EXPRESSION OF INTEREST

Network of Excellence

Semantic Web

Submitted by EEIG-ERCIM,
European Research Consortium
for Informatics and Mathematics

Scientific contact

Julius Stuller
stuller@cs.cas.cz
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodarenskou vezi 2, Prague 8, Czech Republic
E-mail: stuller@cs.cas.cz
Phone: +420 2 6605 3200, Fax: +420 2 8658 5789

Administrative contact

Bruno Le Dantec
Bruno.le_dantec@ercim.org

This Expression of Interest was submitted in response to Call EOI.FP6.2002

7th June 2002

1. Rationale for proposing the research action

The aim of this proposal is to establish Europe as the International scientific leader in the next-generation Semantic Web. The primary goal of the Semantic Web activity so far has been the definition of infrastructure, standards and policies facilitating an explicit description of meaning of Web resources so they can be processed by both humans and automated tools. Complementary to the demand for universal access to information is the ever-increasing need for semantics-based access to services. These efforts towards the next evolution step of the Web have given rise to a large number of research problems that relate to models, architectures, applications and services for the Semantic Web. Research activity has been flourishing around these problems, with numerous impressive results, but still leaving us with an even greater number of challenges and opportunities. Classical data management practices are branching out of traditional frameworks, facing a previously unseen demand for openness, distribution, scalability, interoperability, expressiveness, representativity, flexibility, www-database integration and www-data mining.
The vision of the Semantic Web is a series of layers: URI & Unicode, data representation, knowledge representation, ontology, rules, logic, and culminating in trust and other applications managed by agents.
A concerted European approach, with a multidisciplinary synergy, is required to developing semantic web technology across these layers because otherwise the research base will fracture into independent pieces and will be carried out by isolated communities. These may address individual layers, or take vertical slices through the model for narrow applications that may not interoperate with others. Without a guiding structure, the semantic web developments will not interoperate, which is the primary reason for Semantic Web developments themselves. The strength of the consortium will be its multidisciplinary competences.

2. Objectives

The objectives of the Network are:

1.) To provide ***an integrated pan-European forum*** for research on the formal basis of semantic web technologies which can address the cultural diversity of Europe and the wider world wide web community and prevent both fragmentation and duplication of research in separate countries.
2.) To provide ***demonstrators of the capability*** of the Semantic Web which integrate ontologies, knowledge, data and information across Europe, across European languages and European cultures, and the presentation of this information to users. These can provide clear ***models for European industry*** on how to exploit Semantic Web technologies.
3.) To provide a ***main point of contact in Europe*** for the application development industry which can see that the Semantic Web offers them something, but need guidance on when to enter the marketplace with different technologies, and how their technologies fit into wider developments.
4.) To provide ***migration paths for user organisations across Europe*** from their current closed information systems to join the collective integrated Semantic Web as it develops and as it can provide them with business benefits.
5.) To provide ***advanced training*** for researchers across Europe in Semantic Web technologies so that they can transfer the technology with them into application developer industries and user organisations in commerce and industry, since the transfer by knowledgeable skilled staff has proven to be the best (if not the only) route for technology transfer in the knowledge technologies.

The impact of the World Wide Web on the US economy in 1998 was estimated by the US Federal Reserve as 2% of GDP. Its impact has grown since. The impact of the Semantic Web on the US and Europe is expected to be as great again. Firstly, *if there is not a unified European approach* to the semantic web, European industry will be beaten to the applications by US and Asian competition, *losing out on another IT boom*. Secondly, unless European users in industry and commerce can adopt and assimilate Semantic Web technologies at the right time for their commercial businesses, they will either invest too early and loose the investment, or invest too late and lose market advantage. The network intends to address these issues and should have a ***significant impact on European industry and commerce***.

3. General Approach foreseen to achieve the objectives

- The network will co-ordinate and direct the research of a wide range of European institutions through a steering committee representing the core members, and through the addition of user members to provide requirements for applications and to evaluate prototype applications developed by the network.

A) Steering committee:

Each participating organization in the Consortium will have one representative in the Steering committee. The Steering Committee will be chaired by the Project manager. Full conflict resolution and management procedures will apply to the Steering committee. The Steering committee will be responsible for the planning, and monitoring of the actions of the network.

B) Managing Organisation

The management will be undertaken by EEIG-ERCIM, the European Consortium for Informatics and Mathematics founded in 1992. Through this unique structure, sixteen different national research organisations with strong activity in IT research and development, in as many European countries, are tooled to participate in joint projects, keeping the management simple and effective. ERCIM has its central Office located in France and acts as a front-end to access the scientific expertise of its members (e.g. ERCIM has acted as an evaluation agency for the "InfoDev" World Bank programme). The Office has set up strong internal and external channels for solicitation and dissemination. It also has considerable experience in managing EU-funded projects (more than 30). ERCIM has its own International Fellowship Programme and Internal Mobility scheme. It also organises prospective workshops (like the EU-NSF ones), seminars and conferences. The ERCIM Office has also a very valuable competence in results dissemination, as part of its assets rest with customised web design, set-up and assistance, and the edition of the ERCIM News magazine (over 8000 copies distributed worldwide). This, combined with the sixteen research organisations disseminated across Europe makes ERCIM a key player in European IT research and development, and a reliable foothold for international cooperation.

C) The types of partners involved:

1. The *Core Partners* in the project will be the ERCIM research institutes across Europe. In addition, the OntoWeb thematic network, in which most of the ERCIM institute are involved and that successfully took on its shoulders the structuring of European semantic web community, has been approached and should most probably join this initiative. These partners include all the significant European universities researching in the semantic web, the major research institutes, as well as both small and large IT companies across Europe who can develop and exploit applications.
2. *Assistant Partners* will support the Core Group:
   User organisations in various commercial, government and industry application areas with whom the researchers have previous contact will be called upon to join the network at times during its lifetime in order to ensure the applicability in varied application domains.
3. As analysed elsewhere, the NoE will seek to actively involve additional, recognized experts, R&D groups, and industry into its activities. The Steering Committee will be responsible for the management of the process for the liaison with external National and International relevant bodies and organizations.

- The activities of the network to provide a pan-European virtual laboratory will include**:**

A) Integration Activities

Selection of integration activities: joint training courses, annual summer/winter schools, thematic workshops, electronic communication networks, mobility of personnel, joint studies

B) JPA - Joint Programme of Activities

The following problems are the cornerstone issues for the realization of the Semantic Web:

- Infrastructure for Knowledge Data and Process Mediation (to facilitate automated description, discovery, brokering and composition of e-services)
- Metadata Management (to enable management and maintenance of superimposed resource descriptions and schemas of heterogeneous data sources)
- Ontology Evolution and Metadata Revision
- Ontology Interoperability (to support multiple, heterogeneous ontologies, represented in different languages)
- Information Extraction (to enable (semi)-automation of ontology creation and metadata generation)
- Transactional Aspects for e-services (revision of ACID properties of transactions)
- Formal Foundations for Web Metadata Standards (efficiently implementable formal models - expressiveness vs. efficiency tradeoffs)
- Semantics-aware Query Languages versus Inference Services (to enable sophisticated searching and browsing, declarative content-based access, resource discovery and matchmaking)
- Persistent Storage (to enable the efficient management of voluminous resource description bases and support effective sophisticated querying - schema-specific storage, index structures)
- Extension of database technologies to semantic web

- Presentation of extracted information to the user, tailored to the user's environment, making use of the appropriate media.

1. The network includes several university and research institute partners who will work together on the ***formal basis of the semantic web*** - there are many outstanding research issues about the role of reification, reasoning over the web, managing ontologies, inference engines for the web, tackle heterogeneity and dynamics of the web, etc… that require formal solutions before the semantic web can progress.
2. ***International, Cultural and Semiotic issues*** underlying a semantic web are crucial to get beyond the limitations of IT and logic, and to provide a global system that can be assimilated into all cultures. Centuries of development of communication have lead to sophisticated tools (pragmatics, rhetoric, aesthetics, typography, etc.). Their development, unrelated to information technology, renders their apprehension by computer means very difficult. Research in this area will include extending the traditional linguistic based semantics of the web to address how information, knowledge and data are used.
3. The network partners will together provide a ***European Semantic Web Open Source Software Factory***. The network will encourage open source development of high quality components and act as a non-profit shelter organisation for software development (like Apache). For the web CERN and NSCA have played this role at the beginning. This model could be applied to ontologies as well. These developments could cover API, data structures, transformation environments for semantic web languages, inference engines, annotation software, finally extended to general purpose and demonstrative applications and resources (ontologies, vocabularies).
4. ***Ontology*** is a key factor for enabling interoperability in the semantic web. Ontology is an explicit specification of a conceptualisation. It includes an explicit description of the assumptions regarding both the domain structure and the terms used to describe the domain. Ontologies are central to the semantic web because they allow applications to agree on the terms that they use when communicating. Within a multi-agent system, agents represent their "view of the world" by explicitly defined ontologies. The interoperability of such a multi-agent system is achieved through the reconciliation of these views by a commitment to common ontologies that permit agents to interoperate and cooperate. In that sense, ontology in itself is a step forward towards interoperability in the Internet. In a real open environment, ontologies are developed and maintained independently of each other in a distributed environment. Thus two agent systems may use different ontologies to represent their view of the domain. This is often referred to as ontology mismatch. In such a situation, interoperability between agents is based on the reconciliation of their heterogeneous views. How to tackle with ontology mismatch is still a question under intensive research. First one has to understand the problem of ontology mismatch (semantic heterogeneity) itself. Different classifications of ontology mismatch have been proposed in the literature. The classification is important because it points out what types of mismatches is easy to solve and what are the hard ones. For the easy ones, technology can be developed for mapping and translating among ontologies. For the hard ones, the challenge is to find ways to make simple assumptions, which enable agents to do useful things in practical situations. For dealing with semantic heterogeneity among distributed, autonomous information sources, solutions in the multi database and information systems area have existed for some time. Many of them are based on a database-style modelling of data, global schema, and use of meta-information such as provided by a common ontology or different domain ontologies for a content-based source selection. Others focus on information retrieval techniques for best-match queries, and relevance assessment. Some of the experiences in these areas can be reused.
5. ***Web publishing systems*** have to take into account a plethora of Web-enabled devices, user preferences and abilities. Technologies generating these presentations will need to be explicitly aware of the context in which the information is being presented. Semantic Web technology can be a fundamental part of the solution to this problem by explicitly modelling the knowledge needed to adapt presentations to a specific delivery context.
6. ***Agent technology*** is an additional technology that will enable taking into account distribution of information sources in the Semantic Web through cooperation among software agents (for Web mining, resource discovery, cooperative reasoning, information retrieval, push towards users or applications, etc.)
7. In order to enable scalable creation of ontologies and metadata from huge Web corpora, techniques enabling ***semi-automatic creation of ontologies and metadata*** will be useful. Such techniques can be based on Natural Language Processing Tools or on Machine Learning. We will integrate results from these research areas.
8. ***Corporate Semantic Webs*** or ***Community Semantic Webs*** are a specific, interesting case, as one can rely on some simplifying hypotheses about the context and use of these Semantic Webs local in a company/community or distributed between several companies/communities.

9. Activities to provide ***guidance to user organizations*** including ***assessments of socio-economic impact of Semantic Web***; Provision of ***guidance to SME's on costs and benefits*** of Semantic Web; Provision of ***guidance to SME's on migrating to the Semantic Web***.

C) Training of researchers

Many of the core organisations are Universities, which grant research degrees, or research centres welcoming PhD students. Following the guidance of the network plan, many of these PhD and Masters degrees students will be directed within the research framework for the Semantic web proposed by the Network.

D) Networking with other organisations

The network contains W3C as a member for standardisation, but will also be open to a links with a wide range of other European and international organisations - see section 6b.

E) Dissemination activities

The network will establish a ***web site***, and ***monthly e-mail newsletter*** that will be open to those outside the network as a means of dissemination. The Research members of the network will use the usual channels of publication of ***academic papers*** and ***conference presentations*** at W3C, Semantic Web and other conference series. As the network progresses, it is planned to present more frequently at conferences and ***workshops*** in user domains to illustrate how the semantic web can be applied to provide ***business benefits*** to those organisations. Workshops will also be organised (independent workshops or workshops integrated in relevant conferences dedicated to research carried communities).

It is planned to establish a ***revenue stream*** ensuring the running of the network that will ***support it after the termination of CEC funding***. The exact method is not yet clear, this could involve users of the software factory providing shareware or subscription based income, the formation of a consortium to which members pay fees, or the incorporation of the network and its activities into an already existing consortium (e.g. W3C Europe). The network will have to select from these options when realistic business plans for them have been developed.

4. Timescale, Effort and Costs

The funding arrangements for FW6 still appear unclear. If the network lasts 5 years under CEC funding before becoming self financing and includes ***10 main partners***, ***25 active participant organisations*** and ***10 temporary organisations***, then we should expect about ***100 person years*** of effort per year yielding about ***€ 2 million per year*** or a total of about ***€ 10 million***.

5. Needs and Relevance

The Semantic Web in conjunction with Web Services will offer new forms of doing business and accessing services, which are required to meet the high expectations of citizens for a ***better quality of life*** as they now appreciate the wide range of possibilities, that IST applications and products can offer. The semantic web will be a major advance in global knowledge and information technologies in which Europe must contribute significantly in order to maintain employment, growth, industrial competitiveness and the living standards of its citizens. As use for the Internet grows, greater attention must be given to social changes, to social inclusion and trust, security and privacy that are addressed by the Semantic Web technologies.

If the EU is to achieve its objective *to become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion* set in March 2000 by the European Council, then it must be at the forefront of research, development and application of Semantic Web technologies since they will drive the dynamic knowledge-based economy. The network proposed here contains the main active research and development teams across Europe with experience in Semantic Web technologies having strong links with a wide range of application providers who can be brought into the network to provide exploitation of the IPR to maintain Europe's position as a leader in the knowledge based economy.

To develop a European Research Area, Semantic Web technologies are not only essential as a significant research theme, but they also provide an enabling technology for the communication of Research activities and results around Europe. The network proposed includes the major national ITC laboratories from each European country that will be exposed to

the earliest experience of these technologies in order to speed up their integration, and therefore that of the European Research Area.

In April 2002, only 40% of pages served on the web were in English leaving 60% in other languages. A similar percentage of the web's users and contributors are not native English speakers, nor from an English culture. In order to extend the web to be based on meaning it will be necessary to incorporate the meanings of many nations and cultures. This cannot be done in a single country, and requires a pan-European, if not a global research exercise. Many of the participants in the proposed networks indeed have close research links outside Europe to link into research from other cultures.

6. Excellence

- Under the FP5, IST programme, there are 7 Semantic Web funded projects, although several other related networks and projects which address this theme are financed by other national or transnational sources. Start up companies have been formed to transfer to the market the results issued by these projects, while several of the larger European IT developers have started to investigate Semantic Web tools and applications. This mass of researchers and developers is sufficiently significant and of sufficient excellence to indicate the need for a European network as this expression defines.
- **Multi-disciplinary skills and critical mass** must to be put together to establish Europe as the world leader in Semantic Web.
  a.) Included in the network are the ERCIM partners who have considerable research experience as well as providing pan-European links with developer and user organisations in their countries.
  b.) The OntoWeb thematic network members have been contacted similarly as several other universities and developer organisations, although they have not yet reached a final agreement. Potentially interested people, highly involved in the following research fields - Annotation tools, RDF Transformation,Ontologies, DAML + OIL, Natural language tools - already beeen approached, may join the proposal : Michel Crampes (Ecole des Mines d'Ales), Dave Beckett (University of Bristol), Nigel Shadbolt (University of Southampton), Rudi Studer (University of Karlsruhe), Asunción Gomez-Perez (Universidad Politecnica de Madrid), Dieter Fensel, Frank van Harmelen (VUAmsterdam), Enrico Motta (Open University), Dave Reynolds (Hewlett Packard)….
  c.) Third countries/International organisations
  W3C has agreed to be part of the network, as the international organisation expected to standardise the Semantic Web. Contact has also been made with representatives on ISO committees who could standardise developments. OASIS as the third relevant standards body has not been contacted yet.
  FIPA for agent standards for the semantic Web is another candidate which will be invited to joint the network.

Proposed Consortium (Organisation, Background and Role)

CORE GROUP

	Organisation	Country	Chief Scientist	Area of Excellence	Role in Project
1	European Research Consortium for Informatics and Mathematics (**ERCIM**)	**France**	JE Pin B Le Dantec		Management, Research & Development
2	Central Laboratory of the Research Councils, Rutherford Appleton Laboratory (**CLRC RAL**)	**UK**	B.M.Matthews M.D.Wilson	Metadata, Thesauri, Ontologies	Management, Research & Development
3	Czech Research Consortium for Informatics and Mathematics (**CRCIM**)	**Czech Republic**	Julius Stuller	Agent-based approaches, Conceptual modelling	Research & Development
4	Centrum voor Wiskunde en Informatica (**CWI**)	**Netherlands**	Lynda Hardman	Multimedia in the Semantic Web	Research & Development
5	Foundation for Research and Technology-Hellas, Institute of Computer Science (**FORTH**)	**Greece**	Dimitris Plexousakis	RDF Parser, RDF Query Language	Research & Development
6	Institute National de Recherche en Informatique et en Automatique (**INRIA**)	**France**	Jérôme Euzenat, Rose Dieng	Ontologies, Corporate semantic Webs	Research & Development
7	Norwegian University of Science and Technology (**NTNU**)	**Norway**	Arne Solvberg, Jon Atle Gulla	Conceptual modelling, Text analysis tools	Research & Development
8	Swedish Institute of Computer Science (**SICS**)	**Sweden**	Olle Olsson, Sverker Jansson, Magnus Boman	Personal information monitoring, Trust management	Research & Development
9	Slovak Research Consortium for Informatics and Mathematics (**SRCIM**)	**Slovakia**	Peter Vojtas	Knowledge based systems, Flexible querying	Research & Development
10	Technical Research Centre of Finland (**VTT**)	**Finland**	Raimo Launonen	Multilingual text and Multimedia documents	Research & Development
11	W3C	**France**	D Dardailler D Brivhley		Research & Development