



SIXTH FRAMEWORK PROGRAMME
PRIORITY 2
Information Society Technologies IST



Contract for:

NETWORK OF EXCELLENCE

Annex I – “Description of Work”

Project acronym: SemanticMining

Project full name: Semantic Interoperability and Data Mining in Biomedicine

Proposal/Contract no: 507505

Related to other Contract no:

Data of preparation of Annex I: September 15, 2003, revised October 24

Start date of contract:



Table of Content

1 PROJECT SUMMARY	5
2. PROJECT OBJECTIVES	6
2.1 Application Framework	6
2.2 Integrating Activities	8
3. LIST OF PARTICIPANTS	9
4. RELEVANCE TO THE OBJECTIVES OF THE IST PRIORITY	10
5. POTENTIAL IMPACT	12
5.1 Contribution to Standards	12
5.2 Contribution to Policy Development	13
5.3 Risk Assessment	13
5.4 Impact on the Citizen in the Information Society	14
5.5 Impact on the ICT Industry	15
6. JOINT PROGRAMME OF ACTIVITIES	16
6.A Activities	16
6.1 Integrating Activities	16
6.2 Programme for Jointly Executed Research Activities	19
6.3 Spreading of Excellence Activities	22
6.4 Management Activities	23
6.B Plans	24
6.5 Plans for Dissemination of Knowledge	24
6.6 Gender Action Plan	24
6.7 Raising Public Participation and Awareness	25
6.C Milestones	26
6.8 Major Milestones	26
7. QUALITY OF INTEGRATION	27
7.1 Comments on Quality of Integration	27
7.2 Indicators of Quality	27



8. PROJECT ORGANISATION, MANAGEMENT AND GOVERNANCE STRUCTURE	30
8.1 Governing bodies, roles and responsibilities	30
8.1.1 The co-ordination of the NoE	30
8.1.2 The Assembly	31
8.1.3 The Board	31
8.1.4 Committees	31
8.2 The Consortium Agreement	32
9. JOINT PROGRAMME OF ACTIVITIES FIRST 18 MONTHS	33
9.1 JPA Plan Introduction	33
9.1.1 Risk Analysis and Contingency Plan	33
9.4 Work Package List (18 months)	37
9.5 Deliverables list (18 months)	38
9.6 Work Package Description	40
9.6.1 Project Management and Integration	40
9.6.3 Knowledge Transfer and Dissemination	46
9.6.5 Jointly Executed Research Activities	55
10. PROJECT RESOURCES	63
10.1 Efforts for the full duration of the project	63
10.2 Efforts for the first 18 months	64
10.3 EC contribution	66
10.4 Project management level description of resources and grants	68
11 ETHICAL ISSUES	69
12 OTHER ISSUES	69
APPENDIX A.1	71
Medical Informatics, Linköping University, Sweden	71
Department of Computer Science and Information Systems, Linköping University, Sweden	72
Karolinska Institutet (KI), Stockholm, Sweden	72
Sahlgrenska University Hospital and Göteborg University, Göteborg, Sweden	73
University of Freiburg, Germany	73
Institute for Formal Ontology and Medical Information Science (IFOMIS), University of Leipzig, Germany	74
Department of Computer Science, University of Kiel, Germany	75
Medical Informatics Group, University of Manchester, UK	75
The Information Technology Research Institute, University of Brighton, UK	76
Centre for Health Informatics and Multiprofessional Education (CHIME), University College London, UK	77
Medical Informatics Division, University Hospital of Geneva	77
Public Health and Medical Informatics Laboratory (SPIM), Paris, France	78
Institute of Cognitive Science and Technology (ISTC-CNR), Italy	79
EBI - European Bioinformatics Institute, Hinxton, UK	80



Committee Nomenclature, Properties and Units in Laboratory Medicine (cNPU)	81
The Nordic Terminology Network - Public Health and Welfare Organisations in the Nordic countries	81
National Institute and Library for Health Information, Hungary	82
European Dynamics, Greece	83
Merrall-Ross International Ltd – Cheshire, UK	83
APPENDIX A.2 – A.5	84
APPENDIX B: TABLE OF RESEARCHERS AND DOCTORAL STUDENTS	85

1 Project Summary

The aim of this proposal for an Network of Excellence entitled Semantic Interoperability and Data Mining in Biomedicine is to establish Europe as the international scientific leader in medical and biomedical informatics. The long-term goal of the network will be the development of generic methods and tools supporting the critical tasks of the field; data mining, knowledge discovery, knowledge representation, abstraction and indexing of information, semantic-based information retrieval in a complex and high-dimensional information space, and knowledge-based adaptive systems for provision of decision support for dissemination of evidence based medicine. The proposal is a response to the strategic objectives addressed in the IST call 1, areas 'Semantic-based Knowledge Systems' and 'eHealth'.

The general objective of the network is to bridge gaps in the European research infrastructure and to facilitate cross-fertilisation between scientific disciplines. Traditionally academic departments in the domain have their roots either in computer science, system engineering (including a variety of engineering disciplines) or in a medical or clinical context. The proposed network is composed of partners from these scientific areas, all bringing their experience and in-depths knowledge together into a common framework. An important aspect of this is the merging of medical or clinical informatics and bioinformatics including the new fields of genomics and proteomics. Another bridging activity addressed by the proposal is knowledge-transfer and co-operation between academia and organisations in the health and welfare sector, such as standardisation bodies and public and user-driven health care organisations.

The NoE application is based on the partnership of 25 participants from 11 European countries with 99 identified researchers (25 female) and 31 associated PhD students (10 female).

2. Project Objectives

2.1 Application Framework

The establishment of the Network of Excellence (NoE) 'SemanticMining' is in response to challenges expressed in the Workprogramme of Information Society Technologies, particularly the strategic objectives of Semantic-based Knowledge Systems (IST-2002-2.3.1.7), eHealth (IST-2002-2.3.1.11) and Bioinformatics in the Life Sciences Workprogramme (LIFESCIHEALTH-1.1.4).

The general objective of the NoE instrument (Network of Excellence) is to bridge gaps in the European research infrastructure and to facilitate cross-fertilisation between scientific disciplines. Traditionally academic departments in the domain of interest in this application – namely medical and biomedical informatics – have their roots either in computer science, systems engineering (including a variety of engineering disciplines) or in a medical or clinical context. There are, however, important research areas where the overlap in terms of projects and persons still falls short of what would be desirable, for example between the fields of formal ontologies and computational linguistics on the one hand, and clinical and biomedical informatics on the other. The proposed network is composed of partners from these disciplines, all bringing their experience and expertise together into a joint framework.

The network will specifically contribute to current efforts in the fields of medical informatics and bioinformatics to develop generic methods and tools for managing and interpreting the vast amount of digital data produced by the health care system and by biotechnological laboratories. The network will contribute to the bridging of gaps between so-called clinical and pre-clinical information on the one hand and the development of technological platforms for managing the results of evidence-based medicine on the other, taking full advantage of the potential of the information- and knowledge-based services offered by current technology.

An important aspect of the initiative is to find ways of merging current efforts of medical (or clinical) informatics and bioinformatics (the latter including the new fields of functional genomics and proteomics) as also pointed out in the BIOINFOMED project meeting, November 2002 in Valencia, Spain. Thus one of the primary reasons for the establishment of this NoE is to facilitate networking among representatives of the different communities of computational linguistics, ontology engineering, artificial intelligence, data mining and medical informatics and bioinformatics. An important objective of this NoE is to feed the biomedical informatics community with the results of state-of-the art research in these disciplines. There has been a deficiency of such communication in the past, with a very small overlap between the people attending and presenting research papers at computer science and ontology symposia (e.g. ACL, COLING, IJCAI, AAAI, KR, FOIS, ECAI, SIGIR) and those presenting at biomedical informatics conferences (e.g. MIE, AMIA, MEDINFO).

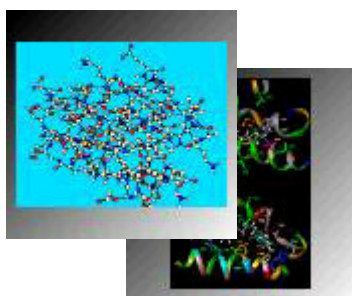
Another bridging activity addressed by this NoE is knowledge transfer and co-operation between academia and organisations and SMEs in the health and welfare sector, including standardisation bodies and the different public and private institutions involved in health care delivery and management. The national institutes and organisations responsible for policy making and quality management with a regulatory and normative function will have an important role to play in the network. We believe that co-operation between these organisations and those involved in research departments needs to be strengthened, both in the early phase of research programme identification and in the later phases of implementation and large-scale evaluation of results and impact. The bridging activities between different levels of the health care system are exemplified in figure 1.

The research carried out under the auspices of this NoE will also address the need for approaches in Europe which will bridge language barriers and facilitate access for non-English native persons to the large scientific corpus of texts written in English.

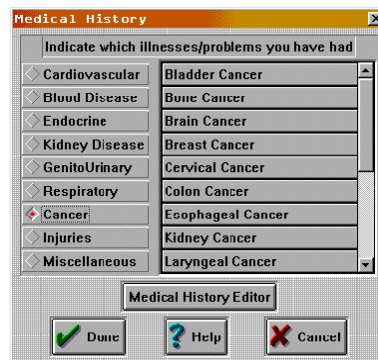
Several of the participating partners in the NoE have a common history in previous EU funded research programmes. However, the list of partners in the proposed NoE has been extended substantially to ensure necessary coverage of the domain. Support for the establishment of a NoE covering the areas of medical informatics, bioinformatics and the standardisation bodies will work against the fragmentation and the under-financing of the field. We strongly believe that a unified effort based on research in the scientific disciplines of medical- and bioinformatics combined with the application of emerging standards and technological infrastructure is a necessary presupposition of the development of a cost-effective distributed health care system in Europe.

The research challenges in focus for the NoE are of specific interest for the health care sector. However, methods and components developed, implemented and evaluated within the health care sector with its high requirements on quality, security, and robustness will also be of general interest and applicable in other domains as well.

The NoE application is based on the partnership of 25 participants from 11 European countries with 99 identified researchers (25 female) and 31 associated PhD students (10 female).



Bioinformatics



Clinical information



Health statistics

Figure 1. The NoE address research issues on three levels; on the pre-clinical level of bioinformatics (functional genomics, proteomics etc.), on the clinical level of primary and secondary health care (hospital information systems, electronic health records etc.), and on the level of health statistics (population-based statistics, epidemiological surveying etc.). The identified areas of research share the basic problem of semantic interoperability, which means that semantics is preserved in communication between users and information systems.

2.2 Integrating Activities

Integrating activities within the NoE will in the early phase focus on ‘get together’ activities through workshops and symposia, where current, predominantly nationally funded, research projects within the network will be presented and analysed. Recent PhD dissertations and on-going PhD thesis work will be presented and discussed. Standards and procedures with respect to the production of PhD dissertations will be compared and possibly harmonised. The set-up and use of a NoE web portal and common database for information exchange, internally as well as externally, will be an integral part of these activities.

The walk-through of current research programmes will then be the basis for the development of joint PhD training programmes, applications for new research projects and identification of further actions supporting the overall objective of network integration. It should be appreciated that the current application with its description of integrating activities and the joint programme of activities is a first version indicating the scope and general objectives of the NoE. The resources associated with the establishment of a NoE is as such a prerequisite for an in-depth analysis of meaningful activities and for further development of informal and formal connections between the actors within the network.

The presentation of the existence and activities of the NoE will be carried out at all relevant meetings and conferences on the national and international levels. Activities for information exchange and knowledge transfer (through joint workshops and meetings) to national institutes and organisations responsible for health and welfare as well as the working groups of the European standardisation organisation (CEN TC 251) will be essential.

The principal components of the integrating activities are:

- ? Socialisation of partners and members of the NoE:
Kick-off meeting, network meetings
- ? Spreading of knowledge and research infrastructure within the NoE:
Tutorials and workshops, common database, web portal
- ? Mobility program:
Cohesive doctoral programs, mobility programs for researchers and PhD students
- ? Common research activities:
Research projects in identified areas of interest
- ? Contribution to standards and to associated improvements in health care policies and practice:
CEN TC251, IMIA, ISO, W3C
- ? Dissemination of results to the scientific community, public domain and industrial partners:
Research papers, conference proceedings, educational materials, patents etc.

Key parameters for follow-up of the progress of the NoE are presented in chapter 7 Quality of Integration..

3. List of Participants

N	Short name	Participant
1a	LIU (IMT)	Linköpings universitet, Dept Biomedical Engineering/Medical Informatics, Sweden
1b	LIU (IDA)	Linköpings universitet, Dept of Computer Science, Sweden
1c	LIU (C-NPU)	Linköpings universitet, C-NPU committee International Federation for Clinical Chemistry and Laboratory Medicine and International Union of Pure and Applied Chemistry
3	KI	Karolinska Institutet, Stockholm, Sweden
4	SU	Västra Götalands läns landsting, Sahlgrenska universitetssjukhuset, Göteborg, Sweden
5	UGOT	Göteborgs Universitet, Dept of Swedish, Sweden
6	UKLFR	Universitätsklinikum Freiburg, Abteilung Medizinische Informatik, Germany
7	UNIFR	Albert-Ludwigs-Universität, Arbeitsgruppe Computerlinguistik, Freiburg, Germany
8	IFOMIS	Universität Leipzig, Institute of Formal Ontology and Medical Information Science, Germany
9	CAU	Christian-Albrechts-Universität zu Kiel, Institut für Informatik und Praktische Mathematik, Germany
10	DIM	Université de Genève, NLP Group - Medical Informatics Division, University Hospital of Geneva, Switzerland
11	UOM	The Victoria University of Manchester, Dept Computer Science, UK
12	UCL	University College London, Centre for Health Informatics and Multiprofessional Education, UK
13	ITRI	University of Brighton, The Information Technology Research Institute, UK
14	INSERM	INSTITUT NATIONAL de la SANTE et la RECHERCHE MEDICALE, Public Health and Medical Informatics Laboratory, UFR Broussais, Paris, France
15	CNR-ISTC	Consiglio Nazionale delle Ricerche, Institute of Cognitive Science, Laboratory for Applied Ontology, Roma, Italy
16	EMBL-EBI	European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK
17	MEDINFO	Országos Egészségügyi Információs Intézet és Könyvtár, National Institute and Library for Health Information, Budapest, Hungary
18	NORDCLASS	Uppsala universitet, WHO Collaborating Centre for Classification of Diseases in the Nordic countries
19	SOS	Socialstyrelsen, Stockholm, Sweden
20	STAKES	Sosiaali- ja terveystieteiden tutkimus- ja kehittämiskeskus, National Research and Development Centre for Welfare and Health, Finland
21	KITH	KITH AS, Norwegian Centre for Medical Informatics
22	NBH	Sundhedsstyrelsen, The National Board of Health, Denmark
24	MRI	Merrall-Ross International Ltd, Cheshire, UK
25	EDSA	European Dynamics Advanced Systems of Telecommunications Informatics and Telematics S.A., Greece

4. Relevance to the Objectives of the IST Priority

It is well known that the health care system is faced with a series of challenges concerning quality and cost-effectiveness. The distribution of health care services in ways which allow the patient to take an active part in relevant decisions and the provision of evidence-based medicine at all levels in the system and the effective use and reuse of information are all key issues for the organisation of health care delivery in Europe. The information and communication technology infrastructure should reflect a view of the health care system as a seamless system where information can flow under the necessary forms of regulation, across organisational and professional – and national – borders.

The need for cross-referencing between biological and clinical information provides another grand challenge. The vast amount of data available in nucleotide and protein sequence databases together with the growing volume of electronically available clinical information composed of images, biosignals, laboratory data, and textual information, call for automated (or at least semi-automated) methods for high-quality indexing, annotation, and cross-referencing through discovery of patterns and relationships. Thus it calls for standardisation and resources for the integration of data derived from divergent sources of the sort which ontology can provide.

The complex application area of distributed health care and the need for cross-referencing between biological and clinical information provide in this perspective a family of general research problems which we believe must be studied in tandem to allow successful systems to be implemented. A list of such related research areas is presented below. The areas share the basic problem of semantic interoperability, which simply means that semantics is preserved in communication between information systems, a condition which should be natural but has proven to be very hard to achieve. It should be emphasised that the identified research areas themselves are thoroughly intertwined even though the corresponding scientific community in Europe is somewhat scattered in different disciplines and distributed in different countries.

Terminology Issues In Health Care

Terminological systems used in health care are systems supporting communication of information through standardised use of language. These systems include thesauri, nomenclatures, classifications, vocabularies and also formally and non-formally represented health-care models. Traditionally such systems have been disseminated through paper-based publications. Now, however, the integration of terminological systems with health care information systems is becoming ever more viable. Thus, the focus of this research is on terminological systems within health care and their use in health care information systems such as documenting and retrieval systems, and for cross-referencing between clinical and pre-clinical terminologies.

Consistent use of language, or more generally the consistent formulation of any information used for communication purposes, is a prerequisite for high-quality classification and for semantic interoperability. An ideal terminological system should thus serve as an inter-lingua in communication between health care workers and between professionals and patients, and permit data comparison across organisational, professional, and even national borders. Concept representation systems based on description logic is a key technology. Description logic, as a knowledge representation formalism based on formal logic, offers a precise language for the engineering of domain ontologies, and provides taxonomic subsumption as a basic inference mechanism. Off-the-shelf implementations exist, and considerable experience has been gained in previous projects in the medical domain, e.g. through the GALEN-project. The health care domain is rich in terminological systems targeted towards different areas and specialities, and several models are proposed as basic reference terminology, possibly providing a common framework for further integration and mapping between specific terminological systems. A great challenge would be to combine the massive coverage offered

by informal medical terminologies (such as ICD10, ICF, ICPC etc.) with the high level of formally solid description logic formalisms. Of specific interest for the realisation of the ideal of a reference terminology based on formal logic is the release of SNOMED CT. The availability of such a large scale reference terminology in all European languages is both scientifically and organisationally a great challenge.

Information Quality

The documentation of clinically useful, patient-specific information is fundamental for the management of health care related problems. In recent years much interest has been devoted to the structure and content of a multi-professional patient record, in part in reflection of the introduction of electronic patient records (EPRs), in part in response to the organisational issues originating from the goal of using the patient record as a means of communicating across organisational and professional borders and as an active instrument in planning and outcome analysis. The EPR has to meet the requirements of quality assurance and evidence-based practice, but at the same time preserve the rich and nuanced character of the patient-centred story. In this perspective, it is well known that the traditional paper-based patient record suffers from a series of limitations.

Consequently, demands for different types of information handling within the health care sector range from clinically useful, patient-specific information to high-level aggregation of data needed for follow-up and statistical reporting. Several coding systems are put to use for these purposes in domains such as diagnoses, health problems, interventions, and procedures. The challenge is to assure high information quality on different levels of abstraction, and to allow aggregation according to different aspects or viewpoints.

Data Mining and Information Retrieval

Considerable progress has been made in the field of text processing and information retrieval during the last decade, but these achievements have not yet reached a level of maturity to be useable at ground level in the health care sector. Successful information retrieval relies on consistent use of an indexing language which must, depending on the actual domain, comprehend terms from each of the existing terminological systems. This research will also address the multilingual need of researchers and medical practitioners Europe, not only in order to help bridge language barriers between different European countries but also to facilitate access to non-English native persons to the large scientific corpus of texts written in English.

Information retrieval techniques include mathematical modelling of the relation between document content, e.g. the content of the patient record, and the end-user query. Generally, the set of terms used for indexing and thereby also for retrieval is a flat list without internal structure apart from what is given lexically. An interesting research perspective is how the structure of conceptual models imposed on the indexing system could be utilized to improve the performance of search engines. Evaluation of the results of information retrieval and text mining requires extensive annotated test corpora of a type which do not yet exist for the medical domain, a deficiency addressed by the NoE.

Bioinformatics - Genomics, Proteomics and Related Fields

The mission of the European Bioinformatics Institute is to ensure that the growing body of information from molecular biology and genome research is placed in the public domain and is accessible freely to all facets of the scientific community in ways that promote scientific progress. Much interest is also expressed from the European Commission in networking of researchers in areas of relevance to the development of new systems assisting in the individualisation of disease prevention, diagnosis and treatment, a challenge that is seriously addressed by the proposed NoE.

A long term goal of the joint activities within the NoE is to develop a common framework for mining of relations in large data volumes composed of signals, symbols, text etc. Such a framework will be critical for knowledge extraction from clinical and pre-clinical databases once account needs to be

taken also of the information explosion of genomics and proteomics. The task of finding means to search automatically for interactions between genes and proteins on the one side, and clinical diagnostic information on the other, is a true challenge of our century.

The Semantic Web

High requirements on semantic interoperability are bound up with the project of the Semantic Web. Web search engines have become extremely powerful in terms of processing speed and information storage, but there still exist fundamental problems concerning web navigation and information retrieval, as also with harvesting the potential of more sophisticated forms of processing of information on the web. The Semantic Web is conceived as a system where the requirement of exact matching between search terms is relaxed while additional behind-the-scenes functions are provided to establish semantic equivalence and other types of associative relations between expressions or concepts. We understand our work as an active contribution to the Semantic Web initiative, which offers a useful framework for adding semantics to web documents by using metadata based on common standards, such as XML for document markup and DAML+OIL for expressing the underlying knowledge. Different ways to generate semantic web compatible documents will be analysed, e.g. the knowledge-based authoring of documents, semantic tagging of unstructured text, etc. In addition work will be encouraged critical of details of the Semantic Web project but embracing its general goals.

Adaptive Interfaces and Ubiquitous Computing

The development of semantic-based and context-aware systems to organise, process and visualise the vast amounts of both individual and general health care information is at the core of the activities of the proposed network. The success in meeting the research challenges addressed by the network will have an impact also on the development of adaptive multimodal interfaces. The distribution of information services over organisational, national, language and other sorts of social and economic borders in Europe requires foundational research on new methods and languages for information and knowledge representation, even more so if the full power of mobile equipment will be utilised, allowing ubiquitous computing with ‘anytime-anywhere inferencing’ to be implemented on a large scale.

5. Potential Impact

5.1 Contribution to Standards

The proposed effort will contribute to the European research area through the establishment of a multidisciplinary network composed of researchers, industry and health care professionals. This network of experts will establish close links to those working formal standardisation efforts ensuring that the research results can be used without delay to influence the formal standards that guide product development and procurement of interoperable solutions. The NoE will in particular work with the European Standards Committee CEN/TC 251 and its four working groups (I) Information Models, (II) Terminology and Knowledge Bases, (III) Security, Safety and Quality, and (IV) Technology for Interoperability. Work in the two first working groups is at the core of the proposed network activities. Co-operation will also be done globally with the ISO/TC 215 and through the IMIA and W3C organisations. Cooperation with WHO collaborating centres and national boards for health care system administration and epidemiological studies is another important target for standards setting which this NoE will address.

5.2 Contribution to Policy Development

The health care system in general is faced with a series of challenges concerning quality and cost-effectiveness. The distribution of cost-effective health care allowing the patient to take an active part in health care decisions is a high priority issue in several European countries. The development of generic methods and components facilitating robust, scalable and easy-to-use products and services in the field of “eHealth” has the potential to effect the health care system as a whole in positive ways and also to assist the health care professional and the patient in coping with risk management and decision making.

The Commission Communication COM (2000) 285 final of 16 May 2000 on the health strategy of the European Community presents several challenges which the Community Health Strategy should address. One of these concerns the *Quality of Health Systems*, which has a major impact on patients’ wellbeing. Member States have been undertaking a wide range of structural reforms to improve the efficiency and effectiveness, including the total management of quality, of their health systems.

On 23 September 2002, the European Parliament and the Council adopted a Decision establishing a programme of Community action in the field of public health (2003-2008) (Decision 1786/2002/CE (OJ L 271/1 of 9.10.2002).)

The general objectives of the programme are:

- ? to improve information and knowledge for the development of public health
- ? to enhance the capability of responding rapidly and in a coordinated fashion to health threats
- ? to promote health and prevent disease through addressing health determinants across all policies and activities.

It was further elaborated in the 2003 work plan that Health information activities involve "action on health information and knowledge, the development of a sustainable information system at EU level will be continued. It involves the definition, collection and exchange of data, building on data that are available or collectable, taking into account of the position in the Member States and applicant countries. The outputs of the system – including reports and analyses focusing on specific population groups or health concerns - will lead to policy spin-offs at Community level."

This NoE will through its research activities contribute to these general objectives of the community facilitating the collection, analysis and presentation of information on public health issues and quality performance of health care services in the different member states. The use of international classifications in health care is a prerequisite for medical information processing and has gained increasing importance for comparative health statistics, quality assurance, clinical research, assessment of productivity and resource utilization for use by the member states public health authorities, researchers and citizens. In addition the activities of this NoE will contribute to the development of more efficient direct care of the individual patients, particularly when several organisations need to co-operate for a continuous seamless care process.

5.3 Risk Assessment

There is obviously a need for ethical considerations for projects related to the use and reuse of health information in general and of patient-specific information in particular. The activities of this NoE are foreseen to have an impact not only on the European Research Area and European industry, but also on the health care professional and the empowered citizen and patient acting according to more or less well defined roles and responsibilities. The contribution of the partners representing public health and welfare organisations will have an essential role in the NoE in bringing ethical considerations into the different activities of the network. Knowledge transfer between the research community and the

standardisation bodies will also contribute to the awareness of security and privacy in relation to health care applications.

Apart from the general awareness of policies and legislation in relation to health care applications, the handling of patient-specific information and the anticipated mapping of pre-clinical and clinical information (the so-called genotype and phenotype information) should be addressed. All data generated through the different research activities or used in demonstrations and made available for the NoE in the common database or publicly available via the website will be anonymous or be subject to techniques for pseudonymisation to provide linkage of data that is not identifiable to a person.

A long-term goal of the NoE is to develop a common framework for mining of relations in large data volumes composed of signals, symbols, text, etc. Such a framework will be critical for knowledge extraction from clinical and pre-clinical databases especially when account needs to be taken also of the information explosion of genomics and proteomics. The task of finding means to search automatically for interactions between genes and proteins on the one side, and clinical diagnostic information on the other, is a true challenge of our century, but also a task where careful ethical consideration is needed. This area will be considered as subject for a dedicated Committee or Work Package during the second period of the network.

5.4 Impact on the Citizen in the Information Society

Medical laypersons are supposed to play an ever more an active role as partners of their physicians in the management of chronic diseases. The Internet has become an outstanding source of health-related knowledge, but both multilinguality and the characteristics of the medical sublanguage make difficult the task of retrieving relevant texts. Europe's diversity is reflected by its multilingual character, which constitutes a cultural asset whose value and importance is becoming increasingly recognised.

Although an increasing number of Europeans – probably more than in any other part of the world are fluent in more than one language, and European schools excel in the value they give to foreign language teaching, linguistic barriers exist. The modern information society, with its ever growing amount of textual information, has not yet developed ready-to-use retrieval tools for searching multilingual document collections. The need for such tools is obvious, not only for targeting texts in English as the *lingua franca* in science and technology but also for targeting languages of the same family (Danish-Norwegian-Swedish; Spanish-Portuguese; Polish-Czech) which can be read but from which information cannot be easily retrieved due to spelling and morphologic differences. Multilingual text retrieval is useful for accessing both scientific and non-scientific literature.

The citizens/patients are playing an ever more an active role as partners of their physicians in the management of chronic diseases. The Internet has become an outstanding source of health-related knowledge, but both language barriers and the characteristics of the medical sub-language make difficult the task of retrieving and understanding relevant texts. This NoE will particularly address tools for providing citizens with automatic assistance in various languages on medical knowledge.

In addition the efforts on providing a common conceptual framework and reference terminologies for describing clinical findings and health care services performance will enable much more information to be presented to the citizens comparing different member states and health care organisations as outlined in the EU public health programme.

The emerging cross-border health care services where particular services such as operations sometimes are performed in another country will be greatly facilitated by a common semantics for health allowing exchange of medical records across language barriers. This semantic interoperability across Europe is also essential for the free movements of the citizens considering requirements for emergency treatment and transfer of clinical information from the home country when temporarily visiting another EU country for work, studies or vacation.

5.5 Impact on the ICT Industry

The development and large scale implementation of useful services in the health care domain could also, as a secondary effect, have an impact on the information and communication technology industry (ICT). It is well known that queries and searches regarding medicine and health care on the Internet are a key topic, and further development of useful and easy to use semantic-based and context aware services could be the basis for development of new business plans for a variety of enterprises all over Europe.

An important frontier for new product research and development for the IT industry as a whole is the successful integration of informatics technologies - semantic or natural language processing based both alone and in combination - to manage and analyse data that is not primarily numeric. However, this is a particularly acute need in the medical and bioscience domains, both of which are characterised by very large quantities of such descriptive data.

The medical and bioscience communities increasingly believe that an absolute requirement and an opportunity exists to link their two domains - to link genotype and phenotype information - as a vital step on the road to the promise of the post-genomic world. However, establishing and maintaining this link has made the need for informatics support yet more acute, and more complex. Supporting this research activity, therefore, will itself create a highly specialised market for new products. Further, it is expected that delivering the fruits of genophenotypic research - for example, software tools to customise clinical treatments according to the patient's known genotype - will create another larger and more general market, aimed at both professional users and the citizen.

By bringing together those at the forefront of European research, this network therefore seeks both to support the development of key informatics technologies to the point where they can be productised, as well as to grow the community of people who understand them. Through the network, the ICT industry will gain access to both expertise and trained personnel in an important emerging technology. Finally, the network seeks to counter the risk that, in the absence of any strong advocacy and research for inherently multilingual informatics solutions, the de facto global solution will be monolingual (i.e. English only).

6. Joint Programme of Activities

6.A Activities

6.1 Integrating Activities

The Joint Programme of Activities (JPA) described in chapter 9 is built up out of the following principal components:

- ? Socialisation of partners and members of the NoE:
Kick-off meeting, network meetings
- ? Spreading of knowledge and research infrastructure within the NoE:
Tutorials and workshops, common database, web portal
- ? Mobility program:
Cohesive doctoral programs, mobility programs for researchers and PhD students
- ? Common research activities:
Research projects in identified areas of interest
- ? Contribution to standards and to associated improvements in health care policies and practice:
CEN TC251, IMIA, ISO, W3C
- ? Dissemination of results to the scientific community, public domain and industrial partners:
Research papers, conference proceedings, educational materials, patents etc.

It should be appreciated that the scope and general objectives of the NoE described herein represent only part of an evolving JPA plan. It should also be appreciated that the funding for a NoE does not realistically support new large scale projects. However, through a series of relatively low cost activities the NoE instrument can contribute in substantial ways to the realisation of more cohesive research activity, to the spreading of excellence and to the generation of new projects at the forefront of international research.

The resources associated with the establishment of a NoE are to be conceived in this light as a prerequisite for the coherent planning of meaningful joint research activities and for the development of meaningful informal and formal connections between the actors within the network.

The activities detailed in our list of work packages are based on an initial identification and analysis of a series of on-going research projects funded through national and EU grants. These research activities are currently somewhat fragmented. The hypothesis is that through cross-fertilisation they will be moulded into a cohesive research framework at the very forefront of international research. This in turn will bring about a snowball effect, in that the process of extended collaboration and enhanced research quality will increase the levels of funding which the partners will be able to obtain in the future.

Integrating activities within the NoE will focus in the early phase on get-together activities through workshops and symposia, where current, predominantly nationally funded, research projects within the network will be presented and analysed. Recent PhD dissertations and on-going PhD thesis work will be presented and discussed. Standards and procedures with respect to the production of PhD dissertations will be compared and possibly harmonised. The set-up and use of a NoE web portal and common database for information exchange, internally as well as externally, will be an integral part of the activities. Tutorials and workshops will be organized both as physical and as virtual meetings, utilising the Internet and web technology.

The existence of the NoE and the results of its work will be presented at relevant meetings and conferences on the national and international levels. In addition we will ensure information exchange and knowledge transfer through joint workshops and meetings organised in conjunction with national

institutes and organisations responsible for health and welfare as well as the working groups of the European standardisation organisation (CEN TC 251).

The program of jointly executed research activities will demonstrate that the NoE, besides favouring the integration of partners, has the effect of stimulating wider joint research activities in the identified areas. All work packages are led by one partner who in co-operation with two or three other partners will be responsible for organisation and realisation of the work package. This co-operative approach will in itself foster co-operation between partners. Different sets of partners will be participating in the different work packages.

Work packages supporting integration during the first phase of the NoE lifecycle are listed below.

<i>WP no</i>	<i>Title</i>	<i>Responsible</i>
1	Project Management	Management Office, Board
2	Strategic planning	Board, Committee
3	Common database/repository	Board
4	Public website	Board, EDSA
5	Planning of network meetings	Board
6	Mobility programs	ITRI
7	Summer school 2004	MEDINFO
8	Participation in standardisation work	KI
9	Dissemination and exploitation	Committee, MRI
10-16	Workshops/tutorials on specific topics: Health care statistics The semantic web Ontology engineering Natural language processing Data mining in EBI databases Text mining and information retrieval The electronic health record	LiU (IMT) LiU (IDA) IFOMIS DIM EBI UKLFR/UNIFR UCL

Table 1. List of integrating work packages

The objective of WP1 is to extend the preliminary analysis of excellence and on-going research activities within the network performed during the pre-proposal phase, and to oversee the overall progress of the integration process among partners. Activity areas already identified and formulated as work packages will be further developed in terms of tasks and mobility plans. Further areas suited for joint research activities will be identified and analysed. An extended inventory of the competences of each partner and the complementarity of these competencies will be prepared in the form of a database made available via the Internet. The expectation is that the NoE will evolve into a self-sustainable network after the EU financing period. The key objective of the strategic planning (WP2) is to facilitate this, as well as to hammer out joint proposals funding for the funding of future work. The key issue for success in terms of a sustainable co-operation among partners will not be the construction of new permanent and inflexible administrative organs, but dynamic and flexible constructs facilitating sharing of resources, mobility of persons among countries and organisations, and the spreading of excellence, the latter being the key driving force for the research community in establishing European research groups at the forefront of international research and development.

The objective of WP3 is to initiate the task of building a common database repository for experimental data (e.g. corpus of annotated medical text), terminological systems and thesauri, formal and application ontologies, as well as associated technical and scientific information useful for the integration of activities in the NoE. Due to the variety in different data types, this task will begin with

the attempt to find solutions to problems standing in the way of the necessary data standardisation. The objective of WP4 is to design and implement a public Website where information is made available about the network as a whole, together with competence profiles for each partner and other relevant public information about the progress of NoE activities. The use of Internet-based technology for supporting and maintaining communication within the NoE (e.g. through mailing lists, email fora, virtual meetings, etc.) will be an integral part of the management of the NoE.

Network meetings (WP5), both physical and virtual, will be essential for the integration process. The objective of this activity is therefore to identify, schedule and assign planning responsibility for a series of network meetings. The general objective of network meetings is to speed up the integration process and to foster co-operation and exchange of know-how and research infrastructure within the network. Network meetings will be of different kinds, comprehending a kick-off meeting, workshops, a summer school, and joint national and international conferences and seminars.

The objective of WP5 is further to identify, schedule and plan participation in regular national and international conferences in the area of interest for the NoE. By co-ordination of NoE meetings and travels with regular conferences and symposia, travel and resource consumption can be minimised. Through co-ordinated and planned participation of NoE members in conferences, and through the organisation of dedicated workshops within the framework of larger conferences, the visibility of the NoE will be enhanced, and cutting-edge research of interest to the NoE framework will be identified and brought to the awareness of all the partners.

The objective of WP6 is to identify opportunities for exchange of researchers and PhD students between partners in the NoE, and to plan and organise these as short-, medium- and long-term visits. The objective of WP6 is further to compile descriptions of recent PhD dissertations and on-going PhD thesis work. Standards and procedures with respect to the production of PhD dissertations will be summarised and compared. The compilation will be used as a basis for a possible process towards harmonisation of procedures and for sharing of senior research staff as co-tutors in doctoral programs handled by the separate universities.

The objective of WP7 is to plan and realise a summer school in 2004, where researchers and doctoral students in the network, perhaps including guests from the US and elsewhere, can gather in one place for an extended period (at least a full week) for an informal program of workshops, tutorials, demonstrations and social activities. A preliminary location of the summer school 2004 is Budapest or Veszprem close to the lake Balaton in Hungary. The planning of annual NoE meetings, e.g. in the form of an annual summer school, is the responsibility of WP5.

Workshops and tutorials will be used as a means for knowledge transfer in the NoE. The objective of WP5 is, as an extension and continuation of the already proposed workshops to further identify material and methods of interest, and to plan and organise workshops or tutorials for spreading of excellence within or without the NoE. Key topics identified as subjects for workshops or tutorials are listed below (WP10-WP16).

- ? Health care statistics
- ? The semantic web
- ? Ontology engineering
- ? Natural language processing
- ? Data mining in bioinformatics databases
- ? Text mining and information retrieval
- ? The electronic health record

Participation of partners from the NoE in standardisation work organised e.g. by CEN TC251, ISO, IEEE, W3C will be encouraged, organised and partly funded. The knowledge transfer will be bi-directional, bringing the results of the different research activities into the standardisation process, but also bringing issues dealt with in the global perspective of European and international standardisation into the framework of the NoE activities. WP8 will specifically be devoted to the planning and realisation of participation in standardisation work.

The different activities within the NoE will produce a large volume of valuable educational material. The sharing and distribution of already existing course material on under-graduate and research education level, as well as newly compiled material produced for the different tutorials and workshops, PhD thesis, research papers, conference proceedings, reviews, reports and NoE deliverables etc. should be analysed. WP9 will analyse ways of dissemination of all these types of material, including identification of target groups (e.g. under-graduate students, PhD students, researchers, health care professionals, health care policy makers, industrial partners, public audience etc.) and different means for spreading and possibly commercialisation of the material.

An inherent problem of information management addressed by the NoE activities is the indexing and retrieval of information. This expertise will be applied to the systematic indexing of material produced within the NoE. One of the SMEs (MRI) is specifically devoted to providing indexing and information retrieval services to international medical publishing houses, and the development of improved medical information extraction and presentation techniques, while another SME (EDSA) has extensive experience from the development and maintenance of public web sites. WP9 will be the responsibility of a dedicated committee with both internal and external members.

For a more detailed description of work packages in terms of content, participants, deliverables and interactions among work packages, see the list of work packages in chapter 9.

6.2 Programme for Jointly Executed Research Activities

The proposed topics for workshops and tutorials are identified as core knowledge areas in response to the main objective of the NoE, that is to establish an integrated network of European research institutes and departments in the forefront of international research. These knowledge areas will then form the basis for joint research activities in strategic areas. Identified areas for joint research activities rely on an analysis of the on-going activities of the separate partners (see Appendix 1 for description of competencies and research activities of NoE partners). The hypothesis is that joining forces will create frontline research in areas of relevance for the European health care system as well as for the European ICT industry. Identified areas for joint research activities are listed in the table below.

The identified research areas all share the underlying problem of semantic interoperability, which means the problem of ensuring that semantics (meaning, understanding) is preserved in communication between users and information systems. This condition has proven to be very hard to achieve, especially in a multi-lingual context of the sort which we find in Europe, and especially within the health care context, with its different levels of specialisation and different users and lack of standardisation in the development of information processing tools and frameworks.

<i>WP no</i>	<i>Title</i>	<i>Responsible</i>
20	Multilingual medical dictionary	UKLFR/UNIFR
21	Ontology engineering	UOM
22	SNOMED CT	KI
23	Health care statistics	LiU (IMT)
24	Data mining and information retrieval	EBI
25	Concept systems for laboratory medicine	LiU (cNPU)
26	The electronic health record	UCL

Table 2. List of joint research areas

Multi-lingual Medical Dictionary

The lack of a large-scale multi-lingual medical dictionary hampers the integration of European research activities in the medical field, and more seriously also the development of multi-lingual information retrieval services. An interesting language technology useful for this problem is corpus-based machine translation. The aim of this project is to develop techniques and systems for lexical data generation from parallel corpora, and to develop and apply methods for evaluation of machine translation systems. Parallel corpora exist e.g. as translations from English to other European languages of the official WHO classifications and some other terminology systems. Several of the NoE partners have extensive experience in multilingual lexical resources and computational lexicography, while others have an interest in applying such tools e.g. for semi-automated translation, semi-automated coding and indexing, and advanced systems for information retrieval.

Ontology Engineering

The terminological systems used in health care and in other domains are designed to support information communication via standardisation in language usage. These systems include thesauri, nomenclatures, classifications, vocabularies and also formal and non-formal frameworks for representing health-care phenomena. Traditionally such systems have been disseminated through paper-based publications. Now, however, the prospect of integrating terminological systems with health care information systems is becoming ever more viable. The field of information systems ontology is divided into three disciplines: investigating underlying ontological theory and philosophy, investigating formal logics for computing over ontologies, and investigating how to construct and maintain ontologies that are functionally useful. All three schools are represented in the NoE. Several network partners have long histories of working in the biomedical ontology field, using various philosophical and formal approaches. The objective is to share understanding across the three ontological disciplines, especially those issues unique to or especially important to the biomedical domain, and to achieve coordinated input into the emerging semantic web ontology standards for both computational logics and ontology authoring environments.

SNOMED CT as Reference Terminology

A large and potentially very interesting terminological system is SNOMED CT which in its new release (CT) includes the former SNOMED III and the Read Clinical Terms. SNOMED CT is intended to serve as a reference terminology by which composite concepts could be defined out of more primitive ones. As a system originating from the College of American Pathologists and the UK Read Clinical Terms, SNOMED CT is developed and maintained in the English language. Partial translations are available e.g. into French, German and Spanish, but the development of a more comprehensive multi-lingual perspective can help to resolve the many problems by which SNOMED CT is still affected and so lead to a more robust large-scale reference terminology for the European health care system. The objective of this research activity is to share experiences and understanding of

the use of large scale reference terminologies in general and SNOMED CT in particular and to identify and resolve the problems in the construction of such systems as they now exist. Moreover, the objective is to encourage sharing of methods and tools for translation, and to co-ordinate the evaluation of SNOMED CT as a reference terminology e.g. for structured data entry applications in different domains.

Data Mining and Information Retrieval

Several of the partners of this NoE have extensive experience from data mining and information retrieval, but there is generally a lack of information exchange between projects and applications in different domains, e.g. image and signal processing, database analysis, and text processing. The overall objective of this research activity is therefore to develop a common understanding and framework for mining of relations in large data volumes composed of signals, symbols, text etc. Such a common framework will be critical for knowledge extraction from clinical and pre-clinical databases

Synergies between medical informatics and bioinformatics in the area of data mining should be emphasised. The EBI is Europe's major molecular biology database producer, bioinformatics service provider, and bioinformatics research centre. The major focus is on development of databases of nucleotide and protein sequences, protein structures, micro-arrays and genome annotation. A network of worldwide collaborators are engaged in these projects, and many other specialised databases (protein interactions, enzymes, immunogenetics, mutations) are actively pursued. More recent exploitation of text resources and ontology development has triggered a variety of biomedical text-mining activities relevant to this network of excellence.

The objective of this activity is to share in-depth knowledge of the content and structure of EBI databases, as well as knowledge of useful methods and tools for data mining in different types of databases. This NoE activity will promote the establishment of new collaboration between research groups of medical informatics and bioinformatics and thereby the authoring of new research applications targeted towards national and European research foundations.

Concept Systems for Laboratory Medicine

The practice of computing in medical laboratories has reached a maturity level that requires clear insights into what the efforts are supposed to achieve in view of seamless communication of laboratory generated data for the benefit of the individual patients. This should be so regardless of administrative, technical, language or cultural boundaries. The availability of electronic patient records will also greatly increase the quality of health care, reduce health care cost and facilitate epidemiological surveys to the benefit of patient's. Thus in developing a generalised architectural build up and adherent structures, there is a need for systems and schemes in each of the medical domains to support and populate the electronic patient record.

The work package is meant to provide a forum for researchers developing connectivity between the bioinformatics databases, the analytical laboratory process and the electronic health care record. The long term objective is to develop concepts that can be understood, aggregated and generalised depending on the viewpoint of various healthcare worker and their IT support systems without compromising the basic meaning of laboratory reports including information based on data produced by bioinformatics.

The Electronic Health Record

Standards for representing and communicating electronic health record (EHR) information are now in their third iteration within Europe. Implementation experience of EHR middleware is growing through the results of successive EU Health telematics R & D projects in the 3rd, 4th and 5th Framework programmes. However, much of this work has focused on the federation of legacy feeder systems, and the interoperability of record servers with clinical applications. Work on integrating the structure and content of individual patient records with terminology servers, clinical ontologies, alerting systems,

guidelines and decision support services is at a much earlier stage. This work is vital if EHRs are to contribute successfully to the practice of evidence based medicine – to support the care of individuals and to enable the evaluation of clinical outcomes, care provision through audit and next-generation research queries.

The practice of evidence-based medicine will require the seamless and standards-based interaction of knowledge, record and inference services. There is presently limited ability to analyse individual patient records that have been federated from multiple sources, or to interrogate population records that are comprehensive, multi-enterprise and possibly organised at the level of a regional health care network. The human and skills network established through this work package will greatly inform the next generation of such EHR demonstrators, will contribute to the development of tools, and will place the individual research teams in a good position to bid for future collaborative projects through recognised research funding bodies.

Health Care Statistics

While several organisations within Europe and throughout the world collect statistical data on health and national health care systems, the general problem is the low reliability of data, conflicting interpretations and definitions and the poor exploitation of the information caused not least by the absence of standardised frameworks for data representation. Statistical data produced in the health care field have to be exploited by at least three different communities: health policy makers, scientific researchers, and the public audience. These communities need different forms of presentation and employ different modes of interpretation of the same data. The two scientific communities who work in this field – drawn from the domains of medical informatics and public health – have little connection and generally poor knowledge about each other. Therefore workshops and tutorials are planned to facilitate joint research activity. The topic of interest covers – among others – the following fields:

Ontology of health indicators

Presentation and representation of data

Tools for navigation among data in a multidimensional space

Quality of data: reliability, accuracy, usefulness

Theory and practice of coding systems.

The objective of this research area is to contribute to the improvement of European Health Statistics by dissemination of related knowledge and by achieving consensus in areas where conflicting practices, theories and interpretations occur. Further to share experience and understanding on ontologies for health indicators, to promote the development of statistical methods for measuring information quality (e.g. reliability), and methods for quantification of semantic distance or imperfect matches. Moreover, the objective is to encourage sharing of data material (e.g. quality registries, coded patient cases) applicable for development and evaluation.

6.3 Spreading of Excellence Activities

The activities related to spreading of excellence are already indirectly described above. In summary the activities can be grouped in the following way.

- ? Spreading of excellence among researchers and PhD students within the NoE.
Specific activities in this category are the Common Database, Public Website, Mobility Program, Summer School, Tutorials and Workshops, and the Joint Research Activities.
- ? Spreading of excellence between academia and standardisation bodies and public health care organisations.
Since the composition of the NoE is a mix of universities and health care organisations (WHO collaborating centre, public health and welfare organisations etc.) the different activities of the NoE will contribute to the exchange of knowledge and perspectives among these partners. A specific activity will focus on communication between the NoE and relevant standardisation bodies.

? Dissemination and exploitation of results.

The NoE has from start two SMEs as partners. It will be the responsibility of a specific committee to work on strategic planning including issues such as long term sustainability of the network, inclusion of new partners and particularly in the following phases inclusion of new SMEs, together with exploitation of results. An inherent issue in relation to exploitation of results is the handling of intellectual property rights which will be addressed in the Consortium Agreement and further elaborated by the committee.

? Public relation.

The dissemination of material produced by the different NoE activities will be the responsibility of a specific committee. The task is here to work in close co-operation with the Public Website group, but also to analyse other ways of publishing and spreading of information about and from the network.

6.4 Management Activities

The management of the NoE will be set up from a 'simplification and maximisation' perspective, which implies simplification of NoE overhead and administration on the one hand, and the maximisation of information and knowledge transfer within the NoE on the other, focusing on co-operative high-level research among European researchers and dissemination and exploitation of results. This balance between simplification and maximisation will be crucial for the success and sustainability of the NoE, which as an EU-financing instrument has to be seen both from the European Commission's perspective (with the overall objective of re-shaping the European Research Area), and from the research community's perspective, where key features will be contribution to and development of research excellence under an atmosphere of trust and transparency and in such a way as to promise decision making which is sufficiently flexible and efficient to promote the development of the NoE and to ensure work of the highest scientific standards.

Thus, the objective of the NoE management will be to facilitate effective administration, transparent and accepted decision making, dynamic and flexible constructs supporting sharing of resources, mobility of persons among countries and organisations, and spreading of excellence. The latter being the key driving force for the research community in establishing European research groups at the forefront of international research and development.

The key elements of the management will be the following.

The Management Office at Linköping University, Department of Biomedical Engineering/Medical Informatics, headed by Dr Hans Åhlfeldt. The Medical Informatics Group at the Department of Biomedical Engineering at Linköping University has a long history in the field of medical informatics and is nationally leading in the field. The group has during the last decade been partners of several EU-funded projects within the Advanced Informatics in Medicine programme. The group has established links with partners in the Nordic Terminology Network, several international departments. Linköping University will support upcoming projects within the sixth framework by an International Office with expertise on legal affairs, international agreements and exchange programs, management and business administration.

The Assembly is the principal institution of the NoE and is composed of one duly authorised representative of each partner. The Assembly will meet at frequency defined in a Consortium Agreement. The aim is to have a decision making body which is sufficiently flexible and efficient to promote the development of the NoE.

The Board is an operative body, working in close co-operation with the Management Office and with close consultations with the committees and work package leaders. The Board is responsible for all pro-active measures concerning defaulting parties and for management of the NoE within the framework set up by the Assembly in accordance with the EU Contract and the Consortium

Agreement. The Board shall constantly seek to ensure close consultation with all partner as well with as with the established committees.

The establishment of Committees and Work packages will ensure a flexible and dynamic steering of the network. The principle difference is that Committees might have also external members (although they can not be funded through the NoE budget without being subcontractors) ensuring external input to and evaluation of the NoE activities, while Work packages are the construct encapsulating defined activities within the NoE. Members from other NoEs in the eHealth area will be considered as candidates as Committee members. Both the Committee and the Work package are constructs with a specific objective and defined duration, while the other constructs described above will be permanent parts of the management of the NoE. All Committees and Work packages will be lead by an appointed co-ordinator or leader.

6.B Plans

6.5 Plans for Dissemination of Knowledge

Internal as well as external knowledge transfer is a major objective behind the NoE. Knowledge transfer will be carried out along the following lines:

- ? communication of research at the forefront between departments of the NoE,
- ? communication of research results to the scientific community through research papers, conference proceedings etc,
- ? communication of research results and state-of-the-art in technology between academia and public health care organisations and ICT-partners,
- ? communication of priorities and policies between health care organisations and academia,
- ? communication of research results to the public.

Apart from knowledge transfer, the NoE will be useful for pooling of resources and sharing of research infrastructure (e.g. databases with research material and software systems). The issues of dissemination and exploitation of results will be the responsibility of two specific work packages (WP7-8). Internal knowledge transfer will be supported through a variety of activities, including integrating activities, workshops and tutorials and joint research programs.

6.6 Gender Action Plan

The research challenges upon which the work of this NoE are focused are of interest not only to the research community and the ICT industry but also to the health care system and to the citizens utilising health care services of the EU. A specific aspect of the research agenda is the development of multi-lingual services, which will bridge language barriers and facilitate access for non-English-speaking persons to the large scientific corpus of texts written in English. By its nature, the content of the research is gender-neutral and does not require a specific analysis of the impact of the work on any particular social group. It is however essential, that the partners ensure through their work that all social groupings are encouraged to participate in the program, particularly those that have traditionally been under-represented in this scientific field. In this way, the network can significantly contribute to the achievement of the European Research Area in promoting gender equity and accessing the untapped potential of all members of our society. Moreover, the issues addressed in the field of health care statistics such as common ontologies for health indicators and quality of health statistics, will contribute to the enhanced reuse of health information for follow-up of the health care system on different levels including aspects of gender and socio-economics in the delivery and consumption of health care resources.

The Network management will encourage all partners to work towards the achievement of the goal of 40-60% participation of any gender in all kinds of positions within the network. Of general interest is the participation of women in scientific research and particularly in the consortium management team. The Network management will have as part of its objectives the tasks of encouraging the participation

of women in the management of the project in such a way as to ensure a balanced representation and the equitable resolution of any gender-related issues that arise. The Network management team will also encourage the distribution of Gender Action Plans that might be available for the different partners in the network. Any specific gender implications identified by the participating public health care organizations during the course of the joint program of activities will be acted upon by the project management.

The consortium will support the findings of the ETAN (European Technology Assessment Network) report and the so called 'Helsinki Group' that recommended the development of indicators on the situation of women in research. The subsequent Gender Watch System operated by the EC will be provided with information based on our monitoring as part of the annual reports that will provide valuable data concerning the progress of women in scientific research.

A specific element of the project action plan relates to the public image presented by the project through its dissemination activities within the international scientific and wider community. All material will give a balanced representation of all social groupings to prevent any gender stereotyping. The partners will consult with relevant external agencies to assess the gender impact of the material used by the network in order to present an unbiased image of the field covered by the project.

According to the Amsterdam treaty the gender dimension must be integrated in all EC policies and cannot be derogated in any community action; equal opportunity policies cannot therefore be sectoral but must be integrated in all activities (gender mainstreaming). In IT, statistics bear witness to a conspicuous under-representation of women, whose total participation in university education reaches only about 20%. Although this figure can be traced to problems at a far younger age it still indicates the perception for women of difficulty in establishing a career in this field. In both the IT sector, as everywhere, vertical discrimination is present, making it difficult if not impossible for women to reach top positions.

The positive actions envisaged by the NoE are the following:

- 1) To achieve the target established at European level for the participation of women in the over all composition of the management.
- 2) To provide a selection mechanism ensuring equal opportunity for bursaries.
- 3) To implement a gender- approach to training, in order to overcome psychological and cultural stereotypes that prevent women from participating with effective parity in the world of ICT.
- 4) To monitor every network activity in order to evaluate the compliance with gender mainstreaming, following the recommendations of the "Helsinki Group".
- 5) To emphasize and valorize the role of women in history where it has been conspicuously omitted in traditional approaches, as far as content of applications and case studies is concerned.
- 6) To give a balanced representation of the genders through all dissemination activities, avoiding gender stereotyping.

6.7 Raising Public Participation and Awareness

As discussed in section 5.4 the citizens/patients are increasingly playing an active role as partners of the health care professionals in the management of their health problems. The Internet has become an outstanding source of health-related knowledge, but both language barriers and the characteristics of the medical sub-language make difficult the task of understanding medical textbooks or the content in the patient record. The emerging cross-border health care services need to be supported by a common

semantics for health allowing exchange of medical records across language barriers. The same challenge apply for follow-up of health status and utilization of health care resources in Europe, where health indicators and health statistics must be comparable. This semantic interoperability across Europe is essential for the free movements of the citizens considering requirements for emergency treatment and transfer of clinical information from the home country when temporarily visiting another EU country for work, studies or vacation. In the section on Ethical consideration issues concerning security and privacy in relation to health care information as well the linkage of clinical and pre-clinical data are discussed, topics which need to be addressed by the research community as well as health care professionals, politicians and the public.

NoE-activities concerning public participation and awareness can be described by the following.

- ? Direct participation of health care workers, patients and citizens in research activities, such as pilot testing and evaluation of developed tools and systems.
- ? Participation of researchers in public debates concerning health care services.
- ? The public bodies in the health care and welfare area participating in the network will particularly play an important role as mediators of research issues and research progress to politicians, health care professionals as well as to the citizens.
- ? The Committee for dissemination of results will also have the responsibility to address the issues of public awareness of network activities.

6.C Milestones

6.8 Major Milestones

Milestone 1 (months 8): Technical infrastructure supporting communication and integration up-and-running.

Quality indicators: Q2 Sharing of research software tools, Q3 Logging of Website
Assessment based on Deliverables D3, D4.

Milestone 2 (months 2/8): Successful realisation of kick-off meeting and summer school

Quality indicators: Q1 Workshops and tutorials, Q6 Mobility programs
Assessment based on Deliverables D1, D5, D7.

Milestone 3 (months 12/24): One/ten co-authored research reports (papers, proceedings etc.).

Quality indicators: Q7 Co-authoring of research papers, Q10 Key characteristics
Assessment based on Deliverables D1, D20-26.

Milestone 4 (months 18): First artefact/model as results of jointly executed research programme, e.g. multi-lingual medial dictionary, principles for a high-level ontology for the biomedical domain, architecture for the semantic-based electronic health record.

Quality indicators: Q8 Participation in standardisation work, Q9 Jointly executed research programme.
Assessment based on Deliverables D8, D9, D20-D26.

Milestone 5 (months 18/30): Long-term integration – evaluation of the NoE progress by Scientific Advisory Committee and Exploitation and Dissemination Committee.

Quality indicators: Q6 Mobility programs, Q9 Jointly executed research programmes, Q10 Key characteristics.
Assessment based on Deliverables: D1, D20-26.

7. Quality of Integration

7.1 Comments on Quality of Integration

“Integration” is an abstraction and as such more difficult to measure and evaluate than a concrete item such as the quality of a laboratory test where there exist precise standards. Quality is also an abstract concept, which relates to the difference between what is expected or stated as a goal and what is achieved. Thus, a general problem with measuring quality, that is the quantification of quality, is the quantification of objectives and standards.

The notion of information quality and quality of terminological systems are addressed in this application where aspects of the former include validity and reliability and aspects of the latter include content coverage, structure and usability. As often experienced in the domain of biomedicine and health care, we lack also in the domain of “integration in the sixth framework” a golden standard, which is why relative measures have to be used (as reliability is often used instead of validity in the absence of any gold standard). Integration is also a dynamic process, why it makes sense to monitor the progress of the NoE by periodic comparison of a series of indicators over the NoE lifecycle.

The overall objective of the NoE instrument is as described in the introduction and the description of integrating activities to re-shape the European Research Area for the benefit of co-operative research of relevance to the European health care sector and ICT industry at the forefront of international research and development. In the following section a list of indicators is presented which will be used for periodic monitoring of the NoE integration process. The indicators will be analysed annually by the Board, so that appropriate modifications of the integration strategies could be decided when necessary.

7.2 Indicators of Quality

Key indicators for follow-up of the progress of the NoE will be:

Q1 Workshops and symposiums

- ? Workshop/kick-off designed for initial familiarization
type, length and location of meetings: to be decided
subjects covered: all
number of participants: at least 1 and an average of 2 from each participating institution, making ca. 32 persons in all
Baseline: N/A
- ? Summer Schools, 1 per year; length 2 weeks; location 2004: Hungary
subjects covered: all
number of participants: at least 1 and an average of 2 from each participating institute, plus external participants, making ca. 50 persons in all
Baseline: N/A
- ? Special Interest Group (SIG)-type workshops, at least 3 per year; length 3 days; location: variable sites
subjects covered: each workshop should cover one or more work packages
- ? Participation in national and international conferences:
Goal: The NoE should be represented with several NoE member-institutions presenting as a unit via workshops, information meetings, and the like in at least 3 national and 1 international conference per year
- ? Board meetings: regular meetings and tele conferences for follow-up of the progress of activities and deliverables (configuration and responsibility of the Board is regulated in

Consortium Agreement).

Q2 Sharing of resources and use of research software tools

Goal: The goal is to develop a common database repository for at least three types of materials: 1) software resources, 2) experimental data, 3) terminological systems and thesauri, as well as associated technical and scientific information useful for the integration of activities in the NoE. At least two resources in each category should be identified during the first year.
Baseline: N/A

Q3 Logging of Website

Goal: The NoE should have its own Portal up-and-running by 31 January 2004. The portal should be updated at least weekly; each constituent institute should update its respective website at least monthly. The portal should include details of all events organized within the framework of the NoE, together with information about other relevant external events, especially within the European Scientific Area. The portal should include a list of FAQs, it should be monitored for statistics of web users by number of searches, user profiles (IP number, domain etc.), user actions (search, download etc.). Use of FAQs should be monitored, and the website updated and refined accordingly.
Baseline: N/A

Q4 PhD-study courses

We will identify all institutions/departments that offer a PhD within the topic areas of the network, noting that in the UK, there's no such thing as a PhD 'course', since students do no actual course-work but proceed directly to dissertation. Opportunities for joint courses will be identified. Given the different national regulations, we envisage that only a small number will actually be achievable. The Summer School in Month 8 will be a critical point bringing this program together.
Number of participants : to be identified later
Baseline: 0
Goal: 8

Q5 Co-tutoring of PhD-students

The Summer School in month 8 will provide many opportunities for co-tutoring PhD students from the institutions in the network, and we envisage that this will result in the identification of opportunities for formal joint supervision of PhD students.
Number of involved universities, departments, senior researchers and PhD students to be identified in WP6
Baseline: 0
Goal: At least 10 at the Summer School; 5 formal joint supervision of dissertations

Q6 Short-and medium-term visits of staff members

medium-term visits = 2 weeks or longer, for preparation of joint publications, student supervision, conference organization etc.
short-term visits = less than 2 weeks, for conferences and workshops, meetings with students
Baseline: assumed 0
Goal: all institutions should be involved; each institute should organize at least 2 and an average of 3 short-term and at least 1 and an average of 2 medium-term visits to other institutes within the NoE per year

Q7 Co-authoring of research papers, reports and educational materials

Participants in all institutions at senior researcher, post doc and PhD student levels should be involved with participants from other institutions in the preparation of co-authored scientific papers (original research paper, reviews, conference proceedings etc.) and reports.

Baseline: assumed 0

Goal: all research institutions should have at least 1 and an average of 2 co-authored scientific paper per year

Q8 Participation in standardisation activities

Participation of NoE members in CEN TC251, IMIA, ISO, W3C working group meetings on national and European level.

CEN/TC 251/WG I and II

Baseline: 5 NoE experts/3 times yearly

Goal: 8 experts/ 3 times yearly

IMIA

Baseline: IMIA WGs 2 experts yearly

Goal: Medinfo 2004 15 NoE attendees (incl. relevant IMIA WGs)

ISO/TC 215/WG 3 Health Informatics/Concept representation and ISO/TC 37 Terminology work

Baseline: TC 215 2 NoE experts 2 times yearly, TC 37 no NoE expert

Goal: TC 215 5 NoE experts 2 times yearly, TC 37 1 NoE expert once yearly

W3C

Baseline: No NoE experts

Goal: 2 NoE experts

Q9 Jointly executed research programmes

Goal: To qualitatively follow the progress of the jointly executed research programme in relation to objectives and time frame set forth. Assessment will be based on task descriptions and deliverables for each work package.

Baseline: N/A

Q10 Key characteristics of NoE partners

Goal: To quantitatively and qualitatively follow the development of each partner (through indicators budget and employees, research funding, production of dissertations, research papers and patents, number of students) under the hypothesis that participation in the NoE will positively strengthen the partners as centres of excellence.

Baseline: N/A

8. Project Organisation, Management and Governance Structure

8.1 Governing bodies, roles and responsibilities

The statutory institutions of the NoE are the Assembly and the Board. The Assembly approves the Committees necessary for the fulfilment of the activities within the NoE. The number and assigned tasks of approved Committees will change during the lifecycle of the NoE, thereby facilitating a dynamic and flexible organisation of the NoE. The Board is an operative body, working in close co-operation with the Management Office and with close consultations with the Committee Co-ordinators.

8.1.1 The co-ordination of the NoE

The Co-ordinator (CO)

In addition to the CO's functions pursuant to the EU Contract, the CO will primarily have the function of serving as chair of the Assembly and the Board, including administration and preparation of minutes and transmission of any documents and information connected with NoE activities.

The Management Office

An authorisation for the submission of a NoE proposal under the sixth Framework Programme and a mandate for co-ordination have been signed by the NoE partners. The latter commits partners to signing a Consortium Agreement as a prerequisite for signing an NoE-contract with the European Commission. The NoE co-ordinator institution will be the Department of Biomedical Engineering/Medical Informatics of Linköping University, headed by Dr Hans Åhlfeldt. A Management Office within Linköping University will be set up with the following functions:

- ? Coordinator's Office, with day-to-day administrative support.
Key personnel: Dr Hans Åhlfeldt (co-ordinator), Hans Gill (local project officer), Jörgen Jonson (financial support)
- ? International Office with experience of and responsibilities for international agreements and exchange programs.
Key personnel: Ann-Christine Comstock, International Office Director, Johan Åkerman, EU-Financed Research
- ? Legal Advisor with experience of and responsibilities for legal affairs relating to the Consortium Agreement and the EU-contract.
Key personnel: Göran Hessling, Chief Legal Advisor
- ? Management advisory group from the Department of Management and Economics, with professional experience of management and business administration.
Key personnel: Magnus Holmström, Head of Department

The Management Office will be responsible for:

- ? Drafting and finalising the Consortium Agreement and signing of the NoE Contract.
- ? Awarding financial assistance under the terms of the Consortium Agreement
- ? Administration of changes in NoE partner list in accordance with the rules set forth in the Consortium Agreement
- ? Liaison between the NoE and the Commission
- ? Overall monitoring of NoE activities and execution of deliverables and milestones

The Committee Coordinator (CCO)

The CCO has the function as the chair for a Committee, including administration and preparation of minutes, transmission of any documents and information connected with the activities in question

between the parties concerned or the CO and transmission of the Deliverables of the parties engaged in the Committee activities.

Work package Leader

Each Work package defined in the Joint Programme of Activities will be led by a Work package Leader. A Work package might be a joint adventure of several parties whose activities are co-ordinated by the Work package Leader. The latter is also responsible for close consultations with the other partners and with the Board, and for ensuring co-ordination with related Work packages.

8.1.2 The Assembly

The Assembly is the principal institution of the NoE and is composed of one duly authorised representative of each partner. The Assembly will meet at frequency defined in a Consortium Agreement. Extraordinary meetings can be called at the request of the Chair or at any other time when necessary at the request by one of the parties. The aim is to have a decision making body which is sufficiently flexible and efficient to promote the development of the NoE.

The Assembly shall be responsible for the overall direction of the NoE. To that end, it shall have the powers to decide upon the strategy of the NoE and the allocation of its budget, making proposals to the parties, deciding upon press releases and joint publications as well as establishing roadmaps with regard to the activities within the NoE, deciding upon the establishment of specific Committees and upon measures in the framework of audit procedures to ensure the effective co-ordination and monitoring of the progress of NoE activities. To form a quorum the Assembly must have a majority of two-thirds (2/3) of its parties represented, directly or by proxy. Where decisions are to be taken unanimously, all parties must be represented at the meeting.

8.1.3 The Board

The Board shall be responsible to the Assembly for setting, approving, directing and monitoring the overall strategic direction of the NoE, in accordance with the decisions made by the Assembly. The Board is responsible for all pro-active measures concerning defaulting parties and for management of the NoE within the framework set up by the Assembly in accordance with the EU Contract and the Consortium Agreement. The Board shall constantly seek to ensure close consultation with all partner as well with as with the established committees.

The Board is chaired by the CO and shall consist in addition of a maximum of six representatives of the partners, elected by the Assembly. The Board must have a majority of two-thirds (2/3) of its members present to form a quorum. Where decisions are to be taken unanimously, all members must be represented at the meeting. A member of the Board cannot be represented by a proxy.

8.1.4 Committees

The NoE needs to be organised in a flexible and dynamic way. The assignment of Committees for specified tasks during the different phases of the NoE lifecycle will ensure that the management of the NoE can be responsive to salient developments within and outside the NoE. Committees already foreseen are listed below.

The Network Integration Committee

The responsibility of this "think-tank" committee will be strategic planning, with the objective of establishing a self sustainable network after the EU financing period. The key deliverable will be letters of intent and commitment from the partners within the network. The constitution of the network in terms of numbers and types of partners will be analysed, as well the progress of the NoE as monitored by the list of indicators identified in section B6.

The Scientific Advisory Committee

The Scientific Advisory Committee is constituted from scientists from the academic, industrial and public health world, both inside and outside the NoE and inside and outside the EU. At least half of the members should come from the NoE, and at least one third should come from outside. The main responsibility of the Scientific Advisory Board will be to provide an annual evaluation of ongoing network activities, and the recommendation to the Assembly on re-direction or re-shaping of research activities.

Exploitation and Dissemination Committee

An Exploitation Committee will be appointed with the responsibilities of working on public relations, dissemination and exploitation of results.

8.2 The Consortium Agreement

Before the EU-NoE Contract is signed, a Consortium Agreement will be finalised and signed by the NoE partners. The process of establishing the Consortium Agreement will be the responsibility of the Management Office. The Consortium Agreement will specify the general conditions of the NoE set forth by the Commission, responding to the items listed in Annex II and III of the Model Contract (Decision C (2003), 799 dated 17.03.03), including implementation and deliverables, financial aspects, and intellectual property rights.

The financial model of the Consortium Agreement will include the following principal components:

- ? Management cost for the Management Office up to 7% of the total budget
- ? Base funding to each partner for consumables and administration (approximately 2%)
- ? Cost for travels and exchange programs for each partner (approximately 15%)
- ? Direct cost for activities performed within Work Packages (approximately 76%)

9. Joint Programme of Activities first 18 months

9.1 JPA Plan Introduction

The Joint Programme of Activities (JPA) is built up of the following principal components:

- ? Management activities including monitoring of the overall integration process, evaluation of progress against milestones and objectives, and long term strategic planning (WP1-2).
- ? Integrating activities including infrastructure for communication, and mobility programs (WP3-7).
- ? Activities to spread excellence. Internal knowledge transfer will be supported through mobility programs among researchers and PhD-student, and workshops and tutorials where front line material in the subject areas of interest will be presented and worked upon (WP10-16).
Dissemination and exploitation of results will be the responsibility of two specific work packages (WP8-9) including contributions to standardisation efforts and to the formulation of health care policies, e.g. CEN TC251, IMIA, ISO, W3C. Dissemination of results to the scientific community and public domain through research papers, conference proceedings, educational material will be an inherent part of these activities.
- ? Common research activities in identified areas of interest (WP20-26).

9.1.1 Risk Analysis and Contingency Plan

The network aims to bring together different academic communities – ontology engineering, language engineering, data mining. There is a risk that understanding each other's domains will take longer than the term of the project, particularly as all the fields are themselves rapidly evolving. In such a cross-disciplinary context, joint projects organized around distributed task forces are the best guarantee for long-term management of the network. Therefore it is to be noted that in each of the core domains identified in the network, leading groups are always back up by at least one group, as can be seen from the work package description.

Mutual understanding of the network members depends on continued participation of at least a core of people from each discipline over a sustained period. Although the network might have a dynamic membership, the board will aim at insuring representativeness of scientific content of the project, together with controlling that efforts needed to educate those who join the network from time to time may distract from the goal of achieving deeper mutual understanding of the core members.

The network does not provide sufficient funding to secure full employment of the core domain experts. Their continued ability to participate is therefore contingent on their ability to acquire local grant income. The follow-up of key characteristics for partners such as research funding, and changes in budget and employees will therefore be performed and acted upon if necessary.

Intellectual Property Rights - the project proposes shared lexical and ontological corpora. A licensing model needs to be devised that allows and encourages this and, more importantly, is still able to interoperate with the existing (often proprietary) resources and tools.

Sensitive data issues. - for the sake of cooperation and cross-fertilization, samples of data are to be shared among partners in the network. Among such shared data, items of the patient records are highly sensitive; therefore each partner will rely on his own national/local regulations and legal/ethical bodies to define pseudonymization methods and guidelines (subsidiary principle). Targeted bilateral agreements will be investigated when necessary to facilitate activities in a given subproject.



The objective of the decision-making mechanisms regulated in the Consortium Agreement is to facilitate the integration process under an atmosphere of trust and transparency. The decision making mechanism need to be both sufficiently flexible and efficient to promote the integration process and to ensure work of highest scientific standards. Although based on the unified UNECA model, the realisation of the managerial structures and decision-making mechanisms set forth by the Consortium Agreement is a critical matter. One of the management activities will therefore be to share experiences and on critical matters seek for advice from the EC project officers, other NoEs and other expertise on management and business administration.



9.2 Planning and timetable

WP no	Month																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	D1.4,5					D1.1,4		D1.4			D1.2,3			D1.4		D1.4		
2							D2											
3						D3.1					D3.2							
4						D4.1										D4.2		
5						D5.1					D5.2							
6			D6.1								D6.2,3							
7				D7.1					D7.2									
8						D8.1												D8.2
9									D9.1								D9.2,3	
10						D10												
11						D11												
12						D12												
13							D13											
14										D14								
15										D15								
16								D16										
20											D20.1					D20.2		
21											D21.1					D21.2		
22											D22.1					D22.2		
23											D23.1					D23.2		
24											D24.1					D24.2		
25											D25.1					D25.2		
26											D26.1					D26.2		



9.3 Involvement of partners in workpackages

WP No	Title	Board	Committee	LIU IMT	LIU IDA	KI	SU/UG OT	UKLFR/ UNIFR	IFOMIS	CAU	DIM	UOM	UCL	ITRI	ACP-HEGP	CNR-ISTC	EBI	MEDINF O	NORDC LASS	LIU C-NPU	MRI	EDSA	
1	Project management	Red	Red																				
2	Strategic planning	Red	Red																				
3	Set-up of common database	Red																					Yellow
4	Set-up of public website	Red																				Yellow	Red
5	Planning of network meetings	Red																					
6	Mobility program	Blue												Red									
7	Summer school 2004	Blue																Red					
8	Participation in standardisation	Blue				Red	Yellow						Yellow										
9	Dissemination and exploitation	Blue	Red																			Red	
10	Workshop on health care statistics	Blue		Red														Yellow	Yellow				
11	Workshop on semantic web	Blue			Red			Yellow		Yellow		Yellow											
12	Workshop on ontology engineering	Blue			Yellow				Red	Yellow		Yellow				Yellow							
13	Workshop on NLP	Blue			Yellow						Red			Yellow									
14	Workshop on datamining in bioinformatics	Blue									Yellow							Red					
15	Workshop on text mining and IR	Blue					Yellow	Red			Yellow			Yellow	Yellow								
16	Workshop on EHR	Blue									Yellow	Yellow	Red		Yellow								
20	Research multilingual med dictionary	Blue		Yellow	Blue	Blue	Yellow	Red			Yellow			Yellow		Blue				Blue			
21	Research ontology engineering	Blue		Blue	Blue	Blue	Blue	Yellow	Blue		Blue	Red		Blue	Blue	Yellow							
22	Research SNOMED CT	Blue		Yellow	Blue	Red	Blue				Blue	Yellow	Blue		Blue							Blue	
23	Research health statistics	Blue		Red	Blue										Blue								
24	Research data mining and IR	Blue		Blue	Blue			Yellow			Yellow	Yellow											
25	Research concept systems in lab medicine	Blue				Blue	Blue										Blue						
26	Research EHR	Blue		Blue	Blue	Blue		Blue	Blue		Blue	Yellow	Red		Yellow	Yellow					Red		
	Lead contractor	Red																					
	Assisting partner	Yellow																					
	Partner	Blue																					
	Follow up, assessment	Grey																					
	Integration activities	Yellow																					
	Dissemination, exploitation	Purple																					
	Knowledge transfer	Green																					
	Joint research areas	Blue																					

9.4 Work Package List (18 months)

Work-package No	Work package title	Lead contractor	Start month	End month	Deliverable
1	Project management	The Board	1	18 (36)	D1
2	Assessment and strategic planning	The Board	7	18 (36)	D2
3	Set-up of a common database	DIM	2	8 (18)	D3
4	Set-up of public website	EDSA	2	8 (18)	D4
5	Planning of network meetings	The Board	1	6	D5
6	Mobility programs	ITRI	3	12	D6
7	Summer School 2004	MEDINFO	4	11	D7
8	Participation in standardisation work	KI	2	18	D8
9	Dissemination and exploitation	MRI	9	18	D9
10	Workshop/tutorial on Health Care Statistics	LiU IMT	5	9	D10
11	Workshop/tutorial on the Semantic Web	LiU IDA	5	9	D11
12	Workshop/tutorial on Ontology Engineering	IFOMIS	5	9	D12
13	Workshop/tutorial on Natural Language Processing in the Biomedical domain	DIM	7	11	D13
14	Workshop/tutorial on Data Mining in Bioinformatics Databases	EBI	10	14	D14
15	Workshop/tutorial on Text Mining and Information Retrieval	UKLFR	10	14	D15
16	Workshop/tutorial on the Electronic Health Record	UCL	7	11	D16
20	Research Activity Multilingual Medical Dictionary	UKLFR	6	36	D20
21	Research Activity Ontology Engineering	UOM	6	36	D21
22	Research Activity SNOMED CT	KI	6	36	D22
23	Research Activity Health Statistics and Semantic Distance	LiU IMT	6	36	D23
24	Research Activity Data Mining and Information Retrieval	EBI	6	36	D24
25	Research Activity Concept Systems for Laboratory Medicine	LiU c-NPU	6	36	D25
26	Research Activity Electronic Health Record	UCL	6	36	D26
	TOTAL				

9.5 Deliverables list (18 months)

Del.no	Del.code	Deliverable title	Delivery date	Nature ¹	Dissemination level ²
1	D1.1	Detailed analysis of research activities)	6	R	RE
2	D1.2	Periodic management report	12	R	RE
3	D1.3	Periodic financial report	12	R	RE
4	D1.4	Quarterly reports	3,6,9,15,18	R	RE
5	D1.5	Project Presentation	3	R	RE
6	D2	Strategic planning	12	R	RE
7	D3.1	Typology of shared resources	6	R	RE
8	D3.2	Report on common database	11	R	RE
9	D4.1	Report on public website	11	R	RE
10	D4.2	Report on public website	17	R	PU
11	D5.1	Network meetings	6	R	RE
12	D5.2	Network meetings	11	R	RE
13	D6.1	Report on doctoral dissertations	6	R	PU
14	D6.2	Report on PhD programmes	11	R	RE
15	D6.3	Report on mobility program	11	R	RE
16	D7.1	Planning summer school	8	R	RE
17	D7.2	Evaluation summer school	11	R	RE
18	D8.1	Participation in standardisation work	6	R	RE
19	D8.2	Participation in standardisation work	17	R	RE
20	D9.1	Format and layout for educational material	11	R	RE
21	D9.2	Report on educational material	17	R	PU
22	D9.3	Report on dissemination	17	R	RE
23	D10	Workshop on health statistics	9	R	RE
24	D11	Workshop on the semantic web	9	R	RE
25	D12	Workshop on ontology engineering	9	R	RE
26	D13	Workshop on NLP	11	R	RE
27	D14	Workshop on data mining	14	R	RE
28	D15	Workshop on information retrieval	14	R	RE
29	D16	Workshop on the EHCR	11	R	RE
30	D20.1	Report on multi-lingual medical dictionary	11	R	RE
31	D20.2	Prototype multi-lingual medical dictionary	17	P	RE

¹ Please indicate the nature of a deliverable using one of the following codes:

- R** = Report
- P** = Prototype
- D** = Demonstrator
- O** = Other

If milestone, indicate with **M**

² Please indicate the dissemination level for deliverables using one of the following codes:

- PU** = Public
- PP** = Restricted to other programme participants (including the Commission Services).
- RE** = Restricted to a group specified by the consortium (including the Commission Services).
- CO** = Confidential, only for members of the consortium (including the Commission Services).



32	D21.1	Report on ontology engineering	11	R	RE
33	D21.2	Report on ontology engineering	17	R	RE
34	D22.1	Report on SNOMED CT – tools and systems	11	R	RE
35	D22.2	Report on SNOMED CT - evaluation	17	R	RE
36	D23.1	Report on health statistics	11	R	RE
37	D23.2	Report on information quality in health registries	17	R	RE
38	D24.1	Report on data mining and IR	11	R	RE
39	D24.2	Report on data mining and IR	17	R	RE
40	D25.1	Report on concept systems for lab.medicine	11	R	RE
41	D25.2	Report on concept systems for lab.medicine	17	R	RE
42	D26.1	Report on EHCR	11	R	RE
43	D26.2	Architecture for semantic-based EHCR	17	R	RE



9.6 Work Package Description

9.6.1 Project Management and Integration

<i>Work Package 1 (WP1): Project Management</i>
Responsible: Management Office, The Board (LIU-IMT, KI, UKLFR, DIM, UOM)
Person months: 8 (1.5, 1.5, 1.5, 1.5, 1.5)
Participants: All partners
Start: Month 1
Duration: 18 months
<p>Objective and Description:</p> <p>The objective is to oversee the progress of the NoE activities through follow up of work package activities and deliverables. Regular communication with work package leaders will be established for follow up of activities according to plan.</p> <p>Periodic activity reports will be produced containing an overview of activities carried out, a description of progress towards objectives and milestones, identification of problems encountered and corrective actions taken.</p> <p>Periodic management reports will be produced including cost statements with justification of resources deployed by partners, and on an annual basis compilation of audit certificates.</p> <p>The objective is further to extend the preliminary analysis of excellence and ongoing research activities within the Network which was carried out already during the pre-proposal phase. Activity areas already identified and formulated as Work Packages will be further developed in terms of a coordination program and a mobility plan. Further areas suited for joint research activities will be identified and analysed. An extended inventory of the competences of each partner and the complementary of these competencies will be made and summed up in a database which will be available via the Internet.</p>
<p>Task:</p> <ol style="list-style-type: none">1.1 Follow up of progress of activities and deliverables1.2 Compilation of periodic reports1.3 Identification of further joint research activities1.4 Identified gaps of competencies1.5 Identification of information suitable for a common database
<p>Deliverable:</p> <ol style="list-style-type: none">D1.1 Report Detailed Analysis of Research Activities m6D1.2 Periodic management report including activity reportsD1.3 Periodic Financial reportsD1.4 Quarterly reportsD1.5 Project Presentation
<p>Interaction with other Work Packages:</p> <p>WP3 Common Database</p> <p>WP4 Public Website</p> <p>WP6 Mobility Program</p>



<i>Work Package 2 (WP2) Assessment and Strategic Planning</i>
Responsible: The Board, Network Integration Committee (LIU-IMT , KI, UKLFR, DIM, UOM)
Participants: All partners
Person months: 8 (1.5, 1.5, 1.5, 1.5, 1.5)
Start: Month 6
Duration: Twelve months
Objective and Description: The objective is monitoring and evaluation of the NoE integration progress in relation to established objectives, milestones and goals for quality indicators. Different management structures will be analysed as well as formal and informal ways of co-operation, together with strengths, weaknesses, opportunities and threats (SWAT-analysis) of the NoE. Legacy constrains need to be analysed and worked on. The evolvement of the network in terms of size and composition of partners and funding will be analysed. The list of quality indicators will be used as input for the analysis. This work activity will be carried out in co-operation with the Scientific Advisory Committee and the Dissemination Committee. The long term objective is to assure a self sustainable network after the E.U. financing period.
Task: 2.1 Analysis of management structures, formal and informal ways of co-operation. 2.2 Analysis of quality indicators 2.3 SWAT-analysis 2.4 Analysis of legacy matters 2.5 Analysis of funding situation
Deliverable: D2 Report Assessment and Strategic Planning m12

<i>Work Package 3 (WP3): Set-up of a Common Database</i>
Responsible: The Board (LIU-IMT , KI, UKLFR, DIM, UOM), EDSA
Person months: 7 (1.1, 1.1, 1.1, 1.1, 1.1, 1.9)
Participants: All partners
Start: Month 2
Duration: 5 months (first version, v.1), thereafter continuous upgrading
Objective and Description: The objective is to initiate the task of building a common database repository for at least three types of materials: 1) software resources, 2) experimental data, 3) terminological systems and thesauri, as well as associated technical and scientific information useful for the integration of activities in the NoE. Due to the variety in different data/software types, this task will begin with the attempt to find solutions to problems standing in the way of the necessary data standardisation. In a preliminary step, we will start sharing samples of data (corpora, databases, lexicons.etc.) from all partners in order to identify regular patterns and possible normalization strategies. For the first year, we will concentrate on identifying a set of software to be shared within the network and plan to collect textual resources from clinical records: thus, copyrights and pseudonymization procedures will have to be solved as sine qua non condition.
Task: 3.1 Analysis of terminological resources and corpora 3.2 Identification and description of sharable data 3.3 Identification and description of sharable software 3.4 Design specification of database including web interface 3.5 Database implementation 3.6 Rules for access and updating 3.7 Organisation for maintenance



Deliverable: D3.1 Report on Typology of shared resources m6 D3.2 Report on Common database m11
Interaction with other Work Packages: WP1 Project Management WP4 Public Website

<i>Work Package 4 (WP4): Set-up of a Public Website</i>
Responsible: EDSA, MRI, The Board (LIU-IMT, KI, UKLFR, DIM, UOM)
Person months: 11 (7.5, 1.9, 0.4, 0.4, 0.4, 0.4, 0.4)
Participants: All partners
Start: Month 2
Duration: 5 months (first version, v.1), thereafter continuous upgrading
Objective and Description: The objective of WP3 is to design and implement a public Website where information is made available about the network as a whole, together with competence profiles for each partner and other relevant public information about the progress of NoE activities. The use of Internet-based technology for supporting and maintaining communication within the NoE through various media such as mailing lists, meeting agenda, conferences, virtual meetings, and newsletters. This task will be an integral part of the management of the NoE. Usual web facilities, with separate public and restricted areas (depending on a set of user profiles) will be set up for both internal management purposes and dissemination of activities outside the network.
Task: 4.1 Design specification 4.2 Realisation of website functionality 4.3 Rules for information structure and information handling 4.4 Organisation for updating and maintenance
Deliverable: D4.1 Report Public Website m11 D4.2 Report Public Website m17
Interaction with other Work Packages: WP1 Project Management WP3 Common Database



<i>Work Package 5 (WP5): Planning of Network Meetings</i>
Responsible: The Board (LIU-IMT , KI, UKLFR, DIM, UOM)
Participants: Input from all partners
Person months: 4 (0.7, 0.7, 0.7, 0.7, 0.7)
Start: Month 1
Duration: Four months, thereafter annual scheduling
<p>Objective and Description:</p> <p>The objective is to identify, schedule and assign planning responsibility for a series of network meetings. The general objective of network meetings is to speed up the integration process and to foster co-operation and exchange of know-how and research infrastructure within the network. Network meetings will be of different kinds: kick-off meeting, workshops, summer school, and joint national and international conferences and seminars.</p> <p>By co-ordination of NoE meetings and travel to external conferences and symposia, resource consumption might be minimised and time of Network participants more rationally allocated. Through co-ordinated participation of NoE members in conferences, the visibility of the NoE will be enhanced, and cutting edge research of interest to the NoE framework can be identified and brought to the awareness of the partners.</p> <p>The objective is also, as an extension to the already identified workshops, to further plan workshops and tutorials for spreading of excellence within the NoE.</p>
<p>Preliminary material:</p> <p>Kick-off meeting January 2004</p> <p>Regular national and international conferences</p> <p>CEN TC251 working group meetings</p> <p>IMIA working group meetings</p> <p>Summer School 2004</p> <p>AMIA Fall Symposium, Washington, November 8-12, 2003</p> <p>Swedish/Nordic Annual Terminology Conference, Stockholm, November 11-12, 2003.</p> <p>KR 2004, 9th International Conference on Principles of Knowledge Representation and Reasoning, Whistler, Vancouver, British Columbia, Canada, June 2-5, 2004</p> <p>ECAI 2004, 16th European Conference on Artificial Intelligence, Valencia, Spain, August 22-27, 2004</p> <p>IJCAI-05, Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 31-August 5, 2005 (IJCAI-03 Acapulco, Mexico, August 9-15, 2003, no conference 2004)</p> <p>AAAI, Nineteenth National Conference on Artificial Intelligence (AAAI-04) and the Sixteenth Innovative Applications of Artificial Intelligence Conference (IAAI-04) will be held July 25-29, 2004 in San Jose, California</p> <p>ISWC2003, 2nd International Semantic Web Conference, October 20-23, 2003, Sanibel Island, Florida, USA</p> <p>MEDINFO 2004, Hilton San Francisco, San Francisco, California, United States, September 7-11, 2004</p> <p>COLING International Conference on Computational Linguistics</p> <p>ACL Conference of the Association for Computational Linguistics</p> <p>EACL Conference of the European Chapter of the Association for Computational Linguistics</p> <p>LREC International Conference on Language Resources and Evaluation</p>
<p>Task:</p> <p>5.1 Scheduling of network meeting first 18 months</p> <p>5.2 Planning kick-off meeting January 2004</p> <p>5.3 Planning of joint NoE/national/international meetings/conferences</p> <p>5.4 Planning of additional workshops and tutorials</p>



Deliverable: D5.1 Report Network Meetings m6 D5.2 Report Network Meetings m11
Interaction with other Work Packages: WP1 Project Management WP4 Public Website

<i>Work Package 6 (WP6): Mobility programs</i>
Responsible: ITRI , The Board (LIU-IMT, KI, UKLFR, DIM, UOM)
Participants: All university partners
Person months: 7 (5.6, 0.2, 0.2, 0.2, 0.2, 0.2)
Start: Month 3
Duration: Nine months
Objective and Description: The objective is to compile descriptions of recent PhD dissertations and on-going PhD thesis work for presentation to the NoE. Standards and procedures with respect to the production of PhD dissertations will be summarised and compared. The compilation will be used as a basis for a possible process towards harmonisation of procedures and for sharing of senior research staff as co-tutors in doctoral programs (although formally handled by the different universities). The objective is further to identify opportunities for exchange of researchers and PhD students between partners in the NoE, and to plan and organise these as short-, medium- and long-term visits.
Task: 6.1 Compilation of recent and ongoing doctoral programs within the NoE 6.2 Compilation of PhD-course material within the NoE 6.3 Description of standards, procedures and quality control with respect to PhD dissertations 6.4 Proposal for co-tutoring of doctoral students and harmonisation of procedures 6.5 Proposal for researcher and student mobility program
Deliverable: D6.1 Report on recent and ongoing doctoral dissertations m6 D6.2 Report on PhD programmes, standards and procedures m11 D6.3 Report mobility program m11
Interaction with other Work Packages: WP4 Public Website WP7 Summer School



<i>Work Package 7 (WP7): Summer School 2004</i>
Responsible: MEDINFO , The Board (LIU-IMT, KI, UKLFR, DIM, UOM)
Participants: All partners
Person months: 5 (3.7, 0.2, 0.2, 0.2, 0.2, 0.2)
Start: Month 4
Duration: Five months
Objective and Description: The objective is to plan and organise a summer school 2004 where researchers and doctoral students in the network, maybe including US guests, gather in one place for an extended period (e.g. a full week) for an informal program of workshops, tutorials, demonstrations and social activities. Preliminary location is Budapest or a resort close to lake Balaton in Hungary.
Task: 7.1 Planning of content, activities, duration, localisation 7.2 Organisation and realisation
Deliverable: D7.1 Summer School 2004 m8 D7.2 Evaluation of Summer School 2004 m11
Interaction with other Work Packages: WP4 Public Website WP5 Network Meetings



9.6.3 Knowledge Transfer and Dissemination

<i>Work Package 8 (WP8): Participation in Standardisation Work</i>
Responsible: KI , SU, UCL, The Board (LIU-IMT, KI, UKLFR, DIM, UOM)
Participants: All partners
Person months: 10 (5.8, 1.9, 1.9, 0.2, -, 0.2, 0.2, 0.2, 0.2)
Start: Month 2
Duration: 17 months
Objective and Description: The objective is to facilitate information exchange and knowledge transfer between the NoE partners and the working groups of the European standardisation organisation (CEN) and of the International standardisation organisation (ISO). The network will establish close links to those working on the standardisation efforts in health informatics organised by CEN/ TC 251 and its four working groups (I) Information Models, (II) Terminology and Knowledge Bases, (III) Security, Safety and Quality, and (IV) Technology for Interoperability. Work in the two first working groups of TC251 is at the core of the proposed network activities. Participation in international working groups such as ISO/TC 215/WG 3 will be encouraged as well as the LOINC and cNPU groups. The NoE shall also maintain contacts with other bodies such as the IMIA working groups on Intelligent Data Analysis and Data Mining (03), Medical Concept Representation (06), Standards in Health Care Informatics (16), Computerised Patient Records (17), and the W3C.
Task: 8.1 Participation in CEN TC251 working groups, primarily WG 1,2. 8.2 Participation in IMIA working groups, primarily WG 3, 6, 16, 17 8.3 Participation in W3C
Deliverable: D8.1 Report Participation in Standardisation work m6 D8.2 Report Participation in Standardisation work m17
Interaction with other Work Packages: WP1 Project Management WP4 Public Website WP5 Network Meetings



<i>Work Package 9 (WP9): Dissemination and Exploitation</i>
Responsible: MRI , Dissemination Committee (LIU-IMT, KI, UKLFR, DIM, UOM)
Participants: All partners
Person months: 9 (3.7, 1.0, 1.0, 1.0, 1.0, 1.0)
Start: Month 9
Duration: Ten months
Objective and Description: The objective is to bring together results from the research projects for analysis of exploitation possibilities. The objective of this activity is further to compile and refine the study material produced during the realisation of workshops and tutorials, and to identify suitable ways of dissemination and publishing of this material.
Task: 9.1 Analysis of exploitation possibilities of research results 9.2 Proposal for structure and format for study material produced by tutorials and workshops 9.3 Compilation of study material and publication on NoE website 9.4 Analysis of potential target groups and media for further publication
Deliverable: D9.1 Report Format and Layout for Study Material m11 D9.2 Report Educational Material m17 D9.3 Report Dissemination and Exploitation m17
Interaction with other Work Packages: WP1 Project Management WP4 Public Website WP10-WP16 Workshops and Tutorials



<i>Work Package 10 (WP10): Workshop/tutorial on Health Care Statistics</i>
Responsible: LiU (IMT) , MEDINFO, NORDCLASS, SOS, STAKES, KITH, NBH
Participants: All partners
Person months: 9 (5.6, 1.9, 0.5, 0.5, 0.5, 0.2, 0.2)
Start: Month 5
Duration: Five months
Objective and Description: The objective is to plan and organise a workshop on morbidity and mortality statistics, quality registries and the use of WHO classifications for comparative health statistics. While several organisations within Europe and throughout the world collect statistical data on health and national health care systems, the general problem is the low reliability of data, conflicting interpretations and definitions and poor exploitation of the information. The overall objective of this activity is to contribute to the improvement of European Health Statistics by dissemination of related knowledge and by achieving consensus in areas where conflicting practices theories and interpretations occur. The topic of interest covers the following fields: Ontology of health indicators Presentation and representation of data Tools for navigation among data in a multidimensional space Quality of data, reliability, accuracy, usefulness Theory, practice history and future of coding systems. Possible event in conjunction with Summer School 2004.
Task: 10.1 Planning of workshop/tutorial 10.2 Realisation of workshop/tutorial 10.3 Study material
Deliverable: D10 Report WP10 m9
Interaction with other Work Packages: WP1 Project Management WP3 Public Website WP4 Network Meetings WP26 Research Activity Health Statistics and Semantic Distance



<i>Work Package 11 (WP11): Workshop/tutorial on the Semantic Web</i>
Responsible: LiU (IDA) , UKLFR, UNIFR, CAU, UOM, STAKES, KITH, NBH
Participants: All partners
Person months: 13 (5.6, 1.0, 0.9, 1.9, 1.9, 0.5, 1.0, 0.4)
Start: Month 5
Duration: Five months
Objective and Description: The objective is to plan and organise a workshop on material relating to the Semantic Web. The Semantic Web initiative offers a useful framework for adding semantics to web documents by using metadata based on common standards, such as XML for document markup and DAML+OIL for expressing the underlying knowledge. Different ways to generate semantic web compatible documents will be presented, e.g. the knowledge-based authoring of documents, semantic tagging of unstructured text etc. Possible event in conjunction with Summer School 2004.
Task: 11.1 Organisation of a workshop/tutorial 11.2 Realisation of workshop/tutorial 11.3 Study material
Deliverable: D11 Report WP11 m9
Interaction with other Work Packages: WP1 Project Management WP3 Public Website WP4 Network Meetings WP15 Workshop/tutorial on Text Mining and Information Retrieval



<i>Work Package 12 (WP12): Workshop/tutorial on Ontology Engineering</i>
Responsible: IFOMIS , CNR-ISTC, UOM, CAU, LIU(IDA)
Participants: All partners
Person months: 13 (5.6, 1.9, 1.9, 1.9, 1.9)
Start: Month 5
Duration: Five months
Objective and Description: The objective is to plan and organise a workshop on ontology engineering. Two schools of thought are gradually beginning to crystallize in the domain of information systems ontology. On the one hand is the school which focuses primarily on the representational adequacy of an underlying ontological theory, leaving for others the task of transforming this theory into working applications. On the other hand is the (much larger) school which focuses primarily on the construction of ontologies as working applications, at the expense of representational adequacy. Both schools are represented in the NoE, providing an opportunity for fruitful debate, competition and co-operation. The presentation of specific ontologies and reference terminologies will include OpenGalen, MedO, Gene Ontology, UMLS and SNOMED CT. Tools for authoring and managing ontologies as well as common terminology services will be covered. Possible event in conjunction with Summer School 2004.
Task: 12.1 Organisation of a workshop/tutorial. 12.2 Realisation of workshop/tutorial 12.3 Study material
Deliverable: D12 Report WP12 m9
Interaction with other Work Packages: WP1 Project management WP3 Public Website WP4 Network Meetings WP21 Research Ontology Engineering



<i>Work Package 13 (WP13): Workshop/tutorial on Natural Language Processing in the Biomedical domain</i>
Responsible: DIM , ITRI, LiU (IDA)
Participants: All partners
Person months: 9 (5.6, 1.9, 1.9)
Start: Month 7
Duration: Five months
Objective and Description: The objective is to plan and organise a workshop on natural language processing. The development and adaptation of basic NLP tools (word segmenter, tagger, chunker, parser) to the medical sub-language serves as a foundation for higher-level semantic and knowledge-base understanding of text. This understanding, qualitatively improves the task of information extraction and information retrieval. In 2002, the first workshop on Natural Language Processing in Biomedical Application was held in Cyprus as an EFMI satellite workshop (http://www.genisis.ch/~natlang/NLPBA02/). This first event gathered about 50 researchers and featured 14 international scientific presentations (11 from all over Europe, 3 from North-America, and 1 from Japan). A selected set of these papers was later published in a special issue of the International Journal of Medical Informatics (Eslevier). Since this date, different forum (ACL, ACM-SIGIR) have organised similar satellites workshops. The objective of this NoE workpackage is to co-ordinate these events in order to organise a workshop on a more regular bases. Ideally, the workshop could be associated with an evaluation forum on a specific task to be defined, in the spirit of the CoNLL (http://ilk.uvt.nl/~signll/shared.html) shared tasks. Co-operation with WP 24 is foreseen. Possible event in conjunction with MEDINFO 2004.
Task: 13.1 Organisation of a workshop/tutorial 13.2 Realisation of workshop/tutorial. 13.3 Study material
Deliverable: D13 Report WP13 m11
Interaction with other Work Packages: WP1 Project Management WP3 Public Website WP4 Network Meetings WP15 Workshop/tutorial on Text Mining and Information Retrieval



<i>Work Package 14 (WP14): Workshop/tutorial on Data Mining in Bioinformatics Databases</i>
Responsible: EBI , DIM
Participants: All partners
Person months: 7 (5.6, 1.9)
Start: Month 10
Duration: Five months
Objective and Description: <p>The EBI is Europe's major molecular biology database producer, bioinformatics service provider, and bioinformatics research centre. The major focus is on development of databases of nucleotide and protein sequences, protein families, protein structures, microarrays and genome annotation. A network of worldwide collaborators are engaged in these projects, and many other specialised databases (protein interactions, enzymes, immunogenetics, mutations) are actively pursued. More recent exploitation of text resources and ontology development has triggered a variety of biomedical text-mining activities relevant to this network of excellence.</p> <p>The Computational Genomics Group at EBI are currently developing research in the fields of genome sequence annotation, classification of protein function, protein sequence motif discovery, data mining, ontologies for molecular biology, metabolic pathways, deep phylogeny, knowledge representation in molecular biology databases, pattern discovery in sequences and transcription.</p> <p>The objective of this activity is to share in-depth knowledge of the content and structure of EBI databases, as well as knowledge of useful methods and tools for data mining and information retrieval. Although well-known in the bioinformatics community, this NoE activity will promote the establishment of new collaboration between research groups of medical informatics and bioinformatics and thereby the authoring of new research applications targeted towards national and European research foundations.</p> <p>Possible event at EBI Nov-Dec 2004.</p>
Task: 14.1 Organisation of a workshop/tutorial 14.2 Realisation of workshop/tutorial. 14.3 Study material
Deliverable: D14 Report WP14 m14
Interaction with other Work Packages: WP1 Project Management WP3 Public Website WP4 Network Meetings WP8 Mobility Program WP15 Workshop/tutorial on Text Mining and Information Retrieval



<i>Work Package 15 (WP15): Workshop/tutorial on Text Mining and Information Retrieval</i>
Responsible: UKLFR , UNIFR, DIM, ITRI, INSERM, SU
Participants: All partners
Person months: 13 (3.0, 2.6, 1.9, 1.9, 0.9)
Start: Month 10
Duration: Five months
Objective and Description: .Information retrieval techniques include mathematical modelling of the relation between document content, e.g. the content of the patient record, and the end-user query. Generally, the set of terms used for indexing and thereby also for retrieval is a flat list without internal structure apart from what is given lexically. An interesting question is how the structure of conceptual models imposed on the indexing system could be utilized to improve the performance of search engines. A common problem in text mining and information retrieval is the lack of a suitable metric to measure the relevance of given documents to a prior user query. One approach to solving this problem is latent semantic indexing, in which a low rank approximation of the term-document-matrix is obtained using singular value decomposition. Recent research has shown that canonical correlation analysis can effectively discover and model relevant relations between different sources of information by maximisation of mutual information. The objective is to plan and organise a workshop on text mining and information retrieval, thereby providing an overview of existing methods and approaches for indexing, of text, mining of relations between documents, and information retrieval as compared to methods applied to material composed of other data types than text (e.g. structured relational databases).
Possible event at EBI Nov-Dec 2004.
Task: 15.1 Organisation of workshop/tutorial 15.2 Realisation of workshop/tutorial 15.3 Study material
Deliverable: D15 Report WP15 m14
Interaction with other Work Packages: WP1 Project Management WP3 Public Website WP4 Network Meetings WP11 Workshop on Semantic Web WP13 Workshop/tutorial on NLP WP14 Workshop/tutorial on Data Mining



<i>Work Package 16 (WP16): Workshop/tutorial on the Electronic Health Record</i>
Responsible: UCL, INSERM, CNR-ISTC, UOM
Participants: All partners
Person months: 11 (5.6, 1.9, 1.9, 1.9)
Start: Month 7
Duration: Five months
Objective and Description: <p>This workshop will aim to bring members of this wide health informatics consortium, and other attendees, to a level of understanding that enables them to understand EHR research results, to collaborate with EHR research groups, to adopt and use next generation EHR systems in their own demonstrator settings, and to contribute positively to the next generation of research challenges in the widescale adoption and exploitation of comprehensive distributed electronic health records.</p> <p>The workshop will include an analysis of the new CEN EHR standards, the HL7 Clinical Document Architecture, and the work of the openEHR Foundation. The NoE partners organising this workshop have a 10-year pedigree in this field and are active in all of the areas mentioned. They already deliver masters programmes on these topics, but will use the resources of this work package to develop a wider range of study materials including case based examples and Web-based interactive tasks with a live EHR server.</p> <p>The workshop will specifically include the early results of WP35, offering a unique understanding of the ways in which EHR services might usefully interact with decision support and knowledge services to support evidenced based health care, clinical audit and next-generation research queries.</p> <p>Possible event in conjunction with MEDINFO 2004.</p>
Task: 16.1 Organisation of workshop/tutorial 16.2 Realisation of workshop/tutorial. 16.3 Study material
Deliverable: D16 Report WP16 m11
Interaction with other Work Packages: WP1 Project Management WP3 Public Website WP4 Network Meetings

9.6.5 Jointly Executed Research Activities

<i>Work Package 20 (WP20): Research Activity Multilingual Medical Dictionary</i>
Responsible: UKLFR , UNIFR, LiU (IMT), DIM, ITRI, UGOT, SU
Participants: KI, LiU (IDA), CNR, KITH, SOS, STAKES, NBH
Person months: 27 (5.0, 2.5, 3.7, 3.7, 3.7, 2.0, 1.7, 1.1, 1.1, 1.1, 0.5, 0.2, 0.2, 0.2)
Start: Month 6
Duration: 30 months
<p>Objective and Description:</p> <p>Cross-language mapping between medical terminologies, cross-language text retrieval and corpus-based machine translation require different kinds of multilingual terminology resources such as lexicons and thesauri, their underlying data structures, corpora, and tools for automated lexical data generation. This WP aims at the integration, adaptation and enhancement of biomedical terminologies. Several approaches will be tested and evaluated. Additional languages are assessed to what extent they fit into existing frameworks, migration routes between terminologies are explored and techniques for semi-supervised lexical acquisition are tested. A joint effort is the design of a common data structure for a multi-purpose medical dictionary covering several languages and domains.</p> <p>We capitalize on the following R&D experiences of the involved partners:</p> <p>Research on machine translation from parallel corpora is currently being carried out e.g. at Linköping University which is currently also constructing a Swedish-English dictionary. Parallel corpora which can be built upon are the translations of the official WHO classifications ICD and ICF, as well as other terminology systems available in parallel languages such as MeSH, ICPC, and a number of classifications of surgical procedures (which was the target domain of the GALEN-IN-USE project).</p> <p>A lexicon which has its scope limited to morpheme-like entities of biomedical terminology is given by the MorphoSaurus subword dictionary, being developed at Freiburg University for German, English, Portuguese and Spanish. It can be used (a) to support automated subword extraction from any texts and also (b) for language-independent automated indexing of existing biomedical terminologies as a support for cross-mapping of corresponding classes in different terminology systems.</p> <p>Two other initiatives propose a common data structure for biomedical lexicons: the "German Specialist Lexicon" (DSL) and the French Medical Language System (UMLF), represented by the Freiburg and Geneva partners. Extending the basic design foundations of the UMLS specialist lexicon in order to meet the needs of morphologically richer languages, the goal is to establish a complete lexical data model by which lexical and morphological knowledge can be represented. In addition, tools can be provided for the automated generation of lexical variants and inflections, and for the mapping between spelling variants, synonyms and abbreviations. Several of the NoE partners have extensive experience in multilingual lexical resources and computational lexicography (e.g. Göteborg, Linköping, Freiburg, Brighton, Geneva). Göteborg is involved in LSP (Language for Specific Purposes) teaching in the medical domain, viz. Medical Swedish for non-Swedish healthcare workers, and is starting a project on the introduction of ICT in these courses, where one of the tools to be developed is an "active dictionary" for supporting learners' reading of medical texts. A similar "active dictionary" will later be tried out for assisting patients in accessing their own patient records and other documents relating to their health.</p>
<p>Task:</p> <p>20.1 Facilitating short study visits of members of each others' groups</p> <p>20.2 Sharing and exchange of methods, materials and collaboration on work in progress</p> <p>20.3 Proposal for a common data structure for a multi-lingual medical dictionary</p> <p>20.4 Generation of multi-lingual medical lexicon in English, German, French, Portuguese, Italian,</p>



Spanish, Swedish in a range of 4.000-40.000 entries per language

Deliverable:

D20.1 Report Multi-lingual Medical Dictionary m11

D20.2 Report Multi-lingual Medical Dictionary m17

Interaction with other Work Packages:

WP1 Project Management

WP2 Common Database

WP3 Public Website

WP4 Network Meetings

Work Package 21 (WP21): Research Activity Ontology Engineering

Responsible: **UOM**, IFOMIS, CNR-ISTC, UKLFR, UNIFR,

Participants: LiU(IDA), LiU(IMT), KI, CAU, INSERM, ITRI, SU, UGOT

Person months: 27 (7.5, 3.7, 3.7, 2.0, 1.7, 1.1, 1.1, 1.1, 1.1, 1.1, 1.1, 0.6, 0.5)

Start: Month 7

Duration: 30 months

Objective and Description:

The field of information systems ontology is divided into three disciplines: investigating underlying ontological theory and philosophy, investigating formal logics for ontologies, and investigating how to construct and maintain ontologies that are functionally useful. All three schools are represented in the network.. Several network partners have long histories of working in the biomedical ontology field, using various philosophical and formal approaches, including:

IFOMIS has pursued the methodology of a reference ontology for the biomedical domain, developing a theory of categories and relations of entities in medicine, and exploring the trade-off between representational adequacy and computational usefulness.

Manchester University continues to build on its prior research both in authoring large scale, formally based biomedical ontologies (e.g. OpenGALEN, Gene Ontology) and also in delivering them (e.g. Gene Ontology Annotation Tool, TAMBIS, PRODIGY, Prescribing Indicators, CLEF). It continues to play a central part in the emergence of Ontology Web Language (OWL) as the World Wide Web Consortium (W3C) standard formal logic for ontologies on the semantic web.

Freiburg University develops ontological foundations for physical structure in biomedicine (modelling anatomy, cell biology), aiming particularly at a better understanding of reasoning over part-whole relationships between structures. It explores also ways to transform semi-formal ontologies into a formally rigid representation.

Laboratory for Applied Ontology (LOA) of CNR - ISTC performs basic and applied research on the ontological foundations of conceptual modelling, exploring the role of ontologies in different fields, such as: knowledge representation, knowledge engineering, database design, information retrieval, natural language processing, and the semantic web. The group draws on strong interdisciplinary expertise combining Computer Science, Philosophy and Linguistics, and relies on logic as a unifying paradigm. On the application side, special emphasis is given to the use of ontologies for electronic commerce, medical information systems, enterprise modelling, integration of lexical resources, and information access to the Web.

The objectives are:



To share understanding across the three ontological disciplines, especially those issues unique to or especially important to the biomedical domain, and to coordinate future research efforts so as to achieve coherent divisions of labour and to avoid duplication of effort.

To coordinate input into standardisation activities relevant to biomedical ontologies, including the emerging semantic web and its associated ontology authoring and delivery environments as well as established international medical informatics standards activities such as ISO, CEN, IEEE and HL7.

To argue the case for developers of bioscience and clinical vocabularies and coding systems to adopt ontology-based domain modelling, and to develop migration pathways that make it practical for them to do so.

To contribute to a consensus on a bio-medical "upper ontology".

To contribute to the convergence of clinical and bio-ontologies.

Task:

21.1 Cross-supervision of PhD candidates

21.2 Facilitating short study visits of members of each others' groups

21.3 Sharing and exchange of ontologies and materials and collaboration on work in progress

21.4 Developing and testing tools for ontology fusion and maintenance

21.5 Developing evaluation standards and benchmarks in order to better compare competing representation languages and ontology paradigms

21.6 Constructing large ontologies by knowledge extraction from informal sources (UMLS, Gene Ontology etc.)

21.7 Organising and attending conferences and standards meetings

Deliverable:

D21.1 Report Ontology Engineering m11

D21.2 Report Ontology Engineering m17

Interaction with other Work Packages:

WP1 Project Management

WP2 Common Database

WP3 Public Website

WP4 Network Meetings



<i>Work Package 22 (WP22): Research Activity SNOMED CT</i>
Responsible: KI , LiU(IMT)
Participants: UOM, SU, DIM, UCL, MRI, KITH, SOS, STAKES, NBH
Person months: 21 (7.5, 3.7, 3.7, 1.1, 1.1, 1.1, 1.1, 0.5, 0.2, 0.2, 0.2)
Start: Month 7
Duration: 30 months
Objective and Description: A large and potentially very interesting terminological system is SNOMED CT which in its new release (CT) includes the former SNOMED III and the Read Clinical Terms. SNOMED CT has approximately 350.000 concepts, 900.000 descriptions, 1.300.000 defined relations, and is organized in multiple root hierarchies. SNOMED CT is intended to serve as a reference terminology by which composite concepts could be defined by primitive ones. SNOMED CT is, as a system originating from the College of American Pathologists and the UK National Health Services Centre for Coding and Classification, developed and maintained in the English language. Partial translations are available in e.g. French, German and Spanish but a more comprehensive multi-lingual perspective must be added to the evaluation of SNOMED CT as a large-scale reference terminology for the European health care system. The objective of this research activity is to share experiences and understanding of the use of large scale reference terminologies in general and SNOMED CT in particular. Moreover, the objective is to encourage sharing of methods and tools for translation, and to co-ordinate the evaluation of SNOMED CT as a reference terminology e.g. for structured data entry applications in different domain.
Task: 22.1 Workshop for the discussion on national evaluations and strategies on SNOMED 22.2 Report on translation tools and experiences 22.3 Report on experiences using SNOMED in various European countries
Deliverable: D22.1 Report SNOMED CT m11 D22.1 Report SNOMED CT m17
Interaction with other Work Packages: WP1 Project Management WP3 Public Website WP4 Network Meetings



<i>Work Package 23 (WP23): Research Activity Health Statistics and Semantic Distance</i>
Responsible: LiU (IMT)
Participants: UOM, INSERM, MEDINFO, SOS, NORDCLASS, STAKES, KITH, NBH
Person months: 12 (7.5, 1.1, 1.1, 1.1, 0.3, 0.2, 0.2, 0.2, 0.2)
Start: Month 7
Duration: 30 months
<p>Objective and Description:</p> <p>Several coding systems are utilised in health care domains such as diagnoses, health problems, and interventions. The challenge is to allow aggregation according to different aspects or perspectives and to assure high information quality on different levels of data abstraction. Information quality is usually analysed in terms of reliability or observer variability, and current methods include kappa-type statistics, loglinear and latent class models for agreement, and frequency distribution display of categories in the coding system.</p> <p>According to kappa, which is the most widely used method, observed ratings are summarised in a contingency table and a measure of agreement that is corrected for the degree of agreement by chance is calculated. This method was first introduced for nominal data but has later been modified to take proximity in ordinal ratings into account, and it has also been generalised to cases with multiple raters. However, further research is needed to develop statistically sound methods for comparing ratings according coding schemes with different structures, i.e. with different numbers of categories or categories of different size.</p> <p>Another major research challenge is to develop methods for quantification of conceptual information embedded in the terminology models into measures of semantic distance. This approach relates to the problem of giving partial credit to less than perfect matches in the statistical methods for measuring reliability in categorical ratings. Generally, the set of terms used for indexing and thereby for retrieval is a flat list without an internal structure apart from what is given lexically. An interesting research perspective is how the structure of conceptual models imposed on the indexing system can improve the performance of search engines and support measurement of information quality. Thus, our objective is to develop methods for quantifying the semantic distance between index terms based on concept models including generic, partitive and associative relationships.</p> <p>The main objective of this research activity is to share experience, understanding and development of statistical methods for measuring information quality, ontologies for health indicators, and methods for quantification of semantic distance. Moreover, the objective is to encourage sharing of data material (e.g. quality registries and coded patient data) applicable for development and evaluation.</p>
<p>Task:</p> <p>23.1 Documentation of problems in European health statistics</p> <p>23.2 Documentation of ontologies for health indicators</p> <p>23.3 Proposal for methods for measuring reliability and semantic distance</p>
<p>Deliverable:</p> <p>D23.1 Report Health Statistics m11</p> <p>D23.2 Report Information Quality in Health Registries m17</p>
<p>Interaction with other Work Packages:</p> <p>WP1 Project Management</p> <p>WP3 Public Website</p> <p>WP4 Network Meetings</p>



<i>Work Package 24 (WP24): Research Activity Data Mining and Information Retrieval</i>
Responsible: EBI , UKLFR, DIM
Participants: UOM, LiU (IMT), LiU (IDA), UNIFR, KITH, SU, UGOT, STAKES,
Person months: 23 (7.5, 3.0, 3.7, 3.7, 1.1, 1.1, 0.7, 0.7, 0.6, 0.5, 0.4)
Start: Month 7
Duration: 30 months
<p>Objective and Description:</p> <p>Several of the partners of this NoE have extensive experience from data mining and information retrieval, but there is a lack of information exchange between projects and applications in different domains, partly due to highly specialised scientific communities focusing on e.g. image and signal processing, database management and text document handling. The overall objective of this research activity is therefore to develop a common understanding and framework for mining of relations in large data volumes composed of signals, symbols, text etc. Such a common framework will be critical for knowledge extraction from clinical and pre-clinical databases</p> <p>A common problem in text and data mining and information retrieval is the lack of a suitable metric to measure the relevance of given documents to a prior user query. One approach to solving this problem is latent semantic indexing, in which a low rank approximation of the term-document-matrix is obtained using singular value decomposition. Previous research from the field of signal theory has shown that canonical correlation analysis (CCA) can effectively discover and model relevant relations between different sources of information by maximisation of mutual information, a concept well known from information theory. Thus, a research challenge is to further develop non-linear CCA and related methods and to combine them with existing methods for text processing with applications to e.g. automatic classification of medical diagnosis from patient records and cross-language information retrieval. Additional problems for medical language processing include establishing the proper scope of automated as opposed to manual vs. semi-automatic indexing, the construction of biomedical lexicons geared towards different languages. Evaluation of the results of information retrieval and text mining requires extensive annotated test corpora of a type which do not yet exist for the medical domain. Another main question is the degree of content abstraction we should aim at. On the one hand the results of text and data mining must be expressive enough that we are able to measure the semantic distance, on the other hand it must be viable with regard to lexical and conceptual coverage as well as computational cost.</p> <p>EBI are currently developing research in the fields of data mining, genome sequence annotation, classification of protein function, ontologies for molecular biology, and knowledge representation in molecular biology databases. Thus, the objective of this activity is, apart from what is expressed above, to share in-depth knowledge of the content and structure of EBI databases, as well as knowledge of useful methods and tools for data mining in these databases. Although well established and already widely used, this NoE activity will promote the establishment of new collaboration between research groups of medical informatics and bioinformatics and thereby the authoring of new research applications targeted towards national and European research foundations.</p>
<p>Task:</p> <p>24.1 Documentation of ongoing research activities with description of methods and material</p> <p>24.2 Pooling of relevant methods and material</p> <p>24.3 Proposal for application for collaborative research grants</p>
<p>Deliverable:</p> <p>D24.1 Report Data Mining and Information Retrieval m11</p> <p>D24.2 Report Data Mining and Information Retrieval m17</p>
<p>Interaction with other Work Packages:</p> <p>WP1 Project Management</p> <p>WP3 Public Website</p>



WP4 Network Meetings

*Work Package 25 (WP25): Research Activity Concept Systems for Laboratory Medicine*Responsible: **LiU(C-NPU)**

Participants: SU, KI, EBI

Person months: 11 (7.5, 1.1, 1.1, 1.1)

Start: Month 7

Duration: 30 months

Objective and Description:

The practice of computing in medical laboratories has reached a maturity level that requires clear insights into what the efforts are supposed to achieve in view of seamless communication of laboratory generated data for the benefit of the individual patients. This should be so regardless of administrative, technical, language or cultural boundaries. The availability of electronic patient records will also greatly increase the quality of health care, reduce health care cost and facilitate epidemiological surveys to the benefit of patient's. Thus in developing a generalized architectural build up and adherent structures, there is a need for systems and schemes in each of the medical domains to support and populate the electronic patient record.

The scientific basis, management and execution of laboratory medicine must thus be supported by computing and communication solutions that take care of the internal data flow of laboratories and communication with requesting parties using concepts that adequately represent examinations and concepts used by the health care providers. This can be achieved by a wide range of Laboratory Information Systems, Request and Report Systems and conceivably other sophisticated subsystems integrated into a health care computing context

The efforts to express properties examined in medical laboratories made by the Committee on Nomenclature, Properties and Units (C-NPU) of the International Federation of Clinical Chemistry (IFCC) and the International Union of Pure and Applied Chemistry (IUPAC), applying principles of metrology as fostered by the Bureau International des Poids et Mesures (BIPM) is normative in character in the interest of promoting sound common concepts (properties) for expressing the result of examinations in laboratory medicine and developing coding schemes that unambiguously carries the result to the doctor requesting the result.

The workpackage is meant to provide a forum for researchers developing connectivity between the bioinformatics databases, the analytical laboratory process and the electronic health care record. The long term objective is to develop concepts that can be understood, aggregated and generalized depending on the viewpoint of various healthcare worker and their IT support systems without compromising the basic meaning of laboratory reports including information based on data produced by bioinformatics.

Task:

25.1 Analysis of ontologies bridging between bioinformatics and clinical laboratories, e.g between the Gene Ontology and Microbiology

25.2 Analysis of ontologies bridging between clinical laboratories and the health care record

Deliverable:

D25.1 Report Concept System for Laboratory Medicine m11

D25.2 Report Concept System for Laboratory Medicine m17

Interaction with other Work Packages:

WP1 Project Management

WP3 Public Website



WP17 Standardisation work

<i>Work Package 26 (WP26): Research Activity Electronic Health Record</i>
Responsible: UCL, UOM, CNR-ISTC, INSERM
Participants: DIM, LiU (IMT), IFOMIS, UKLFR, KI, STAKES, KITH, NBH, SOS
Person months: 25 (7.5, 3.7, 3.7, 3.7, 1.1, 1.1, 1.1, 1.1, 1.1, 0.3, 0.3, 0.3, 0.2)
Start: Month 7
Duration: 30 months
<p>Objective and Description:</p> <p>This work package activity will enable leading-edge research groups active in specifying, developing and deploying EHR services to network and collaborate with other health informatics centres of excellence in the semantic, decision support and data analysis fields. The early work will be informed by the results of WP3, which is specifying the requirements for a project repository of medical records and other knowledge resources. Through a variety of site visits, joint workshops and student projects the requirements, data flows, interfaces and information models needed to enable rich semantic interrogation of patient records will be defined. This will include the support of queries needed to run decision support systems interoperable with EHR services, and for the EHR to include the results of a user's interaction with a guideline system. These activities will foster significant knowledge and skills sharing across consortium members, and at the same time provide a target objective that will be very practical and deliverable.</p> <p>Although significant engineering work cannot be supported through a Network of Excellence, the specification outputs of this work package will be directly applied to enhance existing and practical demonstrators to provide a verification pathway for this work.</p> <p>The practice of evidence-based medicine will require the seamless and standards-based interaction of knowledge, record and inference services. There is presently limited ability to analyse individual patient records that have been federated from multiple sources, or to interrogate population records that are comprehensive, multi-enterprise and possibly organised at the level of a regional health care network. The human and skills network established through this work package will greatly inform the next generation of such EHR demonstrators, will contribute to the development of tools, and will place the individual research teams in a good position to bid for future collaborative projects through recognised research funding bodies.</p>
<p>Task:</p> <p>26.1 Specification of the interoperation of EHR, decision support and medical ontology services</p> <p>26.2 Semantic mining of the EHR: specification of information models and services.</p>
<p>Deliverable:</p> <p>D26.1 Report Information Architecture of the Semantic Electronic Health Record m11</p> <p>D26.2 Report Information Architecture of the Semantic Electronic Health Record m17</p>
<p>Interaction with other Work Packages:</p> <p>WP1 Project Management</p> <p>WP3 Public Website</p> <p>WP8 Standardisation work</p>