assumed -- also in light of the giurisprudence examined -- that this is not strictly necessary.

Believing that there should be a specific constitutional provision means in essence to doubt, if not deny, that at this right, with such characteristics, cannot be configured in our system. Of this, in fact, we should strongly doubt.

References

Bovero, La libertà e i diritti di libertà, in Quale libertà . Dizionario minimo contro i falsi liberali, a cura di Bovero, Roma-Bari, 2004.

Frosini, Il diritto costituzionale di accesso ad Internet, in Il diritto d'accesso ad Internet. Atti della Tavola Rotonda svolta nell'ambito dell'IGF (Roma, 30 ovembre 2010), a cura di Pietrangelo M., Napoli, 2011.

Rozo Acuna, Habeas Data costituzionale: nuova garanzia giurisdizionale del diritto pubblico latinoamericano, in Diritto pubblico comparato ed europeo, 2002.

Zeno-Zencovich, La libertà di espressione. Media,mercato,potere nella società dell'informazione, Bologna, 2004.

Di Lello, Internet e costituzione: garanzia del mezzo e i suoi limiti, in Diritto dell'informazione e dell'informatica, 2007.

Pubusa, Diritto d'accesso ed Automazione, Torino, 2006.

Ferrajoli, Costituzionalismo principalista e costituzionalismo garantista, in Giurisprudenza costituzionale, 2010.

Pino, Diritti e interpretazione, I ragionamento giuridico nello Stato costituzionale, Bologna, 2010.

Corte Costituzionale, 2 aprile 1969, n.84, in *Giurisprudenza costituzionale*, 1969.

Corte Costituzionale, 21 ottobre 2004, n.307, in *Giurisprudenza Costituzionale*, 2004.

Masucci, Erogazione on line dei servizi pubblici e teleprocedure amministrative, in Diritto pubblico, 2003.

Corte Costituzionale, 8 agosto 1977, n.94, in http://www.cortecostituzionale.it.

Corte Costituzionale, 9 giugno 1972, n.105, in http://www.cortecostituzionale.it.

Tribunale di Milano, 12aprile 2010, n.1972, in http://www.altalex.com.

De Marco, L'anonymat sur Internet et ledroit, Montpellier, 2005.

Rodotà, Una Costituzione per Internet?, in Diritto e sfera pubblica nell'era digitale, a cura di Amoretti, Politica del diritto, 2010.

Clement- Shade, The Access Rainbow: Conceptualising Universal Access to the Information/Communication Infrastructure, a cura di Gurstein, Toronto, 2000.

Sadowsib, De Rooij, Smits, State aid, open access and market size: two cases of FTTH, network implantation in duch municipalities, Eindhoven University of technology, 2006.

I-com in to data *Europès Digital Competitiveness Report*, in http://www.i-com.it/AllegatiDocumentiHome/420.pdf.

Biography

Valentina Amenta is Bachelor of Economics at the University of Pisa, thesis title "Consumer protection in electronic commerce. Reflections on the meaning of the distinction B2C-B2B. "Doctorate in Public Law and economics, during which she focused her research on the right of access to the Internet. Subject Expert "Information Technology Law" at the Faculty of Economics in Pisa, conducts didactic support for the two chairs of IT Law and Private Law. Research fellow at the National Research Council, Institute of Informatics and Telematics for research activities in the field of Internet Governance. Participate in the process of the Internet Governance Forum of the United Nations. Member of ISOC Italy.
email: valentina.amenta@iit.cnr.it

The Information Technology in Support of Everyday Activities: Challenges and Opportunities of the Service Oriented Computing

M. Bertolotto, P. Di Giovanni, M. Sebillo, G. Tortora, G. Vitiello

Abstract. *Nowadays, we are witnessing a paradigm shift from traditional monolithic software systems to distributed solutions where the whole computation is partitioned among various modules that can be also located in different geographic locations. Such a change has been surely fostered by the continuous development of the network infrastructures but, more effectively, by the need to easily access information distributed almost everywhere on the planet. In this context the Service Oriented Computing (SOC) has emerged as one of the leading approaches for both the design and implementation of distributed applications. The key concepts of this paradigm are the idea of service, an independent software module that performs certain operations, and the possibility to seamlessly combine such modules in order to offer more sophisticated functionalities. In this paper, after analyzing the main characteristics that made the SOC paradigm a so widespread solution to create a framework for application-to-application interaction, we discuss, as a concrete use case, the development of a software system aimed at helping farmers in Sri Lanka to improve the quality of their cultivations and related earnings.*

Keywords: Service oriented computing, Services composition, Interoperability.

1. Introduction

In recent years, the development of the network infrastructures and the growing importance of the Web and its related technologies have, on one side, changed the way users access information and, on the other side, had a profound impact on the design of computer applications.

One of the most obvious results of this change is the possibility for software components located in different geographic locations to communicate in order to satisfy the requests of final users. To facilitate the design and development of such systems various solutions have been proposed and, among them, the

Service Oriented Computing (SOC) paradigm has emerged as one of the leading approaches to develop distributed applications.

The fundamental idea behind this paradigm is the concept of service, an independent software module able to perform a defined set of operations. A service exposes its capabilities through its public interface that is usually described using public standards and technologies accessible by anyone in a platform independent way. Therefore its functionalities can be invoked by any type of software system: traditional desktop application, mobile applications and even other services. Such a characteristic, in addition to encouraging the development of loosely coupled solutions, facilitates the reuse of functionalities since the capabilities offered by each service can be reused in more contexts and not only in the specific context for which the service has been originally developed. Moreover, the availability of a technology agnostic public interface hides the implementation details to the potential clients letting, for example, an application developed using Microsoft .Net to seamlessly communicate with a service written in Java.

The ability to see services as independent building blocks and the need to only know their public interface to take advantage of their functionalities constitute the basis for the other key idea behind the success of the SOC paradigm, namely services composition, i.e., the ability to compose different services, developed also by different organizations to provide with complex functionalities. However, although the idea behind the services composition is quite simple, as we will see, there are some issues that must be carefully taken into account in order to guarantee a seamless exchange of information.

The remainder of this paper is organized as follows. In Section 2 we briefly describe the main features of the current standards for the development of reusable services. Section 3 discusses the fundamental characteristic of services composition as a mean to realize complex solutions based on the principles of the SOC paradigm. A description of a concrete example of a services-based system is provided in Section 4 and, finally, some conclusions are drawn in Section 5.

2. Augmented reality and new marketing scenarios

As outlined in the introduction, the functionalities of a service are exposed through its public interface. A complete description of such an interface is the only thing that a service client needs to know in order to invoke and use the service features. Once the potential client knows the operations supported by the service, all the communication between the two entities is based on various messages exchange mechanisms, such as the traditional Request-Response pattern.

Such a communication must be supported by a specific middleware that must guarantee some basic activities [Tsalgatidou and Pilioura, 2002]:

- Service creation,
- Service description,

- Service publishing for potential users to locate it,
- Service discovery by potential users,
- Service invocation and binding,
- Service un-publishing in case it is no longer available or needed.

However, since the actual implementation of services and clients might be realized using different platforms and programming languages, the use of proprietary formats for the information exchange is simply unfeasible. Therefore, there is the need to describe the public interface and to provide a messaging framework in a neutral manner using globally accepted standards that must be available on each platform.

In this context, the World Wide Web Consortium (W3C) has defined a series of universally accepted standards based on the use of the Extensible Markup Language (XML), in order to guarantee their independence from a specific platform or technology. In particular, in the following subsections we will briefly describe the SOAP protocol for the exchange of messages and the Web Services Description Language (WSDL) for the description of the service interface. Moreover, in the subsection 2.3, we will compare the W3C services design choices with the service standards used by the Geographic Information Systems (GIS) community.

2.1. SOAP

SOAP is a "lightweight protocol for exchange of information in a decentralized, distributed environment" [Gudgin et al, 2001] and constitutes the backbone of the messaging framework of most service oriented solutions.

It is entirely based on XML and heavily relies on XML technologies, such as XML Schema and XML Namespaces.

The structure of a SOAP message is made up of three elements: the Envelope, the Header and the Body. The Envelope can be seen as the container of the SOAP message. The Header element is used to carry additional information and although optional represents, in practice, a fundamental part for the implementation of a series of additional important protocols, such as those related to the information security and reliability. Finally, the Body element contains the real payload of the exchanged message [Tsalgatidou and Pilioura, 2002] [Erl, 2005]. Such a payload is usually represented by plain XML although binary data (images or PDF documents, for example) can be embedded in a SOAP message using particular techniques, such as the Message Transmission Optimization Mechanism (MTOM) or the Base64 encoding [Powell, 2004].

2.2. The Web Services Description Language

The WSDL [Christensen et al., 2001] is an XML based language for describing W3C services and how to access them. A WSDL document separates the abstract aspects of a service description from the concrete aspects such as the binding with a certain network protocol, thus preserving the public interface of the service by changes in the underlying technology. Seven elements constitute the typical structure of a WSDL document, namely Types, Message, Operation,

Port Type, Binding, Port, Service. The former four constitute the Abstract Description of a service, while the latter three allow the connection of the abstract interface of a service to a real technology and transport protocol, such as the Hypertext Transfer Protocol (HTTP). In the remaining of this subsection we describe in greater detail the role of each element.

The Types element can be seen as the container of the data type definitions used inside the Web Service. The Message element represents the data being communicated. In a WSDL document, it comes after the Type element. Each WSDL document can have one or more Message elements. Each Message has a univocal name and contains one or more children referred to as "Parts". The Parts can be compared to the parameters of a function in a traditional programming language. The data type of a part element can be a simple type or a type defined in the Types element. The Operation element describes, instead, the features that the Web Service will expose. Each Operation element is composed of Input and Output elements, which refer to Messages exchanged during the communication. The Port Type element is a named set of abstract Operations and the abstract Messages involved. Every Port Type has a unique name and is made of several Operations.

The Binding, Port and Service elements describe how the Abstract Description is mapped into a concrete format. The Binding element, in particular, defines message format and protocols details for the Operations and Messages defined by a particular Port Type. The Port element represents an instance of an abstract port (Port Type) obtained as a combination of a binding and a network address and, finally Service, the higher-level element in a WSDL document declares a W3C service as a collection of related Ports.

2.3. Geographic services standards

Despite the wide acceptance of the W3C proposals the GIS community, also for historical reasons, has developed, during time, its own set of standards for the fulfillment of geospatial data oriented services.

In particular, the proposals of the Open Geospatial Consortium (OGC) have become the *de facto* standard for developing distributed geographic applications but, although based on the use of XML for the exchange of data and HTTP as the transport protocol, they are incompatible with the W3C services.

The first important difference between OGC and W3C services is represented by the strong standardization imposed by the OGC about the public interface of a geographic service. Moreover, unlike W3C services each type of OGC service represents a different standard with a fixed public interface. However, to facilitate the development process, the aspects that are common to all types of OGC services are described in the OGC Common Standard specification [Whiteside and Greenwood, 2010]. The Common Standard describes another fundamental difference: the only way for a client to know the capabilities of an OGC service is to parse the XML document returned by the standardized GetCapabilities operation. In the W3C realm the service capabilities are publicly exposed in its WSDL document.

The last difference we mention here concerns the binding type and the binding time of operations. In the OGC services the type of a response message can dynamically vary based on the client requests, while in W3C services the message payload is completely defined at design time [Schaffer, 2008].

Two of the currently most used OGC standards are the Web Map Service standard and the Web Feature Service standard. The former provides an HTTP based interface for requesting georeferenced map images from one or more distributed geospatial databases, while the latter provides interface to access and manipulate geographic features (an abstraction of real world phenomena associated with an Earth location).

3. Services Composition

The general idea behind services composition is quite simple: two or more services are combined together in order to use their functionalities to "build networks of collaborating applications distributed within and across organizations" [Nano and Zisman, 2007]. A composed service can be the result of the composition of elementary services, previously composed services or a combination of the above.

Among the various advantages of the composition we can surely mention a better reuse of the service functionalities since the same service can be used in different contexts, and a better flexibility since the internal representation of a service can be modified (e.g., for refactoring or optimization reasons) without affecting the behavior of the whole system (provided that the service public interface remains unaltered). Therefore, instead of developing a whole application from scratch, an organization can realize its system by composing different types of services renting, for example, third-party services for certain functionalities (e.g., the functionalities offered by a financial entity for the electronic payments) and focusing on the development of those functionalities that represent its core business. [Di Nitto et al, 2008].

Nevertheless, despite the undeniable advantages, services composition is a very challenging task still involving research both in academia and industry even though, due to space constraints, we'll briefly summarize only the most important issues.

First of all, before starting to actually compose the various services, factors such as quality of service, security, response time and cost of the involved services must be carefully evaluated, since they can have a huge impact on the final result of the computation, especially when the involved components are developed by different organizations. Therefore, as a common practice, the service providers and the consumers agree upon a Service Level Agreement (SLA), a contract that disciplines the rights and the obligations of the involved participants [Di Nitto et al, 2008].

As for the real composition, the fact that, now, all the operations and algorithms might be the result of the composition of different services represents the first important issue to deal with. Such a peculiarity implies, as a direct consequence, the need to coordinate the sequence of operations to ensure the

correctness of the computation and avoid inconsistencies [Dustdar and Schreiner, 2005] and constitutes an important difference when the service-based applications are compared with the traditional monolithic systems.

Another important issue is related to the choice of the composition strategy. In fact, the software architect might choose to compose services during the design of the system (Static composition) or during its execution (Dynamic composition). Static composition is a suitable option only when there is the certainty that the involved services will rarely change [Dustdar and Schreiner, 2005].

The change in the structure of a service represents another problem that must be taken into account. Two main types of changes can be identifies [Di Nitto et al, 2008]:

- Shallow service changes when the changes are localized to a single service
- Deep service changes when the changes may affect the entire service chain.

The last fundamental issue we mention here concerns the need to have a commonly accepted "composition model and a language to specify the services involved in the composition" [Dustdar and Schreiner, 2005] that describes in a formal manner and without ambiguity, the sequence of the operations to be performed.

The most widely accepted approaches are the services Choreography and Orchestration.

In the Choreography approach each participant is provided with some rules that must follow. The global behavior of the system is the result of the interaction of the various parties each following its own rules.

In the Orchestration approach, instead, the compositions process is represented by a sequence of steps, conditions and exceptions coordinated by a central controller [Rosen, 2008]. In the Orchestration context, the Web Services Business Process Execution Language (WS-BPEL) represents the *de facto* standard to create applications by composing existing services [Alves et al, 2007]. WS-BPEL is expressed by using XML and heavily relies on several W3C specifications, such as the previously described WSDL.

Several elements contribute to define the definition of a WS-BPEL process. The most important of them are the <PARTNERLINK> elements that identify the various services involved in the process, the <VARIABLES> elements that contain the variables holding the messages exchanged during the composition and the elements used to arrange the execution order such as the <SEQUENCE> element or the <IF> element for the conditional execution.

As a concrete example of services composition, in the next section we briefly describe our choices for what concerns the design of a software system aimed at helping Sri Lankan farmers to improve the quality of their cultivations and related earnings.

4. The Social Life Networks for the Middle of the Pyramid project

The Social Life Networks for the Middle of the Pyramid (SLN4MoP) is an international collaborative research program that aims to provide real-time information to support activities related to livelihood targeted to meet the needs of people living in developing countries. In particular, agriculture in most of those countries represents a major economic sector, employing the largest share of the workforce, while suffering from low productivity. The reasons for this low productivity may include land fragmentation, lack of postharvest infrastructure, weak market linkages, information and knowledge asymmetries (or lack thereof), and, most important, low technology utilization. Therefore, as a pilot study, we have designed and are developing a software system aimed at helping Sri Lankan farmers to improve their rural activities and gain higher incomes.

From a high level point of view, the proposed system presents a traditional client-server architecture: a client sends a request to a remote backend which processes it and sends back the desired results. As for the actual structure of the system and the subdivision of the available functionalities, a three-tier subdivision has been chosen, namely a client tier, a middle tier and a data tier.

For the time being, the client tier mainly consists of applications running on modern mobile devices, although other potential stakeholders, such as government agencies could use different type of applications (e.g., traditional desktop application) for their needs. Using such mobile applications farmers can get detailed information about the best crops to grow according to the type of soil or the expected income, detailed information about the best fertilizers to use, information about the selling price of the selected products and can send data about their choices. Such data, aggregated in an anonymous form, will be processed by the middle tier to update, for example, the average cultivated quantity of a certain crop into a specific area. Detailed information about the design choices concerning the development of the mobile application used by the farmers can be found in [Di Giovanni et al, 2012] [Di Giovanni et al, 2013] [Sebillo et al, 2013].

Different and potentially distributed data sources constitute the data tier, such as information about the farmers, the various types of cultivable crops, the land extension of a farm or the different soil types. Moreover, such an information could be stored using different type formats.

As for the middle tier, the technical choices behind its design have been influenced by several factors. First of all the system must be able to provide not only with different types of information to the various types of potential stakeholders involved in the project, but must also be able to provide the same information with different levels of granularity according, for example, to the specific user's preferences or security policies. Therefore, flexibility is one of the requirements that influenced our choices.

Other important requirements to satisfy were the ability to add new features without affecting the existing components, the independence of the system

functionalities from the specific format of the various data sources and the possibility to seamlessly add new data sources or replace existing one without modifying the behavior of existing implementations. Finally, in order to accelerate the development process, the opportunity to use different development tools from different vendors was a desirable addition.

To satisfy such requirements, instead of developing a single, monolithic application we chose to develop independent and reusable modules for each functionality. We also decided to use a common protocol for the exchange of information among such modules. With these premises, organizing the design of the middle tier around the principles of the SOC paradigm has been a natural step.

After the initial partitioning of the functionalities into a set of services, we further specialized the derived services into three layers of abstraction [Erl, 2005]:

- The application service layer
- The business service layer
- The orchestration service layer

Such a subdivision presents several advantages. Among them we can mention a better reuse of the functionalities, a better maintenance and most important an improved flexibility of the whole system, since new modules can be added or existing modules can be substituted e.g., with a more optimized version without affecting the behavior of the remaining services . Moreover, third party solutions can easily interact with the functionalities exposed by the available services.

The application service layer represents the lower level of the abstraction and includes all those services that expose reusable and solution agnostic functionalities. Such functionalities are used in the upper level, the business service layer and, finally, the services in the orchestration layer organize and compose the services of the lower layers in order to offer the desired functionalities to the final users.

The actual technological platforms we chose for the development of our service oriented system deserve further considerations. The majority of our platform is being developed in compliance with the above mentioned services standards proposed by the W3C. However, although W3C service are one of the best alternatives for the development of enterprise class distributed applications, as we have seen, they are not the best choice for the management of geospatial data.

Geospatial data are of utmost importance for the purposes of our system (for example, the suggestions for a specific user about the best crops to grow are based not only on the current market trends but also on the specific characteristics of the soil where his/her farm is located). Moreover, since the geospatial information could be useful also for third party entities (e.g., a government agency might be interested to visualize the various soil types available in a specific region using a traditional desktop GIS application such as uDIG) we chose to develop the services that deal with geospatial information following the standards proposed by the OGC (in particular, our implementation will be compliant with the Web Feature Service standard).

Unfortunately, as we have seen, despite sharing some common characteristics such as the use of XML for the exchange of information, the W3C and OGC services are incompatible.

Therefore, to guarantee the highest level of interoperability among our various software modules, according to the solutions discussed in literature, we are developing a service wrapper whose purpose is to query the proper OGC service and return the geospatial information in a W3C compliant way keeping the structure of the original W3C or OGC services unchanged.

In particular, the wrapper carries out the following tasks:

1. It receives a SOAP message from a generic client containing a request for a specific geospatial data
2. It translates the SOAP message in a format suitable for the underlying OGC service and sends the request
3. It translates the received response in a SOAP message and sends it to the client.

A detailed analysis of the challenges we are facing during the development of the wrapper can be found in [Ioup et al, 2008].

We briefly mention that such issues can be divided into three main categories ,namely Data handling, Functionalities mapping and Metadata management.

The first type of issue concerns the different data types that can be returned by an OGC service, the second issue deals with the difficulties to correctly expose the interface of an OGC service using a WSDL document and, finally, the third issue relates with the need to deal with the spatial metadata required by each OGC service.

5. Conclusions

The SOC paradigm has radically changed the way distributed applications are designed and developed. The key idea behind this approach is the concept of service, an autonomous software module that combined with other services can be used to create complex solutions. Unfortunately, services composition is a non-trivial task and several aspects must be taken into account. In this paper we have provided an overview of the current main standards used for the actual services development and a description of the most important challenges that arise during the composition task. We have also described some design choices we made for the development of a services-based system aimed at helping Sri Lankan farmers to improve their farming activities.

References

Alves A., Arkin A., Askary S., Barreto C., Bloch B., Curbera F., Ford M., Goland Y., Guízar A., Kartha N., Liu C. K., Khalaf R., König D., Marin M., Mehta V., Thatte S., van der Rijn D., Yendluri P., Yiu A., Web Services

Business Process Execution Language Version 2.0, OASIS, 2007, http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.pdf.

Christensen E., Curbera F., Meredith G., Weerawarana S.,Web Services Description Language (WSDL) 1.1, World Wide Web Consortium, 2001, http://www.w3.org/TR/wsdl.

Di Giovanni P., Romano M., Sebillo M., Tortora G., Vitiello G., De Silva L., Goonethilaka J., Wikramanayake G., Ginige T., Ginige A., User Centered Scenario Based Approach for Developing Mobile Interfaces for Social Life Networks, Proc. of the First International Workshop on Usability and Accessibility Focused Requirements Engineering (UsARE 2012), June 4, 2012, Zurich, Switzerland ISBN 978-1-4673-1846-4, IEEE, 2012 18-24.

Di Giovanni P., Romano M., Sebillo M., Tortora G., Vitiello G., De Silva L., Goonethilaka J., Wikramanayake G., Ginige T., Ginige A., Building Social Life Networks through Mobile Interfaces - the Case Study of Sri Lanka Farmers, in Spagnoletti P. (ed.) Organizational Change and Information Systems, Lecture Notes in Information Systems and Organisation 2, Springer-Verlag Berlin Heidelberg 2013.

Di Nitto, E., Ghezzi, C., Metzger, A., Papazoglou, M., Pohl K., A journey to highly dynamic, self-adaptive service-based applications, Autom Softw Eng (2008) 15: 313–341.

Dustdar, S., Schreiner, W., A survey on web services composition. Int. J. Web Grid Serv. 2005 1(1): 1-30.

Erl T., Service-oriented architecture: concepts, technology, and design, Prentice Hall PTR, Upper Saddle River, New Jersey, 2005.

Gudgin M., Hadley M., Moreau JJ., Nielsen H. F., SOAP Version 1.2. World Wide Web Consortium, 2001, http://www.w3.org/TR/2001/WD-soap12-20010709/.

Ioup E., Lin B., Sample J., Shaw K., Rabemanantsoa A., Reimbold J., Geospatial Web Services : Bridging the Gap between OGC and Web Services, in Sample J.T., Shaw K., Tu S. and Abdelguerfi M. (eds) Geospatial Services and Applications for the Internet. Springer, New York, 2008, 73-93.

Nano E., Zisman A., Realizing Service-Centric Software Systems, Software, IEEE 24,6, 2007, 28-30.

Powell M., Web Services, Opaque Data, and the Attachments Problem, Microsoft Corporation, 2004. http://msdn.microsoft.com/en-us/library/ms996462.aspx.

Rosen M., Orchestration or choreography?. http://www.bptrends.com/ 2008.

Schäffer B., OWS 5 SOAP/WSDL Common Engineering Report, Open Geospatial Consortium Inc, 2008 http://www.opengeospatial.org/standards/dp.

Sebillo M., Tortora G., Vitiello G., Di Giovanni P., Romano M., A Framework for Community-Oriented Mobile Interaction Design in Emerging Regions, Kurosu M. (Ed.)Human-Computer Interaction, Part III, HCII 2013, LNCS 8006, Springer-Verlag Berlin Heidelberg, 2013, 342–351.

Tsalgatidou A., Pilioura T., An overview of standards and related technology in web services. Distrib. Parallel Databases, 12,2-3, 2002, 135-162.

Whiteside A., Greenwood J.,OGC Web Services Common Standard. Open Geospatial Consortium Inc., 2010, http://www.opengeospatial.org/standards/common.

Biographies

Michela Bertolotto received a BSc (1993) and a PhD degree in Computer Science (1998) from the University of Genova, Italy. Subsequently she worked at the National Center for Geographic Information and Analysis and the Department of Spatial Information Science and Engineering of the University of Maine as a postdoctoral researcher. Since September 2000, she has been a faculty member at the School of Computer Science and Informatics at University College Dublin, where she leads the Spatial Information Systems group. Her research interests include web-based and mobile spatial information systems, the semantic geospatial web, spatio-temporal data modelling and mining, geovisual analytics.
email:michela.bertolotto@ucd.ie

Pasquale Di Giovanni is a PhD student in computer science under the supervision of Prof. Giuliana Vitiello and Dr. Michela Bertolotto. His main research interests fall in the fields of Geographic Information Systems and Human Computer Interaction. In particular, his research focuses on the investigation of issues related to the design and development of service-based solutions for geospatial mobile applications.
email: pdigiovanni@unisa.it

Monica Sebillo received a laurea degree in Scienze dell'Informazione from the University of Salerno and a PhD in Applied Mathematics and Computer Science from the University of Naples. She is now an Assistant Professor in Computer Science at the Department of Management and Information Technology (DISTRA) at the University of Salerno. Her current research interests are in Geographic Information Systems and Spatial Databases, Human - GIS Interaction, and Mobile Applications. She is an ACM Senior member, and a GISIG, ICA and AMFM member. She serves or has served on the Program Committee of several international conferences and is author of more than 80 papers in international journals and conference proceedings.
email: msebillo@unisa.it

Genny Tortora is professor of Computer Science at the Department of Management and Information Technology. Her research interests include software-development environments, visual languages, GIS and biometry. She is author of more than 250 papers in scientific journals or proceedings of refereed conferences. Editorial Board Member of high quality international journals. Program Chair and Program Committee Member of relevant international conferences. IEEE Computer Society Senior Member, ACM, EATCS and IAPR Member. Reviewer for international scientific journals, and evaluator of research projects.
email: tortora@unisa.it

Giuliana Vitiello is Associate Professor of Computer Science at the Department of Management and Information Technology. Her present research interests fall mainly in the fields of Human-Computer Interaction and Usability Engineering. She has published several scientific papers in international journals, contributed volumes, and conference proceedings. She is a Senior Member of ACM-SIGCHI and a Member of AMFM/GIS. Since 2002 she is in the Panel of Experts set up by the Italian Ministry of University and Research (MIUR) for the evaluation of industrial research projects.
email: gvitiello@unisa.it

Digital Citizenship and Social Responsibility of Computer Professionals

N. Patrignani, M. De Marco

Abstract. *Digital revolution is taking us into the Information Society so quickly that the entire Information and Communication Technology (ICT) world has no time to reflect on new social and ethical issues. In particular this dramatic change put computer professionals in a position full of opportunities but also with new responsibilities. Computer professionals and Information Systems managers have in front of them a "rainbow" of risks and opportunities, they face a collection of ethical dilemmas and social issues that requires a deep reflection and debate among all stakehoders: universities, ICT professional organizations, ICT industry, policy makers, users and society at large. A code of ethics for computer professionals could be a good starting point for defining a collection of guidelines.*

Keywords: Social responsibility, computer professionals, computer ethics, Code of ethics.

1. Introduction

In this paper we would like to address one of the themes of the AICA 2013 National Congress, "*digital citizenship*", from the particular point of view of computer professionals and Information Systems managers. They are the real people in charge for the design, development and governance of ICT infrastructures at the basis of our Information Society, they are the real providers of this "*digital citizenship*". How are changed their roles and responsibilities among the years? What are the most important changes from *mainframe era* to Distributed Computing era and finally to Cloud Computing (and Big Data) era? What kind of new social issues and ethical dilemmas are they facing everyday? What kind of instruments have they for facing these issues? What kind of relationship should be in place between "*digital citizens*" and the people that are in charge for ensuring that the "*digital platforms*" run smoothly, respect basic

human rights, and minimize the impact on the environment? What are the relationships between the computer professionals (at all levels) inside an organization and the Corporate Social Responsibillity strategy of the organization itself? What kind of support can computer professionals societies (like AICA in Italy, British Computing Society, BCS in UK, etc.) provide to their members? What contribution could provide a *Code of Ethics* for computer professionals? In this paper we would like to set up the stage for finding some possible answers to these questions.

2. From Mainframe to Cloud Computing and Big Data

Computing evolution can be shortly described with three main eras: the (centralized) *Mainframe Computing era*, the *Distributed (Personal) Computing era*, and the *Cloud Computing era*. In each era, computer professionals and Information Systems managers played different roles and faced many responsibilities.

In the (centralized) mainframe computing era (1950s-1970s) the basic technologies were "dumb" terminals (with just input, output and network) connected to big computer (mainframes) inside machine rooms. Any application was designed to run on mainframe platforms. The main issues in front of computer professionals were to properly design the hardware and software platforms to run inside the computer room. The main problems for Information Systems organizations was to control the access to computer rooms (physical and logical) and to guarantee the reliability of the systems. The end users had little or no role in this scenario, they were just "consuming" ICT applications entering their inputs on keyboards or punched cards and visualizing output on printers or screens [Williams, 1997].

In the distributed (personal) computing era (1970s-2010s) the novelty was based on the processing power available to any department and user, and the networking capabilities extended across countries. The computer professionals were involved into the design of "distributed computing" infrastructures and "client-server" applications. The Information Systems role grew in complexity due to the management issues related to networked domains and applications running also at department and user levels [Couloris et al., 2011].

In the Cloud Computing era (2010s-now) the novelty is the re-centralization of computing power and storage ("*in the cloud*"), the decreasing role of end-users' devices (with the local processing and storage capabilities mostly unused), and the network extended at global scale. Computer professionals develop Web-based applications (or simply *apps*) to be "consumed" on touchscreen devices with little or no computing autonomy.

In this scenario the *cloud ICT infrastructure* becomes global with the following characteristics: it is (broadband) *network-based*, the servers are shared-platforms (*multitenancy*), the processing power and storage capabilities are easily scalable (*scalability* and *elasticity*), all resources consumption are *measured* (e.g. for billing purposes) and the users can allocate resources *on-demand* in a *self-service* way [NIST, 2013].

Then there are different Service Models in this scenario:

- *Infrastructure as a Service* (IaaS) where Cloud Providers just rent IT physical components to Cloud Users (in this case Cloud Users organizations have the control of the Applications and shares the Virtual Machines with the Provider, whilst the Provider controls Server, Storage and Network levels);
- *Platform as as a Service* (PaaS) where the Cloud environment is used mainly for developing and testing very large applications (in this case Cloud Users shares the Applications and the Services with the Provider, whilst the Provider has the control of Server, Storage and Network levels);
- *Software as a Service* (SaaS) where the entire application is on the Cloud Provider side (in this case Cloud User organizations have no control on any level, whilst the Cloud Provider controls all the Applications, Services, Server, Storage and Network levels) [Mather et al., 2009].

Of course many organizations are implementing Cloud Computing solutions "inside" thier domain for using ICT resources more efficiently. In these cases the term used is "Private Cloud", an evolution of the "intranet" concept [Nuttgens et al., 2011], there are no data crossing public networks. In these cases, "*data produced and subsequently distributed within an organization is not only a strategic resource to achieve/maintain a competitive advantage but it is also used as a tool to develop and expand the ability of an enterprise to abruptly respond to unexpected generated by the external environment*" [Zardini et al., 2011; p. 390].

In the standard (Public) Cloud solutions, the data of organizations crosses public networks and are stored "outside" the administration border controlled by Information Systems managers. In these cases, often referred as "Public Cloud" there are many interesting and difficult issues that arise. Information Systems and Chief Informaton Officers (CIOs) roles are now very critical since the pressure from Chief Financial Officers (CFOs) about moving ICT services towards the *pay-per-use* model of Cloud Computing is growing: with no more investments needed, the move from CAPEX (Capital Expenditures) model to OPEX (Operational Expenditures) model for ICT services is becoming very attractive for CFOs [Patrignani and Kavathazopoulos, 2013].

The risk for CIOs and IT services is to become just "*service brokers*": to chase the best offers from "Cloud Providers" market and offer them to the organization's employees [Ricciardi and De Marco, 2012]. What will happen to all workers inside the IT organizazions? Of course all the activities related to design, develop, deploy, maintain all IT services (and the computer room itself) will disappear as all those tasks and responsibilities will be shifted outside to Cloud Providers. Even the task of managing the internal workplaces, personal computers, and laptops is disappearing, since many organizations are encouraging employees to use their daily personal device also at work, it is the so called phenomenon of BYOD, Bring Your Own device [Casey, 2013].

But there are also many other important issues related to Cloud Computing that CIOs have to address like:

- *governance* (in particular in the Service Model like SaaS, the Cloud Provider is delegated to control the entire "stack" of ICT layers, Application, Services, Server, Storage, and Network);
- *de-perimeterisation* (the traditional network boundaries between organization domains will disappear, information storage and processing are outside the control of the organization that still keeps only input, output and network, a kind of "return" to mainframe era);
- *contractual obligations* (there will be the case where the Cloud Provider itself is not the owner of the resources it is "selling"; there will be "ICT brokers", in these cases who will be responsible of what?);
- *problems of many-hands* (when there are several administrators - "*many hands*" - that control mission critical infrastructures, what will happen if, for example, a cloud administrator stops a service for maintenance? There will be cases where both sides – Providers and Users of the cloud – must agree with a kind of "*four-hands-authorization*");
- *risk management plans* (in case of an incident – ICT people know that "*something can always go wrong*" – will the cloud be "*traceable*"? It will emerge the need for storing events in encrypted secure logs, time-stamped and digitally signed, and the need to agree on *joint* risk-management-plans);
- *compliance* (some Cloud Users will need, for compliance purposes, to know the physical location of their data);
- *open market* (what will be the data formats? What kind of standards will be adopted for guaranteeing portability? Can Cloud Users easily change Cloud Provider?).

Computer professionals, Information Systems managers, and CIOs has to deeply reflect on all these issues before signing contracts with Cloud Providers. But there are also many others social and ethical issues related to the management of ICT infrastructures in the current Information Society scenario. Here there are some examples:

- *e-democracy* (what is the correct definition of digital citizenship? What are the rights and duties of citizens online, what kind of decisions are we going to make online?) [Ricciardi and Lombardi, 2010];
- *e-Inclusion* (are we providing the proper interfaces also to elderly people or people with disabilites?);
- *digital divide* (is the access to ICT infrastructures available everywhere? For example, in a "*smart city*" environment, who and how will have access to these kind of "*public digital services*"? Is all the data collected by the smart city applications provided to other organizations in an "*open access*" way?);
- what will be the impact of ICT on workplaces, on information systems users?
- how will change the definition and management of intellectual property in the digital world?

- what will be the approach to *privacy* issues when most of the data and applications will be "outside" the organization? [Patrignani and DeMarco, 2012];
- what kind of relability of ICT infrastructures and applications will be guaranteed in the Cloud computing scenario?
- what kind of initiatives and procedures will CIOs put in place in order to minimize the environmental impact of ICT? [Patrignani, 2009];
- how will the organization be prepared to face the "*digital tsunami*" of BigData, where Billions of Gigabytes (Exabytes) will have to be processed and stored?

3. ICT landscape: a network with many stakeholders

In the previous part of the paper we have discussed the complex scenarios in front to computer professionals at all levels. These scenarios include several social and ethical issues. Now we introduce some tools or instruments that could help these people in facing these problems that cross the traditional technology borders and reach in many aspects the "*digital citizens*" and the society. Indeed the traditional strategies of the majority of organizations are concentrated only on business opportunities and on short-term profit goals. ICT was always considered just as a "*business enabler*", until 1990s. Since then, with the growing role of *e-business* and online engagement, transaction, fulfillment and service management, ICT is becoming *the* (*core) business* in itself: without ICT business activities are simply impossible [Rossignoli et al. 2009]. Also, the growing need for organizations to have a clear Corporate Social Responsibility strategy raised the need of a kind of alignment between these strategies and the need to address the social and ethical issues of computing, a clear *computer ethics* strategy. It is now clear how the two domains should be strictly aligned in order of transparently distributing roles and responsibilities inside the organization, and for a complete alignment between *Business Ethics* and *Computer Ethics* strategies [Chartier and Plante, 2013].

Nevertheless computer professionals need some specific tools for facing the immense social and ethical issues and the "rainbow" of risks and opportunities in front of them.

The ICT *stakeholders network* has been proposed as a powerful tool for reasoning and dialoguing about these difficult choices (see fig.1 – Example of ICT Stakeholders Network – applied to the recent Cloud Computing scenario) [Patrignani and Kavathatzopoulos, 2013]. Within this network, it is possible to identify all the stakeholders and relationships related to a specific ICT scenario. The simple construction of this network is already a good help for identifying *conflicts* between stakeholders and missing relationships usually not considered into the customary ICT landscape. In some way it could help also in *ethical decision making* [Laaksoharju, 2010].

Computer professionals and people usually involved just in "technical" decisions are rarely exposed to the concept of artifacts (including information systems), as "*socio-technical systems*", or to the concept that artifacts embed values [Johnson, 1985]. It is difficult for them to see ICT systems as artifacts that at

development-time incorporated the values of the designers, or to see systems as a result of a *Value Sensitive Design* [Friedman, 1996].

This is the main reason of our next proposal: in order to really face their social and ethical responsibilities, as "*digital citizenship providers*", computer professional, or in general, ICT people involved along the entire ICT value-chain need to introduce a reflection inside their professional organizations, this reflection will enrich their "technical" skills with some deontological competences that may prepare the road towards a *code of ethics for computer professionals*.

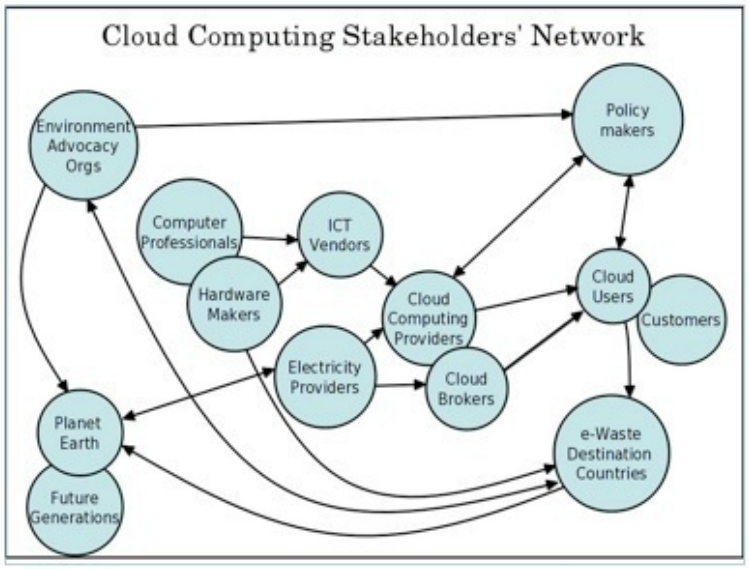


Figure 1
Example of ICT Stakeholders Network

4. Towards a Code of Ethics?

Around the world many computer professionals organizations are providing their members with resources, events, and working groups related to the ethical issues of ICT. One of the most established ones is the Ethics Group inside British Computing Society (BCS) in the UK, one of the oldest ICT professional organizations in the world. The BCS Etihcs Group has a challenging responsibility: "*... is responsible for promoting awareness and engagement with the ethical issues associated with the advancement of Information Technology science and practice and ensuring that Ethics is fully embedded in everything the Institute says and does*" [BCS-a, 2013]. Along the years, they have also defined a "*BCS Code of Conduct*" for providing support to the people in the field [BCS-b, 2013]. One of the oldest Code of Ethics in ICT is the one defined in the 1990s by the Association for Computing Machinery (ACM) [ACM, 2013].

Maybe that these "codes" will not be able to provide answers to any possible question or ethical issue one can face, but for sure they will provide useful

guidelines and suggestions in order to *be prepared*, in order to improve one status from "*technician*" to a real *professional*.

Also in Italy, the "*Italian Computing Society*" (AICA, Associazione Italiana per l'Informatica ed il Calcolo Automatico) has recently setup a Working Group on ICT and Ethics (Progetto ETIC, Etica e Tecnologie dell'Informazione e della Comunicazione), coordinated by professor Ivo De Lotto of the University of Pavia. There are several ongoing activities within this working group and many of them are preparing the background for a discussion about a *code of ethics for computer professionals* also in Italy [AICA, 2013].

5. Conclusions

The evolution of ICT happened so quickly that gave little time to reflections on different points of view with respect to the common vision of the computer as a technology that can solve most of the problems of society. The risk that we face is to be too fascinated by the wave of technological innovation losing sight of the sea of changes and controversial issues below the sea surface. Yet the widespread diffusion of computers in society, and the indispensable role played by networks of computers in almost all economic activities, induces a series of ethical reflections, not just professionals or experts in ICT, but also to end users and the entire society.

In conclusion we would like to propose some simple recommendations:

- for Universities: to introduce interdisciplinary courses in Computer Science and Computer Engineering courses that could prepare the future generations of computer professionals to face also the social and ethical issues related to ICT;
- for Policy makers: encourage and promote public discussions for decisions regarding the use of ICT in critical systems for the society (e.g. "*smart cities*"), encourage the use of technology for improving quality of life, *well-being* and *well-living* of human beings;
- for ICT industry: define clear Corporate Social Responsibility strategies that involve the analysis of the entire stakeholders network of ICT market;
- for computer professionals organizations: inform the public about the intrinsic limits of ICT systems reliability, question the presupposition that technology can (alone) solve all social and political problems, be involved in national and international debates about the (social, ethical, legal) professional issues related to ICT, and design systems in order to minimize the environmental impact of ICT.

With this paper we would like to support a reflection, in particular among computer professionals and people involved in the ICT value-chain, about their social responsibility as main providers of the "digital citizenship" to the society. We would also like to provide a contribution to the preparation of people with complex skills and knowledge, people that are not just ICT experts, but people that are also able to understand and make right evaluations about social and ethical implications of ICT.

References

ACM, ACM Code of Ethics and Professional Conduct, adopted d by ACM Council 16 October 92, available at: http://www.acm.org/about/code-of-ethics (accessed 12 June, 2013).

AICA, Progetto ETIC, Etica e Tecnologie dell'Informazione e della Comunicazione, available at: http://www.aicanet.it/attivita/gruppi-di-progetto/progetto-etic (accessed 10 July 2013).

BCS-a, The Chartered Institute for IT - Enabling the Information Society - Ethics Group, available at: http://www.bcs.org/category/8620 (accessed 12 July 2013).

BCS-b, BCS Code of Conduct, available at: http://www.bcs.org/category/6030 (accessed 12 June 2013).

Casey K., Six Risks Your BYOD Policy Must Address, Information Week, 19 November 2012.

Chartier A., Plante B., IS/IT ethical issues as a Corporate Social Responsibility: revisiting strategic business planning under the lens of IS/IT ethical preoccupations, in Ward Bynum T., Fleishmann W., Gerdes A., Moldrup Nielsen G., Rogerson S. (eds.) The Possibilities of ICT, Proceedings of 13th International Conference ETHICOMP2013, University of South Denmark, 2013.

Coulouris, G., Dollimore J., Kindberg T., Blair G. (2011). Distributed Systems: Concepts and Design (5th Edition), Addison-Wesley, Boston, 2011.

Johnson, D.G., Computer Ethics, 1st Edition 1985, 4th Edition, Pearson International Edition, Prentice Hall, 2009.

Friedman, B., Value Sensitive Design, *Interactions*, November/December 1996.

Laaksoharju M., Let us be philosophers! Computerized support for ethical decision making, Department of Information Technology, Uppsala University, Uppsala, 2010.

Mather T., Kumaraswamy S., Latif S., Cloud Security and Privacy - An Enterprise Perspective on Risks and Compliance, O'Really Media, 2009.

NIST (National Institute of Standards and Technology) Cloud Computing Program 2013, Available at: http://www.nist.gov/itl/cloud/ (accessed 1 July 2013).

Nuttgens M., Gadatsch A., Kautz K., Schirmer I., Blinn N., (eds), Governance and Sustainabilityof Information Systems - Managing the transfer and diffusion of IT, Proceedings of the IFIP WG 8.6 International Working Conference, Hamburg, Germany, September 2011.

Patrignani N., Computer Ethics. Un quadro concettuale, Mondo Digitale, n.3, Settembre 2009.

Patrignani N., DeMarco M., The road to a responsible and sustainable e-Business, Proceedings of the International Conference on e-Business, ICE-B2012, Rome, July 2012.

Patrignani N., Kavathazopoulos I., The Brave New World of Socio-Technical Systems: Cloud Computing, in Ward Bynum T., Fleishmann W., Gerdes A., Moldrup Nielsen G., Rogerson S. (eds.) The Possibilities of ICT, Proceedings of 13th International Conference ETHICOMP2013, University of South Denmark, 2013.

Ricciardi F., Lombardi P., Widening the Disciplinary Scope of eParticipation. Reflections after a Research on Tourism and Cultural Heritage. In: Tambouris E., Macintosh A., Glassey O. (eds.), Electronic Participation. Second IFIP International Conference, ePart 2010. Lausanne, Switzerland. Lecure Notes in Computer Science, Springer, 2010.

Ricciardi, F., De Marco, M., The challenge of Service Oriented performances for Chief Information Officers, in Snene M. (ed.), Exploring Service Science, Third International Conference, IESS 2012, Geneva, Switzerland. Lecture Notes in Business Information Processing, Springer, 2012.

Rossignoli, C., Carugati, A., Mola, L., The strategic mediator: a paradoxical role for a collaborative e-marketplace, Electronic Markets, 19(1), 55-66, 2009.

Williams M.R., A History of Computing Technology, 2nd Edition, IEEE Computer Society Press, Los Alamitos, CA, 1997.

Zardini, A., Mola, L., Vom Brocke, J., Rossignoli, C., The Role of ECM and its Contribution in Decision-Making Processes, Journal of Decision Systems, 19, 4, 389-406, 2010.

Biographies

Norberto Patrignani is Senior Associate Lecturer of "Computer Ethics" at Graduate School of Politecnico di Torino, Expert for the EU Commission at European Research Council (ERC) and Lecturer of “ICT & Information Society" at Catholic University of Milano. From 1999 to 2004 was Senior Research Analyst with META Group (Stamford, USA). From 1974 to 1999 worked at Olivetti's Research & Development (Ivrea, Italy). He graduated (cum laude) in Computer Science at University of Torino and in Electronics (cum laude) from "Montani" Institute of Technology (Fermo, Italy). He is frequently speaker at international conferences, published many articles in international journals and several books on the subjects of responsible innovation and computer ethics.
email: norberto.patrignani@unicatt.it

Marco De Marco. After having served 30 years at the Catholic University of Milan up to the top of the academic career — today is full Professor of Organization and Information Systems at the Guglielmo Marconi University in Rome. Marco De Marco is the author of five books that discuss the development of information systems, the computer industry, and the impact of technology on organizations, as well as the writer of several articles and essays. He is also a member of the editorial board of a number of journals. His major interests are systems development, e-government, programme evaluation, banking information systems, IT and Organizations. For his contribution to the discipline he received in 2010 the award of AIS Fellow.
email: m.demarco@unimarconi.it

The Multidimensional Value of Transparency in Healthcare Organizations Computerization

A. Tommasetti, G. Festa

Abstract. *Healthcare informatics and related computerization represent a fundamental driving force for the evolution and development of healthcare organizations. Health information systems (in the broadest sense, as proposed by AICA with its professional certification denominated 'ECDL Health') are scheduled to become more and more important in the years to come, by virtue of the benefits that business informatics can produce in the health sector. In particular, health computerization can aid in pursuing transparency, added value in terms of health not only as regards performance and responsibilities, but also as regards organizational processes and related information flows. Our study, a conceptual paper, intends to develop a theoretical framework relative to the multidimensional value of health transparency, analyzing health information systems in this perspective, resulting from business computerization; in addition, by means of a content analysis, we recommend a more integrated 'ECDL Health' Syllabus.*

Keywords: information systems, healthcare organizations, transparency

1. Introduction

Current socio-economic scenarios are becoming ever more complex. One of the underlying causes is certainly computerization, not only in terms of hardware and software data processing (the traditional meaning of computer science), but also (and, in our opinion especially) in terms of the organizational change that inevitably derives, characterized by better organization, engineering, proceduralization and, ultimately, process/outcome standardization. Extensive use, at least in the design of the digital organizational structure, of the criterion of discontinuity [De Witt, 2001; Tonti, 2002; Bracchi et al., 2010] is another

significant factor. In conceptual terms, designing and developing an information system (in general) and its computerized parts (in particular) oblige the professionals involved to reinterpret, represent and re-engineer specific flows of information and related organizational processes, with the application of approaches that are both reductionist (*fragmentation*) and systemic (*integration*). Only downstream, therefore, will it be possible (eventually) to embed flows and processes within a computer solution, while developing a deep understanding of the business case in question.

Such efforts inspired by the end user and inevitably, implemented in functional terms could also be preliminary to the engendering of transparency, from a lexical and semantic perspective. A case in point is the absence of information asymmetry, i.e., the non-existence of situations and/or persons in the presence of uneven information alignment. Increased availability of information, however, does not mean transparency merely in terms of information symmetry: the concept of openness, on the contrary, considers other values, such as social (or collective) equifinality and ethics (or at least ethical awareness), encompassing the traditional triple line approach: efficacy, efficiency and economicity (fundamental in a healthcare context) in an all-embracing concept of 'civil health'. Clearly the concept of transparency discussed in our paper is beyond the scope of the privacy of personal and sensitive data referred to individual citizens/users/patients which is regulated by law in Italy (see Legislative Decree no.196/2003).

In IT language, at least as concerns Italian, the term 'transparent', refers to any computer application, designed and/or implemented at any level, in which an operator (potential user) is not required to know how the system really works in order to use it. Transparency in other words, comes very close to usability, i.e. the minimum gap possible between the user model and the design model.

In a parallel manner, for the final beneficiary, i.e. the citizen in the broadest sense, the more a health system is 'transparent' the more 'usable' it is, both in operational and social terms. Health informatics or better health computerization, represents an important contribution in this direction. It can be considered the mental and cultural propensity underpinning the governance and management of health data, information and knowledge in computer terms, and their effects on organizational processes and information flows [Teti and Festa, 2009].

Our study, a conceptual paper, aims at elaborating a theoretical framework for an integrated analysis of the multidimensional value of health transparency deriving from the computerization of health organizations, where transparency is a guiding principle for the design of health information systems. In addition, we intend to investigate the potential dimensions of health transparency within the ECDL Health Syllabus (a standard professional certification, recognized internationally and issued in Italy by AICA, for health informatics). Our aim is to attempt to contribute to the potential integrating of the Syllabus as regards transparency in health informatics, taking into account also recent Italian regulations.

2. The multidimensional value of transparency in the health sector

Neoclassical (or marginal) economics which have impacted most on current interpretations of economic events, defines 'value' as the appreciation of a utility (of time, place, possession and form), which in the case of health, becomes a 'subjective use value' (the exact neoclassical phrase) of the tangible, intangible, financial, professional and technological capacity (the traditional function of production) of health facilities to meet needs/desires in physical, mental, social and environmental terms (in accordance with the definition of 'health' proposed by the World Health Organization). In this sense, healthcare value is considered from a 'technical' point of view [Tramarin, 2002].

In truth, even in classical (although not core as in neoclassical) economics we find the principle of 'utility', emphasizing in particular the concept of marginal utility as the pivot of economic dynamics [Da Empoli, 2012]. In addition, the term 'professional' in the health sector and in our perspective, includes also the personal and public contribution of health communication, considered nowadays fully and rightly an essential component in producing the health service and in providing outcomes.

To legitimize its value, from a 'commercial' point of view, goods should be 'market' orientated or be of interest (and hence value) as an object of exchange: a) different goods have varying values (e.g. a cure for cancer has a completely different value compared to cosmetic treatment); b) some goods have value for some people and not for others: in other words, there may not be the same interest for these assets (cancer treatment is out of context and therefore worthless to a person desiring cosmetic treatment); c) similar goods have different values in different contexts (depending on the customers requiring them).

Undoubtedly, appreciation is not limited merely to economic monetary valuation, classified in terms of the marketing mix as 'price', the measure of commercial exchange, but end up inevitably including other elements intrinsic to exchange (time required to gather information on purchase options, the contextual specifics of purchase and consumption, stress in the decision-making and in potential cognitive dissonance, etc.). Clearly therefore, the same utility, for two different buyers / consumers, may have the same 'price', but be attributed a different 'value', because appreciated in discretionary, subjective and therefore psychological terms.

Finally, in an economic and business perspective, value can be detected even from an ethical and social point of view. A particular example is the 'means-end chain' and its *instrumental* and *terminal* values (Reynolds and Gutman). In this model, 'instrumental' is the value deemed a necessary and/or sufficient condition for achieving a further purpose, while 'terminal' is the value which represents a specific purpose [Lambin, 2008]. In the health sector, for instance, the effective competence of the organization represents a health instrumental value (in ethical terms, a physician has to be competent, so to provide an appropriate service for the citizen), while a terminal value (the terminal value *par excellence*) represents the health of the patient in particular and the health of the

community in general (in social terms, fundamental in all healthcare systems, but especially in systems based on solidarity, as in Italy).

'Transparency' can also be attributed to one or the other category (instrumental or terminal). In the context of the public administration, for example, the absence of even the mere suspicion of corruption is a fundamental attribute of multidimensional public interest (i.e. transparency as a terminal value, with a marked 'social' connotation), but it is clear that especially in an organizational perspective, and particularly in the health services area, it can also be connotated in essence, as instrumental value, improving performance in terms of management (more effective, more efficient, cheaper and generally more 'legitimate' and more 'ethical') of available resources.

Considering a single utility, a citizen of the Campania Region for example may not have a direct interest in how transparent the waiting list of a healthcare facility in the Lombardy Region turns out to be, but she/he could quite rightly have an indirect interest (both 'ethical' and 'social') to the effect that all Italian citizens should benefit from the same treatment, a right sanctioned by the Italian Constitution (art. 3 the right to equal treatment and art. 32 the right to good health). The Italian National Health Service permits (and in some ways offers incentives towards, in terms of competitiveness) regional mobility, whereby a citizen of the Campania or any other Region can expect transparency in the management of a hospital waiting list even in the Lombardy Region.

The example is purely theoretical and obviously is without prejudice. However, the mobility rate indicates respectively a positive and negative trend in the two Regions, North and South. As will be highlighted in this paper, in order to strengthen further the claim of transparency in the national healthcare system, the National Health Service emerges as an integrated system, both at the structural level (hospitals and similar structures), sectorial (in the sense of first, second and third sector) and regional (i.e. from Region to Region within the Italian territory, but even outside national borders, following procedures required by the specific health organization of reference).

From the perspective of the citizen/user/patient, it should be noted that in normative terms, "... transparency is understood as total accessibility, including the publication on the websites of government departments relative to information concerning: the organization; indicators relating to performance management and use of resources; results of measuring and evaluating carried out by the competent bodies in order to encourage widespread forms of monitoring and compliance with principles of good conduct and impartiality. Essential levels of service are provided by the government in accordance with art. 117, second paragraph, letter m) of the Constitution" (art. 11 – "Transparency", paragraph 1 of Legislative Decree no. 150/2009, "Implementation of the law dated 4th March 2009, no. 15, for the optimization of productivity, efficiency and transparency in public administrations) (*our translation, also for what follows)*. More generally, an information system can be envisaged as moving along a continuum: at one end public regulation and at the other, the absence of public regulation (this is never completely the case however, especially in the health sector). In the Italian health system, as mentioned above with regard to the 'sectorial' level, privately run hospitals are numerous (for-profit or non-profit), integrated at different levels with the public

sector. Private healthcare organizations accredited by the National Health Service are institutionally, virtually identical to those of the public sector. In cases where they are not accredited however, authorization is necessary to carry out clinical/healthcare activities, which are subject to a 'publicising'. Clearly, being deprived of 'accreditation' is detrimental both in economic terms and the contrast of interests in favor of the community/society. Thus, information asymmetry may have more space, save positive pressures arising from sense of responsibility, deontology, ethics, etc.

In normative terms, the Legislative Decree no. 33/2013 ("Reorganisation of the rules concerning the obligations of disclosure, transparency and dissemination of information by public authorities") is "... prepared for the implementation of principles and criteria under the delegation of art. 1, paragraph 35 of the Law of 6th November 2012, n. 190, on "Measures for preventing and combating corruption and illegality in public administration" [and] reorders, in a single regulatory body, numerous laws relating to obligations of information, transparency and publicity by public administrations. [...] Some of the most important [...] include: the establishment of the civic right of access, the obligation to prepare and publish a three-year plan for transparency, the obligation to appoint the person responsible for transparency in every administration, a review of the rules on transparency in the patrimony of politicians and public administrators and their nominees, the obligation to define in the home page of the corporate website of each entity a special section called 'transparent administration'" (source: Vv.Aa., "Provisions on transparency in public administrations – Information note on the Legislative Decree no. 33/2013", ANCI, April 2013). Art. 41 of the Decree, in particular, is reserved exclusively for healthcare administrations and agencies (see extract from the Official Journal).

Transparency of the National Health Service.

1. *The administrations and agencies of the national health service, regional health service, including local health authorities and hospitals, agencies and other bodies and public agencies involved in planning and delivery of health services, are liable for compliance of all disclosure requirements, established by law.*
2. *The health agencies and hospitals publish all information and data concerning the procedures for appointing the general director, medical director and administrative director, as well as the positions of head of department and head of simple and complex structures, including notices, the carrying out of procedures, the conferment deeds.*
3. *For medical managers mentioned in paragraph 2, with the exception of those responsible for simple structures, the disclosure requirements laid down in Article 15 are to be applied. For professional activities, pursuant to paragraph 1, letter c) of Article 15 also professional services performed under intramoenia are to be intended.*
4. *It is published annually and updated the list of accredited private health organizations. Also the agreements signed with them are published.*
5. *The Regions include compliance with the requirements established by the law among the requirements for accreditation of health organizations.*
6. *Institutions, hospitals, public and private organizations that provide services on behalf of the national health service are required to specify on their website, in a special section called 'waiting lists', the expected waiting time and the average time to wait for each typology of service provided.*

In general, therefore, it seems possible to articulate the concept of 'value' in at least five fundamental components: technical, commercial, psychological, ethical and social. In the perspective of the present work, we intend to investigate the specific value of transparency in health, in particular when accompanied by computerization, contestually more or less pushed ('low' or 'high') by/of the health organization: this integration is derived from a theoretical framework, in other words a kind of conceptual map, see Fig. 1, which summarizes the modes of evolution – which could also be considered further 'values' – whereby greater (or lesser) transparency in the health organization is achieved, thanks to computerization (e.g. in the presence of scarce computerization, as shown, more opacity results).

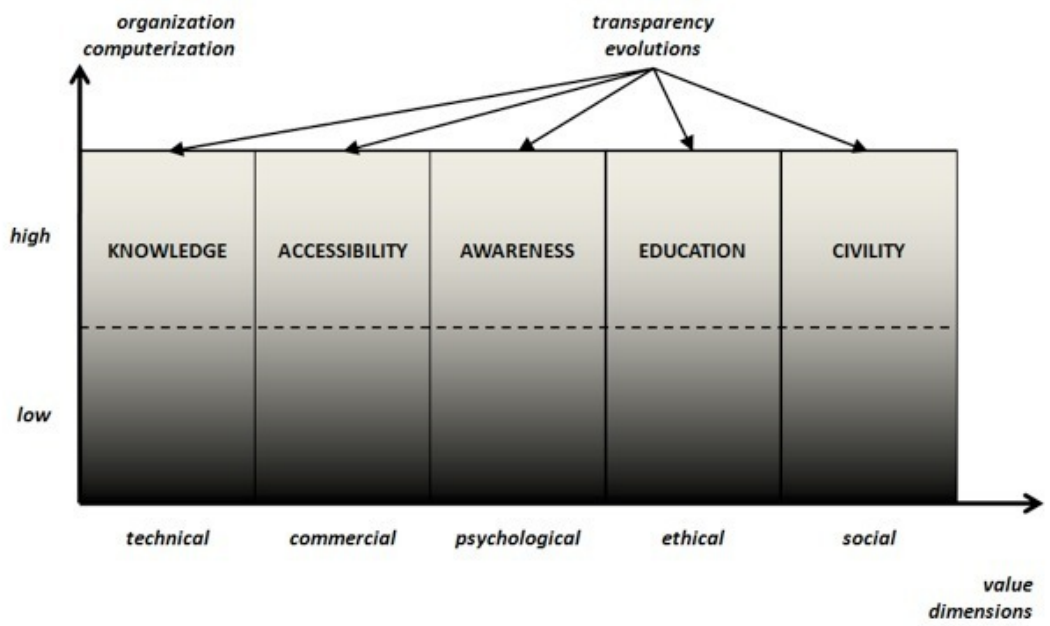


Figure 1
Evolutions / values of transparency in health computerization

Transparency, considered first of all an organizational need, actually becomes a real expectation in more informed modern societies. A classic example is the New York Times new building, designed by the architect Renzo Piano with an extensive use of transparent glass, symbolizing the absolute honesty of the newspaper in favor ot its readers and society in general. Citizens, dynamic, conscious, alert, needing/wanting quality, are investing even more heavily in economic and cultural terms, in correct behavior (*ethics*) and subsidiary (*social*) 'civilized' values. More transparency should result also in less bureaucracy and greater productivity potential, but it should above all, push organizations that are promoting it to operate in a more honest, open and healthy manner.

Without ethics, however, in a concretely modern vision of business scenarios, it seems not possible to aspire to quality, which, in the absence of 'business-generic' ethics, would be limited to mere technical results, not realizing the normal ambitions of entrepreneurial sustainability. Moreover, it is clear that transparency finds an enemy not in bureaucracy (which, mechanical or professional, remains one of the basic organizational configurations), but in the excess of bureaucracy, that emerges when unnecessarily formal, harmfully redundant mechanisms coarsen the efficacy and efficiency of organizational functioning.

3. The multidimensional evaluation of transparency in the health sector

Nowadays, data, information and knowledge have become real assets, probably the most important; thus, they have to be protected, both in the phases of production, storage, distribution and use. They can also evidently, be considered in more economic and business terms, i.e. 'goods', resources, which normally are scarce, later requiring both governance and management. In particular, such goods are intangible resources, with their own value (autonomous and marketable), seeing as, (as mentioned previously) they can be produced, stored, distributed and consumed [Gambaro and Ricciardi, 2003].

It is evident that a person in possession of a particular item of information, i.e. information that would be useful in a given context, has a specific advantage, if compared to an individual in the same situation without the information in question. Information asymmetry, as in this case, is generally a situation of economic injustice, in many contexts perfectly legitimate and profitable [Schilling, 2009]. However, if linked to access to healthcare, it becomes intolerable generally speaking and even more so in a solidarity system such as that in Italy, strictly based, at least in theory, on equity and equality.

It is no coincidence, in fact, that in the studies of general economics information symmetry and asymmetry constitute hypotheses capable of projecting scenarios completely different. Indeed, studies that are based on the perfect symmetry are obvious artifacts of reality, even though, obviously, they constitute a fundamental basis for the development of models of economic functioning and behavior with an undoubted social importance (we may think, for example, of studies on perfect competition, the theories of Modigliani-Miller, etc.).

Such information asymmetry in healthcare can occur at different levels: the most 'intolerable' refer to the opportunity/possibility, for some to use services or have access to services in a more advantageous, 'competitive' manner compared to others (one person booking an x-ray is obviously competing with another patient/citizen in need of the same exam), or, even more generally, with respect to the community. The crux of the question is that in healthcare systems based on solidarity benefit mechanisms for one to the detriment of others (e.g. 'private' healthcare), should not exist, only wellness and welfare. Studies of health economics, although based mainly on negotiation mechanisms of individual utility (i.e. the agency theory, with the general practitioner in the role of *agent* and the patient in that of *principal*), cannot ignore the prospect of equity, especially in terms of allocative efficiency and therefore in the perspective of health planning [Dirindin and Vineis, 2004].

In healthcare, in particular, the cost of information can engender dual degeneration: moral hazard (i.e. defensive medicine) or adverse selection (information is paid for only by disadvantaged subjects) [Ziliotti, 2001; Cappelli and Renzi, 2010]. Without a policy of transparency therefore, traditional economic considerations in terms of uncertainty seem to vert [Sofio Donia, 2001] on the propensity – or otherwise – for risk (for example, in the triage) [Rumiati and Bonini, 2010].

Transparency may also be necessary not only to combat the degenerations of information asymmetry, but also negligence relating to information and communication. In fact, patients very often are not informed about the potential range of services performed by a given structure, obtaining information either from their general practitioners, the mass media or by word of mouth, in order to make delicate decisions. It can also happen, on the other hand, that the well-informed individual has to adopt a policy of transparency in order to benefit (not having, in other words, exclusive rights to the information); or, that the well-informed individual communicates such items of information to others either in order to obtain a filter effect of such knowedge or as a collective experience (e.g. the increasingly popular role of social networks relative to health and healthcare).

In economic terms, the contribution of the transparency can be analyzed firstly in the perspective of the two main objectives of the appropriate functioning of any organization, i.e. efficacy and efficiency, enhanced, in the health sector in general and in the public sector in particular, by 'economicity' [Tommasetti and Cuccurullo, 2004; Marinò, 2008]. It is only common sense that a more transparent organization (in terms of legality, ethics, accountability, meritocracy, competitiveness, etc.) implies greater efficacy and efficiency if compared to those adopting more opaque policies; the reference to 'economicity', in addition, stretches the concept in the direction of universality, equity, justice and so on, all of which are sadly lacking in health systems that lack transparency.

In business terms, however, the concept of transparency becomes an enabling factor for competition, flexibility (both technical and economical), a focus for specialization and so on. Generally speaking, therefore, transparency in (health) markets encourages the integrating of virtuous conflicts of interest, enabling an in-depth understanding of the choice mechanisms of the performance mix.

In the health sector competition always is (should be) internal and not external, i.e. committed to delivering even better performance once a 'basic level of care' is guaranteed, hopefully, standard in benchmarking terms and patient-centred. The inevitable option of internal competition will mean that there could (should) be no competition between good and bad healthcare systems, but only between competitive levels of quality within the same. Transparency, therefore, becomes a necessary step firstly, to bring out the 'good' or 'bad' cases and, secondly, in order to classify or compare cases that are at least 'good practice'.

In the perspective of competition, it seems natural that transparency is an intrinsic virtue for operators who are more competitive (higher performance, more ethical, etc.), or for those whose outcomes/output are quality driven. Better performance (or at least a susceptibility/attitude to transparency) should be properly informed and communicated in a perspective of resonance with the stakeholders of the socio-economic environment (internal and external and, in

the latter case, micro and macro). Such performance also involves rational and emotional behavior that fosters market expectations in line with more transparent health policies in which standards (low or high) should match those of other competitors.

The same healthcare for everyone means that any technical tool (from gauze to syringes, from the stethoscope to the electrocardiograph, from scanner devices to those for computerized axial tomography and so on) are objects with an information content. Over time, they also become more precise and analytical thanks to the greater technology incorporated, especially knowledge input. In a certain sense, therefore, technology produces 'better' information and requires a health technology assessment (HTA). To deal with the evaluation of health technologies, HTA naturally, measures not only the technology itself, but also potential information and its subsequent use in a clinical process [Arduini, 2010]. Thus, it seems logical to apply criteria used for the economic evaluation of HTA also to transparency, envisaged as a factor capable of producing greater efficacy, efficiency and economicity in health care, a process in essence, borrowed from pharmaeconomics.

Some criteria for the appreciation of the (multidimensional) value of transparency in healthcare, according to the basis of representation can also be determined. The individual value of transparency can primarily, be attributed to the absence of asymmetric information: in economic terms, evaluation criteria should coincide with the elimination of collective/social costs equal to the enlarged net present value for the better informed subject. To the individual, in particular, the first three dimensions (at least) represented in the framework, i.e. technical, commercial and psychological, can be ascribed.

'Extended/greater' value could be achieved also within the healthcare organization by simplification and acceleration (with the same or improved standards of safety and quality) in terms of the different operations (internally i.e. bookings for specialist treatment made by indvidual departments to clinics, laboratories, radiology units, operating theatres, etc.) together with greater uniformity (and therefore, greater efficacy, efficiency, economicity) in organizational procedures, quality systems, accreditation paths, etc.

In our present society, characterized by immaterial production, distribution costs (in terms of media disposing of information) are less than production costs (e.g. information transmission via the Internet). This situation is (or at least should be) known to the parties involved, i.e. bidder (individual) and buyer (individual/ collective). In case of asymmetry, however, it is quite possible that those not in possession of the information (or do not know how to use it adequately, resulting in an asymmetry of 'theory' and another of 'practice') will be damaged in economic terms, although the cost could be low if both costs and resources are shared (i.e. with the operator in question having the information, knowing how to use it, but without the necessary resources – physical, financial, etc. – to use it). Moreover, information that becomes public automatically generates transparency and immediately loses value for the individual (agent), becoming even if only potentially, an infinitely greater value for the society.

In fact, whoever produces specific information, if valid and reliable, would most likely suffer no great cost (since the information exists now, it makes no sense to reproduce it more economically, but at most refine it to produce other information). This shifts the perspective of the analysis from the information economy to the knowledge economy, in terms of organization and distribution (i.e. mainly in coordination terms). As in any other economic environment, since information has a

value, producers, intermediaries and users (which may also coincide) are involved: if a user would like to retrieve x information individually, it would involve costs: $C_x \times N$ would be the total cost for all N users who wish to obtain x information individually (e.g. the direct and indirect costs that individual users, not in communication with one another, undergo when attending the Reservation Unified Center for information relative to the waiting list of a radiology department).

In symmetrical terms, once the communication infrastructure, including intermediaries (with the cost compensated for by extent of use), only C_x, would result because the direct costs of distribution of the information would be nearly zero at some point (or at least asymptotically approaching zero), while in a polyarchy structure, as in effect the organizations that produce or consume services and benefits in healthcare are, the number of connections needed to cover all the needs of communication is equal to N (N -1) in the case of one-way flows of information, or N (N -1) / 2 in case of bi-directional flows of information. Clearly enhanced value generally expressed in quantitative terms (units), would ultimately result from related social aspects.

As the complexity of health needs grows, the complexity of clinical information also grows, but this is not necessarily the case of the complexity of the communication infrastructure (e.g. health data, simple or complex, delivered on the same intranet). It is evident, therefore, how much saving both internally and externally, could result from transparency, engendered by the increase of resources available to attribute greater added value to the production of health services.

The 'ethical' value in transparency is as we have said, also fundamental, due to the need for the 'normal' functioning (legal, fair, caring, etc.) for any kind of service in favor of an individual and/or a community. In economic terms, evaluation criteria for this case should be the decisional analysis mechanisms, i.e. ratios of cost/effectiveness, cost/benefits, and cost/utilities. By virtue of transparency, it becomes potentially feasible to combat corruption, cronyism, favoritism (a sort of 'black area') thus increasing economic value in general for the state and for society.

Finally, 'social' value derives from transparency, due to solidarity in favor of any other individual in society. In economic terms, the criterion for assessment could be value for money. The social value of transparency, however, is the founding principle of additional values, instrumental from the 'means-end chain' perspective: i.e. the importance of equal treatment for users/patients or the availability of evidence of clinical practice for health professionals.

Transparency, on the other hand however, is potentially capable of generating not only advantages, but also disadvantages, at least in terms of situations at risk, just like any socio-economic case. In social terms, transparency might also lead negatively to insensitivity towards the individual in situations of distress (for example, inflexible respect of waiting list planning), while in economic terms, negative outcomes from excessive red tape deriving from observing principles of transparency (e.g. undue attention addressed to 'procedures'). Evidently these are degenerations, but unfortunately, quite capable of producing potentially negative effects in highly complex environments and contexts especially from a humanitarian point of view.

Another hypothesis of non negligible risk nowadays concerns the excess of information deriving from transparency, which is one of the 'disasters' of the information society. In healthcare, in particular, the excess of information could also result in litigation or rather the increase in litigation since this is quite widespread in the healthcare field.

4. The 'transparent' design of the health information system

In the perspective of this study, one of the main goals pursued from a multidimensional approach to organizational transparency (in addition to compliance with legislation and ethics) is the greater propensity of the structure for efficacy, efficiency and economicity. In terms of excellence, this means producing more and better, utlilizing fewer resources and resulting therefore in more sustainably. This economic ambition, in a sector as complex as healthcare, is linked above all to the social appreciation of such an approach, oriented to the production and delivery of health [Bellucci and Cardoni, 2008].

Consequently, an enterprise information system, especially for particular organizations such as health care structures, can be envisaged also in terms of transparency of information both internally and towards the outside. Transparency of organizational information to date, especially in the public sector, has always been seen as a kind of 'antidote' to malfunction (which in the case of health becomes specifically 'medical malpractice', not only in terms of outputs, but also in terms of means), but as we have already shown, important benefits are to be gained for organizational economicity, as a sort of positive externality on the contrary.

Assuming the perspective of transparency as a driver for the design of the organization and especially of the information system, it seems possible to delineate a logical and methodological approach that should lead to the transparency of information and, downstream, of the organization, which is the main purpose of our discussion. By following a 'shrimp' approach, very often used in project management, it is easy to argue that one can obtain organizational transparency (necessarily) only in the presence of information transparency; upstream, the latter is a consequence of information sharing, which, in turn, depends increasingly on the computerization of the organization; further upstream, this is a consequence of the implementation of organizational procedures (i.e. in the language of knowledge management, operating transformations from continuous to discontinuous), which, in turn, depend on organizational formalization, which, finally, descends, in the specific case of health and in a transversal vision to different health structures, from natural reference to professional bureaucracy [Mintzberg, 1996].

It is clear, as noted above, that transparency cannot constitute an end value in itself (except perhaps only in the public context, i.e. Legislative Decree no. 150/2009), but in more pragmatic terms, it is merely an instrumental value. Routine operations in healthcare enterprises will inevitably, therefore, have to remain discretionary (i.e. the classic example of urgency in waiting lists), because obviously, the patient and not the disease, or worse the procedure itself, has to be considered. It is equally obvious however, that discretionary procedures cannot become the primacy of the individual. Transparency of internal rules and limits of discretion does not prevent opaque situations, but at least it limits them as far as possible: i.e. 'clinical discretion', a clear example of properly exercised discretion, if justified and transparent, in a rigid system of constraints.

In modern healthcare it is clearly anachronistic to think of organizational functioning or professional behavior inspired by a procedures in the bureaucratic sense, lacking adequate contributions of communication, attention, empathy, etc. At the same time, however, we cannot run the risk of being only on the side of the patient, because we should also be on the side of the healthcare operators, exposed as they are to growing complexity. In this sense procedures, clearly detailed, thought-out, designed, tested, validated and appropriately and promptly updated, are expected to be a fundamental contribution to the good exercise of the profession; also from a psychological point of view, because it reassures the operators in their relationship with their patients. Thus, transparency can obviously help to improve the organizational climate and consequently the psychological and performance profile of the healthcare individual professional.

As nowadays enterprises do not compete with each other as individual companies, but as integrated supply chains, no individual physician can be considered competitive except in her/his healthcare organization taken as a whole. From information to 'transparency' then can the contribution to efficiency be derived not only in terms of the individual enternprise, but also of the healthcare network as a whole. A transparent health information system would be an example of a highly concentrated network structure in terms of resources, performance, outputs and outcomes, and so on. In social terms, it should be considered also as a vector of equality for access to health services in a multidimensional sense: quantity, quality, speed, efficiency, cost, etc., valid for all circumstances and contexts, but especially for health systems based on solidarity (equal access may also mean transparent compiling of waiting lists, adequate information about services and performances, awareness of the mix between internal and external supply – i.e. intramoenia – and so on).

Unfortunately, even so, in the health sector there are boundless levels of information asymmetry (internal and external with respect to organizational confines) not only in 'traditional' scenarios such as labor, credit, insurance, etc. The service sector is certainly that with more abundant information content and therefore information and even more so the absence of information, is a critical resource for health services. In the modern healthcare system, information travels at network speed, via a macroscopic socio-economic supply chain (through, vertical and horizontal subsidiarity, at least in Italy). Without information and communication, a health organization is destined to remain a monolithic entity, not only opaque as to operations, but also isolated in terms of the competitiveness of the health system market (it is likely that having the opportunity for choice, patients would tend to use the most transparent structure).

The push for transparency, however, is amplified by the diffusion of web 2.0 tools, which constantly produce user-generated content, often considered as more reliable sources of information and communication. The reason for such interpretation is obvious: citizen opinion, especially if organized within huge communities, is associated automatically with greater value, because of an innovative word-of-mouth process, counteracting the traditional absence of transparency on the part of official bodies.

In addition, web 3.0 (i.e. the semantic web) is now also on the web horizon. It is bound to represent a real cultural revolution for the Internet in the traditional sense, but it will also provide a decisive push in the direction of transparency. At the basis of this new web philosophy, is data and related content authentication

which should restore reliability to the information found on the Internet, producing in turn even more trust.

Transparency, furthermore, as mentioned in the Introduction to our paper, is fundamental in the sense that we must not only distinguish environments/contexts lacking information/communication from those characterized with a higher intensity or even by an excess of information, the latter ending in effect, in a substantial lack of transparency. A particular example is an article of the Brunetta Reform in Italy (Law 15/2009 and Legislative Decree no. 150/2009), which requires public bodies to communicate by Internet the remuneration paid for appointments conferred in the public sector: the effort is certainly relevant, since it indicates an effective cultural change, but it should be stressed that this form of 'transparency' is often cumbersome, because frequently only available through 'hidden' links or through data published in 'pdf' format, not aggregated/disaggregated for integrated and/or cross analyses, and therefore, not generating widespread data usability.

In such an evolutionary scenario, a significant role for health service computerization can be played by ECDL Health, which is the most recent computer science certification for health professionals (for which in Italy AICA is responsible). Certification is mainly focused on criteria for healthcare data and information processing. Investment in training for widespread appropriate use of health informatics, in the past regulated by law (in Italy the first version of Annex B of the Legislative Decree no. 196/2003, modified in 2012 by the 'Simplifications Decree' of the Monti Government), respond to the fundamental attention that needs to be addressed to human resources, which constantly play a unique role in the organization, the rationale being that the most sophisticated computer system is actually worth very little if used by incompetent staff, inadequately educated and trained for exploiting the technology to the full.

The above considerations as regards transparency can thus be contextualised to ECDL Health as it is our firm conviction that such value, especially by virtue of its multidimensional nature, has a considerable potential precisely in the health sector. In other words, transparency can be considered an important principle for the design of health information systems, consistently generating downstream the need/opportunity for awareness in healthcare workers of the sense and motivation for the appropriate use of daily healthcare informatics. In any case, even beyond the perspective proposed in this study, the re-adjustment of the cultural horizon on the part of the ECDL Health, at least with respect to recent changes in regulations as regards transparency in the Italian Health Service in general, would be essential.

5. A content analysis of the ECDL Health Syllabus

Underpinning the professional certification in question is a body of concepts, knowledge and skills in the field of health information systems which has been formalized within the Syllabus - 1.1 divided into 4 sections: 'Concepts'; 'Hospital tasks'; 'User's skills'; 'Policies and procedures' (a total of 52 items) constituting the assessment framework in order to obtain certification. The Syllabus, a knowledge reference for health information systems, also in an international

perspective, is subject to updating in the event of development of reference skills (computer science, organization, law, etc.).

In 2013 a new version of the Italian Syllabus (1.5) was released. though it is currently still under internal review on the part of the AICA working group on ECDL Health. Syllabus 1.5 retains the division into 4 sections (but from 'Hospital tasks' it evolves to 'Professional duties and responsibilities') and increases slightly the number of items from 52 to 54.

By virtue of this new Syllabus, it was considered useful to make a content analysis of the two versions of the Syllabus (1.1 – and 1.5) to ascertain the extent to which the concept of transparency is taken into account, especially in view of recent regulatory changes in Italy (Law 15/2009, Legislative Decree no. 150/2009, Legislative Decree no. 33/2013). The methodology of the survey was structured as follows:

A. definition of a keyword set, connected to the multidimensional value of transparency studied in this work, traceable in the text;
B. identification and measuring of the (absolute) frequency of occurrences;
C. identification and measuring of the (cumulative) frequency in the reference areas (as regards lexical proximity);
D. comments on findings/outcomes.

By means of this methodology, keywords were considered as follows: terms associated in a general sense to health information transparency, i.e. *transparency, appropriateness, timeliness, accessibility, sharing, confidentiality, identification, recognition, integration*. In addition to nouns, related adjectives and adverbs (from appropriateness, we parsed appropriate and appropriately) Subsequent to initial analysis, it was decided to include also the terms 'access', 'accesses', 'identify' and 'recognize'.

We then proceeded to calculating occurrences, using, due to the limited universe under investigation, standard office automation tools. The results of the analysis are given in Table 1 (the Italian terms are repowered on the basis of number and gender).

We found that:

- the concept of transparency in the strictest sense is never expressly stated in any of the two versions of the Syllabus, as the reference area registers zero occurrences, albeit summing up all the keywords of the area;
- the concept recording the highest number of occurrences, at area level and in both versions, is related to the term 'accessibility', which in some ways obviously recalls transparency, both from the perspective of internal users (health professionals) and that of external users (patients and citizens);
- the concept expressing the most numerous increase in percentage of occurrences, at the area level, is related to the term 'integration', which, however, together with that of accessibility, seems to define at least in the internal perspective of health information systems, the horizon of the 'transparent' organization;

- in the new version of the Syllabus all the concepts analyzed show a number of occurrences equal to or greater than the previous version (the concept of transparency, as commented previously, is absent in both versions).

Analysis of the Occurrences	Syllabus		Areas		Analysis of the Occurrences	Syllabus		Areas	
Keyword	*1.1*	*1.5*	*1.1*	*1.5*	*Keyword*	*1.1*	*1.5*	*1.1*	*1.5*
Trasparenza	0	0	0	0	Riservatezza	5	6	6	8
Trasparente	0	0			Riservato	0	0		
Trasparenti	0	0			Riservata	1	1		
Trasparentemente	0	0			Riservati	0	0		
Appropriatezza	0	0	2	3	Riservate	0	1		
Appropriato	1	1			Riservatamente	0	0		
Appropriata	1	1			Identificazione	0	0	2	6
Appropriati	0	0			Identificato	0	0		
Appropriatezza	0	0			Identificata	0	0		
Appropriatamente	0	1			Identificati	0	0		
Tempestività	0	0	1	1	Identificate	0	0		
Tempestivo	0	0			Identificare	2	6		
Tempestiva	0	0			Riconoscimento	0	1	3	7
Tempestivi	0	0			Riconosciuto	0	0		
Tempestive	1	1			Riconosciuta	0	0		
Tempestivamente	0	0			Riconosciuti	0	0		
Accessibilità	1	1	9	11	Riconosciute	0	0		
Accessibile	0	0			Riconosciutamente	0	0		
Accessibili	0	0			Riconoscere	3	6		
Accessibilmente	0	0			Integrazione	0	2	0	4
Accesso	8	7			Integrazioni	0	1		
Accessi	0	3			Integrato	0	0		
Condivisione	0	0	2	2	Integrata	0	0		
Condiviso	0	0			Integrati	0	0		
Condivisa	0	0			Integrate	0	0		
Condivisi	1	1			Integratamente	0	0		
Condivise	1	1			Integration	0	1		
continues on the other column					**Total**	**25**	**42**	**25**	**42**

Table. 1
Synoptic and comparative table of the results from the content analysis

Prior to its final release, therefore, we consider it useful to include in Syllabus 1.5 references to transparency, now regulated by various laws in Italy specifically referring to the healthcare sector (see Legislative Decree no. 33/2013, article 41). Moreover, it would seem that this adjustment should be oriented (mainly) towards the availaibility of information systems towards the outside, as a 'system of information' [Carignani, 2004]. This observation also stems from the fact that in our opinion, the content analysis has returned positive results regarding the internal perspective, i.e. collaboration between information flows, processes and operators.

6. Conclusions

With the development of computer science and business information systems, information processing, both within and outside the organizations, has become increasingly easier (in the sense of 'user friendly'), widespread and open. In particular, the introduction of the Internet (and the associated solutions of intranet and extranet), with the subsequent development of web 2.0, has actually changed the manner, purposes and in some ways, the 'styles' of using data, information and knowledge, especially in business contexts. IT innovations, moreover, very often become over time, genuine social innovations (e.g. home automation, smartphones and social networks), especially in areas where the human element is prevalent. In this sense, the one sector of maximum relevance is without a doubt the health sector, whose services are by definition, oriented to producing the most human driven of resources, i.e. health. Such technological and sociological combinations need to be constantly informed nowadays of 'values', either in terms of social responsibility, norms or prospects of improvement (efficacy, efficiency and economicity). Among these values, transparency occupies an important role today, especially in the health sector and also in economic terms, in order to ensure sustainability for future generations [Kotler et al., 2010].

In our study, we have considered health transparency from a multidimensionality perspective, theoretically investigating a possible reconstruction of its 'values' to propose an overall framework, in order to highlight the applications from an economic, business and social point of view. Our study, a conceptual paper, has thus focused on the dynamics that generates transparency starting from health service computerization, examining the functioning of healthcare environments, markets and enterprises in the perspective of information management and communication in health, arriving at the definition of a potential principle of transparency for the design of health information systems.

In addition, with reference to ECDL Health, the professional certification in health informatics for which in Italy AICA is responsible, we have also carried out a content analysis on versions 1.1 and 1.5 of the Syllabus, in order to verify the attention addressed to the transparency of health information systems. The survey results are uneven: a direct reference to transparency is lacking in both versions, but there are numerous references to specific 'attributes' of transparency, especially in a strictly computer science sense. There are therefore potential margins of development for the new Syllabus by virtue of a) the recent changes in Italian law, b) the necessity for further openness on the part of healthcare enterprises towards comparison and evaluation processes, and c) the ever-increasing demands, by citizens and operators, for information, participation and commitment.

In conclusion, the content analysis was carried out on a 'desirable' corpus of concepts, knowledge and skills in the field of health informatics, to ascertain the extent to which effective competences, skills and attitudes are practiced and evolve in the routines of health enterprises. A more transparent healthcare organization, based on more transparent health information systems, constantly

provided with appropriate mechanisms for information security [Carignani and Rajola, 2010; Festa and Teti, 2010], represents nowadays a real business 'attractor', certainly in the perspective of the economic value to be generated (due to greater efficacy and efficiency), but especially in the perspective of human, social and civil values, inherent in the health needs of individuals and populations.

References

Arduini R., Economia e gestione delle aziende sanitarie, Angeli, Milan, 2010.

Bellucci A., Cardoni A., Elementi di economia delle aziende sanitarie, Giappichelli, Turin, 2008.

Bracchi G., Francalanci C., Motta G., Sistemi informativi d'impresa, McGraw-Hill, Milan, 2010.

Cappelli L., Renzi M.F., Management della qualità, Cedam, Padua, 2010.

Carignani A. (a cura di), Tecnologie dell'informazione e della comunicazione per le aziende, McGraw-Hill, Milan, 2004.

Carignani A., Rajola F., (a cura di) ICT e sistemi informativi aziendali, McGraw-Hill, Milan, 2010.

De Witt G., Informatica moderna e produttività d'impresa, Angeli, Milan, 2001.

Dirindin N., Vineis P., Elementi di economia sanitaria, il Mulino, Bologna, 2004.

Donia Sofio A., Microeconomia sanitaria e politiche d'intervento, Aracne, Rome, 2001.

Festa G., Teti A., Il contributo della certificazione ECDL Health alla sicurezza dell'informazione in sanità, in Di Resta F., Ferraris di Celle B., (a cura di) La sicurezza nei sistemi informativi sanitari, Edisef, Rome, 2010.

Gambaro M., Ricciardi C.A., Economia dell'informazione e della comunicazione, Laterza, Bari, 2003.

Kotler P., Shalowitz J., Stevens R.J., Turchetti G., Marketing per la sanità, McGraw-Hill, Milan, 2010.

Lambin J.J., Market driven management – Marketing strategico e operativo, McGraw-Hill, Milan, 2008.

Marinò L., La ricerca dell'economicità nelle aziende sanitarie locali, Giappichelli, Turin, 2008.

Mintzberg H., La progettazione dell'organizzazione aziendale, il Mulino, Bologna, 1996.

Rumiati R., Bonini N., Decisioni manageriali, il Mulino, Bologna, 2010.

Schilling M.A., Gestione dell'innovazione, McGraw-Hill, Milan, 2009.

Teti A., Festa G., Sistemi informativi per la sanità, Apogeo, Milan, 2009.

Tommasetti A., Cuccurullo C., Mappe strategiche e relazioni sistemiche tra indicatori di gestione: il caso delle aziende sanitarie locali, in AA.VV., L'evoluzione del controllo di gestione. Modelli ed esperienze, Angeli, Milan, 2004.

Tonti A., La semplificazione dei processi e delle procedure, Egea, Milan, 2002.

Tramarin A., L'ospedale ammalato, Marsilio, Milan, 2002.

Ziliotti M., L'economia dell'informazione, il Mulino, Bologna, 2001.

Biographies

Aurelio Tommasetti is the Rector of the University of Salerno (Italy) and a Full Professor of Accounting. He was visiting professor at the University of Maputo (Mozambique). He has taught several courses, such as financial accounting and management accounting. He has published several articles in national and international refereed journals, Health accounting and controlling are one of his most important research topics.
email: tommasetti@unisa.it

Giuseppe Festa is an Assistant Professor of Management at the University of Salerno (Italy). He has taught several courses, such as economics and management of health organizations and economics and management of information technology. He has published several articles in national and international refereed journals. Health information systems are one of his most important research topics.
email: gfesta@unisa.it

Envisioning Smart Disclosure in the Public Administration

G. Ciaccio, A. Pastorino, M. Ribaudo

Abstract. *Currently, public administration is undergoing significant transformations, driven by a greater demand for transparency and efficiency in a participative framework involving nonprofit organizations, enterprises, and citizens, with the modern network infrastructure as a common medium. The Open Data movement is considered one of the keys to this change. One of the forthcoming evolutions of Open Data is the idea of "smart disclosure" of personal data, mostly managed by public administrations, in order to allow third party applications to provide new personalized online services to individuals and organizations. In this paper we propose a possible implementation of the "smart disclosure" idea, that takes advantage of the OAuth 2.0 authorization framework. OAuth 2.0, if properly implemented, guarantees access to selected personal data upon authorization of the individual data owner. An implementation is presented together with possible use cases.*

Keywords: Open Data, Smart Disclosure, OAuth.

1. Introduction and motivation

The *Memorandum on Transparency and Open Government* signed by the US president Barack Obama [Obama, 2009] has fostered a new era for the public sector in which *transparency, participation, collaboration* and, ultimately, *Open Government* should become central in the democratic decision process. Administrations should become more transparent and promote the use of new technologies to ensure that the data they routinely produce and manage are made available online so that they can be leveraged by any party: enterprises, private citizens, public entities, and other branches of the public administration. This document marked the official onset of the so called *Open Data movement*.

Shortly afterwards, several public administrations in the USA and UK started releasing massive amounts of Open Data in the form of aggregated datasets made available on their websites. The first catalog of Open Data was published in May 2009 by the US Government (http://data.gov), followed by the UK (http://data.gov.uk). The British Government is now at the forefront in Europe, engaging in an unparalleled effort towards widespread adoption of the Open Data paradigm.

Quoting [UK Government, 2012], "*Data is the 21st century's new raw material*": by means of handheld devices, social networks, cash dispensers, credit cards, people are directly or indirectly generating an unprecedented volume of data that is deemed to transform our very lives [Hoffman *et al.*, 2012].

The current wave of Open Data released by public administrations is largely made of formatted datasets of a static nature (i.e., they will not reflect changes occurring after the release date) and concerning diverse fields: politics, traffic and local transportation, tourism and culture, the environment, healthcare and welfare, cartography, and many others. Such datasets are roughly of two kinds, namely: aggregated and anonymized data (e.g. number of children in each school of the region); and identification data of public entities (e.g. names and addresses of restaurants in the region). No data concerning individuals have been released due to obvious privacy issues. As already stated in a position paper of ours [Ciaccio and Ribaudo, 2012], such a lack of *personal data* in the Open Data realm, along with the static nature of the released datasets, are weaknesses of the current wave of Open Data. Without personal data and without timeliness, it is indeed impossible to build useful services tailored to the actual needs of a given individual at a given time.

Many of the data managed by public administrations as well as private entities are of a personal kind. Consider, for instance, the huge amount of personal data contributed to the various online social networks, or the electricity consumption data collected and stored by energy providers, or the telephone and internet data collected by telecommunications companies. As these data are not in the Open Data domain, those public and private entities may act as the "owners" of our data. This means they hold a monopoly on services while we, the legitimate owners of the data, must abide by their terms and conditions concerning how our data are treated and used.

By unleashing personal data "into the wild", such a monopoly would collapse and a new ecosystem of personal services based on these data could flourish. While we cannot expect a corporate giant like Facebook to voluntarily relinquish the personal data contributed by us to their servers, we can easily imagine that a public administration could, and arguably should, do so. One possible scenario would see third parties routinely providing public online services that make use of personal data, and the administrations routinely providing personal data online to registered third parties on behalf of the legitimate data owners (the taxpayers). In such a scenario, the administrations are responsible for ensuring the authenticity and integrity of personal data, preventing any unauthorized access, yet allowing what is called a *smart disclosure* of personal data to the web.

The importance of personal data as an economic asset on its own is now being acknowledged worldwide [Schwab *et al.*, 2013],along with the need to strengthen trust by people in the possible process of smart disclosure to be undertaken by public administrations [Hoffman *et al.*, 2012]. Smart disclosure of personal data is considered a forthcoming process capable of "*enormous economic and civic good opportunities*" [Howard, 2012]. A recent white paper from the UK Government [UK Government, 2012] stresses the importance of

smart disclosure as an enhancement of the current Open Data movement. The 'midata' initiative by the UK Government (www.bis.gov.uk/news/topstories/2011/Nov/midata) [Shadbolt, 2013] and the Smart Disclosure initiative by the White House (www.whitehouse.gov/blog/2012/03/30/informing-consumers-through-smart-disclosure) are two programs aimed at promoting smart disclosure of customer's personal data which are held by companies and providers so as to allow people to make better choices.

It might be argued that adding personal data to the Open Data heap might jeopardize our privacy, if done in the wrong way. However, this risk is also present with the current process of releasing massive anonymized datasets. By definition, these datasets leak personal information, and information from many datasets may be jointly mined in search of individual profiles. The inferred profiles may sometimes be linked to real identities, leading to statistical de-anonymization or "identity disclosure through mosaic effect" [Hoffman *et al.*, 2012]. The whole Open Data movement would immediately come to an end, should these confidentiality concerns prevail over the individual and social benefits of transparency and smart disclosure. A balance between privacy and transparency must clearly be sought, with the information technology playing a key role.

Another criticism is that, once disclosed and no matter how "smartly" this is done, our personal data might be copied somewhere else and we, the legitimate owners, would no longer be able to exert control over the copies. But this indeed holds without disclosure as well, as we currently have no choice but to trust the entity that stores and "owns" our data, without any actual control by us. In addition, due to the lack of smart disclosure, we are forced to input our personal information by hand every time we register for a new online service (and abide by *their* terms and conditions). It would be much easier to refer to a single master copy of our personal data, either centrally stored or scattered across servers in a distributed system, and smartly disclose them to third parties after obtaining their formal commitment online to *our* terms and conditions (for instance, prohibiting unauthorized distribution of copies).

Last but not least, a proper technology for a "sufficiently smart" disclosure of data remains to be identified, along with a number of practical use cases working as an informal definition of what a smart disclosure is. In this paper we propose a few such use cases, and we advocate the use of the OAuth 2.0 authorization framework (http://oauth.net/2/) to achieve smart disclosure. On the basis of this approach, individuals are restored to their role of *resource owners* while administrations (public or private ones) are stripped of their de-facto ownership of personal data and keep a role of bare *resource managers*. A resource owner (namely, an individual) may grant online authorization to any third party application to use a given item of personal data located on a given resource manager, in exchange for a useful personalized online service that the application is expected to provide using that data item. The process of granting authorization is based on unforgeable cryptographic tokens released by a trusted *authorization server* with which individuals, applications, and resource managers, are all registered.

The balance of this paper is as follows. Section 2 introduces some significant examples of implementation of Open Data involving personal data. Section 3 introduces the OAuth 2.0 framework, showing how it can be used to grant access to personal data while preserving individual privacy. Section 4 proposes a few simple use cases in which OAuth 2.0 could be successful in supporting the smart disclosure idea. Section 5 discusses our current PHP implementation of OAuth 2.0, thanks to which we plan to demonstrate suitability with the use cases. Finally, Section 6 suggests possible directions for further research.

2 Smart disclosure stories

Several examples and case studies on the use of the current wave of Open Data do exist but, when the search concentrates on smart disclosure success stories, results become more difficult to discover. In this section we briefly recall some examples we are aware of.

The Department for Business Innovation & Skills in the UK has launched the 'midata' project (www.bis.gov.uk/news/topstories/2011/Nov/midata) which aims to give consumers more control and access to their personal data. The program involves various business sectors including energy, telecommunications, finance and retail. By letting customers access data about their purchasing and consumption habits, and safely add new data and feedback of their own, businesses have the opportunity to create rich, new person-centric applications while consumers can make better consumption decisions and lifestyle choices.

The next two examples are related to the medical context. In the private sector, Microsoft has proposed the HealthVault (www.microsoft.com/engb/healthvault/default.aspx) as "*a trusted place for people to organize, store, and share health information online.*" According to the HealthVault website, Microsoft offers an open platform for security enhanced data sharing amongst health services organizations and citizens. Any information entered into HealthVault can be, with the citizen's permission, re-used across many different apps and supplemented by a growing list of devices. Apparently, the service is now available for US, UK and Germany.

In the public sector an interesting success story is the project known as the Blue Button (www.bluebutton.com). Developed in collaboration with the US Department of Veteran Affairs (VA), the project allows veterans to go to the VA website, click a blue button, and download their personal health records. These records can be individually examined or shared for example with doctors or with third parties applications. The Blue Button download capability can help individuals access their information so they can manage their health care more effectively. We were unable to find technical details concerning how Blue Button is implemented; in particular, we do not know whether the various health data belonging to a given user are routinely copied from the health care centers to the Blue Button server, or are rather left at their originating servers and collected on the user demand; this second approach, however, would be a complete yet simple example of smart disclosure.

A very successful initiative is the Green Button project (www.greenbuttondata.org), similar to 'midata', and part of the White House's Smart Disclosure Program. Thanks to this program, consumers in the US can access and download their energy usage information provided by their utility or retail energy service provider, take advantage of online services and apps, and manage their energy consumption. From the scarce information we were able to retrieve (energy.gov/downloads/nist-green-button-presentation), it seems that the service leverages Apache Wink (http://incubator.apache.org/wink/) for exporting data in a RESTful way via the Atom Publishing protocol [IETF, 2007], plus Spring Social (http://www.springsource.org/spring-social) as the authorization framework for user-controlled smart disclosure. Spring Social leverages OAuth [IETF, 2012], so the overall picture is similar to the one we provide in this paper.

Falcão-Reis and Correia [Falcão-Reis and Correia, 2010] propose coupling Electronic Health Records (EHRs) with an extended version of OpenID (http://openid.net/) [Recordon and Reed, 2006] in an effort to implement a user-controlled system of Health Digital Identity for Portuguese citizens. They also propose leveraging OAuth 2.0 as an authorization technology for user-controlled access to EHRs, thus anticipating smart disclosure in the medical care field.

In all these examples, the recurring theme is that the wealth of personal data contained in medical records, telephone or energy usage reports, or other information sources, present a unique opportunity for software developers to build applications that can truly transform how individuals interact with their data to stay healthy and manage their care, to save energy and therefore money, in other words to improve several aspects of their everyday life.

3. OAuth 2.0

3.1. Goals and current status

OAuth 2.0 (http://oauth.net/2/) [IETF, 2012][Hammer-Lahav, 2010] originated in 2010 from a complete redesign of the previous OAuth 1.0 specification [IETF, 2010]. OAuth 2.0 is too high-level to be defined as a protocol specification. It should be considered as a blueprint of a protocol, within which many implementations are feasible, although possibly not interoperable with one another. This is maybe (and hopefully) just a sign of immaturity of the specification. Nevertheless the interest around this technology is huge: the IETF OAuth working group include members like Google, Facebook, Salesforce, Microsoft, Twitter, Deutsche Telekom, and Mozilla [Hammer-Lahav, 2010], and OAuth 2.0 has already been adopted by Google (http://code.google.com/accounts/docs/OAuth2), and Facebook (http://developers.facebook.com/docs/concepts/login), just to cite a few. We are quite confident that OAuth will soon become the reference technology for authorization and controlled access on the internet.

OAuth stems from a typical Web 2.0 use case with online social networks. Suppose a third party client application is offering a user a personalized service making use of that user's personal features provided by a social network (e.g. pictures, contact list, posting a comment) via a RESTful API. The naïve approach

of requiring users to release their credentials (username and password) to the third party client so that the latter could get those features from the social network is highly risky. A more appropriate solution is to release cryptographic proof of authorization, issued by the user (at least in principle) to the third party application, and to let the latter subsequently spend such authorization proof at the social network API in order to get the required features. In addition, the authorization proof may work as a form of user authentication whenever it grants access to user identification data stored in the social network or elsewhere.

3.2. Abstract protocol

Abstractly, OAuth 2.0 identifies four actors exchanging information in an ordered way (Figure 1). These actors are the *client application* or just "client", the *resource owner* or "owner" for short (or "user" when human), an *authorization server* (AS, the trusted entity), and one or more *resource servers* (RS) hosting data or services to be smartly disclosed. The AS and the various RSs may each belong to a distinct administration domain, but this is not mandatory.

The protocol is started by the client requesting authorization to access a given *resource* (e.g. a set of personal data), subject to a set of constraints called *scope* (e.g. read vs. write access to specific fields of personal data during a specified time window). The resource owner is shown the request and, if they agree with the scope, they may yield an *authorization grant* bound to the resource and scope. Such a grant is then exhibited by the client to the AS, which validates it and returns an *access token* valid for the required resource at a specific RS subject to the scope. The client finally passes the access token to the given RS along with the resource request; upon token validation and scope verification, the RS returns the resource to the client.

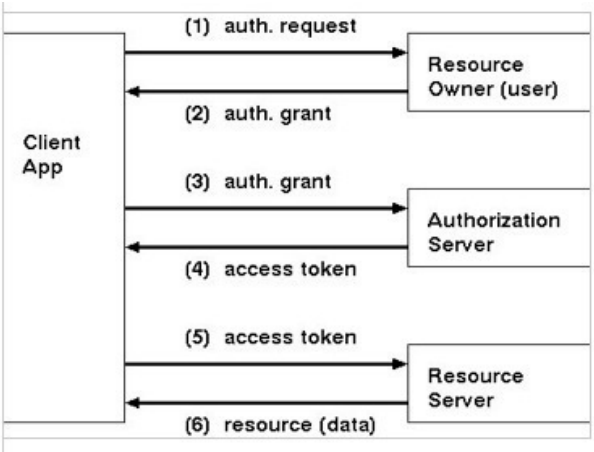


Figure 1
OAuth 2.0 abstract protocol, with the four actors (client application, resource owner, authorization server, and resource server) and their interplay. Actions are ordered by increasing number

3.3. Four ways of obtaining authorization

In practice, the authorization grant can be obtained in four possible ways. These four ways are called *flows* in OAuth jargon (IETF could not call them "protocols" due to the many details that are omitted or left undefined), and are nothing but instantiations of the abstract protocol scheme discussed above.

In the first flow the grant is obtained indirectly, with the AS acting as an intermediary under the control of the resource owner. The flow is depicted in Figure 2. With this flow, the client is typically (but not necessarily) a web application on a remote server, and the resource owner is typically (but not necessarily) a human with a browser initially pointed to the web application. The grant is represented by an *authorization code* issued by the AS after obtaining consent from the resource owner, and is delivered to the client through the browser of the resource owner via HTTP redirection (see the Figure).

In the current IETF draft, TLS protection is not required when the AS redirects the browser to the client after the authorization step; in other words, it is legal for the client redirection URI (Figure 2) to be an HTTP endpoint of the web application instead of an HTTPS one. On the one hand, this shortcoming is meant to cover client applications that are unable or unwilling to provide TLS endpoints (due to lack of resources, for instance). On the other hand, an authorization code sent to a non-TLS endpoint is transmitted in plain text and could therefore easily be eavesdropped. The stolen authorization code could then be used by an attacker client to obtain an access token from the AS. But if the client application has a persistent identity registered at the AS, that is, it shares a secret with the AS and can prove that it holds it (a "confidential" client, in the OAuth jargon), then the AS will prompt the client application to authenticate before converting the authorization code into an access token, thereby preventing the use of stolen authorization codes by rogue clients. As an additional security measure, the access token itself may be bound to the specific client identity (a so called "proof token" [IETF, 2012a], as opposed to an anonymous "bearer token"). In contrast, if the client does not hold a secure and persistent secret registered with the AS (a so called "public" client) then the flow is insecure, unless the client redirection URI is an HTTPS endpoint.

A security analysis of OAuth 2.0 is beyond the scope of this paper, and can be found in a dedicated IETF draft [IETF, 2012a].

In the second flow the authorization grant is *implicit*, that is, no intermediate authorization code is released. As apparent from Figure 3, with this flow the resource owner is typically a user with a browser, and the client is a Java applet or an embedded Javascript code, loaded by the browser from the client's website. The user yields credentials to the AS and grants authorization through the browser, and the applet is then given an access token directly.

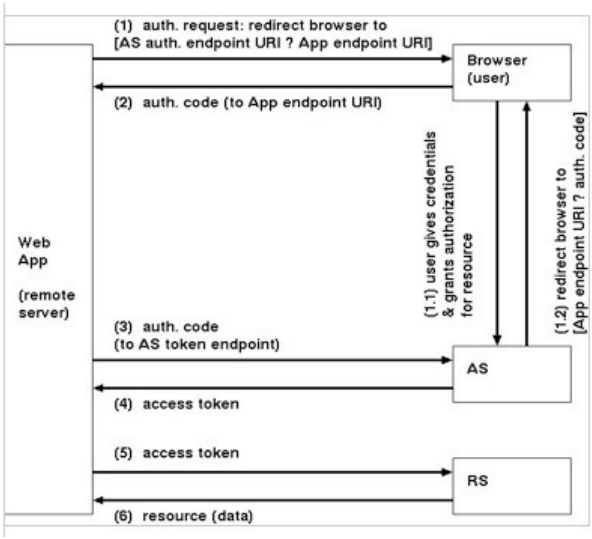


Figure 2
OAuth 2.0 Authorization Code flow, for web applications. The "?" denotes an HTTP GET parameter. See [IETF, 2012] for more details

The Implicit flow is said to be possibly more responsive than the previous one [IETF, 2012]. On the other hand, its typical application scenario is less secure. A client consisting of an applet or embedded script can in principle build a secret, but such a secret is not persistent (it might not survive from one run to another), so it cannot be registered at the AS (the client is "public", in other words) and thus the access token cannot be bound to any specific client identity. In addition, the received access token would be stored in the browser at the user device, whose security level is generally low. The access token could be easily stolen by a malware affecting the user device and then sent out to an attacker elsewhere, who can well make use of it, because it is not bound to any specific client identity.

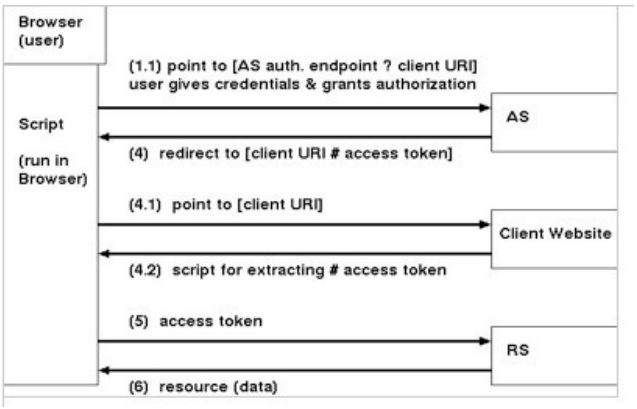


Figure 3
OAuth 2.0 Implicit flow, for applets and scripts running in browsers. The "#" denotes a URI fragment. See [IETF, 2012] for more details

The remaining two flows are more simplified. In the *Resource Owner Password Credential* (ROPC) flow (Figure 4), the client is typically a native application installed in the user device, and the user interacts directly and uniquely with it. The application takes credentials from the user, and sends these credentials to the AS for obtaining an access token. There is no easy way for the user to inspect the access scope being requested by the client to the AS, thus the user must trust the client. A native application can generate a persistent secret so it can have an identity, but such identity is as secure as the user device where the application is installed; so the typical scenario for the ROPC flow is generally vulnerable to, for instance, client impersonation attacks.

Lastly, with the *Client Credentials* flow (Figure 5) the client appears to be resource owner so it sends its own credentials to the AS, and get the access token directly. In fact this straightforward flow requires that the authorization by the actual resource owner is given outside of the OAuth flow (in the jargon of the old OAuth 1.0 protocol, this would be called a "two legged" variant of the protocol).

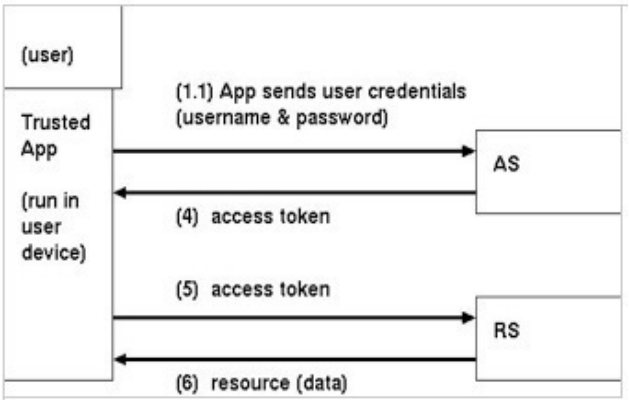


Figure 4
OAuth 2.0 Resource Owner Password Credentials (ROPC) flow, for trusted apps installed in the user device. The user credentials are sent by the app to the AS as proof of authorization grant by the user

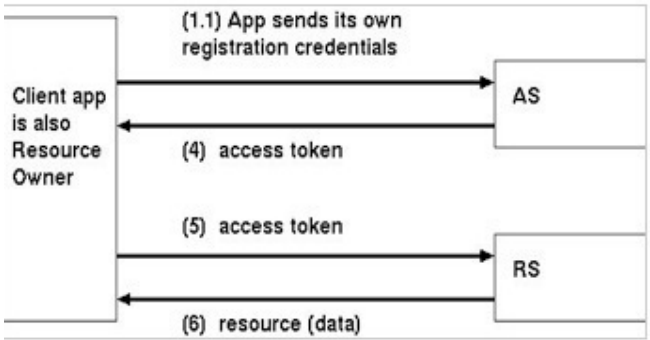


Figure 5
OAuth 2.0 Client Credentials flow, for client applications that are also resource owners. A simplified form of the ROPC flow with no user involved

3.4. Token expiration and refresh

When a resource owner gives authorization, they may specify a scope (what resources and for what kind of access) and also a lifetime (the expiration time of the authorization). So, in principle, an access token might inherit a lifetime equal to the authorized lifetime. In practice, however, each access token might be given a shorter lifetime, after which it must be refreshed, so a form of revocation can be easily obtained by simply refusing to refresh an expired token. More in detail, each new access token can be accompanied by a corresponding *refresh token* (Figure 6). The client can spend the refresh token back at the AS in order to get a new access token (and corresponding new refresh token) for the same resource under the same scope or a narrower one, until the overall authorization lifetime, as originally granted by the owner, has expired. A user could revoke authorization by instructing the AS to refuse refresh tokens from that client for the time being, but nothing can be done to invalidate an access token after it has been issued.

Refresh tokens are not permitted with the Implicit and Client Credential flows. With these flows, instead, an expired access token can only be replaced by a new one after replaying the flow from the beginning.

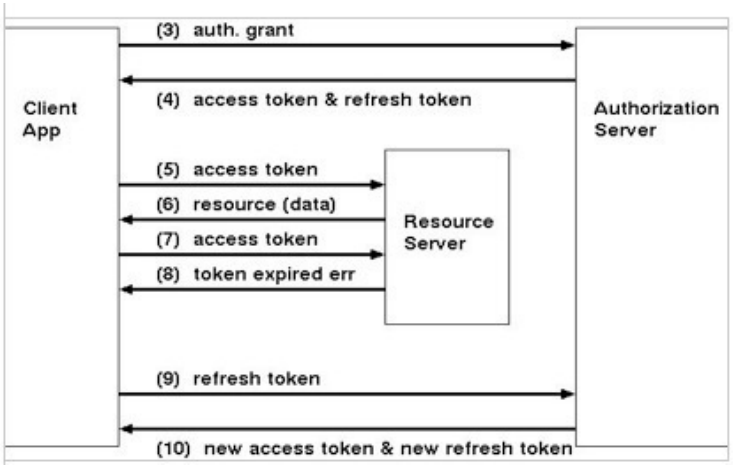


Figure 6
Refreshing expired access tokens in OAuth 2.0.
Actions are ordered by increasing number

3.5. Registration and authentication to the Authorization Server

From the above it is reasonable to deduce that resource owners (users), clients, and RSs, should establish a relation with the AS before engaging in any OAuth 2.0 protocols. The current IETF draft explicitly leaves out of scope (but does not forbid) any interaction between an RS and the AS, and

seemingly ignores the registration of resource owners, although this is indeed a necessary step if the AS is to issue authorization grants on their behalf. Only client registration to AS is explicitly mentioned in the IETF draft. After registration, the client is given a unique identifier that is valid at the AS. "Confidential" clients will be allowed to associate their secret credentials with the identifier at the AS for future authentication.

4. Use cases

This section introduces two possible use cases that show what a smart disclosure is and how to obtain it using OAuth 2.0. For each use case we first informally sketch the problem, then we discuss how OAuth 2.0 could improve current solutions and present aspects which are challenging to the current OAuth 2.0 flows.

4.1. Medical use case

This trivial use case is also found in the work by Falcão-Reis and Correia [Falcão-Reis and Correia, 2010], although they failed to mention some potential challenges arising from it.

Scenario. Alice needs professional medical care and asks for advice; a friend recommends Dr. John Smith. Alice makes an appointment and takes all the medical certificates she can find at home with her. She then summarizes her medical history, presents her problem and listens to the doctor's response.

How OAuth 2.0 could help. In many countries, health records are still predominantly paper records, given to patients after medical examinations. Patients' data are stored in local databases managed by various laboratories, hospitals, or other health care settings, all of which may be using different technologies and data representation that do not usually interoperate; storing data as PDF files is a common practice. Health information is therefore scattered across many places, and accessing the medical history of an individual is a complex task.

A possible solution requires the adoption of a machine-readable representation of medical data, along with the definition of an architecture for retrieving and merging records spread over several databases. In the medical context, standard data representations do exist, for instance the HL7 Clinical Document Architecture, a document markup standard that specifies the structure and semantics of clinical documents for the purpose of exchange. OAuth 2.0 could be the enabling technology of this architecture if

adopted to get authorization grants across different resource servers (the different medical databases). The use of OAuth 2.0 does not require centralizing data into a single repository: data are kept where they have been produced and are accessed upon authorization by the owner.

Figure 7 shows Alice and Dr. Smith. In order to know Alice's medical history, Dr. Smith uses a web application that connects to the distinct databases of hospitals that provide online health data upon verification of access tokens (resource servers Hospital1, Hospital2, Hospitalk in the picture).

Alice is in front of the doctor, so she can give online consent to access her personal resources (arrows numbered 1 and 2) by interacting with the web application being used by the doctor. After receiving an authorization grant, the web application applies for an access token (arrows 3 and 4) which is subsequently used to collect Alice's data stored within the resource servers (arrows 5 and 6), thus building a view of Alice's medical history without resorting to any paper document and, perhaps more importantly, without missing any of Alice's medical records.

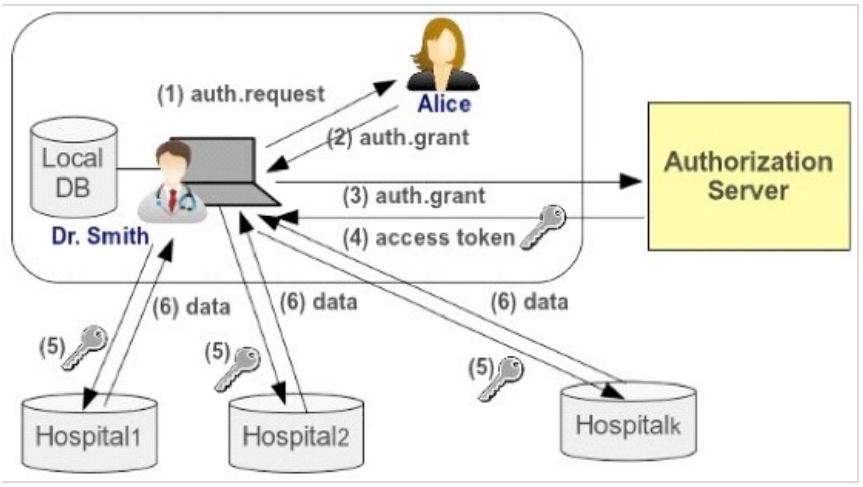


Figure 7
Dealing with distributed health data via OAuth 2.0

Challenges. Alice might not want to show all of her medical records to the doctor. Some illnesses, like HIV infection for instance, pose serious privacy concerns and might not be significant for the specific disease Alice is asking advice for. Scopes in OAuth 2.0 allow to specify which fields in a generic record can be accessed and which cannot, but they are not able to discriminate among distinct instances of the same field in distinct records. A workaround might be to have a specific field for privacy-sensitive information and letting the user decide whether to authorize disclosure of those sensitive fields or not.

In a different scenario, Alice might not be able to give authorization, perhaps because she is unconscious, due for example to a car accident. In this case, there must be another entity which has enough privileges to give authorization without being the data owner. This entity might be the head physician or a close relative of Alice's, and currently this scenario is not explicitly covered by OAuth 2.0, which contemplates a single owner for each data item.

In the medical context it is also very important to have access to aggregated and anonymized datasets for different purposes, for instance to perform statistical analysis for a given disease or to compare the performance of different hospitals. This release of large anonymized datasets is not covered by the current OAuth 2.0 flows which indeed grant access to specific records of an individual user. We will briefly discuss these points in Section 6.

4.2. Tax payment

Scenario. Bob needs to fill in his annual tax return. As usual, he visits his business accountant bringing lots of pieces of paper: his annual salary, medical expenses, documents concerning his properties, other expenses he can deduct from his income, and so on. The other possibility is to use a web application provided by the government and fill in an online form by manually copying all the data printed into the various paper documents.

How OAuth 2.0 could help. Figure 8 shows Bob in front of an innovative online service. He still has access to a web application to fill in the online form for his annual tax payment. This year, however, most of Bob's income information comes directly from various remote databases where it is scattered, upon a simple online authorization by Bob himself. Filling in the form is simpler, quicker and less prone to mistakes.

Bob is the owner of the data. He applies for the authorization grant (arrows 1 and 2) so that the web application he is connected to can directly access his restricted access resources stored in different databases (the Employer Registry, the Estate Registry, the Medical Expenses Registry shown in the figure). He then fills in the form adding only data that could not be obtained via online third parties.

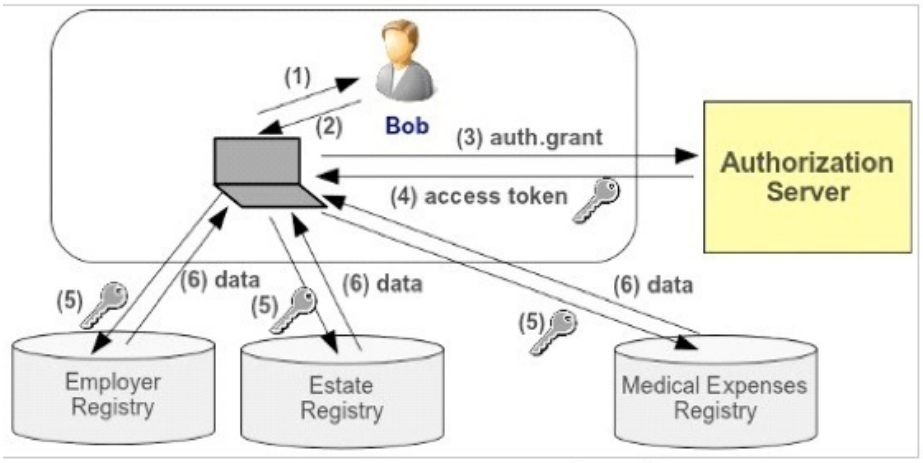


Figure 8

Tax payment via OAuth 2.0

Challenges. Even in this case, we can identify entities that might need to access data without being the owners. Consider for instance the Judiciary or the Inland Revenue Office in case of lawsuits related to tax evasion. Moreover, the release of aggregated data is desirable for statistical purposes in this use case too.

In this scenario there is also another subtle point. The same resource, for instance a receipt of payment, involves two people. For example the receipt for plumbing work carried out in an apartment involves the plumber who did the work and received the money, and the owner of the apartment who paid for the work. For the former the resource represents income, while for the latter it is an expense; there is only one resource but it has two distinct resource owners. How to deal with resources of this type is not explicitly specified by OAuth 2.0 (see Section 6).

5. A working prototype

Currently we have a working implementation of the OAuth 2.0 authorization server, that we have called *OAuthwo*. Due to the high level nature of the OAuth 2.0 framework, many features are left to the implementor's choice. In this section we describe the main features of OAuthwo, especially those ones regarding the mitigation of some security issues and how it handles multiple resource servers with possibly multiple credentials of the same user.

OAuthwo is free software, written in PHP as a Zend (framework.zend.com) module and available for download at GitHub (github.com/andou/oauthwo zend/). Although OAuthwo is written in PHP, we will not bind its description to any specific implementation language.

5.1. Data elements

In OAuthwo the authorization server AS must keep some information about resource owners, clients and resource servers.

5.1.1 Information about clients

Information for a client C retained by the AS are:

- C_{ID} , a unique ID for C
- C_{secret} , a shared secret between AS and C, present if C is a confidential client and used for client authentication
- $C_{profile}$, the profile of C, with possible values web, user-agent or native ([IETF, 2012], Section 2.1)
- $C_{redirectionURI}$, the client redirection URI as specified in Section 3
- C_{name} , a human-readable name for C.

These data are provided by each client during the preliminary client registration phase.

In OAuthwo, public clients as well as confidential clients are required to register their redirection URIs. Each client must register one and only one complete redirection URI, preventing the "Authorization Code Redirection URI Manipulation" and mitigating the "Client Impersonation" and the "Open Redirectors" problems described in [IETF, 2012a].

5.1.2 Information about resource servers

Information for a resource server RS retained by the AS are:

- RS_{ID} , a unique identifier for RS
- RS_{secret} , a secret key shared between AS and RS, to be used for encrypting access tokens; see Section 5.2.2
- $RS_{idMethod}$, the way by which RS identifies users within its realm, e.g. by email, social security number, phone number or others; see Section 5.1.3
- $RS_{endpointURI}$, the URI from which RS's data could be accessed.

Furthermore, AS retains information about allowed scopes on each RS. All these information are gathered during trust establishment between each resource server and the authorization server.

5.1.3 Information about resource owners

In order to deliver user-specific protected resources, a RS employs one out of several available user identification methods (e.g. email, social security number, phone number, etc.). Thus the same user can have different identifiers at different RSs and even another different one at the AS. For example a resource server RS_1 could identify users by email, RS_2 by social security number and RS_3 could use email too, but with different addresses. AS must keep track of the identification and authentication scheme in use at each RS. More precisely, the AS must build and maintain an *identity equivalence class* for each resource owner. This is necessary because, for instance, when the AS issues an access token T to be used at an RS to retrieve information on the resource owner U, T must contain an identification key that is valid at the RS, so that RS could properly identify U.

In order to build the identity equivalence class, during the user registration phase the user is prompted with a list of RSs known by the AS and is asked for an identification key for each of these. The AS should then verify that the specified identification keys do actually belong to the user. Some of the identification keys are easily verifiable online (e.g. email, social network nickname), others may require a third party digital certification (e.g. social security number).

An OAuth 2.0 authorization server only needs to abstractly identify and authenticate resource owners, regardless of any authentication mechanism. Consistently OAuthwo is not bound to any specific login mechanism, to foster its use as an authorization component or plugin inside other software applications with their own login mechanisms.

5.2. Access Tokens

An access token represents the authorization originally given by a resource owner for accessing a certain set of resources, subject to given constraints (scope and lifetime).

5.2.1 Identifier vs self-contained token

While the purpose of access tokens is well defined within the OAuth specification, their format and structure are not. Yet, two main categories can be distinguished. The access token may be an opaque identifier, used by the RS to retrieve the authorization information; or it may self-contain such information in a verifiable form.

The former kind of token does not require resource servers to implement any cryptography and allows for timely revocation of authorization. On the other hand this solution implies a further communication between token receiver and issuer to retrieve authorization information and it is thus less scalable.

The latter kind of token, however, requires resource servers to be able to interpret cryptographic messages and implies the use of short-living access tokens due to impossibility to revoke a token once issued. As the access information is self-contained, this solution requires less communication and is therefore more scalable.

OAuthwo uses self-contained bearer tokens [IETF, 2012b] whose format and content is described below.

5.2.2 Access token format

When a client seeks authorization for protected resources stored in multiple resource servers RS_1 , ..., RS_n , the OAuthwo AS issues a composite access token T made of several chunks T_1 , ..., T_n . The generic chunk T_i is to be spent at resource server RS_i .

Access tokens are served to clients through the access token parameter in the successful response (see [IETF, 2012], section 5.1). In OAuthwo this parameter is a base 64 encoding [IETF, 2006] of a JSON [IETF, 2006a] structure which serializes an associative array where the keys are the URIs of resource servers RS_1 , ..., RS_n and the values are the chunks T_1 , ..., T_n . Each T_i consists of a triple

$< U_i , S_i , P_i >$. S_i , the scope, instructs the client on what actions are allowed on what data at resource server RS_i , using U_i as endpoint and P_i as access token for that particular resource server. P_i is thus a monolithic self-contained access token for a single resource server. Being self-contained, P_i carries all the information needed by the resource server RS_i in order to provide the requested resources, namely:

- $P_{(i,userReference)}$, a unique identifier of the user (resource owner) recognized at RS_i
- $P_{(i,scopes)}$, a space separated list of the scopes allowed by the token
- $P_{(i,validity)}$, the token lifetime.

These information are serialized as claims in a JSON Web Token [OAuth WG, 2012]. To prevent token manifacture and modification [IETF, 2012b], P_i is digitally signed by the AS with a JSON Web Signature [JOSE WG, 2012].

The user identifier $P_{(i,userReference)}$ valid at RS_i is obtained by the AS using the identity equivalence class it maintains (Section 5.1.3). Such identifier must be considered as a sensitive information; to prevent disclosing it to the client [IETF, 2012b], each serialized and signed P_i is finally encrypted with the secret key shared between the AS and RS_i (Section 5.1.2) using JSON Web Encryption [JOSE, 2012a].

5.3. Authorization Codes and Refresh Tokens

In OAuthwo, both authorization codes and refresh tokens are generated by means of random numbers. Such a random number is associated with the client for which it was issued and the resource owner for which the grant was requested, along with the set of yielded scopes and a timestamp used to check time validity. In OAuthwo the generation of this random number relies on a pseudo-random generation of 20 bytes[1], which ensures the generated number to be resonably random and hard to guess by a potential attacker. The probability of an attacker guessing the generated authorization code is 2^{-160} , as required in [IETF, 2012].

[1] *OpenSSL RAND_pseudo_byte(). Description and synopsis of this cryptographic function can be retrieved from http://www.openssl.org/docs/crypto/RAND bytes.html*

Each issued authorization code is short lived (10 minutes) and single use, while each issued refresh token is long lived (5 days) and single use, which means that a new refresh token is issued with every access token refresh response. The previous refresh token is then discarded.

6. Conclusions and further work

In this paper we have made a case for releasing personal data in the Open Data domain by following a smart disclosure approach, and have proposed the adoption of the OAuth 2.0 authorization framework to this end. OAuth 2.0, if properly implemented, guarantees access to selected personal data upon authorization of the individual data owner. We have presented OAuthwo, a free and modular PHP implementation of OAuth 2.0 based on the current IETF draft specification. OAuthwo will allow us and everybody else to play with the OAuth 2.0 technology and implement proof-of-concept smart disclosure systems supporting a number of interesting use cases that leverage personal data securely disclosed as Open Data.

A prof of concept in tax payment domain (Section 4.2) has been developed as a master thesis [Tolstenco, 2013]. The prototype uses OAuthwo to build a web application (https://dione.disi.unige.it) supporting a simplified yet representative example of tax return, namely, the italian *modello 730* (devoid of its graphical skin for simplicity) in the case of a single income plus medical expenses and the property of a single residential apartment, along with anagraphic personal data.

One of our current tasks is to extend the set of use cases. This will allows us to strengthen our claim about OAuth 2.0 being a key technology for smart disclosure. But more important, this will allow us to identify significant scenarios that put the current OAuth 2.0 framework at challenge, ultimately suggesting new features to be added to OAuth 2.0. In Section 4 we already pointed out some of these challenges.

One challenge has to do with single resources owned by multiple users. OAuth 2.0 seems at odds with such a setting. As illustrated in the medical use case, the life of an unconscious patient might depend upon the rescuing physicians being able to access that patient's online medical record, yet the resource owner (the patient) would be unable to provide consent. One possible solution is to allow multiple owners, and then asynchronously warn each of the owners when one of them is granting authorization for their common resources. In some cases the set of owners is static (e.g. the plumber and the householder in the receipt for plumbing work, the citizen and the Judiciary in all tax-related data), but in other cases it is not (e.g. the unconscious patient after an accident and the physicians involved in the rescue). This whole point needs investigation.

Another challenge is related to anonymized datasets. OAuth 2.0 lacks mechanisms for defining what is an anonymized version of a resource. It is unclear when and how a resource owner could authorize access to anonymized resources, and it is also unclear whether or not the resource owner should be able to exert ownership at all over anonymized versions of their data. In any case, authorization should be given ahead of time since the aggregation of

anonymized data might take place at any time in the future and even repeatedly. The aggregation procedure itself might need special primitives for bulk verification of authorization grants, since verification on an individual basis would be highly inefficient when extracting a large anonymized dataset.

Last but not least, the whole OAuth 2.0 mechanism seems unable to support any form of delegation of access rights to third parties. Such delegation is very useful in a number of cases where the data owner needs to grant access right ahead of time, so that a third party can subsequently and asynchronously run an application that makes use of personal data of that owner. Even in a simple use case like tax payment there is a obvious need for the user to be able to delegate a professional for doing the entire job of the annual tax return. It would be impractical for the professional to require the user be present during the procedure just for giving consent to the data accesses being performed. A solution might be found in an evolution of the OAuth 2.0 technology, namely, User-Managed Access (UMA) [UMA WG, 2012]. We are currently investigating on this very point.

References

Ciaccio and Ribaudo, 2012 G. Ciaccio and M. Ribaudo. Open Data for the Masses: Unleashing Personal Data into the Wild. In 8th International Conference on Web Information Systems and Technologies (WEBIST 2012), pages 201-206. SciTePress, June 2012.

Falcão-Reis and Correia, 2010 F. Falcão-Reis and M. E. Correia. Patient Empowerment by the Means of Citizen-managed Electronic Health Records, pages 214-228. Number 156 in Studies in Health Technology and Informatics. IOSPress, June 2010.

Hammer-Lahav, 2010 E. Hammer-Lahav. Introducing OAuth 2.0. hueniverse.com/2010/05/introducing-oauth-2-0/, 2010.

Hoffman *et al.*, 2012 W. Hoffman *et al.* Rethinking Personal Data: Strengthening Trust. Report, World Economic Forum, May 2012. Retrieved Dec. 2013 from www.weforum.org/reports/rethinking-personal-data-strengthening-trust.

Howard, 2012 A. Howard. Data for the Public Good. Strata: Making Data Work. O'Reilly, 2012.

IETF, 2006 IETF. The Base16, Base32, and Base64 Data Encodings. www.ietf.org/rfc/rfc4648, Oct. 2006.

IETF, 2006a IETF. The application/json Media Type for JavaScript Object Notation (JSON). www.ietf.org/rfc/rfc4627, July 2006.

IETF, 2007 IETF. The Atom Publishing Protocol. tools.ietf.org/html/rfc5023, 2007.

IETF, 2010 IETF. The OAuth 1.0 Protocol. tools.ietf.org/html/rfc5849, 2010.

IETF, 2012 IETF. The OAuth 2.0 Authorization Framework. tools.ietf.org/html/draft-ietf-oauth-v2-31, July 2012

IETF, 2012a IETF. OAuth 2.0 Threat Model and Security Considerations. tools.ietf.org/html/draft-ietf-oauth-v2-threatmodel-07, Aug. 2012.

IETF, 2012b IETF. The OAuth 2.0 Authorization Framework: Bearer Token Usage. tools.ietf.org/html/draft-ietf-oauth-v2-bearer-23, Aug. 2012.

JOSE WG, 2012 JOSE Working Group. JSON Web Signature (JWS). tools.ietf.org/html/draft-ietf-jose-json-web-signature-05, July 2012.

JOSE WG, 2012a JOSE Working Group. JSON Web Encryption (JWE). tools.ietf.org/html/draft-ietf-jose-json-web-encryption-05, July 2012.

OAuth WG, 2012 OAuth Working Group. JSON Web Token (JWT). tools.ietf.org/html/draft-ietf-oauth-json-web-token-06, July 2012.

Obama, 2009 B. Obama. Transparency and Open Government, 2009. Retrieved Dec. 2013 from www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment.

Recordon and Reed, 2006 D. Recordon and D. Reed. Openid 2.0: a platform for user-centric identity management. In Proceedings of the second ACM workshop on Digital identity management, DIM '06, pages 11-16, New York, NY, USA, 2006. ACM.

Shadbolt, 2013 N. Shadbolt. Midata: Towards a Personal Information Revolution, pages 202-224. IOSPress, June 2013.

Schwab *et al.*, 2013 K. Schwab *et al.* Personal Data Ownership: The Emergence of a New Asset Class. Report, World Economic Forum, Jan. 2011. Retrieved Dec. 2013 from www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.

Tolstenco, 2013 A. Tolstenco. Architetture web per "smart disclosure": il caso della dichiarazione dei redditi. Master's thesis (in italian), DIBRIS, Università di Genova, July 2013.

UK Government, 2012 Open Data White Paper: Unleashing the Potential. Technical report, Her Majesty Government, UK, June 2012. Retrieved Dec. 2013 from www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential.

UMA WG, 2012 UMA Work Group. User-Managed Access (UMA) Profile of OAuth 2.0. http://tools.ietf.org/html/draft-hardjono-oauth-umacore-06, Dec. 2012.

Biographies

Giuseppe Ciaccio obtained a PhD in Informatics at the University of Genoa, Italy. In 2001 he joined the same university as assistant professor. He is currently working in the field of Open Data and web technologies for the public administration, with emphasis on online authorization protocols. He was formerly interested in online anonymity in decentralized systems.
email: giuseppe.ciaccio@unige.it

Antonio Pastorino got a Degree (Laurea Specialistica) *cum laude* in Informatics at the Univerity of Genoa, Italy, with a thesis on online services based on social networks. He was then appointed a grant for doing research on online authorization for smart disclosure. Currently Antonio is working as web developer at Bitmama, Italy.
email: antonio.pastorino@gmail.com

Marina Ribaudo obtained a PhD in Informatics at the University of Turin, Italy. In 2001 she joined the University of Genova as an Associate Professor. She is currently interested in web technologies, with an emphasis on services and applications for citizens and students. She has been involved in research activities on e-learning, with a special emphasis on users accessibility and, more recently, she has started working in the field of Open Data, with the aim of designing/developing useful yet privacy-preserving services for individuals.
email: marina.ribaudo@unige.it

Sampling Issues and Management Solutions in Internet-Based Market Researches

M. V. Ciasullo, G. Festa

Abstract. *In the broad context of internet-based market research, this paper focuses on research carried out on net users. An analysis of the advantages, but also disadvantages, related to online market research shows that a crucial problem is the 'vagueness' of the representativeness of sample surveys, even affecting the credibility of the online survey. The authors have developed a theoretical / practical framework that integrates contributions from marketing, statistics and information technology, following a logical-methodological path oriented towards reducing uncertainty in online surveys. The model utilizes helpful considerations from inferential statistics and proposes solutions for managing internet samples as virtual communities. The conceptual paper aims to contribute to an advancement in the field of online market research, shifting the focus on the validity / reliability of the investigation from the mere software or statistical tool to a wider analytical internet marketing strategy.*

Keywords: e-research, analytical internet marketing, online sampling.

1. Introduction

The *customer based* approach [Valdani and Busacca, 2000], fundamental for the success and survival of any organization, has become a prevalent trait of corporate behavior nowadays. In this respect needs analyses are essential if we are to grasp a customer's way of thinking and the means by which she/he chooses or otherwise, to buy. At the same time, an ever more globalized economy, characterized by the capacity to rapidly transfer information, delineates not only more attentive and complex customer profiles, but also new styles of consumption based on the capacity of the goods/products to reflect the culture, taste and style of each individual.

The ongoing changes associated to growing markets and their related innovative dynamics in terms of purchasing, impose on businesses constant commitment in the analysis and interpretation of the above phenomena. It is clear therefore, that in the context of studies on marketing management, particular attention has been addressed to the analytical aspect of marketing, for the purpose of analyzing the structure of the market and trends in demand in order to identify the most significant factors that impact on consumer behavior.

Market research, generally intended as systematic and objective analyses for the purpose of gathering relevant information in order to elaborate and put in place interaction strategies with clients and other members of the corporate value system, observe markets and delineate scenarios. Therefore, it is evident how such strategies constitute basic drivers of corporate action, where competitiveness is based on the capacity to conquer and foster ready, exigent and innovative demand.

Within such a complex context, the value of high tech prevalently *internet-based* and capable of contributing significantly to social research and specifically market research finds its place. The Internet is a suitable 'natural' tool for gathering targeted, in-depth and appropriate information relative to demand [Kiang et al., 2000], both in a market research context and prior to that, through pertinent operations of *marketing intelligence* [Kotler, 2001]. In such a perspective, an important element of the web is underlined i.e. its capacity to render communicational processes substantially independent from the limits of time and space. Evidently, this attribute is particularly valuable with respect to research online in that it enables the overcoming of a series of barriers that might exist between researchers and the parties taking part in the survey.

Furthermore, as evidenced in the literature [Cantone et al., 2006; Vescovi, 2007], new information and communication, in particular based on open, universal standards, take on a fundamental role in improving the management of the intra and inter-business relational system and of the underlying processes of value creation, facilitating at the same time, the phases of approach to international culture and markets. Consequently, the successful outcome of market research depends to a great extent on the constant updating combined inevitably with evolving *internet-based* marketing information and, in the context of our study, the strategies of analytical internet marketing in the broadest sense.

2 Internet-based market research categories, taxonomies and critical issues

The combination between marketing and statistics is traditionally at the basis of the skills system regulating market research [Bassi, 2008]. Marketing in this context, uses inferential statistics prevalently in the compiling of the sample for the survey stage in order to extend sample findings to the population. Over time, these two skills have branched into a third discipline i.e. information technology which by virtue of the benefits involved in terms of time and costs, has contributed remarkably, to the two skills mentioned above, to the extent that it can be said that the market research system is governed by the sum of these three disciplines (marketing, statistics and information technology).

The Internet on the other hand, has not only engendered a new series of information technology tools for researchers, but by virtue of its success in 'social' terms has ended by offering the market research field, a new strand of investigation. Thus, it is now possible to distinguish between market research carried out *by means of* the Internet and market research directly in contact with

web users *on* the Internet: the two categories making up the extended family of *internet-based marketing research.*

In the context of the present study, our focus will not be on the former category (*by means of)* primarily in the sense that we consider the Internet exclusively a tool and in the second place because it has already been analyzed by the Authors in a previous work [Festa, 2001b]. In this study we intend to concentrate on the second category (*on the net*), by identifying the main critical issues of reference (i.e. the representativeness of the sample) and proposing where possible, a theoretical analysis that takes into account the integrated contribution of marketing, statistics and IT.

We consider that devising appropriate methodology relative to the issue of *internet-based* market research has to start from a traditional market research approach. This is necessary in order to verify whether the relative 'toolbox' of techniques (i) is suitable for internet use, (ii) applicable if adapted, using IT mode and (iii) applicable in an innovative key, exploiting the enormous potential of the web. Two factors in epistemological terms underpin this point of view [Festa, 2001a]: the 'technological' characteristics of the web platform and the 'social' characteristics of the web users. For the sake of brevity they are analyzed by reviewing the advantages and disadvantages linked to market research on line, without presuming to examine the issue in any depth.

As concerns some of the main advantages, prevalently in terms of efficiency, economy and timesaving (where current IT hardware and software enable elaborations to be made in real time) and reduced costs (not in an absolute sense but if compared to the normal costs of traditional marketing research) the Internet becomes in a certain sense, consecrated as a more efficient channel for data collecting [Weible and Wallace, 1998]. In any event, even more significant qualitative advantages of the Internet result. The potential of the web however, should not be considered merely in terms of stored data but also and above all, with respect to the manner of collecting, which offers the interviewee a more pleasant experience (thanks to multimedia applications), greater independence in terms of time and place of data detection. Another factor is the convenience and accuracy of detection for both parties together with more independence from the context (personal and/or situational for categories of persons such as the disabled, the ill, the secluded or in areas such as war zones or zones where epidemics are rife etc.), thus breaking down psychological and socio-cultural barriers (pivoting on the sensation or mere virtual presence of the other parties) [Schmitz, 2004; Wright, 2005].

The overall sense of convenience and comfort, furthermore, generally produces more satisfying results especially as concerns qualitative research which seems quite a propitious and interesting investigating scenario for *internet-based* market research [Andreani and Conchon, 2001b]. In this respect, new competences are required of researchers e.g. the skills required of a moderator of a group online. Qualitative research aspires to attaining in-depth understanding of consumer behavior by analyzing the aspirations and sensations associated for example, to a category, brand or image [Mariampolski, 2001; Marbach, 2010]. Besides statistic representation

(fundamental in quantitative research), qualitative techniques have the advantage of applicability even in cases where information is difficult to detect, complex to analyze or of problematic interpretation [Andreani, 1998; Tissier-Desbordes, 1998; Andreani and Conchon, 2001a] as they offer the interviewees greater freedom of expression [Prandelli and Verona, 2006]. Thanks to the Internet however, the interviewee can take over 'command' of the interview, thus preventing the interviewer from 'manipulating' the interview in any way. If this were not the case and should the interviewee become aware, a click would be sufficient to log out or even remain online fooling the interviewer with trick responses. Consequently, technological, methodological and above all socio-relational skills in line with the Internet environment will be required of the online qualitative research designer and relative interviewer. For this reason too in our opinion, the most interesting market research on the Internet is qualitative research, precisely because it enables the extrapolating of information which in other contexts an interviewee would be hard put to confess (despite the fact that non-verbal language might easily give him away). It goes without saying however, that many disadvantages can be linked to market research online, fully compensated for nonetheless by the above mentioned advantages. For instance, the extent to which a user is familiar with the 'language' of the keyboard or mouse (i.e. monitoring a chat line) or the loss of information on the context (i.e. a questionnaire online to which one responds either from one's office or from home, being 'bound' naturally enough to the rules of behavior regulating different environments) and not least, to the impossibility of grasping the indications of non verbal language (or even those of language: different way of expression using electronic writing compared to using traditional methods or even voice mode, register or tone of voice etc.). In any event, as emerges from many studies [Howard et al., 2001; Andrews et al., 2003; McDonald e Stewart, 2003; Di Fraia, 2004; Wright, 2005] the main defect of internet based research is linked to the 'vagueness' of the representativeness of the survey sample which can even jeopardize the credibility of the survey itself. As the literature confirms this weakness, the main aim of our conceptual paper is to develop a theoretical / practical framework that can better delineate the boundaries of the online sample representativeness issue, proposing specific guidelines for governing the same deriving from complex analytical internet marketing strategies. In this respect, the implications of the particular insights we offer could be the basis for further research on the subject.

3. Elements characterizing online sample surveys

Analyzing the problems relative to online sample surveys, with a vision based on 'gaps', not all the reference population are necessarily regular users of the Internet (*netizen*); not all can necessarily be *identified / contacted* by the researcher for the survey; even among those contacted, not all of them would necessarily decide to respond (*respondent*); and among these not all would necessarily want to respond honestly (*authentic*) or with trustworthiness (*reliable*), the reasons being the spatial-temporal gap/distance in terms of the survey and sense of 'irresponsibility' socially circulating online. The number of web surfers for instance using free services (e-mail, video sharing, freeware,

etc.) who when requested by the provider to register for commercial purposes, give unreliable or even false personal data in order to overcome the obstacle of registration and to use the online services is huge. Such obstacles tend to create an effective *funnel* which is illustrated in Figure 1.

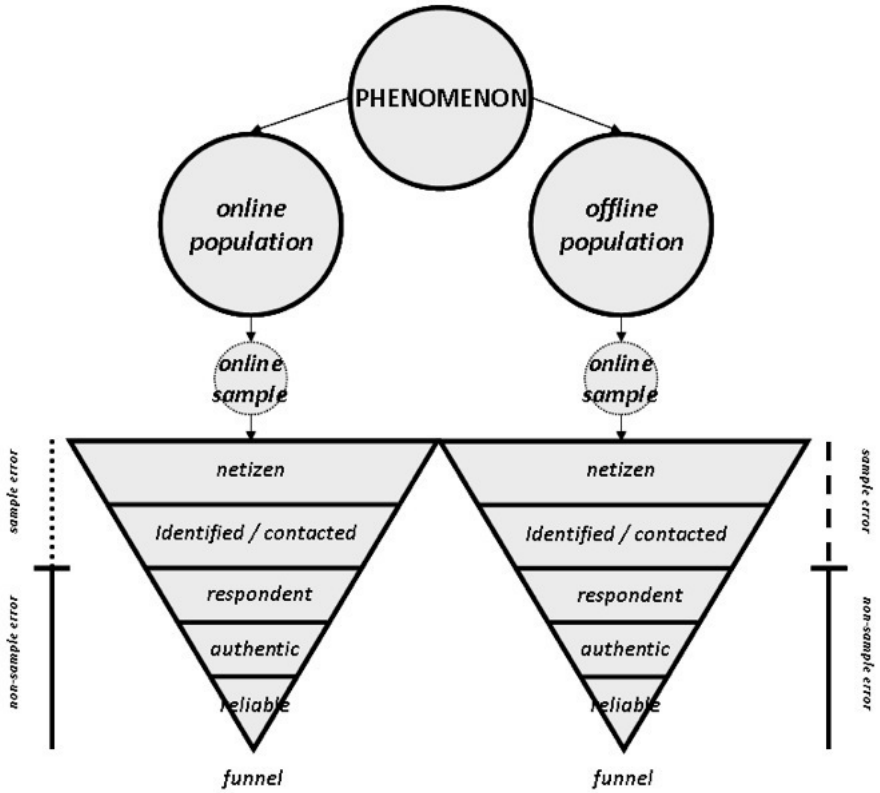


Figure 1
Common sample and non-sample errors in online surveys

The figure clearly shows that internet based market research involving both the online or offline populations, presents identical problems in terms of non sample errors (i.e. errors on the part of interviewers, interviewees, researchers etc.), where *respondent, authentic* and *reliable* have to be governed but only 'from a bottom up perspective' (i.e. the group of individuals that can be identified and contacted on the Internet). Previously, the question of representativeness was posed underpinning the rationale of our study or in other words, the potential (almost certain) gap between the offline population and the sample online. This contributes to rendering the sample error even more complex, as shown by the bold line in Figure 1.

Sample errors strictly speaking are in the main, identical for both samples (this observation constitutes one of the presuppositions for our methodological

approach), as it derives from the application of the central limit theorem, according to which should the number involved in the sample increase, representativeness also increases (or, more precisely, with the increase in the number involved in the sample, distribution of the sample tends towards standard thus enabling reliable predictions once errors and representativeness have been defined) [Natale, 2004]. Nonetheless, it is clear that a sample online is the 'legacy' of a population online, while in the case of a sample online used for representing a population offline, such legacy results inevitably, spurious.

In our view, this appears to be a major issue of online research as not only is it impossible to classify the devices used in *internet-based market research*, as online tools are in constant evolution (technological and social) above all their combination, but from the point of view of marketing, such an effort could in effect, result neither decisive. On the contrary, representation of the sampling problem and relative solution could be devised following a conceptual scheme which starts from the logical-methodological framework of marketing research and proceeds by means of a synoptic vision of the tools and – above all – techniques available for use.

In this context, following a strictly classical approach to marketing research taking into account the flexibility needed to meet various professional scenarios, it is fundamental to start from a solid conceptual basis. As a result, a suitable methodological approach to *internet-based* market research set in the framework of the traditional process of market research, would entail eight stages [Barile and Metallo, 2002]:

1. defining the marketing issue;
2. identifying the aim of the research;
3. formulating the aims of the survey;
4. designing intervention (in terms of quality, times and costs);
5. economic evaluation of scheduled actions;
6. data collecting (primary and/or secondary, from internal and/or external sources);
7. data processing and analyzing (with suggestions for application);
8. final reporting (even for accreditation).

Obviously such an approach cannot be considered mandatory but the advantages in schematic terms are evident. Proceeding with the structuring of the approach, the most delicate phase concerns 'designing intervention'. This phase can be divided into eight sub-phases [*ibid.*]:

1. formulating operative research objectives;
2. planning the survey in practical terms;
3. designing the sample;
4. selecting detecting tools;
5. managing human resources engaged in the survey;
6. scheduling time scales for activities;

7. planning results tabulating and codifying;
8. cost estimating.

This kind of division is typical of the logical-methodological process of market research. The first sub category for instance, (*formulating operative research objectives*) and the last (*cost estimating*) of the 'designing intervention' phase are linked to the previous phase (*formulating the aims of the survey*) and to the subsequent phase (*economic evaluation of scheduled actions*).

Diverse conceptual 'zones' characterized by identical problems both in the *earth-based* and *internet-based* research (i.e. analysis of problematic issues in marketing) emerge clearly from the above. Other 'zones' on the contrary, regard the tools context (i.e. questionnaires online as opposed to paper documents), obviously more advantageous for market research online. Finally a 'zone' characterized by specific issues regards internet-based market research, the most relevant of which is undoubtedly representativeness (as mentioned above and confirmed from a methodological perspective).

Observation (cf. the *funnel*) and methodology (cf. the *literature review*) show that the main issue relative to *internet-based* research consists in the limited representativeness of the survey. To overcome such limit, in our view, it is considered fundamental the support of marketing to inferential statistics, the opposite of what usually occurs in traditional market research where the contribution of statistics is functional to the analysis and eventually to the solution of the marketing problem [Molteni and Troilo, 2003].

Math and statistics which constitute the disciplines contributing most to the study of sample representativeness, cannot but derive from the GIGO rule ('garbage in, garbage out') [Seglin, 1994] and in a certain sense, show no interest for the quality of the starting data. On the contrary, the researcher well aware of the potential of the Internet has to examine this issue in depth in an attempt to overcome contextual limits. Currently, the core issue is the representativeness of the online sample with respect to phenomena (also) offline. While it is evident that survey of a particular phenomenon online could pose serious problems of representativeness if carried out offline, or in other words, involving interviewees who are not normally users of the internet services under analysis.

Paradoxically, therefore, we show that online research, although deeply affected by the 'funnel', cannot be considered unreliable beforehand. It would be epistemologically wrong for example to investigate a user by means of traditional techniques even only in terms of a *concept test*, relative to the experience of browsing in a multimedia virtual store which could even be 3D. Some proposals are developed in this perspective in the concluding part of the paper where the methodology of collecting 'authentic' and 'reliable' data from the *internet sample* conceived as virtual community [Hagel and Armstrong, 1997], is discussed relative to the problem of marketing in question.

4. Potential solutions for the problem of online representativeness

In this section the main characteristics of an adequately represented sample is discussed and accounted for statistically, with, in coherence with our study, the main focus on managerial aspects. Generally, the representativeness of the sample (*n*) with respect to the population (*N*) is delivered, together with the correct calculation of the sample number, by means of probabilistic sampling, in the first place by mere random sampling [Molteni and Troilo, 2003]. Only random sampling (i.e. the odds known and the same for each member of the sample) can guarantee the absence of selection errors (common on the contrary in non probabilistic sampling, i.e. by convenience, opinion, quota). Eliminating selection errors from random sampling together with an adequate sample numbering should ensure the absence of distortion.

In non-probabilistic surveys, furthermore, researchers are not unaware of potential distortions but on the contrary, use them to their own advantage. They consider that a deliberate error of selection enables a more adequate contextual representativeness although this obviously cannot be known beforehand. From random extraction a more representative non-probabilistic sample could emerge, because random with respect to the same (identical) non-probabilistic sample itself. The paradox lies in the fact that the random sample finds its strength not in the result but in the process: furthermore, even a random sample (precisely because it is random) can produce 'exceptional' outcomes.

The sample as mentioned previously, is based on the central limit theorem according to which the increase in number of the sample renders the sample distribution closer to standard thus allowing calculation of the required number of the sample to support a given margin of error (Es_x) and a given trust/ confidence interval (z). The formula set out below illustrates how (see Table 1) the sample number is calculated (a population in excess of 100.000 elements is defined infinite by statistic convention, a phenomenon is binomial if represented in only two modes, continuous in other scenarios) [Barile and Metallo, 2002].

	Finite population	Infinite population
Continuous phenomenon	$n = N z^2 s^2 / [(N - 1) E^2 s_x + z^2 s^2]$	$n = (z s / Es_x)^2$
Binomial phenomenon	$n = N z^2 p (1 - p) / [(N - 1) E^2 s_x + z^2 p (1 - p)]$	$n = p (1 - p) (z / Es_x)^2$

Table 1
Calculating sample number (n) compared to number of population (N)

Estimated accuracy therefore, depends not so much on the fraction of the sample (f = n / N) but rather on the number of the sample by virtue of the strength of the central limit theorem. Such considerations, crucial in sampling theory, enable us to state that the issue of representativeness on the Internet to a certain extent, is inspired more by the practical aspect of the survey rather than

by the scientific aspect of the methodology. In other words, the first tier of the funnel can be ignored, i.e. the non-coinciding of the population under observation with that of the Internet users as long as a sufficiently numerous online sample (updated statistically in terms of mortality rate) can be made up.

In practice, the most relevant issue concerning online representativeness refers to the number of elements of the sample concurring to form an adequately calculated sample number. The fact that such elements of the offline population also apply to online users represents an additional as opposed to penalizing characteristic. On the contrary, recruiting exclusively active online users for a sample of offline population could to some extent be assimilated to non probabilistic sampling [Andreani and Conchon, 2001a], deriving from opinion or convenience (or even by quota: e.g. a survey measuring customer satisfaction of a bank having both traditional branches and internet based services).

In non probabilistic sampling, the underpinning concept of standard error in the calculating of sample number has absolutely no value. Qualitative market research tends prevalently to make use of non probabilistic sampling, which also for this reason seem to be the most practicable online market research survey, driven by definition, not by descriptive but merely exploratory aims. Given that sample representativeness should be governed by the number of internet users surveyed as opposed to the extent of internet users compared to the population, it should not be necessary therefore to extract an online sample from an online population corresponding in part or as a whole to the offline population seeing as the positive confirmation of the characteristics of the population in the survey would be sufficient provided it were adequate in number. The procedure would therefore not be that of extraction (from top to bottom) but of abstraction (from bottom to top).

It is fundamental, therefore, to mitigate data distortions linked to the survey by means of credentials in terms of transparency: characteristics of the population, type of sampling, number of sample, percentage of errors acceptable and estimated reliability. All the other information considered of impact for outcomes should be added to the statistic indications (in terms of research integrations): considerations on the training of the interviewers, on the formulating of the questions, on the classifying of responses and so on, addressing attention to putting in place correctly systems for the eliminating or mitigating of non sampling online error, with particular reference to the 'funnel' phases relative to *respondent*, *authentic* and *reliable* categories.

Such credentials represent not only the 'litmus test' for interpreting the findings of the survey but also and above all, the key for evaluating information obtained. Consequently, the credentials themselves become information for the decision making process. The natural statistic distortion of the sampling, due to the effective distance or gap between population and sample cannot be questioned: in the same way that it stands for *earth-based* research, it certainly stands for *internet-based* (the socio-economic nature is constantly changing and internet users are people). Consequently, the issue of representativeness of the internet population with respect to offline events constitutes the main focus of this study. Our logical-methodological reflections outlined above would seem

to suggest a solution (taking into account characteristics and context involved). By means of a global strategy of analytical internet marketing, aimed at constructing, monitoring and governing the internet samples as if they were virtual communities, would render it possible to discuss useful devices for guaranteeing beforehand the 'clarity' of incoming data, resolving in this way, the further limits imposed (represented by the *funnel*) utilizing our proposed solutions outlined below.

5. Guidelines for managing an internet sample

On the basis of the theories discussed in our study, some fundamental methodological (*strategic*) and applicative (*operative)* characteristics could be delineated not limited merely to appropriate planning but also to the efficient managing of the internet sample for online research. In this sense the idea of a sort of virtual community appears particularly appealing the construction of which should be based on strategic internet marketing as opposed to the perspective of market research in the strictest sense and underpinned (even when analytical) on the binomial 'visibility' and 'attraction' [Metallo and Festa, 2003].

In other words, paradoxically, the focus should not be on the issue of the validity and reliability of online market research as such, but rather that of the number of population online, from which to extract the sample base and subsequently, the sample for the survey. The latter if adequate numerically speaking might be sufficient in terms of representativeness at least from an empirical point of view: from the methodological perspective on the contrary, the above approach might not be practicable, not being in substance possible to proceed on the Internet to a random sampling which is certainly representative of an offline population [Schmitz, 2004]. In this direction, the *funnel* requires constant adjusting and mitigating. In other words, insistence should be not only on the statistic planning/ design of the sample but also and above all, on the correct management of the sample, whereby a series of devices is activated to constantly authenticate interviewee data for the purpose of guaranteeing identity and authenticity in data detection.

However, on the other hand, such barriers could at the same time provoke the opposite effect whereby the interviewee, irritated by being subjected to constant authentication processes, becomes less spontaneous and natural in responding and even tires of being interviewed altogether. To date in any event, this seems to be the only path to follow to ascertain the reliability of the survey online. Various authentication steps are necessary and can be illustrated as follows:

- *registration*: the user has to register with the virtual community (the researcher consequently, right from the start of the survey, has to imagine the sample online of any research online as a virtual community, thus deriving a greater sense of 'belonging' on the part of the interviewee in terms of the positive outcome of the research), compiling a registration form, on the basis of which the researcher can compile a series of preliminary data in order to classify the user in conformity with the principal criteria of segmentation of the community. Such strata are neither static nor dynamic but merely 'virtual' (in the same way that in a database, the *queries*

can be considered virtual tables), exploiting the capacity of the IT system underpinning the virtual community to extract, and filter the most useful segments in real time (OLAP applications). On registration, the system assigns the user a *username* and *password*, giving them authenticated, authorized and responsible access to the system;

- *accreditation*: during the process of data detection, users might be requested not only to authenticate their identity (using the account) but could also be subject to other authorizations through for example a *captcha* (on the access page of the survey) and/or a *password* (in the e-mail of 'invitation to take part in the survey' sent to the user). *Captcha* is the acronym for 'completely automated public Turing test to tell computers and humans apart' to confirm the use of the internet service on the part of an individual and not a machine. The user is required to digit a set number of characters displayed on the page into an empty box for verification purposes. The characters seem distorted but an individual can read them easily (a further aim is to raise the level of interviewee attention). Thus the system guarantees against the user (or the IT system underpinning the virtual community) being unwittingly infected by *malware*, which infecting the platform could create conditions for unreliable responses;
- *verification*: to make sure it is effectively the user involved in the survey, it is vital to oblige the same to provide adequate responses to specific questions that only the user can give. Such questions should be presented in random mode in order to 'oblige' the particular user to interact with the system. The process of user authentication represents a core element of the IT security infrastructure and knowledge based techniques are currently the most frequently used to authenticate user identity [Ellison et al., 2000; Dhamija and Perrig, 2000). Specifically, by means of *knowledge based authentication* (KBA) the user in order to be recognized as such, has to respond accurately to a (standard) question, the response to which has been memorized in the IT system on a previous occasion. This helps to create greater empathy with the user, who should then see herself / himself not as a merely passive 'number' in the battery of interviewees but rather as an individual being asked to interact dynamically. This technique is widely used to allow users to regain access to internet services in the event the password is forgotten. Should the response to the standard query be correct, the automatic system sends an e-mail to the address previously provided by the user reminding her / him of the username and password (or engendering new ones). Some queries in effect could be conceived as 'control' questions in order to obtain the maximum reliability possible in the interviewee's responses [De Luca, 2006].

As mentioned previously, the system of constant authentication could slow down or even irritate the user. However, to date no more incisive or efficient system seems to exist. Other mechanisms of access are in place, based not on what the user 'knows' as analyzed up to now, but rather on what the user 'owns' (e.g. a smart card, a token, etc.) and/or on what the user 'is' (biometric scanning) [Teti and Festa, 2009]. Even in such cases however, a user could easily gain access

to the system for the first time with her / his own credentials and leave someone else to take part in the survey at a later date. The process of constant authentication (registration - authentication - verification) refers consequently to an ongoing verification process relative to user behavior in terms of seriousness (eliminating or mitigating problematic issues concerning the related stages of *respondent, authentic and reliable* categories) of the 'funnel'. As in all monitoring activities, from an operative perspective, performance is slower; at the same time, however, improved quality in terms of the survey are guaranteed. It could happen that some of the selected users abandon the survey or even the internet sample managed as a virtual community: at this point a compromise between sample quality and sample quantity should be sought. For this reason, qualitative research seems more adequate from an internet environment perspective being less sensitive to sampling problems and sample number and being aimed primarily at exploratory and not descriptive aims.

6. Research findings, implications and conclusions

The evolving of IT and in particular, the introduction, diffusion and development of the Internet have had, continue to have and no doubt will have in the future an extraordinary impact on routine economic and social processes. These innovations have naturally impacted on enterprise information systems above all, provoking radical changes in procedural design, working and development (organizational processes) and logics (information flows) called upon to carry out specific business activities.

One of the disciplines involved to a great extent in such change is that of market research which uses the Internet as an additional tool together with its traditional 'toolbox' (research 'using' the Internet) or even as an additional/substitute/ innovative target of individuals to survey (research 'on' the Internet). The most problematic issues from a research perspective derive effectively speaking from research carried out 'on' the Internet where – besides the inevitable need to contextualize adequately new IT tools – potential distance (social, cultural, economic, etc.) has to be taken into account between the population offline and that online, with inevitable repercussions on the appreciation of the representativeness of the online sample compared to the population offline.

In this study, clearly oriented towards research of a conceptual nature, we have defined a logical representation of the main problematic issues of sampling in online research and developed a theoretical-practical framework ('funnel') which takes into account on the whole, the sample and non sample errors which could emerge in methodological terms in the representativeness link (on the one hand) between the population online and sample online and (on the other) between population offline and sample online. By virtue of this scheme, with the support of marketing, statistics and IT resources, we have suggested some empirical solutions for the managing of online representativeness, over and beyond 'substantial' statistics in the case of sampling error and over and beyond 'applied' marketing in the case of non sampling error.

These solutions in any event are useful only if pursued in the prospect of the internet sample of interviewees conceived as a virtual community, considered by

the researcher (and above all by themselves) as members of a specific active online group even if only for the duration of the survey. Thus it is fundamental to devise an analytical internet marketing strategy in order to adapt and maintain the sample / community, resorting to the changing combination of visibility and attraction.

The research implications of the present study seem to concern above all the potential evolution of the *humus* at the foundation of the market research discipline or in other words, the traditional combination of marketing, statistics and (by now) IT, identifying in such mix of skills, grounds for further research both in methodological and applied terms. By virtue of the pervasiveness of the Internet, in depth studies on market research will need in the future to deal with the social evolution of the Net which is already scheduled to meet (more so in the future) the need / opportunity of offering solutions (for the present) and/or far-sighted solutions (for the future), in the same way that the reflective spirit of this study identifies indications for application above all in terms of operative implications and therefore of interest for anyone called upon to carry out research online.

Market research online it has to be noted is still at an extremely early stage. This is generally true for all internet-based environments as concerns values, principles and above all working rules (e.g. netiquette, privacy, social networks, etc.). This characteristic, together with the vertiginous speed of development of internet tools and applications and their potential of use which compress the duration of the *internet year* exponentially, renders already obsolete or at least, risky, any empirical solution (one of the most likely limits of our conceptual paper). For this reason, it is fundamental to devise, contextualize and develop a strict logical-methodological framework such as that underpinning our study. Our approach, starting with the analysis of the 'traditional' methodology of market research, proposes, in the prospect of the Internet and subsequently with regard to the prospect of sampling on the Internet, substantially 'common sense' management solutions typical of business management scenarios. In other words, our study goes beyond the mere analysis of the usefulness of a particular statistic or IT tool, but measures considerations and examines applications observed as respondent to a global analytical internet marketing strategy.

References

Andreani J.C., Conchon F., Gli studi qualitativi in Internet, Micro & Macro Marketing, 1, 2001a, 65-74.

Andreani J.C., Conchon F., Les études qualitatives en marketing, Paris: ESCP-EAP, Les Cahiers de Recherche, 2001b, 01-150.

Andreani J.C., L'interview qualitative en marketing, Revue Française du Marketing, 1988, 3-4.

Andrews D., Nonnecke B., Preece J., Electronic survey methodology: A case study in reaching hard-to-involve Internet users, International Journal of Human-Computer Interaction, 16, 2, 2003, 185-210.

Barile S., Metallo G., Le ricerche di mercato. Aspetti metodologici e applicativi, Giappichelli, Turin, 2002.

Bassi F., Analisi di mercato. Strumenti statistici per le decisioni di marketing, Carocci, Rome, 2008.

Cantone L., Calvosa P., Testa P., Tecnologie ad alta intensità connettiva, relazioni di marketing e processi di creazione di valore per i clienti nei mercati BtB, Mercati & Competitività, 2006, 3, 1.38.

De Luca A., Le ricerche di mercato. Guida pratica e metodologica, Angeli, Milan, 2006.

Dhamija R., Perrig A., D´ej`a Vu: A User Study Using Images for Authentication, Proceedings of the 9th USENIX Security Symposium Denver, Colorado, USA, August 14-17, 2000, 1-15.

Di Fraia G., Validità e attendibilità delle ricerche on line, in Di Fraia G. (ed.), e-Research. Internet per la ricerca sociale e di mercato, Laterza, Rome-Bari, 2004.

Ellison C., Hall C., Milbert R., Schneier B., Protecting secret keys with personal entropy, Future Generation Computer Systems, 16, 2000, pp. 311-318.

Festa G., Il ruolo dei nuovi strumenti informatici nel sistema informativo di marketing, Esperienze d'impresa, 2, 2001b, pp. 67-114.

Festa G., Digital Business English - Glossario ragionato di linguistica d'impresa per la new economy, Edisud, Salerno, 2001a.

Hagel J., Armstrong A., Net gain, Harvard Business School Press, Boston, Massachusetts, USA, 1997.

Howard P., Rainie L., Jones S., Days and nights on the Internet: The impact of a diffusing technology, American Behavioral Scientist, 45, 3, 2001, 383–404.

Kiang M.Y., Raghu T.S., Huei-Min Shang K., Marketing on the Internet — who can benefit from an online marketing approach?, Decision Support Systems, 27, 2000, 383-393.

Kotler P., Marketing Management, Pearson Education Italia, Milan, 2001.

Marbach G., Ricerche per il Marketing, Utet, Turin, 2010.

Mariampolski H., Qualitative Market Research, Sage, London, UK, 2001.

Martone D., Furlan R., Online Market Research Tecniche e metodologia delle ricerche di mercato tramite Internet, Angeli, Milan, 2007.

McDonald H., Stewart A., A comparison of online and postal data collection methods, Marketing Intelligence & Planning, 21, 2, 2003, 85-95.

Metallo G., Festa G., La progettazione dei portali web nell'ottica del customer service, in Barile S., Metallo G. (eds.) Soluzioni problematiche d'impresa. Riflessioni e modalità risolutive, Edizioni Culturali Internazionali, Rome, 2003.

Molteni L., Troilo G., Ricerche di marketing, McGraw-Hill, Milan, 2003.

Natale P., Il sondaggio, Laterza, Bari, 2004.

Prandelli E., Verona G., Marketing in rete, oltre Internet verso il nuovo Marketing, McGraw Hill, Milan, 2006.

Schmitz N., Rappresentatività e campionamenti nelle survey on line, in Di Fraia G. (ed.), e-Research. Internet per la ricerca sociale e di mercato, Laterza, Rome-Bari, 2004.

Seglin J.L., Marketing (la guida al), McGraw-Hill, Milan, 1994.

Teti A., Festa G., Sistemi informativi per la sanità, Apogeo, Milan, 2009.

Tissier-Desbordes E., Les études qualitatives dans un monde post-moderne, Revue Française du Marketing, 1998, 3-4.

Valdani E., Busacca B., Customer Based View: dai principi alle azioni, Conference Proceedings "Le tendenze del Marketing in Europa", Università Ca' Foscari Venezia, 2000, 1-24

Vescovi T., Il Marketing e la rete. La gestione integrata del web nel business. Comunicazione, e-commerce, sales management, business to business, Il Sole 24Ore, Milan, 2007.

Weible R., Wallace J., Cyber research: the impact of the Internet on data collection, Market Research, 1998, 10, 3, 19-31.

Wright K.B., Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services, Journal of Computer-Mediated Communication, 10, 3, 2005.

Biographies

Maria Vincenza Ciasullo is an Associate Professor of Management at the University of Salerno (Italy), where she teaches business administration and corporate governance. Ph.D. in business economics (c/o the University of Naples "Federico II"). Main topics of her studies are management and governance systems in private and public organizations, focusing on corporate social sustainability and business ethics. She successfully combines theoretical and empirical research. She is author of several national and international publications.
email: mciasullo@unisa.it

Giuseppe Festa is an Assistant Professor of Management at the University of Salerno (Italy). He has taught several courses, such as economics and management of health organizations and economics and management of information technology. He has published several articles in national and international refereed journals. Internet marketing is one of his most important research topics.
email: gfesta@unisa.it

Integrating Statistical Data with the Semantic Web: The ISTAT Experience

R.Aracri, S. De Francisci, A. Pagano,
M. Scannapieco, L. Tosco, L. Valentino

Abstract. *The paper describes an experience, made by Istat (Italian National Institute of Statistics), on data dissemination according to Semantic Web technologies, starting from standards adopted in the statistical domain. More specifically, the paper shows the design and implementation issues related to the development of a translator from the SDMX (Statistical Data and Metadata eXchange) data model to the RDF Data Cube vocabulary data model. The effectiveness and efficiency of the translator are validated against real Istat dataset*

Keywords: semantic web, statistical data, linked data.

1. Introduction

"Linked Data" (http://linkeddata.org/) permits to create and interlink arbitrary volumes of structured data across the Web. The Linked Data initiative is made possible by the widespread adoption of Web standards for publishing data according to the Resource Description Framework (RDF) model. RDF allows to uniquely identifying resources on the Web, by means of a specific URI (Uniform Resource Identifier). This feature has several advantages, including (i) the possibility of a direct access to resources via a query language and (ii) the ability to link data together in order to access them in an integrated way (with the clear positive side-effect of higher quality, more information more easily accessed, and so on).

In this paper, we present the first result of the IS-LOD (Istat - Linked Open Data) project. The project started in September 2012, with the general objective of investigating how to integrate the huge amount of data disseminated as Official Statistics with the Semantic Web.

National Statistical Institutes have a consolidated language for representing the statistical data they publish, namely SDMX (Statistical Data and Metadata eXchange) [SDMX-2013]. The focus of the contribution of this paper is on the design and implementation issues related to the development of a "translator" from the SDMX data model to the RDF Data Cube vocabulary data model [RDF-

QB-2013], this latter being the current standardization initiative for statistical data publication based on Semantic Web standards. An extensive validation of the translator against real Istat datasets is a further presented contribution.

2. Background

The increasing Internet penetration and the need for electronic data exchange have generated a proliferation of exchange protocols ad hoc, also for statistical data.

Eurostat maintains a large number of transactions of this type in its role of coordination on national institutes communications and holder of shared data, for this reason it receives data flows from all the National Statistical Institutes of the state community members. In 2001, Eurostat launch the initiative SDMX (Statistical Data and Metadata eXchange) in partnership with other international organizations (Bank for International Settlements, European Central Bank, International Monetary Fund, OECD, United Nations Statistical Division and World Bank). The initiative was intended to ensure that metadata always come along with the data, making the information immediately understandable and useful.

SDMX provides standard formats with content guidelines and an IT architecture for exchange of data and metadata. Organizations are free to make use of whichever elements of SDMX are most appropriate in a given case. To enable the dissemination of the new standard, Eurostat has established this as protocol for flows concerning Eurostat directly, and has recommended it for all statistical data exchanges.

In a few years SDMX has established itself as a standard protocol for statistical data and metadata description. Eurostat also promotes an IT architecture for data dissemination based on SDMX. Istat has currently in place such an architecture, enriched with a service oriented to machine to machine communication, named Single Exit Point (SEP) [Cardacino 2013].

2.1. SDMX Framework

SDMX is a standard, based on XML, designed for the exchange of statistical data. The data structures that are needed to completely describe a statistical phenomenon are grouped into two levels:

- the data set, that is a collection of observations related to the described phenomenon;
- the data structure definition, that is all metadata describing the structure and organization of the data set, the statistical concepts and attached to them code lists used within the data set.

The SDMX Information Model (SDMX-IM) is organized to represent both levels. In particular, the following structures are defined:

- Key Family: ordered list of dimensions, measures and attributes that define the structure of the data set. Each observation in the data must respect this structure;
- Concepts: descriptor of dimensions, measures and attributes (and optional group key) used in the data set;

- Code Lists: define the valid content of each of the concepts used in the data set;
- Data Set: contains data (observations) and related metadata whose content conforms to the specification of the Key Family definition.

The first three structures are grouped into a single data flow that takes the name of Data Structure Definition (DSD). For the Data Set flow there are three possible formats: Generic, Compact and Cross Sectional.

In the Generic and Compact formats, observations are organized by time series. In this case, the difference between the two formats is that the Generic is more verbose than the Compact.

If the time dimension is not present, or data cannot be organized in time series the format to use should be the Cross Sectional. This format is very general and, for each observation block, all the dimension values are explicitly expressed.

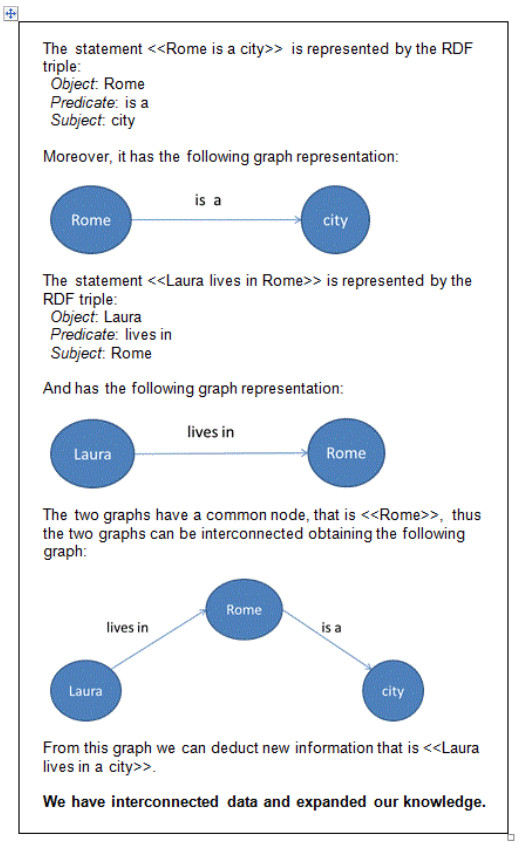


Figure 1
Example of knowledge deduction with data represented via RDF.

2.2. RDF, RDF Schema, OWL, SPARQL

The RDF (Resource Description Framework) [RDF-2004] is a standard W3C data model that has features facilitating data integration even if the underlying schemas differ.

The RDF model is very simple: all objects are represented by URIs and URIs are linked by a simple subject-predicate-object (triple) structure. Using this structure, different phenomena can be represented and data from different Web sources can be mixed and linked.

RDF data can be represented with one of the following serialization formats:

- RDF/XML: a notation XML compliant;
- N-triples: the graph is serialized as a set of triples subject-predicate-object [N-triples-2013];
- Notation3: the graph is serialized describing, one at a time, all theresources and all its properties [N3-2013];
- Turtle (Terse RDF Triple Language) it is a subset of Notation3, this notation is considered the most human readable [Turtle-2013].

RDF Schema is the RDF based standard [RDF Schema-2004] for the definition of an ontology. An ontology is "a formal and explicit representation of a shared conceptualization of a domain of interest" [Gruber, Thomas R. 1993]. The term “formal” indicates the usage of a logical language, thus processable by a machine; the term “explicit” indicates that there is no ambiguity of interpretation; the term “conceptualization” indicates the abstract visualization of the domain of interest; the term “shared” indicates the acceptance of the ontology from the community.

RDF Schema defines the following (main) classes: (i) rdfs:Class, (ii) rdfs:Resource, (iii) rdfs:Literal, (iv) rdfs:Datatype and the following properties: (i) rdfs:subClassOf, (ii) rdfs:subPropertyOf, (iii) rdfs:domain, (iv) rdfs:range.

Besides RDF Schema, the further standard for ontology representation is OWL (Ontology Web Language) [OWL-2012]. The purpose of OWL is identical to RDF Schemas, i.e. to provide an XML vocabulary to define classes, properties and their relationships, however (see Figure 2):

- RDF Schema enables you to express very rudimentary relationships and has limited inference capability;
- OWL enables you to express much richer relationships, thus yielding a much enhanced inference capability.

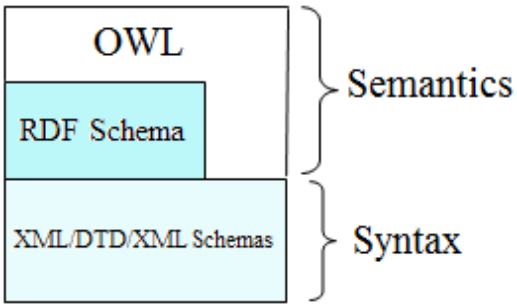


Figure 2
The Ontology Semantic Web Stack

SPARQL (Sparql Protocol And RDF Query Language) [SPARL-2008] is a language with a syntax similar to SQL for querying RDF data and a communication protocol based on HTTP. A SPARQL client can query a SPARQL endpoint with queries on a RDF graph by allowing “graph pattern matching” on RDF data.

2.3. RDF-QB

The specialization of RDF protocol to represent statistical data is RDF Data Cube (RDF-QB) [RDF-QB-2013] witch is a W3C candidate recommendation as of 25 June 2013.

The RDF-QB is based on the SDMX Information Model and on other existing vocabularies listed in Table 1. Like in the other vocabularies, the names of the entities are URIs expressed with a compact notation prefix:localname where the prefix identify a URI namespace and its concatenation with the localname gives the complete URI. The prefix to be used in RDF-QB is qb.

Vocabulary	Namespace (prefix)	Description
SKOS - Single Knowledge Organization	skos	SKOS gives specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web
Void - Vocabulary of Interlinked Dataset	void	RDF schema vocabulary for expressing metadata about RDF DATASETS. It is a bridge between publisher and user of RDF data, with application ranging from data discovery to cataloging and archiving of datasets
FOAF - Friend of A Friend	foaf	RDF-OWL Vocabulary for describing persons, their activities and their relations to other people and objects.
Dublin Core	dc	Basic vocabulary to describe documents.
ORG	org	Core ontology for organizational structures, aimed at supporting linked-data publishing of organizational information across a number of domain
RDF Vocabulary	rdf, rds	A defined set of predicates that can be used in an application

Table 1
RDF-QB Vocabularies

2.4. The Semantic Web

The term “Semantic Web” refers to an evolution of the Web with the aim to give a semantic meaning to data and documents there published in order to make them semantically understandable not only to humans but also from machines (thus, machines will be able to carry out logical deductions applying knowledge representation techniques).

The Semantic Web includes the possibility to interconnect data, and not only documents, represented using well-defined syntaxes making data machine readable.

Figure 3 represents the Semantic Web stack, with seven distinct layers.

The first two layers represent the format and syntax standard to be applied to represent data. Specifically, in the first layer, we find Unicode, an international standard for the representation of all characters, and URI (Uniform Reference Identificator) standard for the unique identification of all resources. In the second layer, we find a block referring to XML, NS (Name Space) and XMLSchema that are three standards to describe data with specific markup syntax.

The subsequent two layers are related to the standards to be used to represent the semantics of data. Specifically, in the third layer we find RDF and RDF Schema.

In the fourth layer, we found the block related to ontology.

The last three layers related to the “knowledge” have not yet been formalized. In particular, the logical level refers to the possibility of creating a unifying logic that hides incompatibilities between lower layers. The level of proof refers to the possibility to check the truthfulness of the new knowledge obtained inferring the semantic relationships that interconnect the data (e.g. the deduction <<Laura lives in a city >> in Figure 1).

Finally, the trust level of knowledge is related to the reliability of knowledge determined by the lower levels of the Semantic Web stack.

Currently, the Semantic Web, in its pure sense, is still far from being realized; what is taking place is the Web of (Linked) Data that is the fundamental layer on which the Semantic Web can be built on.

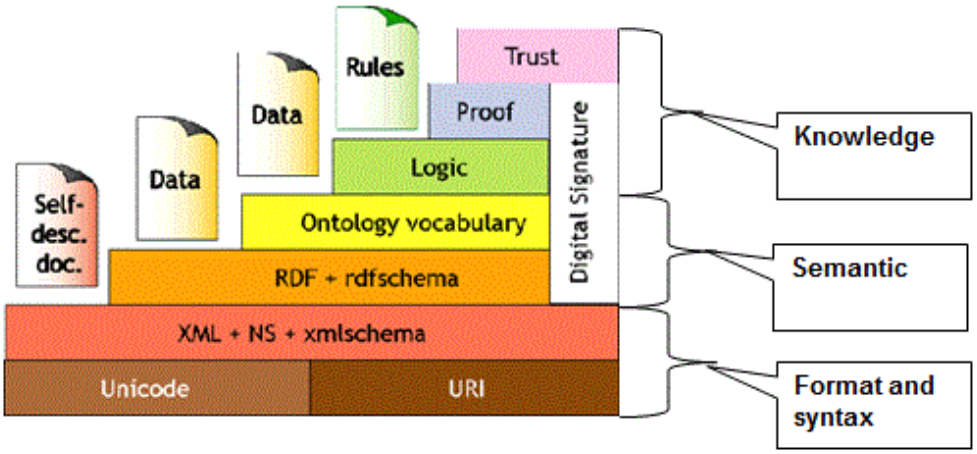


Figure 3
The Semantic Web Stack

2.4.1. Linked Data and Linked Open Data (LOD)

Linked Data paradigm is fully aligned with the Semantic Web: it defines the rules that must be followed to make data on the Web available, easily accessible, and usable by human users and by machines. Thus, data are represented in RDF and a huge number of links are referenced in order to allow a more extensive browsing through the data.

A key step in this project has been the addition of the “Open” adjective to Linked Data making available all datasets and agreeing to publish bigger and more interesting datasets.

The Linked Open Data project has had an immediate and widespread success, thus it has established itself as a standard de facto for the publication of data in the e-government sector. An important example of e-government portal that publishes Linked Open Data is the U.S. data.gov site; further examples are Wikidata and DBpedia.

The LOD project has been developed from the low, mainly from DBpedia project that presents the content of Wikipedia in LOD format.

The fundamental principles on which LOD rely are:

1. URI usage to identify both objects and abstract things/concepts;
2. Usage of the HTTP protocol;
3. Usage of standards as RDF, SPARQL to return information when a URI is dereferenced;
4. Include in the descriptions links to other URI in order to increase the interconnections between data.

Figure 4 shows the Linked Data Cloud that represents graphically the size reached by the currently linked data on the Web.

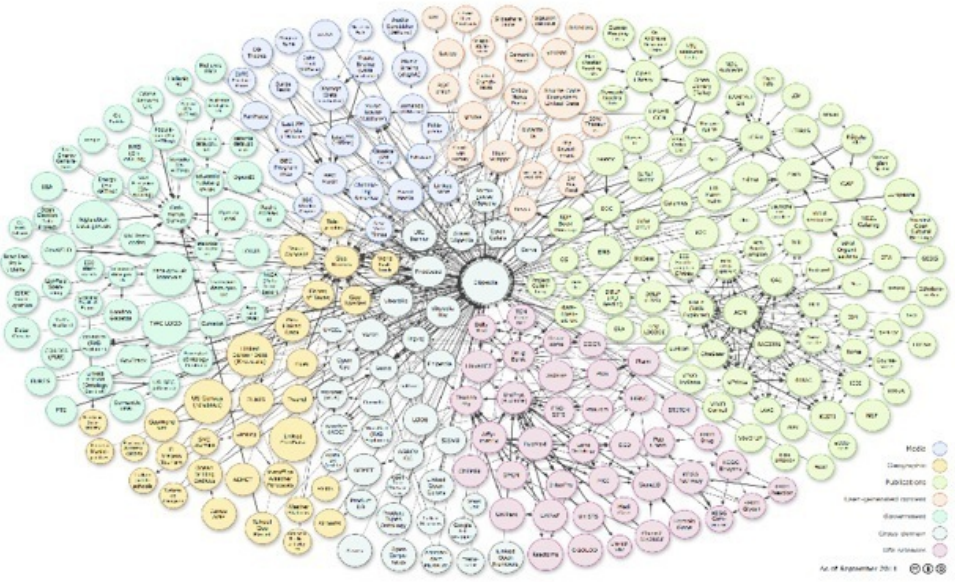


Figure 4
Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. [http://lod-cloud.net/]

2.4.1.1. LOD Five-level model

Tim Berners-Lee has proposed a series of principles to be followed in opening data, and has defined a model, known as the model of the five-star quality for open data on the Web.

The model allows you to understand the level of readability of published data and their ease of access.

It consists of five layers as shown in Figure 5.

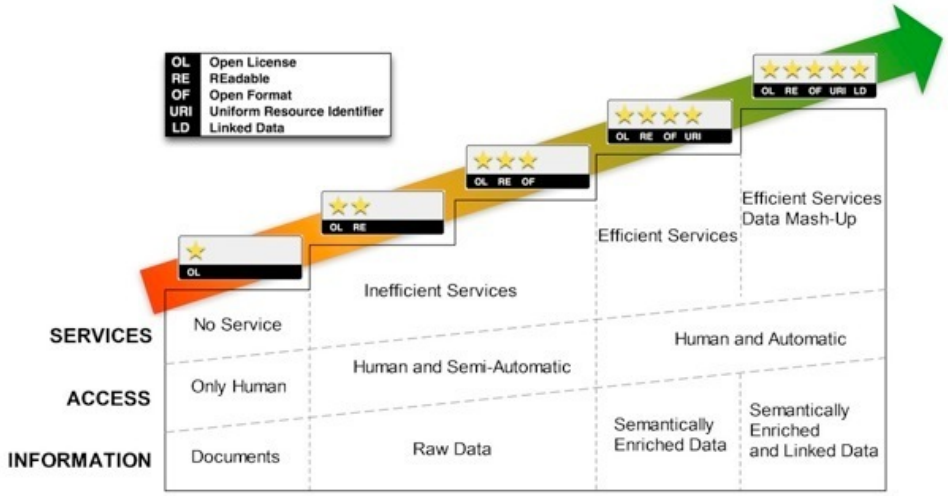


Figure 5
Five-level model for the production and release of open data [AgID-2013]

The five levels are:

	Data available in any format, but with an open license;
	Data available in a format readable by an agent automatically. Typically, fall into this level data in proprietary formats (e.g. Excel)
	Data with the characteristics of the previous level but with a non-proprietary format;
	Data with the characteristics of the previous level but exposed using the standard W3C RDF and SPARQL;
	Data with the characteristics of the previous level but interlinked to data exposed by other people/ organizations.

The three dimensions introduced in Figure 5, with the different degrees of the scale, are described as follows.

INFORMATION

This dimension describes the quality of information provided along with the open data. In this sense we may have:

- *documents*: the data is embedded within documents and then only readable by humans (level 1);
- *raw data*: the data can also be read by a program, but human intervention is a strong need for some of the same processing (levels 2 and 3);
- *semantically enriched data*: the data are semantically described using metadata and ontologies (level 4);
- *semantically enriched and linked data*: the data are semantically described using metadata and ontologies (level 5). Human intervention is minimal and sometimes even absent.

ACCESS

This dimension describes the easiness with which human users and automated agents are able to access the data, and then considers the effort of understanding of the data structure in order to question them and use them properly. The degrees are:

- *only human*: only humans are able to read the documents and then make sense of the data contained on (level 1);
- *human and semi-automated*: automated agents can process data but are not able to understand them, so humans have to write ad-hoc programs for their use (levels 2 and 3);
- *human and automated*: automated agents who know the reference ontology can process data without further human intervention (levels 4 and 5).

SERVICES

This dimension describes the type of services that can be designed and implemented with open data, based on different degrees of efficiency and capacity of the service to exploit information even from different sources.

- *no service:* no service can be enabled from the data contained in the documents, unless significant interventions of human in extraction and data processing (level 1);
- *inefficient services*: ad-hoc applications that use the data can be developed. These should incorporate data inside (levels 2 and 3);
- *efficient services and apps*: applications, also for mobile devices, can be developed using direct access to the Web to find the data of interest (level 4);
- *efficient services and mashups of data*: applications for mobile devices, can be developed by exploiting both direct access to the Web and the extra information captured by the "link" of the relevant data (level 5).

3. The IS-LOD (Istat- Linked Open Data) Project

Istat is the principal provider of open data among all the Italian public administrations, according to an analysis perfomed by Formez [Marras-2013]. In particular, Istat provides more than 600 datasets as open data. Such datasets are mainly published on the I.stat web datawarehouse (http://dati.istat.it), from which it is possible to download datasets in CSV, Excel and SDMX formats. The machine-to-machine communication is managed by a dedicated service named Single Exit Point (SEP), that allows immediate access and reusability of the disseminated data. As anticipated in the Section Background, the SEP system is based on the Eurostat SDMX Reference Infrastructure, hence using SDMX for data and metadata descriptions.

However, Istat open data are at level 3, of Tim-Berners Lee five-star stack [Five-Star-2013]. In order to improve Istat open data dissemination towards level 4 (RDF) and level 5 (LOD), in September 2012 we started the IS-LOD project.

The first objective of this project was to investigate how to publish Istat data taking advantage from the standardization work carried out for the publication of data according to SDMX model and format. It is notable that data expressed in SDMX come with a large amount of metadata that, if reused, will allow us to publish Linked Data of high quality.

3.1 Translator: from SDMX to RDF Data Cube Vocabulary

As mentioned, our objective was to understand how much and how costly was to approach the semantic Web starting from SDMX. Hence, we decided to design and implement an automated translator from the SDMX model and format to RDF Data Cube Vocabulary.

As a first step, we performed an evaluation of the technological environments that could support the implementation of the Translator. Three different solutions were identified and compared, namely:

- R package RSDMX developed at FAO (http://r-forge.r-project.org/R/?group_id=1298). The package could serve the purpose of supporting reading/writing operations for manipulating SDMX files. However, the package is not complete and not actively maintained.
- Apache Jena (http://jena.apache.org/). Jena is a Java framework for reading, processing and writing data in RDF, it also provides a query engine compliant with the latest SPARQL specification. While useful for RDF manipulation, the framework does not cover our further requirement related to manipulation of SDMX documents.
- XSLT (Extensible Stylesheet Language Transformations) technology: useful for specifying transformations between XML files. We found a successful usage of XSLT transformation within the MIMAS Linked data Project [MIMAS-2011], with an objective similar to ours, i.e. mapping data expressed in SDMX format to RDF-QB format.

This XSLT solution was hence selected for our Translator.

As a second step, we designed the mapping transformations between the elements of the SDMX data model and the elements of the RDF-QB data model, as depicted in Fig.6.

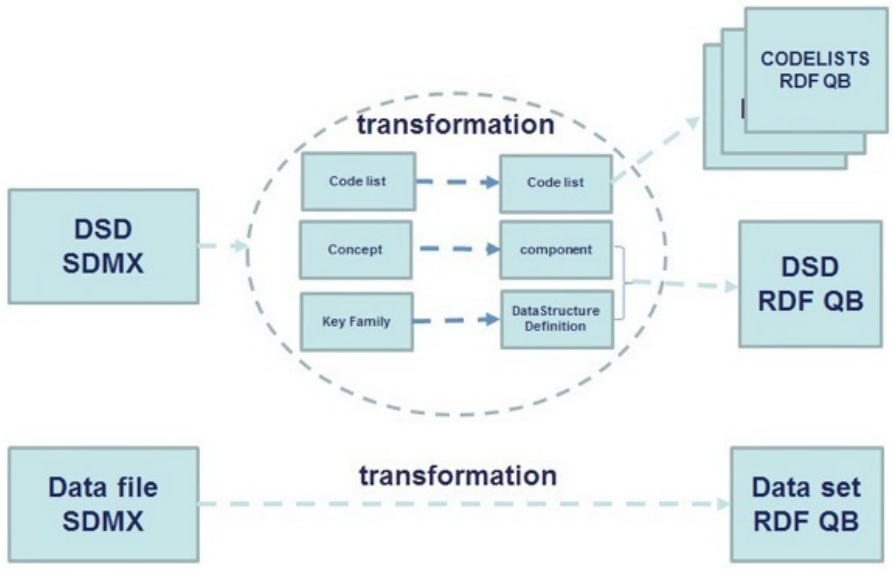


Figure 6
Translator's mapping rules

In the following, we show an example of a mapping rule expressed via XSLT.

Let us consider the component REF_AREA stated in the key family section of the DSD related to "Separate collection indicators" of the Istat data on Environment and Energy Waste:

```
<registry:KeyFamilies>
  <structure:KeyFamily id="SEP_AMBIENTE_RIF_INDRACDIFF" version="1.0" agencyID="IT1" isFinal="false">
    <structure:Name xml:lang="en">Environment and energy - Waste</structure:Name>
    <structure:Description xml:lang="en">Separate collection indicators for the provincial capitals
              Separate collection indicators
              Separate collection kilogram per capita by type of waste</structure:Description>
    <structure:Components>
      <structure:Dimension conceptRef="REF_AREA" conceptSchemeRef="SEP_AMBIENTE" conceptSchemeAgency="IT1"
                           conceptSchemeVersion="1.0" codelist="CL_REFAREA" codelistVersion="1.3" codelistAgency="IT1" />
```

The component REF_AREA stated in the key family refers to the CL_REF_AREA code list:

```
<structure:CodeList id="CL_REFAREA" version="1.3" agencyID="IT1" isFinal="false">
  <structure:Name xml:lang="en">Territory</structure:Name>
```

REF_AREA is defined as follows:

```
<structure:Concept id="REF_AREA">
  <structure:Name xml:lang="en">Geographical reference area</structure:Name>
</structure:Concept>
```

The element under consideration, when described in RDF-QB is more verbose as it must be declared not only as a dimension of the Data Cube Vocabulary but also as an SDMX DimensionProperty and as an SDMX CodedProperty as it refers to a specific code list (in this case "territory" code list):

```
<rdf:Description rdf:nodeID="REF_AREA">
    <qb:dimension rdf:resource="http://dati.istat.it/output/code/territory"/>
    <dc:language>en</dc:language>
    <rdf:type rdf:resource="http://purl.org/linked-data/sdmx#DimensionProperty"/>
    <rdf:type rdf:resource="http://purl.org/linked-data/sdmx#CodedProperty"/>
    <sdmx:codeList rdf:resource="http://dati.istat.it/output/code/territory"/>
    <rdfs:range rdf:resource="http://dati.istat.it/output/code/territory"/>
    <rdfs:label xml:lang="en">Geographical reference area</rdfs:label>
</rdf:Description>
```

XSLT and XPATH statements, included in the portion of the stylesheet showed in the following, write the tags rdf:Description and qb:dimension which refer respectively to the Core RDF vocabulary (http://www.w3.org/1999/02/22-rdf-syntax-ns#) and to the Data Cube Vocabulary (http://purl.org/linked-data/cube#):

3.2 Technological Solution

As described in the previous Section, we used XSLT to convert data expressed in SDMX to RDF-QB. XSLT is a very powerful language xml-based proposed by W3C to apply transformations to XML documents. It is designed to be used as part of XSL, which is a stylesheet language for XML.

In more details, XSL consists of three elements:

- XSL Transformations (XSLT): a language for transforming XML;
- XML Path Language (XPath): an expression language used by XSLT (and many other languages) to access or refer to parts of an XML document;
- XSL Formatting Objects (XSL-FO): an XML vocabulary for specifying formatting semantics.

As XSLT Processor we chose Saxon (http://saxon.sourceforge.net/). It provides implementations of XSLT 2.0, XQuery 1.0, and XPath 2.0 at the basic level of conformance defined by W3C. It is available both for Java and .NET. We used Saxon Java Home Edition, the open source version of the tool. This tool version can be used from the operating system command line or can be integrated into other tools and applications.

For the Translator, an XSLT file, even if generalized, needs the following parameters to be set at runtime for each transformation to be executed:

- Language: indicating the reference language, that is the language in which SDMX code lists, DSDs and data are expressed;
- Alternative language: indicating an alternative language in case no language is specified in SDMX files;
- DSD: containing the path and the name of the DSD file.

We developed a standalone Java application that, before carrying out the transformation, performs a preprocessing of the stylesheet setting the necessary parameters. For accessing, manipulating, and outputting stylesheets we have used JDOM, a Java-based "document object model" for XML files (http://www.jdom.org/).

Finally, as anticipated in the Section Background, RDF can be serialized in different formats, namely: RDF/XML, Turtle, Notation-3 and N-triples. Since the source SDMX data and DSD file are both in XML format, we chose to use the XML serialization of RDF as our primary output. By integrating a Java library provided by any23.org (http://any23.apache.org/) in our Translator, we can also obtain Turtle format as an alternative output.

In Fig. 7 the overall technological architecture of our Translator is shown.

4. Validation and Experiments

The validation of our Translator was carried out according to three steps:

1. syntactic validation;
2. semantic validation;
3. experimental validation.

The validation process was carried out on real Istat datasets, represented in Table 2. The datasets were selected in order to maximize their diversity with respect to both dataset number of observations and number of dimensions.

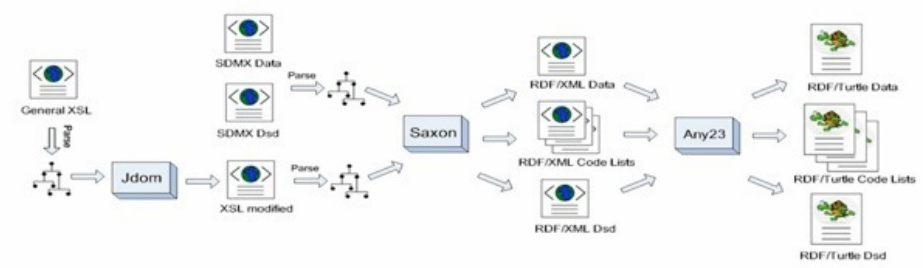


Figure 7
Technological architecture of the translator

In the *syntactic validation* step, the obtained output files, expressed in RDF/XML format, were syntactically validated by creating the RDF triples and their graph representation. To this purpose, we used a free service provided by W3C called Validator (http://www.w3.org/RDF/Validator/) that takes as input RDF/XML document and returns a 3-tuple representation of the corresponding data model and an optional, but very explanatory, graphical visualization of the data model. Fig.8 shows the graphical representation of a simplified DSD containing only the component REF_AREA.

The syntactic validation of our output does not imply that our RDF files are "valid" with respect to the RDF-QB model. Therefore, we undertook a second step of *semantic validation*. To this purpose, we used the RDF Data Platform: Virtuoso Open Link Software (http://virtuoso.openlinksw.com/), and in particular we used the Virtuoso SPARQL entry point. Once uploaded the RDF/XML files (code lists, DSD and data), we examined our data by performing a set of focused SPARQL queries. For example, we issued a SPARQL query extracting the “Percentage of separate collection of waste in Rome”, as shown in Fig. 9:

With respect to the experimental validation, three different versions of the Translator were compared, namely:

- Base: This version was actually a generalization of the stylesheet used by the MIMAS project. Such a stylesheet had been really tailored on a specific case study, the World Development Indicators from the World Bank's Databank service. Hence, the translator performances suffered from such an ad-hoc development of the XSLT code.
- Optimized_1: We proceeded to rewrite the stylesheet of the Base version, to avoid redundancy, and unnecessary cycles minimizing repeated access to the information contained in the DSD file.
- Optimized_2: In this version, a considerable improvement was obtained using a technique that allows to load in memory, as an XSLT variable, an

XML structure that contains information of the DSD avoiding multiple accesses to the file. This version reduced drastically the execution time.

Dataset Name	Size (KB)	Dabase Description	
		#Dimentions	#Observations
Households Economic Conditions and Disparities - Poverty	17	8	160
Households Economic Conditions and Disparities - Housing conditions	60	17	532
Gross earning per full time equivalent unit index - quarterly data	81	10	1.071
Households Economic Conditions and Disparities - Income	159	8	2.152
Services - Tourism	753	9	12.663
Households Economic Conditions and Disparities - Consumer confidence	764	13	10.550
Services - Air transport, rail transport, maritime transport, road transport	791	10	13.104
Environment and energy - Waste	1.338	8	14.302
Consumer price index for the whole nation (NIC - from 2011 onwards)	4.341	8	71.050
Annual national accounts + Quarterly national accounts + territorial Economic Accounts and Aggregates + Annual sector accounts	4.937	12	63.576
Employment	16.367	19	148.400
Households Economic Conditions and Disparities - Consumtions	18.560	12	205.980

Table 2
Datasets used for validation

Table 3 shows the Translator execution times in relation to data flow sizes, for the different Translator versions.

The experiments show that the performance gain from the Base to the Optimized_2 version is of two orders of magnitude.

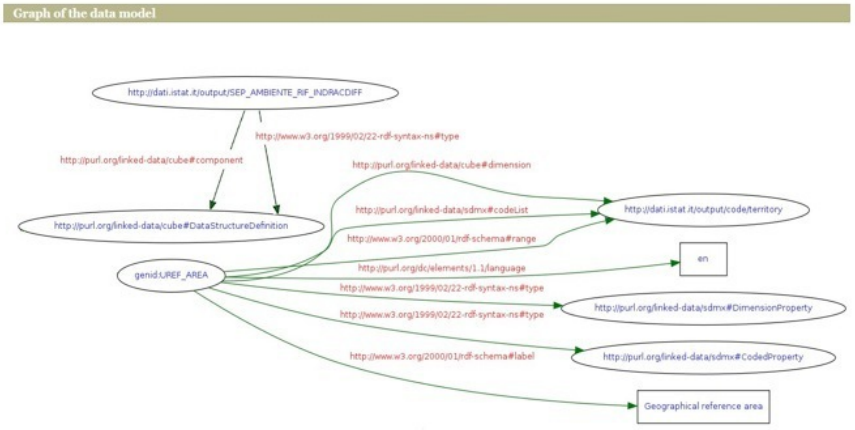
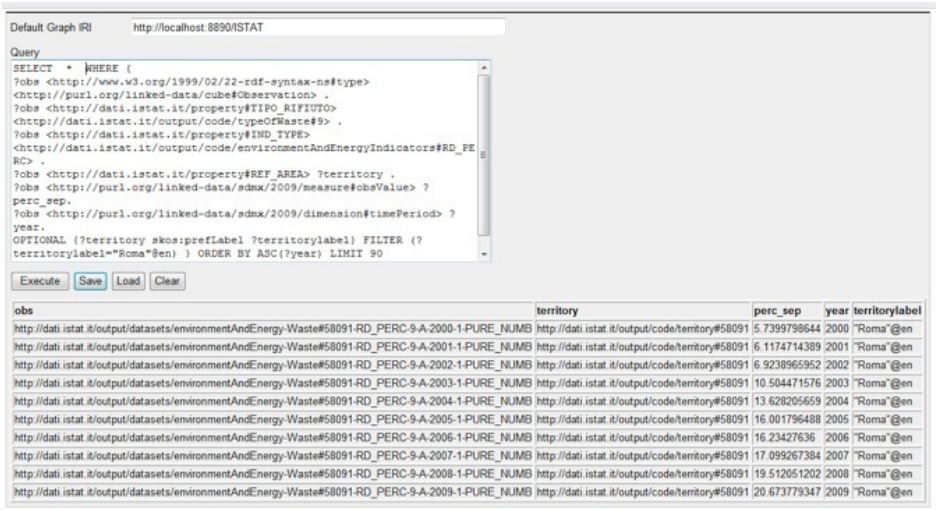


Figure 8
Syntactic validation: RDF triples graph representation of DSD



obs	territory	perc_sep	year	territorylabel
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2000-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	5.7399798644	2000	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2001-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	6.1174714389	2001	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2002-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	6.9238965952	2002	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2003-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	10.504471576	2003	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2004-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	13.628205659	2004	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2005-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	16.001796488	2005	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2006-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	16.23427636	2006	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2007-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	17.099267384	2007	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2008-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	19.512051202	2008	"Roma"@en
http://dati.istat.it/output/datasets/environmentAndEnergy-Waste#58091-RD_PERC-9-A-2009-1-PURE_NUMB	http://dati.istat.it/output/code/territory#58091	20.673779347	2009	"Roma"@en

Figure 9
Semantic validation: example query returning the percentage of separate collection of waste in Rome

5. The IS-LOD Project: Current and Future Work

The actual results of the IS-LOD project tested the viability of disseminating statistical data expressed in RDF-QB starting from SDMX data.

Inevitably, the resulting data are actually documents, and, as such, not directly accessible via related URIs. In future work, the project will investigate how integrating the current Istat dissemination architecture with an architecture fully supporting LOD publication (levels 4 and 5 of Tim-Berners Lee five-star stack).

Also the RDF initiative in Istat is being developed in two additional directions described in the following.

Test Case	Input Size	Base Execution Time	Execution Time after Optim. 1	Execution Time after Optim. 2
Case 1	1,338 KB	6m 7.90s	2m 17.71s	5.60s
Case 2	2,674 KB	12m 32.86s	4m 36.93s	7.87s
Case 3	4,009 KB	18m 42.25s	7m 28.95s	11.00s
Case 4	5,345 KB	23m 49.13s	9m 12.63s	13.97s

Table 3
Translator's execution times for different versions

5.2. Single Exit Point integration

As already mentioned, the translator is currently an offline process that manages SDMX files as input and RDF files as output. The publication phase, in this case, consists of making available RDF-QB files on the Web.

A preferred solution is to make the data directly usable in RDF-QB turning this process into a new output channel of the SEP. This achievement can be obtained in two ways.

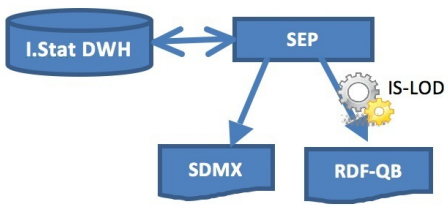


Figure 10
Single Exit Point Integration

The first way is to submit the current SDMX answer to the IS-LOD translator to produce on-line the preferred RDF-QB output format.

The second way is to write a new module of the SEP that, proceeding similarly to the current SDMX, generates a new data flow in RDF-QB format. This new module must use the knowledge of the IS-LOD project as a guide and as a target for output.

5.3. LOD Population Census

In parallel to the Translator project related to its integration with the SEP, a further ongoing project is related to possible RDF data publishing, independently of SDMX.

In particular, we are currently investigating a project aimed at the development of a platform for the LOD dissemination of the population census indicators at the territorial level of census sections.

The idea is to provide both an access point SPARQL, for free querying of census data, stored and linked in a triple store, and a graphical interface for predefined and guided queries which allows a direct and flexible use by different types of users (civil protection, journalists, government agencies, etc.).

References

[SDMX-2013]: SDMX: http://sdmx.org/.

[RDF-QB-2013] RDF Data Cube Vocabulary: http://www.w3.org/TR/2013/CR-vocab-data-cube-20130625/.

[Cardacino-2013] Cardacino A. - Tecnologie web per la diffusione di dati: il Single Exit Point – Newstat n.6 October 2012 – p.4.

[RDF-2004] RDF Specification: http://www.w3.org/standards/techs/rdf.

[N-Triples-2013] W3C, “N-Triples”, http://www.w3.org/TR/n-triples/, 2013.

[Turtle-2013] W3C, “Turtle - Terse RDF Triple Language”, http://www.w3.org/TeamSubmission/turtle/, 2013.

[RDF Schema-2004] RDF Vocabulary Description Language 1.0: RDF Schema: http://www.w3.org/TR/rdf-schema/.

[Gruber, Thomas R. 1993] Gruber, Thomas R. (June 1993). "A translation approach to portable ontology specifications”.

[OWL-2012] OWL 2 Web Ontology Language Document Overview: http://www.w3.org/TR/owl2-overview/.

[SPARQL-2008] SPARQL Query Language for RDF: http://www.w3.org/TR/rdf-sparql-query/.

[Agid-2013] Linee Guida per la valorizzazione de patrimonio informativo pubblico (in Italian): http://www.digitpa.gov.it/notizie/linee-guida-valorizzazione-dei-dati-della-pa.

[Marras-2013]: Marras S.: Open data: Lo scenario di riferimento, XI Conferenza Italiana di Statistica, 2013.

[Five-star-2013]: Five star Open Data, http://5stardata.info/, 2013.

[MIMAS 2011]: MIMAS Project : http://mimasld.wordpress.com/2011/11/25/final-product-post-esds-the-sdmx-to-rdf-process/.

Biographies

Raffaella Maria Aracri works in computer science area since 1999 as software analyst and designer. From 2012 she is a member of the Unit "Software Support to Statistical Processes" of Istat - Italian National Institute of Statistics. She earned a University Degree in Electronic Engineering in 1998 at University of Florence. She is mainly engaged in software development in the context Linked Open Data and Semantic Web.
email: aracri@istat.it

Stefano De Francisci works in ISTAT, in the Information Technology sector, since 1994. He is currently leading the department of "Regolazione e Sviluppo delle Tecnologie ICT". He graduated in 1979 at "La Sapienza" University in Rome with a Master Degree in "Statistical Sciences" and got a Post-Master degree in Operative Research and Decisional Strategies (1992). He has been a lecturer for several courses at "La Sapienza" University from 2004 to 2009. His main interests are in the application of new technologies to statistics, including Big Data and Open Data, and visualization and graphical processing of statistical information.
email:defranci@istat.it

Andrea Pagano works in computer science area since 2000 as Web Technology developer. From 2012 he is a member of the Unit "Software Support to Statistical Processes" of Istat - Italian National Institute of Statistics. He is mainly involved in the Linked Open Data and Semantic Web contexts.
email:andrea.pagano@istat.it

Monica Scannapieco is Head of the Unit "Software Support to Statistical Processes" in the Directorate of Development of Information Systems and Corporate Products, Information Management and Quality Assessment of Istat - Italian National Institute of Statistics. She is a lecturer at SAPIENZA - Università di Roma, where she earned a University Degree in Computer Engineering with honors and a Ph.D. in Computer Engineering. She is author of more than 80 papers mainly on the data quality, privacy preservation and data integration architectures published in National and International contexts. She has been involved in several European and Italian projects on quality of data and data integration.
email:scannapi@istat.it

Laura Tosco is a member of the “Software Support to Statistical Processes” Unit of the Directorate of Development of Information Systems and Corporate Products, Information Management and Quality Assessment of Istat - Italian National Institute of Statistics. She earned the bachelor degree cum laude in Computer Engineering at the Sapienza - Università di Roma. She is co-author of several papers on record linkage and data integration published in national and international contexts.
email: tosco@istat.it

Luca Valentino works in computer science area since 1999. From 2008 in Istat - Italian National Institute of Statistics - where is currently assigned to the unit “Software Support to Statistical Processes”. He earned a University Degree in Mathematics with honors in 1997 at SAPIENZA - Università di Roma. He is expert in data integration procedures, in this area he has been involved in several projects in Istat and he took part in collaborations and conference in international context.
email: luvalent@istat.it

An Overview of Cloud Computing PaaS Offers and Cloud Patterns

B.Di Martino, G. Cretella, A. Esposito

Abstract. *ICloud computing is a very highly interesting topic in both academia and industry and many big players of the software industry are interested to become protagonists of such panorama. This paper presents a comparative overview of cloud services offered by different cloud providers, in particular IBM, Oracle, Amazon and Microsoft, focusing on their PaaS offers and the available tools developed by the vendor themselves to enable their resources and services management. The analysis goes through the comparison of cloud pattern defined from specific vendor and agnostic pattern defined by independent efforts.*

Keywords: e-research, analytical internet marketing, online sampling.

1. Introduction

Cloud computing is a very newsworthy topic in both the academic and commercial world, thus a lot of organizations are involved and interested in the development of solutions and new cloud services. There are different reasons for the raising interest in the development of Cloud solutions. First of all, Cloud Computing platforms offer access to an almost unlimited amount of resources, which can also easily scale according to the user needs, at a reasonable price. Second, clients never purchase or own the resources they use, since they simply rent them: this means that they are completely unaware of how and where these resources are managed and maintained, thus being relived from the relative costs. This model is also known as “Pay As You Go”, since clients have only to pay for the real use of the services and resources they use, not for their management.

Cloud platform vendors generally offer their services according to three different service models: “Infrastructure As A Service” (IAAS), “Platform As A Service” (PAAS) and “Software As A Service” (SAAS).

While SaaS is considered the base service model, offered by most of the Cloud providers, PaaS and SaaS extend and strongly rely on it for their functions. IaaS provides the basic computational, storage and network capabilities on which all services can be built on, so it is more flexible and configurable than other

paradigms, but this also means that clients have to configure everything, from applications to middleware and databases. PaaS provides fully operative computational platforms, which comprehend development and execution environments supporting different programming paradigms and languages, thus making the development process easier and faster, even if with limitations in terms of flexibility. SaaS provides full applications and already configured software, which can be immediately used by clients without installation and customization efforts, and which can be composed to build more complex services.

PaaS model has started developing more recently, if compared to IaaS: the first relevant PaaS provider was Microsoft, with the Windows Azure platform, which was released to the public only on 1st February 2010. Other noticeable organizations, like IBM, Oracle and Amazon, decided to put their effort in developing PaaS platforms for development of Cloud applications, thus enriching the existing offers and services.

In this article we will analyze the different PaaS services offered by the main Cloud providers, namely IBM, Amazon, Oracle and Azure, stressing their differences and describing their approach to the use of Patterns, a fundamental concept inherited from Software Engineering, where it is mainly applied to the development of object-oriented software. In Cloud computing, Patterns can be used to help clients and developers to compose services and functionalities, offered by the same or different providers, in order to create new applications which can fulfill complex requirements or determine new services. As of today, the concept of Cloud Pattern is often used in a misleading way by the different providers: architectural pre-configured cloud solutions, based on specific platforms and thus strongly tied to their particular characteristics, are erroneously labeled as Cloud Patterns.

Here we will make an overview of the different Cloud platforms, and related services, offered as PaaS by different vendors, namely IBM, Oracle, Amazon and Microsoft, and then we will introduce the concept of Cloud Pattern, showing how it is used with different meanings in different contexts.

2. An overview of PaaS Offers

This section provides an overview of the main PaaS offers of cloud provider, highlighting the main distinctive features and capabilities of the different platforms.

2.1. IBM PaaS: SmartCloud Application Services

IBM SmartCloud is an integrated portfolio of services that represents the whole stack of cloud services offered by IBM. It is composed of three families of services: SmartCloud Foundation includes hardware and software that can be used to build private clouds that can be stand-alone or connected to existing private cloud in a hybrid configuration; SmartCloud Services includes the public offers SmartCloud Enterprise and the private service SmartCloud Enterprise+; SmartCloud Solutions provides SaaS services based on applications provided by IBM.

IBM introduced SmartCloud Application Services [SCAS] as part of the SmartCloud portfolio, introducing it as a logical extension of the infrastructural services offered

through SmartCloud Enterprise (SCE) and Enterprise+. In this way IBM defines an integrated PaaS built to maximize the benefits of using cloud for different applications and environments. It can be seen as an additional service provided by SmartCloud Enterprise, which enables all users of SCE to use SCAS immediately, by accessing through the same portal.

The PaaS services are represented through application pattern working on IaaS cloud. IBM patterns are established software components designed to speed up the development and deployment of applications in the cloud, based on predefined architectures.

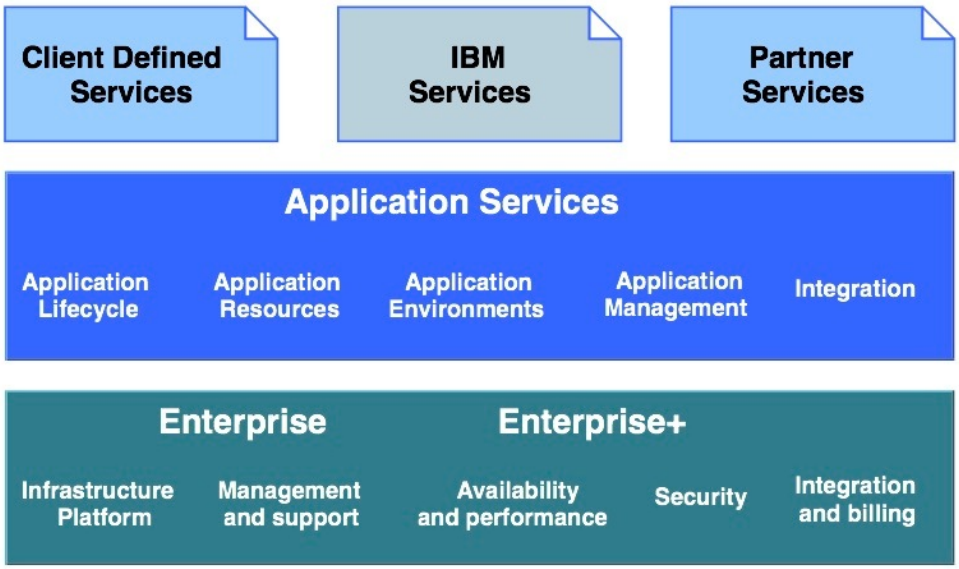


Figure 1
IBM SmartCloud Service

The first two available services are the Collaborative Lifecycle Management and the Workload Service. The Collaborative Lifecycle Management service enables development managers to add users and roles through the SCE portal, thus allowing the monitoring and data availability. It uses the set of development tools of IBM Rational: in particular, the life cycle of the development process is managed through Rational Team Concert, Rational Requirements Composer and Rational Quality Manager. Workload Service provides the ability to adjust the scalability and the management of the applications based on automatic mechanisms defined by policy, through the Workload Deployer, the heart of the service. The offered service can be divided into the five functional areas listed below, which can be used also independently or in a cooperative way.

Application LifeCycle: based on IBM Rational, it includes development and collaboration tools. The development tools the speed up the creation and release of applications, enabling developers to reduce repetitive programming tasks, replacing them with preconfigured templates. The collaboration tools

enable the communication among developers, allowing also non-technical stakeholders to verify the code and keep track of the results. The main objectives of the Application LifeCycle services are to improve agility and innovation in business process, and at the same time improve software quality and reduce errors.

Application Resources: the cloud era has brought to light the idea that IT don't need necessarily to be developed on their own, but companies can benefit from shared services and application packages without losing competitiveness. Application Resources allows cost reduction and simplifies the development of common application resources, such as database services and backup, eliminating the costs related to maintenance and operation, while providing immediate availability and scalability. It provides a centralized and shared service to use common middleware and application resources, such as database as service (based on DB2 Enterprise Server).

Application environments: the service facilitates and automates the deployment of applications through a simplified interface. The user is enabled to select a preconfigured pattern suitable for the type of application to release and to set the appropriate configuration to accomplish the performances required. The Platform is in charge to create and publish the appropriate resources. Integrated control agents keep track of performance, thus enabling automatic management of the resources in order to maintain the required performance levels.

Application Management: the lifeblood of the business for many companies is an enterprise resource planning system such as SAP. In many case the hesitation to move critical systems to the cloud is due to the fear of disruption or malfunction of operations. For these businesses, IBM launched the Application Management Services, focused on the support for third-party applications. The service provides application-specific SAP patterns and configurations that allow users to release directly relevant critical workloads in the cloud.

Integration: to maximize productivity and efficiency, users require access to common components from different applications, thus requiring the publication of a certain number of interfaces. Unfortunately, this integration activity takes time and resources to produce the appropriate code to perform the integration. Integration capabilities of IBM SmartCloud Application Service allow the user you to integrate new and legacy applications without producing additional code, through connectors to existing software and template builders.

2.2. Oracle PaaS

The Oracle Cloud platform [Oracle] is a portfolio of products that can be used to build applications to publish as services on both private and public cloud. The platform is based on the Oracle Grid technologies, as well as on applications that include WebLogic Server, Coherence in-memory datagrid and JRockit JVM. In terms of infrastructure, the platform is based on Oracle IaaS offer that contains Oracle Solaris, Oracle Enterprise Linux and Oracle VM for virtualization, Sun SPARC and Storage. Both IaaS and PaaS services are handled using Oracle Enterprise Manager, which provides an integrated system for the management of the entire development lifecycle of applications.

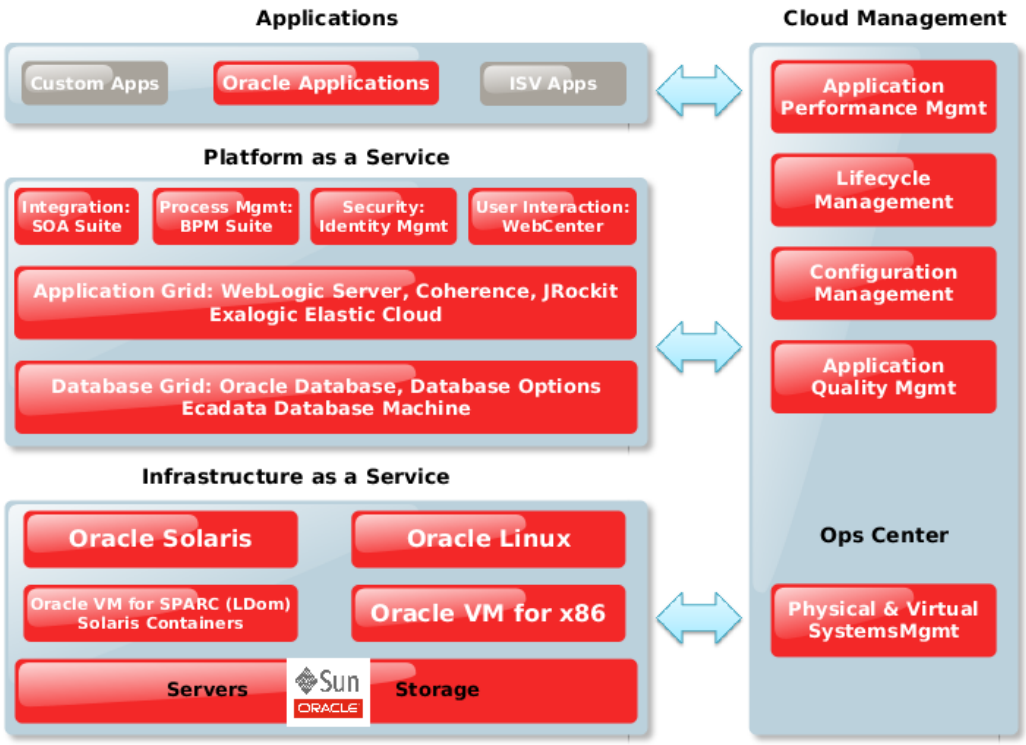


Figure 2:
Oracle PaaS

Below are listed the main services and functionality provided by the Oracle platform.

Storage: Sun Open Storage products combine open source software with industry standard hardware to provide a platform for open and scalable storage. Sun provides virtual networks for large-scale computations through InfiniBand, allowing users to create very large computational grid. The Oracle Exadata Storage Servers provide software intelligence features, with a particular affinity with the Oracle Database.

Server and Operating Systems: Oracle offers several Sun's servers (Netra, Blade Servers, SPARC Enterprise, X64) and support for standard operating systems (Solaris, Linux, Windows), which provide a broad range of optimized physical infrastructure for virtualized and distributed nature of cloud application.

Server Virtualization: Oracle VM provides support for both the x86 and SPARC architecture, making possible the publication of applications in heterogeneous environments. Users can exploit Oracle VM to consolidate servers, release software quickly, recover quickly from system failures and associate ability to workloads.

VM Templates and Assemblies: Oracle VM Templates are virtual machine images that contain enterprise software pre-installed and configured, which can be used to develop, package and deploy applications for faster publication. Oracle VM Templates can speed and simplify application deployments and help reduce the risk of errors in production, development, or test environments. Each VM Template is essentially a software appliance because just like hardware appliances, they are pre-built and very easy to deploy. The next level of this type of application packaging is the concept of VM Assemblies. While software appliances are useful, enterprise applications are not always self-contained, single-VM entities but are sometimes complex, multi-tier applications spanning multiple VMs. There might be multiple VMs in the Web Tier, other VMs in the middle tier, and other VMs in the database tier. There needs to be a way for these multi-VM applications to be packaged for easy deployment. Oracle Virtual Assembly Builder is a tool that takes such a multi-tier, distributed application and packages it up into an assembly that can be reused in a way similar to the way appliances are used. The assembly, like an appliance virtual image, is essentially a file that contains the images of the constituent appliances as well as metadata about appliance configuration, connections, and start up sequence. This technology will be a critical element for creating a library of applications and shared services in either public or private cloud environments.

Database e Storage Grid: Oracle Database has offered grid computing capabilities since the release of Oracle Database 10g in 2003. Since then, Oracle has made continued to enhance the grid capabilities of the database in the areas of clustering with Oracle Real Application Clusters (RAC), storage virtualization and manageability with Automatic Storage Management (ASM) and database performance with In-Memory Database Cache. When lighter weight database services are needed, Oracle Berkeley DB and MySQL are also possible options that are actively developed and supported by Oracle.

Application Grid: similar to the grid architecture in Oracle Database and storage, Oracle Fusion Middleware also supports robust grid functionality in the middle tier with a group of products called Oracle application grid. The key technologies that make up Oracle's application grid are Oracle WebLogic Server as the flagship application server; Oracle Coherence providing in-memory data grid services, JRockit JVM providing lightweight, lightning fast Java runtime environments; and transaction monitoring and management with Oracle Tuxedo.

SOA e Business Process Management (BPM): Oracle SOA Suite provides a comprehensive yet easy-to-use basis for creating the reusable components at the heart of your PaaS private cloud. Rich drag-and-drop SOA component features in JDeveloper and the SCA designer enable rapid creation of components and subsequent composition of those components into applications. Oracle Service Bus provides a simple way to make components available to department application creators using your PaaS cloud. End-to-end instance tracking and Oracle Business Activity Monitoring provide a range of metrics visualizations supporting both the central IT function charged with keeping the PaaS up and running and the departmental application owners concerned with business-level performance indicators. In addition to SOA

components, many enterprises will want to include business process components managed within a unified BPM framework as part of their PaaS. Oracle BPEL Process Manager provides the federation capability to create BPEL process components out of new as well as legacy assets while also supporting the flexibility to enable multiple departments to incorporate PaaS-based BPEL components into their respective workflows.

User Interaction: a centrally-managed library of UI components can give department application owners a great head start in composing their solutions and also gives the central IT function a desirable level of control over consistency across the enterprise's UIs. Oracle WebCenter Suite provides a number of portal and user interaction capabilities that are ideal for creating reusable UI components as part of a PaaS.

Identity Management: Oracle Identity and Access Management Suite provides an ideal facility for managing access and security in a PaaS environment. Oracle Access Manager supports corporate directories and single sign-on. Oracle Entitlements Server provides centralized access control policies for a highly decentralized PaaS environment. Oracle Identity Manager is a best-in-class user provisioning and administration solution that automates the process of adding, updating, and deleting user accounts from applications and directories. Oracle Identity Federation provides a self-contained and flexible multi-protocol federation server that can be rapidly deployed with your existing identity and access management systems.

2.3. Amazon's platform services

Amazon's platform services can be categorized according to their contribution relative to the application architecture

Application Logic-as-a-Service: application logic is typically written by hand in modern programming languages like Ruby, Java, PHP or C#. Each language also has frameworks or libraries that are used to accelerate development. The primary service used to host and execute application logic is Elastic Beanstalk. This service originally focused exclusively on running applications that were written for the Java Virtual Machine and could be executed inside of an Apache Tomcat servlet engine. More recently, the Beanstalk service was extended to support PHP. Beanstalk automatically handles the necessary configurations to launch the software, as well as to ensure scalability, persistence, integration and security.

Database-as-a-Service: Amazon offers three native choices for databases each with their own advantages and disadvantages. The earliest offering was SimpleDB. This solution was introduced as a simple way to store information persistently by using key/value pairs. SimpleDB's claim-to-fame is that it really is easy to use, mostly because it doesn't have many of the more complicated features developers have come to expect in database management systems, whilst providing automatic backup services and scalability of resources. Although SimpleDB satisfied many needs, most business applications used a relational database. Amazon responded with Relational Database Service (RDS). Unlike SimpleDB, RDS is not an encapsulated horizontally scaling system as this would require significant changes to the underlying database engines.

Instead, RDS gives the users the ability to self-service provision a database and configure it to their needs. The service currently supports most of the popular editions and versions of MySQL and Oracle. Users can specify specific configuration settings for their database including the size of the machine (CPUs and Memory), backup & restore options, the ability to auto-patch the database engine, the publishing of monitoring data and high availability features like the auto-recovery of a database system in a remote data center if the original went down. The third database service offered by Amazon is DynamoDB. This offering is considered a "NoSQL" database, which means that it doesn't rely on SQL for data definition or for data manipulation. Instead, DynamoDB offers a schema-less database management system. Many view this offering as a replacement for SimpleDB because it has a superset of the functionality while being delivered in the same encapsulated, horizontally scalable manner.

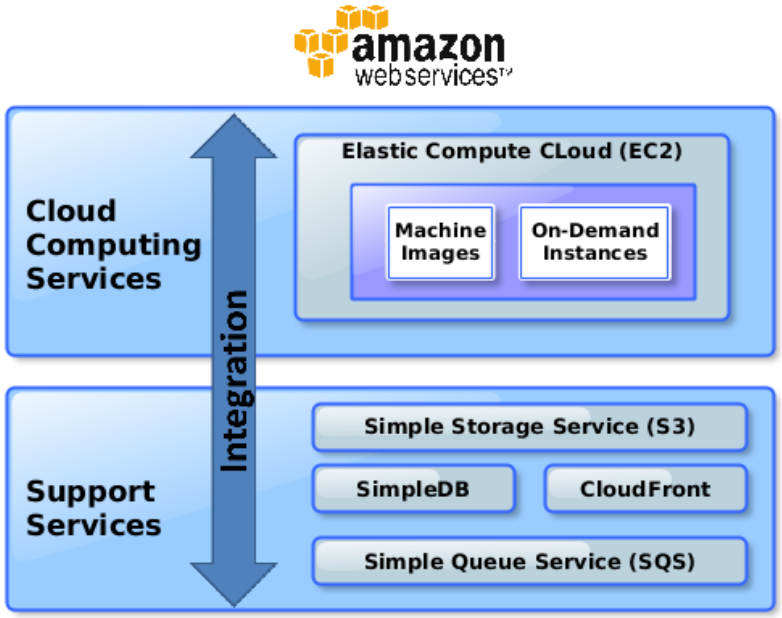


Figure 3
Amazon Application Architecture

Caching-as-a-Service: Amazon offers an implementation of a clustered cache by wrapping one of the most popular open source solutions, Memcached. Users are able to launch a cache via self-service provisioning (API or portal). The Elasticache service offers the ability to associate the caching software with various types of EC2 compute services, allowing the users to manage the scalability of cache resources automatically or based on alarms generated by the service itself.

Integration-as-a-Service: Amazon Web Services currently offers two types of integration services for system-to-system decoupling and messaging. At this time, there is no mechanism to do payload transformations or protocol

mediation. The current services are Simple Notification Service (SNS) and Simple Queue Service (SQS). Simple Notification Service provides pub/sub (publication/subscription) capabilities inside the AWS cloud. The service is an encapsulated, horizontally scalable offering. Developers can call the service via SOAP- or REST-based commands and they specify their delivery protocol of choice. After a message has been placed on a topic, the SNS service sends the message to all subscribers. In its current state, SNS does not offer guaranteed delivery notification by confirming receipt of individual messages, nor does it provide guarantees on the timeliness of delivery. A second integration service offered by AWS is Simple Queue Service, which allows developers to separate two modules from a load-over-time perspective. The message queue enables the communication among different modules which work at different speeds. Using the SQS service, developers can use either a SOAP or RESTful interface to create, delete and inspect queues as well as to add or remove items from a queue. AWS has chosen not to implement advanced queuing capabilities like FIFO (first-in first-out) or priority queues. The assumption is that if users want these features, they will extend the core service to include finer grained management of message arrival and departures.

2.4. Windows Azure execution models

Windows Azure [Redkar et al, 2011] is Microsoft's application platform for the public cloud. This platform can be used in many different ways depending on the developers' needs, covering a wide range of use. One of the simplest is to build a web application that runs and stores its data in Microsoft datacenters, even if the platform can be used just to store data, with the applications that use this data running on-premises (that is, outside the public cloud). Additionally Windows Azure can be used to create virtual machines for development and test or to run SharePoint and other applications.

Windows Azure provides different execution models for applications. Running application on virtual machines is essentially the approach IaaS of Windows Azure, which allows the user to select operating systems from a catalog or upload its own Virtual Hard Drive. This execution model is the less transparent since the management of the development environment and the responsibility for the administration of virtual machines lies to the user.

If the user wants to develop applications or websites and delegate to the provider the care of the administrative part, the solution is the use of Windows Azure Web Site, which offers both a management portal and API. Once a web site is running, the user can add or remove instances dynamically, relying on Windows Azure Web Sites to load balance requests across them. However this solution has many limitations and is not suitable for applications that require, as instance, the installation of arbitrary software.

A solution that offers a good flexibility, still remaining transparent with respect to aspects of reliability and administration, is Windows Azure Cloud Service. This model represents the real platform solution of Windows Azure, allowing the execution of applications and offering support for scalability, reliability, and without caring about administrative details.

An application developed for the Windows Azure platform is linked to two main types of entities somehow combined together: Instances of *web role* (intended primarily for running Web-based applications) and *worker role* (designed to perform a very large amount of computation, such as simulations or testing or to perform a service in background).

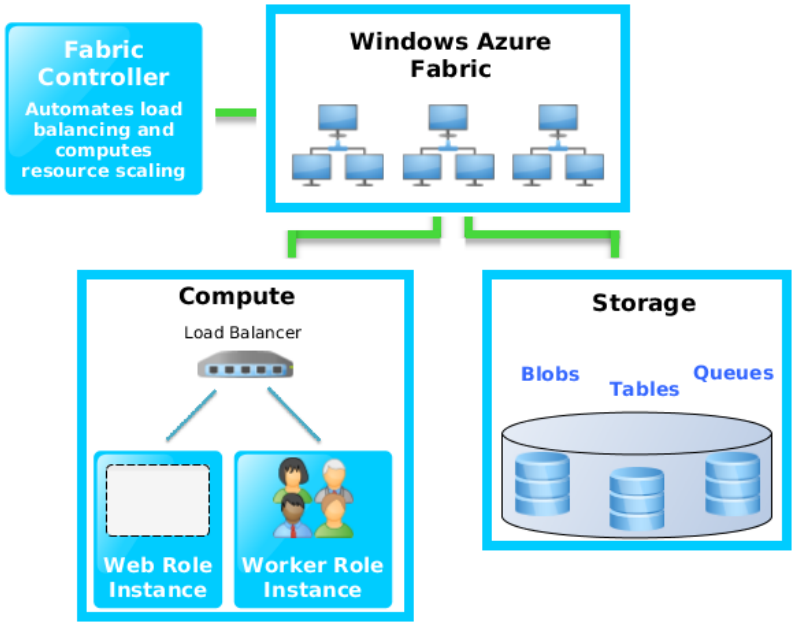


Figure 4
Application architecture in Windows Azure

An application developed for the Windows Azure platform is linked to two main types of entities somehow combined together: Instances of *web role* (intended primarily for running Web-based applications) and *worker role* (designed to perform a very large amount of computation, such as simulations or testing or to perform a service in background).

An application developed with the cloud service execution model is made available generally in a two-step process: first the developer develops and loads the application with the necessary files on the platform and then using the portal runs the application. This division between production and execution is useful for version upgrade, because the new version is updated without service interruption.

The monitoring service of the platform is able to detect failures not only hardware but also application failures and rerun instances of the worker role and web roles failed. Due to the platform characteristic the entire application must be designed and implemented so that it can continue the execution correctly even if any instance of a worker role or web role fails. To ensure this, the application need to remember its state explicitly in the Azure data store, because the state cannot be saved persistently on virtual machines on which the services are running, due to the nature of the service platform. The fulfillment of this constraint makes

applications running on Azure platform easily scalable and resilient to failures. The scalability of applications can be easily operated manually or by setting parameters which drive the scaling of web roles and worker roles based on the workload, such as CPU usage or the status of queues.

Applications need data, and different kinds of applications need different kinds of data. Because of this, Windows Azure provides several different ways to store and manage data. The use of an execution model based on IaaS allows the installation and the use of any type of system, from relational databases to NoSQL solutions. To facilitate the management of the databases Windows Azure offer also for the data store a platform service, in particular SQL Azure for relational database, Tables for scalable NoSQL database and Blob to store unstructured binary data. Each of the three options addresses a different need: relational storage, fast access to potentially large amounts of simple typed data, and unstructured binary storage. In all three cases, data is automatically replicated across three different computers in a Windows Azure datacenter to provide high availability. It's also worth pointing out that all three options can be accessed either by Windows Azure applications or by applications running elsewhere.

Another key point for the execution of applications is the interaction among different components through messaging services. Windows Azure provides two mechanisms for messaging: the first, based on queue (Azure Queue), offers a paradigm publisher / subscriber; the second and more complex, is called Service Bus and allows (in addition to one-to-one and publish-subscribe) to implement mechanisms of subscription to the topic as well as offering other features related to communication security.

The following table presents a comparison among the four cloud platforms.

	IBM	Oracle	Amazon	Azure
Application Development	• Java EE e .NET	• Java EE • SERVLETS, JSF, JPA, EJB, JAX-WS • WebLogic Server.	• AWS SDK for Android, iOS, Java, .NET, Node.js, PHP, Ruby, Python (boto)	•Azure SDK for Python, Java, node.JS and .NET.
Application Life Cycle Management	• Application Life Cycle Management • IBM Rational Requirements Composer • IBM Rational Quality Manager • IBM Rational Design Management • IBM Rational Team Concert	• Jdeveloper Eclipse Net Beans.	• Amazon OpsWorks • Elastic Beanstalk	• Team Foundation Service

	IBM	Oracle	Amazon	Azure
Data Management	• Database As A Service based on DB2Enterprise Edition Server	• Oracle Real Application Clusters • Oracle In-Memory Database Cache • Oracle Berkeley DB	• SimpleDB • Relational Database Service (RDS) • DynamoDB	• SQL Azure • Table • Blob

Table 1
Comparison of IBM, Oracle, Amazon and Microsoft Offers

2. Cloud Patterns

In the current Cloud Computing scenario there exist so many alternatives, services and offers, provided by different vendors, that it can be really difficult to decide which solution or service to choose and exploit in a specific application context. Promoting Cloud Patterns becomes a necessity, born from the need to provide both general and specific solutions to recurring problems in the definition of architectures for Cloud applications. While Design patterns deal with problems related to different aspects of software development, being the structural view of the application one of this aspects, most of the existing Cloud Patterns focus on the architecture of the Cloud solution: this has lead, in many cases, to the development of platform dependent patterns, which can be applied only to a specific platform offered by a specific vendor.

The various existing Cloud Patterns catalogs, given their different nature and objectives, deliver content at different levels of detail and abstraction. Some Patterns, like those presented in [AWSP] e [WAP] are tied to a specific Cloud platform, thus being more detailed in terms of the components they rely on for the implementation, but the solution they provide is strictly dependent on the reference platform and has very poor flexibility.

Catalogs developed during academicals studies and efforts, like those defined in [Fehling et al, 2011], are not tied to industrial solutions: they describe general functionalities and behaviors, and they propose architectural models which are much less bound to specific Cloud platforms, thus resulting in less details and better flexibility.

According to a definition provided by IBM experts [IBMP2], a Pattern is a collection of elements describing a complete and fully functional software solution, which can involve different inter-connected systems or a single entity. All the knowledge needed to create, configure and support every aspect of the solution is already included in the Pattern. In IBM, a Virtual Application Pattern represents a collection of applicative components, behavioral politics and relative links. A workload service which takes advantage of this kind of application can automatically build the infrastructure and middleware resources needed to operate and manage the virtual application, which thus becomes an instance of the pattern. The use of Virtual Application Patterns takes place

through the services offered by a platform, provided by IBM itself, known as Workload Deployer, which can be accessed through both the IBM PureApplication System (Private Cloud) and the Smart Cloud Application Workload Service (Public Cloud), based on the IaaS platform provided by the IBM Smart Cloud Enterprise .

Oracle followed a different direction in the definition of their Patterns: instead of providing already configured solutions, they defined a catalog known as Oracle Fusion Applications User Experience (UX) Design Patterns [Oraclep]. This catalog represents a set containing more than 150 models of interaction between users and data flows, which can be directly applied to Oracle Fusion Applications or used different environments. Oracle are exposes its different pattern according to a common template, which summarizes: the problem addressed by the Pattern and its main objectives, the specific context in which the Pattern is applied, one or more practical solutions to the problem. In particular the template shows design examples that represent possible implementations of the pattern, generally referring to the Fusion middleware, even if their solution is quite general. An internal search engine, which also benefits from an interactive graphical interface, can be used to navigate through the catalog.

The AWS Cloud Design Pattern [AWSP] is a collection of solutions and design ideas that use the AWS Cloud technologies to solve common design problems. These solutions are strongly connected to the AWS platform, so they are particularly well detailed and supported by precise diagrams which describe structure and interactions of the Pattern efficiently. For each Pattern there is a description that indicates the problems leading to the creation of the Pattern and what difficulties can be resolved through its implementation. The implementation itself is described precisely, defining step by step the procedure of application of the Pattern within the AWS, specifying which components should be used and how this should be done.

The set of Patterns provided by Windows Azure [WAP] cannot be considered as a real catalog, but rather as a collection of links to articles or digital publications that describe Cloud Patterns, or that illustrate techniques, based on the services offered by the very Azure platform, for the management of particular Cloud functionalities. The Patterns presented by Microsoft are all strictly bound to the possible implementations on the Azure platform, to the point that their description is presented directly through the service, functionality or tools, provided by Azure itself, that solve the problem addressed by the Pattern.

The vision of Patterns of the various providers is therefore different: providers like IBM see Patterns as templates of applications which can be personalized through policies and constraints, on the basis of existing pre-configured solutions; Oracle patterns in general are not really meant for the cloud, but come from collections of recurrent solutions applied in the Oracle Fusion Applications ; Amazon offers a catalog of Patterns that are expressed in terms of proprietary services which can be difficult to generalize to other Cloud platforms; finally Microsoft describes Pattern as solutions to problems that can be they are already solved by their platform in a transparent way .

3. Conclusions and future work

From the brief overview of the PaaS services we have just provided, it is clear how vast and varied the offers of different Cloud vendors are. This diversity is reflected in the profound differences between the various Cloud Patterns' catalog that, if in some cases are neutral with respect to existing platforms, placing themselves at a rather high level of abstraction, in others are deeply tied to them. Future work could focus on the formalization, through a uniform and shareable language, of both proprietary and non-proprietary Cloud Patterns, as well as on the definition of a more general class of Patterns, which could comprehend both categories, in order to make it easier to trace solutions applicable to specific problems.

4. Acknowledgments

This work has been supported by PRIST 2009, “Fruizione assistita e context aware di siti archeologici complessi mediante dispositivi mobili”.

References

[AWS] An Overview of the Amazon PaaS. Transcend Computing. 2012.

[AWSP] “Aws cloud design patterns”, http://en.clouddesignpattern.org.

[Fehling et al, 2011] Fehling, C.; Leymann, F.; Mietzner, R.; Schupeck, W. A Collection of Patterns for Cloud Types, Cloud Service Models, and Cloud-based Application Architectures. Technical Report No. 2011/05; University of Stuttgart: Stuttgart, Germany, 2011.

[IBMP] IBM Design Patterns, www-01.ibm.com/software/ucd/designpatterns.html.

[IBMP2] http://expertintegratedsystemsblog.com/index.php/2012/07/getting-back-to-the-basics-what-is-a-pattern/.

[Oracle] George Demarest, Rex Wang : Oracle Cloud Computing, Oracle White Paper, http://www.oracle.com/us/technologies/cloud/oracle-cloud-computing-wp-076373.pdf.

[Oraclep] Oracle Fusion Applications User Experience Patterns and Guidelines, http://www.oracle.com/webfolder/ux/applications/fusiongps/patterns/index.htm.

[Redkar et al, 2011] Redkar, Tejaswi, Guidici. Windows Azure Platform. Apress,2011.

[SCAS] http://www.ibm.com/cloud-computing/us/en/paas.html.

[WAP] Windows Azure Architecture and Patterns, http://www.windowsazure.com/en-us/develop/net/architecture/.

Biographies

Beniamino Di Martino is full professor of Information Systems at the Second University of Naples (Italy). He participated to various research projects supported by national and international organizations, with role of Unit Responsible. He is author of 8 international books and more than 200 publications in international journals and conferences. He is Editor of two international journals and editorial board member of many international journals. His research interests include: Knowledge Discovery and Management, Semantic Web and Semantic Web Services, Semantic based Information Retrieval, Cloud Computing, High Performance Computing and Architectures, Mobile and Intelligent Agents and Mobile Computing, Reverse Engineering.
email: beniamino.dimartino@unina.it

Giuseppina Cretella is a PhD Student of Computer and Electronic Engineering at the Department of Industrial and Information Engineering at the Second University of Naples. She received her Master Degree in Computer Engineering in 2011. She is involved in research activities dealing with Semantic Web and Semantic Web Services, Knowledge Discovery, Reverse Engineering and Cloud Computing. She participated in research projects supported by international and national organizations, such as: mOSAIC Cloud FP7 project, CoSSMic Smart Cities FP7 and Cloud@Home.
email: giuseppina.cretella@unina2.it

Antonio Esposito graduated in July 2013 with a master thesis on the recognition of Design Patterns from UML documentation through semantic approaches. Currently he is a Ph.D. student at the Second University of Naples and he is working on extending his previous research to analyse source code for patterns recognition. His main interests are Software Engineering, Cloud Computing, Design and Cloud Patterns, Semantic based Information Retrieval.such strategies constitute basic drivers of corporate action, where competitiveness is based on the capacity to conquer and foster ready, exigent and innovative demand.
email: antonio.esposito7@studenti.unina2.it

The European Strategy for Cloud Computing: Harmonization of Technical and Legal Rules

Caterina Flick, Vincenzo Ambriola

Abstract. *Cloud computing is an emerging technology that aims to reduce the cost of software services and resources. The basic idea is to move computational power and data storage from the client to the server, thus improving quality of service, enhancing safety levels, and allowing a better allocation of resources. The widespread diffusion of Internet and the rapid acceleration of the use of smart mobile devices have shown that cloud computing solutions are not only possible but somehow inevitable. In this new scenario the European Union has developed a strategy for adopting cloud computing and exploiting its potential. In this paper we present the theme that characterizes this strategy: harmonization of technical and legal rules. The focus on legal rules is of paramount importance for the profound impact that cloud computing can have on our society. We will discuss the following aspects related to data stored in the cloud: access and integrity, property, storage and transfer. Particular attention will be devoted to the adoption of cloud computing in the public administration.*

Keywords: Cloud computing European Union strategy, standards and rules, public administration.

1. Introduction

Cloud computing is the centralization of infrastructures, platforms, and programs and their redistribution to end-users through Internet. The centralization of data repositories and processes for their provision allows such a scale economy that even large companies alone cannot achieve. For this reason, the adoption of cloud computing can lead to substantial savings in IT budgets, and pave the way for the solution of technical and economic problems related to existing IT systems.

Although one of the fastest growing IT industries in the world, cloud computing is a technology almost unknown to the majority of European citizens: less than a quarter of them, in fact, uses cloud services. In addition, its use is much more common for personal needs than for business reasons.For many companies, on the other hand, the transition to cloud computing is perceived as an important opportunity to reduce the costs of IT infrastructures. The anticipated savings would mainly derive from the ability to purchase computing resources and

services from third parties, without having to purchase and maintain expensive and complex IT systems. According to a study commissioned by Microsoft [Forum PA, 2012], the revenues from cloud computing business may be high (832 billion Euro over the next three years) and new jobs could have a significant impact on employment (152,000 in Italy, with an increase of 125%).

The sustainability of cloud computing depends on simplification and harmonization, both in technical and legal terms: tools easily accessible, flexible and affordable contractual instruments, full control of data (for reasons of security and privacy), effective exercise of rights.

2. The European strategy for cloud computing

The strategy adopted by the European Commission on September 27, 2012 has the aim to accelerate and increase the use of cloud computing. This strategy is the result of a wide-ranging analysis of the political aspects, regulations, and technology in the member countries, followed by an extensive consultation aimed at identifying all the potential offered by cloud computing. The strategy is divided in three main objectives.

The first objective aims to simplify and eliminate the "jungle" of technical standards, with the aim of allowing users to enjoy interoperability, data portability and reversibility. The Commission will work with the support of ENISA and other organizations to provide assistance to the development of voluntary certification schemes, providing a list by 2014 (the list of high-level criteria and the first list of certification schemes has been released by CERT-SIG on November 2013). The second objective is to develop security models and conditions of use that are accessible (fair), uniform, and easy to apply and interpret. The third objective is to identify a European partnership to drive innovation and growth of the public sector. This will be achieved by bringing together experts from industry and the public sectors that will work on the definition of common requirements for the purchase of cloud computing services, in an open and fully transparent way. The public sector, in particular, has a key role in the development of the cloud computing market. In the presence of a fragmented market, in fact, this requirement will have a minimal impact on the public sector, since the integration of services would be low and the citizens would not get the best results at the lowest cost (best quality/price ratio).

Neelie Kroes, European Commissioner for the Digital Agenda, has repeatedly stressed the importance of cloud computing for European economic growth and announced a European Cloud Partnership (with an initial investment of 10 million Euro) to start building a solid foundation for the joint procurement of cloud by the public authorities. According to data provided by the Commission, this opportunity can generate 800,000 jobs and economic benefits for more than 200 million Euro. The indirect benefits induced by the adoption of cloud computing are endless, especially for public services, which can be made more efficient and affordable, with great benefits for the European population.

The existence of a digital market, globally integrated and cohesive, will give users the freedom to choose between different services. In addition, the

European service providers can take advantage of these opportunities to grow even beyond the European borders. In the coming months, the European Commission should indicate a harmonized regulatory framework, which aims to promote the use of cloud computing in enterprises and public administrations. The difficulty in finding such a framework depends on the multiplicity of visions and different laws between member states, in particular on important issues related to security, privacy, and transparency.

Among others, the European Commission should focus on legal issues involving privacy, data retention, applicable law, liability, and consumer protection. Other important issues that need be taken into account are those related to interoperability, standardization, data portability. As regards the public sector, in addition, there is a need to define precise rules for choosing suppliers and entrusting the management of services.

3. Simplification of technical standards

The simplification of technical standards is intended to provide users with interoperability, data portability, and reversibility. A change of this kind, however, must be accompanied by proper regulation of the rules of intellectual property and licensing. The existence of a multiplicity of technical standards, combined with the use of rigid rules on intellectual property, may in fact lead to restrictive effects on competition, either by preventing the lowering of prices, limiting or controlling production, market, innovation, and technological development.

First, if cloud computing producers are involved in a conflict in the context of anti-competitive standards there is a risk that this will lead to less price competition, facilitating agreements between parties who are outside the market.

Secondly, the identification of technical standards (often very detailed) for a product or service may limit development and technological innovation. At the moment in which a standard is being developed, alternative technologies can compete for inclusion in the standard. Once a technology has been chosen and the standard has been identified, different technologies and manufacturers may face unfair restrictions and could potentially be excluded from the market.

Thirdly, standardization can lead to results which are contrary to competition, where they place limits on some players with respect to the capacity of effective access to the results of the process that led to the identification of the standard.

In recent years, the European Commission has revised the rules for horizontal cooperation agreements in the light of the European competition law, including the guidelines on the implementation of Article 101 of the Treaty on the Functioning of the European Union. The Commission has also addressed the issue of licenses for ICT and finally, in November 2012, the implementation of Open Source standards. A working group for the detection of certification schemes started to work in February 2013.

Simplification requires special attention. For portability services, simplification is the ability to seamlessly migrate applications, virtual machines, and data from a cloud computing environment to another. To make this possible it is necessary that the two environments (departure and arrival) be highly interoperable,

because the same application environment can be replicated by different suppliers. Interoperability, however, is also the opportunity to share the same management tools, virtual machines, and other resources, including a plurality of service providers and cloud computing platforms. Portability should make data and programs "understandable" even by a receiving system, made available by another cloud service provider, regardless of the specific characteristics of the hardware and software platforms used.

In late 2013 the CERT-SIG (ENISA) elaborated a list of guiding principles for certification schemes for cloud providers. These principles are connected with – *inter alia* – technological neutrality, global standards and affordability. The definitive list of schemes should be released on 2014.

4. Consistent, secure, and simple legal rules

During the European Conference on Cloud Computing, which was held on March 7, 2013 in Brussels, was raised the question of defining a legal framework to create a market for cloud computing services. The obstacles to a uniform regulation of the terms and conditions of the contract for the provision of these services derive, in fact, from the differences in this area between national laws. It is therefore necessary to develop a model of contractual terms that cover matters not governed by European legislation in the field of contracts, such as data retention after contract expiry, access and integrity of data, location and data transfer, data ownership, direct and indirect responsibility in the management of cloud computing service from suppliers and sub-contractors.

5. Data access and data integrity

The availability of data, and accessibility at any time, requires the assurance of quality standard of the connectivity provided by Internet. Without this quality requirement, cloud computing services may be degraded by traffic peaks, or even made unavailable by abnormal events (failures, for example). This aspect depends only in part by the cloud computing providers, as it involves the highest level of political bodies and government, responsible for the organization, the diffusion, and the management of Internet.

The accessibility of data requires to preserve its integrity, with an explicit focus on deletion or damage. It is also necessary to take appropriate security measures to protect data confidentiality, especially with regard to visibility or use by unauthorized persons. In a nutshell, from the legal point of view, it is necessary to refer to the rules on privacy, or to the Legislative Decree no. 196/03 and a number of European directives, in particular Directive 95/46/EC and the recently adopted Regulation 611/2013, in force since August 25, 2013, on mandatory reporting of violations of privacy.

Data confidentiality largely depends on the security mechanisms used by the supplier. This technical aspect, however, does not relieve public and private entities that use a cloud computing service to manage their information assets or, even more, data and information on behalf of third parties, from their responsibilities. It is implicitly stated that they must take appropriate measures to

ensure the security of data and information handled by these services. In the case of treatment of personal data, the entity in charge of a cloud infrastructure continues to be responsible for the adoption of security measures, as it holds the position of "owner" of the treatment. At the same time, all the obligations related to information security associated with the commission of crimes during the execution of a service, which lead to a direct responsibility in charge of providing the service, are part of the measures envisaged by the 2001 European Convention on Cybercrime.

The provider of a cloud computing service must guarantee data security in a transparent manner, taking care of correct data transmission and storage, including the adoption of safe and regular back-ups. Technical and organizational infrastructure aspects of a cloud computing framework, including its overall design, the development process of provided services, the configuration of the transmission systems, the adoption of specific contracts with users and sub-contractors, the systems that control access requests, have an essential role in the objective of ensuring data security.

The need to protect data from intruders and unauthorized uses imposes a limit to the amount of information about users, traffic, property. This could be in conflict with the need both to make data rapidly and continuously available to users, and to react quickly in an emergency. This requirement is also in contrast of the need for confidentiality (or rather the interest in not being controlled) of those who are authorized to access. In case of system faults, in fact, the accuracy and timeliness of diagnosis and remedy are closely related to the completeness of the available information. In addition, as many faults and anomalies involve multiple vendors, the resolution of a problem may require cooperation between several parties, potentially residing in different states.

In summary, a system that safeguards data security and confidentiality has two objectives: on the one hand it must be able to detect an attack as quickly and accurately as possible and should react just as quickly, to maintain the levels of quality of services provided, on the other it must ensure the confidentiality of the users involved and whose actions are monitored. A reduction of user data can increase the level of confidence, ensuring infrastructure recovery operations. Similarly, increasing the available data, a high level of protection can be easier to accomplish. A solution that that solves this dilemma is based on the use of advanced encryption techniques.

6. Data property

Moving data to a cloud service provider requires to clearly define the notion of data. This is an issue that directly involves both intellectual rights and industrial property, with some complications that arise from the difficulty of identifying the place where data are stored.

With particular reference to creative works, still ruled by the Italian law on copyright 633/1941 which reflects only in part the characteristics of Internet, it is imperative to point out that dematerialization of media storage and preservation of these works can raise new and relevant questions. When creative works are

transmitted to a cloud computing provider (on a physical media that he owns) the way these works are used immediately changes, since they could be made available to multiple users through an on-demand access from multiple locations, with the permission of the author, but also to his knowledge.

On the other hand, the concentration of data in the hands of individual market participants, who are not involved in providing creative works on Internet but that simply allow their diffusion, requires, in order to protect the author rights, to break down the evident state of irresponsibility of the service provider. Without taking into account this aspect, service providers would not respond for content uploaded by users. For the author of a creative work would be difficult and expensive to identify those who, in theory, have violated his rights (an activity made even more complex by coordination with the legislation on privacy). The conflict between different interests has been brought in an Italian court, with the so called *Peppermint versus others*. In this case, the argument was the protection of copyright, owned by Peppermint, against the protection of the privacy of Internet users.

In the absence of regulatory changes, including the obligation on a provider of telecommunications services, there is only a duty to inform the supervising judicial or administrative authority, in relation to reliable information received in respect of the infringement of copyright accomplished through telecommunications network. There is no obligation to terminate the service provided by providers, even when there is evidence of unlawful acts whose effect is to make available some content that infringes copyright.

Another issue of great importance about data ownership reflects the requirements related to personal data processing, both during the contractual relationship and at its termination, in relation to the compulsory destruction of data held by the supplier. In general, it is sufficient to say that the principles established for the protection of personal data must be adapted for the use of cloud computing services, i.e. taking into account the problems associated with the peculiarities of such systems. More specifically, it is necessary to address the issue of the obligations in charge of the supplier, treating this actor as someone other than the owner of data, and appointed by him responsible, with respect to the discretionary powers resulting from the treatment of the data. The issue needs to be addressed with particular attention when the data entrusted by the user of the infrastructure provider are related to third parties: this is the case, for example, of public administrations, which collect citizens' data, as well as those of its employees.

7. Location and data transfer

The theme of the placement and transfer of data is of utmost importance, especially when computing resources are physically allocated to countries located outside of the boundaries of the European Union.

Community legislation regarding the protection of personal data - already mentioned above - allows the transmission of data to a third country only if it provides an adequate level of protection (possibly with the consent of the

national authority that will evaluate the adequacy of the protection that the parties are preparing to ensure on the basis of negotiated agreements agreed, in accordance with current European legislation on the processing of personal data). The controls and formalities that must precede data transfer will be easier if there is a direct relationship between the user and the supplier or sub-supplier, resident outside the European Union. They will be more complex when multiple parties are involved in the provision of cloud computing services.

Another aspect that deserves attention is related to the possibility that large amount of data, provided by different owners with different interests, are treated by a limited number of multinational companies, with the possible risk of commingling and conflicts (the so-called big data).

8. Direct and indirect responsibility of suppliers and sub-contractors

The importance of regulating in detail the terms and conditions of the service is of utmost importance for the role that different actors play in the adoption of cloud computing services: service provider, providers of Internet access. Each of them need to receive adequate information and contractual guarantees, on the quality parameters of the service provided.

During the migration to cloud computing services, in fact, the user becomes completely dependent on the adequacy of the quality level of the suppliers. Each (even temporarily) unavailability or inefficiency of the services can have a significant negative impact and result not only in economic losses, but also in considerable damage to the user image. Accordingly, it is essential to introduce contractual clauses that provide for the payment of compensation, which describe, with the utmost precision, the performance expected by the user and to clarify how certain benefits are of crucial interest.

In general, providers of cloud computing services are similar to other service providers, whose obligations are essentially based on the Legislative Decree no. 70/2003 governing electronic commerce, in the transposition of the European Directive 2000/31/EC and, with regard to consumer protection, the Legislative Decree no. 206/2005, also involving the community. Also in general, specific measures must be taken into account as the provision of on-line services is a complex contractual operation, divided in two phases: the signing of the contract and subsequent execution. The contract, in fact, will evolve throughout the duration of the relationship and also after its termination. The supplier's obligations depend on the type of services offered and the activities carried out under the contract.

Liability arising from the provision of digital preservation must be provided by an analysis of many aspects that are reflected in the terms of contractual responsibilities placed in the hands of various actors: the service provider, any intermediaries (which contribute to the delivery of the final service), the role responsible for the preservation of data (for example, the legal entity which has been delegated parts of the process, including those relating to information storage in the cloud computing infrastructure). Storing data in different

geographical locations may have implications for all these actors of the applicable law in the event of a dispute between the data owner and the supplier, and in relation to specific national law governing data treatment, storage, and security. Therefore, in order to properly manage the contract with the cloud computing provider of services related to digital preservation it is crucial to apply the concept of "intellectual interoperability" between lawyers, archivists, and those who, as managers (of internal or external organizations), supervise the process adopted for the preservation of digital documents.

A good contract for the provision of cloud services (and digital preservation of documents) must therefore be the result of the application of the rules governing the liability of the supplier and the processes of safety and protection of the confidentiality of information, stating how and by whom data safety will be ensured. When establishing contractual provisions, the parties must also take into account all the obligations required by specific guidelines. For example, there are specific obligations that define the irresponsibility of senior roles with regard to the commission of crimes, through the provision of a suitable organizational model and the adoption of specific procedures for the optimization of the service, to be shared with the supplier. The contract will have to consider and provide for the application of international and ISO standards.

The contracts may have specific clauses to ensure confidentiality obligations, whose violation involves the payment of penalties. To prevent unauthorized or not allowed access it is also necessary to plan the use of both encryption, for data subject to transit operations, and adequate authentication systems, in order to guarantee with certainty the identity of those entitled to access data. The issue of cross-border data flows requires to set at least some minimum guarantees, in order to meet the requirements of Community law.

Data portability and interoperability of infrastructure with the computing resources of the users should be among the requirements to be accepted and respected by the supplier. To define any liability profiles of the entity that manages the data, it is necessary to clarify the exact legal nature of provider of cloud computing services. Even in the case of limited autonomy, the supplier should not be considered a controller but a processor, especially when the methods of data management are agreed between the user and the supplier through specific clauses.

A service provider is responsible for the management and its actions are limited to only certain data. In general, the provider does not have specific and appropriate skills to play a predominant role in their treatment. Like controllers, the provider maintains, however, autonomy and responsibility on traffic data related to the circulation of information, infrastructure or to the user's computing resources.

The division of responsibilities between the supplier and the user cannot be rigidly fixed. Existing models of cloud computing services can be easily integrated with each other, either when the supplier provides an integrated service, or when multiple providers compete to offer a complete service. Suffice it to say, for example, to the case where the provided service involves

processing and data management (storage, copying, transmission of data to third parties) where one has to distinguish between the operations of data processing carried out directly by the computer and those established by a human operator. Only in this case it is possible to recognize responsibility profiles, which affect other transactions related to the service provider.

Particular attention should be paid in the event that the supplier is entitled to honor the services contract with third parties, even if only for the management and the allocation of physical resources on which data resides. In this case, provision should be made in the contract specifications warranties relating to the sub-contract, requiring the supplier to notify this decision to the user.

The law applicable to the provision of cloud computing services is rooted to the physical place where the supplier is established. If the offender is a company based in the European Union this does not create any obstacle, as it applies to Community law. However, the problems on the legislation applicable in the event of a dispute between the service provider and the user can be included in more complex terms, the occurrence of situations in which the service provider interacts with other entities.

In a cloud computing infrastructure data are often stored in different data centers, which can be physically located in different countries. The service provider may also use third parties to exchange computational resources (for example, if the supplier does not have enough available capacity in terms of storage media and relies on others as well). Exchanges between several parties determine a continuous data flow, making it difficult to identify who handles them at any given time, nor their exact location. The user is then limited to access to the service and the provider will be in charge to retrieve his data. A plurality of contractual relationships also occur when the supplier provides a service which, in turn, is obtained from other suppliers, always in the cloud infrastructure.

Therefore, since different subjects may intervene in the management of the infrastructure it is necessary to regulate the cases where the supplier uses third-party vendors who do not reside in the European Union. There are many possible solutions ranging from (a) the use of Community clauses type between the user of the services and the sub-supplier, (b) the warrant issued by the user to the supplier, so that the latter directly enters into agreements with the sub-supplier, (c) the provision of specific contractual arrangements between the parties.

The signing a contract cannot, however, be sufficient for the user to be risk-free, since suppliers can contract certain activities (including management of physical resources that hold data) and can in turn be subject to corporate events (mergers and acquisitions) which lead to significant changes, such as, for example, the registered office and, consequently, the applicable regulations. In summary, even after signing a valid contract, it may be difficult for the user to request compliance with the obligations that have been covered, especially when the supplier does not reside in the same country.

9. Conclusions

The technological frontier moves quickly, changing lifestyles and assumptions previously considered absolute. In the case of cloud computing, the effects are still visible and not perceived. As citizens, we take for granted many of the services offered on Internet (home banking, reservation systems and online purchase of tickets, e-mail, remote data storage) without knowing that they rely on cloud infrastructure. More and more our society depends on these technologies.

The different speed of adjustment of the legal body of rules to the technological progress, increases the risk (but often the certainty) that the introduction of new services is not accompanied by a sound legislation that protects the interests of consumers but also of those who have invested in technology. Perhaps this is the novelty of the third millennium: a world in which the law must pursue reality and not vice-versa.

References

Bollier D., The promise and peril of big data, The Aspen Institute 2010.

Coleman N., Borrett M., Cloud security, who do you trust?, IBM 2010.

Data Protection Working Party – Art. 29, Working Document 1/2009 on pre-trial discovery for cross border civil litigation, 2009.

DigitPA, Recommendations and proposals on the use of cloud computing in the public administration (Raccomandazioni e proposte sull'utilizzo del cloud computing nella Pubblica Amministrazione, in Italian), ver. 2.0, 2012.

ENISA, Certification in the EU Cloud Strategy, 2013.

ENISA, Priorities for research on current and emerging network technologies, 2012.

ENISA, Security and resilience in governmental clouds: Making an informed decision, 2011.

Enter the Cloud, Cloud survey 2013: The state of cloud computing in Italy (Cloud survey 2013: lo stato del cloud computing in Italia, in Italian), www.enterthecloud.it, 2013.

European CIO Association, Users recommendations from the European CIO Association for the success of the cloud computing in Europe, ARES 2012.

European Commission, Commission plans guide through global Internet policy labyrinth, ec.europa.eu/digital-agenda 2013.

European Commission, Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard

to the processing of personal data and on the free movement of such data (COM(2012) 11 final), 2012.

European Commission, *A digital agenda for Europe* (*Un'agenda digitale per l'Europa*, in Italian), COM(2012) 245 def/2, 2010.

Flick C., Ambriola V., Data in the clouds: Legal aspects of cloud computing and application to the public administration (Dati nelle nuvole: aspetti giuridici del cloud computing e applicazione alle amministrazioni pubbliche, in Italian), Federalismi.it, 6, 2013.

Forum PA – Quaderni, PA on the cloud : G- Cloud: Gaining efficiency to innovate and reduce costs (La PA sulla nuvola: G-Cloud: innovare per guadagnare efficienza e ridurre i costi, in Italian), 2012.

Gantz. J.F., Minton S., Toncheva A., *Cloud computing's role in job creation*, IDC 2012.

Mantelero A., Outsourced computing processes and cloud computing: Management of personal and business data (Processi di outsourcing informatico e cloud computing: la gestione dei dati personali e aziendali, in Italian), Dir Inf. 2010, 673.

Schubert L., The Future of cloud computing. Opportunities for European cloud computing beyond 2010, Expert Group Report, 2009.

Biographies

Caterina Flick is a lawyer qualified in white collars and business, ITC law and cybercrime, privacy. Temporary professor of ITC law and privacy (University of Pisa, Siena, Lumsa of Rome and, since 2013, UTIU), she collaborates on research projects, participates to conferences. She is author of publications. She is consultant and lecturer at public authorities, including the Authority for the protection of personal data. Delegate at the FAO on behalf of IFWLC, she is also member of Committees for equal opportunities. Awarded in 2013 as *Excellent woman of Rome*.
email: c.flick@nmlex.it

Vincenzo Ambriola is full professor at the Department of Computer Science of the University of Pisa. Author of more than 100 scientific publications, his current research interests are in e-government, software engineering, programming languages.
email: ambriola@di.unipi.it

The NeuViz Data Visualization Tool for Visualizing Internet-Measurements Data

G. Futia, E. Zimuel, S. Basso, J.C. De Martin

Abstract. *In this paper we present NeuViz, a data processing and visualization architecture for network measurement experiments. NeuViz has been tailored to work on the data produced by Neubot (Net Neutrality Bot), an Internet bot that performs periodic, active network performance tests. We show that NeuViz is an effective tool to navigate Neubot data to identify cases (to be investigated with more specific network tests) in which a protocol seems discriminated. Also, we suggest how the information provided by the NeuViz Web API can help to automatically detect cases in which a protocol seems discriminated, to raise warnings or trigger more specific tests.*

Keywords: Data visualization, network performance, big data.

1. Introduction

The Internet is a cornerstone of our societies and has been enabling unprecedented levels of social interaction, content sharing, business creation, as well as innovation in many fields. As Frischmann argues convincingly, one of the main reasons why the Internet is so relevant for us is that the Internet is an *infrastructural resource*, i.e., a shared piece of infrastructure that is typically managed as a *commons* in a non-discriminatory way [Frischmann, 2012].

However, the Internet is not an infrastructural resource as a fact of nature, or because of an immutable, technological law; the current status of the Internet is, instead, the consequence of specific choices, both private and public, that could very well change over time. For example the policy decision of who (the State or the Internet Service Providers) should finance (and under which conditions) the so-called 'Next Generation Networks' (NGNs) has the potential of radically changing the landscape.

In fact, many parties (including the authors of this contribution) believe that, if States allow the Internet Service Providers (ISPs) to implement premium services to collect more money and finance NGNs, the infrastructural-resource characteristics of the Internet may become less relevant, and the Internet may lose part of its *generativity* (i.e., the property of enabling more and more people to write and distribute software and/or media content [Zittrain, 2009]).

To be fair, there is little empirical evidence supporting most policy positions on both sides of the debate. On the one hand, for instance, it is hard to prove empirically ex ante that allowing ISPs to implement premium services will reduce the generativity of the Internet. On the other hand, there is surprisingly little evidence backing the 'bandwidth hogs' argument (i.e., the argument that there is a little number of people that consume most bandwidth). The Internet policy debate, in general is so ill informed by poor data, by missing data, and by data provided by one single stakeholder that – we agree with Palfrey and Zittrain – there is a need for more, better data to anchor the debate to solid foundations and move forward [Palfrey and Zittrain, 2011].

This is indeed starting to happen: more and more network measurement tools and visualizations, in fact, are being developed by researchers and companies worldwide. Many of such tools and visualizations are hosted by Measurement Lab [MLab], an umbrella project run by the Open Technology Institute and the PlanetLab Consortium, and supported by academic partners and companies such as Google.

In this paper, in particular, we propose NeuViz (Neubot Visualizer), an architecture that allows us to process and visualize the data collected by Neubot, the network neutrality bot [Basso et al, 2011a], one of the tools hosted by Measurement Lab. Neubot – a project of the Nexa Center for Internet & Society – is a centrally-coordinated bot that runs the in background on the user computer and periodically runs network-performance tests that currently emulate HTTP and BitTorrent, and, in future, will emulate other protocols, such as the uTorrent Transport Protocol (uTP) [Norberg, 2009].

The purpose of NeuViz is to visualize and navigate Neubot data through its Web user interface, to search for cases (to be investigated with more specific network tests) in which a protocol seems discriminated. Also, NeuViz is designed to help, in the future and with a more advanced Neubot architecture, to automatically detect cases in which a protocol seems discriminated, to raise warnings or trigger more specific tests.

Many existing visualization architectures are based on cloud services and allow one to query the data on demand using SQL-like query languages; compared to such visualization tools, NeuViz is much more optimized for the specific purpose of visualizing network measurement data. We designed, in fact, a robust, scalable backend architecture to support special-purpose, complex data analysis, in which the query (or the filtering algorithm) is executed in advance on the network-experiments dataset, and in which the result is stored in one (or more) NoSQL database(s), for fast data access.

We evaluate our work by loading into NeuViz the results of two Neubot network tests (Speedtest and BitTorrent) collected in the January 2012 - May 2013 period. We show that NeuViz helps us to effectively navigate Neubot data to identify cases in which a protocol seems discriminated. Also, we suggest that the information provided by the NeuViz Web API can help to automatically detect cases in which a protocol seems discriminated.

The rest of this paper is organized as follows. In Section 2 we describe related network measurement tools and visualizations. In Section 3 we describe Neubot and the Neubot data that we used in this paper. In Section 4 we describe the NeuViz architecture. In Section 5 we describe our implementation choices. In Section 6 we describe what we learnt from browsing Neubot data with NeuViz. In Section 7 we draw the conclusions, and we describe future developments.

2. Related Work

In this section we mention the related tools and visualizations. Some of the tools that we mention (including Neubot) are hosted by Measurement Lab (M-Lab) [Dovrolis et al, 2010], a distributed server platform that also provides advanced services (e.g., the possibility of querying the hosted-tools data using BigQuery, a RESTful service to query big datasets using an SQL-like query language [BigQuery], and the possibility of measuring TCP state variables by using the instrumented Web100 TCP/IP Linux stack [Mathis et al, 2003]).

2.1. Network-Measurement Tools

In this section we mention four tools similar to Neubot: Glasnost, the Network Diagnostic Tool, SpeedTest.net, and Grenouille.

Glasnost is a client-server browser-based Java applet developed by the Max Planck Institute for Software Systems and maintained by the Measurement Lab community. Glasnost compares a certain protocol flow (e.g., BitTorrent, Emule) with a reference flow to detect traffic shaping and its cause (e.g., the port number, the payload). Glasnost flags a network path as shaped if repeated tests show that (i) the path is non-noisy and (ii) the application-level speed of the protocol flow is 20% (or more) lower than the one of the reference flow [Dischinger et al, 2010].

The Network Diagnostic Tool (NDT) is network-measurement Java applet that measures the download and upload speed between the user computer and a Measurement Lab server [Carlson, 2003]. During the measurement, the server uses the modified Web100 Linux TCP/IP stack to expose the state variables of TCP during the transfer. In addition to the Java applet a NDT command-line application is also available.

The well-known SpeedTest.net web site [SpeedTest] provides a network-measurement, flash-based test that relies on many parallel HTTP connections to estimate the download and upload broadband speed of the user's connection, using a methodology that is documented, e.g., in "Understanding Broadband Speed Measurements" [Bauer et al, 2010].

Grenouille is a network measurement tool that measures the round trip time, the download speed, and the upload speed [Grenouille].

Differently from Glasnost, Speedtest.net, and NDT (which run on-demand tests), Neubot and Grenouille run tests in the background; however, Neubot uses diverse protocols, while Grenouille focuses on the performance only.

2.2. Network-Measurement Visualizations

In this section we mention six visualizations similar to NeuViz: the visualizations of the Syracuse University School of Information studies, the world map created by Open Knowledge Foundation, the two tools proposed by Measurement Lab, the visualization of data collected by SpeedTests.net, and the visualization of the data collected by Grenouille.

The Syracuse University School of Information Studies developed three visualizations of the data collected by Glasnost [SyracuseVis]: an interactive table that shows which ISPs seem to shape (or block) BitTorrent; a visualization that displays the "top throttlers" ISPs from 2009 to 2012; a visualization that shows alleged BitTorrent shaping (or blocking) in selected countries.

Michael Bauer, data wrangler at the Open Knowledge Foundation, created a visualization of Glasnost data as well, which shows on the world map the percentage of tests that Glasnost detected as shaped [OkfnVis]. The user can filter the dataset to show only the results that are related to a single protocol emulated by Glasnost, e.g., HTTP, BitTorrent, eMule.

The Measurement Lab team developed a visualization of NDT data that shows many indexes (e.g., the number of tests, the download and the upload speed, the round trip time) on the world map [MLabVis]. Such visualization allows one to aggregate the data by ISP and by geographical dimension (country, region/state, city), and it also allows one to compare the performance of multiple ISPs at different geographical levels.

Dominic Hamon, a software engineer at Google and Measurement Lab, developed visualizations (and a video) that show, on the world map, a point indicating the latitude and the longitude of each client that runs a test towards a Measurement Lab server, using NDT data and BigQuery [BigQueryVis].

Visualizations of the data collected by SpeedTest.net can be browsed online and downloaded from the NetIndex.com website [NetIndex].

Data collected by the Grenouille tool can be browsed online through the visualization available at the Grenouille website [Grenouille].

Similarly to the NDT visualizations NeuViz is based on the world map; however, NeuViz is optimized for complex data analysis and uses precomputed data, while the NDT visualizations are more interactive and fetch the data from BigQuery on demand. Also, the aim of NeuViz is similar to the aim of the Glasnost visualizations; both, in fact, intend to make access networks more transparent by, respectively, showing anomalies and alleged shaping.

3. Neubot and Neubot data

In this section we describe Neubot and the Neubot data that we use in this paper.

3.1. Description of Neubot

Neubot is a free-software Internet bot that performs active, lightweight network-performance tests [De Martin and Glorioso, 2008; Basso et al, 2010; Basso et al, 2011a]. Once installed on the user's computer, Neubot runs in the background and every 30 minutes performs active transmission tests with servers hosted by Measurement Lab. To coordinate the botnet composed of all the Neubot instances worldwide, there is the so-called *Master Server*, which suggests each Neubot the next test to run as well as the default test parameters. Currently, the Master Server does not optimize the suggestions returned to each Neubot; however, as we will show the information returned by NeuViz could help the Master Server to implement more dynamic policies.

Neubot implements three network performance tests: Speedtest, BitTorrent, and RawTest. Speedtest measures the network performance using the HTTP protocol, BitTorrent measures the network performance using the BitTorrent protocol, and the RawTest test measures raw, TCP-level performance (hence the name of the test). In this paper we only describe the Speedtest and the BitTorrent tests, because we are mainly interested to use NeuViz to find cases in which a protocol seems discriminated.

3.1.1 The Speedtest Test

Speedtest is an HTTP-based test – originally inspired to the test of SpeedTest.net, hence the test name – that downloads and uploads data using a single HTTP connection [Basso et al, 2011b]. The test measures the download and the upload speed at the application level. Also, the test estimates the base Round Trip Time (RTT) using as a proxy the time that the connect system call takes to complete (later indicated as connect time). The test transfers a number of bytes that guarantees that each phase of the test (download, upload) lasts for about five seconds.

3.1.2 The BitTorrent Test

The BitTorrent test is, in principle, similar to the Speedtest test, except that it uses the BitTorrent peer-wire protocol [Cohen, 2009] instead of the HTTP protocol.

As Speedtest does, the BitTorrent test transfers a number of bytes that guarantees that each phase of the test (download, upload) lasts for about five seconds.

However, while Speedtest makes a single GET request for a large-enough amount of data, BitTorrent – to better emulate the BitTorrent protocol – downloads many small chunks in a request-response fashion and, to approximate a continuous transfer, makes many back-to-back requests at the beginning of the test.

3.2 Data Preprocessing and Publishing

Measurement Lab (which hosts Neubot on its servers) periodically collects the Neubot experiments results saved on its servers and publishes such results on the Web [MLabData] under the terms and conditions of the Creative Commons Zero 1.0 Universal license [CC0]. We mirrored the data provided by Measurement Lab, and we converted such data to CSV format, generating CSV files that contain one month of data each. To prepare this paper, we imported into NeuViz the CSV files from January 2012 to May 2013 (reading 5,383,376 test, from 4,037 Neubot clients worldwide, for a total of 1.5 GB) [NeubotData].

Each CSV file contains the following fields (the type is indicated in parentheses): client address (str); connect time, in second (float); download speed, in byte/s (float); Neubot version (str); operating system platform (str); server address (str); test name (str: “speedtest” or “bittorrent”); timestamp of the test, i.e., the number of seconds since 1970-01-01 00:00 UTC (int); upload speed, in byte/s (float); unique identifier of the Neubot instance (str).

4. Description of the NeuViz Architecture

Fig. 1 shows the NeuViz architecture, which is a pipeline that processes data provided by *Producers*, and which organizes the data such that *Consumers* can visualize (or further process) such data. The pipeline is composed of a *Backend* and a *Frontend*: the Backend receives data from many Producers and processes such data to allow for efficient visualization; the Frontend is a Web interface that visualizes the data. In the middle there is a *Web API*.

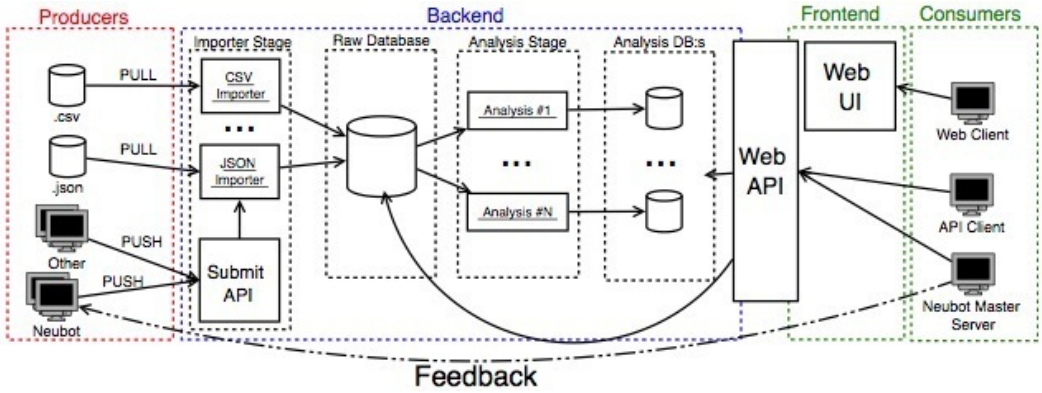


Figure 1
NeuViz Architecture

4.1 The Producers

As a first approximation a Producer is a static dataset. For example, in this paper we used Neubot data expressed in CSV format and in the future we may want to import datasets from other projects (e.g., SpeedTest.net) and encoded in other formats (e.g., JSON).

NeuViz also includes a Submit API, which allows network-experiment tools (e.g., Neubot and possibly other network-measurement tools) to push the result of

their experiments just after the experiments are run. We added the Submit API because we want to create a feedback loop in which data posted by Neubot is processed by NeuViz and consumed by the Master Server to provide better suggestions to Neubot instances.

4.2 The Backend

The Backend receives network-experiments data from many sources and organizes such data for an efficient visualization. As Fig. 1 shows, the Backend is composed of two processing stages, each followed by a database stage. The first processing stage is the *Importer Stage*, which receives data from many sources, normalizes the data, and writes the data into the *Raw Database*. The second processing stage is the *Analysis Stage*, which reads data from the Raw Database, analyzes the data to compute aggregate metrics, and saves the aggregate metrics into one or more *Analysis Databases*.

In the following sections we discuss the stages of the Backend, starting from the Importer Stage.

4.2.1 The Importer Stage

The Importer Stage organizes data coming from many sources (and possibly represented using different formats) into a single database. There is one Importer Module for each network measurement tool and data format. To make an example, if we want to use NeuViz to visualize SpeedTest.net data (expressed in CSV format) and Neubot data (expressed in CSV and JSON format), we need to write three Importer Modules: one for the SpeedTest.net data and two for the Neubot data (the former for the CSV and the latter for the JSON format).

The Submit API design reflects the fact that there is an Importer Module for each network measurement tool and data format. The basic API request to store the result of a new experiment, in fact, is a POST request to this URI: "/neuviz/1.0./import/<tool>/<params>", where <tool> is the name of the tool that produced the piece of data (e.g., "neubot"), and where the Content-Type HTTP header must reflect the data type (e.g., "application/json"). The problem of whether (and how) to authenticate the measurement tool submitting the data is not discussed in this paper.

The Importer Stage does not reduce all the input data to the same schema (be it a real SQL schema or not), because such transformation is not practical. The input data schema, in fact, depends on which metrics the specific network experiment measures; therefore, this stage just enriches the data with geographical information (if needed), converts the data into a common, database-dependent format (e.g., JSON), and writes the data into the Raw Database.

4.2.2 The Raw Database

The Raw Database receives heterogeneous data organized in a uniform format (e.g., JSON) by the Importer Stage. As said before, it is not practical to reduce all the input data to the same schema, suggesting that the Raw Database could be easily implemented using NoSQL (e.g., MongoDB [Mongo]).

A possibly-conflicting requirement for the Raw Database is that, in addition to being able to store heterogeneous data, the Raw Database shall also be

scalable-enough to handle continuous streams of data posted on the Submit API by, at least, Neubot and possibly by other network measurement tools.

4.2.3 The Analysis Stage

The Analysis Stage is a collection of Analysis Modules that periodically fetch data from the Raw Database and process it to produce the aggregate data needed for the visualizations. To start off we plan to implement two different visualizations: one that shows a given performance metric (e.g., the median download speed) on the world map and that allows the user to zoom and see the same performance metric on a smaller geographic scale (i.e., country, province, city); the other that shows a given performance metric in function of the time.

As far as functional requirements are concerned, the Analysis Stage needs to process data in a scalable way, because we need to process multiple times the raw data stored in the Raw Database. Also, the Analysis Stage should minimize the computational cost of adding the results of new experiments to NeuViz.

4.2.4 The Analysis Databases

The Analysis Databases are a number of (conceptually-separated) databases that store data which is ready to be visualized on the NeuViz Frontend with minimal computational cost. We want, in fact, to allow the user to visualize and browse the data as seamlessly as possible.

4.3 The Web API

The Web API connects the Backend and the Frontend. The Frontend, in fact, uses the Web API to retrieve the data that should be visualized by a Web client through the Neuviz Web interface. However, also other clients can access the Web API to extract information from the collected data.

The Web API typically returns the Analysis Database data, because NeuViz is optimized to store and quickly return the results of the data analyses. However, in cases in which the cost of processing the Raw Database data on the fly is negligible, the Web API will access directly the Raw Database data and will compute the result on the fly. This is represented in Fig. 1 by an arrow that goes from the Web API to the Raw Database.

In this paper we do not discuss whether and how the access to the API should be restricted. This will possibly be the subject of a future work.

4.4 The Frontend and the Consumers

The Frontend is a Web interface that visualizes the data stored in the Backend.

The typical (and default) Consumer is of course a Web client that uses the NeuViz Web interface, but also other clients can consume the available data. In particular an interesting, already-planned reuse of the Web API is the following: we plan to modify the Master Server to retrieve data from the Web API, process the data, and adapt accordingly the suggestions the Master Server provides to Neubot instances (e.g., if there are few Neubot instances in a specific geographical area, the Master Server suggests to perform tests more frequently in that area).

5. Implementation Choices

In this section we describe the implementation of the first NeuViz prototype [NeuVizGit], and we explain our implementation choices.

5.1 The Importer Stage

We implemented the Importer Stage step using a Python command-line script that accepts in input a CSV file. In our tests we imported and normalized 1.5 GB of Neubot data (using CSV files), from January 2012 to May 2013, and we stored the data into a MongoDB database. We run the code on a laptop with an Intel Core i7 CPU at 2.0 Ghz, with 8 GB of RAM, and a 256-GB SSD, running GNU/Linux 3.5.0. The Python code is designed to execute both on a common computer and in a cloud environment, if needed: to this end we divided the Importer and the Analysis code into a *map* step and a *reduce* step.

We also used the GeoLite Free Database to retrieve geo-information from the client IP address, using MongoDB to store the geographic information [GeoLite]. As explained in the GeoLite website, when the database is not up-to-date, the geolocation loses 1.5% of accuracy each month because IP addresses are re-assigned. To minimize the damages caused by out-of-date GeoLite databases, we never used databases older than two months.

5.2 The Raw Database

We implemented the The Raw Database using MongoDB, a NoSQL database very often deployed in big data scenarios [Moniruzzaman and Akhter, 2013]. We exploited the indexes feature of MongoDB to speed up the query execution, processing about 5.3 million of samples in less than 60 minutes.

As noted above, the code is written in a way that potentially allows us to use MapReduce techniques on cloud services [MapReduce], should we need to do that. However, especially during the development of the initial prototype, we didn't used MapReduce, because a single NoSQL database allowed us to perform queries on demand and retrieve data immediately (which is not, of course, possible in a cloud-based MapReduce scenario).

5.3 The Analysis Stage

We implemented a prototypal Analysis Module, written in Python, to retrieve and process data from the MongoDB database and create our world map visualization, and we are also working on another Analysis Module that will generate data for the visualization that shows a given performance metric in function of the time.

The Analysis Stage that we implemented outputs a JSON file in which the information is aggregated at the geographical level (countries, and cities), at the temporal level (hour of the day), and at the business level (ISP). Therefore, the Web interface receives in input, for BitTorrent and Speedtest, the median value of the upload speed, of the download speed, and of the connection time of a specific country or city, and their ISPs, in a precise hour of the day. We decided to use the median, which is a common index used to analyze network traffic, to avoid the risk that few outliers could dominate our index.

We also computed the number of Neubot instances (per country, city, ISP) as well as the number of Neubot tests (per country, city, ISP). Since the IP address can vary over time, we identified a Neubot instance by using the (Neubot ID, IP address) tuple. The number of Neubot instances and the number of tests can be used to understand the geographical distribution of Neubot clients and the network traffic produced by each Neubot.

5.4 The Analysis Databases

We generated a JSON file for each month of the Analysis Stage. The collection of these files can be considered to be the Analysis Databases. However, these JSON files can also be stored in a MongoDB to retrieve the data according with different parameters or different search query. Data could also be stored in the cloud when scalability needs occur.

5.5 The Web API

To access the NeuViz API, the user sends the following HTTP/1.1 request: GET /neuviz/1.0/<viz>/<params>, where <viz> is the name of the visualization, and <params> is a placeholder for (possibly-empty) parameters. The returned JSON contains a recursive set of dictionaries that represent the geographical dimension (country, city), the time dimension (hour of the day) and the business dimension (ISP). The leaves are dictionaries that contain the following hour-wide median statistics for the Speedtest and the BitTorrent tests: download speed, upload speed, connection time, number of Neubot instances, number of tests. The geographical (country, city), the time (hour of the day), and the business (ISP) dimensions of data is shown in Fig. 2.

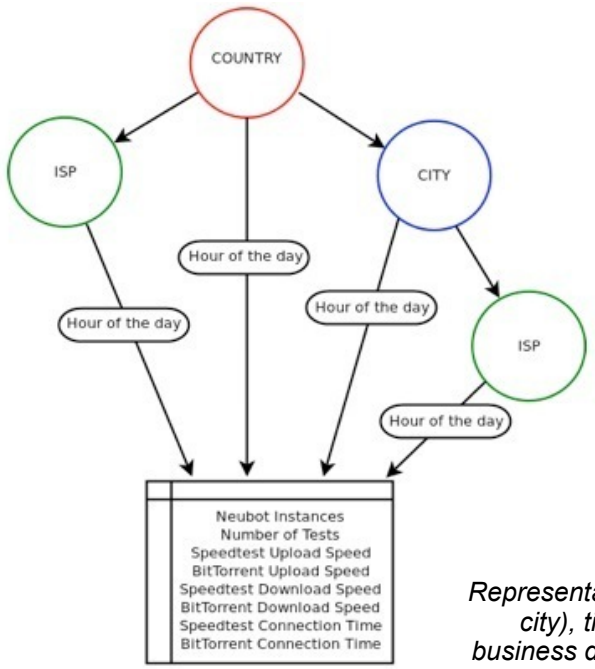


Figure 2
Representation of geographical (country, city), time (hour of the day), and business dimensions (ISP) of JSON file.

5.6. The Frontend

The Web interface, written using D3.js [D3], allows the user to explore different network measurement performances at different geographic dimensions (country, cities, and ISPs). For simplicity, and since it does not seem to cause any performance issue, we currently use the Web interface to compute some statistics, e.g., the difference between the median Speedtest download speed and the median BitTorrent download speed that we use in Section 6.2 to compare the performance of BitTorrent and Speedtest.

6. Results

In this section we report what we learnt from using NeuViz to browse Neubot data, both in terms of number of tests and in terms of performance.

6.1. Number of Neubot Tests

Fig. 3 shows the visualization of the number of tests per country and per hour. The alpha channel of the country color indicates the median number of tests per country. The visualization, in particular, shows the median number of tests performed between 9:00 PM and 10:00 PM (local time) in April 2013. The selected country is Canada, in which the median number of tests performed is indicated by the number in the bottom right corner (1084).

By selecting other countries in the visualization, we have seen that the countries with more median tests per hour between 9:00 PM and 10:00 PM in April 2013 are: the US (4223); Italy (2866); Germany (2285); and Canada (1084). Other countries have less tests per hour.

The availability of the number of tests per country is interesting because, by knowing the number of tests per country, the Master Server could maximize the test coverage; e.g., it can increment the frequency of testing on countries where there are few Neubot users.

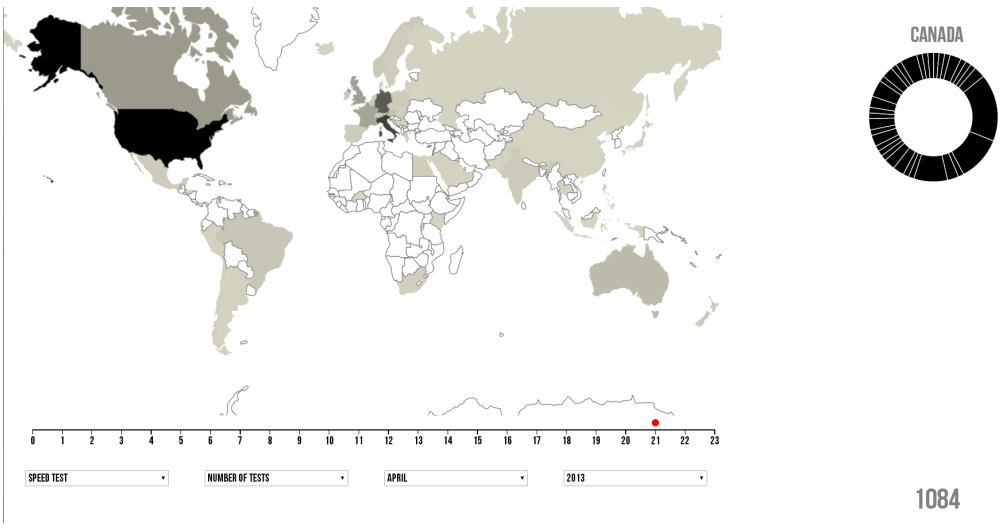


Figure 3
NeuViz interface of the worldwide map for Neubot data of April 2013

6.2. Comparison of Speedtest and BitTorrent performance

Before studying the visualization that shows the difference between the Speedtest and the BitTorrent test download and upload speeds, we checked whether the Speedtest and the BitTorrent connect times were 'comparable'. To this end we arbitrarily define ‘comparable’ two median connect times whose difference is smaller than five milliseconds in our experience a reasonable threshold for this kind of analyses.

The visualization of the difference between the median BitTorrent connect time and the median Speedtest connect time shows, surprisingly, that in Italy such difference is always positive and often greater than five millisecond (i.e., the Speedtest connect time is typically lower). Italy is the only country in which, for 2013 data, we noticed this behavior.

Also we noticed interesting things from the comparison of the median upload speed in countries in which the median connect times are comparable. We noticed, in fact, that in 2013 the median upload difference between Speedtest and BitTorrent in Canada was very often positive, while the same difference was very often negative in the US (see Fig. 4).

Moreover, when comparing the download speeds in countries in which the connect times are comparable, we also noticed that the US Speedtest download speed is always lower (in median) than the BitTorrent one for every hour of the day and for every month of 2013. Interestingly, instead, the download speeds are comparable in Italy, in which – as we have seen – there is a connect time bias in favor of Speedtest.

The above observations lead us to speculate that: (a) BitTorrent is slightly faster than Speedtest; (b) in Italy the two tests are comparable because of the connect-time bias that we observed; (c) the BitTorrent upload speed seems to be discriminated in Canada. Of course, these are only hypotheses that need to be verified (or contradicted) by more detailed experiments.

6.3. Concluding Remarks

Despite being still in beta stage, NeuViz allowed us to discover the three diverse network anomalies we described in Sect. 6.2. In the future, a more advanced Master Server could learn, from the NeuViz API, about similar anomalies and ask Neubot instances that are near the anomalies to gather more information needed to investigate the anomalies (e.g., one could capture packets to gather RTT samples useful to understand whether there is a connect-time bias).

7. Conclusion and Future Work

In this paper we described NeuViz, an architecture that allows us to process and visualize the data collected by Neubot, the active, network-measurement tool developed by the Nexa Center for Internet & Society. The purpose of NeuViz is to visualize and navigate Neubot data through its Web user interface, to search for cases (to be investigated with more specific network tests) in which a protocol seems discriminated.

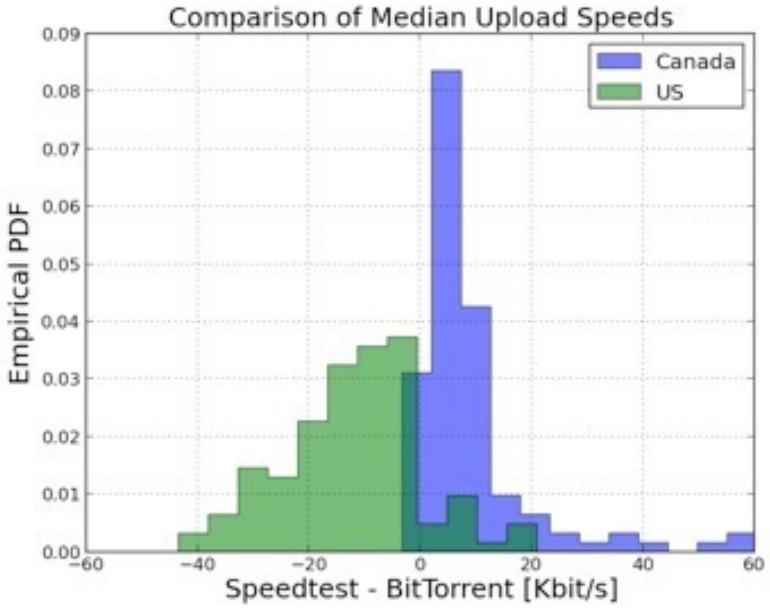


Figure 4
The Empirical Probability Density Function (PDF) of the difference of the median upload speed of US and Canada

Differently from other visualization architectures NeuViz is much less flexible and much more optimized, on purpose. NeuViz, in fact, executes the queries in advance and the result is stored into one or more NoSQL databases (using MongoDB), for fast data access. The Backend of NeuViz, written in Python, is structured to ease the task of porting it to a cloud-based MapReduce solution, for future scalability. The Web interface Frontend of NeuViz shows a world-map-based visualization of Neubot results implemented using the D3.js library.

To evaluate NeuViz we loaded one-year-and-a-half records collected by two network tests periodically run by Neubot, called Speedtest (based on HTTP) and BitTorrent. We showed that NeuViz effectively helped us to identify cases (to be investigated with more specific network tests) in which a protocol seems discriminated. In our discussion we also suggested how the Web API of NeuViz can help to automatically detect cases in which a protocol seems discriminated, to raise warnings or trigger more specific tests (by cooperating with the Master Server of Neubot). As part of our future work we plan to extend NeuViz to automatically raise warnings and to cooperate with the Master Server of Neubot to trigger more-specific network experiments.

Acknowledgments

The first prototype of the NeuViz project has been developed as final project of the BigDive course 2013 [BigDive]. We would like to thank Christian Racca of the TOP-IX Consortium and all the staff and teachers of the BigDive course for their support during the development of this project.

References

[Basso et al, 2010] Basso S., Servetti A., De Martin J. C., Rationale, Design, and Implementation of the Network Neutrality Bot, in Proc. of Congresso Nazionale AICA 2010, L'Aquila, 2010.

[Basso et al, 2011a] Basso S., Servetti A., De Martin J. C., The network neutrality bot architecture: A preliminary approach for self-monitoring of Internet access QoS, in Proc. of the Sixteenth IEEE Symposium on Computers and Communications, Corfu, Greece, 2011.

[Basso et al, 2011b] Basso S., Servetti A., De Martin J. C., The hitchhiker's guide to the Network Neutrality Bot test methodology, in Proc. of Congresso Nazionale AICA 2011, Torino, 2011.

[Bauer et al, 2010] Bauer S., Clark D., Lehr W., Understanding broadband speed measurements, in Proc. of Telecommunications Policy Research Conference, 2010.

[BigDive] Big Dive course website, from http://www.bigdive.eu.

[BigQuery] Google BigQuery, from http://developers.google.com/bigquery/.

[BigQueryVis] Hamon D., Visualizing M-Lab data with BigQuery, from http://dmadev.com/2012/11/19/.

[Carlson, 2003] Carlson R., Developing the Web100 Based Network Diagnostic Tool (NDT), In Proc of the Passive and Active Measurement Conference, 2003.

[CC0] Creative Commons Zero 1.0 Universal License, from http://creativecommons.org/publicdomain/zero/1.0/.

[Cohen, 2009] Cohen B., The BitTorrent Protocol Specification, from http://www.bittorrent.org/beps/bep_0003.html.

[D3] D3.js – Data Driven Documents, from http://d3js.org/.

[De Martin and Glorioso, 2008] De Martin J.C., Glorioso A., The Neubot project: A collaborative approach to measuring internet neutrality, in Proc. of the IEEE International Symposium on Technology and Society, Fredericton, Canada, 2008.

[Dischinger et al, 2010] Dischinger M., Marcon M., Guha S., Gummadi K. P., Mahajan R., Saroiu S., Glasnost: Enabling End Users to Detect Traffic Differentiation, in Proc. of USENIX Symposium on Networked Systems Design and Implementation, 2010.

[Dovrolis et al, 2010] Dovrolis C., Gummadi K. P., Kuzmanovic A., Meinrath S., Measurement Lab: Overview and an Invitation to the Research Community, ACM SIGCOMM Computer Communication Review, 40, 3, 2010, 53–56.

[Frischmann, 2012] Frischmann B. M., Infrastructure: The Social Value of Shared Resources, Oxford University Press, 2012.

[GeoLite] GeoLite Free Database, from http://dev.maxmind.com/geoip/legacy/geolite/.

[Grenouille] Grenoulile.com website, from http://grenouille.com/.

[MapReduce] Amazon Elastic MapReduce service, from http://aws.amazon.com/elasticmapreduce/.

[Mathis et al, 2003] Mathis M., Heffner J., Reddy R., Web100: Extended TCP Instrumentation for Research, Education and Diagnosis, ACM SIGCOMM Computer Communication Review, 33, 3, 2003, 69–79.

[MLab] Measurement Lab website, from http://www.measurementlab.net/.

[MLabData] Measurement Lab data, from http://measurementlab.net/data.

[MLabVis] Broadband performance using NDT data, from http://goo.gl/m9WbS (google.com/publicdata/explore/...).

[Moniruzzaman and Akhter, 2013] Moniruzzaman A. B. M., Akhter H. S., NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison, International Journal of Database Theory and Application, Vol. 6, No.4, 2013.

[Mongo] MongoDB, from http://www.mongodb.org/.

[NetIndex] Net Index by Ookla company, from http://www.netindex.com/.

[NeubotData] Neubot Measurement Lab data mirror, from http://data.neubot.org/mlab_mirror/.

[NeuVizGit] NeuViz GitHub repository, from https://github.com/neubot/neuviz.

[Norberg, 2009] Norberg A., uTorrent transport protocol, from http://www.bittorrent.org/beps/bep_0029.html.

[OkfnVis] Network neutrality map using Glasnost data, from http://netneutralitymap.org/.

[Palfrey and Zittrain, 2011] Palfrey J., Zittrain J., Better Data for a Better Internet, Science, 334, 6060, 2011, 1210-1211.

[SpeedTest] SpeedTest.net website, from http://www.speedtest.net/

[SyracuseVis] Deep packet inspection stats using Glasnost data, from http://dpi.ischool.syr.edu/MLab-Data.html.

[Zittrain, 2009] Zittrain J., The future of the Internet--and how to stop it., Yale University Press, 2009.

Biographies

Giuseppe Futia is communication manager of the Nexa Center for Internet & Society, Politecnico di Torino (DAUIN), Italy, since February 2011. He holds a Master Degree in Media Engineering from Politecnico di Torino. Since 2008, he collaborates with the Italian newspaper “La Stampa”, especially on Internet & Society topics. Giuseppe holds data analysis and data visualization skills, useful to both sustain the outreach of some of the Nexa projects, and to support research in the field of open data.
email: giuseppe.futia@polito.it

Enrico Zimuel is a software engineer since 1996. He works in the R&D department of Zend Technologies, the PHP Company based in Cupertino (USA). He did research in algorithms and data structures at the Informatics Institute of the University of Amsterdam. He is an international speaker about web and open source technologies. He got a B.Sc. honors degree in Computer Science and Economics from the University “G.D’Annunzio” of Chieti-Pescara (Italy) and he studied at the NKS school of Stephen Wolfram at the Brown University (USA).
email: enrico@zend.com

Simone Basso is a research fellow of the Nexa Center for Internet & Society at the Politecnico di Torino (DAUIN), Italy, since 2010, where he leads the research and development of the Neubot software project on network neutrality. His main research interests are network performance, network neutrality, TCP, Internet traffic management, peer to peer networks, and streaming. He is currently a PhD student at the Department of Control and Computer Engineering of Politecnico di Torino, where he received the Bachelor's (in 2006) and the MoS degrees (in 2009).
email: simone.basso@polito.it

Juan Carlos De Martin is faculty co-director of the Nexa Center for Internet & Society at the Politecnico of Torino (DAUIN), Italy, where he teaches computer engineering and digital culture. He is also faculty fellow at the Berkman Center of Harvard University and senior visiting researcher at the Internet and Society Laboratory of Keio University. Juan Carlos De Martin is a member of the Institute of Electrical and Electronic Engineers (IEEE) and he serves as member of the Scientific Board of the Institute of the Italian Encyclopedia Treccani.
email: demartin@polito.it

Cloud Agency: a Guide through the Clouds

R. Aversa, L. Tasquier, S. Venticinque

Abstract. *Cloud Computing is a new technology that has rapidly established itself in computer science since it allows the usage of huge computing resources that are dynamically allocated in order to satisfy user's needs and that are accessible as a service through a remote interface, such as a web browser. Cloud resources can be distributed in different places and such distribution is made transparent to the user; this one does not need to worry about data replication and/or maintenance of the infrastructure but he/she benefits only of the required services, by exploiting the Pay-Per-Use business model. In this context, the interoperability among different providers becomes critical due to the vendor lock-in problem. Here we present a multi agent system that accesses, on behalf of the user, the utility market of Cloud computing to maintain the best resources configuration that satisfies the application requirements. It also offers management and monitoring facilities for the Cloud infrastructure in order to guide the user in all the phases of the application's lifecycle. Together with the platform, we present client-side tools that can be used to orchestrate agents' based services.*

Keywords: IaaS cloud interoperability, mobile agents, cloud monitoring.

1. Introduction

Cloud Computing is a new computing paradigm whose adoption has been so fast that everyone is already in the Cloud without being aware of it. Web application mailbox, for editing documents or drawing, for joining social network are simple examples of services by which everyone, using a browser, creates and manages his/her information without knowing where applications execute and data reside. Cloud Computing allows to deliver computational resources across the network by a *pay per use* business model. This approach provides to customer the possibility to spend money just for the current needing both from personal and commercial usage. In fact it is not necessary to buy licenses or to invest for big hardware infrastructures which become obsolete, need to be hosted somewhere, are not scalable and must be managed with relevant effort.

On the other hand the *pay-per-use* business model allows for the possibility to change the Resource Providers when a more convenient offer is found. At server side the Cloud paradigm founds on the virtualization technology. Virtualization allows to Cloud providers for building a computing infrastructure that is independent from the underlying hardware. Hardware and software of virtual machine can be configured according to the customer needs. The resource can be easily replicated, moved, also by hot migration, to different hardware to optimize the usage of physical infrastructure, to improve reliability and availability.

The delivery of a computing infrastructure as it happens for any service utility, like water and electricity, has been referred as *Infrastructure as a Service*, whereas Platform as a Service and Software as a Service are offered to developers and final users. When service providers choose to move to IaaS Cloud they could realize that the provisioning and the management workflow of the computing infrastructure really change and a number of new issues arise. The lack of interoperability among different technological Cloud solutions and the limited portability of Cloud applications is a relevant problem raised to the attention of the scientific community. It is known as *lock-in* problem. Even when the service developer is able to overcome these difficulties, by making technical choices that are independent with respect to the Cloud provider, it is not easy to discover and retrieve the available Cloud proposals, to check if they can accomplish the service requirements, and also to compare each other. Currently there is not a common ontology for describing service terms and service levels, neither in a formal way nor through natural language. Other issues regard the management of the acquired resources. Also in this case the lack of a wide adopted standard for service at Cloud infrastructure level (IaaS) affects the chance of opting for a different commercial or technological solution. In fact the use will have to change both management tool and methodology. In particular here we deal with monitoring of resource utilization. This problem has been extensively investigated with the perspective of the resource provider, which aims at optimizing the utilization of its physical resources in order to improve its own service level and to increase its profit. However monitoring needs to be addressed with a different perspective in the case of a service provider that stocks computing resources through the Cloud market. Cloud customers cannot check the compliance of the Service Level Agreement (SLA) trusting the monitoring service of the same provider, who has a conflicting interest ensuring the guarantees on service levels it provides. Besides Cloud customers needs to detect under-utilization and overload conditions.

In this paper we present a set of tools which allows the user for orchestrating agents based services, which support discovery, brokering, management and monitoring of Cloud resources. We describe how these services can be used to execute a workflow for Cloud governance that allows for vendor agnostic provisioning, deployment, management and monitoring Cloud services at Infrastructure level.

2 Running application over IaaS Clouds

Ideally, the life cycle of a Cloud application that is running by using a Cloud infrastructure is divided into three phases (Figure 1):

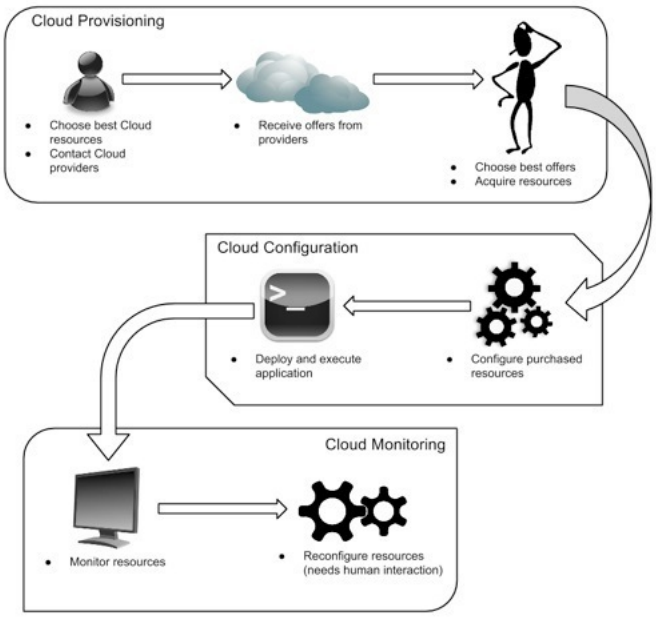


Figure 1
Deployment and execution workflow by using IaaS Cloud

1. *Cloud Provisioning*. The user has to choose the best Cloud resources for his/her application (e.g. VMs, storages, etc.). After that he/she has to select the best IaaS provider basing his/her thinking on a lot of parameters (cost per use, amount of VM memory, storage's size, bandwidth, etc.). Many times this reasoning is too difficult because each provider offers its resources highlighting different characteristics and parameters. This happens because the Cloud vendors haven not a common and standardized interface to describe the resource parameters, making the comparison among same resources an hard job.
2. *Cloud Configuration*. After selecting the best resources for his/her application, the Cloud customer needs to sign an SLA with a Cloud vendor. Once this has been done some management activities are carried before deploying applications. For instance OS images have to be attached and the purchased VMs have to be started. For this reason, the cloud user/developer has to know the allowed actions for that resource and the service interface for that Cloud provider. In fact the same resource purchased from a different provider have different interface and different supported functionalities. At this point the Cloud application can be deployed and executed.

3. *Cloud Monitoring*. Here Cloud users configure a network of probes that collects measures about the performance parameters of the Cloud resources. To get an up to date knowledge of Cloud performance and an history of the Cloud behavior it needs to periodically compute a set of performance indexes and to set up some triggers which notify critical conditions. In fact it would be useful to know if the workload of the infrastructure is different from the one foreseen, in order to avoid saturation or under-utilization of Cloud resources. This information is necessary to design effective reconfigurations of the infrastructure, in order to better adapt it to the current application requirements and to optimize performances and costs.

The complete lifecycle will be handled by the proposed multi-agent system, whose design and exposed services will be described in Section 5; the access to the agents' environment will be allowed by the OCCI/RESTful Interface that is in charge of handling the incoming requests, translating these ones to an agent common language and providing a reply to the user (Section 6). The services' invocation will be made easier by introducing a set of client APIs and tools, that will be detailed in Section 7. Experimental results and conclusion will be presented in Section 8 and 9.

3. Cloud research context

According to [Buyya et al.(2009)] a market-oriented resource management is needed in order to regulate the supply and demand of Cloud resources, providing feedback in terms of economic incentives for both Cloud consumers and providers, and promoting QoS-based resource allocation mechanisms that differentiate service requests based on their utility. The current Cloud computing technologies offer a limited support for dynamic negotiation of SLAs between participants. There are no mechanisms for automatic allocation of resources to multiple competing requests. Furthermore, current Cloud computing technologies are not able to support customer-driven service management based on customer profiles and requested service requirements. The work presented in [Kertesz et al.(2009)] represents a first proposal to combine SLA-based resource negotiations with virtualized resources in terms of on-demand service provision. The architecture description focuses on three topics: agreement negotiation, service brokering and deployment using virtualization. It involves multiple brokers. A Cloud multi-agent management architecture is proposed in [Cao et al.(2009)]. A simpler agents based architecture has been proposed in [You et al.(2009)]. Preliminary investigations by the authors on related topics have been presented in [Aversa et al.(2010)]. SLA@SOI is the main project that aims (together with other relevant goals) at offering an open source based SLA management framework. It will provide benefits of predictability, transparency and automation in an arbitrary service-oriented infrastructure, being compliant with the OCCI standard. In [Edmonds et al. (2011)] an extension of OCCI [Metsch et al.(2010)] for IaaS provisioning is described. The infrastructure specification extends the OCCI core model. It extends the *Resource* entity to represent various infrastructure-related resources and the *Link* to represent concrete relationship between infrastructural resources.

As in our approach, extension is used to improve interoperability but, in this case, it doesn't add new functionalities. Another extension of OCCI to support mechanisms for Monitoring and SLA agreement negotiation is described in [Metsch et al.()]. dynamically over time the system configuration, and so the distributed monitoring must adapt itself quickly to the new requirements.

Besides the aforementioned problems, another open issue in Cloud Computing is about the monitoring of the infrastructure. In order to check or guarantee an agreed SLA at IaaS level, it is necessary to monitor performance and quality indexes, to enforce the agreed service terms. Traditional monitoring technologies for single machines or Clusters are restricted to locality and homogeneity of monitored objects and, therefore, cannot be applied in the Cloud in an appropriate manner [Emeakaroha et al.(2010)]. At the state of the art there are many tools which provide Cloud monitoring facilities, like Rackspace Cloud Monitoring[1], Nimsoft Monitor[2], Monitis[3], Opnet[4], RevealCloud[5]. All of them are proprietary solutions and do not aim at defining a standard for monitoring.

4. mOSAIC

The mOSAIC project (EU FP7-ICT programme) [Di Martino et al.(2011)] intends to progress the state-of-the-art in Cloud Computing by creating, promoting and exploiting an open-source Cloud application programming interface and platform targeted for developing multi-Cloud oriented applications and to negotiate Cloud services according to application requirements. The main benefit of using the mOSAIC software package will be a transparent and simple access to heterogeneous Cloud computing resources and avoidance of lock-in proprietary solutions. The mOSAIC project started on September 2010 and ended on February 2013 and it was coordinated by the Second University of Naples: it benefited of the efforts of a multi-national consortium, composed by Second University of Naples (Italy), Institute e-Austria Timisoara (Romania), European Space Agency (France), Terradue (Italy), AITIA (Hungary), Tecnalia (Spain), Brno University of Technology (Czech Republic), University of Ljubljana (Slovenia), XLAB (Slovenia).

The main components developed within the mOSAIC project are: the Software Platform, the Semantic Engine and the Cloud Agency (whose description will be detailed in Section 5).

The *Software Platform* represents the execution environment for mOSAIC applications. It supports application deployment and execution and hosts

[1] http://www.rackspace.com/cloud/monitoring/

[2] http://www.ca.com/us/lpg/nimsoft.aspx

[3] http://www.monitis.com/

[4] http://www.opnet.com/

[5] http://copperegg.com/

services that take the form of mOSAIC Components. It provides public components called Drivers in order to facilitate the applications' access to Cloud resources at the low level through Connectors. Other Platform components are not public services but contribute to the overall expected behavior from the mOSAIC infrastructure related to scalability, availability, autonomy, etc. The mOSAIC programming model promotes few basic principles. One of the main mOSAIC concepts is that an application is a set of Cloud building blocks connected together. A building block is any identifiable entity inside the Cloud environment that is either a Cloud component or a Cloud resource. A Cloud component is an entity which is controlled by the user, configurable, exhibiting a well defined behavior, implementing functionalities and exposing them to other application components, and whose instances run in a Cloud environment consuming Cloud resources. A Cloud resource is a typical infrastructure resource provisioned by an IaaS provider.

The *Semantic Engine* [Cretella et al.(2012)] is an important mOSAIC tool supporting the user in selecting Cloud APIs components and functionalities needed to build a Cloud application, and provides a list of needed resources to be acquired from the Cloud providers. It utilizes semantic descriptions of the Cloud providers' resources and services, and available mOSAIC components and functionalities offered through APIs.

On the premises of mOSAIC project, other European projects started aiming at using and improving the mOSAIC functionalities: the FP7 project SPECS (Secure Provisioning of Cloud Services based on SLA management) [spe()], starting in Autumn of 2013, continues to enhance the mOSAIC's SLA framework, as long as the mOSAIC's software platform plays the role of run-time environment in the model-driven engineering project named MODAClouds [mod()]. Furthermore mOSAIC agent and semantic technologies will be exploited within the CoSSMic project [cos()] that aims at developing ICT tools needed to facilitate the sharing of renewable energy within a neighborhood.

5. Cloud Agency

Cloud Agency [Venticinque et al.(2012)] is a multi agent system (MAS) that accesses, on behalf of the user, the utility market of Cloud computing to maintain always the best resources configuration that satisfies the application requirements. It is in charge to provide the collection of Cloud resources, from different vendors, that continuously meets the requirements of users' applications. According to the available offers, it generates a service level agreement that represents the result of resource brokering and booking with available providers. The user is able to delegate to the Agency the monitoring of resource utilization, the necessary checks of the agreement fulfillment and eventually re-negotiations. Cloud Agency will supplement the common management functionalities which are currently provided by IaaS Private and Public infrastructure with new advanced services, by implementing transparent layer to IaaS Private and Public Cloud services. Cloud Agency will support the Cloud user in two different scenarios. In the *Deployment* scenario CA supports the discovering and provisioning of the available resources needed to run Cloud

applications. In this case the mOSAIC user is negotiating, by the Cloud Agency, the resources it needs in order to run his applications. In order to propose to the user the best offer of resources, that fits his requirements at best, the CA will use a Brokering Module that chooses among the available offers the best one. Furthermore for configuration and start of resources it will provide management facilities. In the *Execution* scenario it allows to monitor and eventually to reconfigure Cloud resources according the changed requirements of the Cloud Application. More specifically, during the execution Cloud Agency allows the user for the Monitoring of the infrastructure in terms of resource utilization and for the definition of some strategy of autonomic reconfiguration. Reconfiguration can use management facilities by stopping, starting, moving instances, but it could ask for provisioning of additional resources. The *Provisioning Service* provided by CA allows the mOSAIC user to discover, acquire, and set-up resources for deployment of his applications. The result of Provisioning is a set of Cloud resources. They are described, together the offered service levels, terms of services and other information in a Service Level Agreement (SLA). The *Resources Management Service* is used both for deployment and for execution. In fact we need to configure and start resources before to start the application, and we need to start/stop/migrate and reconfigure in general the resource dynamically during its utilization. The *Resource Monitoring Service* is used to get an up to date knowledge of the performance figures of the Cloud Infrastructure, and an history of the Cloud behaviour. The mOSAIC user can start performance meters which are running in the Cloud itself. The *Resource Reconfiguration Service* is necessary in order to avoid saturation or under-utilization of resources, when the workload of the infrastructure is different from the one foreseen.

Cloud Agency architecture is depicted in Figure 2.

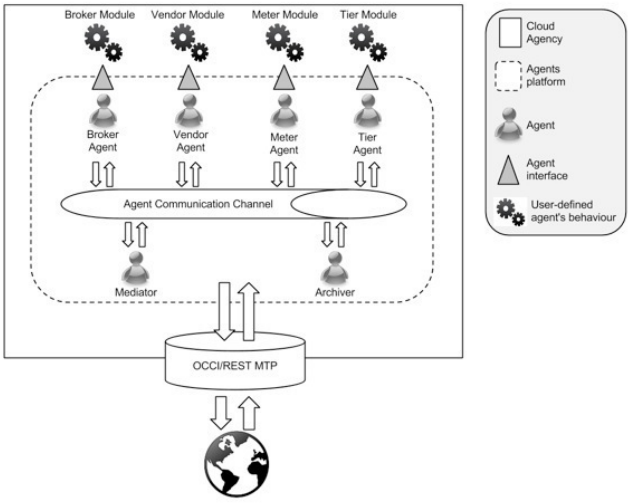


Figure 2
Cloud Agency architecture

One of the leading agents that composes Cloud Agency is the *Broker Agent* that receives the list of those resources that the mOSAIC application needs for its deployment and execution, asks to providers for available offers, brokers the best one and allows for closing the transaction. *Vendor Agents* implement a wrapper for a specific Cloud: they are used to get an offer for resource provisioning, to accept or refuse that offer, to get information about a resource or to perform an action on it (start, stop, resume). About Cloud monitoring, *Meter Agents* perform measures of performance indexes at IAAS level and return their values to the *Archiver Agent*. The Archiver collects the measures and maintains statistics about the mOSAIC system. For the reconfiguration service the *Tier Agent* has been developed: it is triggered by the Archiver and uses policies defined by the user to apply the necessary reactions. The *Mediator Agent* receives requests directly from the user: it is in charge of starting new transactions for provisioning by creating Broker agents; it also starts new Tiers and Meters, and returns information about Cloud Agency configuration and services.

In order to allow a straightforward specialization of agents' behaviors it has been defined abstract interfaces for resource management, monitoring and the other services that the Cloud Agency offers. These abstract interfaces can be implemented by developers who want to extend Cloud Agency by new monitoring technologies, by supporting more Cloud provider and so on. These implementations become plugins of CA, which are automatically used by agents through their general interfaces.

6. A standard interface

Cloud Agency interface provides an asynchronous APIs in order to access the Cloud Agency services. To address this issue Use Cases are designed in terms of Service Requests, Events and Callbacks. Access to Cloud Agency services will be enabled by *HTTP RESTful interface*. Asynchronous requests are used to ask the Cloud Agency for something to be executed. For example to start a Negotiation, to accept or to refuse an SLA, to change a Policy, etc.. They are not-blocking invocations. Execution is started remotely, but the client can continue to run. Completion or failures of requests are notified at client side. Clients are in charge to handle incoming events. Synchronous requests are available to get information. For example clients can ask for reading an SLA, the status of a negotiation, to get the list of vendors, or the list of resources. Queries are synchronous, they return immediately the response if it is available, an exception otherwise. An OCCI compliant message transfer protocol (OCCI-MTP) allows the communication between the client and the Cloud Agency. The main tasks of OCCI-MTP are:

- the verification of the request's correctness;
- the translation from REST request to ACLMessage and vice versa;
- the forwarding of the client's request to the involved agent.

On asynchronous requests an acknowledgment is returned after that the correctness of the message is verified. The response is provided by Cloud

Agency on another connection when the result is available (e.g. when an virtual machine has been started). On synchronous requests the response is always available and returned (e.g. the virtual machine configuration). The management of asynchronous requests is described in Figure 3. When the OCCI-MTP component receives an asynchronous request, it sends a message to the client notifying the successful request's verification and closes the connection. After that it translates the HTTP message in an ACLMessage and forwards it to the involved agent. Since the agents can dynamically change in the Cloud Agency (e.g. a broker agent lives only for a transaction), the OCCI-MTP recovers the correct agent dynamically from database by using an internal coding. When a response is available, the agent calls the OCCI-MTP to send it to the client. The OCCI-MTP translates the ACLMessage in an HTTP message and opens a new connection by using the address provided by the client in the request, sending on it the Cloud Agency response.

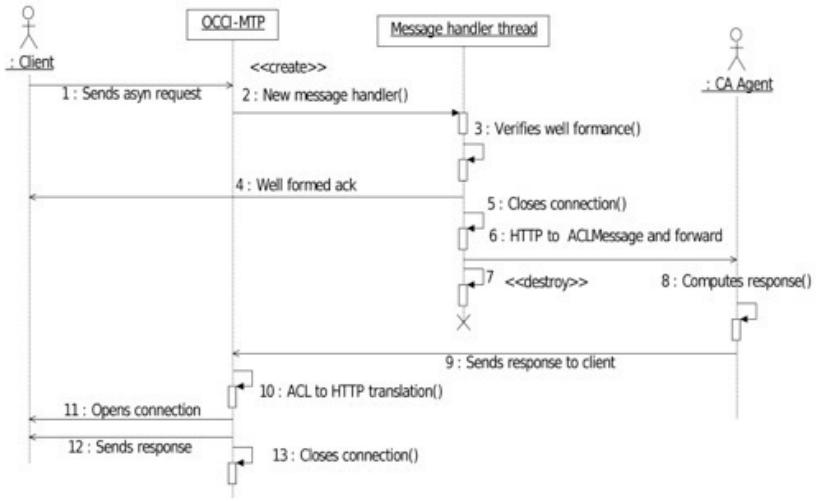


Figure 3
OCCI-MTP ASR

The handling of synchronous requests is shown in Figure 4. When the OCCI-MTP receives a synchronous request, it keeps alive the connection to the client until the response is not available. To do that, the particular thread forwards the request to the involved agent and saves its ID into a common data structure. After that it suspends itself keeping the connection alive. When the response is available, the OCCI-MTP is invoked by the Cloud Agency. The OCCI-MTP translates the ACLMessage in an HTTP message, searches for the involved thread into the common data structure and wakes up it. The thread reads the response, sends it to the client on the old connection and closes this one.

An example of a OCCI-MTP request message is shown in Figure 5.

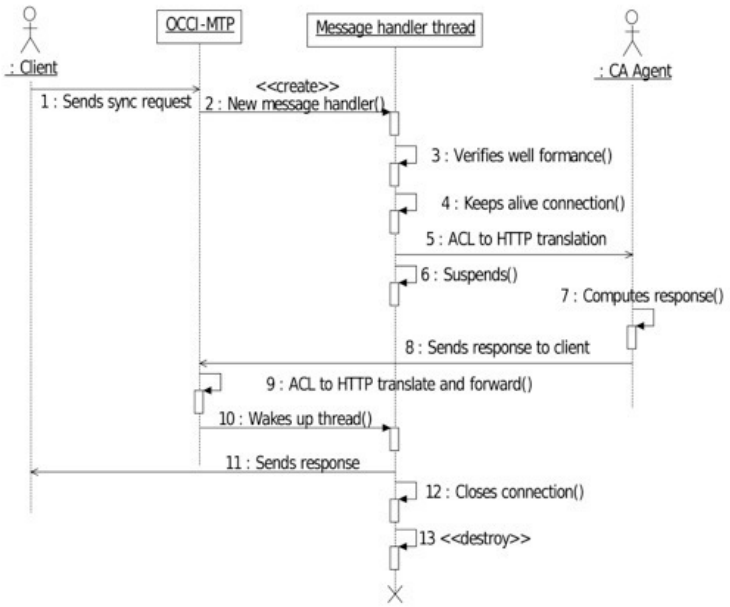


Figure 4
OCCI-MTP SSR

```
POST /cfp HTTP/1.1
User-Agent:curl/7.21.0 [...]
Host: [...]
Accept: */*
Cookie: [...]
Content-Type: text/occi
Reply-To: [host:port/CallBackHandler]
Category: cfp;scheme=[...];class="entity";
Content-Length: ...
\n\r
```

Figure 5
OCCI-MTP Request example

Java API are provided to developers for using agents based services by the provided interface with a object oriented and event-driven programming model. Stubs and callbacks allows for sending requests and handling asynchronous notifications.

- A stub offers a set of methods to invoke the Cloud Agency services. These methods wrap the REST requests in order to query the Cloud Agency RESTFull interface and to perform actions on Cloud resources or to obtain information about the state of the Cloud infrastructure. At the state-of-the-art of the development API client there are four classes that implement the

CAStub class and that provide methods to invoke provisioning, management, monitoring and reconfiguration services;

- An adapter is an handler for the asynchronous messages coming from the Cloud Agency, according to its event-driven architecture. When a new message arrives, it is forwarded to the adapter that implements the particular service listener. The user can implement autonomic reactions by adding his/her own new adapter by the core API.

7. Cloud Agency Application Tools

A Command Line Interface offers to the user a variety of commands to invoke the Cloud Agency's services. It follows the user in all the phases of the deployment and of the execution of his/her own application by giving him/her the possibility to book and manage the Cloud resources in a flexible and simple way. The CA-CLI helps the mOSAIC developer in the deployment process, beginning from the brokering of the best resources for his/her application. The application opens a console that starts a listener to handle the notifications sent by Cloud Agency and allows for the execution of a number of commands. The management is vendor agnostic, in the sense that the user asks for performing a specific operation (start, stop, restart, etc.) on any given resource.

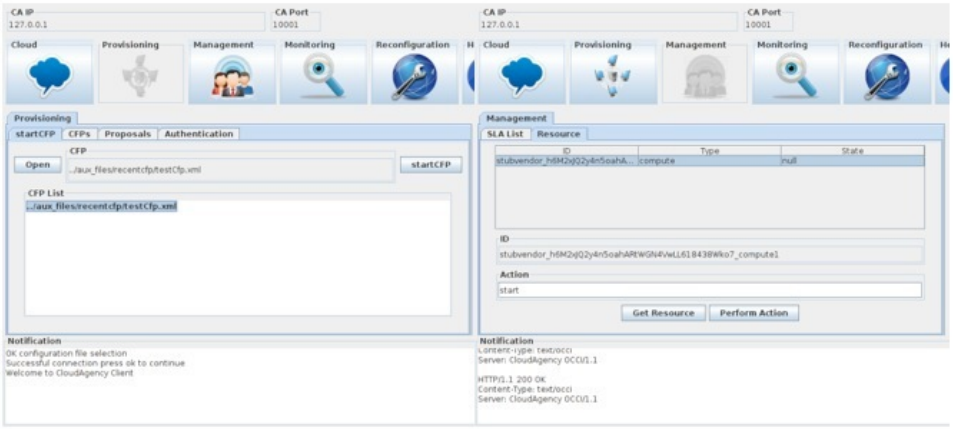
A Graphical User Interface (CA-GUI) is another tool that helps the user during the deployment and execution of his/her mOSAIC application. The functionalities offered by the CA-GUI are basically the same of the CA-CLI ones. Of course, the graphical interface is more powerful than the command line to take under control all the stuff during the provisioning phase: the editing and the listing of the *Call For Proposals (CFPs)*, the providers' proposals or SLAs, the acquired resources and their state. It also provides some additional functionalities in order to simplify the Cloud management, such as the listing of the available vendors, the start/stop of an available VM, the attach and detach of an available storage and so on.

In Figure 6 (a) it is shown how the CA-GUI appears. On top of the interface there is the location of the connected Cloud Agency instance. Just below this one there are some buttons, each one representing a Cloud Agency service and allowing the user for easily performing provisioning, management, monitoring and reconfiguration operations. Moreover, the *Cloud* button provides several functionalities to manage the CFPs and to get information about Cloud Agency status. By clicking a button, a new panel appears, which is customized by the particular operations that the selected service allows. The CA-GUI handles both ASR and SSR in order to get the requested information and/or to perform an action. On the bottom of the window the raw notification messages are displayed.

The management console, shown in Figure 6 (b) allows the user for starting/stopping VMs, for loading and attaching VM images, for deploying and executing applications and so on.

The monitoring console allows for the configuration of the monitoring infrastructure on the acquired resources. For each resource is possible to select a set of measures, each one supported by a specific by Meter Agent. When a

set of measures has been selected and the monitoring configuration is finished, the Cloud Agency creates a new Meter Agent sends it to the target resource. At this point the Meter Agent gets the measures and sends them to the Archiver Agent, that stores them and is able to compute metrics on performance information on user's demand.

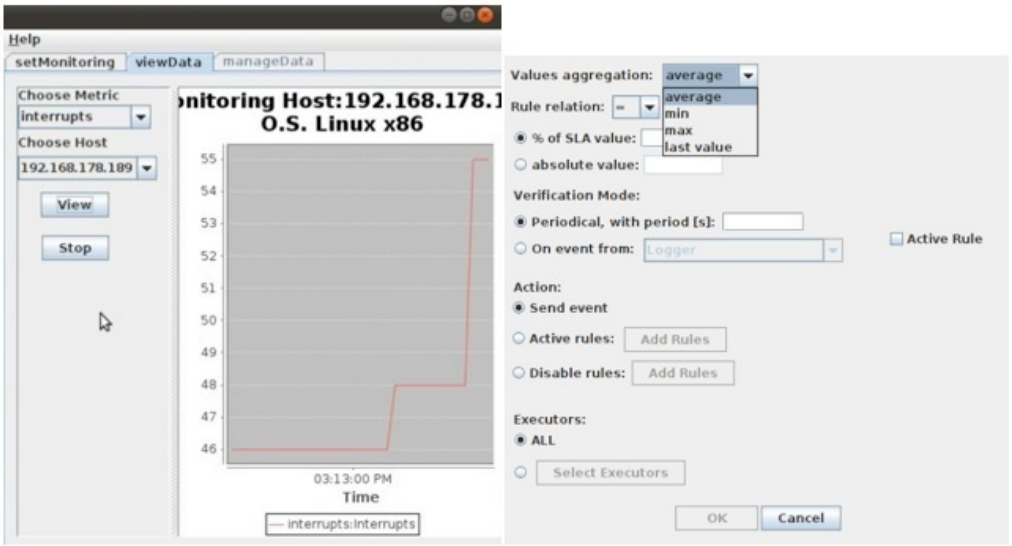


(a) GUI tab for provisioning **(b) GUI tab for management**

Figure 6
Graphical User Interface for Cloud Agency Client

After that the application has been deployed. As regards the mOSAIC developer, he/she can take under control the infrastructure performances by using the monitoring tool that can be started by the CA-GUI. It allows for the visualization of a list of available measures, as it is shown in Figure 7 (a), or for setting up the computation of some metrics about performance indexes. When a new metric is created, the developer can read synchronously the last value of that index or can create a trigger to be notified asynchronously according to a specific time period or when a critical condition is verified as it is shown in Figure 7 (b). An example of available metric is the average value of a measure that is periodically computed and that is notified when it is out of a certain range.

When a trigger is activated by the verification of a critical condition on a resource's parameter, the user can decide to be notified about the verified event or to activate/deactivate other rules previously defined and related to other resource's parameters and/or other resources. So the developer can set up a complex trigger by composing some simple ones.



(a) Visualization of performance's indexes

(b) Creating a trigger on a resource's parameter

Figure 7
Monitoring Tool for Cloud Agency

8. Performance evaluation

In order to evaluate the scalability of the proposed architecture we set up a testbed, by using a single machine hosting the Cloud Agency environment, that analyzes the behaviour of the agents involved in the provisioning service. The testbed aims at demonstrating how multiple requests can be distributed to agents which are able to start or migrate themselves on different computing resources.

A client application submits a workload composed of multiple provisioning requests. The client sends 100 provisioning requests in a closed loop. It means that the client starts a new transaction after that the previous one has been closed. All protocol steps are performed automatically on behalf of the user and the client handles all event occurrences notified by the provisioning service. After an acceptance, the SLA notification closes the transaction. The described scenario has been tested with a different number of vendors and a different number of concurrent clients.

Figure 8 shows the results of the experiment. The chart proves that the mean time for closing a transaction grows linearly with the number of concurrent clients and it is not significantly affected by the number of the involved vendors: this result implies a good scalability of the platform.

In Figure 9 the percentage of time needed for message processing and delivery decreases when the workload increases. Most of the time is spent by agents which can be replicated and executed on different hosts.

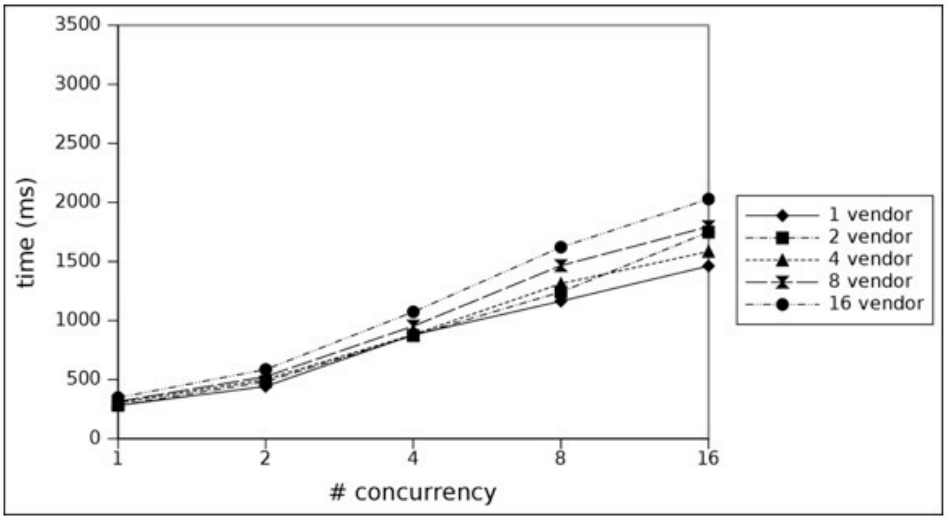
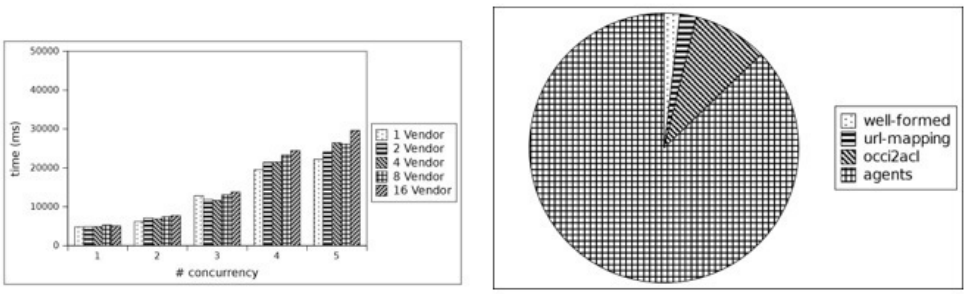


Figure 8
Average time for provisioning transaction



(a) Total time for 100 transactions **(b) Time distribution for message processing with high concurrency**

Figure 9
Scalability analysis results

9. Conclusion

Within the activities of the european research project FP7 mOSAIC it has been designed and developed a multi-agent platform for provisioning, management, monitoring and reconfiguration of Cloud resources. In this work it has been presented the agents' infrastructure, by detailing the roles of each entity and the access methodologies to the offered services. Furthermore it has been presented an API and tools that allows the orchestration of the agent services according with the lifecycle for the governance of Cloud resources. It has been described how to access the provisioning and management functionalities and how to configure and use the monitoring infrastructure. In particular, the alert

triggering functionality at the occurrence of specific conditions could be useful in order to implement reconfiguration policies based on the verification of some critical conditions on one or more parameters of the Cloud infrastructure. In future works it will be possible using this functionality to design and develop autonomic agents for resources' reconfiguration in order to balance the infrastructure and solve the detected critical conditions.

References

[cos()] CoSSMic project. URL http://www.cossmic.eu.

[mod()] MODAClouds project. URL http://www.modaclouds.eu.

[spe()] SPECS project. URL http://www.fp7-specs.eu.

[Aversa et al.(2010)] Rocco Aversa et al. Cloud agency: A mobile agent based cloud system. In *Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on*, pages 132–137. IEEE, 2010.

[Buyya et al.(2009)] Rajkumar Buyya et al. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation computer systems*, 25(6): 599–616, 2009.

[Cao et al.(2009)] Bu-Qing Cao et al. A service-oriented QoS-assured and multi-agent cloud computing architecture. In *Cloud Computing*, pages 644–649. Springer, 2009.

[Cretella et al.(2012)] Giuseppina Cretella et al. Towards a semantic engine for cloud applications development. In *Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on*, pages 198–203. IEEE, 2012.

[Di Martino et al.(2011)] Beniamino Di Martino et al. Building a mosaic of clouds. In *Euro-Par 2010 Parallel Processing Workshops*, pages 571–578. Springer, 2011.

[Edmonds et al.(2011)] Andrew Edmonds et al. Open Cloud Computing Interface in Data Management-Related Setups. In *Grid and Cloud Database Management*, pages 23–48. Springer, 2011.

[Emeakaroha et al.(2010)] Vincent C Emeakaroha et al. Low level metrics to high level SLAs-LoM2HiS framework: Bridging the gap between monitored metrics and SLA parameters in cloud environments. In *High Performance Computing and Simulation (HPCS), 2010 International Conference on*, pages 48–54. IEEE, 2010.

[Kertesz et al.(2009)] Attila Kertesz et al. An SLA-based resource virtualization approach for on-demand service provision. In *Proceedings of the 3rd international workshop on Virtualization technologies in distributed computing*, pages 27–34. ACM, 2009.

[Metsch et al.()] Thijs Metsch et al. SLA Agreement, Negotiation, Execution and Monitoring using OCCI.

[Metsch et al.(2010)] Thijs Metsch et al. Open Cloud Computing Interface–Core. In *Open Grid Forum, OCCI-WG, Specification Document.*, 2010.

[Venticinque et al.(2012)] Salvatore Venticinque et al. Agents based cloud computing interface for resource provisioning and management. In *Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on*, pages 249–256. IEEE, 2012.

[You et al.(2009)] Xindong You et al. Ras-m: Resource allocation strategy based on market mechanism in cloud computing. In *ChinaGrid Annual Conference, 2009. ChinaGrid'09. Fourth*, pages 256–263. IEEE, 2009.

Biographies

Rocco Aversa graduated in Electronic Engineering at University of Naples in 1989 and received his Ph.D. in 1994. He is Associate Professor in Computer Science at the Second University of Naples. His research interests include: the use of the mobile agents paradigm in the distributed computing; the design of simulation tools for performance analysis of parallel applications; the development of middleware to enhance the Grid and Cloud computing platforms. He is Associate Editor of International Journal of Web Science. He participated to various research projects supported by national organizations and by EC in collaboration with foreign academic institutions and industrial partners.
email: rocco.aversa@unina2.it

Luca Tasquier is PhD Student in Computer and Electronic Engineering at Department of Industrial and Information Engineering of the Second University of Naples. He received his Master Degree in Computer Engineering in 2011. He is involved in research activities dealing with parallel and cloud computing and mobile/intelligent agents for distributed systems. He participated to research projects supported by international and national organizations.
email: luca.tasquier@unina2.it

Salvatore Venticinque is Assistant Professor at Second University of Naples since 2006. He has been lecture of "Computer Programming" and "Computer Architecture" in regular academic courses. He is involved in research activities dealing with parallel and Grid computing and Mobile Agents programming for distributed systems. He is author of more than 100 publications in international journals, books, and conferences in collaboration with national research organizations and foreign academic institutions. He participated to research projects supported by international and national organizations.
email: salvatore.venticinque@unina2.it

In Store Augmented Reality: Retailing Strategies for Smart Communities

M. T. Cuomo, D. Tortora, G. Metallo

Abstract. *The multi-channel strategy becomes crucial for the competitiveness of retail system; it is based on open trustworthy relationships, on the support of new communication tools (e.g. social media), and on innovative devices (24/7 connections), which extend possibilities, process and moment of purchasing. Thanks to the support of innovative information technologies, the store constitutes a privileged area of integration between real and digital, where brand management strategies confront with new social spaces. Augmented reality add different degrees of information to the consumers' sense. So as the augmented reality reshapes the commercial area, providing it for contextual information and activable by potential customers when needed; at the same time, the integration with the mobility reconfigures the mode of use, organizing new opportunities of connections with the user. In addition, the generation of contents both bottom-up and top-down makes the individual from tryer into buyer into advertiser, through social networking, generating greater value experiences and, therefore, additional sales.*

Keywords: Open Data, Smart Disclosure, OAuth.

1. Introduction

The capacity for "contextualizing" – purchase offers, consumer goods, content value of brands and interactions between individuals/consumers and the product – is a distinguishing trait of the pervasive retailing *formula;* 'pervasive' - no longer invasive - in the sense that it is activated by the consumer. Consumers use augmented reality to identify and define new opportunities of interaction. In other words, they find advertising or commercial activity (or are contacted by other users) through geo-localized information, multimedia content or indications from websites and social networks (Riva, 2010; Cuomo et al., 2011), from instructions relative to the product displayed in the store, from tweets, tagging objects, noting news in and on places that particularly impact on consumer perceptions, within a new system of social sense making. Moreover, by fully comprehending the strategic capacity of such changes and potential evolution – where trust relationships based on open mode relations, new communication tools (e.g. social media), innovative devices and 24/7 connections (the mobiles

user is by definition always on), extending possibilities, processes and time of purchase, making multi-channeling a fundamental factor for competitiveness in the *retail* context – can be fully understood only in the light of re-thinking the store as socially constructed context, capable of creating innovative consumer relations. The retail scenario is thus transformed from a space, or an area in which place elements prevail to a place designated by the practices and experience that defines it (Accoto e Mandelli, 2012).

2. Augmented reality and new marketing scenarios

A billion smartphones predisposed for augmented reality by 2016, compared to the 150 million currently in use (Editoriale, 2013a) in the retail commerce sector, already constitute devices capable of governing enabling mechanisms, destined to create/stimulate stories about the products, excite interest and generate shelf emotions/experiences, creating new modes of interaction between individuals and the surrounding reality in terms of consumer experience. At the same time, the technological evolution (Fig. 1) – both as concerns the introduction of analogical elements (sound recognition, computer vision) within the digital world and the creation of inter-connecting reality in the digital world, semantic analysis and concision of available information in the real world – and the proliferating of applications with easily adaptable interfaces, signify and services ever more specialized and customized (Frà et al., 2011).

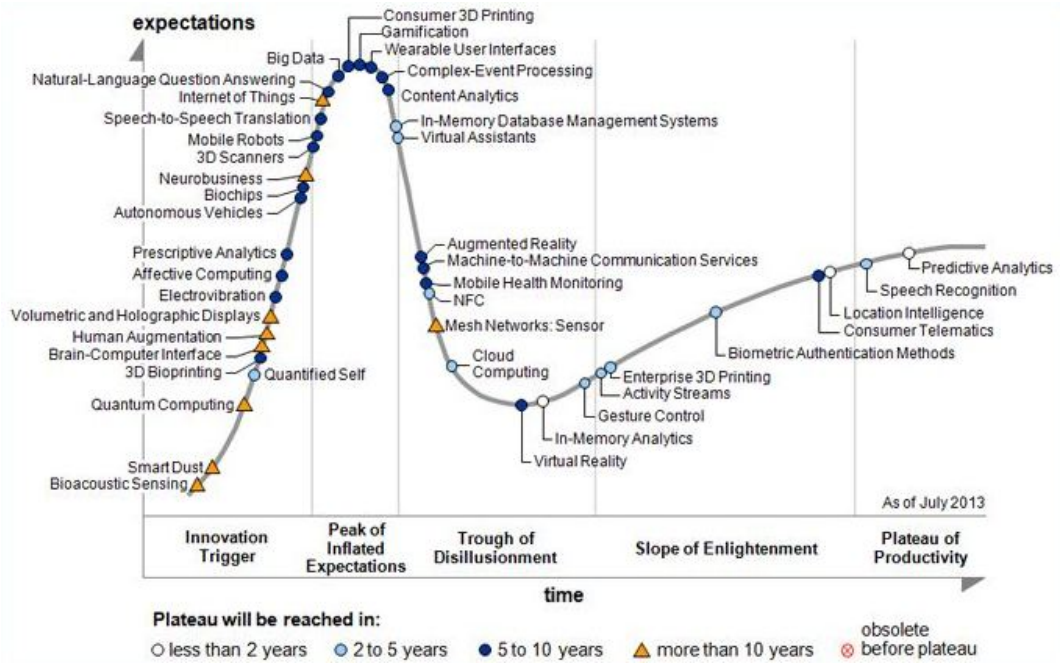


Figure 1
Gartner's technology hype-cycle

In fact, thanks to the support of such innovative information technologies (since the 1990s, Quick Response Codes have conveyed quantities of information on the goods present in the retail area, with direct links to the manufacturer's website – Editoriale, 2013b) the retail store has become a privileged context of integration between real and digital, generating narrative levels to which brand management strategies are applied and where new relational spaces and sociality (Mandelli, 2011) are created.

First appearing in the literature in the 1940s but becoming widespread only in the late 1980s, the term augmented reality (AR – used in 1992 by the researcher Thomas Preston Caudell, a Boeing engineer, to describe a system of new generation technology for assembling and installing electric cables in airplanes – Caudell e Mizell, 1992) indicated «a term for a camera enhanced view of a physical real-world environment, where virtual elements are merged with the real-life scenes creating a "mixed reality" of virtual elements and the real world. The "virtual elements", given their nature, can consist of anything: 3D models, video, web information…anything. The point here is that your mind is the only boundary» (Trubow, 2011a, p. 4). In other words, a monitor and a video are sufficient to integrate real life into a virtual context.

In any event, although seen as the extreme end of a *continuum*, compared to virtual reality (VR), in which information added or removed electronically is predominant and the consumer is led towards a "cancelling" perception of his emotional exploring of the context or situation, in AR, purchaser perceptions are solicited by the addition of information and emotional levels (through multimedia content, i.e. video, audio and animations), enjoyed "unrestrictedly", and at the same time, transferable to other users – once modified by the former, who has meanwhile in his turn, added virtual content. Such content in transit towards social media platforms, generates links with the brand – adding "self produced" value to a shopping experience of a full immersion type (Pine e Gilmore, 2000; Tortora, 2007).

Evident output defines this model of interactive marketing – based on the contribution of the community of reference, decreeing obsolete the centrality of mass media in communications of product/brand – and redesigning the strategic interest surrounding the creation of interactive and social experience thus conferring on the product greater value (Mardegan et al., 2012; Arvidsson e Giordano, 2013). The product in other words is deflected from its strictly economic value as it is «created within a personal experience that is shared and which embraces the dimensions of identity, the feeling for and belonging to a community» (Riva, 2012, p. 214). Brand equity is thus redefined (Cuomo et. al, 2009). As a result, in order to survive in this "new age of marketing" businesses have to identify strategies of (digital) engagement, i.e. change their interaction tactics with consumers within a graded scale of persuasion, relations, experience and sharing (Scatena e Mardegan, 2012). Nowadays, brand knowledge, reliability, communication and widespread diffusion are no longer sufficient to guarantee a brand's success. Consumers require total involvement in the consumption experience, and the governance of brand interaction has to be redefined ad hoc in the specific socio-constructed spaces of the environment, technology, service and sociality mix (Fig. 2).

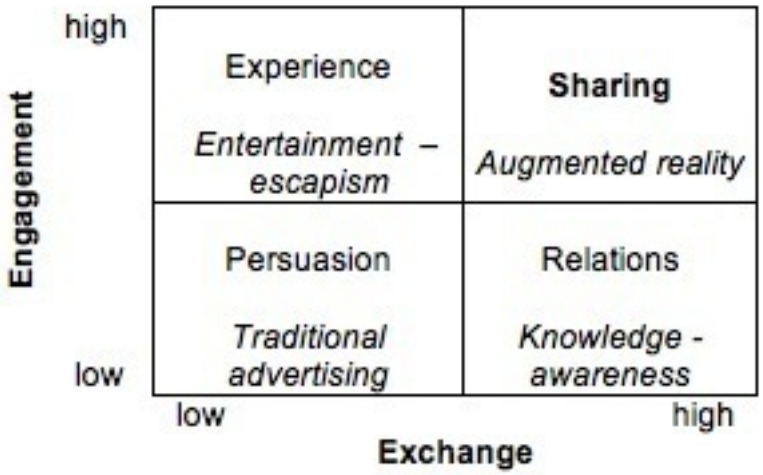


Figure 2
Tools for promoting consumer involvement

On the other hand, recent research (Scatena e Mardegan, 2012) indicates a positive and significant correlation between the levels of *engagement* promoted by a brand and annual turnover and profits achieved by the firm, especially during a recession. Consequently, AR offers marketers new opportunities of approaching and involving clients particularly in branding terms. Customization and exclusivity of the message and the timeliness, placing and contextualizing of the user experience are already becoming essential elements of the media platforms for repositioning products in the sense and value making process.

The opportunity to benefit from augmented reality applications using the standard Internet browser suggests a series of new scenarios in corporate-client relations (Trubow, 2011b, p. 5 – Fig.3).

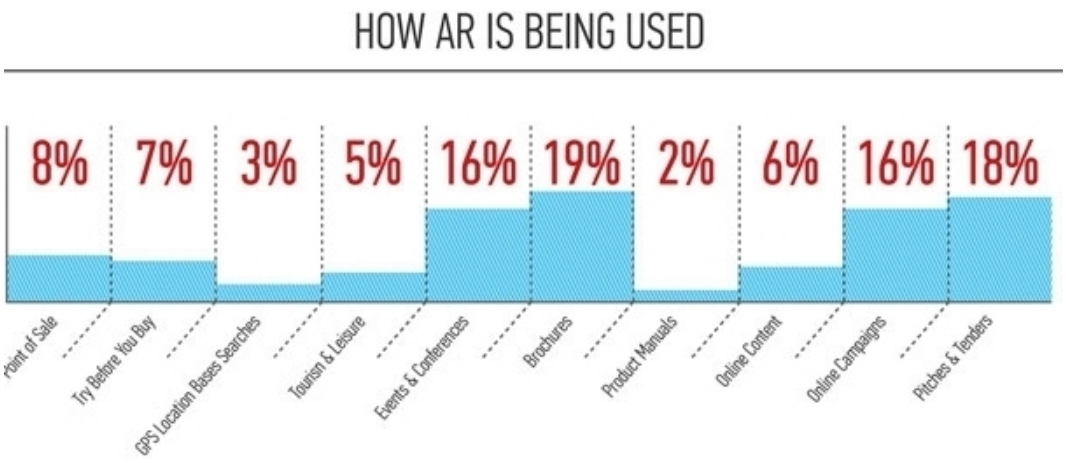


Figure 3
Uses based on current adoption by major brands

Emblematic in this sense is the development recorded of augmented advertising i.e. AR advertising and promoting of the main corporate brands. Reference – in order to fully comprehend the difference in terms of participation and the cognitive and emotional impact of the message, is to the recent application of AR Aurasma (relative to the app economy, Kim et al., 2011) where real world images are recognized and overlapped into real time by means of a virtual layer reproducing multimedia content relative to the captured images: thus when a registered user for instance, frames and snaps a poster of a film with his smartphone, Aurasma recognizes the image (which has to be included the database of the tags Super Anywhere of the application) and reproduces the trailer of the film. Users can create personal *anywheres,* associating to the photo of a place or object, multimedia content on the theme to share it with other users (Filardo et al., 2012).

Retailers are now resorting frequently to AR "to bring to life the retail store". In effect, AR technology finds varied application in the selling process (Trubow, 2011a – Fig.4).

Figure 4
Augmented reality and sales strategy

Providing new impulse to the interpretation of the AIDA model, as concerns attracting consumer interest for the product by means of a selling proposition which shows its "real" problem solving capacity, the consumer is supported in the buying and consumption process. One of the most relevant advantages of AR lies in its capacity for "direct interaction" with the product, an aspect traditionally linked to the presence both of consumer and product in a real life context i.e. the store, and such interaction satisfied by contact with the same. Many studies point out that by touching an object, customer intention to buy is increased as well as their willingness to pay a higher price (Trubow, 2011a). By means of AR it is possible within the retailing area, to increase product-consumer interaction even more as opposed to mere physical contact, where full immersion in the experience is facilitated (Addis, 2005) thanks to the support of additional information, movement, details on the assembling of the parts,

indications of what the product is made of, how it works and, the contextualization or preview of the product, whereby the potential purchaser becomes interpreter and co-designer and not merely the beneficiary of the end product (Tab.1).

LEGO	Lego uses various terminals of augmented reality – Lego Digital Box – which enable customers to see all the elements of toy in detail in 3D, merely by placing the box in front of a video camera (www.korus.fr).
BRAUN	To launch the new models of its Series 5 razors, Braun used AR for the first time to explore the product virtually by means of gestures. The application, available on their native website, captures hand movements with a webcam and uses them to control a 3D model of the razor without the aid of a mouse. The user can also set in motion content for individual parts of the new models with access to a virtual shopping trolley (www.realta-aumentata.it/home.asp)
Mercatone Uno	By downloading the free application Aurasma Lite and framing the cover of the 2012 catalogue and the other pages marked with the logo "A", the catalogue comes to life, with video and images narrating the Mercatone Uno world. News, details and extra content are provided by means of direct interactive consultation by smartphone (www.realta-aumentata.it/home.asp).
i'm lovin' it®	To resolve the problem of criticism relative to the quality labeling and tracking of food used by the Australian multinational, an AR application has been devised called "Track My Mac", which shows customers by means of creative features in movement relative to basic ingredients, information on the farmers. Well known images and GPS are exploited for this purpose (www.realta-aumentata.it/home.asp).
kinder®	To enrich consumer experience both as concerns the retail store and to enhance the range of surprises, Kinder devised supplementary toys for children for Easter 2013, exclusively for some particular supermarkets. Near the Kinder GranSorpresa Easter stand, a panel and monitor were installed. One of the Kinder GranSorpresa eggs was put in front of the webcam and surprise, surprise, was transformed into one of the surprises the egg itself contained, to emphasize Kinder excellence in terms of quality and value. Each Kinder GS is a world of surprises: a real opportunity for children to "lose themselves" in play (www.saporinews.com).
Galeries Lafayette	During the event "fashion 24h/24, 9-12.10.2012, Galeries Lafayette, Lille, installed interactive changing rooms near their counters for customers to try on in real time, a number of outfits virtually. The virtual mirror in the changing rooms recomposed the image of the client standing in front of it, automatically detecting movement. By means of a simple gesture, the customer could change outfits, colours or styles, while the clothes themselves changed or adapted according to the customer's size, shape and gestures (www.korus.fr).

Table 1
Augmented in-store experience: some examples

Utilized as a part of an advertising strategy or to stimulate customers to come to a store, or even to increase brand awareness and customer loyalty, dilating the purchasing experience into environments of mixed reality, AR tends to redesign retail space by promoting a different mode of customer perceptions and sense making, influenced closely by cultural and social processes, including elements of an anthropological or semiotic nature. Stand alone installations in a retail store for example, are stands where customers and other subjects not present on the scene, interact virtually and exchange information (value exchange). In this sense reality is socially augmented and the shopping experience itself also becomes a social experience.

The retail arena as the context in which a shopping experience is made, is part of a dual category: situational, in which signals in an analogical form (e.g. recognized images or sounds) become the input for subsequent elaborations that tell us what is happening, what elements are present in reality and what information corresponds to the virtual reality surrounding the consumer i.e. inbound movement and what is conversational, where information comes from the social media, concentrating on communication surrounding reality, i.e. outbound movement (Frà et al., 2011). In any event, while in inbound mobility(towards the subject) firms continue to hold a pre-eminent role, predisposing and guiding the subject-context interaction process, as concerns the outbound transfer, the control of mass media communication content by an ever growing number of actors widespread in spatial terms and/or close knit in virtual communities has become the norm – human broadcasters (Riva, 2012). Thus the confines of customer experience are redefined and extended. Links (mobility), creation of multimedia content (above all photos and videos) by sharing opinions on the product, reviews and comments on specific content) the use of technology to reinforce the sense of belonging to the community of reference, characterizing the consumer-product relation for the "C Generation" (De Felice, 2011). Notwithstanding, above all with reference to mobile marketing, but applicable to the shopping experience promoted by full immersion technologies generally, numerous studies highlight how convenience – linked to the hedonistic value that can be derived – is still a fundamental factor in using "augmented" services and as concerns mobility (Fig. 5).

The co-existing of functional and emotional factors is acknowledged in the studies on the mobile internet, mobile data service, mobile multimedia service and location-based mobile service (Varnali e Toker, 2010).

Furthermore, precise components defining customer experience in full immersion stores can also be classified (see Table 2 – Pantano e Servidio, 2012, p. 283).

In short, the creative force of both bottom-up (governed by firm or brand) and top-down (creation and sharing of multimedia content) thanks to a digital grammatical basis that guarantees interaction between the parties, transforms the subject-consumer:

- from *Trier,* using AR to test the augmented product before buying in augmented contexts in terms of quality, problem solving capacity and promise,
- to *buyer*, relying on the initial value proposition, contextualized experience, aggregation dynamics and at the same time, privacy and safety of transactions,
- to *advertiser*, thanks to human broadcasting communication which via social networking, participates in the creating and spreading of the brand story, spontaneously motivating and supporting (if results are satisfactory) other potential *triers-buyers-advertisers*, thus generating on the whole, added value experience and as a consequence, more sales for the firm.

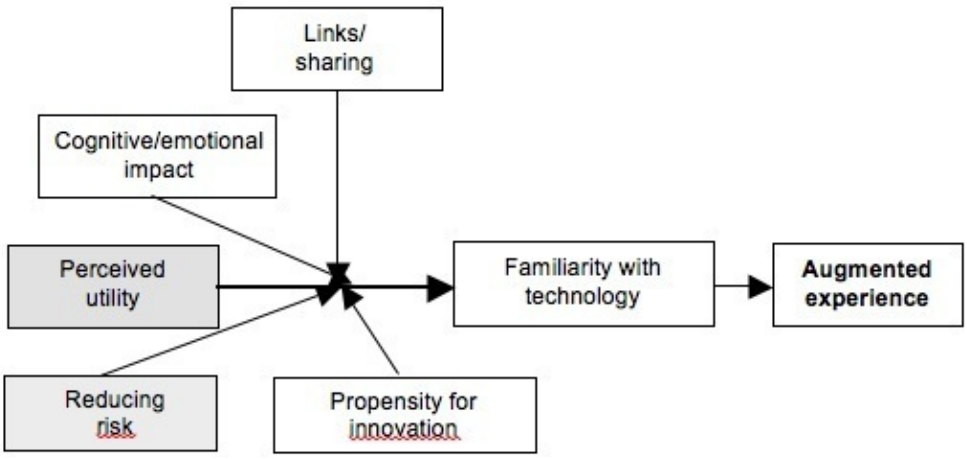


Figure 5
A conceptual model for using AR

Facilities	Fast response, secure transaction, system flexibility, entertainment
Product information	Product variety, detailed product information, personalized information
Service	Product selection assistance, virtual sales person, online support
Convenience	Details about the firm, navigational efficiency, more realistic navigation and interaction
Appearance	Pleasant, attractive, more realistic appearance
Institutional factor	Requirements related to consumer's profile, firm's information, consumer's position tracer while in the immersive store

Table 2
Full immersion stores

3. Critical reflections and implications for future research

As we have said, full immersion technologies and digital tools especially those which are always on, have become an "exciting" opportunity for promoting retail stores, for tracing or tracking purposes and for increasing customer attendance. AR, above all by virtue of the widespread diffusion of mobile devices, represents an ever greater means (multiplying factor) of access both to brands and to or for consumers (Rohm et al., 2012). It should be noted however, that their use could conceal negative impacts that it might be worthwhile to indicate at this stage.

Above all, additional information, animations, detail, technical analysis etc., characterizing contexts of augmented reality could easily generate informational entropy (chaos) for the consumer, impacting negatively on his/her decision-making processes, thus distorting to a certain extent, the brand/consumer link. In this context, the fundamental role of the brand as a mark of guarantee or reliability capable of simplifying the buying process, should not be ignored. On the contrary in a context of AR where the amount of information hinges on the excessive, together with the incapacity or difficulty in selecting pertinent information, could have deviating or even counter-productive effects. It might therefore, be opportune to structure "augmented reality stores" articulated on a series of factors effectively capable of attracting clients and supplying them with relevant informational content, necessary and efficacious for an extremely positive shopping experience. In short, AR cannot be considered a vital tool to be used at any cost!

Furthermore, the limits of using AR that could derive from the self-production of information on the part of consumers should be highlighted, taking into account the concomitant risk of "manipulations/alterations" of such information in *social* environments, the effect of the generating force of content created and shared *(*top-down logic*)*. For AR to add impulse to a value experience, it has to (i) concern themes which are relevant for the buyer, (ii) provide immediately recognizable incentives-benefits, (iii) be simple to use regardless of the technological limits typical of mobile devices (Prunesti e Lalli, 2011), and (iv) promote cooperation and synergies relative to all the actors involved in the value chain with the focus remaining on firm/customer interaction.

However it is fundamental for organizations to have a solid theoretical basis underpinning their defining of a model of augmented (in store) customer experience, linked to an efficient monitoring system of the additional data selected on the basis of the consumer context (situation, event, etc).

At the same time, the applicability of AR to the diverse types of purchasing generally should be weighed up. Intuitively, it can be stated that AR is more effective in scenarios of problematic or complex purchasing, while for the other kinds- e.g. spur of the moment purchases- "Diminished Reality" (DR) might be preferable

Furthermore, the implications should be considered of putting in place augmented/diminished scenarios for identifying market segments/targets when selecting high/low levels of informational content (Fig. 6):

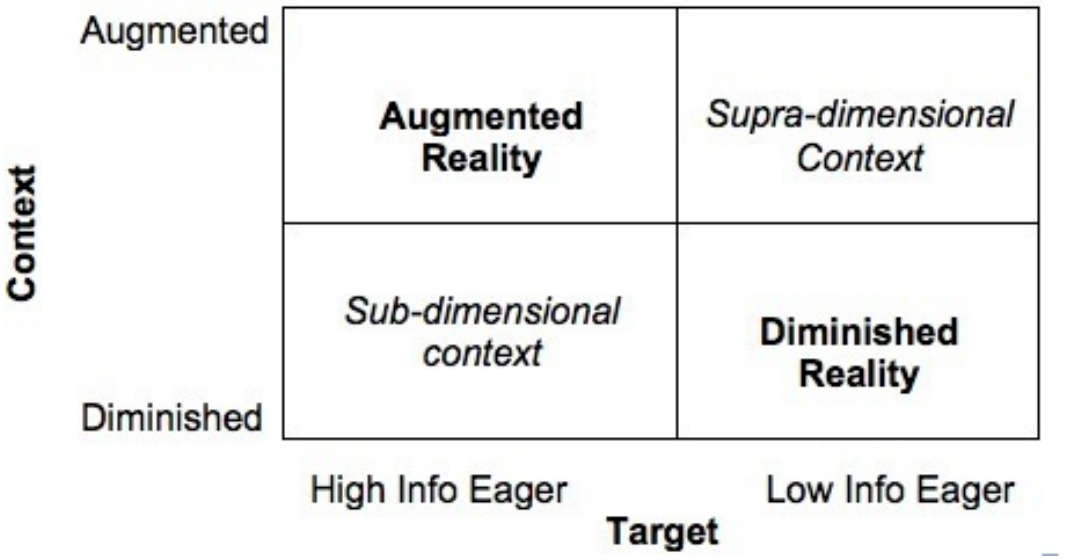


Figure 6
Market segments in AR/DR contexts

In conclusion, AR is in any event "a tool" within a wider and a more complex brand's overall customer communication ecosystem, in which creativity, technology and content have to combine traditional and innovative forms and relative tools to generate efficacious customer engagement (Schultz e Block, 2011). In this context, consequently, scientific research should proceed to addressing the empirical validation of the conceptual models proposed (outlined in this as in other studies) in order to offer a valid theoretical foundation for augmented marketing planning and processes.

References

Addis, M., L'esperienza di consumo. Analisi e prospettive di marketing, Pearson- Prentice Hall, Milano, 2005.

Arvidsson, A., Giordano, A., Societing reloaded. Pubblici produttivi e innovazione sociale, Egea, Milano, 2013.

Caudell, T.P., Mitzell, D.W., Augmented reality: an application of heads-up display technology to manual manufacturing processes. System Sciences, 2, 1992, 659-69.

Cuomo, M.T., Metallo, G., Tortora, D., Opportunità, limiti e criticità dei social network. Esperienze d'impresa, 2, 2011, 25-48.

Cuomo, M.T., Metallo, G., Tortora, D., Testa, M., Kitchen, P.J., Building brand equity: the genetic coding of Mediterranea brands. Euromed Journal of Business, 4, 3, 2009, 237-253.

De Felice, L, Marketing conversazionale: dialogare con i clienti attraverso i social media e il Real-Time Web di Twitter, FriendFeed, Facebook, Foursquare, Il Sole 24 Ore, Milano, 2011.

Editoriale, Realtà Aumentata e QR code a confronto. www.arnews.it, 25 gennaio, 2013b.

Editoriale, Realtà aumentata: suo impiego nel mondo retail. www.puntodivendita.info, 11 febbraio, 2013a.

Filardo, V., Messina, M., Bortolussi, S., Marino, S., Megna, L. (a cura di), Augmented reality. Nuove applicazioni, nuove soluzioni, in La comunicazione multimediale, 2012, http://www.arproject.altervista.org/intro.html.

Frà, C., Lamorte, L., Martini, G., Dall'augmented reality al check-in. Notiziario tecnico Telecom Italia, 3, 2011, 20-31.

Kim, H.W., Lee, H-L., Son, J.E., An exploratory study on the determinants of smartphone app purchase, in Proc. 0f the 11th International DSI and the 16th APDSI Joint Meeting, Taipei, Taiwan, July 12 – 16, 2011, 1-10.

Mandelli, A., Accoto, C., Social Mobile Marketing. L'innovazione dell'ubiquitous marketing con device mobili, social media e realtà aumentata, Egea, Milano, 2012.

Mandelli, A., Processes of Value Creation in Markets as Mediated Conversations.Working paper, 2011.

Mardegan, P., Riva, G., Pettiti, M., Mobile Marketing: la pubblicità in tasca, Lupetti Editore, Bologna, 2012.

Pantano, E., Servidio, R., Modeling innovative points of sales through virtual and immersive technologies. Journal of Retailing and Consumer Services, 19, 2012, 279–286.

Pine, J.B. II, Gilmore, J.H., L'economia delle esperienze. Oltre il sevizio, Etas, Milano, 2000.

Prunesti, A., Lalli, F., Geolocalizzazione e mobile marketing. Fare business con le App e i social game, Franco Angeli, Milano, 2011.

Riva, G., Digital Marketing 2.0. Multicanale, Sociale, Esperienziale, Mobile. Micro & Macro marketing, 2, 2012, 213-218.

Riva, G., I social network, Il Mulino, Bologna, 2010.

Rohm, A.J., Gao T., Sultan, F., Pagani, M., Brand in the hand: A cross-market investigation of consumer acceptance of mobile marketing. Business Horizons, 55, 2012, 485-493.

Scatena, S., Mardegan, P., Mobile Marketing: stato dell'arte e applicazioni pratiche. Micro & Macro marketing, 2, 2012, 219-235.

Schultz, D.E., Block, M.P., Understanding customer brand engagement behaviors in today's interactive marketplace. Micro & Macro marketing, 2, 2011, 227-244.

Tortora, D., Experience marketing e creazione di valore. Relazioni e interazioni tra consumatore, offerta e contesto, Giappichelli editore, Torino, 2007.

Trubow, M., Augmented reality marketing strategies: the how to guide for marketers. Hidden Creative Ltd, www.hiddenltd.com, 22 marzo, 2011b, 1-18.

Trubow, M., Sales technology: selling with augmented reality. Hidden Creative Ltd, www.hiddenltd.com, 5 settembre, 2011a, 1-14.

Varnali, K., Toker, A., Mobile marketing research: The-state-of-the-art. International Journal of Information Management, 30, 2010, 144-151.

Biographies

Maria Teresa Cuomo, PhD. She is Associate Professor and teaches Management & Business Development and Marketing at the University of Salerno and Global Business at the University of Milan "Bicocca". Her numerous research interests range from Marketing & Market Research to International Management & Finance. She has published several papers as well as numerous studies. She regularly participates at international Conferences as speaker. She is Member of the Business School's PhD's Joint Teaching Committee and of international editorial board journals. She is also widely engaged in Applied Research and Consultancy.
email: mcuomo@unisa.it

Debora Tortora, PhD. She is currently a research fellow in Marketing at University of Salerno. Her research interests are focused on Consumer Behavior, Brand Management and Corporate Reputation, as shown by several articles published on Italian and International Reviews and Conference Proceedings. In addition to this she carries out consulting activity for local companies.
email: dtortora@unisa.it

Gerardino Metallo. He is Full Professor of Management & Business Development and Finance at the University of Salerno. His research interests range widely from Business & Finance issues, where he has published numerous works, and more general Management, where he has published several articles in International Journals and numerous other studies. Besides his commitment to academic and Scientific Research, he is also widely engaged in Applied Research and Business Consultancy.
email: gemetall@unisa.it

Haptic Rendering of Deformable Surfaces in Medical Training

A. F. Abate, A. Casanova, M. Nappi, S. Ricciardi

Abstract. *Haptic systems applied to medical simulation and training can provide users with crucial kinaesthetic and tactile info otherwise impossible to convey. As the typical objects involved in this kind of simulation are very often deformable tissues, one of the main challenges in this area is the perceptually believable reproduction of these structures through haptic rendering techniques. In this paper we propose the use of a colour or grey scale bitmap to associate local deformability info to the geometry. This approach allows the visual-haptic engine to modulate the resistance to compression exerted by the simulated tissues based to a local parameter. By using multiple layers of textures or animated textures, locally non-linear or even dynamic behaviours can be simulated with a low computational load. Preliminary experiments based on a Immersion Cyberforce hand-based haptic device are encouraging.*

Keywords: haptic systems, haptic rendering, simulation and training.

1. Introduction

Haptic systems have a remarkable potential for many applications, specially those involving a specific tactile know-how. Indeed, medical applications such as tele-surgery or surgical simulators may particularly benefit from haptic interfaces, but their efficacy depend on the realism of the visual-haptic perceptions provided to the system's user. To this regard the simulation of the contact with different structures or tissues characterized by different deformability (i.e. reaction to the contact forces) can represent one of the key advantage of haptic based interaction compared to conventional visual interaction, as it provides user with a level of info not derivable otherwise. Unfortunately, contact modeling in medical simulation is a challenging problem. The way contacts are handled plays a very important role in the overall behavior of the interacting objects. The kind of contact model adopted (including friction or not), highly influences the post-impact motion of the interacting objects. In most simulators with haptic feedback, the collision response of soft tissues with a virtual surgical instrument is assumed to be very local: the interaction only consider a single point [Mahvash and Hayward, 2004] [Mendoza et al., 2002].

The most popular approach is the penalty method which consists in defining a contact force f = k at each contact point where is a measure of the interpenetration between a pair of colliding objects, and k is a stiffness parameter. This stiffness parameter must be large enough to avoid any visible inter-penetration, however, its value cannot be determined exactly. In addition, if an explicit time integration scheme is used, and k is large, very small time steps are required to guarantee the stability. The quick growth of energy in the haptic control loop induced by the method often leads to excessive damping in the provided solutions. A possible improvement over the penalty method can be achieved through the use of an implicit integration scheme [Meseure, 2003]. Yet, solving the resulting stiff and non-smooth system can be computationally prohibitive when the objective is to reduce as much as possible the inter-penetration distance. Some methods, developed for force feedback applications are based on the avoidance of visible inter-penetration through constraint-based techniques. The collision can be prevented by geometrical deformation constraints like in [Picinbono, 2002] or [Forest et al., 2004], or by god-object [Zilles and Salisbury, 1995] and proxy [Barbagli et al., 2003] methods. Another way is the use of Lagrange multipliers, which are appropriate for handling bilateral constraints [Galoppo et al., 2006]. However, contacts between objects intrinsically define unilateral constraints, which means that physics is not always verified when using techniques based on Lagrange multipliers. As a consequence, colliding objects could stay artificially stuck at the end of the time step. Improvements over constraint-based techniques are possible by using a Linear Complementary Problem (LCP) formulation. The solution of the LCP gives an accurate description of the contact forces needed to zero out the interpenetration, and prevents objects to stick together [Pauly et al., 2004]. By expanding the LCP, or by using a non-linear solver, the formulation can be extended to model both static and dynamic friction for rigid [Anitescu et al., 1999] and deformable [Duriez et al., 2006] objects. Computationally efficient methods for solving linear complementary problems are proposed [Murty, 1997], thus making such approaches appealing even for interactive simulations.

Anyway, though contact modeling is a major concern in haptic systems, most of the methods described so far compute reaction force according to a stiffness parameter often defined at an object level (e.g. the whole object surface has the same stiffness) and even in case local stiffness parameters are available they approximate the real surface behavior with a very coarse granularity. The same consideration may be applied to the law which approximates the typically non-linear behavior of the object's surface when compressed: whatever the law adopted, it is the same for the whole surface.

Unfortunately, the consistency of many organic tissues is complex to describe at a global level and is often related to their healthy or pathological condition, providing crucial info to the specialist during palpation. As one of the main aims of visual-haptic simulators is to replicate the most faithfully possible the perceptual aspects of the interaction, we believe that a greater attention should be put to the simulation of those surface characteristics which may enhance the "haptic knowledge" of the system's users.

2. Map-based approximation of deformable surfaces

We propose a simple yet effective and efficient way to represent detailed information on local stiffness of a simulated surface by means of a dedicated bitmap providing a much greater accuracy than usual object-level attributes or even coarse array of parameters.

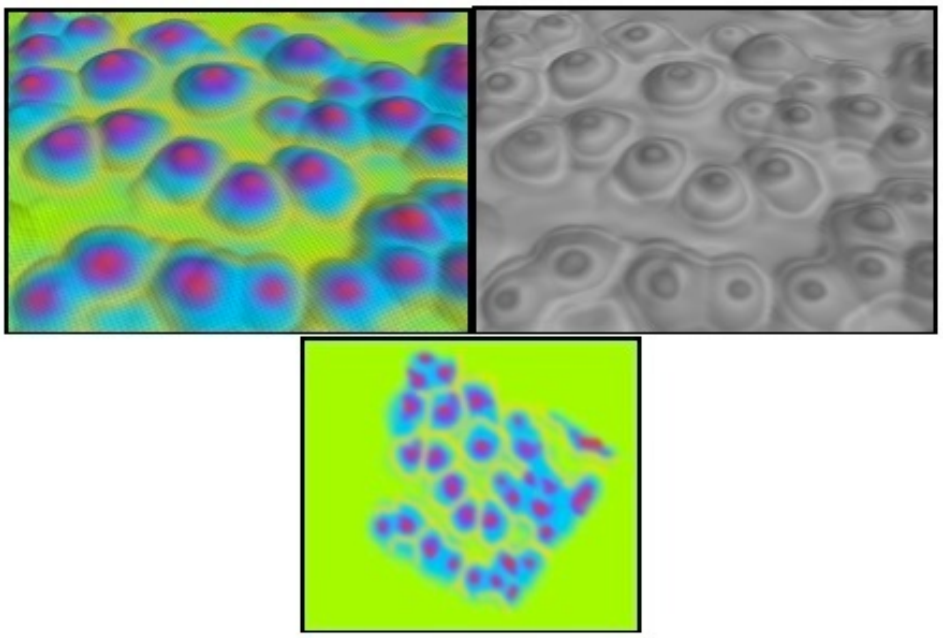


Figure 1
Color and grayscale deformability maps

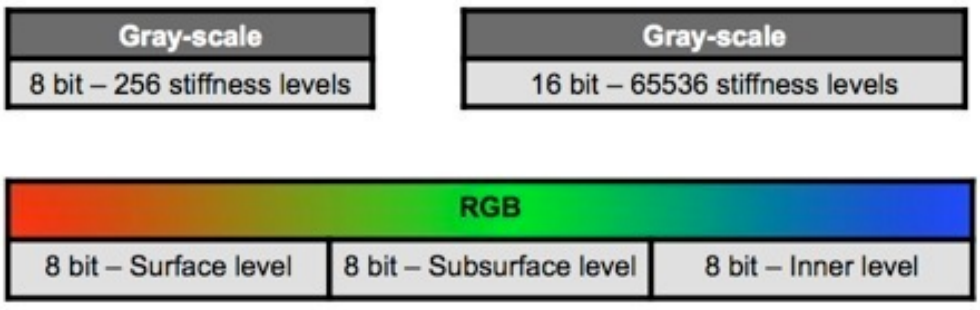


Figure 2
Various arrangements for stiffness encoding in bitmaps, by grayscale or colour maps. The former provide local stiffness at a pixel level, the latter exploits three values per pixel to modulatet stiffness according to three level of depht.

The basic idea behind this proposal is to exploit texture mapping (typically used to simulate visual properties such as ambient and diffuse color, transparency,

roughness, shininess, etc.) to associate local deformability data to 3D geometry. The local surface stiffness encoded by color at a texel level is then exploited by the haptic renderer to modulate the contact force feedback due to local tissue deformation resulting from interaction. More precisely, the deformability map is associated to mesh vertices through mapping coordinates in the form (u, v), previously projected onto the surface. The additional info can be represented through each pixel's RGB channels in a color texture or, in the simplest case, in a grayscale image, according to different arrangements offering a great flexibility of use (see Fig. 1). A deformability can be produced by a data driven methodology (for instance from image processing of diagnostic data or procedurally from anatomical models) or even by hand, by means of a 3D paint application.

In its simplest form an 8 or 16 bit grayscale image may encode the local stiffness parameters required to compute the reaction force at a texel level, thus providing a range of 256 or 65536 stiffness levels with a spatial granularity only depending by image's resolution. A 24 bit color image is able to arrange a more articulated set of data. Indeed it can store three 8 bit wide layers of stiffness data, enabling to simulate a non-linear surface behavior at a texel level. In other terms, it is possible to exploit the three (or even four if a 32 bit image is used) stiffness values associated to a given (u, v) position on the surface as a discrete approximation of the stiffness measured at a progressively greater depth (see Fig.2). Where required, by using multiple textures the approximation can be easily improved. As the deformability of a surface involves both visual and haptic feedback during simulation, but not necessarily these two channels (which in most applications are decoupled due to different frame-rate requirements) are supposed to share the same stiffness coefficient, mixed visual-haptic stiffness data may be embedded in the same 16, 24 or 32 bit deformability map (e.g. 8 bit visual + 8 bit haptic, or 16 bit visual + 8 bit haptic, or even 16 bit visual + 16 bit haptic). Moreover dynamic modification of the local stiffness can also be rendered by means of animated bitmaps or implementing a real time procedural processing of bitmap pixels in the RGB color space (e.g. a color shift may simulate a modification of the surface's deformation capability, see Fig.3). An additional advantage of this technique is related to its ultra-fast computing by means of modern GPUs. Indeed these vector processors can effectively process multiple very large (up to 8K pixel wide) textures on a single pass due to their highly parallel cores, while their dedicated VRAM (up to 4GB wide) could arrange many texture layers for each object in the virtual environment.

3. First experiments

The proposed technique has been experimented on a visual-haptic platform which is part of a wider research project aimed to the training of obstetricians to delivery [Abate et al., 2010]. A CyberForce® hand-based force feedback system by Immersion Corporation has been utilized in order to provide the user with haptic sensations during the simulated intervention. The Haptic Sub-System translates in terms of force feedback the output of the visual-haptic rendering engine, but it can also act as an input interface for function selection and command triggering. As motivated in *Section II.* the CyberForce® force feedback system by Immersion Corporation has been selected to provide the user with haptic sensations during the simulated intervention. The Cyberforce is composed by an articulated exoskeleton ending with a CyberGrasp system including a CyberGlove 22 sensors dataglove (see Fig. 4 and 5). The CyberGrasp provides the transmission of contact sensations to the operator's fingertips while the whole CyberForce simulates the weight and the inertia of the manipulated objects within its operative volume.

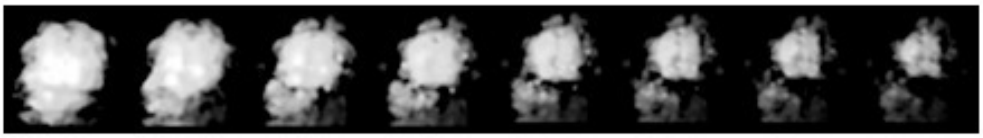


Figure 3
Animated map implementing spatial and temporal modification in surface stiffness

The weight of the grasped object can be rendered by the CyberForce through the application of a force on the back of the user's hand. The operator can, therefore, estimate the effort required to perform a particular task or the resistance opposed by a solid or deformable object with specific mass [Chryssolouris et al., 2000] since the penetration among objects gets cancelled by the reaction forces activated in relation to contacts among an polygonal object and the virtualized hand [Colgate et al., 1993]. This system is able to provide up to 8.8 N of force to the hand (through the CyberForce) and up to 12 N to each finger (through the CyberGrasp), values adequate to reasonably imitate the real forces involved during the delivery. It can also inherently measures hand's position and rotation (6 DOFs) with an accuracy of approx. one tenth of millimeter and one tenth of degree respectively. The main system limit, at least for our context of interest, is the lack of a torque feedback on the wrist and on the arm joints as well. This means that it can't replicate the resistance due to the rotation of an object (e.g. the child's neck) approximating its joint limits.

The test bed hardware included a dual quad-core Intel Xeon processor based Mac Pro workstation from Apple Inc., equipped with 8 Gigabytes of RAM and an Nvidia Quadro 5600 graphics board with 1,5 Gigabytes of VRAM. Overall the system's testers were positively impressed by quality of deformation of the simulated tissues (child's head and body, mother's perineal area), even if the

quantity of deformation was sometimes inadequate to imitate actual plastic phenomenon due to the compression exerted on the child's head. Nevertheless the system capability to provide (via the deformability map and the related haptic rendering) different perceptions of the deformability of the structures handled through the haptic device, was appreciated and considered one of the strong point of the on-going research for a realistic delivery simulator.

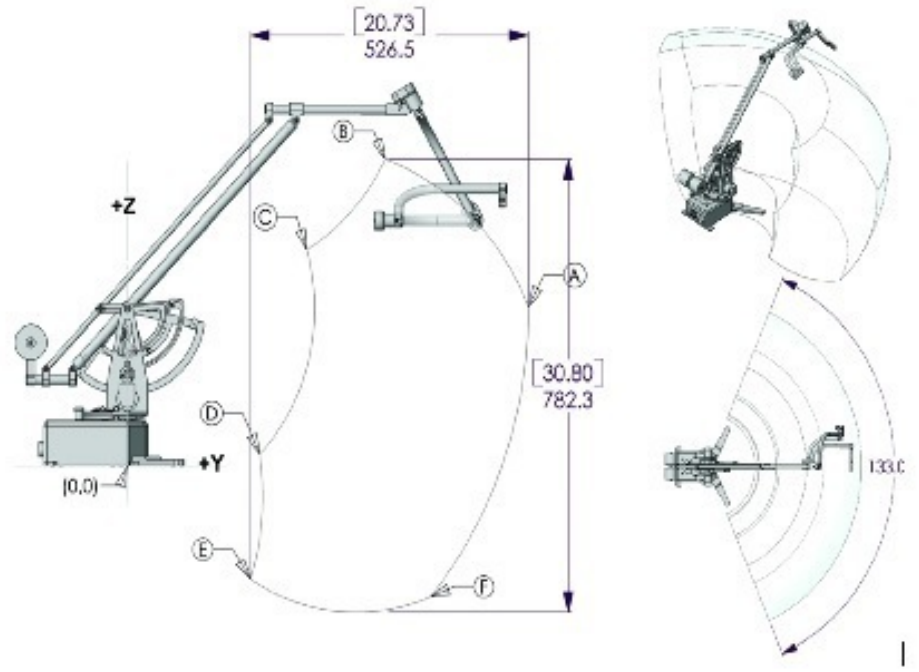


Figure 4
The Immersion CyberForce articulated exoskeleton and its operative volume.

4. Conclusions and future research

We presented a method to improve haptic simulation of deformable surfaces based on the use of a colour or grey-scale bitmap to associate local deformability info to the geometry. The proposed approach is computationally efficient as it involves simple calculations and it is inherently suited to be executed on GPUs. Its flexibility enable the approximation of non linear behaviour in the surface stiffness at a pixel level as well as for its dynamic modification. Besides performing an extensive testing of the proposed technique in different environments, we are exploring the possibility to use a 24 bit depth image to store a quantized stiffness vector (each RGB component is a vector component, instead of a single scalar value) which could be useful for improve surface deformation calculations during inter-object contact.

Figure 5
A view of the visual-haptic delivery simulator test-bed, featuring an Immersion CyberForce hand-based haptic device.

References

A.F. Abate, G. Acampora, V. Loia, S. Ricciardi and A. Vasilakos, A Pervasive Visual-Haptic Framework For Virtual Delivery Training, IEEE Transaction on Information Technology in Biomedicine, Volume 14, Issue 2, 2010, pp. 326-334.

M. Anitescu, F. Potra, and D. Stewart. Time-stepping for three-dimentional rigid body dynamics. Computer Methods in Applied Mechanics and Engineering, 1999, (177):183–197.

F. Barbagli, K. Salisbury, and D. Prattichizzo. Dynamic local models for stable multi-contact haptic interaction with deformable objects. Haptic Interfaces for Virtual Environment and Teleoperator Systems 2003, pages 109–116.

Chryssolouris G, Mavrikios D, Fragos D, Karabatsou V, A virtual reality-based experimentation environment for the verification of human-related factors in assembly processes. Robot Comput Integr Manuf, 2000, 16(4):267–276.

Colgate JE, Grafing PE, Stanley MC, Schenkel G (1993) Implementation of stiff virtual walls in force-reflecting interfaces. In: Proceedings IEEE virtual reality annual international symposium (VRAIS), Seattle, 1993, pp 202–208.

C. Duriez, F. Dubois, A. Kheddar, and C. Andriot. Realistic haptic rendering of interacting deformable objects in virtual environments. IEEE Transactions on Visualization and Computer Graphics, 12(1):36–47, 2006.

C. Forest, H. Delingette, and N. Ayache. Surface contact and reaction force models for laparoscopic simulation. In International Symposium on Medical Simulation, June 2004.

N. Galoppo, M. A. Otaduy, P. Mecklenburg, M. Gross, and M. C. Lin. Fast simulation of deformable models in contact using dynamic deformation textures. In SCA '06, pages 73–82, Switzerland, 2006. Eurographics Association.

M. Mahvash and V. Hayward. High-fidelity haptic synthesis of contact with deformable bodies. IEEE Computer Graphics and Applications, 2004, 24(2):48–55.

C. Mendoza, K. Sundaraj, and C. Laugier. Faithfull force feedback in medical simulators. In International Symposium in Experimental Robotics, volume 8. Springer, 2002.

P. Meseure. A physically based virtual environment dedicated to surgical simulation. In Surgery Simulation and Soft Tissue Modeling , 2003, pages 38–47.

K. Murty. Linear Complementarity, Linear and Nonlinear Programming. Internet Edition, 1997.

M. Pauly, D. K. Pai, and L. J. Guibas. Quasi-rigid objects in contact. In SCA '04, Eurographics Association, 2004, 109–119.

G. Picinbono, J.-C. Lombardo, H. Delingette, and N. Ayache. Improving realism of a surgery simulator: linear anisotropic elasticity, complex interactions and force extrapolation. Journal of Visualisation and Computer Animation, 2002, 13(3):147–167.

C. B. Zilles and J. K. Salisbury. A constraint-based god-object method for haptic display. In IEEE IROS '95: Proceedings of the International Conference on Intelligent Robots and Systems, 1995, pages 31–46.

Biographies

Andrea F. Abate received the Ph.D. degree in Applied Mathematics and Computer Science from the University of Pisa, Italy, in 1998. He now serves as Associate Professor of Computer Science at the University of Salerno. His research interests include computer graphics, virtual and augmented reality, haptics systems , human-computer interaction, biometrics and multimedia databases. He is currently the Co-Director of the VR_Lab - Virtual Reality Lab at the University of Salerno.
email: abate@unisa.it

Andrea Casanova is an assistant professor at Department of Computer Science - University of Cagliari. Lecturer of “Biometric Authentication” and “Software Engineering” for the course of studies of Computer Science and “Medical Informatics” for the course of studies of Medicine. Member of the GIRPR (Italian Research Group in Pattern Recognition) and MILab (Medical Image Laboratory, University of Cagliari). Current research interests are in the field of Biometric Authentication, Haptic System, Image Analysis and Processing, Human-Computer Interaction, VR/AR , Computer-Aided Diagnosis.
email: casanova@medicina.unica.it

Michele Nappi received the laurea degree (cum laude) in computer science from the University of Salerno, Italy, in 1991, the m.sc. degree in information and communication technology from I.I.A.S.S. "E.R. Caianiello," in 1997, and the Ph.D. degree in applied mathematics and computer science from the University of Padova, in 1997. He is currently an associate professor of computer science at the University of Salerno. His research interests include pattern recognition, image processing, image compression and indexing, multimedia databases and biometrics, human computer interaction, VR\AR.
email: mnappi@unisa.it

Stefano Ricciardi received the Laurea degree in Computer Science and the Laurea degree in Informatics from the University of Salerno. He has been co-founder/owner of a videogame development team focused on 3D sports simulations. He is currently a researcher at the Virtual Reality Lab of the University of Salerno. His main research interests include virtual and augmented reality, biometry, haptics systems and human-computer interaction. He is author of about sixty research papers including international journals, book chapters and conference proceedings.
email: sricciardi@unisa.it

A Layout-Analysis Based System for Document Image Retrieval

G. Pirlo , M. Chimienti, M. Dassisti, D. Impedovo, A. Galiano

Abstract. *This paper presents new system for document image retrieval, based on layout-analysis. The system, that is well suited for commercial form retrieval, uses Radon Transform for layout description and Dynamic Time Warping for document image matching. The experimental results, that were conducted using real and simulated data sets, demonstrate the proposed approach is effective and robust with respect to different classes of commercial forms and shifted/rotated document images.*

Keywords: Document management, Document Image Retrieval, Mathematic Morphology, Radon Transform, Dynamic Time Warping.

1. Introduction

Document retrieval is a very a critical task of current document management systems due to the exponential growth of the number of documents available in databases and digital libraries. Traditional document retrieval systems – based on set-theoretic, algebraic and probabilistic models - require a document to be present in text form and the querying method is based on a specific textual content in the document [Doermann, 1998; Manning et al., 2009]. Whatever the model used, text-based document retrieval systems require a document in text form, since the search for similar documents is based on comparing the textual contents. As a consequence, a preliminary stage of image to text conversion by an Optical Character Recognizer (OCR) is required when a document is in image form. OCR is a time-consuming error-prone process, specifically in the case of multi-lingual/multi-font documents and poor-quality document images [Marukawa

et al., 1997; Taghva et al., 1996; Lorpesti, 1996], as discussed in comprehensive surveys on this topic [Doermann, 1998; Mitra and Chaudhuri, 2000].

Along with the spreading of multimedia documents, it is useful to search a document on the basis of its structure and not only on the basis of its textual content. In such cases, methods adopted for document retrieval use feature vectors in which each feature is extracted from a specific region of the document image. For instance, some researchers used a static zoning strategy for document image decomposition to extact a fixed-size feature vector from the document image. In this approach, a regular grid is superimposed to the document image in order to extract regional characteristics [Tzacheva et al., 2002]. In another approach, a hierarchical zoning strategy was proposed to overcome the problem of optimal grid selection, in order to face with the treatment of set of documents of different characteristics [Duygulu and Atalay, 2002]. A system that extracts text lines and describes the layout by means of relationships between pairs of these lines was also discussed in the literature [Huang et al., 2005], whereas some researchers used Brick Wall Coding Features (BWC) features to represent bounding boxes of the words [Erol et al., 2008]. Although the features are scale invariant and robust to slight perspective distortion, the accuracy of their system is very low. In addition the method does not work correctly when documents are written in languages such as Japanese and Chinese, in which words are not separated. Several approaches can also be combined to identify a document, such as barcode, micro optical patterns, encoding hidden information, paper fingerprint, character recognition, local features and RFID. Owing to utilize SIFT. Unfortunately, the retrieval process is time consuming and requires special equipment [Liu and Liao, 2011].

In this paper a Layout-based Document Image Retrieval (LDR) system is presented, that is specifically devoted to commercial form processing, such as invoices, waybills, receipts, etc., in which layout is strongly characterized by a grid-structure. In fact, in these particular cases, traditional document-image approaches are not effective since they are not able to describe documents on the basis of the grid-based structure. In the first step the system uses a technique based on mathematical morphology for removing textual components from the document image and for extracting the grid-based structure in the document layout. Subsequently Radon Transform is used to obtain the feature vector characterizing the specific grid structure of the document. Dynamic Time Warping (DTW) is finally adopted to perform document matching.

The paper is organized as follow. The architecture of the system is presented in Section 2. Section 3 describes the preprocessing phase, which uses operators of mathematical morphology. The feature extraction phase is presented in Section 4. In Section 5 the matching process based on DTW is discussed while the decision combination process is illustrated in Section 6. The experimental results are reported and discussed in Section 7. Section 8 presents the conclusion of the work and highlights some directions for further research.

2. The Radon Transform for Layout-based Document Image Retrieval

The LDR system presented in this paper is based on three main phases: Acquisition and Preprocessing; Feature Extraction; Matching. After document image acquisition, the document is preprocessed and transformed by Radon Transform [Pirlo et al., 2013a,b]. The features extracted are then stored in the reference database in the enrollment stage. In the running stage, an unknown document is first scanned and preprocessed, successively the features are extracted compared to the those stored into the database. The matching module performs matching by Dynamic Time Warping (DTW) and outputs the ranked list of similar documents. More precisely, the input document is acquired as a standard 256 gray-level – 100dpi PDF file. Figures 1 shows an input document concerning a real invoice.

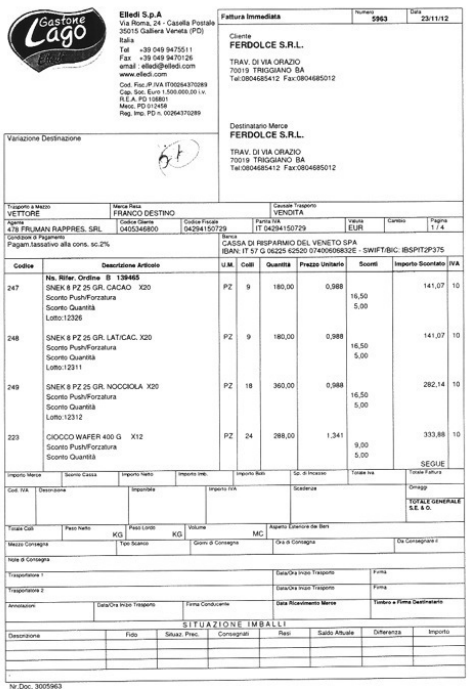


Figure 1
Input document image I=I(x,y)

Successively, after noise removal, document is resampled to 100 dpi and grid-based structure is extracted by mathematical morphology [Serra, 1982]. More precisely, let I=I(x,y) be the document image (1xX , 1yY) and let be

- B_{hor} the horizontal structure element defined as (see Figure 2a): B={(-s,0), …, (-1,0), (0,0), (1,0), … , (s,0)};
- B_{ver} the horizontal structure element defined as (see Figure 2b):

B={(0,-s), …, (0,-1), (0,0), (0,1), … , (0,s)} ;

being s a small positive integer which determine the size of the structure element.

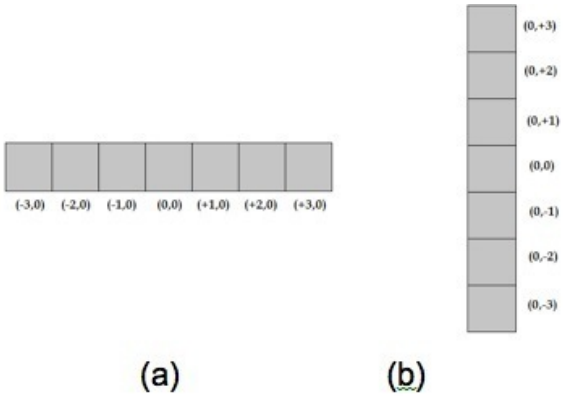


Figure 2
Structure elements (s=3): (a) B_{hor} , (b) B_{ver}

In the preprocessing phase from the image I(x,y) two filtered images $I_{hor}=I_{hor}(x,y)$ and $I_{ver}=I_{ver}(x,y)$, which contains respectively horizontal and vertical segments, are obtained by a closure operator as follows (see Figure 3):

$$I_{hor} = I \bullet B_{hor} = (I \oplus B_{hor}) \ominus B_{hor} \quad (1a)$$
$$I_{ver} = I \bullet B_{ver} = (I \oplus B_{ver}) \ominus B_{ver} \quad (1b)$$

being "•" the closure operator, while "⊕" and "⊖" indicate respectively Minkowski sum and difference.

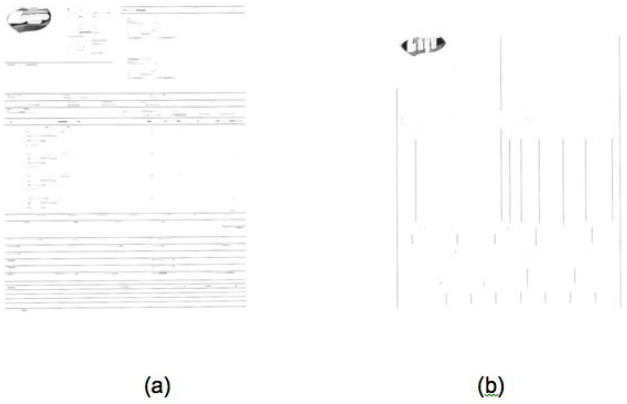


Figure 3
Example of filtered images: (a) I_{hor} , (b) I_{ver}.

Finally, $I_{hor}(x,y)$ and $I_{ver}(x,y)$ are combined to reconstruct the preprocessed image I^* according to XOR operator:

$$I^* = I_{hor} \text{ XOR } I_{ver} \qquad (2)$$

Figure 4 shows an example of document image after preprocessing.

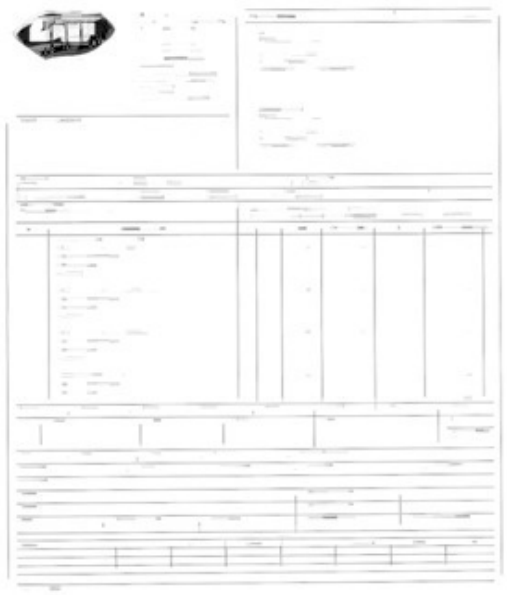


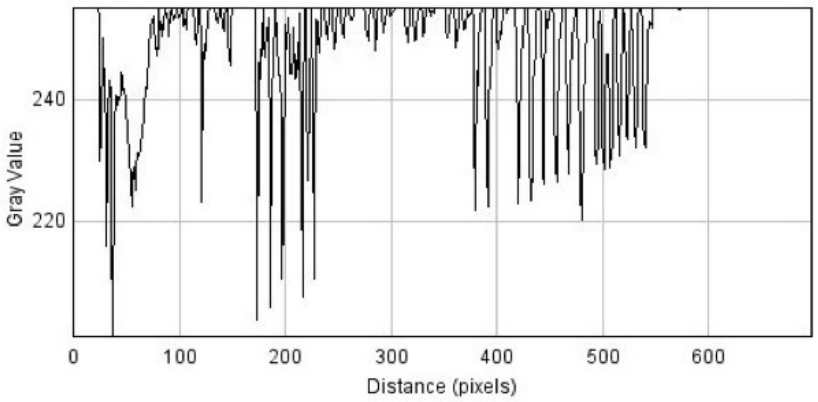
Figure 4
The preprocessed image I^*

In the feature extraction step, in order to extract grid-based layout document images, the Radon Transform was considered. It is worth noting that the Radon Transform was extensively used in image analysis and has a number of important applications, like those related to MRI and computed tomography [Cormack, 1983; Deans, 1983]. The complete description of the Radon Transform is beyond the scope of this paper (see further details in [Jafari-Khouzani and Soltanian-Zadeh, 2005; Seo et al., 2004]). For the aim of this paper we only remind that the Radon Transform computes projection sum of the image intensity along a oriented at line ($\rho - x\cos\vartheta - y\sin\vartheta$) =0, for each ϑ and ρ. More precisely the Radon Transform of a function $I^*(x,y)$ in an Euclidean space is defined by [Hjouj and Kammler, 2008]:

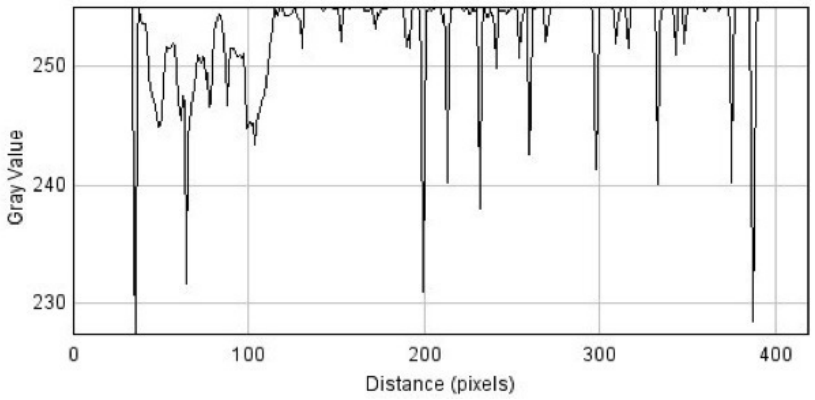
$$S_{\vartheta,\rho} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} I^*(x,y)\cdot\delta(\rho - x\cos\vartheta - y\sin\vartheta)dxdy \qquad (3)$$

where the d (r) is Dirac function, which is infinite for argument zero and zero for all other arguments (it integrates to one).

Therefore, computing the Radon Transform of a two dimensional image intensity function $I^*(x,y)$ results in its projections across the image at arbitrary orientations ϑ and offsets p Figure 5 presents the results of the Radon Transform applied to the preprocessed image I^* for the parameter values related to horizontal (ϑ=0 , p=0 and vertical (ϑ=/2 , p=0)projections.



a. Horizontal Projection



b. Vertical Projection

Figure 5
Feature extraction by Radon Transform

Dynamic Time Warping (DTW) is used for matching the feature vectors extracted by the radon transform from two document images. More precisely, let be F^r, S^t the feature vectors of M elements extracted from the document images I^{*r} and I^{*t}, a warping function between S^r and S^t is any sequence of couples of indexes identifying points of S^r and S^t to be joined [Salvador and Chan, 2004; Lemire, 2009]:

$$W(S^r,S^t)=c_1,c_2,\ldots,c_K, \quad (4)$$

where $c_k=(i_k,j_k)$ (k,i_k,j_k integers, $1 \le k \le K$, $1 \le i_k \le M$, $1 \le j_k \le M$). Now, if we consider a distance measure $d(c_k)=d(z^r(i_k), z^t(i_k))$ between elements of S^r and S^t, we can associate to $W(S^r,S^t)$ the dissimilarity measure

$$D_{W(S^r,S^t)} = \sum_{k=1}^{K} d(c_k) \quad (5)$$

The DTW detects the warping function $W^*(S^r,S^t)= c^*_1,c^*_2,\ldots,c^*_{K^*}$ which satisfies the condition of [Salvador and Chan, 2004]:

- Monotonicity (i.e. $i_{k-1} \le i_k$, $j_{k-1} \le j_k$ for $k=2,...K$) (6a)
- Continuity (i.e. $i_k - i_{k-1} \le 1$, $j_k - j_{k-1} \le 1$ for $k=2,...K$) (6b)
- Boundary (i.e. $i_1 = 1$, $j_1 = 1$ and $i_K = M$, $j_K = M$) (6c)

and which provides the distance value between S^r and S^t defined as [Salvador and Chan, 2004; Lemire, 2009]:

$$D_{W^*(S^r,S^t)} = \min_{W(S^r,S^t)} D_{W(S^r,S^t)} \quad (7)$$

The value in eq. (7) represents the similarity between the document images I^{*r} and I^{*t}. Therefore, given a document image as input, the matching module will outputs the ranked list of the k top similar document images retrieved from the database.

The matching procedure provides two distance-based ranked lists of documents, obtained respectively from horizontal and vertical projections. The decision making process obtains the final decision combining the two ranked lists using the Borda-count strategy [Kittler et al., 1998; Xu et al., 1992]. According to this strategy, let $D=\{D_1, D_2,\ldots, D_k,\ldots, D_K\}$ be the set of K documents enrolled into the system for reference and D" the unknown input document. Furthermore, let be:

- $L^h : < D^h_1, D^h_2, \ldots, D^h_k, \ldots, D^h_K >$ the ranked list of documents obtained from the match of the horizontal projection (D^h_k D, for k=1,2,,,,,K and D^h_{k1} D^h_{k2} for k_1 k_2);
- $L^v: < D^v_1, D^v_2, \ldots, D^v_k, \ldots, D^v_K >$ the ranked list of documents obtained from the match of the vertical projection (D^v_k D, for k=1,2,,,,,K and D^v_{k1} D^v_{k2} for k_1 k_2);

The Borda-count approach assigns to each reference document D_k a confidence score $S(D_k)$ defined as [Ho et al., 1994]:

$$S(D_k) = S^h(D_k) + S^v(D_k) \quad (8)$$

being $S^h(D_k)=K-i$, if $D_k = D^h_i$; $S^v(D_k)=K-j$, if $D_k = D^v_j$.

Hence, the final list of ranked documents is

$$L^* : < D^*_1, D^*_2, \ldots, D^*_k, \ldots, D^*_K > \qquad (9)$$

so that D^*_{k1} precedes D^*_{k2} in L^* if and only is $S(D_{k1})$ $S(D_{k2})$, and – of course - $_1$ is the top candidate document [Ho et al., 1994].

3. Experimental Results

Two datasets of documents were considered for the test. The first dataset concerns real documents, the second dataset concerns simulated documents. The first dataset contains 33 commercial forms belonging to 16 different categories. Figure 6 shows some examples of commercial forms in the dataset. In this case they belong to the category n. 1.

Figure 6
Examples of commercial forms of the same category

Documents were scanned (100dpi , 256 gray-level) and preprocessed. Finally they were stored into a database along with the values of the Radon Transform concerning the horizontal ($S_{0,0}$) and vertical ($S_{0,/2}$) projection. Table 1 reports the number of forms for each category.

Category	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Number of documents	9	4	3	3	2	2	1	1	1	1	1	1	1	1	1	1

Table 1
Dataset of Real Documents

In the testing phase the leave-one-out method was considered to verify the effectiveness of the system. In order to estimate the quality of the ranked list provided by the system for a given query, the Average Normalized Rank (ANR) was adopted, defined as follows [Huang et al., 2005]:

$$ANR = \frac{1}{N \cdot N_w} \cdot \sum_{i=1}^{N_w} \left(R_i - \frac{N_w + 1}{2} \right) \qquad (10)$$

being

- N the number of documents in the set,
- N_w the number of relevant documents (for the given query) in the set,
- R_i is the rank of each relevant document in the set.

It is worst noting that ANR ranges in [0,1]:

- ANR=0 means that relevant documents are at the top of the ranked list (right position);
- ANR=1 means that relevant documents are at the bottom the ranked list (wrong position).

Figure 7 shows the experimental results. They demonstrate that the proposed approach is very robust with respect to different categories of documents. On average the value of ANR is equal to 0.08. Furthermore, 26 cases out of 33 the ANR is less than 10%, whereas only in one case it is greater than 0.5.

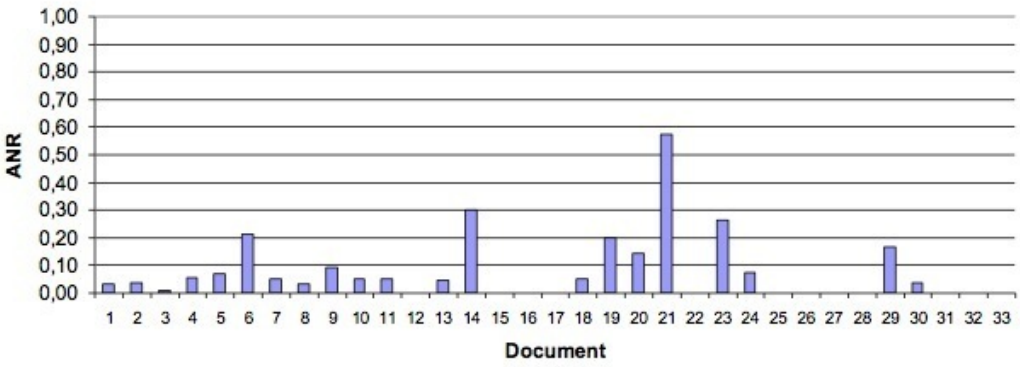


Figure 7
Real dataset: standard documents

In order to estimate the robustness of the new approach two additional tests have been carried out using shifted and rotated document images as input.

When shifted documents are fed into the system the experimental results are shown in Figure 8. In this case a shift of 5 pixel is considered in the four main

directions and the average result is computed. Also in this case the value of ANR is equal to 0.08, on average.

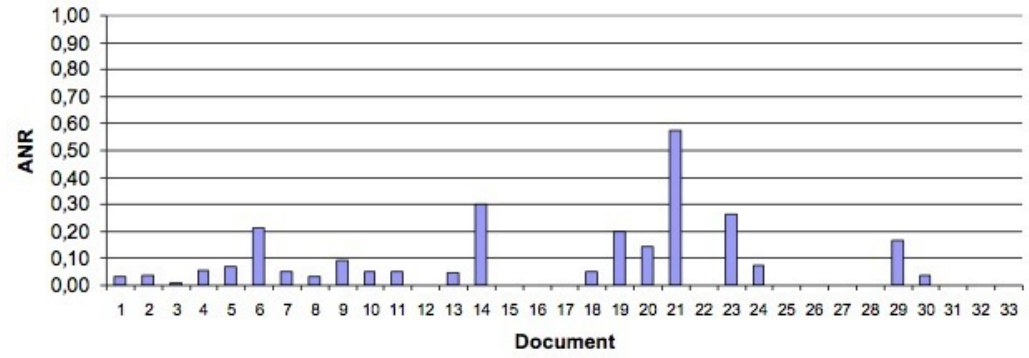


Figure 8
Real dataset: shifted documents

Conversely, Figure 9 shows the results when rotated document images are fed into the system. In this case a rotation of 2° clockwise and anticlockwise is considered and the average result is computed. In this case the value of ANR is equal to 0.14 on average.

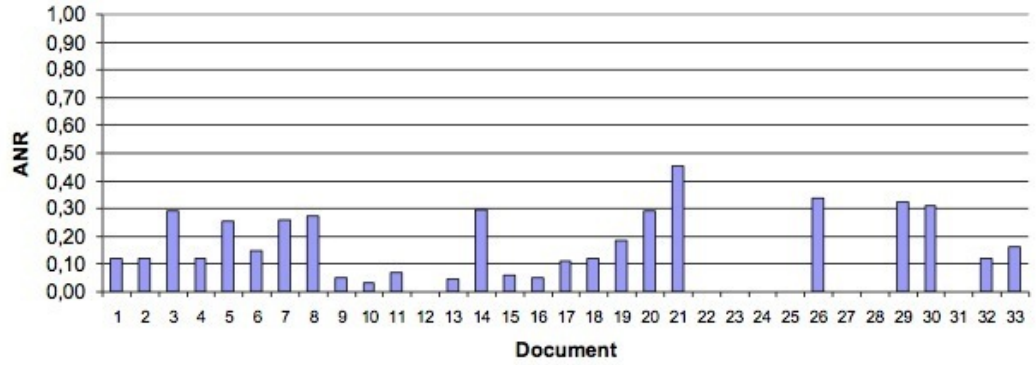


Figure 9
Real dataset: rotated documents

Figure 10 shows the results when document images are shifted and rotated before to be fed into the system. In this case a shift of 5 pixels and a rotation of 2° clockwise and anticlockwise are considered. In this case the value of ANR is equal to 0.15 on average.

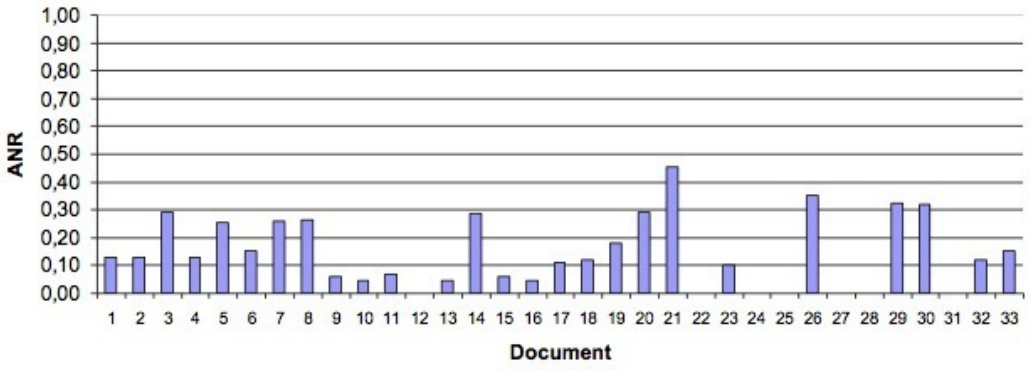


Figure 10
Real dataset: shifted and rotated documents

The second dataset contains 100 simulated documents belonging to 16 different categories. Table 2 reports the number of forms for each category.

Category	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Number of documents	12	11	10	10	10	8	8	6	6	3	3	3	3	3	2	2

Table 2
Dataset of Synthetic Documents

Also in this case, the leave-one-out method was considered to verify the effectiveness of the system.

Figure 11 shows the experimental results. More precisely it shows the ANR for each category of documents. In this case it results that the value of ANR is equal to 0.17, on average.

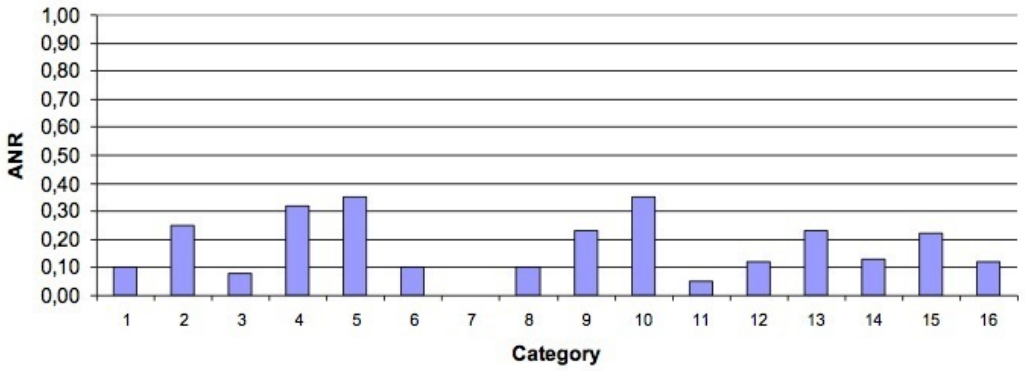


Figure 11
Simulated dataset: standard documents

Figure 12 shows the experimental results, when shifted documents are considered. Also in this case a shift of 5 pixel is considered in the four main directions and the average result is computed for each category of documents. The value of ANR is equal to 0.17, on average.

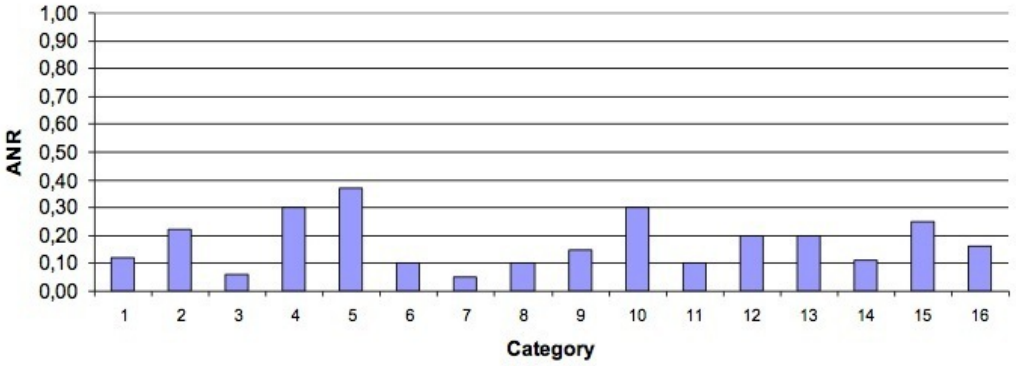


Figure 12
Simulated dataset: shifted documents

Conversely, Figure 13 shows the results when rotated document images are fed into the system. In this case a rotation of 2° clockwise and anticlockwise is considered and the average result is computed. In this case the value of ANR is equal to 0.27 on average.

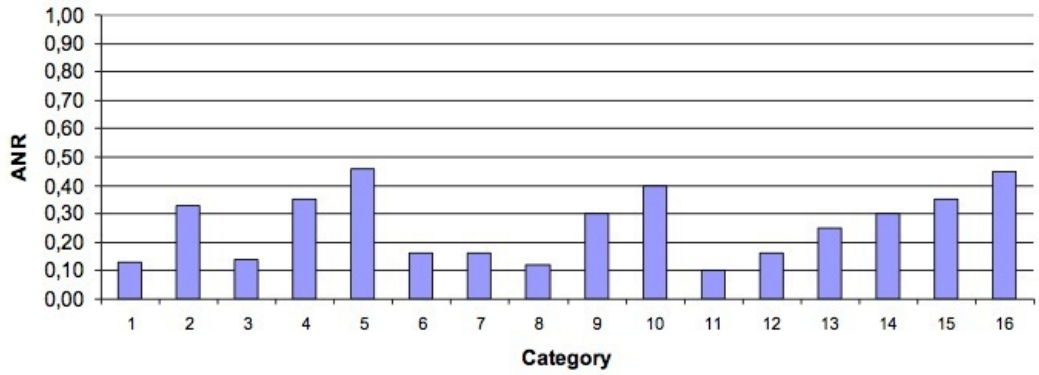


Figure 13
Simulated dataset: rotated documents

Figure 14 shows the results when document images are shifted and rotated before to be fed into the system. In this case a shift of 5 pixels and a rotation of 2° clockwise and anticlockwise are considered. In this case the value of ANR is equal to 0.29 on average.

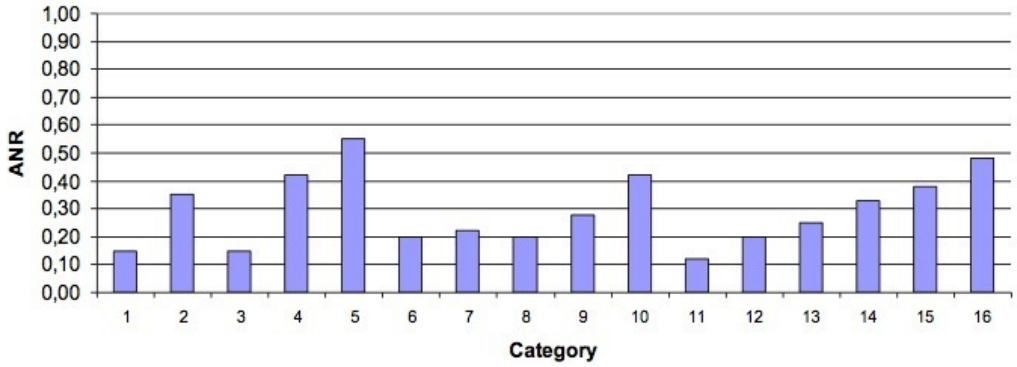


Figure 14
Simulated dataset: shifted and rotated documents

4. Conclusion

In this paper a new system for layout-based document image retrieval was presented. The system was specifically designed for retrieval of commercial forms as invoices, waybills and receipts, to optimize document management and sustainability. It used a morphologic filtering technique and the Radon Transform to obtain multiple document image descriptions. Document matching was then performed by Dynamic Time Warping whereas the Borda-count decision combination strategy was used to combine multiple decisions.

The experimental results, carried out on a dataset of real commercial documents, demonstrate the effectiveness of the proposed solutions and the robustness also with respect to shifted and rotated input document images.

References

Lopresti, D., Robust Retrieval of noisy text, *Proc. of ADL'96,* 1996, 76–85.

Cormack, A. M., Computed tomography: Some history and recent developments, *Proc. Symposia in Applied Mathematics*, 27, 1983, 35–42.

Deans, S. R., The Radon Transform and Some of Its Applications, New York: Wiley, 1983.

Doermann, D. , The Indexing and Retrieval of Document Images: A Survey, *Computer Vision and Image Understanding,* 70, 3, 1998, 287–298.

Duygulu, P., Atalay, V., A Hierarchical Representation of Form Documents for Identification and Retrieval, *International Journal on Document Analysis and Recognition,* 5, 1, 2002, 17–27.

Erol, B., Ant´unez, E., Hull, J. J., Hotpaper: multimedia interaction with paper using mobile phones, *Proceeding of the 16th ACM international conference on Multimedia*, 2008, 399–408.

Hjouj, F., Kammler, D. W., Identification of Reflected, Scaled, Translated, and Rotated Objects From Their Radon Projections, IEEE Trans. Image Processing, 17, 3, 2008, 301-310.

Ho, T.K., Hull, J.J., Srihari, S.N., Decision combination in multiple classifier systems, IEEE Trans. Pattern Anal. Mach. Intell., 16, 1, 1994, 66–75.

Huang, M., Dementhon, D., Doermann, D., Golebiowski, L., Document ranking by layout relevance, *Proc. 8th ICDAR*, 2005, 362–366.

Jafari-Khouzani, K., Soltanian-Zadeh, H., Radon Transform orientation estimation for rotation invariant texture analysis, *IEEE Trans. Pattern Anal. Mach. Intell.*, 27, 6, 2005, 1004–1008.

Kittler, J., Hatef, M., Duin, R.P.W., Matias, J., On combining classifiers, IEEE Trans. on Pattern Analysis Machine Intelligence, 20, 3, 1998, 226-239.

Lemire, D., Faster Retrieval with a Two-Pass Dynamic-Time-Warping Lower Bound, Pattern Recognition, 42, 9, 2009, 2169-2180.

Liu, Q., Liao, C., PaperUI, Proceeding of the 4th International Workshop on Camera-Based Document Analysis and Recognition, 2011, 3–10.

Manning, C. D. , Raghavan, P. , Schütze, H., An Introduction to Information Retrieval, Cambridge Press, 2009.

Marukawa, K., Hu, T., Fujisawa, H., Shima, Y., Document retrieval tolerating character recognition errors - Evaluation and application, *Pattern Recognition,* 30, 8, 1997, 1361-1371.

Mitra, M., Chaudhuri, B., Information retrieval from documents: A Survey, *Information Retrieval,* 2, 2/3, 2000, 141–163.

Pirlo G., Chimienti M., Dassisti M., Impedovo D., Galiano M., “Document Image Retrieval by Layout Analysis”, Atti del Congresso Nazionale AICA 2013, Frontiere Digitali: Dal Digital Devide alla Smart Society, 2013, pp. 644-653.

Pirlo G., Chimienti M., Dassisti M., Impedovo D., Galiano M., “Layout-Based Document-Retrieval System by Radon Transform Using Dynamic Time Warping”, Proc. International Conference on Image Analysis and Processing (ICIAP 2013), Petrosino (Ed.), Naples, Sept. 11-13, 2013, , LNCS 8156, pp. 61–70.

Salvador, S., Chan, P., Fast DTW: Toward Accurate Dynamic Time Warping in Linear Time and Space, *Proc. KDD Workshop on Mining Temporal and Sequential Data*, 2004, 70-80.

Seo, S., Haitsma, J., Kalker, T., Yoo, C. D., A robust image fingerprinting system using the Radon transforms, *Signal Process.: Image Commun.*, 19, 4, 2004, 325–339.

Serra, J., Image Analysis and Mathematical Morphology, Academic Press, 1982.

Taghva, K., Borsack, J., Condit, A., "Evaluation of model-based retrieval effectiveness with OCR text", *ACM TOIS,* 14, 1, 1996, 64–93.

Tzacheva, A., El-Sonbaty, Y., El-Kwae, A., Document Image Matching Using a Maximal Grid Approach, *Proceedings of the SPIE Document Recognition and Retrieval IX,* 2002, 121-128.

Xu, L., Krzyzak, A., Suen, C. Y., Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition, IEEE Transaction on Systems, Man and Cybernetics, 22, 3, 1992, 418-435.

Biographies

Giuseppe Pirlo received the Computer Science degree cum laude in 1986. Since 1991 he has been Assistant Professor at the University of Bari, where he is currently Associate Professor. His interests cover the areas of document processing and pattern recognition, biometry, computer arithmetic and e-learning. He has developed several scientific projects and published over 200 papers. Giuseppe Pirlo is associate editor of IEEE Trans. on Human Machine Systems and reviewer for IEEE T-PAMI, IEEE T-SMC, IEEE T-EC, PR, IJDAR, IPL. He was general co-chair of ICFHR 2012 and EAHSP 2013. He was guest editor of "*Handwriting Recognition and other PR Applications*" of the PR Journal and "*Handwriting Biometrics*" of the Biometrics Journal. He is IEEE and IAPR member.
email: giuseppe.pirlo~uniba.it

Michela Chimienti was born in Bari, Italy, in 1978. She received the degree in Management Engineering and her Ph.D. in Advanced Production Systems from the Polytechnic University of Bari (Italy), respectively in 2006 and 2010. In 2010 she has been visiting lecturer at the University of Kentucky (USA) at the Institute for Sustainable Manufacturing. Since 2012 she is Scientific Director of Laboratorio Kad3, a research organisation in mechanical engineering and environmental protection areas. Her interests cover the areas of sustainable manufacturing processes and systems, economic models for sustainable manufacturing, renewable energy systems, and interoperability techniques.
email: m.dassisti@yahoo.it

Michele Dassisti is Associate Professor of Manufacturing Systems and Technologies at Politecnico di Bari where he teaches Industrial Quality Management and Sustainable Process Continuous-Improvement at Faculty of Management Engineering. He conducts research at DMM (Department for Mechanics, Mathematic and Management), leads also as a scientific responsible of two network industrial research-laboratories. His research field aims at continuous improvement of sustainability and innovation of several conventional and non conventional manufacturing processes and technologies. Several innovative applications of advanced statistical methods, advanced mathematical functions, Artificial Intelligence techniques (neural networks,

expert systems, genetic algorithms, fuzzy logic, predicate logics, ontology), Decision Making techniques, interoperability techniques, discrete event simulators, Finite Element Analysis. He is author or co-author of more than 60 papers. He is currently Founders of the Italian Pole of the INTEROP-VLab.It (Virtual Laboratories for Interoperability) and member of the IFAC Technical Committee 5.3 ''Enterprise Integration and Networking''.
email: m.dassisti@yahoo.it

Donato Impedovo received the MEng degree 'summa cum laude' in Computer Engineering in 2005 and the PhD degree in Computer Engineering in 2009 both from the Polytechnic of Bari (Italy). Impedovo received the II level Master Degree in technologies for remote sensing in 2011 from the University of Bari. Impedovo is co-author of more than 30 articles in these fields in both international journals and conference proceedings. He received 'The Distinction' for the best young student presentation in May 2009 at the International Conference on Computer Recognition Systems (CORES – endorsed by IAPR), and the award for the Nereus-Euroavia Academic competition on GMES in October 2012. Impedovo serves as reviewer for the Elsevier Pattern Recognition journal, IEEE THMS, IET Journal on Signal Processing and IET Journal on Image Processing and for many International Conferences including ICPR and ICASSP. He is IAPR and IEEE member.
email: donato.impedovo@dyrecta.com

Angelo Galiano received the MSc degree in Education in 2009 from Unipegaso University (Naples, Italy). He has worked since 1996 in the field of ICT for many company (including Enel spa) covering roles of gradually increasing importance. Galiano founded, on 2001, Dyrecta an ICT SME, successively grown in DyrectaLab when, on 2011, the company obtained the recognition of certified private research laboratory by MIUR (Italian Ministery of Research). Dr. Galiano is CEO of DyrectaLab, he is scientific coordinator as well as member of the scientific commitee of many (public and private) industrial research projects. His research interest are in the field of HCI, Computer Vision and Augmented Reality.
email: maurizio.galiano@dyrecta.com

A Biometric Authentication System Based on Face Recognition and RFID tags

F. Battaglia, G. Iannizzotto, L. Lo Bello

Abstract. *Authentication systems usually adopt either the conventional identifier-password paradigm or different kinds of tokens (e.g., badges, keys). However, passwords can be disclosed while being input and tokens can be stolen and used by impostors. As a result, in the last decades biometric techniques were developed to identify a user through physiological features that cannot be stolen or counterfeited. However, even those techniques have their flaws, and for this reason recent research addressed the combination of multiple identification factors. In this context, this work proposes VisilabFaceRec, a multi factor authentication system based on the combination of a dual-stage cascading classifier for biometric identification (face recognition) with an encrypted RFID tag for token-based authentication. Unlike other approaches in the literature that propose a centralized database for storing biometric data, with serious risks regarding user privacy, our work avoids a centralized database and stores sensitive data in the RFID, thus also making the system performance independent of the total number of subjects enrolled. The proposed architecture is able to simultaneously minimize the False Acceptance Rate and the False Rejection Rate, thanks to an innovative approach for the calculation of the decision thresholds for the two discriminators. VisilabFaceRec has been realized on a commercial board for embedded computing and proven to be able to run in near real-time. The paper describes the system architecture and the algorithm used to jointly determine the couple of decision thresholds for the cascading classifiers, and proposes a performance evaluation, in terms of both accuracy and speed, on a well-known and publicly available face database.*

Keywords: Biometrics, security, authentication, RFID tags.

1. Introduction

Access control (or authentication) mechanisms aim to guarantee only authorized users the access to a given resource or service at any time, while blocking impostors.

The simplest authentication mechanism is password protection [Jain et al, 2006]. However, the need for systematically typing a password to obtain access to resources or restricted areas (i.e., buildings, offices, devices) is a tedious and inefficient operation, that is also prone to serious security flaws, due to the possibility that the password be either read by someone else (e.g., while it is being typed) or involuntarily disclosed.

Although tokens (such as, badges or keys) offer effective security features (e.g., encryption, public/private keys, etc.), their use is not the ultimate security solution either, as a token is not exclusively bound to its owner and therefore could be used by an impostor.

For this reason, a significant effort has been made over the years to develop authentication systems based on biometric data, which offer the advantage of being always available at the very place where the user to be authorized is while being quite difficult to counterfeit.

Although several features can be used for biometric recognition (e.g., digital footprint, iris, hand shape, voice), face recognition offers multiple advantages. First, face recognition does not require expensive sensors. Second, there is no need for physical contact between user and sensor. Finally, it can be used also for video-surveillance and in the non-decisive phases of judicial investigation.

However, when authentication techniques are not combined with other approaches, they suffer from the limitation of requiring a database, either centralized or distributed, to maintain the biometric data of all the users to be authenticated, as such a database allows performing authentication through the comparison between the biometric characteristics of the user and those that are stored in the database. Moreover, it has not been proven yet that automated face recognition is able to achieve 100% accuracy for an arbitrarily large database and this fact raises doubts about the adoption of this technology.

An approach that is being broadly investigated nowadays is Multi Factor Authentication, which combines multiple techniques to obtain more reliable results.

For instance, in the past some solutions that integrate RFID tags (i.e., radio-frequency tokens) with face recognition systems were proposed [Min et al, 2011] [Jing et al, 2009] [Nguyen et al, 2012]. However, these solutions maintain the set of poses of the authorized users in a centralized database and store in the RFID tags only very few data, such as, the user identifier (declared identity).

Furthermore, these methods require, for a single claimant, a large number of poses to be acquired at high resolution under different lighting conditions and stored in the database [Nguyen et al, 2012].

These approaches raise problems, not only due to the database size, but also due to legal issues: The authorities, both at a national and supranational level,

have stated several times that biometric data are personal data, therefore it is in general not advisable to maintain them in a centralized database run by the service provider, unless there exist important and proven security reasons for doing so (principle of finality, necessity, proportionality). In addition, a clear preference towards *serverless* solutions, in which the biometric data are stored on a chip that the authorized user is able to take with them, was given [Art. 29 WP, 2003].

Authentication systems like the one proposed in this paper, i.e., combining RFID tags with face recognition and based on a serverless architecture, are quite innovative compared to the current state of the art. For instance, in [Meng et al, 2010], the authors propose an embedded system that stores in the RFID tag only the set of the *n* principal decomposition components (PCA) [Pentland et al., 1991] associated to the owner's face. However, *for both the enrolment and authentication stages* the PCA representation requires the availability, local to the authentication system, of a set of images (*Eigenfaces*) which depend on the whole of the images of the enrolled subjects. Moreover, every time a new subject is enrolled, the set of Eigenfaces changes and therefore must be recalculated. Unfortunately, when the set of Eigenfaces changes, also the set of principal components of each enrolled user changes, therefore all RFID tags must be redistributed. The paper does not clearly state where the Eigenfaces are stored (either in the tag or in the local memory of the authentication device) and what happens when a new subject is enrolled.

Furthermore, none of previously cited works was tested for robustness against the intrusion of impostors by simulating a significant number of illegal access attempts. The systems were optimized for recognition accuracy or processing speed (for example, in [Jing et al, 2009] the system was tested on three subjects only and an average false acceptance rate of 3.89% was obtained).

The system presented in this paper, called *VisilabFaceRec,* integrates RFID recognition technology with a face recognition system, in such a way that the resulting system provides the following properties:

- The user biometric data are encrypted and stored in the RFID badge; therefore, there is no need for centralized databases or for arbitrarily large databases.
- The algorithms adopted allow for high accuracy, although the amount of stored data is compliant with the capacity of a commercial RFID. This significantly reduces costs and architectural complexity compared to other solutions based on centralized databases.
- The system is fully scalable, as the operations of adding or deleting a subject in the list of the enrolled users do not require any recalculation or the redistribution of the tags.

The main contribution of this work is therefore an authentication system that, combining biometric and RFID authentication, does not need a centralized database, thus avoiding any problems related to privacy issues. The proposed system obtains good results, and its performance in terms of accuracy is independent of the number of subjects to be authenticated. This solves a

significant problem found in other approaches in which the reliability of the results decreases when the number of subjects grows. Finally, the system works on very low resolution face images (e.g. 40x30 pixels), so the image acquisition can be performed by simple, cheap, off-the-shelf cameras and the communication with the RFID tag is reasonably fast.

As will be shown in the following, the proposed authentication system is based on a *two-cascaded discriminator* stage that is able to realize simultaneously the minimization of the False Acceptance Rate (proportion of impostors accepted) and of the False Rejection Rate (proportion of genuine claimants rejected), thanks to an innovative approach for the calculation of the decision thresholds for the two discriminators.

The proposed system also offers, in addition to the already listed advantages, the appealing possibility of being implemented on embedded devices at relatively low cost.

The paper is organized as follows. Sect. 2 describes the VisilabFaceRec architecture, discussing the rationale behind the design choices and providing details on the mathematical formulation of the approach devised for determining the optimal thresholds for the two cascaded discriminators. Sect. 3 presents experimental results obtained by testing the performance of VisilabFaceRec using a well-known publicly available face database and evidences the accuracy and speed of the system. Finally, Sect. 4 provides conclusions and directions for future work.

2. VisilabFaceRec

Before a subject can be authenticated, they must be enrolled, i.e. a set of images of their face must be recorded together with their identity information. In VisilabFaceRec the images of the enrolled subject are saved in an RFID tag and the enrollment step is a one-time process.

In the authentication step, when a subject to be authenticated approaches the camera, a sequence of images of their face is acquired and, at the same time, the content of the RFID tag is read. The acquired images are processed one at a time. If the similarity level between at least one of the acquired faces and the images retrieved from the RFID is sufficient, the subject is authenticated.

The adopted approach (see Fig. 1) is composed of two cascaded authentication stages, optimized for low-resolution images which can be saved in the small memory available in commercially available RFID tags (4-32 KB). The first stage is based on the 2D Principal Components Analysis (2DPCA) algorithm [Yang et al, 2004] and operates a first discrimination of the input images. The second stage operates only on the images which pass the first stage and exploits a *Scalar Image Feature Transform* (SIFT) [Lowe, 2004] to produce a final selection.

2DPCA is a template-matching algorithm originally proposed for face recognition. Given a collection B of *m* face images (poses) belonging to *p* different *known subjects* (classes) with *k* poses for each class, 2DPCA receives as an input a pose belonging to an *unknown subject* and selects from the collection the pose most similar to it, thus identifying the claimant.

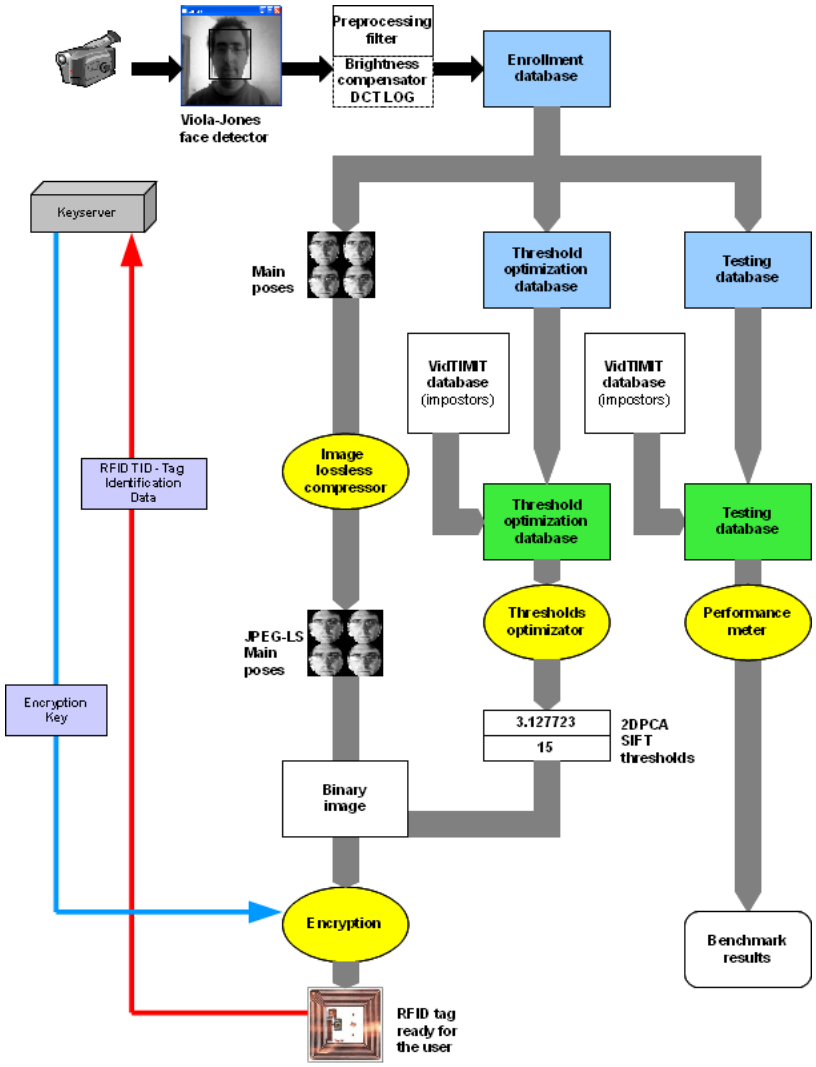


Figure 1
Schematization of the VisilabFaceRec enrolling process

Within 2DPCA an image I of size *w*h* pixels can be decomposed into a set of z_{2DPCA} <=w vector components (called *decomposition vectors*):

$$(\mathbf{w}_0, \mathbf{w}_1, ..., \mathbf{w}_{z_{2DPCA}-1}) \qquad z_{2DPCA} \leq w \quad \mathbf{w}_s \in R^h \tag{1}$$

where the number of components z_{2DPCA} is chosen according to the fraction of the original information associated to the image, which we want to retain.

We therefore can decompose each pose in the collection B of known subjects and the image I of an unknown subject, and then select the pose in B which is most similar to I by minimizing the *distance in the feature space* (DIFS) defined in (2):

$$DIFS(\mathbf{I},B) = d(\mathbf{I},\mathbf{W_i}) = \sum_{s=0}^{z_{2DPCA}-1} \left\| \mathbf{w}_s - (\mathbf{w}_i)_s \right\|_2 \qquad i \in \{0...(m-1)\} \qquad (2)$$

where $\mathbf{W}_i$ is the i-th pose from B, $\mathbf{w}_s$ is the s-th component vector of the unknown pose I, $(\mathbf{w}_i)_s$ is the s-th component vector of $\mathbf{W}_i$ and *m=p*k*.

The 2DPCA decomposition of the image I in respect to the collection B is obtained by multiplying the matrix A containing the pixel intensities of I by each one of the z_{2DPCA}<=w main eigenvectors of the covariance matrix $\mathbf{G}_t$ (of size w*w**)** defined in (3):

$$\mathbf{G}_t = \frac{1}{m}\sum_{j=0}^{m-1}(\mathbf{A}_j - \overline{\mathbf{A}})^T(\mathbf{A}_j - \overline{\mathbf{A}}) \qquad (3)$$

where $\mathbf{A}_j$ is the matrix containing the pixel intensities of the j-th pose from the collection B.

In our approach B contains only the poses of the enrolled subject (p=1 and therefore m=k), B is named *main poses database* and it is possible to verify if a newly acquired image I belongs to the same subject by simply applying a decision threshold ρ to the distance DIFS defined by (2). However, in single threshold systems, False Acceptance Rate (FAR) and False Rejection Rate (FRR) cannot be minimized at the same time [Flach, 2003]. The decision threshold ρ must therefore be determined as a trade-off between the two minimization objectives. Moreover, due to (3) depends on the poses of the collection B and therefore is specific for each collection. In Section 2.1 we describe a solution for the two problems introduced above.

The second stage of our authentication algorithm is based on the SIFT algorithm, which belongs to the class of feature-matching algorithms. When applied to a couple of images, it determines a set of *keypoints* in both images and applies a robust and reliable criterion to establish a correspondence between couples of matching keypoints from the two images [Lowe, 2004]. For each acquired image which passes the 2DPCA stage, and the corresponding most similar pose from the collection B, the SIFT stage determines the number *s* of matching SIFT features and, if this is higher than a decision threshold $\mathbf{o}_{SIFT}$, the acquired image passes also the second stage and the unknown subject is authenticated. Otherwise, the subject is rejected.

2.1. The Enrollment phase

A new subject is enrolled by acquiring $n_{enrolldb}$ images of their face (poses). Each pose is then decomposed according to 2DPCA and the resulting representations are clusterized by applying a K-means algorithm using the distance (2). For each one of the $n_{main.poses}$ clusters obtained, the pose closest to the centroid of the cluster is selected, thus creating a set of *main representative*

poses of the subject. Those poses compose the collection B for a specific subject and are compressed with a lossless algorithm (LS-JPEG) and saved in the RFID of the subject.

Unlike other solutions proposed in literature, our approach uses several poses for each subject, coded at 256 gray levels. We save in the tag the images and not their 2DPCA representations (1) because saving a number z_{2DPCA} of floating point components suitable for our authentication purposes would take more space than the original 8 bit gray level images. We therefore prefer to recalculate on-the-fly the covariance matrix (3) and the corresponding eigenvectors every time the RFID tag is read. As shown in Section 3, the time required by such recalculation does not affect severely the execution time of the whole authentication process.

As stated earlier, VisilabFaceRec uses two cascaded authentication stages which each require a custom threshold for each collection of representative poses. We therefore need a couple of thresholds (ρ_{2DPCA}, σ_{SIFT}) for each subject. Such thresholds are saved together with the main representative poses of the subject in the RFID tag. Determining such thresholds is not a trivial task because the outcome of the 2DPCA stage affects the behavior of the SIFT stage. In the following we describe how we determine the two thresholds and in Section 3 we show some relevant experimental results which support the suitability of the described solution.

We define the *threshold optimization database* (TOdb) composed of p_{TOdb} rows and k_{TOdb} columns. The first row is composed of the $k_{TOdb}= n_{enrolldb} - n_{main.poses}$ poses of the enrolled user. The other rows contain k_{TOdb} poses of (p_{TOdb}-1) *impostors*, i.e. people not belonging to the set of enrolled users. In our case we took the impostors from the VidTIMIT public domain database [Sanderson, 2002]. We calculate the covariance matrix (3) from the set of main representative poses of the subject being enrolled and apply the 2DPCA authentication stage, *using the calculated matrix*, to the poses of the TOdb. For each pose of TOdb we calculate the DIFS distance (2) from the set of main representative poses of the subject. Finally, we assign to d_{max} the maximum distance value among all the calculated distances. We can then draw the *Primary Receiver Operating Characteristic* curve (ROC). On a (FRR,FAR) plane, the primary ROC shows the behavior of False Rejection Rate (FRR) and False Acceptance Rate (FAR) for the case where we select a given value of ρ(ρ in [0..d_{max}]) as a threshold for the first authentication stage (note that we always have FRR(ρ=d_{max})=0 and FAR(ρ=0)=0). The primary ROC is a parametric curve where each value of ρ corresponds to a point P and vice-versa.

We therefore choose as a threshold for the first stage the value ρ_{2DPCA} which minimizes the following cost function (4):

$$C(\rho) = C_{FRR} FRR(\rho) + C_{FAR} FAR(\rho) \qquad (4)$$

where C_{FAR} and C_{FRR} are the costs heuristically assigned, respectively, to the case of false acceptance and false rejection (see Section 3 for an example of how they can be assigned). In Fig. 2, on the right, the calculated ρ corresponds to the point $P(\rho_{2DPCA})$.

So far, we have determined the threshold for the first stage and still need to determine the corresponding threshold for the second authentication stage.

For each pose of the TOdb which passes the first stage (i.e. for which $DIFS(I_n,B)<\rho_{2DPCA}$) we consider the corresponding most similar pose (according to the previous stage) in the set of main representative poses of the subject. We then calculate the number *s* of matching SIFT features for each couple so determined and assign to s_{max} the maximum value of *s* over the whole set of poses analyzed. Now we can draw a second graph, named *secondary ROC*, which shows the behavior of FRR and FAR of the aggregated system (i.e. the system composed of the two cascaded authentication stages) when the threshold σ of the second stage varies in the interval $[0..s_{max}]$. In other words, the secondary ROC is a parametric curve describing the performance of the aggregated system when σ varies from 0 to the maximum number of features calculated according to threshold optimization database (note that a pose is accepted by the second stage if the number of detected matching features is *higher* than σ).

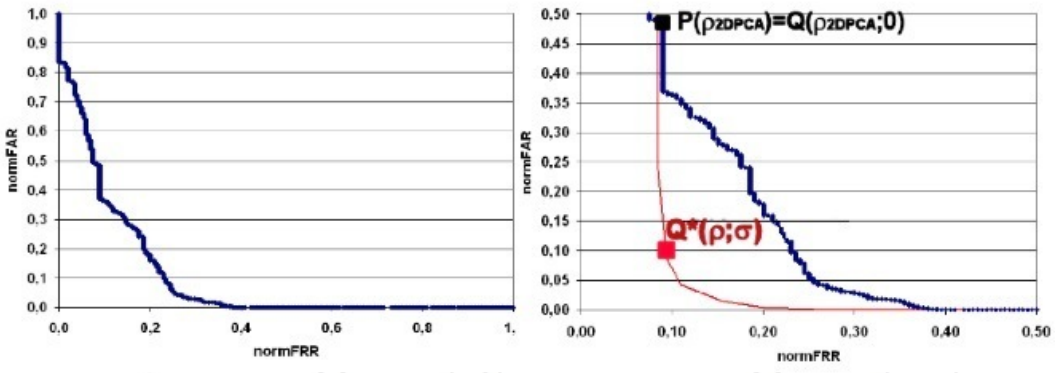


Figure 2
Primary ROC curve (left) and secondary ROC curve (right)

Note that when σ=0, the SIFT stage is "*turned off*" and all images pass through it, so FAR and FRR for this case are the same as those of a single stage 2DPCA system. Therefore, the secondary ROC always intersects the primary ROC in the point $Q(\rho_{2DPCA};0)=P(\rho_{2DPCA})$.

Finally, in order to minimize at the same time both FAR and FRR we choose for s the value which minimizes the function (5):

$$\sqrt{[FRR^*(\rho_{2DPCA};\sigma)]^2+[FAR^*(\rho_{2DPCA};\sigma)]^2} \qquad \sigma \in [0..s_{\max}] \qquad (5)$$

where FRR* and FAR* are FRR and FAR normalized in the interval [0..1], respectively. The corresponding point in Fig. 2 is $Q^*(\rho_{2DPCA}; \sigma_{SIFT})$. The couple $(\rho_{2DPCA}, \sigma_{SIFT})$ so determined is saved in the RFID tag together with the collection of main representative poses of the subject. The data is AES-128 encrypted for security reasons and the encryption key is different for each tag. Indeed, should we use the same key for all the tags, an impostor might find the correct value of the key and create a false RFID tag with their poses and arbitrary identity data (*spoofing attack*). Instead, in VisilabFaceRec every tag is encrypted with a different key $K(x_{EPC}, x_{ID})$ which is generated during the enrolment phase and associated unambiguously to the two codes x_{EPC} and x_{ID}. The first code, x_{EPC}, is determined by the Tag Identification Data, which is unique to each RFID tag, and the second code is the identification code assigned to the enrolled subject by the issuing agency. The association between the encryption key K and the couple (x_{EPC}, x_{ID}) is maintained by a centralized keyserver which does not introduce privacy issues because it does not contain biometric data.

During the authentication phase, the TID data and the x_{ID} are read from the tag and the encryption key is retrieved from the key server. Note that an impostor would not be able to retrieve the correct encryption key from the server because their (x_{EPC}, x_{ID}) couple would not correspond to any entry in the keys database.

2.2. The authentication phase

The authentication algorithm is depicted in Fig. 3.

During the authentication phase a sequence of images of the unknown subject is acquired. In each image, the region containing the face of the subject is detected through a face detection algorithm [Viola and Jones, 2001], cropped and scaled in order to normalize its size. The normalized face region is then compensated for brightness [Chen et al, 2006]. The sequence of images is then passed to the two cascaded authentication stages. At the same time, the content of the RFID has been read and the covariance matrix (3) calculated. If at least one of the images passes both stages, the subject is authenticated.

3. Performance Evaluation

In order to evaluate the FAR and FRR performance of our approach we adopted the “A priori performance type A” procedure based on the VidTIMIT database and proposed in [Sanderson, 2002]. The VidTIMIT database contains a set of poses belonging to 43 different subjects from two different classes, 35 *true claimants* and 8 *impostors*. Tests have been performed with two different configurations, using 2 and 4 main representative poses for each user, respectively.

For each configuration, the poses of each subject are grouped in 3 sessions. The first session is used to create a model of the subject, by determining a set of 2 or 4 main representative poses (according to the configuration of the test) through the clustering procedure described in Section 2.1. The second and third sessions are used for a two-phase testing procedure as follows.

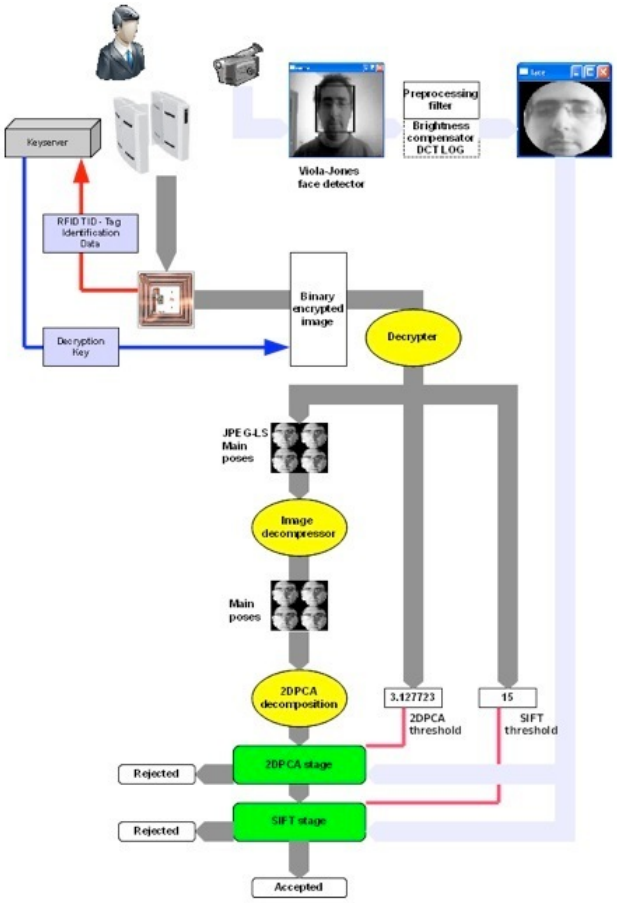


Figure 3
Schematization of the authentication process

During the first phase (phase A) the second session acts as a threshold optimization database (TOdb) for the procedure described in Section 2.1 and the third session is used to test the authentication performance and to measure the FAR and FRR with the thresholds (ρ_{2DPCA}, σ_{SIFT}) just calculated.

During the second phase (phase B) the roles of the second and third sessions are swapped and the procedure of phase A is repeated. Then the FAR and FRR measurements thus obtained are averaged in order to produce a single couple (FAR,FRR) for each subject.

Finally, the values avgFAR and avgFRR are calculated as the averages of FAR and FRR over all the subjects. The couple (avgFAR, avgFRR) constitutes an estimate of the system performance at the given resolution and configuration (2 or 4 main representative poses). For each one of the 35 true claimants we thus simulate 204 authorized authentication trials and 1632 unauthorized (i.e. impostors) authentication trials. The tests were performed at 4 different resolutions (20x15, 40x30, 80x60 and 160x120 pixels) and the number of vector

components for the 2DPCA stage was set as $z_{2DPCA}=w$, i.e. to its theoretical maximum. We now need to assign a value to the ratio C_{FAR}/C_{FRR} (see (4)), which also affects the performance of the system as shown in Fig. 4, drawn for a resolution of 80x60 pixels.

In order to find the value to the C_{FAR}/C_{FRR}, which maximizes the performance of the system, we report the values of normalized FAR and normalized FRR (FAR* and FRR*, respectively) on the plane [0,1]x[0,1] of Fig. 5, thus obtaining a point in the plane for each value of the ratio C_{FAR}/C_{FRR}, representing the performance obtained with that value.

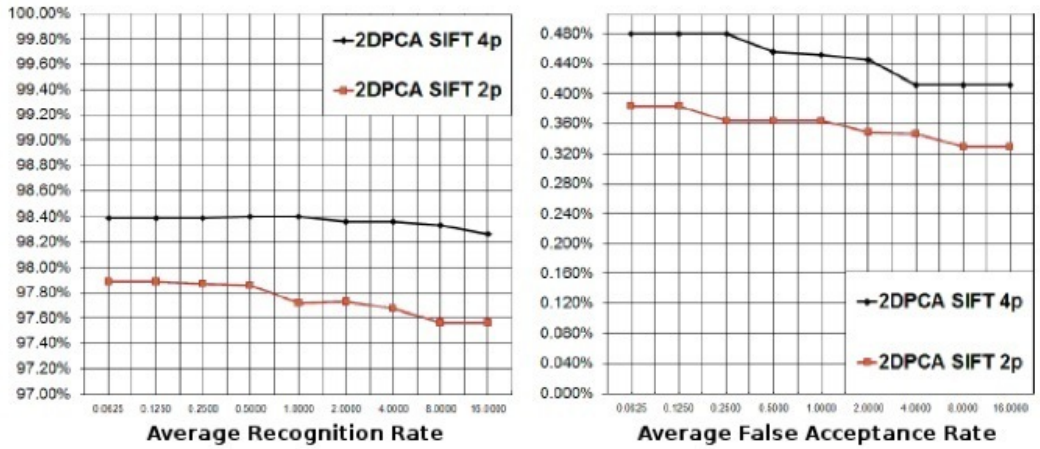


Figure 4
Average recognition rate(left) and average FAR (right) for different values of the C_{FAR}/C_{FRR} ratio. Resolution 80x60 pixel

Our target is to minimize both FAR and FRR, therefore we choose the value for the ratio corresponding to the point closer to the origin of the axes in the plane.

As shown in Fig. 5, the optimal ratio changes according to the number of main representative poses (actually, 2 and 4 for the two configurations). We choose to assign as a common value the average of the optimal values (i.e. 3), assuming that the performance would not be sensibly affected. Similarly, we assumed that the value chosen for the resolution 80x60 would also be acceptable for lower resolutions.

Fig. 6 shows the performance of VisilabFaceRec with changing resolution and two different configurations (2 and 4 main representative poses). The proposed system produces a recognition rate RR>97.69% at all resolutions higher than 20x15 pixels and a false acceptance rate FAR<0.45% for resolutions equal to or higher than 80x60 pixels.

In our two-stage architecture the SIFT stage allows a reduction of FAR while keeping a high value of the recognition rate RR (RR=1-FRR). In order to assess the actual improvement produced in terms of FAR while keeping constant RR, compared to the single-stage architecture, we designed the following experiment.

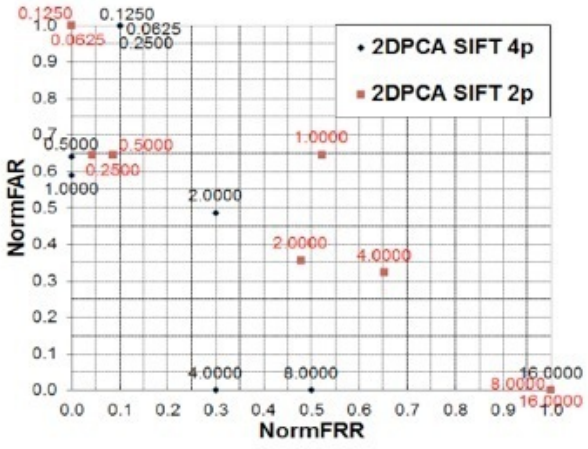


Figure 5
FAR and FRR* for different values of the C_{FAR}/C_{FRR} ratio at a resolution of 80x60 pixels and for 2 and 4 main representative poses for each subject*

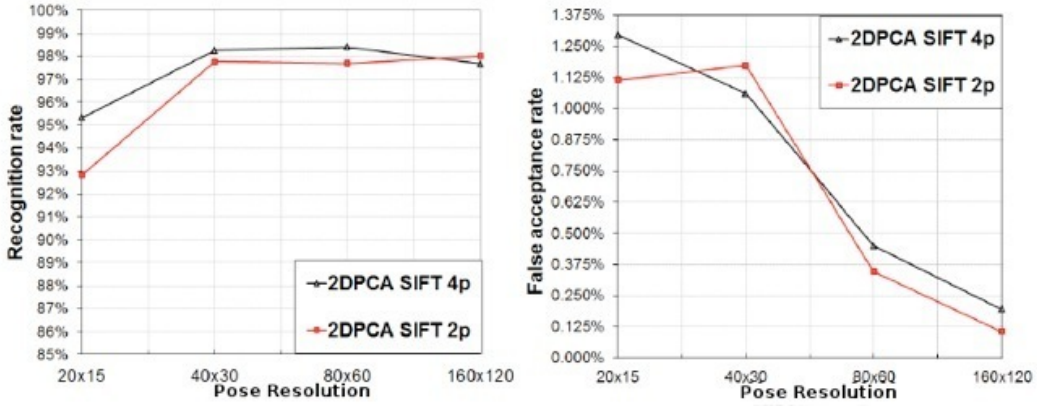


Figure 6
Average recognition rate (left) and FAR (right) vs. resolution

For each subject in the VidTIMIT database we consider the couple ($exFRR^{dblstage}$, $exFAR^{dblstage}$) (expected FRR, expected FAR) produced by the two-stage system according to the procedure in Section 2.1 and also the primary ROC corresponding to the single-stage 2DPCA system. Let us choose a new threshold $\mathbf{p}_{2DPCA}{}^{onestage}$ in such a way that:

$$exFRR^{onestage}(\rho_{2DPCA}^{onestage}) = exFRR^{dblstage}(\rho_{2DPCA}^{twostage}; \sigma_{SIFT}^{twostage}) \qquad (6)$$

In other words, we configure a 2DPCA single stage system so that we obtain the same *expected recognition rate* (exRR = 1-exFRR) than the two-stage system.

Then we run the single-stage system on the testing database in order to produce the actual (i.e., not *expected*) values of the single-stage RR and FAR. The result is shown in Fig. 7.

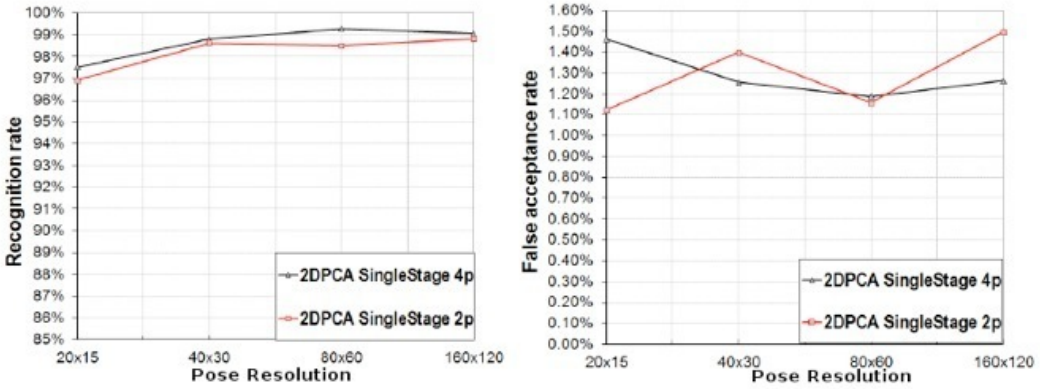


Figure 7
Average RR (left) and average FAR (right) versus pose resolution for the single stage system, while imposing the same expected FRR than the corresponding two-stage 2DPCA-SIFT system

Owing to (6), the recognition rate (Fig. 7, on the left) for the single stage system appears very similar to that of the corresponding two-stage system. Indeed, the single stage performs slightly better, because the second stage, besides rejecting a number of impostors (thus reducing the FAR), also rejects a small minority of true claimants (thus reducing the RR).

The actual improvement in the two-stage over the single-stage is shown by the FAR values (Fig. 7 on the right). The single stage produces a FAR sensibly higher than the two-stage for all resolutions higher than 20x15 pixels (below this resolution the SIFT stage does not operate correctly): Such results demonstrate the correctness of our approach.

Fig. 8 shows the minimum authentication time t_{min} of VisilabFaceRec and its components. The reported values were measured on an ASRock E350M1 board equipped with an AMD Fusion dual core processor running at 1.6GHz and an Inpinji Speedway RFID reader. The system uses high capacity RFID tags produced by Tego Inc.

The authentication time is minimum if the user is authorized immediately after the RFID is read, i.e. when the authentication of the very first image of the claimant succeeds. In this case we have:

$$t_{min}=t_{rfid}+t_{fdet}+t_{fspace}+t_{frec}+t_{sift} \qquad (7)$$

where:

t_{rfid} is the time needed for RFID tag reading, binary decryption, and decompression of the user main poses.

t_{fdet} is the time needed by the Viola-Jones cascade classifier for the detection of the user face.

t_{fspace} is the time needed to calculate the covariance matrix G_t according to (3), its eigenvalues and eigenvectors, and the 2DPCA decomposition vectors of the main poses according to (1).

t_{frec} is the time needed for the 2DPCA decomposition of the acquired image and its comparison with the decomposition vectors of the main representative poses retrieved from the RFID (according to (2)).

t_{sift} is the time needed for the calculation of the SIFT features of the image acquired and the application of the Lowe criterion.

At the resolution of 80x60 pixels, using the configurations with 4 or 2 main representative poses, the authentication requires 13.09 seconds and 6.72 seconds respectively. The dominant component of the authentication time, however, is the time t_{frid} needed to transfer the data from the RFID tag to the board and decode (i.e. decrypt and decompress) it. After the first read, if the authentication of the first image of the claimant fails, the next trials require much shorter times, $t_{min^*} = t_{min} - t_{rfid}$ (0.36 s and 0.29 s respectively), because the RFID data is temporarily cached. As, usually, the first trial is performed when the claimant is still approaching the camera, the first, slower trial in most cases does not significantly affect the perceived speed of the authentication system.

		4p 20x15	4p 40x30	4p 80x60	4p 160x120	2p 20x15	2p 40x30	2p 80x60	2p 160x120
RFID tag reading time t_{rfid} (s)	Avg	1.0072	3.6983	12.7322	50.1272	0.5517	1.9477	6.4282	25.1802
	StdDev	0.0787	0.2021	0.6568	2.5363	0.0395	0.1156	0.3333	1.2846
Face detection time t_{fdet} (s)	Avg	0.0605	0.0650	0.0638	0.0640	0.0605	0.0650	0.0638	0.0640
	StdDev	0.0062	0.0067	0.0071	0.0067	0.0062	0.0067	0.0071	0.0067
Face space rebuilding time t_{fspace} (s)	Avg	0.1208	0.1593	0.2549	0.5812	0.1129	0.1320	0.1887	0.3940
	StdDev	0.0066	0.0093	0.0172	0.0540	0.0048	0.0064	0.0178	0.0787
Face recognition time t_{frec} (s)	Avg	0.0002	0.0007	0.0031	0.0133	0.0002	0.0007	0.0028	0.0122
	StdDev	0.0000	0.0000	0.0001	0.0004	0.0000	0.0000	0.0001	0.0004
SIFT feature computation time t_{sift} (s)	Avg	0.0060	0.0147	0.0393	0.1088	0.0057	0.0143	0.0397	0.1097
	StdDev	0.0014	0.0019	0.0059	0.0218	0.0013	0.0018	0.0059	0.0214
$t_{min} = t_{rfid} + t_{fdet} + t_{fspace} + t_{frec} + t_{sift}$		1.1948	3.9382	13.0933	50.8946	0.7309	2.1597	6.7232	25.7601
$t_{min}^* = t_{fdet} + t_{fspace} + t_{frec} + t_{sift}$		0.1876	0.2398	0.3611	0.7674	0.1793	0.2120	0.2950	0.5799

Figure 8
Operating times of VisilabFaceRec (in seconds)

For a given configuration (number of main representation poses and resolution), the data transfer and decoding time t_{rfid} can be considered as almost constant, as it mainly depends on the transfer rate between the RFID tag and the processing board. As it is shown in Fig. 8, the other time components are characterized by small average values and very small standard deviations in all the tests we performed, therefore we suggest that our approach can be successfully used also for near real-time applications.

4. Conclusions

This paper presented VisilabFaceRec, a multi factor authentication system for controlling the access to services and restricted areas, which combines RFID tags and biometric recognition (face recognition) for the sake of improved accuracy, reliability and privacy. The system is specifically devised to work with very low resolution images, thus allowing for the storing of sensitive user data (e.g., face images) directly into the RFID tag, without the need for a centralized database. To the best of our knowledge, there are no other systems that obtain similar results, in terms of FAR/FRR balance, when working with the same resolution and operating constraints (i.e., using RFIDs to store sensitive data). The proposed system was realized and tested on a commercial board for embedded systems. The obtained execution times are short enough to be suitable for adoption in applications, such as access control of restricted areas, which cannot tolerate long authentication times or afford expensive hardware. Future work will deal with further improvements in the calculation of the two decision thresholds for the cascading stages and with the assessment of other algorithms for the implementation of the two stages with the aim to further increase the reliability of the authentication while reducing the execution time.

References

[Art. 29 WP, 2003]. Data Protection Working Party, Working document on biometrics. 12168/02/EN WP 80. Available online at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2003/wp80_en.pdf.

[Chen et al, 2006] Chen W., Er M.J., Wu S., Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain, IEEE Transactions on Systems, Man, and Cybernetics, 36, 2, 2006, 458-466.

[Flach, 2003] Flach P.A., The Geometry of ROC Space: Understanding Machine Learning Metrics through ROC Isometrics, Twentieth Intnl. Conf. on Machine Learning, 2003, 194–201.

[Jain et al, 2006] Jain A.K., Ross A., Pankanti S., Biometrics: A Tool for Information Security, IEEE Transactions on Information Forensics and Security, 1, 2, 2006.

[Jing et al, 2009] Jing B.Z., Yeung D.S., Ng W.W.Y, Ding H.L., Wu D.L., Wang Q.C., Li J.C., RFID Access authorization by face recognition,

Eighth Intnl. Conf. on Machine Learning and Cybernetics, Baoding,2009, 302-307.

[Lowe, 2004] Lowe D., Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60, 2004, 91-110.

[Meng et al, 2010] Meng X.L., Song Z.W., Li X.Y., RFID-Based Security Authentication System Based on a Novel Face-Recognition Structure, WASE International Conference on Information Engineering, 1, 2010, 97-100.

[Min et al, 2011] Min D.G., Kim J.W., Jun J.S., The Entrance Authentication System in Real-Time using Face Extraction and the RFID Tag, Intnl. Conf. on Ubiquitous Computing and Multimedia Applications, 2011, 20-24.

[Nguyen et al, 2012] Nguyen T.D., Quang L.D., Van N.C., Thanh L.T., Hoang T. M., De Souza-Daw T., An efficient and reliable human resource management system based on a hybrid of face authentication and RFID technology, Fourth Intnl. Conf. on Communications and Electronics (ICCE) 2012, 333-338.

[Pentland, 1991] Turk M., Pentland A., Eigenfaces for Recognition, Journal of Cognitive Neuroscience, 3, 1, 1991, 71-86.

[Sanderson, 2002] Sanderson C., The VidTIMIT Database, Available online at: http://publications.idiap.ch/index.php/publications/show/710.

[Viola and Jones, 2001] Viola P., Jones M., Rapid Object Detection Using a Boosted Cascade of Simple Features, IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii, 2001, 511-518.

[Yang et al, 2004] Yang J., Zhang D., Frangi A.F., Yang J.Y., Two-dimensional PCA: a new approach to appearance-based face representation and recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 26, 1, 2004, 131-137.

Biographies

Filippo Battaglia received the M.S. degree in Electronic Engineering from the University of Messina, Italy, in July, 2008 and the Ph.D. degree in Information Technology from the University "Mediterranea" of Reggio Calabria, Italy, in February, 2013. His research interests include artificial vision, voice synthesis, operating systems, micro-optoelectronics and communication systems.
email: filbattaglia@libero.it

Giancarlo Iannizzotto received the M.D. degree in Electronic Engineering from the University of Catania, Italy, in 1994 and the Ph.D. in Computer Science from the same University in February, 1998. From 1996 to 2006 he was Assistant Professor at the Faculty of Engineering, University of Messina, Italy. From 2006 to 2012 he was Associate Professor at the same Faculty. Currently he is Associate Professor at the Department of Cognitive Sciences, Education and Cultural Studies at the University of Messina. His research activity is in the fields of Computer Vision, Artificial Intelligence, Human-Computer Interaction and Human Factors in ICT.
email: ianni@unime.it

Lucia Lo Bello received the M.S. degree in Electronic Engineering and the Ph.D. degree in Computer Engineering from the University of Catania, Italy, in 1994 and 1998, respectively. She was a Visiting Researcher with the Department of Computer Engineering, Seoul National University, Korea (2000-01). She is currently an Associate Professor with tenure with the Department of Electrical, Electronic and Computer Engineering, University of Catania. Her research interests include real-time systems, wireless networks and sensor networks, factory communications, and embedded systems. She published more than 120 technical papers on peer-reviewed international conferences, books, and journals.
email: lucia.lobello@dieei.unict.it