Johanna: A Knowledge Centered Infrastructure for Teleorganizations

Mauro Gaspari[1], Davide Guidi[1], Andrea Petrucci[3], Lucio Picci[2]
[1]*Dipartimento di Scienze dell'Informazione, University of Bologna, Via Mura Anteo Zamboni, 7, Bologna, 40127, Italy*
Tel: +390512094875, Fax: + 390512094510, Email: gaspari@cs.unibo.it
[2]*Dipartimento di Scienze Economiche, University of Bologna, Strada Maggiore, 45, Bologna, 40100, Italy*
Tel: +390512092608, Fax: + 390512092664, Email:lucio.picci@unibo.it
[3]*CSP s.c.ar.l., Via Livorno 60, Torino 10144 Italy*
Tel: +390114815176, Fax: + 390114815001, Email:Andrea.Petrucci@csp.it

Abstract: The Johanna project develops an open framework to support organizations with collaboration technologies in information technologies - rich environments. Its main features are represented by compliance with the Semantic Web guidelines, usability, attention to users' needs and open source technologies. In this paper we describe the project's goals, its architecture, its current status, and a roadmap for future developments.

1. Introduction

Johanna is an open framework providing organizations with knowledge-centered communication and collaboration technologies.

Knowledge is playing an increasingly important role in the Internet that, under the Semantic Web research agenda, is evolving toward an open and services-oriented architecture. Within this context, services are being extended with semantic information, and sophisticated methods for retrieving and composing them will be available in the future. Organizations aiming to operate in this scenario should consider knowledge not only as a tool to manage their internal competences and procedures, but also as a key feature of their interface with respect to the outside world.

Johanna provides a set of tools aiming at that goal. Its core includes a knowledge based description of an organization that automatically turns into running programs: databases, Web interfaces, and Web services. E-mail and Web services are the basic mechanisms to enable the integration of other collaborative technologies. E-mail is the basic tool for communication, and Web services provide both human readable information and machine-understandable information, according to the principles of the Semantic Web.

Johanna basic philosophy is best described by a simple motto: ``usability and users' feedback''. Usability is achieved following a non invasive approach. While the knowledge coded into the system could support a variety of synchronous and asynchronous communication tools, at the current stage of development, the user exploits e-mail as the main communication tool, and he is not constrained to a particular email program, desktop environment and/or application program.

Johanna supports collaborative work by providing the "glue" to integrate a set of collaboration components, and without requiring a fixed working environment. Thus Johanna differs from tools such as Lotus Notes [1] or Novell Groupwise [2], that constrain the user to adopt a set of groupware (often proprietary) tools.

"User's feedback" is seen as a key element for success. Too often collaborative environments that are successful from a technological perspective, ultimately fail because

users do not materialize. Users' feedback helps in securing the consensus of potential users, and it is guaranteed by the continuous monitoring of users' needs, as already experimented in a field test - the first Johanna based site has been delivered to an academic organization. Such a monitoring is carried out on the technology, in order to improve features, interfaces, etc., and, most important, on the users reactions to the technology.

In this paper we describe the project's goals firmly rooting the project to a solid analytical framework in the theory of complex organizations. Then we present the Johanna architecture, its current status, and a roadmap for future developments.

2. Objectives: Organizations, Teleorganizations and the role of Johanna.

We provide two different arguments to explain Johanna's objectives. We start with a bottom-up view, to show that in the "belly" of many organizations some current practices would be carried out more easily by means of collaborative tools of the type that we propose. After that, we look at matters from the top, by defining the concept of "teleorganization". We will show that the concept of teleorganization is useful both to highlight the requirements for using Johanna, and also to provide a broader framework for its analysis.

Let us consider a generic organization where all its members have developed the habit to check their e-mail boxes at least a few times during their work day, and they also regularly access the Web. Within similar organizations, where there is an expectation that anybody can be reached by e-mail, such messages are used for activating procedures that are part of the functioning of the organization. That is, e-mail is used not just in semi-official contexts, such as within work groups or advisory committees, but also in official communications, and this is true regardless of the presence of digital signatures.

As an example, and as an anticipation to the application that will be described in a further section, at the University of Bologna official Faculty's meetings used to be called by means of closed envelope traditional messages signed by the Dean. Today, such meetings are regularly called using email, with no (digital) signature, that is, based on the good faith of the senders and of the receivers, and on a degree of trust on the security of the communication tools that are used (or, more to the point, to the consideration that the stakes are not high enough to encourage unlawful violations of the integrity of the network or of the messages).

A big part of modern life, it has been remarked, is today spent in meetings. This piece of popular wisdom reflects the fact that, in any complex organization, there are many organs: boards of directors, committees, working groups, etc. Within the organizations satisfying the requirements that we have specified, e-mail is today widely used for the purpose of making these organs work. At a simpler level, it is used for scheduling purposes, as we have seen, sometimes substituting the traditional official letters calling a meeting. Also, it is used to exchange documents and opinions, and for preparatory work that a face-to-face meeting is then called to refine or, sometimes, just to ratify. Moreover, e-mail is used massively to organize and implement communication with the outside environment. Contacts with vendors, customers, even the press, are increasingly carried out in this way.

How does all this happen in real life organizations? Typically, in a very decentralized and inefficient manner. When sending an e-mail message to many people, most times people use the "cc" (or "carbon copy") field of the header of their email message. More refined users have learned the use of so called "aliases", and group aliases, that allow to do the copying of many addresses once and for all (as long as the list does not change). In both cases, different people needing a given list of email addresses, are responsible for setting it up and for its upkeep.

A further level of refinement is the use of mailing list software such as Majordomo, Mailman or Listproc, to name a few popular choices. Such software centralize the list of

addresses for a given mailing list, thus avoiding the duplications that we have described. Also, they typically provide useful features, such as Web archives, easy management of subscriptions, and monitoring of malfunctioning. These operations are typically carried out by a so called "list master", a person in charge of the upkeep of the mailing list, and typically in power to decide who is in and who is out.

However, mailing lists software, while being a very useful tool, are only a partial solution to the problem. Let us consider an example to clarify the point. Suppose that at "We Do It, Inc.", a private firm where everybody has frequent access to email, there are 50 mailing lists, and that Mr. Smith – an employee -, as part of his work duty, is in 20 of them. Now suppose that Mr. Smith leaves his job. He then has to be cancelled by 20 mailing lists, possibly by 20 distinct listmasters.

The solution to this problem in principle is quite simple: "We Do It, Inc." should put the information about the people who work there, appropriately structured, in a database, and the 50 mailing lists should be the result of appropriate queries to that database. When Mr. Smith leaves "We Do It", he also leaves the data base, and, by such an act, he automatically does not satisfy the query for *any* mailing list. Quite obviously, the same database can be queried in order to publish desired personal information on a Web site, such as a directory of people by office or job.

Although some mailing list managers, for instance Sympa [3], provide facilities to retrieve users information and mailing lists from a database, the integration of the mailing list services with the underlying database is loose. The database is seen just as an external data repository and is not fully integrated with the provided communication facilities. And, viceversa, a database is not aware that the mailing list manager exists. A better solution would be a framework where a database is fully integrated with the mailing lists manager. As soon as "We Do It" database schema is designed, all the relations which are necessary to the mailing-list manager would be ``automantically" generated, and, viceversa the mailing list manager should be (automatically) tailored to the specific "We Do It" scenario.

This is a first description, bottom up, of what Johanna does: organize information in a data base which is tightly integrated with a mailing list manager (and, in principle, to a host of other communication tools). The design of a database schema automatically generates a connection with a mailing list manager, and appropriate queries decide who should receive a given message. Part of the story, then, is about tightly integrating databases, mailing list managers and eventually other Web services into a unifying framework.

Before we move on to the other part of the story, it is useful to consider the matter further: the issue is not just to allow people in organizations to be more efficient, by centralizing several repositories of information, structuring information appropriately and, in this way, avoiding duplication of effort and guaranteeing a better upkeep of the information itself. With Johanna, the people in the organization are allowed to do things that they could not have done before. In order to see that, let us proceed with our example, to consider the interaction of the organization with its outside environment.

Mr Smith does not leave "We Do It" (maybe because its employer finally decided to adopt Johanna). Suppose that he is in charge of sales, and assume that, following the opening of a new local branch, "We Do It" decides to offer a special rebate for its products to the people in that area. Suppose further that "We Do It" owns a vast list of e-mail addresses, maybe of people who accessed its Web side and declared some interest. Mr. Smith contemplates about writing to the people who live nearby the new branch. If the information about them were not collected for that purpose, it may prove impossible, or very costly, to do it (maybe the right information are there, but are held by Mr Jones, who does not like Mr Smith, and querying the database is not straightforward anyway). On the other hand, if the information were collected within the Johanna framework, Mr Smith could send the desired message to the appropriate people with a few clicks of his mouse.

Johanna, then, is not just a way to "rationalize" e-mail (and other communications tools) usage". In fact, the basic idea is much more general, and has to do with the construction of a structured repository of the information that is relevant for an ample set of organizations' procedure, and of the tools to use such information appropriately. Johanna is then a "machine" to entertain relations within the organization, and between the organization and its outside environment.

We have mentioned "an ample set of organizations' procedure", but so far we have come short of declaring that the purpose of Johanna is to describe the whole organization. The scope of Johanna is an important issue when we define the domain of application of the project. We clarify it by means of a "top down" approach.

Researchers have long studied the impact of new technologies on organizations. While it can be safely said that the nature of an organization is never neutral to technological change, there is a presumption that change and innovations in the field of information technology are of particular relevance, given that they have an impact on an essential feature of organizations - the flow of information, so much so, that an organization can be seen, and described, from the point of view of the information flow within it and between it and its environment [5].

While there is a wide consensus on the fact that the new information technologies are already shaping a new organizational world, it is not easy to describe any definite pattern, besides a general increment in "interconnection" among the different atoms of the organizational world. [6]. Here we need not take a stand in this debate, and we simply describe such a process of technology adoption, using the term "virtualization", that we define as the process by which more and more things within a given organization are done without the need for-face-to meetings.

We also introduce a terminal point for such a process, by defining the concept of "teleorganization": an organization where *all* transactions are carried out thanks to the Internet. A teleorganization is an "ideal type", that is, an abstraction, and we need not worry whether any such organization exists.

An organization can always be described in terms of the information flows within it, and between it and the outside environment [5]. A teleorganization, from the point of view of information management, is also a schema describing a flow of information within an organization, and between an organization and its outside environment. In a teleorganization, we assume that such information flows are all Internet based: e-mail messages, Web services, etc., and are implemented, and accessed, in a variety of ways: from desktop computers, personal digital assistants, mobile phones.

What is then Johanna within such a teleorganization "ideal type"? It is the technology that allows such information flows to be effectively managed. Since the description of this flows is a description of a given teleorganization, Johanna represents the organization and its working, and, in a sense, Johanna *is* the teleorganization. More to the point, the *knowledge* representation whose encoding Johanna allows, *is* the teleorganization.

A teleorganization, as we stressed, is an ideal construct and not a description of a given existing organization. Real organizations, at a given point in time, will be a sort of combination of traditional components, where face-to-face interactions matter, and of a teleorganization – its virtual part. A "virtualization strategy" represents the process by which an organization becomes more similar to a teleorganization. This is what we mean when we observe real life organizations investing in Internet technologies and trying to modify their oganizational practices accordingly.

Within this framework, Johanna represents two things. At a given point in time, Johanna, when used, represents that part of the organization which is already virtual, and it describes the set od "teleorganizational practices" already in place. Moreover, Johanna is the set of technologies and practices that embody a given virtualization strategy. The set of

practices, and not just technologies: behind Johanna there is an idea of how technology enabled change management works. That is, Johanna is also a methodology – explained in the full-length version of this paper, supported by a set of technologies, to implement what has become fashionable to call “change management”.

3. The Johanna Framework

Johanna is a generic framework that can be adopted by a wide range of organizations needing support for e-mail based collaborative work and Web services. The delivery of a Johanna based site to an organization follows a knowledge-centered approach aiming to speed-up the development process.

An advanced knowledge-based editor (Protégé-2000 [4]) allows the designer(s) to give a high level description of the organization, and of its relations with the outside environment, providing an easy to use graphical user interface. This description is then exported into the Ontology Web Language (OWL) [8] describing the working of the organization, the desired flows of information within the organization and between the organization and its environment. The OWL description contains several Johanna specific classes, objects and relations and constitutes the ``Johanna Meta Language" (JML).

Given the JML description, an interactive installation process generates the glue procedures integrating Johanna's main components. These are an easy to use graphical user interface - the Johanna GUI - running on a Web server (Apache-Tomcat), using a database (Postgres or Mysql) for storing information on the users and the entities of the application domain, and a mailing-list manager (Sympa [3]). Both the Johanna GUI and the mailing-list manager use the database as a source of information. Other components, such as notification of relevant messages via mobile telephone, could also be included, but they are not at this stage of the project.

Johanna exploits a modular architecture. A core,``the Johanna core", provides the glue to integrate other collaborative technologies. The core is based on the integrated use of e-mail and Web services and on JML, and it allows the user to easily configure and customize a Johanna application. This infrastructure is the basis for the integration of other tools supporting different forms of collaboration.

3.1 The Johanna Meta Language

The JML allows the designer to customize in a machine-understandable format the main components of Johanna with the information relevant to a particular application domain. To simplify the modelling task of an organization, a generic template is included in the Johanna distribution. The designer should extend this template with all the information related to the particular domain, using the namespace provided with it. A snapshot of Protégé-2000 working on a generic organization is shown in figure 1.

The left part of Figure 1 shows some class describing an example organization. The *jObject* class and all its subclasses (*jFillInBlank, jSelectFromFile* and *jSelectFromSql*) are provided by the namespace. These three classes refer to object types - use for storing data in the database - that are threat differently by the Johanna GUI: *jFillInBlank* means that any content can be entered in this field, *jSelectFromFile* means that a drop down menu will appear, whose values are retrieved from a file, one per line. Finally the *jSelectFromSql* will shown another drop down menu, whose values are grabbed from a query to some table of the database.

Any object created by the designer must be an instantiation of a subclass of *jObject*. Every object has different properties that describe itself. Those properties may be related to the database (i.e. if a specific field represents a primary key) or to the Johanna GUI (i.e. a regular expression associated to an input field that control user input). The *Preferences*

class stores some preference related to the project, like the backend used for the database: Mysql or Postgres.

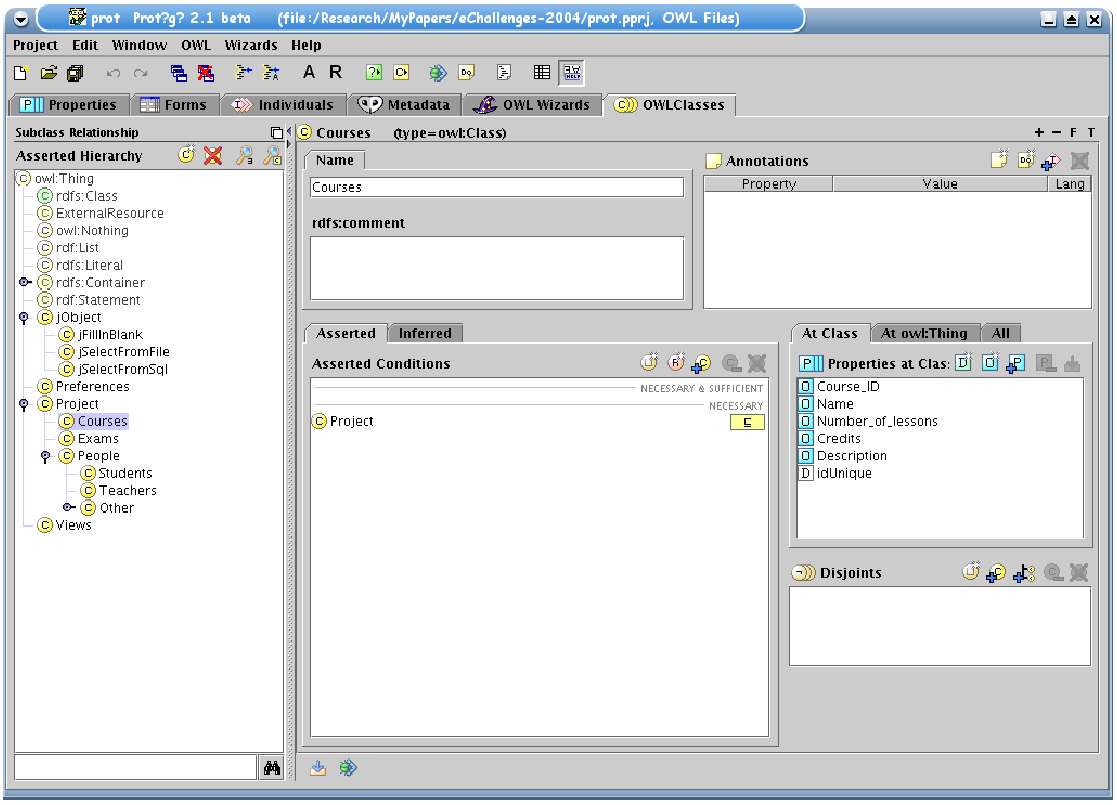

Figure 1

The Project class stores all the information about the specific organization. The *courses* object is selected, and all its properties are shown in the right part: *Course id, Name, Number of lessons, Credits, Description, [idUnique]).*

Finally the *views* class is used to store different profiles for creating several layer of visibility of the database.

3.2 The Johanna Graphical User Interface

Both the database and the graphical user interface are automatically generated from the JML. With the Johanna GUI the user can administrate the content of the database and manage the configuration of Johanna itself. Built upon the Apache web server, the Johanna GUI is still reachable from any host connected to the Internet. However permission to view only selected information is achieved through the "views" configuration. Different profiles my be created, and any physical people registered in the database can be associated to one or more profiles. A snapshot of the Johanna GUI when inserting data for the *course* object is shown in figure 2. In the left part of the figure, the structure of the organization that refers to the previous example is shown, meanwhile in the right part all the properties of the courses object are displayed. The three icons in the Main Menu allows the user to insert, modify/search and delete instances of object in the database. In the Insert Course menu the *Description* field is an instance of *jSelectFromFile*, where all others are instances of *jFillInBlank*. The green dot near *Number of lessons* means that this field is optional, while a red dot means that the field is required.

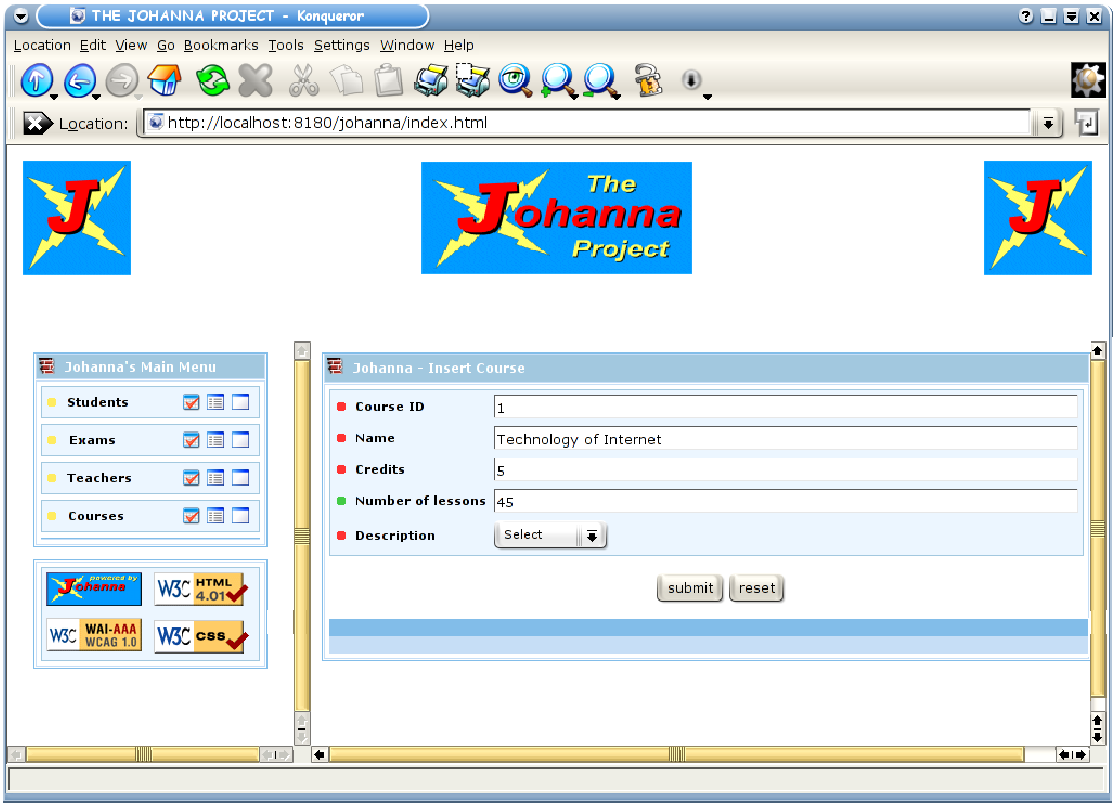

Figure 2

3.3 The Johanna Web Services

Johanna provides a set of Web services. Although the role of Web services is to make accessible the Johanna functionalities to external agents, they also constitute an application programming interface to extend Johanna with new functionalities. As an example, any functionality provided by the Johanna GUI can be exploited as a web service. Useful public services, currently under development, enable the search for particular information over a portion of the database.

4. Results

As a way to illustrate the benefits provided by Johanna, we briefly illustrate its more coherent application to date. In 2000, the University of Bologna started a new three year Bachelor Degree in Internet Economics ("Corso di laurea in Economia di Internet", http://www.ei.unibo.it). Over the last three years, Johanna has been used – and in fact, it was originally developed – to serve the needs of such an organization.

Such needs can be easily summarized. The main actors involved are: instructors, students, administrative personnel, potential students, and firms (where students are required, in their third year of study, to do an internship). An appropriate coding of the information has allowed for easy communications between instructors and their students (to each course there corresponds a mailing-list, where all students who are enrolled and have not yet passed the final exam are automatically subscribed). Such mailing lists are sometimes used for simple purposes, such as class rescheduling, etc.; some instructors are using them for what we could dub "soft e-learning": discussions on material covered in class, etc.. Other mailing lists are in use to deal with other organizational needs:

communications among instructors, various committees, and communication with the listed actors.

There are around 300 students enrolled in the program, and 25 instructors, 10 adiministrative personnel, currently actively using the system. A measure of Johanna's success is in the simple fact that people are *really* using it, because they find it useful and because it is simple: for example, an instructor knows that in order to communicate to his/her students he/she only has to write an email message to a mailing list, whose name derives from the official denomination of the course he/she's teaching.

Additional functionalities are realized exploiting the Web services interface using the encoded information in other ways. For example, the administrators of the Degree in Internet Economics are provided with a particularly useful Web-based system providing a host of descriptive statistics on grade averages, students progresses, etc.. Such a system is widely used for monitoring and planning purposes.

5. Conclusions

Johanna has been delivered and used in several academic organizations and projects. All the Johanna components are open source software systems, and the Johanna interfaces are distributed under the GPL agreement themselves.

Future development will integrate other collaborative technologies and tools, such as a chat manager or a sms messenger. We also plan to enhance Johanna investigating several directions: the integration of a distributed authentication procedure (based on LDAP), and the addition of new modules, such as a content manager and the integration with a Geographical information system (GIS).

An important issue concerns the analysis of the organizational needs that are expressed by the users, that we perform by using the analytical tools of the theory of complex organizations, in order to guarantee the necessary consensus to the organizational change that Johanna helps in delivering. To this respect, we emphasize that Johanna is an example of a technology that enables organizational change, where by "technology" we do not only mean the software, but also the set of codified practices that allow for the planning of the information flow and for the continuous monitoring of users' need.

Such a methodology requires an appropriate considerations of what each Johanna's user is allowed to do within the system. A solution, currently under study, is required presenting flexibility and granularity of users' rights.

References

[1] M.Falkner. How to Plan, Develop and Implement Lotus Notes 5.0. JohnWiley&Sons, 2000.

[2] S.B.Rogers. Novell's Groupwise 6 User's Handbook. JohnWiley&Sons, 2001.

[3] S.Aumont and O.Salaun. Smime and sympa mailing lists manager. In Linux Expo Paris. 2001.

[4] J.Gennari, M.A.Musen, et al. The Evolution of Protege: An Environment for Knowledge-Based Systems Development. Technical Report SMI-2002-0943. Stanford Medical Informatics.

[5] Morgan, Gareth. 1986, Images of Organizations,…

[6] Picci, Lucio, 1999, La Sfera Telematica, Baskerville, Bologna.

[7] Visani, Erika, 2004, La valutazione della didattica nell'università italiana. Il caso del Corso di Laurea in Economia di Internet. Dissertation in Political Science, University of Bologna.

[8] S. Bechhofer, et al., editors M.Dean, G.Schreiber OWL Web Ontology Language Reference. http://www.w3.org/TR/2004/REC-owl-ref-20040210