

Editoriale

Qualche considerazione sulla ricerca scientifica e su come gestirla

Dovendo pianificare un programma di ricerca avanzata in campo scientifico/tecnico, uno dei problemi principali da affrontare è quello di immaginare quali siano gli argomenti più promettenti per una ricerca di base e quali le aree applicative che richiedono un maggiore apporto di tecnologie avanzate. In una discussione alla quale ho partecipato qualche tempo fa in ambito europeo, uno degli argomenti emersi è stato quello degli sviluppi futuri della Internet-of-Things e delle ricerche a questa associate. Devo ammettere che sono rimasto piuttosto imbarazzato e perplesso di fronte a queste questioni di previsioni scientifico/tecnologiche, soprattutto pensandole in una prospettiva storica; detto in poche parole, mi sembra che il problema stia in questi termini: *mentre la ricerca procede in modo bottom-up, la sua gestione dovrebbe procedere in modo top-down*. Ciò causa quello che gli elettrotecnici chiamano un *disadattamento di impedenza*, che rende la pianificazione e il finanziamento della ricerca scientifico/tecnologica un'attività difficile e piena di rischi.

Di fatto, il processo di scoperta scientifica parte dal “nulla” e cerca qualche cosa che precedentemente non si conosceva; il caso ha un posto rilevante nella scoperta scientifica, ma – citando Louis Pasteur – “il caso coglie le menti preparate”. Alcune delle più importanti scoperte scientifiche dei secoli scorsi, anche lasciando perdere la storia della mela di Newton, furono il frutto di eventi casuali tra loro collegati in modo corretto da vari gruppi di ricerca: Fleming scoprì l'effetto antibatterico del Penicillium e Chain e Florey avevano la tecnologia per produrre la Penicillina come farmaco; il Sildenafil veniva studiato nell'ambito della cura delle malattie coronariche, ma è diventato universalmente noto come Viagra per le disfunzioni erettili. Penzias e Wilson stavano ripulendo un'antenna a microonde dai nidi di alcuni uccelli, sospettati di essere l'origine di un noioso rumore di fondo, quando vennero a conoscenza degli studi teorici di un gruppo di ricercatori dell'Università di Princeton sulla radiazione prodotta dal Big Bang e la Radiazione di Fondo a 3 °K divenne un'importantissima scoperta cosmologica. Possiamo pertanto concludere che la ricerca scientifica procede



attraverso il collegamento da parte di “menti preparate” di un certo numero di scoperte indipendenti.

Anche la tecnologia si sviluppa spesso lungo direzioni imprevedute, anche se in modo più “continuo”. La scoperta del Transistore è un buon esempio: nato all’inizio per miniaturizzare prodotti di consumo, quali ad esempio le radioline portatili, è diventato il componente fondamentale per lo sviluppo di circuiti integrati sempre più potenti, che comprendono milioni di transistori su di un unico chip, oggi usati in smartphone e calcolatori portatili più potenti dei mainframe degli anni '50 e in ogni tipo di apparecchiature digitali. E tornando all’argomento di Internet e delle sue applicazioni, il progetto ARPAnet nacque per permettere ai ricercatori di livellare il carico computazionale di picco sulle macchine locali distribuendolo tra le altre macchine collegate in rete (workload sharing). Tuttavia la rete divenne presto il *servizio postale* dei ricercatori, al punto che a qualche burocrate venne l’idea di far pagare il francobollo ad ogni messaggio scambiato! La rete ARPA ha costituito la spina dorsale di Internet e ” l’invenzione” di MOSAIC come interfaccia amichevole ha aperto l’era del WEB con le annesse applicazioni creative che influiscono continuamente sulla vita di ogni giorno.

Un secondo punto sollevato durante la discussione ha riguardato il rispettivo ruolo nell’attività di ricerca dell’accademia e delle aziende; in genere, si tende a pensare che le istituzioni accademiche siano la sede per sviluppare la ricerca di base mentre le aziende si occupano di ricerca applicata e finalizzata. In realtà le cose non vanno sempre così: Penzias e Wilson, al tempo della loro scoperta, stavano lavorando ai Bell Labs, emanazione di una compagnia telefonica americana, e ricercatori dei laboratori IBM di Zurigo hanno ottenuto in due anni consecutivi il Premio Nobel per la Fisica, mentre l’interfaccia MOSAIC, che diede origine al World Wide Web, fu concepita al CERN, laboratorio per lo studio della fisica delle particelle. Ciò dimostra anche che le relazioni tra scienza e tecnologia sono cambiate nel tempo e che i loro confini sono diventati labili: il quadro tradizionale monodirezionale, nel quale la scienza è il motore principale della scoperta di conoscenza e del progresso della tecnologia, mentre quest’ultima è solamente utile alla scienza nel fornirle gli strumenti necessari, è diventato bidirezionale in quanto, come si è visto, la tecnologia stessa può portare alla scoperta scientifica.

Fin qui per quanto riguarda “scoperta” e “invenzione”, ma che cosa si può dire sul lato gestionale? Per dimostrare quanto questo sia un lavoro difficile e a rischio di errori madornali ricorrerò ancora alla storia con due casi emblematici: i) la famosa, anche se dubbia, affermazione di Thomas J. Watson Sr., capo di IBM, nel 1943 “penso che ci sia un mercato mondiale per non più di cinque calcolatori”; ii) lo sfortunato destino della Divisione Elettronica Olivetti, dove, a cavallo tra gli anni '50 e '60, sono state prodotte alcune macchine con tecnologie molto avanzate per l’epoca e dove fu progettata e prodotta la Programma 101 - il primo personal computer -, che non trovò alcun supporto dalle autorità politiche ed economiche nazionali.

E’ pertanto molto difficile costruire categorie con le quali classificare e valutare a priori il merito delle attività di ricerca. D’altra parte, politici e manager devono



pur avere degli elementi sui quali potersi basare per finanziare la ricerca, specialmente quando i finanziamenti disponibili sono limitati.

A mio avviso, occorre fornire direttive sulle aree applicative che hanno maggior bisogno di progredire con lo sviluppo di conoscenze scientifico tecnologiche già acquisite (p. e.: sanità, ecologia, sviluppo urbano,...) e verso le quali indirizzare una parte dei finanziamenti, ma occorre anche dedicare una quantità non trascurabile di risorse alla ricerca di base non finalizzata, che sola può portare a salti qualitativi in aree ancora sconosciute. Sfortunatamente il ritorno di investimento della ricerca di base non è garantito e, in ogni caso, è molto più lento di quello della ricerca finalizzata, ma governanti illuminati devono avere il coraggio di rischiare!

Vorrei concludere con un detto popolare: “La Scienza cerca in una stanza buia un gatto nero; la Filosofia cerca in una stanza buia un gatto nero che non c’è; la Religione cerca in una stanza buia un gatto nero che non c’è ... e lo trova”. Pertanto cerchiamo di aver fede nelle “menti preparate” dei nostri governanti e dei manager della ricerca!

Fabio A. Schreiber
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano

La Nascita della Filosofia Digitale

G.O. Longo, A. Vaccaro

Sommario. *Dopo una rassegna storica delle correnti e degli autori principali che hanno contribuito alla nascita della filosofia digitale, si illustrano gli apporti fondamentali di Edward Fredkin, Gregory Chaitin e Stephen Wolfram, che propugnano la visione di una realtà costituita di informazione e animata dalla continua esecuzione di algoritmi, tra i quali sono fondamentali gli automi cellulari. Infine si illustra la nuova immagine dell'universo che scaturisce da questa visione, in cui l'unica legge è la computazione: essa si incarna in un "paradigma pancomputazionale", secondo cui tutto computa, tutto è frutto di computazione e il cosmo stesso è un Grande Computer.*

Keywords: Algorithm, Cellular automata, Complexity, Computation, Information

1. Introduzione

Dopo decenni di filosofie nichiliste, deboli o decostruzioniste, quando ormai i filosofi sembravano aver rinunciato definitivamente allo sguardo globale sul Tutto, accontentandosi di veder chiaro in qualche settore d'indagine limitato (filosofia del linguaggio, filosofia della mente...), ecco comparire una filosofia dalle sembianze antiche, compiuta in sé stessa, con una sua ontologia, con un'idea forte del divenire e, nientemeno, con una sua metafisica. Tuttavia è bene precisare subito che solo le sembianze sono antiche: i contenuti di questa filosofia sono decisamente nuovi, poiché appartengono all'ambito dell'innovazione per eccellenza, quella tecnologica, e, più specificamente, all'area delle tecnologie dell'informazione. Infatti l'impulso decisivo al costituirsi del nuovo quadro della realtà è stato fornito dal computer, il quale sotto questo profilo si presenta come una vera e propria "macchina filosofica", capace di

suggerire con forza persuasiva che tutte le grandezze della natura sono finite e discrete, e possono quindi essere rappresentate esattamente mediante quantità intere, escludendo dunque ogni variabile infinita, infinitesima, continua o localmente indeterminata e soggetta alla casualità. Inoltre queste grandezze fisiche corrispondono a nient'altro che a configurazioni di bit e la loro evoluzione temporale è governata da processi computazionali. Si apre così lo scenario della *filosofia digitale*, che presenta una forte originalità non solo quanto al ricco patrimonio di idee che racchiude, ma perfino nel configurarsi della sua genesi. E questa genesi occorre innanzitutto illustrare.

2. I prodomi

La storia della filosofia ci ha abituato a uno svolgimento piuttosto ripetitivo: ogni sistema di pensiero si è presentato, in genere, con una paternità certa, indicata dal nome del fondatore (che quasi sempre, con l'aggiunta del suffisso “-ismo”, si trasforma nel nome del sistema: platonismo, pitagorismo, hegelismo...); con un'opera fondamentale, che assurge a testo di culto per gli adepti; e infine con una certa continuità dialettica sia nei confronti delle questioni che il sistema eredita dal passato e a cui offre nuove risposte, sia rispetto ai contributi dei seguaci che approfondiranno, più o meno fedelmente, gli insegnamenti del fondatore.

La filosofia digitale non rispetta questo schema caratteristico. Anche nelle sue origini essa ostenta, quasi con fierezza, i caratteri peculiari del pensiero che incarna, cioè quella cultura digitale che, in ambiti disciplinari sempre più numerosi, fa preferire l'ipertestualità al fluire lineare, il modello a rete e l'interconnessione rispetto alla concatenazione causale-temporale, la condivisione e il *copyleft* piuttosto che l'autorialità tradizionale... In particolare la nascita della filosofia digitale non è contrassegnata da una data esatta e dal nome di un autore, bensì da un insieme di eventi variamente dislocati che si richiamano, interagiscono e si rafforzano a vicenda.

2.1 I nodi più antichi

I nodi più antichi della rete che ha dato forma alla filosofia digitale hanno un'accentuata connotazione tecnico-scientifica. Sono da considerarsi tali: 1) la pubblicazione, nel maggio 1943, di *A Logical Calculus of Ideas Immanent in Nervous Activity*, in cui il neuropsichiatra Warren Mc Culloch e il matematico Walter Pitts profilano, per la prima volta, l'analogia tra il neurone e un dispositivo logico a due stati, dando avvio all'altalenante – e affascinante – storia del connessionismo; 2) la nascita ufficiale della cosiddetta “Teoria dell'informazione”, riconducibile alla pubblicazione in due puntate, nel luglio e nell'ottobre del 1948, di *A Mathematical Theory of Communication*, in cui Claude Shannon conia tra l'altro l'espressione “bit”, acronimo di *binary information unit*, per definire l'unità di misura della quantità d'informazione (mentre in altri contesti “bit” è contrazione di *bi(nary digi)t* o cifra binaria); 3) l'esordio della cibernetica con la prima delle dieci storiche *Macy Conferences*, tenutasi a New York nel marzo 1946, cui parteciparono, tra gli altri, studiosi del calibro di John von Neumann, Claude Shannon, Heinz von Foerster, Gregory Bateson, Kurt Lewin, i citati Pitts e Mc Culloch e Norbert Wiener che, di per sé, con il suo 4)

Cybernetics: Or Control and Communication in the Animal and the Machine del 1948, costituisce un altro punto di riferimento notevole per la filosofia digitale. In *Cybernetics*, Wiener sostiene di aver individuato nell'informazione il minimo comun denominatore dei sistemi fisici, biologici, sociali e artificiali.

Nel regno della biologia, la tesi di Wiener viene ad essere avvalorata, circa cinque anni più tardi, da *Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid* di James Watson e Francis Crick, un lavoro pionieristico che impone la metafora (che è più di una metafora) del genoma come deposito di informazioni codificate, con tanto di destinatario (le proteine) e di canale o messaggero (l'Rna). Anche il dominio della fisica, da parte sua, si lascia permeare in maniera sempre più convinta dall'intuizione di Wiener man mano che approfondisce l'indagine sulla meccanica quantistica: John Archibald Wheeler, nel suo autobiografico *Geons, Black Holes & Quantum Foam*, sintetizza la presa di potere dell'informazione con una confessione emblematica: «Più rifletto sul mistero dei quanti e sulla nostra singolare capacità di comprendere il mondo in cui viviamo, più mi persuado che la logica e l'informazione possono avere un ruolo basilare nelle fondamenta della teoria fisica»[12]. La sua convinzione, espressa anche a nome di un nutrito novero di studiosi nel settore della quantistica, viene riassunta nella celebre formula: *it from bit*, che esprime la primazia del *bit*, dell'informazione, rispetto all'*it*, cioè alla materia, che da quella deriverebbe.

Nella rete della filosofia digitale che comincia così a configurarsi vengono ad aggiungersi, negli anni Ottanta, quelli che potrebbero essere definiti due nuclei fondamentali.

2.2 I nuclei fondamentali

Nel maggio 1981, nel corso del primo convegno di "Fisica e computazione", tenutosi nel Centro Congressi del Massachusetts Institute of Technology, alcuni dei personaggi più autorevoli della fisica del tempo, a partire da Richard Feynman, cui era stata affidata la prolusione dei lavori, enunciano una metafora ardita: l'Universo sarebbe un Grande Computer. Edward Fredkin – la mente organizzativa del simposio – aveva fatto di tutto per conferire prestigio e autorevolezza all'evento e, in questo senso, avere l'introduzione di Feynman era davvero il massimo. Fredkin conosceva personalmente piuttosto bene Feynman. Nel suo *Digital Mechanics* ricorda che nel 1962, quasi trentenne, prese il coraggio a due mani e, insieme con l'amico Marvin Minsky, bussò alla porta del fisico, già celebre, per esporgli la sua ossessiva e bizzarra idea della meccanica digitale. Egli – scrive Fredkin – ascoltò con gentilezza e attenzione, ma si raccomandò, alla fine, di non parlare con nessuno del suo interesse per tale prospettiva. L'effetto di quella conversazione, seguita da molte altre, si riscontra tuttavia, due anni dopo, nel saggio di Feynman *Il carattere della legge fisica*: «Mi lascia sempre perplesso il fatto che, secondo le leggi della fisica come le comprendiamo oggi, sia necessaria una macchina calcolatrice e un numero infinito di operazioni logiche per individuare ciò che accade in una regione dello spazio o in un intervallo di tempo piccoli a piacere... Così ho fatto spesso l'ipotesi che, alla fine, la fisica non avrà più bisogno di un'enunciazione

matematica, e si scoprirà che le sue leggi sono semplici, come una scacchiera di automi cellulari. Si tratta, tuttavia, di una speculazione come un'altra»[5].

Michael Dertouzos, nella sua presentazione al discorso inaugurale di Feynman al convegno del 1981, fece menzione di questo sentimento ambivalente del relatore nei confronti dell'idea della fisica come processo computazionale e del suo rifiuto a partecipare ai lavori se essi si fossero svolti all'insegna del titolo originario, "Modelli computazionali della fisica", che, per questo motivo, fu trasformato nel più vago "Fisica e computazione". A questo punto, però, l'oratore interruppe vivacemente il presentatore con un espressivo: «Ebbene, da allora ho cambiato idea!»[6], dando inizio alla sua magistrale lezione *Simulating Physics with Computers* che – spiegò – non mirava alla questione se il computer potesse compiere previsioni sugli eventi fisici come già faceva per le condizioni meteorologiche, bensì, molto più profondamente, se al cuore della fisica stesse la medesima legge del divenire che muove un calcolatore elettronico, ovvero un processo computazionale. Le forti parole di Feynman infusero molto coraggio ai contributi che seguirono, e a questa determinazione contribuì molto anche Wheeler, sottoscrivendo esplicitamente l'idea nel suo *The Computer and the Universe*.

L'intervento dell'informatico tedesco Konrad Zuse, *The Computing Universe*[14], sorprese per le sue dirompenti aperture. Anche Zuse – geniale inventore, negli anni Quaranta, del primo computer programmabile e del primo linguaggio di programmazione di alto livello – era una "scoperta" di Fredkin, il quale, con il suo gruppo di lavoro, aveva tradotto il pionieristico *Rechnender Raum* in *Calculating Space*, riscattandone l'autore dall'isolamento in cui l'avevano relegato l'esito della guerra e le idee troppo avanzate per il mondo accademico europeo dell'epoca. Invitato al Mit, Zuse aveva esposto concetti che molto a lungo aveva tenuto per sé: l'inspiegabile ma evidente corrispondenza tra il comportamento delle particelle digitali (bit) e quello delle particelle subatomiche; la quantizzazione sempre più convalidata delle grandezze fisiche, con il conseguente abbandono del "dogma" del *continuum* in natura; la progressiva coincidenza, nella terminologia e nei concetti, della teoria dell'informazione e della fisica (com'era avvenuto, per esempio, con "informazione" ed "entropia"); il ruolo degli automi cellulari, già in grado di risolvere equazioni differenziali, nell'esplorazione dell'Universo... Al termine del convegno, erano tutti molto più convinti che l'asserzione "il Cosmo è un Grande Computer" non fosse da intendere come una semplice metafora, bensì come un valido strumento euristico.

Tommaso Toffoli, stretto collaboratore di Fredkin, contribuì notevolmente alla diffusione dell'idea formulandola così: «In un certo senso, da miliardi di anni la Natura sta continuamente computando il proprio stato successivo: tutto quello che dobbiamo fare – e, in effetti, *possiamo* fare – è farci dare un passaggio da questa enorme computazione in corso e cercare di scoprire in quali punti essa si avvicina al luogo dove vogliamo andare noi»[11]: questa è diventata la citazione preferita di tutti coloro che vogliono far riferimento alla tesi dell'Universo computante. Per il resto, fu un assemblaggio quasi "cubista" di riflessioni sul concetto di "informazione" nelle sue dimensioni storiche, cosmologiche, fisico-matematiche, teoriche, e così via. Lo scienziato della Nasa e dell'Ibm Rolf Landauer asseriva perentoriamente che "l'informazione è fisica" e

il matematico Gregory Chaitin, nel suo ambito di competenza, elaborava l'idea secondo cui "tutto è algoritmo".

Il secondo nucleo che negli anni Ottanta contribuì alla nascita della filosofia digitale fu il primo convegno interdisciplinare sulla sintesi e la simulazione dei sistemi viventi, tenutosi nel settembre 1987 a Los Alamos sotto la direzione di Christopher Langton, fondatore riconosciuto dell'area di ricerca denominata *Vita Artificiale*. Come il convegno del Mit rappresenta il maggior contributo alla genesi della filosofia digitale sotto il profilo scientifico-concettuale, così il convegno di Los Alamos ne costituisce la spinta maggiore sotto l'aspetto tecnico-intuitivo. Come il primo tendeva a mostrare quanto la natura fosse vicina, nel suo intimo funzionamento, al procedere del computer, così il secondo lasciava intravedere quanto certi processi elaborati dal computer fossero sorprendentemente simili ai fenomeni della vita reale. Al confluire dei due percorsi scattava la scintilla della filosofia digitale.

Furono le immagini di Vita Artificiale sugli schermi dei computer, dunque, a impressionare i convegnisti radunatisi a Los Alamos. A metà anni Ottanta, in effetti, sarebbe stato difficile non rimanere stupiti e sconcertati osservando come veniva simulata la crescita delle piante: da un esile fusto spuntavano via via ramoscelli e, su questi, fiori e frutti, in un divenire indistinguibile da quanto accade in natura. Oppure ammirando una simulazione ancor più famosa, dovuta a Craig Reynolds, in cui uno stormo di uccelli (più precisamente uccelloidi o *boids*) spiegava un volo geometrico in un ambiente costellato di ostacoli di vario tipo, che costringevano lo stormo a rompere temporaneamente l'ordine del volo per aggirarli, salvo riorganizzarsi subito dopo ogni ostacolo. E lo stupore e gli interrogativi non potevano che aumentare quando si veniva a sapere che i comportamenti di quelle pianticelle e degli stormi di *boids* non erano frutto di un sistema di istruzioni astruso, ingombrante e di alto livello, ma scaturivano da regole di base molto semplici. La simulazione al calcolatore si proponeva dunque come terzo strumento conoscitivo accanto alla deduzione aristotelica e all'induzione baconiana. La via inaugurata dal gioco *Life* (in sostanza un automa cellulare) ideato nel 1970 da John Conway, incrociandosi con la teoria dell'auto-riproduzione non-biologica inaugurata anni prima da John von Neumann, con la scoperta del valore euristico delle simulazioni e con altre intuizioni limitrofe, stava dando forma ad un paesaggio che si estendeva oltre l'ambito delle indagini settoriali meramente tecniche per raggiungere il livello di una vera e propria concezione del mondo e della vita.

3. Gli esponenti e le idee principali

Ognuno dei principali autori che orbitano intorno alla filosofia digitale approfondisce con acutezza e originalità alcuni aspetti del ricco repertorio concettuale che essa presenta: i concetti-cardine di "informazione", "algoritmo", "computazione", "automa cellulare" e così via. Dalla composizione di questi apporti individuali scaturisce il grande quadro della filosofia digitale. È dunque opportuno seguire gli apporti di alcuni di questi esponenti primari per poi concludere con un tentativo di sintesi.

3.1 Edward Fredkin, la *Digital Philosophy* e l'informazione

È a Edward Fredkin che tocca il posto d'onore tra gli autori del nuovo sistema di pensiero, fosse solo perché è lui a coniare l'espressione "filosofia digitale" quando si rende conto che il nucleo concettuale della sua impostazione originaria si è ampliato a tal punto da rendere assolutamente inadeguata la definizione precedente di "fisica digitale". Oltre a ciò, nel 2002 Fredkin redige un testo dal titolo molto esplicito, *Introduction to Digital Philosophy*, per quanto, sin dalla prefazione, l'autore ammetta che il piano dell'opera è ben inferiore alle potenzialità del titolo. Sebbene, infatti, la filosofia digitale possa fornire risultati fecondi in vari ambiti scientifici, l'indagine condotta in questo libro si limita intenzionalmente ai soli processi fondamentali della fisica. Di conseguenza, l'*Introduction* finisce col riesporre, in forma ampliata, quanto Fredkin aveva già presentato in *Digital Mechanics* nel 1990, in *A New Cosmogony* del 1992 e, soprattutto, in *Finite Nature* dello stesso anno, ovvero la tesi che tutte le grandezze presenti in natura sono finite e discrete, e possono quindi essere rappresentate esattamente mediante quantità intere, escludendo ogni variabile infinita, infinitesima, continua o localmente indeterminata e soggetta alla casualità. Una volta presentato anche il secondo principio basilare, ovvero che le suddette grandezze fisiche corrispondono a configurazioni di bit e che la loro evoluzione temporale è governata da processi computazionali, ecco che l'autore può erigere una minuziosa architettura, comprendente i momenti, le forze, l'inerzia, l'energia, la teoria dei campi discreti e molti altri aspetti, che vengono esaminati rigorosamente e introdotti nel quadro della fisica digitale.

Allargando tuttavia l'orizzonte oltre l'ambito disciplinare specifico, l'apporto maggiore di Fredkin consiste in un'asserzione ontologica: l'informazione è il principio primo della realtà, il suo elemento costitutivo. In altri termini: dove Pitagora poneva i numeri e Leibniz immaginava le monadi, ecco che Fredkin colloca l'informazione. E questa è, senza dubbio, un'asserzione filosofica. «Non credo che esistano oggetti quali gli elettroni e i fotoni, o cose che siano sé stesse e nient'altro. Credo che vi sia un processo d'informazione, e che i bit, quando si trovano in certe configurazioni, si comportino come le cose che chiamiamo "elettrone" o "atomo d'idrogeno" o in altri modi ancora»[7], dichiara Fredkin; e se questo enunciato restringe troppo l'obiettivo al campo della fisica, non manca, subito dopo, la dovuta estensione: «Esistono tre grandi domande filosofiche: Che cos'è la vita? Che cosa sono la coscienza, il pensiero, la memoria e simili? Come funziona l'Universo? Il punto di vista informazionale le concerne tutte e tre»[7].

Così il quadro è completo: l'informazione è alla base della realtà materiale, che fin dal tempo dei presocratici costituisce il campo d'indagine privilegiato della filosofia, ma è anche alla base della realtà mentale del soggetto che si pone la domanda e investiga. In parole ancora più trasparenti, l'informazione (questa volta intesa in senso semantico) è anche alla base della formulazione della verità, il cui possesso dovrebbe acquietare la sete di conoscenza che muove la ricerca. L'informazione è insieme l'oggetto e il soggetto. Informazionale è la natura della verità: secondo Fredkin tutto si muove all'interno di questo circolo.

3.2 Gregory Chaitin, la metafisica digitale e l'algoritmo

Quando ci si comincia a chiedere quale sia la vera essenza di un oggetto e quale sia la natura di una legge fisica, è segno che ci si sta allontanando dal sicuro porto della scienza per dirigersi verso più incerti approdi metafisici. Gregory Chaitin, ricercatore eclettico ma soprattutto matematico, non esita a intraprendere questa rotta. Il vento che lo sospinge è, anche nel suo caso, dichiaratamente digitale.

L'elemento che ha reso necessario il ritorno a una riflessione metafisica oramai tramontata da tempo dal panorama filosofico (specie di quello analitico) è stato il computer. Il computer inteso non come macchina fisica per calcolare, bensì – dice Chaitin – come «nuovo e meraviglioso concetto filosofico e matematico»[1]. Il computer diventa concetto filosofico-matematico in quanto conferisce significati prima inimmaginabili ai verbi “comprendere” e “conoscere”. Dopo l'avvento del computer, dichiara Chaitin, comprendere l'essenza di un ente consiste nell'identificarne il contenuto d'informazione algoritmica, ovvero il programma informatico che restituisce, sullo schermo, la forma dell'ente considerato; inoltre, dopo l'avvento del computer, conoscere la legge fisica che regola il *divenire* di un fenomeno equivale a individuare il processo computazionale che simula, sullo schermo, l'evoluzione di quel fenomeno.

Per quanto concerne l'essenza di un ente si può sollevare subito un'obiezione, sulla base della differenza tra un oggetto reale e la simulazione dello stesso oggetto; ma qui subentra un fondamentale assioma digitale, secondo cui l'essenza si identifica con la forma, l'organizzazione, la configurazione, il *pattern* dell'ente, e non certo con gli anonimi, indiscernibili, sostituibilissimi elementi materiali. Del resto, chiunque in filosofia abbia trattato di “essenza”, dalla tradizione platonica a quella aristotelico-tomista giù giù fino a Hegel, sa bene che ciò per cui una cosa è quello che è, e non un'altra cosa è da riconoscere nell'elemento intelligibile-formale. Dopo l'avvento del computer, tale elemento intelligibile-formale coincide con lo “schema informazionale”.

Per quanto riguarda il concetto di “legge”, l'elaborazione di Chaitin assume una coloritura ancora più digitale. Dato che l'essenza di un ente, e quindi anche di un fenomeno, è la sua descrizione informazionale, in codice binario, si intenderà per legge di quel dato fenomeno il programma informatico capace di generarne la simulazione esatta. Più il programma è breve (Chaitin dice *elegante*) – e quindi maggiore è la compressione dei dati – migliore è la legge. L'impossibilità di ottenere una compressione dei dati equivale a un'assenza della legge: in altri termini, ci si trova di fronte a un evento casuale.

Siccome con Chaitin siamo in ambito digitale, ma pur sempre metafisico, è inevitabile che egli rivolga un omaggio a Gottfried Wilhelm Leibniz, il filosofo che ha percorso i tempi ed è stato l'unico pensatore digitale prima della comparsa del computer: «La teoria migliore è il più piccolo programma per computer che riproduca esattamente i dati empirici. Questo è ciò che sostiene Leibniz, secondo cui la scienza è possibile e il mondo è comprensibile precisamente perché Dio minimizza la complessità delle leggi di natura e insieme massimizza la ricchezza e la diversità dell'Universo determinate da queste leggi. Nel mio modello, sia le

leggi di natura sia l'Universo che ne risulta sono rappresentati come stringhe finite di bit... Le leggi sono un programma e l'Universo il suo output»[7].

3.3 Stephen Wolfram, *Un nuovo tipo di scienza e gli automi cellulari*

Stephen Wolfram rappresenta il caso di un *enfant prodige* che non ha tradito le attese e, da grande, è diventato, per così dire, un *homme prodige*. A sedici-diciassette anni pubblica articoli di fisica subatomica su riviste autorevoli come "Nuclear Physics" e, intorno ai quaranta, nel 2002, pubblica in proprio il suo monumentale *A New Kind of Science*. Nel frattempo ha fondato un'azienda multimilionaria che ruota intorno a *Mathematica*, un software di sua invenzione diventato strumento imprescindibile di ogni istituzione scientifica e didattica di alto livello.

Wolfram non soffre di immodestia: «Circa tre secoli fa – scrive all'inizio della sua maestosa opera – la scienza fu trasformata da un'idea nuova e fondamentale, secondo cui per descrivere il mondo naturale si potevano usare regole espresse da equazioni matematiche. Il mio scopo in questo libro è dare avvio ad un'altra trasformazione di portata simile»[13]. La scienza nuova che Wolfram inaugura è basata su un tipo di regole molto più generali, incorporabili in programmi per computer piuttosto semplici. Non c'è ragione per credere – osserva Wolfram – che i sistemi che osserviamo in natura debbano seguire soltanto le regole della matematica tradizionale; l'unica ragione per cui finora è sopravvissuta tale convinzione è il limite dell'intelligenza umana: la nostra capacità di calcolo (il nostro cervello) non può materialmente spingersi oltre una determinata soglia. Questa stessa soglia, però, non delimita affatto le potenzialità dei computer, i cui programmi possono attuare varietà enormi di regole.

Il dato più sorprendente per Wolfram, tuttavia, è che, procedendo a una verifica sistematica di tale varietà di regole, si vede che sono proprio i programmi più semplici a rendere ragione dei fenomeni complessi della natura. Ecco la scoperta che conduce alla necessità di ripensare il funzionamento dei processi naturali, ecco la soluzione del mistero più inavvicinabile del mondo naturale: come esso faccia a produrre, senza sforzo apparente, realtà tanto complesse. La necessità di un nuovo tipo di scienza è giustificata con una serie di ragioni. Per qualche tempo – osserva Wolfram – la scienza (la matematica) tradizionale ci ha illuso di poterci misurare con l'enigma della complessità, ma ciò che funziona nella spiegazione e previsione dei moti planetari non funziona per comportamenti enormemente più complessi, quali i regimi di turbolenza dei fluidi o i meccanismi di crescita di varie specie di piante. La nuova scienza (e la nuova matematica) invece – a giudizio del suo autore – può affrontare anche fenomeni di complessità enorme. Non solo. La scienza tradizionale ha affrontato i problemi con un metodo rigorosamente riduzionista (da molti considerato l'unico metodo davvero scientifico), secondo cui le proprietà di un sistema derivano dalle proprietà dei suoi componenti: ma il passaggio dal comportamento dei componenti al comportamento del sistema rimane spesso un problema insolubile, soprattutto per la presenza di proprietà emergenti, dovute alle interazioni tra i componenti. La nuova scienza affronta con successo anche questo passaggio. E ancora: mentre la scienza tradizionale presta molta

attenzione alla differenza tra le caratteristiche dei sistemi fisici, chimici o biologici, la scienza nuova ci fa scoprire che le stesse forme di comportamento compaiono universalmente, a prescindere dalla natura dei sistemi. E, guarda caso, ricorrono anche nei programmi informatici semplici.

Tali programmi semplici, a cui il testo rimanda costantemente, sono gli *automi cellulari*, al cui studio e contemplazione Wolfram ha dedicato dieci anni (di notti) della sua vita, per poterne scandagliare tutti i segreti e tutte le prerogative. E tra tutti i 256 automi cellulari unidimensionali di tre celle da lui studiati ve n'è uno, identificato dalla regola di trasformazione numero 30, che è assunto ad emblema della scienza nuova perché, pur non seguendo una legge più complicata degli altri 255, produce senza sosta schemi estremamente irregolari e complessi. La sensazionale scoperta di Wolfram consiste in sostanza in questo: partendo da programmi e da condizioni iniziali decisamente semplici, senza inserire alcun elemento di complessità, ne emerge spontaneamente un comportamento altamente complesso: l'evoluzione degli automi cellulari porta alla comparsa di figure identiche alle strutture microscopiche delle venature delle foglie, ai disegni sulle conchiglie o alle forme di gorgi, vortici, esplosioni e nebulose in via di formazione (si noti che vi è una certa analogia con i frattali, strutture molto complesse anch'esse generate da regole ricorsive semplicissime).

Le conseguenze di tale scoperta trovano applicazione non solo nei sistemi fisici, ma anche in quelli biologici. Infatti il mondo è uno e una sola dev'essere la legge del suo divenire: la computazione. Con buona pace della scienza tradizionale, costretta a formulare leggi diverse per ogni diverso dominio della realtà. Anche il quadro dell'evoluzione rimane profondamente coinvolto da queste idee rivoluzionarie, che offrono una spiegazione alternativa a quella darwiniana tradizionale: e tutto, si potrebbe dire, a causa dello stupefacente automa cellulare numero 30 che, appropriatamente, Michael S. Malone asserisce essere per Wolfram «l'equivalente del fringuello di Darwin e del pisello dolce di Mendel»[9]. Insomma, il mondo è uno e la computazione è la sua legge.

4. La filosofia digitale è una vera filosofia

Le riflessioni di Fredkin, Chaitin e Wolfram sono le espressioni forse più esplicite di un movimento che ha ormai oltrepassato la soglia delle intuizioni sporadiche per consolidarsi in un sistema filosofico dai tratti ben delineati e dagli orizzonti considerevolmente estesi: un sistema filosofico, dunque, con cui è necessario confrontarsi seriamente, anche perché la sua parabola sembra possedere ampi margini di ascesa.

4.1 Ontologia digitale

L'annuncio che l'informazione è il principio ultimo e definitivo della realtà non è proclamato solo dalle voci dei tre solisti sopra menzionati. Al contrario esso è sorretto, per così dire, da un coro che si leva da più ambiti disciplinari e che trova facile eco in primo luogo nel settore delle tecnologie avanzate, dove un nutrito gruppo di esperti concordano con quanto lo storico e preveggenete cofondatore del *Mit Media Lab*, Nicholas Negroponte, ripete da tempo, ovvero che «il passaggio dagli atomi ai bit... è irreversibile e inarrestabile»[10]. Ma si leva

anche dall'area degli studi sul cervello e la mente: per esempio David Chalmers raccoglie con minuziosa attenzione tutte le novità del settore nel suo sito consc.net e può concludere che «l'informazione è una candidata naturale ad avere un ruolo fondamentale anche nella teoria della coscienza»[2]. Il messaggio, peraltro, circola, in maniera piuttosto inattesa, anche nell'ambito delle ricerche evoluzionistiche, come conferma nel suo celebre *L'orologio cieco* un autore non certo creazionista come Richard Dawkins: «al cuore di ogni cosa vivente non c'è fuoco, né alito caldo, né una "scintilla di vita", bensì informazione, parole, istruzioni. Se si vuole ricorrere a una metafora, non si deve pensare a fuochi e scintille e respiro. Si pensi, invece, a un miliardo di caratteri discreti, digitali, incisi su tavolette di cristallo. Se si vuol comprendere la vita, non si deve pensare a gel e a poltiglie vibranti e palpitanti, bensì alla tecnologia dell'informazione»[3]. Dopo una dichiarazione del genere ogni altra attestazione risulterebbe pleonastica.

Da tutto ciò sembra potersi concludere che, per tutti questi autori, alla base della realtà (anzi, di ogni realtà: naturale, artificiale, mentale ...) sta l'informazione. Conformandoci alla tradizione – e al linguaggio – della filosofia classica, l'informazione è l'*arché*, il principio primo che costituisce l'origine e il fondamento del Tutto. La sua natura è immateriale, anche se non possiede quei requisiti tipici che la tradizione assegnava al primo polo nei dualismi Spirito-Materia e Soggetto-Oggetto: si potrebbe parlare quasi di un'immaterialità neutra, o laica. Con buona parte della tradizione, invece, la filosofia digitale condivide l'idea che sia proprio tale elemento immateriale a svolgere la funzione di "sostanzializzazione" della realtà materiale: così è stato per il numero di Pitagora, per l'idea iperuranica di Platone, per la monade leibniziana e via dicendo.

Tentare di approfondire ancora l'analisi dell'informazione – oltre la sua funzione nella realtà – magari cercando di darne una definizione, sarebbe improprio perché, come sottolinea Fredkin, l'elemento ultimo non può, per sua natura, essere ricondotto ad altro. Illuminante è tuttavia la definizione di sapore operativo che ne dà Gregory Bateson: "l'informazione è una differenza capace di generare una differenza", e ancora: "l'unità d'informazione è la più piccola differenza capace di generare una differenza". Si è portati a immaginare che dal vuoto cosmico, omogeneo e indifferenziato, sia scaturito un bit, la minima unità di realtà-informazione, in senso sia ontologico sia percettivo. Da quel bit fu l'esplosione della realtà. Da allora, quell'Evento primordiale si rifrange, a livello semantico o sintattico, in ogni ordine di esperienza.

4.2 L'unica legge è la computazione

Nella filosofia digitale i bit sono gli unici elementi costitutivi della realtà; allo stesso modo la computazione è l'unica legge che regola il divenire della realtà, a tutti i livelli. Il principio secondo cui la computazione è il vero motore del cosmo è così potente da generare tre verità fondamentali: a) tutto computa, b) tutto è prodotto dalla computazione, c) tutto può essere trasformato in un dispositivo che computa.

La formula "tutto computa" racchiude in particolare la suggestiva immagine del cosmo come Grande Computer e si esplicita, in forma più tecnica, nella

dinamica degli automi cellulari. Vengono alla mente le parole di Toffoli, secondo cui da miliardi di anni la Natura non cessa di computare il proprio stato successivo. Se l'idea appare troppo astratta, per ricondurla al concreto basta pensare alle ricerche sul cosiddetto *natural computing* in cui si cerca di carpire alla Natura i più intimi segreti computazionali per trarne vantaggio a livello ingegneristico sotto il profilo sia dell'economicità sia dell'estetica. Ciò presuppone non solo che in Natura esistano modelli computazionali, ma addirittura che essi superino, sotto molti aspetti, quelli concepiti dalla mente umana. Tra le fonti d'ispirazione del *natural computing*, menzioniamo la computazione neurale, la cui potenza di calcolo in parallelo è ancora ben al di là della portata delle ricerche informatiche attuali; la logica genotipo-fenotipo, per l'ottimizzazione degli algoritmi genetici ed evolutivi; la cosiddetta *swarm intelligence*, o dello sciame, cioè l'intelligenza collettiva degli insetti sociali (api, formiche), così efficace nel determinare il comportamento complessivo in base a leggi locali semplicissime, come ha ampiamente dimostrato la simulazione dei *boids* di Reynolds. All'asserzione "tutto computa", la filosofia digitale ne affianca un'altra: "tutto è frutto di computazione". Quindi non solo tutti gli attori, materiali o animali o artificiali, della realtà, compresa la Natura stessa, computano, ma essi stessi (attori e Natura) sono prodotti di computazione. All'interno di questo principio si aprirebbe la delicata questione del cosiddetto "evoluzionismo digitale", l'unica alternativa scientifica al darwinismo che la storia abbia saputo produrre. Il tema è cruciale e sarebbe scorretto liquidarlo in poche battute. Qui si può solo annotare che il movimento di ricerca in questa direzione sta infoltendo le proprie schiere, lanciando segnali di una possibile svolta epocale.

Al "tutto computa" e al "tutto è frutto di computazione" si aggiunge infine la postilla non del tutto trascurabile secondo cui "tutto può essere trasformato in un computer (secondo la prospettiva ingegneristica umana)". Che la Natura nel suo complesso computi è la prima certezza della filosofia digitale, ma che ogni elemento del Grande Computer Cosmico possa essere separato dal suo contesto originario per essere a sua volta considerato un computer "indipendente" sta diventando sempre più chiaro grazie agli esiti assai promettenti del *molecular computing*. Questo settore fu avviato intorno alla metà degli anni Novanta quando il matematico Max Adleman mostrò al mondo il suo computer a DNA, ove i dati erano codificati tramite filamenti di materiale genetico e le operazioni erano eseguite da altre strutture biomolecolari. E se vertiginose sono le aperture del computazionismo molecolare, abissali potrebbero essere definite quelle della computazione quantistica e di altri settori di ricerca affini.

4.3 Filosofia digitale e scienza

L'intreccio tra filosofia digitale e scienza ha molte trame. Se ogni sistema filosofico, soprattutto a partire dall'età moderna, si è confrontato in modo più o meno deciso con la scienza, oggi le interrelazioni in questo senso sembrano essere obbligatorie. Da una parte gli scienziati tradizionali (ovvero gli scienziati *tout court*), anche quelli meglio disposti nei confronti del nuovo sistema di pensiero, sono portati a sospendere il giudizio, perché l'idea che l'informazione sia il principio primo del cosmo, seppur suggestiva, dev'essere ancora verificata

con metodi scientifici. D'altra parte filosofi e scienziati "digitali", come ad esempio Wolfram, osservano che si è sì in dovere di avvalorare tale filosofia tramite una ratifica scientifica, purché essa sia esercitata non secondo i canoni della vecchia logica fisico-meccanicistica del Cosmo come Grande Orologio, bensì con i criteri del nuovo tipo di scienza, una scienza info-computazionale. In effetti sarebbe paradossale tentare di validare una filosofia che vuol fondare una nuova scienza con i criteri scientifici che si vogliono superare.

In questo spirito, l'esperta di *computer science* Gordana Dodig-Crnkovic ha proposto un utile *Dialogo su due sistemi del mondo: info-computazionale e meccanicistico*[4], che imposta bene il confronto e, sin dal titolo, suggerisce che potremmo trovarci su un tornante di storia della scienza simile a quello della rivoluzione galileiana. Nella visione meccanicistica, le entità ontologiche fondamentali sono elementi materiali, il metodo è riduzionista, la realtà oggettiva è indipendente dal soggetto conoscente, il quale è un osservatore esterno al sistema. Nella concezione info-computazionale, invece, gli elementi ultimi sono immateriali, il metodo è emergentista, l'oggetto e il soggetto sono coinvolti in una circolarità ineludibile e si influenzano reciprocamente.

Quest'ultima proprietà ci pone dinanzi alla questione determinante del rapporto tra filosofia digitale e scienza o, più in generale, tra filosofia (o metafisica) e scienza: l'oggettività del reale è un dato che la scienza dà per scontato e sul quale fondare la propria indagine oppure è un'ipotesi bisognosa, come tutte le altre, di verifica empirica? Che la realtà abbia natura materiale e oggettiva è un punto di partenza (assunzione metafisica) o un punto d'arrivo, cioè l'unica ipotesi sopravvissuta al vaglio selettivo di molte candidate? In altre parole, si ripropone il classico dilemma se a fondare la scienza debba essere un'assunzione filosofica (un "pre-giudizio" filosofico, più o meno consapevole, del ricercatore) o, al contrario, se sia compito della scienza confermare una filosofia; se sia l'intelletto, nell'atto conoscitivo, a dipendere dal materiale fornito dalla sensazione o se, viceversa, siano le percezioni ad essere fondamentalmente influenzate dalle categorie dell'intelletto.

La posizione della filosofia digitale è, dinanzi a tali interrogativi, piuttosto ben delineata. È innegabile che la filosofia digitale stia insinuando in maniera sempre più penetrante taluni dei propri concetti nella scienza tradizionale (il concetto di informazione sopra ogni altro), e consolidi le nuove discipline che sull'informazione già basano i propri sviluppi per costituire in seguito uno strato culturale comune su cui fondare un nuovo tipo di scienza.

4.4 Filosofia digitale ed etica

Poiché ambisce a presentarsi come una concezione compiuta e strutturata, la filosofia digitale non poteva trascurare l'ambito morale. È Wolfram, in particolare, a occuparsene, esprimendo la convinzione che la sua nuova scienza offrirà non solo, sul piano teorico, un inaspettato allargamento delle conoscenze, ma anche, sul piano pratico, un incredibile aumento dei beni materiali e degli strumenti per ottenerli. Il potenziamento tecnologico metterà a disposizione di noi esseri umani una strumentazione capace di costruire pressoché tutto ciò che desideriamo. Potendo fare di tutto, vedremo il campo della nostra libertà

ampliarsi fino ai suoi confini estremi. A quel punto, sottolinea Wolfram, sarà necessario pensare in termini di “evoluzione dei fini umani”. In altri termini, occorrerà non solo impegnarsi a individuare le strategie migliori per conseguire gli obiettivi tradizionali del “bene”, del “giusto” e simili, ma occorrerà anche accettare che questi stessi valori tradizionali siano superati, in seno a un’umanità più libera e quindi più “umana”, da valori eticamente più alti e, al momento, non immaginabili. La previsione ha naturalmente sapore utopico, ma è comunque molto suggestiva. Essa introduce anche nella sfera dell’etica – che da più parti viene considerata, purtroppo, “stagnante” – un elemento di dinamicità, che si traduce nell’invito ad adottare una prospettiva di ampio orizzonte e a vedere nella tecnologia un mezzo di crescita morale e di liberazione. A differenza delle filosofie del Novecento che, in genere, consideravano la tecnologia solo come fonte di alienazione, ansia e spersonalizzazione, la filosofia digitale potrebbe restituire all’umanità la fiducia in sé stessa e nelle proprie scelte, inquadrandole in un robusto riferimento etico per evitare che le innovazioni tecno-scientifiche siano guidate solo dall’inventiva e dall’ambizione dei ricercatori e dalla ricerca del profitto da parte delle aziende.

5. Conclusione

Come è sempre accaduto per le tecnologie importanti, anche il computer ha avuto un effetto deflagrante sulla nostra visione del mondo: da quando esso è comparso, ai nostri occhi tutto è diventato calcolo e informazione e codice. Ma solo ai nostri occhi? Il computer è solo una metafora, per quanto potente e capace di accendere la nostra mente suscitandone l’immagine ineludibile, ma non per questo necessariamente veritiera, di una natura digitale; oppure è il rivelatore attendibile di quella natura? La filosofia digitale risponde che la Natura non solo *si può interpretare* in termini digitali, ma è digitale: come se il calcolatore fosse una sonda che con un’estremità pesca nell’inconscio più vero del cosmo e con l’altra sollecita la nostra mente ad accogliere quelle verità. Questa conclusione, di carattere metafisico, si raccorda bene con la scoperta della fisica quantistica, che ha intaccato la concezione del *continuum* classico. Ma anche se a volte si può avere la sensazione vertiginosa di toccare “con mano” la natura discreta del reale, bisogna sempre diffidarne: la distanza kantiana tra fenomeno e noumeno non è superabile. Forse è proprio vero che tutto è interpretazione: un’interpretazione basata sulla nostra natura psicofisica, sulle caratteristiche della nostra biologia, sulle illusioni dovute alla nostra posizione particolare nella storia evolutiva, alla taratura dei nostri sensi, alla scala dei nostri strumenti, alle nostre esperienze infantili.

Ma nonostante questi scrupoli doverosi, molti sono i pensatori, matematici, informatici, filosofi e ingegneri che nutrono una fede incrollabile nella natura digitale dell’universo e sono disposti a reinterpretare in questa chiave tutte le conquiste della scienza, compresa la parabola dell’evoluzione biologica, che non scaturirebbe più, come nella vulgata neodarwiniana, da un rimbalzare di caso e necessità, bensì dall’inesorabile dispiegamento di una famiglia di automi cellulari. La svolta discreta sarebbe la rivelazione della vera natura del reale: non più un universo soggetto ai capricci della probabilità, che tanto avevano

turbato Laplace, giù giù fino a Boltzmann e Einstein, ma un mondo ordinato e deterministico. Niente più caso, niente più disordine, solo la libertà esercitata nei limiti autoimposti dalla disciplina di una natura autoorganizzata.

Un'ultima considerazione. Poiché l'universo è un sistema complesso (il più complesso che esista), forse non basta una sola descrizione, condotta da un solo punto di vista, ad esaurirne la ricchezza. I sistemi complessi si possono descrivere a molti livelli, da molti punti d'osservazione e con molti linguaggi diversi, e ciascuna di queste descrizioni fornisce un elemento di "verità": quindi affiancare alle due descrizioni note – quella continua della fisica classica e quella probabilistica della meccanica quantistica – la descrizione della filosofia digitale ci permette di avvicinarci a un quadro asintoticamente veritiero.

0

1

0

1

0

Appendice

Obiezioni e risposte

La filosofia digitale è oggetto di alcune obiezioni, suscitate dalla perentorietà con cui essa propone, specie in certe formulazioni, la sua visione del mondo. Questo spirito radicale, alieno da mediazioni e compromessi, sembra voler disorientare il lettore in modo da spingerlo a riesaminare certe questioni di fondo intorno alla realtà, che di solito sono date per scontate, ma che di fatto poggiano su basi incerte. Ecco alcune di queste obiezioni, con le possibili repliche.

1. *L'ipotesi del bit come unità fondamentale della realtà dev'essere ancora verificata scientificamente.*

È vero. La questione è stata affrontata, problematicamente, nel paragrafo 3.3.

2. *L'ipotesi dell'informazione come fondamento della realtà è metafisica e non scienza.*

In parte è vero: i filosofi digitali non hanno difficoltà ad ammettere che la teoria contiene una componente metafisica, ma sostengono che si tratta di assunzioni metafisiche con un alto tasso euristico ed esplicativo nei confronti del mondo fisico. Ogni teoria, del resto, compreso il realismo scientifico, poggia su basi metafisiche.

3. *Per esistere e per propagarsi l'informazione ha sempre bisogno di un supporto materiale o energetico. Quindi non può essere il principio primo*

Su questo punto i filosofi digitali si dividono: alcuni si limitano ad assegnare all'informazione lo stesso statuto di principio primo attribuito nella tradizione fisica alla materia e all'energia, istituendo una sorta di arché trinitaria; mentre altri, più radicali, considerano l'informazione l'unico principio primo, capace di generare, in certe sue configurazioni, la materia stessa. Il problema comunque non è risolto: se l'informazione non può essere ridotta al suo supporto, essa non può neppure prescindere. Si pensi alla computazione in un computer, che opera su entità fisiche ma che ha senso solo in quanto queste entità sono portatrici di informazione. La posizione moderata, trinitaria, sembra più accettabile di quella radicale.

4. *La filosofia digitale pone l'immateriale a fondamento della realtà, quindi è una nuova forma di spiritualismo.*

I filosofi digitali accettano la prima parte dell'affermazione, ma non la seconda. Essi anzi si pregiano di aver individuato una terza dimensione ontologica, che si colloca nel mezzo del tradizionale dualismo di Spirito e Materia. Oggi i fisici parlano di "universo informato" per esprimere la sostanziale interconnessione quantistica inseparabile di tutto l'universo, che sarebbe una rete cosmica di relazioni (e la relazione è un concetto squisitamente informazionale). Inoltre lo sviluppo recente della fisica teorica ha rinsaldato la convinzione che le proprietà della realtà ultima siano rispecchiate nelle strutture della matematica, cui la filosofia digitale

aggiunge la dinamica della computazione. Immaterialità non spirituale, dunque, bensì matematica e infocomputazionale.

5. *Essa contraddice il "rasoio di Occam".*

Tutt'altro, ribattono i filosofi digitali. Il rasoio si applica non all'esuberanza delle forme naturali, bensì alle leggi fondamentali che le producono e che, stando alla visione di Wolfram, si riducono a un piccolo numero di programmi semplici (automi cellulari).

6. *Contraddice il senso comune.*

Che l'universo sia un Grande Computer è in effetti contrario al senso comune. Ma il progresso della scienza è sempre stato all'insegna di un allontanamento progressivo dal senso comune e dalla percezione immediata, altrimenti saremmo ancora fermi alla fede in una Terra piatta e immobile al centro dell'universo.

7. *Implica che il mondo che vediamo è solo una simulazione.*

È un po' la visione di cui si è fatto portatore il film *Matrix*. Ma per la filosofia digitale non viviamo dentro una simulazione, bensì dentro una computazione. Non esiste nulla di cui la realtà sarebbe una simulazione: l'universo non ha modelli da replicare, ma semplicemente esegue il suo programma. Quanto alla simulazione, è proprio la natura computante comune all'universo e al computer che consente di simulare (riprodurre) in quest'ultimo un numero crescente di fenomeni del primo.

Riquadro: Gli automi cellulari e il solitario Life

Un automa cellulare (AC) è una struttura matematica dinamica che può descrivere l'evoluzione dei sistemi complessi discreti in matematica, in fisica e in biologia. Un AC ha come base una griglia suddivisa in celle di estensione (finita) arbitraria. Ciascuna cella può assumere due stati, diciamo bianco e nero. Lo stato iniziale dell'AC è definito dallo stato iniziale delle sue celle. Ad ogni istante di una successione temporale discreta le celle cambiano stato secondo un insieme di regole fissato (che varia a seconda dell'AC considerato). La regola con cui una cella modifica il proprio stato dipende dallo stato delle celle ad essa vicine. La successione degli stati di un AC può avere vari esiti (una configurazione fissa, oscillante, periodica, aperiodica).

Il gioco (solitario) *Life*, di John Conway (1970), il primo AC di larga popolarità, dimostra che alcuni sistemi artificiali manifestano certe caratteristiche degli organismi viventi. All'inizio tutte le caselle della griglia di questo *Gioco della Vita* sono bianche. Poi l'unico giocatore annerisce alcune celle, creando la prima configurazione o pattern. I pattern successivi si ottengono esaminando per ciascuna casella il colore delle otto caselle ad essa adiacenti (le sue vicine) e nell'applicare la regola seguente: se una casella bianca possiede esattamente tre vicine nere, diventa nera; se una casella nera ha due o tre vicine nere, resta nera; in tutti gli altri casi, la casella esaminata diventa o resta bianca. Poi si riapplica la stessa regola alla nuova situazione. La sequenza delle configurazioni dà la sensazione di un movimento continuo. A seconda dello stato iniziale compaiono forme sorprendenti chiamate "aliante", "veliero", "girandola", "semaforo". Molto interessante è l'aliante o *glider*, un oggetto formato da cinque celle nere che si sposta attraverso la scacchiera e che riassume periodicamente la forma iniziale. È da sottolineare che le regole del solitario non prescrivono il movimento delle figure, il quale semplicemente si manifesta. Un *cannone* è una configurazione che emette di continuo alianti.

Certi AC generano forme sempre più complesse a partire da regole semplicissime. La complessità scaturisce dall'applicazione ripetuta delle regole, non dalla loro complessità: dunque non è vero che un fenomeno complesso sia per forza retto da meccanismi soggiacenti altrettanto complessi. Nel suo *A New Kind of Science*, Stephen Wolfram fornisce un'analisi esauriente dei 256 AC unidimensionali binari costituiti da terne di

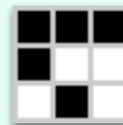


Figura 1
Un aliante di Life

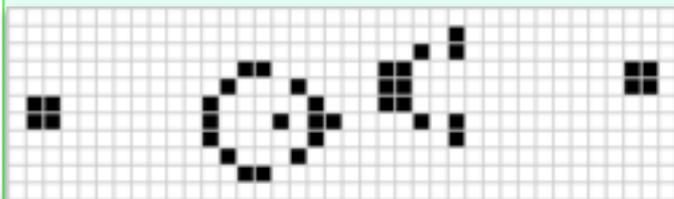


Figura 2
Un cannone spara-alianti

celle disposte in orizzontale. Lo stato della cella centrale dipende dallo stato delle due celle adiacenti: una regola potrebbe stabilire che la cella diventa nera se almeno una delle sue vicine è nera; un'altra che la cella diventa nera se entrambe le sue vicine sono nere.

Poiché per tre celle si hanno in tutto 8 possibili configurazioni binarie, e poiché per ciascuna di queste configurazioni si può scegliere il valore binario della cella centrale, le scelte possibili sono 2 elevato all'ottava potenza, cioè 256, e Wolfram le ha analizzate tutte. A ciascuna di queste scelte, o regole, corrisponde un AC, che può essere designato con un numero progressivo da 0 a 255 scritto in binario, quindi da 00000000 a 11111111. Per esempio la trentesima regola è individuata dal numero 30 che in binario è 11110: in testa si scrivono tre zeri per arrivare alla lunghezza 8. Quindi l'AC n. 30 è descritto dalla regola: se la cella centrale x è bianca e quella alla sua destra parimenti bianca, al passo successivo x diventa del colore della cella alla sua sinistra; in tutti gli altri casi diventa del colore opposto. Le altre 255 regole sono simili.

Figura 3
La regola dell'AC n. 30



Nello schema unidimensionale di tre celle scelto da Wolfram, ad ogni passo scandito dal segnatempo la configurazione iniziale si trasforma in un'altra configurazione (o nella stessa, a seconda della regola applicata), dando luogo a una successione di configurazioni situate una sotto l'altra e lo scorrere del tempo si manifesta come una cascata di configurazioni. Tra gli AC più interessanti si trovano il n. 110, che è in grado di svolgere le funzioni della Macchina di Turing Universale, quindi di effettuare ogni computazione possibile; e l'AC n. 30.

Quest'ultimo fornisce un risultato sorprendente: esso sviluppa forme sempre nuove e anche dopo milioni di passi la sequenza non si ripete mai, generando parvenze di regolarità e situazioni di apparente casualità. A tratti compaiono sullo schermo figure identiche a certe forme naturali: la struttura dei fiocchi di neve, i disegni di una conchiglia, le macchie del manto del leopardo... Se, come sostiene Wolfram, la comparsa sullo schermo, a partire da regole e condizioni iniziali semplicissime, di forme complesse identiche a quelle naturali non è una pura coincidenza, si potrebbe concludere che la struttura degli AC rispecchia la struttura delle forme naturali e che le regole che li governano sono le stesse che governano i processi fisici e biologici del nostro Universo. In breve: la Natura computa.

Figura 4
Dall'alto in basso: le prime 21 configurazioni dell'AC n. 30



Riferimenti

- [1] Chaitin, G. (2010). *Metabiology: Life as Evolving Software*, Course Notes, chapter 3: Is the world built of information? Is Everything software?, Lecture, The Technion, Haifa. <http://www.cs.auckland.ac.nz/~chaitin/metabiology.pdf> (ultimo accesso maggio 2014).
- [2] Chalmers, D. (1995). "Facing up to the Hard Problem of Consciousness", *Journal of Consciousness Studies*, 2, (3), 200-219.
- [3] Dawkins, R. (1986). *The Blind Watchmaker*, Norton; t. i., L'orologiaio cieco. Creazione o evoluzione?, Mondadori, 2003.
- [4] Dodig-Crnkovic, G., Muller, V. (2011). "A Dialogue Concerning Two World Systems: Info-computational vs Mechanistic" in Dodig-Crnkovic, G., Burgin, M. (eds.) *Information and Computation*, World Scientific.
- [5] Feynman, R. (1965). *The Character of Physical Law*, Mit Press; t. i., *La legge fisica*, Boringhieri, 1971.
- [6] Feynman, R. (1982). "Simulating Physics with Computers", *International Journal of Theoretical Physics*, v. 21, nn. 6/7, June, 467-488.
- [7] Fredkin, E. (1990). "Digital Mechanics: An Informational Process Based on Reversible Universal Cellular Automata", *Physica D*, 45, 254-270. http://www.digitalphilosophy.org/dm_paper.htm (ultimo accesso maggio 2014).
- [8] Longo, G. O., Vaccaro, A. (2013). *Bit Bang. La nascita della filosofia digitale*, Apogeo Education.
- [9] Malone, M. S. (2000). "God, Wolfram and Everything Else", *Forbes ASAP*, 27 Nov., 162-180.
- [10] Negroponte, N., (1995). *Being Digital*, Alfred A. Knopf; t. i., *Essere digitali*, Sperling & Kupfer, 2004.
- [11] Toffoli, T. (1982). "Physics and Computation", *International Journal of Theoretical Physics*, v. 21, nn. 3/4, April, 165-175.
- [12] Wheeler J. A., Ford, K. W. (1998). *Geons, Black Holes & Quantum Foam: A Life in Physics*, W. W. Norton & Co.
- [13] Wolfram, S. (2002). *A New Kind of Science*, Wolfram Media Inc.
- [14] Zuse, K. (1982). "The Computing Universe", *International Journal of Theoretical Physics*, v. 21, nn. 6/7, June, 589-600.

Biografie

Giuseppe O. Longo, ingegnere e matematico, è professore emerito di Teoria dell'informazione all'Università di Trieste. Romanziere, drammaturgo, traduttore, divulgatore scientifico e attore, è interessato alla comunicazione in tutte le sue forme.
E-mail: giuseppe.longo41@gmail.com

Andrea Vaccaro, filosofo e teologo, indaga le propaggini più avanzate della filosofia contemporanea (postumano, immortalismo, cyber-filosofie) e su esse ha pubblicato vari studi. E' docente presso l'Istituto Superiore di Scienze religiose di Firenze.
E-mail: andreapaolovaccaro@gmail.com

Usare i Social Media in Applicazioni Predittive

C. Francalanci, P. Giacomazzi, A. Poli

Sommario. *Sui social media vengono ogni giorno espresse opinioni su fatti, marchi, persone, prodotti e servizi di ogni tipo. Secondo la teoria della “wisdom of the crowd”, l’opinione media di un numero sufficientemente elevato di individui rappresenta una buona approssimazione della realtà. Se l’opinione riguarda il futuro, la media delle opinioni può forse essere utilizzata a fini predittivi. La letteratura sul tema e le diverse esperienze fatte in quest’ambito forniscono un interessante quadro delle opportunità e dei limiti dell’uso del parlato sui canali sociali a fini predittivi, mostrando tuttavia molti risultati sorprendentemente positivi.*

Keywords: Predictive analytics, Social media, Twitter

1. Introduzione

Sui social media vengono ogni giorno espresse opinioni su fatti, marchi, persone, prodotti e servizi di ogni tipo. Tali opinioni sono pubblicate in modo spontaneo, rispondendo a un innato bisogno di socializzazione e condivisione del proprio *status*. Diversi canali sociali rispondono a diversi bisogni emotivi o pratici. Ad esempio, Facebook, il canale sociale più diffuso in assoluto, supporta la socializzazione coi propri amici che, in generale, corrispondono a vere relazioni di amicizia nel mondo reale, passate o presenti. Su Twitter gli utenti condividono fatti che ritengono interessanti, rispondendo a un più generale bisogno di informazione. I forum settoriali, ad esempio quelli sulla salute, riuniscono persone accomunate da uno stesso interesse o da uno stesso problema, che è per loro così importante da rappresentare di per sé l’origine della necessità di socializzazione.

Sui canali sociali le opinioni non sono sollecitate da domande fatte da chi ha interesse nel conoscere le risposte. Ad esempio, nel settore turistico le persone condividono esperienze e si danno reciprocamente consigli senza che un operatore turistico abbia loro inviato un questionario per misurare l'indice di gradimento di una destinazione turistica e dei suoi servizi. Le persone parlano dunque di ciò che realmente le appassiona, di quello che ritengono interessante o utile per sé e per altri, di quello che li ha profondamente soddisfatti o chiaramente delusi, con una sincerità spesso facilitata dalla privacy della quale si gode o si può godere nella maggior parte dei canali sociali.

È pur vero che molte delle opinioni espresse sono in realtà di natura pubblicitaria e vengono immesse in rete, a volte in massa, da profili aziendali gestiti con l'obiettivo di sfruttare le capacità relazionali dei canali sociali a fini promozionali. Ad esempio, i profili di leader aziendali, quali Stefano Gabbana (@stefanogabbana) o Valentino Rossi (@ValeYellow46) su Twitter, rappresentano veri e propri canali di comunicazione con i fan, la cui efficacia risiede in parte nella capacità offerta dal canale sociale di relazionarsi con i *follower* in maniera più vivace, meno impersonale, e disinvoltamente innovativa. Lo stile di comunicazione, l'unicità dei contenuti, la velocità con la quale fatti e notizie vengono condivisi, rendono i canali sociali interessanti e inducono alla lettura e alla condivisione. I lettori sono consapevoli che il fine ultimo dei messaggi è spesso promozionale, ma se il contenuto è interessante e originale si lasciano comunque coinvolgere, purché ciò che viene comunicato sia vero e abbia un carattere di autenticità. Negli esempi citati, è chiaro che per Stefano Gabbana comunicare coi fan è parte del mestiere, ma se ciò che viene detto corrisponde alla personalità di Stefano Gabbana dimostrandosi coerente con il suo *stile*, i lettori dimostreranno comunque interesse, in un clima di partecipazione e sincera approvazione.

In sintesi, la gente, cioè la *crowd*, si fida dei canali sociali. La convinzione più o meno espressa, ma largamente condivisa, è che se qualcuno diffondesse un messaggio falso, la massa dei lettori è talmente ampia che arriverebbe la smentita da parte di qualcun altro. Se, viceversa, tutti condividono un particolare contenuto supportandolo, allora quel contenuto è ritenuto *vero*. La letteratura indica due fenomeni come i principali responsabili di questo atteggiamento: la *wisdom of the crowd* e l'assenza sui social media del fenomeno della *sospensione di coscienza* che caratterizza invece i tradizionali canali di marketing.

Secondo la teoria della *wisdom of the crowd*, l'opinione media di un numero sufficientemente elevato di individui rappresenta una buona approssimazione della realtà, in generale migliore di quella che può fornire un singolo esperto. Questa teoria ha avuto origine all'inizio del secolo scorso ed è comunemente associata ad un esperimento condotto dallo statistico Francis Galton nel 1907 [1], ampiamente citato come un interessante aneddoto che ha originato un'ampia letteratura in diversi ambiti, dalla psicologia, all'economia. Nell'esperimento in questione, Galton riporta i risultati di un'indagine empirica condotta a una fiera di paese. Egli mostra come la media di un campione statisticamente significativo di valutazioni del peso di un bue fatte da non esperti era più vicina al peso reale del bue di tutte le stime fatte da esperti prese individualmente. Il suo esperimento è stato largamente discusso e criticato.

Tuttavia, secondo la teoria della *wisdom of the crowd*, la media delle opinioni pubblicate sui canali sociali può costituire un dato di grande interesse e se l'opinione riguarda il futuro, può forse essere utilizzata a fini predittivi.

L'assenza del fenomeno della sospensione di coscienza rafforza ulteriormente l'interesse dell'informazione proveniente dai canali sociali. I media tradizionali sono basati sul paradigma del *broadcasting*, ovvero sono monodirezionali e gli ascoltatori non hanno (quasi) possibilità di replica. Questo ha creato nel tempo il fenomeno della sospensione di coscienza, secondo il quale gli ascoltatori apprezzano un messaggio in broadcasting anche se sanno che non è necessariamente vero. Alcune pubblicità hanno un notevole successo di pubblico, indipendentemente dalle caratteristiche del prodotto o servizio al quale fanno riferimento e con il quale hanno talvolta una relazione molto remota (vedi, ad esempio, la nota pubblicità della Coca Cola [2] o di Mentadent [3]). Queste pubblicità sono talmente piacevoli da guardare che, indipendentemente dalla loro relazione con il prodotto, vera o falsa che sia, inducono il pubblico ad apprezzare la forma senza porla in relazione con la sostanza spesso lontana dalla realtà. In altre parole, il pubblico sospende la propria coscienza, non si concentra sul fatto che il contenuto dello spot pubblicitario è essenzialmente falso (una palese esagerazione della realtà) e apprezza la piacevolezza della storia raccontata.

Questo fenomeno di sospensione di coscienza non vale, in generale, per i social media. Sui canali sociali si pretende che ciò che viene detto sia *vero* e, se non lo è, si utilizza la possibilità di replica lasciando commenti negativi che, nel numero e nella diretta espressione della negatività di opinione sono spesso distruttive dell'immagine del marchio pubblicizzato (vedi [4]). La letteratura riporta molti casi di pubblicità che hanno avuto un buon riscontro in televisione e un pessimo riscontro sui social, dimostrando il rischio di passare da *above the line* (la televisione) a *below the line* (i social).

La fiducia nella veridicità complessiva dell'informazione proveniente dai canali sociali e nella sua capacità di fornire valutazioni medie affidabili rappresenta un presupposto fondamentale delle applicazioni di *predictive analytics*. Se, viceversa, non si ritiene che l'opinione media espressa sui canali sociali rappresenti una buona valutazione della media del fenomeno di interesse, la correlazione fra tale opinione e la futura evoluzione del fenomeno perde il suo fondamento teorico. La robustezza del legame teorico fra opinione media e evoluzione del corrispondente fenomeno di interesse è, a nostro avviso, un fattore determinante del successo o dell'insuccesso dei tentativi pratici di utilizzo dell'informazione sociale a fini predittivi.

2. L'informazione sociale

È naturale domandarsi quali fenomeni sia possibile predire sulla base del parlato sociale. Non esiste una risposta generale a questa domanda, ma è possibile, e relativamente facile, escludere alcuni ambiti. In generale, è possibile utilizzare il parlato sociale a fini predittivi solo se del fenomeno in oggetto si parla a sufficienza. Se invece i volumi sono scarsi, la media delle opinioni non rappresenta un indicatore statisticamente significativo e non può essere ritenuta affidabile, a prescindere da qualunque considerazione teorica.

Quale sia il valore di volumi di parlato necessario dipende poi dalla complessità del fenomeno di interesse. Occorre osservare che i volumi di parlato possono essere elevatissimi. Ad esempio, i volumi di parlato in lingua inglese sul solo canale Twitter sul brand "Nutella" sono circa un milione di tweet al mese, quelli sul brand "Londra" quasi due, i volumi in lingua italiana sul brand "Milano" sono circa 100.000 tweet al mese. Benché Twitter rappresenti il canale con i più elevati volumi di parlato, esistono una miriade di canali sociali concettualmente importanti con volumi singolarmente più bassi, ma complessivamente comunque significativi (ad esempio, TripAdvisor nel settore turistico o Healthboards nel settore salute).

La *wisdom of the crowd* concentra il suo interesse sulle opinioni dei non esperti, ovvero della massa che liberamente condivide il proprio pensiero sui diversi social media. La vasta letteratura sull'ascolto del parlato sociale indica con ampio consenso che le persone tendono a parlare spontaneamente di ciò che li riguarda da vicino e di cui hanno esperienza diretta (vedi, ad esempio, [5]). Essi parleranno, ad esempio, di prodotti e servizi dei quali sono utilizzatori, mentre non parleranno di prodotti e servizi a monte nella filiera produttiva, a meno che questi non siano in qualche modo visibili e sperimentabili. La conseguenza di questo dato di fatto è che il parlato sociale esiste con volumi spesso significativi per i prodotti e servizi classificabili come *B2C* (business to consumer), mentre non esiste o ha volumi molto bassi per prodotti e servizi classificabili come *B2B* (business to business).

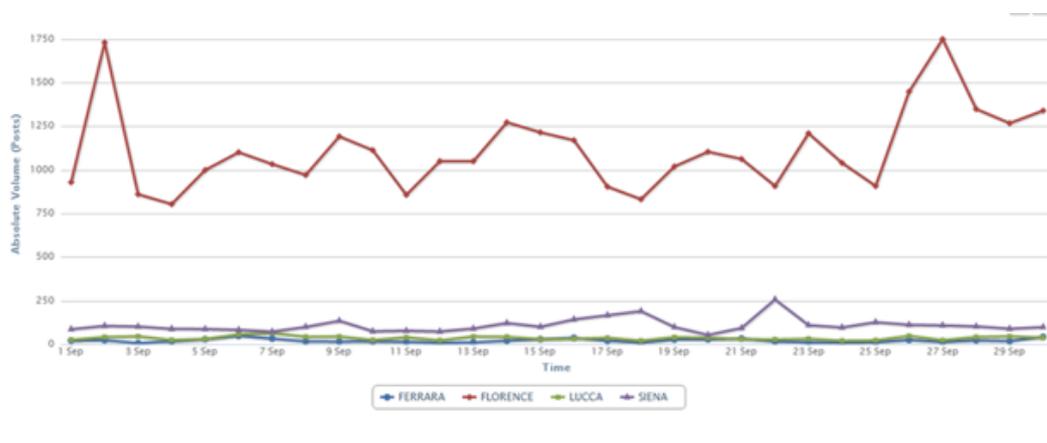


Figura 1
Volumi di parlato su Twitter nel settore turistico (esempio)

Questa suddivisione fra B2C e B2B è criticabile e approssimativa, ma aiuta a farsi un'idea a priori e non crearsi aspettative in ambiti nei quali le persone non forniscono spontaneamente opinioni. È importante notare che anche nel B2B può esserci informazione, ma normalmente non è di tipo sociale. Ad esempio, sugli impianti di refrigerazione industriale è possibile trovare molta informazione sul Web, specialmente sui siti di informazione e istituzionali, mentre è quasi nullo il parlato sociale. L'obiezione secondo la quale su Twitter si può trovare del

parlato su questo argomento non considera la distinzione fra parlato sociale e parlato degli esperti. I tweet di giornalisti che parlano, anche su Twitter, di eventi e notizie che riguardano gli impianti di refrigerazione industriale non possono essere considerati come un campione significativo di opinioni di non-esperti o di generici utenti degli impianti industriali. Viceversa, se esistesse un forum di scambio di opinioni fra persone che lavorano in aziende utenti degli impianti di refrigerazione e se il numero di tali persone fosse sufficientemente elevato, il loro parlato potrebbe costituire il campione di non esperti interessante ai nostri scopi. Tuttavia, tale tipo di forum esiste molto raramente e, se esiste, presenta volumi di parlato che non possono essere considerati statisticamente significativi.

Vogliamo anche osservare che anche nel B2C i volumi possono essere bassi o non essere statisticamente significativi. Ad esempio, il volume di parlato su prodotti elettronici non è sempre elevato. Mentre si parla tanto di smartphone, si parla poco di prodotti di elettronica domestica, quali ad esempio quelli di Bticino. La conseguenza è che, ferma restando la regola generale, i volumi devono essere sempre correttamente valutati e analizzati a priori, con una chiara distinzione fra il parlato degli esperti e quello dei non esperti.

Infine, i volumi possono avere ampia variabilità anche in uno stesso settore. A titolo d'esempio, la Figura 1 riporta i volumi di parlato nel tempo per alcune città italiane, mostrando come essi non siano elevati per tutte le destinazioni turistiche di interesse, ma solo per alcune.

3. Valutazione delle opinioni e sentiment analysis

Una delle principali sfide dell'utilizzo del parlato sociale a fini predittivi è la sua interpretazione. Nella maggior parte dei casi, le opinioni vengono espresse in linguaggio naturale. La difficoltà di interpretazione del linguaggio naturale rende particolarmente preziose le opinioni espresse come voto numerico su un oggetto di interesse, come un hotel o una autovettura, e su alcune sue caratteristiche predefinite, quali la pulizia delle stanze di un hotel o la piacevolezza della vista. I meccanismi di voto sono infatti implementati dalle imprese sui propri siti per ottenere facilmente una valutazione dei propri prodotti e servizi. Le valutazioni numeriche sono facili da interpretare e da aggregare per ottenere valutazioni medie rappresentative dell'opinione di una campione significativo di persone.

Tuttavia, tali valutazioni numeriche aggregate spesso non sono condivise e, quando lo sono, mostrano tutti i limiti dell'approccio a domande chiuse con il quale esse sono ottenute. Gli utenti sono supportati nell'esprimere opinioni su caratteristiche predefinite dei prodotti e servizi delle aziende, che possono non includere gli aspetti dei prodotti e dei servizi sui quali l'utente vorrebbe poter esprimere le proprie opinioni. Questo dato di fatto in alcuni casi è gestito affiancando alla richiesta di valutazioni numeriche la possibilità di esprimere commenti liberi. Ad esempio, TripAdvisor supporta entrambe le modalità e, in effetti, i commenti liberi sugli hotel sono numerosi e contengono indicazioni articolate e precise sull'esperienza di chi li ha scritti. Tali indicazioni sono generalmente ritenute molto utili, ma nella pratica sono poco fruibili a causa del tempo e dello sforzo richiesto per leggerle e farsene una sintesi [6].

Le opinioni espresse in forma testuale rappresentano la gran parte dei contenuti sui social più diffusi, si pensi a Facebook e Twitter. Tali opinioni devono essere analizzate e strutturate per poter essere utilizzate in modelli matematici predittivi che, ovviamente, hanno bisogno di dati numerici in input. La trasformazione dei dati testuali in dati strutturati facilmente traducibili in dati numerici rappresenta a tutt'oggi una sfida.

Per comprendere la complessità del problema, facciamo riferimento a un esempio pratico. Supponiamo di voler effettuare una previsione sui volumi di vendita della *Ford*. L'assunzione di base che possiamo fare è che se la *crowd* parla tanto e positivamente dei vari modelli di auto Ford, allora è probabile che la Ford abbia in futuro un buon fatturato. Per tradurre questa assunzione in un semplice modello predittivo occorrono:

- Una serie storica di dati di fatturato, ad esempio mensile, della Ford.
- Una serie storica dei volumi di parlato sui principali canali sociali, ad esempio i volumi mensili di Tweet che riguardano la Ford.
- La valutazione dell'opinione espressa sulla Ford da ciascun post.

Supponiamo di avere accesso ai dati di fatturato mensile della Ford. Tali dati sono per natura quantitativi e possono essere facilmente utilizzati in un modello quantitativo. I dati sui volumi mensili di Tweet possono essere ottenuti scaricando tramite le API di Tweeter o di un suo rivenditore autorizzato (ad esempio, Gnip) tutti i Tweet che contengono la parola chiave "Ford" nell'intervallo di tempo considerato. Una volta ottenuti i dati grezzi, occorre eliminare i Tweet che pur contenendo la parola chiave "Ford" non fanno in realtà riferimento alla casa produttrice di automobili, ma ad altri oggetti del mondo reale, ad esempio Harrison Ford, l'attore e qualunque altro omonimo di Ford. Questa operazione si chiama *disambiguazione*. Si tratta di un'operazione molto importante e che non può essere tralasciata, poiché i dati non pertinenti possono rappresentare una percentuale molto rilevante del totale dei volumi, spesso superiore al 50%.

Sui dati rimanenti, occorre poi valutare l'opinione espressa. Questa valutazione prende il nome di *sentiment analysis*. L'obiettivo ultimo della sentiment analysis è comprendere se un testo, tipicamente un post condiviso su un canale sociale, riporta un'opinione positiva, negativa o neutra su un oggetto di interesse. Un testo può esprimere una opinione positiva o negativa esplicita sull'oggetto di interesse. Ad esempio, un post che dichiara che "le auto Ford sono inaffidabili" esprime un'opinione chiaramente negativa sulla sicurezza delle Ford. L'opinione però può essere espressa anche in forma più implicita tramite la condivisione di un fatto che riguarda la Ford. Ad esempio, un post che dichiara che "le auto coinvolte in incidenti autostradali sono per il 15% prodotte dalla Ford" fornisce un fatto negativo sulla sicurezza delle auto Ford, pur non esprimendo esplicitamente un'opinione. Considerare questo tipo di post come neutrali sarebbe evidentemente un errore ai fini predittivi. Questa varietà delle modalità linguistiche con cui le opinioni possono essere espresse è una delle difficoltà tecniche principali della sentiment analysis.

Ulteriore difficoltà è poi creata dal fatto che le opinioni sono espresse in maniera articolata, facendo riferimento a specifiche caratteristiche degli oggetti dei quali si parla. Ad esempio, le persone si lamentano se la batteria della loro auto si rompe prima del tempo, oppure se i consumi sono considerevolmente più elevati di quelli nominali, eccetera. Per utilizzare questo dettaglio a fini predittivi, occorre capire se le lamentele, o l'entusiasmo, sono allineate alla media di mercato, oppure rappresentano una oggettiva debolezza del brand in esame che può riflettersi in una diminuzione del fatturato. Inoltre, occorre selezionare le caratteristiche di interesse del prodotto o servizio che si ritiene possano avere un valore predittivo.

Tra l'altro, visti in maniera relativa, i volumi stessi hanno un valore predittivo. Volumi di parlato in diminuzione possono infatti essere un'indicazione di un minor interesse da parte del mercato in un determinato prodotto o servizio. L'attenzione ai volumi da parte delle numerose agenzie di comunicazione è proprio legata alla correlazione e potenziale impatto economico fra l'attenzione ricevuta dai diversi brand sui canali sociali e il loro successo di mercato.



Figura 2
Tag cloud relativa al brand MSC Crociere (marzo 2013, Twitter)

La Figura 2 riporta, a titolo d'esempio, una tag cloud del parlato su Twitter su MSC Crociere. I colori indicano il sentiment con il quale si parla dei diversi concetti, in rosso e viola i tag molto e blandamente negativi, in verde e giallo quelli molto e blandamente positivi. Il sentiment generale su MSC crociere è molto positivo (in verde), ma si notano alcune negatività, come, ad esempio, la negatività (in viola) su «Costa Fortuna» dovuta nello specifico a sporcizia,

cabine brutte, orari di imbarco impossibili (in relazione soprattutto a crociere in Sud America). In figura, i tag neutri sono in grigio, evidenziando una prevalenza di commenti neutrali che, in effetti, rappresenta una caratteristica generale del parlato sociale. Dal punto di vista predittivo, questo si traduce in una percentuale piuttosto piccola dei post con sentiment e una conseguente necessità di utilizzare intervalli temporali ampi nei modelli quantitativi, per poter contare su volumi di sentiment statisticamente significativi.

4. Applicazioni direzionali: la previsione dei volumi di vendita

L'informazione proveniente dai social media e il suo sentiment possono essere ritenuti indicativi dell'orientamento delle future decisioni di acquisto da parte degli utenti. A tal fine, i volumi del parlato ed il sentiment associato possono essere impiegati per predire i volumi di vendita. La disponibilità di previsioni di vendita può consentire di allocare al meglio i fattori produttivi al fine di rispondere alla variazione della domanda; al contempo, la disponibilità di un modello che correla il parlato sociale alle vendite può suggerire strategie di marketing che agendo sulla leva dei social media cerchino di incrementare i volumi di vendita.

Numerose analisi sono state portate avanti nel settore cinematografico, laddove si è cercato di prevedere il successo del lancio dei film sulla base dei commenti e delle recensioni su forum, blog e altri canali sociali, analizzati sia prima che durante e dopo l'uscita nelle sale cinematografiche. L'ampia disponibilità di informazioni sugli incassi ha reso possibili le analisi da parte di numerosi ricercatori. Come risultato di queste analisi vi è stata la proposta di una serie di sistemi predittivi che usano diversi approcci, ad esempio tecniche di regressione lineare, classificatori k-NN, e reti neurali. I modelli predittivi individuati basano i loro risultati talvolta sui volumi del parlato, talvolta sul sentiment, e talvolta su un uso combinato di sentiment e volumi. Non vi è quindi un accordo in merito alla predominanza della significatività dei volumi piuttosto che del sentiment, ma anzi appare che modelli differenti riconoscano un potere predittivo differente a queste due variabili fondamentali.

Un settore particolarmente interessante è quello della telefonia mobile. Numerosi appassionati discutono quotidianamente, con alti volumi di parlato in particolare su forum e blog specializzati, in merito a telefoni cellulari, smartphone, tablet ed altri dispositivi elettronici mobili. Inoltre, sezioni apposite o interi forum sono dedicati alle discussioni inerenti i servizi e le tariffe degli operatori di telefonia mobile. Gli utenti confrontano e commentano le diverse offerte e promozioni ed inoltre si esprimono in merito alla qualità del servizio offerto, riportando i casi di disservizi subiti. Questi messaggi possono essere anticipatori della volontà degli utenti di cambiare operatore mobile, laddove si siano individuate migliori offerte o si stiano riscontrando disservizi eccessivi. Allo stesso tempo, i messaggi sono in grado di influenzare i lettori, i quali si costruiranno un'opinione sulle diverse offerte e sulla qualità dei servizi disponibili, portando a future decisioni di acquisto. Sulla base di queste ipotesi, appare interessante verificare la correlazione delle vendite con il parlato sui forum tematici. In [7], gli autori prendono in considerazione i dati dei cambi di operatore con portabilità del

numero telefonico (MNP, Mobile Number Portability). In un mercato saturo, queste informazioni sono una buona approssimazione delle acquisizioni/perdite di clienti da parte degli operatori. Sono stati quindi considerati i flussi netti di clienti per ogni singolo operatore e al contempo analizzate le discussioni sui forum in merito all'operatore mobile preso in esame. La Figura 3 mostra un esempio di confronto su un intervallo di 16 settimane tra i flussi netti dei clienti in portabilità di un operatore mobile italiano (asse *sales*) ed il sentiment normalizzato dei messaggi, relativo allo stesso operatore, inseriti sul forum del sito web *telefonino.net*.



Figura 3
Relazione fra sentiment e vendite (settore telefonico, 2012)

Come si può notare, la curva delle vendite cresce, con un ritardo di circa una settimana, in seguito ad un miglioramento del sentiment. Ciò mostra che il miglioramento del sentiment è indicatore di un futuro incremento delle vendite. La presenza cioè di messaggi favorevoli ad una offerta, tariffa o servizio è un indice del gradimento, che si traduce a breve nello spostamento dei clienti verso un operatore. La decrescita del sentiment, invece, si accompagna anch'essa ad un calo delle vendite, ma è in alcuni casi anticipatoria e in altri casi posticipata rispetto ai passaggi di operatore. Il disagio dei clienti si manifesta infatti a volte in anticipo rispetto alla loro decisione di abbandono, mentre a volte impiega più tempo per essere osservato.

5. Applicazioni nel settore delle scommesse: il caso del calcio

Il calcio e lo sport in generale sono popolari argomenti di discussione su forum e social network da parte di numerosi appassionati. Una parte di questi appassionati scommette – più o meno regolarmente – denaro sull'esito di singoli incontri sportivi o campionati. Esistono inoltre forum pubblici dedicati alla discussione di strategie di scommessa prendendo in esame singoli incontri sportivi, ad esempio *Mondo Scommesse*, *Pronostitalia*, *Pronostigoal* e il forum di

Bet4Win in Italia, e *Punterslounge* e *OLBG* nel Regno Unito. Mentre su questa categoria di forum sono possibili discussioni articolate e motivate delle diverse strategie di scommessa, su Twitter si trovano tipicamente messaggi relativi alle gare e alle squadre, ma con osservazioni generiche e non finalizzate all'individuazione di strategie di scommessa. Inoltre, mentre sui forum specializzati sono prese in considerazione anche scommesse elaborate – vale a dire diverse dalla semplice individuazione del risultato finale “1X2” – i commenti su Twitter in genere forniscono solamente indicazioni di sentiment a favore o a sfavore di una squadra, senza indagare o fornire indicazioni sull'esito di eventi più articolati, come il numero di goal, i risultati del primo tempo, o il risultato esatto. Attraverso una metodologia di classificazione dei messaggi sui forum e l'analisi empirica dell'impatto della lettura di questi messaggi da parte di uno scommettitore, in [8] si osserva come il giocatore sia indotto ad alzare la propria soglia di rischio, ovvero a decidere di scommettere su eventi a bassa probabilità e alta quota di vincita. Ciò si verifica in quanto gli scommettitori esperti che partecipano ai forum tendono a non considerare gli eventi altamente probabili e poco remunerativi, ma al contrario forniscono, per le medesime gare, motivazioni articolate in merito all'opportunità di scommettere su altri eventi più remunerativi.

L'uso di Twitter finalizzato all'individuazione di una strategia di scommessa si è invece focalizzato sull'individuazione dell'esito finale “1X2”, prendendo in esame un campione di squadre di calcio (Inter, Juventus, Milan e Napoli per il campionato italiano e Arsenal, Chelsea, Manchester City e Manchester United per quello inglese). Sono stati presi in esame i tweet relativi a queste squadre nei periodi precedenti e seguenti le gare. Da ogni tweet, attraverso un classificatore automatico, sono stati estratti i sentiment – positivi, negativi o neutrali – espressi nei confronti delle squadre citate. Prendendo in considerazione i risultati reali delle partite, e confrontati con le aspettative di risultato sulla base di un indicatore di difficoltà delle gare, sono state condotte delle verifiche sulla significatività statistica di una serie di ipotesi che coinvolgono i volumi dei messaggi osservati su Twitter e il sentiment misurato. Si è dapprima indagato quale impatto i risultati effettivi delle gare hanno avuto sui messaggi di Twitter. I risultati hanno mostrato che l'esito di un evento non ha alcun impatto sul sentiment misurato su Twitter, mentre invece è possibile affermare che l'esito degli incontri influenza sensibilmente i volumi, in particolare un evento negativo comporta la diminuzione dei volumi dei tweet relativi alla squadra, ed eventi positivi comportano talvolta variazioni in positivo e talvolta variazioni in negativo dei volumi dei messaggi. In secondo luogo, si è indagato in merito al potere predittivo dei tweet, confrontando i messaggi antecedenti le gare con i risultati effettivi. I risultati mostrano come a seguito di un aumento della percentuale dei tweet di tipo positivo e una contemporanea diminuzione di quelli di carattere negativo, l'esito dell'evento successivo è in genere positivo. Ciò significa che le buone sensazioni delle persone nei confronti di una squadra sono spesso giustificate. Allo stesso modo, anche un aumento di tweet positivi e un piccolo aumento di post negativi porta statisticamente ad un esito positivo per la squadra. Al contrario, si è dimostrato che a seguito di un aumento della percentuale dei tweet negativi e una contemporanea diminuzione di quelli

positivi, si verifica un evento negativo per la squadra. I volumi, ai fini della predizione, non sono invece stati riconosciuti come significativi. Si è osservato quindi che, in talune situazioni, è possibile trarre indicazioni in merito all'esito delle partite analizzando i messaggi che gli appassionati di calcio postano su Twitter, in particolare analizzando le variazioni del sentiment. E' possibile quindi individuare le gare per le quali si verificano le ipotesi qui descritte, al fine di determinare la puntata con esito più probabile. Tuttavia, le quote offerte dai bookmaker già incorporano, al fine di ridurre il rischio d'impresa, il sentiment degli scommettitori, portando ad offrire quote più basse, confrontate alla loro probabilità, per gli eventi che si ritengono possano ricevere maggiori scommesse. Questo fenomeno potrebbe quindi vanificare la maggiore accuratezza delle predizioni calcolate con i social network, abbassando sensibilmente i ritorni economici derivanti dalle scommesse.

Un analogo esperimento è stato condotto da un blogger americano [10], che ha utilizzato il parlato presente su Facebook, Twitter, blog e forum al fine di prevedere i risultati delle partite del campionato di football NFL (National Football League). L'esperimento ha coinvolto le 32 squadre del campionato e per ogni messaggio è stato assegnato, con tool automatici di analisi del testo, un valore di sentiment. Sono stati successivamente individuati alcuni indicatori che, per ogni squadra, calcolano, ad esempio, la soddisfazione degli utenti, la percentuale di conversazioni relative alla squadra rispetto alle conversazioni totali, o la percentuale di conversazioni con sentiment positivo/negativo rispetto alle conversazioni totali. Questi indicatori vengono quindi utilizzati per confrontare le squadre che stanno per sfidarsi, ed attraverso un algoritmo di decisione viene fornito il pronostico della squadra vincente. Questa metodologia ha consentito di prevedere correttamente 137 risultati a fronte di 111 errori durante la stagione 2012/13.

6. Applicazioni in finanza: il caso del trading

L'idea che si possano fare previsioni dei prezzi di borsa ha origine con le prime critiche all'*ipotesi dei mercati efficienti*. Assumendo che il flusso delle notizie non si interrompa mai, che le notizie abbiano un impatto diretto sul prezzo dei titoli, e che le notizie non siano predicibili, ne deriverebbe infatti che le variazioni di prezzo siano casuali e imprevedibili, in quanto i prezzi rifletterebero pienamente tutta l'informazione conosciuta.

Si è fatta strada inoltre l'idea che i consumatori, gli investitori e i manager siano in qualche modo guidati dall'umore della massa, anche detto *social mood*, il quale influenza le loro decisioni. Infatti, gli indizi forniti da altre persone influenzano le nostre opinioni, facendo nascere e propagare una visione condivisa della realtà (vedi, ad esempio, [11]).

L'uso del sentiment per ottenere informazioni su un titolo quotato è un'attività praticata correntemente ai fini dell'investimento azionario e nel mercato delle valute. Sono stati a tal fine sviluppati prodotti di analisi automatica delle notizie che convertono le news finanziarie in indici di sentiment che possono essere utilizzati dagli analisti finanziari per prendere decisioni di investimento. I leader in questo settore sono Ravenpack, Thomson Reuters e Alexandria. Questo

approccio tuttavia ha due principali difficoltà tecniche: in primo luogo l'estrazione del sentiment da un frammento di testo solleva il problema di adottare tecniche efficienti e precise di elaborazione del linguaggio naturale, ed in secondo luogo le variazioni di sentiment devono poter essere ricondotte alle variazioni di prezzo dei titoli, attraverso un modello appositamente definito.

Prima della diffusione dei social network, i lavori finalizzati all'individuazione del potere predittivo si sono focalizzati su gruppi di discussione online e siti di informazione, mentre negli ultimi sei anni si è iniziato ad investigare in merito all'utilizzo dei social network come fonte di *sentiment universale*, da utilizzare per correlare e predire vari indicatori economici. I risultati ottenuti non sono univoci, ma una larga parte dei ricercatori concorda con il fatto che la presenza di volumi di messaggi particolarmente elevati rispetto alla media su un dato forum o social network, si correli significativamente a variazioni elevate del valore dei titoli, ad un aumento della volatilità, e a un aumento del volume degli scambi. Controversa è invece la possibilità di fornire indicazioni in merito alla direzione della variazione del prezzo, e quindi di poter conseguire un vantaggio economico dall'uso dei modelli predittivi, sebbene alcuni studi affermino di aver raggiunto accuratezze nelle predizioni tra il 60 e l'80%, mediante la classificazione con modelli bayesiani, alberi di decisione e altre tecniche di machine learning [12].

Un tool software comunemente usato per l'estrazione del sentiment dalle news è SentiWordNet [13], il quale è liberamente installabile ed utilizzabile gratuitamente. Il tool, che si presta a numerosi utilizzi, in questo ambito è stato spesso impiegato per calcolare il numero di commenti positivi, negativi e neutrali per i titoli azionari di riferimento, sulla base delle parole presenti nel testo. Queste informazioni, assieme ai prezzi di borsa, sono quindi usate per la costruzione dei modelli predittivi, da utilizzare poi per prevedere gli andamenti futuri.

Un importante studio di Bollen et al. [14] ha coinvolto i messaggi del social network *Twitter*, usati per correlare i valori dell'indice americano *Dow Jones Industrial Average* (DJIA). La collezione dei messaggi giornalieri è stata classificata con tool automatici sulla base di 6 differenti dimensioni di umore: *calma*, *allerta*, *sicurezza*, *importanza vitale*, *gentilezza* e *felicità*. Trasversalmente si è identificata la polarità emozionale nelle due classi *positiva* e *negativa*. Come risultato è stata individuata una significativa relazione di causalità tra la dimensione *calma* e i movimenti di prezzo dell'indice DJIA nei 5 giorni successivi. Inoltre, attraverso un modello a reti neurali, la dimensione *calma* è stata usata al fine di predire l'indice DJIA, usando i dati dei precedenti 3 giorni. Il modello ottenuto ha mostrato una capacità predittiva della direzione dell'indice pari a circa l'87%.

Ad oggi esistono imprese specializzate nella sentiment analysis, come *MarketPsych* o *AlphaGenius*, che utilizzano le numerose teorie e tecniche di finanza comportamentale, data mining, machine learning e linguistica computazionale per estrarre conoscenza finanziaria rilevante dalle più possibili fonti, come social network, chat, forum, siti di informazione e siti aziendali. Alcune piattaforme di trading inoltre incorporano tecnologie che forniscono agli investitori la possibilità di creare legami sociali e di replicare le strategie di trading altrui.

Molto ampio è inoltre l'interesse dei gestori di hedge fund, i quali hanno iniziato ad affiancare ai tradizionali indicatori finanziari alcuni indici di sentiment osservato sui social media. A titolo di esempio si cita la pionieristica *Derwent Capital Markets* che nel 2011 ha lanciato sul mercato un fondo speculativo che usava Twitter per prendere decisioni di investimento.

In [9], è descritta una metodologia automatizzata per l'individuazione di modelli di correlazione tra i messaggi su Twitter e alcuni indici finanziari settoriali. In particolare, il lavoro si è concentrato sul settore automobilistico coinvolgendo i post relativi ai costruttori di auto e impiegando un indice finanziario che racchiude i relativi titoli (DJTATO, Dow Jones Automobiles & Parts Titans 30 Index). Il sistema utilizza un tool proprietario per l'estrazione del sentiment e la classificazione in varie dimensioni di analisi di rilevanza economica (ad esempio vendite, assunzioni, licenziamenti, prodotti, fabbriche, acquisizioni, fusioni, e management) ed esplora il comportamento di una serie di modelli predittivi messi a confronto per individuare dinamicamente quelli che mostrano una migliore storia di successo nelle predizioni. Il sistema è stato usato per prendere due decisioni giornaliere di acquisto o vendita dell'indice finanziario, mettendolo a confronto con modelli predittivi che usano la sola autocorrelazione dei prezzi e non prendono in esame il parlato sociale.



Figura 4
Simulazione di investimento basata su informazione sociale

In Figura 4, è mostrata la simulazione di investimento su un intervallo temporale di circa 5 mesi. A fronte di una perdita di circa l'8% dell'indice considerato, la metodologia di investimento, con l'utilizzo dei dati sociali, ha consentito di ottenere un ritorno del 26% con una volatilità del 12%. Questo dato va confrontato con la metodologia applicata in assenza dell'informazione sociale, la

quale ha prodotto un ritorno del 9%, mostrando un significativo valore aggiunto ottenibile dal suo sfruttamento.

Vale la pena ribadire che l'affidabilità delle predizioni non può prescindere dalla presenza di volumi statisticamente significativi del parlato sui social media. Perciò, l'efficacia delle predizioni sarà ricercabile solo in settori di mercato laddove vi sia una larga utenza che fornisce il proprio contributo nei social media, escludendo quindi le imprese che vendono prodotti e servizi classificabili come B2B (business to business).

Conclusioni

I contributi scientifici sull'utilizzo dell'informazione sociale a fini predittivi sono numerosi e distribuiti in diversi ambiti. Questo dimostra non solo una naturale curiosità nei confronti dei nuovi media che tanto appassionano i loro utilizzatori, ma anche un sincero interesse per un fenomeno che potrebbe avere interessanti applicazioni di mercato. Tuttavia, l'evidenza fornita dai lavori di ricerca finora svolti non si può considerare conclusiva e mostra ancora alcuni risultati non completamente coerenti e convincenti. Un'applicazione aziendale estensiva dei modelli e delle tecniche qui discussi richiede l'integrazione fra l'informazione sociale e le altre più consolidate fonti informative aziendali. Un'azienda che analizza con interesse i risultati dei modelli predittivi sociali non può non domandarsi se, in ultima analisi, i dati reali di vendita non rappresentino un indicatore più diretto e affidabile del successo dei prodotti aziendali e delle aspettative di vendita future. E' pur vero che i social media forniscono indicazioni sui *trend* emergenti, i cosiddetti segnali deboli che indicano un fenomeno nascente non ancora riscontrabile dai soli dati oggettivi di vendita. Tuttavia, se e come tali segnali deboli si possono integrare nei più consolidati processi di pianificazione aziendale resta oggetto di interessanti futuri lavori di ricerca.

Ringraziamenti

Gli autori desiderano ringraziare Francesco Merlo, collega e designer di alcune delle infografiche mostrate in questo articolo.

Riferimenti

- [1] Francis Galton (1907). "Vox Populi," *Nature*, 75(1949), 450-451, March.
- [2] <https://www.youtube.com/watch?v=yRqUTA6AegA>, 16 Dicembre 2013.
- [3] <https://www.youtube.com/watch?v=BCSwjYsy0QQ>, 22 Aprile 2013.
- [4] <http://vimeo.com/16143908>, 24 Ottobre 2010.
- [5] L. Bruni, C. Francalanci, P. Giacomazzi, F. Merlo, A. Poli (2013). "The Relationship Among Volumes, Specificity, and Influence of Social Media Information," *International Conference on Information Systems 2013* (full paper), Milano, Dic.
- [6] R. Pagano, P. Cremonesi (2014). "Summarization of restaurant reviews," *Advanced User Interfaces*, Milano.

- [7] L. Rasina (2012). "Un'architettura per la predizione delle vendite basata su dati di sentiment," Politecnico di Milano, Tesi di Laurea Magistrale, Dic.
- [8] C. Consolandi (2012). "Una metodologia decisionale per le scommesse sportive basata su social media," Politecnico di Milano, Tesi di Laurea Magistrale, Dic.
- [9] A. Maggioni, L. Mazzoni (2012). "Design and validation of a forecasting trading system based on Twitter," Politecnico di Milano, Tesi di Laurea Magistrale, Dic.
- [10] Esposito, Jeff. Social Media's NFL Week 1 Picks. [Online] Jeffesposito.com, 8 Settembre 2011. <http://jeffesposito.com/2011/09/08/social-medias-nfl-week-picks/>.
- [11] Nofsinger, J. R. (2005). Social mood and financial economics. *Journal of Behavioral Finance*, 6 (3): 144-160.
- [12] Sehgal, V. and Song, C. (2007). SOPS: Stock prediction using web sentiment. *Proceedings of the 7th IEEE. International Conference on Data Mining Workshops, ICDM Workshops '07*: 21-26.
- [13] Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th Conference on International Language Resources and Evaluation, LREC '10*: 2200-2204.
- [14] Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2 (1): 1-8.

Biografie

Chiara Francalanci è Professore di Sistemi Informativi del Politecnico di Milano, dove si è laureata in Ingegneria Elettronica nell'ottobre 1991. Durante gli studi di dottorato è stata ricercatrice ospite per un periodo di due anni presso Harvard Business School. Ha scritto numerosi articoli sulla progettazione dei sistemi informativi e sul valore economico dell'informatica, svolto attività di consulenza nel settore finanziario e manifatturiero, sia in Italia che negli Stati Uniti, è editor del Journal of Information Technology e senior editor delle AIS Transactions on Enterprise Systems.

E-mail: chiara.francalanci@polimi.it

Paolo Giacomazzi è professore di Multimedia Internet al Politecnico di Milano, dove si è laureato con lode in Ingegneria Elettronica. Ha svolto un periodo di ricerca presso la University of Mississippi, presso il National Center for Wireless Communications. Ha svolto attività di ricerca e consulenza nel settore delle reti di telecomunicazione fisse e mobili. Attualmente, si occupa di sistemi previsionali relativi sia al traffico Internet, sia ad altri domini applicativi fra i quali i social media, la grande distribuzione, e il mass market in generale.

E-mail: giacomaz@elet.polimi.it

Alessandro Poli è nato a Cremona nel 1981. Si è laureato al Politecnico di Milano in Ingegneria Informatica e ha ottenuto il titolo di dottore di ricerca in Ingegneria dell'Informazione dal Dipartimento di Elettronica, Informazione e Bioingegneria del Politecnico di Milano, dove attualmente è impiegato come assegnista di ricerca. I suoi interessi accademici sono rivolti alle reti peer-to-peer video streaming, all'ottimizzazione di infrastrutture ICT, all'analisi semantica dei social media per la reputazione dei brand e la predizione, e all'analisi del mercato delle mobile app.

E-mail: alessandro.poli@polimi.it

Archivi Digitali per Grandi Masse di Dati

Il caso degli archivi televisivi

G. Dimino, M. Montenovo

Sommario. *In una società nella quale la massima parte dell'informazione nasce ormai digitale, l'archiviazione di dati in forma binaria rappresenta un problema di fondamentale importanza e criticità.*

Delle molte preoccupazioni sollevate sulla durata e garanzia di conservazione delle informazioni digitalizzate, il presente articolo si concentra sulle moderne strategie ed architetture dei grandi sistemi di archiviazione indirizzate, in particolare, ad assicurare la persistenza ed il recupero dell'informazione nel tempo. Attraverso tali tecniche si rende decisamente meno critica la scelta dei supporti di memorizzazione ottimali da sempre considerati il costituente di gran lunga più importante delle archiviazioni di lunga durata.

Gli archivi multimediali rappresentano in questo panorama un'applicazione particolarmente importante e la loro evoluzione viene approfondita nel corso del presente articolo.

L'approfondimento delle caratteristiche degli archivi per grandi masse di dati per altre importanti applicazioni quali gli archivi per OTT, gli archivi Cloud e quelli più specificatamente per Big Data vengono rimandati ad altra occasione.

Keywords: Multimedia archives, Storage technologies, SOA

1. Premessa

Il presente articolo analizza le moderne tecnologie ed architetture degli archivi digitali per la conservazione e restituzione di grandi masse di dati (vedi Riquadro 1). In primo luogo se ne mettono in luce i principi generali per poi analizzare, come caso d'uso specifico, quello delle applicazioni audiovisive con una breve panoramica al nuovo archivio *master* RAI in via di realizzazione.

Obiettivo dell'articolo - oltre all'analisi critica del recente stato dell'arte - è anche quello di dare una risposta, per quanto limitata ad alcuni singoli aspetti, alle molte preoccupazioni degli studiosi sulla "memoria digitale" ed in particolare alla criticità conseguente alla sua eccessiva "volatilità".

Se da un lato, infatti, la Società digitale produce un eccesso di informazione (spesso ridondante od inutile od addirittura negativo), non sembra che fino al recente passato essa sia stata in grado di dare una risposta convincente nell'elaborazione di tecniche per la conservazione a lungo termine della nostra "memoria".

Le **cause di criticità** sono chiaramente molteplici: supporti fisici, interoperabilità dei lettori e loro obsolescenza, linguaggi di lettura e standard in continua evoluzione, capacità informatica di interrogazione degli archivi cui si aggiungono, per la stessa volatilità intrinseca del dato digitale, altre cause non tecniche quali facilità di cancellazioni di archivi per motivi politici o terroristici o di presunta opportunità o per errori veri e propri anche operati da un singolo individuo.

Nello specifico, una delle più oggettive e pressanti preoccupazioni è sempre stata quella relativa ai **supporti di memorizzazione** che, anche con nuove frontiere della tecnologia (ad es. il disco ottico) hanno sempre deluso le aspettative dimostrando una vita nel tempo limitata (misurabile in anni o al massimo in qualche decina di anni ma non certamente in secoli). Le nuove tecnologie di archiviazione, come si vedrà nel corso dell'articolo, rendono meno critico questo problema rivoluzionando i criteri seguiti nei classici sistemi di archiviazione e, soprattutto, rendono rara una possibile perdita di informazioni a seguito di un degrado del supporto digitale. Tutto questo avviene a spese, tuttavia, di un **processo continuo di manutenzione** che rappresenta di per sé un vincolo non trascurabile e, allo stato attuale della tecnica, è il tributo che va pagato alla limitatezza dei supporti digitali.

2. Requisiti fondamentali di un archivio conservativo

Un archivio deve rispondere a due requisiti fondamentali: garantire la **conservazione** dei contenuti e favorirne la **fruizione**. Un curatore di archivio ha pertanto il compito di attuare tutte le misure necessarie ad evitare la perdita o la distruzione di ogni singolo materiale audiovisivo ma allo stesso tempo di metterlo a disposizione degli utenti nei modi e nei formati più adatti ad agevolarne il riutilizzo. Queste due esigenze sono spesso in contrasto tra di loro e quindi la progettazione di un archivio richiede scelte attente e ponderate. Vediamone più in dettaglio le problematiche principali.

Due fattori di rischio specifici della conservazione sono **l'invecchiamento dei materiali** che costituiscono il supporto su cui è memorizzata l'informazione e **l'obsolescenza degli apparati** richiesti per la sua riproduzione.

Non esistono supporti immuni agli agenti atmosferici e all'usura dovuta all'utilizzo, anche se certamente alcuni formati sono più robusti di altri. **L'invecchiamento** dei supporti può essere rallentato effettuando la conservazione in luoghi sicuri tenuti in condizioni climatiche ottimali, come viene fatto talvolta per archivi cinematografici, ma questo comporta ovviamente costi estremamente elevati, specie in presenza di collezioni di vaste dimensioni e rende lento e complesso l'accesso ai contenuti. In teoria sarebbe possibile

generare copie di utilizzo separate dalla copia master conservata in teca ma questo rende ancora più complessa ed onerosa l'organizzazione dell'archivio.

L'obsolescenza tecnologica dei riproduttori diventa un fattore determinante di rischio, in quanto la scomparsa dal mercato di riproduttori e delle parti di ricambio relative mina la sostenibilità di un archivio basato sul mantenimento in servizio di apparati vetusti la cui complessità non ne permette una facile riproduzione al di fuori delle logiche di mercato, anche solo per i fini di conservazione. Questo problema non può che acuirsi col passare del tempo in quanto i sistemi di archiviazione moderna diventano sempre più complessi mentre i cicli di vita dei prodotti su cui si basano si riducono continuamente.

Per tutte le ragioni dette in precedenza, l'adozione di formati e tecnologie standardizzati assume un'importanza sempre crescente: la longevità di standard di mercato consolidati è generalmente superiore a quella di formati proprietari e al contempo si può contare su documenti pubblici che li descrivano in dettaglio e ne permettano, in casi estremi, una completa re-implementazione.

Per **evitare la perdita di contenuti** dovuta all'invecchiamento dei supporti e delle tecnologie, la strategia normalmente impiegata consiste nella migrazione periodica dei contenuti verso formati e supporti di memorizzazione allo stato dell'arte. Ciò tuttavia non è sempre possibile senza perdita di parte dell'informazione. A tale riguardo, ad esempio, i **supporti audiovisivi** analogici sono quelli più critici, in quanto la migrazione comporta un riversamento del segnale elettrico con conseguente riduzione del rapporto segnale-rumore ed è per questa ragione che, ormai la migrazione avviene sempre verso formati digitali. Il cambio di paradigma che sta avvenendo, oggetto principale del presente articolo, è il passaggio da supporti video dedicati (videocassette) a supporti generici di memorizzazione informatica, perdendo quindi la corrispondenza biunivoca supporto-contenuto.

I responsabili della conservazione si trovano spesso a dover effettuare scelte strategiche complesse relativamente all'allocazione delle risorse finanziarie ed umane disponibili, sempre limitate, nel tentativo di trovare il bilanciamento ottimale tra la tendenza a conservare tutti i contenuti senza distinzione, col rischio di non aver risorse sufficienti a mettere in atto misure efficaci, e la capacità pratica di gestire in modo appropriato solo un sottoinsieme dell'archivio, privilegiando così la salvaguardia del materiale di maggior pregio a scapito di altro ritenuto meno importante.

Nel caso particolare degli archivi audiovisivi, la fruizione dei contenuti non comporta problematiche meno importanti, come già analizzato in [1]. Un archivio non può prescindere dal mantenimento di un catalogo dettagliato e consistente di quanto immagazzinato, o tutti gli sforzi di salvaguardia dei supporti risulteranno vanificati dall'impossibilità di accedere a contenuti specifici o semplicemente a conoscerne l'esistenza.

Ma l'individuazione dei contenuti di interesse non è che il primo passo, a cui segue l'accesso al contenuto audiovisivo vero proprio. Qui entrano in gioco parametri prestazionali che determinano in gran parte il grado di complessità tecnologica richiesta all'infrastruttura d'archivio, ovvero il tempo necessario a

soddisfare una richiesta di accesso, il numero di utenti contemporanei che possono effettuare richieste o i formati in cui il contenuto audiovisivo può essere consegnato all'utente. Nel caso di archivi tradizionali, basati sulla conservazione con supporti analogici, queste operazioni richiedono tempi dell'ordine di ore o addirittura giorni, mentre per i sistemi più moderni basati su tecnologie informatiche - dove il video è immagazzinato sotto forma di file - l'accesso può avvenire con tempistiche dell'ordine dei secondi tramite un qualunque dispositivo connesso su rete dati.

3. Dalla conservazione del supporto alla conservazione del contenuto

La trasposizione in formato digitale di contenuti originariamente in formato analogico, cioè la trasformazione dei contenuti in *file*, ha come principale conseguenza lo **spostamento del focus dalla cura del supporto alla cura del contenuto**. Soddisfatto il prerequisito di disporre di un file, ossia di un oggetto informatico, è infatti possibile sfruttare tutto l'insieme delle tecniche e degli accorgimenti che la tecnologia informatica mette a disposizione per evitare la perdita di informazioni, strategie ovviamente valide anche nel caso di informazioni native digitali.

In particolare due sono le principali **criticità dei supporti** per gli archivi:

- il degrado nel tempo
- la rottura o il guasto

La **prima criticità**, il degrado nel tempo, è correlata sia al deterioramento del supporto propriamente detto sia alla grande velocità con cui l'informatica evolve che porta, come conseguenza nel lungo periodo, alla difficoltà di mantenere mezzi elaborativi (apparati e software) con caratteristiche coerenti con una corretta lettura del contenuto memorizzato.

Poiché un file può essere copiato infinite volte senza perdita di informazione e di qualità, lo sforzo viene rivolto non più alla conservazione dei supporti (con un orizzonte temporale dell'ordine delle decine di anni) quanto piuttosto ad una migrazione "continua" dei contenuti da supporti vicini alla obsolescenza a supporti di nuova generazione.

Questo approccio che oggi può sembrare ovvio ha richiesto parecchio tempo prima di essere accettato fra gli archivisti, una comunità tradizionalmente incline ad effettuare scelte conservative, ma ha avuto l'enorme vantaggio di consentire l'evoluzione di supporti, apparati e software, in una logica sistemica bilanciando i maggiori costi introdotti dai processi di copiatura con i vantaggi tecnici ed economici derivanti dall'impiego degli apparati più moderni con decisivo vantaggio sul "*Total Cost*" di un archivio.

La migrazione dei contenuti dai supporti obsoleti - o prossimi al fine vita (stimato statisticamente) - verso supporti di più recente generazione diventa quindi il meccanismo principale attraverso cui si attua la conservazione realizzata attraverso un processo di manutenzione continua integrato nell'architettura dello stesso archivio (e non evento straordinario).

La capacità di stimare quando si renda opportuno attivare il processo di migrazione è un fattore chiave. A tale proposito si possono identificare due criteri decisionali principali: lo **stato di salute di ciascun supporto** e **l'obsolescenza tecnologica** di una classe di supporti, dei relativi apparati e del software di gestione. Il primo criterio comporta parecchie criticità legate alla precisione con cui si riesce ad effettuare diagnosi precoci sull'effettivo stato di salute dei supporti. Il secondo criterio invece richiede scelte a tavolino basate principalmente su fattori economici e valutazioni sul già menzionato "Total Cost". Infatti, la manutenzione con apparati tecnologicamente allo stato dell'arte è significativamente meno costosa rispetto ad apparati obsoleti, permette l'impiego di supporti a più alta densità di memorizzazione (e quindi con minori unità da monitorare), ha prestazioni superiori, garantisce maggiore compatibilità con sistemi più recenti ed evita la necessità di provvedere ad effettuare scorte di parti di ricambio prima che diventino irreperibili sul mercato.

La **seconda criticità di base dei supporti** è legata alla loro **rottura o guasto**. Le tradizionali strategie impiegate in campo informatico per la protezione dei dati nel caso degli archivi sono diventate via via meno praticabili a causa della crescita della densità di memorizzazione e del conseguente allungamento dei tempi di ripristino aprendo la strada a nuove strategie meno limitanti. La più banale consiste ovviamente nella duplicazione dei supporti *on-line* ed è di fatto utilizzata laddove il costo della duplicazione rappresenti una entità accettabile. Tuttavia esistono molti casi dove è necessaria una maggiore accuratezza della scelta. Grazie alle reciproche convergenze tra informatica e telecomunicazioni, anche a seguito di analoghe tecniche sperimentate con grande successo nelle telecomunicazioni negli anni 80 per la trasmissione dati su mezzi trasmissivi rumorosi, sono stati avviati studi e depositati brevetti che hanno dato il via a prodotti di mercato. Ci si riferisce in particolare alla tecnologia di memorizzazione cosiddetta Raid (vedi Riquadro 2).

Al di là della complessità della tecnologia stessa e delle sue numerose varianti, di cui solo alcune hanno avuto successo di mercato, l'idea base per far fronte a rotture o guasti senza duplicazione dei supporti consiste nel memorizzare l'informazione segmentandola su diversi supporti. In caso di guasto di uno dei supporti l'informazione può essere ricostruita utilizzando opportuni algoritmi di codici correttori. E' chiaro che tale tecnica consente di limitare i costi aggiuntivi rispetto ad una totale duplicazione dei supporti.

4. Il modello OAIS per archivi dati digitali

Il *Consultative Committee for Space Data Systems* (CCSDS), organizzazione che comprende le principali agenzie aerospaziali mondiali, ha definito un modello di riferimento per l'organizzazione di archivi di dati digitali finalizzati alla conservazione a lungo termine, denominato **Open Archival Information System (OAIS)** [2]. Il modello OAIS è stato concepito per essere il più generico possibile e fornisce una serie di "recommended practice" utili alla progettazione di sistemi di archiviazione di grandi masse di dati, con particolare enfasi alle problematiche della conservazione a lungo termine.

Questo modello ha validità generale e si applica anche perfettamente alle esigenze degli archivi “master” per applicazioni audiovisive approfonditi nel seguito.

L'elemento principale del modello è l'**Information Package (IP)**, ovvero il contenitore di ciascun contenuto oggetto di archiviazione. Un IP è costituito da due tipi di informazioni chiamate *Content Information (CI)* e **Preservation Description Information (PDI)**. Le CI sono costituite dalla sequenza binaria che rappresenta il contenuto da archiviare vero e proprio, mentre le PDI sono tutte le informazioni (metadati) necessarie alla interpretazione e utilizzo del contenuto, come schematizzato in Figura 1.

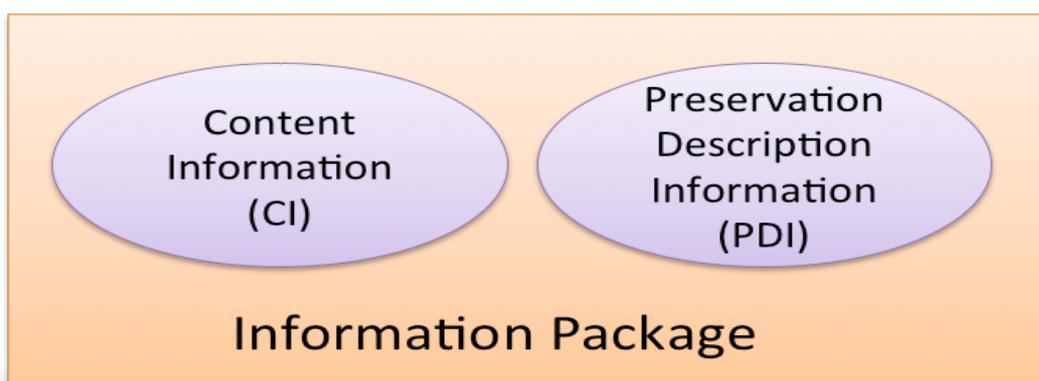


Figura 1
“Information Package” nel modello OAIS

Questo principio è fondamentale: non ha senso archiviare contenuti digitali senza corredarli di tutti i metadati che si possano rendere necessari alla loro corretta fruizione. Le informazioni PDI vengono classificate in cinque categorie:

- 1. Provenienza**
Metadati che descrivono la sorgente da cui è stato derivato il contenuto, ad esempio un nastro video analogico oppure un altro contenuto digitale sottoposto ad un certo processamento;
- 2. Contesto**
Informazioni relative a come il contenuto in oggetto è correlato ad altri contenuti esterni, come la ragione per cui è stato generato, da chi, in quale occasione, ecc ...;
- 3. Identificazione**
Uno o più codici univoci che permettono l'identificazione certa del contenuto all'interno del dominio di riferimento (ad es. il codice ISBN nel dominio librario);
- 4. Fissità**
Metadati che permettono di verificare l'integrità del contenuto, ad esempio una firma digitale o un Codice a Ridondanza Ciclica (CRC);

5. Diritti di Accesso

Informazioni relative al proprietario del copyright e ai diritti di accesso a disposizione del proprietario del contenuto, ad esempio, nel caso in cui il contenuto sia costituito dalla copia di un film commerciale, chi detiene il copyright sull'opera e quali diritti di uso sono a disposizione dell'archivio.

Il modello funzionale su cui si basa l'operatività dell'archivio è illustrato nella Figura 2.

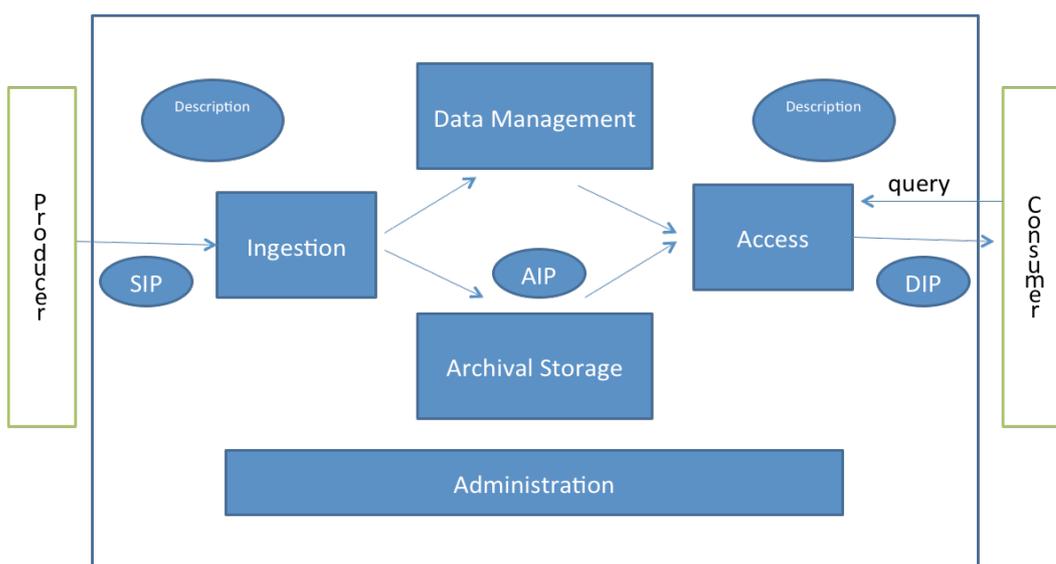


Figura 2
Modello funzionale OAIS

Gli attori coinvolti sono il **Producer**, utente che chiede l'archiviazione di un contenuto nuovo, il **Consumer**, utente che utilizza qualche contenuto dell'archivio e il **Manager**, gestore dell'archivio. Quando un *Producer* chiede l'archiviazione di un nuovo contenuto, fornito all'archivio sotto forma di **Submission Information Package (SIP)**, viene attivato il processo di **Ingestion**, durante il quale il SIP viene trasformato nell'**Archival Information Package (AIP)**, l'oggetto corrispondente nel formato adatto all'archiviazione.

Il processo include varie fasi tra cui la verifica di integrità dei dati, la verifica della presenza dei metadati essenziali e la loro eventuale integrazione, il controllo di qualità del contenuto audiovisivo e se necessario la sua transcodifica nel formato file di archiviazione.

L'AIP viene quindi memorizzato permanentemente nell'**Archival Storage** in Figura 2. I metadati, oltre che inseriti nell'AIP, vengono anche inviati al sistema di **Data Management**, un database tramite il quale avviene la ricerca dei contenuti da parte dei *Consumer*. Il sistema di *Data Management* serve anche per le operazioni di gestione dell'archivio da parte del *Manager*, come la verifica

periodica dell'integrità dei dati contenuti nello *storage* e la gestione del ciclo di vita del contenuto. In seguito ad una ricerca nel sistema di *Data Management*, reso disponibile tramite un portale di accesso (*Access* in Figura 2) e verificati i diritti di accesso al materiale selezionato, l'Utente (*Consumer*) riceve i contenuti voluti sotto forma di **Dissemination Information Package (DIP)**, una versione dell'AIP appositamente creata in funzione delle necessità del *Consumer* e dei diritti di accesso posseduti. Il DIP differisce generalmente dall'AIP in quanto alcuni metadati potrebbero essere omessi, perché non essenziali alla fruizione o protetti, e il contenuto, a seconda delle politiche di accesso e delle caratteristiche del dispositivo di fruizione, potrebbe essere fornito in un formato a qualità ridotta rispetto al master presente in archivio.

5. “Archival Storage”: i nuovi supporti di memorizzazione e lo storage gerarchico

Uno dei vantaggi della informatizzazione degli archivi risiede nella possibilità di immagazzinare i contenuti su dispositivi di memorizzazione di uso generale, che essendo inseriti in un mercato vasto e in continua espansione, sono permanentemente oggetto di ricerca e ottimizzazione da parte di molte università e industrie. Le prestazioni dei dispositivi di memorizzazione informatici, sia in termini capacitivi che prestazionali, crescono in modo sostenuto mentre i prezzi puntano costantemente al ribasso.

Il costo dello *storage* in un archivio conservativo rappresenta una delle principali voci di spesa, quindi molta attenzione viene data alla scelta delle tecnologie più opportune. Possiamo classificare i dispositivi di memorizzazione secondo due macro categorie: **dispositivi prestazionali** e **dispositivi capacitivi**.

I **dispositivi prestazionali** vengono progettati per fornire prestazioni di punta, sia in termini di velocità di lettura e scrittura sequenziale (parametro indicativo del tempo richiesto a soddisfare una richiesta) che di numero di operazioni di lettura o scrittura casuali al secondo (parametro correlato al numero di richieste contemporanee che il dispositivo può soddisfare). La capacità di immagazzinamento dati del dispositivo viene considerata una caratteristica secondaria.

Tra i dispositivi prestazionali i più performanti sono i **Solid State Disk (SSD)**, memorie allo stato solido senza parti in movimento. Le prestazioni offerte dagli SSD sono nettamente superiori a quelle degli *Hard Disk* meccanici tradizionali (HD), ma i costi sono ancora relativamente elevati. Non avendo parti meccaniche sono meno soggetti a guasti dovuti all'usura o a urti rispetto agli HD ma soffrono di problemi legati al numero limitato di riscritture possibili su ciascuna cella di memoria. Trattandosi di una tecnologia relativamente giovane (almeno rispetto agli *hard disk* meccanici) il margine di sviluppo è elevato e parecchi analisti prevedono che soppianteranno gli HD nella maggior parte delle applicazioni nel giro di pochi anni. Una seconda classe di dispositivi prestazionali è quella degli **HD a 10.000 e 15.000 giri al minuto**. Questi dispositivi sia in termini prestazionali che di costo si collocano a un livello intermedio tra gli SSD e gli HD capacitivi.

Nella categoria dei **dispositivi capacitivi** rientrano i dispositivi progettati per ottenere la massima capacità di immagazzinamento, anche a scapito della velocità di accesso. Fra questi rientrano **gli HD con fattori di forma da 3,5' e da 2,5'** e velocità di rotazione di 5400 e 7200 giri al minuto e i **nastri dati**. Gli HD coniugano elevate capacità di immagazzinamento a costi bassi, grazie alla loro ampia diffusione nel mondo informatico.

I nastri dati, tra cui quelli della **famiglia LTO** sono i più diffusi, comportano tempi di accesso notevolmente più lenti, anche se inseriti in librerie robotizzate, ma presentano il vantaggio rispetto agli HD di poter essere tenuti fuori linea su scaffale, al riparo da shock elettrici (ad es. uno sbalzo della tensione di alimentazione) e quindi offrono la longevità massima. Offrono anche il costo di immagazzinamento più basso.

La Tabella 1 presenta un confronto delle varie classi di dispositivi.

Classe	Nastri LTO 6	HD capacitivi	HD prestazioni	SSD
Velocità lettura sequenziale	160 Mb/s	130 Mb/s	200 Mb/s	500 Mb/s
Velocità scrittura sequenziale	160 Mb/s	130 Mb/s	200 Mb/s	500 Mb/s
Velocità lettura casuale	-	120 io/s*	300 io/s	75000 io/s
Velocità scrittura casuale	-	160 io/s	360 io/s	36000 io/s
Capacità	2,5Tb	6Tb	1 Tb	1Tb
Costo per Gb**	€ 0,02***	€ 0,04	€ 0,18	€ 0,50

Tabella 1
Parametri tipici

Un archivio conservativo spesso richiede uno spazio di memorizzazione (*storage*) dell'ordine di parecchi PetaByte (1 PetaByte = 1024 TeraByte), che quindi sarà costituito da diverse migliaia di dispositivi. Per sfruttare in modo ottimale le caratteristiche delle diverse classi di dispositivi e contenere i costi di realizzazione dell'infrastruttura, lo *storage* è organizzato in maniera gerarchica,

* Numero di operazioni di lettura/scrittura al secondo

** Valori indicativi a Giugno 2014

*** Non include il costo del lettore

ovvero i contenuti ad accesso più frequente verranno posizionati su dispositivi più veloci (e costosi) mentre la gran parte del materiale sarà disponibile solo attraverso dispositivi capacitivi ad accesso più lento. Copie di *backup* e di *disaster recovery* possono invece essere stoccate su nastri inseriti in librerie automatizzate o addirittura su scaffale.

Tipicamente un sistema di *storage* è suddiviso in tre livelli logici: **On-line**, **Near-on-Line**, **Off-line**.

Il **livello On-line**, costituito da dischi, e' a sua volta suddiviso in due sottolivelli: il primo, basato su dispositivi prestazionali, e' specializzato nel soddisfare il grosso delle richieste di accesso - che si suppone riguardino una parte limitata dell'intero archivio - ed eventualmente l'elaborazione/adattamento dei media per il sistema richiedente. Il secondo, invece, basato su dispositivi capacitivi mantiene la rimanente parte dei contenuti che si ritiene debbano essere sempre disponibili in linea.

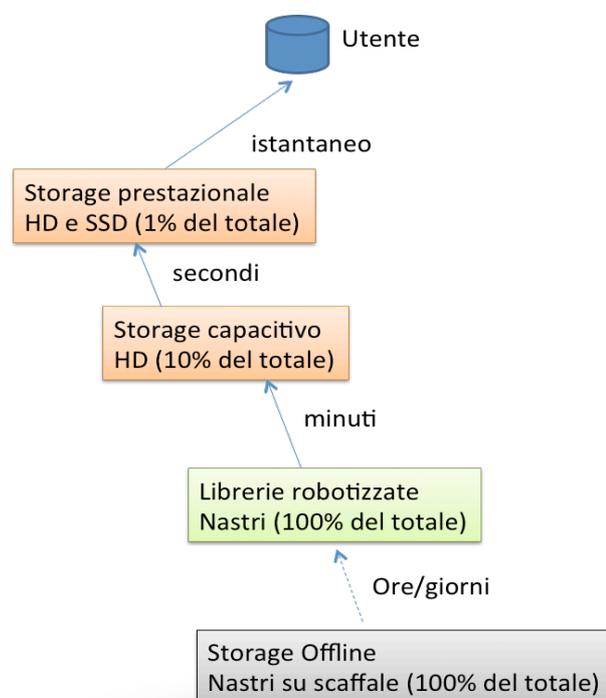


Figura 3
Esempio di sistema gerarchico di archiviazione

Il **secondo livello Near-on-Line** contiene l'intera collezione di contenuti su nastri gestiti all'interno di librerie robotizzate, che ne permettono l'accesso nel giro di pochi minuti, in funzione del carico del sistema e del numero di lettori installati all'interno delle librerie. Quando viene richiesto un contenuto non presente sul sistema *On-line* viene generata una richiesta al sistema *Near-on-Line* che provvede a copiare il contenuto sul sistema *On-line*, da cui è reso accessibile all'utente.

Il **terzo livello Off-line** e', di fatto, costituito dal deposito (eventualmente automatizzato) di una **copia di sicurezza** dei nastri contenuti nelle librerie, per il cui ripristino occorrono tempi lunghi, anche giorni. Viene utilizzato solo in caso di necessità in seguito a cancellazione accidentale della copia in libreria (*disaster recovery*). In Figura 3 viene schematizzata una configurazione tipica di **storage gerarchico**, indicando anche una stima dei tempi di accesso attesi da ciascun livello.

6. Criteri per la selezione dei formati di memorizzazione digitale per l'Archival Storage

La Library of Congress Americana ha identificato una serie di fattori di sostenibilità per valutare l'idoneità di un **formato di memorizzazione** digitale ad essere impiegato in un archivio, validi indipendentemente dalla categoria di contenuti archiviati [3].

- **Divulgazione** della specifica tecnica che permette di interpretare il formato e accessibilità a strumenti *software* e *hardware* per la verifica dell'integrità del contenuto;
- **Adozione** del formato da parte di un elevato numero di comunità di riferimento;
- **Trasparenza** della rappresentazione digitale, ovvero semplicità di interpretazione del formato tramite ispezione diretta (es, un testo in formato ASCII è più *trasparente* di un testo criptato);
- **Auto-documentazione**. L'inclusione nel formato dei metadati necessari alla sua interpretazione, senza dover dipendere da informazioni immagazzinate altrove (ad es. in un database);
- **Dipendenze esterne**. La necessità di disporre di hardware o software particolari per poter riprodurre il contenuto ne aumenta il rischio di compatibilità con le piattaforme future;
- **Impatto dei brevetti**. Le questioni legali connesse all'utilizzo di brevetti associati alla tecnologia sottostante al formato possono ostacolare lo sviluppo di strumenti di decodifica e utilizzo dei formati da parte di terze parti;
- **Meccanismi di protezione tecnica** come crittografia, protezione anticopia, o associazione del contenuto ad un particolare dispositivo fisico possono rendere problematica la generazione di copie di sicurezza.

Questi fattori di sostenibilità non vanno intesi come prescrizioni ma piuttosto come elementi per valutare il rischio di perdita di informazioni associato alla scelta di un determinato formato di memorizzazione. Sarà compito del manager dell'archivio scegliere la combinazione rischi/benefici più appropriata alle finalità specifiche dell'archivio.

Ad esempio, nel caso degli archivi audiovisivi, il segnale audiovisivo codificato propriamente detto viene incapsulato in un opportuno contenitore - comunemente denominato *file wrapper* - che contiene anche tutti i metadati essenziali al suo utilizzo. I criteri e le specifiche sono indicati in Appendice.

7. Archival Storage: architetture moderne per l'archiviazione multimediale

Le caratteristiche principali di un sistema di archiviazione multimediale non possono prescindere dai seguenti principi:

- Contenuti sempre in linea
- Gestione automatizzata
- Modularizzazione dei componenti
- Processo di migrazione integrato nel sistema

Il costo decrescente dei dispositivi di *storage* ha reso economicamente conveniente la realizzazione di archivi di dimensioni elevate (dell'ordine delle centinaia di migliaia di ore) dove tutti i contenuti siano accessibili in modo automatizzato (senza l'intervento umano per la movimentazione di supporti) in tempi dell'ordine dei secondi (*storage On-line*) o al più dei minuti (*storage Near-on-Line*). Questa caratteristica, oltre che a migliorare sensibilmente i tempi di reazione dell'archivio alle richieste degli utenti rispetto agli archivi tradizionali basati sull'archiviazione di supporti audiovisivi, permette una gestione sofisticata delle risorse tarabile dinamicamente in funzione delle condizioni di utilizzo dell'archivio. Le architetture gerarchiche permettono di ottimizzare i costi dello *storage* tramite l'organizzazione su diversi livelli prestazionali. La collocazione dei contenuti su ciascun livello non deve essere effettuata manualmente ma il sistema, in base a una serie di regole impostate, replica autonomamente i contenuti di maggiore utilizzo su memorie ad accesso più rapido. Poiché i dispositivi di *storage* sono soggetti a guasti, dovuti a usura, invecchiamento o cause esterne, i contenuti devono essere sempre archiviati in copie multiple. Il minimo è costituito da una copia nel sistema *Near-on-Line* e una copia nel sistema *Off-line*, ma spesso, dato il basso costo dei nastri, si preferisce avere anche una ridondanza completa del sistema *Near-on-Line* o tramite un secondo sistema gemello o addirittura sul sistema Online. Durante i tempi morti, viene effettuato un monitoraggio continuo dello stato di salute di ciascun supporto, tramite opportuni software diagnostici, in modo da rigenerare automaticamente le copie dei contenuti che fossero divenute indisponibili a causa di guasti o che si trovino su dispositivi considerati ad alto rischio di guasto.

Un archivio deve anche avere risorse di calcolo in grado di effettuare operazioni come cambio formato e controllo di qualità sui contenuti. Queste operazioni si rendono necessarie sia in fase di *Ingestion*, quando nuovi contenuti vengono inseriti in archivio sia quando i contenuti vengono forniti agli utenti.

Le tecnologie informatiche sono in costante evoluzione e la progettazione di un archivio risente necessariamente delle possibilità tecniche offerte dalla tecnologia. Per fare un esempio, le prestazioni oggi offerte dagli SSD erano impensabili solo 10 anni fa e avere l'intero archivio disponibile su tecnologie *On-line* comportava costi proibitivi. La capienza di un HD nel giro di 10 anni è aumentata di circa 10 volte e così pure quella dei nastri LTO. L'aumento della capienza dei dispositivi fa sì che il numero totale di elementi che devono essere aggregati si riduca proporzionalmente, provocando una generale

semplificazione dell'architettura dell'intero sottosistema. Di conseguenza un sistema deve essere progettato in chiave evolutiva e consentire un adeguamento tecnologico dei suoi componenti durante il suo ciclo di vita. In pratica è necessario suddividere il sistema in moduli indipendenti, in modo che l'adeguamento di un modulo non impatti sull'operatività degli altri. Inoltre aspetto molto importante è la virtualizzazione delle interfacce che consente di rendere i vari moduli indipendenti dalle tecnologie sottostanti; il comportamento di un modulo deve essere, dal punto di vista esterno, funzionalmente identico e indipendente dal cambio della tecnologia. La sostituzione di supporti di *storage* obsoleti con altri più moderni non genera in questo modo ripercussioni onerose sulle altre componenti del sistema.

Lo schema di principio di un archivio multimediale si adatta bene al modello di riferimento OAIS descritto nel §4. Il cuore del sistema è costituito dal **Media Asset Management (MAM)**, il componente che gestisce il ciclo di vita dei contenuti e sottende a tutti i flussi di lavorazione. Al **MAM** si interfacciano i moduli di **Storage**, che contengono i file audiovisivi, l'**Ingestion**, che si occupa della normalizzazione dei nuovi contenuti archiviati, l'**Access**, che permette il reperimento dei contenuti da parte degli utenti ed il **Catalogue**, sistema di ricerca e visione in bassa qualità (Figura 4).



Figura 4
Scomposizione logica di un sistema di archiviazione

La metodologia di progettazione denominata **Service Oriented Architecture (SOA)** rappresenta lo stato dell'arte nella integrazione dei sistemi informatici. Sviluppata per risolvere le problematiche di interoperabilità nei sistemi gestionali, si è oggi affermata anche nel mondo dei *Media* [5]. La metodologia SOA è stata sviluppata per guidare i progettisti nella realizzazione di sistemi snelli che siano semplici da gestire e riconfigurare, a prescindere dalla loro

complessità intrinseca. Grande attenzione viene dedicata alla scomposizione del sistema in unità funzionali indipendenti con alto potenziale di riusabilità. In altre parole SOA spinge alla individuazione e fattorizzazione di moduli e funzioni che ricorrono in diverse parti del sistema. Questi moduli vengono detti *Service*, e costituiscono i mattoncini tramite i quali l'intero sistema viene realizzato.

Un modulo **Service** deve esporre un'interfaccia pubblica ben definita e formalizzata. Essa contiene tutte le informazioni necessarie per poter utilizzare la *Service* ed è il suo unico punto di connessione verso il mondo esterno. Per convenzione due *Service* che esponano la stessa interfaccia sono da considerarsi intercambiabili, quindi l'evoluzione tecnologica può agevolmente essere introdotta in modo trasparente su ogni singola *Service* senza impattare sull'intero progetto. Un altro concetto importante richiesto dalla metodologia SOA è la completa separazione dello strato implementativo delle *Service* dalla gestione dei flussi di lavoro che le utilizzeranno. A questo modo si migliora la flessibilità ed adattabilità del sistema alle condizioni dinamiche dell'ambiente.

Per favorire lo sviluppo di un mercato di componenti per la manipolazione di contenuti multimediali che aderisca alla metodologia SOA e che presenti interfacce standardizzate, l'*European Broadcasting Union* (EBU) in collaborazione con l'*Advanced Media Workflow Association* (AMWA) ha definito uno standard, denominato **Framework for Interoperable Media Services (FIMS)**, che prescrive il comportamento e l'interfaccia dei moduli più utilizzati nella realizzazione di sistemi di produzione e archiviazione multimediale [6], tra cui acquisizione, trasformazione, copia e immagazzinamento. Come si può vedere in Figura 5 il ruolo dello standard FIMS (in arancione nella Figura) è quello di fornire un livello di astrazione tra la logica di controllo del sistema (orchestrazione) e l'implementazione dei servizi sottostanti.

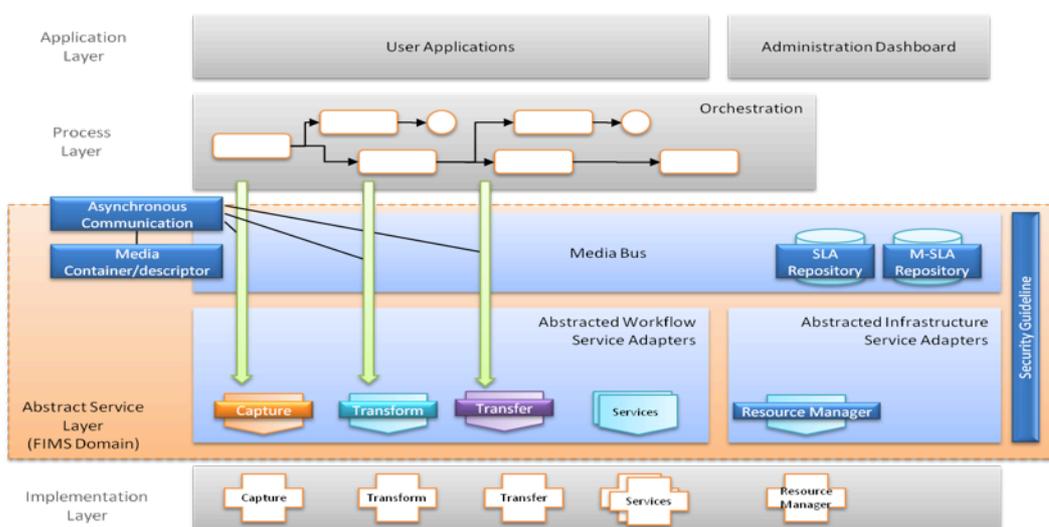


Figura 5
Architettura di riferimento di FIMS.

8. Un esempio di applicazione: il nuovo archivio “master” RAI

La ristrutturazione degli archivi RAI è iniziata nel 1997, con il progetto Catalogo Multimediale, descritto estensivamente in [1], il cui scopo è fornire uno strumento omnicomprensivo di ricerca e selezione dei contenuti di archivio, basato su tecnologie multimediali che permettano la visione dei contenuti senza necessità di accesso all’archivio master.

Per quanto il progetto di ristrutturazione degli archivi coprisse tutte i vari aspetti di archiviazione e prevedesse, in particolare, anche la creazione di una teca *master* digitale, la scelta definitiva che si fece allora fu di mantenere i *master* su cassette video tradizionali, sia per ragioni economiche e di complessità nella realizzazione di una teca master informatica con le tecnologie del tempo, sia perché più funzionali alla normale operatività di produzione. Sono comunque state effettuate tutte le operazioni di riversamento necessarie per portare i contenuti originati su formati video obsoleti (nastri aperti “2 Pollici”, “1 Pollice” e BVU) sui supporti attuali digitali (IMX).

Oggi, grazie ai progetti **T3** (*Transition To Tapeless*) e **NIPS** (*News Integrated Production Systems*), la catena produttiva di programmi e notiziari si sta evolvendo verso la **completa informatizzazione** con la conseguente eliminazione dei supporti video tradizionali. Le condizioni sono quindi mature perché anche l’archivio master si trasformi secondo le linee guida descritte in questo articolo e diventi un’infrastruttura sempre più integrata nel processo produttivo. È perciò stato avviato il progetto **DigiMaster**, relativo al riversamento su file di tutti i contenuti dell’attuale archivio master, le cui quantità sono riassunte in Tabella 2, ed è in corso un’estensione del sistema T3 che lo renderà in grado di gestire l’intero patrimonio audiovisivo. Il piano prevede il completamento dei riversamenti e dismissione dell’archivio attuale entro il 2020.

Tipologia supporto	Quantità	Formato
Pellicola 16 e 35 mm	880.000	Rullino Film
Betacam	890.000	Cassetta
D2	220.000	Cassetta
IMX	510.000	Cassetta

Tabella 2
Consistenza dell’archivio televisivo master RAI

Il sistema T3 è stato concepito come infrastruttura di collegamento fra le varie isole di produzione digitale già esistenti, dalla post-produzione alle *Newsroom*, come schematizzato in Figura 6. È costituito da due moduli integrati: il *Media Asset Manager* (MAM) fornisce servizi di persistenza dei contenuti, sia durante il ciclo di lavorazione che a lungo termine, mentre la *Media Factory* gestisce flussi e risorse computazionali per operazioni di adattamento dei contenuti e

controllo di qualità relativamente a processi aziendali come il collaudo tecnico dei contenuti d'acquisto o la preparazione per la messa in onda dei programmi. Il MAM è responsabile per la gestione dello *storage* fisico, denominato *Asset Repository*, un sistema di storage gerarchico a sua volta costituito da un sottosistema *On-line* basato su hard disk, organizzati in strutture RAID e un sottosistema *Near-on-Line* basato su librerie robotizzate di nastri LTO6. Tutti i *file system* dell'archivio *On-line* di T3, sia quelli realizzati su dischi veloci che quelli su dischi capacitivi, svolgono sia un ruolo di archiviazione che di trasformazione, pertanto sono stati configurati in RAID5, come meglio descritto nel Riquadro 2. Per ragioni di sicurezza dei dati l'*Asset Repository* è interamente ridondato in due installazioni gemelle in due locazioni indipendenti. Con l'espansione prevista per ospitare l'archivio master digitale ciascun *Asset Repository* sarà costituito da uno spazio *Near-on-line* di 15 PetaByte espandibile fino a 50 PetaByte e di uno spazio *On-line* di almeno 2 PetaByte.

La progettazione è stata effettuata secondo la metodologia SOA e si basa sull'impiego di protocolli standard come FIMS e MXF.

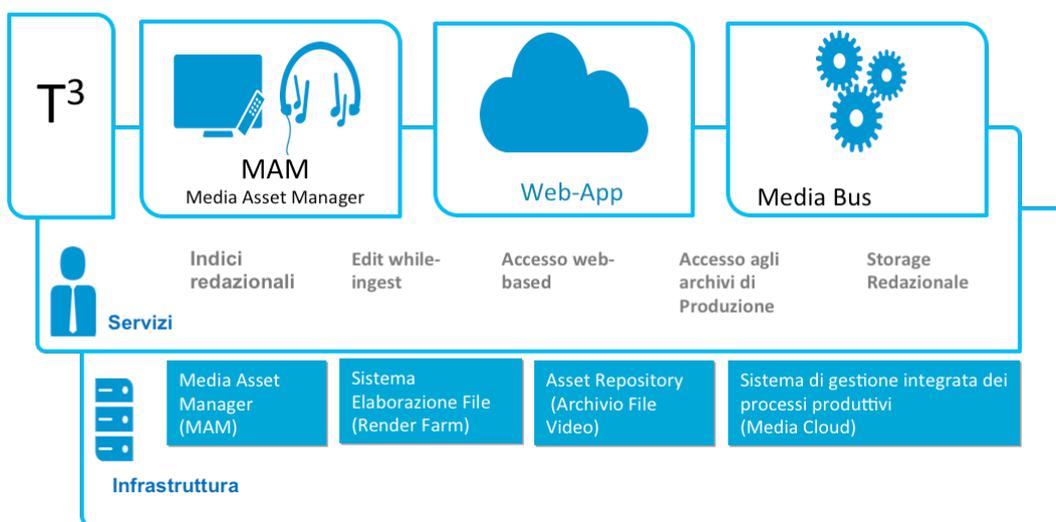


Figura 6
Ambiente operativo del sistema T3

Il progetto *DigiMaster* si avvale di tecnologie messe a punto all'interno di progetti finanziati dalla Comunità Europea (*PrestoSpace*¹, *PrestoPrime*²) e dal Fondo Europeo di Sviluppo Regionale (*Adam*³), nel corso dei quali sono state studiate le problematiche relative al corretto trattamento dei supporti video analogici e alla loro digitalizzazione, con particolare attenzione al controllo di qualità. È stata messa a punto una filiera altamente integrata che include le

¹ <http://www.prestospace.org/index.it/html>

² <http://www.prestoprime.org/>

³ <http://www.adam-project.eu/>

operazioni di bonifica del magazzino (eliminazione di doppi e materiale di scarso interesse), pulizia nastri, digitalizzazione, controllo di qualità del riversamento, integrazione della documentazione per il Catalogo Multimediale e consegna del nuovo file master al sistema T3. I supporti video su cassetta vengono riversati tramite l'utilizzo di un sistema robotizzato che sottende alle operazioni di pulizia e riversamento nastri in modo autonomo senza necessità di intervento dell'operatore. Ciascuna linea è in grado di lavorare sino a 10.000 supporti all'anno. I nastri sono tutti a definizione standard (720x576@25i formato 4:2:2) e verranno convertiti in file in formato MXF D10⁴. Le pellicole invece richiedono una lavorazione manuale che consiste nella pulizia su tavolo passafilm, eventuale risincronizzazione dell'audio su moviola e digitalizzazione tramite film scanner. Il materiale 16 mm verrà digitalizzato ad alta definizione (1920x1080@25p 4:2:2) in formato MXF XDCAM⁵. Le pellicole 35 mm saranno invece digitalizzate in UHD (3840x20160@25p 4:2:2) in un formato file non ancora stabilito.

Dal processo di bonifica ci si aspetta che il materiale risultante da digitalizzare ammonti a circa 300.000 ore. Poiché un'ora sia in formato D10 che XDCAM richiede circa 30 GB di storage, l'occupazione totale all'interno dell'infrastruttura T3 sarà di circa 9 PetaByte (da moltiplicare per il numero di copie di sicurezza effettuate).

9. Conclusioni

Nel presente articolo è stato descritto lo stato dell'arte relativo alla organizzazione degli archivi digitali con particolare riferimento a quelli conservativi. Poiché la tecnologia nel campo informatico avanza molto rapidamente, è lecito aspettarsi che anche le architetture per l'archiviazione ne risentiranno pesantemente. L'utilizzo di metodologie di progettazione come SOA che favoriscano il disaccoppiamento dei componenti è perciò fondamentale per garantire la longevità dei sistemi.

Non deve inoltre essere sottovalutato l'impatto che le nuove architetture che compongono il *cloud computing* avrà sui sistemi di archiviazione. Il *cloud computing* introduce nuovi concetti come "portare le risorse di calcolo vicino ai dati" o "archiviare oggetti anziché file", sicuramente rilevanti ed innovativi. Numerosi progetti di ricerca sono attivi in questo contesto con lo scopo di valutare i pro e i contro dell'approccio *cloud*, che, accanto ai vantaggi attesi come scalabilità e flessibilità dell'infrastruttura, pongono anche delle problematiche nuove, in particolare sulla sicurezza dei dati e sulla connettività. Questo argomento potrà essere oggetto di un futuro articolo di approfondimento.

⁴ Il formato prevede che il video sia codificato in MPEG2 intraframe a 50 Mb/s e contiene sino a 8 tracce audio non compresse. Vedere norma SMPTE D10.

⁵ video è codificato in MPEG2 con predizione interframe a 50 Mb/s e contiene sino a 16 tracce audio non compresse. Vedere norma SMPTE RDD9.

Appendice

La creazione dell'AIP (Archival Information Package) di un segnale audiovisivo

Com'è noto, un segnale audiovisivo viene rappresentato nel dominio digitale come una sequenza di immagini, ciascuna delle quali è costituita da una matrice di punti (pixel) descritti da tre componenti che ne determinano luminanza (Y) e crominanza (Cr e Cb). Le componenti di crominanza possono essere sottocampionate rispetto alla luminanza in quanto l'occhio umano è meno sensibile all'informazione di colore che di luminosità. I formati di sottocampionamento più utilizzati sono denominati **4:4:4**, ovvero nessun sottocampionamento viene applicato (utilizzato per lo più nella grafica), **4:2:2**, le due componenti di crominanza vengono sottocampionate di un fattore 2 nella direzione orizzontale (formato video per produzione televisiva) e **4:2:0**, le due componenti di crominanza vengono sottocampionate di un fattore 2 sia orizzontalmente che verticalmente (formato per la distribuzione televisiva verso l'utente). La tabella A1 descrive i parametri utilizzati nelle applicazioni televisive a norma europea.

	4:4:4	4:2:2	4:2:0
Definizione standard (SD)	Y 720x576@25i 8 bit/pixel Cr 720x576@25i 8 bit/pixel Cb 720x576@25i 8 bit/pixel	Y 720x576@25i 8 bit/pixel Cr 360x576@25i 8 bit/pixel Cb 360x576@25i 8 bit/pixel	Y 720x576@25i 8 bit/pixel Cr 360x288@25i 8 bit/pixel Cb 360x288@25i 8 bit/pixel
Alta definizione (HD)	Y 1920x1080@25i 8-10 bit/pixel Cr 1920x1080@25i 8-10 bit/pixel Cb 1920x1080@25i 8-10 bit/pixel oppure Y 1280x720@50p 8-10 bit/pixel Cr 1280x720@50p 8-10 bit/pixel Cb 1280x720@50p 8-10 bit/pixel	Y 1920x1080@25i 8-10 bit/pixel Cr 960x1080@25i 8-10 bit/pixel Cb 960x1080@25i 8-10 bit/pixel oppure Y 1280x720@50p 8-10 bit/pixel Cr 640x720@50p 8-10 bit/pixel Cb 640x720@50p 8-10 bit/pixel	Y 1920x1080@25i 8 bit/pixel Cr 960x540@25i 8 bit/pixel Cb 960x540@25i 8 bit/pixel oppure Y 1280x720@50p 8 bit/pixel Cr 640x360@50p 8 bit/pixel Cb 640x360@50p 8 bit/pixel
Ultra Alta Definizione (UHD)	Y 3840x2160@50p 10 bit/pixel Cr 3840x2160@50p 10 bit/pixel Cb 3840x2160@50p 10 bit/pixel	Y 3840x2160@50p 10 bit/pixel Cr 1920x2160@50p 10 bit/pixel Cb 1920x2160@50p 10 bit/pixel	Y 3840x2160@50p 10 bit/pixel Cr 1920x1080@50p 10 bit/pixel Cb 1920x1080@50p 10 bit/pixel

Tabella A1. Formati video televisivi.

I parametri sono indicati nella forma WxH@F dove W è la risoluzione orizzontale, H la risoluzione verticale e F la frequenza di quadro. La lettera i o p indica rispettivamente se la scansione è interlacciata o progressiva

L'audio associato può essere nei formati mono, stereo o multicanale (5.1). Ciascun canale audio è composto da una sequenza di campioni codificati con parole a 16 o 24 bit e una frequenza di campionamento di 48kHz. In produzione normalmente gli effetti audio e il parlato vengono tenuti su tracce separate e possono essere presenti più lingue. Quindi il numero di tracce audio presenti può essere notevolmente superiore a quanto richiesto dal formato finale (8 o 16 tracce audio sono valori comuni).

Per ridurre lo spazio di memorizzazione il segnale viene compresso con tecniche di codifica entropica. Poiché vengono utilizzati per ragioni di efficienza sistemi di codifica con perdita, quali MPEG2, MPEG4-AVC o JPEG2000, la compressione implica che parte dell'informazione contenuta nel segnale originale venga rimossa permanentemente. Un segnale video contiene molta informazione irrilevante per la percezione umana (l'occhio non è un trasduttore perfetto) e i sistemi di codifica allo stato dell'arte operano principalmente su questa, tuttavia un fattore di compressione troppo elevato produrrà difetti visibili sul segnale e potrebbe pregiudicarne il futuro utilizzo.

In un archivio televisivo possono essere conservati contenuti audiovisivi a diversi livelli di qualità. I contenuti di pregio, prodotti internamente e di cui si detengono tutti i diritti di sfruttamento, vengono di norma conservati alla qualità master, ovvero alla qualità a cui sono stati prodotti. Contenuti d'acquisto, di cui si hanno diritti limitati, o provenienti dalla digitalizzazione di supporti analogici obsoleti possono invece essere archiviati ad un livello di qualità inferiore, applicando un maggiore fattore di compressione del segnale. Analogamente, accanto al master è possibile conservare per ragioni pratiche altre versioni del contenuto a qualità inferiore o a risoluzione ridotta, ad esempio a scopo di visione (selezione rapida dei contenuti da acquisire dall'archivio) o per un utilizzo specifico, come l'alimentazione di servizi di video on demand.

Come prescritto dal modello OAIS, il segnale audiovisivo codificato (vedere l'Appendice per una descrizione dei formati audiovisivi) deve essere incapsulato in un opportuno contenitore che unitamente alla sequenza di bit che costituisce il segnale, contenga anche tutti i metadati essenziali al suo utilizzo, comunemente denominato *file wrapper*. Esistono svariati formati di *file wrapper* con caratteristiche e finalità differenti. Il formato file in cui il contenuto viene memorizzato (AIP nella terminologia OAIS) sarà quindi costituito dal formato del *file wrapper* e dal formato di codifica del contenuto.

Il formato di *file wrapper* più utilizzato nelle applicazioni televisive professionali, sia di archiviazione che di produzione, è lo standard SMPTE MXF [4]. Questo standard definisce la struttura di un formato contenitore per materiali audiovisivi e relativi metadati flessibile ed estensibile. I dati video e audio possono essere organizzati in vari modi per facilitare operazioni di montaggio piuttosto che di streaming o per gestire in un solo file edizioni diverse dello stesso prodotto (es, colonne sonore alternative). È inoltre possibile definire schemi di metadati specifici per particolari applicazioni. Il formato è organizzato a profili di differente complessità e può essere esteso tramite la definizione di standard SMPTE complementari che specificano aspetti particolari, ad esempio la mappatura di un particolare formato di codifica video nella struttura dell'MXF. La struttura di base dell'MXF è esemplificata in Figura A1.

È possibile inserire metadati sia all'inizio del file (*Header*) che in coda (*Footer*). Può essere presente un indice (*Index Table*) per accedere rapidamente a parti significative del contenuto ed è possibile mappare in vari modi le tracce audio e le tracce video. Nell'esempio mostrato è presente una sola traccia video (*Picture Item*) a cui sono associate due tracce audio (*Sound Element*). Dopo ciascuna

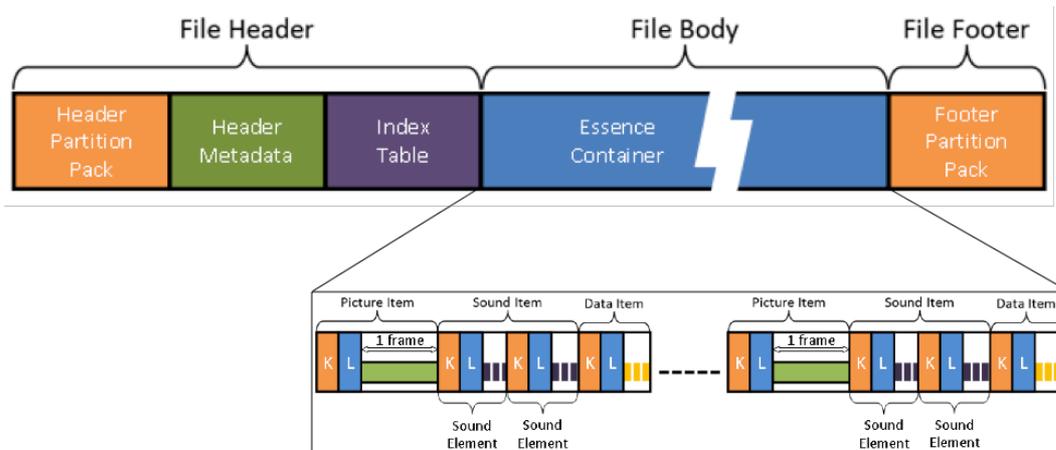


immagine del video è presente il frammento di audio corrispondente (configurazione ottimale per la riproduzione del contenuto).

Figura A1
Schema di principio del formato MXF

Riquadro 1 - I diversi tipi di archivi per grandi masse di dati**a) Archivi audiovisivi o multimediali**

Questa classe di archivi può considerarsi rappresentativa della categoria di archivi definiti conservativi che rappresenta quella maggiormente approfondita nel presente articolo. La finalità è quella di preservare per moltissimi anni, rendendole accessibili ed utilizzabili in ogni momento, collezioni organizzate di opere audiovisive di vario genere quali opere audio/radiofoniche, televisive ed opere cinematografiche.

Un tema centrale di tale classe di archivi è l'applicazione di tecniche di compressione del segnale in modo da mantenere in dimensioni gestibili la quantità di dati da memorizzare e da movimentare durante l'accesso ai contenuti. In particolare, la copia master viene normalmente salvata in formati che implicano bit rate anche 10 volte superiori a quelli usati durante la distribuzione del segnale verso l'utente mentre per la cinematografia digitale per la copia master non si applica alcuna compressione.

Per i grandi broadcaster televisivi, la capacità di questi archivi può andare da 2 PetaByte On-line e 10-15 PetaByte Near-on-line a 8-10 PetaByte On-line e 150-200 PetaByte Near-on-line.

b) Archivi Cloud (per servizi "Cloud" propriamente detti)

Questi tipi di archivi sono caratterizzati da contenuti informativi a carattere individuale (e non di alto interesse per l'insieme della collettività) messi a disposizione da Provider di telecomunicazioni (in genere a pagamento) con opportuni spazi di memorizzazione. Sono rivolti ad un mondo essenzialmente consumer (migliaia/milioni di utenti) ma anche professional per aziende medio-piccole mettendo a disposizione spazi di memorizzazione di cui è fondamentale l'accesso anche in mobilità.

In questa classe di archivi i dati non sono strutturati, non rappresentano una collezione, non richiedono una persistenza eterna, ma l'aggregato porta generalmente a dimensioni estremamente elevate con problemi non banali di estensibilità, velocità di accesso ai dati, garanzie concordate con gli utenti, ecc ...

Il motivo per cui gli archivi di tipo cloud stanno acquisendo importanza è legato, in prevalenza, alla competitività e flessibilità del loro costo, liberando l'utente dall'onere di possedere e gestire in proprio l'infrastruttura di archiviazione.

Per un grande Provider di servizi Cloud, la capacità di questi archivi può andare da 8-10 PetaByte On-line e 80-100 PetaByte Near-on-line fino a 20-40 PetaByte On-line e 250-300 PetaByte Near-on-line.

c) Archivi OTT (per servizi "Over The Top")

Sono gli archivi (quasi sempre distribuiti geograficamente e spesso free) basati generalmente su tecnologie cloud, nati per i vari servizi OTT quali motori di ricerca (Google, ecc), relazionali (Facebook, ecc), commercio elettronico (Amazon, ecc), distribuzione musicale, cinematografica ed altri nati per i vari servizi OTT. Sono rivolti ad un mondo essenzialmente consumer e sono resi disponibili, nell'ambito di opportune regole, agli utenti per i contenuti cui accedere (clip, immagini, video) o per i contenuti che gli stessi utenti hanno affidato per la conservazione (foto, ecc) con servizi aggiunti applicativi (Dropbox, Netflix, Spotify, Instagram, Sky drive, ecc).

Per un grande provider di servizi OTT, la capacità di questi archivi è prevalentemente On-line e può andare da 40 PetaByte a 80-100 PetaByte.

d) Archivi Big Data

Con il termine Big Data ci si riferisce a raccolte di dati che debbono essere elaborati per successive analisi. Per tali archivi quello che maggiormente interessa è l'informazione ottenuta nella correlazione statistica di migliaia/milioni di dati non strutturati piuttosto che i dati singoli. Questa particolare finalità richiede strumenti molto differenti da quelli tradizionali in tutte le fasi del processo: dall'acquisizione al risultato finale ad iniziare dai costi di archiviazione richiesti per i dati origine che devono essere molto bassi.

I grandissimi archivi di "eventi" probabilistici tipo quello del CERN di Ginevra possono ricondursi a tali tipi di archivi.

Riquadro 2 - Tecniche moderne per integrità nello storage dei dati

Dalle classiche ed elementari tecniche informatiche di protezione, molti passi avanti sono stati fatti nella protezione dei dati, in particolare per applicazione agli archivi.

L'impiego di tali tecniche avanzate nell'unità di *storage* (*Archival Storage*) degli archivi dipende dalla progettazione dell'intero sistema e può o meno essere impiegata a seconda di vari tipi di considerazioni.

Con la denominazione RAID acronimo di *Redundant Array of Independent Disk* ci si riferisce ad una tecnica di *storage* virtualizzato dei dati che combina una molteplicità di supporti in una unica unità logica con lo scopo di aumentare le ridondanze e/o le prestazioni del singolo supporto utilizzato. I dati sono distribuiti su più supporti con l'aggiunta di supporti di ridondanza (complessivamente una matrice monodimensionale di supporti) che sono utilizzati per la correzione degli errori. I differenti "modelli" sono definiti RAID 0, RAID 1 etc. in accordo con vari livelli sempre più sofisticati di possibili implementazioni. I differenti modelli hanno lo scopo di dare maggior enfasi alle prestazioni dell'unità di memorizzazione stessa come affidabilità, disponibilità, velocità di risposta, capacità.

In estrema sintesi e semplificazione, per la memorizzazione di una parola di 8 bit la tecnica consiste nel disporre di una pila di nove dischi e di porre il primo bit sul primo supporto, il secondo sul secondo e così via. Il supporto aggiuntivo, il nono, conterrà il bit di parità. In caso di guasto di uno degli otto supporti sarà facilmente ricostruibile il valore del bit perso utilizzando l'informazione di parità. Risulta intuitivo dedurre che disponendo di un maggior numero di supporti aggiuntivi e rendendo più complessi gli algoritmi di ricostruzione sia possibile rimediare a maggiori quantità di errori.

I modelli più recenti impiegano tecniche combinate di "*time interleaving*" e "*space interleaving*" in associazione a codici correttori Reed Salomon per ricostruire i bit persi nell'eventuale guasto parziale o totale di un disco.

Nel 1986 fu brevettato un modello da un ricercatore dei laboratori IBM che fu successivamente definito RAID 5. A partire da metà anni 90 l'utilizzo commerciale di *Array Disk* si è andato ampliando moltissimo per vari tipi di archivi. Negli anni successivi gli sforzi dei produttori si sono orientati a ottimizzare i tempi di ricostruzione della informazione e a ridurre il degrado delle prestazioni nel tempo di ricostruzione. Ad esempio una problematica sollevata è costituita dal fatto che la matrice dei dischi è costruita con la stessa tecnologia e opera per la stessa quantità di tempo; di conseguenza la probabilità che il guasto possa avvenire in almeno due dischi con elevata vicinanza temporale può risultare significativa.

Le caratteristiche delle differenti tipologie di RAID sono riassunte nella tabella seguente.

Caratteristiche per tipologia di RAID	RAID 0	RAID 1	RAID 5	RAID 6	RAID 1+0
Numero minimo di dischi richiesti	2	2	3	4	4
Fault Tolerance (numero di dischi sacrificabili senza perdita di dati)	0	1	1	2	1
Capacità residua	Originale	0.5	4/5	4/6	0.5
Velocità di accesso (read/write)	Fast read and write	Fast read, slow write	Fast read, medium write	Fast read, slow write	Fast read, medium write
Costo di implementazione	Dischi economici, controller molto costosi	dischi costosi	medio (software)		dischi e controller costosi

La scelta di un tipo rispetto a un altro dipende dal ruolo che si vuole dare allo specifico file system. A titolo di esempio si possono identificare le seguenti tipologie di RAID a seconda del contesto:

RAID 0, Produzione (Film) e trasformazione

RAID 1, Sistemi Operativi

RAID 5, Acquisizione e trasformazione

RAID 6, Archivio Conservativo

RAID 1+0, Produzione (News) e trasformazione

Tipicamente lo stesso archivio ricopre più ruoli pertanto, in esso, coesistono più tipologie di RAID.

Riferimenti

- [1] “Le Teche Audiovisive Nell’Era Digitale”, rivista Mondo Digitale, Supplemento al n.1, Marzo 2005
- [2] Consultative Committee for Space Data System, “Reference Model for an Open Archival Information System (OAIS)”, Recommended Practice CCSDS 650.0-M-2, 2012 <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- [3] Library of Congress, “Sustainability of Digital Formats Planning for Library of Congress Collections”, <http://www.digitalpreservation.gov/formats/index.shtml>
- [4] Society of Motion Picture and Television Engineers, “SMPTE 377M-1: The MXF File Format Specification”, 2009
- [5] J. Footen e J. Faust, “The Service-Oriented Media Enterprise: SOA, BPM, and Web Services in Professional Media Systems”, 2008, Focal Press
- [6] European Broadcasting Union, “EBU Tech 3356: FIMS Media SOA Framework V.1.1”, 2014, <http://www.fims.tv>

Biografie

Giorgio Dimino si è laureato in ingegneria elettronica presso il Politecnico di Torino nel 1987. Dal 1988 è ricercatore presso il Centro Ricerche e Innovazione Tecnologica (CRIT) della RAI. Attualmente è responsabile per le attività di supporto alle ingegnerie relative all’innovazione tecnologica. I campi di ricerca che coordina, nell’ambito del CRIT, sono relativi all’applicazione delle tecnologie informatiche nella catena di produzione televisiva, alla progettazione di archivi multimediali e alla valutazione della qualità tecnica dei segnali audiovisivi, in particolare applicata ai nuovi formati ad altissima definizione (UHD). È attivo in vari enti di standardizzazione internazionali.

E-mail: giorgio.dimino@rai.it

Marco Montenovo ha sempre ricoperto il ruolo di architetto, focalizzando i suoi interessi nella trasformazione delle tecnologie della produzione e distribuzione TV e le nuove modalità di fruizione di TV interattiva e partecipata. Ricopre dal 2010 il ruolo di DM-CDN Practice Principal in Hewlett Packard Italiana quale responsabile del gruppo che si occupa della progettazione ed esecuzione di progetti in ambito Digital Media in Italia, Francia, Grecia, Egitto ed Israele. In passato e' stato World-Wide di Global Domain Architect and Global Practice Principal, dove ha definito l'architettura di riferimento delle soluzioni Digital Media per HP e lo sviluppo del team di architetti a supporto dei progetti internazionali. Nel 2004 Marco ha creato il Digital Media Lab di Roma dove sono stati realizzati diversi prodotti e soluzioni per il mercato italiano ed internazionale che rimane ancora sotto al suo diretto controllo. Marco e' di classe 1966, ha effettuato studi classici ed e' laureato in Ingegneria Elettronica.

E-mail: marco.montenovo@hp.com

Smart City: Tecnologie, Architetture e Servizi

F. Ferrero, A. Pacifici, A. Vesco

Sommario. *L'Unione Europea intende accelerare il percorso di sviluppo delle città verso una prospettiva di triplice sostenibilità (sociale, economica e ambientale).*

Molte opportunità si presentano nel percorso verso tale obiettivo per le Imprese, la Pubblica Amministrazione, le Università ed i Centri di Ricerca. Inoltre le Tecnologie dell'Informazione e della Comunicazione (ICT) giocano un ruolo fondamentale per la realizzazione di città intelligenti e sostenibili. L'articolo analizza i principali ambiti applicativi delle Smart City, ribadendo la necessità di un approccio integrato alla pianificazione e gestione urbana.

Keywords: Smart City, ICT, Architectures, Smart Services, Sustainability

1. Introduzione

Il termine *Smart City*, o città intelligente, non è facilmente caratterizzabile. Non esiste una definizione universalmente accettata, ma un'autentica moltitudine di definizioni, non sempre coerenti, che tendono a mettere in luce aspetti differenti di un problema con molte sfaccettature [1]. Nonostante questa proliferazione di interpretazioni e definizioni, l'essenza del problema *Smart City* è relativamente semplice: dal momento che siamo in presenza di una fortissima concentrazione della popolazione nelle aree urbane, le città sono i luoghi del pianeta nei quali si concentra ormai la maggior parte del consumo di energia e di risorse non rinnovabili. Ciò implica che le innovazioni che ci devono guidare verso un nuovo modello di sviluppo sostenibile devono essere sperimentate prima di tutto all'interno delle città, perché è lì che possono indurre i maggiori benefici. *Smart* sarà pertanto quella città capace di attivare un approccio integrato e interdisciplinare alla progettazione e al funzionamento dei suoi meccanismi, con

l'obiettivo di realizzare, col tempo, il paradigma della triplice sostenibilità: ambientale, sociale ed economica.

È proprio il perseguimento di un modello di sviluppo sostenibile che ha spinto la Commissione Europea a dedicare alle *Smart City* una delle più importanti iniziative industriali del proprio Piano Strategico sulle Tecnologie per l'Energia, più noto come *SET-Plan* [2] e in seguito ad annunciare la creazione di una *European Innovation Partnership* dedicata alle *Smart Cities and Communities* [3], con lo scopo di accelerare la realizzazione e l'adozione su vasta scala di soluzioni innovative che si collocano all'incrocio di tre filiere strategiche dell'evoluzione urbana: Energia, ICT e Mobilità.

2. Ambiti Applicativi delle Smart City

Gli ambiti applicativi delle *Smart City* sono molteplici e correlati tra loro. In accordo alle priorità definite da diverse amministrazioni andremo di seguito ad approfondire gli ambiti applicativi dell'energia, dell'edilizia, della mobilità, dell'acqua, della salute e dell'*e-government*.

Un ambito applicativo molto importante è quello dell'*energia*, che deve necessariamente diventare sostenibile nelle sue fasi di produzione, distribuzione e consumo. In questa direzione si muovono tutti gli sforzi impiegati per lo sviluppo della cosiddetta *smart grid*, una rete intelligente di distribuzione che fa largo uso di un'infrastruttura di comunicazione per controllare efficacemente la distribuzione elettrica. La *smart grid*, inoltre, integra e controlla la produzione da fonti rinnovabili, che sono caratterizzate da un funzionamento intermittente, ma che sono vitali per il raggiungimento della sostenibilità nella fase di produzione. Questa capacità di integrazione e controllo delle fonti rinnovabili diventa ancora più centrale dato il trend di diffusione della microgenerazione distribuita da parte dell'utente, che implica la massimizzazione dell'autoconsumo e la possibilità di vendere alla rete l'energia in eccesso. Guardando al tema dell'energia dal lato degli utenti, diventa rilevante lo *smart metering*, ovvero la misurazione in tempo reale dei consumi energetici. Questa misurazione da un lato migliora, negli utenti, la consapevolezza dei propri consumi e potenzialmente li invoglia ad assumere un comportamento più sostenibile dal punto di vista economico e ambientale; dall'altro lato aiuta il produttore nella pianificazione e nel controllo della produzione energetica. È bene sottolineare che gli stessi trend sono riscontrabili anche nell'ambito dell'energia termica e del gas.

La centralità dell'*edilizia* come ambito applicativo è fuori dubbio. Tutti noi, per diversi motivi, passiamo gran parte del tempo all'interno di edifici. Questi devono essere progettati oppure ricondizionati per ridurre l'impatto ambientale, specialmente in termini di consumi elettrici e termici. Le attività in questo ambito puntano al miglioramento dell'isolamento termico e della schermatura solare, all'utilizzo dell'energia del sole per la produzione di energia elettrica così come all'uso della geotermia per il riscaldamento, all'efficienza dei sistemi di illuminazione, all'uso razionale delle risorse idriche e alla corretta gestione della ventilazione naturale e del raffrescamento passivo. È bene richiamare i sistemi di *smart metering* per la sensibilizzazione alla riduzione dei consumi e un più efficiente collegamento alla *smart grid*. Ragionando sulla relazione tra edilizia e

lavoro risulta immediato introdurre il fenomeno degli *smart hub*, ovvero di edifici dedicati al telelavoro e al *coworking*. La prenotazione delle postazioni di lavoro all'interno di tali edifici può essere fatta da remoto dall'azienda o dal lavoratore. I centri sono normalmente equipaggiati con tecnologie di videocomunicazione ad alta definizione e altri strumenti di collaborazione, per consentire l'interazione con colleghi e clienti. Gli spazi dedicati al telelavoro possono essere affiancati da servizi di accoglienza (es. asili per i figli dei dipendenti, catering, palestre, ecc.) che trasformano questi *smart hub* in veri e propri centri di inclusione sociale, specialmente se collocati in aree periurbane che sarebbero altrimenti abbandonate durante le ore lavorative. L'interazione tra lavoratori di aziende diverse può inoltre favorire la nascita di nuovi business e generare idee innovative. Infine, grazie alla riduzione degli spostamenti e all'utilizzo di tecnologie evolute per la gestione dell'energia all'interno degli *smart hub*, si possono ridurre i consumi energetici, i costi a carico dei lavoratori e delle aziende, e le emissioni inquinanti, con notevoli benefici economici e ambientali.

Un ambito applicativo centrale è quello della *mobilità* dei cittadini e delle merci. La mobilità è in costante evoluzione nella nostra società. È evidente che la transizione verso un modello sostenibile di mobilità non è fatta solamente di veicoli e infrastrutture, ma anche di innovazione tecnologica dei servizi di supporto. Quando si parla di mobilità sostenibile, immediatamente si pensa al veicolo elettrico. Quest'ultimo rappresenta sicuramente un'ottima soluzione per ridurre le emissioni inquinanti sul luogo di utilizzo, e per ridurre *tout court* se l'energia che lo alimenta è prodotta da fonti rinnovabili, ma è solo un tassello di una soluzione più complessa. Per ridurre veramente la congestione delle città e le esternalità negative per l'ambiente il trasporto pubblico diventa un tassello fondamentale, così come le soluzioni di *car sharing* e di *car pooling* e il supporto alla mobilità ciclabile e pedonale. Tali soluzioni devono essere integrate nell'area urbana e facilitate attraverso politiche e servizi di infomobilità che aiutino e involino il cittadino a farne uso. Un tema molto importante è la multimodalità, ovvero la possibilità di compiere un tragitto mediante l'uso di mezzi diversi. La mobilità multimodale richiede sistemi integrati di pianificazione dei tragitti e soluzioni di bigliettazione integrata. Guardando infine alla mobilità delle merci è evidente che la logistica dell'ultimo miglio ha un impatto enorme sull'ambiente urbano. In questo campo sono allo studio tutta una serie di strategie che puntano ad ottimizzare la consegna per ridurre l'impatto del trasporto merci sul traffico urbano, in termini di inquinamento e congestione, senza incidere sulle condizioni di lavoro dell'operatore o anzi migliorandole.

Un ambito applicativo di fondamentale importanza è poi quello dell'*acqua*. L'acqua è una risorsa vitale e non deve essere sprecata, quindi la sua distribuzione a livello urbano deve essere monitorata in tempo reale per essere ottimizzata e per rilevare possibili perdite nella rete. Ricadono inoltre in questo ambito tutte le soluzioni per la gestione intelligente delle acque piovane e di eventuali inondazioni, purtroppo sempre più frequenti nelle città attraversate da fiumi e corsi d'acqua.

Nell'ambito della *salute*, la telemedicina rappresenta una particolare modalità di erogazione dell'assistenza sanitaria, attraverso cui si rendono disponibili le

risorse mediche di base e specialistiche a pazienti che non possono accedere direttamente ai servizi sanitari per difficoltà di varia natura. È indubbio che la sua adozione possa portare a una riduzione dei costi diretti e indiretti delle prestazioni. Gli sviluppi in questo settore consentono la digitalizzazione degli strumenti, che producono dati numerici che possono essere gestiti e trasmessi ad appositi centri servizi nei quali vengono monitorati e analizzati per verificare automaticamente il superamento di determinate soglie e generare degli *alert*, nonché memorizzati per produrre una cartella clinica elettronica del paziente. Tipiche applicazioni di telemedicina riguardano la medicina d'urgenza, la diffusione capillare sul territorio di competenze specialistiche, l'esecuzione di monitoraggi nell'ambito domiciliare, la didattica e la diffusione telematica dell'informazione riguardante i servizi sanitari.

Per quanto concerne l'*e-government*, una declinazione *smart* del governo della città prevede la creazione di punti di presenza virtuali della Pubblica Amministrazione (decentrati o addirittura sul computer del cittadino), in qualsiasi luogo e mediante l'uso di diversi dispositivi, allo scopo di migliorare la diffusione spaziale e temporale dei servizi, riducendo allo stesso tempo i costi. Un tipico scenario applicativo è il dislocamento di chioschi o uffici dotati di tecnologie di videocomunicazione ad alta definizione per l'interazione con operatori remoti. Queste soluzioni consentono all'amministrazione di concentrare e ottimizzare gli uffici comunali, e ai cittadini di poter accedere a determinati servizi o a personale esperto vicino alla propria residenza. I chioschi o gli uffici remoti possono essere dotati di strumenti per la stampa e la condivisione di documenti e altri contenuti multimediali, e sono normalmente videosorvegliati. Naturalmente un'applicazione efficiente di questo concetto deve basarsi su un'avanzata ingegneria di processo che garantisca un accesso ai servizi semplice e robusto e un aggiornamento dei dati tempestivo ed efficace. Di recente, con il termine *e-government*, ci si riferisce anche a quelle azioni di coinvolgimento dei cittadini per indurli a una *partecipazione attiva* alle attività di governo attraverso l'uso delle nuove tecnologie di comunicazione.

È evidente che questi ambiti applicativi hanno una ricaduta gli uni sugli altri e non possono essere gestiti e sviluppati in modo verticale. Una città *smart* deve analizzare le proprie capacità, i risultati che vuole raggiungere nei diversi ambiti e le risorse necessarie. La pianificazione prima, la gestione e lo sviluppo delle soluzioni *smart* dopo devono dunque prendere in esame la totale complessità dell'area urbana ed essere svolte in modo integrato per razionalizzare gli sforzi e massimizzare il beneficio per la collettività. L'ICT gioca un ruolo decisivo per la realizzazione delle soluzioni *smart* e soprattutto funge da collante tra le diverse infrastrutture in modo da abilitare una gestione integrata dell'area urbana.

3. Tecnologie ICT Abilitanti

Le tecnologie dell'informazione e della comunicazione che abilitano una pianificazione e gestione della *Smart City* basata sull'estrazione di valore dei dati relativi al contesto urbano sono molteplici. Queste tecnologie stanno acquisendo sempre maggiore importanza nella vita di tutti i giorni, sia nella fase di produzione che in quelle di centralizzazione ed elaborazione dei dati urbani. I

dati, se correttamente analizzati, sono fondamentali per supportare la presa delle decisioni e per sviluppare e fornire nuovi servizi ai cittadini. I dati stessi possono essere messi a disposizione della cittadinanza attraverso interfacce navigabili che ne aiutano la fruizione.

La raccolta dei dati urbani è tipicamente implementata attraverso le *Wireless Sensor Network* (WSN) [4]. Queste reti di sensori sono caratterizzate da un *range* (m) e da una capacità di comunicazione (bit/s) ridotti rispetto alle comuni *Wireless Local Area Network* (WLAN) [5]. Queste ultime sono talvolta utilizzate per interconnettere nuvole di WSN disperse sul territorio. Una WSN abilita una misurazione capillare e pervasiva dei sistemi urbani grazie alla sua architettura fortemente distribuita. L'architettura prevede un numero elevato di nodi sensore e nodi di interconnessione che trasferiscono le misure al nodo *gateway* connesso alla rete Internet. Il paradigma delle WSN ha portato negli anni alla realizzazione della cosiddetta *Internet of Things* (IoT), ovvero alla capacità di far comunicare attraverso Internet dispositivi eterogenei, che non sono nati per comunicare tra loro.

I dati possono però anche essere generati e/o misurati direttamente dai cittadini che si muovono all'interno del contesto urbano. In questo caso si parla di *sensing* partecipativo o *crowdsensing* [6]. I dati vengono raccolti attraverso gli *smart phone* in possesso dei cittadini o attraverso dispositivi portatili e/o indossabili, ovvero sensori appositamente studiati e dotati di molteplici funzionalità di misura e comunicazione. Il termine *crowdsensing* è altresì pertinente anche in quelle situazioni in cui con modalità attiva un cittadino condivide dati più complessi come il verificarsi di un evento urbano, ad esempio, condizione del traffico.

In altri casi, sistemi informativi diversi misurano e mettono a disposizione attraverso interfacce più o meno standard una serie di dati che possono essere integrati con nuove misure sul campo. In questi casi diventa importante l'uso di *middleware*, ovvero di sistemi di traduzione che permettono di leggere e integrare dati provenienti da fonti eterogenee.

I dati sono spesso arricchiti da informazioni di posizione e tempo ricavate dai *Global Navigation Satellite System* (GNSS) come il *Global Positioning System* (GPS) [7] o, in un prossimo futuro, Galileo [8]. Per quanto l'importanza della posizione sia chiara a tutti, il tempo di misura è anch'esso molto importante perché permette di analizzare sulla stessa scala temporale dati provenienti da punti diversi della città.

Per essere elaborati, i dati misurati dai diversi sensori devono essere centralizzati: qui entra in gioco il ruolo principale dell'infrastruttura di rete. Là dove sia presente, i dati possono essere trasferiti attraverso un accesso xDSL. In tutti gli altri casi si utilizza l'accesso mobile offerto dalle reti UMTS [9], LTE [10] e, in un prossimo futuro, 5G. In campo elettrico, i dati provenienti dagli *smart meter* vengono trasferiti tramite la tecnologia *Power Line Communication* (PLC) [11] che sfrutta la rete elettrica stessa come mezzo di comunicazione.

I dati collezionati sul territorio sono raccolti, salvati ed elaborati in sistemi basati sul paradigma del *cloud computing* [12]. Con questo termine s'intende un

insieme di tecnologie che consentono l'accesso a risorse (es. CPU, RAM, *storage* e applicazioni) in modo elastico, scalabile e *on demand*. Il *cloud* è dunque un insieme distribuito di server interconnessi che gestiscono servizi, eseguono applicazioni e archiviano dati in modo totalmente trasparente all'utilizzatore. Con l'ausilio del *cloud computing* lo sviluppo degli strumenti che abilitano una corretta gestione dei dati acquisiti è più efficiente. L'obiettivo è di far sì che il dato fornisca un reale supporto alle decisioni e sia fruibile dai cittadini, in accordo con il paradigma *open data*, in modo semplice ed intuitivo. In molti casi si usano i Sistemi Informativi Territoriali (GIS) che permettono di visualizzare i dati in modo geo-referenziato su cartine reali del contesto urbano. I Sistemi di Supporto alle Decisioni (SSD) permettono di aumentare l'efficacia delle analisi poiché forniscono supporto a tutti i soggetti che devono prendere decisioni strategiche di fronte a problemi complessi. La funzione principale di un SSD è di estrarre, da dati eterogenei, in poco tempo e in modo versatile, le informazioni utili ai processi decisionali.

4. Architettura di Riferimento

Analizziamo di seguito un'architettura proposta dall'Agenzia per l'Italia Digitale [13] per abilitare crescita e diffusione di servizi intelligenti in ambito *Smart City*. La definizione di un'architettura standard permette di superare le limitazioni che caratterizzano molte soluzioni tecnologicamente avanzate oggi sul mercato, ovvero la verticalità, l'impossibilità del riuso e la loro natura *vendor lock-in*, che consente solo al produttore di effettuare modifiche ed evoluzioni.

L'architettura di riferimento in Figura 1, presenta diversi livelli, ciascuno dei quali corrisponde ad una fase della gestione dei dati e delle informazioni. Nella fase di raccolta e invio dati tutte le informazioni, raccolte da fonti eterogenee, sono immagazzinate e convogliate verso un livello intermedio di elaborazione, dove sono analizzate e incrociate per creare intelligenza, fornire supporto alle decisioni e sviluppare nuovi servizi o forme di controllo del sistema urbano. Nella fase di distribuzione dati, infine, un flusso informativo, possibilmente in tempo reale, o comunque caratterizzato da un ritardo accettabile, su servizi, eventi e attività della *Smart City*, è reso disponibile a diversi utilizzatori. Gli utilizzatori possono essere sistemi e/o persone di natura pubblica o privata.

Le tecnologie analizzate nella Sezione 3 si integrano all'interno dell'architettura, e ne implementano i diversi blocchi. Ad esempio, la raccolta dei dati è comunemente implementata attraverso WSN o *smart meter* interconnessi tra loro oppure attraverso meccanismi di *crowdsensing* da *smart phone*. L'invio dei dati è poi implementato attraverso tecnologie di comunicazione a più lungo raggio con accesso mobile come UMTS, LTE e 5G oppure con tecnologie WLAN se il punto di connessione alla rete Internet si trova nelle vicinanze della zona di misura. La fase di elaborazione e di distribuzione sono implementate in *cloud computing* per ragioni di efficienza e scalabilità. Le applicazioni implementate in *cloud* sono dunque in grado di estrarre valore dai dati, offrire supporto alla decisione e alla pianificazione e presentare in modo semplice ed efficace dati urbani eterogenei.

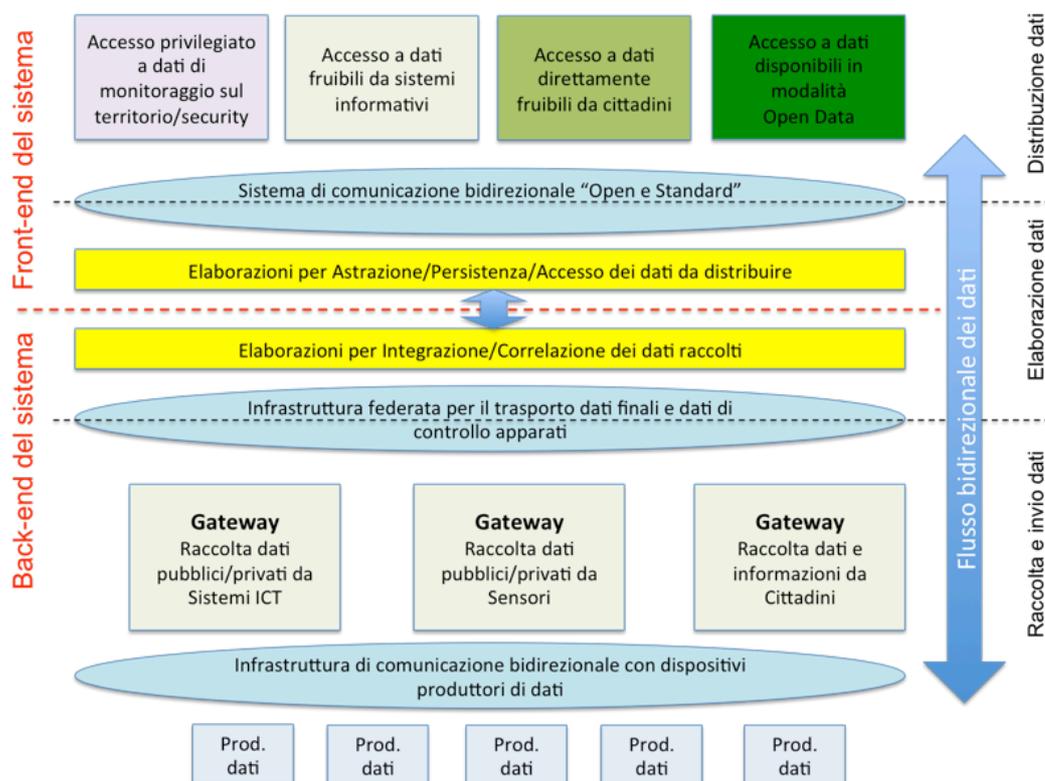


Figura 1
Architettura di riferimento per la Smart City. Fonte: [13]

In accordo a quest'architettura, la *Smart City* è un "sistema di sistemi", che si costituisce in modo dinamico in conformità a tre principali modelli di integrazione:

- Modello di integrazione tra servizi: prevede di integrare servizi esistenti al fine di crearne di nuovi a valore aggiunto e produrre informazioni arricchite;
- Modello di integrazione dei dati: consente di aggregare, trasformare, e processare dati provenienti da fonti eterogenee e distribuite, non necessariamente connesse tra loro, al fine di produrre dati a valore aggiunto e informazioni arricchite;
- Modello di integrazione partecipativo: consente di sfruttare le informazioni e i dati generati dai cittadini che utilizzano sempre più massicciamente servizi social.

5. Servizi

L'architettura e la sua implementazione con le tecnologie d'informazione e comunicazione è finalizzata all'erogazione di nuovi servizi per il cittadino e per la città, perseguendo l'obiettivo di crescita sostenibile dal punto di vista economico, sociale e ambientale. Di seguito analizzeremo alcuni servizi significativi negli ambiti applicativi evidenziati.

Nell'ambito applicativo dell'*energia* si stanno sviluppando servizi per il risparmio energetico e per un uso sostenibile delle risorse. Un esempio calzante è dato dalle piattaforme Enercloud [14] ed ENERGY@Home [15]. Enercloud punta alla realizzazione di un sistema per l'ottimizzazione dei consumi energetici imputabili all'illuminazione pubblica. Il servizio gestisce in modo intelligente la commutazione On/Off dell'illuminazione, facendo uso di sistemi embedded a basso costo e di un'applicazione di *back office* in *cloud* per la gestione remota degli impianti e la memorizzazione ed analisi dei consumi. ENERGY@Home nasce da un progetto di collaborazione avviato nel 2009 da quattro aziende (Electrolux, Enel, Indesit e Telecom Italia) con l'obiettivo di sviluppare una piattaforma di comunicazione tra *smart device* ed elettrodomestici di ultima generazione in ambito domestico. Questa piattaforma ha sviluppato un servizio innovativo per il controllo degli elettrodomestici che consente di risparmiare sulla bolletta elettrica e ridurre l'impatto ambientale dei consumi residenziali.

Nell'ambito dell'*edilizia* i servizi offerti sono molteplici. Il progetto *Lyon Smart Community* [16], nato dalla collaborazione tra il Comune di Lione e l'agenzia giapponese per l'innovazione NEDO, coinvolge il quartiere *Confluence* della città di Lione e mira a farne entro il 2016 un distretto ad alto livello di innovazione e sperimentazione. Le attività di progetto mirano alla costruzione e rivalorizzazione di un insieme di edifici in grado di generare più energia rispetto al proprio fabbisogno, grazie ad esempio all'installazione di pannelli solari sui tetti e a tecniche di costruzione che minimizzano gli sprechi di energia. L'energia in eccesso è immagazzinata e utilizzata per servizi complementari come la ricarica di veicoli elettrici messi a disposizione in modalità *car sharing*. Tutti gli edifici facenti parte di questo quartiere avranno la possibilità di gestire in tempo reale i consumi di luce, gas e acqua minimizzandone gli sprechi. Una piattaforma di controllo sarà in grado di prevedere i consumi, dunque di gestire le risorse per bilanciare domanda e offerta di luce, gas e acqua.

Per ciò che concerne la *mobilità*, si stanno diffondendo sempre più servizi di *car sharing* di tipo *free floating* [17], ovvero servizi in cui l'utente può utilizzare i veicoli messi a disposizione in maniera libera, senza dover recuperare e consegnare il veicolo in luoghi prestabiliti, come le stazioni di servizio. Questa modalità innovativa di *car sharing*, è abilitata dall'ICT e dalle tecnologie *embedded* a bordo veicolo. Il servizio permette di localizzare e prenotare i veicoli messi a disposizione attraverso il *web* o con applicazioni per *smartphone*, conoscere la quantità di carburante presente nel serbatoio, o la carica delle batterie in caso di veicoli elettrici, far inserire un *feedback* da parte degli utenti sullo stato di pulizia della macchina e permettere l'apertura delle portiere senza dover consegnare le chiavi del veicolo.

Focalizzando la nostra attenzione sul trasporto pubblico locale, notiamo come si stiano diffondendo servizi di bigliettazione integrata, ad esempio il sistema BIP [18] nella regione Piemonte, per l'utilizzo di diversi mezzi di trasporto con un solo biglietto, il cui credito è eroso in base ai servizi utilizzati. In questo modo il cittadino ha una carta, ricaricabile o collegata al proprio circuito bancario, con la quale può usufruire di diversi servizi di mobilità come treni, autobus urbani e suburbani, parcheggi e, in un futuro prossimo, *car sharing* o *bike sharing*.

Questa integrazione di mezzi di pagamento per l'utilizzo di diversi servizi di mobilità evidenzia come nella *Smart City* la multimodalità dei trasporti sia la chiave per gli spostamenti urbani da parte del cittadino.

Nell'ambito applicativo dell'*acqua*, si trovano servizi integrati per la gestione ottimale delle acque in ambito urbano. La rete di drenaggio urbano rappresenta un'infrastruttura fondamentale per la città; su di essa insistono le acque meteoriche derivanti dalle aree direttamente connesse, come le coperture degli edifici e le aree asfaltate. Il progetto "Servizio di gestione integrata e sostenibile del ciclo acqua-energia nei sistemi di drenaggio urbano" [19] intende realizzare un servizio per il controllo degli afflussi alla rete di drenaggio, il controllo del carico inquinante generato, i benefici termo-energetici e le potenzialità di riuso delle acque meteoriche. Il progetto *UrbanWater* [20] mira a creare una piattaforma *web-based* per la gestione delle acque urbane; la piattaforma incorpora delle soluzioni di *metering* avanzato e una misura di consumo dell'acqua in tempo reale. La finalità è quella di agevolare il *decision maker* nelle sue decisioni manageriali sulla gestione di questo bene prezioso, integrando un modello che considera la disponibilità delle acque in funzione delle previsioni meteorologiche, delle riserve idriche in determinate zone geografiche e della stima di domanda da parte degli utenti.

Altri esempi di servizi a vero valore aggiunto per il cittadino e per l'amministrazione pubblica sono quelli nell'ambito della *salute*. Un esempio è il servizio offerto dall'orologio da polso ADAMO (Autonomia al Domicilio Assistenza e Monitoraggio) [21], ideato da Fondazione Torino Wireless e Istituto Superiore Mario Boella e prodotto da Caretek Srl. L'orologio monitora 24 ore su 24 la salute del paziente, garantendo interventi rapidi ed efficaci in situazioni di emergenza, grazie all'integrazione con servizi di telesoccorso e telemonitoraggio. L'orologio misura costantemente alcuni parametri vitali del paziente che lo indossa come la temperatura cutanea, i movimenti della mano e del polso, la rimozione dal polso, parametri ambientali quali temperatura e luminosità e, con l'attivazione da parte dell'utente, il battito cardiaco. Tutti i dati sono inviati alla stazione che li elabora, riconosce situazioni anomale, quali cadute, immobilità sospette e stati febbrili e trasmette eventualmente un segnale di allarme alla centrale operativa che può prontamente intervenire.

Per quanto concerne l'ambito dell'*e-government* ci preme citare il servizio offerto da *OPML Visual Lab* [22], il laboratorio sulla rappresentazione grafica dei dati sull'occupazione realizzato dall'Istituto Superiore Mario Boella e dall'Osservatorio sul mercato del Lavoro della Provincia di Torino. *OPML Visual Lab* è uno spazio comune nel quale esperti di grafica interattiva, di *open data* e di dinamiche occupazionali hanno interagito per realizzare un primo set di quattro strumenti che consentono a singoli cittadini, operatori professionali del mercato del lavoro, decisori pubblici e ricercatori di accedere in maniera intuitiva al patrimonio di conoscenza proveniente dalle comunicazioni obbligatorie sui rapporti di lavoro. *OPML Visual Lab* è quindi al contempo uno strumento di valorizzazione della *Public Sector Information*, di supporto alle decisioni e di *empowerment* dei cittadini, tre elementi chiave della *Smart City* del futuro.

6. Conclusioni

L'analisi delle tecnologie, dell'architettura di riferimento e dei servizi *smart* nei campi dell'energia, dell'edilizia, della mobilità, dell'acqua, della salute e dell'*e-government* ha evidenziato come i diversi ambiti applicativi della *Smart City* siano strettamente correlati tra loro. Per questo motivo una città che vuole essere *smart* deve essere capace di adottare un approccio integrato alla pianificazione e alla gestione del contesto urbano, abbandonando la cultura dei silos che caratterizza purtroppo molte amministrazioni. In questa sfida le Tecnologie dell'informazione e della comunicazione si rivelano uno strumento fondamentale, capace di rappresentare, analizzare e misurare in modo nuovo fenomeni complessi, ricavando intelligenza dalla correlazione di informazioni apparentemente indipendenti.

Riferimenti

- [1] Hollands, R. G. (2008). "Will the real smart city please stand up?", *City: analysis of urban trends, culture, theory, policy, action*, 12, 3, 303-320.
- [2] European Commission (2009). "Investing in the Development of Low Carbon Technologies".
- [3] <http://ec.europa.eu/eip/smartcities/> (ultimo accesso marzo 2014)
- [4] IEEE 802.15.1 (2005). "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks".
- [5] IEEE 802.11 (2012). "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications".
- [6] Raghu, G., Fan, Y., Hui, L. (2011). "Mobile Crowdsensing: current state and future challenges". *IEEE Communications Magazine*, 49, 11, 32-39.
- [7] Kaplan, E.D., Hegarty, C.J. (2006). *Understanding GPS: principles and applications*, House Inc.
- [8] http://www.esa.int/Our_Activities/Navigation/The_future_-_Galileo/What_is_Galileo (ultimo accesso marzo 2014)
- [9] International Telecommunication Union (1999). "Global standard for 3G".
- [10] 3rd Generation Partnership Project (2011). "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Overall Description".
- [11] IEEE 1901 (2011). "Broadband Over Power Line Standard".
- [12] National Institute for Standards and Technology (2010). "Cloud Architecture Reference Models: A Survey".
- [13] Agenzia per l'Italia Digitale (2012). "Raccomandazioni alla Pubblica Amministrazione per la definizione e sviluppo di un modello tecnologico di riferimento per le Smart City".
- [14] <http://www.provincia.torino.gov.it/ambiente/energia/progetti/Enercloud/index> (ultimo accesso marzo 2014)

- [15] <http://www.energy-home.it/SitePages/Home.aspx> (ultimo accesso marzo 2014)
- [16] <http://www.grandlyon.com/Lyon-Smart-Community.5518.0.html> (ultimo accesso marzo 2014)
- [17] <http://www.car2go.com> (ultimo accesso marzo 2014)
- [18] <http://bip.piemonte.it> (ultimo accesso marzo 2014)
- [19] <http://www.ponrec.it/open-data/progetti/scheda-progetto?ProgettoID=5224#Descrizione> (ultimo accesso marzo 2014)
- [20] <http://urbanwater-ict.eu/the-solution-2/> (ultimo accesso marzo 2014)
- [21] <http://www.adamo-vita.it> (ultimo accesso marzo 2014)
- [22] <http://www.opmltorino.it> (ultimo accesso marzo 2014)

Biografie

Francesco Ferrero è Responsabile del Programma Strategico *Smart City* dell'Istituto Superiore Mario Boella (http://www.ismb.it/smart_city). Il programma intende fondere le tradizionali competenze dell'Istituto con nuove competenze non tecnologiche (es. urbanistiche e socio-economiche) e promuove l'uso dell'ICT come fattore abilitante di un approccio integrato alla pianificazione e all'amministrazione delle città, con l'obiettivo della triplice sostenibilità (ambientale, sociale e finanziaria). Francesco è membro dell'*International Steering Committee* della *World Smart Capital Initiative* e del gruppo di lavoro *ICT4Smart Cities* della *Smart Cities Stakeholder Platform* europea.

E-mail: fferrero@ismb.it

Andrea Pacifici si è laureato in Ingegneria delle Telecomunicazioni presso il Politecnico di Torino nel 2007. Ha lavorato per sei mesi presso l'Istituto Superiore Mario Boella (ISMB) occupandosi dello *scouting* di nuove tecnologie nel settore bancario, smaterializzazione della carta, riconoscimento della firma su *tablet* e timbro digitale. Da Ottobre 2007 a Giugno 2013 ha lavorato nell'IT Finance di FGA come *project manager*. A Luglio 2013 è rientrato in ISMB nel Programma Strategico *Smart City*, dove si occupa di mobilità urbana, *smart parking* e *car sharing* oltre ad un continuo *scouting* delle tecnologie per veicoli elettrici di nuova generazione.

E-mail: pacifici@ismb.it

Andrea Vesco si è laureato in Ingegneria delle Telecomunicazioni nel 2003 ed ha conseguito il dottorato di ricerca in Ingegneria Informatica e dei Sistemi nel 2009 presso il Politecnico di Torino. Nel 2010 ha iniziato il suo percorso presso il *Networking Lab* dell'Istituto Superiore Mario Boella (ISMB). Le sue ricerche si sono concentrate sulla qualità del servizio in reti a pacchetto, reti di accesso wireless e network-on-chip. Dopo dieci anni di esperienza e diverse pubblicazioni nel settore del *networking*, nel 2013, ha avuto l'opportunità di concentrare le sue ricerche sul tema delle *Smart City* con particolare attenzione alla mobilità sostenibile e all'efficienza energetica. Andrea Vesco è membro della *Smart City Stakeholder Platform* europea.

E-mail: vesco@ismb.it

Non solo un traduttore di formule: Il FORTRAN e i suoi 60 anni

C. Petrocelli

Sommario. Il 10 novembre del 1954, IBM divulgò un rapporto intitolato “Specifications for the IBM Mathematical FORMula TRANslating System, FORTRAN” che segnò la nascita di un linguaggio di programmazione ancora oggi utilizzato. A distanza di 60 anni si ripercorrono le tappe della sua diffusione.

Keywords: John Warner Backus, FORTRAN, IBM, History of programming languages

1. Gli anni '50: Programmazione sinonimo di pseudocodifica

Il periodo compreso tra la fine della II Guerra Mondiale e la metà degli anni Cinquanta, fu caratterizzato fondamentalmente dalla nascita dei primi calcolatori elettronici a programma memorizzato, secondo i modelli proposti da Turing e von Neumann, per i quali la memoria della macchina avrebbe dovuto essere utilizzata per le elaborazioni delle istruzioni dei programmi e contemporaneamente dei dati in esse coinvolti. Ma, in quegli anni, la *programmazione* era per lo più un'arte misteriosa, una materia arcaica che solo in pochi avevano svelato facendo uso di tecniche ed inventiva che certo non garantivano leggibilità ed adattabilità dei programmi stessi. Questi ultimi erano infatti funzionanti solo ed esclusivamente sulla macchina per cui erano stati codificati e venivano impostati secondo il criterio personale del programmatore il quale, spesso, doveva far uso della 'soluzione del momento' per la produzione di un codice efficiente.

L'inevitabile ridondanza che ne era conseguita, aveva cominciato a far riflettere sulla necessità di definire e formalizzare degli standard al fine di ottimizzare gli sforzi di ogni programmatore, consentendogli di concentrarsi solo ed

esclusivamente sul problema da risolvere. Bisognava pertanto, in primo luogo semplificarne il processo di scrittura utilizzando, ad esempio, un linguaggio di più alto livello rispetto al codice macchina e contemporaneamente rendere il codice 'usabile', cioè standardizzarlo per altre macchine o applicazioni.

In questo clima, tra il 1952 ed il 1956, iniziarono a circolare espressioni quali *pseudocodice*, *codice automatico*, *programmazione automatica*, con le quali si tendeva a far riferimento a quelle istruzioni che era possibile 'automaticamente' far tradurre alla macchina nella corrispondente forma binaria, permettendo al programmatore di svincolarsi dall'architettura di riferimento.

Quando John Warner Backus (*Riquadro 1*), nella primavera del 1949, prossimo alla laurea in matematica, si trovò a visitare un centro di calcolo dell'IBM nel quale erano in corso le procedure di test di uno dei primi calcolatori elettronici, il Selective Sequence Electronic Calculator¹ (SSEC), si rese immediatamente conto che, se la macchina fosse stata dotata di un sistema in grado di permettere ai programmatori di ridurre il tempo impiegato per l'impostazione di un problema, sarebbe notevolmente migliorata sia la qualità che l'affidabilità del software generato. A seguito della sua proposta, i dirigenti del New York Scientific Computing Service dell'IBM, diedero il via alle sperimentazioni per lo *Speedcoding System* e nel gennaio del 1953 ne affidarono la direzione generale a John Sheldon e la supervisione tecnica proprio a John Backus.

Lo *Speedcoding* fu il nome dato ad un sistema inteso come combinazione integrata di un'ampia gamma di computer digitali ed un metodo mediante il quale questi computer potevano essere programmati per risolvere problemi scientifici e di ingegneria. La famiglia presa come riferimento fu quella del 701², una macchina con memoria a tubi a vuoto³, spesso definita come una «memoria elettrostatica»: per garantire prestazioni veloci ed affidabili, si impose che lavorasse internamente utilizzando il sistema numerico binario, ma coloro che

¹ Lo SSEC non era un calcolatore nel senso moderno del termine: non aveva memoria per l'immagazzinamento dei dati ed i programmi venivano inseriti utilizzando un nastro di carta perforato.

² La prima linea di produzione dell'IBM riguardò il «modello 701» che fu disegnato da Nathaniel Rochester nel 1952 e fu progettato per le esigenze di calcolo del Pentagono, impegnato nella guerra in Corea: per questo motivo fu definito *defence calculator*. Ne vennero prodotti ed installati solamente diciannove esemplari nell'arco dei suoi tre anni di vita.

³ «Vacuum tube»: dispositivo di amplificazione o di interruzione che opera controllando un fascio di elettroni nel vuoto. Nella forma base, consiste di un filamento costituito da una resistenza riscaldata elettricamente, un catodo collegato al filamento, una griglia ed una piastra. Il filamento produce elettroni che vengono «raccolti» dal catodo ed inviati in direzione della piastra. Se la griglia è negativa rispetto al catodo, gli elettroni vengono respinti e solo un numero molto basso di essi raggiunge la piastra. Se la griglia è positiva, gli elettroni passano attraverso la piastra e la corrente fluisce nel circuito collegato all'uscita. Il tubo funziona come dispositivo di interruzione, con una debole corrente nella griglia che attiva e disattiva una corrente relativamente elevata nella piastra. Se la tensione sulla griglia è contenuta in una banda tale che il suo livello più alto non causa il passaggio di tutta la corrente nella piastra, il tubo funziona come un amplificatore: più alto è il segnale di entrata, più alto è il segnale di uscita.

usavano lo *Speedcoding* non risentivano in alcun modo di ciò poiché tutti i codici venivano introdotti e stampati solo in decimale.

I programmi potevano essere scritti ed immessi nel computer in vari modi, ma solitamente le istruzioni venivano lette su pacchetti di schede perforate. Per istruire la macchina ad eseguire calcoli, apposite unità di registrazione andavano ad incidere su nastri magnetici la sequenza del codice di calcolo, mentre le schede, su cui erano rappresentati dati ed istruzioni, venivano piazzate nel corrispondente lettore. La sola pressione del bottone di avviamento della macchina faceva sì che il programma venisse memorizzato insieme ai dati ed alle istruzioni di inizio della procedura di calcolo, rispettando rigidamente i canoni e le regole imposte dal modello di von Neumann. Da quel momento in poi, le operazioni che il computer eseguiva erano completamente automatiche e tutti i componenti della macchina erano sotto il controllo del programma.

2. Le prime fasi del progetto FORTRAN

Forte dell'esperienza messa a punto con lo *Speedcoding*, Backus fu stimolato a migliorare ulteriormente il sistema di rappresentazione delle istruzioni, cercando di portare la scrittura delle stesse ad un livello di astrazione più alto, un livello espressivo che avrebbe dovuto allontanare il linguaggio dai dettagli minuziosi imposti dalle operazioni svolte dai circuiti digitali, per avvicinarsi alle capacità espressive umane.

L'intuizione di Backus rappresenta probabilmente il punto di svolta nello sviluppo dei linguaggi di programmazione e questo verosimilmente perché, per la prima volta, egli sposta il confine tra l'uomo e la macchina nella direzione dell'essere umano: getta le fondamenta della programmazione ad un livello più alto, livello che andrà crescendo e si perfezionerà con la stessa rapidità con cui si sono percepite le mutazioni tecnologiche delle architetture di calcolo.

Backus presentò al suo manager IBM, Cuthbert C. Hurd, la sua idea di progetto per l'implementazione di un "sistema di programmazione avanzato" che fosse in grado di ovviare alle difficoltà di scrittura di codice a basso livello dell'IBM 704 ed ai problemi di correttezza e di efficienza dei programmi relativi.

Il *Programming System Department* dell'IBM con a capo Cuthbert Hurd, Charles De Carlo e John McPherson, valutarono positivamente le sue proposte tanto che, per i cinque anni successivi al 1953, gli assegnarono i finanziamenti necessari senza domandare mai nulla su come procedeva il lavoro.

Per qualche tempo, a questo progetto non venne dato neanche un nome e alla fine Backus, non sapendo come identificare il linguaggio che stava progettando, prese l'abitudine di riferirsi ad esso come ad un sistema di "traduzione di formule" (*FORmula TRANslating system*). Nel novembre del 1954, il «*Fortran team project*⁴» divulgò un rapporto intitolato *Specifications for the IBM*

⁴ Il gruppo, diretto da Backus, era composto anche da R.J. Beeber, S. Best, R. Goldberg, H.R. Herrick, R.A. Huges, L.B. Mitchell, R.A. Nelson, D. Sayre, P.B. Sheridan, H. Stern ed I. Ziller. Roy Nutt, dipendente della United Aircraft Corporation, si unì in seguito al gruppo e si dedicò principalmente all'implementazione di tutte le specifiche di input/output del linguaggio. Si veda [22] e [5].

Mathematical FORMula TRANslating System, FORTRAN⁵ (Figura 1) [15], contenente l'idea progettuale di una prima versione del Fortran, identificata con il nome *FORTRAN 0*. Il rapporto evidenziava come i sistemi precedenti offrivano o un codice semplice, a discapito di una esecuzione lenta, oppure una codifica più complessa per esecuzioni veloci. Con il *Fortran* si mirava invece a raggiungere una situazione di compromesso [21, p. 241]:

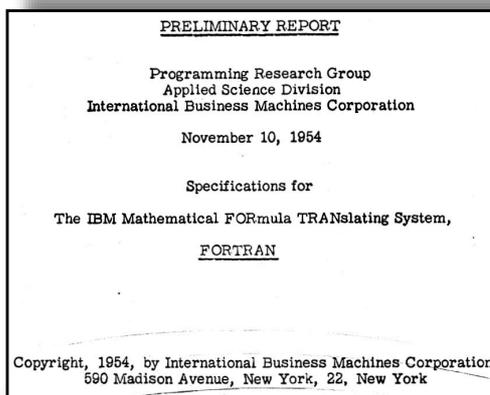


Figura 1
Primo Report sul FORTRAN

«L'IBM Mathematical *FORMula TRANslating System*, o brevemente *Fortran*, comprenderà un grosso insieme di programmi che abiliteranno l'IBM 704 ad accettare una concisa formulazione di un problema in termini di notazione matematica e a produrre automaticamente un programma veloce per la sua risoluzione.» [21 e 32 p.143]

L'IBM 704 era una macchina dotata di una unità di elaborazione, una unità magnetica centrale di memorizzazione, tamburi e nastri magnetici, un punzonatore ed un lettore di schede, una stampante ed una *console* composta anche da un monitor⁶. Questo calcolatore ad alta velocità veniva controllato da un programma, residente nella sua memoria interna, composto da istruzioni ad un solo indirizzo. La macchina supportava istruzioni in virgola mobile e consentiva anche l'indicizzazione delle operazioni. Fu disegnata per eseguire elaborazioni rapide, anche con elevate quantità di dati, richieste nella ricerca scientifica, per l'ingegneria, e, in generale, per la produzione di programmi finanziari, industriali e militari. [26] e [6]

Sebbene i tempi per il completamento del progetto furono stimati in circa sei mesi, non ci fu notizia del linguaggio fino a che, nell'ottobre del 1956, venne pubblicato il primo manuale di riferimento [17] nel quale l'IBM precisò che il compilatore sarebbe stato distribuito alla fine dello stesso anno. In realtà, questo impegno non venne mantenuto perché il Dipartimento di Ricerca lo rese disponibile solo nell'aprile del 1957⁷.

⁵ Si noti la sigla IBM nel titolo del rapporto. In effetti il FORTRAN sarà sempre considerato un prodotto IBM e negli anni a venire il successo di IBM andrà di pari passo con quello del FORTRAN.

⁶ Ogni registro di memoria era costituito da un gruppo di 36 nuclei magnetici. Vi erano soltanto due unità di memoria, ognuna delle quali era costituita da 4096 registri. Per ampliare lo spazio disponibile la macchina era stata dotata di memorie ausiliarie costruite mediante dei tubi magnetici. Le «parole», o unità fondamentali di informazione, erano costituite da 36 numeri binari (bit). Per una descrizione più approfondita, si veda [16].

⁷ «Il tempo richiesto e gli sforzi necessari superarono la stima iniziale – situazione comune per chi lavora in questo campo!» [12] e [9]

Il linguaggio non ebbe un immediato successo, al contrario, non destò neppure una minima curiosità. Probabilmente questo era dovuto al fatto che non ne venne mai data una completa ed esplicita definizione, tranne quella informale, che si poté solo desumere dai vari manuali di riferimento. Tuttavia, un articolo scritto da Grace E. Mitchell e pubblicato dall'IBM nel 1957⁸ [24], fu il primo documento che incuriosì la comunità scientifica, tanto che da quel momento ci fu una richiesta sempre crescente nei confronti del linguaggio appena sviluppato. Lo scopo dell'articolo della Mitchell fu quello di illustrare il *Fortran* a scienziati, ingegneri ed a tutti coloro che non erano «specialisti» in materia informatica, fornendo una descrizione «passo passo» di ogni sua singola parte.

Facendo un paragone tra le specifiche preliminari del linguaggio, già pubblicate nel 1954, e quelle evidenziate nel manuale rilasciato nell'ottobre del '56, ci si accorge che effettivamente ci furono pochissime differenze tra una versione e l'altra, sebbene, inevitabilmente, fu comunque praticata qualche piccola variazione⁹. Malgrado fossero stati realizzati, nel corso degli anni successivi al 1954, molti cambiamenti ed aggiunte al linguaggio, e di conseguenza ai relativi manuali, la struttura di base e l'intento dei progettisti furono soltanto ampliati piuttosto che alterati o modificati. William P. Heising dell'IBM Corporation, nel 1963, cioè circa dieci anni dopo la diffusione della prima versione del software, espresse le seguenti importanti valutazioni:

«Il completamento della versione originale del *Fortran* per il 704 rappresenta un significativo successo per la storia della programmazione ed ha indubbiamente avuto influenza per gli sviluppi successivi. Termini comuni quali *programma sorgente*, *programma oggetto* e *macchina oggetto* sono passati nel vocabolario della programmazione grazie all'uso che se ne è fatto nel manuale originale del *Fortran*.» [12]

3. La potenza di un nuovo linguaggio: il FORTRAN I

La prima versione del linguaggio subì alcuni cambiamenti durante il periodo di progettazione ed implementazione, periodo che cominciò nel gennaio del 1955 e continuò sino a che non fu distribuito il compilatore nell'aprile del 1957 [33, p. 45]. Queste variazioni progettuali spinsero il «gruppo» ad identificare la nuova versione con il nome *FORTRAN I*.

⁸ Il testo era suddiviso in tre sezioni, ognuna delle quali descriveva un sottoinsieme del linguaggio, accompagnato da numerosi esempi. La prima edizione apparve nell'autunno del 1957; una edizione rivista e corretta venne pubblicata nel mese di marzo del 1958.

⁹ Le specifiche preliminari del linguaggio indicavano che erano permesse soltanto funzioni con nomi di una o due lettere; nella successiva implementazione risultò invece possibile definire nomi di funzioni che iniziavano con un carattere alfabetico e che potevano essere seguiti da due o più lettere. Oltre a questo, dal rapporto iniziale, venne eliminata la possibilità di mettere insieme numeri ed espressioni (questa opportunità sarà presente soltanto in seguito, con l'ALGOL, ma non ce ne sarà traccia nelle versioni successive del Fortran), quella di utilizzare nell'*if* il paragone tra due variabili anziché soltanto un confronto con «zero», e quella di dichiarare esplicitamente il «range» dell'istruzione *do* come una coppia di istruzioni. [32, p. 144].

«[...] fu probabilmente il più complesso linguaggio di programmazione mai prodotto fino a quel tempo ed il fatto che il sistema comprendeva circa 25.000 linee di codice può indicare una misura della sua vastità.» [12]

Per il *FORTRAN I* vennero previste istruzioni di input/output, nomi di variabili di uno o due caratteri, funzioni definite utilizzando tre o più caratteri, espressioni ricorsive, sottoprocedure definibili dal programmatore, formule aritmetiche, l'istruzione di selezione IF e l'istruzione ricorsiva DO. [5, p.71]

L'istruzione logica IF richiedeva l'uso di operatori relazionali del tipo «>» oppure «<», ma il modello 704 non contemplava questi caratteri, per cui gli operatori di questo tipo vennero rimpiazzati utilizzando una «selezione aritmetica» all'interno dell'istruzione IF, così espressa:

IF (espressione aritmetica) N1, N2, N3

dove N1, N2, ed N3 rappresentavano tre istruzioni. Se il valore dell'espressione era negativo, il salto portava all'istruzione etichettata N1; se era zero si saltava ad N2; se era maggiore di zero ad N3. Questa tipologia di rappresentazione è ancora presente nelle versioni più moderne del linguaggio.

Analogamente, per l'istruzione iterativa DO, la sua codifica fu:

DO N1 variabile = primo valore, ultimo valore

dove N1 indicava l'istruzione terminale del ciclo, la «variabile» rappresentava un controllo, mentre «primo valore» e «ultimo valore» (che potevano essere variabili o costanti intere positive) segnalavano, rispettivamente, il parametro iniziale e quello finale del ciclo¹⁰.

Per il *FORTRAN I* non furono previste istruzioni di tipizzazione dei dati: le variabili che avevano nome che cominciava con le lettere I, J, K, L, M ed N furono classificate come variabili di tipo intero, tutte le altre come variabili reali in «virgola mobile». In aggiunta, venne anche incluso un programma di «editing» per consentire eventuali modifiche sul codice. Infine, uno dei nastri magnetici utilizzato per la compilazione venne invece adoperato solo ed esclusivamente come strumento per l'input dei programmi.

La progettazione del linguaggio venne organizzata per sezioni:

«L'intero progetto fu portato avanti grazie ad una stretta collaborazione tra gruppi autonomi e separati costituiti da una, due o, al massimo, tre persone; ogni gruppo era responsabile di una *sezione*: si sviluppavano e concordavano gradualmente le specifiche di input ed output con gli altri gruppi per delineare precisamente il confine tra le varie sezioni.» ([5, p.73])

Nella prima sezione fu prevista una lettura dell'intero codice del programma sorgente e, di conseguenza, la compilazione delle sole istruzioni che potevano

¹⁰ Ad esempio, l'esecuzione dell'istruzione «DO n1 i = k1, k2», faceva sì che il gruppo di istruzioni, detto gruppo DO, a partire da quella immediatamente successiva fino a quella contraddistinta da n1, veniva ripetuto più volte. La prima volta, alla variabile «i», veniva assegnato il valore k1; successivamente si incrementava «i» di 1 fino a che non si raggiungeva o superava k2; non appena tale condizione si verificava, il controllo del programma passava all'istruzione immediatamente successiva ad n1. [33, pp.61-62]

tradursi a questo livello con il successivo salvataggio, in apposite tabelle, di quelle non compilate¹¹. Proprio per questa sua impostazione di compilazione, il *FORTRAN I* venne definito ad «un solo passo», nel senso che il compilatore «vedeva» il programma sorgente solamente una volta. Harlan Herrick fu il responsabile della creazione della maggior parte delle tabelle, Peter Sheridan della compilazione di tutte le espressioni aritmetiche¹² e Roy Nutt della compilazione di tutte le istruzioni di input/output.

Per la seconda sezione si pensò all'analisi delle istruzioni *DO* ed alla generazione del codice corrispondente alle variabili indicizzate, procedendo poi con l'esame delle tabelle prodotte in «sezione 1», anche al fine di ottimizzare gli accessi alla memoria. L'architettura di questa parte del compilatore e tutte le tecniche utilizzate, furono ideate da Robert Nelson ed Irving Ziller i quali progettaron e programmarono l'intera sezione.

La «sezione 3» fu progettata solo ed esclusivamente per convertire l'output delle precedenti nella forma richiesta per le analisi successive: il risultato non fu altro che un *merge* degli output delle sezioni precedenti in un unico programma scritto in codice macchina del 704. Il progetto e la programmazione di questa sezione furono seguiti da Richard Goldberg, un matematico che vi ci lavorò a partire dal novembre del 1955.

Qualche mese dopo, Lois Mitchell Haibt si unì al gruppo del *Fortran* per progettare e programmare la sezione numero 4. In questa fase venne anche contemplata l'analisi statistica della frequenza di esecuzione delle parti del programma sorgente attraverso una sua simulazione. Il programma venne diviso in «blocchi base», blocchi non contenenti istruzioni di salto, e su di esso venne eseguita un'analisi per determinare il percorso e la frequenza di esecuzione degli stessi blocchi (con l'istruzione *FREQUENCY*) per farne un uso più mirato nella sezione successiva¹³. Infine, per minimizzare il tempo necessario per il caricamento e la memorizzazione dei dati e delle istruzioni nei registri di indirizzamento, si pensò di raccogliere informazioni circa il loro uso e di far corrispondere ai registri reali alcuni registri simbolici. [1]

Nella quinta sezione sarebbe poi avvenuta un'ulteriore trasformazione del programma: un codice, che faceva riferimento ad un numero illimitato di registri indice, veniva tradotto in una versione corrispondente facente riferimento ai soli tre registri del 704. Questa sezione fu interamente progettata e programmata da Sheldon Best, che fu «prestato» al gruppo da Charles Adams del Digital Computer Laboratory del MIT. A partire dal 1955, Best disegnò una delle più intricate e complesse sezioni del compilatore ritenuta tra le più solide, innovative

¹¹ Il traduttore creava innanzitutto una sequenza di istruzioni aritmetiche dalle quali venivano eliminati i calcoli ridondanti (infatti il termine «*common subexpression elimination*» apparve di frequente nelle prime documentazioni). [1]

¹² Peter Sheridan ideò ed implementò un numero di trasformazioni ottimali sulle espressioni che vennero utilizzate anche nelle versioni successive del Fortran. [34]

¹³ Per un approfondimento sull'uso di queste tecniche si veda [7].

e significative per le tecniche di gestione degli indirizzi e dei registri di memoria¹⁴.

Nell'ultima sezione, la sesta, si assemblava il programma finale in un programma binario rilocabile; in questa fase veniva prodotta una mappa del programma e dei dati, sunto dell'output in *Fortran*. Roy Nutt, che disegnò e programmò l'intera sezione, scrisse anche tutti i programmi per la gestione dell'I/O.

Dalla primavera del 1956 ai primi mesi del 1957 si effettuarono i test di controllo:

«[...] fu un periodo molto emozionante; spesso rimanevamo attoniti nel verificare ciò che il compilatore eseguiva ed i cambiamenti che portavano alla definizione di un programma oggetto efficiente.» [5, p.74]

Nell'estate del 1956, mentre sembrava vicino il completamento del progetto, si paventò, per la prima volta, un problema che ancora oggi è tenuto in particolare conto: la *stesura della documentazione*. È questo probabilmente l'ulteriore salto di qualità e l'importante tassello che si aggiunge nella direzione della standardizzazione dei linguaggi: la programmazione non è più il semplice mettere insieme istruzioni di codice sorgente ma si identifica con processi più complessi tra cui quello di fornire una documentazione che ne descriva non soltanto i requisiti di sistema ma anche un manuale di utilizzo per l'abbattimento dei tempi di comprensione del software.

Il 15 ottobre del 1956, con una copertina piuttosto grossolana, scritto completamente a mano, con margini molto ampi ed un numero di pagine abbastanza ridotto (soltanto 51¹⁵), fu distribuito il primo *Programmer's Reference Manual* (Figura 2) [17], che rappresentò, per molto tempo, l'unico esempio di manuale di linguaggio di programmazione.

Una delle caratteristiche più rilevanti del *FORTRAN I* non fu descritta nel manuale di riferimento, poiché venne aggiunta al sistema dopo la sua pubblicazione. Questa opzione forniva la possibilità di definire una funzione nel programma sorgente mediante l'uso di una istruzione specifica. Questa istruzione doveva necessariamente precedere il resto del programma e si presentava come un'istruzione di assegnazione che riportava, nella sua parte sinistra, la definizione del nome della funzione e dei suoi argomenti, e, nella parte destra, un'espressione coinvolgente tali argomenti.

¹⁴ Richard Goldberg descrive dettagliatamente questa sezione in un articolo pubblicato nel 1984. La sua familiarità con le tecniche ivi utilizzate ha una spiegazione nel fatto che, insieme a David Sayre, ne esaminò e testò il funzionamento. [10]

¹⁵ La descrizione del linguaggio occupava 21 pagine, 11 erano dedicate alle istruzioni di input/output e le restanti contenevano esempi.

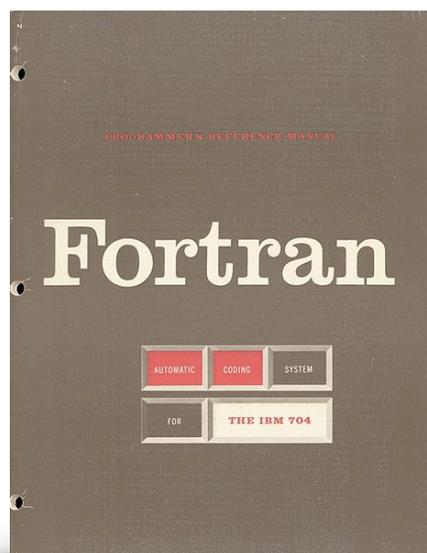


Figura 2
Programmer's Reference Manual
(1956)

L'IBM commercializzò il 704 corredato del compilatore *Fortran* di cui distribuì una versione aggiornata del suo manuale di riferimento:

«Quando distribuimmo il sistema, consegnammo anche un *Preliminary Operator's Manual* datato 8 aprile 1957. Esso conteneva la lista completa dei simboli dell'intero compilatore più altre informazioni sul sistema e sulla diagnostica¹⁶.» [18]

4. FORTRAN II

L'utilizzo del *FORTRAN I* evidenziò una serie di piccoli errori: fu un periodo difficile, sia per gli utenti che per gli autori, che terminò soltanto quando questi errori, almeno i più frequenti, furono riconosciuti e corretti. Le modifiche fatte sul linguaggio portarono alla stesura della una nuova versione del compilatore *FORTRAN II*, che fu ufficialmente rilasciata dal Programming System Department dell'IBM nel giugno del 1958. [12]

In questa versione del linguaggio vennero coperti molti dei «buchi» della versione precedente¹⁷ ed aggiunte alcune significative funzionalità, la più importante delle quali fu la compilazione indipendente di ogni singola sottoprocedura o funzione. Senza questa possibilità, per ogni cambiamento apportato ad un programma sarebbe stato necessario fare una intera ricompilazione.

Il *FORTRAN II* fu per la maggior parte progettato da Backus, Nelson e Ziller. Mitchell, con la collaborazione di Benyce Brady e LeRoy May, si occupò della scrittura del nuovo codice¹⁸.

Una cosa interessante da segnalare è che né l'IBM, né alcun membro del gruppo che sviluppò questa nuova versione del compilatore, pubblicarono mai un manuale di riferimento ufficiale nel quale si faceva uso esplicito della parola «*Fortran II*». In verità questa versione fu distribuita così velocemente rispetto alla precedente che non si ebbe il tempo di valutarne le migliorie. [30, p.148]

5. FORTRAN III

Durante la fase di sviluppo del *FORTRAN II*, Irving Ziller cominciò a pensare ad un sistema di compilazione più completo ed avanzato rispetto alla versione precedente.

Nella terza versione del *Fortran*, infatti, Ziller aveva previsto la possibilità di scrivere le istruzioni del linguaggio unitamente a quelle del codice macchina

¹⁶ Il manuale era composto da cinque pagine contenenti istruzioni generali e 32 pagine per la gestione degli errori di terminazione di compilazione. Alcuni dei messaggi che il compilatore inviava in caso di errore, citavano: «errore del programma sorgente, arrestare la macchina, correggere la formula in questione e far ripartire l'elaborazione del programma», oppure «parentesi insufficienti o ridondanti nella formula aritmetica o in una formula di tipo IF».

¹⁷ «[...] alcuni difetti progettuali del sistema, che avevamo riscontrato ma che non avevamo avuto il tempo di correggere, furono costantemente sottoposti alla nostra attenzione.» [5, p.76]

¹⁸ Per un approfondimento sulle caratteristiche tecniche del linguaggio FORTRAN II si veda [32].

dell'IBM 704 (le istruzioni simboliche del 704 potevano avere variabili *Fortran* come «indirizzi»). La versione «III» permetteva anche l'uso di istruzioni algebriche booleane, il trattamento delle informazioni algebriche, la capacità di passare sottoprogrammi come argomenti e l'inclusione di istruzioni in linguaggio macchina «in linea».

Anche questa *release* fu solamente una versione di “passaggio”: il sistema fu implementato dallo stesso Ziller con la collaborazione di Nelson e Nutt. Non fu mai distribuito ma, per tutto il 1959 ed i primi mesi del '60, se ne resero disponibili solo 20 installazioni per uso interno della società, accompagnate da una breve documentazione. [19] Queste venti versioni divennero operative per gli hardware IBM 709/7090¹⁹, per i quali si utilizzarono solo le istruzioni compatibili con il 704²⁰.

Tra il 1958 ed il 1959 il «progetto *Fortran*» venne affidato all'Applied Programming Department. La responsabilità del progetto non fu più di John Backus, ma di William Heising, direttore della sezione che curava la parte della programmazione dell'IBM 7090. Data la sua ovvia conoscenza del codice della macchina, l'IBM gli affidò anche il progetto relativo allo sviluppo del *Fortran* per quel modello. Seguirono gli anni di transizione tra il *FORTTRAN II* ed il *FORTTRAN IV*. [14]

6. Gli anni '60: la standardizzazione del linguaggio

Nel 1962 l'IBM pubblicò un bollettino che anticipava la distribuzione di un linguaggio, che prenderà poi il nome di *FORTTRAN IV*, «indipendente dalla macchina», un linguaggio, cioè, che «non doveva ricordare il codice macchina».

Tutte le istruzioni proprie del *FORTTRAN II* che coinvolgevano, ad esempio, i nastri per l'input/output ed i tamburi di memoria del «704», vennero definitivamente eliminate rendendo il *FORTTRAN IV* incompatibile con la sua precedente versione «II».

Le funzionalità offerte da questa nuova versione, furono sostanzialmente legate al tipo di dati che si potevano utilizzare (LOGICAL, DOUBLE PRECISION, COMPLEX, REAL, INTEGER ED EXTERNAL), alle espressioni booleane implementate come argomento dell'operatore IF, alle funzioni e sottoprocedure aventi nomi passati come argomenti di altri sottoprogrammi ed, infine, alle nuove istruzioni DATA e BLOCK DATA.

L'IBM adottò definitivamente il *Fortran* come linguaggio di programmazione per le sue macchine, tanto che venne utilizzata una politica aziendale di vendita, suggerita per la verità da Backus già nel 1954, secondo la quale «ogni calcolatore IBM doveva essere dotato di un sistema simile al *FORTTRAN*». [21, 241-242] e [9] Difatti, nel 1958, il compilatore *FORTTRAN* venne rilasciato per l'IBM 709 e 650 e, nel 1960, per i modelli 1620 e 7070.

¹⁹ Il modello 709 aveva la stessa architettura dell'IBM 704 ma consentiva di effettuare molte più operazioni di input/output e aveva prestazioni migliori rispetto al modello precedente. L'IBM 7090 fu la versione «transistorizzata» del 709. [8]

²⁰ Ad esempio non fu più possibile utilizzare le istruzioni «in linea» del linguaggio macchina viste le differenze esistenti tra il modello 704 e il 709.

Per evitare che si venisse a creare una certa confusione tra le versioni diverse del *Fortran* e i computer su cui potevano “girare”, l’IBM pubblicò un manuale che includeva un elenco di istruzioni in corrispondenza delle quali veniva specificato il modello di macchina su cui potevano essere implementate. [20]

Furono questi gli anni di maggiore popolarità del *Fortran*: in aggiunta alle versioni «ufficiali», la maggior parte dei produttori di computer e quasi tutte le «software house» cominciarono a produrre la loro personale versione, alcune volte aggiungendovi delle peculiarità non presenti in quelle distribuite dall’IBM²¹ (vedi Riquadro 2).

Nel 1964, nel suo articolo *The Various FORTRANS*, H. Oswald indicò i risultati di una sua ricerca, segnalando l’esistenza di 43 differenti versioni del linguaggio²². L’esistenza di un così elevato numero di «dialetti», rese molto problematico eseguire il trasferimento di programmi da una macchina all’altra, tanto che, nel 1963, l’American Standards Association (ASA, ora conosciuta come American National Standards Association) avvertì la necessità di rilasciare il suo primo «Standard for a Programming Language» a cui venne attribuito semplicemente il nome *FORTRAN*²³.

Nel 1960 l’American Standards Association, Sectional Committee X3 for Computers and Information Processing, aveva già costituito una sottocommissione, la X3.4 Sectional Subcommittee, per determinare uno standard tra i linguaggi di programmazione comuni. Il 17 maggio del 1962, la X3.4 diede l’avvio ai lavori del gruppo X3.4.3-FORTRAN, per esaminare le proposte riguardanti la standardizzazione del linguaggio *Fortran*.

W.P. Heising ha reso pubblici i documenti ufficiali prodotti dal «Subcommittee» dai quali si desume che:

«L’obiettivo dell’X3.4.3 Working Group dell’ASA è produrre un documento, o una serie di documenti, che definiranno l’ASA Standard per il linguaggio *Fortran*. Il risultante linguaggio standard sarà chiaramente riferito a ciò che era chiamato *Fortran* nel passato. I criteri utilizzati per considerare e valutare gli elementi dei vari linguaggi dovranno comprendere (non in ordine di importanza):

- a. facilità d’uso
- b. compatibilità con le precedenti versioni
- c. estensione dell’applicazione
- d. possibilità di ampliamento
- e. facilità di implementazione corrispondente ad una compilazione ed una esecuzione efficienti.

²¹ Nel 1961 il compilatore fu reso disponibile anche per lo UNIVAC e, nello stesso anno, l’ATLAC, una imitazione del *Fortran*, fu utilizzata sul Philco 2000. [27]

²² La situazione era tale che gli utenti non erano affatto sicuri che un valore computato su una macchina sarebbe stato lo stesso se prodotto da una computazione su un’altra macchina. [28]

²³ [13], riportato integralmente anche in [32, pp. 154-566].

Il *FORTRAN Standard* faciliterà i trasferimenti da macchina a macchina dei programmi scritti in ASA Standard FORTRAN. Lo Standard servirà come documento di riferimento sia per gli utenti che vorranno raggiungere questo obiettivo, sia per i produttori che programmeranno il prodotto riferendosi ad esso.» [32, pp. 154-566]

L'ASA Working Group, alla fine dei suoi lavori, decise di produrre due standard, il *FORTRAN* e il *Basic FORTRAN*, corrispondenti, rispettivamente, al *FORTRAN IV* e al *FORTRAN II*. Questi standard furono approvati il 7 marzo del 1966 e divennero ufficialmente ASA Standards X3.9-1966 (*FORTRAN* [34]) e X3.10-1966 (*Basic FORTRAN* [36]).

7. Controversie, compromessi e modernizzazioni

L'evoluzione del *Fortran* continuò fino a che, nel 1978, venne rilasciata un'ulteriore versione standardizzata, il *FORTRAN 77* [2], che conservò la maggior parte delle caratteristiche del *FORTRAN IV* con l'aggiunta della gestione delle stringhe di caratteri, delle istruzioni di controllo dei cicli con operatori logici e della clausola opzionale *else* combinata con l'*IF*.

Il *FORTRAN 90* [3] è il nome dato all'ultima versione del *Fortran*. Estremamente differente dalla precedente, questa *release* includeva principalmente operazioni per la gestione degli «array²⁴» e nuove istruzioni di controllo²⁵. Con il *FORTRAN 90* si sviluppò una nuova concezione del linguaggio, secondo la quale da esso potevano essere eliminate tutte quelle istruzioni o caratteristiche ereditate dalle versioni precedenti, ritenute obsolete dal programmatore (ad esempio, la gestione aritmetica dell'*ELSE IF*).

Anche se ci sono stati molti cambiamenti e aggiunte nel corso degli anni, la struttura di base del linguaggio e gli intenti di sviluppo sono stati estesi piuttosto che alterati. Di conseguenza, gli obiettivi che si possono desumere già dal manuale originale relativi alle potenzialità e agli intenti del primo FORTRAN, sono chiari e limpidi oggi come sessanta anni fa, quando il primo rapporto fu scritto. Il completamento del linguaggio originale per il 704 costituisce dunque un risultato di estrema importanza nella storia della programmazione e ha indubbiamente influenzato gli sviluppi successivi.

In realtà, è assai difficile comprendere il successo globale del Fortran in un momento storico nel quale "il modo" in cui le macchine vengono utilizzate è radicalmente cambiato e gli obiettivi e la logica di programmazione si plasmano e si adeguano ad esigenze completamente diverse. Quello che sicuramente va detto è che gran parte del suo successo è dovuto al suo essere il primo

²⁴ *Allocati e deallocati dinamicamente in memoria mediante operatori appositi: è una svolta radicale dal primo Fortran che gestiva solamente dati statici.* [33, pp. 47-48].

²⁵ *L'istruzione case usata per istruzioni con selezione multipla, exit per uscire dal ciclo forzatamente e cycle per spostare il controllo all'inizio del ciclo.*

linguaggio ad alto livello ampiamente utilizzato²⁶. Esso rappresenta cioè la *lingua franca* del mondo dell'informatica, il linguaggio "di strada" che sopravvive e sopravviverà perché è parte straordinariamente utile di ciò che vi è di "vitale" nella programmazione.

Si possono valutare le grosse potenzialità solo se lo si connette strettamente al suo utilizzo e non propriamente alla sua tecnologia: il contributo più rilevante è sicuramente l'essere stato un linguaggio «usabile» su qualsiasi hardware e aver avuto una grammatica di riferimento molto semplice e di facile comprensione. [11]

Il limite più grande è forse stato quello di essere un linguaggio di programmazione estremamente efficace per la risoluzione di problemi scientifici numerici ma poco adattabile per altri usi: la formulazione di programmi relativi a quesiti di altra natura ne implicava, purtroppo, un uso non appropriato e poco elegante.

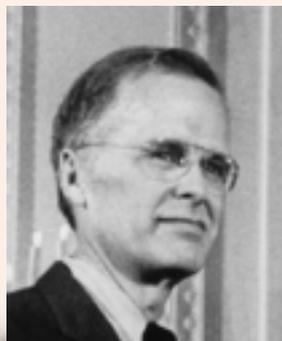
Emblematiche sono le parole di Jane Sammet che sintetizzano chiaramente le inutili pecche che si attribuiscono a questo linguaggio:

«[...] la maggior parte degli svantaggi del *Fortran* sono radicati nella sua stessa popolarità: le tante persone che lo adoperano avrebbero piacere ad utilizzarlo per ogni tipo di problema. Quando non soddisfa le loro necessità, ingiustamente, lo criticano.» [32, p.157]

²⁶ Per una completa descrizione dell'evoluzione e dello sviluppo dei linguaggi di programmazione si veda [29].

Riquadro 1 - John Warner Backus

Nato a Philadelphia nel 1924, la sua famiglia benestante gli permise di frequentare la prestigiosa Hill School di Pottstown – Pennsylvania. Non fu un buon allievo e gli anni trascorsi alla Hill School furono contrassegnati da una serie di fallimenti. Ciononostante, riuscì a diplomarsi nel 1942 e ad iscriversi alla University of Virginia, al corso di chimica, solo per seguire la volontà del padre. Anche in questo caso, non fu tra gli studenti più diligenti, al contrario, alla fine del secondo semestre, venne espulso dall'università.



J.W. Backus
1924 - 2007

Nel 1946 si trasferì a New York City dove prese in affitto un piccolo appartamento. Grande estimatore della buona musica, non potendosi permettere di acquistare un sistema Hi-Fi, si iscrisse ad una scuola di tecnico radiofonico per imparare a costruirne uno. Presso questa scuola ebbe l'occasione di aiutare un istruttore ad eseguire i calcoli matematici complessi per la messa a punto di un sistema di amplificazione. In questa occasione si rese conto della sua attitudine e del suo interesse per la matematica e decise di intraprendere questi studi iscrivendosi alla Columbia University di New York.

Nel 1959 ebbe l'idea di introdurre una notazione sintattica formale per la descrizione delle regole grammaticali di un linguaggio di alto livello (Backus-Naur Form: J.W. Backus, *The syntax and semantics of the proposed international algebraic language of Zurich-ACM-GAMM Conference*, Proc. International Conf. Information Processing, UNESCO Paris 1959, pp. 125-32).

Negli anni '70 lavorò alla ricerca di metodi per migliorare la programmazione, sviluppando quella che poi è stata definita la cosiddetta *Programmazione Funzionale*.

Nel 1976 ha ricevuto la medaglia nazionale della scienza e nel 1977 il Premio Turing con la seguente motivazione:

«Per i suoi profondi, autorevoli e durevoli contributi al progetto di sistemi di programmazione ad alto livello, in special modo attraverso il suo lavoro sul FORTRAN e per l'originale e influente proposta di metodi formali per la specifica di linguaggi di programmazione.»

Ha lasciato l'industria dell'informatica nel 1991.

Riquadro 2 - Fortran Implementations (1954 - 1965)

J. A. N. Lee · Pioneer Day

List of FORTRAN implementations

Date	Name	Machine	Author(s)	Location
1954-57	FORTRAN (06)	IBM 704	Backus et al. ^a	IBM
1957	FORTTRANSIT	IBM 650	Bemer, Alexander, Pessin, Hemmes, May	IBM
1958	650 FORTRAN	IBM 650	Pessin, Wu	IBM
1958	FORTRAN	IBM 709	"	IBM
1958	FORTRAN E	IBM 704	Backus et al.; Mitchell, Sheridan, Brady, May	IBM
1958	FORTRAN II	IBM 704	Ziler, Nelson	IBM
1960	FORTRAN	IBM 1620	Laffan	IBM
1960	ALTAC	Philco 2000	Rosen, Goldberg	Philco
1960	FORTRAN	IBM 7070	Alexander et al.	IBM
1960	FORTRAN	IBM 705	Brittenham, Horning, Kuss, Matheny, Miller, Seldon, Williams	IBM, GUIDE
1960	FORTRAN	CDC 1604	Cray	CDC
1960	Honeywell Algebraic Compiler	?	Opler	CUC
1961	FORDO	IBM 1620	Davidson, McClure	Wisconsin
1961	FORTRAN	H-290	Hankins	HIS
1961	GOTRAN	IBM 1620	Laffan, Resta	IBM
1961	Automath-800	H-800	Opler, King, O'Conner, Beeber, Hopkins, Brestwick	CUC
1961	FORTRAN I	UNIVAC	?	Rem-Rand
1961	FORTRAN E	LARC	CSC	Rem-Rand
1961	UT FORTRAN	IBM 1620	Lee, Field	Toronto
1961	AFIT FORTRAN	IBM 1620	Pratt	Wright Patterson AFB
1961	1401 FORTRAN	IBM 1401	Haines et al.	IBM
1961	FORTRAN	B5000	?	Burroughs
1961	FORTRAN IV	IBM 7090J4	Larner	IBM
1961	FORTRAN II	RCA 301	Hux et al.	RCA
1962	FORTRAN IV	IBM 7040J4	Medlock	IBM
1962	FORTRAN II	Univac LARC	Erdwinn, Ferguson, Gatt, Malenkoff, Nutt, Richards	CSC
1962	Automath-400	H-400	Boris and Cahill	HIS
1962	FORTRAN II	RCA 301	Hux et al.	RCA
1963	FORTRAN IV	Univac 1107 ^a	Anderson, England, Erdwinn, Ferguson, Gatt, Kinney, Martin, Nutt, Pemine	CSC
1963	FORTRAN IV	UNIVAC III	Balke, Harkins, Nutt, Peterson, van der Wouw	CSC
1963	Automath-1800	H-800/1800	Chang	HIS
1963	Automath-1400	H-1400	Roth	CDC
1963	FORTRAN	CDC 160A	Neuhaus	
1963	FORTRAN S1	IBM Stretch	Glennie	UKAEA (AWRE)
1963	FORTRAN IV	IBM 7030	Bishop, Webster	IBM
1963	KINGSTRAN	IBM 1620	Field, Jardine, Lee, Lee, Robinson	Kingston, Ont.
1964	FORTRAN S2	IBM Stretch	Glennie	UKAEA (AERE)
1964	FORTRAN IV	Philco 2000	Brown, Rice, Richards, Sidrane	CSC
1964	HARTRAN	Atlas (F)	Pyle	UKAEA (AERE)
1964	FORTRAN	Orion (F)	Taylor, Harrigan	Rutherford Lab. (UK)
1964	FORTRAN	SDS910	Dunlap, Ryan	Digtek
1964	FORTRAN	CDC 3000L	Schumacher	CDC
1964	FORTRAN	CDC 3600	Tiede	CDC
1964	Fast FORTRAN	CDC 3600	?	Mich. State
1964	FORTRAN E	RCA 3301	Hux et al.	RCA
1964	FORTRAN IV	RCA 601	Best	Decision Systems
1965	PUFFT	IBM 7094	Rosen et al.	Purdue
1965	FORTRAN IV	IBM 7030	MacIntosh, Matheny, McPherson, Peterson, Sexton, Stephens	CSC

10 · Annals of the History of Computing, Volume 6, Number 1, January 1984

Bibliografia

1. Allen F.E., "A Technological Review of the Early FORTRAN Compilers, in "Early Days of FORTRAN", *Annals of the History of Computing*, 6 (1), Jan. 1984, pp. 22-26.
2. ANSI, *American National Standard Programming Language FORTRAN*, ANSI X3.9-1978, American National Standards Institute, New York 1978.
3. ANSI, *American National Standard Programming Language FORTRAN 90*, ANSI X3.198-1992, American National Standards Institute, New York 1992.
4. Backus J.W., "Programming in America in the 1950s", in [23], pp. 130-131.
5. Backus J.W., "The History of Fortran I, II, and III", *IEEE Annals of the History of Computing*, 20(4), Oct.-Dec. 1998, pp.68-78.
6. Bright H., "FORTRAN Comes to Westinghouse-Bettis, 1957", *Annals of the History of Computing*, 1(1), Jul. 1979, p.74.
7. Cocke J., Schwartz J.T., *Programming Languages and Their Compilers*, Courant Institute, NYU, New York 1970, p. 511.
8. Davis G.B., *An Introduction to the IBM 7090/94 Computer*, McGraw-Hill, New York-San Francisco-Toronto-London-Sydney, 1965.
9. Friedman W., "From Babbage to Babel and Beyond: A Brief History of Programming Languages", *Computer Languages*, 17 (1), 1992, p. 7.
10. Goldberg R., "Register Allocation in FORTRAN I", in "Early Days of FORTRAN", *Annals of the History of Computing*, 6 (1), Jan. 1984, pp. 19-20.
11. Greenfield M.N., "The Impact of FORTRAN Standardization", in "Istitutionalization of FORTRAN", *Annals of the History of Computing*, 6 (1), Jan. 1984, p. 33.
12. Heising W.P., "Fortran", *Communications of the ACM*, 6 (3), Mar. 1963, pp. 85-86.
13. Heising W.P., "History and Summary of FORTRAN Standardization Development for the ASA", *Communications of the ACM*, 7 (10), Oct. 1964, p. 590.
14. Heising W.P., "The Emergence of FORTRAN IV from FORTRAN II, in "Istitutionalization of FORTRAN", *Annals of the History of Computing*, 6 (1), Jan. 1984, pp. 31-2.
15. IBM, "Preliminary Report, Specifications for the IBM Mathematical FORMula Translating System, Fortran", *Programming Research Group, Applied Science Division*, IBM Corporation, New York, 10 Nov. 1954, pp. 1-29.
16. IBM, "IBM Electronic Data-Processing Machines – Type 704, Manual of Operation: Preliminary Edition", *International Business Machines Corporation*, 1954.

http://www.cs.virginia.edu/brochure/images/manuals/IBM_704/IBM_704.html
(visto 12 marzo 2014)

17. IBM, *The FORTRAN Automatic Coding System for the IBM 704 EDPM*, Programmer's Reference Manual, IBM Corporation, form n. 32-7026, Oct. 1956.

18. IBM, Preliminary Operator's Manual for the Fortran Automatic coding System for the IBM 704 EDPM, Programming Research Dept., IBM Corporation, New York, 8 Apr. 1957.

19. IBM, *Description of Source Language Additions to the Fortran II System*, Programming Research, Additions to Fortran II, IBM Corporation, New York 1958, (Distributed to users of Fortran III, 12 pp.).

20. IBM, *FORTRAN (General Information Manual)*, Data Processing Division, IBM Corporation, F28-8074, White Plains, N. Y. 1961, p. 65.

21. Knuth D.E., Pardo L.T., "The Early Development of Programming Languages", in [23], pp.197-273.

22. Lee J.A.N., "Pioneer Day, 1982", *Annals of the History of Computing*, 6 (1), Jan. 1984, pp.8-11.

23. Metropolis N., Howlett J., Rota G.C. (eds.), *A History of Computing in the Twentieth Century*. A collection of essays, Academic Press, New York, 1980.

24. Mitchell G.E., Programmer's Primer for FORTRAN Automatic Coding System for the IBM 704, 32-0306-1, IBM Corporation, 1957.

25. Murray Hopper G., "The First Bug", *Annals of the History of Computing*, 3 (3), Jul. 1981, pp.285-286.

26. Nutt R., "Compiler Techniques Available in 1954", in "Early Days of FORTRAN", *Annals of the History of Computing*, 6 (1), Jan. 1984, pp.20-2.

27. Olsen T.M., "Philco/IBM Translation at Problem-Oriented, Symbolic and Binary Levels", *Communications of the ACM*, 8 (12), Dec. 1965, pp. 762-7688.

28. Oswald H., "The Various FORTRANS", *Datamation*, 10 (8), Aug. 1964, pp. 25-9.

29. Petrocelli C., *Dalla programmazione automatica alla compilazione : Plankalkul, Fortran, Algol e Lisp*, Pensa multimedia, Lecce, 2003.

30. Rabinowitz I.N., "Report on the Algorithmic Language FORTRAN II", *Communications of the ACM*, 5 (6), Jun. 1962, pp. 327-37.

31. Ridolfi P., Coén H., *Come programmare con il Fortran*, Franco Angeli Editore, Milano, 1980.

32. Sammet J.E., *Programming Languages: History and Fundamentals*, Prentice-Hall series in Automatic Computation, Englewood Cliffs, 1969.

33. Sebesta R.W., *Concepts of Programming Languages*, Addison Wesley, Boston-San Francisco-New York-London-Toronto-Sydney-Tokyo-Singapore-Madrid-Mexico City-Munich-Paris-Cape Town-Hong Kong-Montreal, 2002, 5th ed.

34. Sheridan P.B., "The Arithmetic Translator-Compiler of the IBM Fortran Automatic Coding System", *Communications of the ACM*, 2 (2), Feb. 1959, pp. 9-21.

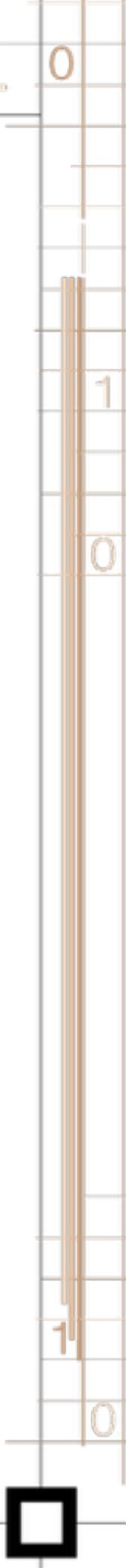
35. *USA Standard FORTRAN*, United States of American Standards Institute, USAS X3.9-1966, New York, Mar. 1966.

36. *USA Standard Basic FORTRAN*, United States of American Standards Institute, USAS X3.10-1966, New York, Mar. 1966.

Biografia

Carla Petrocelli è docente di Storia e Fondamenti di informatica presso l'Università degli Studi di Bari Aldo Moro. Studiosa di linguistica computazionale e di storia dell'informatica, la sua attività di ricerca si è concentrata in particolare sull'analisi di opere scientifiche del '600 investigate con tecnologie computazionali, e sulla evoluzione dei sistemi di calcolo automatici e dei linguaggi di programmazione ad essi applicati.

E-mail: carla.petrocelli@uniba.it



Rubrica

Nuovo Impulso per la Sicurezza delle Informazioni con la Revisione delle Norme

A. Piva - A. Rampazzo

Le informazioni sono cruciali per il funzionamento e talvolta persino per la sopravvivenza di un'organizzazione, pertanto la loro gestione e tutela sono aspetti cruciali soprattutto quando si tratta di dati critici in termini di proprietà industriale e di tutela degli stakeholders.

E' necessario proteggere le informazioni, attraverso un opportuno Sistema di Gestione per la Sicurezza delle Informazioni, da accessi non autorizzati e dal rischio che esse vengano corrotte o rese non disponibili, cosa che avrebbe un impatto negativo su diversi aspetti del business.

Introduzione

Il valore delle informazioni gestite da un'organizzazione richiede un continuo impegno per misurare, controllare e migliorare la sicurezza dei servizi offerti, garantendo il rispetto delle norme e delle direttive per la tutela della riservatezza, integrità e disponibilità delle informazioni stesse. Per garantire la protezione del patrimonio informativo aziendale ogni organizzazione dovrebbe aver definito un proprio Sistema di Gestione per la Sicurezza delle Informazioni (SGSI)¹, modellato sulla base di uno standard. In questo contesto l'elemento centrale è costituito dalla "Gestione del Rischio", che permette di analizzare i rischi tramite la loro identificazione, stima e misurazione, di individuare le vulnerabilità che potrebbero compromettere la riservatezza, l'integrità e la disponibilità delle informazioni, di definire le contromisure per contrastare le minacce ed infine di pianificare gli interventi da attuare per la riduzione dei rischi stessi. Un'attenzione particolare è posta nell'adozione di misure organizzative

¹ Vedi articolo "La sicurezza delle informazioni e le Norme ISO 27000" pubblicato su "Mondo Digitale" nr. 3 di settembre 2008



che rendano efficaci gli interventi tecnici, nella strutturazione dei processi di lavoro, nell'individuazione di soluzioni tecnologiche innovative, nella formazione e nella sensibilizzazione del personale in merito ai problemi della sicurezza. L'attuazione del Sistema di Gestione per la Sicurezza delle Informazioni porta alla definizione di ruoli, di responsabilità e di regole specifiche della sicurezza, di attività di pianificazione e realizzazione di politiche e procedure, nonché alla verifica dei processi attuati secondo quanto previsto dal modello indicato dallo standard.

Cosa sono i sistemi di gestione della sicurezza delle informazioni

Il Sistema di Gestione della Sicurezza delle informazioni (Information Security Management System - ISMS) è regolato dalle norme della famiglia ISO 27000 e nasce allo scopo di custodire e proteggere l'insieme di informazioni di una organizzazione.

Il Sistema di Gestione della Sicurezza dell'informazione è dunque fondamentale per proteggere i dati ed evitare possibili violazioni e divulgazioni di notizie private. Un'organizzazione dovrebbe dunque tenere in particolare considerazione la sicurezza del proprio protocollo comunicativo attraverso sistemi di gestione della sicurezza delle informazioni validi, efficaci ed efficienti sia nel confronto dell'interno che dall'esterno.

Pertanto ISMS sono necessari per monitorare il flusso delle informazioni, per proteggerle e controllarle in qualunque momento anche dal punto di vista degli accessi alla rete. Essi possono quindi essere fondamentali anche per migliorare e accrescere l'efficienza della propria azienda.

Sistemi di gestione della sicurezza delle informazioni: le norme ISO/IEC 27001 e ISO/IEC 27002

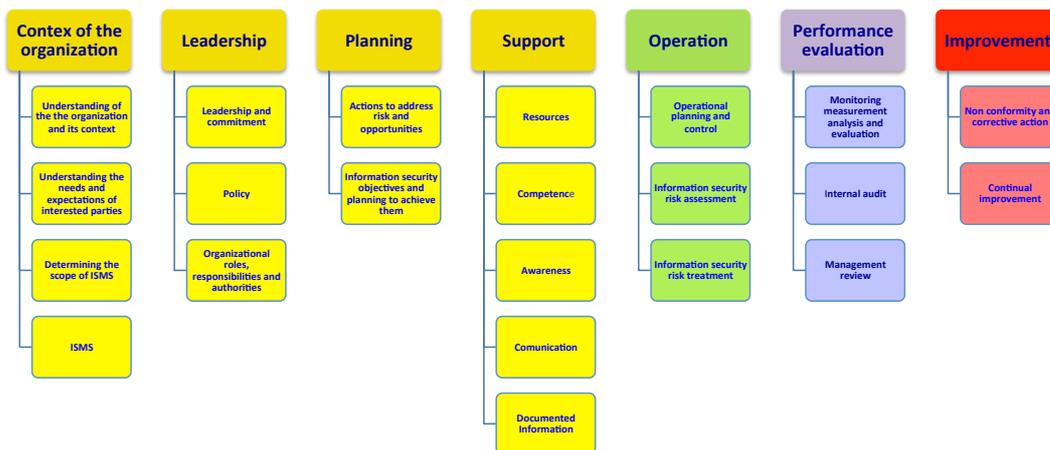
La norma ISO/IEC 27001 è adatta a diverse tipologie di aziende ed organizzazioni ed è una certificazione importante soprattutto quando l'azienda divulga o utilizza un ampio numero di informazioni anche con soggetti terzi.

La nuova ISO/IEC 27001 rivista e pubblicata nella sua seconda versione il 1 ottobre 2013, segue le nuove direttive definite dalla ISO e descritte nel MSS HLS (Management system standards - High level structure). Il primato in questo senso va alla ISO 22301:2012, essendo stata la prima norma pubblicata con la nuova strutturazione dei contenuti.

L'obiettivo di questo modello uniforme è sostanzialmente l'allineamento di tutte le norme dei sistemi di gestione ad una medesima organizzazione dei contenuti, avviando così il progetto di integrabilità concettuale degli schemi. L'integrabilità di fatto, sempre possibile in linea teorica, deve essere oggetto di valutazione da parte delle singole organizzazioni interessate, anche per individuare le migliori modalità applicabili.

Sinteticamente i contenuti della nuova ISO/IEC 27001, espandendone i punti principali dell'indice (si veda figura 1 struttura della norma), si possono riassumere:

- Il contesto dell'organizzazione - Capire l'organizzazione ed il suo contesto - Comprendere le necessità e le aspettative delle parti interessate - Determinare il campo di applicazione del sistema di gestione per la sicurezza delle informazioni - Sistema di gestione per la sicurezza delle informazioni
- Guida e direzione (Leadership) - Guida, direzione e impegno - Politica - Ruoli, responsabilità e poteri dell'organizzazione
- Pianificazione - Azioni per fronteggiare rischi e opportunità - Valutazione del rischio relativo alla sicurezza delle informazioni - Trattamento del rischio relativo alla sicurezza delle informazioni — Obiettivi per la sicurezza delle informazioni e piani per conseguirli
- Supporto - Risorse - Competenze - Consapevolezza - Comunicazione - Informazioni documentate - Creazione e aggiornamento - Controllo delle informazioni documentate
- Operatività - Pianificazione e controllo operativo - Valutazione del rischio relativo alla sicurezza delle informazioni - Trattamento del rischio relativo alla sicurezza delle informazioni
- Valutazione delle prestazioni - Monitoraggio, misurazione, analisi e valutazione - Audit interni - Riesame della Direzione
- Miglioramento - Non conformità e azioni correttive - Miglioramento continuo
- Annex A - Riferimenti alla ISO/IEC 27002



Struttura della norma ISO/IEC 27001:2013

Si può quindi notare una nuova organizzazione delle tematiche con alcune novità. A titolo di esempio: si parla di informazioni documentate e non più di procedure documentate e registrazioni; le azioni preventive sono state eliminate, perché incluse nelle "azioni per fronteggiare rischi e opportunità"; la valutazione e il trattamento del rischio sono presenti sia nella pianificazione del Sistema sia nella sua operatività.

Non è sicuro che l'eliminazione del concetto di azione preventiva possa essere del tutto un beneficio, ma l'attuazione efficace del sistema di gestione è di per sé l'origine della prevenzione di qualsiasi possibile fattore di instabilità organizzativa.

Va evidenziato il forte richiamo alla comprensione del "contesto" nel quale opera l'organizzazione ed alle aspettative delle parti interessate, che dello stesso sistema possono essere le promotrici.

Con la precedente edizione della norma questo aspetto era poco sviluppato, concentrando da subito l'attenzione in modo troppo immediato sui beni e sulle pratiche di gestione della sicurezza. Oggi l'esigenza di definire le finalità, le opportunità ed i rischi relativi al sistema di gestione nel suo complesso, sia strategico aziendale che tecnico, risulta ben chiara, e permetterà di focalizzare con maggiore efficacia ed efficienza lo sviluppo dei controlli di sicurezza non solo da un punto di vista tecnico, ma anche e soprattutto da un punto di vista organizzativo e gestionale.

Vi è un nuovo approccio anche alla gestione del rischio: uno dei cambiamenti più importanti della norma è quello di definire un nuovo approccio per l'applicazione di valutazione del rischio, sia nella fase di "pianificazione" che nella fase di "attuazione". I requisiti di valutazione del rischio sono generici e sono allineati alla norma ISO 31000:2009². La gestione del rischio è un obiettivo a cui ogni impresa attenta agli aspetti preventivi dovrebbe tendere e che ogni cliente dovrebbe pretendere, in particolare in settori caratterizzati da alta vulnerabilità. La Linea Guida ISO 31000:2009 ci propone un modello di gestione del rischio e di integrazione dello stesso nel sistema di gestione aziendale. Essa è applicabile a tutte le tipologie di rischio (da quelli strategici a quelli operativi, valutari, di mercato, di compliance, di paese, ecc.).

Pertanto, non è più necessario individuare le attività, le minacce e le vulnerabilità al fine di individuare i rischi. Se la metodologia di valutazione del rischio utilizzata per l'organizzazione utilizza questo metodo e funziona, è possibile mantenerla e non c'è bisogno di cambiarla. Tuttavia, se si vuole, ci sono metodi alternativi che possono essere perfettamente validi da usare, che non utilizzano asset, minacce e/o vulnerabilità per identificare i rischi.

E' stata introdotta anche una nuova funzione nel processo di valutazione del rischio, che è il proprietario del rischio (o risk owner).

Complessivamente (si veda figura 2 confronto requisiti), l'edizione 2005 della ISO/IEC 27001 aveva 102 requisiti obbligatori contenuti nelle clausole da 4 a 8

² Nel novembre del 2010 è stata pubblicata la norma UNI ISO 31000:2010 "Gestione del rischio", traduzione italiana della corrispondente norma internazionale ISO 31000 del novembre 2009.

mentre ora nella versione 2013 sono stati aggiunti 28 requisiti, portando ad un totale di 130 requisiti presenti nelle nuove clausole da 4 a 10.

ISO/IEC 27001:2013			
§ requisito	Descrizione	Controlli	
Mandatory	4	Contesto dell'organizzazione	8
	5	Leadership	19
	6	Pianificazione	39
	7	Supporto	28
	8	Operatività	9
	9	Valutazione delle prestazioni	29
	10	Miglioramento	26
		Controlli	148
ISO/IEC 27001:2005			
§ requisito	Descrizione	Controlli	
Mandatory	4	Sistema di Gestione per la Sicurezza delle Informazioni	50
	5	Responsabilità della direzione	18
	6	Audit interni del SGSI	4
	7	Riesame del SGSI da parte della direzione	16
	8	Miglioramento del SGSI	14
		Controlli	102

Confronto requisiti ISO/IEC 27001

Sono state revisionate anche le norme ISO/IEC 27000 e la ISO/IEC 27002.

Tutte le definizioni presenti nella versione precedente della norma ISO/IEC 27001 sono state eliminate, e quelle che sono ancora rilevanti, sono state trasferite alla norma ISO/IEC 27000 pubblicata nel mese di gennaio 2014 e scaricabile gratuitamente per uso personale all'indirizzo: <http://standards.iso.org/ittf/licence.html>

Questo trasferimento ad una unica norma garantisce la coerenza dei termini e delle definizioni a tutti gli standard della famiglia ISO/IEC 27000.

La revisione della ISO/IEC 27002 ha previsto una nuova struttura dell'Annex A (si veda figura 3 - confronto Annex A): i controlli dell'annex A diminuiscono rispetto alla versione del 2005, passando da 133 a un totale di 114, vale a dire 19 controlli vengono rimossi in questo allegato. Tuttavia, nonostante questa diminuzione, il nuovo annex A che era composto nella versione 2005 da 11 domini, da A.5 a A.15, nella nuova versione 2013 è composto da 14 domini da A.5 a A.18.

Nella nuova versione sono stati separati due domini e creato uno nuovo su "Provider Relations" in risposta alla popolarità del "Cloud Computing" e gli sforzi per proteggere le catene di approvvigionamento.

Questo miglioramento della struttura dell'annex A, ha un impatto sulla chiarezza e l'allineamento con le politiche, i processi e le procedure aziendali esistenti.

Annex A ISO/IEC 27001:2013				
Area controllo	Descrizione	Obiettivi	Controlli	
Discrezionale	A5	Politica per la sicurezza	1	2
	A6	Organizzazione della sicurezza delle informazioni	2	7
	A7	Sicurezza delle risorse umane	3	6
	A8	Gestione dei beni	3	10
	A9	Controllo degli accessi	4	14
	A10	Crittografia	1	2
	A11	Sicurezza fisica ed ambientale	2	15
	A12	Sicurezza delle operazioni	7	14
	A13	Sicurezza delle comunicazioni	2	7
	A14	Acquisizione, sviluppo e manutenzione dei sistemi	3	13
	A15	Rapporti con i fornitori	2	5
	A16	Gestione degli incidenti relativi alla sicurezza delle informazioni	1	7
	A17	Aspetti relativi alla sicurezza delle informazioni nella gestione della continuità operativa	2	4
	A18	Conformità	2	8
Controlli		35	114	
Annex A ISO/IEC 27001:2005				
Area controllo	Descrizione	Obiettivi	Controlli	
Discrezionale	A5	Politica per la sicurezza	1	2
	A6	Organizzazione della sicurezza delle informazioni	2	11
	A7	Gestione dei beni	2	5
	A8	Sicurezza delle risorse umane	3	9
	A9	Sicurezza fisica ed ambientale	2	13
	A10	Gestione delle comunicazioni e dell'operatività	10	32
	A11	Controllo degli accessi	7	25
	A12	Acquisizione, sviluppo e manutenzione dei sistemi informativi	6	16
	A13	Gestione degli incidenti relativi alla sicurezza delle informazioni	2	5
	A14	Gestione della continuità operativa	1	5
	A15	Conformità	3	10
Controlli		39	133	

Confronto Annex A ISO/IEC 27001

Oltre ad avere cambiato il numero di controlli, un altro cambiamento importante di cui all'annex A ha a che fare con l'applicazione: nella nuova versione della norma ISO 27001 non è più necessario "selezionare" i controlli invece, le organizzazioni devono "stabilire" quali sono i controlli necessari, come parte del trattamento dei rischi, e di confrontare i controlli con l'annex A, al fine di garantire che non venga dimenticato alcun controllo importante.

Nel complesso, gli attuali Sistemi di Gestione della Sicurezza delle Informazioni non dovranno essere completamente reingegnerizzati per soddisfare i nuovi requisiti, anche se saranno, da un lato necessarie, e dall'altro lato possibili, delle modifiche significative a quanto attualmente implementato dalle diverse organizzazioni.

Comunque da questa revisione ne consegue la necessità di rimodulare le attuali Dichiarazioni di Applicabilità (SoA) che verrà comunque facilitata dalla presenza di un'apposita tabella di correlazione scaricabile all'indirizzo: <http://www.jtc1sc27.din.de/sbe/wg1sd3>

Impatto sulle organizzazioni già certificate

Tutti questi cambiamenti avranno un impatto su migliaia di aziende già certificate: dalla pubblicazione della norma ISO/IEC 27001:2013 tutte le organizzazioni avranno un periodo di tempo di due anni per provvedere all'adeguamento del Sistema di Gestione. L'adeguamento può essere fatta su qualsiasi revisione di audit annuale programmata, o da una verifica speciale concordata con l'Organismo di Certificazione.

Dalla prima pubblicazione della norma ISO/IEC 27001 nel 2005, il numero di aziende certificate è passato da 5.800 a circa 20.000 a fine del 2012 (survey ISO 2012). Tra i paesi con le aziende certificate in questo standard sono inclusi Giappone, Romania, Cina, Inghilterra, India e Italia. L'Italia ha circa 500 aziende certificate.

Certamente la revisione di queste norme porterà un nuovo impulso alle organizzazioni nella gestione della sicurezza delle informazioni un tema di vera attenzione in un continuo scenario nel quale le informazioni sono diventate la moneta del nuovo millennio.

Bibliografia

- ISO/IEC 27000:2014 Information technology -- Security techniques -- Information security management systems -- Overview and vocabulary
- ISO/IEC 27001:2013 Information technology -- Security techniques -- Information security management systems -- Requirements
- ISO/IEC 27002:2013 Information technology -- Security techniques -- Code of practice for information security controls

Biografie

Antonio Piva, Laureato in Scienze dell'Informazione, Vice Presidente dell'ALSI (Associazione Nazionale Laureati in Scienze dell'Informazione ed Informatica) e Presidente della commissione di informatica giuridica. Ingegnere dell'Informazione, docente a contratto di diritto dell'ICT, qualità e comunicazione all'Università di Udine. Consulente su Governo Elettronico, Agenda Digitale ed innovazione nella PA locale, Auditor Sistemi informativi e 231, è consulente e valutatore di sistemi di qualità ISO9000, Privacy e Sicurezza presso Enti pubblici e privati. Ispettore AICA presso scuole ed enti di formazione. Membro del Consiglio Nazionale del Forum Competenze Digitali, è Presidente della Sezione Territoriale AICA del Nord Est.

E-mail: antonio@piva.mobi

Attilio Rampazzo, CISA CRISC, C|CISO CMC consulente di Sistemi Informativi e Sicurezza delle Informazioni in primaria azienda di Servizi Informatici italiana. Ha maturato un'esperienza pluriennale nello sviluppo e conduzione di progetti informatici in ambito bancario e finanziario, nei quali la qualità e la sicurezza hanno ricoperto un ruolo determinante.

E' Vice Presidente di AICA sez. Nord Est, AICQ Triveneta e CISA Coordinator e Research Director in ISACA Venice chapter. Svolge attività come Valutatore di Sistemi di Sicurezza delle Informazioni e di Sistemi di Gestione dei Servizi (cert. AICQ Sicev) presso CSQA Certificazioni. Trainer accreditato APMG per ITIL, ISO 20000 e Cobit 5.

Socio AICA, AICQ, ISACA Venice chapter, ASSOVAL, ANIP.

E-mail: attilio.rampazzo@gmail.com