Preface

INTRODUCTION

This is the fifth book of the series Advances in Information Technology and Web Engineering series containing updated articles published in volume 5 (2010) of the International Journal of Information Technology and Web Engineering, with the title of “Models for Capitalizing on Web Engineering Advancements: Trends and Discoveries.” This preface reports on current and future trends in Information Technology and Web engineering, such as the evolution of Web usage, tools and methods for linking data over the Web, cloud computing models and approaches, and finally, Information Technology infrastructure to support these new developments. Several suggestions on application scenarios, challenges, and research directions are interspersed based on these new trends.

Web 2.0 is a term used to describe the plethora of websites that exists currently addressing the needs of Internet users in the cyberspace where they can network and participate in a more interactive way. Examples of Web 2.0 based technologies are Flickr, YouTube, Twitter, Facebook, where users can share photos, videos, and interact with each other, and Wikipedia, a place where users can help to contribute to an article’s content either by editing or adding to it. Blogging is also included in the Web 2.0 family. Compared to the conventional fashion of publishing, it allows readers to share their views by commenting on it. Recently there’s a discussion of the third wave, the Web 3.0 to be followed by Web 4.0.

WEB 3.0

Web 3.0 based applications are expected to be a virtual reality location where users can try several applications interactively such as mapping and gaming. An example would be the Second Life, where more than 1 million players, including offline merchants, participate. Another application where map application, geographic information system (GIS), and global positioning system (GPS) are linked together to draw maps for a particular location for the purpose of urban planning and development, areas relocations, and utilities infrastructures. Figure 1 depicts the evolution of Webx.x from the first Web1.0, to Web 2.0, to current and future Web 3.0, and the future Web 4.0. The left side of the line shows tools, while the bottom of the line gives concepts and method.

The social graph just connects people; the semantic graph connects people, companies, places, interests, activities, projects, events, groups, multimedia, documents, Web pages, services, products, and emails. Figure 2 presents example of semantic graphs of linked data.
Figure 1. The road of the Web 2.0 (Source: www.novaspivack.com, used by permission)

Figure 2. Sample Semantic Web graph (Source: www.novaspivack.com, used by permission)
This new trend in data links uses richer semantics to enable better search mechanisms, effective targeting of marketing advertisements, smarter collaboration among different people and groups, deeper integration of linked data, richer contents by accessing data from different sources, and enhanced personalization and profiling through intelligent interfaces. In order to accomplish such links, semantic Web should depart from the traditional method on linking Web pages through a layer of metadata on the top of the Internet to the method of having the meaning reflected in the data itself. In other words, data = metadata. For example, data and Web pages and sites about Amman should use the same word “Amman” in URL names as well as in database files. The new semantic databases should permit the retrieval of all information and data related to “Amman” following the execution of a query.

The pros for the semantic Web approach include: better and more precise query execution, smarter applications with less development effort, ease of discovering and sharing of linked data between applications, and facilitates the integration of both unstructured and structured data retrieval. The cons include the need to learn new tools and technologies, metadata has to be generated or extracted, and ontologies need to be created and agreed upon.

Two paths exist to attach semantics to data and information (www.novaspivack.com, accessed October 22, 2011): bottom-up standard-based, and top-down services and tools based. The first one requires adding semantic databases all over the Web so every website becomes semantic, and the tools used are mainly Resource Description Framework (RDF) (www.w3.org/RDF/) and Ontology Web Language (OWL) (www.w3.org/TR/owl-ref/). The top-down approach, on the other hand, generates automatically metadata for vertical domains, creating services that provide this as an overlay to non-semantic Web. To achieve these objectives, new tools are used such as Twine (twine.com – later bought by Evri), Freebase (Freebase.com), and Evri (Evri.com).

The last future phase in the Web 4.0 of 2020+ is intelligence and personalization of the Web using reasoning and artificial intelligence, which leads to the creation of the intelligent Web, and where eventually the Web learns and thinks collectively. Future advances in IT and Web engineering volumes will report on new trends in this domain.

CREATING KNOWLEDGE OUT OF INTERLINKED DATA

This section reports on collaborative project in information and communication technologies that started in 2010 in Europe (http://lod2.eu).

The LOD2 is a large-scale integrating project co-funded by the European Commission within the FP7 Information and Communication Technologies Work Programme (Grant Agreement No. 257943). Commencing in September 2010, this 4-year project comprises leading Linked Open Data technology researchers, companies, and service providers from across 7 European countries and is coordinated by the AKSW research group at the University of Leipzig.

The Internet as we know it is merely a virtual library where documents are linked to each other; it is a Web of documents for humans to read. The W3C-hosted Linking Open Data initiative (LOD, http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData) now aims to extend the Web of documents with a Web of data by publishing and interlinking open data sources on the Web based on well-established standards such as RDF and URIs. A Web comprising linked data in addition to linked documents brings the advantage that the content becomes machine-readable. Computer and software can interpret the meaning of Web content instead of just offering it to the user to read. As a result, more
intelligent search engines can combine information from different sources, mash-ups integrating heterogeneous information can be more easily built, and many more, currently unforeseen creative uses of information on the Web become possible.

With more than 20 billion facts already published as Linked Open Data (LOD), the Data Web is not just a vision, but becoming reality right now. For example, all BBC programming, Wikipedia as a structured knowledge base (DBpedia), and statistical information from Eurostat and the US census are, in addition to hundreds of other datasets, readily available on the Web of Data.

Over the past 3 years, the semantic Web activity has gained momentum with the widespread publishing of structured data as RDF. The Linked Data paradigm has therefore evolved from a practical research idea into a very promising candidate for addressing one of the biggest challenges in the area of intelligent information management: the exploitation of the Web as a platform for data and information integration in addition to document search. To translate this initial success into a world-scale disruptive reality, encompassing the Web 2.0 world and enterprise data alike, the following research challenges need to be addressed: improve coherence and quality of data published on the Web, close the performance gap between relational and RDF data management, establish trust on the Linked Data Web, and generally lower the entrance barrier for data publishers and users. With partners among those who initiated and strongly supported the Linked Open Data initiative, the LOD2 project aims at tackling these challenges by developing:

1. Enterprise-ready tools and methodologies for exposing and managing very large amounts of structured information on the Data Web.
2. A test-bed and bootstrap network of high-quality multi-domain, multi-lingual ontologies from sources such as Wikipedia and OpenStreetMap.
4. Standards and methods for reliably tracking provenance, ensuring privacy and data security, as well as for assessing the quality of information.
5. Adaptive tools for searching, browsing, and authoring of Linked Data.

LOD2 will integrate and syndicate linked data with large-scale, existing applications and showcase the benefits in the three application scenarios of media and publishing, corporate data intranets, and e-Government. The resulting tools, methods, and data sets have the potential to change the current Web 2.0 to Web 3.0.

**LOD2 Partners**

The partners in LOD2 project include:

**Universität Leipzig:** Project coordination; develop knowledge structuring and enrichment algorithms as well as browsing, visualization and authoring interfaces; collaborate with OKFN in order to employ LOD2 results in the PublicData.eu use case.

**Centrum Wiskunde & Informatica:** CWI will be primarily involved in WP2 and work together with OpenLink on improving RDF data management with state-of-the-art database research approaches. CWI will be involved with a minor stake in WP5 in order to evaluate and adapt browsing and navigation in large-scale knowledge bases.
Exalead: Exalead will contribute and advance components of its search engine infrastructure, with emphasis on semantic linked data and search on linked data. Exalead search technology will be adapted and integrated as a component in the LOD2 Stack. An open search API will be developed to browse and access to the semantic linked data. As an (application) service provider in corporate environments, Exalead will lead the specification, setup, and implementation of the enterprise use case (WP8). In addition to this, Exalead will provide a prominent channel for exploiting this use case and the outcomes of the LOD2 project as a whole.

Freie Universität Berlin (FUB): Freie University Berlin will bring in expertise, tools and outreach capabilities to LOD2: (1) FUB has developed and maintains the Silk – Link Discovery Framework and Link Quality Assurance Workbench, which will be significantly extended and integrated into the LOD2 Stack. (2) FUB has developed D2R Server, the most widely used tool for publishing relational databases as Linked Data on the Web. D2R Server will be used for the domain complementation task in WP3 and will be included together with Pubby and Silk into the LOD2 Stack (WP6). (3) Within WP10 Training, Dissemination, community building, FUB will use its existing community building (initiator of W3C LOD) and outreach capabilities (Linked Data on the Web (L4DOW) workshop series, Semantic Web Challenge competition series) to maximize the impact of LOD2.

Digital Enterprise Research Institute: NUIG will guarantee technical excellence in reliable large-scale data processing with the same practices which have been daily driving the works behind the Sindice and Sig.ma projects. NUIG will provide the relevance, feasibility, and consensus of the initiative thanks to the continuous interaction between the Linked Data community and the Linked Data Research Centre, a cross institute initiative.

Open Knowledge Foundation: Adaptation of the LOD2 Stack for PublicData.eu Support with legal mechanisms for knowledge sharing and with open standards; research and consultation with end-user communities; and dissemination and communication with relevant knowledge users and providers.

OpenLink Software: OpenLink contributes in particular to developing the scalable LOD2 knowledge store (WP2); to track and infer about data provenance and reliability; to support personalized views on knowledge and spatial data; alert on data; and to contribute to standardization activities regarding the integration of semantic and spatial technologies.

Semantic Web Company: SWC adopted the tool Poolparty, a self-developed modeling tool for corporate thesauri, as a component for the LOD2 stack. SWC will provide its expertise in technology assessment and business development when it comes to evaluate the economic rationale, organizational effects, and the commercial potential of semantic Web technologies. SWC will investigate governance and regulatory issues such as competition, IPR issues, and privacy, thereby contributing to the use cases in WP 7, 8 & 9.

TenForce: TenForce brings in LOD2: (1) thorough expertise in industrial implementation of taxonomies and metadata for automatic categorization and content management, (2) hands-on experience in conducting large-scale projects in this matter, such as portals for Wolters Kluwer Europe and the European Commission, (3) the capacity to deliver product quality.

Wolters Kluwer: WKD will primarily work on adapting and evaluating the LOD2 Stack for media and publishing, as well as contribute to the PublicData.eu use case, due to the experience as a publisher with governmental information.
LOD2 Technology Stack Projects

The LOD2 technology stack projects include:

**Comprehensive Knowledge Archive Network (CKAN):** CKAN is a registry or catalogue system for datasets or other “knowledge” resources. CKAN aims to make it easy to find, share, and reuse open content and data, especially in ways that are machine automatable.

**D2R Server:** D2R Server is a tool for publishing relational databases on the Semantic Web. It enables RDF and HTML browsers to navigate the content of the database, and allows applications to query the database using the SPARQL query language.

**DBpedia Extraction:** DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. It currently already contains a tremendous amount of valuable knowledge extracted from Wikipedia. The DBpedia knowledge base will be used for evaluation LOD2’s interlinking, fusing, aggregation, and visualization components. The DBpedia multi-domain ontology will be used as background-knowledge for the LOD2 applications (WP7, WP8 and WP9), and as an alignment and annotation ontology for LOD in general.

**DL-Learner:** DL-Learner is a tool for supervised Machine Learning in OWL and Description Logics. It can learn concepts in Description Logics (DLs) from user-provided examples. Equivalently, it can be used to learn classes in OWL ontologies from selected objects. It extends Inductive Logic Programming to Descriptions Logics and the Semantic Web. The goal of DL-Learner is to provide a DL/OWL-based machine learning tool to solve supervised learning tasks and support knowledge engineers in constructing knowledge and learning about the data they created.

**MonetDB:** MonetDB is an open-source high-performance database system that allows to store relational, XML, and RDF data, downloadable from monetdb.cwi.nl. While being well-known for its columnar architecture and CPU-cache optimizing algorithms, the crucial aspect leveraged in the scope of this project is its unique run-time query optimization framework, which provides a unique environment to crack the recursive-correlated-self-join queries caused by semantic Web queries to triple stores.

**OntoWiki:** OntoWiki is a tool providing support for agile, distributed knowledge engineering scenarios. It facilitates the visual presentation of a knowledge base as an information map with different views on instance data. It enables intuitive authoring of semantic content, with an inline editing mode for editing RDF content, similar to WYSIWIG for text documents.

**PoolParty:** PoolParty is a thesaurus management system and a SKOS editor for the Semantic Web including text mining and linked data capabilities. The system helps to build and maintain multilingual thesauri providing an easy-to-use interface. PoolParty server provides semantic services to integrate semantic search or recommender systems into systems like CMS, DMS, CRM, or Wikis.

**SemMF:** SemMF is a flexible framework for calculating semantic similarity between objects that are represented as arbitrary RDF graphs. The framework allows taxonomic and non-taxonomic concept matching techniques to be applied to selected object properties. Moreover, new concept matchers are easily integrated into SemMF by implementing a simple interface, thus making it applicable in a wide range of different use case scenarios.

**Sig.ma:** Sig.ma is a tool to explore and leverage the Web of Data. At any time, information in Sigma is likely to come from multiple, unrelated websites – potentially any website that embeds information in RDF, RDFa or Microformats (standards for the Web of Data). Sig.ma is a semantic Web browser as well as an embeddable widget and also provides a Semantic Web API.
Silk Framework: The Silk Linking Framework supports data publishers in setting explicit RDF links between data items within different data sources. Using the declarative Silk - Link Specification Language (Silk-LSL), developers can specify which types of RDF links should be discovered between data sources as well as which conditions data items must fulfill in order to be interlinked. These link conditions may combine various similarity metrics and can take the graph around a data item into account, which is addressed using an RDF path language.

Sindice: Sindice is a state of the art infrastructure to process, consolidate and query the Web of Data. Sindice collates these billions of pieces of metadata into a coherent umbrella of functionalities and services.

Sparallax: Sparallax is a faceted browsing interface for SPARQL endpoints, based on Freebase Parallax. This demonstrator showcases the benefits of intelligent browsing of Semantic Web data and represents a good starting point for LOD2 interfaces developed in WP 5.

Triplify: Triplify provides a building block for the “semantification” of Web applications. As a plugin for Web applications, it reveals the semantic structures encoded in relational databases by making database content available as RDF, JSON, or Linked Data. Triplify makes Web applications more easily mashable and lays the foundation for next-generation, semantics-based Web searches.

OpenLink Virtuoso: Virtuoso is a knowledge store and virtualization platform that transparently integrates data, services, and business processes across the enterprise. Its product architecture enables it to deliver traditionally distinct server functionality within a single system offering along the following lines: Data Management & Integration (SQL, XML and EII), Application Integration (Web Services & SOA), Process Management & Integration (BPEL), and Distributed Collaborative Applications. The open-source data integration server and the highly efficient and scalable RDF triple store implementation in Virtuoso will be the basis for the knowledge store component in the LOD2 Stack.

WIQA: The Web Information Quality Assessment Framework is a set of software components that empowers information consumers to employ a wide range of different information quality assessment policies to filter information from the Web. Information providers on the Web have different levels of knowledge, different views of the world, and different intentions. Thus, provided information may be wrong, biased, inconsistent, or outdated. Before information from the Web is used to accomplish a specific task, its quality should be assessed according to task-specific criteria.

Examples of Projects

The following is a list of projects LOD2 selected from 2010:

1. Umweltbundesamt GmbH (Environment Agency Austria), Austria - Team contact: BastiaanDeblieck, Tenforce, Belgium

Abstract: The Federal Environmental Agency of Austria (UBA) is the leading expert organization for all environmental issues and media. It works for the conservation of nature and the environment and thus contributes to the sustainable development of society. Its core tasks include the monitoring, management and evaluation of environmental data. The UBA intends to learn how Linked Open Data can help them to aggregate, share and publish data. As the UBA primarily deals with measurements, statistics, research results and geo-location, they face the following specific challenges:
• How to align meta-data
• How to connect data in general
• How to geo-locate the data

2. Greater London Authority, U.K. - Team contact: Hugh Williams, OGL, U.K.

Abstract: The Greater London Authority (GLA) is home to the Mayor of London and the London Assembly. As part of its commitment to openness and transparency, the GLA has published a number of datasets in their currently available formats on its data store site (http://data.london.gov.uk/). They would now like to consolidate this effort around a concrete data model enabling a general deployment as Linked Open Data. Initially, the Greater London Assembly members’ data is being considered for the Proof of concept dataset, with some of the challenges to be addressed being:

• Is this a suitable dataset for initial publication?
• How do we model the data?
• What do we do with the temporal aspect of the data?
• What is the best format to present this data to a community of developers?
• What is the technology for publishing linked open data?

3. Deutsche Bibliographie, HistorischeKommission, Germany - Team contact: Thomas Riechert, ULEI, Germany

Abstract: The German Biography is an online project of the Historical Commission at the Bavarian Academy of Science. The original print version of two biographical lexica contains information about 47,000 biographies, including 45,000 additional persons and over 12,000 places. Funded by the German Research Foundation (Deutsche Forschungsgemeinschaft), some 55 volumes in print have already been digitalized and tagged according to TEI-P5, while all persons have been aligned to the open data authority file PND. This information is publicly available at http://www.deutsche-biographie.de/.

The aim of the current project with PubLink is to provide metadata about the individual biographies to enable the visualization of interpersonal relations, for instance. The publication of the metadata in RDF will make the retrieval of such information and inference of new statements not only possible, but also very flexible. Likewise, the integration of the biographical metadata into the Linked Data Cloud will also enhance the use of the biography for researchers. In this way, it is also an example of how European cultural heritage is merged into the digital world. The project is supported by PubLink in the transformation and publication of the data as RDF and helps to foster the establishment of a knowledge engineering methodology.

You can see the result of the support activities by LOD2/PUBLINK here: http://ndb.publink.lod2.eu.

4. InstitutoCanario des Estadística - ISTA, Canary Islands - Team contact: Michael Hausenblas, NUIG/DERI, Ireland

Abstract: The InstitutoCanario de Estadística (ISTAC) http://www.gobiernodecanarias.org/istac is the central organ of the regional statistical system and official research center of the Canary Islands. ISTAC is extending their Dissemination Environment JAXI-2 with Linked Data capabilities. JAXI-2,
based on a combination of Tomcat, Alfresco and an Oracle DB, allows the publication of meta-report statistical resources based on PC-Axis. The main open questions are around conversions of statistical data from PC-Axis and in future from SDMX as well as integration aspects into the JAXI-2 workflow.

5. Digital Agenda Scoreboard, Belgium - Team contact: BastiaanDeblieck, Tenforce, Belgium

**Abstract:** The EC Directorate General Information Society and Media (DG Infso) is one of the larger DGs in the European Commission and aims at supporting the development and use of information and communication technologies (ICTs) for cultural, societal and economic benefits. The PubLink project with the DG Infso has mainly focused on the publication of the Digital Agenda Scoreboard as Open Data with a flexible but pragmatic visualization through the support from the LOD2 consortium. This statistical information had previously been published as a report in PDF format. Yet, this format used to have extremely limited capabilities for reuse and browsing.

In May 2010, the European Commission adopted the Digital Agenda for Europe (DAE) - a strategy to take advantage of the potential offered by the rapid progress of digital technologies. The DAE is part of the overall Europe2020 strategy for smart, sustainable and inclusive growth. The Digital Agenda contains commitments to undertake 101 specific policy actions (78 actions to be taken by the Commission, including 31 legal proposals, and 23 actions proposed to the Member States) that are intended to stimulate a virtuous circle of investment in and usage of digital technologies. It identifies thirteen key performance targets to show whether Europe is making progress in this area. The present Scoreboard only addresses policy actions planned for the last twelve months in the Digital Agenda.

Readers can see the result of the support activities by LOD2/PUBLINK at: [http://ec.europa.eu/information_society/digital-agenda/scoreboard/graphs/index_en.htm](http://ec.europa.eu/information_society/digital-agenda/scoreboard/graphs/index_en.htm)

**First Release of the LOD2 Stack**

The LOD2 consortium announced the first release of the LOD2 stack available at: [http://stack.lod2.eu](http://stack.lod2.eu). The LOD2 stack is an integrated distribution of aligned tools that support the life-cycle of Linked Data from extraction, authoring/creation over enrichment, interlinking, and fusing, to visualization and maintenance. The stack comprises new and substantially extended existing tools from the LOD2 partners and third parties.

The LOD2 stack is organized as a Debian package repository making the tool stack easy to install on any Debian-based system (e.g. Ubuntu). A quick look at the stack and its components is available via the online demo at: [http://demo.lod2.eu/lod2demo](http://demo.lod2.eu/lod2demo). For more thorough experimentation a virtual machine image (VMware or VirtualBox) with pre-installed LOD2 Stack can be downloaded from: [http://stack.lod2.eu/VirtualMachines/](http://stack.lod2.eu/VirtualMachines/). More details and the instructions on installing the LOD2 Stack from scratch are available in the HOWTO Start document. The first release of the LOD2 stack contains the following components (available as Debian packages):

- LOD2 demonstrator, the root package (LOD2)
- Virtuoso, RDF storage and data management platform (Openlink)
- OntoWiki, semantic data wiki authoring tool (ULEI)
- SigmaEE, multi-source exploration tool (DERI)
• D2R, RDF wrapper for SQL databases (FUB)
• Silk, interlinking engine (FUB)
• ORE, ontology repair and enrichment toolkit (ULEI)

Online services were integrated into the LOD2 Stack: PoolParty (taxonomy manager by SWCG) and Spotlight (annotating texts w.r.t. DBpedia by FUB). The LOD2 Stack also makes use of dataset metadata repositories, such as thedatahub.org and http://publicdata.eu. Selections of datasets have been packaged and are available in the LOD2 stack repository.

The LOD2 stack is an open platform for Linked Data components. LOD2 welcomes new components. Detailed instructions how to integrate your component into the LOD2 Stack as Debian package are available in the HOWTO Contribute. For assistance or any questions related to the LOD2-stack contact support-stack@lod2.eu. From now on we will regularly release improved and extended versions of the LOD2 Stack. Major releases are expected for Fall 2012 and 2013. Just to notice that leading Web 3.0 technologies are combined in the LOD2 projects into the coherent LOD2 stack (e.g. DBpedia, Virtuoso, Sindice, Silk).

**LOD2 in a Nutshell**

Several challenges still face successful implementation of LOD-based projects. These include:

1. **Coherence**: Relatively few, expensively maintained links
2. **Quality**: Partly low quality data and inconsistencies
3. **Performance**: Still substantial penalties compared to relational
4. **Data consumption**: Large-scale processing, schema mapping and data fusion still in its infancy
5. **Usability**: Missing direct end-user tools

While these challenges will continue to be addressed by new development in LOD2 tools (figure 3) and applications, the last one will lead the way to the next generation of Internet use: the Web 4.0. By adding intelligence to the front end and have knowledge and intelligence built into the Web, a new Semantic Intelligent Web will emerge. Intelligent software agents will be faster to develop than currently done, leading to lean and light software that will be easier to surf the Internet and interact with other agents.

This preface identifies several challenges facing the realization of Web 4.0 semantic Web of knowledge and intelligence (K&I):

• Identifying relevant K&I
• Capturing such K&I
• Validation and verification of K&I to build trust in using Web 4.0
• Human factors of experts inclination not to share K&I
• Tools to codify K&I for the creation of semantic Web of K&I
• Dealing with K&I decay and maintenance

**LOD2 Major use Cases**

The following is a brief description of the three major application scenario of LOD2:
Use Case I – Media & Publishing: Large amounts of data resources from the legal domain are used to test and explore the commercial value of linked data in media and publishing. This data will be interlinked and merged automatically. Data from external sources will be used to semantically enrich the existing datasets. Adequate licensing and business models are also investigated with respect to the management of interoperable metadata.

Use Case II – Enterprise Data Web: Linked Data is a natural addition to the existing document and Web service intranets and extranets. Corporate data intranets based on Linked Data technologies can help to substantially reduce data integration costs. Using the LOD2 Stack for linking internal corporate data with external references from the LOD cloud will allow a corporation to significantly increase the value of its corporate knowledge with relatively low effort.

Use Case III – Linked Governmental Data: The project will showcase the wide applicability of the LOD2 Stack through the design, specification, implementation, testing, and user evaluation of a case study targeting ordinary citizens of the European Union. LOD2 will establish a network of European
governmental data registries in order to increase public access to high-value, machine readable data sets generated by European, national as well as regional governments and public administrations. The semi-automatic classification, interlinking, enrichment and repair methods developed in LOD2 will create a significant benefit, since they allow governmental data to be more easily explored, analyzed and mashed together.

This preface recommends scenarios for implementing Use Case II and III. The Enterprise Use Case may comprise supply chain, integration of different enterprise systems, and virtual enterprises. The government Use Case may include:

- Linking entry/exit border points of a particular country or group of related countries, such as airports and land border entry points.
- A group of government units belonging to a major ministry, such as defense and foreign affairs may link its data for faster access and security reasons.
- Custom border points may be linked to identify potential risk of material and product movement between states or countries in the same region.

CURRENT TECHNIQUES FOR LINKING DATA

In pre-LOD era, Web services (WS) technologies were used to link loosely coupled data in order to facilitate access over the Internet. Web services linking strategies include native language-based, such as XML and JEE2, and standard-based.

The core standards are related to the three basic functions of defining, publishing, and accessing WS:

- **Accessing**: SOAP (Simple Object Access Protocol)
- **Registering**: UDDI (Universal Description, Discovery and Integration) OASIS latest release 3.0
- **Describing**: WSDL (Web Services Description Language)

Extended standards supporting composite WS include:

- Messaging
- Business processes and workflow
- Database update of transactions

Other standards for supporting WS functionally include:

- Security
- Management
- Reliability
- Addressing
- Ontology, Semantics, and Metadata, and finally
Companion standards to support platform functionally:

- **Portal technology:** Coverage of portal technology may be found at net.educause.edu/ir/library/pdf/pub5006k.pdf
- **Grid computing and enterprise grid:** Coverage of grid computing may be found at www.oracle.com/us/technologies/grid/index.html
- **Enterprise Service Bus (ESB):** Complete coverage of ESB technology may be found at http://go.techtarget.com/r/15236108/10927021/6
- And all of the above could be stored and accessed through cloud computing platform

A WS may deal with one application, such as car reservation, or composite WS that deals with more than one application, such as making car, airline, restaurant, and hotel reservations in one WS application. This preface is concerned with custom developed WS applications, such as E-business, supply chain, virtual enterprises, and enterprise computing, rather than the general purpose WS applications, such as weather, currency conversion, and the like.

Figure 4 contains a proposed framework for describing the relationships among these standards:

**CLOUD COMPUTING (CC)**

The discussion on data linking approaches will not be complete without exposing new trends in cloud computing.

Cloud computing comprises three service models:

- **Software as a service (SaaS),** which delivers device-independent Web apps with Web services extensions;
- **Platform as a service (PaaS),** which hosts a development environment for mashing up composite processes and apps; and
- **Infrastructure as a service (IaaS),** which deploys high-end apps configured for elastic computing resources.

Implementing cloud computing may follow one the following approaches:

- **Public clouds** where all services provided by CC service provider
- **Private clouds** where a particular company or enterprise uses CC as a host while custom development or deployment of applications by company IT personnel.
- **Community clouds** where a limited geographical area is serviced by a CC service provider using one or more of the above stated services or as a private cloud service.
- **Hybrid clouds** where required services are deployed in more than one of the above preceding models. For example, common data center may be deployed on public clouds with some non-critical applications, critical applications deployed on private cloud, and selected services deployed and provided to other parties as a community cloud.
In such environments, cloud computing architecture may expand or shrink as needs increase or diminish. This dynamic elasticity dictates different design approaches for developing new applications and scaling and re-engineering existing applications to fit the cloud platform. A new research paradigm has emerged to deal with the selecting and evaluation of different alternatives for mapping CC services to models in light of current and future applications of a particular company. Method of evaluation may use cost/benefit analysis and multi-attribute utility models such Analytical Hierarchical Process (AHP). AHP may use software program such as ExpertChoice (www.expertchoice.com)

**Current Trends in Cloud Computing**

- Merging the clouds with ESB messaging

For example, the Microsoft Azure cloud supports relational as well as non-relational data. One new Azure aspect is represented in the Windows Azure Service Bus September Release. This software is intended to help developers build distributed and loosely-coupled applications in the cloud, as well as hybrid applications across on-premises and the cloud. Enhancements enable asynchronous cloud eventing, event-driven SOA, and advanced intra-app messaging.

In this environment, many future cloud computing applications will be completely new. However, others will attempt at relocating current systems on the clouds. For example, many enterprises want to deploy the data transformations and message brokering being done in today’s land data centers, and place that on a cloud platform. Full ESBs on cloud have been discussed. While it is not alone, Microsoft’s effort to meet architects’ messaging needs places it in the forefront of cloud vendors (Vaughan, 2011).
Security

Gartner views cloud identity management as having three different aspects. One would be identity management to the cloud – being able to send something from the enterprise to the cloud. The second would be identity management from the cloud – being able to send something that exists somewhere else, to your organizations. And the third would be identity management within the cloud to cloud (Earls, 2011).

iPlatform as a service

Gartner proposed Integration Platform as a service (iPaaS) to support services integration on the cloud. These include cloud to on-premises, cloud to cloud, on-premises to on the premises, and E-commerce B2B integration. In addition to a set of integration services, iPaaS provides cloud-based services aimed at enabling design time and runtime governance of the integration artifacts, such as process models, composition, transformation and routing rules, service interface definition, service level agreement, and policies utilized to address specific integration issues (http://go.techtarget.com/r/15209212/10927021/8).

LINKING CLOUDS TO INTEGRATION APPROACHES

In this section, the preface proposes a framework for linking integration approaches with the cloud computing platform, as in the figure 5.

IT INFRASTRUCTURE FOR THE CLOUD AND DATA INTEGRATION

This section contains current development in IT infrastructure that support cloud computing environments and data integration strategies. Both hardware made by IBM and Oracle employs blades to achieve scalability, virtualization, and elasticity. However, Oracle used Sum blades, while IBM uses its own.

IBM zEnterprise Hardware and System z Software (www.ibm.com)

Highlights of zEnterprise include:

- A “System of Systems,” integrating leading technologies from IBM to dramatically improve productivity of today’s multi-architecture data centers and tomorrow’s private clouds
- First-of-a-kind design that embraces multiple technology platforms—mainframe, UNIX® and x86, integrated within a centrally managed unified system
- Unique hybrid computing capabilities powered by the industry’s premier enterprise server, offering breakthrough innovation, virtualization and unrivalled scalability, reliability, and security
- Rapidly deploy services using prepackaged solutions and preintegrated technologies designed to meet the needs of specific workloads
The demands of customers, partners and employees to smarter computing systems—systems that raise the bar on efficiency, performance and cost savings while lowering management complexity. Furthermore, it extends the strengths and capabilities of the mainframe—such as security, fault tolerance, efficiency, virtualization and dynamic resource allocation—to other systems and workloads running on AIX® on POWER7, and Microsoft Windows or Linux® on System x®. The zEnterprise System includes a central processing complex (CPC)—either the zEnterprise 196 (z196) or the zEnterprise 114 (z114), the IBM zEnterpriseBladeCenter Extension (zBX) with its integrated optimizers and/or select IBM blades, and the zEnterprise Unified Resource Manager (http://www-03.ibm.com/systems/z/hardware/zenterprise/).

New IBM Software for System z enables hybrid and cloud environments for smarter computing, enabling clients to:

- Handle all variety of workloads
- Address the full application lifecycle
- Exploit the zEnterprise hardware and OS capabilities
- Modernize any environment

Figure 5. A proposed framework for linking cloud computing to integration strategies
System z software announcements address critical new and modern workloads (http://www-01.ibm.com/software/os/systemz/announcements/):

- **Multiplatform development and transaction processing**: Accelerate agility with intelligent application development and management
- **Business Process Management**: Agile processes and decisions to optimize business performance
- **Virtualization, optimization and risk management**: Consolidation to reduce cost, complexity and help align IT resources
- **Data Warehousing and Business Analytics**: Integrating and transforming data into trusted information, helping organizations better understand, anticipate and shape business outcomes
- **Cloud computing and Social Business**: Cloud applications to elevate business performance and productivity; Social Business solutions that enhance collaboration and community, both internally and externally

zEnterprise configured an entry level offer for deploying an Infrastructure as a Services (IaaS) Cloud delivery model for zLinux environments that enables the provisioning of zLinux images (under z/VM) through a self-service portal quickly.

*zEnterprise™ Starter Edition for Cloud deployments offers:

- Advanced automation and optional monitoring to dramatically speed new service provisioning (measured in minutes) reducing datacenter operations costs
- The industry’s highest RAS and most efficient virtualization to ensure multi-tenancy cloud deployments are continuously available
- The industry’s most secure platform (EAL 5) protecting customer and corporate data in a shared cloud infrastructure

Cloud Computing with IBM System z addresses the need to manage large data driven workloads is achieving broad adoption of cloud computing. By infusing clouds with security and manageability, companies are provided with the agility to move quickly in highly competitive environments; to activate and retire resources as needed; to manage infrastructure elements in a dynamic way; and to move workloads for more efficiency—while seamlessly integrating with their traditional computing environment. IBM Cloud Computing on System z is transforming the business, delivering proven qualities of service, advanced workload optimization, and efficient resource consolidation.

**Oracle Cloud Suites (www.Oracle.com)**

Oracle’s strategy is to offer a broad portfolio of software and hardware products and services to enable public, private and hybrid clouds, enabling customers to choose the right approach for them. Unlike competitors with narrow views of the cloud, Oracle provides the broadest, most complete, and integrated cloud offerings in the industry. Oracle offers the following cloud models:

**Public clouds (PC)**: in PC the application services area offers Fusion customer relationship management (CRM), Fusion human capital management (HCM), and social network for enterprise collaboration. It offers the following as platform services: Java and its database.
**IaaS:** For Infrastructure as a Service (IaaS), Oracle offers a complete selection of computing servers, storage, networking fabric, virtualization software, operating systems, and management software. Unlike other vendors with partial solutions, Oracle provides all the infrastructure hardware and software components needed to support diverse application requirements. Oracle’s robust, flexible cloud infrastructure supports resource pooling, elastic scalability, rapid application deployment, and high availability. The unique ability to deliver application aware virtualization and management integrated with compute, storage, and network technologies enables the rapid deployment and efficient management of public and private IaaS. The Sun ZFS Storage Appliance seamlessly integrates with Oracle VM and VMware features providing a powerful breakthrough solution for deploying storage in a cloud computing infrastructure.

**PaaS:** The Oracle Platform as a Service (PaaS) provides a shared and elastically scalable platform for consolidation of existing applications and new application development and deployment. The Oracle PaaS platform delivers cost savings through standardization and higher utilization of the shared platform across multiple applications. The Oracle PaaS also delivers greater agility through faster application development leveraging standards-based shared services, and elastic scalability on demand. The Oracle PaaS includes database services based on Oracle Database and Oracle Exadata Database Machine, as well as middleware service based on Oracle Fusion Middleware and Oracle Exalogic Elastic Cloud. With engineered systems such as Exadata and Exalogic providing extreme performance and efficiency for mixed workloads, Oracle provides the best foundation for PaaS.

**Cloud Management:** Oracle Enterprise Manager is Oracle’s complete cloud lifecycle management solution. It is the industry’s first complete solution including self-service provisioning balanced against centralized, policy-based resource management, integrated chargeback and capacity planning and complete visibility of the physical and virtual environment from applications to disk.

**Cloud Integration:** As more organizations use a mix of private and public cloud services they need a better way to integrate those services into an effective, secure hybrid cloud environment. Oracle takes a unified approach to loading and replicating data, as well as integrating transactions and business processes to ensure that organizations retain optimal control, fast time-to-market, and flexibility within their infrastructure. At Oracle, this solution for cloud integration includes Oracle SOA and Oracle Data Integration products. Oracle’s cloud integration is the only solution with the following design principles:

- **Unified:** Comprehensive and unified set of integration components seamlessly integrate on-premise and public cloud applications and services
- **Proven:** Deployed by thousands of leading organizations to ensure high reliability, real-time performance, and trusted integration
- **Open:** Leverages existing investments in Oracle database, middleware, applications and hardware systems, while working with third party cloud applications

Cloud security: Oracle can uniquely safeguard information and allow organization to benefit from the reduced costs and complexity of consolidation and cloud computing. Areas include Data Security, Identity Management, and Governance, Risk, and Compliance – offering solutions you can rely on to deploy private clouds, public clouds, or outsource, with the following benefits:

- **Complete:** Provides a comprehensive set of solutions to mitigate threats across your databases and applications
• **Proven:** Deployed by thousands of leading organizations to address compliance for multiple government and industry regulations
• **Cost-Effective:** Leverages existing investments in Oracle database, middleware, applications and hardware systems

**GRID COMPUTING AND CLOUD COMPUTING**

Redhat announced a PaaS offering called open-shift, followed by the release of Jboss Enterprise Data Grid 6, cloud integration is greatly enhanced. The enterprise data grid (EDG) is a solution that (Sharma, 2011):

• Is cloud ready
• Comes with highly scalable distributed data cache
• Will reduce response times in applications
• Will provide additional failure resilience

Given the fact that EDG is based on infispan and pieces borrowed from NoSQL etc, it supports mutitenacy, scalability, elasticity and distributed code execution. With these features, the EDG manages to achieve cloud readiness within its core, the architecture.

This preface presented current and future trends of information technology and Web engineering, such as Web 3.0 and Web 4.0, linking data strategies with coverage of a major project in Europe, cloud computing. Throughout the preface, the discussions included a proposed framework for linking clouds to integration strategies, challenges facing the realization of knowledge and intelligent semantic Web, and several suggested applications and scenarios to the implementation of clouds and data integration strategies.

This volume contains 16 chapters classified into four sections, each containing multiple articles as follows:

**Section 1:** Web Engineering Trends: Clouds, Location-aware, and Agents

This section contains three articles on using cloud computing for Web collaboration, improving location aware and search engine retrieval systems, and using ontologies to improve software agent communication.

**Section 2:** Web Engineering Discoveries

With five articles, coverage in this section include proposing a language for knowledge discovery in a semantic Web, searching beyond Web pages into deep Web retrieval system, discovering the rippling effects in Web application projects, developing finer garbage collection in object oriented systems, and modeling of ranking and selection of integrity tests in distributed databases.
Section 3: Web-engineered Applications

These applications include micro payment of peer to peer systems, virtual telemedicine and virtual telehealth, handling exceptions in concurrent workflows, and geographic information retrieval and text mining on Chinese tourism Web pages.

Section 4: Web-based Technologies for Improving QoS

This last section contains article dealing mainly with Quality of Service for Multimedia and real-Time services and in packet system as managed by third party, discussion on tool to assist in selecting COTS components to improve quality of software performance, and considerations of physical affordance to improve the design of multifunctional mobile devices.

Ghazi I. Alkhatib
Princess Sumaya University for Technology, Jordan

REFERENCES


