UNIVERSITÀ DEGLI STUDI DI TRENTO

Facoltà di Economia
Corso di Laurea specialistica in Net Economy
Tecnologia e Management dell'informazione e della conoscenza

Tesi di Laurea specialistica

Riassunto

METODI E MODELLI PER L'AGGREGAZIONE, L'ANALISI E LA CERTIFICAZIONE DI SORGENTI DATI ETEROGENEE E DISTRIBUITE

IL CASO DI TRENTINO RISCOSSIONI S.P.A.

Relatore:
Prof. Paolo Giorgini

Correlatore:
Dott. Marco Combetto

Laureando:
Stefano Brida

Anno Accademico 2008-2009

RIASSUNTO

Nelle moderne realtà aziendali è inevitabile che differenti parti dell'organizzazione utilizzino sistemi diversi per produrre e memorizzare i dati necessari per lo svolgimento delle loro attività. Soltanto combinando le informazioni provenienti dai diversi sistemi l'organizzazione potrà accedere al pieno valore dei dati che possiede. La domanda di prodotti di integrazione dei dati consistenti, flessibili e performanti è oggi in continua crescita, la complessità e la rapidità di evoluzione dei business moderni rende vitale per l'azienda disporre di dati integrati, aggiornati e di qualità. Nel contesto attuale grandi quantità di dati sono memorizzate in database relazionali, tuttavia, esiste una pluralità di altre fonti di dati, più o meno strutturati (documenti di testo, fogli di calcolo, ecc.), che con l'avvento del Web si è ulteriormente allargata. Per poter competere nel mercato attuale le imprese hanno la necessità di accedere alle informazioni presenti in tutte le basi dati possedute internamente, attraverso una vista unica integrata e consolidata, inoltre, la necessità di accedere a dati esterni (dati sul Web) si fa sempre più pressante.

L'obiettivo di questa tesi è l'investigazione e lo studio dettagliato delle diverse metodologie e tecnologie esistenti a supporto dell'integrazione di sorgenti di dati eterogenee e distribuite in un ambiente unico, omogeneo e storicizzato. In uno scenario contraddistinto dalla presenza di diverse tecniche di integrazione dei dati sono state analizzate le principali metodologie di Data Integration, descrivendo funzionalità e caratteristiche delle diverse tecnologie presenti sul mercato. Definiti ed esaminati i principali prodotti di integrazione, viene analizzato l'aspetto economico di tali applicativi attraverso un'analisi delle diverse tipologie di costo generate da un progetto di Data Integration. La tesi, oltre a descrivere gli aspetti positivi e i punti di forza che caratterizzano le metodologie di integrazione consolidate, evidenzia gli aspetti più critici ed accenna al possibile impiego di metodologie di nuova generazione, in particolare di tecnologie semantiche, che, sebbene ancora immature, sembrano essere promettenti e si pongono l'obiettivo di superare i limiti tipici delle tecniche tradizionali.

In seguito all'esposizione dello stato dell'arte della disciplina della Data Integration, la tesi descrive e analizza l'impiego di una metodologia specifica in un caso reale nel settore della gestione delle entrate, con particolare riferimento agli aspetti di accertamento dei tributi, dominio applicativo caratterizzato da alta variabilità ed eterogeneità dei dati. Nello specifico, viene valutato l'utilizzo di alcuni strumenti a supporto della realizzazione di una base dati integrata, omogenea e storicizzata, con il fine di giungere ad una soluzione implementabile in tempi medio-brevi e con un chiaro valore aggiunto per l'utilizzatore finale (committente del progetto). In questo senso i risultati ottenuti dalle attività di sperimentazione sono stati giudicati buoni dal committente e sono oggi il punto di partenza per la realizzazione di un

sistema di integrazione che porterà in produzione ciò che si è appreso dall'esperienza sperimentale. Il lavoro pratico svolto nell'ambito di questa tesi da un lato ha permesso di riconoscere i punti di forza e gli aspetti positivi delle metodologie di integrazione tradizionali, dall'altro ha contribuito ad evidenziarne i limiti e le criticità, determinando l'avvio di una fase di valutazione di un possibile impiego di metodologie innovative che, attraverso l'applicazione di tecnologie semantiche, promettono di superare alcuni limiti delle tecniche di integrazione tradizionali.

La tesi è frutto del lavoro effettuato nell'ambito dello stage svolto presso Informatica Trentina S.p.A. nel periodo compreso tra marzo e agosto 2009. In particolare, nel corso dello stage sono state effettuate attività a supporto del progetto di sperimentazione che ha coinvolto Trentino Riscossioni S.p.A., Informatica Trentina S.p.A. e il Comune di Folgaria. Nell'ambito di questa sperimentazione è stata testata, su un ambito specifico dell'accertamento tributario, una soluzione di Data Warehousing dotata di funzionalità di integrazione automatica dei dati, tramite l'utilizzo di tecniche di inferenza statistica, e di strumenti a supporto della gestione della qualità del dato.

La tesi si suddivide in due sezioni. Nella sezione I vengono analizzate le principali metodologie di integrazione, proponendo uno stato dell'arte della disciplina che parte dalle origini dei sistemi di gestione dei dati, descrive le principali metodologie consolidate, per arrivare all'analisi dei metodi più innovativi e recenti che si propongono sul mercato. Nella sezione II vengono invece descritte nei dettagli le attività di sperimentazione relative all'integrazione dei dati riguardanti il Comune di Folgaria, mettendo in luce pregi e difetti della situazione sperimentata, i vantaggi ottenibili dalla realizzazione di una base dati integrata e le criticità che si presentano nel realizzare soluzioni di questo tipo, sia a livello generale sia in riferimento al contesto operativo di Trentino Riscossioni.

La prima parte della tesi si apre con l'analisi dell'evoluzione dei sistemi di gestione dei dati. Partendo dalla nascita dei primi sistemi in grado di memorizzare e gestire i dati sono state ripercorse le principali tappe della disciplina del Data Management, mettendo in luce le maggiori evoluzioni tecnologiche che hanno permesso la realizzazione dei moderni strumenti di Data Integration. Tale analisi permette di identificare quelle che sono le principali metodologie di integrazione dei dati: (a) *Data Consolidation*, consiste nel trasferire tutti i dati dai sistemi sorgente ad un database unico, centralizzato e integrato sul quale verranno sviluppati applicativi di analisi dei dati, (b) *Data Propagation*, a differenza dell'approccio precedente non viene creato un database unico ma più database ognuno dei quali serve i bisogni di uno specifico ambito applicativo, (c) *Data Federation*, non prevede lo spostamento fisico dei dati in un database unificato, ma la realizzazione di una vista unificata virtuale, recuperando i dati dai singoli database di origine per via telematica. Tali metodologie di integrazione trovano applicazione nei primi prodotti di Data Integration. Il concetto di Data Consolidation viene infatti applicato nella tecnologia del Data Warehouse, che, tramite

l'utilizzo di Data Mart dedicati alle singole applicazioni, evolve nell'ambito della Data Propagation. L'approccio definito Data Federation da invece vita alla tecnologia dei database federati. Per ognuna delle tecnologie individuate sono descritte le architetture utilizzate, gli schemi dei dati e ne vengono analizzati pregi e difetti cercando di capire in quali contesti applicativi una metodologia sia migliore dell'altra. In base all'analisi condotta l'approccio di Data Integration più completo ed affidabile appare essere quello del Data Warehouse. Tuttavia, le evoluzioni più recenti evidenziano come il Web e le sue tecnologie siano sempre più utilizzate per la gestione e la fornitura di dati tramite web-service rendendo di fatto sempre più numerose e variegate le fonti di dati esistenti. In un contesto di questo tipo, con migliaia di fonti di dati differenti, una soluzione di Data Warehousing diventa irrealizzabile e l'approccio federato si dimostra senza dubbio preferibile.

La tesi prosegue con l'analisi delle evoluzioni più recenti dei prodotti di integrazione dei dati, analizzando il total cost of ownership dei moderni software di Data Integration. Se in un primo momento (anni '90) le tecnologie viste nella prima parte della tesi erano vendute come prodotti singoli, specializzati per quella particolare metodologia di integrazione, l'evoluzione del mercato negli anni 2000 ha portato alla creazione di prodotti di Data Integration composti da un insieme di tecnologie e metodologie differenti che operano in un ambiente unico. I moderni prodotti di integrazione presentano infatti un ampia gamma di funzionalità, tra le quali le più importanti sono le seguenti: acquisizione di dati per la realizzazione di Data Warehouse e analisi di Business Intelligence, creazione di Master Data Stores integrati, trasferimenti e conversioni/trasformazioni di dati, sincronizzazione di dati tra applicazioni differenti, creazione di database federati partendo da sorgenti multiple e distribuite, fornitura di servizi sui dati tramite architetture Service-Oriented (SOA), fornitura di servizi secondo il modello Software-as-a-Service (SaaS), aggregazione di dati strutturati con dati non strutturati. Inoltre, tali prodotti forniscono funzionalità aggiuntive, come ad esempio: gestione della data quality, funzionalità di data clearing, data modeling, metadata management, ecc.

Viene quindi proposta un'analisi del mercato relativo a tali tipologie di prodotti, distinguendo l'approccio commerciale dall'approccio open source che recentemente sta ottenendo un discreto successo nel mercato della Data Integration. Di seguito vengono analizzate funzionalità e caratteristiche dei principali prodotti commerciali (IBM, SAP-Business Objects, Informatica, Microsoft, Oracle e Pervasive Software) e di alcuni prodotti open source (Talend Integration Studio, SnapLogic DataFlow, CloverETL, XAware, Apatar). Questa parte della tesi si chiude con un'analisi dei limiti di questa tipologia di prodotti (costi elevati, scarsa flessibilità, complessità dei prodotti, modelli dei dati scarsamente rappresentativi, ecc.). Limiti che vengono accentuati dalle recenti evoluzioni del contesto informatico (aumento esponenziale della quantità di dati, dati sul Web, alta variabilità dei moderni contesti di business, ecc.), rendendo di fatto tali prodotti deboli sotto alcuni aspetti. Per questo la tesi si concentra successivamente sull'analisi di metodologie e tecnologie di integrazione innovative,

che siano in grado di semplificare le operazioni di integrazione dei dati e di superare i limiti tipici dei prodotti tradizionali. È quindi proposta un'analisi delle tecnologie semantiche che, tramite l'applicazione dei linguaggi tipici del Semantic Web, cercano di proporsi sul mercato come soluzioni potenti e innovative nell'ambito della Data Integration, sia come soluzioni alternative ai prodotti tradizionali sia come elementi complementari. In particolare, è condotta un'analisi sul modello semantico dei dati (modello a grafo), evidenziandone le differenze e i vantaggi rispetto al classico modello relazionale, viene introdotto il concetto di ontologia, cercando di capire in che modo potrebbe risultare utile in operazioni di integrazione dei dati e viene fatto accenno ai concetti di schema matching, semantic matching e ontology matching. L'analisi di tali metodologie/tecnologie semantiche permette quindi di definire un modello di architettura semantica per l'impresa, architettura che consentirebbe di superare i tipici problemi di integrazione di dati, assicurando l'interoperabilità tra i diversi componenti del sistema informativo aziendale. Tuttavia, tali tecnologie non hanno ancora raggiunto la piena maturità e sono viste ancora con diffidenza dalle aziende. Alcuni prodotti di integrazione semantica sono comunque già presenti sul mercato e di alcuni di questi è proposta una breve analisi (Expressor Integrator, DataLens System, Progress DataXtend Semantic Integrator). La prima sezione della tesi si conclude con un'analisi dei punti di forza delle tecnologie semantiche (modello dei dati flessibile, abilitazione del ragionamento automatico, integrazione di informazioni e interrogazione dei dati facilitate), potenzialità che garantiscono il superamento dei limiti dei prodotti tradizionali.

La seconda parte della tesi si apre con l'analisi del contesto di riferimento nel quale si sviluppa il progetto sperimentale di Data Integration. Vengono descritte caratteristiche e attività delle società che prendono parte alle attività di sperimentazione, Trentino Riscossioni S.p.A. e Informatica Trentina S.p.A., con particolare riferimento al progetto di sviluppo del sistema informativo a supporto delle attività di riscossione e gestione delle entrate. Trentino Riscossioni si presenta come una società giovane e in fase di espansione, operativa da poco più di due anni si aspetta di raggiungere la piena operatività entro fine 2010. Trentino Riscossioni si configura come una società "in-house" della Provincia Autonoma di Trento ed è dotata di un modello di governance particolare che rafforza il controllo dei soci (gli Enti che affidano servizi) sull'operato dell'azienda. Il contesto di business nel quale opera la società di riscossione è complesso e altamente variabile; un contesto che obbliga la società ad adattare i propri processi di funzionamento alle costanti evoluzioni in tema di tributi e gestione delle entrate. Per questo motivo, Trentino Riscossioni si pone come obiettivo di medio-lungo periodo quello di dotarsi di un sistema informativo sofisticato e innovativo, che sia in grado di raccogliere in una base dati integrata tutti i dati e le informazioni sulle quali si originano processi e attività dell'azienda. Tale sistema informativo, in particolare la base dati integrata, dovrà essere sufficientemente flessibile per adattarsi senza difficoltà alle frequenti variazioni del contesto di business in cui l'azienda opera. La realizzazione di una base dati integrata

sulla quale sviluppare le diverse attività di Trentino Riscossioni (ma anche di altri enti e comuni trentini) permetterà di ottenere una serie di vantaggi. In particolare, Trentino Riscossioni si aspetta di aumentare efficacia ed efficienza dei processi di gestione delle entrate, di garantire maggiore equità fiscale nel territorio trentino (e di conseguenza una riduzione delle tariffe) e di poter contare su un sistema in grado di facilitare i rapporti con il cittadino.

La tesi prosegue quindi con l'analisi del caso di reale applicazione, essenzialmente basato sulla metodologia del Data Warehousing, ma con alcuni aspetti innovativi legati all'incrocio dei dati tra sorgenti multiple mediante inferenza statistica, sperimentato con mano e sul campo con i dati del Comune di Folgaria. La sperimentazione ha visto l'applicazione di tale metodologia di integrazione nel contesto tributario locale che coinvolge Trentino Riscossioni e i 223 comuni trentini (la sperimentazione si è focalizzata su un singolo Comune ma sono coinvolte anche basi dati provinciali che raffigurano l'intera realtà trentina). In particolare, è stata sperimentata l'integrazione delle seguenti basi dati: Catasto Urbano, Catasto Fondiario, Catasto Metrico, Piano Regolatore Generale (PRG), Registro Imprese Locale (Parix), Database I.C.I., Database T.A.R.S.U., Catasto Elettrico, Catasto acquedotto, Anagrafe Residenti e Toponomastica. Il processo di integrazione dei dati è costituito da una serie di fasi tra loro collegate. Il punto di partenza è costituito dalle basi dati di origine che, a seguito di un'analisi della qualità dei dati ed a una normalizzazione dei tracciati, sono state trasferite in un ambiente temporaneo per poi essere sottoposte ad integrazione (tramite un apposito modulo che utilizza l'inferenza statistica per il matching dei record), per poi confluire nel database integrato finale. Il risultato del lavoro svolto è stato valutando secondo indicatori differenti: innanzitutto il sistema è strutturato per tenere traccia di tutte le bonifiche automatiche e manuali effettuate sui dati, in modo da poter estrarre facilmente dei flussi di dati bonificati da restituire all'ente come ritorno del processo di integrazione. Questo innesca il circolo virtuoso tra Enti e Trentino Riscossioni con la conseguenza di portare ad una migliore qualità del dato sia nei database degli Enti che nel Data Warehouse di Trentino Riscossioni. Nei dettagli, l'integrazione automatica dei dati ha permesso di recuperare automaticamente il 38% dei codici fiscali mancanti in determinate banche dati. In secondo luogo, la bontà del processo di integrazione è stata valutata in un particolare contesto operativo: l'accertamento ICI. In seguito alla definizione delle variabili e delle procedure che governano tale dominio è stata realizzata una procedura di analisi automatica dei dati in grado di restituire una lista di contribuenti identificabili come potenziali evasori. Il confronto tra procedura automatica e attività manuali di verifica ha permesso di affermare la bontà del lavoro di integrazione svolto. In particolare, per il 55% delle posizioni contributive è stato possibile calcolare con certezza che hanno pagato l'importo corretto. Questo è già un risultato importante in quanto consente al personale che si occupa dell'attività di accertamento di tralasciare il controllo di tali posizioni per concentrare le proprie analisi su coloro che non

pagano il giusto. Le attività di sperimentazione hanno permesso alla società di valutare vantaggi e svantaggi di una soluzione di questo tipo, considerando anche i costi che un tale progetto potrebbe avere qualora si coinvolgesse la totalità dei comuni trentini. Nonostante alcuni limiti riscontrati nelle fasi di collezione, scambio ed elaborazione dei dati precedenti alla fase di integrazione e alcune lacune nell'interfaccia e funzionalità offerte dal sistema, i risultati conseguiti con questa sperimentazione sono stati giudicati buoni e hanno contribuito ad estrarre dai dati, in maniera automatica, un rilevante valore aggiunto nel particolare contesto operativo dell'accertamento dell'Imposta Comunale sugli Immobili (ICI). I risultati conseguiti testimoniano inoltre la bontà del motore di integrazione che, nonostante una qualità del dato non sempre opportuna, ha permesso l'integrazione automatica della quasi totalità dei record disponibili, riconducendo il lavoro di bonifica manuale soltanto a qualche centinaio di casistiche. L'esito positivo della sperimentazione ha determinato il passaggio del progetto ad una fase successiva, ancora in fase di definizione, che nel corso del 2010 dovrà portare a regime il sistema attraverso l'integrazione dei dati degli enti che hanno affidato i dati a Trentino Riscossioni e lo sviluppo di applicativi di analisi in grado di supportare le attività della società di riscossione.

Tuttavia, a fronte delle problematiche individuate nel sistema sperimentato (scarsa flessibilità del sistema, difficoltà di espansione futura, operazioni di mappatura ed elaborazione dei dati macchinose), Trentino Riscossioni ha deciso di intraprendere una sperimentazione parallela all'attuale, sperimentando l'utilizzo di tecnologie più recenti e innovative che si pongono l'obiettivo di superare i limiti tipici dei sistemi relazionali, garantendo al sistema maggiore flessibilità e adattabilità a situazioni nuove che rappresenterebbero un notevole valore aggiunto per sistemi operanti in contesti incerti come quello di Trentino Riscossioni.

Nell'ultima parte della tesi viene quindi analizzata la possibile implementazione di tecnologie innovative che potrebbero risultare utili nel contesto di Trentino Riscossioni, contesto caratterizzato da alta variabilità ed eterogeneità di formati e schemi di dati, cercando di capire quali nuove tecnologie possono portare alla realizzazione di un sistema di integrazione più flessibile ed espandibile rispetto alla tecnologia consolidata sperimentata, valutando inoltre se tali tecnologie possono essere complementari e non sostitutive nei confronti della soluzione sperimentata. In particolare, per una delle tecnologie innovative individuate è già stata intrapresa una nuova sperimentazione con l'obiettivo di verificare l'effettiva utilità dell'utilizzo di un approccio semantico nel contesto operativo di Trentino Riscossioni e più in generale nell'ambito del sistema informativo provinciale. Si tratta della tecnologia semantica sviluppata dal consorzio OKKAM nell'ambito di un progetto europeo di cui l'Università di Trento è coordinatore. Trentino Riscossioni ha individuato nel progetto OKKAM un sistema in grado di rispondere non solo alle proprie necessità, ma di rivestire un ruolo centrale anche a livello provinciale, nell'ottica di costituire un sistema informativo provinciale in grado di favorire l'integrazione di dati provenienti da fonti differenti. La collaborazione tra Università,

Provincia e Trentino Riscossioni garantisce le competenze e le risorse necessarie per esplorare ed eventualmente intraprendere strade innovative, nell'ottica della realizzazione di un sistema informatico flessibile, espandibile e in grado di abilitare l'interoperabilità tra i diversi sistemi informativi (provinciali, comunali, ecc.). La ricerca di soluzioni innovative e di valore è quindi un'attività molto importante non solo per Trentino Riscossioni ma anche per gli altri Enti trentini, di conseguenza la ricerca di nuove soluzioni da sperimentare rivestirà un ruolo di primo piano nel progetto di integrazione dell'azienda al fine di realizzare un sistema in grado di collocare il Trentino in una situazione di vantaggio nel panorama nazionale che va verso l'adozione del federalismo fiscale.

BIBLIOGRAFIA

[1] Frank Hayes, *The Story So Far*, 2002 [Online]. http://www.computerworld.com/s/article/70102/The_Story_So_Far

[2] Cark W. Olofson, *Worlwide Database Management Systems 2009-2013 Forecast and 2008 Vendor Shares*, IDC Market Analysis, 2009.

[3] William Harvey Inmon, *The Evolution of Integration*, Inmon Consulting Services, 2008.

[4] Vincent McBurney, *Data Integration*, 2009 [Online]. http://it.toolbox.com/wiki/index.php/Data_Integration

[5] Ted Friedman, Mark A. Beyer, e Andreas Bitterer, *Magic Quadrant for Data Integration Tools*, Gartner Research, 2008.

[6] Vincent McBurney, *Data Integration Techniques*, 2009 [Online]. http://it.toolbox.com/wiki/index.php/Data_integration_techniques

[7] Mark R. Madsen, *The Role of Open Source in Data Integration*, 2009, Third Nature Technology Report.

[8] William Harvey Inmon, "What is a Data Warehouse?," in *Prism*, vol. 1, no. 1, 1995.

[9] Ralph Kimball e Margy Ross, *The Data Warehouse Toolkit*, 2nd ed. Wiley Computer Publishing, 2002.

[10] Michael Read, *A Definition of Data Warehousing*, 2009 [Online]. http://www.intranetjournal.com/features/datawarehousing.html

[11] William Harvey Inmon, *Building the Data Warehouse*, 3rd ed. Wiley Computer Publishing, 2002.

[12] Fon Silvers, *Building and Maintaining Data Warehouse*. Auerbach Publications, 2008.

[13] Ralph Kimball, *The Data Warehouse Lifecycle Toolkit*. Wiley, 1998.

[14] William Harvey Inmon, "Data Mart Does Not Equal Data Warehouse," in *DM Review*, 1998.

[15] Dennis Heimbigner e Dennis McLeod, "A federated architecture for database systems," in *AFIPS National Computer Conference 1980*, Anaheim (California), 1980, pp. 283-289.

[16] Dennis Heimbigner e Dennis McLeod, "A Federated Architecture for Information Management," in *ACM Transactions on Office Information Systems*, vol. 3, no. 3, pp. 253-278, 1985.

[17] Alon Halevy et al., "Enterprise Information Integration: Successes, Challenges and Controversies," in *ACM SIGMOD Conference*, Baltimore, Maryland, USA, 2005.

[18] Laura Haas, Eileen Lin, e Mary Roth, "Data Integration through database federation," in *IBM Systems Journal*, vol. 41, no. 4, 2002.

[19] Michael Stonebraker, *To ETL or federate? That is the question*, 2007 [Online]. http://databasecolumn.vertica.com/2007/12/to-etl-or-federate.html

[20] Jim Harris, *All I Really Need To Know About Data Quality I Learned In Kindergarten*, 2009 [Online]. http://www.ocdqblog.com/home/all-i-really-need-to-know-about-data-quality-i-learned-in-ki.html

[21] Colin White, *Developing a universal approach to cleansing customer and product data*, 2008, SAP BusinessObject White Paper.

[22] Thomas C. Redman, *Data Driven*. Boston, Massachusetts Harvard Business Press, 2008.

[23] Leo L. Pipino, Yang W. Lee, e Richard Y. Wang, *Data Quality Assessment*, Communications of the ACM, 2002.

[24] Jonathan Geiger, "Data Quality Management: The Most Critical Initiative You Can Implement," in *SUGI 29 Proceedings*, Montréal, Canada, 2004, pp. Paper 098-29.

[25] Ted Friedman, *Data Integration Technology and Architecture: Building Your Data Circulatory System*, Gartner Summit Events, 2007.

[26] Ted Friedman, *The Benefits (and Challenges) of Deploying Data Integration Tools*, 10 agosto 2007.

[27] SAP Press Conferences, *SAP to Acquire Business Objects in Friendly Takeover*, 2007 [Online]. http://www.sap.com/about/press/businessobjects/20071007_005046.epx

[28] Ted Friedman, Mark A. Beyer, e Andreas Bitterer, *Magic Quadrant for Data Integration Tools*, Gartner Research, 2007.

[29] Oracle Press Release, *Oracle Buys Sun*, 20 aprile 2009 [Online]. http://www.oracle.com/us/corporate/press/018363

[30] Oracle Press Release, *U.S. Department of Justice Aprroves Oracle Acquisition of Sun*, 20 agosto 2009 [Online]. http://www.oracle.com/us/corporate/press/029738

[31] Jordan Robertson, *EU objects to Oracle's takeover of Sun*, in Yahoo! Tech News, 10 novembre 2009 [Online]. http://tech.yahoo.com/news/ap/20091110/ap_on_hi_te/us_tec_oracle_sun

[32] IBM Press room, *IBM to Acquire Cognos to Accelerate Information on Demand Business Initiative*, 2007 [Online]. http://www-03.ibm.com/press/us/en/pressrelease/22572.wss

[33] SAP Press Conferences, *Business Objects to Acquire Firstlogic, Inc.*, 2006 [Online]. http://www.sap.com/about/newsroom/businessobjects/20060208_005974.epx

[34] Oracle Press Release, *Oracle to Acquire BEA Systems*, 16 gennaio 2008 [Online]. http://www.oracle.com/corporate/press/2008_jan/bea.html

[35] Zoomix Press Release, *Microsoft Signs Agreement to Purchase Data Quality Start-up Zoomix*, 14 luglio 2008 [Online]. http://www.zoomix.com/pressreleases_article.asp?id=26

[36] Vincent McBurney, *Gartner Versus Open Source Data Integration*, 2009 [Online]. http://it.toolbox.com/blogs/infosphere/blog-fight-gartner-versus-open-source-data-integration-29169

[37] Talend, *The Top 10 Reasons for Choosing Open Source Data Integration*, 2009.

[38] The Yankee Group Report, *Uncovering the Hidden Costs in Data Integration*, 2004.

[39] Philip Howard, *Comparative costs and uses of Data Integration Platforms*, 2008, Research paper by Bloor Research.

[40] David Waddington, *Ten Tips for Selecting a Data Integration Tool. Total Cost of Ownership Really Matters*, 2008, Research from Tyson Consulting.

[41] Neil Raden, *Data Integration: What's Next?*, 1 maggio 2008, Smart (enough) Systems Report.

[42] Patrick Ziegler e Klaus Dittrich, "Three Decades of Data Integration: All Problems Solved?," in *18th IFIP World Computer Congress*, Toulouse, France, 2004.

[43] Neil Raden, *The Emerging Role of Semantic Technology in the Enterprise*, 2006.

[44] Tim Berners-Lee, James Hendler, e Ora Lassila, "The Semantic Web," in *Scientific American Magazine*, 2001, http://www.scientificamerican.com/article.cfm?id=the-semantic-web.

[45] Emanuele Della Valle, Irene Celino, e Dario Cerizza, *Semantic Web - Modellare e condividere per innovare*, 1st ed. Pearson Addison/Wesley, 2008.

[46] Mitchell Ummel et al., "The Rise of the Semantic Enterprise," in *Cutter IT Journal*, vol. 22, no. 9, pp. 1-37, 2009.

[47] Thomas Gruber, "A Translation Approach to Portable Ontology Specifications," in *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220, 1993.

[48] Rudi Studer, Richard Benjamins, e Dieter Fensel, "Knowledge engineering: principles and methods," in *IEEE Transactions on Data and Knowledge Engineering*, vol. 25, no. 1-2, pp. 161-197, 1998.

[49] Neil Raden, *Semantic Integration: Tapping the Full Potential of Enterprise Data*, 2006.

[50] John Talburt, *Entity Resolution vs. Entity Identification*, 2009 [Online]. http://www.identityresolutiondaily.com/533/entity-resolution-vs-entity-identification/

[51] John Talburt, *The Myth of Matching: Why We Need Entity Resolution*, 2009 [Online]. http://identityresolutiondaily.com/493/the-myth-of-matching-why-we-need-entity-resolution/

[52] Peter Mork, Arnon Rosenthal, Joel Korb, e Chris Wolf, *The Harmony Integration Workbench*, MITRE, 2009.

[53] Peter Mork et al., *The Role of Schema Matching in Large Enterprises*, MITRE, 2009.

[54] H. Wache et al., *Ontology-Based Integration of Information: A Survey of Existing Approaches*, 2001, Intelligent Systems Group, Center for Computing Technologies, University of Bremen, Germany.

[55] Pavel Shvaiko e Jerome Euzenat, *Ten Challenges for Ontology Matching*, Dipartimento di Ingegneria e Scienza dell'Informazione, 2008.

[56] Semyon Axelrod, *MDM is Not Enough - Semantic Enterprise is Needed*, 20 marzo 2008 [Online]. http://www.information-management.com/specialreports/2008_69/10000964-1.html

[57] Oracle, *Oracle Database 11g Semantic Technologies, Semantic Data Integration for the Enterprise*, 2009, An Oracle White Paper.

[58] Expressor Press Release, *expressor 2.0 announcement*, 12 maggio 2009 [Online]. http://www.expressor-software.com/expressor-v2-announcement.htm

[59] Expressor Software, *Expressor™ - redefining data integration* , 2009, Expressor 2.0 product overview.

[60] Philip Howard, *Expressor Software*, 2009, A InDetail Paper by Bloor Research.

[61] Expressor Software, *expressor/Netezza intelligent load & go: low-cost, high-performance data warehousing*, 15 novembre 2009 [Online]. http://www.expressor-software.com/ilag-netezza.htm

[62] Trentino Riscossioni S.p.A., *Il modello di Governance*, 2006.

[63] Claudio De Stasio, *Compendio di diritto tributario*, 1st ed., 2004.

[64] Raffaello Lupi, *Diritto Tributario. Oggetto economico e metodo giuridico nella teoria della tassazione analitico-aziendale*. Giuffrè Editore, 2009, pp. 1-27, 139-213.

[65] Roberto Martini e Luigi Lovecchio, *ICI: guida ragionata all'applicazione dell'imposta*. Sistemi Editoriali, 2003.

[66] Heiko Stoermer, 6 marzo 2009, *More about the OKKAM Project* [Online]. http://www.okkam.org/okkam-more

[67] Paolo Bouquet, Heiko Stoermer, Daniele Cordioli, e Giovanni Tummarello, "An Entity Name System for Linking Semantic Web Data," in *Proceedings of the Linked Data on the Web Workshop CEUR*, 2008.

[68] Stefano Brida, Marco Combetto, Silvano Frasson, e Paolo Giorgini, "Tax&Revenue Service scenario for Ontology Matching," in *The Fourth International Workshop on Ontology Matching*, Washington, DC, USA, 25 ottobre 2009. http://www.dit.unitn.it/~p2p/OM-2009/om2009_poster3.pdf