Nuovi modelli dell'ICT – Web 2.0 e oltre

Web 2.0, web semantico, grid computing, virtualizzazione … un'opportunità per le imprese strutturate

di *Antonio Candiello*[1]

Premessa (Abstract)

Web collaborativo, *social networks*, *wiki*, *blogs*, *community* virtuali, architetture *Ajax*, linguaggi xml, infrastrutture *peer to peer (*p2p) e sistemi di connettività come Skype: il modello "Web 2.0" è in una fase di massima popolarità. Non bastasse questo cataclisma tecno-mediatico, il sistema sta ora muovendo verso il nuovo mondo del "Web 3.0", con il suo raffinato e un po' criptico arsenale fatto di ontologie, web semantico, OWL e motori inferenziali, che ci consente di recuperare su un piano più maturo le linee di frontiera dell'intelligenza artificiale che avevamo lasciato orfane negli anni '70.

Il tutto si innesta su un altrettanto mobile (e "sociale") scenario di tecnologie architetturali ormai in piena transizione verso i modelli di *sharing* computazionale, che stanno sottraendo all'*High Performance Computing* (HPC) lo scettro del calcolo verso modelli più distribuiti e collaborativi basati su comunità, come il *cloud computing* di Amazon, la *grid* del calcolo creata dalla comunità dei fisici delle alte energie, fino ai progetti di "offerta volontaria di cpu time" del circuito Boinc/Seti@Home, variante "computing" dei sistemi paritari di scambio file inaugurati da Napster.

Di fatto, questa è la considerazione, stanno ormai giungendo a consolidamento un insieme di strumenti sul fronte dell'*Information & Communication Technology* (ICT), che a questo punto offrono la possibilità di incidere in maniera sostanziale nell'evoluzione dei nostri sistemi socio-economici, bussando alla porta delle aziende. Ora bisogna concentrare gli sforzi sull'aspetto chiave: la *corretta declinazione* di queste strategie di facilitazione tecnologica (siano esse web 2.0, web 3.0, grid/hpc o altro) nei giusti ambiti settoriali industriali e/o della pubblica amministrazione.

Introduzione

Il "percorso parallelo" che si è strutturato nel tempo tra qualità ed informatica fonda le sue motivazioni sulla base di una *convergenza* incardinata sul comune obiettivo di sfruttare al meglio le potenzialità di un approccio *strutturato*, *trasparente* e *flessibile* nella gestione dei processi aziendali.

L'impostazione che suggeriscono le norme di gestione della qualità e più in generale le norme volontarie create per garantire una efficace gestione dei sistemi aziendali per le "tante qualità" ormai richieste alle imprese (dei prodotti, dei servizi, dell'ambiente, dell'etica, della sicurezza, …) è infatti centrata su un insieme di strategie organizzative e strutturali che ben si appaiano con la crescente esigenza di favorire una sempre più estesa informatizzazione delle aziende e dei suoi meccanismi operativi interni ed esterni:

a) una chiara definizione dei *processi* aziendali, delle loro interrelazioni e della relazione di questi con le funzioni stabilite;
b) una classificazione delle *funzioni* aziendali, mansioni e correlate competenze richieste per l'accesso al ruolo;
c) la strutturazione per *procedure* dei documenti aziendali che rendano chiare le operatività aziendali e le modalità operative esistenti;

---

[1] Consulente di processi di innovazione e qualità, ricercatore attivo su *grid computing*, *eGovernment* e *semantic web*. *Email*: candiello@gmail.com.

d) un insieme di *registrazioni* (ormai sempre più "informatizzate") che diano la misura di come i processi evolvono in azienda e di quali siano gli esiti dei controlli correlati;
e) la presenza di un comprensivo modello di gestione di *documenti* e registrazioni, che tenga *up-to-date* il sistema e garantisca la diffusione di informazioni aggiornate e validate;
f) il mantenimento di un opportuno insieme di *workflow* per la gestione di relazioni strutturate tra funzioni aziendali, personale, clienti e fornitori per i processi operativi, autorizzativi, afferenti a processi interni, esterni o misti.

La strutturazione operativa che tale impostazione comporta nelle aziende (generalmente) certificate le rende significativamente favorite, in quanto ambito di applicazione privilegiato da parte delle tecnologie informatiche, che possono in tal modo trovare la migliore espressione in termini di efficacia d'uso senza richiedere preventivamente onerosi processi di *business reengineering* all'organizzazione esistente.

Quanto detto vale in termini generali: in realtà, l'esperienza ha dimostrato una oggettiva difficoltà, anche in aziende ben strutturate, nell'applicazione di tecnologie informatiche che possiamo definire *hard*, a causa della rigidezza che hanno gli strumenti applicativi tradizionalmente attivati in azienda. Questi sono, tipicamente, i sistemi aziendali integrati (denominati *Enterprise Resource Planning*, ERP) e le loro evoluzioni "lato cliente" come i sistemi di gestione del cliente (denominati *Customer Relationship Management*, CRM) o altre tecnologie similari, applicate sul lato della fornitura e sui processi di progettazione dei prodotti.

L'ICT *flessibile* del Web 2.0

È in questo contesto che entrano prepotentemente le più recenti tecnologie, con un approccio *soft* nell'applicazione ai contesti sociali ed aziendali e le potenzialità nel facilitare i processi aziendali e i meccanismi di relazione semi-strutturata (tipicamente sociali) che rendono difficile le implementazioni dell'informatica *hard*.

Il fenomeno più conosciuto riguarda la diffusione dei sistemi basati su tecnologie web collaborative, dove l'utente, secondo un modello spesso portato in grafica ad anelli concentrici:

- produce informazione (*user-generated content*),
- valida / certifica le informazioni di altri utenti (*ratings, reviews*),
- fruisce dell'informazione così consolidata (*user*),
- dimostra attenzione/interesse per alcuni ambiti specifici (*taste*).

Si tratta in qualche modo dell'evoluzione dei modelli d'uso delle *communities of practices* di Wenger (Lave e Wenger, 1991), noti nel nostro paese soprattutto per l'applicazione a contesti associativi professionali (l'Aicq fornisce un ottimo esempio di comunità professionale wengeriana che vive superando i confini aziendali riunendo specialisti ed esperti della materia).

Se le comunità di pratica creano le condizioni per l'accumulazione della conoscenza e la sua strutturazione in determinati oggetti di conoscenza in grado di renderla persistente, i modelli riconducibili agli odierni *social networks*, che siano esterni, interni o "trasversali" ai confini aziendali, dimostrano una dinamicità maggiore specificamente in relazione agli strumenti interattivi, più semplici da usare rispetto agli anni '90 ma contemporaneamente più potenti e flessibili.

I contesti di applicazione più noti di questa nuova modalità sono sicuramente quelli "esterni", come l'enorme bacino dei blogs (oltre 70 milioni al 2007), la grande Wikipedia (oltre due milioni di articoli al 2007), o i grandi bacini come come mySpace (oltre 100 milioni di utenti), youtube (oltre 100 milioni di video/views/day), Facebook ed altri ancora.

Essi hanno dimostrato l'efficacia del modello "2.0" del web, che talvolta viene esemplificato con il seguente insieme di strati concettuali:

- **valori**: utente come generatore di contenuti, intelligenza collettiva, *perpetual beta*, estrema facilità d'uso;
- **applicazioni**: *blog*, *wiki*, *podcast*, *rss feeds*, *tagging*, *social networks*, motori di ricerca, giochi online multiplayer;
- **tecnologie**: *Ajax*, xml, *Open Api*, *web services Soap/SOA*.

Di fatto si è spostata l'attenzione dalle *tecnologie* come soggetto agli *utenti* che tali tecnologie utilizzano. La scoperta è che le tecnologie sono estremamente efficaci quando si tratta di facilitare la relazione sociale per la costruzione di contesti virtuali collettivi. L'utente non è più tale, il termine stesso dimostra la sua inadeguatezza, diviene *co-producer*.

Il *web semantico*, lanciato da Tim Berners Lee in un famoso articolo di *Scientific American* [2] parte da dove il Web 2.0 è arrivato. La promessa è quella, fondamentale, di poter finalmente interpretare il contenuto del web senza il rischio di fare quella confusione che fanno abitualmente i motori di ricerca tra termini che hanno significati diversi in settori differenti.

Lo strumento per ottenere questo risultato sono le cosiddette *ontologie*, ovvero ambiti tematici logici autonomi, che potranno consentire alle applicazioni di fare determinati ragionamenti, sulla base di una conoscenza strutturata preesistente (appunto, le ontologie). È l'"anello mancante" dell'intelligenza artificiale anni '70, che, ricordiamo, muoveva unicamente su di un piano *epistemologico*, quindi trascendendo dalla conoscenza accumulata. Si pone rimedio pertanto ora a questa passata disattenzione (che era dovuta a inerenti limitazioni tecnologiche e logico-formali dei sistemi di quegli anni), con il vantaggio di disporre di un insieme di solide tecnologie per la costruzione di semantiche e costrutti logici, basati su XML e da questo evoluti, come il linguaggio di "triple logiche" soggetto/predicato/oggetto del *Resource Description Framework* (RDF) [14], o le sue più recenti evoluzioni come il *Web Ontology Language* (OWL) [13], adatto a strutturare le ontologie sia in varianti inferenziabili che in varianti non computabili.

L'opinione che possiamo portare è quella che questi strumenti "semantici", più focalizzati ed in grado di gestire in qualche misura le complessità del linguaggio, siano *più adatti ai contesti aziendali*, per loro natura territorio "naturale" di ontologie settoriali (ogni professione porta con sé un insieme di logiche implicite e specifici vocabolari "per addetti ai lavori"). L'applicazione di strumenti di web semantico dovrebbe pertanto trovare nel contesto delle aziende maggiore affinità d'uso, più di quanto non sia invece avvenuto per gli strumenti di classe "Web 2.0", come *blogs*, *social networks*, p2p, spesso anzi osteggiati dalle imprese perché ritenuti troppo "aperti" e distraenti dai veri obiettivi dell'organizzazione. In ragione di ciò, suggeriamo a coloro che hanno responsabilità nei sistemi di gestione aziendale di valutare in anticipo strumenti/esempi di web semantico allo scopo di individuarne potenziali ambiti di applicazione nel proprio contesto organizzativo.

L'ICT *virtuale* del multi-*computing*

Parallelamente alla trasformazione delle tecnologie di interazione applicativa e del web sta avvenendo una *mutazione basilare* nei meccanismi "tecnici" di fornitura delle risorse di calcolo, forse in una modalità meno evidente, ma parimenti rivoluzionaria in quanto sta stravolgendo la stessa "fisicità" a cui l'informatica ci ha abituato.

Il fondamento strutturale per tale trasformazione ha le sue radici tecnologiche nella rapidità di crescita delle performance di rete, che supera sistematicamente la già rapida crescita della potenza di calcolo e di *storage* disponibile nei sistemi a parità di costo. La conseguenza è che i sistemi di calcolo e le risorse di memoria divengono ogni anno sempre "più prossimi" anche se geograficamente distanti. Questo fenomeno non solo sta rendendo i modelli di condivisione delle risorse più competitivi sul piano economico dei sistemi centralizzati di supercalcolo, ma sta

portando la disponibilità di risorse di calcolo abbondanti ed a basso costo fuori dai ristretti circuiti di chi se lo poteva permettere.

Ecco quindi che il vecchio modello del *supercomputing*, tipicamente basato su costosi mainframe proprietari, rapidamente obsolescenti, lascia il passo a sistemi basati su standard industriali ad alta scalabilità che ora vengono definiti *High Performance Computing* (HPC), per evidenziare le *capacità* di tali sistemi, più che le caratteristiche tecnologiche. Il sistema più performante secondo le ultime news è rappresentato dal "RoadRunner" di IBM che ha raggiunto il *petaflop* di potenza di picco (*peta* = 10^{15}).

Stanno d'altra parte fornendo molta eccitazione al mercato, per la grande flessibilità d'uso, alcuni interessanti modelli di utilizzo delle risorse di calcolo che Amazon (cfr. l'unità *Amazon Web Services*, AWS, [1]) sta portando in avanguardia. Si tratta del cosiddetto *cloud computing*, le cui basi tecnologiche sono la virtualizzazione delle risorse di calcolo (*Elastic Computing Cloud*, EC2) e la facilitazione dell'accesso allo storage (*Simple Storage Service*, S3) tramite un mix di interfacce costruite attorno agli innovativi protocolli *Rest/ROA* e complementate dalla connessione ai circuiti p2p come *BitTorrent*. Diverse aziende stanno considerando la possibilità di gestire la scalabilità "improvvisa" delle risorse tramite servizi come questi.

Ma, soprattutto, emergono nuovi modelli di *grid computing* per la condivisione dei centri di calcolo da parte di più istituzioni/aziende, mirando ad una consistente (di parla di alcuni ordini di grandezza rispetto agli HPC) riduzione degli oneri di acquisto e di gestione, secondo la semplice strategia che vede una singola realtà organizzativa accedere alla capacità aggregata di risorse rese disponibili da centinaia di realtà similari nelle situazioni di bisogno (condizioni di *peak usage*). In questo modo, il network mantiene lo spazio aggregato per soddisfare la domanda di picco delle singole istituzioni o, più precisamente, per soddisfare le esigenze di opportuni raggruppamenti *cross-border* di utenti denominati *Organizzazioni Virtuali* (VO), accomunati da un comune bisogno di risorse di calcolo per un contesto finalizzato.

La trasformazione che la grid può indurre sui sistemi di calcolo aggregato potrebbe avere forti analogie con l'impatto che i protocolli tcp/ip ebbero nelle reti pre-internet: la maturazione dei protocolli standard e di *middleware stack* a codice aperto come il software *gLite* europeo potrebbe trasformare i miliardi di variamente incompatibili sistemi computazionali presenti sul pianeta in un sistema pienamente interoperabile accessibile da una classe di applicazioni di nuova generazione: un vero e proprio "web del calcolo". Secondo quanto riporta Ian Foster [7], la grid è definibile come:

- un'infrastruttura hardware e software che fornisce accesso flessibile, consistente, pervasivo ed economico accesso a capacità computazionali consistenti (prima definizione, del 1998),
- un meccanismo di condivisione coordinata di risorse e di *problem solving* per organizzazioni virtuali che attraversano confini istituzionali (definizione "sociale" del 2000);
- un modello coordinato di risorse computazionali non soggette ad un controllo centralizzato, che utilizza protocolli ed interfacce standard, aperti, generali, e consente l'erogazione di definite misure e modalità di qualità di servizio (definizione "operativa" del 2002).

Ricordiamo in questo contesto che la più ampia infrastruttura di grid computing "di produzione" ovvero utilizzata quotidianamente per il calcolo, è la grid europea di *Enabling Grids for E-sciencE* (EGEE) www.eu-egee.org [5], una struttura federata di risorse eterogenee finalizzata al calcolo scientifico che comprende oltre 120 istituzioni, 250 siti in 48 paesi, oltre 68000 cpu disponibili per circa 8000 utenti 24 ore al giorno, 7 giorni alla settimana., che è pronta a digerire i previsti 17 *peta*bytes che verranno prodotti ogni anno dall'acceleratore di particelle del Cern a partire da quest'anno.

Non solo. La griglia europea è attualmente in fase di transizione per divenire la prima infrastruttura di calcolo distribuito su base continentale (cfr. www.eu-egi.org, [5]), ponendo le condizioni per una piattaforma tecnologica europea di calcolo utilizzabile per la ricerca ma anche

quale prima sperimentazione di un "circuito tecnologico" di nuova generazione nel quale possano operare le imprese più avanzate e pronte a cogliere nuove opportunità di sviluppo.

Conclusioni

Confidiamo, grazie a questo breve ma denso percorso nel mondo delle tecnologie ICT emergenti, di aver suscitato qualche interesse a valutarne l'applicabilità potenziale nel proprio ambito.

L'ICT diviene *flessibile* e si adatta al nostro modo di ragionare, ci facilita nuovi modelli relazionali e ci aiuta ad interagire costruttivamente con gli altri, mantenendo la persistenza semi-strutturata di quanto (socialmente) costruito. L'ICT diviene persino *virtuale*, non si fa vedere ma ci dà il suo prezioso supporto anche in condizioni di forte variabilità di consumo di risorse, evolve in un sistema infrastrutturale onnipresente ed accessibile, scalabile e capace.

Il centro del sistema però, paradossalmente, si sposta sempre più dalla tecnologia alla persona ed all'organizzazione, che insieme costituiscono il vero *soggetto* artefice di un uso *creativo* e *finalizzato* degli strumenti a disposizione, che tornano ad essere *oggetti*, non più dominanti e complicati, come nei dialoghi tra *freaks* appassionati, ma *diffusi* e *facili* da utilizzare, come è necessario in una società tesa ad evolvere continuamente i propri modelli operativi a vantaggio di persone, imprese ed istituzioni.

Glossario

Ajax: *Asynchronous JavaScript and XML*, architettura per la creazione rapida di applicazioni web semplici e facili da usare;

API: *Application Programming Interface*, modello di interfaccia software esposta alle applicazioni "utenti";

AWS: *Amazon Web Services*, unità organizzativa di www.amazon.com per i nuovi servizi avanzati;

Blog: modello d'uso e di interfaccia di grande popolarità per inserire individualmente contenuti in logica cronologica sotto forma di "diario pubblico";

Byte: composta di 8 *bits*, rappresentà l'unità di riferimento per lo *storage*;

Cloud: (Computing), tecnologia per l'accesso remoto a risorse di calcolo "a consumo". Il modello fa largo uso delle tecniche di *virtualizzazione* delle risorse di calcolo;

Cpu: *Central Processing Unit*, unità di calcolo "atomica" dei computer;

CRM: *Customer Relationship Management*, letteralmente "Gestione del Cliente", che in IT indica una classe di strumenti creati allo scopo;

EC2: *Elastic Computing Cloud*, tecnologia di www.amazon.com per fornire calcolo a richiesta a partire da una "nuvola elastica" di risorse disponibili;

EGEE: *European Grid for E-sciencE*, circuito europeo di riferimento per il grid computing scientifico (cfr. www.eu-egee.org);

EGI: *European Grid Initiative*, iniziativa per un sistema di calcolo basato su griglie condivise a livello europeo con caratteristiche di persistenza nel tempo "infrastrutturale" (cfr. www.eu-egi.org);

eGov: (*eGovernment*) sigla che denota il complessivo ambito delle tecnologie IT per la Pubblica Amministrazione;

ERP: *Enterprise Resource Planning*, sistemi informatici gestionali orientati alla produzione di largo uso presso le imprese;

Flop(s): *floating-point operations* (*per second*), unità di riferimento per il calcolo in virgola mobile;

Freak: termine comunemente utilizzato per indicare gli appassionati di Information Technology;

Grid: (Computing), tecnologia per la condivisione (*sharing*) delle risorse di calcolo che consente a gruppi trasversali di utenti (le VO) di accedere con grande efficienza ad elevate potenze aggregate solo nei momenti di bisogno;

HPC: *High Performance Computing*, sigla che individua una strategia che privilegia l'accentramento di consistenti risorse di calcolo per il calcolo tecnico e scientifico;

HTTP: *hypertext transfer protocol*, protocollo standard per il web su internet;

ICT: *Information & Communication Technology*, sigla che indica il complesso delle attività nel settore dell'informatica (hardware, software, servizi) e telecomunicazioni (telefonia fissa e mobile, connettività);

Kilo, Mega, Giga, Tera, Peta, Hexa …: qualificatori progressivamente crescenti (10 ad esponente 3, 6, 9, 12, 15, 18 …) per specificare misure molto grandi;

Ontologia: sistema concettuale autoconsistente, che nell'accezione informatica è derivata dalla filosofia e costituisce la "base di conoscenza" (da creare inizialmente e da aggiornare) dei sistemi semantici;

OWL: *Ontology Web Language*, linguaggio basato su XML per la definizione semantica di *ontologie*;

P2P: *peer to peer*, punto a punto, architetture multipolari per lo scambio di risorse (storage, computing, files, banda, informazioni) attraverso le reti informatiche. L'esempio più conosciuto è Napster;

RDF: *Resource Description Framework*, linguaggio adatto alla definizione di "metadati" per la classificazione delle risorse del web. Basato su XML, il sistema si basa su "triple logiche" e consente di attribuire significati semantici alle risorse taggate;

Rest/Roa: *Representational State Transfer / Resource Oriented Architecture* classe di protocolli semplificati per la gestione di risorse di memoria basate su http. Spesso portato a raffronto con Soap/Soa;

RSS: *Really Simple Syndication*, un protocollo per distribuire *headlines*, notizie, avvisi di aggiornamento di contenuti nel web;

S3: *Simple Storage Service*, tecnologia di www.amazon.com per fornire memoria di massa (*storage*) a richiesta secondo un modello via *http* semplificato tramite i protocolli Rest/Roa;

Soa: *Service Oriented Architecture*, modello architetturale dove tutti i sistemi software espongono "servizi" accessibili via web con protocolli specifici;

Soap: *Simple Object Access Protocol*, interfaccia di comunicazione tra basata su XML (per il contenuto) e http(s) per il trasporto;

Social Network: sistemi interattivi per facilitare le relazioni sociali virtuali e reali;

Tag: nell'IT, si intende con tale termine un attributo utile alla classificazione tematica;

VO: *Virtual Organization*, comunità d'uso delle reti di calcolo condivise;

Web 2.0: denota il contesto delle nuove applicazioni che affidano quote crescenti di responsabilità all'utente ("*user as a producer*");

Web 3.0: step più avanzato del "Web 2.0", intende sottolineare il passaggio da uno logica tecnico/testuale ad una logica semantico/concettuale;

Web Semantico: modello del web basato sui concetti/significati, come contrapposto al sistema "testuale" nel quale siamo immersi quotidianamente (cfr. le ricerche testuali in Google);

Web Service: interfaccia di base per rendere i servizi interoperabili via web;

Wiki: modello di interfaccia adatto alla gestione condivisa di grandi basi di conoscenza enciclopedica (cfr. www.wikipedia.org);

XML: *eXtensible Markup Language*, standard di base per un'intera classe di linguaggi "tagged" che hanno inteso superare la dominanza di "aspetto" intrinseca in HTML.

Riferimenti

[1] Amazon Web Services (AWS), aws.amazon.com/ec2; EGEE NA1, "Comparative study: Grids and Clouds, evolution or revolution?", EGEE-II INFSO-RI-031688, 27/5/2008;

[2] Berners-Lee T., Hendler J. and Lassila O. "The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities", Scientific American, May 2001;

[3] Bushell S., "Enterprise 2.0 - What is it good for? A 12-step guide to getting the most out of Web 2.0 tools and making it safe-for-purpose", CIO, 6/5/2008;

[4] Davies J., Fensel D., van Harmelen F. (ed.) (2003), "Towards the semantic web: ontology-driven Knowledge Management", Wiley, Chichester, Uk;

[5] Enabling Grids for E-sciencE (EGEE), www.eu-egee.org, European Grid Initiative, www.eu-egi.org, Italian Grid Initiative, www.italiangrid.org;

[6] Feigenbaum, Lee, Ivan Herman, Tonya Hongsermeier, Eric Neumann, and Susie Stephens. "The Semantic Web in Action." Scientific American, vol. 297, Dec. 2007, pp. 90-97;

[7] Foster, I. "What is the grid? a three point checklist", GridToday, July 22, 2002;

[8] Hinchcliffe D., "The story of Web 2.0 and SOA continues", on ZDNet "Enterprise 2.0", May 17th, 2007;

[9] Lave J., Wenger E. (1991), Situated Learning: Legitimate Peripheral Participation, Cambridge University Press, New York;

[10] MacManus R., "10 Semantic Apps to Watch", Read Write Web, 29 Nov. 2007;

[11] Mika P. (2007), "Social Networks and Semantic Web", Springer, NY,;

[12] Osimo D., "Web 2.0 in Government: why and how?", JRC Scientific and technical reports, 2008;

[13] W3C, "Web Ontology Language (OWL)", http://www.w3.org/2004/OWL;

[14] W3C, “Resource Data Framework (RDF)”, http://www.w3.org/RDF;
[15] W3C, “Semantic Web case studies and use cases”, http://www.w3.org/2001/sw/sweo/public/UseCases.