A Semicolon project deliverable.

Survey on vocabulary and ontology tools

Including a methodology for comparing tools

www.semicolon.no

Version 1.0

Date 2013-09-20

Date first issue

2013-09-20

Report nr

Work verified by

Vibeke Dalberg

Report title

Survey on vocabulary and ontology tools. Including a methodology for comparing tools and corresponding test results.

Summary:

The eGov research project Semicolon (www.semicolon.no) has identified a need to pilot vocabulary / ontology tools with workflow support. This document describes the pilot prerequisites: (i) a survey of relevant pilot tools (candidates), (ii) a method for comparing tools according to defined needs, and (iii) the results of the tool evaluation according to the methodology. The chosen pilot software are: Collibra Business Glossary (by Collibra), Enterprise Vocabulary Net (by Topbraid), Pool party (by Semantic Web Company GmbH) and Semantic MediaWiki (open source).

Distribution:

☒ Open distribution

☐ No distribution/ confidential

☐ No distribution/ confidential

☐ Project internal distribution only

Authors:

Per Myrseth
Chief Specialist in Information Risk Management
Department: Intelligent Networks & Communication
DNVKEMA Norway

Erlend Øverby
Senior Advisor
Karde, Norway

Jim J. Yang
Head of Unit for Classifications,
Department Statistics and Classifications
Norwegian Directorate of Health.

Key words

INDEX

1 Introduction

1.1 Background

The eGov research project Semicolon (www.semicolon.no) has identified a need to pilot vocabulary / ontology tools with workflow support. By running a pilot we will build knowledge of opportunities and obstacles related to eGov service engineering and governance. Prerequisites for running such a pilot are (i) a survey of relevant pilot tools (candidates), (ii) a method for comparing tools according to defined needs, and (iii) perform the tool evaluation according to the methodology. This document contains the answers to prerequisite i, ii and iii mentioned.

The Norwegian Brønnøysund Register Centre (brreg.no), develops and operates many of the nation's most important registers and electronic solutions. One example is Altinn, the Norwegian public reporting portal. It has been operating for approximately 10 years and is a core infrastructure portal for information flow between businesses and the public sector. As an important user of vocabularies and ontologies, this portal and all the public agencies connected to it are dependent on good tools and procedures for both vocabulary / ontology engineering and governance.

Since 2001 the Brønnøysund Register Center has been running different versions of the "Norwegian Semantic Repository of Electronic Services" called SERES. SERES is an eGovernment vocabulary / ontology repository and a framework for establishing semantic interoperability. For each term in SERES, the repository aims at containing definitions and links to relevant/connected terms. According to the 5 Star Scheme for linked open data suggested by Tim Berners-Lee [9], SERES can solve the additional metadata criteria mentioned. The LOD design principles have been used for SERES. In order to make the various terms in SERES linkable, a rest based architecture that provides look up functionality and a SERES URI (GUID) has been piloted [8]. SERES acts as an online dictionary which responds according to LOD design principles, e.g. with RDF or HTML.

The top level of SERES terms are conceptual terms. For these terms we seek an editor, search, navigation and publishing tool to meet the needs described in a set of defined use cases.

In human to human communication there is a tradition to establish dictionaries, grammar rules and guidelines for good usage of a language. When computer systems communicate and automate work processes, the computer systems must act correctly upon data. The programmers developing the computer system will have a hard time to understand the full legal, business, semantic and technological impact of the exchanged data, stored data, compiled data and reused data. The dictionaries and the information models made based on dictionaries is a contract between business responsible and the computer scientists. In this picture the dictionary is the common ground for legal, business and computer scientists. A terminology tool will ensure consistent process and methodology usage, systematic workflow, support acceptance procedures and quality criteria, build common knowledge and offer a publication portal for terms and the meaning of these terms.

1.2 Contributors to this report

In addition to the authors important contributors to this report have been:

- Jenny Linnerud
    - Senior adviser, Statistics Norway
- Jostein Ven

    - Senior Adviser, Section of Standardisation, Norwegian Directorate of Health.
- Stine Strømsnes Kvaløy
    - Part of the core Semicolon project team.
    - Department for Altinn and Semantic register for electronic collaboration (SERES), The Brønnøysund Register Centre.
- Joachim Mørk-Eidem
    - Part of the core Semicolon project team
    - Department for Altinn and Semantic register for electronic collaboration (SERES), The Brønnøysund Register Centre
- Even Thorbergsen
    - Part of the Semicolon pilot team
    - Department for Altinn and Semantic register for electronic collaboration (SERES), The Brønnøysund Register Centre
- Guillermo Vega Gorgojo
    - Part of the Semicolon project
    - Research group for Logic and Intelligent Data, Department of Informatics, University of Oslo

1.3 Terms, definitions and acronyms

Terms and definitions

API	An Application Programming Interface (API) is an abstraction implemented in software that defines how others should make use of a software package such as a library or other reusable program. APIs are used to provide developers access to data and functionality from a given system. (W3C – Linked Data Glossary)
Controlled vocabulary	Carefully selected sets of terms that are used to describe units of information; used to create taxonomies, thesauri and ontologies. In traditional settings the terms in the controlled vocabularies are words or phrases, in a linked data setting they are normally assigned unique identifiers (URIs) which in turn link to descriptive phrases. (W3C – Linked Data Glossary)
Data	Is e.g. number, text, binary object like pictures and sound captured in a form and format making it suitable for storage and multiple forms of usage and exchange. Data as an abstract concept can be viewed as the lowest level of abstraction from which information and then knowledge are derived. (inspired by OAIS Reference Model definition)
Data modelling	Data modelling is a process of organizing data and information describing it into a faithful representation of a specific domain of knowledge. (W3C – Linked Data Glossary)
Government Open Data	Is data compiled and maintained by / or on behalf of a public body. The intellectual property rights, access mechanisms, quality, confidentiality, formats and description of data are such that data is suitable for reuse.
Concept	A type of thing that exists in a domain, can be identified with one or more terms/identifiers and it has properties. Equivalent to the term class as it is used in informatics. Inspired by *"Ontology" in the W3C – Linked Data Glossary*
Information governance	A holistic approach to managing and leveraging information for business benefits and encompasses information quality, information protection and information life cycle management. (IBM) Information governance is setting the rules and goals for how to perform information management.
Information management	A method of using technology to collect, process and condense information with a goal of efficient management. Most large enterprises have a central information management function to facilitate this coordination. The primary technologies

	needed are contained in a set of modelling tools that either have or interface to a production-worthy repository where the information is stored and managed. (Gartner – IT Glossary)
Linked Data	A pattern for hyperlinking machine-readable data sets to each other using Semantic Web techniques, especially via the use of RDF and URIs. Enables distributed SPARQL queries of the data sets and a browsing or discovery approach to finding information (as compared to a search strategy). Linked Data is intended for access by both humans and machines. Linked Data uses the RDF family of standards for data interchange (e.g., RDF/XML, RDFa, Turtle) and query (SPARQL). If Linked Data is published on the public Web and has a licence permitting reuse, it is generally called *Linked Open Data*. Inspired by (W3C – Linked Data Glossary)
Linked Data Principles	Provide a common API for data on the Web which is more convenient than many separately and differently designed APIs published by individual data suppliers. Tim Berners-Lee, the inventor of the Web and initiator of the Linked Data project, proposed the following principles upon which Linked Data is based: 1. Use URIs to name things; 2. Use HTTP URIs so that things can be referred to and looked up ("dereferenced") by people and user agents; 3. When someone looks up a URI, provide useful information, using the open Web standards such as RDF, SPARQL; 4. Include links to other related things using their URIs when publishing on the Web. (W3C – Linked Data Glossary)
Ontology	A formal model that allows knowledge to be represented for a specific domain. An ontology describes the types of things that exist (classes), the relationships between them (properties) and the logical ways those classes and properties can be used together (axioms). (W3C – Linked Data Glossary)
Linked Open Data / Open Linked data	Linked open data is (i) data licensed under one of several open licenses permitting reuse, (ii) published on the public Web and (iii) follow the linked data principles.
Schema	Schema refers to a data model that represents the relationships between a set of concepts. Some types of schemas include relational database schemas (which define how data is stored and retrieved), taxonomies and ontologies.
Taxonomy	Is a classification scheme for e.g. animate objects, inanimate objects, places, concepts, events, properties, and relationships. Taxonomy often has hierarchy relationship. Taxonomies are considered narrower than ontologies since ontologies apply a larger variety of relation types.
Term	An entry in a Controlled Vocabulary, Schema, Taxonomy or Ontology. It is a word/phrase or a symbol used to identify a concept and its definition. Inspired by (W3C – Linked Data Glossary)
Terminology	The set of terms/ the vocabulary of technical terms used in a particular field, subject, science, or art. Usage is very similar to vocabulary. Some subject domains prefer to use terminology to vocabulary.
Vocabulary	Is a general term for more or less structured list of terms. Usage is very similar to terminology.

Controlled Vocabulary, Schema, Taxonomy, Terminology, Vocabulary and Ontology all have the purpose of structuring terms and concepts into a model. This model is suitable for management and

to establish a common understanding of terms and concepts used in a domain. Different domains use these terms (controlled vocabulary etc.) differently and this causes confusion in interoperability efforts.

Acronyms

OWL	Web Ontology Language. OWL is a family of knowledge representation and vocabulary description languages for authoring ontologies, based on RDF and standardized by the W3C. (W3C – Linked Data Glossary)
RDF	Resource Description Framework. A family of international standards for data interchange on the Web produced by W3C. Resource Description Framework (RDF) is based on the idea of identifying things using Web identifiers or HTTP URIs, and describing resources in terms of simple properties and property values. (W3C – Linked Data Glossary)
SKOS	Simple Knowledge Organisation System. Is a vocabulary description language for RDF designed for representing traditional knowledge organization systems such as enterprise taxonomies in RDF. (W3C – Linked Data Glossary)
SPARQL	(A recursive acronym.) SPARQL Protocol and RDF Query Language define a query language for RDF data, analogous to the Structured Query Language (SQL) for relational databases. (W3C – Linked Data Glossary)
UML	Unified Modelling Language, by Object Management Group.
URI	Uniform Resource Identifier. A global identifier that may or may not be resolvable on the Web. URIs play a key role in enabling Linked Data. URIs can be used to uniquely identify virtually anything including a physical building or more abstract concepts such as colours. See also URL. (W3C – Linked Data Glossary)
URL	Uniform Resource Locator. A global identifier for Web resources. A URL is resolvable on the Web and is commonly called a "Web address". URLs and URIs are standardized by the W3C and the IETF. All HTTP URLs are URIs however, not all URIs are URLs. (W3C – Linked Data Glossary)

2 Methodology

This chapter describes our input sources, the sequence of our work and the criteria we use to evaluate candidate tools. A high level figure of the process is shown below. The evaluation process is performed by the authors of this document. If a tool does not meet one or more criteria the issue is logged and the tool is not evaluated further.

The work started in April 2013 and the first step was to choose and describe a set of use cases. These are described in Appendix: use cases. We performed a simple literature study, collected experience and searched for tools. In advance to Semtechbiz 2013 in June the use cases and main evaluation criteria were described. Discussions with vendors and users at Semtechbiz were very helpful.

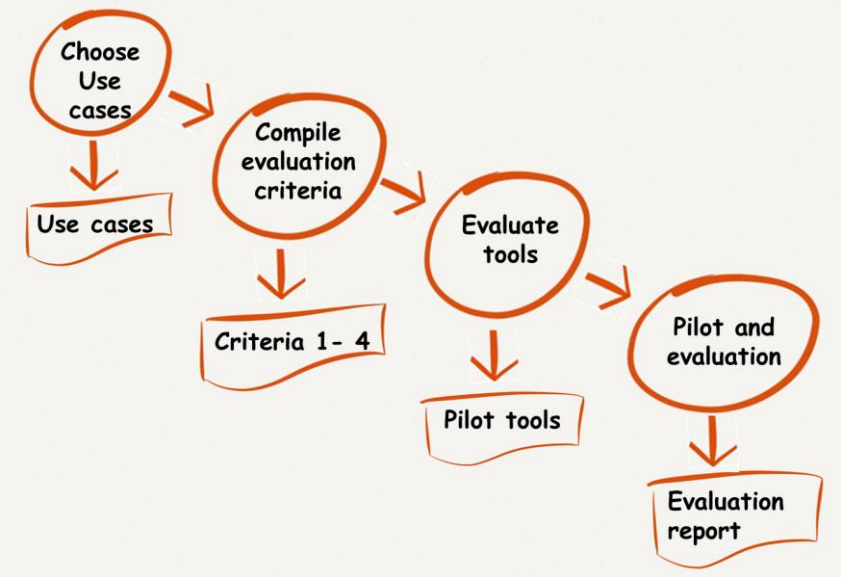

Figure 1, Project activities and deliverables

The deliverables and processes shown in the figure are further described below.

Semicolon is a research project and this survey and the upcoming pilots have got much bigger attention than anticipated and timing related to procurement plans at several government bodies for vocabulary tools seem to be very good. Based on this, even as a research project, we have to take into consideration Norwegian and EU procurement regulation. This means that we will pilot 2-4 tools and have a well defined and a distant relationship to all vendors during our piloting process.

2.1 Use cases

During the spring a group of government bodies was invited to discuss relevant use cases. The government bodies and the project submitted use case drafts and these were quality assured in a separate workshop. The use cases are:

- Use case 1: Terminology/vocabulary maintenance at a government body
- Use case 2: Publishing terminology/vocabulary
- Use case 3: Register a concept and its relationships
- Use case 4: Browse and view concepts from another government body
- Use case 5: Compare concepts

The use cases are described in more detail in Appendix: Use cases.

2.2 Evaluation criteria

The evaluation criteria are made by input from:

- The use cases
- Project scope, context, limitations and investment & lifetime cost
- Existing software the pilot tools needs to be integrated with/ synchronise models with
- Need for workflow support
- Offering of relevant tools
- Literature. , e.g.
    - Evaluation of tools for the Semantic Repository for Electronic Services (SERES) [2][3][5][4]. Conclusion was Magic Draw as main editor and a multiuser UML repository. Currently the repository is based on Adaptive.
    - Evaluation of vocabulary tools [10], which listed a set of requirements and which concludes on use of PoolParty. Candidate tools were Lexaurus (Lexaurus Editor og Lexaurus Bank), PoolParty Thesaurus Management System and SKOSed for Protégé
    - Decision Framework for Evaluating Metadata Repositories [7] describing a framework for setting up hierarchies of evaluation criteria and how to handle the weighting of the criteria.

The evaluation criteria are grouped and the evaluation process uses the criteria in sequence. The groups are as follows:

- Criteria 1: General criteria for vocabulary and concept management tools
- Criteria 2: Project specific requirements
- Criteria 3: Tool category properties
- Criteria 4: Detailed level criteria for tool and vendor

Separate chapters below describe these criteria in detail. A prerequisite for using these criteria is knowledge within:

- Technical, semantic, organisational and juridical interoperability issues.
- Terminology, vocabulary and/or ontology engineering, management and governance.
- Software engineering and system integration.

Some of the criteria listed below are general for most tool evaluation processes and some are project specific. If other projects reuse our criteria then they should tailor the criteria to meet their own project goals, context, use cases and limitations.

Tool vendors, academia and standardisation bodies use different words for the concept & vocabulary engineering and management topic. Based on this knowledge we have searched for tools for (i) terminology or vocabulary management, (ii) concept and ontology management or (iii) classification management.

2.2.1 Evaluation Criteria 1 (General criteria for vocabulary and concept management tools)

The search process for candidate tools has used criteria 1, described below, as evaluation criteria.

Criteria group	Criteria description, Criteria 1
C-1	1. Open and easily accessible information should document that the tool can meet the criteria listed in Criteria 1, Criteria 2, Criteria 3 and Criteria 4.
C-1	2. Tools dedicated for concept and terms/vocabulary management with the purpose to achieve concept openness and semantic interoperability
C-1	3. The tool must a. Be applicable for more than one domain terminology b. Handle multiuser issues c. Give workflow support
C-1	4. The tool must be able to handle relationships between (i) terms, (ii) concepts, (iii) definition, and combinations of i – iii.
C-1	5. Cost of client side must be zero or close to zero. (Anticipate that the system architecture has a server/repository and a client/GUI.)

Supplementing explanation to [C-1 : 3] above: There are many variants of repository model complexity and the variants listed below are only examples and not a complete list of alternatives for the pilot:

a) Controlled vocabulary, with simple or not well defined relationship types between concepts
b) Taxonomy (with hierarchical relations) and other types of defined relationships between concepts
c) Taxonomy as part of a structured graph with formally defined relationships
d) Variant of bullet b or c above + linking between concept repositories.

In this survey we do not include: Advanced formal ontology tools (since they will be too advanced for the user community), tools for DB design, traditional information modelling tools, general diagram tools, general UML editors and tools packaged in enterprise suites.

For each bullet and sub-bullet, the evaluation is scored as:

- No support
- Some support
- Good support
- Very good support, and or interesting addition

2.2.2 Evaluation Criteria 2 (Project specific requirements)

Criteria 2, the project specific requirements are as follows:

Criteria group	Criteria description, Criteria 2
C-2	1. Methodologies, content and model standard compliance a. DIFI metadata specification (in Norwegian only) [6] b. Terminology and concept life cycle methodology (in Norwegian only) [18]

C-2	2. Adaptive XMI profile or Adaptive API support, alternatively by tailored transformations.
C-2	3. Payment to vendor. Price and life time cost for pilot software and paid configuration, tailoring or consultancy etc. (Semicolon resource man-hours not included in this criteria.)

For each bullet and sub-bullet, the evaluation is scored as:

- No support
- Some support
- Good support
- Very good support, and or interesting addition

2.2.3 Evaluation Criteria 3 (Tool category properties)

Criteria 3, the tool category properties are as follows:

Criteria group	**Criteria description, Criteria 3**
C-3	1. Bundling issues a. Standalone vocabulary tool able to utilize/integrate with text mining/ data analytics features for making initial ontology and/or ontology learning. (best score) b. Standalone vocabulary tool c. Part of software development suite or requires other preinstalled and / or costly software
C-3	2. Runtime category, client a. Web-browser based client, no installation (best score) b. Web-browser based client, with installation c. Separate software client
C-3	3. Ontology engineering support a. Manual engineering or by import, plus text mining or data analytics features for making initial ontology and or ontology improvements (best score) b. Manual engineering or by import
C-3	4. Repository architecture a. Search, concept comparison and interlinking between repositories. (best score) b. Interlinking between logical repositories. c. One logical repository with several domain vocabularies
C-3	5. Terminology and code list tool a. Support for terms with code lists with versioning (best score) b. Support for terms with code lists c. Only support for term
C-3	6. Little need for tailoring and configuration before pilot can be tested to meet the use cases. Installation, configuration setup and import of one or more

	vocabularies believed to be: a. no more than 1 day of work (best score) b. no more than 3 days of work c. More than 3 days of work

To pass this evaluation scoring open and easily accessible information should indicate tool capabilities to support the requirements above. Best score gives 5 points. Example of how this could be used in the scoring process is as follows:

1. Bundling issues a. Standalone vocabulary tool able to utilize/integrate with text mining/ data analytics features for making initial ontology and/or ontology learning. (best score) b. Standalone vocabulary tool c. Part of software development suite or requires other preinstalled and / or costly software	5 points 2 points 1 point

2.2.4 Evaluation Criteria 4 (Detailed level criteria for tool and vendor)

Criteria 4, the detailed level criteria for tool and vendor are as follows:

Criteria group	**Criteria description, , Criteria 4**
C-4	1. Workflow support a. Collaboration support b. Enable tailoring of workflow for a set of processes (see the use cases) including: i. drafting process with commenting functionality ii. review process with commenting functionality iii. quality assurance iv. publishing internally and externally c. Tracing of status and reporting of workflow, workflow analysis
C-4	2. Integration capabilities a. Import and export of concepts models with relevant history b. Ability to be used as frontend to one or several concept repositories in one or more IT-security zones/ distributed environments etc.
C-4	3. Usability a. Easy to learn and use. b. Open access to updated tutorials and user documentation c. Easy to tailor and develop to meet end user needs, super user needs and IT- administrator needs d. Easy to integrate to meet enterprise needs e. Multilingual, one concept may have several terms in more than one language. f. Concepts with code lists

	g. Visualisation and filtering of i. Concepts and their relations ii. Changes and graph evolution timeline iii. Graph visualisation capabilities h. Support for automatic or suggestion based linking of word used in definitions to concept.
C-4	4. Information governance a. Creation, update, terminate and versioning (life cycle management) of terms, concepts, relations, subsets, domain graphs, and graph-subsets etc. b. Graph consistency handling and checking c. Data quality methodology and functionality, graph quality measurements d. Use global identifier for concepts e. Support for i. Cross model referring. Other internal models ii. Cross model referring to external models, ISO standards, statistic definitions etc. iii. Relations to source/sources of concept and definition, e.g. source to laws
C-4	5. Repository administration a. User administration and role based access b. Statistics and reporting c. Administrative grouping of concepts (domain, data steward, abstraction level etc.) d. Error warning, tracing and handling e. Tool governance and life cycle management f. Tailoring of tool concept meta model
C-4	6. Publishing capabilities a. Easy to publish, export, prepare for advanced search etc., the whole or part of the concept model b. Machine to machine look up c. Give open access to public published concepts, e.g. Look up functionality from other tools and information resources. E.g. offering metadata for the linked data cloud.
C-4	7. Search capabilities a. Traditional search, e.g. exact match on term name according to fields in repository schema b. Free text search, synonyms, antonyms etc., and Natural Language Processing (NLP) capabilities (e.g. find correlations even if singular and plural forms of word are used, and much more.) c. Graph search with use of logical constraints (includes bullet a and b)
C-4	8. Vendor properties (these properties are also valid for open source initiatives) a. Trust b. Solidity

	c. Tool strategy d. Upgrade road map e. License philosophy f. Local representatives g. Support capacity and quality h. Consultants for software tailoring
C-4	9. Commercial issues a. Licence cost b. Tools short and long term cost c. Intellectual property rights of content / concepts (most relevant for hosted services)
C-4	10. Tool community a. Number of users b. Level of activity c. Kind of offerings like, tool support, methodologies, tutorials, mailing lists, gatherings etc.
C-4	11. Methodologies, content and model standard compliance a. ISO 704, Terminology work – Principles and methods b. ISO 1087-1, Terminology work - Vocabulary c. ISO 11179, Information technology – metadata registries (MDR) d. ISO 20943-1 Information technology – Procedures for achieving metadata registry (MDR) content consistency e. Dublin core f. W3C Provenance
C-4	12. Interface compliance, formats and protocols a. Formats i. SKOS ii. OWL iii. RDF iv. XMI (UML export format) b. Protocols and API i. OWLAPI ii. Web services
C-4	13. Runtime issues a. Multiuser handling b. IT risk issues like confidentiality, integrity, accessibility, scalability etc. c. IT platform and architecture d. User accounts e. Single sign on f. User tracing

It is recommended to perform the scoring according to the criteria above in a spread sheet. Scoring scale can be:

Rating	Meaning of rating
1	zero or very weak support
2	weak support
3	support
4	good support
5	good support with supplements

The scoring is summed to main bullet level, meaning that the category titles 1, 2, 3 etc. will receive the scoring. For our project, usage of detailed scoring and documentation for all sub-bullets is too time consuming and difficult without very thorough investigations and piloting of the different tools.

3 Evaluation results

3.1 Evaluation of criteria 1

The input to the candidate list comes from:

- Input from colleagues and project partners
- Info gathering at Semtechbiz 2013[1] and discussions with vendors and conference participants.
- Online search for vocabulary tools, see the references: [12],[13],[14],[15] and [16].

The different online list of ontology tools contains both recent and discontinued tools, and for our purpose discontinued tools are not of interest. Each tool below is described in a separate chapter in Appendix: Tool description.

The evaluation of Criteria 1 gave the following result:

#	Tools	Criteria 1, evaluation result	Comment
1	**Adaptive Business Glossary Manager** By Adaptive	Stop	We did not find sufficient open information on functionality or architecture to proceed.
2	**Anzo, Operational Metadata Management** by Cambridge Semantics	Short listed	
3	**Business Glossary** by IBM	Short listed	
4	**Business Information Modeler** by Kalido	Stop	We did not find sufficient open information on functionality or architecture to proceed.
5	**Collibra Business Glossary** by Collibra	Short listed	
6	**Enterprise Vocabulary Net** by Topbraid	Short listed	
7	**Lexaurus** by Knowledge Integration	Short listed	

[1] http://semtechbizsf2013.semanticweb.com/

8	**Ontotext**	Stop	We did not find sufficient open information on functionality or architecture to proceed.
9	**Pool party** by Semantic Web Company GmbH	Short listed	
10	**SemanticXpress**	Short listed	
11	**Semantic MediaWiki**	Short listed	
12	**SKOSed for Protégé**	Short listed	
13	**VocBench**	Short listed	
14	**WebOntoStudio** By Semafora Systems GmbH	Short listed	
15	**Webprotege 2.0** by Stanford University	Short listed	

If a tool fails to meet one or more criteria, we have not continued to evaluate the rest of the criteria for that tool.

3.2 Evaluation of criteria 2

The evaluation of Criteria 2 gave the following result:

Tools	Criteria 2, evaluation result	Comment
Anzo, Operational Metadata Management by Cambridge Semantics	Stop	The part of the tool used for ontology engineering requires Microsoft Excel on each client.
Business Glossary by IBM	Short listed	
Collibra Business Glossary by Collibra	Short listed	
Enterprise Vocabulary Net by Topbraid	Short listed	
Lexaurus by Knowledge Integration	Short listed	
Pool party by Semantic Web Company GmbH	Short listed	
SemanticXpress	Short listed	
Semantic MediaWiki	Short listed	
SKOSed for Protégé	Short listed	
VocBench	Short listed	
WebOntoStudio By Semafora Systems GmbH	Short listed	
Webprotege 2.0 by Stanford University	Short listed	

3.3 Evaluation of criteria 3

The evaluation of Criteria 3 gave the following result:

Tools	Criteria 3,	Comment

	evaluation result	
Business Glossary by IBM	Stop	Seems to be bundled with other software and could not find a trial license.
Collibra Business Glossary by Collibra	Short listed	
Enterprise Vocabulary Net by Topbraid	Short listed	
Lexaurus by Knowledge Integration	Short listed	
Pool party by Semantic Web Company GmbH	Short listed	
SemanticXpress	Stop	Unclear if SemanticXpress is a separate tool or a rebranding/ bundling of WebOntoStudio.
Semantic MediaWiki	Short listed	Anticipate medium or high effort to set up a pilot environment suitable for piloting the use cases and meeting the user needs.
SKOSed for Protégé	Short listed	
VocBench	Short listed	
WebOntoStudio By Semafora Systems GmbH	Short listed	
Webprotege 2.0 by Stanford University	Stop	Anticipate medium or high effort to set up a pilot environment suitable for piloting the use cases and meeting the user needs. (To owl focused user interface to meet user group and use cases.)

3.4 Evaluation of criteria 4

The evaluation of Criteria 4 gave the following result:

Tools	Criteria 4, evaluation result	Comment
Collibra Business Glossary by Collibra	Short listed	
Enterprise Vocabulary Net by Topbraid	Short listed	
Lexaurus by Knowledge Integration	Stop	Uncertain commercial issues and vendor properties.
Pool party by Semantic Web Company GmbH	Short listed	
SKOSed for Protégé	Stop	Lack of active community and tool releases.
Semantic MediaWiki	Short listed	
VocBench	Stop	Unclear tool strategy and vendor properties. The VocBench project themself seems to evaluate commercial off-the-shelf products.
WebOntoStudio By Semafora Systems GmbH	Stop	Too OWL focused user interface to meet user group and use cases.

3.5 Tools to be piloted

Tools	Comment input to piloting process
Collibra Business Glossary by Collibra	Could not download a trial license without submitting information to vendor.
Enterprise Vocabulary Net by Topbraid	Price for pilot unclear.
Pool party by Semantic Web Company GmbH	May-be there is only a net account offer for test purposes.
Semantic MediaWiki	This pilot is run in parallel at the Brønnøysund register centre but measured against the same use cases and methodology.

4 Further work

Further project activities are:

- A common approach for how to measure use case compliance will be made.
- The project will install the suggested pilot tools and run the pilots with one or more terminologies during fall 2013.
- After the pilot period a short experience report will be made.
- The relevance and structure for the criteria 1-4 will be evaluated, and is planned to be published
- An experience paper may be published.

In the end there are several possible practical outcomes of this work. We believe the main alternative outcomes are a combination of:

1. Brønnøysund Register Center uses the knowledge:
    a. for internal purposes and may develop their own vocabulary management tool or extend some existing tools.
    b. as input to a public tender process for either a COTS tool or as input to a software development delivery procurement based on open source or COTS.
2. Other public bodies and companies use the evaluation criteria and pilot evaluation report as a basis for their own development or public tender processes.
3. The evaluation criteria and the knowledge built during the pilots are compiled in a report or paper.

4.1 System description of pilot

A conceptual figure of how pilot tools are integrated with SERES is shown below.

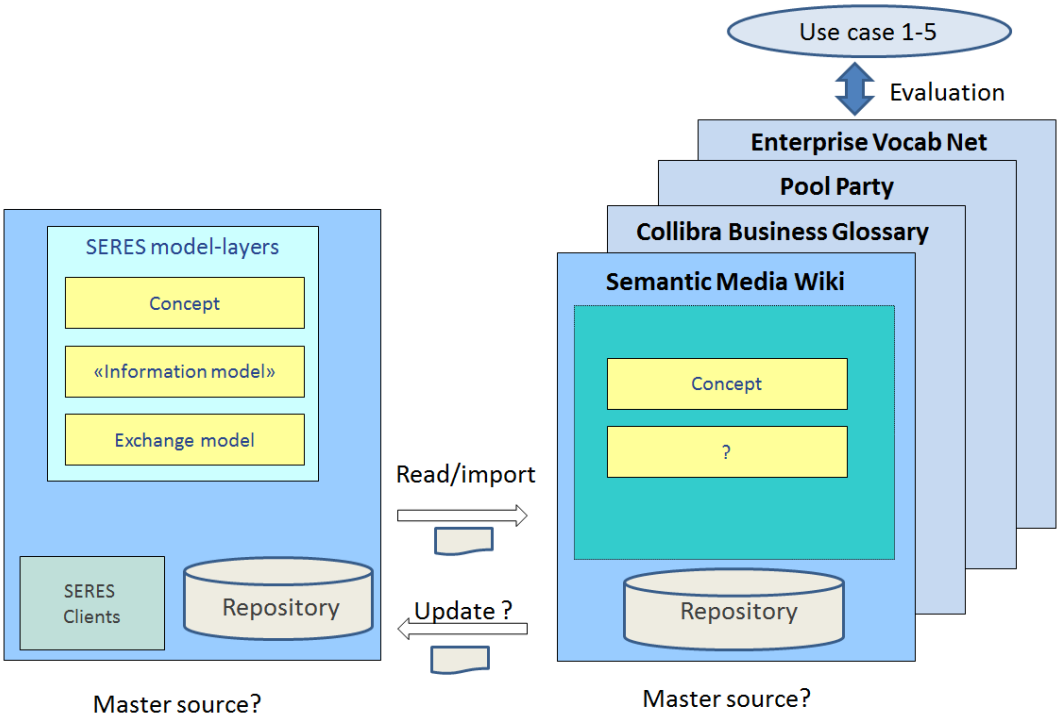

Figure 2, Pilot evaluation set up

Each tool will be evaluated against the use cases listed in Appendix Use Cases.

5 References

All websites and material have been visited during May-June 2013.

1. An Environment for Merging and Testing Large Ontologies. Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR2000). Breckenridge, Colorado. April 12 - 15, 2000. http://disi.unitn.it/~accord/RelatedWork/Matching/McGuinnessKR.pdf
2. Coordinating public reporting obligations by utilizing of reference data and related services. (in Norwegian only). Offentlig innrapportering basert på samordning, metadata og tjenester fra OR. Skisser over mulige realiseringer av veien videre for OR. Desember 2003. Arne Jørgen Berre, Arve Larsen, Per Myrseth, Espen Slotvik og Terje Bertelsen.
3. Survey on UML tools. (In Norwegian only) Markedsundersøkelse av UML-modelleringsverktøy, DNV rapport 2007. Hjørdis Hoff og Martin Skjæveland.
4. SERES requirements, conceptual model and functionality. (In Norwegian only)
5. Evaluation of SERES by the Norwegian Tax Authority. (In Norwegian only). Terje Kolbu, Geir Myrind and Tor Murvold.
6. Norwegian national standard for describing concepts. (In Norwegian only) Standard for begrepsbeskrivelser. Versjon 1.0. DIFI 2012-05-11 http://standard.difi.no/filearchive/2012-05-13-mal-begrepsbeskrivelser-1-0.pdf
7. Decision Framework for Evaluating Metadata Repositories. Published: 17 October 2011. Gartner. http://www.gartner.com/id=1824215
8. National Master Data as 5 Star Linked Open Data. By Per Myrseth, Jens Kilde Mjelva, David Norheim and Thom-Kåre Granli. IFIP EGOV 2012. 3-5. September 2012, Kristiansand, Norway.

9. Tim Berners-Lee, Is your linked open data 5 star? Amendment 2010. http://www.w3.org/DesignIssues/LinkedData.html, visited April 2012
10. Evaluation of vocabulary tools, UDI 2011. In Norwegian only. (Original title: "Evaluering av forvaltningsverktøy")
11. SKOS Implementation Report, May 19th 2009, http://www.w3.org/2006/07/SWD/SKOS/reference/20090315/implementation.html
12. Wikipedia ontology editors. http://en.wikipedia.org/wiki/Ontology_editor
13. A Comparative Study Ontology Building Tools for Semantic Web Applications. By Bhaskar Kapoor and Savita Sharma. International journal of Web & Semantic Technology (IJWesT) Vol.1, Num.3, July 2010.
14. OpenStructs Tech Wiki. Ontology tools. http://techwiki.openstructs.org/index.php/Ontology_Tools
15. Sweet Tools – Simple List, by Michael K. Bergman. http://www.mkbergman.com/sweet-tools-simple-list/
16. Semantic Web Development Tools. http://www.w3.org/2001/sw/wiki/Tools
17. Selection Criteria for Business Glossary Tools, by Sunil Soares. June 26, 2013. http://www.dataversity.net/selection-criteria-for-business-glossary-tools/
18. Terminology and concept life cycle methodology. 2011, by Vibeke Dalberg, Per Myrseth and Jim Yang. http://www.semicolon.no/?p=682

6 Appendix: Tool Descriptions

Read comment: Text in this chapter is cut and paste from web-pages, presentations and or documents, see source for each tool.

6.1 Adaptive Business Glossary Manager

Vendor	Adaptive
Source	http://www.adaptive.com/products/adaptive-business-glossary-manager/
Software version or data visited web site	Adaptive Business Glossary Manager. Visited august 2013.

Adaptive Business Glossary Manager™ is a web-based platform used to acquire, organize, analyze and distribute knowledge about an organization's data assets.

Adaptive Business Glossary Manager:

•Offers a lightweight business-oriented interface;

•Centrally manages business terms across multiple domains;

•Declares accountability and stewardship of terms;

•Automates change approval processes and policies to accelerate creation of common vocabulary;

•Reduces time to find business term definitions making IT and business project work more efficient; and

•Enables mapping between business terms and industry ontologies to support benchmarking and gap analysis.

6.2 Anzo, Operational Metadata Management

Vendor	Cambridge Semantics
Source	http://www.cambridgesemantics.com/solutions/smart-enterprise-data-management#metadata-management
Software version or data visited web site	Anzo EDM solution. Visited 26. june 2013

Anzo Metadata Management solutions capture rich, expressive metadata about your key business entities and data elements. Anzo lets both data stewards and business analysts manage concepts, relationships, vocabularies, taxonomies, and thesauri by tracking metadata elements' approval and usage life cycles, promoting reuse via semantic/conceptual search, and allowing you to capture arbitrary extended attributes about any metadata element. Anzo also lets you operationalize your metadata by using it to directly drive forms, visualizations, analytics, and data-integration processes.

6.3 Business Glossary

Vendor	IBM
Source	http://www-01.ibm.com/software/data/infosphere/business-glossary/features.html
Software version	8.7

Features:

- **Manage business terms and categories:** Business Glossary provides a dedicated, web-based user interface for creating, managing, and sharing a controlled vocabulary, including batch editing capabilities. Terms represent the major information concepts in your enterprise and categories are used to organize into hierarchies.
- **Manage stewardship:** Stewards are people or organizations with responsibility for a given information asset. Using Business Glossary administrators can import steward profiles from external sources, generate and edit profiles in the Web interface, and create relationships of responsibility between stewards and business terms or any of the artifacts managed by IBM InfoSphere Information Server.
- **Customize and extend:** Needs around business metadata tend to differ from one enterprise to the next. For this reason, there is no "one size fits all" meta-model. In addition to being able to customize the entry page to the application, administrators can extend the application with custom attributes on both business categories and business terms.

- **Collaborate:** It is not enough to simply document business metadata. This information is active in the enterprise with open access to all members of business and development teams. IBM InfoSphere Business Glossary provides a collaborative environment in which users can evolve this important information asset as the business changes and adapts to market conditions, shifting customer needs and competitive threats.
- **Contextual search and visibility business term definitions:** Business Glossary Anywhere is an application independent search / pop-up box that can be called from any application (Excel, data modeling tools, reporting applications, Microsoft Word, etc.) that provides instant access to Business Glossary terms, taxonomies and stewards.
- **Simply Browse:** Business Glossary browser is an intuitive read-only web-based interface requiring no training to utilize. Business users can search and explore the common controlled vocabulary and relationships, identify stewards responsible for assets and provide direct feedback.
- **Stronger data governance:** Users can take ownership of business metadata by importing terms from external sources, authoring and editing terms in the Web interface, classifying terms into categories, and relating terms to more technical artifacts managed by the IBM InfoSphere Information Server.
- **Accountability and responsibility** can be assigned, supporting enterprise data governance and Electronically Stored Information (ESI) requirements.
- **Improved productivity:** Administrators can tailor the tool to the needs of their business users.
- **Increased collaboration:** users are able to add annotations to business terms and categories as well so other team members downstream in project cycles have a greater understanding of the context of information.
- **Greater trust in information:** business users now have immediate, in-context access to terms and definitions, facilitating a greater understand of and trust in the information they rely on to make critical business decisions.

6.4 Business Information Modeler

Vendor	Kalido
Source	http://www.kalido.com/business-information-modeling.htm
Software version	

The Kalido Business Information Modeler is a transformational approach to designing, deploying and maintaining your BI infrastructure. At the heart of the Business Information Modeler is a user interface for drawing, structuring, and defining the attributes of types of business entities. The ability to print and export the model information to the popular PDF format also improves communication between architects and business users.

Instead of modeling data and their structures, the Business Information Modeler allows you to model the actual parts of your business – customers, products, assets, transactions, even people – and define how you view the information in business terms, not technical ones.

6.5 Collibra Business Glossary

Vendor	Collibra
Source	http://www.collibra.com/products/business-semantics-glossary
Software version or data visited web site	26. June 2013

Different terms mean different things depending on the context. For example: Is “Customer” the same for Finance as it is for Sales? Do the same rules apply? What are the policies? Which underlying business objectives are they representing? Who are the owners, stewards and stakeholders?

A business glossary is a key component to tackle the questions above. The collaborative and iterative creation of such an inventory is a mandatory first task for the core data governance team. Once this process is up and running you have a solid foundation to move data governance forward.

Data Directory

A true data directory also includes technical metadata: which fields, columns and records are out there, and what do they mean? What is the type of the data element, and is there a limit on the length? In which system or application is the field contained? What data structures is it part of?

Features:

- Advanced search
- Ownership & responsibility
- Templating
- Collaborating and workflow
- Control & track changes
- Hierarchy management & semantic modelling
- Import and export
- Traceability and impact analysis
- Mapping

6.6 Enterprise Vocabulary Net

Vendor	Topbraid
Source	http://www.topquadrant.com/solutions/ent_vocab_net.html
Software version	TopBraid EVN

TopBraid Enterprise Vocabulary Net (TopBraid EVN) is a web-based solution for simplified development and management of interconnected controlled vocabularies. It supports business stakeholders who need to collaborate on defining and linking enterprise vocabularies, taxonomies, thesauri and ontologies used for information integration, customization and search.

Key Features of TopBraid EVN

- **Flexible Data Model** Ability to define and use any number of custom attributes and relationships
- **Vocabulary Editing** Cloning, merging, repositioning and re-numbering of concepts; global edit operations for selected groups of concepts
- **Search** Simple lookups and advanced search, tree and list views, visual graph exploration and query
- **SKOS** Built-in support for all standard SKOS attributes and hierarchical, associative and equivalency relationships
- **Automatic processing** Ability to create custom validation rules and automated script processing via SPARQL Rules and SPARQLMotion; SKOS and OWL rules and constraints ready to use out of the box
- **Import/Export** Import/Export from RDBMs, RSS feeds, spreadsheets, XML, SPARQL endpoints, RDF and OWL
- **Audit Trails** Every change is logged and time stamped, change history can be searched, changes can be rolled back if desired
- **Unlimited Work-in-Progress Copies** Virtual work-in-progress copies of vocabularies allow parallel development of versions and enable controlled publishing, review and approval workflow
- **Impact Analysis** Ability to query and create longitudinal analysis on any aspect of the vocabularies
- **Role-based Access Control** Ability to define view-only, edit and manage roles for each vocabulary and for the individual working versions of a vocabulary
- **Reporting** Built-in reports for hierarchy exports, spell checking, constraint violations and graph statistics Advanced query-building tools and reporting through graphical interfaces.
- **Merging** RDF standard universal identifiers provide easy “hooks” for merging vocabularies
- **Systems Integration** Integrate with existing enterprise or vocabulary management systems via Web Services interfaces and APIs
- **Rich Web Interface** Types (classes), properties and instances can be defined in an intuitive web-based system that supports drag and drop, autocompletion and rich text editing
- **Customizability** On-the-fly creation of customized user interfaces that can meet the needs of each user group
- **Open Architecture and Standards** External data need not be transformed into a proprietary schema. TopBraid EVN offers native support for RDF, OWL and SPARQL

- **Enterprise-ready** Scalable and robust architecture with DBMS deployment (choice of relational and RDF databases) with LDAP integration for access control

6.7 Lexaurus (Lexaurus Edior og Lexaurus Bank)

Vendor	Knowledge Integration
Source	http://www.k-int.com/products/lexaurusbank
Software version	Lexaurus bank and editor

Lexaurus Bank is a powerful terminology management system for publishing term or concept based vocabularies, concept schemes, data definitions, taxonomies and thesauri.

It facilitates the import, export and cross-mapping of terminology structures in a variety of formats currently supporting SKOS, VDEX (token and nested), XVD, ESD Toolkit, Zthes,i2b2 (ontology and concept) and Snomed.

Lexaurus Bank can also synchronise with one or more Lexaurus Editors to provide collaborative editing of terminology information, including multiple users working on the same structure.

It is completely multilingual both in terms of the data it contains and the user interface for the application.

Lexaurus Editor key features:

- **Easy to use** The editing interface can be used for developing all forms of terminology structures, including poly-hierarchical taxonomies, faceted thesauri and ontologies.
- **Multilingual** Lexaurus can create and edit multilingual structures and the user interface canalso be presented in different languages.
- **Customisable** The metadata used to describe terms or concepts is configurable and schema-driven. This allows the creation of custom properties and data types allowing you to manage your data in the way that suits your organisation.
- **Customisable relationships** In addition to all common thesaural relationships users can define their own relationship types.
- **Format independent** supporting both concept and term-based formats. Please ask about support for your format of choice.
- **Drag and drop** for quick, simple modification of hierarchies, re-use of different terminology sets and mapping between vocabularies.
- **Customisable reporting** langauge, content and styling customisation is supported.
- **Synchronising with remote Bank(s**) for distributed terminology management and team editing.
- **Tracking** of remote and local changes to facilitate change merging for team development.
- **Full history** including specific edit operations and rollback to any point (equivalent to ‘n level undo’).

- **Export and import of translation data** as Excel to allow translations to be edited in a spread sheet and then re-imported.

6.8 Ontotext

Vendor	Ontotext AD
Source	http://www.ontotext.com/products
Software version	

Ontotext develops, manages and supports the technologies behind the following products:

- OWLIM is an industrial-scale semantic database, using Semantic Web standards for inference and integration/consolidation of heterogeneous data.
- KIM Platform is a semantic search engine, using text analysis to provide hybrid queries involving structured data and inference.
- Semantic Biomedical Tagger is an information extraction system, designed to create semantic annotations in biomedical texts using more than 100 different semantic types.
- Web Mining Framework is a comprehensive, efficient web intelligence and web search platform.
- PROTON is an upper-level light weight ontology, used for semantic search and annotation and as a reference layer to access LOD.

6.9 Pool party

Vendor	Semantic Web Company GmbH
Source	http://www.poolparty.biz/portfolio-item/vocabulary-management/
Software version	

Our solution approach

In order to get (enterprise) vocabularies widely accepted the costs for the creation and development of such thesauri and vocabularies have to stay as low as possible. This can be achieved if thesaurus managers get support by appropriate methods and software tools to produce high-quality semantic metadata built upon open standards. In case the enterprise (or domain-specific) thesaurus is built upon W3C's Simple Knowledge Organization System (SKOS) it can also build the core of an organization's knowledge graph to be reused by many other applications. In addition, built-in text analytics, several importers and linked data enrichment tools help to extend the enterprise vocabulary further and further while keeping the efforts as low as possible. A comprehensive library of quality- and validity checks makes sure that the outcome will meet the highest demands for quality. Putting an enterprise vocabulary to the right place means, that it should be reused by other applications as often as possible. Several standard APIs allow quick integration as well as complex queries over the resulting knowledge graph.

- Enterprise vocabularies fully compatible with W3C's semantic web standards (SPARQL, RDF, SKOS)

- Ready to be used within a linked data enterprise architecture
- Highly comfortable thesaurus editor, fully web-based with hundreds of features
- Importers for legacy data sources
- Integrations with frequently used enterprise systems like Sharepoint, Confluence or Drupal
- Facilities to enrich thesauri with terms from document collections and linked open data

Used methods, technologies and standards

- PoolParty Thesaurus Server
- Simple Knowledge Organization System (SKOS)
- PoolParty Knowledge Modeling Approach
- Linked Data enrichment
- Data importers and text analytics
- Thesaurus Quality and Validity Checker (qSKOS)

6.10 Semantic MediaWiki

Vendor	open-source extension to MediaWiki
Source	http://semantic-mediawiki.org/wiki/Semantic_MediaWiki
Software version	Semantic MediaWiki 1.8.0.5

Semantic MediaWiki introduces some additional markup into the wiki-text which allows users to add "semantic annotations" to the wiki. While this first appears to make things more complex, it can also greatly simplify the structure of the wiki, help users to find more information in less time, and improve the overall quality and consistency of the wiki. Here are some of the benefits of using SMW:

- **Automatically-generated lists.** Wikis tend to contain many aggregated lists; Wikipedia itself has thousands, like "List of metropolitan areas in Spain by population". Those lists are prone to errors, since they have to be updated manually. Furthermore, the number of potentially interesting lists is huge, and it is impossible to provide all of them in acceptable quality. In SMW, lists are generated automatically like this. They are always up-to-date and can easily be customised to obtain further information.

- **Visual display of information.** The various display formats defined by additional extensions, such as Semantic Result Formats and Semantic Maps, allow for displaying of information in calendars, timelines, graphs and maps, among others, providing a much richer view of the data than simple lists would.

- **Improved data structure.** MediaWiki wikis tend to make heavy use of categories for structuring data. While these are generally helpful, consider the category on Wikipedia called "1620s deaths"; if the information in these pages were stored using SMW, these categories could be replaced by simple semantic values, reducing the need for a complex classification system. In addition, if semantic markup within the wiki is stored within templates, otherwise known as semantic templates, a wiki can easily gain a solid data structure. And the Semantic Forms extension lets administators create forms for adding and editing the data within semantic templates, thus making the addition of semantic information possibly even easier and more straightforward than regular wiki text.

- **Searching information.** Individual users can search for specific information by creating their own queries, supported via extensions like Halo and Semantic Drilldown.

- **External reuse.** Data, once it is created in an SMW wiki, does not have to remain within the wiki; it can easily be exported via formats like CSV, JSON and RDF. This enables an SMW wiki to serve as a data source for other applications, or, in the case of enterprise usages, to take over the role that a relational database would normally play. Through the use of the External Data extension, SPARQL, and other tools, one SMW-based wiki can even use the data from another, eliminating the need for redundancy between wikis. You can also query SMW's data from outside the wiki, via the API or an RDF triplestore.

- **Integrate and mash-up data.** Data contained in an SMW installation does not have to be an isolated store of information. Extensions such as Data Import, Data Transfer and External Data empower you to integrate external data (coming e.g. from legacy systems, web services or linked data sources) and interrelate it with existing semantic data in the wiki. Thus, an SMW-powered wiki can serve as a central information hub in an IT landscape.
- Simple data export through standard functionality: Special:Export and Special:ExportRDF
- Utilize categories and properties to enrich functionality

6.11 SemanticXpress

Vendor	Semafora systems GmbH
Source	http://www.semafora-systems.com/en/solutions/semanticxpress/
Software version	

SemanticXpress is an ontology-based solution for company-wide vocabulary-management. The existing terminology from the different departments is linked to the company's vocabulary using so-called mappings. Users can access and interact with SemanticXpress using a standardized, company-wide Web interface. External applications can also access the stored vocabularies (for example, to leverage the Intranet search of a company). Therefore SemanticXpress provides Web services that are suitable for easily connecting external applications.

Features:

- Use available vocabularies for Life Science and Healthcare (UMLS, e.g. MeSH, NCI, ChEBI, ICD-10, GO or SNOMED)
- Integration with internal and external vocabularies via graphical mappings
- Application-specific subsets and term extensions
- The faster provision of information via Web services
- Scalable and highly performant
- Central cockpit for the entire organization:
- "We speak one common language!"
- Dynamic information management
- Semantic search based on the stored ontologies

6.12 SKOSed for Protégé

Vendor	Open source community

Source	http://code.google.com/p/skoseditor/
Software version	

SKOSEd is a plugin for Protege that allows you to create and edit thesauri (or similar artefacts) represented in the Simple Knowledge Organisation System (SKOS).

6.13 VocBench

Vendor	Food and Agriculture organisation of the United Nations
Source	http://aims.fao.org/tools/vocbench-2
Software version	VocBench 1.3.1

VocBench is a web-based, multilingual, editing and workflow tool that manages thesauri, authority lists and glossaries using SKOS. Developed by FAO and its partners, VocBench is designed to meet the needs of semantic web and linked data environments. VocBench provides tools and functionalities that facilitate both collaborative editing and multilingual terminology. It also includes administration and group management features that permit flexible roles for maintenance, validation and quality assurance.

The VocBench community is growing and today includes FAO's Fisheries and Aquaculture Department and the data.fao.org project, the European Commission Publications Office and the European Environment Agency. FAO's instance of VocBench kindly hosted by FAO Centre of Excellence MIMOS Berhad currently manages the AGROVOC thesaurus, the Biotechnology Glossary and other bibliographic metadata.

The latest production release is VocBench 1.3.1 which can be downloaded here (please note that it is not particularly easy to install and that we do not provide support of any kind.) The team is currently completing a major rewrite that shares services with University of Rome at Tor Vergata's Semantic Turkey, a Mozilla plugin for semantic annotation and ontology enrichment. This version (2.0) will be released in the spring of 2013 and features many improvements such as native SKOS support, support for multiple triple-stores and OSGi compliance.

6.14 WebOntoStudio

Vendor	Semafora Systems GmbH
Source	http://www.semafora-systems.com/en/products/ontostudio/
Software version	Onto studio and Web onto studio

OntoStudio is the most widespread commercial modelling environment for creating and maintaining ontologies. It stands out due to its comprehensive functions in intuitive ontology modelling. OntoStudio is also able to import many structures, schemas and models. Some of OntoStudio's most important functions are the mapping tool, which can be used to match heterogeneous structures quickly and intuitively, the graphic rule editor which specialists can use to model complex correlations or the integrated test environment that assures the quality of the modeling at all times.

With the help of OntoStudio, several editors can use the OntoBroker Collaboration server to simultaneously create and enhance ontologies. The queries created can be exported as a Web service and integrated into any applications.

Functions:

- The easy connection of databases and knowledge bases using a graphical mapping tool.
- The export of self-provided queries to the ontology as a Web service Enhanceable with additional plug-ins.
- The editing of OWL, RDF(S), RIF, SPARQL and ObjectLogic ontologies.
- The collaborative development of ontologies using the OntoBroker
- Enhancement Collaboration server.

Web OntoStudio is a lean version OntoStudio which can be easily used via every browser and is hence very flexible. It is ideal for large distributed teams who edit ontologies collaboratively.

Web OntoStudio offers you the following functions:

- Creation of classes and instances
- Definition of properties (attributes and relations) at schema level
- Assignment of values to properties at instance level

Web OntoStudio and OntoStudio both access the same ontology server. If users need to create rules that go beyond the Web OntoStudio functions, these can be created using OntoStudio and saved to the ontology server. All of the changes are immediately visible for all other editors.

In addition, there are the following import options:

- UML 2.0
- Database schemas (Oracle, MS-SQL, DB2, MySQL)
- Excel tables
- Outlook E-Mails
- Folder structures of the file system

6.15 WebProtégé 2.0

Vendor	WebProtégé is being actively developed by the Protégé team at the Stanford Center for Biomedical Informatics Research.
Source	http://protegewiki.stanford.edu/wiki/WebProtege http://protegewiki.stanford.edu/wiki/WebProtegeAdminGuide http://protegewiki.stanford.edu/wiki/WebProtegeUsersGuide
Software version	WebProtege build 103

WebProtégé is an open source, lightweight, web-based ontology editor. WebProtégé provides a friendly and highly configurable user interface that can be adapted for the use of domain experts. It has support for form-based editing and full-fledged collaboration.

- A Web-based application
    - edit ontologies in your Web browser
    - nothing to install
- Supports distributed editing
    - multiple editors can make changes at the same time
- Includes many collaboration features
    - discussion, watches, feeds

Useful features for collaboration:

- Tools for discussion and reaching consensus
    - Add notes to ontology entities (classes, properties, individuals, axioms)
    - Add reviews and change proposals anywhere in the ontology
    - Document the decision process and final decisions
- Complete Change history
    - Establish provenance
    - Retrieve ontology snapshots at any time
    - Implement different conflict resolution mechanisms
- Personalized views of an ontology based on:
    - User's role and tasks
    - User's level of expertise
- User roles and access control
    - Fine-grained control for editing and viewing rights
    - Sharing of ontologies
- Publishing released versions of an ontology in a central location, e.g. a repository
- Scalability, reliability and robustness

7 Appendix: Use cases

The Semicolon project has made a set of use cases as basis for the vocabulary tools. The use cases are briefly described in the chapters below.

The use cases are:

- Use case 1: Terminology /vocabulary maintenance at a government body
- Use case 2: Publishing terminology / vocabulary
- Use case 3: Register concept and its relationships
- Use case 4: Browse and view concepts from another government body
- Use case 5: Compare concepts

7.1.1 Use case 1: Terminology /vocabulary maintenance at a government body

Purpose of process	Maintain and publish a catalogue of concepts according to a set of governance and modelling principles. This supports a higher goal of open and efficient eGovernment to meet the need and expectations from citizens and private sector.
Sub -Processes	• Receive candidate concept or relation • Evaluate if concept is within domain, a new concept, an update, etc

	• QA of concept: Assure that concept is consistent with other relevant concept within the domain. • Suggested statuses: candidate, accepted, alternative concept, recommended used, not accepted, terminated, etc (align with ISO 11179) • Evaluate how other concepts, relations and usage of other concepts is influenced of this update. Use case 3: Register a concept and its relationships Use case 2: Publishing terminology /vocabulary Use case 4: Browse and view concepts from another government body Use case 5: Compare concepts
Could be initiated by	• Valid need from own organisation or change in responsibilities • Updated or new legislation or compliance demand • Open government and transparency • Configuration management of concepts • Software development
Roles	• Terminology steward • Concept coordinator • Legal responsible in organisation • Information architect • Service line owner

Functionality needed	• Create concept entry
	• Read concepts
	• Update concepts
	• Search, navigate, compare concepts
	• View concept and related attributes
	  ○ Single, group of, graph
	  ○ Change record for one, set of concepts
	• Export to other tools
	• Change status on concepts
	• Maintain group consistency of concepts.
	  ○ Accept change only if changes to a whole set of concepts can be committed.

Output from process	• Concepts ready to be published as part of a model. The model management is based on a configuration management regime supported the vocabulary tool.

Dependency to other use cases, details in context, activity sequences, error handling etc is not described in these short version of the use cases.

7.1.2 Use case 2: Publishing terminology /vocabulary

Purpose of process	Transparency, openness and increased public understanding of the government body effort and goals.
Sub-Processes	Publishing and marketing of concepts
Could be initiated by	• Concepts ready to be publish • Request for transparency
Roles	• Terminology steward • Concept coordinator • Information architect

Functionality needed	• Publish concepts • Search, navigate, compare concepts • View concept and related attributes ○ single, group of, graph • Change record for one, set of concepts • Export concept to other tools
Output from process	Open accessible domain concepts

7.1.3 Use case 3: Register a concept and its relationships

Purpose of process	A prerequisite to enable maintenance and publishing of concepts used in a government body work processes, IT-systems and guiding of public users of eGov services.
Sub-Processes	
Could be initiated by	Valid need from own organisation and vocabulary maintenance process.
Roles	• Terminology steward • Concept coordinator • Legal responsible in organisation • Service line owner • Information architect
Functionality needed	• Search, navigate • View concept and related attributes ○ single, group of, graph • Create concept and link concepts together • Validate consistency of graph • Export to other tools • Enter comments on related concepts
Output from process	A new concept entered into a well-defined maintenance regime.

7.1.4 Use case 4: Browse and view concepts from another government body

Purpose of process	Find, build knowledge and make decisions on (i) concept alignment and harmonisation and (ii) eGov service development.
Sub-Processes	• Choose relevant domain and government body • Examine concepts
Could be initiated by	Need to understand a public agency opportunities and limitations related to eGov service development and governance.
Roles	• Terminology steward • Concept coordinator • Service line manager • Citizen or user of open data

Functionality needed	• Drilldown on a domain model for a government body
	• Search, navigate, compare concepts
	• View concept and related attributes
	  ○ single, group of, graph
	  ○ change record for one, set of concepts
	• Export to other tools

Output from process	• List of relevant concepts

7.1.5 Use case 5: Compare concepts

Purpose of process	Make high quality concepts with relations to other relevant concepts. Examine opportunities for merging and sharing data based on concepts search.
Sub-Processes	• Perform concept harmonization
Could be initiated by	• Increase quality in concept model. • Understand similarities, overlaps and errors in own model or how a domain model refer to another domain model. • Examine the fundament for sharing or merging data. • A need for aligning or harmonising concepts or the legal corpus the concepts are derived from.
Roles	• Terminology steward • Concept coordinator • Legal responsible in organisation • Information architect • Service line owner
Functionality needed	• Drilldown on a domain model for a government body • Search, navigate • View concept and related attributes ○ single, group of, graph • Visualize how concepts overlap, has similar properties, differ, etc. • Export to other tools
Output from process	A list of related concepts according to drilldown and search criteria.

8 Appendix: Candidate terminology/ vocabulary sources to be tested in the pilot

Suggested vocabularies to be tested in the pilot are:

- Vocabularies from e.g. Norwegian tax authority, The Norwegian Labour and Welfare Administration, Statistics Norway
  - Business register (ER)
  - Addresses
  - Person related concepts
  - Concepts used in the EDAG project

Additional vocabulary candidates:

- SSB vocabulary, as open data from www.ssb.no open data sources
- The UDI vocabulary hosted in Pool Party
- Norwegian directorate of health
- The Norwegian Labour and Welfare Administration information model concepts
- www.More.no concept repository for Norwegian Municipalities