

## Web technologies in a collaborative platform for clinical trials

DOI: 10.3395/reciis.v3i4.236en



*Miriam Akemi  
M. Capretz*

Departamento de Engenharia Elétrica e de Computação, University of Western Ontario, Ontario, Canada  
mcapretz@eng.uwo.ca



*Maria  
Beatriz Felgar  
Toledo*

Instituto de Computação, Unicamp, Campinas, Brazil  
beatriz@ic.unicamp.br

*Marcelo Fantinato*

Universidade de São Paulo, São Paulo, Brazil  
m.fantinato@usp.br

*Shuying Wang*

University of Western Ontario, Ontario, Canada  
swang259@uwo.ca

*Olga Nabuco*

Centro de Tecnologia da Informação Renato Archer-CTI, Campinas, Brazil  
olga.nabuco@cti.gov.br

*Rodrigo Bonacin*

Centro de Tecnologia da Informação Renato Archer-CTI, Campinas, Brazil  
rodrigo.bonacin@cti.gov.br

*Itana Gimenes*

Universidade Estadual de Maringá, Maringá, Brazil  
itana@din.uem.br

*Diego Zuquim G. Garcia*

Unicamp, Campinas, Brazil  
diego.garcia@ic.unicamp.br

*David Allison*

University of Western Ontario, Ontario, Canada  
dallison@uwo.ca

*Marcos Rodrigues*

Centro de Tecnologia da Informação Renato Archer-CTI, Campinas, Brazil  
marcos.rodrigues@cti.gov.br

*Emma Chen Sasse*

Unicamp, Campinas, Brazil  
emma.sasse@uol.com.br

*Americo Brigido Cunha*

Centro Federal de Educação Tecnológica do Rio de Janeiro-CEFET, Rio de Janeiro, Brazil  
americobc@gmail.com

## Abstract

Web technologies have changed software development. The changes affect a full range of applications as well as the way users interact with computers. In the health domain, clinical research demands a lot of investment, effort and information in order to safely commercialize a new drug. The WebInVivo project aims at providing automated support for clinical research based on Web technologies. It includes mechanisms for sharing and reusing clinical trial information, such as protocols, protocol data, workflows and workflow metadata and for controlling the protocol life cycle, from modeling to execution. In this project, knowledge from the biomedical area permeates three segments of Brazilian society: (a) research and development, (b) health agents, and (c) the population. This knowledge will be made available through social networks for these segments of Brazilian society.

## Key words

web technologies; service-oriented computing; workflow management; social networks

## Introduction

Service-oriented computing (PAPAZOGLU & VAN DEN HEUVEL, 2007), with its application in web service technologies (ALONSO et al., 2004), has changed the way software systems are structured and developed. A full range of applications in business, government, health, education, culture and entertainment can take advantages of web technologies, such as functionality reuse, cost reduction and operational efficiency. Furthermore, web service technologies have also been widely adopted because of the availability of many standards, for instance the Web Services Description Language - WSDL (WSDL, 2007) and the Web Services Business Process Execution Language - WS-BPEL (ALVES et al., 2007).

Web 2.0 technologies (O'REILLY, 2005), including social networks and wikis, have enhanced the interactivity and user-generated content found on the original web (BERNERS-LEE, 1999), promoting the development of new types of applications and facilitating the way users interact with computers.

The Semantic Web (BERNERS-LEE et al., 2001) allows the Internet to be used by computer programs as well as by humans, relying on knowledge formalizations known as ontologies (GRUBER, 2007). It can be viewed as a conceptual model that has logic formalizations rigorously describing the world. This model allows the web to be used not just as a set of documents that are interrelated through hyperlinks, but also as a semantic network of concepts.

In the health domain, the low process predictability and the loose coupling of organizational units, support the argument of the collaborative platform potential in organizations.

The predictability of processes in clinical research can be classified as unframed, ad hoc and loosely framed, as defined by Dumas, Aalst, and Hofstede (2005). These processes are very distinct from the business process approach of tightly framed activities, which are common in profitable organizations. Another important characteristic of healthcare organizations is the fact that they involve the "modular, loosely coupled fashion where separate and semi-autonomous work units specialize in different areas of care delivery" (PINELLE & GUTWIN, 2006). The loosely coupled relationship between units results in a limited flow of information both throughout the organization and among research groups. The degree of understanding and accountability in the control and planning of activities is enhanced by the visibility available in the collaborative platform (DUMAS et al., 2005).

Well-known organizations, such as the Organisation for Economic Cooperation and Development (OECD) and the European Union's Seventh Framework Programme, are unanimous in proposing solutions that emphasize the need for managing and sharing the knowledge generated in the course of Research and Development (R&D) for new medicine (CALLAN & GILLESPIE, 2007; IMI, 2008). These organizations suggested creating a global virtual network, featuring a sustainable information architecture for the sharing and exchange of knowledge, data and research tools, while safeguarding the intellectual property of the research objects.

The WebInVivo project (authors' names removed, 2009; authors' names removed, 2009) focuses on clinical research and its steps to develop new drugs. Considering that clinical research

is a long-duration and expensive process, the WebInVivo software tools have been proposed to optimize the use of resources, to minimize efforts in the research for new drugs, to facilitate the information flow among researchers and to reduce time for drug commercialization.

The WebInVivo project uses web service technologies and Web 2.0 collaborative tools to meet the needs of clinical research, in particular research that supports the sharing of knowledge during experiments. The knowledge in biomedical sciences permeates three groups of Brazilian society with distinct characteristics and needs. Firstly, it affects the R&D teams working on the discovery and development of new treatments and medications. Secondly, it involves the team of health workers, such as doctors, nurses, administrators, technicians and educators. Finally, it impacts the population exposed to diseases and lacking information about them. In this project, knowledge will be available through social networks for the three mentioned segments of society according to their needs and educational levels. This challenge creates new initiatives in the areas of health R&D and of knowledge management respecting the diversity of users involved in the process of producing and consuming information.

Some projects such as MyExperiment (ROURE et al., 2008) support experiments using a Web 2.0 architecture for domains other than clinical research. Arden (CLAYTON et al., 1989), GLIF (PELEG et al., 2004) and Gaston (CLERCQ & HASMAN 2004) target at medicine applications but are not based on the service-oriented paradigm and Web 2.0 facilities for collaboration and sharing. WebInVivo goes a step further towards these new technologies. It proposes a layered architecture where each layer involves a set of components. Some components were implemented and more details can be found in papers referred to in this article.

Overall, the article proposes an architecture based on web technologies that provides support for clinical research. The remainder of this article is

structured as follows. Section 2 presents the basic concepts related to the project, while the WebInVivo architecture is discussed in Section 3. The related work and benefits of a web-based architecture are presented in the following sections. Finally, the findings and current status of the project conclude the article.

## Background

Clinical research is a critical stage in the development of drugs, and thus it requires support for the sharing and reuse of knowledge in order to accelerate the progress of experiments and optimize resources. The WebInVivo platform was proposed to meet these needs; it provides the translation of clinical research procedures into scientific workflows and the sharing of information through social networks. The basic concepts of this platform are described in this section.

## Clinical trial

The drug life development process, depicted in Figure 1, consists of several phases. One of these phases is the clinical trial, which contains a rigorous sequence of steps, each of which must be measured and registered. Phase 1 evaluates the effects of the new drugs in healthy volunteers. In Phase 2, the drug is tested for safety/efficacy in a population of patients. This may last up to two years. This phase produces comparative information about the safety of the new drug and its effectiveness. Only about one-third of experimental drugs successfully complete Phase 2. In Phase 3, the trial involves a larger test population of patients with the new drug in comparison with the standard therapy or a placebo. This phase provides a more thorough understanding of the drug effectiveness and the possible adverse reactions. Phase 3 may last several years. Seventy percent to ninety percent of drugs entering this phase successfully complete it. Depending on the results, the drug may or may not be commercialized.

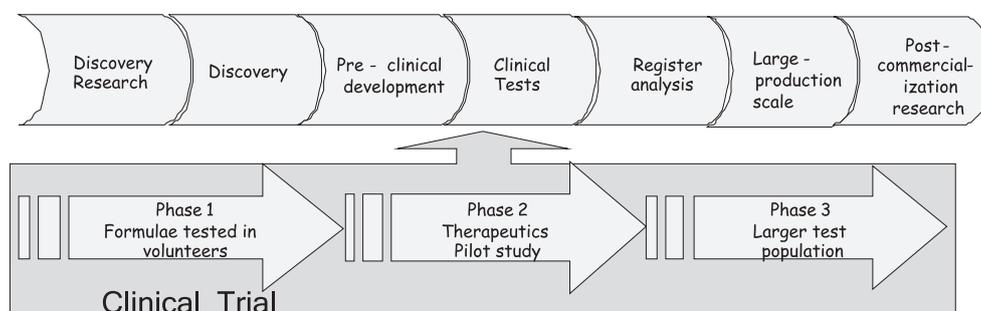


Figure 1 - Clinical trial phases.

The phases of the clinical trial must be clearly described, have scientific rigor, be verifiable and any changes have to be reported. In order to assure that these requirements are met, a protocol should be produced, approved and registered before the initiation of the trial. A clinical trial involves information collection, registration of treatment evolution and data analysis. The entire process could benefit from the development and use of novel software tools to enter data, to collect data, to follow the procedures and to visualize the results.

The envisioned solution comprises the automatization of protocol execution and the support for collaboration among possible stakeholders including the government, regulatory bodies, researchers, health agents and citizens. Collaboration is essential in order to join efforts, accelerate research processes and optimize available resources. Novel software tools ensure a greater level of data reliability and provide a more robust support for protocol execution.

### *Clinical protocols and workflows*

A clinical protocol is a study plan on which all clinical trials are based. The plan should be carefully designed in order to safeguard the health of the participants as well as to answer specific research questions. A protocol must describe the roles of the participants in a specific trial, the schedule of tests, procedures, medications and dosages, and the length of the study. In a clinical trial, the participants following a protocol have regular appointments with the research staff to

monitor their health and to determine the safety and effectiveness of their treatment. Moreover, all visits and procedures must be scheduled in advance. Currently, the protocol is written as a form, however, it can be modeled as a workflow in order to provide an overview of the procedures involved in the trial.

A workflow describes a set of activities designed to achieve a business goal. It contains an accurate description of the activity sequence that must be followed in order to achieve the particular goal. Specifically, the implementation of the workflow activities involves both individuals and computer systems. Many tools were already developed to support workflow management, which can be adapted and reused in the specific area of clinical research.

An example of a protocol for the study of a particular drug is shown in Table 1. According to this protocol, there are different procedures that apply to different phases of the study, some of which recur in more than one phase. These procedures can involve only the investigator, or they can also involve the patients participating in the study, thus requiring the exchange of information between various parties.

The first two phases in Table 1 are modeled as a workflow, using the Business Process Modeling Notation (BPMN) in Figures 2(a) and 2(b). The first phase represents the 'screening' phase, while the second phase is labeled the 'inclusion plus the first dose' phase. In each phase, tasks, gateways and sequence flows are modeled for each role, the investigator and the participant, as well as the

Table 1 - Protocol flow chart

Study Phases	Screening	"Attack" doses				Monthly doses		End of study
Visits	Screening	Inclusion + 1 <sup>st</sup> Dose	2 <sup>nd</sup> dose	3 <sup>rd</sup> dose	Evaluation of response	Odd months	Even months	Evaluation of response
Date	14 days before Start	Start	1 day after start	1 day after start	5 to 8 days after start	Months after start	Months After start	30 days after last dose
Eligibility	X	X						
Physical examination and history	X					X	X	X
Concomitant medications	X	X	X	X	X	X	X	X
Informed consent	X							
Pain assessment		X			X	X	X	X
Quality of life scale		X				X		X
Toxicities		X	X	X	X	X	X	X
Experimental drug infusion		X	X	X		X	X	

message flow between them.

Workflow technology has recently attracted a substantial amount of interest in the scientific community, in particular in the case of scientific workflow systems (TAYLOR et al., 2006). These systems allow for sharing and collaboration among scientists in the development of scientific experiments (ROURE et al., 2008). Workflow Management Systems (WfMS) provide automated support for workflow design, execution and monitoring. In addition, they interpret process definitions, interact with human actors and invoke tools and applications for performing workflow activities.

### Web 2.0 and Social Networks

A concepção original da Internet, tal com The original vision of the web, as proposed by Berners-Lee (1999), allows human interaction with a library of documents interconnected by hypermedia links. Computer programs such as browsers do the presentation of these documents, called web pages, whereas humans are responsible for linking the web pages and interpreting their content. In contrast to this previous vision, the Semantic Web (BERNERS-LEE et al., 2001) supports machine analysis and interaction among programs. This model allows the web to be used

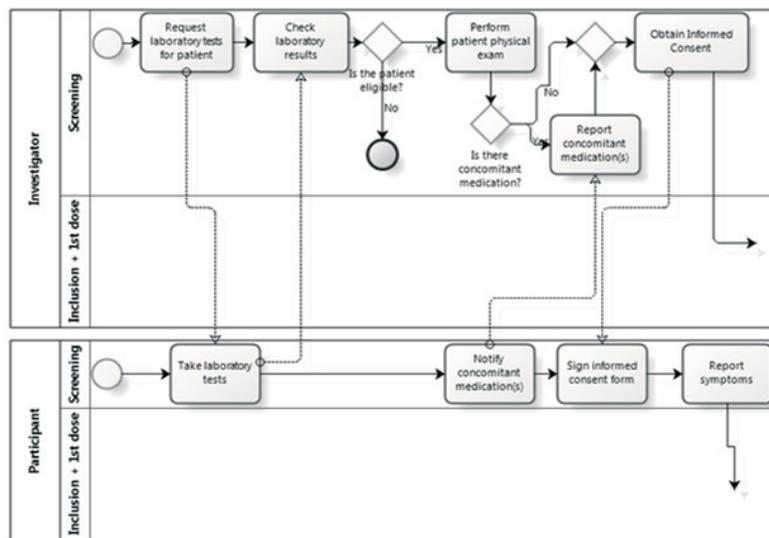


Figure 2 (a) - The protocol as a workflow in BPMN.

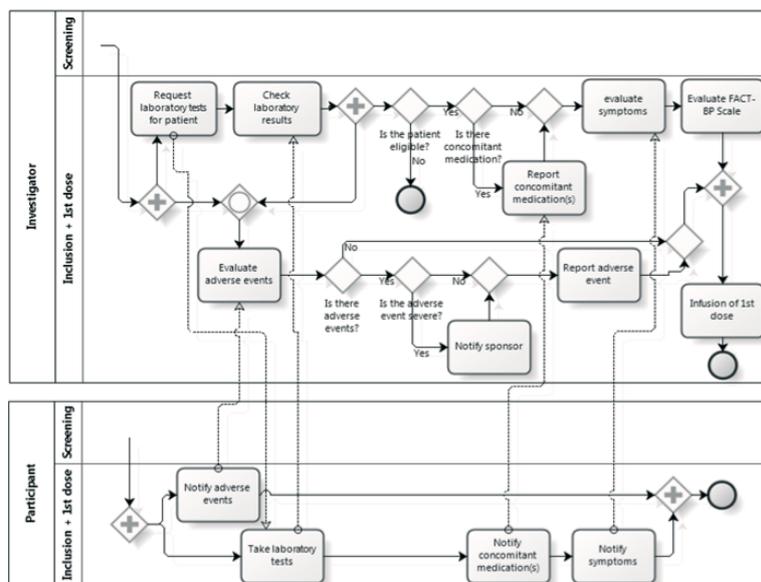


Figure 2 (b) - The protocol as a workflow in BPMN.

not just as a set of documents that are interrelated through hyperlinks, but also as a semantic network of concepts.

Web 2.0 technologies (O'REILLY, 2005), such as blogs, discussion forums, social networks and wikis, have enhanced the interactivity and user-generated content on the original web. Blogs allow individuals to produce a diary of items, to which readers can add comments, in reverse chronological order. Discussion forums offer opportunities for democratic discussions about various subjects. Moreover, social networks enable individuals to participate in similar interest groups with which they can interact online and share media. Finally, wikis allow users to cooperatively generate content by updating web pages.

A social network consists of nodes, representing individuals or organizations, which are interconnected by one or more specific types of relationships, such as friendships, interests and preferences. Specifically, the network is responsible for sharing ideas among people who have common interests, goals and values. In the medical context, but for domains different from clinical research, there are also examples of social networks, such as PatientsLikeMe (<http://www.patientslikeme.com>), which is for sharing information about diseases. Also, DailyStrength (<http://dailystrength.org>) is a site for emotional support, and Sermo (<http://www.sermo.com>) is an online forum for doctors to exchange professional opinions.

## The WebInVivo Architecture

The WebInVivo architecture supports a collaborative research environment structured in four layers, as depicted in Figure 3. These layers are:

- *Social Networks*: connects the three groups of Brazilian society targeted in this project.
- *Human-Computer Interface*: provides adaptation according to the use and requirements of specific user categories.
- *Application Logic*: consists of protocol, workflow and service management components.
- *Content and Resources*: includes wikis, protocols, workflow instances and ontologies.

The following subsections describe each of these layers.

### The social networks layer

The Social Networks layer, depicted in Figure 4, facilitates the work of three groups of Brazilian society, and includes: the Researcher Network, the Health Agent Network and the Citizen Network. The concept of "The Long Tail" (O'REILLY, 2005; ROURE et al., 2008) could be applied to these networks as the digital assets of clinical research, including protocols, workflows, web services, articles and training videos, will be developed and shared by geographically separated researchers and health workers. They will use the results as procedures for prevention, diagnosis and treatment. The "network effects" concept

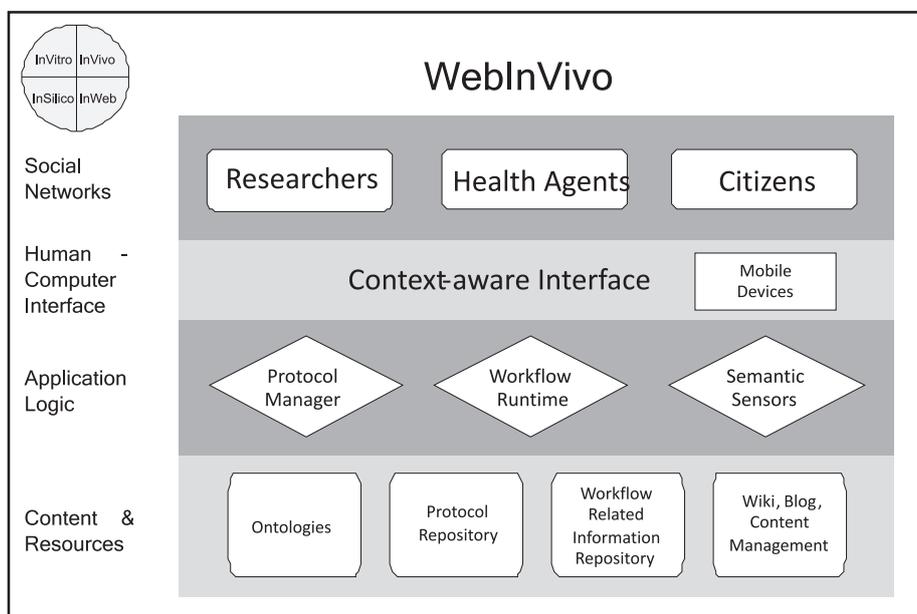


Figura 3 - WebInVivo architecture.

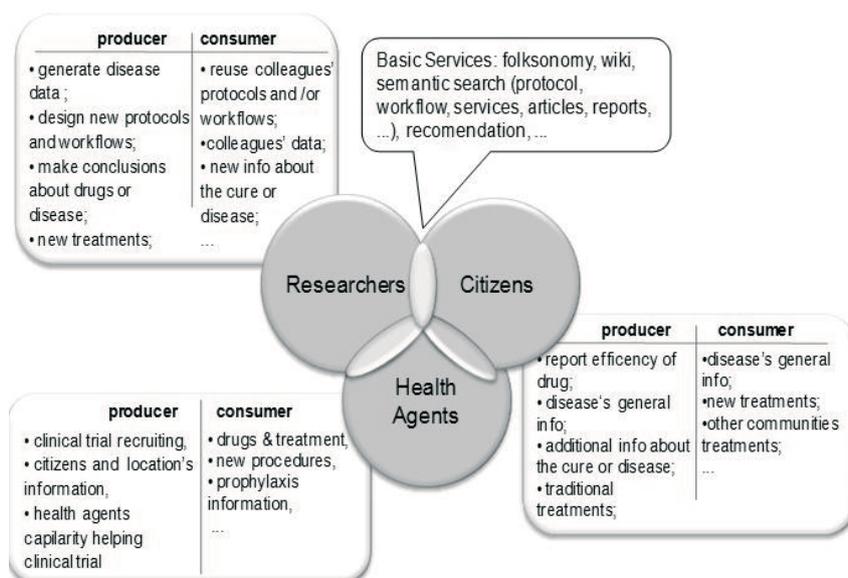


Figure 4 - The three social networks.

(O'REILLY, 2005; ROURE et al., 2008) may also be applied as participants of the Researcher Network use digital assets and aggregate data as a side effect of use. Labels are used for associations, which allow the creation of knowledge on the basis of usage. The Citizen Network should have national coverage of accessible information for prevention and treatment, with multi-channel access, including digital TV and mobile devices (SORIANO et al., 2008).

The Researcher Network will be based on the Semantic Web technology. It will enable protocol sharing, workflow sharing, workflow execution, data sharing, resource storage and search. Moreover, the Health Agent Network will have partial access to results about treatment and prevention. It may also contribute to the workflow design and protocol review by adding comments in the corresponding wikis. Finally, the Citizen Network will have access to information about prevention and treatment. It may also directly participate in research trials as agents of workflows, respond to questionnaires, and, in the future, interact through specific sensors.

The interconnections in the social networks are supported by our proposed architecture, which is based on the service-oriented paradigm. Services supporting folksonomy, semantic searches, workflows, services, articles and data permeate the networks, providing a facility to share the generated knowledge. These networks will be interconnected, generating events and alerts about crisis situations: epidemics, weather conditions, vector infestation (dengue, malaria, etc). New treatments and new

drugs can be part of the events as well. Collateral effects can be reported and reach the specific community related to it.

The assets produced and consumed by the researchers, health agents and citizens are highlighted in the rounded rectangles of Figure 4. In particular, researchers produce data about diseases, and design and develop protocols, workflows, web services and clinical trial databases; conversely, they may consume other researchers' assets. Health agents collaborate in clinical trials and search for information about new drugs, procedures and treatments. Citizens also collaborate in clinical trials, producing supplementary information, and they will search for general information about diseases.

The activities related to producing and consuming resources are supported by services in the WebInVivo architecture. These include design and runtime clinical trial support services, protocol and workflow design tools, clinical trial process management, and knowledge management support services, such as recommendation systems, folksonomy, and semantic search.

### *The human-computer interface layer*

The proposed social networks include users with various educational levels and languages, who may experience different environmental conditions and/or have different means of access. The Citizen Network can be used by any Brazilian citizen, thus it needs to take into account the vast differences in socio-economics, culture, geography, knowledge and access to technology. Social indicators presented by the Committee of

Entities Combating Hunger and for Life (COEP<sup>1</sup> in its Portuguese acronym) show that in 2003, 53 million Brazilians, or 34% of the population, lived below the poverty line. Furthermore, approximately 33 million Brazilians, or 26% of the population, are functional illiterates<sup>2</sup>, meaning that they are at least fifteen years old with less than four years of formal education. Moreover, a census conducted in 2000 found that 24.5 million Brazilians, or 14.5% of the population, have impairment<sup>3</sup>. This aspect is related to one of the current grand challenges in computer science research in Brazil for the years of 2006 to 2016, as pointed out by the Brazilian Computer Society fourth challenger: “Participative and universal access to knowledge for the Brazilian citizen” (SBC, 2006).

Thus, methods of design that result in alternative interfaces should be investigated in order to ensure that all Brazilian citizens can access and participate in this network. The use of flexible interfaces has been pointed out in the literature as one alternative to design universal accessible human-computer interfaces (SAVIDIS & STEPHANIDIS, 2004). This alternative is explored in our context in order to minimize the access barriers to the Citizen Network toward a universal accessible network. Although some barriers cannot be totally removed by using flexible interfaces (i.e. computer and internet access), they can be attenuated by adapting human-computer interfaces according to the individuals, thus enabling the access to vast and diverse users as the Brazilian citizens.

In the Health Agent Network there are also different yet equally arduous problems. Unlike the Brazilian citizens, the users in this network are professionals, and usually have at least some contact with technology. Nevertheless, the interface must also be flexible for the users to report the outcomes of the clinical trials by describing precise results in various contexts and situations. In addition, mobility and multi-device support are key requirements for this network.

As far as the Researcher Network is concerned, a major challenge is to provide interfaces that support the description of the protocols with the necessary formalism, without imposing restrictions, difficulties or additional work. In general, protocols are described using a technical language from the medical viewpoint, but they are still natural language descriptions. Terms from common ontologies are used in order to maximize the knowledge shared through the network. Appropriate interfaces that search for and reuse clinical trials based on these ontology descriptions are provided to the researchers.

Sekeres et al. (2008) present various problems related to reporting Clinical Trial Registers. However, they also emphasize the importance of the reported clinical trials to stakeholders, including patients and healthcare professionals. Therefore, WebInVivo aims to explore new solutions for interactions between humans and computers that enhance the quality of the information shared through the proposed networks. In this context, these interactions pose several challenges, such as the fact that users are not restricted to computer-literate professionals, developers are not aware of their target users, and artifacts are no longer bound to the technological specifications of a pre-defined interaction platform (SAVIDIS & STEPHANIDIS, 2004).

In this scenario, interfaces are not static artifacts designed for well-known users; rather, they should be adapted to various contexts. Consequently, the proposed platform explores and improves techniques for context-aware interfaces (MORAN & DOURISH, 2001). The context-aware applications collect and use information regarding the context in which provision of services is appropriate to particular people, place, time and events. In WebInVivo, the context is not restricted to the physical aspects of the interaction; it also considers the semantic, pragmatic and social aspects. Semantic aspects denote the meaning of the shared information, whereas pragmatic aspects refer to the intention of use and social aspects refer to conventions that delineate the interaction. We argue that it is possible to improve the quality of the interaction by analyzing, understanding, and simulating situations of flexibility in the user interface. Semiotics can provide us with theoretical and methodological fundamentals for this task.

Norms are a key concept that can help us understand intentions, communication, conversations, negotiations, beliefs, expectations, commitments, contracts, law, culture, as well as business (LIU, 2000) in semiotic terms. A norms-based simulation is proposed as a way of exploring the personalization of the interface to the usage context. Users and domain specialists provide and maintain the norms used in the simulation in order to personalize the interface.

Figure 5 shows an overview of an architecture that was constructed to allow flexibility in web applications. This architecture has two core components: NBIC (Norm-Based Interface Configurator) and ICE (Interface Configuration Environment). The architecture as a whole is based on the service-oriented paradigm, and the webservice technology. In the architecture, NBIC provides web services to manipulate norms and rules (associated

to an ontology) and to compile them. ICE has services to maintain the context information and to interpret norms (which describe the context of use) at runtime. The changes themselves are done by the Tailoring Framework, which receives an action plan as a result of the norms interpretation and modifies the interface using dynamic pages. Further details about the architecture can be found in authors' names removed (2007), and the fundamentals from the Semiotics and Human-Computer Interaction viewpoint, as well as case studies using this architecture are described in authors' names removed (2009).

We propose to adapt and expand this architecture in order to provide interfaces adapted to each social network. The proposal is to specify norms associated to features of each social network in addition to norms based on individual preferences and context. The following steps can be followed to perform the adaptation (according to Figure 5):

1. Administrators or Domain Specialists access norms modelers to specify norms. For example: "If a Researcher accesses the Social

Network then the advanced search options, using parameters from the Researcher domain vocabulary (of the Researcher area), will be enabled.";

2. This norm is stored at the NBIC;

3. The norm is translated to a computer interpretable language;

4. The interface has specific tags that mark points where the page can change;

5. The perception mechanism informs the context of use. For example: "The user is a biological engineer who is using a mobile device to access the system";

6. ICE fires the norms and infers an action plan. The action plan is a set of changes in the user interface;

7. The action plan is interpreted by the action mechanism. For example: "enable a button, change an interface division size, and include a new code from an Extensible Markup Language (XML) file".

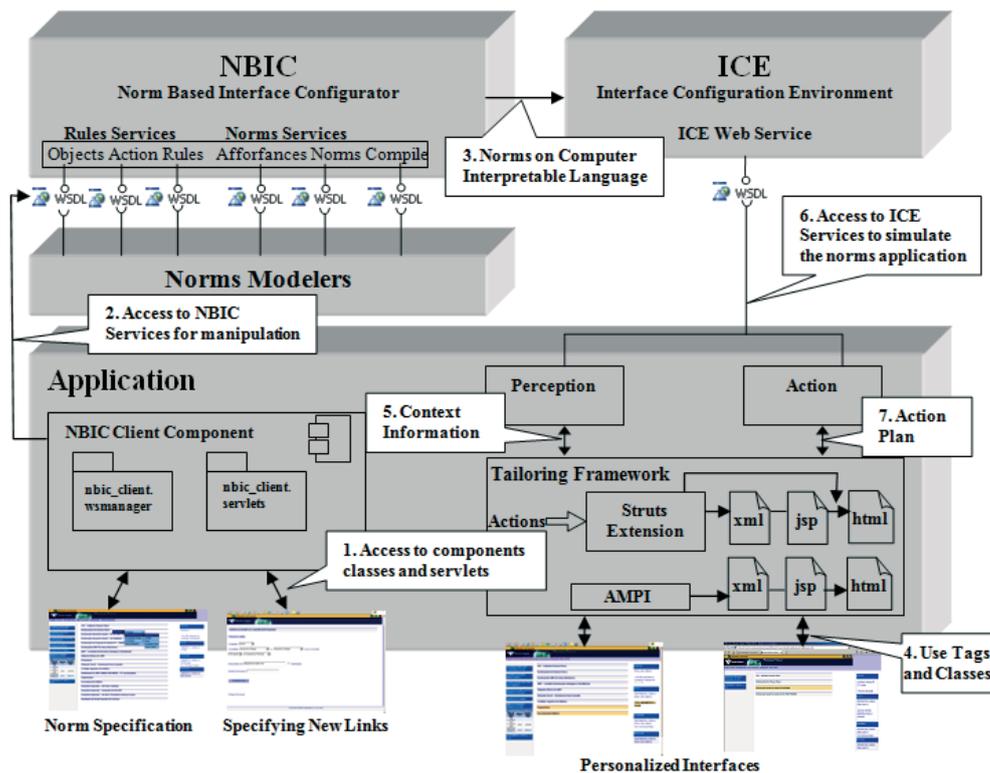


Figure 5 - Overview of the flexibility architecture.

## ***The application logic layer***

The Application Logic layer offers several services for enabling cooperation, communication, as well as knowledge sharing and construction among users in the social networks. The clinical trials will require protocol specification, protocol mapping, workflow reuse, and service discovery and composition.

### *Protocol discovery and specification*

Researchers and health agents perform experiments according to protocol specifications. Consequently, a Protocol Manager will be developed to allow the discovery and specification of protocols. This tool will enable the complete construction of new protocols or the specialization of existing ones. The language for protocol specification will be based on BPMN in order to allow for its mapping to the BPMN notation. The Protocol Manager will offer support for the link among elements in protocol specifications and ontologies. This leads to a semantics-enriched specification, which is important for the specialization of existing protocols, as it enables the consideration of semantic aspects of protocol specifications when discovering existing protocols.

### *Protocol mapping to workflow*

Protocols are mapped to workflows so that they can be executed and controlled by the platform.

Due to the complexity involved in the conception of workflows, it is important to enable the reuse of them. Reuse minimizes errors and reduces costs, as a workflow can be constructed from an existing one already designed and tested by specialists. When a workflow is well-established, it can be shared by specialists willing to perform similar experiments (ROURE et al., 2008). Confidence in the use of workflows can be improved by allowing users in the social networks to recommend workflows and to see their reputation through the networks.

One way of achieving the reuse of workflows is to use generic templates from which specific workflows can be instantiated. However, rather than simply using templates, our approach (authors' names removed, 2008) is based on a more systematic technique that improves the overall costs and benefits. In particular, the product line and software factory approaches provide concepts and mechanisms to the domain engineering unit, which facilitates the design of workflows based on feature models. Hence, specific workflows can be generated by configuring generic ones. Our aim

is to apply this technique for modeling common parts and different elements of workflows in order to enable their configuration in specific cases (authors' names removed, 2008).

Although this approach is available for use in an isolated form, including a toolkit to provide automatic support for it, it must still be integrated to the other components. The toolkit includes tools that support: the generation and configuration of feature models and the exportation of XML files; the generation of workflow templates, specified in the WSDL and WS-BPEL languages, based on such feature models; and, the instantiation of different workflows based on the feature model configurations.

### *Service composition*

Our approach considers a workflow as a service composition. To allow for the composition and execution of services, the proposed platform provides a service composition engine and a broker (authors' names removed, 2008). The service composition engine uses the broker to select service implementations in a registry according to service functionalities, semantic descriptions and Quality of Service (QoS) policies (VEDAMUTHU et al., 2007).

Service composition focuses on the combination of web service semantics and service registries, such as the Universal Description, Discovery, and Integration - UDDI (2004), in order to support dynamic service discovery. For example, the Web Ontology Language for Services - OWL-S (2004) is a semantic web service language using upper ontology. OWL-S provides discovery and logical reasoning in the service profile model and process ontology in the process model.

Overall, we propose three types of semantic web services composition for clinical research: workflow, ontology, and privacy.

*Workflow-based service composition.* Service composition can be achieved by using the semantics of the process. For example, by publishing the semantic process descriptions to the service registry, the broker can discover a relationship between the generic templates and executable processes. In particular, we use a semantic process profile as a semantic description of the process activities associated with a web service. It introduces the Input, Output, Precondition and Effect (IOPE) descriptions for process activities. Moreover, the semantic process profile can use a domain-specific ontology as a parameter for describing the activity requirements. When a process requirement originates from the service requester, the broker

retrieves the generic template and its related semantic process profiles. Subsequently, the process profiles are sent to the service matchmaker to find relevant services. Therefore, although the clinical researchers may use different workflows, they can publish these workflows as services and compose the resulting services using a generic workflow template.

*Ontology-based service composition.* As a popular semantic form of knowledge representation in clinical research, ontologies are used to represent data models with the aim of interoperating medical information. However, directly mapping the ontologies, in particular the local ontologies that represent the data models of applications, is complex. Therefore, we have introduced a reference ontology-based approach that uses a combination of web services. In this approach, each application can publish its own ontologies as web services, and the other applications or services can retrieve these data models and find the correspondence between local ontologies through reference ontology matching. This approach can increase the number of possible service compositions, since even if applications have different data models, they can still interoperate if a correspondence between their ontologies is discovered. Thus, data generated by different applications as a result of their execution can be integrated in order to complete a service composition. This is accomplished by considering the semantic information offered by the reference ontology, which allows for the mapping between data models represented by different ontologies. Therefore, data integration in the clinical research is realized through semantic web service composition.

*Privacy-based service composition.* The use of non-functional characteristics to compose services is important in clinical research. Privacy, for example, is an important aspect in the context of this project. Service capabilities and consumer requirements regarding privacy should be considered in order to include privacy in service composition. In this approach, services must have privacy policies. These policies define how the services protect the privacy of the service consumers. Moreover, consumers need to specify policies describing their privacy preferences.

To specify privacy policies, the Web Services Policy Framework - WS-Policy (VEDAMUTHU et al., 2007), which is a standard for web service policy specification, can be used. It offers a policy model to define different policy types. Specifically, a privacy vocabulary has to be defined. This vocabulary, along with the policy model, can be

used to define privacy policies, which include elements such as *collector*, *what*, *purpose*, *retention* and *recipient* (authors' names removed, 2009).

The first of these elements, *collector*, allows the service provider to state the name of the organization or party who will be collecting the data. On the other hand, the service consumer uses the *collector* element to state who it will allow to collect its data. Moreover, the second element, stated as *what*, allows the provider and consumer to outline what data they want to collect and provide respectively. The third element is *purpose*, which enables the provider to describe the reasons for collecting the data. Conversely, the consumer uses the *purpose* element to describe the allowable reasons for collecting their information. *Retention*, the fourth element, allows both parties to signify the length of time for which they want the data to be stored. The fifth and final element is *recipient*, which enables the service provider to list all the parties to which it could pass the information it has collected. Together, the five privacy elements create a single privacy rule; a privacy policy can consist of one-to-many rules.

An example of a service privacy policy in WS-Policy is illustrated in Figure 6. The policy indicates that all providers involved in the service execution (Line 06) may use all data provided to the service (Line 04) to enhance, evaluate or review it (Line 05).

```

01 <Policy>
02   <ExactlyOne>
03     <All>
04       <InputParameters>
05         <Develop/>
06         <Executors/>
07       </InputParameters>
08     </All>
09   </ExactlyOne>
10 </Policy>

```

Figure 6 - Example of a privacy policy.

Thus, the policy of each potential component service and the preferences of a consumer can be matched to build a service composition at the privacy level.

### **Content and resources layer**

In terms of semantics, ontologies support the Application Logic and Human-Computer Interface layers. For instance, a domain ontology

can be part of the Researcher Network interface, providing a specialized vocabulary and the reuse of a UMLS (Unified Medical Language System) thesaurus (UMLS, 2008). In this case, ontologies will be either developed or reused and adapted.

When a researcher enters information, opening a new clinical trial and naming the disease and the experiment, an information pop-up window can notify the researcher of experiments in the same domain. Services that support the development of the protocol, thus alerting a researcher about available workflows and web services, will be part of this layer. These services will enable the possible reuse of existing knowledge in the networks, which results from a search and classification service that runs in background.

The clinical research process comprises a sequence of steps, which can be described as a chain, and contains a specific vocabulary. The result of a step can be the input for another or for parallel steps. This method will result in an ontological process description, with input and output parameters as well as events, which can be linked to others. This creates relationships that are also described by ontologies.

Ontologies are also necessary to connect the different social networks. Semantic equivalence can be used to discover distinct vocabularies with the same meaning. Semantic sensors, including previously known words, group of words and synonyms, can interconnect the networks, sending an alert among them. Ontologies will be also conceived to add information to information, such as tagging and recommending documents or workflows.

## Related work

Computer-Supported Collaborative Work (CSCW) initiatives to support scientific interactions and collaboration are not new although the large-scale use of these systems still faces several barriers among researchers. In clinical research, the skepticism is complemented by the complexity and privacy requirements during the research development. Firstly, CSCW researchers have focused on the modeling of “office work” to improve the collaborations during the execution of organizational processes (LUBICH, 1995). In business organizations, the social network applications are designed to improve the non-formalized interactions that are not supported by the workflow systems. In clinical research, it would be useful if a workflow model, or part of it, could be handled as communications and interactions are in the “regular” social network applications.

As WebInVivo, other breakthrough initiatives have been developed in order to promote and facilitate the collaboration among researchers. One good example is the MyExperiment (ROURE et al., 2008) that has expanded the use of social networks and workflow systems into the research development process. The “spectrum” tasks supported by MyExperiment include well-structured procedures and protocols as well as the ones that are created during the research development process.

The social network MyExperiment (ROURE et al., 2008) shares scientific workflows using the Web 2.0 architecture. Workflow search and exposition are based on Web 2.0 techniques such as tags, recommendations, citations, reviews and blog discussions. Furthermore, the use of wikis in the biomedical area is becoming widespread in order to facilitate protocol sharing. One example of this is OpenWetWare (OPENWETWARE, 2008), built by non-specialists on Web 2.0 and the service-oriented paradigm. WebInVivo also uses Web 2.0 techniques and the service-oriented paradigm, but expands on these techniques by using domain ontologies within the Web 2.0 architecture.

Web semantics with rule languages, inference models and web services, assist the characterization of information based on a tagging ontology, which allows filtering and content recommendation (GRUBER, 2007). UMLS (2008) is another example of knowledge sharing based on domain ontologies and knowledge tools, such as a thesaurus, that exist in a web environment. This work proposes an ontology-based approach for collaboration. WebInVivo employs this approach, together with the workflow technology and the service-oriented paradigm, in order to provide a comprehensive framework for clinical research.

One of the first projects dealing with protocol modeling in computational systems was Arden (CLAYTON et al., 1989). Other life cycle phases, including protocol test and execution, have been studied in GLIF (PELEG et al., 2004) and Gaston (CLERCQ et al., 2004). Although WebInVivo deals with protocol life cycle, it is based on more advanced technologies, such as the service-oriented paradigm and Web 2.0 facilities for collaboration and sharing. In fact, the direction towards a service-oriented architecture has already been stated by European (IMI, 2008) and North-American (KAWAMOTO & LOBAC, 2007) agencies.

## Benefits from web technologies

The proposed architecture, based on web technologies, can offer a collaborative research

environment with the possibility of sharing knowledge and assets among users in the social networks.

WebInVivo can provide contributions for clinical research. First, due to protocol reuse, there is a time and cost reduction in the production of new drugs. Also, by using workflow technology, responsible actors can be notified about deadlines and other activities associated with protocol execution. In addition, there is a reduction in error as well-established workflows can be shared to perform similar experiments. Finally, there is increased control and reliability with patients participating in clinical trials through the social network.

There are a number of contributions associated with social networks, one of which is the construction and dissemination of knowledge through selectively interconnected social networks. In addition, the architecture results in a promotion of the population affected by diseases as well as the detection of problems and risk areas through reports by health agents and by the population. This leads to the extension and reuse of existing protocols by researchers and doctors. Lastly, it results in collaboration among research groups and the promotion of e-learning courses.

WebInVivo will provide additional contributions, including interfaces that adapt to different contexts. The researchers can describe the protocols with the necessary formalism through an interface, without imposing restrictions or additional work during this task. Also, it extends the service-oriented paradigm for social network applications and the collaborative research environment using web semantics techniques. Since the clinical trials require protocol specification, protocol mapping, workflow reuse, and service discovery and composition, the platform enables cooperation, communication, as well as knowledge sharing and construction among users in the social networks. The final contribution involves ontologies that support interoperability among the different social networks.

## Conclusions and future work

The WebInVivo platform aims at supporting clinical trials and enabling the sharing of protocols in order to perform clinical experiments. Some of the relevant contributions in IT are going to be related to mechanisms that map protocols to workflows, reuse workflows and compose workflow activities implemented by web services. Moreover, social networks will facilitate the construction and sharing of knowledge about prevention and

treatment in an accessible way to different segments of Brazilian society.

In this article, we identify clinical research requirements and propose a service-based architecture for a collaborative environment that enables information sharing through social networks.

WebInVivo is a proposal whose implementation is going to include the integration of existing components that implement functionalities of the Application Logic layer. Moreover, its other layers have been proposed in order to support the specialization of the Application Logic layer to the domain of clinical trials. The components that must be implemented so that each layer can perform the required functionalities have been described in this article.

Future work includes extending the platform to allow individuals access to the social networks through mobile devices such as personal digital assistants and cell phones. To meet this requirement, the infrastructure should be adapted to provide configurability and customizability for various types of mobile devices and applications, as well as resource awareness and runtime adaptability for managing dynamic changes in the environment. Specifically, middleware techniques for mobile devices will be considered in order to achieve these goals (authors' names removed, 2007).

This work has been carried out by universities and a research center and financed by Brazilian and Canadian agencies in the form of graduate scholarships.

## Notes

1. <http://www.coepbrasil.org.br/>
2. <http://portal.mec.gov.br>
3. <http://mj.gov.br>

## Bibliographic references

ALLISON, D. S.; EI YAMANY, H.; CAPRETZ, M. A. M. Metamodel for Privacy Policies within SOA. Proceedings Of The 5th International Workshop on Software Engineering for Secure Systems (SESS'09), Vancouver, Canada, May 19, 2009.

ALONSO, G.; CASATI, F.; KUNO, H.; MACHIRAJU, V. Web services: Concepts, Architectures, and Applications. Springer Verlag, Berlin, 2004.

ALVES, A. et al. Web Services Business Process Execution Language Version 2.0 Specification. OASIS, Apr. 2007. <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>, accessed in 01/2008.

- BERNERS-LEE, T. Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor. Orion Business Books. <http://www.w3.org/People/Berners-Lee/Weaving>, 1999.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American*, 284(5):35-43, 2001.
- BONACIN, R.; BARANAUSKAS, M. C. C.; dos SANTOS, T. M. A semiotic-based framework for the development of tailorable Web applications. *Lecture Notes in Computer Science*, 4553:859-868, 2007.
- BONACIN, R.; BARANAUSKAS, M. C. C.; LIU, K.; SUN, L. Norms-based simulation for personalized service provision. *Semiotica, Journal of the International Association for Semiotic Studies*, 175(1/4):403-428, 2009.
- CALLAN, B.; GILLESPIE, I. The Path to New Medicines. *Nature*, 449(7159):164-5, 2007.
- CLAYTON, P.D.; PRYOR, T.A.; WIGERTZ, O.B.; HRIPCSAK, G. Issues and structures for sharing knowledge among decision-making systems: The 1989 Arden Homestead Retreat. In: Kingsland LC, editor. In *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care*, pages 116-21. IEEE Computer Society Press, 1989.
- CLERCQ, P.; HASMAN, A. Experiences with the Development, Implementation and Evaluation of Automated Decision Support Systems. *Stud Health Technol Inform*, 2004, [http://www.openclinical.org/gmm\\_gaston.html](http://www.openclinical.org/gmm_gaston.html), accessed on 10/2008.
- DUMAS M.; AALST, W., HOFSTEDE A. (eds) *Process-Aware Information Systems: Bridging People and Software Through Process Technology*. Wiley InterScience, Hoboken, NJ, US, 2005.
- FANTINATO, D. Z. G.; TOLEDO, M. B. F.; GIMENES, I. M. S. WS-Contract Establishment with QoS: an Approach Based on Feature Modeling. *Int. J. Cooperative Inf. Syst.*, 17(3):373-407, 2008.
- FANTINATO, M.; GIMENES, I. M. S.; TOLEDO, M. B. F. A Product Line for Business Process Management. In *Proceedings of the 12th International Software Product Line Conference*, Limerick, Ireland, 2008.
- GARCIA, D. Z. G.; TOLEDO, M. B. F. A Privacy Framework for Web Services. In *Proceeding of the Web2Touch, IEEE 11th International Conference on Computational Science and Engineering*, 2008.
- GRUBER, T. Ontology of Folksonomy: A Mash-up of Apples and Oranges. *International Journal on Semantic Web & Information Systems*, 3(2), 2007.
- KAWAMOTO, K.; LOBAC, D. Proposal for Fulfilling Strategic Objectives of the U.S. Roadmap for National Action on Decision Support through a Service-oriented Architecture Leveraging HL7 Services. *J Am Med Inform Asso.*, Mar-Apr; 14(2), 2007.
- LIU, K. *Semiotics in information systems engineering*. Cambridge: Cambridge University Press, 2000.
- LUBICH, H. TOWARDS. a CSCW Framework for Scientific Cooperation in Europe (Lecture Notes in Computer Science). Berlin: Springer-Verlag, 1995.
- MORAN, T; DOURISH, P. Introduction to the special issue on Context-Aware Computing. *Human-Computer Interaction*, 2(3), 2001.
- OpenWetWare, [http://openwetware.org/wiki/Main\\_Page](http://openwetware.org/wiki/Main_Page), accessed in 10/2008.
- O'REILLY, T. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, accessed in January 2009. <http://www.oreillynet.com/go/web2>, 2005.
- OWL-S Web Ontology Language for Web services <http://www.w3.org/Submission/OWL-S/>, 2004.
- PAPAZOGLU, M.P; VAN DEN HEUVEL, W.-J. Service Oriented Architectures: Approaches, Technologies and Research Issues. *VLDB Journal*, 16(3):389-415, 2007.
- PELEG, M. et al. The InterMed Approach to Sharable Computer-Interpretable Guidelines: a Review. *J Am Med Inform Assoc.*, Jan-Feb; 11(1), 2004.
- PINELLE, D., GUTWIN C., Loose Coupling and Healthcare Organizations: Deployment Strategies for Groupware. *Journal of Computer Supported Cooperative Work* 15:537-572. 2006.
- ROCHA, T; ARNTSEN, A.-B.; EIDSVIK, A. K.; TOLEDO, M. B. F.; KARLSEN, R.. Promoting Levels of Openness on Component-Based Adaptable Middleware. In *Proceedings of the 6th Workshop on Adaptive and Reflective Middleware*, Newport Beach-CA, United States, 2007.
- ROURE, D.; GOBLE, G.; STEVENS, R. The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflows, <http://www.myexperiment.org/>, accessed in 10/2008.

- SAVIDIS, A.; STEPHANIDIS, C. Unified user interface development: the software engineering of universally accessible interactions. *Universal Access in the Information Society* 3:165-193. 2004.
- SBC. Grand Challenges in Computer Science, 2006 – 2016. <http://www.sbc.org.br/>, accessed in 06/2009.
- SEKERES, M. et al. Poor Reporting of Scientific Leadership Information in Clinical Trial Registers. *PLoS ONE*, 3(2) 2008.
- SORIANO, J. et al. Enhancing User-Service Interaction Through a Global User-Centric Approach to SOA. In *Proceedings of the Fourth International Conference on Networking and Services*, 2008.
- TAYLOR, I.; DEELMAN, E.; GANNON, D.; SHIELD, M. (Eds.). *Workflows for e-Science*. Springer, 2006.
- IMI - The Innovative Medicines Initiative (IMI) Research Agenda, [http://imi.europa.eu/docs/imi-gb-006v2-15022008-research-agenda\\_en.pdf](http://imi.europa.eu/docs/imi-gb-006v2-15022008-research-agenda_en.pdf), accessed in 10/2008.
- TOLEDO, M. B. F.; NABUCO, O.; RODRIGUES, M.; GARCIA, D. Z.; CAPRETZ, M. A. M.; FANTINATO, M.; GIMENES, I.; BONACIN, R.; GUERRA, A.; ROCHA, T.; VIANA, L. A SOA-based Collaborative Environment for Clinical Trials on Neglected Diseases. In *Proceedings of 4th International IEEE Workshop on Service Oriented Architectures in Converging Networked Environments*, in conjunction with The IEEE 23rd International Conference on Advanced Information Networking and Applications (AINA-09) Bradford, UK, May 26-29, 2009.
- TOLEDO, M. B. F.; NABUCO, O.; RODRIGUES, M.; CAPRETZ, M.; FANTINATO, M.; GIMENES, I.; BONACIN, R.; GUERRA, A.; ROCHA, T.; JINO, M. *Gestão do Conhecimento e da Colaboração em Pesquisa Clínica*, In *Proceedings of II Seminário sobre Grandes Desafios da Computação no Brasil*, Manaus, Brasil, March 3-4, 2009.
- UDDI Universal Description, Discovery and Integration, Version 3.0.2, 2004.
- UMLS - Unified Medical Language System <http://kswebp1.nlm.nih.gov/uPortal/frame.jsp?myvt-frame=http://www.nlm.nih.gov/research/umls/documentation.html>, 2008.
- VEDAMUTHU, A. S. et al. *Web Services Policy 1.5 - Framework*. W3C, Sep. 2007. <http://www.w3.org/TR/2007/REC-ws-policy-20070904/>, accessed on 06/2009.
- WSDL - Web Services Description Language. Version 2.0, W3C Recommendation, June 2007. <http://www.w3.org/TR/wsdl20/>



## Sobre os autores

### *Miriam A. M. Capretz*

Dr Miriam A. M. Capretz is an Associate Professor and Associate Chair – Graduate in the Department of Electrical and Computer Engineering at the University of Western Ontario, Canada. Before joining the University of Western Ontario, she was an Assistant Professor in the Software Engineering Laboratory at the University of Aizu, Japan. Dr Miriam Capretz received her BSc and MEdSc degrees from UNICAMP, Brazil and her PhD from the University of Durham, UK. She has been involved with organization of several workshops and symposiums as well as has been serving as program committee member in several international conferences. She was a Program Co-Chair of the IEEE Workshop Web2Touch 2009 – living experience through web (W2T) in 2008 and 2009 and was the Program Chair of the IEEE Symposium on Human and Socio-Cultural Service Oriented Computing 2009. She has been working in the software engineering area for more than 25 years. Her current research interests include service oriented architecture, ontology and semantic integration, business process management, software security and grid computing.

### *Maria Beatriz Felgar de Toledo*

Maria Beatriz Felgar de Toledo is an associate professor in the Institute of Computing, University of Campinas (Unicamp), Brazil. She received her M. Sc. in Computer Science from Unicamp and her Ph. D. degree from Lancaster University, UK (1992). She has organized the three first editions of the Brazilian Workshop on Business Process Management since 2007. She is a committee member of the Brazilian Symposium on Information Systems, she organized the IEEE Web2Touch workshop in 2008 and 2009, and the IEEE Symposium on Human and Socio-Cultural Service Oriented Computing, held in Toronto, Canada, in September 2009. Her main research interests are advanced transaction models, service oriented computing, grid computing, business process management systems, knowledge base systems for clinical research and cultural heritage.