INFORMATION SOCIETY TECHNOLOGIES
(IST)
PROGRAMME

Project IST-2001-33562 MoWGLI

Report n. D0.c
Second Self-Assessment Report

Main Author:
A. Asperti

Project Acronym: MoWGLI
Project full title: Mathematics On the Web: Get it by Logic and Interfaces
Proposal/Contract no.: IST-2001-33562 MoWGLI

Contents

1 Project Meetings and Information flow

1.1 Project Meetings

During the second reporting period only one meeting has been held, in Sophia Antipolis. One meeting was held in Bologna immediately before the end of the first reporting period (18-19 February 2003), and another meeting was held in Saarbrücken on March 22-24 2004 (so, immediately after the end of the reporting period).

The meeting in Sophia Antipolis was held by INRIA on September 15th - 16th, 2003. We had 15 participants, so subdivided: UNIBO (4), INRIA (5), DFKI (1), KUN (3), MPG (2), Trusted Logic (0).
There have been also several personal visits or bilateral meetings among project partners, in particular:

- Herman Geuvers and Iris Loeb (Nijmegen) visited Bologna on 7-10 July and 7-13 respectively;
- Claudio Sacerdoti Coen (Bologna) visited Paris (INRIA) for the "Math in Coq Meeting", Paris, France, 23-24 April 2003.
- H. Barendregt, F. Wiedijk, B. Spitters, I. Loeb and L. Cruz-Filipe (Nijmegen) visited Paris (INRIA) for the "Math in Coq Meeting", Paris, France, 23-24 April 2003.
- Luca Padovani and Claudio Sacerdoti Coen (Bologna) visited Sophia Antipolis on 17-20 April 2003;
- Claudio Sacerdoti Coen (Bologna) and Edoardo Gimenez (Trusted Logic) met in Paris on September 23th, 2003.
- Hugo Herbelin (INRIA Paris) has visited Nijmegen on 16-20 February 2004.

1.2 Project Synergy

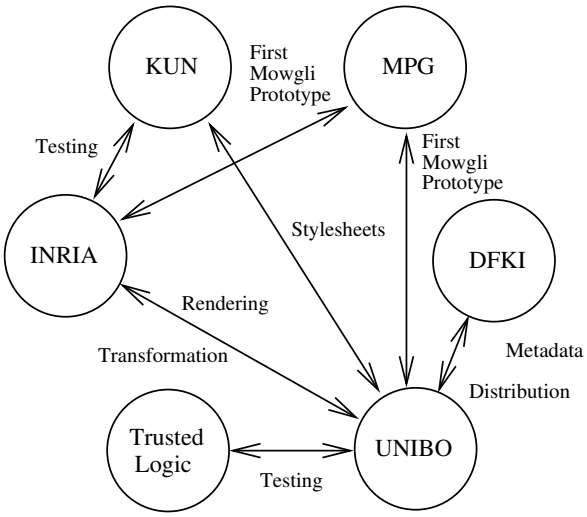
Figure 1.2 describes the main synergies between the different partners, up to now (thick lines). In particular, there has been a strong interaction on the following themes:

Stylesheets One of the main activity during this reporting period has been the study and the development of presentational stylesheets, both for expressions and proofs (tasks T2.4-5). The work has been performed in cooperation between Bologna and Nijmegen, anticipating some of the issues of the validation task 6.2.

Metadata Metadata modeling has been a joint effort between Saarbrücken and Bologna.

Mowgli Prototypes The development of the first and second Mowgli Prototypes have required the cooperation of most partners, and in particular of all those partners involved in the development of semantic based services: Bologna, INRIA, Saarbrücken and MPG.

Exportation and rendering The release of Version 8 of Coq, comprising major syntactical and semantical modifications to the core logical calculus, required a global revisitation of the exportation module and the transformation stylesheets, done in collaboration within INRIA and Bologna.



Testing Pilot application 6.1, concerning the development of an interactive algebra textbook, with fully formalized content, required a deep collaboration between the INRIA Coq-team and Nijmegen.
For Pilot application 6.3, Trusted Logic started the validation of Coq2XML, the tool translating formal models written in Coq into XML format developed at Bologna. This validation work created a stimulating synergy between the two sites.

2 Work progress overview

2.1 Specific Objectives for the reporting period

According to the contractual workplan, the activity of the second reporting period has been mainly focused on the following Workpackages and Tasks:

WP0 Project Management
WP2 Transformation (Tasks T2.4, T2.5 and T2.6)
WP3 Metadata (Task T3.2)
WP4 Interfaces (Tasks T4.2, T4.3 and T4.4)
WP5 Distribution (Tasks T5.1 and T5.2)
WP6 Testing and Validation (all Tasks)
WP7 Information Dissemination and Exploitation

During the second year we planned to perform, in parallel, four major activities:

1. study and development of presentational stylesheets, both for expressions and proofs, and automatic extraction of metadata (tasks T2.4-6);
2. architectural design and implementation of the consultation engine (task T4.2), and of the functionalities for assisted annotation in natural language of the documents (Task T4.3);
3. finalize the first prototype of a LaTeX based authoring tool (Task T4.4)

4. overall architectural design and first prototype implementation of the distribution model (task T5.1-T5.2);

In parallel with these activities, we also started an extensive validation of MoWGLI's achievements, by means of three major pilot applications:

1. Formalization of a full undergraduate course in algebra or analysis for didactic purposes.
2. Formalization of (part of) the process of loading, verifying and executing an applet into a smart card.
3. Make maximal use of content marked-up articles in a solely electronic scientific physics journal. This will allow us to demonstrate the benefits of content mark-up for search, retrieval, and re-use of mathematical content, and user customizable content presentation.

2.2 Overview of the progress

All Milestones have been successfully accomplished, and all reports/prototypes have been delivered at the due time, respecting the prefixed timetable.

We are particularly proud of this fact, considering some unexpected difficulties, mostly related to the new release of the COQ Proof Assistant (V.8). This was a major release, with substantial modifications at all levels of the system, from the kernel to the concrete syntax, to the library of formal theorems itself. As a consequence, a huge amount of work has been devoted to update the exportation module of Task 2.1 (and other components) to the new version.

2.2.1 Presentational Stylesheets and the Consultation Engine

The work has mostly been a joint cooperation between UNIBO, NIJMEGEN and INRIA. It consisted in the development of presentational stylesheets transforming the content (OMDoc-like) description of Task 2.2-3 into several rendering formats, comprising HTML and MathML, and leading to the first MoWGLI prototype supporting browsing, rendering and consultation.

The main bulk of the work was the pretty printing of proofs, the suitable rendering of mathematical notation, a bit of dynamic HTML (Javascript) for implementing telescopic navigation inside proofs, and the creation of the structure of hyperlinks supporting browsing and navigation inside the repository.

More than 8000 lines of XSLT have been written, organized in about 20 files covering basic notation for different areas of mathematics (arithmetics, analysis, trigonometry, set theory, etc.). The stylesheets have been schematically described in a report accompanying deliverable D2e-f.

The result of the full transformation process can be consulted on-line at the pages

`http://mowgli.cs.unibo.it/library`
`http://helm.cs.unibo.it/library.html`

It is worth to mention that stylesheets are applied *on the fly*; in spite of the fact that even the best available implementations of XSLT (libxslt[1]) is much less efficient than one could reasonably expect, the elaboration time is still within the limits of a typical web transaction (few seconds).

[1]It is worth to mention that our extensive and somehow extreme use of xslt allowed to spot and to solve several implementation problems of libxslt, especially related to the management of namespaces.

2.2.2 LaTeX-based authoring tool

The work (Deliverables D4d and D4e) consisted in the development of an authoring tool for scientific documents (Hermes) supporting manual and automatic generation of Content-MathML, and automatic generation of Presentation-MathML, enabling an author to add and recover semantic depth and clarity to LaTeX written documents.

The prototype implementation of Hermes has the following components:

- a set of helper LaTeX macros, which allows the author to disambiguate the meaning of the mathematical expressions he writes, while allowing some choices for the presentation; this set is included by the author in the originally written LaTeX document. A LaTeX run on the macro-enriched document will output a 'semantic dvi' file (a dvi file containing 'special' annotations of various combinations of graphical and non-graphical symbols in the source).
- a scanner, written in flex, which extracts from the resulting dvi file the semantic tokens seeded by the macro collection above and sends them to the parser below.
- a parser, written in bison, which creates the final content-oriented XML representation.
- an XSLT stylesheet, which transforms the content-oriented XML document into a renderable, cross-referenced, document.

The first Hermes prototype (D4d) handled consistently only those mathematical expressions in LaTeX which contain structures covered by the current Content-MathML standard (version 2.0). The refined prototype (D4e) handles arbitrary mathematical expressions in LaTeX by first encoding them into Presentation-MathML and then refining expressions which have a clear meaning to Content-MathML while the text in the sources is converted into Unicode.

2.2.3 Distribution model

The work, mostly performed by DFKI, consisted in the architectural design of an infrastructure for collaborative content management and version control for structured mathematical knowledge. This will enable multiple users to work jointly on mathematical theories with minimal interference.

In particular, report D5a describes an API covering all functionalities needed to realize a CVS-like version control and distribution model. The proposed architecture extends the CVS architecture in two ways, motivated by the specific needs of distributed management of structured mathematical knowledge on the Internet. On the one hand the one-level client/server model of CVS is generalized to a multi-level graph of client/server relations, and on the other hand the underlying change-detection tools take the math-specific structure of the data into account.

2.2.4 Validation

The second Hermes prototype (deliverable D4e) was tested on a 700 pages AMSTeX volume of mathematical abstracts provided by ZentralBlatt MATH, and as a consequence, the ZentralBlatt MATH group wants to try Hermes for a prototype migration from the current AMSTeX sources to XML+MathML. It was also tested on a review article from the Living Reviews

database and the result of the conversion is available on the Living Reviews site along with the Hermes source distribution documentation. The AEI group is in permanent contact with various other individual users interested in using Hermes to put their work on the web reliably.

The style sheets and the consultation engine have been tested by the Nijmegen team on the C-CoRN repository of formalized mathematics (C-CoRN stands for 'Constructive Coq repository at Nijmegen'). The presentational stylesheets developed in Bologna have been tested and improved by evaluating the output that is obtained by applying these stylesheets to the Coq-XML output. The Nijmegen team has also made experiments in producing course notes based on a library of formal mathematics. This was done by presenting IDA (the 'interactive course notes on algebra', developed at the Technical University of Eindhoven, a subsite of Nijmegen) using the web services developed in Bologna, which allows to consult and view the formal C-CoRN definitions and proofs underneath.

Trusted Logic has started the validation of Coq2XML, the tool translating formal models written in Coq into XML format developed by Bologna.

The work developed by Trusted Logic in the last months concerned two main activities.

First, Coq2XML has been successfully installed in Trusted Logic site. This step implied the resolution of some technical problems for fitting the tool into an industrial environment. On one hand, Coq2XML has been conceived for working on an open environment, where remote connections and information transfer is not restricted at all. On the other hand, Trusted Logic is subject to strong restrictions concerning IP protection and the management of security sensitive information for its clients. Technical problems have been overcame with the help of the Bologna site.

Second, Trusted Logic has started the formal modeling in Coq of one of the case studies that will be used to test Coq2XML. The models concern some of the security policies that shall be enforced by a smart card implementing GlobalPlatform. GlobalPlatform (http://www.globalplatform.org) is an industrial standard specifying the card management services provided by a multi-application smart card, like loading of a new application on the smart card, starting the execution of an installed application, modifying one of the cryptographic keys contained in the smart card, etc. Those models will be supplied as input to Coq2XML in order to produce part of the documents required to certify a smart card platform following the Common Criteria standard (http://csrc.nist.gov/cc/). The Common Criteria is an international certification scheme concerning the security of information technology products. This standard is supported by the United States and several countries in the European Union, like France, Germany and the United Kingdom. The highest assurance levels that can be reached in a Common Criteria certification require the developer of the IT product to provide a mathematical model of the security policies it enforces.

2.3 Deliverables

All Milestones have been successfully accomplished, respecting the prefixed timetable. All deliverables have been submitted in due time, or within a short delay. The following is the list of Deliverables for the period covered by this report.

n.	Deliverable Title	WP no.	Lead Partic.	Estim. person-month	Type	Security	Deliv. month
0.c	Second Self-Assessment Report	0	UNIBO	1	R	Pub	24
2.e	Presentational Stylesheets (formulae)	2	KUN	4	P	Pub	18
2.f	Presentational Stylesheets (proofs)	2	KUN	6	P	Pub	18
2.g	Tools for metadata extraction	2	UNIBO	4	P	Pub	18
4.b	**First MOWGLI Prototype** (browsing, rendering and consultation)	4	INRIA	15	P	Pub	18
4.c	Prototype for assisted annotation	4	UNIBO	7	P	Pub	18
4.d	LaTeX-based authoring tool	4	MPG	9	P	Pub	18
4.e	Extended LaTeX-authoring prototype	4	MPG	8	P	Pub	24
5.a	Overall Architectural Design of the distribution Model	5	DFKI	11	R	Pub	18
5.b	**Advanced MOWGLI Prototype** (distribution)	5	DFKI	16	P	Pub	24
R =Report, P =Prototype, Pub=Public							

3 Cost Breakdown

The summary of the costs for the reporting period is shown in Fig. 1. There is no sensible departure between the expected and the actual costs. The (unexpected) costs for Durable equipment for the site of Bologna is due to a misunderstanding already explained in the first progress report (in the proposal, the full cost of the equipment was charged entirely on the first year).

	UNIBO		INRIA		DFKI		KUN		MPG		Trusted Logic		Total	
Cost Category	Est	Act	Est	Act	Est	Act	Est	Act	Est	Act	Est	Act	Est	Act
Personnel	31208	37771	35582	72769	85462	73840	40151	50504	68424	68369	0	0	260827	
Durable equip.	0	2852	0	0	500	114	0	0	0	0	0	0	500	
Subcontracting	0	0	0	0	0	0	0	0	0	0	0	0	0	
Travel and subs.	7500	5994	7232	1984	8000	1817	10000	4025	11000	4390	1000	?	44732	
Consumables	0	0	0	0	0	0	0	0	0	0	0	0	0	
Computing	0	0	0	0	0	0	0	0	0	0	0	0	0	
Prot.of know.	0	0	0	0	0	0	0	0	0	0	0	0	0	
Other costs	5500	5561	985	1000	0	0	0	0	800	0	0	0	7285	
Subtotal	44208	52178	43799	75754	93962	75771	50151	54529	80224	72759	1000	?	313344	
Overheads	8842	10435	61599	89200	18792	15154	10030	10906	16045	14552	0	0	115308	
Total	53050	62613	105398	166928	112754	90925	60181	65435	96269	87311	1000	?	428652	

Figure 1: Costs in Euro for the reporting period: 03/2003-03/2004

4 Dissemination of Results

4.1 MoWGLI presentations

The Dissemination of Results is progressing particularly well, especially thanks to the intensive work of Prof. Bernd Wegner, Chair of the Project Exploitation Board.

The following is a list of the main talks and presentations of the Project held at International Events:

- A. Asperti. "An overview of MoWGLI". Mathematics on the Semantics Web, Monet Workshop. Technische Universiteit Eindhoven, Eindhoven (NL).
- R. Anghelache. "Hermes: A content oriented LaTeX to XML conversion/authoring tool"
  OpenMath Thematic Network Meeting, International University of Bremen, Bremen, Germany, November 2003
- R. Anghelache. "Hermes: A content oriented LaTeX to XML+MathML conversion/authoring tool" Albert Einstein Institute, Germany, March 2004
- Henk Barendregt (Nijmegen, NL) Computations via proofs, Mathematics, Logic and Computation, ICALP satellite workshop in honour of De Bruijn's 85th birthday, 4-5 July 2003, Eindhoven NL.
- L. Cruz-Filipe, Program Extraction from Large Proof Developments 11/9/2003, TPHOLs 2003, Rome, Italy.
- L. Cruz-Filipe, H. Geuvers, C-CoRN: The Constructive Coq Repository at Nijmegen, 30/4/2003, TYPES 2003 Workshop, Torino, Italy.
- L. Cruz-Filipe, What do we want from a documentation tool? 24/4/2003, Math in Coq Meeting, Paris, France
- H. Geuvers. The use of types to formalize mathematics. Workshop Mathematics on the Semantic Web, May 12-14 2003, Eindhoven, NL.
- H. Geuvers. C-CoRN, the Constructive Coq Repository at Nijmegen. Calcul formel, algorithmes certifis, preuves constructives Meeting held in the Centre International de Rencontre Mathmatiques de Luminy, 12-16th January, 2004.
- M. Jost, B.Wegner. EMIS 2003 - eine weltweite Kooperation zur elektronischen Mathematik-Information Sektionsvortrag, 9. Kongress der IuK-Initiative Universitt Osnabrck, March 2003.
- P. Libbrecht, "A presentation architecture for personalized content", Adaptive Hypermedia Workshop at the World Wide Web Conference, May 20 2004, Budapest.
- M. Kohlhase, " Towards Collaborative Content Management And Version Control For Structured Mathematical Knowledge", Mathematical Knowledge Management Conference, Bertinoro, Feb 2003.

- B. Wegner. MoWGLI - Mathematics in the semantic web. NA-MKM session at the Joint Mathematics Meeting 2004 Phoenix, Arizona, January 2004.
- B. Wegner. DML - un proyecto para instalar un archivo digital de matemticas de todo el mundo Plenary lecture, MATHINFO 2003 Universidad Holgun, Cuba, April 2003
- B. Wegner. Recent Achievements of European Projects for Digitising MathematicsPlenary lecture, Crimea 2003 Sudak, Crimea, June 2003
- B. Wegner. The way of mathematics into the information society Invited lecture Chulalongkorn University, Bangkok, July 2003
- B. Wegner Web Information on Mathematics and Symmetry Plenary Lecture, V: International Society for Symmetry in Culture and Science, Symmetry Festival Budapest, August 2003
- B. Wegner EMIS - La biblioteca electrónica, las bases de datos y los proyectos Invited Lecture, Escuela de Verano: El futuro delas publicaciones electrónicas San Lorenzo de El Escorial, Spain, Seprember 2003
- B. Wegner Proyectos Europeos para el establecimiento de la Sociedad de Información en Matemáticas Invited Lecture, Real Sociedad Matemtica Espaola, Encuentro de las Sociedades Matemticas Latinoamericanas Santiago de Compostela September 2003
- B. Wegner Proyectos Europeos para el establecimiento de la Sociedad de Informacin en Matemticas Plenary Lecture, APPOPT VI, Len, Nicaragua, March 2004-04-12
- B. Wegner Mathematics on its way into the information society Keynote speech, International Congress on Mathematics and its Applications Kuwait University, April 2004-04-12
- F. Wiedijk, Formal proof sketches & Lamport-style proofs, TYPES 2003, Torino, 2003-05-03.
- A. Asperti, The Science of Equality: Some Statistical Considerations on the Coq Library. Mathematical Knowledge Management Symposium, Edinburgh, November 25-29 2003.

4.2 Publications

Many scientific and technological aspects of the Projects have been presented to International Conferences, testifying the high technical quality of the work performed so far. The following is a list of main scientific publications related to MoWGLI for the current reporting period:

1. A.Asperti, B.Wegner. MoWGLI - An approach to Machine-Understandable Representation of the Mathematical information in Digital Documents. Electronic Information and Communication in Mathematics. ICM 2002 Satellite Conference, Beijing, China (revised paper). LNCS 2730, 2003.
2. A.Asperti, F.Guidi, L.Padovani, C.Sacerdoti Coen, I.Schena. Mathematical Knowledge Management in HELM. Annals of Mathematics and Artificial Intelligence 38(1):pp.27-46. May 2003.

3. F.Guidi. Searching and Retrieving in Content-based Repositories of Formal Mathematical Knowledge. Ph.D. Thesis in Computer Science, University of Bologna, Technical report UBLCS 2003-06, March 2003
`ftp://ftp.cs.unibo.it/pub/techreports/2003/2003-06.ps.gz`

4. F.Guidi, C.Sacerdoti Coen. Querying Distributed Digital Libraries of Mathematics. In Calculemus 2003, Aracne Editrice S.R.L., Therese Hardin e Renaud Rioboo editors, ISBN 88-7999-545-6, pp. 43–57.

5. L.Padovani. MathML Formatting. Ph.D. dissertation, University of Bologna, Technical Report UBLCS-2003-3, February 2003
`ftp://ftp.cs.unibo.it/pub/techreports/2003/2003-03.ps.gz`

6. C.Sacerdoti Coen, S.Zacchiroli. Brokers and Web-Services for Automatic Deduction: a Case Study In Calculemus 2003, Aracne Editrice S.R.L., Therese Hardin e Renaud Rioboo editors, ISBN 88-7999-545-6, pp. 43–57.

7. P. Audebaud, L. Rideau. TexMacs as Authoring Tool for Publication and Dissemination of Formal Developments. In Proceeding of User Interfaces for Theorem Provers (UITP'03), p. 14-32, 2003, Technical Report Nr.189, Institut fur Informatik, Albert-Ludwigs-Universitat Freiburg, ISBN 88-7999-547-2.

8. Y. Bertot, F. Guilhot, L. Pottier. Visualizing Geometrical Statements with GeoView.In Proceeding of User Interfaces for Theorem Provers (UITP'03), p. 33-45, 2003, Technical Report Nr.189, Institut fur Informatik, Albert-Ludwigs-Universitat Freiburg, ISBN 88-7999-547-2.

9. F. Guilhot. Formalisation en Coq d'un cours de gometrie pour le lyce. Journes Francophones des Langages Applicatifs, INRIA, Janvier 2004.

10. Romeo Anghelache: *Towards Collaborative Content Management And Version Control For Structured Mathematical Knowledge*, Second International Conference on Mathematical Knowledge Management (MKM 2003). LNAI Springer Verlag

11. W3C MathML Working Group: *Mathematical Markup Language (MathML) Version 2.0 (2nd Edition)* W3C Recommendation, 2003

12. G. Goguadze, E. Melis, V. Izhutkin, and Y. Isulanov. Interactively learning operations research methhods with activemath. In H.G.Bock, W. Domschke, R. Fahrion, M. Juenger, H. Kogelschatz, G.D. Liesegang, G. Reinelt, F. Rendl, and G. Waescher, editors, *Operations Research 2003, Annual Conference of the german Operations Research Society*, page 159, Heidelberg, 2003.

13. A. Gonzalez-Palomo, P. Libbrecht, and C. Ullrich. A presentation architecture for individualized content. In Paul de Bra, editor, *Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems, AH2003*, pages 87–98, 2003. See also about `http://wwwis.win.tue.nl/ah2003/`.

14. P. Libbrecht, C. Ullrich, and S. Winterstein. An efficient presentation-architecture for personalized content. In R. Tolksdorf and R. Eckstein, editors, *Proceedings of Berliner XML Tage 2003*, pages 379–388, 2003.

15. E. Melis, G. Goguadze, P. Libbrecht, and C. Ullrich. Wissensmodellierung und -nutzung in ActiveMath. *KI*, 1:12–18, 2003.

16. E. Melis and C. Ullrich. How to teach it – Polya-scenarios in activemath. In U. Hoppe, F. Verdejo, and J. Kay, editors, *AI in Education, AIED-2003*, pages 141–147. IOS Press, 2003.

17. E. Melis and M. Weber. Lessons for (pedagogic) usability design of elearning systems. In *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (eLearn-2003).*

18. F. Wiedijk, "Formal proof sketches", In: Wan Fokkink, Jaco van de Pol (eds.), 7th Dutch Proof Tools Day, Program + Proceedings, CWI, Amsterdam, 2003.

19. F. Wiedijk, "Bewijzen nalopen met de computer, de QED utopie" (Checking proofs with a computer, the QED utopia), Pythagoras, 43(3), 4-7, december 2003, Dutch

20. M. Giero, F. Wiedijk, "MMode, a Mizar Mode for the proof assistant Coq", report NIII-R0333, University of Nijmegen, 2003.

5 WorkPackages' evaluation parameters and assessment criteria

According to the contractual workplan, the activity of the second reporting period has been mainly focused on the following Workpackages ans Tasks:

WP0 Project Management
WP2 Transformation (Tasks T2.4, T2.5 and T2.6)
WP3 Metadata (Task T3.2)
WP4 Interfaces (Tasks T4.2, T4.3 and T4.4)
WP5 Distribution (Tasks T5.1 and T5.2)
WP6 Testing and Validation (all Tasks)
WP7 Information Dissemination and Exploitation

5.1 Transformation

This work package was devoted to the transformation (via stylesheets) of a low level, content description of mathematics (understandable by automatic applications for the mechanization of mathematics) into a human-readable presentational format. It covered both statements and proofs.

The deliverables fall in three categories:

1. XML exportation module for the COQ Proof Assistant (D2.a).

2. XSL Stylesheets and DTS's (D2.b-f).

3. Tools for automatic extraction of metadata (D2.g).

5.1.1 XML exportation module for the COQ Proof Assistant

The Exportation Module is by now a standard component of the Coq official distribution. In particular, it has been largely revisited and improved in view of the release of Coq V8. According to our own evaluation parameters, the task has been *very successful*.

5.1.2 XSL Stylesheets and DTS's

More than 15000 lines of XSLT have been developed, in charge of the full transformation process. In addition, tools for managing the complexity of this process have been developed, spanning from the automatic generation of notational stylesheets starting from a meta-specification of mathematical symbols, to software and web-services managing cascading transformations.

The application of the stylesheets, even on very large proofs, takes just a few seconds, testifying the actual feasibility of this technology (that was one of our main evaluation criteria).

Both Nijmegen and Trusted Logic started an extensive validation of the transformation technology, according to the pilot applications 6.1 and 6.3, respectively (see section 2.2.4).

Summing up, our preliminary evaluation of the task is extremely positive.

5.1.3 Tools for automatic extraction of metadata

As explained in deliverable D2g several prototype tools for metadata extraction have been developed in various languages, the only one providing reasonable performances being a C implementation based on the `flex` lexical analyzer.

According to the evaluation criteria this task is successful.

5.2 Metadata

Most part of MoWGLI's indexing and retrieving functionalities are based on the definition of a sophisticated metadata model and its exploitation via a suitable query language. The main bulk of these metadata are automatically extracted from documents, taking advantage of the content encoding of the mathematical information. These metadata are meant to give an approximation of the actual content (such as, say, list of identifiers in critical positions inside statements or expressions) suitable for fast searching and retrieving operations. Tools and languages exploiting these metadata have been developed as well (see e.g. the "on line" interface on the MoWGLI Web-site). These tools allow the user to ask typical queries based on content patterns. Answers are both accurate and fast, thus achieving our main evaluation criteria. The task is a clear success.

5.3 Interfaces

During this reporting period, the activity in this WorkPackage was mainly focused on Task4.2 (Consultation Engine) and and Task4.4 (LaTeX based authoring tool).

5.3.1 Web-based browsing and consultation interface

One of the key points of the MoWGLI project is to provide access to repositories of mathematical information by means of standard, commonly used Web browsers. The Web interface was conceived as a simple integration of components already developed in other MoWGLI

tasks, augmented with the requested features. Among the main components of this architecture there are the transformation stylesheets (applied "on the fly"), and the query engine.

The interface attempts to provide a uniform access to the different parts of the MoWGLI library, comprising both formal mathematical developments (from the COQ library) and physical papers (from Living Reviews), meeting the most important of our self-assessment evaluation criteria. As a proof of concept, and compared to the complexity of the goal, our achievements are an undeniable success.

In particular, other projects and communities have recently started to get interested and to adopt MoWGLI technologies, such as the NuPRL Group of James Caldwell.[2] A link to a working prototype can be found at: `http://helm.cs.unibo.it/NuPRL/index.html`. The prototype allows browsing an XML version of the NuPRL library, hosted at the University of Wijoming and rendered on-the-fly by the MoWGLI web-services running on the MoWGLI server in Bologna. The stylesheets used to provide mathematical notation to the NuPRL library share a large amount of code with the stylesheets used for rendering the Coq library, and they are now maintained by the NuPRL team.

5.3.2 LaTeX-based authoring tool

This tool, called Hermes, provides functionalities for creating content-based mathematical information from standard digital repositories by means of a suitable LaTeX-based authoring system.

ZentralBlatt for Mathematics has expressed a clear interest in migrating their database of mathematical abstracts to XML/MathML using Hermes. We are currently testing the prototype for this purpose (more on the final report).

5.4 Distribution

This WorkPackage was devoted to the overall architectural design of the distribution model, its implementation and integration with the consultation engine. The expected functionalities were just meant to provide a *physically* distributed library with a single *logical* view and significant searchability features.

This has been done and works in a pretty satisfactory way. Some partners have however raised the need for a more sophisticated architecture, comprising versioning, that goes far beyond the goals of MoWGLI. The new architecture has been already anticipated in deliverable D5a, and consists of an infrastructure for collaborative content management and version control for structured mathematical knowledge allowing multiple users to work jointly on mathematical theories with minimal interference.

In particular, report D5a describes an API covering all functionalities needed to realize a CVS-like version control and distribution model. The proposed architecture extends the CVS architecture in two ways, motivated by the specific needs of distributed management of structured mathematical knowledge on the Internet. On the one hand the one-level client/server model of CVS is generalized to a multi-level graph of client/server relations, and on the other hand the underlying change-detection tools take the math-specific structure of the data into account.

Although the implementation of the above architecture goes far beyond the aims of MoWGLI, its design is already a major achievement of our project.

[2] `http://www.cs.cornell.edu/Info/Projects/NuPrl/nuprl.html`