On-To-Knowledge Dissemination and Use Plan

Hans Akkermans (VUA, editor)

with contributions from

John Hughes, John Davies (BT)
Robert Engels (CognIT)
Frank van Harmelen, Jos van der Meer (AIdministrator)
Ulrich Reimer (Swiss Life)
Hans-Peter Schnurr, York Sure, Rudi Studer (AIFB)
Fredrik Ygge (EnerSearch)

Executive Summary.
On-To-Knowledge EU-IST-1999-10132 Project Deliverable D34
(WP9, due month 6), 30 June 2000.

Document Id.	VUA/OTK/2000/D34/v1.0
Project	On-To-Knowledge
Date	30 June 2000
Distribution	consortium, EC

On-To-Knowledge Consortium

This document is part of a research project partially funded by the IST Programme of the Commission of the European Communities as project number IST-1999-10132. The partners in this project are: Vrije Universiteit Amsterdam VUA (coordinator, NL), University of Karlsruhe (Germany), Schweizerische Lebensversicherungs- und Rentenanstalt/Swiss Life (Switzerland), British Telecommunications plc (UK), CognIT a.s. (Norway), EnerSearch AB (Sweden), AIdministrator Nederland BV (NL).

Vrije Universiteit Amsterdam (VUA)
Division of Mathematics and Informatics W&I
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands
Tel: +31 20 4447718, Fax: +31 20 872 27 22
Contactperson: Dieter Fensel
E-mail: dieter@cs.vu.nl

Schweizerische Lebensversicherungs- und Rentenanstalt / Swiss Life
Swiss Life Information Systems Research Group
General Guisan-Quai 40
8022 Zürich
Switzerland
Tel: +41 1 284 4061, Fax: +41 1 284 6913
Contactperson: Ulrich Reimer
E-mail: Ulrich.Reimer@swisslife.ch

CognIT a.s
Busterudgt 1.
N-1754 Halden
Norway
Tel: +47 69 1770 44, Fax: +47 669 006 12
Contactperson: Bernt. A.Bremdal
E-mail: bernt@cognit.no

AIdministrator Nederland BV
Julianaplein 14B
3817 CS Amersfoort
The Netherlands
Tel: +31 33 4659987, Fax: +31 33 4659987
Contactperson: Jos van der Meer
E-mail: Jos.van.der.Meer@aidministrator.nl

University of Karlsruhe
Institute AIFB
Kaiserstr. 12
D-76128 Karlsruhe
Germany
Tel: +49 721 608392, Fax: +49 721 693717
Contactperson: R. Studer
E-mail: studer@aifb.uni-karlsruhe.de

British Telecommunications plc
BT Adastral Park
Martlesham Heath
IP5 3RE Ipswich
United Kingdom
Tel: +44 1473 605536
Fax: +44 1473 642459
Contactperson: John Davies
E-mail: John.nj.Davies@bt.com

EnerSearch AB
Carl Gustafsväg 1
SE 205 09 Malmö
Sweden
Tel: +46 40 25 58 25; Fax: +46 40 611 51 84
Contactperson: Hans Ottosson, Fredrik Ygge
E-mail: hans.ottosson,fredrik.ygge@enersearch.se

Contents

Chapter 1

On-To-Knowledge dissemination approach

1.1 On-To-Knowledge Vision and Advances

Semantic Information Resources: the Knowledge Web. In the Information Society, a crucial need is support for information and knowledge exchange at a large scale. Currently, knowledge workers suffer from information overload, and they waste a lot of time in searching, navigating, and browsing information resources. The On-To-Knowledge vision is that this problem can be significantly alleviated by automated forms of semantic information processing, that is, by electronic methods that are capable to filter and handle *meaning* in large repositories of documents. This would enable a much more selective and meaningful user access to electronic documents. Ultimately, the aim is that all electronic documents in intranets and the World-Wide Web are standardly processed in terms of their content and meaning to information users *by default*, so that we can rightfully speak of a Semantic or Knowledge Web.

On-To-Knowledge Contributions. On-To-Knowledge will make a major step into this direction by developing a set of tools that perform various aspects of semantic information processing: (1) automatic concept extraction and annotation of documents; (2) ways to represent, share and analyze the semantics of document domains; (3) meaning-based querying of document contents. The key technology to achieve this are *ontologies*. The architecture of the On-To-Knowledge toolset is depicted in Figure 1.1.

In a nutshell, the key advance by On-To-Knowledge is that it enables to *get meaning-based answers* to user questions that are put to information resources. This is a significant improvement over the keyword-based search and information retrieval techniques that are the state of the art today.

The above is abstracted from our eBeW-2000 European conference contribution, which is added to this deliverable as an Appendix, and in itself constitutes an important and early dissemination action by the On-To-Knowledge consortium.

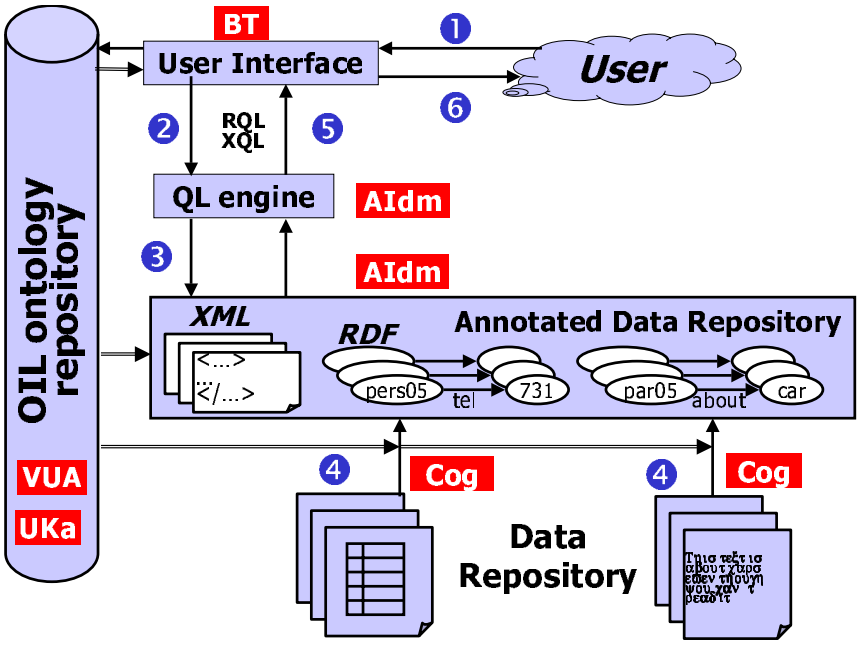

Figure 1.1: On-To-Knowledge toolset and use architecture.

1.2 Dissemination strategy

For the On-To-Knowledge consortium as a whole, there are four main target groups for dissemination and use:

(1) The international industrial and academic RTD community at large.

(2) Formal and informal networks of organizations and projects known to have related interests and activities.

(3) Industrial companies and clients that will directly benefit from the use of the On-To-Knowledge tools.

(4) Standardization bodies that can further the realization of the Semantic or Knowledge World-Wide Web.

In view of this, the following consortium-wide dissemination activities will be undertaken:

1. Dissemination to international RTD groups in industry and academia through publications and presentations. The eBeW2000 conference presentation is a clear example of this, see the Appendix.

2. Links and dissemination to related EU-IST projects will be maintained (e.g. IBROW, MK-BEEM, ENRICH).

3. As a further step, an Ontology thematic network of excellence is being prepared (OntoWeb). This includes both academic research institutions and industrial companies and businesses. See *http://www.cs.vu.nl/~ying/OntoWeb/*, IST-2000-25056.

4. It also involves developing links to related work in the US (e.g. DARPA programme DAML, Stanford, USC, IBM).

5. The On-To-Knowledge consortium intends to contribute to the ongoing development of WWW standards by participating in the relevant workgroups of the W3C consortium, especially those related to the Semantic Web and RDF. This activity is particularly based on OIL, the ontology-based inference layer for the web and its foundation in XML/RDF, as developed and tested by On-To-Knowledge.

6. Dissemination to the general public is done first of all through the On-To-Knowledge website. Also, a general project booklet has been produced and is now distributed by the partners at various national and international events, meetings and conferences. Furthermore, a general On-To-Knowledge electronic slide presentation is under preparation.
   See the deliverables D30 and D33, and *http://www.ontoknowledge.org/*.

7. As part of the workplan, two On-To-Knowledge workshops especially focused on dissemination to an industrial audience are planned (see deliverables D31 and D32).

In addition to these consortium-wide dissemination actions, each individual partner is concerned with dissemination to and use by specific target groups. The company-specific dissemination and use plans are presented in the next Chapter.

Chapter 2

Individual Partner Contributions to Dissemination and Use

2.1 Free University Amsterdam VUA

Partner role and intended results. VUA is a university leading in computer science and business informatics research. Apart from being responsible for the overall project coordination, its main role in On-To-Knowledge is to develop the ontology-based inference layer for the web called OIL.

Target groups for dissemination and use. As a university institution, the main target groups for our dissemination and use activities are: (1) the relevant international research communities in academia and industry; (2) the students (Master as well as PhD degree) we teach; (3) companies with which we have collaborative relationships in applied research and/or business education. As in all our projects, this is achieved by international publications, presentations and talks; courses and degree work in our regular curricula; and by bilateral and multilateral project collaborations and business courses for external organizations.

Dissemination actions up to Milestone 1 (month 8, August 2000). Current dissemination activities include:

- Accepted international publications (several are jointly with other On-To-Knowledge partners):
  - On-To-Knowledge: Ontology-based Tools for Knowledge Management (European Conference on Electronic Business and Electronic Work eBeW-2000, Madrid, Spain, October 2000; see Appendix).
  - OIL in a Nutshell (European Knowledge Acquisition Conference, EKAW-2000, Juan-les-Pins, France, October 2000).
  - OIL and UPML: a Unifying Framework for the Knowledge Web (Workshop on applications of ontologies and problem solving methods at the European Conference on AI,

Berlin, August 2000).

  - Book by Dieter Fensel to appear at Springer-Verlag, Berlin, D, 2000, entitled: Ontologies — Silver Bullet for Knowledge Management and Electronic Commerce.

- Presentations and talks:

  - Seminar presentation for Vuture.net - Amsterdam Centre for Electronic Business Research (Amsterdam, NL, 20 April 2000, Dieter Fensel and Frank van Harmelen).
  - Presentation at the Enersearch board of directors (shareholding companies) (Malmö, Sweden, 10 May 2000, Hans Akkermans), see further the EnerSearch section below.
  - International Dagstuhl seminar Semantics on the Web, Germany, March 2000 (co-organized by Dieter Fensel, presentation by Frank van Harmelen and Ian Horrocks).
  - International Dagstuhl seminar Interdisciplinary Approaches to Knowledge Management, Germany, July 2000 (co-organized by Rudi Studer (AIFB), presentation by Hans Akkermans).
  - Tutorial at the European Conference on Artificial Intelligence ECAI-2000, Berlin, D, August 2000: T09 — Seeking Information: Methods from Information Retrieval and Artificial Intelligence (tutorial by Dieter Fensel and Mounia Lalmas).

- VUA is leading the mentioned initiative on a European thematic network on Ontologies called OntoWeb (Dieter Fensel, Ying Ding, et al., see *http://www.cs.vu.nl/~ying/OntoWeb/*, IST-2000-25056).

- VUA has been the responsible partner for the On-To-Knowledge website and the general project brochure (deliverables D30 and D33, *http://www.ontoknowledge.org/*).

Planned future dissemination activities. Future dissemination activities of VUA include:

- Further scientific publications and presentations. An example is “OIL: a case-study in extending RDF-Schema”, submitted to the Workshop on The Semantic Web in Lisbon, Portugal.

- VUA regular curriculum. A course in Knowledge Management and Modelling has been introduced as a regular subject in the VUA Business Informatics (2nd year) and Computer Science curricula (Hans Akkermans, Dieter Fensel). This course will be given in Spring 2001 for the first time, and attention is paid to On-To-Knowledge as a special topic.

- Research education. Two PhD students at VUA will do their degree work partly within On-To-Knowledge (Michel Klein, Borys Omelyanowski). Also Master students will have the opportunity to do their degree work in On-To-Knowledge.

- International and Business education. VUA staff is also teaching at international and business courses in Knowledge Engineering and Management, usually based on our international research in the field. Examples are:

  - Regularly, specially tailored company courses (“master classes”) in this field are given (Hans Akkermans, Frank van Harmelen).

  - Master course on Knowledge Management in Ronneby, Sweden (Fall 2000, Hans Akkermans).
  - Invited contributions to master education in knowledge engineering at UPM Madrid in October 2000 (Hans Akkermans).

  Where appropriate, such courses are a good channel to disseminate relevant results of On-To-Knowledge.

- Exchanges with other EU-IST projects such as IBROW and MKBEEM, are already ongoing. VUA is also planning for new European projects where On-To-Knowledge results will be employed, especially in relation to the Information Retrieval community and approaches (Ying Ding, Dieter Fensel), and advanced electronic markets in e-Business (Hans Akkermans). Here, our interdisciplinary collaboration between IT, economics/business administration and law within Vuture.net, the Amsterdam Centre for Electronic Business Research of VUA, is a relevant channel for dissemination, as this Centre has a significant and growing body of external industry and business relationships.

- VUA is a member of the W3C Consortium working groups on the Semantic Web.

2.2 British Telecom

Partner role and nature of intended results. BT's Knowledge Management Research Group is a recognised leader in providing solutions to support collaborative working and information retrieval. For On-To-Knowledge BT is providing the Search Facility (WP 4.1), the Knowledge Sharing Facility (WP 4.2) and the Call Centre Case Study (WP 7).

Target groups for dissemination and use. The Knowledge Management Research Group at BT has a long history of publications and presentations at important conferences and we intend to "carry on business as usual" to the benefit of On-To-Knowledge. Where relevant, we will use these platforms to disseminate information on On-To-Knowledge.

We often give demonstrations to external visitors to BT Adastral Park at our customer centre facilities. We will use this as an opportunity to disseminate information on On-To-Knowledge's activities, where appopriate.

As BT is such a large organisation, *internal dissemination* is also important. We will present details of the project within our internal publications and maintain links with teams with a relevant interest.

Dissemination actions up to Milestone 1. As BT's On-To-Knowledge activities have not had milestones prior to month 8 we have not had much to publish so far, but this will commence once we have got closer to meeting our milestones. We expect to be presenting papers at conferences such as ECAI, SIGIR, KME, etcetera.

John Hughes attended the Schemas Workshop in Bath, UK, using the opportuntity to spread the word on On-To-Knowledge and to distribute booklets (a brief report is available from BT).

A number of recent research papers are clear indicators as to how BT's dissemination of On-ToKnowledge results will take place in the form of scientific and professional publications:

- Supporting Virtual Communities of Practice , John Davies, to appear in "Industrial Knowledge Management", Springer-Verlag, 2000.
- Concept Lattices for Knowledge Management, Uwe Krohn, John Davies & Richard Weeks, BT Technology Journal, October 1999.
- Building virtual communities, John Davies, Knowledge Management Vol. 3(7), April 2000.

Planned future dissemination activities. Publications and presentations at conferences is a regular occurence for us and we will make use of this platform to publicise On-To-Knowledge's activities. In the near future John Davies will be presenting at the following conferences:

- International Knowledge Management Summit, Toronto, October 2000
- Knowledge Management Europe 2000, Brussels, November 2000
- The Knowledge Intranet: Cornerstone of successful e-business, London, October 2000, Unicom Seminars.

Furthermore, we expect to publish at least one article in the BT Technology Journal, and at least two visiting research fellows to contribute to the work on On-To-Knowledge.

2.3 Swiss Life

Partner role and nature of intended results. Swiss Life is an industrial partner who is mainly interested in applying On-To-Knowledge results. To this end, case studies are done. A specific research interest of Swiss Life is the integration of different knowledge sources in an Organizational Memory. This will be pursued within Swiss Life's IT Research & Development Group.

Target groups for dissemination and use. In addition to the external research communities, the main target groups for dissemination and use are all organizational units of Swiss Life, including its subsidiaries and its branches in other European countries. Furthermore, Swiss Life is supporting student education in various ways.

Dissemination actions up to Milestone 1. A one-day lecture on Organizational Memories was held as part of the TIAS (Business School) course on Knowledge Management of the University of Tilburg, The Netherlands.

In cooperation with the University of Karlsruhe (Prof. Rudi Studer) a Diploma Thesis on"Learning of Ontologies" is carried out at Swiss Life. This is done together with the IST Project "Mining Mart" (IST-1999-11993).

A presentation of the On-To-Knowledge project and the case studies was given to the corporate board of Swiss Life.

Planned future dissemination activities. A lecture on Organizational Memories will be given at the University of Konstanz (Germany) in the Winter terms of 2000/2001 and 2001/2002.

A tutorial on Organizational Memories will be given at the Int. Conf. on Enterprise Information Systems, 2001 (ICEIS 2001).

Presentations of case study results will be given to to various organizational units of Swiss Life, including its subsidiaries in Switzerland and its branches in other European countries.

Of course, case study results will be presented at various conferences and Workshops on Knowledge Management, as well as in journals.

2.4 EnerSearch

Partner role and nature of intended results. EnerSearch is a virtual, distributed organization with corresponding high needs for efficient knowledge management. The role of EnerSearch in the project is to evaluate the usefulness of the developed tools in a virtual organization setting.

Target groups for dissemination and use. The results from these evaluations are of key interest to the owners of EnerSearch and a dissemination within these companies is hence very natural. The owners of EnerSearch that will be engaged in the evaluation of the tools are:

- ABB Automation Products (SE)
- ECN (NL)
- Electricidade do Portugal (PO)
- Iberdrola (ES)
- IBM Utilities and Services (US/International)
- PreussenElektra (D)
- Sydkraft (SE)

Dissemination actions up to Milestone 1. The project has (in co-operation with VUA) been presented and discussed at the EnerSearch general board meeting in May 2000. Direct dissemination and use results of this meeting were:

- IBM offered to bring the On-To-Knowledge project into contact with their groups doing related RTD work in the US.
- ECN expressed interest in a semantic analysis of their website with the On-To-Knowledge tools.
- There will be a meeting with Iberdrola (one of the biggest utilities in Europe and the largest one in Spain) in Madrid later this year to discuss a joint On-To-Knowledge evaluation action.

The project has also been presented at different workshops for various energy utilities and authorities in Sweden.

Planned future dissemination activities. The evaluation case study is part of a PhD student program supported by EnerSearch and will be documented accordingly, in research papers to be presented at workshops and conferences.

The project results will be further disseminated to the EnerSearch owners in the board of directors meetings, and also to interested utility companies in workshops and seminars that are regularly organized by EnerSearch.

The objective is that the tools are considered successful during the evaluation phase and so will become an integrated part of the EnerSearch information system, used on a daily basis by the major European companies mentioned above.

2.5 Institute AIFB, University of Karlsruhe

Partner role and nature of intended results. The Institute AIFB from University of Karlsruhe is a university research institution leading in the area of Knowledge Management and Ontologies. Main results in On-To-Knowledge are a methodology for KM practitioners and a tool for the modelling of ontologies (OntoEdit).

Target groups for dissemination and use. The dissemination strategy aims mainly at research publications, international events, international presentations, and education (courses on Knowledge Management).

Dissemination actions up to Milestone 1.

- "5. AIK Symposium — Wissensmanagement", Karlsruhe, Germany, on May 5th, 2000; organised by Institute AIFB in cooperation with AIK e.V. (Verein für Angewandte Informatik Karlsruhe e.V.)
  - Presentation "Wissensvernetzung - eine neue Generation von Wissensmanagement-Lösungen", Hans-Peter Schnurr (AIFB)
  - Presentation "Intelligente Techniken für das Wissensmanagement", Dr. Steffen Staab (AIFB)
- Staab, S. and Maedche, A., "Ontology Engineering beyond the Modeling of Concepts and Relations", to appear: Proceedings of the ECAI'2000 Workshop on Applications of Ontologies and Problem-Solving Methods, edited by R. Benjamins, A. Gomez-Perez, N. Guarino, M. Uschold. Berlin, Germany, 2000
- Sure, Y., Maedche, A., Staab, S., "Leveraging Corporate Skill Knowledge — From ProPer to OntoProPer", submitted to: PAKM2000 — Third International Conference on Practical Aspects of Knowledge Management, October 30th — 31st, 2000, Basel, Switzerland
- "Dagstuhl Seminar Knowledge Management", Schloss Dagstuhl, International Conference and Research Center for Computer Science, Dagstuhl, Germany, July 10th - 14th; partly organised by Institute AIFB; further details:
  http://www.dagstuhl.de/DATA/Title/00281.html

  - Presentation “Using Metadata and Ontologies for Knowledge Management”, Dr. Steffen Staab (AIFB)

Planned future dissemination activities.

- “Ontology Learning ECAI-2000 Workshop”, Berlin, Humboldt University, Germany, on August 22nd, 2000, part of the “14th European Conference on Artificial Intelligence ECAI-2000”, Berlin, Humboldt University, Germany, on August 20th — 25th, 2000; the workshop is partly organised by Institute AIFB; further details:
  http://ol2000.aifb.uni-karlsruhe.de/.
- “WM2001: First Conference on Knowledge Management” held in Baden-Baden, Germany, on March 14th — 16th, 2001; organized by Institute AIFB in cooperation with AIK e.V. (Verein für Angewandte Informatik Karlsruhe e.V.); further details:
  http://wm2001.aifb.uni-karlsruhe.de/.
- Institute AIFB will be a partner in the OntoWeb Network (IST-2000-25056); further details:
  http://www.cs.vu.nl/~ying/OntoWeb/.

2.6 CognIT

Partner role and nature of intended results. CognIT’s role in the On-To-Knowledge project is to lead the work on (semi-)automated information extraction from structured and unstructured texts. CognIT’s main results in the On-To-Knowledge project are a tool for concept extraction from unstructured texts and a wrapper tool for extraction of information from structured texts. CognIT will also support the tool-related exploitation of the project results.

Target groups for dissemination and use. The dissemination strategy of CognIT consists of reports, contributions to papers, presence at conferences and exhibitions, and targeting of potential users of the project results, as new client prospects for the CognIT tools.

Dissemination actions up to Milestone 1. Current dissemination activities are:

- Publications: AISC2000, plus white papers.
- Presentations: MLnet-II exhibition (at ECML-2000), CeBIT 2000 Hannover, ECAI-2000 workshop presentation in Berlin, Dagstuhl seminar on Knowledge Management (Germany, July 2000).

Planned future dissemination activities. Future dissemination activities will be:

- Publications: planned are two more white papers and possibly one or more conference contributions.

- Presentations/participation: CeBIT-2001 exhibition (presentation?), Expo-2000 (poster?), CoLing 2000 (Luxembourg), several presentations in EU and USA.
- Education: courses on Knowledge Management at Polytechnic in Narvik, Norway.
- Possibly, dissemination of projects results through a coop-partner in Germany.
- Participation in preparations of an Ontology Network of excellence (OntoWeb).

Finally, CognIT is running and maintaining a machine with an On-To-Knowledge installation of CognIT's CORPORUM/BIP technology for use and reference by the On-To-Knowledge partners and their dissemination activities.

2.7 AIdministrator

Partner role and nature of intended results. AIdministrator is an industrial partner in the On-To-Knowledge consortium in the role of technology provider. We are providing the following technologies to the consortium:

- robust parsing technology;
- tools for interactive ontology definition;
- a fast classification engine for such ontologies;
- visualisation technology for the resulting populated ontology.

The intended results for AIdministrator are:

1. improved tools for ontology acquisition (collaboration with CognIT)
2. improved tools for large-scale semantic data storage and querying
3. improved visualisation techniques (collaboration with BT)
4. show-case applications of this technology (collaboration with Swiss Life)
5. prototype tools for our next generation of semantic site-maps, navigation tools and query-visualisers.

Target groups for dissemination and use. The target groups for our dissemination activities are the market segments described in Figure 2.1.

Dissemination actions up to Milestone 1. Our dissemination strategy consist of wide-audience events as well as targeting specific organisations. Dissemination activities up to milestone 1 (month 8, August 2000) have been:

- Contacts with specific organisations concerning potential application of relevant technology:
  - KPN Research
  - portal provider PZ (www.pz.nl)
- Wide-audience activities to publicise On-To-Knowledge activities:
  - presentation and workshop for students and staff of VU Amsterdam
  - presentation for staff at University of Bremen
  - demonstration at international seminar on The Semantic Web
  - presentation at VanWoodman seminar on Improving Knowledge Productivity (NL, 15 June 2000)
- Presentation to the industrial board of the IBROW EU-IST project No. 1999-19005.
- Publications (jointly with other On-To-Knowledge partners):
  - OIL in a nutshell (European Knowledge Acquisition Workshop, Juan-les-Pins, France, October 2000)
  - On-To-Knowledge: Ontology-based Tools for Knowledge Management (Conference on Electronic Business and Electronic Work 2000)
  - OIL and UPML: a Unifying Framework for the Knowledge Web (Workshop on applications of ontologies and problem solving methods at the European Conference on AI, Berlin, 2000)
  - Knowledge Representation on the Web (IEEE Internet Computing)
  - OIL: a case-study in extending RDF-Schema (submitted to workshop on The Semantic Web, Lisbon)
  - participation in RDF-Schema mailing list at W3C
  - Enrollment as member of OntoWeb consortium

Planned future dissemination activities. Intended activities in the following months are:

- further presentations for both academic and industrial audiences (e.g. presentation for Dutch Graduate School of Information and Knowledge Systems SIKS);
- further collaboration with case-study partners external to the consortium.

Aidministrator			
Service	Timing	Market Segment	Example Customers
Transaction based e-commerce services: • construction of specialised site-maps • visualisations, search-engines • consistency-checks • Website-audits consistency-checks, Website-audits) specialised site-maps, visualisations, search-engines, consistency-checks, Website-audits)	Short term	1. Business with a Web-presence 2. Information Service providers	Large: Elsevier, VNU, Bertelsman Small: Meteo Consult, Dutch Tourist Board
Product	Timing	Market Segment	Example Customers
Standalone Software packages for any of the above	Medium Term	The above two, plus 3. Businesses with large intranets.	The above, plus Philips, ASML, Shell, Unilever
Product	Timing	Market Segment	Example Customers
Integrated toolkit for information repository management	Medium Term	Large companies and organisations	• Banking and Insurance: ABN/AMRO, RABO bank, ING, Fortis, Generale Bank. • Pensiun Funds: PGGM, ABP • Industry: DAF trucks, Phillips, ASML

Figure 2.1: Target groups for dissemination and use of AIdministrator.

Chapter 3

Appendix

On-To-Knowledge Project contribution to be published and presented at the European e-Business and e-Work Conference (eBeW-2000, Madrid, 18-20 October 2000).

On-To-Knowledge: Ontology-based Tools for Knowledge Management

Dieter Fensel, Frank van Harmelen, Michel Klein, Hans Akkermans
Free University Amsterdam VUA, Division of Mathematics and Informatics
De Boelelaan 1081a, NL-1081 HV Amsterdam, The Netherlands
http://www.ontoknowledge.org ; contact: dieter@cs.vu.nl
and
Jeen Broekstra, Christiaan Fluit, Jos van der Meer, *AIdministrator, The Netherlands*
Hans-Peter Schnurr, Rudi Studer, *AIFB, University of Karlsruhe, Germany*
John Hughes, Uwe Krohn, John Davies, *BT, Ipswich, UK*
Robert Engels, Bernt Bremdal, *CognIT, Oslo, Norway*
Fredrik Ygge, *Enersearch AB, Gothenburg, Sweden*
Thorsten Lau, Bernd Novotny, Ulrich Reimer, *Swiss Life, Zürich, Switzerland*
Ian Horrocks, *University of Manchester, UK*

Executive Summary. *On-To-Knowledge*, the European EU-IST project No. 10132, builds an ontology-based tool environment to speed up knowledge management, dealing with the large numbers of heterogeneous, distributed, and semi-structured documents typically found in large company intranets and the World-Wide Web. Results aimed for by the project are: (1) a toolset for semantic information processing and user access; (2) OIL, an ontology-based inference layer on top of the World-Wide Web; (3) an associated methodology; (4) validation by industrial case studies. This paper gives an overview of the *On-To-Knowledge* approach to knowledge management.

1. Heterogeneous Information Resources Need Semantic Access

Support for information and knowledge exchange is a key issue in the Information Society. The exponential growth of online information on intranets and the Web leads to information overload. To cut down on the time wasted in searching and browsing, and reduce associated user frustration, much more selective user access is needed. This is possible by automatic meaning-directed or semantic information processing of online documents. The European IST project *On-To-Knowledge* builds tools that achieve this.

Ontologies [1,2] are the key technology used to describe the semantics of information exchange. Defined as "specifications of a *shared conceptualization* of a particular domain", they provide a shared and common understanding of a domain that can be communicated across people and application systems, and thus facilitate knowledge sharing and reuse. Ontologies will play a key role in growth areas such as knowledge management [3,4] and electronic commerce. In the US, funding agencies have recognized the importance of these issues by setting up the DAML program (http://www.darpa.mil/iso/ABC/BAA0007PIP.htm) that aims at machine-processable semantics of information resources accessible to agents.

The World-Wide Web (WWW) has drastically boosted the availability of electronic information. Already the present first generation of the web has changed our daily practice, and these changes will become even more significant in the near future. However, the Web itself has to change if it is to reach the next level of user service [5]. Currently, the Web is

an incredibly large, (mostly) static information source. The main burden in information access, extraction, and interpretation still rests with the human user. Document management systems now on the market have severe weaknesses:

- *Searching information*: Existing keyword-based search also retrieves irrelevant information that uses a certain term in a different meaning, and misses information when different terms with the same meaning about the desired content are used.
- *Extracting information*: Currently, human browsing and reading is required to extract relevant information from information sources since automatic agents do not possess the commonsense knowledge required to extract such information from textual representations, and they fail to integrate information spread over different sources.
- *Maintaining* weakly structured text sources is a difficult and time-consuming activity when such sources become large. Keeping such collections consistent, correct, and up-to-date requires mechanized representations of semantics that help to detect anomalies.
- *Automatic document generation* would enable adaptive websites that are dynamically reconfigured according to user profiles or other aspects of relevance. Generation of semi-structured information presentations from semi-structured data requires a machine-accessible representation of the semantics of these information sources.

Tim Berners-Lee coined the vision of a *Semantic Web* that provides much more automated services based on machine-processable semantics of data, and on heuristics that make use of these metadata. The explicit representation of the semantics of data accompanied with domain theories (i.e., ontologies) will enable a Knowledge Web that provides a qualitatively new level of service. It weaves together a net linking incredibly large parts of human knowledge and complements it with machine processability. Various automated services will support the human user in achieving goals via accessing and providing information present in a machine-understandable form. This process will ultimately lead to a highly knowledgeable world-wide system with specialized reasoning services that support us in many aspects of our daily life, becoming as central as access to electric power.

The competitiveness of companies depends heavily on how they exploit their corporate knowledge and memory. Most information in modern electronic media is mixed-media and rather weakly structured. This holds for the Internet but also for large company intranets. Finding and maintaining information is a hard problem in weakly structured representation media. Increasingly, companies realize that their intranets are valuable repositories of corporate knowledge. But with the now rapidly increasing volumes of information, turning this into useful knowledge has become a major problem. *Knowledge Management* is about leveraging corporate knowledge for greater productivity, value, and competitiveness [3,4]. Due to Internet-enhanced globalization, many organizations are increasingly geographically dispersed and organized around virtual teams. Such organizations need knowledge management and organizational memory tools that encourage users and foster collaboration while capturing, representing and interpreting corporate knowledge resources and their meaning. *On-To-Knowledge* provides the tools to speed up knowledge management in large distributed organizations, by applying ontologies to electronic information as a basis for semantic information processing and fast, meaning-directed user access.

2. Tool Environment for Ontology-based Knowledge Management

The *On-To-Knowledge* architecture and all its major components are shown in Figure 1. To illustrate these components and their interactions, we present a simple querying

scenario, where a user poses a query to the system that must be answered on the basis of a set of weakly structured data sources in a repository.

Two example scenarios in our case studies are:

- Querying a skills description repository on the Swiss Life intranet, where this repository is filled with a large variety of weakly structured documents (CV's, project descriptions, course descriptions, etc.).
- Locating the material that is required to answer a query at one of the helpdesks operated by BT. Again, much of the relevant material is very heterogenous in nature: technical specifications of devices, previous fault reports, customer data, etc.

The sequence of numbers in Figure 1 indicates the steps that must be taken in order to perform any of the above queries. Each of the components in Figure 1 is based on existing solutions already provided *On-To-Knowledge* consortium partners [6-10]. The red labels indicate which partner is providing crucial technology to each of the steps of the scenario.

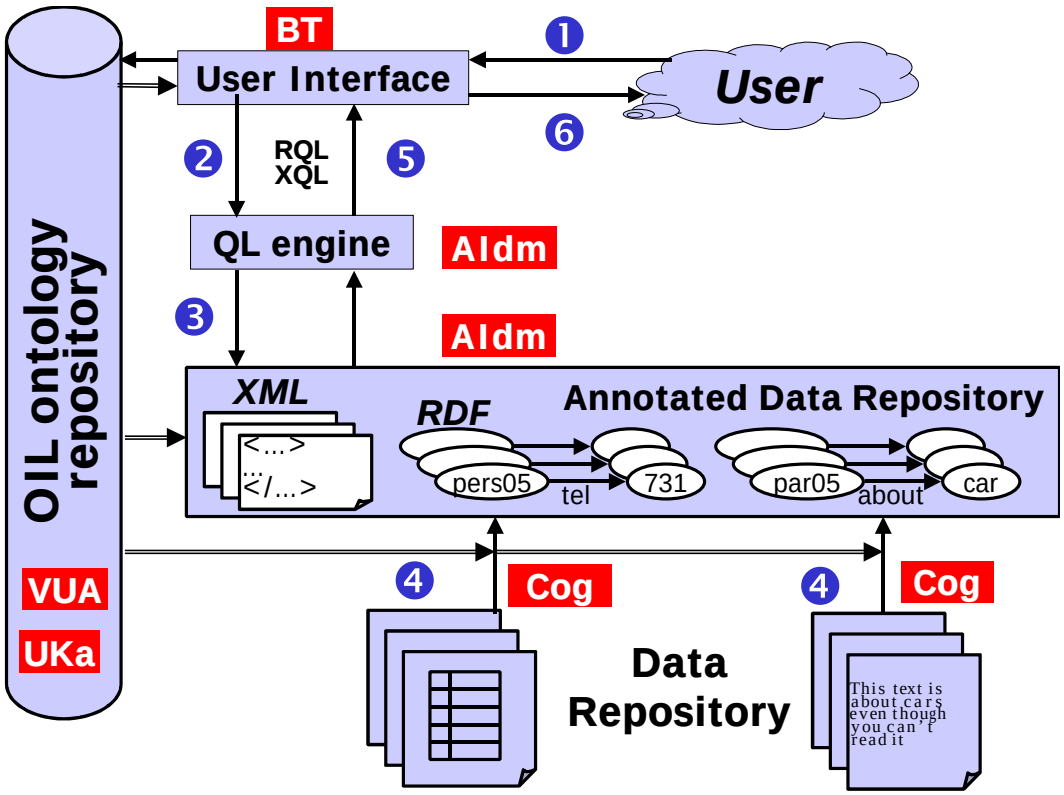

Figure 1. The On-To-Knowledge toolset for ontology-based knowledge management.

Step [1]. The system interacts with a user in order to elicit a specific query to be answered. Both the interaction with the user and the resulting query are entirely in terms of a domain-specific ontology, expressed in the OIL language developed within the consortium (see Section 3). The required ontologies are constructed using tools such as *OntoEdit* (http://ontoserver.aifb.uni-karlsruhe.de/ontoedit/), developed by the University of Karlsruhe [10]. Such an ontology-based user interaction has as main advantage that the user is shielded from any document-specific representation details, and can instead communicate in meaningful domain-specific terms. Furthermore, it makes the system much more robust against changes and variability in the formats of the underlying documents.

Step [2]. The user interaction results in a specific query to be answered by the data repository layer. We rely on the *Resource Description Framework* (RDF, http://www.w3c.org/Metadata/) currently being developed by the World-Wide Web consortium (W3C), to structure the data repository and to express queries over this

repository. The required translation from OIL-based user interaction to RDF-based queries is feasible because OIL is itself definable in terms of RDF-Schema definitions.

Step [3]. The consortium is developing an RDF query engine to efficiently process queries over medium-size data-repositories (with up to a million RDF triples in the repository). The University of Karlsruhe's *SilRI engine* (http://www.aifb.uni-karlsruhe.de/~sde/rdf/) is a starting point for such an engine. Besides RDF, XML may also be used to represent part of the semantically annotated data in the repository. In that case, an XML query language like XQL or XML-QL forms the basis for an XML-based query engine.

Step [4]. Of course, the above steps all assume that the data repository is filled with data that is annotated with sufficiently rich semantic information. Furthermore, the annotations must be related to the ontological vocabulary that was the basis for the original user query. Different technologies will be exploited to provide these annotations, depending on whether we are dealing with weakly structured data sources, or data sources that consist of free text only. In the first case, we will use wrapper technology in the vein of Jedi or W4. In the second case, the *Corporum* technology from CognIT ([9], http://www.cognit.no/) is the main platform for concept extraction from free text. Other tools will be based on automated summarization technology as developed for *ProSum* by BT [7,8].

Steps [5,6]. After the RDF query has been executed over the data repository, the resulting information is communicated to the user. Again, this must be done using an ontology-based vocabulary. Furthermore, powerful graphical visualizations of query results in the context of large data sets are developed. Examples of such visualizations are the semantic sitemaps produced by the *WebMaster* tool of AIdministrator [6] (for some results see Section 4).

3. OIL: Inference Layer for the Semantic World-Wide Web

The technical backbone of *On-To-Knowledge* is the use of ontologies for the tasks of meaningful information access, integration, and mediation. A major result from the *On-To-Knowledge* project is OIL (the ***O**ntology-based **I**nference **L**ayer*) [11]. OIL is a representation and inference layer on top of the Web that is based on ontologies. It unifies three important aspects provided by different communities: (i) formal semantics and efficient reasoning support as provided by Description Logics, (ii) epistemologically rich modelling primitives as provided by the Frame community, and (iii) a standard proposal for syntactical exchange notations as provided by the Web community.

- **Description Logics (DL).** DLs describe knowledge in terms of concepts and role restrictions that are used to automatically derive classification taxonomies. The main effort of the research in knowledge representation is in providing theories and systems for expressing structured knowledge and for accessing and reasoning with it in a principled way. OIL inherits from DL its *formal semantics* and the *efficient reasoning support* developed for these languages. In OIL, *subsumption* is decidable and with the developed *FaCT* engine we provide an efficient reasoner for this.
- **Frame-based systems.** The central modelling primitives of predicate logic are predicates. Frame-based and object-oriented approaches take a different point of view. Their central modelling primitives are classes (frames) with certain properties called attributes. Many other refinements of these constructs have been developed leading to the success of this modelling paradigm. Therefore, OIL incorporates the *essential modelling primitives* of frame-based systems into its language. OIL is based on the notion of a concept and the definition of its superclasses and attributes. Relations can also be defined not as an attribute of a class but as an independent entity having a certain domain and range. Like classes, relations can be arranged in a hierarchy.

- **Web standards: XML and RDF.** Modelling primitives and their semantics are one aspect of an ontology-based exchange language. Given the dominance and importance of the WWW, the syntax of such a language must be formulated using existing web standards for information representation. As already proven with XOL (http://www.ai.sri.com/pkarp/xol/), XML can be used as a serial syntax definition language for ontology-based information exchange. OIL is closely related to XOL and can be seen as an extension of it. Another candidate for a web-based syntax for OIL is RDF and RDFS. The RDF framework for the encoding, exchange, and reuse of structured metadata provides a means for adding semantics to a document without making any assumptions about the structure of the document. RDF schemas (RDFS) provide a basic type schema for RDF. Objects, Classes, and Properties can be described. In relation to ontologies, RDF provides two important contributions: a standardized syntax for writing ontologies, and a standard set of modelling primitives like instance-of and subclass-of relationships. Therefore, OIL offers two syntactical variants: one based on XML schema and one based on RDF schema.

Why not Ontolingua? Ontolingua (http://ontolingua.stanford.edu/) is an existing proposal for an ontology language. It has been designed to support the design and specification of ontologies with a clear logical semantics based on KIF. Ontolingua extends KIF with additional syntax to capture intuitive bundling of axioms into definitional forms with ontological significance; plus a Frame Ontology to define object-oriented and frame-language terms. The problem with Ontolingua is its high expressive power provided without any means to control it. Not surprisingly, no reasoning support has been provided for Ontolingua. OIL takes the opposite approach. We start with a very simple and limited core language. The web has proven that restriction of initial complexity and controlled extension when required is a very successful strategy. OIL takes this lesson to heart.

The use of OIL is currently evaluated in two running IST projects *On-To-Knowledge* and Ibrow (http://www.swi.psy.uva.nl/projects/ibrow/home.html). In *On-To-Knowledge*, OIL will be extended to a full-fledged environment for knowledge management in large intranets and websites. Unstructured and semi-structured data will be automatically annotated, and agent-based user interface techniques and visualization tools will help the user to navigate and query the information space. Here, *On-To-Knowledge* continues a line of research that was set up with SHOE and Ontobroker [5]: using ontologies to model and annotate the semantics of information resources in a machine-processable manner.

4. Business Applications in Semantic Information Access

Industry case studies. *On-To-Knowledge* is carrying out three industrial case studies to evaluate the tool environment for ontology-based knowledge management (Section 2) and the associated web inference layer OIL (Section 3). These case studies are chosen such that they ensure a broad coverage, involving three different industry sectors (insurance, telecom, energy) in three different countries, and facing different knowledge management problems.

Swiss Life: organizational memory. Swiss Life's vision is to build an *organizational memory* with an intranet-based portal. Three case studies explore the problem space:

1. A skills database contains a large variety of structured and unstructured documents like CVs, recruitment profiles, course and project descriptions. Today these documents do not exist or are not integrated into a single repository. Furthermore, there is no common vocabulary (i.e. ontology) that guarantees a unified usage and understanding of the documents, resulting in insufficient retrieval results.

2. Information about an insurance product comprises documents for sales persons, for training purposes, about performing office tasks, etc. This information is created in different places, in different formats and often not distributed to the right places.
3. The IAS document ("International Accounting Standards") is part of the global Swiss Life Intranet, called "GroupNet". The document's 1000 web pages make it very hard to find relevant passages, even though there is a division into chapters and sections.

BT: call centres. Call Centres are an increasingly important mechanism for customer contact in many industries. Every transaction should emphasize the uniqueness of both the customer and the customer service person. To do this one needs effective knowledge management [7, 8]. This includes knowledge about the customer but also knowledge about the customer service person, so that the customer is directed to the right person to answer their query. This knowledge must also be used in a meaningful and timely way. One or more of BT's own Call Centres will be targeted to identify opportunities for effective knowledge management using the *On-To-Knowledge* tools. More specifically, call centre agents tend to use a variety of electronic sources for information when interacting with customers, including their own specialized systems, customer databases, the organization's intranet and, perhaps most importantly, case bases of best practice. The *On-To-Knowledge* techniques provide an intuitive front-end to these heterogeneous information sources, to ensure that the performance of the best agents is transferred to the others.

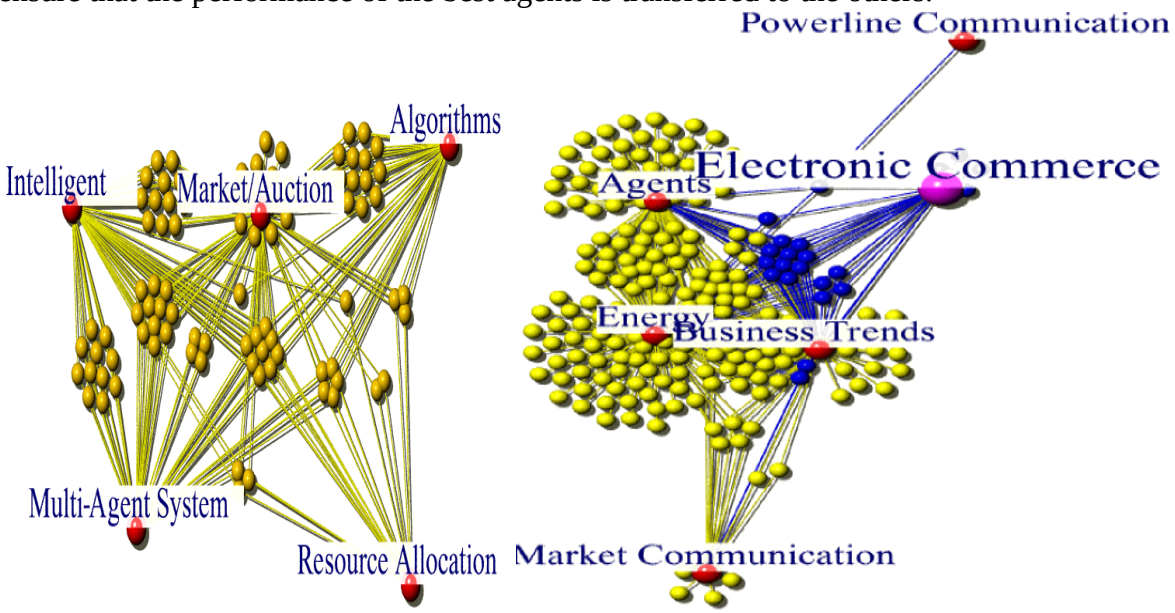

Figure 2. Automatically generated semantic structure maps of the EnerSearch website.

EnerSearch: virtual enterprise. EnerSearch is a virtual organization researching new IT-based business strategies and customer services in deregulated energy markets (e.g., [12], see further http://www.enersearch.se). Its research affiliates and shareholders are spread over many countries: its shareholding companies include IBM (US), Sydkraft (Sweden), ABB (Sweden/Switzerland), PreussenElektra (Germany), Iberdrola (Spain), ECN (Netherlands), and Electricidade do Portugal. Essentially, EnerSearch is a knowledge creation company, knowledge that must be transferred to its shareholders and other interested parties. Its website is one of the mechanisms for this. However, it is rather hard to find information on certain topics – the current search engine supports free text search rather than content-based search. Therefore, the EnerSearch case study evaluates the *On-To-Knowledge* toolkit to enhance knowledge transfer to (1) researchers in the EnerSearch virtual organization in different disciplines and countries, and (2) specialists from shareholding companies interested in getting up-to-date information about R&D results on IT in Energy.

Some first results of *On-To-Knowledge* techniques are shown in Figure 2. It shows two semantic structure maps of the EnerSearch website, produced by the WebMaster tool of AIdministrator [6], and based on a domain ontology concerning important IT in Energy research topics. Every node represents a webpage (that can be directly opened in a browser by clicking on the node); edges denote hypertext links. Left, we see a map of subtypes (subtopics) of agent research by EnerSearch. It is easy to see how subtopics are related and find the relevant webpages. Right, we see an interactively generated sitemap showing how the topic of e-commerce intersects with other topics (dark blue nodes). A nice feature of the visualization is that spatial proximity correlates very well with semantic closeness.

Methodology. In addition to the toolset and the OIL language, *On-To-Knowledge* is developing an associated methodology for ontology-based knowledge management. Input to this are existing European research results, such as the *CommonKADS* approach to knowledge engineering and management [3], experiences from knowledge-based software engineering [12] and tool development [5-10], ontology composition [2] and information retrieval techniques [14], and feedback from industry case studies. The methodology will also cover how to develop the business case for ontology-based knowledge management.

Conclusion. World-Wide Web and company intranets have boosted the potential for electronic knowledge acquisition and sharing. Given the sheer size of these information resources, there is a strategic need to move up in the data – information – knowledge chain. As a necessary step, *On-To-Knowledge* provides innovative tools for semantic information processing and thus for much more selective, faster, and meaningful user access.

References

[1] D. Fensel: *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, Berlin, D, to appear (2000).

[2] P. Borst, J.M. Akkermans, and J.L. Top: Engineering Ontologies, *International Journal of Human-Computer Studies* **46** (1997) 365-406.

[3] A.Th. Schreiber, J.M. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. Van De Velde, and B. Wielinga: *Knowledge Engineering and Management.* The MIT Press, Cambridge, MA, 2000.

[4] U. Reimer (Ed.): *Proc. 2nd Int. Conf. on Practical Aspects of Knowledge Management* (PAKM'98), Basel, Switzerland, October 1998. URL: http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-13/.

[5] D. Fensel, S. Decker, M. Erdmann, H.-P. Schnurr, R. Studer, and A. Witt: Lessons learnt from Applying AI to the Web. *Journal of Cooperative Information Systems*, to appear (2000).

[6] F. van Harmelen and J. van der Meer: WebMaster: Knowledge-based Verification of Web-pages. In *Proceedings 2nd Int. Conf. on The Practical Applications of Knowledge Management* (PAKeM99, London, UK, April 1999), pp. 147-166. The Practical Applications Company, Blackpool, UK, 1999.

[7] J. Davies, S. Stewart, and R. Weeks: Knowledge Sharing over the World Wide Web, *WebNet '98*, Florida, USA, November 1998. (also at http://www.bt.com/innovation/exhibition/knowledge_management/).

[8] J. Davies: Supporting Virtual Communities of Practice, in R. Roy (Ed.): *Industrial Knowledge Management*, Springer-Verlag, London, forthcoming (2000).

[9] B. Bremdal, F. Johansen, C. Spaggiari, R. Engels, R. Jones: Creating a Learning Organisation Through Content-Based Document Management, OECD HRP Meeting, Loen, NO. CognIT Report, Oslo, May 1999.

[10] A. Maedche, H.-P.Schnurr, S. Staab, and R. Studer: Representation Language-Neutral Modeling of Ontologies. In: Frank (Ed.), *Proceedings German Workshop Modellierung 2000*. Koblenz, D, April 2000.

[11] I. Horrocks, D. Fensel, J. Broekstra, S. Decker, M. Erdmann, C. Goble, F. van Harmelen, M. Klein, S. Staab, and R. Studer: *The Ontology Inference Layer OIL*, On-To-Knowledge EU-IST-10132 Project Deliverable No. OTK-D1, Free University Amsterdam, Division of Mathematics and Informatics, Amsterdam, NL, 2000. Available from http://www.ontoknowledge.org/oil.

[12] F. Ygge and J.M. Akkermans: Decentralized Markets versus Central Control - A Comparative Study, *Journal of Artificial Intelligence Research* **11** (1999) 301-333. (Also available from http://www.jair.org/).

[13] J. Angele, D. Fensel, and R. Studer: Developing Knowledge-Based Systems with MIKE, *Journal of Automated Software Engineering* **5** (1998) 389-418.

[14] Y. Ding, G. Chowdhury, and S. Foo: Bibliometric Information Retrieval System (BIRS): A Web search interface utilizing bibliometric research results. *Journal Am. Soc. for Information Science*, to appear (2000).