SWAD–Europe Deliverable 12.1.1: Semantic web applications – analysis and selection Appendix B – Application Survey

Project name:
Semantic Web Advanced Development for Europe (SWAD-Europe)
Project Number:
IST-2001-34732
Workpackage name:
12.1 Open Demonstrators
Workpackage description:
☞http://www.w3.org/2001/sw/Europe/plan/workpackages/live/esw-wp-12.1.html
Deliverable title:
Semantic web applications - analysis and selection
URI:
☞http://www.w3.org/2001/SW/Europe/reports/hp-applications-selection.html
Authors:
☞Dave Reynolds , Steve Cayzer, Ian Dickinson, HP Laboratories, Bristol, UK
Paul Shabajee, Graduate School of Education and ILRT, Bristol, UK

Abstract:
This report concerns the selection of two open demonstrator applications designed to both illustrated the nature of the semantic web and to explore issues involved in developing substantial semantic web applications given the current state of the art.

This report section is Appendix B of the main report and contains an informal survey of some 60 semantic web applications or proposed applications.
Status:
Completed.
Comments on this document are welcome and should be sent to ☞Dave Reynolds or to the ☞public-esw@w3.org list. An archive of this list is available at ☞http://lists.w3.org/Archives/Public/public-esw/

B Application survey

The appendix contains a summary of an RDF database of notes and example applications and suggestions which was developed during the course of this work. The survey is in no way comprehensive - there are many applications that we have missed or not had time to create a record for. Even for those that have been captured our short descriptions may well not do justice to the depth of research and development activities involved. Think of these as an example flags in a map of semantic web applications, not as in depth reviews.

Catalogue/thesaurus management -

- ☞Catalogue Management
- ☞Catalogue integration
- ☞DMOZ - Directory Mozilla - open directory
- ☞Thesaurus management

Data dependent agents -

- ☞Financial Assistant
- ☞ITTalks
- ☞Jema
- ☞Shopping assistants
- ☞Virtual Travel Agent

Data integration -

- ☞B2B web service mediation
- ☞Catalogue integration
- ☞Database integration example
- ☞Financial Portals

- ☞Gene Ontology
- ☞Mozilla
- ☞eScience Data Grids

Knowledge formation -

- ☞Assumption tracker
- ☞Bibliography workbench
- ☞ClaiMaker/Scholonto
- ☞Community Arkive
- ☞Community formation
- ☞DMOZ - Directory Mozilla - open directory
- ☞Ideas workbench
- ☞SWAP - semantic web and peer-to-peer

Knowledge management -

- ☞Community formation
- ☞Community portals
- ☞Helpdesk support
- ☞OntoShare - community of practice support
- ☞PatMan
- ☞PlanetOnto
- ☞Sun GKE
- ☞ePerson

Metadata for annotation and enrichment -

- ☞Annotea
- ☞Assumption tracker
- ☞Bibliography workbench
- ☞Community Arkive
- ☞Community bookmarking
- ☞Distributed topic portals
- ☞EARL
- ☞Gene Ontology
- ☞MIT/HP SIMILE project

Metadata for discovery and selection -

- ☞B2B trading market-places
- ☞DCMI registry
- ☞Edutella
- ☞HP Portal
- ☞MIT/HP SIMILE project
- ☞MUSE
- ☞Recommendation Networks
- ☞Scholnet
- ☞SeLeNe
- ☞Semantic tagging
- ☞Sun GKE
- ☞Web service description and discovery

Metadata for media and content -

- ☞Adobe XMP
- ☞Arkive internal
- ☞MIT/HP SIMILE project

Ontology Management -

- ☞APECKS
- ☞Gene Ontology
- ☞KAON

- ☞SWAP - semantic web and peer-to-peer

Personal information management -

- ☞Bibliography workbench
- ☞Community bookmarking
- ☞Event tracking
- ☞Genealogy assistant
- ☞Haystack
- ☞Ideas workbench
- ☞Jema
- ☞Mozilla
- ☞ePerson

Semantic indexing -

- ☞Community Arkive
- ☞Community portals
- ☞Context aware links
- ☞Curriculum Online
- ☞Distributed topic portals
- ☞HP Portal
- ☞ITTalks
- ☞Museum portals
- ☞MusicBrainz
- ☞PlanetOnto
- ☞Score
- ☞Semantic tagging
- ☞TAP semantic search

Syndication Category -

- ☞Event tracking
- ☞Rich Site Summary/RDF Site summary
- ☞Syndication

APECKS

Summary	Collaborative Ontology management tool
Description	APECKS stands for "Adaptive Presentation Environment for Collaborative Knowledge Structuring". The central goal is the collaborative management of ontologies. Experts produce personal ontologies, which can be compared, and their degree of consistency can be used to assist the experts to construct a richer ontology. Hence, the tool enables discovery of consensus, conflict (same name, different concept), correspondence (different name, same concept) and contrast. Once discovered, any discrepancies can be corrected, or indeed they might point to a difference in world view that is worth exploring. Hence,the tool also encourages & supports asynchronous discussions.
Technical issues	- While conflict and correspondence are simply resolved by changing concept labels (or shuffling instances), contrast is more difficult both to explain to the user and to resolve - APECKS works best if t he schemas are highly similar - "about" the same thing, and at approximately the same level of granularity
Comments	APECKS uses a Frame representation schema (which may provide import support for other ontologies).
Category	☞Ontology Management
Reference	☞http://ksi.cpsc.ucalgary.ca/KAW/KAW98/tennison/

Adobe XMP

Summary	Common format for representing metadata embedded with media files
Description	Adobe eXtensible Metadata Platform (XMP) is a common format for embedding metadata inside media files. It is based on RDF (with a few restrictions to simplify processing) and is thus extensible and supports data in externally defined RDF schemas. The framework itself is available open source and Adobe plan to support it across all

	Adobe applications (currently supported by Photoshop 7.0, Acrobat 5.0, FrameMaker 7.0, GoLive 6.0, InCopy 2.0, InDesign 2.0, Illustrator 10 and LiveMotion 2.0).
Comments	By embedding the data within the application files, XMP ensures that the data/metadata binding is preserved as the data passes through databases or publishing systems. This uniform and extensible approach should support repurposing and archiving of media objects and automation of publishing workflows.
Application status	existing
Category	☞Metadata for media and content
Link	☞http://www.adobe.com/products/xmp/main.html

Annotea

Summary	General web page annotation infrastructure
Description	Annotea is a W3C project to demonstrate the use of RDF for annotating web pages. Annotations are elements of contributed text that anyone, i.e. not just the original author, may add to a web page. Annotea defines a protocol for accessing the annotations from an annotation server, and a RDF format for storing the annotation text.
	Annotea requires an annotation-aware browser to work. Currently Annotea is built-in to the W3C's reference/experimental browser Amaya. When browsing a web page, Amaya will query the annotation server that the user has set up (may be local to a workgroup, or may use the public W3C server), and add a graphic to the rendered page indicating that an annotation exists. Annotations may be specific to a region of a page; regions are identified using ☞XPointer. Once an annotation has been defined, other annotation users can reply to it, similar to a threaded email discussion.
Technical issues	+storing potentially large volumes of RDF annotations +providing simultaneous query and update to the RDF store, possibly with a high transaction rate for queries +what to do with annotation collections that age (for example, the page structure changes so the XPointer path is no longer valid)
User value	+ knowledge dissemination by allowing multiple contributors to add to a public web page + broadening the available information about a topic *in situ*, rather than via a search engine
Application status	Deployed prototype server at W3C, with support in Amaya browser. Ongoing projects to develop Annotea clients (see specific examples).
Category	☞Metadata for annotation and enrichment
More information	☞http://www.w3.org/2001/Annotea/
Reference	☞http://www.w3.org/2001/Annotea/Papers/KCAP01/annotea.html
Specific example	☞http://annozilla.mozdev.org/ ☞http://www.jibbering.com/snufkin/

Arkive internal

Summary	Metadata for internal management of a rich multimedia archive (including technical, content and provenance metadata)
Description	The ☞Arkive project aims to collect and preserve digital images and recordings of endangered species and make them available to educationalists, researchers and the public. It is an initiative of the Wildscreen Trust and numbers HP among its sponsors.
	At present information on the media accessions (technical information, rights information, provenance information, basic categorization) is stored internally using a customized database schema. This is an example large scale repository and content management problem where RDF could be applied for representation of the metadata. This might have advantages - easier extension, simpler processing of imported metadata, ability to export subsets of the metadata to support external user communities.
	This specific example is included as a concrete instance of the general class of metadata for media repositories and content management systems.
Technical issues	- scale and performance of metadata store - evolution of classification terminologies - metadata for time-based media
Application status	suggestion only
Category	☞Metadata for media and content
More information	☞http://www.arkive.org.uk/

Assumption tracker

Summary	Trace provenance of pieces of information underlying a multimedia archive
Description	In existing content management systems there is support for repurposing media assets so that a change in one asset can be propagated. However, in developing complex sites we find that a single "piece of information" will be used as an assumption or basis for generating many assets - text in a document, statistic in a graph, narration on a video. For complex repositories there is thus value in tracking these individual "pieces of information" and linking them to the assets that are derived from them to support both construction and maintenance. The RDF level of the semantic web offers a common format for representing such loosely defined "pieces of information" and the ontology level offers a tool for representing the domain model through which such items can be indexed and structured.
Application status	suggestion
Category	☞Knowledge formation ☞Metadata for annotation and enrichment
Reference	*The Capture and track of "Pieces of Information": Necessary Requirements for Educational and Rich Repurposing Architectures*, Paul Shabajee and Dave Reynolds, 2002, submitted.
Related example	☞Community Arkive

B2B trading market-places

Summary	Semantically rich description of products/services both offers and requests to support discovery and matching
Description	Various forms of B2B marketplaces have been developed which enable buyers of products or services to find suitable sellers, negotiate deals and perform transactions. Both the offer advertisements and the buyer requirements require complex descriptions of features and components and the process of matching these requires at least subsumption reasoning and often partial matching. Work has been done to apply DAML+OIL to the task of representing the offers and requirements and thus be able to exploit description logic reasoning engines to perform the matchmaking.
Technical issues	- partial matching of DL descriptions
Application status	research prototypes
Category	☞Metadata for discovery and selection
Reference	*Description Logics for Matchmaking of Services*, Javier Gonzalez-Castillo, David Trastour, Claudio Bartolini, ☞HPL-2001-265.

B2B web service mediation

Summary	Allow web services to be composed by translating terms (both shallow syntax and deep structure) between different service interfaces
Description	The web services vision includes the notion that consumers of web services will be able to solve complex problems by composing services together. However, the services to be composed many not directly plug together. The issue of message and envelope formats does not concern us - common standards or simple syntactic translation should enable message level interoperability. The harder issue is how to marry the data structures through which the different services communicate. In some cases a compatible conceptualization might be in use (e.g. two services might share a common notion of a "purchase order") and the problem is again syntactic. However, more commonly the domain model will be different, or impose different constraints (number of line items allowed for example) or refer to different vocabularies (for example different product catalogue schemes). The semantic web, especially the ontology layer, could be used to describe the semantics of these data structures and the underlying conceptual models to permit a mediation layer to compose services together. Full service composition will require other functionality, in particular the ability to mediate services with different conversation models (different message sequences).
Application status	active research
Category	☞Data integration
Link	☞http://swws.semanticweb.org/

More information	☞http://www.cs.vu.nl/~dieter/wese/
Reference	

B2C service mediation

Summary	Personal agent that can combine multiple web services to meet an single customer's needs
Description	
Technical issues	
Comments	
User value	
Application status	
Category	☞Data integration
Link	☞file:rdf/b2cMediation.rdf
More information	☞file:rdf/b2cMediation.rdf
Reference	
Related example	☞http://www.hpl.hp.com/semweb/2002/10/semweb-app-schema#foo.rdf
Specific example	☞http://www.hpl.hp.com/semweb/2002/10/semweb-app-schema#foo.rdf

Bibliography workbench

Summary	Tools to manage and categorize personal bibliography databases and link them across communities
Description	Personal citation bibliographies are often more than just a collection of references to be included in papers. They represent the selective reading trail of the researcher and so offer a personal view onto a collection of related research topics. They include comments, evaluations and classifications which enrich the references themselves. It would be valuable to directly capture the relationships between papers ("supports", "refutes", "supersedes") so that the whole collection becomes greater than the sum of the parts. RDF would have value in building a personal information management tool that support bibliography management. The big advantage would appear when such tools were linked together, to allow personal collections to be merged across a community while still allowing local differences in structure. It would also be valuable if the subject hierarchies themselves could be extended by the community and related to centralized term Thesauri without undue load on individual researchers. This dual management of metadata (the bibliograpies) and term vocabularies across a collaborating network of researchers does seem to illustrate many of the facets of the semantic web vision.
Technical issues	- distribution solution to allow local collections to be effectively aggregated - decentralized vocabulary management and evolution
Application status	suggestion
Category	☞Personal information management ☞Metadata for annotation and enrichment ☞Knowledge formation
Related example	☞ClaiMaker/Scholonto

Catalogue Management

Summary	Representation and management of taxonomic schemes, e.g. for product catalogues
Description	There are many situations where "things" need to be tagged with their classification in some taxanomic classification scheme and where that classification scheme evolves. For example classification of products into standard catalogues (e.g. ☞UNSPSC, ☞NAICS and ☞Eclass) which are required for some eCommerce systems (e.g. Commerce One). This problem is related to thesauri for digital libraries and thesauri of controlled terms for government document repositories. What makes this a challenging problem is that the thesauri change (UNSPSC changes regularly with categories disappearing, categories appearing, and entire branches moving to different parts of the tree) and the same items may need to be classified into many catalogues. The challenge is to reduce the cost of managing this. For example, a deeper ontology might enable terms to classified once and then have the classification resued by mapping the deeper ontology onto specific taxonomies as they arise and evolve.

Technical issues	- efficient processing of large vocabularies - automated mapping techniques for mapping between vocab versions. - automated classification. - development of concept ontolo gies for use as stable intermediaries
Application status	active research
Category	☞Catalogue/thesaurus management
Related example	☞Catalogue integration ☞Thesaurus management ☞MIT/HP SIMILE project

Catalogue integration

Summary	Combining multiple catalogues (e.g. software product catalogues) to provide indexing of a set of products or services.
Description	Product catalogues are large vocabularies which classify products and services and assign them unique codes and a position in some taxanomic structure. Some catalogues are specific to specific vendors or narrow industries but there are several large product encoding standards in common use such as ☞UNSPSC, ☞NAICS and ☞Eclass. A significant issue in business integration applications is to map different product catalogues together. A market intermediary may receive product information in several catalogue forms and yet needs to provide a common access solution for the customers. It may do this by attempting to translate all the catalogues into a common unified catalogue or by translating queries into the different source taxonomies. The semantic web ontology layer might provide a means for capturing partial mappings between catalogues. In particular, the a common ontology mapping representation would allow such partial mappings to be discovered and reused.
Technical issues	- ontology mapping discovery - integration of partial ontology mappings - adaptation of ontology mappings to track ontology evolution
Application status	active research
Category	☞Data integration ☞Catalogue/thesaurus management
Reference	E.g. *Integrating Vocabularies: Discovering and Representing Vocabulary Maps*, Borys Omelayenko, Procs. First International Semantic Web Workshop, Sardinia, 2002. *A Data Integration Framework for e-Commerce Product Classification*, Sonia Bergamaschi, Francesco Guerra, Maurizio Vincini, Procs. First International Semantic Web Workshop, Sardinia, 2002.

ClaiMaker/Scholonto

Summary	Make explicit the claims made by scientific papers and the network of relationships (refutes, supports ...) between them
Description	The usability of research papers on the Web would be enhanced by a system that explicitly modeled the rhetorical relations between claims in related papers. ClaiMaker is a system for modeling readers' interpretations of the core content of papers. It provides tools to build a Semantic Web representation of the claims in research papers using an ontology of relations. The system can then be used to make inter-document queries.
Technical issues	The technical issues include: • How do you extract claims from a paper (perhaps the author must do this, but then there is a higher barrier to entry). • The ontology for scholarly discourse (claims, refutes, support etc) may not be stable over time. • The semantic links might be time dependent - a theory may reconcile two previously apparently contradictory findings. • A simple modeling be may insufficient: eg '[Paper] :claims [XYZ]' is not necessarily a simple (uncontroversial) ground fact • Claims can be grouped into a set. Linking those sets may be problematic if set membership changes over time.
Comments	Searching for papers using free text search is clearly limited. Even citations aren't ideal (for example, you don't know whether the authors agreed or disagreed with the authors they cite). However, if we can represent a paper by the claims that it makes, then it is possible to link this into a network of claims (made by other papers). This is done using

an ontology of rhetorical relations (supports, refutes etc). Thus one particular use case might be:

1. Find claims made by this paper.
2. Extend concepts (those claims positively linked) to provide an 'intellectual framework' for this paper.
3. Find claims challenging the intellectual framework of this paper.

Claims can also be collected into related 'sets' which can themselves be linked together.
    Future work envisaged by the group include query (both intuitive, for the novice, and rich, for the expert) Also, visualisation of the claim network

Category	☞Knowledge formation
Reference	Gangmin Li, Victoria Uren, Enrico Motta, Simon Buckingham Shum and John Domingue. ClaiMaker: Weaving a Semantic Web of Research Papers. LNCS . 2002; 2342:436 ff. Notes: Knowledge Media Institute, The Open University, Milton Keynes, MK7 6AA, UK {g.li, v.s.uren, e.motta, s.buckingham.shum, j.b.domingue}@open.ac.uk http://kmi.open.ac.uk/projects/scholonto/

Community Arkive

Summary	Community annotation and structuring of rich multimedia content using emergent ontologies
Description	☞Arkive is a collection of media related to endangered and UK species. Many different user communities (teachers, lecturers, researchers) have different needs for how to access such a collection. In particular, to index items according classifications of habitat, behaviours and species taxonomies and to link this information to other information sources (both external and developed by the communities themselves). The key insight to here is that rather then expecting a globally accepted ontology for things like habitat, we imagine each community wanting to use (or develop or at least enrich) different ontologies based on their different needs. Specialists in slime molds have different needs from lecturers looking for materials on bird behaviours. Yet there will be situations where these communities overlap, where it would be valuable to have a common view of the same set of resources.
Technical issues	- ontology reuse to enable convergent ontology development - ontology transformation for mapping different schemes - coping with inconsistency - will need to be able to store different mutually inconsistent views of the world.
Application status	suggestion
Category	☞Semantic indexing ☞Knowledge formation ☞Metadata for annotation and enrichment
Link	☞http://www.arkive.org.uk/
Related example	☞MIT/HP SIMILE project ☞Catalogue Management

Community bookmarking

Summary	Sharing web bookmarks (with categorization and annotation) across a user community
Description	A user's collection of web references (bookmarks or favorites) is not just a useful index for that user but is potentially a valuable community resource. By combining the annotated bookmarks of colleagues with similar interests one may be able to discover new sites and information sources relevant to a given problem with more focus than possible through generic search tools. "Harness the information filtering capability of other people in the system to separate the signal form the noise." The semantic web could be used to implement a decentralized bookmark sharing network to support exchange of annotated and classified personal bookmarks.
Application status	active
Category	☞Personal information management ☞Metadata for annotation and enrichment
Reference	*Using Memex to archive and mine community Web browsing experience* , Soumen Chakrabarti et al, WWW9, Amsterdam, May, 2000. *Aggregating Recommendations using RDF*, Libby Miller, ILRT. ☞http://ilrt.org/discovery/2000/09/rudolf/recommender.html

Related example	☞On Line Bookmarks ☞NeuroGrid surfind servlet
Specific example	☞ePerson

Community formation

Summary	Explore and exploit social networks for finding experts and fellow practitioners (c.f. foaf)
Description	A common theme in knowledge management is how practitioners can find other practitioners with similar interests with whom to exchange information and collaborate. In large organizations there are often individuals and groups working in overlapping areas who could benefit from sharing of experiences and tools but who are unaware of each other. Similarly individuals looking to solve a problem may have difficulty finding an expert in the appropriate areas to help them. Many large organizations have centralized databases in which individuals can register their skills and interests to assist with this. The semantic web could offer a decentralized solution to the same problem in which individuals could post information on the skills, interests, and others they already work with. Tools could then walk this *personal semantic web* to answer such "find an expert on..." questions. One significant example of this in the RDF world already is the "friend of a friend" ☞schema developed by Dan Brickley and colleagues. This supports explicit "knows" links between people which is valuable both to the users (for exploring social networks) and to the tools (as a means to discovering and aggregating foaf descriptions).
Technical issues	- decentralized management of skill/interest ontologies
Application status	active
Category	☞Knowledge management ☞Knowledge formation
Related example	☞http://rdfweb.org/ ☞ePerson ☞OntoShare - community of practice support

Community portals

Summary	Semantic indexing applied to topic and community specific portals
Description	Community portals are web sites which serve as portals for the information needs of particular communities. Typically they provide access to a range of content - links, documents, future events, tools - of interest to that community. Several groups are developing semantic-web based tools to allow the content of such portals to be organized using community-specific ontologies to support improved query facilities, richer navigation and reduced management costs. A common feature of such portals is that the information is largely contributed by the community itself and not by a centralized organization.
Application status	several existing, active and in research
Category	☞Semantic indexing ☞Knowledge management
Reference	*AI for the web - ontology-based community web portals*, Steffen Staab et al, AAAI 2000/IAAI 2000, Austin, July 2000. *Querying RDF Descriptions for Community Web Portals*, Greg Karvounarakis et al, Proc. BDA'2001, Agarid, Maroc, 2001. *SEAL - Tying up information integration and web site management by Ontologies*, Alex Maedche, et al, IEEE Data Engineering Bulletin, 2002.
Related example	☞Community Arkive ☞OntoShare - community of practice support
Specific example	☞PlanetOnto

Context aware links

Summary	Ability to pop up information and links relevant to terms in a web page, based on the semantics rather than just text of those terms
Description	When browsing a web page a user is able to select an word or a phrase and find links to definitions of the word or other semantically relevant links. For example a plain English

word might be simply looked up in a dictionary, the phrase "Grateful Dead" might be recognised as that of a music group and would link you to RollingStone.com. This would happen based on the local viewing tool exploiting vocabularies and ontologies discoverable over the semantic web - the page author does not explicitly embed these links in the page.
The name and examples are drawn from a paper by Joshua Allen.

Application status	proposed
Category	☞Semantic indexing
Reference	*Making a semantic web*, Joshua Allen, Feb 2001. ☞http://www.netcrucible.com/semantic.htm
Related example	☞TAP semantic search

Curriculum Online

Summary	Semantic indexing and related metadata to link Educational resources to UK National Curriculum requirements via a concept ontology
Description	Standardised format for metadata describing education resources and how they relate to the national curriculum. This data is then used to power portals which allow educators to find resources (both web-based and normal commercial resources) which meet a given need. There is already a UK Government online portal which links the National Curriculum to available learning resources with a new version "Curriculum Online" due to go live around now. For "Curriculum Online" there is a well defined metadata scheme based on the IEEE Learning Objects Model with 59 elements, 21 mandatory. There is a defined binding for this both into both XML and into RDF. For many of the elements there are then controlled vocabularies. In particular there is a defined structure of around 2000 concepts and themes that can be mapped onto National Curriculum requirements. The XML binding just represents these as strings and has no structure, but the RDF binding creates URIs for all of them and links them into categories. There is a SOAP service for uploading and managing your metadata entries in the catalogue. There is also an attempt to create a more widespread UK standard that can be used by web sites that might become part of the "National Grid for Learning". This extends the core Curriculum Online components. The Virtual Teacher Centre is both an example of applying this and source of the standards documents. This recommends how web sites should self-describe the content they are offering and place that description in an rdf file accessible from the site.
Technical issues	Unclear there are any deep issues to be solved here. The data is aggregated so no distributed query required (c.f. Edutella). The portals are, currently at least, yahoo style navigation rather than structured search forms so query formation doesn't appear to be an issue. For Curriculum Online the trust is handled by password protection of the SOAP upload port. Mapping to other concept classifications might be of interest.
Application status	existing active
Category	☞Semantic indexing
Link	☞http://www.dfes.gov.uk/curriculumonline/tech.shtml ☞http://vtc.ngfl.gov.uk/metadata/
Related example	☞TAP semantic search

DCMI registry

Summary	Metadata for management of an information repository
Description	The DCMI metadata registry is a project of the OCLC research team. It holds information on the Dublin Core vocabulary and the relationships between the vocabulary terms, and does so across 23 languages. It uses RDF for representing the metadata and can export the information in RDF format.
Application status	existing
Category	☞Metadata for discovery and selection
More information	☞http://www.oclc.org/research/projects/dcmi_registry/index.shtm
Reference	☞http://dublincore.org/dcregistry/index.html

DMOZ - Directory Mozilla - open directory

Summary	A structured directory of web resources, manually indexed by a large team of volunteer editors
Description	The ☞Open Directory project is a Yahoo!-like index of web references categorized into a directory-like structure. The topic structure itself, as well as the classification and description of the contained references, are contributed by thousands of volunteer editors worldwide. DMoz.org make the raw data, both structure and content data, available for free in RDF format. As of April 2002 this was a 16 million statement RDF data source containing information on some 3,005,746 web pages categorized into 428,590 categories.
Technical issues	- scalable storage and retrieval of RDF
Comments	Note that the data is not strictly standard RDF but can be made so with the aid of few UNIX scripts.
Application status	existing
Category	☞Knowledge formation ☞Catalogue/thesaurus management
Link	☞http://dmoz.org/rdf/

Database integration example

Summary	Integration of multiple database to allow cross-database query
Description	Enterprise information systems often comprise several different corporate databases developed by different groups, at different times, for different custom purposes. Situations arise when to answer a specific query information from several databases needs to be combined. Unless the databases were derived from some common underlying schema such integrated query requires more than access-level integration, it requires translation of terms (relationship names, category labels, object identifiers etc) i.e. semantic level integration. A specific example of this, which has been tackled using semantic web techniques, is work by Boeing Corporation on integration of aircraft maintenance and aircraft design databases. This used RDF as a common data model, a query language on top of RDF to express the target query and a mediator architecture to translate the query both syntactically (e.g. to SQL) and semantically (to the data model of the target database). The common domain ontology was expressed using RDFS and then mapped to the different database schemas.
Technical issues	- cross-database query optimization
Application status	existing
Category	☞Data integration
Reference	*RDF Representation of Metadata for Semantic Integration of Corporate Information Resources*, Tom Barrett et al, Proc. WWW-2002, Hawaii, 2002. ☞http://www.cs.rutgers.edu/~shklar/www11/final_submissions/paper3.pdf

Distributed topic portals

Summary	Enable different groups to create and host components of a distributed and semantically indexed topic portal
Description	In technology and research domains it is common for specialist groups to create local topic-specific portals which provides access to technical information, papers, events, web references relevant their domain. These are not simply web sites for that specialist group but represent that group's view onto the entire set of resouces relevant to the domain. Sometimes a single such topic portal will serve all the needs of a given community. More commonly many research groups will develop their own portals which overlaps with other existing sites but also add specialist information of particular interest to them. If the content metadata in these portals were expressed in a common form using either a shared or an interoperable set of ontologies for information categorization, then it would be possible to create an overarching virtual portal. This would give a common view of the entire communities information resources without any single group having to take the burden of creating and managing a main central repository.
Application status	suggestion
Category	☞Semantic indexing ☞Metadata for annotation and enrichment
Related example	☞Community portals

☞Edutella

EARL

Summary	Collection and annotation of results of accessibility testing on web applications
Description	EARL is a W3C-defined language to record the results of *evaluations* (i.e. tests) applied to web resources. A typical use case, though not the only application, is to record the results of applying tests for conformance to accessibility standards to web resources. EARL is an RDF application - that is, the EARL language is an RDF schema, defining a namespace and well-known properties for EARL's intended uses. A goal in EARL is to be able to declare not just the result, but the provenance for it: who claims the result, what the testing method was, when the test was run, etc. Thus the RDF statements that record the actual result of the test (e.g. page `http://foo.com/bar.html` passes Test01) are reified, and made the subject of `earl:asserts` statements.
Technical issues	+ recording provenance of test results + building a shared space of types of tests and results so that data data can be re-used by other tools
Comments	Current released version of the schema is EARL 0.95, dated December 2001. Further versions are under active development.
User value	Consistent way of describing results that can be automatically generated and processed by other tools.
Category	☞Metadata for annotation and enrichment
More information	☞http://www.w3.org/2001/03/earl/
Specific example	☞http://www.w3.org/2001/03/earl/0.95#

Edutella

Summary	P2P infrastructure for distributed RDF query applied to metadata on Educational resources
Description	Edutella is a project which explores many of the issues to do with community annotation, focusing on the realm of educational metadata. The goals of the project cover query, replication, mapping, mediation, and annotation. It is quite a large, collaborative project; the project partners are CID (from Stockholm), Stanford (esp Stefan Decker), Hannover, Karlsruhe. PADLR is the encompassing proposal

Comments	The main focus appears to be on the query level, which has has various levels of expressibility. The query is accomplished using the JXTA methodology. The syntax is basically datalog (ie property based) though with different levels of expressibility:
	
	• RDF-QEL-1: Conjunction
	• RDF-QEL-2: As above + disjunction
	• RDF-QEL-3: As above + negation of literals (ie Datalog)
	• RDF-QEL-4: As above + recursion (transitive closure) (ie SQL3)
	• RDF-QEL-5: As above + recursive (in language directly).
	
	The P2P architecture is based on JXTA.
	   The replication is primarily replication of metadata (not data).
	   The mapping refers to mapping of vocabularies.
	   The mediation refers to joining of metadata. There is a separate engine called AMOS-II for this, which uses a common data model ECDM.
	   The annotation principle refers to the idea that people can 'annotate anywhere'.

Application status	Ongoing
Category	☞Metadata for discovery and selection
Reference	Wolfgang Nejdl, Boris Wolf Changtao Qu Stefan Decker Michael Sintek Ambjörn Naeve Mikael Nilsson Matthias Palmér Tore Risch. EDUTELLA: A P2P Networking Infrastructure Based on RDF. WWW2002; Honolulu, Hawaii, USA. ACM 1-58113-449-5/02/0005.

Event tracking

Summary	Unified feeds of event information, generic or specialist (c.f. ITTalks)
Description	One particular form of metadata that can be shared over the semantic web is information on future events of interest to some community - concerts, conferences, paper deadlines,

meetings, presentations. Event information is interesting in being of modest semantic complexity and yet giving useful opportunities for automated processing - reminder of a submission deadline for a relevant conference, checking for conflicting appointments, linking to travel planning service.

Outside of the semantic web there is much work on event sharing and tracking. For example, there is an ☞RSS event module and IETF standards for calendaring information (iCalendar - RFC2445, with a IETF draft XML serialization ☞xCalendar). Apart from using RDF as a composable data format for event data the semantic web could provide the basis for categorization of events (ontologies of event types, of topics, of related events like submission deadlines) that could support richer automated processing.

Within the SWAD-E project there is activity looking at the value RDF brings to this area, see ☞calendaring workshop.

Application status	active
Category	☞Personal information management ☞Syndication Category
Link	☞http://www.w3.org/2001/sw/Europe/events/200210-cal/
Related example	☞ITTalks

Financial Assistant

Summary	Advise an individual on investment decisions and financial management options.
Description	If all the information on an individual's finances were available in common form (see ☞Financial Portals) then a personal agent could use this, along with global finance information and rule modules to offer advice on personal financial management.
Comments	This is a classic, old fashioned, expert system style of application. The semantic web would help enable it by providing common global formats for access to the required raw data. In future the rules/logic layer of the semantic web might allow exchange of advisor rule sets. The difficulties are also just the same as such expert system solutions - correctness of rule sets, maintenance given the fast pace of change of legal constraints and best practice, liability, the desire of the experts to maintain their own income streams ...
Application status	proposed
Category	☞Data dependent agents
Related example	☞Financial Portals

Financial Portals

Summary	Aggregate personal financial information into one place
Description	A typical (affluent, western) individual's finances are often spread across several organizations (banks, stock investments, employer etc). Each of these organizations will often provide web-based access to the individual's account but this access is different for each organization. A financial portal harvests all the information relevant to an individual from all the relevant online sources and integrates them into a single view to simplify financial management. Were each of the financial institutions to adopt a common access protocol and format such portals would be more customizable and more effective. RDF is a vendor-neutral candidate for the common format with its support for aggregation of information items and has already been used in at least one product - ☞iSoco's GetSee (believed to be the leading Spanish Financial Portal).
Comments	One difficulty is that the financial institutions have have no interest in supporting such a common access mechanism, it weakens their customer relationship while conferring no advantages directly to them. This means that current solutions tend to work by web scraping and so the particular format used is neither particularly relevant nor visible. Yodlee, for example, is a success financial portal which (as far as we are aware) does not use RDF and it is hard to see how switching of to RDF would benefit them.
Application status	existing
Category	☞Data integration
Reference	☞http://www.isoco.com/en/content/solutions/solution_getsee.html ☞http://www.yodlee.com/
Related example	☞Financial Assistant

Gene Ontology

Summary	Semantic markup of gene data - ontology creation, community annotation and semantic query
Description	There is currently a great deal of interest in bioinformatics; essentially, management information systems for molecular biology. There is a problem both with the vast amount of data and also with the distributed and diverse nature of the information. Most of the data is gene (or protein) sequence information, some of it is annotated. Portals exist, some for retrieval of sequences, some for annotations. The most relevant area for the semantic web is applications which deal with these annotations - manage them, use them for retrieval and for 'semantic matching' (as opposed to sequence matching) There are also a whole range of ontologies available, the most public being the Gene Ontology (and hence there is work in ontology management).
Technical issues	- sheer scale (and diverse nature of) data - decentralized management of annotations - ontology management
User value	A general observation is that shared proteins have similar effects in different organisms. Therefore, information (perhaps marked up as annotations) on genes in 'model' organisms can be really useful for hypothesizing about genes in higher organisms, including humans. Annotations, particularly marked up in a controlled vocabulary (or better, grounded in an ontology) provide a means of doing semantic matching (as well as sequence matching). Related values (perhaps not so relevant to Semantic We) include prediction of structure/chemical interactions and sequence classification
Application status	Application area, with a variety of examples in varying stages of maturity.
Category	☞Data integration ☞Metadata for annotation and enrichment ☞Ontology Management
More information	☞http://www.geneontology.org/
Reference	NM Luscombe; D Greenbaum, and M Gerstein . What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med. 2001; 40:346-58. ☞CiteSeer The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. . Nature Genetics. 2000; 25:25-29. ☞PDF R. Stevens, P. Baker S. Bechhofer G. Ng A. Jacoby N. W. Paton C. A. Goble and A. Brass. TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources. Bioinformatics. 2000; 16(2):184-186. P.W.Lord, J. R. Reich A. Mitchell R. D. Stevens and T. K. Attwood and C. A. Goble. PRECIS: An Automated Pipeline for Producing Concise Reports About Proteins. IEEE International Symposium on Bio-informatics and Biomedical engineering: IEEE press; 2001 N ov: 59-64. Dowell, R. A Distributed Annotation System [Tech Report #01-07]: Washington University in St Louis; 2001 Feb A master's project presented to the Department of Computer Science.
Specific example	☞Sequence Retrieval System, a portal onto a number of sources for sequence data. ☞Entrez, a portal which provides related genes (related by annotation, where text is matched up using a taxonomy, MESH) The ☞TAMBIS project, a portal which allows conceptual queries to be mapped onto a common programming language for access to disparate data sources. PRECIS, a pipeline for annotations (ie gathering, filtering and grouping annotations to present a single, consistent view of annotations harvested from a variety of sources There are lots of gene ontologies to choose from, including: • Gene Ontology: From the ☞Gene Ontology Consortium. • The TaO: An ontology developed as part of the ☞TAMBIS project at Manchester (UK), using description logic (and the OIL editor). • Lots of other ontologies to choose from - Riboweb, EcoCyc, Schulze-Kremer

Genealogy assistant

Summary	Semistructured data approach to annotation of Genealogy information, incremental construction of relationship graphs
Description	Searching for and organizing family history information is a very popular activity, purported to be the second most common consumer use of the web. Many personal genealogy database tools exist but all of them impose a fairly rigid picture for how the information is to be recorded. In particular, they place the family tree first and link the

discovered items directly into that tree. In contrast in many situations it would be preferable to collect the items of evidence in a semi-structured pool and over time link them to the deductions that they support. RDF would provide a good representation for such semi-structured, incrementally organized, data. In particular, its direct support for recording of provenance as well as the data itself (through use of reification) would be valuable.

Application status	proposed
Category	☞Personal information management

HP Portal

Summary	Use RDF for metadata on intranet resources and on user roles to automate intranet portal customization
Description	The HP intranet portal, like many employee portals in large organizations, collects data from many different sources. A single publishing system integrates the output from several content management systems and has to deliver views of the information customized to the role and individual preferences of a wide variety of users. Semantic web technologies could play several roles in this. Firstly, RDF offers a vendor-neutral format for the metadata to be managed by the publishing system - giving a common format into which the different proprietary content management system metadata can be translated. Secondly, the roles and requirements of different user communities and how they map to the different document categories could be explicitly represented by a domain ontology rather than indirectly represented by rules or scripts - thus easing maintenance and extensibility.
Application status	proposal
Category	☞Semantic indexing ☞Metadata for discovery and selection
Related example	☞Sun GKE

Haystack

Summary	Example of personal information management with a sophisticated semantic-driven UI
Description	Haystack seeks to offer a complete management environment for personal information. The Haystack work represents all structured information in a common semi-structured data format (RDF). They have a rich user interface that is driven from both the RDF data itself and from ontology or schema files defining the data models. In Haystack the type of the item being displayed drives the rendering of items and the mapping from category to view is declaratively defined. Haystack supports traditional personal information management (PIM) tasks, such as calendar and task management, as well as snippet and document discovery. Haystack views present UI elements that are mapped to heterogeneous data objects in what they call the Semantic User Interface. They associate views with individual pieces of information, and attempt to keep these views synchronised with changes to the underlying data. A view 'selector' (aka factory) is used to select a view (or views) for a data type. "Stacker" (item collection) views are used to present collections of items, although currently only list views appear to be supported.
Technical issues	The heavy use of RDF for configuration (Haystack has ~1 Million RDF statements concerned with UI state) has obliged the Haystack team to build a special purpose RDF database
Comments	- The Haystack project is based at MIT and is led by Prof. David Karger - A primary focus for Haystack is the "semantic UI" approach to information management tools. - RDF is used extensively for configuration; for example, ~45,000 queries are required to generate a non-trivial UI
Application status	Active.
Category	☞Personal information management
More information	☞http://haystack.lcs.mit.edu/
Reference	David Huynh, David Karger, and Dennis Quan. Haystack: A Platform for Creating, Organizing and Visualizing Information Using RDF. Draft submission to journal 2002. ☞PDF
Related example	☞ePerson

Helpdesk support

Summary	Example of knowledge management, indexing of past cases, product information and defect tracking to support helpdesk staff
Description	Another example of knowledge management. In solving customer problems helpdesk staff need access to well indexed databases of past case solutions, diagnosis advice and product information. In some situations the helpdesk staff themselves may add new case solutions to the database, in others a centralized team will create and maintain the helpdesk knowledge base. The primary semantic web related aspect to this is the use of domain ontologies to give a richer index and structuring of the case information. However, this usage is primarily within a single system and it is less clear if this is just use of ontology technology or whether there is a significant web aspect.
Application status	active
Category	☞Knowledge management

ITTalks

Summary	Web portal for IT seminars; emphasizing agents, personalization and DAML+OIL
Description	IT talks is a web portal for talks, seminars and colloquia, It uses DAML+OIL for both personal profiles and agent communication. The idea is that you can find talks that match your interest, availability and geographical location. You can also personalize your page ("My ITTalks"), subscribe to an RSS feed, and create your own portal. Future scenarios include agents which update your calendar, gather additional information about the selected talk and participate in collective decision making (eg who in your department should go?)
Technical issues	There is a potential issue around ontology definition and metadata creation. To deal with the first, the team have decided on a centralised ontology to represent talks, primarily a topic hierarchy based an the ACM CCS system. Of course, other attributes such as location, time and title are also defined. The second problem here of course is that not all talks may have this information. Some can be extracted automatically, using an information extraction technique (ITtalks use Lockheed Martin's AeroText system). A talk author can also be assisted in marking up a talk entry (they use a standard bag of words classifier to suggest ACM topic based on the abstract). Some of the more speculative scenarios push on quite deep research issues in agent communication and collective decision making.
Comments	ITTalks is a collaboration between many researchers, primarily at UMBC (Tim Finin, hence the emphasis on agents).
Application status	Existing (but unclear how mature or widely used).
Category	☞Data dependent agents ☞Semantic indexing
Reference	Cost, R. S.; Finin, T.; Joshi, A.; Yun Peng; Nicholas, C.; Soboroff, I.; Chen, H.; Kagal, L.; Perich, F.; Youyong Zou, and Tolia, S. ITtalks: a case study in the Semantic Web and DAML+OIL. IEEE Intelligent Systems . 2002; 17 (1):40-47.

Ideas workbench

Summary	Support for incremental capture, structuring and linking of the pieces of information collected in early stages of an information rich activity
Description	In the early stages of a creative project (for example, a research project, development of a documentary film, investigation into an new application space) it is common to rapidly accumulate a lot of references and facts which need capturing. The challenge is that at this stage in the project the ideal structure for indexing and organizing these facts, documents, and media objects is unclear. Even the types of annotation needed is unclear. One could imagine an ideas workbench which would exploit the free-form, semi-structured data, capabilities of RDF to allow immediate capture of such data and then support incremental construction of organization and indexes to help the the user built a view of the domain. This would include tools for discovering existing relationships and classification schemes that might be relevant, ability to incrementally add relationships between ideas, tools for generating and exploring different clusterings and views onto the collected data. At later stages in the project the collected data may be used as part of a more conventional content management system, bibliography database or community portal.

	However, in the early stages existing content management systems are far too rigid to offer any support.
Technical issues	- ontology evolution - clustering and concept formation
Application status	proposal
Category	☞Knowledge formation ☞Personal information management
Reference	Similar issues were mentioned during a talk on the creation of the BBC TV series *Walking with Beasts* in the session "Re-pursing old content for new and future media" at IBC 2002, Amsterdam, 15th September 2002.
Related example	☞Assumption tracker

Jema

Summary	An example of applying semantic web to workflow problems targeted at W3C process support
Description	Jema is an RDF application developed by Brian McBride to support W3C working group process. It provides representations for issues lists and actions to support tracking of progress and creation and tracking of agenda items. It links to both email lists and IRC chat and can be consulted during a weekly teleconference to assist with agenda management.
Comments	Data being managed is similar to that within Personal Information Management applications, but this example has a more active workflow element and some degree of data integration.
Application status	existing
Category	☞Personal information management ☞Data dependent agents
Reference	Demonstrated at W3C Technical Plenary, April, 2002.

KAON

Summary	Karlsruhe ontology management tool (aka Ontobroker)
Description	Kaon covers a number of components as follows: ● KAON API: programmatic access to KAON ontologies and instances ● RDF API: Provides programmatic access to RDF models. ● ENGINEERING SERVER A storage mechanism for KAON ontologies ● RDF SERVERProvides persistent storage of RDF models in relational databases, ● KAON PORTAL: A simple tool for generating multi-lingual, ontology-based Web portals. ● OI-Modeler: a tool for ontology creation and evolution. ● TEXT-TO-ONTO: Supports semi-automatic creation of ontologies by applying text mining algorithms. Currently includes term extraction algorithm, concept association extraction algorithm and ontology pruning algorithm. ● RDF CRAWLER: A simple tool for syndication of RDF-based content from a set of Internet pages. Kaon adopts a slightly different modelling approach to UML, Flogic, OIL and RQL. In the Kaon OI model, an entity may be viewed as a context or instance in a context-dependent manner. The OI model is a 'user' level mapping (ie corresponds to the model the user has in their head), as opposed to the logical (ER) model and the physical (RDB) model. In the latter (physical) model, the scheme varies slightly from RQL (where one table is assigned per concept) for efficiency (table creation is very expensive, and the total data stored is not expected to be large). It is also different to UML (it doesn't have methods), F-Logic (it is more tractable, especially wrt axioms) and description logics, including OIL (it is more intuitive).
Technical issues	Ontology mapping is required (which is tackled using OI-modelling). The implementation of OI-model in Kaon requires editing, storage and DB Mapping.
Comments	Ontology management, with many components - eg visualisation, text-to-onto (extraction of metadata), support for ontology refinement/validation etc.
Application status	Existing application
Category	☞Ontology Management
More information	☞http://kaon.semanticweb.org/

Reference	B. Motik, A. Maedche and R. Volz. A Conceptual Modeling Approach for building semantics-driven enterprise applications. Proceedings of the First International Conference on Ontologies, Databases and Application of Semantics (ODBASE-2002); California, USA. Springer; 2002. L. Stojanovic, A. Maedche, B. Motik, and N. Stojanovic, "User Driven Ontology Evolution Management," Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management EKAW, Madrid, Spain, 2002.
Related example	Any of the ontology management tools. (Ontoedit, OILEdit) The metadata extraction is a common theme coming up in other applications (eg SCORE, ITTalks)

MIT/HP SIMILE project

Summary	Semantic web enhanced digital library
Description	The Simile project is tasked with augmenting the DSpace digital library system with to support richer and more flexible metadata. DSpace, like many digital library systems, has one primary data model (how items are grouped into collections) and one primary metadata model (Dublin Core). By using semantic web approaches to the metadata it should be possible to support more diverse metadata formats. In particular, to allow individual user communities to create their own augmented formats the support their specific needs.
Technical issues	The technical issue include: ● support multiple metadata format simultaneously ● support metadata interoperability (format mapping, provenance issues) ● use metadata to control dissemination policies ● distribution issues (performance if distributed, control if not)
Application status	Ongoing project, started 2002
Category	☞Metadata for annotation and enrichment ☞Metadata for discovery and selection ☞Metadata for media and content
Related example	☞Museum portals ☞Catalogue Management ☞ClaiMaker/Scholonto ☞Scholnet

MUSE

Summary	Business selling metadata on media - music, films
Description	Muse UK is a commercial provider of metadata for media (music, film). It provides (or "provided", current status unclear) comprehensive and clean metadata which was often more accurate than that of the media suppliers. We believe that the metadata database may have been in a form of RDF.
Application status	unclear
Category	☞Metadata for discovery and selection
Reference	For example, minor reference in ☞http://www.wipo.org/eng/meetings/1999/acmc/2_1-04.htm

Mozilla

Summary	Uses RDF within browser to represent all content - email, bookmarks etc
Description	The Mozilla browser uses RDF internally as a common format to represent the many different "data sources" that can be viewed through the browser. In particular, it is used to represent bookmarks, history, search results, file systems, ftp, sitemaps, email headers etc. A common set of tools can then be used to collect and merge this RDF and use it to drive flexible UI displays - for example, the Mozilla side bar. Having constructed this general RDF backend it can then to used to access and display metadata from external sources. For example, Mozilla documents mention a SmartBrowsing facility that aggregates data from trusted-third party metadata sources to provide page annotations - though the status of that particular example is unclear.
Application status	existing
Category	☞Data integration

	☞Personal information management
Link	☞http://www.mozilla.org/rdf/doc/
More information	☞http://books.mozdev.org/chapters/ch10.html

Museum portals

Summary	Semantic indexing, provenance and other metadata for tracking and describing museum artifacts
Description	Museums have a strong need for detailed metadata describing the artifacts they hold. Apart from internal uses of tracking and audit this metadata can be exploited to provide rich museum web sites (a key part of the museum's role of public education and access) and to augment the live exhibits (when electronic presentation means are available). Such metadata can also be used to link exhibit information to other Museum activities such as shops and online commerce. There is significant interest and activity in using the semantic web in this area.
Comments	This domain has much higher requirements for provenance tracking. There is the need to cope with the fact that the artifacts themselves change (e.g. being moved or renovated) and the descriptions of artifacts change (a new appraiser claims it is not a genuine Van Gogh after all). The ABC/Harmony model provides one solution to that. There is some suggestion that the webness of semantic web could be relevant here. For example, linking information on the same artifact/creator/topic from many sources - the AMICO case study mentioned in Eric Miller's talk (see below) is an example.
Application status	active
Category	☞Semantic indexing
More information	☞http://www.cimi.org/ ☞http://metadata.net/harmony/
Reference	*Weaving Meaning: the W3C's Semantic Web Initiatives* Eric Miller, Museums and the Web 2022. ☞http://www.archimuse.com/mw2002/abstracts/prg_175000640.html
Related example	☞Community portals

MusicBrainz

Summary	Community generation of music metadata - open std variant on CDDB
Description	This project provides a set of tools for a community to annotate music files (for example, by title and artist). The idea is to provide a similar functionality to CDDB, but to address its drawbacks. Firstly, the database is open (and therefore cannot be taken into private ownership). Secondly, the annotations are public (and can therefore be easily corrected).
Technical issues	With shared ownership of metadata, how to distinguish between a simple 'correction' and a genuine 'disagreement'. Provenancing may be an issue.
Application status	Existing application
Category	☞Semantic indexing
Reference	Swartz, A. MusicBrainz: a semantic Web service . IEEE Intelligent Systems. 2002; 17 (1):76-77.

OntoShare - community of practice support

Summary	Tool to assist an active community to exchange problem case materials and solutions - case study from OnToKnowledge project
Description	A core problem in knowledge management is how to support *communities of practice* to more effectively share and exchange knowledge. Many groups are exploring ways the semantic web could help with this by providing an explicit representation of the ontology of the community's domain of interest and annotating the documents to be shared with appropriate classification metadata. OntoShare is a tool developed by BTexact Technologies as part of the OnToKnowledge project. Practitioners use the tool to add documents to a shared community store along with manual annotations, semi-autmatic classifications into the community ontology and an automatic summarization. The tools use this metadata to recommend the new document to other community members based on their interest profiles.
Technical issues	- automated support for ontology evolution - automated classification against an ontology

Comments	A particularly interesting aspect of the tool is that if the user overrides the automated classification of the document that information is then used to help evolve the ontology.
Application status	active
Category	☞Knowledge management
Reference	*OntoShare: Using Ontologies for Knowledge Sharing*, John Davies, Alistair Duke, Audrius Stonkus, International Workshop on the Semantic Web, Hawaii, May, 2002.
Related example	☞Community formation

PatMan

Summary	Knowledge management in the medication domain using ontologies for both organizational and medical knowledge
Description	One of several projects which have applied Knowledge Management and workflow principles to support clinical practice. A key feature of this project was to not just encode clinical guidelines but to adapt them to context by exploiting organizational information. Both the clinical guidelines and organizational information were encoded using ontologies.
Comments	This is just one example of many knowledge management projects that exploit ontologies for indexing and structuring information, along with workflow support (using Petri nets in this case). No direct connection to the semantic web other than to reinforce that the use of ontologies to encode semantics in support of retrieval and knowledge management is not particular to the semantic web.
Application status	existing
Category	☞Knowledge management
Link	☞http://aim.unipv.it/projects/patman/

PlanetOnto

Summary	Application of ontology indexing and enrichment of documents to news stories
Description	Web-based news portal that indexes news items related to and of interest to the ☞Knowledge Media Institute (KMi). Combines ontologies describing KMi projects, organizational structure and the nature of academic events and news stories to provide rich indexing, search and navigation of items.
Comments	The KMi have a highly expressive ontology language (OCML) and a web-accessible database of linked ontologies. They have developed a large number of knowledge management applications which combine ontologies for representation of domain and organizational knowledge with automated classification, together with tools for structured discourse, workflow and other activities relevant to knowledge management.
Application status	existing
Category	☞Knowledge management ☞Semantic indexing
Link	☞http://kmi.open.ac.uk/projects/planetonto/
Related example	☞PatMan

Recommendation Networks

Summary	Infrastructure for sharing opinions, ratings and other recommendation information with support for reputation of the recommenders
Description	When choosing a product or service then direct recommendation from people who have tried that product or service is valuable. When this comes from friends and colleagues it is particularly valuable and trusted but, unless you have a very large social network, there will not be advice available for every product. Recommendation Networks are a generic term for mechanisms by which individuals can contribute opinions, experiences and recommendations into a pool so that other individuals can then exploit this collective experiences. The semantic web could provide an infrastructure for such recommendation networks. Exploiting URI's for representation of the products or services (an issue in itself), RDF as the representation of the product annotations and ontologies both for the types of annotation (e.g. review-based-on-direct-experience) and the nature of the product/service (to support search and selective subscription).

Technical issues	- reputation management for assessment of information quality - unambiguous identifiers for products and services
Comments	This application could be very valuable but does suffer from some non-technical challenges: - How to gain critical mass? Once data on a sufficient range of products/services is usefully available this way the service becomes valuable and attracts contributions but the initial bootstrapping is hard. - Some existing commercial sites (notably Amazon) do this for their topic area and gain value from it (customer loyalty, improved sales). The benefit to them of opening up this data in such a way that it could be exploited by competitors is not clear. - Much of the information is aimed at people and not machines and so is largely text based and lacking in structured metadata.
Application status	proposed
Category	☞Metadata for discovery and selection
Reference	☞London Open Network ☞Amazon customer reviews ☞http://www.epinions.com/
Related example	☞Semantic tagging

Rich Site Summary/RDF Site summary

Summary	RSS is an XML format for syndicating content. It provides facilities for listing channels of syndicated information, and items within that channel. The items may describe meta-data, content data or both.
Description	RSS is a family of XML specifications for encoding syndicated information flows. The two versions of RSS currently in circulation are RSS 1.0 and RSS 2.0. They are not directly related to each other, though they descend from a common root. Nor are they compatible (they use, for example, different root elements for the XML document). Efforts are underway, however, to help both formats to co-exist reasonably peacefully. RSS 1.0 is an RDF application. RSS 2.0 is a pure XML application. Hence the acronym is expanded to *Rich Site Summary*, or *RDF Site Summary* among other variations. Both RSS 1.0 and 2.0 are primarily designed to be lightweight publication formats. The choice of RDF properties or XML elements is chosen to represent that goal. RSS 1.0, however, has an explicit concept of *modules*, that plug-in to the standard to provide for more representation that is part of the core. For example, there is no provenance meta-data in RSS 1.0, but there is a Dublin Core module defined as a standard extension. Conceptually, RSS defines channels of information, along which individual items flow. An item may contain a link to another web resource, plus various metadata and optionally content. There are no formal relationships between channels. However, it is common for blogs to be partitioned into categories, where an item might be part of multiple categories and form part of that site's overall flow.
Technical issues	The main technical issue for building on RSS will be to move beyond the lightweight publication flavour of the standards, in a way that doesn't break the existing aggregators and other RSS readers (if we want to be able to re-use these tools). Even so, new tools will be needed that support a more semantic-web flavour to RSS channels and items.
User value	For RSS-classic, the key user values have been easy web-site updating (i.e. lightweight publication) and the use of aggregators to give the appearance of push channels.
Category	☞Syndication Category
More information	☞RSS 1.0 spec ☞RSS 2.0 spec ☞Syndication disc ussion group
Related example	☞Syndication

SWAP - semantic web and peer-to-peer

Summary	EU-IST project developing a p2p solution for knowledge management supporting implicit and emergent ontologies
Description	SWAP is a EU funded project, investigating the application of peer-to-peer computing, combined with semantic web technology, to the knowledge management problem. The vision is very similar to that of ePerson, including emphasis on decentralised knowledge and ontology creation, distributed knowledge finding, and emergent semantics.

	Partners include University of Karlsruhe, VU Amsterdam, Meta4 (Spain), Empolis (UK and Poland), Dresdner Bank (Germany) and IBIT (Spain).
Technical issues	+ using semantic web techniques to improve querying (e.g. resolving synonyms and homonyms, query narrowing and widening) + sharing knowledge at less than file granularity + reconciling indepen dently developed ontologies, supporting "self-updating" ontologies + issues arising from multi-lingual knowledge bases
User value	+ assistance with knowledge management tasks: greater efficiency by locating work-relevant knowledge in the organisation with less effort + p2p, so no dependency on a central server
Category	☞Knowledge formation ☞Ontology Management
More information	☞http://swap.semanticweb.org/public/index_html.htm
Reference	☞http://swap.semanticweb.org/public/publicat.htm/Publications/SWAPpresentation.pdf

Scholnet

Summary	Digital library supporting communication and collaboration within networked scholarly communities
Description	Scholnet is an EU R&D project developing a digital library infrastructure to support the communication and the collaboration within networked scholarly communities. The digital library will provide traditional digital library services in addition to support for non-textual data types, hypermedia annotation, cross-language search and retrieval, and personalized information dissemination. This testbed will be used t o demonstrate how an enhanced digital library can enable members of a networked scholarly community to learn from, contribute to, and collectively build upon the community's discipline-oriented digital collections. It will be built as an extension of the ERCIM Technical Reference Digital Library (ETRDL) The intention is that the tool provides support for community annotation of digital media. Cross-lingual support and notification mechanisms are planned.
Technical issues	Various issues around the support of annotations (distributed, multiply provenanced, on part of a digital media item, access control issues)
Comments	Scholnet runs from 1 November 2000 to 30 April 2002. It is supported by the IST Programme of the European Union (project no. IST-1999-20664). The Scientific coordinator is IEI-CNR; administrative coordinator ERCIM. The members are: ● CNR (Scientific Coordinator, Italy) ● INRIA (France) ● GMD (Germany) ● FORTH (Greece) ● SICS (Sweden) ● Lorasi (France) ● Univ. of Masaryk (Czech Republic) ● ERCIM (Administrative Coordinator, France) A prototype was planned for mid-2002. Not clear whether a public version is/will be available (no link available from web page as of Nov 2002).
Application status	Unclear. Project finished.
Category	☞Metadata for discovery and selection
More information	☞http://www.ercim.org/scholnet/
Related example	☞Arkive internal

Score

Summary	Voquette's Score product
Description	The SCORE product is a commercial semantic web product. It : ● extracts metadata (from a variety of sources, web, database, XML feeds), ● fills in missing fields, classifies (by standard bagging/boosting methods), ● resolves ambiguities and presents to user. The components include the use of metadata to describe syntactic (eg language/format)

	and semantic (domain-specific) information about a document. The normalization of ontologies involves merging concepts. Semantic search provides both disambiguation of words and a useful associative element (one gets related content not explicitly asked for).
Technical issues	The usual issues about ontology management are relevant here.
Comments	Voquette is the company. Comes from the University of Georgia.
Application status	Commercial Application
Category	☞Semantic indexing
Reference	Sheth, A.; Bertram, C.; Avant, D.; Hammond, B.; Kochut, K., and Warke, Y. Managing semantic content for the Web . IEEE Internet Computing. 2002; 6(4):80-87.

SeLeNe

Summary	EU-IST project on self e-learning networks - using RDF for metadata on educational resources with personalized views
Description	SeLeNe is a EU-IST project developing technology to support distributed repositories of educational resources. The key emphasis is on using semantic web techniques for dynamic integration of metadata from heterogeneous and autonomous educational resources, and for creating personalised views over this "Knowledge Grid". It is a one year feasibility study.
Technical issues	- semantic reconciliation of metadata standards - use of schema information to support automatic generation of metadata from learning resources - reformulation of unstructured queries into structured queries - change propagation from source learning objects to derived learning objects
Application status	research investigation
Category	☞Metadata for discovery and selection
Link	☞http://www.dcs.bbk.ac.uk/~ap/projects/selene/homepage.html
Related example	☞Edutella

Semantic tagging

Summary	Semantic tagging (product, opinions, ratings) for purchase support (Gartner)
Description	Semantic "tags" give information on products and services, both descriptive and qualitative (opinions, experiences) which help purchasers find and select items and help providers make better recommendations based on user interest profiles. Gartner predict this will affect $90b b2c and $350b b2b transactions by 2008 and create a new billion dollar industry to collect/organize/sell these tags 2008-20011 [0.6 probability]. This is a generic category rather than a single specific application but is more tightly defined that the categories like "metadata for description".
Category	☞Semantic indexing ☞Metadata for discovery and selection
Reference	*Trend Watch: Four key trends for 2002-2012* Gartner Group, Gartner Symposium ITXPO, San Diego, April, 2002

Shopping assistants

Summary	Personal agents to support or automate purchases, able to help with product location, price comparison and access to ratings and opinions
Description	The process of purchasing products or services across the web comprises many stages each with its own challenges. Firstly, there is finding possible suppliers for the product which involves matching your requirement or description of the product against the way the products are described by the suppliers. Secondly, there is gathering and comparing information on price, availability (and other factors like Terms & Conditions) across many suppliers. Thirdly, there is the desire to combine this with the opinions, recommendations and experiences of other purchasers of similar products or users of the same suppliers in order to make a decision. Finally, there is the mechanics of making the purchase, and tracking delivery. The last stage is already well-enough handled on the web but the first few phases of information gathering, assimilation, decision making often leave a lot to be desired. They are well handled in specific product areas (e.g. books) but not in less mass-market areas

(try buying a non-standard size of "rat cage" for example) and often the best recommendation information is scattered across many specialist bulletin boards and discussion groups. There have been a large number of shopping bots developed (see ☞botspot). However, these are usually forced to use either web-scraping or proprietary protocols to aggregate price information which results in them only existing for highly popular products and even then coverage can be poor or uneven.

The notion of the semantic web shopping assistant is to combine agent technology for seeking out and organizing the relevant information with the semantic web as an infrastructure over which the product, pricing, features, availability and opinion information could be made uniformly available.

Comments	As in ☞Recommendation Networks the challenges here are primarily economic and social rather than technical. How to reach an initial critical mass? What incentive do suppliers and existing mediaries have to participate in such an open network? How to develop effective uniform names for the products and services? Despite these challenges this is a genuine data integration problem of relevance to many consumers and is one of the few semantic web applications that is is reasonably compelling to non-specialists.
Application status	proposed
Category	☞Data dependent agents
Link	☞http://www.botspot.com/search/s-shop.htm
Related example	☞Recommendation Networks ☞Semantic tagging

Sun GKE

Summary	Metadata and vocabularies for digital assets (intranet application)
Description	Management and distribution of corporate digital assets by consistent use of Dublin Core and specialized RDF schemas. Developed by SUN Global Knowledge Engineering group for intranet use.
Application status	active
Category	☞Metadata for discovery and selection ☞Knowledge management
Link	☞http://www.w3.org/Talks/2002/10/16-sw/slide16-0.html
Related example	☞HP Portal

Syndication

Summary	Metadata for syndicating content, aggregate and distributing topic specific feeds, RSS
Description	Syndication is the general name for the re-use of content from one originator, in a publication by another publisher. In the (dead-tree) print media industry, and on television networks, syndication is a high dollar-value activity. The name has been adopted on the WWW to refer to any re-use of content from one site by another. Sometimes these are paid-for transactions, but more often syndication is a means of flowing small items of interesting comment around the web. This lighter-weight content flow is facilitated by the use of web standards such as HTTP for transmission, and lightweight XML standards for encoding the syndicated content. The primary encoding for such kinds of lightweight publishing and syndication is RSS (Really Simple Syndication, Rich Site Summary or RDF Site Summary). Various incompatible versions of the RSS standard exist, and its history, though short, is filled with contention and disagreement. Currently, the only version of RSS that is RDF-based (or even RDF compatible) is RSS 1.0. Note that RSS specifically depends on the XML serialisation of RDF; various RSS processors exist that treat the RDF XML syntax, and in particular the RSS schema, as their target grammar. A specific use of RSS that has grown in popularity over recent years is *blogging* (from "web-logging"). Blogging is the publication of online chronologically-organised journal entries, using tools that emphasise ease of content generation for the non-expert. As a result, a huge number of diverse blogs exist on every imaginable subject. Some are of very high quality. The phenomenon of *blog-rolling* arises from the referencing of one blog from another, allowing syndicated content to flow around a complex network. The blogger network is increasingly supported by underlying protocols such as the Blogger API, and SOAP.
Technical issues	As currently practiced, syndication is a solved problem. Interesting possibilities arise, however, if we imagine extending syndication to structured, semantically rich objects, then the key technical issues are: discovery and monitoring of relevant content in a

highly-dynamic decentralised network, and reconciling independently generated classifications of content (i.e. ontology matching again). Further research will also be needed to understand the best ways to present semantically-rich structures to users in a variety of presentation contexts.

User value	+ lightweight ways of publishing semantically marked-up content for re-use + discovery and re-use of knowledge from a decentralised, dynamic community + support for building meta-networks, suc h as shared-interest communities
Category	☞Syndication Category
More information	☞Syndication discussion group
Related example	☞http://www.w3.org/TR/SOAP/ ☞http://plant.blogger.com/api/index.html
Specific example	☞Rich Site Summary/RDF Site summary

TAP semantic search

Summary	Augment search (internet or specific portal) by searching on inferred semantics of the search terms

Description	The TAP project offers a vision for how the semantic web could function and then offers an example type of application. It is the latter that is most relevant here.
	The philosophy is that the semantic web is a about structured objects (people, places and events) and not about text strings. So that searches can use semantic terms and the things that come back can have more structure (e.g. a search for a CD could pull back information on how to buy the CD, pointers to eBay auctions, calendar details of upcoming gigs etc). The notion of semantic search or activity based search is to infer from a normal text search what the underlying user activity is and what the semantic terms involved are. Thus a free text search on "Yo Yo Ma", as well as pulling back normal Google results, would realize "Yo Yo Ma" probably refers to the classical musician and thus would also search sources like Amazon, TicketMaster etc for relevant information.
	    This "would realize ..." bit above is implemented by having a knowledge base of concepts and looking up the text terms in that knowledge base to see if they appear as, say, the dc:title of some concept.
	    At the technical infrastructure level then TAP offers a simple standard way of doing remote access to semantic web data (the GetData interface), a notion of address by reference to circumvent the problem of universal naming (similar mechanics to the ePerson QBE approach), a caching architecture, a server implementation (Apache mod) which supports aggregation, a client library and an initial knowledge base.

Application status	active
Category	☞Semantic indexing
Link	☞http://tap.stanford.edu/tap/ss.html
Related example	☞Semantic tagging

Thesaurus management

Summary	Representation and management of rich taxonomic schemes and controlled vocabularies to support evolution and integration
Description	Thesauri have been an important component of online database searching within the library community for many years and are now considered useful for the online Web-based search community as well. As a simple form of ontology, they play an important role in the indexing of Web-based documents, adding a certain amount of semantic information. There has been substantial work on RDF Thesauri, for example the DESIRE project defined a standard set of conceptual relationships typical of controlled vocabularies such as thesauri, classification systems and organised metadata collections; this set of relationships was encoded into RDF and used in the SOSIG and LIMBER projects. The challenges at this stage are to show that RDF is a useful encoding for thesauri, and to show how to migrate existing thesauri to RDF.
Technical issues	- relationship to semantic web ontologies - multi-linguality - ISO-compatibility - term cross-mapping
Application status	active
Category	☞Catalogue/thesaurus management
Reference	Already an application with the SWAD-E project (from which this description is

	derived). In particular ☞work package 8.
Related example	☞Catalogue Management

Virtual Travel Agent

Summary	Travel planning agent which functions by composition of component services
Description	Another example problem in which data from multiple sources can be combined, with some planning and agent technology, to deliver a useful service is travel planning. Organizing travel typically requires making a set of bookings with different providers with different constraints and complex interactions (e.g. hotel availability may affect location, which may change car hire and flight arrangements). Human travel agents perform this task but there has been a long standing interest in the agent (especially the multi-agent) community on applying automated constraint satisfaction and planning techniques to create a virtual travel agent. This might have advantages over existing solutions of being able check across a much wider range of vendors. This is also an oft-cited example for composition of web services where the individual providers expose web services which either a user agent or an intermediary then composes to create a virtual travel agent service. The sort of information to be exchanged in making such a service practical is well-suited to semantic web representation.
Technical issues	- distributed constraint satisfaction - planning - representation of extensible qualitative features
Comments	The semantic web could offer some advantages over existing proprietary or web-scraping approaches as the data infrastructure for this application. Firstly, it offers an open vendor-neutral format which might improve the chances of uptake. Secondly, the extensibility of semantic web representations would make it possible for providers to distinguish themselves by marking up their offerings with specialist features which customised user agents could then take into account.
Application status	proposed
Category	☞Data dependent agents
Reference	E.g. *Towards Desktop Personal Travel Agents*, D. Ndumu, J. Collis and H. Nwana, BT Technological Journal 16 (3), 69-78, 1998. E.g. *Task-Structure Based Mediation: The Travel-Planning Assistant Example*. Q. Situ and E. Stroulia. In the Proceedings of the 13th Canadian Conference on Artificial Intelligence, 14-17 May 2000, Montréal, Québec, Canada, 400-410, Vol.1822 Lecture Notes in Computer Science, Springer Verlag.

Web service description and discovery

Summary	Metadata describing the intent and function of web services to allow people or agents to discover sets of services meeting a given need
Description	The web services vision calls for services to be discoverable both by developers and by automated systems performing service mediation. Current web service description approaches such as ☞UDDI given basic information on service endpoints and some categorization support (along with free-text descriptions). Semantic web ontologies could potentially support richer and more extensible descriptions. For example, supporting multiple cross-linked categorization schemes, service parameterization, description of quality-of-service and related "non-functional" features. Furthermore, description logics offer an approach to formulating queries for matching services in terms of description classes which can match a specific service against a more general constraint description.
Application status	proposed
Category	☞Metadata for discovery and selection
More information	☞http://swws.semanticweb.org/
Related example	☞B2B web service mediation ☞B2B trading market-places

ePerson

Summary	ePerson - community information management tools

Description	The HP Labs ePerson activity has developed a set of community information management tools using the semantic web stack. The term *ePerson* refers to an active information store that provides some representation of an individual (their interests, information they wish to share). These stores are connected peer-to-peer so that they can be searched and combined. The aim being to support knowledge management (both community formation and information sharing aspects) but using a decentralized (*person-centric*) approach. The initial test application was a tool for storing, organizing and sharing small items of information (called the *Snippet Manager*). This includes tools for management of web bookmarks (bi-directional access to browser bookmarks, RDF queriable copy of the DMOZ data set, ability to share bookmark classification schemes) but is also able to store arbitrary web-accessible content and associated searchable metadata. RDF is used for all information storage, an RDF query paradigm "query by example" is used for all distributed searches, DAML+OIL is used to represent the schema and vocabularies for both internal machinery and user visible data.
Application status	existing
Category	☞Personal information management ☞Knowledge management
Reference	*The ePerson Snippet Manager: A Semantic Web Application*, HP Labs Technical report, NN-2002. [URL to be supplied].
Related example	☞Haystack

eScience Data Grids

Summary	Support data integration challenges for eScience grids - covers metadata for resource discovery and schema integration for integrated query
Description	eScience refers to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists. The Grid is an architecture proposed to create a reality of such a vision for eScience. It will provide an infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources. This resource includes computational systems and data storage and specialized experimental facilities. Such a proposal has clear requirements for data integration, intelligent query and metadata use that would seem to provide a good fit for semantic web technologies.
Technical issues	Primarily the challenges of working with massive datasets, plus the challenges associated with navigation, query and visulaisation.
Comments	The proposed Grid will act as an enabler for a whole host of projects, from model visualisation to image processing & fluid dynamics. Some projects will require more use of metadata/data integration than others.
User value	Platform enabling a number of data-intensive research projects.
Application status	Infrastructure available, series of workshops ongoing to define and plan actual projects.
Category	☞Data integration
More information	☞http://www.niees.ac.uk/