DigiCULT

Towards a Semantic Web for Heritage Resources

Thematic Issue 3

May 2003

Information Society Technologies

Towards a Semantic Web for Heritage Resources

Thematic Issue 3

CONTENT



Introduction and Overview

By Guntram Geser

Philosophy in Discussion With a Philosopher

Function and focus

DigiCULT, as a support measure within the Information Society Technologies Programme (IST), will for a period of 30 months (beginning March 2002) provide a technology watch mechanism for the cultural and scientific heritage sector. Backed by a network of peer experts, the project monitors, discusses and analyses existing and emerging technologies likely to bring benefits to the sector.

To promote the results and encourage early take-up of relevant technologies, DigiCULT has put in place a rigorous publication agenda of seven Thematic Issues, three in-depth Technology Watch Reports, as well as the DigiCULT.Info e-journal, pushed to a growing database of interested persons and organisations on a regular basis. All DigiCULT products can be downloaded from the project Website http://www.digicult.info as they become available. The opportunity to subscribe to the DigiCULT.Info is also found here.

March 2003 saw the release of the first DigiCULT Technology Watch Report. This report covers the topics Customer Relationship Management, Digital Asset Management Systems, Smart Labels and Smart Tags, Virtual Reality and Display Technologies, Human Interfaces, and Games Technologies. Addressing primarily technological issues, it serves as a guide to what a heritage institution needs to consider when buying into one of these technologies.

In comparison with the Technology Watch Reports, the Thematic Issues focus more on the organisational, policy, and economic aspects of the technologies under consideration. They are based on the expert round tables organised by the DigiCULT Forum secretariat. In addition to the Forum discussion, they provide opinions of other experts in the form of articles and interviews, case studies, short descriptions of related projects, together with a selection of relevant literature.

Topic and challenge

This third Thematic Issue addresses the questions: What is the Semantic Web? What will it do for heritage institutions? And what is the role of certain languages, in particular XML and RDF?

In short, the Semantic Web vision proclaims a Web of machine-readable data which allows software agents to automatically carry out rather complex tasks for humans. Key to realising this vision is semantic interoperability of Web resources. Yet, such interoperability is not the primary goal of heritage institutions (and intelligent software agents are not readily at hand).

What the institutions are looking for are new ways of providing scholarly and non-expert users (e.g. school classes, lifelong learners) with access to their collections and related knowledge. This goal can be accomplished, for example, through online collections

and exhibitions that not only display objects and simple descriptions (drawn from metadata), but also allow for understanding relationships between objects (created by semantically interrelated metadata). The Semantic Web community promises to assist in achieving this goal, but the challenge for the heritage institutions would be to first implement the necessary data infrastructure.

The challenge for the Semantic Web expert round table was, or at least the DigiCULT Secretariat thought it was, not to run into a debate between 'theory' and 'practice'. In other words, between what academic Semantic Web scholars and what practitioners from heritage institutions think needs to be accomplished, what is feasible and affordable, and where to concentrate efforts. For the discussion, XML seemed to provide a good starting point. XML, on the one hand, is increasingly considered by heritage institutions as a key standard for publishing metadata on the Web; on the other hand, it is a major building block for the Semantic Web. It proved different, in a positive sense. In the discussion, wide use of XML was taken for granted, while the key area of interest that surfaced and was seen to be most fruitful to explore was... ontologies.

Overview

Setting the context for this Issue, the position paper looks into the requirements for achieving the goals of the Semantic Web, and assesses whether the available technologies will be able to deliver on what the advocates of the Semantic Web envisage, as well as whether the cultural heritage sector is in a position to take substantial steps towards semantic interoperability. It concludes with the argument that the sector is more likely to be left behind, due in particular to the fact that for the institutions the rewards for the necessary investments are still too nebulous.

Janneke van Kersen from the Dutch Digital Heritage Association, in her interview with the DigiCULT Journalist, suggests that, despite the cloudy Semantic Web horizon, there are medium-term benefits to be gained for heritage institutions in taking steps towards the vision. And she states that it is up to associations like hers, together with larger institutions, to take the lead in this, prove that proposed solutions work, and support smaller institutions in taking advantage of them. On the other hand, Nicola Guarino, in his interview, believes that reaching the 'real' Semantic Web lies in taking 'the fundamental route' of implementing generic ontologies, based on linguistics and logics, within the Semantic Web fabric. He also claims that even incremental progress along this path can have remarkable pay-offs.

Michael Steemson's summary of the Darmstadt Forum illustrates that the Semantic Web topic resembles a labyrinth, with currently no definite map or Ariadne's Thread at hand. Building on the many technologies the Forum participants mentioned as some of the labyrinth's angles, we have added to the summary a list of resources related to these technologies.

In an effort to raise the veil of mystery surrounding the Semantic Web, this issue includes an example from the sector on the implementation of semantic interoperability of metadata, combined with a primer that explains core building blocks such as XML, RDF and ontologies. While a detailed primer of, for example, RDF would alone exhaust the limits of this issue[1], the goal here is to deliver an 'all-inclusive' primer within the space permitted, with all the inevitable limitations this entails. The primer attempts to provide a general understanding of the Semantic Web architecture, without obliging the reader to wander through the long and perplexing corridors of language specifications.

Finally, we want to thank the Koninklijke Bibliotheek, National Library of the Netherlands, for their kind permission to use selected images from their collection of illuminated medieval manuscripts.[2] We hope you will appreciate the little narratives they represent within the overall fabric of this DigiCULT Thematic Issue.

[1] cf. F. Manola, E. Miller: RDF Primer (W3C Working Draft, 23 January 2003), http://www.w3.org/TR/rdf-primer/

[2] See their online collection of such images at: http://www.kb.nl/kb/manuscripts/, which offers advanced search and presentation features.

Position Paper

By Seamus Ross

Genesis – The Creation: Division of Light and Darkness

Tim Berners-Lee and his colleagues at W3C have recognised that the real benefits of the web-based information revolution will come from enabling the interoperability of content. The current generation of web delivery is, they have argued, designed for human users who struggle to make effective use of the billions of pages of information currently accessible. When we search for something at the moment, we sometimes discover suitable candidate information but more often than not this is far from being the case. More than this, the entire process of searching, discovery, and use is designed to be driven by humans. When we discover one piece of the puzzle we need manually to position that information so that it can help us to search out the next piece of the puzzle. We find that Darmstadt is near Frankfurt. Then we find that there are flights from Glasgow to Frankfurt, and there is a bus from Frankfurt Airport to Darmstadt. Then I search for timetables, make manual comparisons and decide which times best meet my requirements. In the Shangri-La that is the Semantic Web my 'agent' would recognise from its regular review of my diary that I needed to be at a meeting in Darmstadt on the 21st of January 2003 and it would search out the options, analyse the timetables, identify the optimum travel arrangements, book my non-smoking hotel accommodation, and order the taxi to take me to the airport. (It might even check the weather forecasts and warn me to bring particular types of clothing.) Certainly, to make this happen there has to be a fundamental shift in the way data, information, and knowledge are represented on the web.

The proliferation of web-based resources makes finding what you are looking for increasingly difficult. According to Internet user studies, in 1996 50% of Internet users reported spending time looking for information without finding it, but by 2002 only about 40% of users ended their 'searching sessions' unsuccessfully. At first glance we might conclude that web discovery tools have improved and/or the information searching skills of users have improved. Over the past seven years the quantity of content has mushroomed, the search tools have become more efficient, developers approach the use of meta-tags more effectively, and anecdotal evidence suggests that the searching techniques of users have become more sophisticated. We should continue to be surprised by the high failure rate and wonder why it remains proportionally so high as the numbers of users have grown to nearly 600 million. In reality, there is just too much content available. It is poorly described. It is not interconnected. Search engines themselves are blunt instruments. Most users of the web do not have very mature searching strategies and rarely use even the blunt instruments as effectively as they might. A solution is to make more of the information capable of discovery, interpretation, and reuse by automated information processing tools themselves. However the current ways content is represented on the web makes it nearly impossible for machines to search the web meaningfully and effectively – even with the limitations of their skills and tools humans are better at searching the web than the most powerful of the current generation of agents. The emergence of the Semantic Web would solve this problem.

The web has made us realise the tremendous potential of digital resources and made them widely available. Content as presented on the web currently is mute. By adding descriptive information to content and resources, and representing both the descriptive information and the content in well-defined, consistent, and structured ways, 'mechanised agents' could be enabled to use web information 'intelligently'. Tim Berners-Lee, Jim Hendler and many other researchers believe that commercial and public sector institutions are increasingly recognising the benefits of ensuring that their content is adequately represented so that it is visible and discoverable within the context of the Semantic Web.

The Semantic Web will enable the heritage sector to make its information available in meaningful ways to researchers, the general public, and even its own curators. The public will be able to plan visits to institutions by, for example, dynamically relating opening times to public transport schedules. Use information to discover whether or not that Vase in the attic or basement is really Ming as their grandmother claimed by comparing it to the holdings of heritage institutions across the world. Curators will benefit from the ability to define an exhibition and have the entire process from the identification of the pieces to be shown in the exhibition to the production of the catalogue and publicity material automatically handled by their 'exhibition agents'.

Towards an Interoperable Semantic Web for Heritage Resources

Delivering the Semantic Web to the heritage sector depends upon (a) the syntactical and semantic mark-up of content, (b) the development of better knowledge analysis and modelling tools, (c) widespread adoption of interoperable knowledge representation languages, and (d) the construction of suitable ontologies. In most of this the heritage sector is lagging behind. We have not yet successfully represented sufficient quantities of our data in ways that makes it accessible to human web users, let alone in ways that would make it feasible for 'mechanised agents' to reason about in meaningful ways. 'Languages for representing data and knowledge are an important aspect of the Semantic Web' (Klein, 2001: 26). The languages that are currently the focus of the most substantial discussion, such as the RDF, DAML+OIL, and OWL[1] do not necessarily provide a suitable framework for delivering the Semantic Web. This point has been increasingly argued in the literature although in practice we still tend to emphasise the possibilities of representation mechanisms such as RDF(S) because it provides a flexible and extensible mechanism to represent metadata. A debate is raging about which language should be used to represent semantics on the web. Resource Description Framework (RDF), an XML based mechanism for expressing metadata, has been put forward at the basic level, but there is a growing body of opinion that indicates it does not have the richness that is necessary to make a suitable language. One of its shortcomings is that it cannot support syntax. In response other languages such as DAML+OIL have been developed. As an indication of the current levels of flux, in a fundamental paper, Patel-Schneider and Siméon from Bell Labs Research remark that '...there is a semantic discontinuity at the very bottom of the Semantic Web, interfering with the stated goal of the Semantic Web: If Semantic languages do not respect World-Wide Web data, then how can the Semantic Web be an extension of the World-Wide Web at all?' (2002a, 147).

Genesis – The Creation: Division of the Waters Above and Below the Firmament

The strength of XML is that it does not, itself, constrain how the data will be interpreted. While XML does not imply a specific interpretation of the data, how the material is marked up does constrain how it can be used. Fallside (2001) has made plain the weaknesses of using DTDs as a way of specifying semantic properties in XML (eXtensible Markup Language). XML Schemas offer a solution to these weaknesses especially where those weaknesses arise from representational problems. On the other hand the hierarchical nature of XML does not fit all domains, it 'does not encode the data's use and semantics' and DTDs and XML Schemas do not specify the data's meaning although they do specify the names of elements and attributes. Will the Semantic Web produce different levels of sophistication in the representation of data and knowledge in the web-world? If it does will this create a patchy representation of web information that will make the Semantic Web of limited value?

[1] See the 'Semantic Web Terms and Reading List' in this Issue.

Ontologies – the Jewels of the Semantic Web

For the Semantic Web to succeed it will require not only modelling languages, such as XML, RDF and OWL, but it will also require methodologies for extracting and defining the knowledge that is to be represented. Decades of research and commercial attempts to exploit the knowledge-based systems have demonstrated the complexity of knowledge modelling. Until there is such a methodology the possibilities of XML (or any other technology) as a knowledge representation language will not be achieved.

The success of the Semantic Web will depend heavily upon the creation of suitable ontologies. To avoid adding new variants to definitions we will follow James Hendler's definition of ontology as 'a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic' (Hendler, 2001: 30). One of the major hurdles facing us in building the Semantic Web is the lack of suitable ontologies. Languages such as OWL enable ontologies to represent 'class taxonomies' and provide mechanisms to enable their rapid development. For example, concepts and relationships can be established, such as 'watercolour is a type of painting' or 'a necklace is a type of jewellery'. But what about their multilingual capabilities? An ontology may well know that a 'watercolour is a painting', but it does not necessarily mean that it knows that an 'aquarelle is a type of painting' or that a 'watercolour is a type of peinture'. In addition, and probably first we need to consider:

- Can we cost the creation of appropriate ontologies for the heritage sector?
- How can we prioritise the ontologies that are needed? (e.g. which ones should the heritage sector develop and which ones will we be able to borrow from other sectors?)
- What heritage-based organisations should focus on ontology creation?
- Ontologies often fail to be interoperable. What solutions are there to this problem and how can they be made to work effectively?
- Does OWL (W3C's Web Ontology Semantic Markup Language for publishing and sharing ontologies) provide a suitable mechanism for ontology creation for the heritage sector?

Gomez-Perez and Corcho (2002) in an analysis of 'Ontology Languages for the Semantic Web' found that the measure of expressiveness in the current generation of ontology creation languages is a spectrum from XOL, RDF(S), SHOE, OML, OIL to DAML+OIL at the richest end of the scale. Indeed in their experience, while any of these languages will work for very simple ontologies, any attempt to use a weak language to create a complex ontology will fail.

Proof and trust is emerging as another central issue. How do we know that what our agent has discovered through its trawl of the Semantic web can be trusted. Even in the case of ontologies how should we decide whose ontology to trust? This is especially important where the two ontologies may conflict with one another. Similarly we are faced with the difficulties of ensuring and maintaining semantic integrity and a lack of methods for testing its presence.

Legitimising the Semantic Web Investment

Heflin and Hendler (2001) make the valuable proposal that semantic markup should be seen as one aspect of webpage design. This in their view would go a long way to ensuring that the costs of this mark-up (and the underlying information analyses that is necessary to make it happen) were met at the appropriate stage of process of putting material up on the web. However, Heflin and Hendler's proposal that semantic mark-up should be embedded into web-page design fails to recognise that the fundamental fabric of the web is changing. For this to happen we need a stronger argument for the benefits that such investment will bring to the heritage sector.

Haustein and Pleumann (2002) have noted that the successful development of the World Wide Web benefited from two factors: 'Participation was simple, and the results of effort were immediately visible to the creator'. As they argue, while these two success criteria, best classified, in my view, as ease of use and instant gratification, were characteristic of the WWW, they are not embedded into the fabric of the Semantic Web. The Semantic Web is hard and rewards are neither immediate nor assured. While in the long term it may bring tremendous benefits, the near-term take-up will be slow.

At least three other factors contributed to fostering the success of the web. Firstly, the early web developments concentrated on content creation and not on the creation of representation languages. The initial instantiation of HTML was simple, but it worked and material tagged using it remained accessible. We were not forced to dispose of work that we had 'webised' unless we wished to replace its representation with more sophisticated ones. Secondly, the value of the content that we put up increased as more users put up content of their own because the additional content attracted more users. Thirdly, to benefit you did not need to generate a lot of content, a very little would do and you could incrementally add more later. Slowly heritage institutions found ways to take advantage of the opportunities offered by the web, there are still many small and medium sized heritage institutions that have not.

Indeed the heritage sector is likely to be left behind because the financial rewards for creating the mark-up necessary to make the Semantic Web a reality are only evident to the commercial sector. There can be little doubt that the access to and understanding of the heritage would benefit from a world in which the vision of the Semantic Web were realised. But this is not the first information technology for which the benefits were promising. Even very simple strategies such as the use of databases to enable collection description have been shown over a period of nearly thirty-five years to bring benefits to the heritage sector institutions through better knowledge about, care of, and access to their collections. In the ALM sector only libraries can be said to have fully taken advantage of the technology to describe their collections and even here a close look shows that this has not covered all their holdings and not every institution. For instance, few libraries in the UK have online catalogues of their pre-1700 items and almost none have accurately described their photographic holdings at anything deeper than collection level. The same can be said of museums where descriptions are limited, except of course at the major institutions. In 1997 a survey in the United Kingdom showed that small and medium-sized institutions were struggling to participate in the computer-based description of their holdings. This was even before they considered putting the output of those holdings online. I would argue that this should hardly be surprising as the heritage sector has already been left behind in the development of online information in the web-world. Too few institutions have too little visible content that is actually usable. If the heritage sector is to make a near term contribution to the development of the Semantic Web it is going to be very moderate. It is very unlikely that developments will be related to reasoning about the heritage in the ways considered by Amann, et al. (2002). The ALM sector is more than likely to participate in the development of the Semantic Web through the creation of semantic mark-up of information about access arrangements, such as opening hours and details of facilities. This information is more likely to be useful to the tourism agent described by Tim Berners-Lee in his 2001 *Scientific American* article. While this may be a very positive way of integrating the heritage into the semantic web it does not maximise the potential benefits.

Genesis – The Creation: Division of Sea and Earth, Creation of Trees and Plants

The days when a curator who wishes to hold an exhibition on the representations of Salome since the 15th century will be able to 'load' an agent with the request to identify, select, negotiate the loan of and arrange the transportation of the key 100 works of art are a long way off. The fundamental descriptions of holdings are not currently available, where they exist they are not online, and certainly have not been semantically encoded to make them usable by our Salome agent. For those who have worked on Knowledge Representation the vision of the Semantic Web holds promise. Knowledge Representation is hard, especially if you intend any particular representation to be usable by others either as a decision making resource or as for research purposes. Efforts in the 1980s and early 1990s, such as those in archaeology failed. The reasons Knowledge Representation failed to achieve its promise ranged from the poor quality of knowledge extraction strategies, the lack of fundamental representation methodologies, the limited applicability of methods to knowledge domains, the problems of boundary constraint and creep, to the high costs of developing applications. The Semantic Web could breathe new life into this earlier promise by providing ways to carve up the problem while bringing us immediate successes.

Conclusion

Over the next five years the possibilities offered by the Semantic Web will bring little near term benefit for the heritage sector unless that sector co-ordinates its efforts to ensure that the fundamental building blocks that are necessary for the Semantic Web to be a success are put in place. We need some quick wins. A quick win involves identifying a domain that every institution can be encouraged to represent semantically and the placing in the public domain of a 'personalisable agent' that can take advantage of these semantics. Three factors could underpin a quick win: (a) a narrowly restricted knowledge domain of real public value; (b) an accessible and narrow ontology; and (c) a personalisable tool for processing knowledge.

Ultimately the same factors that constrain the heritage's sectors ability to take full advantage of the Web will constrain the penetration and pervasiveness of the Semantic Web in the heritage sector. The success of the Semantic Web in the heritage sector depends upon its adopting a XML based approach and a significant experiment that demonstrates its benefits to the wider community. Even for all its weaknesses the Semantic Web offers a tantalising solution to the problem of information overload created by the web and the heritage sector needs to address how it can take advantage of the opportunities it offers.

Bibliography:

Amann, B., Beeri, C., Fundulaki, I., and Scholl, M., 2002: Ontology-Based Integration of XML Resources, in: I. Horrocks and J. Hendler (eds.), The Semantic Web - ISWC 2002. Berlin: Springer, 117-131.

Berners-Lee, T., Hendler, J., and Lassila, O., 2001: The Semantic Web, in: *Scientific American*, May 2001.

Candan, K.S., Liu, H., and Suvarna, R., 2001: Resource Description Framework: Metadata and Its Application, in: *SIGKDD Explorations*, 3.1, 6-19, http://www.acm.org/sigs/sigkdd/explorations/issue3-1/candan.pdf

Doan, A., Madhavan, J., Domingos, P., and Halevy A., 2002: Learning to Map between Ontologies on the Semantic Web, in: *Proceedings WWW2002*, 7-11 May 2002 (Honolulu), 662-673.

Gómez-Pérez, A., and Corcho, O., 2002: Ontology Languages for the Semantic Web, in: *IEEE Intelligent Systems*, January/February 2002, 54-60.

Haustein, S., and Pleumann, J., 2002: Is Participation in the Semantic Web Too Difficult?, in: I. Horrocks and J. Hendler (eds.), The Semantic Web -ISWC 2002. Berlin: Springer, 448-453.

Heflin, J., and Hendler, J.: A Portrait of the Semantic Web in Action in: *IEEE Intelligent Systems*, March/April 2001, 54-59.

Hendler, J., 2001: Agents and the Semantic Web, in: *IEEE Intelligent Systems*, March/April 2001, 30-37.

Hendler, J., Berners-Lee, T., and Miller, E., 2002: Integrating Applications on the Semantic Web, in: *Journal of the Institute of Electrical Engineers of Japan*, 120.10, 676-680.

Klein, M., 2001: XML, RDF, and Relatives, in: *IEEE Intelligent Systems*, March/April 2001, 26-28.

McGuinness, D. L., Fikes, R., Hendler, J., and Stein, L. A., 2002: DAML+OIL: An Ontology Language for the Semantic Web, in: *IEEE Intelligent Systems*, September/October 2002, 72-80.

Patel-Schneider, P., and Siméon, J., 2002a: Building the Semantic Web on XML, in: I. Horrocks and J.Hendler (eds.), The Semantic Web –ISWC 2002. Berlin: Springer, 147-161, http://www-dbs.cs.uni-sb.de/lehre/ss03xml-seminar/Material/PS02.pdf.

Patel-Schneider, P., and Siméon, J., 2002b: The Yin/Yang Web: XML Syntax and RDF Semantics', in: *Proceedings WWW2002*, 7-11 May 2002 (Honolulu), 443-453.

Renear, A., Dubin, D., Sperberg-McQueen, C.M., and Huitfeldt, C., 2002: Towards a Semantics for XML Markup, in: *DocEng'02*, 8-9 November 2002 (McLean, VA). ACM Publication.

Shah, U., Finin, T., Joshi, A., Cost R. S., and Mayfield, J., 2002: Information Retrieval on the Semantic Web, in: *CIKM'02*, 4-9 November 2002, 461-468.

Development of the Semantic Web Must Begin at the Grass Roots Level

An interview with Janneke van Kersen, Dutch Digital Heritage Association, The Netherlands

By Joost van Kasteren

To be successful, the Semantic Web for the cultural heritage sector will have to develop from the grass roots level. A top-down approach whereby institutions have to squeeze themselves into a certain format is not going to work.' Janneke van Kersen has strong views on the initiatives that are currently being undertaken to develop a Semantic Web. 'They are not going to work for the cultural heritage institutions if you do not take into account the position that they are in. Especially not if the institutions are forced to overhaul their digitisation projects completely.'

Kersen graduated in Art History and did a postgraduate course on Historical Information Processing. Since 1999 she has been a consultant with the Dutch Digital Heritage Association (Vereniging DEN), which supports cultural heritage institutions large and small in developing strategies to face the digital future.

The key objectives of the DEN are to assist institutions in digitising and documenting their collections according to high quality standards, and assuring cross-domain and cross-institutional access to heritage information in a context-rich, structured environment. The methods used to realise these objectives are: knowledge dissemination, best practice and standardisation. The Association propagates open standards like XML, OAI and Dublin Core (qualified).

The DEN has approximately 60 member institutions, among them most of the large heritage institutions of The Netherlands. It provides access to the databases of the member organisations through the portal http://www.cultuurwijzer.nl. The Cultuurwijzer (culture pointer) to the collections uses the Aqua Browser to search for terms in a non-hierarchical associative way. Databases can also be accessed through subject fields based on the Dublin Core standard. Research is carried out to apply the Art and Architecture Thesaurus in a post-coordinative way, using it as an additional search aid.

Kersen: 'The Dublin Core has some drawbacks but it is one of the few international standards available for exchange of information. Mapped to the 5 Ws: who, where, what, when and why, it turns out to be a nice tool for interoperability across the databases of heritage institutions. Of course, we have to accept a certain kind of fuzziness and lack of precision compared with domain-specific access at the institutional level.'

According to Kersen, a real Semantic Web is still a long way off. 'We simply do not have the tools yet for a meaningful exchange and representation of information. XML and RDF do not provide the interoperability that is needed. On the other hand, I do not believe in developing a fundamental ontology to give meaning to information on the Net. It looks to me like the 18th-century endeavour to write an encyclopaedia that contains all the knowledge in the world. I am afraid it does not work that way. A lot of knowledge, even scientific knowledge, cannot be described in a logical way. Especially in the arts a lot of "knowledge" is the result of heuristics and associative thinking. Apart from that there is the practical problem that cultural heritage institutions do not have the money and the staff to describe their

collections anew in a way that fits the ontology. That is just too much work.'

Kersen thinks the Semantic Web will grow gradually from the grass roots level onward. Within the Dutch Digital Heritage Association she can point at several initiatives. Considered on their own, they might not seem like much, i.e. in relation to the gargantuan task of developing a Semantic Web, but when they are combined a certain pattern begins to emerge.

There is, for instance, a project, Sitechecker/RDF, which will look into ways in which RDF can be used for describing content for Web-based delivery. Furthermore, several standardisation projects are running that enable the participating institutions to develop the description of their specific knowledge domain, e.g. graphic domain, religion, art history. The formal and semantic mapping schemes used in these projects will include Dublin Core, Encoded Archival Description (EAD), IMS Learning Resources Metadata Specification, Art & Architecture Thesaurus as well as the CIDOC reference model. Kersen: 'At the moment, most of the reference terms are developed at the level of institutions, which means their use is limited to a certain domain. Or, to put it another way, every domain is developing its own dialect. Maybe the development of a combined reference scheme will be a step towards a Semantic Web.'

Another important project is the development of a scheme for description at the collection level in order to offer a clearer and more hierarchical access to heritage collections. This project has its roots in the Dutch project for collection level description, MUSIP (Museum Inventarisation Project). The description scheme will be broadened to make it available for other heritage institutions as well, for example archives.

The description and results of these projects and the programme lines will be made available on the Web site of the Vereniging DEN: http://www.den.nl. As Cultuurwijzer is used as a proof of concept, the results will be directly accessible at http://www.cultuurwijzer.nl. (Kersen kindly invites interested parties to put questions directly to their organisation.)

Projects are always carried out in co-operation with the member organisations. Kersen: 'We first try it ourselves until we are sure it works; a "proof of concept", you could say. These tests are overseen by a small group of automation experts working for our member organisations. The next step is to test the method on a larger scale with some of our member organisations. A larger working group oversees these tests. Only then is the method released to our member organisations. The advantage is that smaller member organisations can ride on the experience of the larger ones. By taking a step-by-step approach we also enhance the level of commitment. You could say we are providing some order in the information chaos that exists on the Internet. A few small steps along the long road towards the Semantic Web.'

Vereniging Digitaal Erfgoed Nederland: http://www.den.nl
Cultuurwijzer: http://www.cultuurwijzer.nl

DigiCULT's Expert 13 Tangle with the Semantic Web

By Michael Steemson

Genesis – The Creation: Birds and Fishes

It was really very kind of the Moderator. He asked the cultural heritage experts: 'What is the message that we should give our scientific writer who will do a write-up of this meeting for the Thematic Issue?' Their inclinations had not always been clear. But the question focused minds and they were certain, now.

'I would put my money on the Semantic Web', said two of them, not quite in unison. 'The Semantic Web is a direction, it is like North. You go North but you never arrive and say "here it is". This is the Semantic Web', said another.

The course of the debate at the Darmstadt DigiCULT Forum had not always been so direct. It had started with the Position Paper's dismaying thought that 'the limited understanding of information processing in the heritage sector almost makes the Semantic Web an impossibility to apply'.

It had touched on the semantics of Simeon poetry, art works of the biblical seductress Salome, weather forecasts for the northern English city of York, and the revolutionary theories of 16th-century Italian astronomer Galileo. But the experts agreed, finally, that the cultural heritage sector needed the Semantic Web... and that a good deal of education and guidance would be required to make it appreciate that need.

The experts numbered 13, lucky for some, at this, the third DigiCULT Forum of the European Union's technology watchdog for cultural and scientific heritage institutions. In the previous 12 months, other forum groups had discussed authenticity and integrity for digitisation programmes and, later, digital asset management systems. Now the Darmstadt 13 - historians, language and information technology scientists, academics and publishers - were looking even further down the information autobahn to the vision of WWW inventor, Englishman Tim Berners-Lee, who sees a new kind of automated Web that learns and understands each user's particular requirements and delivers complete, reliable, tested information sets.

In a co-authored May 2001 *Scientific American* article[1], Mr Berners-Lee imagines a family facing the horrors of re-scheduling its lives around a mother's unexpected illness. The sons and daughters rely on Semantic Web 'agents', small executable Web files, to search online medical records, hospital bed lists, transport timetables, doctors' appointment books, road condition reports and home diaries to find treatment, plan travel and re-arrange personal engagements to fit the emergency.

The vision requires huge world-wide investment in time and effort creating countless 'ontologies' containing, perhaps, XML (eXtensible Mark-up Language) and RDF (Resource Description Framework) data to which the electronic 'agents' could refer for understanding before applying to specially formatted Web pages for the information.

The Berners-Lee et al. dazzling forecast is: 'The Semantic Web will enable machines to comprehend semantic documents and data, not human speech and writings... Properly designed, the Semantic Web can assist the evolution of human knowledge as a whole.'

The Dazzling Prospects

Dazzling it is and the Darmstadt 13 were attracted. But they were not blinded. Moderator Dr Seamus Ross, the Director of Glasgow University's Humanities Advanced Technology and Information Institute (HATII), suggested that, while the current WWW content got a lot of 'bang' for its development dollars, the Semantic Web needed huge, expensive content before it could work well. Application of the Berners-Lee ideas to cultural heritage use was a long way off, he thought, and wondered: 'Is there enough benefit from the Semantic Web in the near term to make it a realisable dream 50 years down the road?'

Italian National Research Council Applied Ontologies Laboratory director, Nicola Guarino, has been working on the subject for 12 years and he knows the difficulties. He said: 'This is the ideal view which Tim Berners-Lee has: machines which work for you, your proxy which works for you, performing these dynamic connections for the Web which preserve meaning. It is pretty ambitious, but this is his idea. I would be happier if, rather than using an automatic proxy, we could just let people establish these dynamic connections using their brain and the Web. This is already something that is not easily done.'

Austria's Wernher Behrendt had encountered other snags. At Salzburg Research, the secretariat for the DigiCULT Forums, he co-ordinates another European Commission IST project, CULTOS (Cultural Units of Learning - Tools and Services). He conceded: 'There is a 50-year research vision behind the issue of the Semantic Web,' and went on, 'but there are incremental steps that, with good utility, can be built in a reasonable time. One of the intellectual challenges is to break the vision into these manageable steps.'

Language Representation Hitches

A CULTOS group had, Behrendt explained, taken one of these incremental steps and built an ontology[2] for digitised works of art. It had encountered problems with language representation like: 'Are there tools to support knowledge representation language? Are the users then actually able to work usefully with that? Can we incorporate the multimedia authoring component, where people who have not built the

[1] T. Berners-Lee, J. Hendler, O. Lassila: The Semantic Web. In: Scientific American, May 2001, http://www.sciam.com/2001/0501issue/0501berners-lee.html

[2] Ontology n. 1. Philosophy. The branch of metaphysics that deals ith the nature of being. 2. Logic. he set of entities presupposed by a theory. Collins English Dictio-ary, Third edition, Glasgow, 1991.

Cultural Units of Learning – Tools and Services (CULTOS)
CULTOS is an RTD project, co-funded by the European Commission under the Information Society Technologies (IST) Programme, which will run until October 2003. The application domain of CULTOS is intertextual studies in literature and arts. The project is developing a multimedia authoring and presentation environment that allows scholars to make the different relationships between cultural works explicit in a way that approximates contextualisation in interpretative processes. The result of the authoring processes is multimedia objects called 'intertextual cultural threads'. These are based on 'EMMOs', a novel type of structured multimedia object containing expert knowledge conforming to current and emerging standards such as XML/SMIL (with interactive extensions), MPEG-7 and RDF.
http://www.cultos.org

ontology themselves should then use it to combine multimedia assets with each other?'

The Dutch had doubts, too. Dr Janneke van Kersen is an art historian with Digital Heritage Netherlands (Digitaal Erfgoed Nederland, http://www.den.nl), where an XML-based content management system is being combined with a Resource Description Framework (RDF) to join databases from several cultural heritage institutions. She told the experts: 'I need to be assured that we will be able to build a layered structure that is equally applicable to each knowledge domain. Furthermore, I think that the cultural heritage sector is too much of a niche market to develop this.'

Her countryman, Dr Frank Nack, from the national research institute for mathematics and computer science, CWI (Centrum voor Wiskunde en Informatica, http://www.cwi.nl) in Amsterdam, works with a multimedia and human computer interaction group. His concern: 'Our group believes in the Semantic Web, but we needed some mechanisms to structure the information so that various groups can work with it. What we came up with was the belief that you can classify the user at a particular time. But that was simply not good enough.'

The group had found that users change their requirements widely and these shifts were invisible to a system. 'Humans can look at material one day and the next day they look at the same stuff differently and describe it differently because they are in a different mood', he said. He characterised the problem as: 'Now I would like to see something for my work, and now I want to be entertained. Which means I would like to access the information differently.'

He had one other worry: Webised mixed media. He said: 'This discussion has been heavily linguistic based which I can understand because most people do still think of the Web as text driven. The issue of describing various media items that are not text will, I think, very soon become important for the Semantic Web. We had better start thinking about that too.'

The Galileo Conundrum

The Institute and Museum of Science History in Florence, Italy, has tried to create an ontology around the works and sciences of its city's famous son, the revolutionary astronomer, mathematician and physicist, Galileo Galilei (1564-1642).[3] But it ran into difficulties when it came to the radical changes in theory that he created.

Institute relational database expert, Andrea Scotti, told the Forum: 'Galileo's scientific theory negates another scientific theory. This negating or development of theories was very difficult to represent in the ontology when dealing with the time factor. That is central to historical documentation but representational time was not part of the process available to us.'

Dr Costis Dallas, the Athens chairman of the European communication and technology group Critical Publics, thought the Florence museum's project was very ambitious. The time argument was difficult because 'of course, it isn't possible to represent time properly within a relational database'. But there were mechanisms - he mentioned software by the Virginia, US, IT group Telos[4] - that better represented issues of time.

Italian National Research Council's Nicola Guarino chipped in: 'The CIDOC[5] reference models have partial answers to these questions.'

Dr Dallas went on 'I do not believe that the whole exercise is futile but we found that in practice you cannot make a subject language for everybody. It has to be for a community of users. If you provide them with a richer representation, for instance if they can know that this is a person and this person lived in a place, then users will have a much richer experience.'

[3] The museum and Web site are rich resources for the life and work of Galileo, http://galileo.imss.firenze.it/
[4] Telos Corporation, http://www.telos.com/, Ashburn, VA, US.
[5] CIDOC: International Committee for Documentation of the International Council of Museums, http://www.willpowerinfo.myby.co.uk/cidoc/#CIDOCe (ICOM-CIDOC). Forum for documentation interests of museums and related organisations, one of 25 international committees of the International Council of Museums (ICOM).

Capitals and Acronyms

The Forum moved on to a discussion of the science and software behind the Semantic Web. Moderator Seamus Ross started by questioning whether 'key thinkers' were 'missing a fundamental point that Web pages are dead, that database-driven Web pages are the future, that people are going to stop making Web pages and make databases.'

He recalled that James A. Hendler, co-author with Tim Berners-Lee in the Semantic Web paper, and a professor in the Department of Computer Science at the University of Maryland, had recommended semantic representation as part of any Web pages. Dr Ross commented: 'This notion of the Semantic Web is not going to work with these databases. Is this true, or not?'

Bert Degenhart-Drenth, the managing director of Netherlands ADLIB Information Systems (http://www.nl.adlibsoft.com), thought it was true. But it was a practical problem that would be solved by projects like the Web Services of the Open Archives Initiative (OAI).

Several members spoke of difficulties created by dynamic Web pages generated from ASP databases. Paul Miller, the UK Interoperability Focus for the University of Bath's UKOLN (formerly the UK Office for Library Networking) project, said the UK's

Open Archives Initiative (OAI) Protocol for Metadata Harvesting

The Open Archives Initiative (OAI) is a mainly US-based group of people and organisations that evolved out of a need to increase access to scholarly publications through interoperable digital repositories. Support for the OAI's goals comes from the Digital Library Federation, the Coalition for Networked Information, and from a NSF Grant. One of its major achievements is an application-independent interoperability framework based on metadata harvesting: the OAI Protocol for Metadata Harvesting.

The OAI Protocol is based on the standard Web protocols http and XML, and employs Dublin Core (unqualified) as metadata standard. Heritage organisations who have systems that support the OAI Protocol can expose metadata about the content in their repository, i.e. allow service providers to harvest the data for services such as search engines. In the OAI Protocol, the XML schema is used at two levels: to define the format of responses to all OAI Protocol requests, and to define the format of metadata streams embedded in the GetRecord and ListRecords responses. In both cases, the goal is to provide a mechanism for data validation.

http://www.openarchives.org

The OAI-PMH, version 2.0, released in June 2002, can be found at http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

See also: John Perkins: A New Way of Making Cultural Information Resources Visible on the Web: Museums and the Open Archives Initiative. *Museums and the Web Conference 2001*, http://www.archimuse.com/mw2001/papers/perkins/

Open Archives Forum Project

http://www.oaforum.org

The Open Archives Forum Project is a two-year Fifth Framework Programme IST accompanying measure that will run until September 2003. The Forum is building a Web-based database on OAI-related projects, software, implementations and services, and supports the information exchange between OAI user communities.

Their surveys provide good insight into the status of uptake of the OAI in Europe. For an overview of the results, see: S. Dobratz, B. Matthaei: Open Archives Activities and Experiences in Europe. An Overview by the Open Archives Forum. In: *D-Lib Magazine*, Vol. 9, No. 1, January 2003, http://www.dlib.org/dlib/january03/dobratz/01dobratz.html

Recently, one of their workshops, 'Providing Access to Hidden Resources' (Lisbon, December 2002), targeted the libraries and archives communities. Requirements, standards, best practice, and solutions to interoperability problems of these communities were analysed and compared with the features provided by the OAI Protocol for Metadata Harvesting. The Tutorial 'OAI and OAI-PMH for Beginners' and other presentations can be found at: http://www.oaforum.org/workshops/lisb_programme.php

Office of the e-Envoy[6], the agency in charge of Britain's e-Government programme, drove it with a Web site called Govtalk[7]. 'One big ASP database', he called it. It carried the Government's interoperability standard framework with a URL of: '... a great, big, long string, something, something dot ASP'. That is how this key piece of legislation is referred to, and next week it will be something else!'

Dr van Kersen thought the new Web Ontology Language with the transposed acronym, OWL, would help. Italian ontology researcher Nicola Guarino discussed the European Commission's On-To-Knowledge project, its RDF tool Sesame[8], and the Ontology Inference Layer (OIL) language.

On-To-Knowledge
On-To-Knowledge is an IST RTD project that was completed in June 2002. The project developed tools and methods for supporting knowledge management in large and distributed organisations. The technical backbone of On-To-Knowledge was the use of ontologies for the various tasks of information integration and mediation. For the project's many results, see their tools repository, project deliverables and publications at:
http://www.ontoknowledge.org/pub.shtml
See also the On-To-Knowledge book 'Towards the Semantic Web. Ontology-driven Knowledge Management'. J. Davies, D. Fensel, F. van Harmelen (eds.). John Wiley, December 2002.

Dr Ross described the US Defense Department's DARPA (Defense Advanced Research Projects Agency) Mark-up Language (DAML) programme and its DAML+OIL variant. Mr Degenhart-Drenth highlighted the importance of protocols used in Web Services, such as SOAP (Simple Object Access Protocol) and UDDI (Universal Description, Discovery and Integration), and also pointed to a SPECTRUM XML standard for museums that he helped write. Then there was the moviemakers' audio-visual search standard MPEG-7 and the SMIL (Synchronized Multimedia Integration Language). No one mentioned SHOE (Simple HTML Ontology Extensions), which was surprising, as the discussion became more and more alphabetic and upper case.

Ontology tutorial in 800 words

Nicola Guarino brought the discussion back on track. The ontology expert delivered a fascinating, impromptu 800-word dissertation on ontology genetics.

Ontologies, he said, started because it was realised that controlled vocabularies, which worked well enough for limited periods, needed something extra to make them really useful. They needed clarification of intended meaning. This could be achieved in much the same way as dictionaries did it, by reference to other more basic terms. This was, he said, the key point.

'Ontologies can work if the basic terms are really used in a principled way. There is a hidden assumption here that it is, indeed, possible to express the meaning in terms of a relatively small set of primitive terms.'

He explained further: 'There are general terms that have a universal meaning. Take the term "part", for instance, or "set". Or take temporal relations, "before". Suppose I have two different periods, the Renaissance and another period, and suppose I say that this period comes "before" the other one, do you exclude the case whereby the two periods overlap or not? This is just a matter of stipulation; this is a general term that is not domain specific. I can simply stipulate exactly whether the "before" relationship between the two intervals includes the case of overlapping or not. And I can do that by means of axioms. Once you clarify the meaning of the basic terms, topological relations, mereological[9] relations, dependence relations, these kinds of things, then you have the basic vocabulary that helps you to introduce more domain-related things. And this is what people are doing in the area of what are called "Foundational Ontologies"... and this is what I am doing. I believe this is the only way to solve the problem of semantic interoperability. So, not just controlled vocabularies but vocabularies that are formally defined in minimal terms.'

Now it was clear, but would it be available to heritage institutions? Dr Ross wanted to know if fundamental ontologies of use to the heritage sector already existed. What would they be? Before any

[6] Office of the e-Envoy, http://www.e-envoy.gov.uk/
[7] Govtalk, http://www.govtalk.gov.uk/
[8] Sesame environment, http://www.ontoknowledge.org/tools/factsheet/Sesame.html
[9] Mereology n. The formal study of the logical properties of the relation of part and whole. Collins English Dictionary, Third edition, Glasgow, 1991.

development could begin, basic classifications and methodologies would be required to form a foundation for the work.

Italian online publisher, Marco Meli, a manager of the EU's MESMUSES (Metaphors for Science Museums) project, insisted: 'You need a clear definition in this particular domain. What are the key words, the terms?'

Mr Guarino answered: 'The key concepts and the key relations.'

The elusive Semantic Grail

By now, the Darmstadt 13 were beginning to realise they were not having much luck with their search for the Semantic Grail. Their discussion became a little tetchy.

Someone talked about 'meta-ontology', and another growled: 'The term "meta" is abused'. 'Outdated technical optimism' was mentioned. 'What is your alternative?' someone else wanted to know.

'I don't have one.'

'Is this the way forward?'

MESMUSES –
Metaphors for Science Museums

MESMUSES is an IST RTD project that will run until July 2003. It aims at designing a general method and supporting tools to produce knowledge maps for use in self-learning environments of science museums. In the project, a knowledge map is defined as a set of related concepts and facts that is offered to learners with some guidance or suggestions on possible itineraries that they may follow to explore the knowledge space.

The method and tools developed in MESMUSES are being tested and validated by two large science museums, the Cité des Sciences et de l'Industrie in Paris and the Istituto e Museo di Storia della Scienza in Florence, which provide access to their digital catalogues. Both museums are developing knowledge maps and itineraries on different themes in Biology (Genome) and Physics (Galileo and the laws of motion).

Project Web site:
http://cweb.inria.fr/Projects/Mesmuses/

See also: M. Meli: Knowledge Management: a new challenge for science museums. In: *Cultivate Interactive*, Issue 9, 7 February 2003, http://www.cultivate-int.org/issue9/mesmuses/

'What part of the problem would that solve?'

'We will know when we have tried that out.'

'Here we are approaching a scary field.'

Civility and peace were restored as Frank Nack, the CWI Netherlands scientist, introduced the thought: 'There are ontologies for art and they are very old and well crafted. There are very clear rules about why they did what they did because they have worked on them for a thousand years. What you could suggest is that we strip down to the basics for one field, say art, and apply it to all the other fields we have in cultural heritage, architecture, film, whatever, all working with very different substances.'

Seamus Ross added: 'So we need one fundamental ontology on which we can build all the others.'

The Luck Changes

The luck of the 13 was beginning to change. The Athenian heritage informatics expert Dr Dallas described work among his company clients on developing 'an upper ontology'. All the issues the Forum had been discussing, what to do about time, a basic concepts process, agents, and so on were being examined. They were beginning to develop 'something very much like a thesaurus' with term expansion that created sub-categories of relationships. He called it 'generic layering', a process that could identify the 'generic grammar' of relationships within a specific domain – art history, for example.

'This is useful', he said. 'This way, we can create Web systems that present an association of content for users that is meaningful to them. Let's say a "cultural" meaning.'

Nicola Guarino went further. He believed that the International Council of Museums' Committee for Documentation (CIDOC) Conceptual Reference Model (CRM) was the 'best starting point' for the heritage community. The CIDOC CRM is the result of 10 years' work by a standards working group. The

CIDOC CRM: 'The Semantic Glue'

The 'CIDOC object-oriented Conceptual Reference Model' (CIDOC CRM) was developed by the ICOM/CIDOC Documentation Standards Group. Since September 2000, the CIDOC CRM is being developed into an ISO standard.

'The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework to which any cultural heritage information can be mapped. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice in conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives.'

http://cidoc.ics.forth.gr/what_is_crm.html

CHIOS - Cultural Heritage Interchange Ontology Standardization project

Since June 2001, the work of the CIDOC CRM Special Interest Group has been supported by CHIOS, a two-year project which receives funding from the Fifth Framework IST Programme. The CHIOS consortium forms an integral part of the CIDOC CRM Special Interest Group which, by organising shared meetings, represents the interests and requirements of the cultural heritage community to the ISO Working Group (ISO/TC46/SC4/WG9).

http://cidoc.ics.forth.gr/chios_iso.html

model is under review for adoption as an International Standards Organisation (ISO) publication.

Mr Guarino was enthusiastic: 'I am not biased on this. I am a reviewer on the CHIOS project that supports this proposal for an ISO standard, and I am amazed by the fact that this standard is more or less principled. On the one hand the authors really strive to get these principled things, and on the other hand they have extensively accounted for existing practice. It is the result of a large community of work. It is not perfect but it really is a starting point for this community.'

He affirmed Dr Ross's delighted question: 'So this is an ontology we can borrow?'

And the Italian expert had more to add: 'The question before was "how can we be sure that the principled things can really solve our problems?". I do not have a crisp answer, but I do have some evidence that even a tiny result on the foundational side has a high pay-off. So you do not need to solve all the foundational issues.'

The distinctions between an object and its role or individual and classes of items were delicate but, once understood, could lead to significant data improvement, he said, adding: 'Take, for instance, the distinction between object and event. This is only one tiny distinction but it is so fundamental that, once you understand it, you can save time in developing your own application ontology. Tiny conceptual progress does have a high pay-off. This is why I believe it is useful.'

There were still one or two doubts, but CULTOS project co-ordinator Wernher Behrendt tidied up with a daring stance: 'Let me be a heretic for a second. How many of us have an operating system other than Windows? What I am saying is that standardisation often helps. Even if it is not the best standard, it does help get people working together... It is perfectly fair to define a standard for the world, now. There will be a lot of discussion but it will wind down to a few constructs. It is a better method of getting an ontology accepted than having ontologies mushrooming all over the place that must then be integrated.'

The Darmstadt 13 were pleased. They had a model. They had 'Web Services' stepping stones to work across. They weren't going to fall into the trap of insisting just yet that the Semantic Web was important for the heritage sector, but they wanted an education process for unconvinced curators.

They needed someone to make the first ontology move. Seamus Ross suggested asking the J. Paul Getty Trust (http://www.getty.edu). They needed automated tools for testing ontologies. They did not want to be delivered into the entertainment industry, but saw benefit in what University of Florence Associate Professor, Franco Niccolucci, called 'cultural entertainment'.

So, where would the experts put their money? asked Mr Behrendt. 'On there not being any benefits in the Semantic Web for the cultural heritage sector or there being some benefits in building such things, whatever they may be?'

Amsterdamer Frank Nack was in no doubt: 'It is going to happen. It will probably look very different from how we imagine it right now, but it is going to happen.' His countryman, ADLIB chief Bert Degenhart-Drenth, thought so too: 'We have put our money there. All our applications work with XML. ' And Dr van Kersen agreed: 'I would put my money on the Semantic Web.'

That seemed to make it game, set and match.

Semantic Web Terms and Reading List: A-X

Compiled by Guntram Geser and Michael Steemson

In the Summary we have provided information boxes on projects but not on the many Semantic Web standards, technologies, etc. mentioned in the Forum discussion. The projects included are only a small fraction of many ongoing activities and are related mainly to the cultural and scientific heritage community. Links to many important Semantic Web development projects can be found at their community portal: http://www.semanticweb.org

The following guide points to resources and readings on terms mentioned in the Forum Summary. It is not intended to provide a comprehensive list of Semantic Web materials. Rather it represents different entry points and levels to this topic.

Genesis – The Creation: Stars and Fishes

Annotation & Authoring
'How to annotate and have fun', http://annotation.semanticweb.org/faq/faq2, and other papers. See also http://annotation. semanticweb.org/tools. Institute for Applied Informatics and Formal Description Methods, University of Karlsruhe, Germany.

The Annotea Project, a 'Live Early Adoption and Demonstration (LEAD)' project of the World Wide Web Consortium (W3C) collaboration environment with shared annotations, www.w3.org/2001/Annotea/

CIDOC CRM
The CIDOC Conceptual Reference Model was an important reference point in the Forum discussion (see also the information box on the CIDOC CRM in the Summary). CIDOC, the Comité International pour la Documentation, is part of the International Council for Museums (ICOM). Its Web site can be found at: http://cidoc.ics.forth.gr
A report on CIDOC´s work on the CRM is provided in 'The CIDOC Conceptual Reference Model: A Standard for Communicating Cultural Contents' by Nick Crofts, Martin Doerr and Tony Gill, in: *Cultivate Interactive*, Issue 9, February 2003, http://www.cultivate-int.org/issue9/chios/
See also their tutorials at http://cidoc.ics.forth.gr/tutorials.html

DAML - DARPA Agent Markup Language
The Defense Advanced Research Projects Agency (DARPA) is the central research and development organisation for the US Department of Defense. Its DAML programme is developing a language and tools to facilitate Semantic Web concepts: http://www.daml.org

'Why Use DAML?' Adam Pease, white paper, *Teknowledge*, 10 April, 2002, http://www.daml.org/2002/04/why.html

DAML+OIL
DAML+OIL is a product of the DARPA Joint United States/European Union ad hoc Agent Markup Language Committee. The committee created a language with the best features of SHOE, DAML, OIL and several other markup approaches. It is a Web ontology language (latest release, March 2001), expected to provide a basis for future Web standards for ontologies. See: http://www.daml.org/2001/03/daml+oil-index.html

OIL - Ontology Inference Layer
OIL is a proposal for a Web-based representation and inference layer for ontologies. It uses a layered approach to defining a standard ontology language. Each layer adds functionality and complexity to the previous layer. This is done such that machines that can only process a lower layer can still partially understand high-level ontologies. See: http://www.ontoknowledge.org/oil/

A white paper on OIL functions: 'An informal description of Standard OIL and Instance OIL', by the On-To-Knowledge group led by Department of Computer Science, University of Manchester, UK, 28 November 2000, www.ontoknowledge.org/oil/downl/oil-whitepaper.pdf

Ontologies
Ontology research: Laboratory for Applied Ontology (LOA), Institute of Cognitive Sciences and Technology (ISTC), http://www.ladseb.pd.cnr.it/infor/ontology/ontology.html

Dieter Fensel: Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. New York, Springer, 2001.

'Ontology Infrastructure for the Semantic Web', includes detailed subject bibliography. WonderWeb Project, Department of Computer Science, Victoria University of Manchester, UK, http://wonderweb.semanticweb.org/deliverables/documents/D15.pdf

Standard Upper Ontology (SUO): An upper ontology for data interoperability, information search and retrieval, automated inferencing, and natural language processing. IEEE Standard Upper Ontology (SUO) Working Group, http://suo.ieee.org/

OWL - Web Ontology Language
The Web Ontology Language is a semantic markup language for publishing and sharing ontologies on the World Wide Web. OWL is developed as a vocabulary extension of the Resource Description Framework (RDF) and is derived from the DAML+OIL Web Ontology Language. For the development of this language, see the documents of the Web Ontology (WebOnt) Working Group, http://www.w3.org/2001/sw/WebOnt/

See also OWL Web Ontology Language Reference (W3C Working Draft 31 March 2003) at http://www.w3.org/TR/owl-ref), and OWL Web Ontology Language Guide (W3C Working Draft 31 March 2003), http://www.w3.org/TR/owl-guide/.

For an understanding of the goals, requirements and usage scenarios for a Web ontology language, see: 'Web Ontology Language (OWL) Use Cases and Requirements' (W3C working draft, 31 March 2003), http://www.w3.org/TR/webont-req/

RDF - Resource Description Framework
See Cultural Heritage Semantic Web Example & Primer, pp. 32-34.

Semantic Web
'The Semantic Web', Tim Berners-Lee with James Hendler and Ora Lassila, *Scientific American*, May 2001, http://www.sciam.com/2001/0501issue/0501berners-lee.html

'Enhanced: Science and the Semantic Web', J. A. Hendler, in: *Science magazine*, Volume 299, Number 5606, 24 January 2003, pp. 520-521.

'Peer-to-Peer: The Infrastructure for the Semantic Web', Stanford University. The Semantic Web as the next evolutionary step of the Internet, http://p2p.semanticweb.org

The Semantic Web Community Portal: SemanticWeb.org, currently operated by three research groups: The Onto-Agents and Scalable Knowledge Composition (SKC) Research Group at Stanford University, the Ontobroker-Group at the University of Karlsruhe, Germany, and the Protégé Research Group at Stanford University, http://www.semanticweb.org/

The W3C Semantic Web Activity Statement explains the consortium's plans in the areas of enabling standards (driven by the RDF Core and Web Ontology Working Groups), education and outreach (RDF Interest Group), as well as co-ordination and advanced development, http://www.w3.org/2001/sw/Activity/

See also Kim Veltman's warning of what he sees to be too narrow a definition of the Semantic Web, one that will not allow the historical dimension, the richness of cultural expression, the unique, and the diversity of interpretations to be adequately dealt with. Cf. K. Veltman: Challenges for a Semantic Web (July 2002), http://www.cultivate-int.org/issue7/semanticweb/

SHOE – Simple HTML Ontology Extensions

SHOE was one of the first ontology-based markup languages developed for use on the World Wide Web. It is a small extension to HTML that allows Web page authors to annotate their Web documents with machine-readable knowledge. See: http://www.cs.umd.edu/projects/plus/SHOE/

SMIL - Multimedia on the Web

The Synchronized Multimedia Integration Language (SMIL, pronounced 'smile') enables authoring of interactive audiovisual presentations. SMIL is typically used for 'rich media'/multimedia presentations which integrate streaming audio and video with images, text or any other media type. SMIL is an HTML-like language and may be written using a simple text editor. W3C: Synchronized Multimedia, http://www.w3.org/AudioVideo/

SPECTRUM-XML DTD

SPECTRUM (Standard Procedures for Collections Recording Used in Museums) was created by the mda (http://www.mda.org.uk). It is a guide to good practice for museum documentation that describes procedures for documenting objects and the processes they undergo, as well as the necessary information that needs to be recorded to support the procedures. For SPECTRUM, an XML Document Type Definition has been produced which serves as a system-neutral interchange format for museum data.

'SPECTRUM: The UK Museum Documentation Standard' is available in its second edition; see: http://www.mda.org.uk/spectrum.htm

For a description of the creation, structure and deployment of the SPECTRUM-XML DTD, see: Bert Degenhart-Drenth: Building on the mda SPECTRUM-XML DTD for Collections Management Data Interchange. *Museums and the Web Conference 2001*, http://www.archimuse.com/mw2001/papers/degenhart/degenhart.html

Web Services

Although the concept of Web Services is not a new one, the definition of Web Services standards allows the wider use of these services. Currently three main protocols are used in the context of Web Services: UDDI (Universal Description, Discovery and Integration), a registry system to find resources and Web Services; WSDL (Web Service Description Language), an interface description language; and SOAP (Simple Object Access Protocol), the communication protocol for Web Services. All three protocols are based on XML.

For a description of the many ways in which XML can enhance Web Services, see: http://www.w3.org/2002/ws/Activity

World Wide Web

Looking back as well as into the future: 'Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor'. Tim Berners-Lee, with Mark Fischetti. Harper, San Francisco, 1999.

XML – eXtended Mark-up Language

See Cultural Heritage Semantic Web Example & Primer, pp. 27-30.

For the work done at W3C within the XML activity, see: XML Working Groups, http://www.w3.org/XML/; the XML Specifications can be found at http://www.w3.org/XML/Core/

'The bane of my existence is doing things that I know the computer could do for me.' – Dan Connolly, The XML Revolution, October 1998. http://www.nature.com/nature/webmatters/xml/

Semantic Web Should be Based on Well-founded Ontologies

An Interview with Nicola Guarino, Laboratory of Applied Ontology, Italy

By Joost van Kasteren

'The Semantic Web as it is advocated by people like Tim Berners Lee and James Hendler does not take enough advantage of the experience built up in knowledge engineering and conceptual modelling. There is this anarchistic idea of the Web as a place where everyone can do his or her own thing. I have no problem with that; a lot of people are able to find what they want on the Web. But if you want real interoperability, with search engines that can grasp the intended meaning of information, that approach falls short. To create a real Semantic Web we have to develop and use well-founded generic ontologies, based on linguistics and logics.'

Nicola Guarino has clear views on the Semantic Web and its development. He is a senior researcher at the Institute for Cognitive Sciences and Technologies in Italy, where he leads the Laboratory for Applied Ontology. Since 1991 he has played an active role in the Artificial Intelligence community in promoting the interdisciplinary study of ontological foundations of knowledge engineering and conceptual modelling. Guarino: 'In our Laboratory the focus is on content and not so much on representation. The use of ontologies is unavoidable when referring to content. People do it implicitly all the time when they are communicating and trying to understand each other. If we want machines to understand each other, in other words real interoperability, we need to make these ontologies explicit in an unambiguous way.'

An ontology is a hierarchical description of the relations between concepts in a certain domain plus an unambiguous description of the concepts themselves. As they are created for a certain domain, ontologies often fail to be interoperable, because of the ambiguity that results from the use of the same terms for different concepts (and vice versa) between different domains. The term 'net' for instance has quite a different meaning for Web designers and fishermen. That is why there is a need for well-founded generic ontologies. An example of a generic ontology is the term 'part', which can have different meanings both within a domain and between domains. For instance, the violist plays a part in the orchestra. His finger is part of him. Can his finger be part of the orchestra? According to Guarino, this is a genuine ontological problem that can only be solved by giving an unambiguous meaning to the term 'part'.

Another example, cited by Guarino, is the term 'in'. What exactly are you describing when you say the spoon is in the cup? Does it mean that the spoon is totally embedded in the cup or is it only partly in the cup? Guarino: 'These examples seem trivial, but if you want real interoperability between different knowledge domains you will have to prevent the problems that come with the ambiguity of day-to-day language.'

In this respect Guarino thinks it is a drawback that computer science curricula scarcely ever contain an introduction to ontological foundations of conceptual modelling. 'Students learn all about Java, HTML and C++ and name all the other languages, and they also learn how to use these. But when they graduate they hardly know a thing about formal ontology. I really think people should know more about the work on ontology that has been done in philosophy. It is certainly not much harder to acquire than, say, studying differential equations, or learning how to use Java.'

It seems as if it is an enormous job to develop well-founded generic ontologies, but it is not as enormous a task as it appears. Guarino: 'I would say that a few dozen would get you on the way nicely. But you have to take the fundamental route. At the moment development of the Semantic Web is driven by the need for short-term results. Hence, interoperability is realised by putting the right tags on the information. That is not what I call semantics; that is syntax. XML and RDF are very useful for this, but they fall short when you want to create a real Semantic Web.'

Laboratory of Applied Ontology, http://ontology.ip.rm.cnr.it

A Cultural Heritage Semantic Web Example & Primer

By Guntram Geser

The Semantic Web, according to a statement of its well-known advocate Tim Berners-Lee, is a vision of 'a distributed machine which should function so as to perform socially useful tasks'.[1] This machine should allow intelligent software agents to understand semantic relationships between Web resources in order to seek relevant information and perform transactions for humans.

Contrasted with the existing, human-readable Web, the Semantic Web is envisaged as a Web of machine-readable data that will be based on 'languages for expressing information in a machine processable form'.[2] Key to an understanding of the Semantic Web, therefore, is how these languages function, how information is expressed in order that computers can automatically process Web sources and assist in making the Web more useful for humans. The aim of this chapter is to provide an overview of the Semantic Web concept by describing its general architecture, i.e. the interplay of its languages.

The chapter has two interrelated parts. Part 1 describes a Finnish project that strives to build the foundations for the "Finnish Museums on the Semantic Web" (FMS), a future semantic museum portal. This part consists of the information boxes on the following pages, which briefly describe the necessary elements and steps in the set-up of the FMS system. It is recommended to start by reading this description (see also graphic 3 on page 36 which provides an overview of the set-up of the FMS system). It should be helpful in gaining a general understanding of how semantic interoperability of, and new ways of interacting with, semantically marked-up cultural heritage information can be realised.

Part 2, the texts below the information boxes, is a primer that explains terms used in part 1 which represent core elements of the Semantic Web architecture, as well as providing illustrative examples. The explanations are not intended to give in-depth definitions of these elements; such definitions are provided in the relevant W3C specifications. The examples have been kept as simple as possible but build on each other. In this way, we will develop a (fictitious) Website, http://www.m-i.org, that provides semantically enhanced access to such marvellous medieval images as the ones we have used to illustrate this Thematic Issue.

How to Make Collection Metadata of Museums Semantically Interoperable on the Web – The "Finnish Museums on the Semantic Web" (FMS)

The Semantic Web concept is visionary, and there are dedicated people, also in the heritage sector, who are trying to make it a reality. In our example, a group of researchers and technology developers, who work at the University of Helsinki and the Helsinki Institute for Information Technology, are translating the Semantic Web vision for a future semantic museum portal.

The group's two-year project will run until spring 2004, and is being carried out in co-operation with, and with funding from, major organisations including

[1] Tim Berners-Lee: Interpretation and Semantics on the Semantic Web (1998), http://www.w3.org/DesignIssues/Interpretation.html

[2] Tim Berners-Lee: Semantic Web Road Map (1998), http://www.w3.org/DesignIssues/Semantic.html

[3] Robert DuCharme, http://lists.xml.org/archives/xml-dev/ 200211/ msg00190.html

FMS Documents: A short presentation is provided in: Eero Hyvönen, et al.: Cultural Semantic Interoperability on the Web: Case Finnish Museums Online, tp://iswc2002.semanticweb.org/posters/hyvonen_a4.pdf; for detailed descriptions, see: Vilho Raatikka, Eero Hyvönen: Ontology-based Semantic Metadata Validation; and Hyvönen, Eero et al.: Semantic Interoperability on the Web: Case Finnish Museums Online. Both texts can be found in: Towards the Semantic Web and eb Services. Proceedings of the ML Finland 2002 Conference, http://www.cs.helsinki.fi/u/eahyvone/xmlfinland2002/ProceedingsXML2002-final.pdf

Espoo City Museum, Helsinki University Museum, National Board of Antiquities, Nokia, TietoEnator and the National Technology Agency (TEKES).

The major goals of the project are to make collection metadata, which stem from heterogeneous databases, semantically interoperable on the Web, and to provide facilities for semantic browsing and searching in the combined knowledge base of the participating museums.

The project's vision is called "Finnish Museums on the Semantic Web" (FMS), and its architecture allows for all Finnish museums to join in. However, in an approach of starting small but ambitious, the project is at present using the collection databases of two museums, the Espoo City Museum and the National Museum of Finland. Furthermore, the implementation is currently restricted to only one part of the collections - textiles.

In order to reach the FMS system's goal of making the museums' metadata semantically interoperable on the Web, the data must be harmonised on the syntactic and semantic level. For this harmonisation, the eXtended Markup Language (XML) and the Resource Description Framework (RDF) are being used, of which RDF is the key language for achieving semantic interoperability of the heterogeneous sets of metadata.

RDF and Metadata - A Natural Fit

An observer of the diffusion of the Resource Description Framework (RDF) into various domains, Robert DuCharme, has commented: 'I still find it a little ironic that while RDF has gotten so much publicity as a technology for warm and fuzzy AI (Artificial Intelligence) pie-in-the-sky technology, it's gotten most of its traction in the mundane world of metadata.[3]

Yet, given the importance of metadata for the Semantic Web vision in general, it does not come as a surprise that metadata of key information communities belong to the first of RDF's intended uses. RDF seems to gain momentum in particular among the library and other communities that use Dublin Core.

The actual W3C RDF Primer (Working Draft 23 January 2003), edited by Frank Manola and Eric Miller, labels RDF as 'an ideal representation for Dublin Core information' and describes Dublin Core as one of their 'RDF in the field' examples. (Cf. http://www.w3.org/TR/rdf-primer/)

At the Dublin Core Metadata Initiative (DCMI), 'Expressing Simple Dublin Core in RDF/XML' was announced as a DCMI Recommendation in October 2002, 'the first in a series of recommendations for encoding Dublin Core metadata using mainstream Web technologies', i.e. XML/RDF/XHTML. 'Expressing Qualified Dublin Core in RDF/XML' is currently a Proposed Recommendation. Cf. http://dublincore.org/groups/architecture/

Syntactic Transformation /1: Creating the XML Documents

In the FMS system, the eXtended Markup Language (XML) is used as the data transfer format. This transfer format enables the system to make use of the data originally stored in the museums' heterogeneous collection databases. Therefore, each museum participating in the FMS initiative provides the relevant collection data as an XML document repository.

In a process of syntactic harmonisation, the data from a museum's collection database are retrieved and transformed to an XML format conforming to the XML Schema of the FMS initiative.

The data to be published are read from the database through a 'view', which helps create the XML format. The view is a queryable interface, a virtual table that results from an SQL query, which may join multiple tables of the database. Through the view, the data are queried so that the rows of the tables are grouped by collection items. For each item, the set of rows is combined into a single XML document.

XML

XML is a markup language for describing data. It is a language created to allow anyone to design the structure of their own documents. An XML document contains text that consists of markup in the form of tags and plain text between them, the latter being just pure information (for example, <creator>Alexander Master</creator>). XML tags are not predefined; everyone can define his or her own tags.

XML shares the syntax and bracketed tags of the well-known HyperText Markup Language (HTML), but XML serves a different goal. While HTML is used to define the layout of pages on the WWW, XML is used to define the content of documents; for example, to specify that an area of text is the name of a creator.

XML allows for creating markup (e.g. <creator>) that seems to carry some semantics. However, for a computer a tag like <creator> carries as much semantics as a tag like <H1>. A computer simply does not know what a creator is and how the concept creator is related to other concepts (e.g. manuscript). For an XML processor, <H1> and <creator> or <manuscript> are all equally (and totally) meaningless. XML is all about describing data; on its own it does not do anything. There needs to be a processing program that uses the markup to interpret the various pieces of elements.

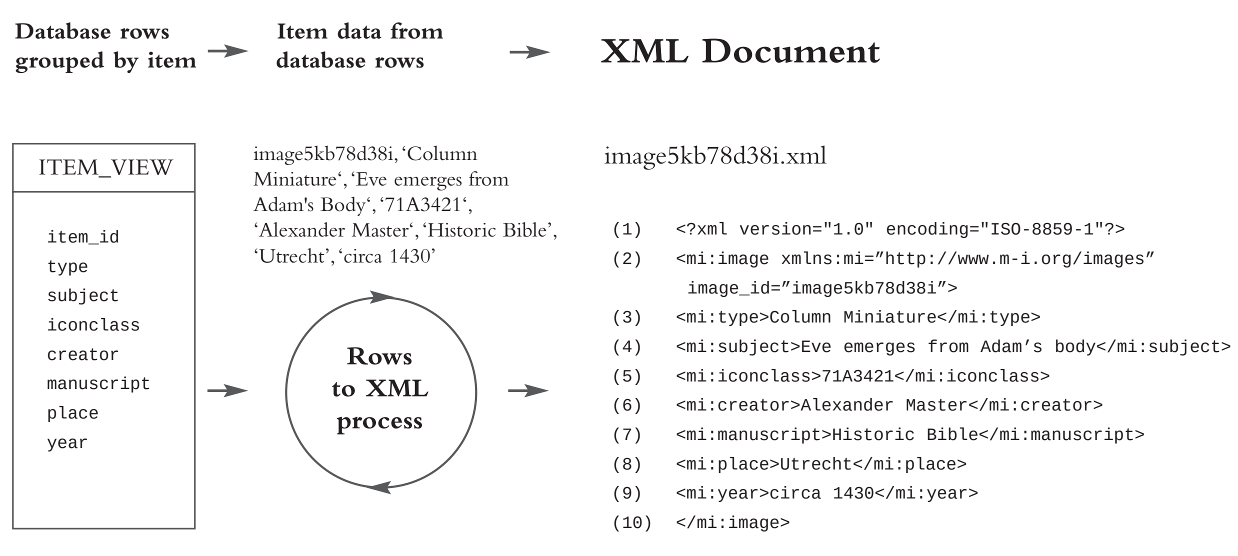
The graphic below illustrates the database rows to XML process, as described in the info box. It provides a very simple example of an XML document that describes some data for one of the medieval column miniatures from the Koninklijke Bibliotheek, The Hague, which we were permitted to use for illustrating this Thematic Issue. It includes the Iconclass classification for this image: 71A3421 Eve emerges from Adam's body (for the hierarchical path of this classification, see the section on ontologies).

Short explanations for the XML document (image5kb78d38i.xml) shown in the graphic below:

A well-formed XML document is one that conforms to the XML syntax rules, of which we would like to highlight the following:

(1) The document must begin with the XML declaration, which defines the XML version and the character encoding used in the document: In the example below we use <?xml version="1.0" encoding="ISO-8859-1"?>, i.e. the document conforms to the 1.0 specification of XML and uses the ISO-8859-1 (Latin-1/West European) character set.

(2/10 a) The XML document must contain a single tag pair to define a root element, in our example <mi:image> </mi:image>.



Graphic 1: Database rows to XML process

(2b) Namespace: Since element names in XML are not fixed, name conflicts can occur when different documents use the same names describing different types of elements. To prevent such conflicts, a unique namespace should be defined using a Uniform Resource Identifier (URI). An XML namespace is a collection of names that are used as element types and attribute names (cf. http://www.w3.org/TR/REC-xml-names/). Our default namespace in the start tag of the root element is xmlns:mi=”http://www.m-i.org/images”.

The namespace prefix mi (for medieval images) functions as a placeholder for the namespace name. It needs to show up in all element tags (e.g. <mi:type> <mi:/type>).

(2c) image_id="image5kb78d38i": This is the id attribute which contains the unique identifier of the data source record.

(3-9) All other elements must be within the root element, and can themselves have sub-elements (child elements) which must be properly nested within their parent element. Our elements do not have sub-elements.

Other syntax rules are, for example: all start tags must match end-tags; because XML tags are case sensitive (i.e. the tag <mi:Creator> is different from the tag <mi:creator>, they must also be written with the same case; all elements must have a closing tag; all attribute values must be within quotation marks (e.g. "image5kb78d38i").

Syntactic Transformation /2: The XML Schema

In order to allow for syntactic harmonisation, the XML documents of the museums should conform to the XML Schema of the FMS initiative. Therefore, the museums use the initiative’s XML Schema when they create their XML documents for validating them against the Schema. If the documents are valid, the process can continue to the semantic level.

XML Schema

The XML Schema defines the building blocks of an XML document, including:

- elements and attributes that can appear in a document;
- which elements are child elements, as well as their order and number;
- whether an element is empty or can include text;
- the data types for elements and attributes;
- as well as default and fixed values for elements and attributes.

XML with an XML Schema is designed to be self-descriptive. One of the greatest strengths of XML Schema is that it allows for data typing. The most common data types are xs:string, xs:decimal, xs:integer, xs:boolean, xs:date, xs:time. In the example below, which is the XML Schema for the XML document (image 5kb78d38i.xml) shown in graphic 1, we only use the data type xs:string. This data type is used for values that contain character strings.

```
(1)  <?xml version="1.0"?>
(2a) <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
(2b) targetNamespace=”http://www.m-i.org/images”
(2c) xmlns=”http://www.m-i.org/images”
(2d) elementFormDefault="qualified">
(3)   <xs:element name="image">
(4)    <xs:complexType>
(5)     <xs:sequence>
(6)        <xs:element name="type" type="xs:string"/>
(7)        <xs:element name="subject" type="xs:string"/>
(8)        <xs:element name="iconclass" type="xs:string"/>
(9)        <xs:element name="creator" type="xs:string"/>
(10)      <xs:element name="manuscript" type="xs:string"/>
(11)      <xs:element name="place" type="xs:string"/>
(12)      <xs:element name="year" type="xs:string"/>
(13)     </xs:sequence>
(14)    <xs:attribute name="image_id" type="xs:string"
         use="required"/>
(15)   </xs:complexType>
(16)  </xs:element>
(17) </xs:schema>
```

Short explanations:

(1) The XML declaration, which states that the document conforms to the 1.0 specification of XML.

(2a) Determines that the elements and data types that are used to construct the schema come from the W3C’s XML Schema namespace. Consequently, each

of the elements and data types in the schema has the prefix xs, which identifies them as belonging to the vocabulary of the XML Schema language rather than the vocabulary of our (fictitious) organisation M-i.org.
(2b) Indicates that the elements defined by this schema come from our http://www.m-i.org/images namespace.
(2c) Is our default namespace.
(2d) Demands that any elements used by an XML document which were declared in this schema must be namespace qualified.
(3/16) The parent element for the data typing of the image descriptions we provide at http://www.m-i.org/images. It is (4) defined as a xs:complexType, i.e. it contains child elements (6-12), which are (5/13) surrounded by an xs:sequence element that defines an ordered sequence of these elements.
(6-12) The child elements, which in our example are simple types because they do not contain other elements. They define various elements of our XML documents, e.g. "image", to be of the data type xs:string.
(14) Furthermore, the Schema determines that for the element xs:element name="image" there is a required attribute "image_id" of the data type="xs:string" (for example, image5kb78d38i).

Major Benefits of XML

In the context of the Semantic Web, XML provides an interoperable syntactical foundation upon which solutions to the issues of representing relationships and meaning can be built. We also want to highlight the many benefits of XML that are adding to its rapid uptake in the first place, and might in the longer term be supportive in realising the Semantic Web vision on a broader scale.

XML is one of the most important standards developments in recent years. It is an international, universal, non-system and non-application specific data exchange standard. XML is international, because it employs Unicode. This means that there is no restriction to the western alphabet, but Arabic, Chinese, Greek, Hebrew, Thai, etc. can be easily integrated.

XML is non-system specific, because it is an open standard set by the World Wide Web Consortium (W3C). As such, there is no owner of XML. All the major software suppliers support it; it can be used on any computing platform: from Windows and MacOS to Linux. This makes it easier for organisations to change systems or combine different systems.

XML is also non-application specific, i.e. it can be used in various applications such as data exchange, data harvesting, Web site management, etc. XML is gaining ever-wider acceptance in many application domains including, in particular, the cultural heritage community.

Bear in mind also, that the major collection management software producers have implemented support for XML in their systems, enabling, for example, the integration of data from different collections and their combination over the Web.

XML allows for multi-channel publishing, i.e. with XML it is easy to produce different products or services from digital cultural heritage assets. Once the data are structured in XML, they can be displayed across a variety of media using an associated style sheet that contains the display information.

Finally, XML can be used to create new languages. For example, the Wireless Markup Language (WML), which is used to markup Internet applications for handheld devices, is written in XML.

Ontological concepts
The goal of the FMS project is to make metadata of the museums' textiles collections semantically interoperable on the Web. In order to achieve such interoperability, an ontology is being designed that describes the common (lower-level) ontological concepts in this domain of knowledge.[4]

Ontology

In the Semantic Web architecture, the semantic relationships are not embedded but explicitly represented by an ontology or, rather, an interrelated set of ontologies. In fact, the wide array of information residing on the Web and the perceived need to make it more machine-processable have acted as a strong impetus for the development of ontology languages.

Yet, what is an ontology? An abstract definition of an ontology is that it describes a formal, shared conceptualisation of a particular domain of interest, for example cultural heritage objects held in art museums. In particular, an ontology allows for constraining, expressing and analysing the intended meaning of the shared vocabulary of concepts and relations in a domain of knowledge.[5]

If these concepts and relations are formalised to a high degree, the domain has at hand a major building block for developing semantically aware information systems. With Semantic Web technologies, the domain ontology can be made available on the network, cross-referenced with upper-level and

[4] The available documents (in English) on the FMS initiative state that their ontology is being created using RDF Schema (RDFS). To develop a fully fledged ontology, advanced languages such as DAML+OIL or Web Ontology Language (OWL) would be required.

[5] For more elaborate and formal descriptions, see Tom Gruber: What is an Ontology? (1995), http://www-ksl.stanford.edu/kst/what-is-an-ontology.html; Nicola Guarino: Ontology-Driven Conceptual Modelling, part 1-3 (2002), http://ontology.ip.rm.cnr.it/Tutorials/

other domain ontologies,[6] and remote applications or intelligent software agents can refer to it when they interact to provide a certain information service.

The degree of formalisation of concepts and their relations varies considerably between different domains of knowledge. At the lower end one finds lexicons and simple taxonomies (i.e. an ordered classification system where terms are related hierarchically). At the middle level one might place thesauri, i.e. controlled vocabularies that are structured to show relationships between terms and concepts, and, for example, allow for retrieving them from a database. At the high end of formalisation of knowledge there are axiomatised logic theories. Such theories include rules to ensure the well-formedness and logical validity of statements expressed in the language of the scientific discipline.

In the cultural heritage sector, a powerful IT-supported example of a hierarchical classification system is Iconclass. This supports the documentation of images, in particular art historical images, by providing a systematic collection of 28,000 ready-made definitions of objects, persons, events, situations and abstract ideas that can be the subject of an image. The definitions consist of an alphanumeric classification code and its textual correlate.[7]

For example, for the image we have described in XML in graphic 1, the Iconclass definition is: 71A3421 Eve emerges from Adam's body. The Medieval Illustrated Manuscripts Website of the Koninklijke Bibliotheek, The Hague, has an Iconclass Browser in place[8] that provides the hierarchical path for this concept in the classification system:

7 Bible
  71 Old Testament
    71A Genesis from the creation to the expulsion from paradise, and later years of Adam and Eve
      71A3 creation of man; the Garden of Eden (Genesis 1:26-2)
        71A34 creation of Eve
          71A342 Eve is fashioned from Adam's rib
            71A3421 Eve emerges from Adam's body

An outstanding example of a controlled vocabulary is the Art & Architecture Thesaurus (AAT), one of the Getty Research Institute's Vocabulary Databases. It is a structured vocabulary of more than 125,000 terms and other information about concepts that are used for describing fine art, architecture, decorative arts, archival materials, and material culture.[9]

The W3C's Resource Description Framework (RDF) Schema Specification 1.0, in a section on its scope, mentions concept navigation, and states: 'Thesauri and library classification schemes are well known examples of hierarchical systems for representing subject taxonomies in terms of the relationships between named concepts. The RDF Schema specification provides sufficient resources for creating RDF models that represent the logical structure of thesauri (and other library classification systems).'[10]

Yet, for realising a full-blown cultural heritage ontology for the Semantic Web, there are currently limitations on both sides. On the one hand, hierarchical classification systems and structured vocabularies do not lend themselves easily to rich inter-linking of conceptual 'trees'.

A major step further in this direction is the "CIDOC object-oriented Conceptual Reference Model" (CRM).[11] This provides an ontology of 81 classes and 130 properties, which describes in a formal language concepts and relations relevant to the documentation of cultural heritage.

On the other hand, RDF Schema has limitations when it comes to expressing complex ontological relationships. New languages based on description logics are being developed. These include DAML+OIL and the upcoming Web Ontology Language (OWL), which are capable of fully describing ontologies.

Also worth highlighting is that tools for ontology building are proliferating at the present time. These ontology editors need to be carefully assessed as their capabilities differ considerably.[12]

[6] Upper-level ontologies describe the basic concepts and relationships invoked when information about any domain is expressed in natural language.

[7] For in-depth information, see the official Iconclass Website: http://www.iconclass.nl

[8] Medieval Illustrated Manuscripts Website, http://www.kb.nl/kb/manuscripts/ browser/. The subject access system for the Website was conceived by Mnemosyne Partners, building on the Iconclass classification system nd technologies. See the valuable information they provide at: http://www.mnemosyne.org/business/mss/tempex.html

[9] A detailed description of the AAT is provided at: http://www.getty.edu/research/tools/vocabulary/aat/about.html

[10] http://www.w3.org/TR/000/CR-rdf-schema-20000327/

[11] See the information box at the end of the Forum discussion and the sources mentioned in the Semantic Web Terms and Reading List.

[12] Michael Denny: Ontology Building: A Survey of Editing Tools (06-11-2002) http://www.xml.com/pub/a/2002/11/06/ontologies.html

Semantic Transformation /1:
The RDF Data Model
Progressing towards semantic interoperability, the metadata in the XML documents are now transformed into RDF statements corresponding to the RDF data model. With these so-called RDF 'triples', the XML metadata elements are mapped to the RDF classes and properties, which are defined by the RDF Schema of the FMS initiative (see section /3).

'XML is nothing more than a way to standardize data formats.... This is not to underplay XML's importance. A data-format standard makes all of the more glamorous technologies possible, and RDF is the leading example of the benefit that comes once the data format has been standardized. Many proclaim that RDF is really the XML's killer app, and with good reason. Despite all this, RDF remains somewhat obscure. This is mainly because at its core RDF is very abstract, very dry, and very academic.'
Uche Ogbuji: An introduction to RDF (2000), http://www-106.ibm.com/developerworks/library/w-rdf/?dwzone=xml

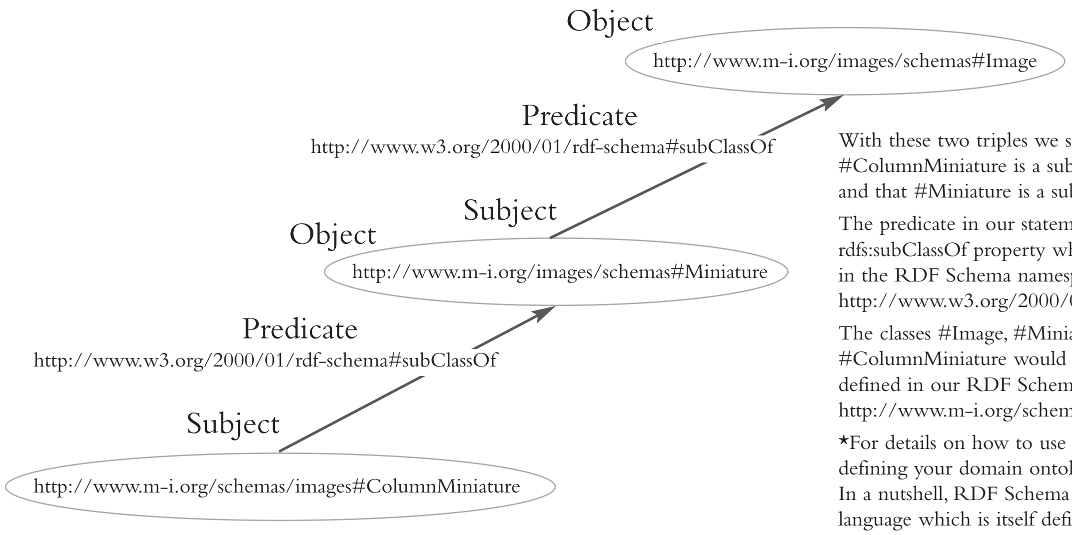
RDF Data Model
In order to make Web resources semantically interoperable, we need resources that provide machine-understandable information about themselves. In the Semantic Web architecture, these statements are built by using the Resource Description Framework (RDF).
RDF defines a data model for the statements describing typed relationships between uniquely identified sources. RDF distinguishes between:

| **resources**: familiar examples are, for example, a Web page, electronic document or digital image, but in RDF also entities that are not 'network retrievable', e.g. museums, curators or bound medieval manuscripts, can be resources.
| **properties**: these identify a specific aspect, characteristic, attribute, or relation used to describe the resource.
| **statements**: these associate a value for a named property with the resource.

Hence, RDF provides a model for describing relationships between resources in terms of named properties and values. The RDF data model intrinsically supports only binary relations. Its base element is the 'triple', which takes the form of subject, predicate, object: a resource (the subject) is linked to another resource (the object) through an arc labelled with a third resource (the predicate). The semantics of a triple clearly depends on the property used as predicate.

A convenient way to visualise this is to draw nodes for subject and object and an arrow between them for the predicate (see graphic 2). In this labelled directed graph, subject and predicate (property) are Uniform Resource Identifiers (URIs), and the object is either a URI or a literal (which is drawn as a box). Everything in RDF can be represented by a graph with nodes and arcs, and the data model allows for using the same URI as a node and as an arc label. To represent RDF statements in a machine-processable way, RDF builds on XML. With RDF/XML, a specific XML markup language, RDF information can be represented and ex-changed between machines.

Graphic 2: RDF Data Model



With these two triples we state that #ColumnMiniature is a subclass of #Miniature, and that #Miniature is a subclass of #Image.

The predicate in our statements is the rdfs:subClassOf property which is predefined in the RDF Schema namespace http://www.w3.org/2000/01/rdf-schema.*

The classes #Image, #Miniature and #ColumnMiniature would also need to be defined in our RDF Schema namespace http://www.m-i.org/schemas/images.*

*For details on how to use RDF Schema for defining your domain ontology see section /3. In a nutshell, RDF Schema is a 'higher-level' language which is itself defined using RDF!

Semantic Transformation /2: Creating and Validating the RDF Statements

In the mapping process, an editor tool (in the FMS case, a tool developed by the project team called Meedio) receives as input the XML documents and assists in transforming them into semantically valid RDF statements (instance descriptions). The tool serves as instance editor, which provides a convenient way of finding and selecting from the XML metadata elements correct instance values for a particular property. The editor tool also serves as semantic metadata validator. When the museum cataloguer saves the set of RDF statements corresponding to an XML document, the semantics of these statements are validated against the property constraints of the FMS ontology. The result of a successful mapping and validation process is a unique set of RDF triples, called the RDF card.

RDF Mapping Rules

When RDF is used to define the meaning of XML metadata elements, a set of mapping rules is created. A mapping rule is a template of RDF triples where XPath expressions are used to identify the actual element values. XPath is a language for addressing parts of an XML document, and was designed for use in XML parsing software (XSLT, XPointer, and others).

When applying such a rule to an XML document, the XPath expressions are instantiated with matching element values. If the rule matches, the RDF template evaluates to a set of RDF triples where XPath expressions are substituted by the corresponding values of the XML elements.

For example, by applying the template rule
<image5kb78d38i/type, mi:hasCreator, /image5kb78d38i/creator>
to the XML document described in the section on XML the following result would be obtained:
<Image Miniature, mi:hasCreator, ‘Alexander Master‘> .

mi:hasCreator is an example of a RDF property. Such properties are explained in section /3: RDF Schema.

Note: Due to the limited space permitted, we do not address issues of term mapping. This is an important aspect of the mapping and validation process carried out in the FMS project. Working with metadata from different museums, they need to deal with partly different terminologies. Their technical solution to synonymous terms (i.e. different terms referring to the same concepts) is to attach synonym sets to the FMS ontology classes. With situations where polysemous terms occur (i.e. the same terms refer to different concepts), the editor tool cannot cope, and the cataloguer needs to select the correct interpretation.

Semantic Transformation /3: The RDF Schema (RDFS)

The shared ontology for the textiles domain is created by using Resource Description Framework Schema (RDFS). An RDF Schema is a tool for indicating the classes of resources one wants to describe as well as for defining the properties used to describe those resources. Furthermore, class/sub-class relationships and property/sub-property relationships can be defined. The museums are mapping their metadata to the classes and properties defined by the RDF Schema of the FMS initiative. Thereby, they are making the meaning of the metadata explicit and representing them in a harmonised uniform way.

RDF Schema

In section Semantic Transformation /1 we have described the data model provided by RDF for expressing statements about Web resources. But we also need a vocabulary for the RDF statements, namely classes and properties defined with RDF Schema (RDFS).

In brief, the RDF Schema mechanism provides a pre-defined vocabulary, a basic type system that can be used in creating domain-specific schemas. Its role is to allow for declaring metadata properties (e.g. for ‘type’, ‘subject’ or ‘creator’), to define the classes of resources they may be used with, to restrict possible combinations, and to detect violations of those restrictions.

Defining classes

With RDF Schema (RDFS), Web resources can be defined as instances of one or more classes. In addition, classes can be organised in a hierarchical fashion. As we hold a collection of digital images drawn from illustrated medieval manuscripts, we first need to define a class of things that are images. In RDF Schema, a class is any resource having an rdf:type property whose value is the RDFS-defined resource rdfs:class.

So, using the basic RDF data model we define: mi:Image [resource] rdf:type [property] rdfs:Class [value]. The self-defined prefix mi (for medieval images) stands for the URI reference of our RDF Schema namespace http://www.m-i.org/schemas/images.

In our image collection we have various special kinds of digitised images, such as column miniatures, decorated initials, schematic drawings, etc. To distinguish, for example, the miniatures, first we need to define a general class Miniature and subclasses of miniatures, e.g. a subclass ColumnMiniature:
mi:Miniature rdf:type rdfs:Class
mi:ColumnMiniature rdf:type rdfs:Class

Secondly, we need to define that mi:ColumnMiniature is a subclass of mi:Miniature, and that mi:Miniature is a subclass of mi:Image, for which we use the predefined rdfs:subClassOf property:
mi:Miniatures rdfs:subClassOf mi:Image
mi:ColumnMiniature rdfs:subClassOf mi:Miniature

As the rdfs:subClassOf property is transitive, this means that mi:ColumnMiniature is also implicitly a subclass of mi:Image.

Graphic 2 on page 32 visualises this with the nodes and arcs of the basic RDF data model.

Defining properties

In order to make the meaning of our metadata (i.e. 'type') explicit, we need to be capable of declaring specific properties that characterise the classes of things we hold at http://www.m-i.org, e.g. digital images of medieval column miniatures.

Basically, RDF schema defines properties in terms of the classes of resources to which they apply. This is the role of the rdfs:domain and rdfs:range mechanisms.

rdfs:range

The range constraint defines the class or set of classes whose instances can be values of a particular property. If we want to define the property mi:hasType, we must describe this resource (which we locate at http://www.m-i.org/schemas/images) with an rdf:type property whose value is rdf:Property:
mi:ColumnMiniature [resource] rdf:type [property] rdf:Property [value].

The following RDF statements indicate that mi:ColumnMiniature is a class, mi:hasType is a property, and RDF statements using the mi:hasType property have instances of mi:ColumnMiniature as values:
mi:ColumnMiniature rdf:type rdfs:Class
mi:hasType rdf:type rdf:Property
mi:hasType rdfs:range mi:ColumnMiniature

rdfs:domain

The domain constraint restricts the set of classes whose instances may have a particular property attached to them. If we want to indicate that the property mi:hasType applies to instances of class mi:ColumnMiniature, we would write:
mi:ColumnMiniature rdf:type rdfs:Class
mi:hasType rdf:type rdf:Property
mi:has Type rdfs:domain mi:ColumnMiniature

Benefits of RDF

In a SearchWebServices.com definition of RDF, some benefits of RDF are mentioned:

- 'By providing a consistent framework, RDF will encourage the providing of metadata about Internet resources.
- Because RDF will include a standard syntax for describing and querying data, software that exploits metadata will be easier and faster to produce.
- The standard syntax and query capability will allow applications to exchange information more easily.
- Searchers will get more precise results from searching, based on metadata rather than on indexes derived from full text gathering.
- Intelligent software agents will have more precise data to work with.'[13]

This is a well-crafted listing of RDF benefits, from provision and exchange of better metadata to agents working with them, hopefully for the benefit of humans. But, as explicitly stated by SearchWeb-Services.com, these are only potential benefits, i.e. they depend on the level of actual uptake of RDF.

13 whatis.com: searchWebServices.com Definitions - Resource Description Framework, http://searchwebservices.techtarget.com/sDefinition/0,,sid26_gci213545,00.html (last updated: July 27, 2001).

Generating and Using the Knowledge Space

The RDF cards represent the original XML documents at the semantic level. The union of such RDF cards constitutes a knowledge base, which is a harmonised semantic representation of the underlying heterogeneous databases.

However, so far the RDF instance descriptions have not left the museum. The museum has complete control of the information it wants to publish, and it does not need to allow the FMS system access to its internal database system. The RDF data are placed in a public directory on the museum's WWW server.

The Web crawler of the FMS system harvests the instance descriptions from the different museums, and the system combines them into an RDF repository. This repository is a large semantic graph that consists of the shared ontology and metadata.

How does a user now search and navigate in this knowledge space? In the FMS system, this is implemented by a server-side software, called Ontogator. Based on the semantic graph, this software dynamically generates semantic linkages for the user's Web browser.

One way of using the FMS system is view-based filtering. The user can select classes of resources from the ontology, and the system finds the instances that match the selected class restrictions. By constraining classes (views) further, the collection instance data searched for are eventually found.

The software also supports topic-based navigation by providing semantic links between topics of interest, the creation of which is based on the collection domain ontology and the related metadata of the collection records. This means that the links also provide the user with an impression of the wider context and pragmatics of the objects in the museums' collections.

From human users to software agents

As described in the Finnish Museums on the Semantic Web example, the RDF repository is a large semantic graph that consists of the shared ontology and metadata of the participating museums. Such a repository can be queried and the results, a set of pointers to the relevant resources, can be accessed using Web browsers. The opportunities provided by a system like the one developed by the FMS initiative (e.g. topic-based navigation) are at present restricted primarily to human users.

The Semantic Web vision includes intelligent software agents which 'understand' semantic relationships between Web resources and seek relevant information as well as perform transactions for humans.[14]

This software would be capable of autonomous action, i.e. could run without direct human control or constant supervision, and ideally is very flexible in doing this. Characterisations of this flexibility include actions that are 'reactive, 'proactive', and 'social' (see below).

While the basic idea of agents is very intuitive and appealing, the actual theory is complex, the tools are immature, the solutions small and prototype-based. In fact, as a parallel distributed systems technology, agents belong to the most complex class of software technology.

However, this primer will conclude with a summary of what an intelligent software agent is and what such a software would generally be capable of doing. This should also serve as an indication of how great the challenge for research and technological development is to make the full Semantic Web vision a reality.

Intelligent Software Agents

The following definitions are taken from Michael Wooldridge's introduction to multiagent systems[15]:

Agent:

'An agent is a computer system capable of autonomous action in some environment'.

Intelligent agent:

'An intelligent agent is a computer system capable of flexible autonomous action in some environment'.

Flexible autonomous action:

'By flexible autonomous action, we mean reactive, proactive, social.'

- Reactivity: 'A reactive system is one that maintains an ongoing interaction with its environment, and responds to changes that occur in it (in time for the response to be useful)'.
- Proactiveness: 'An agent serves a purpose, and therefore exhibits goal-directed behaviour, including the capacity to recognise opportunities for useful courses of action'.
- Social ability: 'Social ability in agents is the ability to interact with other agents (and possibly humans) via some kind of agent communication language, and perhaps cooperate with others'.

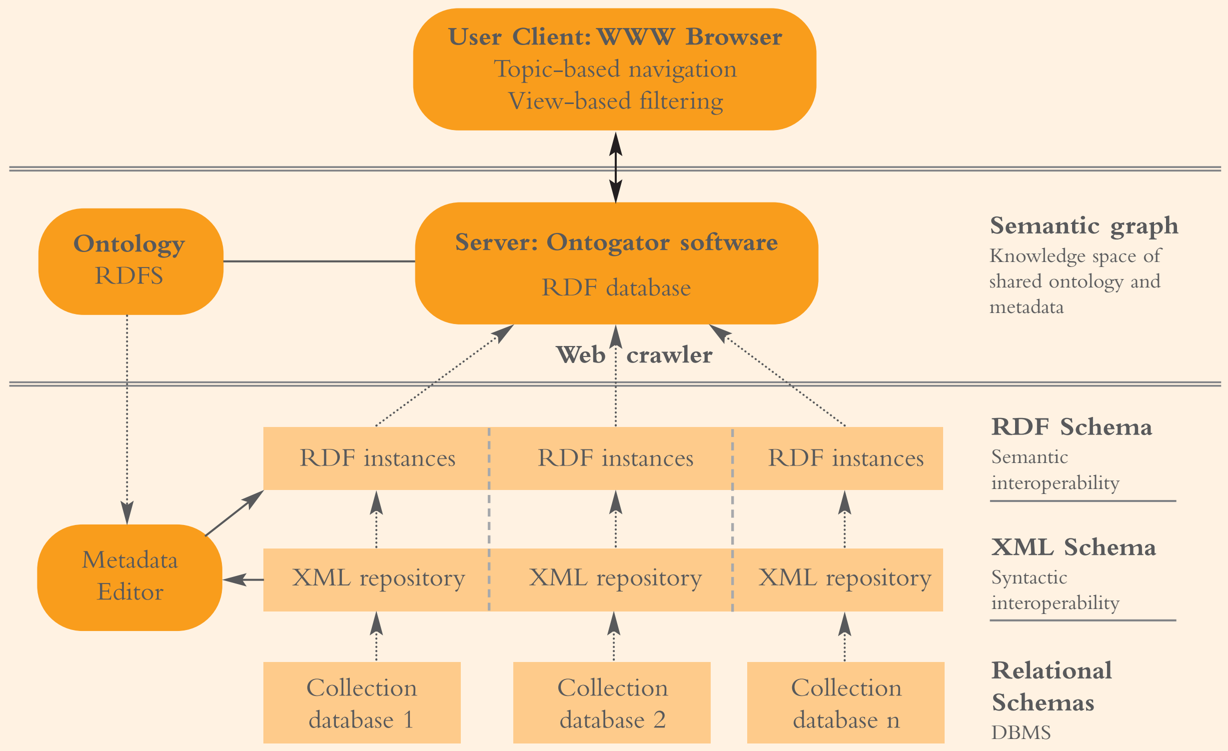
Desirable further properties of agents are:

- Mobility: the ability to move around an electronic network;
- Rationality: an agent will act in such a way that it does not prevent itself from achieving its goals (as far as this is possible with a limited set of beliefs representing its world knowledge);
- Learning: an agent will improve its performance over time.

14 T. Berners-Lee, J. Hendler, O. Lassila, *Scientific American*, May 2001, http://www.sciam.com/2001/0501issue/0501berners-lee.html

15 M. Wooldridge: An Introduction to Multiagent Systems. Chichester: Wiley 2002, and http://www.csc.liv.ac.uk/~mjw/pubs/imas/

The Finnish Museums on the Semantic Web: Overview of the System's Set-up



Graphic 3: Set-up of the Finnish Museums on the Semantic Web System

Resources:

In the primer no references are made to the documents of the World Wide Web Consortium (W3C). All relevant W3C recommendations can be found at http://www.w3c.org.

Of the wealth of introductory materials on XML and RDF available on the Web, the following in particular are useful to consult for further details:

http://www.w3schools.com/xml/
http://www.w3schools.com/schema/
http://www.w3.org/TR/rdf-primer/

K. S. Candan, H. Liu, R. Suvarna: Resource Description Framework: Metadata and its Applications. In: *SIGKDD Explorations*, Vol. 3.1 (2001), 6-19.

Pierre-Antoine Champin: RDF Tutorial (2001), http://www710.univ-lyon1.fr/~champin/rdf-tutorial/rdf-tutorial.html

S. Decker, M. P. Mitra, S. Melnik: Framework for the Semantic Web: An RDF Tutorial, http://www.ida.liu.se/~asmpa/courses/sweb/rdf/rdf_tutorial.pdf

Martius h̄t dies xxxi luna xxx

iii	d		Albini epī ꝛ cf. mem̄ B
	e	vi	
xi	f	v	
	g	iiii	
xix	a	iii	
viii	b	ii	Incipit vii embolismus.
	c	Nonas	Saint thomas daquiū totū duplex
xvi	d	viii	
v	e	vii	
	f	vi	
xiii	g	v	Claues pasche.
ii	a	iiii	Gregorii pp̄. dup'
	b	iii	
x	c	ii	Clamum x̄l
	d	Idus	Equinoctium ĩnale.
xviii	e	xvii	Aprilis.
vii	f	xvi	
	g	xv	Sol in ariete.
xv	a	xiiii	
iiii	b	xiii	
	c	xii	Benedicti abb' simplex.
xii	d	xi	p̓mum pascha.
i	e	x	
	f	ix	
ix	g	viii	Annunciacio dn̄ica. totum duplex.
	a	vii	
xvii	b	vi	
vi	c	v	B
	d	iiii	
xiiii	e	iii	
iii	f	ii	Hore noctis xii. diei xii.

The Darmstadt Forum Participants

Wernher Behrendt, Salzburg Research, Austria
Wernher Behrendt is a Senior Researcher at the Sun Technology and Research Excellence Center at Salzburg Research (Austria) working on multimedia middleware and interoperation issues. He holds an MSc in Cognitive Science from Manchester University and has more than 10 years' experience in near-to-market IT research. From 1989 to 1995 he was a Senior Research Associate in the Informatics Department at Rutherford Appleton Laboratory (UK), working on embedded knowledge based systems in distributed multimedia presentation systems. From 1995 to 1998 he was a Senior Research Associate at Cardiff University (UK), working on interoperation between heterogeneous information systems. Mr Behrendt has held courses in Computer Science and has worked on projects ranging from software engineering methods and quality assurance to legacy system reengineering using migration methods and distributed systems middleware.
E-mail: wernher.behrendt@salzburgresearch.at

Paolo Buonora, Archivio di Stato di Roma, Italy
Paolo Buonora holds a degree in Philosophy from the University of Rome 'La Sapienza' (1976). He worked in the Italian State Archive Administration from 1978, where he was first involved in editing the Guida Generale degli Archivi di Stato italiani. From 1986 he worked in the Soprintendenza archivistica per il Lazio, surveying audiovisual archives, municipal archives; from 1989 to 1991 at the Perugia University, engaged in a doctoral research in 'Urban and rural history'; and from 1991 to 1994 again in the Soprintendenza archivistica per il Lazio. After 1994 he worked in the Archivio di Stato di Roma, where he was responsible for the photograph service and several working groups on informatics application in archival documentation. From 1997 until the present time he has planned and directed the Imago II project in the Archivio di Stato di Roma.
See: http://www.asrm.archiv.beniculturali.it/sid/imago/IMAGOIIen.html

Samuel Cruz-Lara, University of Nancy 2 and LORIA, France
Samuel Cruz-Lara obtained a Master's degree in Computer Science in 1984 (University Henri Poincaré, Nancy 1) and a PhD degree in Computer Science in 1988 (National Polytechnic Institute of Lorraine). The central topic of his PhD thesis was the generation of integrated development environments by using attribute grammars.
He is currently Associate Professor at the University of Nancy 2 (Institute of Technology, Computer Science Department) and permanent Researcher at LORIA (Lorraine Laboratory for Research in Computer Science and its Applications – UMR 7503 – CNRS – INRIA – Universities of Nancy). He is a member of the 'Language and Dialogue' team and has conducted several research activities on distributed software architectures and textual linguistic resources management. He is currently working in the context of distributed architectures and multimedia resources management. Dr Cruz-Lara has participated in several projects, in particular CNRSSILFIDE and MLIS-ELAN, and he is at present co-leader of the 'Digital Museum' project. This is a joint project between LORIA and the National Chi-Nan University. The 'Digital Museum' project is sponsored by the National Science Council of the Republic of China (Taiwan) and supported by INRIA (France).

Costis Dallas, Critical Publics SA, Greece
Costis Dallas is Chairman and Senior Researcher of Critical Publics (http://www.criticalpublics.com), a group of companies active in the field of strategic communications, creative design and technology.
He is currently a Lecturer in the Department of Communication and Mass Media of Panteion University, and has over 15 years of research and professional experience in hypermedia applications, human factor issues and cultural information systems. Dr Dallas has been co-founder and Executive Vice-President of ISP Hellas Online SA, co-founder and Chair of the Multimedia Working Group of the International Council of Museums (CIDOC/ICOM), Head of Documentation and Systems of the Benaki

Museum, General Director of the Foundation of the Hellenic World, Special Secretary of the Greek Ministry of Education in charge of libraries, archives and instructional technologies, and Special Advisor to the Greek Foreign Minister on cultural and information technology issues.
E-mail: info@criticalpublics.com

Bert Degenhart-Drenth, ADLIB Information Systems BV, The Netherlands
Bert Degenhart-Drenth is the founder and general manager of ADLIB Information Systems BV (http://www.nl.adlibsoft.com), a leading company in the field of library, museum and archive automation. Although his background is in electronic engineering, he became involved in a museum automation project as early as 1983. Degenhart-Drenth worked for the MARDOC foundation in Rotterdam for three years, setting up one of the first integrated museum automation systems in The Netherlands. After that he joined Databasix, producer of the ADLIB software. In 1991, Degenhart-Drenth led a management buy-out to form Databasix Information Systems, which was renamed later as ADLIB Information Systems.
ADLIB Information Systems is now the market leader in museum automation in the Benelux region and has more than 1000 customers in the libraries, museums and archives field. ADLIB is a CIMI member and has, as such, been involved in the CIMI Z39.50 and Dublin Core test beds. Degenhart-Drenth has been a core member in the development of the Spectrum-XML schema and collaborates in the EMII-DCF project. Relevant ADLIB projects include: CIMI Dublin Core test bed - ADLIB hosts the CIMI Dublin Core test bed and has developed a database application for this. The Open Archives Initiative Protocol version 1.0 is now available for this database, in addition to the 'standard' HTML/XML access. SPECTRUM-XML - A project of the UK mda, together with a team of software vendors, to produce a schema for exchange of data which contains information elements from the UK Spectrum standard. This will be implemented in the ADLIB Museum software. Internet Gelderse Musea (http://www.igem.nl) and Maritiem Digitaal (http://www.maritiemdigitaal.nl): Two Web-based projects that make the data from multiple museums (including their library data) available on the Web, based on a three-tier implementation with XML as the data exchange mechanism. E-mail: bert@nl.adlibsoft.com

Nicola Guarino, Italian National Research Council, Italy
Nicola Guarino is a senior researcher at the Institute for Cognitive Sciences and Technologies of the Italian National Research Council, where he leads the Laboratory for Applied Ontology. He graduated in Electrical Engineering from the University of Padova in 1978. He has been active in the ontology field since 1991, and has played a leading role in the AI community in promoting the study of the ontological foundations of knowledge engineering and conceptual modelling under an interdisciplinary approach. His current research activities involve formal ontology, ontology design, knowledge sharing and integration, and ontology-based metadata standardisation. He is general chairman of the International Conference on Formal Ontology in Information Systems (FOIS), and associated editor of the International Journal of Human-Computer Studies. He has published more than 60 papers in scientific journals, books and conference proceedings, and has been guest editor of three special issues of scientific journals related to formal ontology and information systems. He is involved in various projects related to ontologies and the Semantic Web, including WonderWeb and OntoWeb.
E-mail: Nicola.Guarino@ladseb.pd.cnr.it

Janneke van Kersen, Digital Heritage Association, The Netherlands
Janneke van Kersen graduated in Art History and took part in a postgraduate programme in Historical Information Processing. After her graduation in 1992 she worked in the field of Humanities and Computing. She held several posts at Utrecht

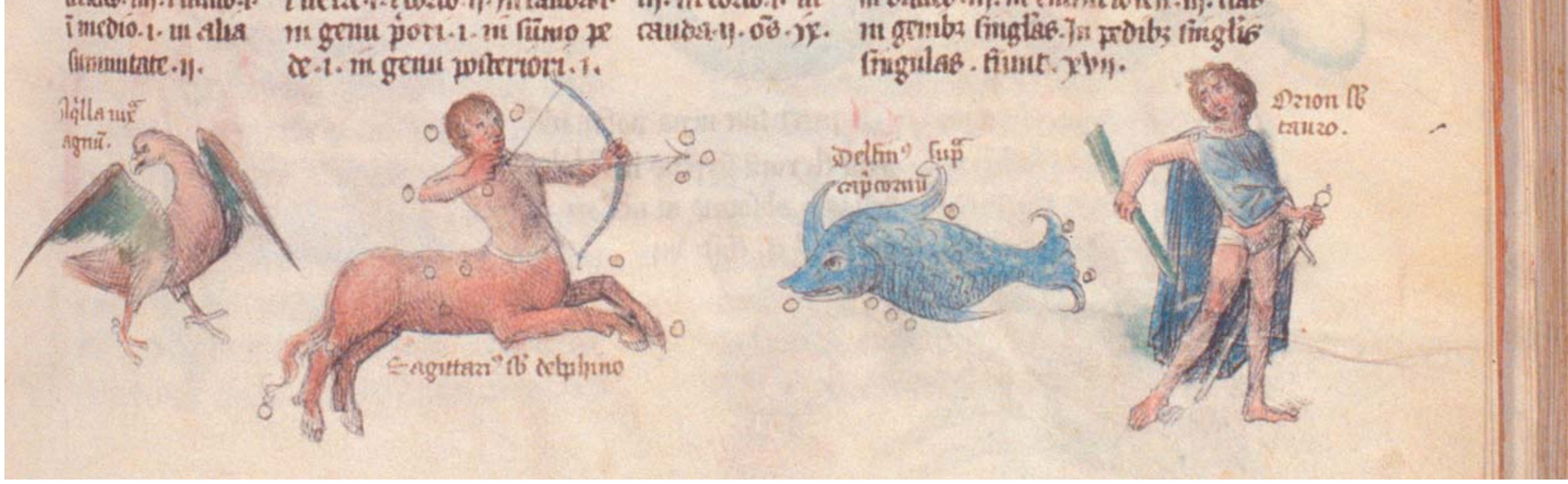
University, as a teacher in the Department of Humanities and Computing and more recently she was responsible for the realisation of computer-aided applications for the Department of Art History. Furthermore she worked at the Netherlands Historic Data Archive and taught in the postgraduate programme on Historical Information Processing at Leiden University.
Since October 1999 Kersen has been working as a consultant with the Dutch Digital Heritage Association. The consulting focuses on digitisation, standardisation, metadata, and education & ICT. The main goal of the organisation is to provide access to distributed databases of cultural heritage organisations such as museums, archives, archaeological organisations, monuments and special collections of libraries in a context-rich and XML-based environment. Interoperability from both a technical and an organisational perspective is therefore a major issue. The Association offers access to heritage information to the general public at http://www.cultuurwijzer.nl and customised access for the educational field at http://www.cultuurwijs.nl.
E-mail: janneke.van.kersen@den.nl

Marco Meli, EDW International, Italy
Marco Meli is CEO and co-founder of EDW International (http://www.edw-international.com), Milan, Italy, a company providing leading corporate publishing and content management applications based on XML and related standards. Meli has long experience in content/document management and in multimedia creation and production. He is a member of the Organisation Group of XML Italia VP of SGML UG Italia. He has given a number of presentations at SGML and XML related conferences in Italy, Europe and the US. Meli is also editor of a column on new facets of publishing in *Graphicus magazine*. He has been involved in cultural projects since 1996, and acts as a reviewer of Information Society Technologies (IST) projects for the European Commission.
E-mail: meli@edw.it

Paul Miller, UKOLN, UK
Dr Miller holds the post of Interoperability Focus at UKOLN (UK Office for Library and Information Networking, http://www.ukoln.ac.uk/). Interoperability Focus is jointly funded by the Joint Information Systems Committee (JISC) of the UK's Further and Higher Education Funding Councils and Resource: the Council for Museums Archives and Libraries. The post is responsible for exploring, publicising and mobilising the benefits and practice of effective interoperability across diverse information sectors, including libraries and the cultural heritage and archival communities. Dr Miller sits on a number of relevant committees, including the Executive Committee of the CIMI Consortium, the Advisory Board of the Dublin Core Metadata Initiative (DCMI), and the Metadata Working Group of the UK Government's Office of the e-Envoy.
E-mail: P.Miller@ukoln.ac.uk

Frank Nack, CWI, UK
Dr Frank Nack is a senior researcher at CWI, currently working within the Multimedia and Human-Computer Interaction group. He obtained his PhD on 'The Application of Video Semantics and Theme Representation for Automated Film Editing' at Lancaster University, UK. The main thrust of his research is on video representation, digital video production, multimedia systems that enhance human communication and creativity, interactive storytelling and media-networked oriented agent technology. He is an active member of the MPEG-7 standardisation group where he served as editor of the Context and Objectives Document and the Requirements Document, and chaired the MPEG-7 DDL development group. He is on the editorial board of *IEEE Multimedia*, where he edits the Media Impact column.
E-mail: Frank.Nack@cwi.nl

Franco Niccolucci, Florence University, Italy
Franco Niccolucci has a background in mathematics and computer science and is at present a professor at the University of Florence, where he lectures in the



Faculty of Architecture and at the School of Archaeology, and at the University of Basilicata where he lectures in the Faculty of Cultural Heritage. He is also director of the Laboratory for Virtual Archaeology and Digital Culture in the Prato campus of the University of Florence. He is a member of several professional associations and is the Italian representative on the international Steering Committee of CAA, the association for Computer Applications to Archaeology. His interests concern virtual archaeology and the use of multi-media in cultural heritage communication, the subjects of the two most recent books edited by him. His present research deals with virtual reality models, the use of XML for archaeological and historic data management, and in general the impact of informatics on archaeological theory and method. He has been a member of the scientific committees of the most recent international conferences on these subjects and will chair the 2004 CAA International Conference. See also:
http://www.geog.port.ac.uk/hist-bound/people/niccolucci.htm

Seamus Ross, HATII, University of Glasgow, UK

Dr Seamus Ross is Director of Glasgow University's Humanities Advanced Technology and Information Institute (HATII). He is also Director of ERPANET (Electronic Resource Preservation and Network) (IST-2001-32706), a European Union funded accompanying measure to enhance the preservation of cultural heritage and scientific digital objects. Previously he was Assistant Secretary for Information Technology at the British Academy, and before that worked for a company specialising in expert systems and software development, as a software engineer and then in management. He researches, lectures, and publishes widely on information technology and digital preservation. Dr Ross acts as ICT advisor to the Heritage Lottery Fund and is a monitor for a number of large ICT-based projects in the UK. He is a member of a number of international organisations including the DLM-Monitoring Committee of the European Commission, the Research Libraries Group's PRESERV Working Group on Preservation Issues of Metadata, and InterPARES (as well as Co-Chair of its European Team).
E-mail: S.Ross@hatii.arts.gla.ac.uk

Andrea Scotti, Institute & Museum for the History of Science, Italy

Andrea Scotti graduated from the University of Bologna, Department of Philosophy, in 1983. From 1985 to 1995 he carried out several research projects involving the cataloguing of scientific manuscripts in Israel (Hebrew University of Jerusalem), Czechoslovakia (Karol University of Prague), Hungary (Szecheny National Library of Budapest), and Germany (Institut für Geschichte der Naturwissenschaften, Munich). His numerous research activities concentrate on software and programming for library databases.

Since 1996 Scotti has been Director of the General Digital Catalogue of the Scientific Manuscripts located at the Central National Library in Florence developed in co-operation with the Istituto e Museo di Storia della Scienza, the National Central Library, and under the auspices of the Italian Ministry for Culture. He is in charge of the work the Institute and Museum of the History of Science carries out for the MESMUSES project (01/02/01-30/07/03), funded by the European Commission under the Information Society Technologies (IST) Programme. The project aims at designing and experimenting with metaphors for organising, structuring and presenting the scientific and technical knowledge offered to the public, implementing Semantic Web technologies.
See: http://cweb.inria.fr/Projects/Mesmuses
E-mail:
scottian@galileo.imss.firenze.it

DigiCULT: Project Information

DigiCULT is an IST Support Measure (IST-2001-34898) to establish a regular technology watch that monitors and analyses technological developments relevant to and in the cultural and scientific heritage sector over the period of 30 months (03/2002-08/2004).

In order to encourage early take up, DigiCULT produces seven Thematic Issues, three Technology Watch Reports, along with the newsletter DigiCULT.Info.

DigiCULT draws on the results of the strategic study 'Technological Landscapes for Tomorrow's Cultural Economy (DigiCULT)', that was initiated by the European Commission, DG Information Society (Unit D2: Cultural Heritage Applications) in 2000 and completed in 2001.

Copies of the DigiCULT Full Report and Executive Summary can be downloaded or ordered at http://www.digicult.info.

For further information on DigiCULT please contact the team of the project co-ordinator:

Mr. Guntram Geser,
guntram.geser@salzburgresearch.at
Phone: +43-(0)662-2288-303

Mr. John Pereira, john.pereira@salzburgresearch.at
Phone: +43-(0)662-2288-247

Salzburg Research Forschungsgesellschaft
Jakob-Haringer-Str. 5/III
A - 5020 Salzburg Austria
Phone: +43-(0)662-2288-200
Fax: +43-(0)662-2288-222
http://www.salzburgresearch.at

Project Partner:
HATII - Humanities Advanced Technology and Information Institute
University of Glasgow
http://www.hatii.arts.gla.ac.uk/
Contact: Mr. Seamus Ross, s.ross@hatii.arts.gla.ac.uk

The members of the Steering Committee of DigiCULT are:

Philippe Avenier, Ministère de la culture et de la communication, France
Paolo Buonora, Archivio di Stato di Roma, Italy
Costis Dallas, Critical Publics SA, Greece
Bert Degenhart-Drenth, ADLIB Information Systems BV, The Netherlands
Paul Fiander, BBC Information & Archives, United Kingdom
Peter Holm Lindgaard, Library Manager, Denmark
Erich J. Neuhold, Fraunhofer IPSI, Germany
Bruce Royan, Concurrent Computing, United Kingdom

DigiCULT Thematic Issue 1 - Integrity and Authenticity of Digital Cultural Heritage Objects builds on the first DigiCULT Forum held in Barcelona on May 6th, 2002, in the context of the DLM-Conference 2002.

DigiCULT Thematic Issue 2 – Digital Asset Management Systems for the Cultural and Scientific Heritage Sector builds on the second DigiCULT Forum held in Essen, Germany, on September 3rd, 2002, in the context of the AIIM Conference @ DMS EXPO.

DigiCULT Thematic Issue 3 - Towards a Semantic Web for Heritage Resources builds on the third DigiCULT Forum held on January 21st, 2003, at Fraunhofer IPSI, Darmstadt, Germany.

DigiCULT Thematic Issue 4 will follow the fourth DigiCULT Forum on Learning Objects, that will take place at the Koninklijke Bibliotheek - National Library of the Netherlands, The Hague, on July 2nd, 2003.

IMPRINT

This Thematic Issue is a product of the DigiCULT Project (IST-2001-34898)

Authors:
Guntram Geser, Salzburg Research
Joost van Kasteren, Journalist
Seamus Ross, University of Glasgow, HATII
Michael Steemson, Caldeson Consultancy

Images:
Images for this Thematic Issue have been provided by and are reproduced with kind permission of the Koninklijke Bibliotheek – National Library of the Netherlands, The Hague, Netherlands.

Graphics & Layout:
Jan Steindl, Salzburg Research

ISBN 3-902448-00-8
Printed in Austria.

IMAGES

Augustine: La Cité de Dieu (Book 1-10).
Paris, c. 1400-1410. Volume I.
Image on p. 5 from Fol. 264r, size 80x75,
illuminator: Orosius Master a.o.

Bible Historiale. Paris, c. 1320-1340. Volume I.
Image on p. 7 from Fol. 2r, size: 45x50,
illum.: 'Sub-Fauvel' Master.

Historic Bible. Utrecht, c. 1430. Volume I.
Images on pp. 8, 10, 14, 28
from Fol. 3r, 3v, 4v, 7v, size: 55x85 to 60x85,
illum.: Alexander Master o.a.

Psalter. Breviary of St. Bridget. Den Bosch,
Monastery Marienwater, Bridgettines, 1468.
Images on pp. 11, 13,
Wednesday, Matins: Invitatorium,
from Fol. 219r

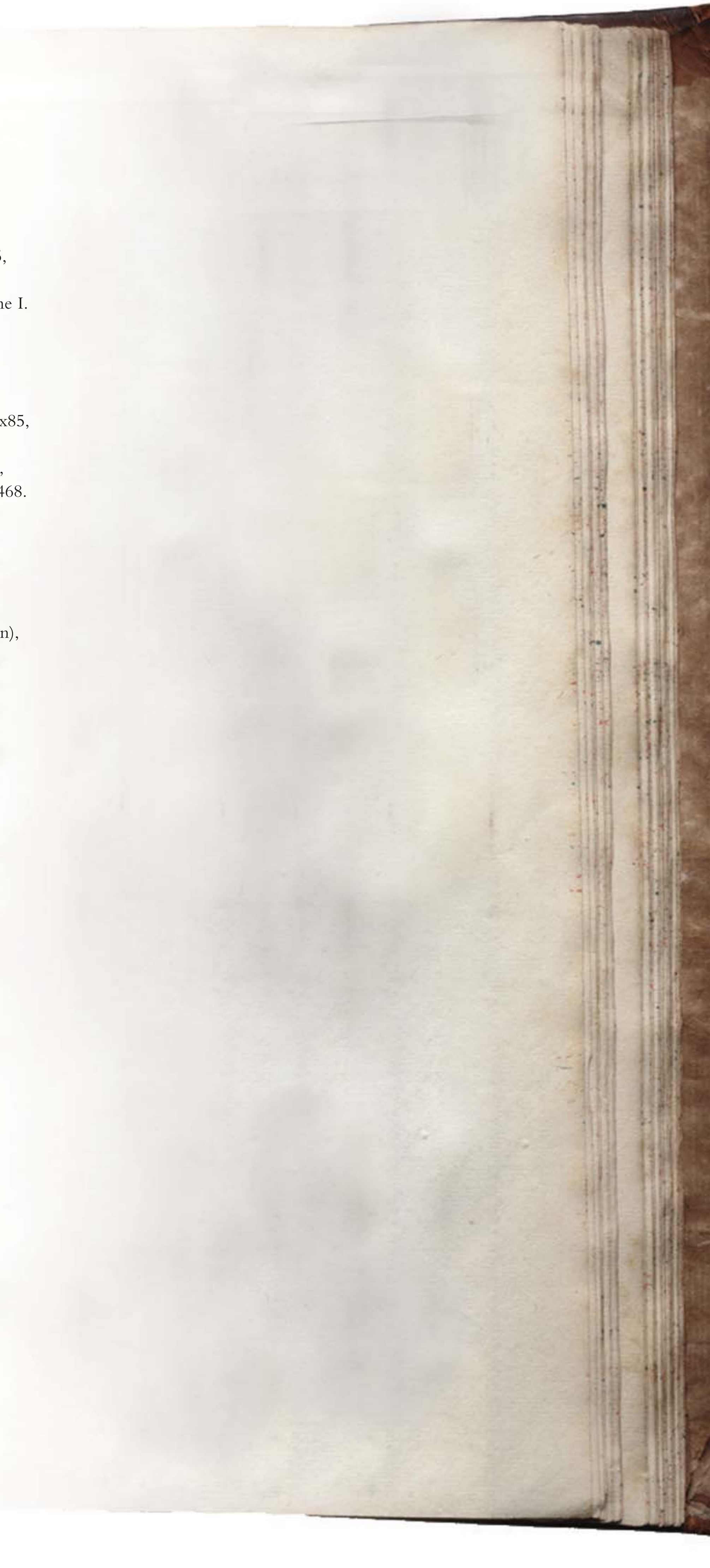
Jacob van Maerlant: Der Naturen Bloeme.
Flanders, c. 1350.
Images on pp. 15 (Cerilius), 17 (Fastaleon),
18 (Draco), 19 (Zitiron)
from Fol. 104rb1, 106rb2, 124r, 111ra,
size: 40x55 to 50x55

Jacob van Maerlant: Spieghel Historiael.
West Flanders, c. 1325-1335.
Image on p. 21 from Fol. 4va1,
size: 45x55.

Lambert of St. Omer: Liber Floridus.
Lille and Ninove, 1460.
Images of Signs of the Zodiac
on pp. 22, 23, 24, 38, 39, 40, 41

Psalter. Normandy, c. 1180.
Image on p. 26 from Fol. 3v,
size: 160x125.

Breviary. Cambray(?), c. 1275-1300.
Images on pp. 27, 29, 31, 33, 34, 35, 37
from Fol. 211r, size: 205x135.



DigiCULT Consortium:

salzburg research

UNIVERSITY of GLASGOW

www.digicult.info

ISBN 3-902448-00-8