INFORMATION SOCIETY TECHNOLOGIES (IST) PROGRAMME

Contract for:

Shared-cost RTD (FET Programme)

Annex 1 - “Description of Work”

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

Date of Preparation of Annex 1: 17 october 2001

Operative commencement date of contract:

Contents

1 Project summary

1.1 Objectives

The World Wide Web is already the largest resource of mathematical knowledge, and its importance will be exponentiated by emerging display technologies like MathML. However, almost all mathematical documents available on the Web are marked up only for presentation, severely crippling the potentialities for automation, interoperability, sophisticated searching mechanisms, intelligent applications, transformation and processing. The goal of the project is to overcome these limitations, passing form a machine-readable to a machine-understandable representation of the information, and developing the technological infrastructure for its exploitation. MOWGLI builds on previous "standards" for the management and publishing of mathematical documents (MAthML, OpenMath, OMDoc), integrating them with different XML technology (XSLT, RDF, ...).

1.2 Description of work

The goal of the project is to provide a comprehensive description, from content to metadata, of a given field of knowledge (in our case mathematics), in order to enhance its accessibility, exchange and elaboration via the World Wide Web. MOWGLI will make an essential use of standard XML technology and aspires to become an example of "best practice" in its use, and a leading project in the new area of the Semantic Web.
In particular, we shall deeply explore the potentialities of XML in the following directions:
Publishing. XML offers sophisticated publishing technologies (Stylesheets, MathML, SVG, ...) which can be profitably used to solve, in a standard way, the annoying notational problems that traditionally afflict content based and machine-understandable encodings of the information.
Searching and Retrieving. Metadata will play a major role in MOWGLI. New W3C languages such as the Resource Description Framework or XML Query are likely to produce major innovative solutions in this field.
Interoperability. Disposing of a common, machine understandable layer is a major and essential step in this direction.
Distribution. All XML technology is finally aimed to the access of the Web as a single, distributed resource, with no central authority and few, simple rules.

MOWGLI builds on the solid ground already provided by previous European projects (Such as OpenMAth and Euler) and several XML dialects for the management of mathematical documents (MathML, OpenMAth, OMDoc, ...). All these languages cover different and orthogonal aspects of the information; our aim is not to propose a new standard, but to study and to develop the technological infrastructure required for taking advantage of the potentialities of all of them.

1.3 Milestones and expected results

First MOWGLI prototype (month 18), supporting browsing, rendering and on-line consultation of large repositories of (content-based) mathematical knowledge. The translation from content to presentation will be done via suitable notational stylesheets.
Advanced MOWGLI prototype (month 24), supporting distribution, indexing, searching and retrieval (based on a sophisticated metadata model).
Final MOWGLI prototype (month 30). Result of validation.

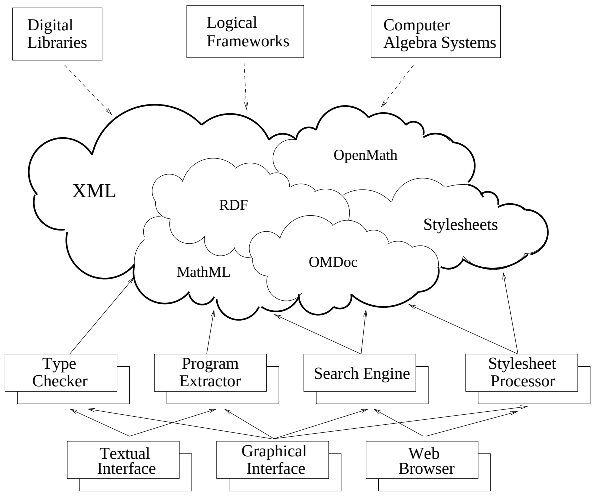
2 Project objectives

The new frontier of Content Based Information Systems is the so called "Semantic Web" (see [3]). Associating meaning with content or establishing a layer of machine understandable data will allow automated agents, sophisticated search engines and interoperable services and will enable higher degree of automation and more intelligent applications. The ultimate goal of the Semantic Web is to allow machines to share and exploit knowledge in the Web way, i.e. without central authority, with few basic rules, in a scalable, adaptable, extensible manner. However, the actual development of the Semantic Web and its technologies has been hindered so far by the lack of large scale, distributed repositories of structured, content oriented information. The case of Mathematical knowledge, the most rigorous and condensed form of knowledge, is paradigmatic. The World Wide Web is already now the largest single resource of mathematical knowledge, and its importance will be exponentiated by the emerging display technologies like MathML. However, almost all mathematical documents available on the Web are marked up only for presentation (in this respect, current practice in MathML improves on, but does not fundamentally differ from the older paper-oriented markup schemes like LaTeX or Postscript). A consequence of this is that the online material is machine-readable, but not machine-understandable, severely crippling the possibility to offer added-value services like

- Preservation of the real informative content in a highly structured and machine understandable format, suitable for transformation, automatic elaboration and processing.
- Cut and paste on the level of computation (take the output from a Web search engine and paste it into a computer algebra system).
- Automatic proof checking of published proofs
- Semantical search for mathematical concepts (rather than keywords).
- Classification: given a concrete mathematical structure, is there a general theory for it?

Due to its rich notational, logical and semantical structure, mathematical knowledge is thus a main case study for the development of the new generation of semantic Web systems. The aim of the proposed project is both to help in this process, as well as pave the way towards a really useful virtual, distributed, hyper-textual resource for the working mathematician, scientist or engineer. All modern sciences have a strongly mathematicised core, and will benefit. The real market and application area for the techniques developed in this project, apart from the obvious realm of education, lies with high-tech and engineering corporations that rely on huge formula databases. Currently, both the content markup as well as the added-value services alluded to above are very underdeveloped, limiting the usefulness of the vital knowledge. The infrastructure and knowhow needed for mining this information treasure and obtaining a competitive edge in development is exactly what we are attempting to develop in our project.

Several languages have been already proposed for the management of mathematical information on the Web, both for publishing, communication and archiving purposes: most notably, MathML [11], OpenMath [15], OMDoc [14]. Other languages must be also considered for definition and specification of Metadata, such as the Dublin-core System [5], or the Resource Description Framework [16]. All these languages, which tend to cover different and orthogonal aspects of the management of mathematical documents, must be eventually taken into account for the ambitious goal of our project. One of our aims is actually the definition of a modular architecture which could exploit the distinctive potentialities of each one of these languages, integrating them into a single application. The integration is in this case facilitated by the fact that all the languages mentioned are particular instances of XML, providing the opportunity to use standard XML technology, and in particular XSL Transformations or stylesheets [18], to pass from one language to the other.



The fact of encoding also the microscopic, logical level of mathematics opens the possibility to have completely formalised subsystems of the library, which could be checked automatically by standard tools for the automation of formal reasoning and the mechanisation of mathematics (proof assistants and logical frameworks [8, 9]). At the same time, any of these tools could be used as an authoring system for documents of the library, by simply exporting their internal libraries into XML, and using stylesheets to transform the output into a standard, machine-understandable representation, such as MathML content markup or OpenMath.
The precise formal content can still be preserved by the machinery of Xlinks [17]. Moreover, stylesheets can be also used to solve the annoying notational problem that usually afflicts formal mathematics, providing a simple way for adding user-defined styles and notations.

So, our approach leads to a natural integration of proof assistant tools and the Web. In this integration, the emphasis is just on "content": we do not try to link directly the specific applications to the Web, that would be a major mistake, for obvious modularity reasons. On the contrary, we adopt XML as a neutral specification language, and then we merely work on XML-documents, forgetting the underlying application. In this way, similar software tools can be applied to different logical dialects, regardless of their concrete nature. Moreover, if having a common representation layer is not the ultimate solution to all inter-operability problems between different applications, it is however a first and essential step in this direction. Finally, this "standardisation" process should naturally lead to a substantial simplification and re-organisation of the current, "monolithic" architecture of logical frameworks. All the many different and often loosely connected functionalities of these complex programs (proof checking, editing, search and consulting, program extraction, and so on) could be clearly split in more or less autonomous tasks, and could be developed by different teams, in totally different languages. This is the new, "content-based" architectural design of future systems.

3 Participant list

Part. Role	Part. no.	Participant name	Participant short name	Country	Date enter Project	Date exit Project
C	1	Università degli Studi di Bologna	UNIBO	Italy	Start of project	End of project
P	2	Institut National de Recherche en Informatique et Automatique	INRIA	France	Start of project	End of project
P	3	German Center for Artificial Intelligence	DFKI	Germany	Start of project	End of project
P	4	Katholieke Universiteit Nijmegen	KUN	Netherlands	Start of project	End of project
P	5	Max-Plank-Gesellshaft zur Foerderung der Wissenshaften e.V.	MPG	Germany	Start of project	End of project
P	6	Trusted Logic S.A.	Trusted Logic	France	Start of project	End of project

4 Contribution to programme/key action objectives

The Project contributes to build a user friendly information society, and in particular it meets the following general objectives of this program:

1. (multimedia content) confirming Europe as a leading force in this field, realizing the potential of its creativity and culture;

2. (essential technologies and infrastructure) enabling technologies which are the foundations of the information society, driving their development, enhancing their applicability, and accelerating their take up in Europe.

As a matter of fact, the project is based in an essential way on the use of most part of the recent recommendations of the World Wide Web Organization for Web publishing and human-computer interaction (XML, XSL, XLL, Namespaces, MathML, RDF, etc.). In particular, we aim to prove how all these specifications naturally fit together, when trying to build a full, integrated description (comprising content, notation, metadata, etc.) of a given field of knowledge. At our knowledge, the project is the first of the kind, and could become a paradigmatic example in the integrated use of these technologies.

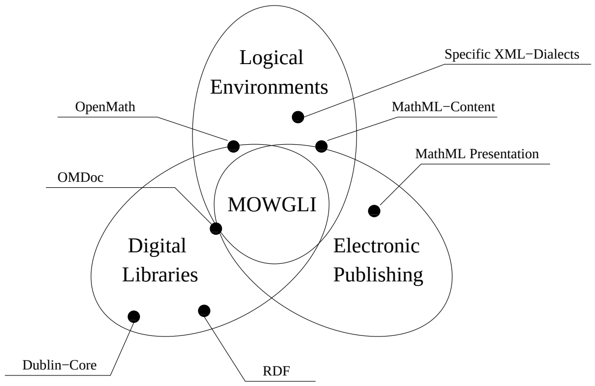
The project also addresses most of the issues of the multimedia content key action, namely: electronic publishing, digital heritage and cultural content, education, information access, filtering and handling. Actually, all these aspects are and must be covered in our project, in order to reach our objectives. In particular, the educational potential of our system should not be neglected either: it could become an essential tool for a wider and more friendly dissemination of mathematical knowledge. For instance, if supported by a suitable technology, proving theorems in a proof assistant could be as amusing as playing a video game. We imagine bunches of young researchers contributing to the free development of the library for the mere gratification of seeing their name as actual editor (or, why not, original author) of a specific fragment.

Finally, the project is particularly related to the specific key-action III.2.3 (access to scientific and cultural heritage). In fact, the aim of our system is exactly to improve access by students and professionals to the fast-growing mathematical knowledge base, allowing mathematical documents to be retrieved, served, and processed directly on the Web. Moreover, our system is meant to be compatible with most of the existing tools for the mechanisation of mathematics and the automation of formal reasoning (proof assistants and logical frameworks). The possibility to build coherent sub-libraries of formal mathematical developments would provide an essential (and unique) added value to the library itself, making of Europe a leader in this area.

Maybe, having the possibility to process, analyse and elaborate mathematical structures as data, the time will come when we shall finally be able to start a completely new and exciting field of research on mathematics: namely a scientific, empirical study on the *real* structure of mathematical entities, and the 'way of thinking' of mathematicians.

5 Innovation

The main technical novelty of the project is in its synergy of different scientific communities and research topics: digital libraries, Web publishing, logical environments.



From the point of view of Web publishing, our project is the first attempt to provide a comprehensive description, from content to metadata, of a given field of knowledge (in our case mathematics), in order to enhance its accessibility, exchange and elaboration via the World Wide Web. To this aim, we shall put to use most of the technologies recently introduced by the W3C: XML, DOM, XSL, XLL, Namespaces, MathML, RDF, etc. From this respect, the project is first of all a complex test for all these technologies, and should hopefully become an example of 'best practice' in their use. Note that the final architecture is likely to be extendible to other fields of structured information: the emphasis on mathematics is motivated by the fact that, due to its complex interplay between content, structure and notation, it provides a major case study for Web-based information systems (it is not a case that MathML has been one of few instances of XML completely developed under the aegis of the World Wide Web Consortium).

From the point of view of digital libraries, our work is aimed at exploiting all the potential functionalities offered by the Web, and in particular a more integrated use of its browsing and searching facilities. The library is not merely seen as a more or less structured collection of texts, but as a virtual structure inside which we can freely navigate, jumping for instance from an entity to its definition, or peeping inside some information at deeper and deeper levels of details (such as different levels of detail of a proof). This is similar to what we currently do with HTML texts, but in order to enhance the effectiveness of the consultation, we clearly need a good metadata model of the information. Moreover, in such an integrated view, it is hardly conceivable to just apply some 'general purpose' metadata model (like the Dublin Core system, say): the metadata model must be eventually specialised to the actual structure of the information it is supposed to model (and more structure we have on the information, more relevant metadata we can usually infer on the document). For instance, metadata could contain the whole signature of a given module of mathematical knowledge. The usual motivation for keeping metadata simple and general is that it is usually difficult to add this information by hand; but in our case a large part of the metadata is supposed to be extracted automatically by the (structured) text itself, allowing for pretty complex metadata models.

Finally, a main aspect of our project is the integration with current tools for the automation of formal reasoning and mechanisation of mathematics (proof assistants and logical frameworks). This integration has a mutual benefit. From the point of view of the mathematical library, the first and fundamental role of these systems is that of providing friendly authoring tools (for instance, our 'core' library will be automatically extracted from existing libraries of these systems). The relevance of this point should not be underestimated: as a matter of fact, the main reason for the failure of complex markup modellings is usually the lack of suitable authoring tools (it is often painful to add the markup by hand). Of course, they can also provide other functionalities (like automatic proof checking) on fragments of the library (typically, the fragments generated by the tool itself, in its specific logical dialect). These additional functionalities may be especially relevant for industrial applications, e.g. in the context of IT security evaluation standards like the Common Criteria standard ([4]). In its highest assurance level, this standard requires the development of formal models of the IT product under evaluation, as well as mechanized proofs that it meets its security objectives. Such models and proofs must be published in a format that can be easily readable and understood by security evaluators. Hence, there is a strong need from software industry to be able to produce such documentation directly from the models introduced in the proof assistant, and to link it with documents describing the IT product, etc.

On the other side, there is a compelling need of integration between the current tools for automation of formal reasoning and mechanisation of mathematics and the most recent technologies for the development of Web applications and electronic publishing. XML, which is rapidly imposing as a pivotal technology in the future development of all Internet applications, and the main tool for representation, manipulation, and exchange of structured information in the networked age, looks as a natural, almost mandatory, choice for modelling the information.

In this way, we just obey to the very primitive commandment of the Web: make your information available. Currently, libraries in logical frameworks are usually saved in two formats: a textual one, in the specific tactical language of the proof assistant, and a compiled (proof checked) one in some internal, concrete representation language. Both representations are obviously unsatisfactory, since they are too oriented to the specific application: they restrict the access of the libraries to the users of the given application, and at the same time they are too sensible to the evolution and the maintenance of the application itself. On the other side, as soon as the information is put in a standard format on the Web, *any* kind of research becomes virtually possible, and *anybody* could start developing his own *spider* for implementing his own searching requirements. This is clearly a major improvement w.r.t. the present situation. Currently, you must not only rely on the searching facilities offered by the specific applications, but even if you would wish to implement your own searching algorithm, you would be prevented by the simple reason that the information is not accessible (in any reasonable sense of the word).

The project builds on the solid ground provided by several existing XML-based languages for the management of mathematical documents such as MathML, OpenMath and OMDoc. Each of these markup languages covers a different aspect of the information. Our aim is not to propose a new language, but to study and to develop the technological infrastructure required to integrate all these languages together, in order to take advantage of the specific features of each of them.

6 Community added value and contribution to EC policies

The Project relies in an essential way on the convergence of information processing, communication and new media, meeting the main policy issue of the IST programme. In fact, due to its very nature, mathematical information is particularly suited to an integrated analysis under the different perspectives of elaboration, communication and (hyper)media publishing, providing a main arena for innovative experiments and solutions, especially in view of the new technological frontier of the Semantic Web.

The Project also tries to answer to a clear need for wider interoperability and coherence in the realm of mathematics over the Web, especially for educational, scientific and commercial purposes.

Electronic textbooks are rapidly becoming a main tool for education and knowledge dissemination. Electronic textbooks must be interactive, allowing intercommunication between the text and scientific software and graphics. This is very hard to achieve starting form a mere presentational description of the information, clearly requiring a sophisticated semantical description of the content.

Similarly, the academic and commercial research communities generate large volume of dense scientific material. Increasingly, research publications are being stored in databases, especially for those areas of physics and mathematics where academic journal prices have been growing at an unsustainable rate. Still, however, the information is stored in a format which is machine readable but not machine understandable; in particular it is not suitable for any form of elaboration other than its rendering. MOWGLI's content and metadata levels are exactly meant to address complex elaboration requirements, and to facilitate the maintenance and operability of large document collections, for which automatic searching and indexing are crucial.

Corporate and academic scientists and engineers also use technical documents in their work to collaborate, to record results of experiments and computer simulations, and to verify calculations. The Web is, potentially, the natural media for sharing this information; MOWGLI is meant to provide the basic technology to transform this potentiality into an actual possibility.

Commercial publishers are also involved with mathematics on the Web at all levels from electronic versions of printed books to interactive textbooks and academic journals. In this case, MOWGLI offers support for advanced capabilities, such as browsing, interactivity, presentational and stylistic customisation, and advanced searching and retrieving features.

The project deals with problems traditionally belonging to different scientific communities: digital libraries, Web publishing, automation of mathematics and computer aided reasoning. Any serious solution needs a coordinated effort of all these groups and a synergy of their different expertise. The members of the consortium have been also carefully chosen for their experience in the above mentioned areas. From this interaction, we expect to develop new technologies and solutions, *fostering innovation* towards the construction of the Semantic Web.

MOWGLI contributes in the *spread of information and know-how*, since new information technologies will be applied to realms such as education and publishing, which are traditionally not very inclined to innovation. More *competitiveness* will be acquired both by the end users of the system and by the technology providers, which will eventually profit by the expertise gained by applying new and emerging technologies to the huge and compelling problem of the management of mathematical documents over the Web.

By opening new perspectives in the domains of interactive publishing and education, MOWGLI suggests *new job-profiles* in these domains, and *increases job opportunities* in the Information Society sector.

Finally, the *social and cultural cohesion* of Europe is eventually strengthened by the creation of a large distributed repositories of scientific knowledge.

7 Contribution to Community Social Objectives

The quality of life, in a civilised country, strictly depends on the quality of its services, and in particular on the attention devoted to those services aimed to preserve, increase and make accessible to a wider audience its cultural and scientific heritage. Building a user friendly Information Society, with particular emphasis on digital heritage, cultural content and education is indeed a main social objective of the European Community.

The new Information Society is essentially based on the convergence of information, communication and networking technologies and takes advantage of infrastructures like the Internet and the Web.

Our project builds on these grounds, to create the technological infrastructure required for the creation and maintenance of a digital knowledge base of *structured* mathematical information, universally and seamlessly accessible to all people, and in particular to students and professionals, through interoperable, dependable and affordable products and services.

From the educational point of view, our system could easily become a main tool for a wider and more friendly dissemination of mathematical knowledge. Indeed, its interactive nature, and the possibility to access single information units, and possibly applying them, provides a more operational and far less abstract comprehension of mathematical entities, and should naturally induce the user to play with the knowledge base, assembling components in the development of new theorems and results.

From the point of view of employment and development of individual skills, there are two different aspects to be considered, according to the *objectives* and the *methodology* of the project.

Form the methodological point of view, the project makes an essential use of technologies which are the very foundations of the information society, driving their development, enhancing their applicability, and accelerating their take up in Europe. In particular, we shall build on most of the recent recommendations of the World Wide Web Consortium (W3C), like XML, DOM, XSL, XLL, MathML, RDF, etc. testing their applicability to the definition of a comprehensive, integrated description of a given field of knowledge.

This kind of tools and techniques is of crucial importance in the development of the Web and of the Information Technology of the next years, and our project could play a major role in their dissemination in Europe, and in training people in their use.

On the other side, MOWGLI itself could become a major source of inspiration for the re-invention of existing activities, and in particular for the exploitation of new business in the publishing market. Currently, digital journals just offer purely textual objects: our project could help to define new market possibility for content-based publishing (*i.e.* structured, possibly formal mathematical developments), with all the potentialities offered by this approach, from browsing facilities, to the personalisation of the style, from enhanced forms of searching, to more or less arbitrarily complex forms of elaboration.

Our system also opens new perspectives on the mechanisation of mathematics and the automation of formal reasoning. The growing complexity of advanced technological projects (in the areas of electronic and avionic engineering, for example) has recently arose a renewed interest in formal methods. From this respect, our project contributes to build an essential infrastructure for the exploitation of these methods, providing a major help for professionals, and fostering their skills.

Let us finally remark that the project has no negative impact on the natural environment. We could even claim a few benefits, related to the electronic distribution of documents, such as the minimisation in the use of means of conveyance and the saving of paper.

The possible customisation via suitable style-sheets of *structured* electronic documents could also meet the needs of particular users, such as disabled (in particular, disabled students) or elderly (as retired researchers who would like to continue their professional activity), providing at the same time comfortable working conditions (*e.g.* at home). Another general design requirement is the ability to render mathematical material in other media such as speech or braille, which is extremely important for the visually impaired.

In conclusion, the aim of our work is to contribute to the creation of the next generation of user-friendly, dependable and interoperable general-interest services, meeting user demands for flexible access, for everybody, from anywhere, at any time.

8 Economic development and Scientific and Technological Prospects

8.1 Identification of exploitable results

MOWGLI is meant to develop, evaluate, fine-tune and deliver a complete prototype supporting friendly and interactive access to huge, distributed repositories of content-based mathematical knowledge. The purpose of the project is to overcome some of the main obstacles that currently hinder a wider dissemination of Information Technologies in the mathematical and research community (comprising both professionals and students). This is intended as a public service for the whole community of users, and we expect no direct commercial profit by our work.

Of course, the core technologies and languages developed in MOWGLI may be exploited in a lot of different ways, according to the nature and exploitation plan of each partner.

In particular, it should be clear that having a coherent, distributed library of structured mathematical knowledge, saved in a clearly defined, well documented and application independent format, is just the starting point of a wider, potentially enormous process. Building on the grounds provided by MOWGLI, a lot of different services can be imagined, and independently developed by different organisations. Inside the project, we shall just give a few hints in this direction, supporting advanced and interactive forms of searching and navigation, as well as automatic validation facilities, for suitable fragments of the library.

8.2 Target Market

The main target market of MOWGLI is education. Here, the potentiality offered by a semantic description of the information for the development of interactive, electronic textbooks are enormous. In particular, the MOWGLI prototype could become an essential tool for a wider and more friendly dissemination of mathematical knowledge.

A second, potential market is Publishing. Digital libraries currently suffers from clear limitations due to the machine-readable but not machine understandable encoding of the information. No other kind of elaboration than displaying or printing is usually possible. There is no possibility of cut-and paste, formal checking, complex searching. For instance, we cannot write a spider going around searching for a proof with a given template, because there is not structure, and in particular no semantics inside the text.

A third market are industries specialised in formal verification and security. These industries are usually asked to check a given system or protocol for security issues or formal validation of functional/temporal behaviour. The result of the formalisation work and its automatic validation is typically in an internal format which is obscure even to experts. Studying the possibility to automatically generate a more friendly presentation in usual mathematical notation is thus a major topic.

8.3 Consortium's Dissemination and Exploitation Strategy

According to the Project Workplan, a considerable effort will be devoted to the dissemination/exploitation of the results. During the first six months, the consortium will undertake the development and execution of a workplan to achieve successful dissemination of the project results. In particular, the very nature of the project suggests to involve the largest community of professionals during the modelling phase. Here, we shall be open to any contribution, even soliciting them, still keeping final technical decisions inside the Project.
The foreseen dissemination steps include:

1. Preparation of an "information pack", including a folder and CDROM with demos.
2. Presentation of MOWGLI to national and international conferences on I.T.Technologies.
3. Presentation of MOWGLI to thematic conferences on Education, Digital Libraries, Formal Verification and so on.
4. Advertising to actors/institutions which may be direct users or may channel the information to final users (such as the World Wide Web Consortium).
5. Technical papers and articles to be sent to specialised reviews.
6. Management of the MOWGLI Web site.
7. Dissemination of technical achievements in specific scientific communities, such as XML-related interests groups.

Management of the Exploitation and Dissemination activities is a specific Work-package that will be lead by the Project Exploitation Board. It is planned that all partners will nominate qualified persons as exploitation managers to coordinate their own exploitation activities. The Exploitation Board will be in charge of the preparation of the exploitation and dissemination plan.

9 Workplan

9.1 Workpackage list

The work plan is structured in the following work packages and Tasks.

WP0 Project Management

WP1 State of the art and Requirements Analysis

- **T1.1** Mathematics and the Web;
- **T1.2** Structured and Formal Mathematics;
- **T1.3** Metadata;
- **T1.4** Searching and Retrieving;
- **T1.5** Distribution;
- **T1.6** Document Authoring

WP2 Transformation

- **T2.1** XML Exportation;
- **T2.2** Stylesheets to Intermediate Representation;
- **T2.3** Proof Transformations;
- **T2.4** Automatic extraction of metadata;
- **T2.5** Presentational Stylesheets (HTML/MathML);
- **T2.6** Automatic Generation of Proofs in Natural Language;

WP3 Metadata

- **T3.1** Use, meaning and classification;
- **T3.2** Modelling;

WP4 Interfaces

- **T4.1** MathML rendering/browsing engines;
- **T4.2** Consultation Engine (archiving, searching and retrieving);
- **T4.3** Assisted Annotation;
- **T4.4** LaTeX-based authoring tool;

WP5 Distribution

- **T5.1** Architectural Design of the Distribution Model;

T5.2 Prototype implementation;

T5.3 Integration with the Consultation Engine;

WP6 Testing and Validation

T6.1 Pilot Application: education;

T6.2 Pilot Application: certified code for Java cards;

T6.3 Pilot Application: electronic publishing

WP7 Information Dissemination and Exploitation

9.2 General description

The previous work packages are not strictly sequential.

WP1 should be reasonably short; apart from a few topics requiring a deeper analysis (Tasks 1.3-5), this phase is essentially meant to rapidly reach a good level of inter-operability among the different sites.

Most part of the work is based on the possibility to have at our disposal, and as soon as possible, large collections of documents encoded with semantic markup. One strategy is the import of material (e.g. journal articles) written in LaTeX. The development of a suitable LaTeX based authoring tool (Task 4.4) will need to begin immediately, as an appropriate semantic encoding in LaTeX has to be developed first. The delivery of the first prototype of the authoring tool is scheduled for month 18.

A more rapid way to get meaningful repositories of fully structured mathematical knowledge is by exporting them from the available libraries of Logical Frameworks and Proof Assistants (Task 2.1). The intelligence contained in the exported XML files should reflect the requirements defined in the previous work package, requiring a deep analysis of the markup model. After six months from the beginning of the project we plan to have a first prototype of the Exportation Module and a first draft of the Document Type Descriptor for the low, logical level. This is our first Milestone.

At this point we may start the study of the intermediate format of the information, and the implementation of the stylesheets performing the transformation (tasks T2.2-3). This part of the work is expected to be essentially completed after one year (second Milestone). Since a strong feedback is expected with presentational issues, we plan to begin the development of presentational stylesheets around month 9. In turn, the need of rapidly have at our disposal good presentational engines suggests to begin their development as soon as possible (task 4.1).

In parallel with these transformation issues, we shall start the study and classification of metadata, and their concrete modelling.

Summing up, at the end of first year we plan to have:

- a formal Document Type Descriptor of the intermediate level;
- a bunch of stylesheet performing the transformation to intermediate representation, both for formulae and proofs;
- a detailed report on metadata;
- a first prototype model of metadata (in RDF format);

- a prototype MathML-viewer.

During the first half of second year we shall 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, requiring the metadata model), and of the functionalities for assisted annotation in natural language of the documents (task T4.3, requiring both the MathML-viewer, and a detailed description of the intermediate level).
3. overall architectural design and first prototype implementation of the distribution model (task T5.1-T5.2).
4. finalize the first prototype of a LaTeX based authoring tool (Task 4.4)

The second half of the second year is devoted to the completion of the previous tasks, and to their integration inside a single, compound application. Around this time we shall also start a detailed validation of the application, according to three pilot applications:

1. Formalisation of a full undergraduate course in algebra or analysis for didactical purposes.
2. Formalization of (part of) the process of loading, verifying and executing an applet into a smart card. This application will provide a case study close to both information technology (IT) industry and Computer Science research, where the presentation and layout needs are not exactly the same as in Mathematics. The example concerns the representation of different abstract state machines, transition systems, typing calculus, and program code. Such concepts are pragmatic use cases of the formal concepts that usually appears in security evaluations of IT products and Computer Science articles.
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 customisable content presentation. Several articles will be processed to test scope, functionality, and user friendliness of the authoring tool developed in Task 4.4. The tool will be refined and the mathematical semantics covered extended. The suite of articles will be used to show benefits of automated annotation and cross-linking between related mathematical concepts.

The last six months are mainly devoted to testing, debugging, validation, dissemination of results and exploitation plans.

9.3 Work Package descriptions

9.3.1 Project Management

work package number: 0 - Project Management
Start date or starting event: month 0 - month 30

Participant Number:	1	2	3	4	5	6
Person-month per participant:	8	0	0	0	0	0

Objectives: General project management and coordination. Knowledge and skills transfer between consortium members. Relation to the European Commission.

Description of work: The Project management will be assured by the following relevant roles: a) Project Manager, b) Exploitation manager, c) Work-package Leaders d)Technical Contributors, and by the following Project Bodies: a)Project Coordination Committee (PCC), b) Project Exploitation Board (PEB) and e) Work-package Teams.
The **Project Manager** chairs the PCC. The mandate of the PCC is to represent the Project, report to the Commission, monitor overall performance of the project, ensure accomplishment of the technical objectives, promote project visibility, promote dissemination of project results in the relevant international forums, promote acceptance of project results, administer project resources and monitor project spending. The **Exploitation Manager** will be responsible for coordinating dissemination and exploitation activities undertaken by the Project Exploitation Board (PEB) in close cooperation with the Project Coordination Committee (PCC).
Information flow within the Project will be ensured by exchange of internal technical papers, notification of relevant new publications technologies or standards, and reports from external meetings. All technical documentation generated by the project will be exchangeable in electronic format, according to a set of guidelines to be agreed at project start-up. The project Manager will enforce adherence to these guidelines. Only strictly formal correspondence will be exchanged by ordinary mail and telefax. Urgent correspondence over e-email will be sent with a request for explicit acknowledgement.
The Coordinating Partner will be responsible to prepare and maintain a Web page of the project and a CVS repository (also available via Web).

Deliverables: Cost Statements and Project Reports (month 12, 20, 30).

D0.a Self Assesment parameters and criteria.

D0.b First Self-assesment Report.

D0.c Second Self-assesment Report.

D0.d Final Self-assesment Report.

Milestones and expected results: Main milestones are the periodic meetings, at month 6, 12, 20, 24, 30.

9.3.2 Requirement Analysis

work package number: 1 - Requirement Analysis
Start date or starting event: month 0 - month 6

Participant Number:	1	2	3	4	5	6
Person-month per participant:	3	4	3	6	3	1

Objectives: Definition of the application scenarios. Precise articulation of all the functionalities required by the system, and all possible expected interactions with documents. Overall requirements of the distribution model of the library.

Description of work: The work plan is naturally organised in subtasks according to the different basic kind of interactions and manipulation to be considered, namely:

T1.1 Mathematics and the Web. State of the art, standards and tools.

T1.2 Structured and Formal Mathematics. Delineation and layering of Semantic Components. Requirements for the interaction with tools for the automation of formal reasoning.

T1.3 Metadata. Classification and data mining for content-based mathematical documents, and key architectural guidelines for the metadata model.

T1.4 Searching and Retrieving. State of the art, use cases and application scenarios.

T1.5 Distribution. Distributed document repositories and peer-to-peer interoperability.

T1.6 Document Authoring. State of the art, use cases and application scenarios.

Part of the activity, covering Tasks T1.1-2-6 should be concluded within the first three months, with the preliminary report D1.a. The main aim of this phase is to rapidly reach a good level of inter-operability among the different sites, by implementing a suitable politic of knowledge and skills transfer between the members of the consortium (short visits and/or small thematic workshops, according to the case).

However, some delicate issues, such as Metadata, Searching and Retrieving and Distribution (Task T1.3-4-5) will eventually require a deeper analysis (deliverables D1.b and D1.c).

WP1 will be eventually closed during the first meeting of the Project (month six), when all the reports will be discussed and approved.

Deliverables:

D1.a Preliminary Report on Application Scenarios and Requirement Analysis.

D1.b Structure and Meta-Structure of Mathematical Documents.

D1.c Distributed Electronic Libraries: development, archiving, retrieving.

Milestones and expected results: The preliminary report D1.a is not a real milestone: it should be considered as a first internal draft summarising the main functionalities required by the system and the basic intelligence to add to documents, as markup and/or metadata, to meet these requirements. All these issues will be eventually detailed in D1.b and D1.c, which are real Milestones, since the rest of the work will be largely driven by them.

9.3.3 Transformation

work package number: 2 - Transformation
Start date or starting event: month 3 - month 21

Participant Number:	1	2	3	4	5	6
Person-month per participant:	13	4	3	11	3	0

Objectives: This work package is devoted to the complex issue of transforming a low level, content description of mathematics (understandable by automatic applications for the mechanization of mathematics) into a human-readable presentational format. It covers both statements and proofs. The transformation will be decomposed in a sequence of intermediate steps, for modularity reasons. All transformations will be implemented by means of XSLT-stylesheets. Stylesheets will be simple, modular, and easily composable. All the transformation process should be independent from any specific application.

Description of work: The work package is articulated in the following tasks.

T2.1 XML exportation. The task is devoted to the translation of the standard library of the COQ Proof assistant into a suitable XML dialect, and to the definition of a low-level DTD for the terms of the Calculus of Inductive Construction (the logical system used by COQ).

T2.2 Stylesheets to intermediate representation. Implementation of a bunch of stylesheets transforming the low-level logical description of COQ-expressions into a "standard" intermediate, content-level representation such as MathML content.

T2.3 Proof transformation. Similar to the previous task, but for proofs. The delicate point, here, is the fact that no "standard" intermediate representation currently exists, and thus it has to be defined.

T2.4 Automatic extraction of metadata. Relevant metadata such as list of identifiers in critical positions inside statements can be automatically extracted from the fully structured representation of mathematical objects. This information can then be exploited for searching and retrieving. The precise list of metadata will be defined in Work Package 3.

T2.5 Presentational Stylesheets. Implementation of a bunch of stylesheets transforming the intermediate content representation into a suitable rendering format (MathML presentation, HTML, etc.)

T2.6 Automatic Proof Generation in Natural Language. Similar to the previous task but for proofs. In this case, a fully automated approach is unlikely to produce really satisfactory results, and the process should be possibly integrated with some mechanism for interactive annotation (see Task 4.3).

Deliverables:

D2.a Exportation module.

D2.b Document Type Descriptors.

D2.c-d Stylesheets to intermediate representation (formulae and proofs).

D2.e-f Presentational Stylesheets (formulae and proofs).

D2.g Tools for automatic extraction of Metadata.

Milestones and expected results: The exportation module D2.a. is our first milestone: without a large amount of available documents it would be impossible to test the transformations. Similarly, without a precise definition of the intermediate language, and a large sample of documents in this format (D2c-d) we cannot start to seriously address the presentational issue. Note that the intermediate language is the real core of the whole project.

The development of presentational stylesheets also depends in an essential way on the development of rendering/browsing engines for the chosen presentational language (in particular, for MathML).

For the end of month 18, we expect to have a first working prototype of the whole application.

9.3.4 Metadata

work package number: 3 - Metadata
Start date or starting event: month 6 - month 21

Participant Number:	1	2	3	4	5	6
Person-month per participant:	3	3	8	3	3	0

Objectives: Delineation of the basic intelligence to be considered for encapsulation in metadata, in order to meet the needs delineated during Requirement Analysis (WP1). Definition and Development of a specific Markup Model in RDF.

Description of work: The precise definition of metadata and their actual Markup Model are essential aspects for implementing the main functionalities of the library, and especially for archiving, searching and retrieving issues. The work will be articulated in two main, almost sequential, tasks:

T3.1 Use, meaning and classification. This task, based on the previous documents D1 and D2, is aimed to provide a precise definition and classification of all metadata required for an effective and efficient management of the library.

T3.2 Modelling. This is devoted to the definition of a precise markup model. To this aim, we plan to use the Resource Description Framework of W3C.

Deliverables:

D3.a Metadata for Mathematics;

D3.b Metadata Model;

Milestones and expected results: The definition of the Metadata model is a main milestone, since the architectural design and implementation of the consultation engine (Task 4.2) will be essentially based on it. Similarly, it is required for Task 2.4 (automatic extraction of metadata).

9.3.5 Interfaces

work package number: 4 - Interfaces
Start date or starting event: month 3 - month 24

Participant Number:	1	2	3	4	5	6
Person-month per participant:	16	16	5	4	14	0

Objectives: This Work-Package is devoted to the design and the implementation of the interfaces to the library, covering rendering, browsing, searching and retrieving functionalities. Some additional functionalities for authoring purposes (such as computer assisted annotation of proofs) will be also taken into account.

Our privileged rendering language will be MathML, which is likely to be rapidly adopted as the main language for representing mathematical notation on the Web.

Description of work: The work is organised in the following tasks:

T4.1 MathML rendering/browsing engines. Our privileged rendering language will be MathML. In order to be able to test the presentational stylesheets, (Task 2.5) we need a MathML compliant browser, that will be developed as a part of the project (starting from a previous prototype).

T4.2 Consultation Engine (archiving, searching and retrieving). The task is devoted to the architectural design an implementation of the main functionalities for the consultation and the management of the library.

T4.3 Assisted Annotation. This cover an additional functionality of the interface, aimed to support the user in the process of annotating a proof in natural language.

T4.4 LaTeX-based authoring tool. A tool supporting automatic generation of Content-MathML from a suitably (macro-)enriched version of LaTeX.

Deliverables:

D4.a MathML Rendering/Browsing engine;

D4.b First MOWGLI Prototype;

D4.c Prototype functionalities for assisted annotation.

D4.d LaTeX-based authoring tool (first prototype).

D4.e Refined and extended protoype of the LaTeX-based authoring tool

Milestones and expected results: The main milestone is the release of the first MOWGLI prototype, at month 18.

9.3.6 Distribution

work package number: 5 - Distribution
Start date or starting event: month 18 - month 30

Participant Number:	1	2	3	4	5	6
Person-month per participant:	6	5	8	3	5	0

Objectives: Overall architectural design of the distribution model, its implementation and integration with the consultation engine.

Description of work: The work is articulated in three, conceptually sequential tasks:

T5.1 Architectural Design of the Distribution Model. The big issue it to find the right compromise between two opposite requirements: *distribution* (in the sense of the Web: few rules, no central authority) and *coherence* (coherence between different copies of a same document and global management of the library as a single rational development). Other architectural problems to be solved are the management of Uniform Resource Identifiers, their mapping to Uniform Resource Locators, and the integration of databases in the distribution model. The final aim is to have a *phisically* distributed library with a single *logical* view.

T5.2 Prototype implementation. First prototyping implementation of the distribution layer.

T5.3 Integration with the Consultation Engine. First implementation of the library as a distributed repository. Distribution should be completely transparent to users of the Consultation Engine.

Deliverables:

D5.a Overall Architectural Design of the Distribution Model;

D5.b Advanced MOWGLI Prototype.

Milestones and expected results: The main milestone is the release of the advanced MOWGLI prototype, for the fourth meeting of the project, at month 24. The advanced prototype will integrate the previous one with the new distribution facilities offered by the distribution layer.

9.3.7 Testing and Validation

work package number: 6 - Testing and Validation
Start date or starting event: month 12 - month 30

Participant Number:	1	2	3	4	5	6
Person-month per participant:	5	6	3	17	14	9

Objectives: The WP intends to measure the system suitability and scalability and the satisfaction level of users with the service.

Description of work: Large scale testing and validation will start after the release of the first MOWGLI prototype, at month 18. We shall consider three main validation tests:

T6.1 Education. Full development of a fragment of the library covering a typical undergraduate course in algebra of analysis.

T6.2 Certified code. The aim is to be able to present the formalization and the demonstration of some security properties related to the code embedded into a smart card. The presentation must be in a format understandable by the company in charge of the evaluation of the code and in accordance with the Common Criteria standard.

T6.3 Electronic Publishing. The aim is to test the LaTeX-based authoring tool and to demonstrate how an electronic physics journal benefits from the exploitation of content markup in journal articles.

Deliverables:

D6.a Validation 1: Education.

D6.b Validation 2: Documents describing the formalization of a security function embedded into a smart card and the proof that the code meets its security objectives.

D6.c Validation 3: Journal interface with added functionalities due to content markup, e.g. search interface

D6.d Final MOWGLI Prototype

Milestones and expected results: We expect a lot of feed-back during this phase, that will be taken into account for the definition of the advanced and final versions of the MOWGLI prototype.

9.3.8 Information Dissemination and Exploitation

work package number: 7 - Information Dissemination and Exploitation Plan
Start date or starting event: month 3 - 30

Participant Number:	1	2	3	4	5	6
Person-month per participant:	1	0	0	1	6	0

Objectives: The work package aims to:

1. involve the largest community of professionals in the modelling phase.
2. promote dissemination of project results in the relevant international forums.
3. promote the use of the service, both for educational and professional purposes. This phase will start after the release of the first prototype, at month 18.

Description of work: The dissemination of results is in charge of the Project Dissemination & Exploitation Committee, which is responsible to prepare, during the first six months, a detailed dissemination work-plan (D2), and to follow its execution during the Project.

Information dissemination will take place via professional journal articles, presentations at conferences, international news groups, specific interest groups and so on. Relevant reports of the projects will be made publicly available on the World Wide Web. For advanced releases, we also plan to prepare an "information pack", including a folder and CDROM with demos.

Deliverables:

D7.a Dissemination and Use Plan;

D7.b Technological Implementation Plan;

Articles Scientific Publications on professional journals and conference proceedings.

Milestones and expected results: Criteria for evaluating MOWGLI's dissemination and its technological impact will be detailed in the Dissemination and use Plan. The Project Dissemination & Exploitation Committee will be in charge to monitor the state of advancement of the Project according to these criteria.

9.4 Deliverable list by WorkPackage

n.	Deliverable Title	WP no.	Lead Partic.	Estim. person-month	Type	Secu-rity	Deliv. month
0.a	Self-Assesment parameters and criteria.	0	UNIBO	2	R	Pub	6
0.b	First Self-Assesment Report	0	UNIBO	1	R	Pub	12
0.c	Second Self-Assesment Report	0	UNIBO	1	R	Pub	20
0.d	Final Self-Assesment Report	0	UNIBO	1	R	Pub	30
1.a	Preliminary Report on Application Scenarios and Requirement Analysis	1	KUN	6	R	Pub	3
1.b	Structure and Metastructure of Mathematical Documents	1	DFKI	8	R	Pub	6
1.c	Distributed Digital Libraries: development, archiving, retrieving.	1	MPG	6	R	Pub	6
2.a	XML Exportation Module(s)	2	INRIA	4	P	Pub	6
2.b	Document Type Descriptors	2	INRIA	2	R	Pub	12
2.c	Stylesheets to Intermediate representation (formulae)	2	UNIBO	7	P	Pub	12
2.d	Stylesheets to Intermediate representation (proofs)	2	UNIBO	9	P	Pub	12
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	INRIA	4	P	Pub	18
3.a	Report on Metadata for Mathematical Libraries	3	DFKI	9	R	Pub	12
3.b	Metadata Model	3	DFKI	11	R	Pub	12
4.a	First Prototype implementation of Rendering Engines for MathML	4	UNIBO	16	P	Pub	12
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
6.a	Validation 1: MOWGLI for education	6	KUN	16	R	Pub	30
6.b	Validation 2: MOWGLI and smart card security	6	Trusted Logic	9	R	Pub	30
6.c	Validation 3: Journal interface	6	MPG	14	P	Pub	30
6.d	**Final MOWGLI Prototype**	6	UNIBO	12	P	Pub	30
7.a	Dissemination and Use Plan	7	MPG	3	R	Pub	6
7.b	Technological Implementation Plan	7	MPG	3	P	Pub	30

R =Report, *P* =Prototype, Pub=Public

Work package list and team involvement

UNIBO = 1
INRIA = 2
DFKI = 3
KUN = 4
MPG = 5
Trusted Logic = 6

Overview over the Work packages

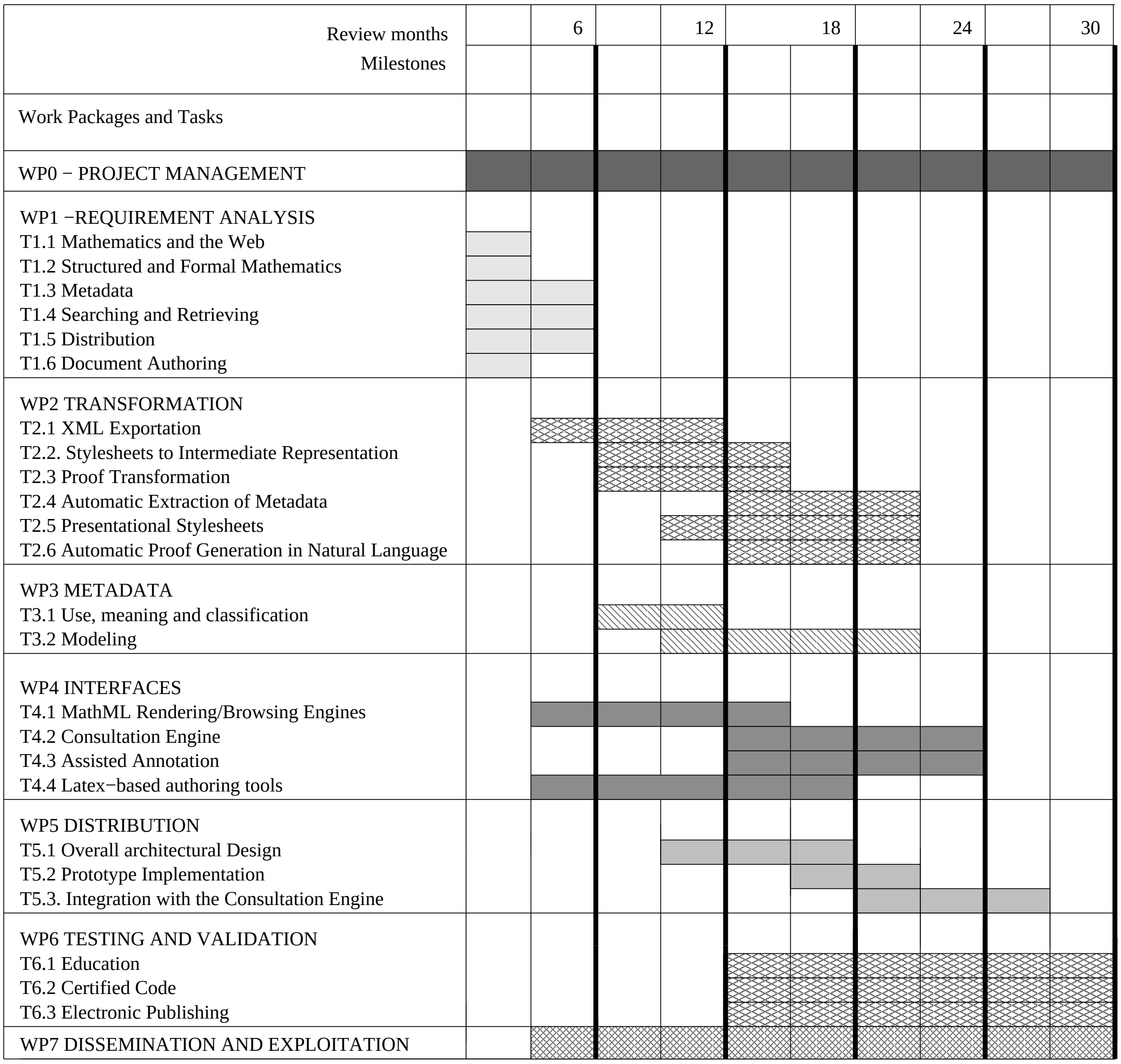
n.	Title	Lead Contr.	Person-months	Start-month	End-month	Deliverables
0	Project Management	1	8	0	30	a,b,c,d
1	Requirement Analysis	4	20	0	6	a,b,c
2	Transformation	1	34	3	21	a,b,c,d,e,f,g
3	Metadata	3	20	6	21	a,b
4	Interfaces	2	55	3	24	a,b,c,d,e
5	Distribution	3	27	9	27	a,b
6	Testing and Validation	4	54	12	30	a,b,c,d
7	Dissemination & Exploitation	5	8	3	30	a,b
TOTAL			226			

Involvement by Research Team (leader in boldface)

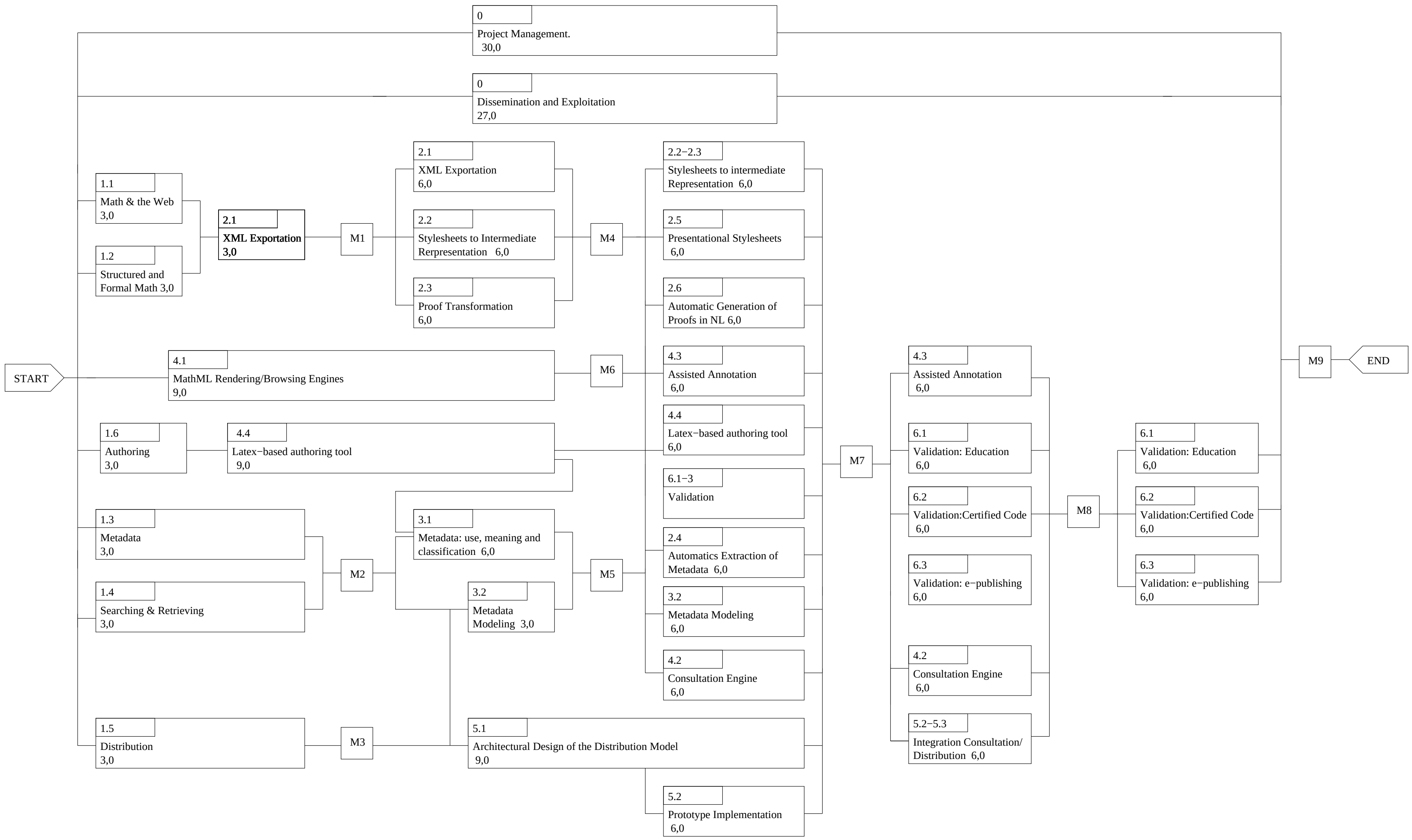
Team	Work Package								total
n.	0	1	2	3	4	5	6	7	
1	**8**	3	**13**	3	16	6	5	1	55
2	0	4	4	3	**16**	5	6	0	38
3	0	3	3	**8**	5	**8**	3	0	30
4	0	**6**	11	3	4	3	**17**	1	45
5	0	3	3	3	14	5	14	**6**	48
6	0	1	0	0	0	0	9	0	10
total	8	20	34	20	55	27	54	8	226

The total of 226 m.m. comprises 59 m.m. of permanent staff work for the four sites under an Additional Cost Regime: UNIBO (17 m.m), DFKI (9 m.m), KUN (15 m.m) and MPG (18 m.m).

9.5 Project Planning and timetable



9.6 Graphical presentation of project components



9.7 Project Management

The Project management will be assured by the following relevant roles: a) Project Manager, b) Exploitation manager, c) Work-package Leaders d) Technical Contributors and by the following Project Bodies: a) Project Coordination Committee (PCC), b) Project Exploitation Board and c) Work-package Teams.

Project Manager The mandate of the Project Manager is to represent the Project, report to the Commission, monitor overall performance of the project, ensure accomplishment of the technical objectives, promote project visibility, promote dissemination of project results in the relevant international forums, promote acceptance of project results, administer project resources and monitor project spending.

Exploitation Manager S/he will be responsible for coordinating exploitation activities undertaken by the Project Exploitation Board (PEB) in close cooperation with the Project Coordination Committee (PCC).

Work-package Leaders They are responsible for the performance of Work-packages, and they must be committed to the Work-package for at least five man months per year.

Technical Contributors Technical Contributors are specialised staff provided by the partners. They could be either people active on a stable basis into a Work-package, or people who contribute from time to time on specific tasks related to the work-plan (external contributors).

Project Coordination Committee The PCC is chaired by the Project Manager and constituted by one representative from each Partner, and is the main decision body of the Consortium. It is in charge of all formal decisions regarding technical direction of the work, proper interrelation between Work-packages, relations with IST Office, policies for promotion of results, administrative arrangements. In view of minimising management overhead and project costs, partners providing the Project Manager, and the Work-packages Leaders are required to appoint the same person as their official representative in the Project Coordination Committee.

Project Exploitation Board The PEB will be chaired by the Project Exploitation Manager, and composed by the Exploitation Managers of each partner. It will monitor the state of the art in the domains of content-based information technologies, W3C activities, Web publishing, and digital libraries. It will be responsible for planning the dissemination and exploitation strategies for the consortium, to be approved by the PCC.

Work-package Teams They are constituted of the technical contributors. Each team will be chaired by a Work-package Leader, and it will be in charge of carrying out the technical work described in Section 9.

9.7.1 Decision Process

Decisions will normally be taken by seeking consensus. However, after a reasonable amount of time has been allowed to illustrate and defend of conflicting positions, in order to avoid deadlock in project operational progress, the approval of two-third majority of the partners will be sufficient. If the decision being taken is unacceptable to partners found in the minority positions, the resolution of the conflict will be elevated to each partner's higher executive level. If no resolution is possible, than the standard "red flag" procedure will be used, as a last resort. Either the PCC or the Project Manager can initiate the conflict resolution procedure.

9.7.2 Information Flow

Information flow within the Project will be ensured by exchange of internal technical papers, notification of relevant new publications technologies or standards, and reports from external meetings. All technical documentation generated by the project should be exchangeable in electronic format, according to a set of guidelines to be agreed at project start-up. The project Manager will enforce adherence to these guidelines. Only strictly formal correspondence will be exchanged by ordinary mail and telefax. Urgent correspondence over e-mail will be sent with a request for explicit acknowledgement.
The Coordinating Partner will be responsible to prepare and maintain a Web page of the project and a CVS repository (also available via Web). The CVS repository is meant to keep the trace of the overall progress of the Project, and it will be used by the PCC to monitor the state of advancement of the work. Each partner periodically submit reports to the prime contractor, listing all technical contributions, publications, meeting attendance and other information which may help in understanding the provided effort and cost figures.
The Project will establish a high quality Web site for internal and external communications. It will be operational as from one month of the commencement date and will be maintained during the project duration. The "members only" area will include e.g. all private deliverables, progress and management reports, and cost statements. The open area will include the project presentation as defined in Appendix 2 of the "Guidelines for Contract Preparation" document, all public deliverables, and any relevant information suggested by the Project Exploitation Board.

9.7.3 Quality Procedures

For all deliverables, the following review procedure will be adopted: release by the Work-Package Leader, two-week review period for comments by the PCC, two-week amendment period to incorporate PCC recommendations, one-week balloting period for approval by the PCC.

9.7.4 Confidentiality and IPR handling

Matters related to Confidentiality and IPR handling will be defined in the "Consortium Agreement". As regards literature papers originated from work in the Project, there will be two categories of papers:

1. Project Endorsed Publications, requiring notification and approval by the PCC.

2. Project-related publications, only requiring notification to the PCC. They must contain an acknowledgement of support by the project Consortium and, if required by the PCC, a disclaimer that the views are not necessarily those of the Consortium.

Concerning possible contributions to Standards, they will be directly submitted by project members.

10 Clustering

Not applicable.

11 Other contractual conditions

11.1 Other significant project costs

One category of other significant project costs is due to “audit certificates” for INRIA and DFKI.

UNIBO (coordinator) requires 11.25Keuro to cover the membership fee to the World Wide Web Consortium (5000 dollars per year) for the duration of the Project. We recall that membership is required in order to participate to some activities and Working Groups of strategic relevance for MOWGLI, such as the MathML or RDF WG's. Establishing a tight cooperation with the World Wide Web Consortium is moreover a crucial point of our dissemination and exploitation plan.

Finally, a small amount of money is required to cover conference fees.

11.2 Travel costs outside MS/AS

Some travels outside MS/AS will be possibly required for dissemination purposes. This covers both the participation to International Conferences outside Europe to present scientific works related to the Project, participation to relevant International Meetings (such as AC meetings of the W3C), and possibly travels aimed to establish cooperations with other related projects outside MS/AS. For obvious reasons, a detailed list of these travels cannot be presently defined, and the cost figures in the CPF's are purely indicative.

References

[1] Communication of the ACM. Special Issue on Digital Libraries. April 1998.

[2] *Digital Libraries in Computer Science: the MeDoc Approach.* Lecture Notes in Computer Science 1392, Springer Verlag. 1998.

[3] Tim Berner's Lee. *The Semantic Web.* W3C Architecture Note, 1998. `http://www.w3.org/DesignIssues/Semantic.html`

[4] Common Criteria for Information Technology Security Evaluation. Part 1,2, 3. August 1999, Version 2.1, CCIMB-99-031.

[5] The Dublin Core Metadata Inititiative. `http://purl.org/dc/`

[6] C.LaGuardia, B.A.Mitchell eds. *Finding Common Ground: Creating the Library of the Future without Diminishing the Library of the Past.* New York: Neal-Schuman Publishers, Inc.1998.

[7] Goldfarb and Prescod. *The XML Handbook.* The C.F.Goldfarb Series on Open Information Management. Prentice Hall. 1998.

[8] G. Huet, G. Plotkin (eds). *Logical Frameworks.* Cambridge University Press. 1991.

[9] G. Huet, G. Plotkin (eds). *Logical Environments.* Cambridge University Press. 1993.

[10] M.Leventhal, D.Lewis, M.Fuchs. *Designing XML Internet Applications.* The C.F.Goldfarb Series on Open Information Management. Prentice Hall, ISBN 0-13-616822-1. 1998.

[11] Mathematical Markup Language (MathML) Version 2.0. W3C Proposed Recommendation, January 2001. `http://www.w3.org/TR/2001/PR-MathML2-20010108/`

[12] Megginson.*Structuring XML Documents.* The C.F.Goldfarb Series on Open Information Management. Prentice Hall. 1998.

[13] E. Maler, J. El Andaloussi. *Developing SGML DTD's. From text to model to markup.* Prentice Hall, ISBN 0-13-309881-8, 1996.

[14] OMDoc: A Standard for Open Mathematical Documents. `http://www.mathweb.org/omdoc/`

[15] The OpenMath Society. `http://www.nag.co.uk/projects/openmath/omsoc/`

[16] Resource Description Framework (RDF) `http://www.w3.org/RDF/`

[17] XML Linking Language (XLink). `http://www.w3.org/XML/Linking`.

[18] XSL Transformations (XSLT) Version 1.0. W3C Recommendation 16 November 1999. `http://www.w3.org/TR/xslt`

A Consortium Description

The Consortium is composed of the following partners:

- University of Bologna (Italy), Department of Computer Science `http://www.cs.unibo.it`. Responsible: A.Asperti.
- INRIA Rocquencourt (France) `http://www.inria.fr`. Responsible: H.Herbelin.
- German Research Center for Artificial Intelligence GmbH, DFKI (Germany). Responsible: M.Kohlhase.
- Katholieke Universiteit Nijmegen, (The Netherlands) `http://www.cs.kun.nl`. Responsible: H.Geuvers.
- Max Planck Institute for Gravitational Physics (Albert Einstein Institute/AEI). `http://www.aei.mpg.de`. Responsible: B.F.Schutz.
- Trusted Logic (France) `http://www.trusted-logic.fr`. Responsible: D.Bolignano.

The Consortium has been built with the aim to join some essential knowhow in different areas of I.T. related to the creation and maintenance of a digital library of structured mathematical knowledge.

As described in the Technical Annex, MOWGLI is meant to develop the technological infrastructure required to integrate existing Markup languages and standards such as MathML, OpenMath or OMDoc, covering different aspects of mathematical intelligence, into a single application. Expertise on these languages and the related technologies is respectively provided by the following partners:

MathML Department of Computer Science, University of Bologna, member of the World Wide Web Consortium and of the W3C Working Group on MathML; Lemme-Project at INRIA Sophia-Antipolis.

OpenMath University of Eindhoven (sub-site of Nijmegen): Professor Arjeh Cohen is one of the leaders of the OpenMath initiative, and MOWGLI is expected to take the maximum profit from the successful OpenMath Esprit project, no.24969.

OMDoc DFKI. Dr.Michael Kohlhase is the main auhtor of OMDoc; DFKI has a long research tradition in the management of mathematical knowledge bases, metadata, searching and retrieval issues.

More generally, the Department of Computer Science in Bologna has a long experience in XML-related technology, and in particular in their application to the particular domain of mathematical developments, as testified by the "Hypertextual Electronic Library of Mathematics" (HELM) Project (`http://www.cs.unibo.it/helm`). A main component of HELM is the GtkMathView widget (`http://www.cs.unibo.it/helm/mml-widget/index.html`), a C++ rendering engine for MathML that will be distributed as an official package of the next Debian release of Linux.

Similarly, the Lemme Project in Sophia-Antipolis has a large experience of edition of mathematical objects. It develops the graphical environment Pcoq, dedicated to the development of mathematical proofs, using the Coq proof assistant. Among many features, Pcoq has a sophisticated two dimensional formula and natural language proof edition component, allowing intuitive and powerful interactions. Built on the Figue environment, Pcoq can be made compatible with MathML. The Pcoq interface is intensively used by teams whose research activity concerns the certification of mathematical algorithms.

DFKI will contribute requirements and metadata from the viewpoint of educational applications including search functionalities. It will actively work on presentational transformations, the generation of proofs in natural language as well as on knowledge bases for mathematical knowledge DFKI intends to exploit the results of the MOWGLI project in pilot applications in current and planned research and in projects for the prototypical implementation of intelligent environments for learning of mathematics. In particular, the knowledge representation for mathematics on the Web is important for such Web-based systems. Knowledge bases that provide a common repository and ontology for mathematical knowledge are indispensible in systems that integrate various systems working on mathematical knowledge. DFKI also has a fierce interest in pushing and leveraging the quality of standardisation efforts within the worldwide initiative of the Semantic Web education systems and electronic publishing.

In order to immediately dispose of a large repository of structured mathematical information, the consortium comprises the developers of one of the most successful proof assistant tools currently available: the Coq proof engine (`http://pauillac.inria.fr/coq/`) of INRIA-Rocquencourt. The Coq standard library includes more than thousand lemmas and theorems and the whole number of statements proved by users is evaluated to hundred thousands, covering arithmetics, algebra, analysis and computer science. We expect to integrate the current different ways of browsing, searching and rendering Coq mathematical developments into a coherent and Web-oriented architecture open to the Coq user community and beyond.

An alternative route for the creation of content-based mathematical information from standard digital repositories by means of a suitable LaTeX-based authoring system will be explored by the Albert Einstein Institute (AEI) in Golm (Germany). AEI publishes a solely electronic review journal, *Living Reviews in Relativity* on the Web, which provides refereed, regularly updated review articles on all areas of gravitational physics. Since its release in January 1998 the journal has become a primary entry point for students, lecturers and researchers alike for up-to-date information on the current status of research in gravitational physics. Moving this unique repository and communication forum of current physical and mathematical knowledge in relativity to content mark-up, making it available for semantic search, and for re-use and evaluation e.g. in math algebra systems motivates the involvement in the MOWGLI project. The journal will develop a LaTeX based authoring tool interfacing with MOWGLI, and serve as a showcase to demonstrate how content-mark-up in mathematics improves the usability and information depth of electronic science journals.

The AEI will be supported by the newly founded Center for Information Management (CIM) of the Max Planck Society. The CIM has been set up by the Society to support researchers and research processes in the area of information management. The objectives of the project include coordination of existing activities within the Society and implementation of a strategy to develop electronic research archives. The current Managing Editor of the AEI's electronic journal Living Reviews in Relativity has been appointed executive director of the CIM (starting from 1 Sep 2001) and will be in charge of the project management for Tasks 4.4 and 6.3 of the proposal. The CIM will be in an excellent position to promote dissemination and use of the project results within the Max Planck Society. It will further give technical support to the Dissemination Manager in providing the MOWGLI website.

Professor Wegner, Scientific Coordinator of EMIS (European Mathematical Information Service), will also provide a main liaison with previous and successful European Projects on digital libraries and metadata, such as EULER (`http://www.emis.de/projects/EULER`) and the TRIAL Solution project (`http://www.trial-solution.de`). In particular, all the achievements of these Projects will be integrated inside MOWGLI, as far as the respective teams will agree to this. Moreover, in his quality of Scientific Coordinator of EMIS, member of the advisory board for MATHDI, and Chairman of the Electronic Publishing Committee of European Mathematical Society, Professor Wegner is an excellent candidate to organise the information dissemination and exploitation activities for the project.

Some of the partners of the consortium will primarily work in the development of pilot applications. In particular, the Department of Computer Science of the University of Nijmegen will apply MOWGLI's technologies to the development of an "electronic book", covering a typical undergraduate course in Algebra or Analysis. The Department of Computer Science in Nijmegen has a lot of experience in formal mathematics and theorem proving. Notably, the group has done large theory developments in the theorem prover Coq. (The FTA project: `http://www.cs.kun.nl/gi/projects/fta/`) Eindhoven University of Technology, a sub-site of Nijmegen, has expertise in OpenMath and in using WWW technology for educational purposes. This has resulted – among other things – in 'IDA', the interactive course notes in algebra (`http://www.win.tue.nl/ ida/`), where a combination of HTML and applets is used to present the mathematics. Jointly, Nijmegen and Eindhoven have experience in combining theorem provers and computer algebra packages, notably Coq and GAP.

Trusted Logic (France), which is specialized in secure and validated solutions for open systems, aims to present the formalization and the demonstration of some security properties related to the code embedded into a smart card. The presentation must be in a format understandable by the company in charge of the evaluation of the code and in accordance with the Common Criteria standard.

A third pilot application is the semantic markup of the Journal *Living Reviews in Relativity* published by AEI-Golm, already mentioned above.

Description of the Participants

A.1 University of Bologna, Department of Computer Science

Dipartimento di Scienze dell'Informazione
Università degli Studi di Bologna
Via di mura Anteo Zamboni VII, 40127, Bologna, ITALY.
http://www.cs.unibo.it

The Department of Computer Science of Bologna is the only educational institution in Italy to be affiliated to the World Wide Web Consortium (and one of the few members of this category in Europe). This affiliation testifies the interest, both technical and didactic, traditionally devoted by our Department to Web technologies, Internet and, more generally, distributed computing. This is joined to a solid expertise in programming languages, algorithms, theory and formal methods, that provides a very stimulating and dialectical environment for research. The Department is in charge of an undergraduate Program in Computer Science, with more than three hundreds new students a year, and a graduate Ph.D. Program, currently comprising sixteen students.

The HELM Project The "Hypertextual Electronic Library of Mathematics" Project (see `http://www.cs.unibo.it/helm`) is active in Bologna since 1999. Its aim is the development of a suitable technology for the creation and maintenance of a virtual, distributed, hypertextual library of formal mathematical knowledge. As a subsidiary goal, HELM is meant to integrate the current tools for the automation of formal reasoning and the mechanisation of mathematics (proof assistants and logical frameworks) with the most recent technologies for the development of Web applications and electronic publishing, taking advantage of the potentiality offered by XML Technologies. The Project is developed in tight cooperation with the W3C MathML Working Group, which we are a member of.

Technical Contributors The people involved in the project are:

- Andrea Asperti (Full Professor, Member of W3C Advisory Committee)
- Irene Schena (Ph.D. Student, Member of MathML-WG)
- Luca Padovani (Ph.D Student)
- Ferruccio Guidi (Ph.D. Student)
- Claudio Sacerdoti Coen (Ph.D. Student)

In case of approval, we expect to offer a two-year Post-Doc position on project funding.

ANDREA ASPERTI (COORDINATING PROPOSER)
Born 05/09/1961.
Full Professor of Formal Languages and Compilers
Member of the Advisory Committee of the World Wide Web Consortium.
Dip. di Scienze dell'Informazione, Via di Mura Anteo Zamboni 7, Bologna, Italy.
Phone: +39 51 35 45 05.
E-mail: `asperti@cs.unibo.it`

Andrea Asperti was born in Italy, in 1961. He was awarded a Ph.D in Computer Science by the University of Pisa in 1989. In the same year he obtained a post-doc position at INRIA-Rocquencourt, where he was employed as a researcher in 1991. In 1992 he obtained the chair of Formal Languages and Compilers at the Department of Computer Science of the University of Bologna, becoming Full Professor in year 2000. He is currently teaching courses on Programming Languages, Theoretical Computer Science and Human-Computer Interaction.

Positions and Management Experience

- Member of the Advisory Committee of the World Wide Web
- Former Director of the the Undergraduate Program in Computer Science of the University of Bologna.
- Responsible of the Bologna site for the European TMR Project "LINEAR".
- Former Responsible of the Bologna site for the WG-21836 CONFER-II.

- Main Responsible for the National Research Project "Linear Logic and beyond".

Research Interests The research interests of Andrea Asperti are focused on forms, formats and methods of knowledge representation and elaboration. This covers Markup Languages, Information Processing, Programming Languages, Interpreters and Compilers, Type Systems, Higher-order Logics, Linear Logic, Category Theory. He is author of over 40 international publications on formal aspects of the theory of Computing, and several books.
He has been member of the Programme Committees of several International Conferences, comprising CONCUR'98, RTA'99, LICS'99, PPDP'00, CSL'01, MKM'01 (First International Workshop on Mathematical Knowledge Management).
His recent scientific activity has been mostly focused on the HELM Project, aimed to build the technological infrastructure for an Hypertextual Electronic Library of Mathematics.

Selected Publications

- A.Asperti, L.Padovani, C.Sacerdoti Coen, I.Schena. *XML, Stylesheets and the re-mathematization of Formal Content*. Proceedings of "Extreme Markup Languages 2001 Conference", August 12-17, 2001, Montr'eal, Canada.
- A.Asperti, L.Padovani, C.Sacerdoti Coen, I.Schena. *HELM and the semantic Math-Web*. Proceedings of the 14th International Conference on Theorem Proving in Higher Order Logics (TPHOLS 2001), 3-6 September 2001, Edinburgh, Scotland.
- A.Asperti, L.Padovani, C.Sacerdoti Coen, I.Schena. *Formal Mathematics on the Web*. Proceedings of the Eighth International Conference on "Libraries and Associations in the Transient World: New Technologies and New Forms of Cooperation", June 9-17, 2001, Sudak, Autonomous Republic of Crimea, Ukraine.
- A.Asperti, L.Padovani, C.Sacerdoti Coen, I.Schena. *Formal Mathematics in MathML*. First MathML International Conference, October 20-21, 2000, Urbana-Champaign, IL, USA.
- A.Asperti, G.Longo. *Categories, Types, and Structures. An introduction to Category Theory for the Working Computer Scientist*. Foundation of Computing Series, Massachusetts Institute of Technology Press, ISBN 0 262 01125-5. 1991.

A.2 INRIA

Institut National de Recherche en Informatique et Automatique (INRIA)
Domaine de Voluceau
78153 Rocquencourt Cedex
France

INRIA (National Institute for Research in Computer Science and Control) is a French public-sector scientific and technological institute operating under the dual authority of the Ministry of Research and the Ministry of Industry. INRIA's missions are "to undertake basic and applied research, to design experimental systems, to ensure technology and knowledge transfer, to organise international scientific exchanges, to carry out scientific assessments, and to contribute to standardisation".

The research carried out at INRIA brings together experts from the fields of computer science and applied mathematics covering the following areas: Networks and Systems; Software Engineering and Symbolic Computing; Man-Machine Interaction; Image Processing, Data Management, Knowledge Systems; Simulation and Optimisation of Complex Systems.
INRIA gathers in its premises around 2 100 persons including 1 600 scientists , many of which belong to partner organisations (CNRS, industrial labs, universities) and are assigned to work in common "projects". On INRIA's budget, around 500 full-time equivalent R&D positions can be accounted for.
A large number of INRIA senior researchers are involved in teaching and their PhD students (about 550) prepare their thesis within the different INRIA research projects (currently 74).
Its budget is roughly 90 MEuro, 20% of which comes from research and development contracts, royalties and sales. Industrial relations are strategic for INRIA:

Industrial contracts and European Projects. Numerous industrial partners contract with the Institute for collaborative research. They are French or foreign companies, of all sizes. 400 such contracts are presently active. Roughly 40% of these contracts are European funded ones. Since 1984, 250 European Framework-Programme (FP) projects have been executed.

Technology companies. As the ultimate step in technology transfer, researchers are party to the setting up of companies in order to implement their technology on the market. Thirty seven spin-off companies have been created since 1984. In 1999, INRIA has launched two subsidiaries to promote high-tech start-up companies: INRIA-TRANSFERT deals with early accompaniment of the future companies, whereas I-SOURCE GESTION provides for "seed-money".
INRIA is a member of ERCIM EEIG, European Research Consortium for Computer Science and Mathematics. Outside Europe, INRIA also has a significant activity: it has created joint research laboratories (Russia and China), signed cooperation agreements (NSF, India, Brazil, etc.) and promotes intensive scientific exchanges.
INRIA Web: http://www.inria.fr/

A.2.1 The Lemme project

The purpose of the Lemme project is to introduce and develop formal methods for use in writing scientific computing software. In scientific computing, algorithms and mathematics are at the forefront. We are thus developing tools and methods to help producing correct programs starting from the usual mathematical descriptions of data, algorithms, properties and proofs, structured into four research themes:

- Proof environments (development of the Pcoq system in Java and its compatibility with XML/MathML).
- Formalisation of mathematical theories (algebraic geometry, elementary algebra and analysis).
- Certified implementation of scientific computing algorithms (computer algebra, arithmetics, logic).
- Proofs on semantics of programming languages (Javacard).

The project belongs to the European working group Types, and to the French action AOC (Arithmétique des Ordinateurs Certifie). It keeps up industrial collaborations with Dassault-Aviation (program proof environments), Alcatel Space Industry (certified numerical code), and GemPlus (Javacard certification), and also collaborates with teachers at university on the use of formal proofs and Web-based environments in mathematics courses.

People of Lemme involved in the MOWGLI proposal are: Laurence Rideau (researcher), Hanane Naciri (PhD student), Loïc Pottier (researcher), Yves Bertot (researcher), Ahmed Amerkad (engineer), Laurent Théry (researcher) and Laurent Chicli (PhD student).

A.2.2 The LogiCal project

The LogiCal team of INRIA is working on theoretical and practical aspects of mathematical proofs. It develops the Coq proof assistant, an implementation of an expressive formalism called Calculus of Inductive Constructions. Coq is used both for development of formal mathematics and for certification of programs, especially protocols and critical systems.

The LogiCal project is a joint project with University Paris 11. It is involved in the European TYPES working group and in several French actions. Especially, it is involved in the S-Java action aiming at certifying safety properties for JavaCard programs, in a project aiming at certifying algorithms used in computer algebra systems. LogiCal collaborates also on proof automation with France Telecom.

Technical Contributors of LogiCal team The people involved in the project are:

- Hugo Herbelin (Associate Professor at University Paris 10, seconded at INRIA)
- David Delahaye (Ph.D. Student)

In case of approval, we expect to offer a three-month training course on project funding.

Hugo Herbelin
Born June 9, 1967.
Associate Professor at University Paris 10, seconded at INRIA.
Domaine de Voluceau
78153 Rocquencourt Cedex
France

Phone: +33 (0) 1 39 63 53 77.
E-mail: Hugo.Herbelin@inria.fr

A.3 German Research Center for Artificial Intelligence GmbH, DFKI

The German Research Center for Artificial Intelligence GmbH, DFKI was founded in 1988. Today, DFKI is one of the largest non-profit contract research institutes in the field of innovative software technology based on Artificial Intelligence (AI) methods. DFKI is focusing on the complete cycle of innovation - from world-class basic research and technology development through leading-edge demonstrators and prototypes to product functions and commercialization. Based in Kaiserslautern and Saarbrücken, the German Research Center for Artificial Intelligence ranks among the important "Centers of Excellence" worldwide. The key directors of DFKI are Prof. Dr. Wolfgang Wahlster (CEO) and Dr. Walter G. Olthoff (CFO).

DFKI's mission is technology transfer, that is to move innovations in AI as quickly as possible from the lab into the marketplace by maintaining research projects at the forefront of science. Expertise and experience of the company DFKI has strong expertise in each of the following domains:

- Information Management and Document Analysis (Director: Prof. Dr. Andreas Dengel)
- Intelligent Visualization and Simulation Systems (Director: Prof. Dr. Hans Hagen)
- Deduction and Multiagent Systems (Director: Prof. Dr. Jörg Siekmann)
- Language Technology (Director: Prof. Dr. Hans Uszkoreit)
- Intelligent User Interfaces (Director: Prof. Dr. Wolfgang Wahlster)

DFKI is involved in numerous industrial, academian projects including projects in the current EC IST programme dealing with research and development in the broad areas of intelligent interface agents, and multiagent systems for applications in supply-chain management, virtual enterprises, e-commerce and advanced information systems. The partners of the DFKI are leading large-scale concerns such as DaimlerChrysler, SAP, and Alcatel, plus the two universities of Kaiserlautern and Saarbrücken. In the range of medium-sized firms INSIDERS, IDS Scheer, Tecmath and KIBG are to be added. Finally, the two most important large-scale research centers on international level, i.e. the Fraunhofer Society and GMD, round off the circle of partners. DFKI is part of several European Networks of Excellence such as AgentLink and CompuLog.

Recently, an Education Technology Group of DFKI, headed by Prof. J.H. Siekmann, has been established. It has carried out basic research and applications in several fields of AI, including presentation planning (for education material), user modeling, proof planning, knowledge representation (for educational and mathematical Web-documents) and integration of (mathematical services). Its main prototypical product so far has been the Web-based, user-adaptive, generative learning environment ActiveMath that integrates several external services. Members of the group are actively involved in the international academic life by organizing or contributing to conferences and workshops.

PROF.DR. JOERG SIEKMANN
Prof. Dr. Joerg Siekmann is Professor of Computer Science and Director of the Deduction and Multiagent Systems research department at the DFKI. He studied at the University of Goettingen (Mathematics and Physics, 1972) and the University of Essex (M.Sc. in Computer Science, 1973, and Ph.D. in Computer Science, 1976). From 1976 to 1983 he was a research assistant at the University of Karlsruhe. From 1983 to 1991 he was Professor of Computer Science at the University of Kaiserslautern and since 1991 he has a joint position as professor for Computer Science at the Universitaet des Saarlandes and as one of the directors at the DFKI. He has published widely in AI, Automated Reasoning and Unification Theory and served on many programme committees and was programme chairman of various conferences (such as CADE, KI etc.). He is editor of several logic and AI oriented scientific journals (such as Journal of Automated Reasoning, Journal of Artificial Intelligence, Journal of Logic and Computation etc). His biography is included in most international Whos Who, as one of the

founders of AI in Germany.

PD.Dr. Erica Melis
Erica Melis is a Senior Researcher at the German Research Center for Artificial Intelligence (DFKI GmbH). She was a research scientist at the School of Computer Science of Carnegie Mellon University and at the Department of AI at the University of Edinburgh. Later she has been an Assistant Professor at the University of Saarland and is member of a Sonderforschungsbereich in Saarbruecken. She served on numerous programme committees. Erica Melis is the editor of proceedings and the co/author of numerous scientific papers in journals, conferences, and collections in areas such as deduction, planning, case-based reasoning, intelligent tutor systems. Currently, her research interests focus proof planning and on the application of Artificial Intelligence technology in intelligent learning environments. She is a member of AAAI and GI.

A.4 Nijmegen (& Eindhoven)

Subfaculteit Informatica
Faculteit Natuurwetenschappen, Wiskunde en Informatica
Katholieke Universiteit Nijmegen
Toernooiveld 1
6525 ED Nijmegen
The Netherlands
http://www.cs.kun.nl

The Sub-faculty of Computer Science at the University of Nijmegen hosts a broad experience in logic, formal methods and theorem proving. The Faculty of Mathematics and Computer Science of Eindhoven University of Technology is strong in computer algebra, theorem proving and applying Web technology to mathematics. Nijmegen and Eindhoven have a long history in cooperation on topics related to this FET proposal, notably type theory, theorem proving and combining various computer mathematics applications, especially using OpenMath. This cooperation was mainly taking place between the research groups of Geuvers and Barendregt in Nijmegen and the research group of Cohen in Eindhoven.

The research group of Geuvers and Barendregt is part of the EC sponsored Thematic Network "TYPES" (IST-1999-29001) and of its ancestor, the EC Working Group "Types for Proofs and Programs", which testifies there interest in theorem proving, especially using type theory based theorem provers. The FTA project (Fundamental Theorem of Algebra), started in 1999 and to be finished in 2001, has as its main goal to formalize (in Coq) a large body of undergraduate mathematics (algebra and analysis), culminating in a proof of the fundamental theorem of algebra. The formalization of the mathematics is now finished and the next step is to make the formalization accessible and usable by others, preferably through the World Wide Web. The research group in Nijmegen and the research group of Cohen in Eindhoven are both part of the EC Working Group Calculemus, which aims at bridging the gap between different mathematical computer applications, like computer algebra systems and theorem provers. One of the vehicles for doing so is the definition of OpenMath as an intermediate language for the exchange of mathematical objects among computer applications. The research group of Cohen is part of the IST Thematic Network "OpenMath" (IST-2000-28719) and its ancestor, the OpenMath Esprit project (`http://www.openmath.org/`).

Technical Contributors The people from Nijmegen and Eindhoven involved in the project are:

- Herman Geuvers (Associate Professor, Nijmegen)
- Arjeh Cohen (Full Professor, Eindhoven)
- Henk Barendregt (Full Professor, Nijmegen)
- Milad Niqui (Ph.D. student, Nijmegen)
- Jasper Stein (Ph.D. student, Nijmegen)
- Luis Cruz-Filipe (Ph.D. student, Nijmegen)
- Freek Wiedijk (Post Doc researcher, Nijmegen)
- Dan Synek (Programmer, Nijmegen)
- Ernesto Reinaldo Barreiro (Ph.D. student, Eindhoven)

In case of approval, we expect to offer a 30-months Ph.D. position on project funding.

HERMAN GEUVERS (NIJMEGEN-EINDHOVEN SITE LEADER)
Born 19/05/64
Associate Professor in Foundations of Mathematics and Computer Science
Faculteit NWI, KUN, Toernooiveld 1, 6525 ED Nijmegen, NL
Phone: +31 243 652603
E-mail:`herman@cs.kun.nl`

Herman Geuvers studied Mathematics at the University of Nijmegen and got his Ph.D. in Mathematics and Computer Science in 1993 at the same University. In the same year he became assistant professor in computer science at the Eindhoven University of Technology in the Formal Methods group. From January 1st 2000, he is associate professor at the Department of Computer Science of the University of Nijmegen in the Foundations group. He is currently teaching in Formal Languages and Computability and Type Theory.

Positions and Management Experience

- Responsible for the Nijmegen-Utrecht site of the EC Thematic Network "TYPES" (Computer Assisted Reasoning Based on Type Theory), IST-1999-29001.
- President of the education committee of the Sub-faculty of Computer Science at the University of Nijmegen.
- Former member of the Management Board of the Dutch research school IPA (Institute for Programming Research and Algorithmics).

Research Interests The research interests of Herman Geuvers are: Formalization of Mathematics, Interactive Theorem Proving, Higher-order Logics, Communicating Formal Mathematics, Type Theory and λ-calculus. His recent scientific activities range from the study of formal theories (especially typed λ-calculi) to doing large theory developments in theorem provers, notably the formalization of the fundamental theorem of algebra in Coq.

Selected Publications

- H. Geuvers, E. Barendsen, Some logical and syntactical observations concerning the first order dependent type system lambda P, *Mathematical Structures in Computer Science*, vol. 9-4, 1999, pp. 335 – 360
- H. Geuvers, F. Wiedijk, J. Zwanenburg, Equational Reasoning via Partial Reflection, in *Theorem Proving for Higher Order Logics, TPHOL 2000*, Portland OR, USA, eds. M. Aagaard and J. Harrison, LNCS 1869, pp. 162 – 178.
- M. Oostdijk and H. Geuvers, Proof by Computation in the Coq system, to appear in *Theoretical Computer Science*, 2001.
- H. Barendregt and H. Geuvers, Proof Assistants using Dependent Type Systems, to appear as a chapter of the *Handbook of Automated Reasoning*, eds. A. Robinson and A. Voronkov, Elsevier 2001.
- P.A.M. Seuren, V. Capretta and H. Geuvers, The logic and mathematics of occasion sentences, to appear in the *Journal of Linguistics and Philosophy*, 2001.

A.5 MPG & TU Berlin

A.5.1 Max Planck Institute for Gravitational Physics

Albert Einstein Institute
Am Mühlenberg 1
14476 Golm, Germany
http://www.aei.mpg.de

Since 1998 the Max Planck Institute for Gravitational Physics, which is devoted to basic research in relativity, publishes one of the most innovative electronic science journals, *Living Reviews in Relativity* (http://www.livingreviews.org). Its mission is to build the journal into a primary reference point in the field, exploiting new and cutting edge Web technologies to maximise use, transparency, and depth of the information provided. The Institute has a number of staff who are exceptionally skilled in computing, and many of them contributed to the development of software that has made the journal's Web site one of the most advanced in the world of electronic publishing. Further background information on the journal can be found at http://www.livingreviews.org/Project/index.html. Fast and effective international dissemination of the Web journal is provided by the European Mathematical Society's Information Service, which maintains a network of more than 30 mirror servers worldwide.

The Institute has close connections to the newly founded Center for Information Management of the Max Planck Society that will act as a central service to the more than 80 Max Planck Institutes, to improve their ability to access and to publish information on the Internet. It is further cooperating with the leading European publisher of original research in gravitational physics, the Institute of Physics Publishing (Bristol, UK) with its journal *Classical and Quantum Relativity.*
The Max Planck Institute for Gravitational Physics will join forces with Professor Bernd Wegner at the Technical University in Berlin. Both parties are currently cooperating in the dissemination and mirroring of electronic journals like *Living Reviews in Relativity.*

Technical Contributors People involved in the project are:

- Bernard Schutz (Institute Director and Editor-in-Chief)
- Ian Kelley (Programmer)
- N.N. (Ph.D. Students and Post-Docs)

In case of approval we expect to offer a full scientific programmer position for a period of 2-years.

Prof. Bernard Schutz (born in 1946 in the USA) is director of the Max Planck Institute for Gravitational Physics (Albert Einstein Institute), heading the Astrophysical Relativity Division of the Institute, and Professor at Cardiff University. In acknowledgement of pioneering scientific work in gravitational physics he has been elected Fellow of the American Physical Society. He invented the "Living Reviews" concept in 1996 and acts as the journal's Editor-in-chief. Bernard Schutz has served on the Editorial Board of *Classical and Quantum Gravity* between 1988 and 1990, and since 1998 is member of the Editorial Board of the monograph series Studies In High Energy Physics, Cosmology and Gravitation (Institute of Physics Publishing, Bristol, UK).

- Schutz, B.F. *A First Course in General Relativity.* Cambridge University Press, Cambridge (1985).
- Schutz, B.F. *Geometrical methods of mathematical physics.* Cambridge University Press, Cambridge, (1980).
- Wheary, J., and Schutz, B. *Making an Electronic Journal Live.* Journal of Electronic Publishing, vol.3, is. 1, September (1997). (http://www.press.umich.edu/jep/03-01/LR.html)
- Wheary, J.; Wild, L.; Schutz, B. F. and Weyher, C. *Thinking and Developing Electronically*, Journal of Electronic Publishing, vol.4, is. 2, December (1998). (http://www.press.umich.edu/jep/04-02/wheary2.html)

A.5.2 Center for Information Management in the Max Planck Society

c/o Max Planck Society
Hofgartenstrasse 8
PF 10 10 62
D-80084 Munich, Germany
http://www.zim.mpg.de (not yet available)

The mission of the new Center for Information Management of the Max Planck Society is to allow the Society with its more than 80 specialized research Institutes to keep pace with and eventually play a leadership role in Europe in the management of scientific information. It will be staffed with 6 employees and start operating in September 2001. The project will be advised by a commission of MPG Institute directors and Rick Luce, the Director of the Los Alamos National Laboratory's "Library Without Walls".
Technical Contributors

- Theresa Velden (Executive Director, CIM)

- N.N. (Technical support)

Theresa Velden (born in 1970 in Germany) graduated in Physics from Bielefeld University in 1997 after having completed a diploma thesis on Mathematical Relativity at the Max Planck Institute for Gravitational Physics. Since 1998 she has been leading the editorial team of *Living Reviews in Relativity* as Managing Editor, organizing the editorial process and supervising the development of software for electronic publishing and journal management.

A.5.3 TU Berlin

Fachbereich Mathematik
Technische Universitt Berlin
Strae des 17. Juni 135
D - 10623 Berlin
`http://www.emis.de`

The group at TU Berlin will be lead by Prof.Dr. Bernd Wegner. He is professor for mathematics with a full position at the mathematics department. Under his guidance TU Berlin is partner in several projects and enterprises dealing with information and communication in mathematics:

He is

- editor-in-chief of Zentralblatt MATH, the most comprehensive literature information service in mathematics, with Web access under EMIS,
- member of the advisory board for MATHDI, an information data base on education in mathematics,
- Scientific Coordinator of EMIS, the European Mathematical Information Service, providing a mathematics portal with an electronic library as the main content,
- leader of the TU-group for the EULER-project, which has developed a prototype for an integrated access to Web-based mathematical documents, funded by Telematics for Libraries, and supported by a small TAKEUP-project for the transition to a regular Web service,
- Scientific Director of the LIMES-project (Large Infrastructures in Mathematics - Enhanced Services) which is designed to transform Zentralblatt MATH into European cooperation with Web-based input structures, funded by the Fifth Framework Programme,
- Director of the ERAM-project (Electronic Research Archive in Mathematics), funded by Deutsche Forschungsgemeinschaft and designed to build up a digital archive of classical mathematics, capturing the Jahrbuch ueber die Fortschritte in database as a pre-Zentalblatt access facility to the archive,
- Chairman of the Electronic Publishing Committee of EMS (European Mathematical Society),
- member of the Database Committee of the EMS

- associated with project Euclid (Cornell University, Ithaka, U.S.A.), establishing a non-profit (electronic) publication facility for mathematics,
- member of the board of IWI (Institute for Scientific Information in Osnabrueck).

He has participated with one or more talks at about 20 events on electronic information and communication in 2000, including continental congresses in Lisbon, Barcelona, Rio de Janeiro, Havanna, Moscow, and Manila. He has been elected as member of the scientific committee for several conferences on the subject in 2001: (Crimea, Athens, Coimbra, Linz, Guatemala City) and received invitations for plenary lectures at additional conferences.

These activities give an excellent background to organise the information dissemination and exploitation activities for the project.

A.6 Trusted Logic

Created in January 1999, TRUSTED LOGIC S.A. is a french start-up company, which presents a unique combination of expertise in embedded software challenges and in formal methods analysis. TRUSTED LOGIC offers a wide range of efficient and secure solutions for smart cards and terminals in the following areas : open systems for smart cards (Java Card,Windows SmartCards, Multos), open systems for terminals (mobile phones, PDA, bank terminals, etc), e-commerce, remote secured access to intranets, and secured networks infrastructures (home or car control).

The development methodology, inside Trusted Logic S.A., includes a permanent concern of quality and security aspects (formal specification and proofs) that provides a software which is efficient, reliable and ready to face high level Common Criteria evaluations when needed.

The main actors in the smart card and in the embedded software industry are among its customers : Sun Microsystems Inc., Bull, Gemplus, Oberthur, Schlumberger, ST Microelectronics, VISA International, Carte Bleue, GIE Cartes Bancaires, ActivCard, Proton World International, etc.

Dominique Bolignano

Dominique Bolignano is Chairman and Chief Executive Officer of Trusted Logic. He also teaches security and languages as a part-time full professor (Professeur Associ) at the Universit Paris-Dauphine. From 1996 to 1999, Dominique Bolignano worked with Dyade, a joint venture between Bull and INRIA (French National Research Institute in Computer Science and Control), as Director, Responsible for technology transfer in security, electronic commerce, and smart cards, and Project Director in the following areas : formal verification of electronic commerce protocols (and more generally cryptographic protocols), evaluation or design of security architectures (mainly for secure embedded systems, such as smart cards, payment terminals, GSM phones, etc.), development of highly secure Java Virtual Machines (Bull Odyssey JavaCard, payment terminals), formal methods for ITSEC and Common Criteria security evaluations (participation to various security evaluations including some on smart cards operating systems). Before that, Dominique Bolignano was the head of a research group in the Corporate Research Centre of Bull and he has acquired, from 1982 to 1996, a deep experience in industry software development and formal verification systems (distributed systems, security protocols). Dominique Bolignano takes part in the Java Card Forum. He also is a Member of the newly created IFIP Working group on Foundations of Security Analysis and Design. He has been an invited speaker at MFPS'98, CAV'99, CARI'98, CIRM'98, JFMM'98, RTCS'96, Cartier'96, FemSys'97, and to many security or formal methods symposia and presented invited tutorials at Forte'98, Cartier'96, Forte'93. Dominique Bolignano also acted as a PhD advisor, as a reviewer and evaluator for European Community programs. He is a reviewer in several journals (Journal of Theoretical Computer Science (TCS), TSI, ...) and has been a reviewer and/or a member of the Program Committee in international conferences (SESS'93, FME'96,VDM 91 et FORTE 92, SaS'94, ...)