SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

INFRA-2007-1.2.2 - Deployment of eInfrastructures for scientific communities

Grant agreement for: Combination of Collaborative projects & Coordination and support actions

Annex I - "Description of Work"

Project acronym: e-NMR
Project full title: Deploying and unifying the NMR e-Infrastructure in System Biology
Grant agreement no.: 213010

Date of preparation of Annex I (latest version): 26/03/2009
Date of approval of Annex I by Commission: 1/4/2009

List of Beneficiaries

Beneficiary Number *	Beneficiary name	Beneficiary short name	Country	Date enter project	Date exit project
1 (Coordinator)	JOHANN WOLFGANG GOETHE-UNIVERSITAET FRANKFURT AM MAIN	BMRZ	Germany	1	36
2	Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine Paramagnetiche	CIRMMP	Italy	1	36
3	UNIVERSITEIT UTRECHT	BCBR	The Netherlands	1	36
4	ISTITUTO NAZIONALE DI FISICA NUCLEARE	INFN	Italy	1	36
5	European Molecular Biology Laboratory	EMBL	Germany	1	36

Contents page

PART A

A1. Budget breakdown and project summary

A.1 Overall budget breakdown for the project

A3.2: What it costs

Project Number [1]	213010	Project Acronym [2]	e-NMR

One Form per Project

Participant number in this project [9]	Participant short name	Estimated eligible costs (whole duration of the project)						Total receipts	Requested EC contribution
		RTD (A)	Coordination (B)	Support (C)	Management (D)	Other (E)	Total A+B+C+D+E		
1	JWG	422,400.00	83,588.00	0.00	18,649.00	55,950.00	580,587.00	0.00	447,300.00
2	CIRMMP	484,400.00	137,720.00	0.00	74,560.00	56,000.00	752,680.00	0.00	549,800.00
3	UU	493,333.00	108,038.00	0.00	18,800.00	18,800.00	638,971.00	0.00	446,999.00
4	INFN -INFN	422,336.00	111,478.00	0.00	18,649.00	55,950.00	608,413.00	0.00	465,900.00
5	EMBL	184,400.00	0.00	0.00	1,700.00	0.00	186,100.00	0.00	140,000.00
TOTAL		2,006,869.00	440,824.00	0.00	132,358.00	186,700.00	2,766,751.00	0.00	2,049,999.00

A.2 Project summary

Grant agreement Preparation Forms

EUROPEAN COMMISSION

7th Framework Programme on Research, Technological Development and Demonstration

Combination of Collaborative Project and Coordination and Support Action

A1: Our Project

Project number (1)	213010	*Project acronym (2)*	e-NMR

ONE FORM PER PROJECT

GENERAL INFORMATION

Project Title (3)	Deploying and unifying the NMR e-Infrastructure in System Biology			
Starting Date (4)	01/11/2007			
Duration in months (5)	36	*Call (part) identifier (6)*	FP7-INFRASTRUCTURES-2007-1	
Activity code(s) most relevant to your topic (7)	INFRA-1.2-02	-	-	
Free keywords (8)				

Abstract(9) (max. 2000 char.)

NMR plays an important role in life sciences (bio-NMR), and structural biology in particular, at both European and international levels. An I3 initiative is operative in the field, which provides access to NMR instrumentation and pursues technical advancement (EU-NMR). In addition, the EC has funded a Coordination action (NMR-Life) aimed at the establishment of common experimental approaches and at the spreading of best experimental practices across Europe. Altogether, these two initiatives provide a reference point for the large majority of European scientists with an interest in bio-NMR.
In parallel, European developments in the area of Research Infrastructures in the past years resulted in a leading edge high-speed research network covering all Europe and in a overlaying production Grid infrastructure, realized by projects as EGEE/EGEE II. This integrated network and processing/storage environment – e-Infrastructure - provides a platform for new methods of global collaborative research – e-Science.
The main objective of this project is to optimise and extend the use of the NMR Research Infrastructures of EU-NMR through the implementation of an e-Infrastructure in order to provide the European bio-NMR user community with a platform integrating and streamlining the computational approaches necessary for bio-NMR data analysis (e-NMR). The e-NMR infrastructure will be based on the Grid infrastructure.
The project will also tackle the following objectives: i) establish a human collaboration network between the bio-NMR and the e-Infrastructure scientific communities; ii) assess the state-of-the-art of the computational aspects of bio-NMR; iii) implement and make available state-of-the-art computational methods. A broad range of networking activities will focus on monitoring, dissemination and outreach, training, hands-on workshops. The development and enforcement of operational and organizational schemes and policies will also be addressed.

A.3 List of beneficiaries

Beneficiary Number	Beneficiary name	Beneficiary short name	Country
1 (Coordinator)	JOHANN WOLFGANG GOETHE-UNIVERSITAET FRANKFURT AM MAIN	BMRZ	Germany
2	Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine Paramagnetiche	CIRMMP	Italy
3	UNIVERSITEIT UTRECHT	BCBR	The Netherlands
4	ISTITUTO NAZIONALE DI FISICA NUCLEARE	INFN	Italy
5	European Molecular Biology Laboratory	EMBL	Germany

PART B

B1. Concept and objectives, progress beyond state-of-the-art, S/T methodology and work plan

B.1.1 Concept and project objective(s)

Nuclear Magnetic Resonance (NMR) plays an important role in life sciences, and structural biology in particular, at both European and global levels. The application of NMR to the investigation of biological systems, using high resolution spectrometers equipped with high-field magnets (typically 15 T to more than 20 T) is generally called bio-NMR. In order to boost European research in this burgeoning field of science, a network of research infrastructures (EU-NMR, contract no.: 026145) has been established for providing experimental access and technical advancements in bio-NMR. The same research infrastructures constitute the core of a Coordination action (NMR-Life, contract no.: 018758) that involves in total 10 European research teams at the forefront of bio-NMR with aim to contribute to the establishment of common experimental approaches and to the spreading of newly developed best practices. Altogether, these two initiatives provide a reference point for the large majority of European scientists working directly in the field of bio-NMR or using bio-NMR as a tool for their research in biology and biomedicine. Indeed, these scientists perform all or part of their measurements at the EU-NMR Research Infrastructures, participate in yearly EU-NMR user meetings, and are involved in the dissemination activities of NMR-Life. Transnational access and coordinative/networking actions similar to EU-NMR and NMR-Life, all focused on the experimental aspects of bio-NMR, have been in place since the mid-nineties. The entire European bio-NMR community is thus quite closely integrated for all that concerns the experimental aspects and protocols applied.

In parallel, European developments in the area of Research Infrastructures in the past years resulted in a leading edge high-speed research network with full administrative and operational support (GÉANT + NRENs, covering all Europe) and in an overlaying production Grid infrastructure, realized by projects as EGEE/EGEE II. EGEE II is currently the largest in the world and enables distributed computing power and storage sharing across geographical and administrative domains for a variety of Virtual Organizations. This integrated network and processing/storage environment – e-Infrastructure - provides a platform for new methods of global collaborative research – e-Science.

The main objective of this project is to optimise and extend the use of the EU-NMR Research Infrastructures through the implementation of an e-Infrastructure, e-NMR. This e-Infrastructure will provide the European bio-NMR user community with a platform that integrates and streamlines the computational approaches necessary for bio-NMR data analyses. The e-NMR infrastructure will be based on the Grid infrastructure.

e-NMR will address the extensive fragmentation of computational methods adopted by individual teams in bio-NMR research, which at present prevents full interoperability among different laboratories and detailed comparison of results derived from bio-NMR data, thereby limiting the scientific impact of European bio-NMR research. More in detail, the main goal of e-NMR is to address all computational aspects of NMR, from data acquisition, processing and analysis to structure computations and validation. e-NMR will first integrate all NMR research infrastructures in a pilot NMR-Grid, which will be connected later on to European Grids such as the EGEE. Reaching this main goal will require porting the major NMR applications to e-Science, solving all technical problems related to middleware services functionality and developing a specific, general NMR application framework. In parallel to the implementation of the e-NMR infrastructure, we will tackle also the following objectives: i) establishing a human collaboration network between the bio-NMR and the e-Infrastructure scientific communities; ii) assess the state-of-the-art in structural bio-NMR; iii) implement state-of-the-art computational methods within the

NMR application framework and extend these methods where required. A broad range of networking activities will focus on monitoring, dissemination and outreach, training, and hands-on workshops, in order to maximize the collaboration with and impact on the international bio-NMR and e-Science communities. Dissemination and training activities will also contribute to spreading the use of best computational practices in the European scientific community. The development and enforcement of operational and organizational schemes and policies, which is crucial for the implementation of the pilot NMR-Grid infrastructure, will also be addressed within the networking activities of the project.

This project will be aimed at achieving the following measurable S&T objectives:

- *Objective 1:* Deploy and support an interoperable Grid infrastructure at the three NMR infrastructures to be later integrated with EGEE
- *Objective 2:* Assess the state-of-the-art in computational bio-NMR and promote knowledge and spreading of best computational practices
- *Objective 3:* Identify and promote a sustainable NMR technical framework for e-Science
- *Objective 4:* Develop an integrated e-NMR computational platform

The achievement of objective 1 will be measured initially by the completion of the installation of the middleware at the three NMR research infrastructures, planned for the first year of activity, and later on by the number of services provided by the e-NMR platform as well as by the evaluation of the effective use of the infrastructure, e.g. through the number of jobs run or the CPU time used. The success of the infrastructure will be measured, when close to project completion (that is, in the last year of the project), by the number of new components deployed. The completion of the installation of the middleware at the three NMR research infrastructures will constitute a milestone of the project (M2.1).
With respect to the second S&T objective of the project, its advancement will be measured by the number of assessment activities carried out, which will take the form of events where groups of experts will be convened to test innovative and/or optimized procedures on well-defined bio-NMR problems. The successful organization of the first of this event will be a milestone for the project (M1.1). For the spreading of best computational practices, we will act at two levels. First, we will seek connections with other initiatives and networks focused on the computational aspects of bio-NMR or making extensive use of highly automated protocols for the analysis of bio-NMR data. In parallel, we will establish dissemination mechanisms to provide training and increase the awareness of recent innovation by the community. The success of this activity will be measured by the participation of bio-NMR scientists - as well as of researchers more specialized in the field of e-Science - to training sessions and to web-based discussions. The number of training materials (e.g. demos, tutorials, etc.) made available will be another indicator of the degree of achievement of the objective.
Objectives 3 and 4 are closely interlinked. Their achievement will logically start with the identification of NMR applications suitable for porting to the Grid (milestone 3.1), and will proceed through the identification and development of the enhancements of the middleware need to support the selected applications (milestone 2.2). These tasks will naturally lead to making optimized state-of-the-art protocols for many widespread bio-NMR applications available on the Grid. The number of applications that are selected, the number of middleware enhancements that will be proposed or carried out, and, most importantly, the number of protocols and software tools that will be made available are all suitable measures for these objectives. Selected applications will be collected in a dedicated software directory, which will be made available (milestone 3.3) in order to encourage and facilitate other NMR centers to join the e-NMR Grid. In the course of the development of the e-NMR platform, we will also provide support for developers to allow easy porting of their software

to the Grid. The release of a set of recommendations for this purpose will constitute a milestone of the project (M3.2).

The successful completion of the present project will open up an entirely new scientific community for the adoption of e-Infrastructures. The bio-NMR community is in fact already relatively well organized for all that concerns the experimental aspects involved with the technique. However, it dramatically lacks organization in the realm of its computational aspects, which are nevertheless of the highest importance for the extraction of high-quality information from the raw experimental data, and the subsequent derivation of biologically meaningful information. Therefore, the bio-NMR community is optimally suited to support the consolidation and expansion of e-Science, while the e-Infrastructure model is optimally suited to satisfy the computational needs of the European bio-NMR community. At the same time, the present project will encourage standardization of the computational approaches applied by the bio-NMR community and thus will foster interoperability among the NMR research infrastructures. This will enhance trans-national access to the existing NMR infrastructures, as well as improve the quality and impact of access to smaller, national or regional level, bio-NMR centres. The availability of the e-NMR platform will also be important to expand the scientific impact of research activities carried out with the use of bio-NMR. Thanks to availability of certified, well-tested protocols, e-NMR will give more visibility to the results that are obtained, and will strengthen the confidence of the broader scientific communities in the outcomes of the data analyses that are performed by individual researchers. On the medium- to long-term timeframe, new users from neighbouring disciplines, who have up to now been discouraged from getting involved in bio-NMR because of the interoperability and streamlining difficulties in the current computational protocols for the analysis of data, will be attracted to the field. The application of e-Infrastructure technologies to bio-NMR will progress along with optimization of the middleware. This optimization will be carried out in tight connection with the relevant European e-Infrastructure initiatives, thereby creating a tight link between these different scientific communities.

B.1.1.1 Global view of the activities

Gantt Chart

	3	6	9	12	15	18	21	24	27	30	33		36
WP0: Project coordination	M0.1												M0.2
Task 0.1				D0.1				D0.2					D0.3
Task 0.2													
Task 0.3													
WP1: Monitoring, Standardisation and Outreach						M1.1 M1.2							
Task 1.1			D1.1		D1.2	D1.3				D1.4			
Task 1.2											D1.6		
Task 1.3								D1.7			D1.5		D1.8
WP2: e-NMR Grid deployment and operation				M2.1				M2.2	D2.4				D2.5
Task 2.1				D2.1				D2.8	D2.9				
Task 2.2						D2.2					D2.6		
Task 2.3							D2.3						D2.7
WP3: Design and development of the e-NMR Grid platform													D3.6
Task 3.1		D3.1		M3.1									
Task 3.2				M3.2		D3.3		D3.9		D3.7			
Task 3.3			M3.3			D3.5							D3.8
Task 3.4					D3.2					D3.4			

D0.1, D0.2, D0.3: Organization of annual meeting and preparation of report
D1.1 e-NMR database model (a document describing it will be publicly released)
D1.2 Test data sets for software development and testing (a document describing them will be publicly released)
D1.3 Assessment of state-of-the-art: first workshop for assessing the state-of-the art in computational aspects of bio-NMR
D1.4 Assessment of state-of-the-art: second workshop for assessing the state-of-the art in computational aspects of bio-NMR
D1.5 Training of bio-NMR users: training sessions for bio-NMR users at the yearly meeting of the EU-NMR large facility and global evaluation of all training activities performed
D1.6 Report on connections established
D1.7 and D1.8: annual reports on the dissemination of e-NMR results in the bio-NMR and e-Science communities
D2.1 e-NMR grid infrastructure operational
D2.2 Assessment of the e-NMR infrastructure and GOC report
D2.3 New components' integration and deployment report

D2.4 Report on e-NMR infrastructure state-of-the-art toward integration with EGEE/EGI
D2.5 Final report on e-NMR infrastructure usage and integration with EGEE/EGI (update of D2.4)
D2.6 Second Assessment of the e-NMR infrastructure and GOC report (iteration of D2.2)
D2.7 New components' integration and deployment updated report (iteration of D2.3)
D2.8 Survey of the interest and demands of the user community as well as of potential industrial stakeholders with respect to the provision of fee-based services and/or training activities
D2.9 Evaluation of the results of the survey and implementation of consequent actions

D3.1: Release of first set of use-cases (a document describing them will be publicly released)
D3.2: Overview of existing validation software able to deal with nucleic acids
D3.3: Request for Enhancements of gLite to support bio-NMR applications
D3.4: Structure validation suite for DNA/RNA (a document describing it will be publicly released)
D3.5: First set of online tutorials (3) describing in details the optimal use of the e-NMR platform for solution NMR
D3.6: Integrated e-NMR platform
D3.7: Updated RFE of gLite to support bio-NMR applications
D3.8: Online tutorials describing in details the optimal use of all elements of the e-NMR platform for solution NMR (iteration of D3.5)
D3.9: Analysis of security issues brought about by the user community and by industrial stakeholders

M0.1: Organization of kick-off meeting
M0.2: Organization of concluding general meeting
M1.1: Report on the results of the first assessment workshop based on the standardized test data sets constituting D1.2
M2.1: Grid infrastructure operational
M2.2: First integration of JRA developed components
M3.1: List of NMR applications suitable for GRID porting
M3.2: Recommendations (software requirements) for NMR software developers to allow easy porting to the e-NMR grid
M3.3: Description of e-NMR software installation structure

Description of interlinks between workpackages

Besides the workpackage specifically dedicated to the management of the project (WP0), we have planned three other workpackages that run in parallel and deal individually with the three main aspects of the present initiative: networking (WP1), service (WP2) and joint research activities (WP3). Workpackages are further split into tasks, in order to allow the managing committee to follow the progress of the project more easily. All workpackages start, with one or more tasks, at the very beginning of the project.
In the first year, the service activities will involve mainly the deployment of the e-NMR grid infrastructure (D2.1). In the meantime, selected tools will be made available for remote usage, while the networking and joint research activities will create a reference for the sharing of data and applications not only among the project members but also with software users and developers outside the present partnership. We will initially assemble use-cases for the applications that will constitute the e-NMR platform (D3.1 and M3.3) which will be made available also through the e-NMR database (D1.1). The latter database will not only contain the use-cases but will also be a repository for data generated/elaborated within e-NMR service activities (WP2). The initial use-cases of D3.1 and other data from the e-NMR database will evolve into a suite of test data sets for software developers that will be exploited in the context of WP1 (D1.2). An assessment workshop will be organized in the second year to evaluate different software tools and calculation protocols

based on the test datasets (M1.1 and D1.3).These activities will provide input for the development of new middleware components specifically designed for the needs of e-NMR (D3.3 and D3.7), which will also improve the service provided (D2.3 and D2.7). The feedback from the user community, as well as that from potential industrial stakeholders, will be collected and analysed in the light of defining actions to guarantee the sustainability of eNMR (D2.8 and D2.9) Prior to the integration of these new components in e-NMR, we will have set up the Grid Operation Center, enabling full exploitation of the grid infrastructure (D2.2 and D2.6). Task 3.4 will in parallel provide indications on problems lacking appropriate software tools for their resolution (D3.2), which in conjunction with the recommendations that will be issued to NMR software developers to allow easy porting to the e-NMR platform (M3.2), will boost the portfolio of scientific applications that can be dealt with through the e-NMR grid. The performance of e-NMR itself will also improve during the project through the development of new/ameliorated calculation protocols in WP3 for inclusion in the grid. These protocols will be tested in, and/or will result from, the assessments workshops of WP1 (D1.3 and D1.4). In order to promote the rapid and successful uptake of NMR Grid technology, we will strive to collaborate closely with the industry as the NMR technology provider. To this end, their needs in terms of data security will be explicitly analyzed (D3.9). Optimal solutions to common problems of the bio-NMR community will be developed in WP3 and made available in the form of online tutorials (D3.5 and D3.8). This material will be also of use for the training activities that will be carried out in WP1 (D1.5). In parallel, in the last year of activity, we will also seek integration with EGEE or other related follow-up European projects, such as EGI, (D2.4 and D2.5). In particular for the latter aim, the e-NMR grid nodes will integrate with their respective national grid initiatives. The integration with EGI will exploit our experience of interaction with the eScience community accumulated during the project (D1.6).

Overall Deliverable List

Del. No.	Deliverable name	WP no.	Lead beneficiary	Estimated person-months	Nature	Dissemination level	Delivery date (proj. month)
D3.1	Release of first set of use-cases	3	1,2,3	9	O	PU	Month 6
D1.1	e-NMR database model	1	2,3	12	R	PU	Month 12
D0.1	Organization of first annual meeting and reporting	0	1	8	R	PU	Month 12
D2.1	e-NMR grid operational	2	4	18	R	PU	Month 12
D1.2	Test data sets	1	1,2,3	12	R	PU	Month 15
D3.2	Overview of software for nucleic acids	3	1,2,3	9	R	PU	Month 15
D1.3	First assessment of state-of-the-art	1	2, 3	14	R	PU	Month 18
D2.2	Assessment of the e-NMR infrastructure and GOC report	2	4	4	R	PU	Month 18
D3.3	Request for Enhancements of gLite to support bio-NMR applications	3	4	18	R	PU	Month 18
D3.5	Online tutorials	3	1,2,3	18	R	PU	Month 18
D2.3	New components' integration and deployment report	2	4	2	R	PU	Month 21
D0.2	Organization of second annual meeting and reporting	0	1	8	R	PU	Month 24
D1.7	Annual report on the dissemination of e-NMR results in the bio-NMR and e-Science communities	1	2,4	1	R	PU	Month 24
D2.8	Survey of the interest and demands of the user community as well as of potential industrial stakeholders with respect to the provision of fee-based services and/or training activities	2	1,2,3	6	R	PU	Month 24
D3.9	Analysis of security issues brought about by the user community and by industrial stakeholders	3	4	6	R	PU	Month 24
D2.4	Report on e-NMR toward integration with EGEE	2	4, 1	8	R	PU	Month 27
D2.9	Evaluation of the results of the survey and implementation of consequent actions	2	1,2,3	6	R	PU	Month 27
D1.4	Second assessment of state-of-the-art	1	2, 3	14	R	PU	Month 30

D3.4	Structure validation suite for DNA/RNA	3	2	12	R	PU	Month 30
D3.7	Updated RFE of gLite to support bio-NMR applications	3	4	18	R	PU	Month 30
D1.5	Training of bio-NMR users	1	1	12	R	PU	Month 33
D1.6	Report on connections established within Task 1.2	1	1	18	R	PU	Month 33
D2.6	Second Assessment of the e-NMR infrastructure and GOC report	2	4	5	R	PU	Month 33
D3.8	Online tutorials	3	1,2,3	18	R	PU	Month 36
D2.7	New components' integration and deployment updated report	2	4	3	R	PU	Month 36
D0.3	Organization of third annual meeting and reporting	0	1	8	R	PU	Month 36
D1.8	Annual report on the dissemination of e-NMR results in the bio-NMR and e-Science communities	1	2,4	1	R	PU	Month 36
D2.5	Final report on e-NMR	2	4, 2	12	R	PU	Month 36
D3.6	Integrated e-NMR platform	3	1,2,3	82	R	PU	Month 36
			Total	324			

- Deliverables of the type O will be accompanied by public documentation.
- The estimated person-months include work subsequent to the release of the deliverable that will be needed to incorporate suggestions/amendments/improvements from the users or other researchers

Tentative schedule of project reviews			
Review no.	**Tentative timing, i.e. after month X = end of a reporting period**	***planned venue of review***	***Comments , if any***
1	After project month: 12	**Frankfurt**	
2	After project month: 24	**Florence**	

Highlights of project advances will be submitted quarterly to the Project Officer in charge.

Project Effort Form - indicative efforts per activity type per beneficiary

Activity Type	BMRZ	CIRMMP	BCBR	INFN			TOTAL ACTIVITIES
Joint Research (RTD) activities							
WP3: Design and development of the e-NMR Grid platform	46	50	58	30			184
Total 'JRAs'	46	50	58	30			184
Networking (Coordination) activities							
WP1: Monitoring, standardization and outreach	12	16	18	18			64
Total 'NAs'	12	16	18	18			64
Support Activities							
WP name							
Total 'Support'							
Consortium Management activities							
WP0: Management	9	9	3	3			24
Total 'Management'	9	9	3	3			24
Service (Other) activities							
WP2: e-NMR Grid deployment & operations	10	10	10	22			52
Total 'SAs'	10	10	10	22			52
TOTAL BENEFICIARIES	77	85	89	73			324

B.1.2 Progress beyond the state-of-the-art

The bio-NMR scientific community - Nuclear Magnetic Resonance (NMR), together with X-ray crystallography, is the major technique capable of providing information on the three-dimensional structures of molecules. Since now more than 20 years, NMR has acquired the capability of solving the structures of molecules of increasing complexity (macromolecules), notably proteins and nucleic acids. The European NMR community in Life Sciences is composed of a large number of research teams that tackle a vast portfolio of different biological problems. This is achieved by investigating the structural and dynamic properties of macromolecules (and of their adducts) in solution or, since a few years, in the solid state, or in both states. These research teams in turn directly collaborate with an equivalent or larger number of scientific groups that are mainly skilled in molecular and cell biology and provide samples for the experimental measurements carried out by the NMR spectroscopists. The outcomes of these studies are of interest for an even wider community of scientists, because of the biological and biomedical implications of the results attained. Structural Biology provides key insight into the reactions in our cells and is of paramount importance in translational research towards the development of new drugs.

The European bio-NMR community comprises about a hundred research groups, ranging in size from a few to several tens of individual researchers. The whole community of stakeholders, as described above, therefore includes several thousand scientists all across Europe. The I3 project EU-NMR is focused on the provision of NMR measurement time and scientific expertise to European researchers who need access to the best available equipment and most sophisticated know-how to carry out their experiments of choice. About a hundred projects are carried out yearly through this access scheme at the research infrastructures involved in EU-NMR (Figure 1). In

Figure 1: Geographical distribution of European NMR Large Scale infrastructure users (FP6 EU-NMR) and number of projects in the period March 2006- February 2007.

addition to the EU-NMR project, the EC-funded Coordination Action NMR-Life brings together the above-mentioned research infrastructure operators and a wide representation of the scientifically most important European research teams working in bio-NMR. This partnership is committed to closely monitoring the development of bio-NMR techniques and technologies, with particular focus on the aspects relevant for the investigation of macromolecular adducts.

Computational requirements of a typical NMR project - NMR is making an important contribution in life sciences by enabling the structural and dynamic characterization of biomacromolecules and of their interactions. The typical steps from sample preparation to data interpretation, through experimental data collection, involve a significant amount of computational work. As such, NMR has always been associated with, and driving new developments in, the high-end computational field. Next to providing information at a structural level, NMR is unique in its ability of characterizing dynamical properties of molecules, an information that is often used in computational chemistry for force field development and validation and is required for the full understanding of enzyme catalysis, also in conjunction with simulations based on time dependent *ab-initio* molecular dynamics. One can distinguish in NMR a number of computational tasks each with various data storage and computational requirements.

The first step involves the processing of the raw time-domain NMR data with multidimensional fast Fourier transform algorithms and computational intensive reconstruction techniques such as Maximum Entropy methods. The size of the resulting processed data (high dimensional spectra) typically exceeds by one to two orders of magnitude the amount of raw data. The processed data are then analyzed with computer-assisted graphical software to interpret the spectra and to extract structural restraints. The latter are used together with chemical and physical knowledge in optimization algorithms to find the arrangement of atoms that fulfil the experimental data. For this purpose, techniques based on molecular dynamics simulation in combination with simulated annealing protocols are typically used. Since such algorithms are heuristic in nature and the experimental data limited, several hundreds of independent calculations are typically performed, highly suitable for distributed computing. Data analysis and optimization steps are tightly intertwined in a cyclic manner and several rounds of computations are thus needed.

The state-of-the-art - The size of the community working on NMR in Structural Biology worldwide is of about 200-300 research groups having the capability of solving NMR structures of biological macromolecules, plus a similar number of biology groups, within or outside the former groups, providing the biological material. The throughput per group is widely variable, ranging from less than one up to several tens of structures solved per year (http://www.rcsb.org/pdb/home/home.do). Collectively, the EU-NMR research infrastructures participating in the EC Trans-national Access programs since FP3, plus their associates, provide the majority of the European throughput. The overall procedure for bio-NMR data analysis leading to protein structure determination is common to the various research teams in the field, as it has been largely developed between the late 80's and the 90's (culminating with the Nobel prize to Kurt Wüthrich in 2002). Nevertheless, most NMR laboratories apply different computational protocols (and thus software tools) to analyze the data and calculate the three-dimensional structures of biological macromolecules. Indeed, most of the protocols used have related and/or common modules, but their interoperability is virtually zero. In part, this is due to the fact that there is no "optimal protocol". The performance of the different computational approaches used in the structural NMR field is somewhat dependent on the experimental systems studied and on the expertise of the researchers. The consequences of this state of affairs entail that

- it is difficult to compare results from different laboratories
- there is not a univocal mode to validate results against raw data

- it is difficult to implement, and thus spread, new algorithms

In this project, we want to set up a platform incorporating and integrating the most widespread calculation modules, and covering all the steps from the raw NMR data to the final 3D structure. Such a platform should make it easy to incorporate any significant advances, including alternative computational approaches, as well as new validation tools. The existence of a large user base will contribute to the rapid spread in the use of this platform. For all this to work, we have to provide adequate computing power, which is beyond what can be provided by a single NMR centre, and resources that are scalable with the increasing needs. Therefore, we will set up a European NMR grid based on the combination of advanced bio-NMR expertise with Grid infrastructures. The synergistic combination of the expertise of the major European NMR research infrastructures with computational science centres will result in the unified e-NMR infrastructure. This will perfectly complement the Integrated Structural Biology Infrastructure foreseen by the ESFRI Roadmap. The e-NMR project aims to build an interoperable Grid infrastructure with EGEE to support the test and deployment of enhanced NMR scientific applications. Identification of relevant user applications, dissemination, and training efforts will be oriented to define a sustainable framework. This initiative is of strategic importance in order to reinforce the collaborative NMR research efforts at European level and to enhance their productivity and impact by taking advantage of the recent deployment of the EGEE enhanced e-Infrastructure.

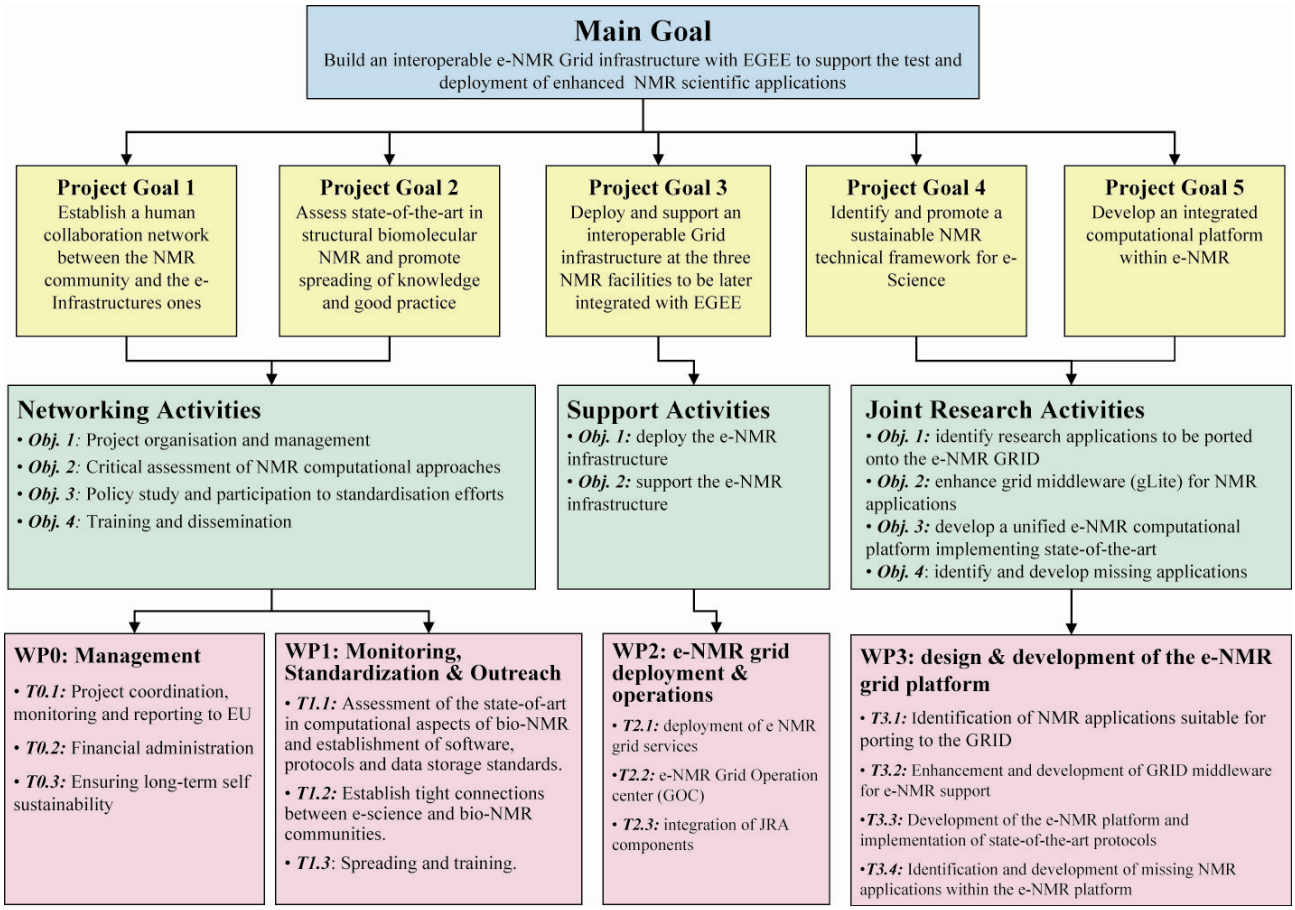
Access to advanced computing technologies is a top priority for the European bio-NMR community, for at least three reasons:

i) to be able to implement more thorough methods for data evaluation, analysis and interpretation, also exploiting at best semi- and fully automated procedures;

ii) to develop, test, and make available to the entire scientific community “golden standards” and best-practice computational approaches in order to enhance the quality of scientific research by the European bio-NMR community.

iii) to increase the impact of NMR in life sciences by speeding up the structural analysis work; a structure provides only the first and necessary step toward characterization of function and interactions

Achieving these general goals will contribute to fostering an efficient European integration in the computational domain.



Baseline of the project and performance/ research indicators

As already mentioned, the present project starts from the observation that the bio-NMR scientific community has already at its disposal a huge variety of software tools that typically cover one-two steps of the complicated workflow that brings from raw NMR data to interpreted data and publishable results. Due to the fragmented development of these programs, both from the geographical and the time point of view, there is poor interoperability between them. Consequently, individual research groups normally adopt their own ensemble of programs to cover the entire workflow, with varying degrees of human intervention needed and with their own choices of parameters for the same program. Both the need for human intervention and the fact that not all the values of the parameters are always made explicit (or, sometimes, known to the users themselves) make the results obtained from a given set of raw NMR data poorly reproducible by different research groups.

The baseline of the project is thus the absence of standardized, highly or fully automated protocols for elaboration of NMR data allowing:

1. minimal or no human intervention
2. full awareness and traceability of all choices of parameter sets preformed
3. straightforward comparison of alternate calculation schemes
4. easy implementation by individual research teams
5. online access

Regarding the NMR data themselves, some proposals for standards are already in place, mainly thanks to the efforts of the largest bio-NMR data repository, the BioMagResBank, but they do not cover all facets of bio-NMR spectroscopy nor are all correctly and thoroughly implemented in the most used software tools.

Starting from the above baseline, the present project aims at the:

- implementation of a grid-based platform enabling calculations with NMR data
- optimisation of the performance of programs in the grid environment (e.g. optimised calculation parameters)
- implementation of automated, transparent workflows involving different programs covering all steps from data to interpreted results
- development and comparison of alternate protocols
- optimisation of middleware for NMR applications
- development of tutorials and test cases
- involvement of the software developers' community in software assessment exercises using standardized, publicly available test data
- improvement of the state-of-the-art in NMR computational aspects, particularly with respect to user-friendliness to non-experts

Cumulatively, the above-mentioned results will allow the bio-NMR community to get access to a flexible, highly automated platform that will enable them to analyse their data with state-of-the-art, extensively tested methods. The "barrier" for the use of this platform will be reasonably low, thanks to the training opportunities provided and the demos and tutorials that will be made available.
The project will also contribute to improving of the state-of-the-art of the computational aspects of bio-NMR. This will be achieved thanks to the networking activities and, in particular, thanks both to the test data sets that will be provided to software developers and to the events devoted to software assessment that will be organized.

The following measures can be used as criteria and performance/ research indicators to evaluate the progress and success of the project during its duration:

- Deployment of Grid middleware
- Number of use-cases provided
- Number of programs implemented
- Number of programs embedded into automated protocols
- Number of test data sets released
- Number and attendance of assessment events
- Usage of e-NMR grid
- Number and attendance of training events
- Tutorials/demos made available

B.1.3 Networking Activities and associated work plan

B.1.3.1 Overall strategy and general description

Setting up an adequate human collaboration network is an essential factor for the success of this project. The setup involves connecting the staff researchers of NMR infrastructures, the bio-NMR user community, the teams involved in on-going NMR research and development efforts, and the researchers and technicians dealing with the e-Infrastructure. Indeed, any e-Infrastructure initiative like the present one should involve different partners, computing infrastructure centres, and researcher communities, including final users of NMR computing applications. The partnership of research teams in this proposal constitutes a core balanced with respect to the target scientific communities and will be able to guarantee, through its already established relationships, the widest dissemination and outreach of the project's activities. The e-Infrastructure will be based on the use of the GÉANT network, which offers a state-of-the-art framework, of the computing infrastructure of the European NMR research infrastructures, which are very close to final users in this research

field, and of the general EGEE e-Infrastructure. In particular, the expertise on grid operations of INFN, one of the European centres already active in EGEE, will support the bio-NMR emerging activities across the European e-Infrastructure.

In order to support the setup and operation of this human collaboration network, the following objectives and corresponding measurable indicators have been identified for the whole of Networking Activities foreseen in the project:

- *Objective 1.1:* Set up the structure of the collaboration network, with an adequate reflection at the Project Management level. As indicators of the progress achieved, a full description of the management structure will be put in place, including composition, meetings, actions on issues and results taken along each reporting period. **(WP0)**
- *Objective 1.2:* Organization of a community-wide critical assessment of NMR computational approaches. To carry out this assessment activity, experts in the field of computational procedures applied to bio-NMR will be brought together to test innovative and/or optimised approaches on defined NMR computational problems. These events will constitute a means to compare and assess the various methods. In this way, we will be able to identify software tools and protocols to be included in service and joint research activities within the present project. In addition, the assessment efforts will serve also the ultimate goal to keep the e-NMR platform always updated in terms of the latest developments and of state-of-the-art procedures. Such developments will be readily made available to users. The number and periodicity of events will be used as an indicator of the state of advancement of this objective. **(WP1)**
- *Objective 1.3:* study, design and adoption of policies accepted by all the partners, regarding the shared use of e-NMR. This objective will be tackled in close contact with the scientists and management of other large international e-science projects, and in particular with EGEE. The policies will be discussed and presented at relevant international forums, and discussed with national level authorities. Progress indicators include the participation in policy forums (including OGF, European Grid Conference, ICT and RI Concertation Meetings and especially eIRG), adoption of policies by the partnership and reflection at national and international level. In addition, we will seek coordination with other initiatives and networks relevant to the computational aspects of bio-NMR, such as CCPN and CCPN-Grid, EC-funded STREP projects, US structural genomics consortia making use of NMR and other relevant activities world-wide. **(WP1)**
- *Objective 1.4:* establish adequate support and dissemination mechanisms for the e-NMR infrastructure, provide training and promote the dissemination of good computational practices in bio-NMR. A Collaborative Forum will be established to support tight collaboration, including control and evaluation for this activity. The reference of collaborative actions (from mailing lists, portal repositories, videoconferencing, on-line demos and tutorials, to meetings and workshops), their adequacy, the participation by the community and, possibly, evaluation reports will be used to measure the status of this objective. Training courses for existing and potential users will be organized also to enhance the dissemination of best practices within the e-NMR user community, possibly in conjunction with EU-NMR user meetings or NMR-Life workshops. Software departments of the manufacturer of NMR machines will be consulted to contribute to the e-NMR platform in a mutual way. Thus, the long-term sustainability of e-NMR grid infrastructure in Europe will be secured as well as the NMR community – the potential customers for both – is addressed more rapidly. **(WP1)**

In order to achieve the above objectives, the networking activities are divided into two workpackages, namely *Project Management* (WP0) and *Monitoring, Standardisation and Outreach* (WP1).

Project Management, WP0 - WP0 is dedicated to the global coordination of the project, and to the project setup. The project will be managed by a Managing Committee (MC), formed by the Project Coordinator and by the coordinators of individual workpackages. In total, the MC will thus comprise four individuals. The representative from Beneficiary 1 – BMRZ- will be responsible for monitoring the development of workpackage activities with respect to what is planned, reviews reports from all actions undertaken, and is responsible for producing and submitting to the EC periodic progress reports for the whole project. (**T0.1, D0.1-0.3**). Beneficiary 1 will overview all legal and financial matters related to the project as well as for financial reporting to the EC. (**T0.2)**. The consortium will strive to guarantee self sustainability of e-NMR in the medium and long term. It is foreseen that approximately two FTEs will be needed after the end of the e-NMR project to guarantee the maintenance and update of the e-NMR platform. Therefore, the MC will seek contacts with national funding bodies, potential industrial partners, national e-infrastructure operators and other possible stakeholders, in order to devise means for technical and financial support and to set up collaborations. A crucial step in this direction, also in the light of the organization of the EGI project, is the integration of each of the three NMR nodes of eNMR into their respective National Grid Initiatives (NGIs). Partner 4 will be an important guide in this process. Interactions will also be established with the Life Sciences specific support center (SSC) (**T0.3).** Additional strategies to guarantee the long-term sustainability of e-NMR include:

1. seeking an agreement with the main manufacturer of NMR spectrometers to directly link their software to the e-NMR platform, after completion of the latter, so that the analysis of newly acquired data can be seamlessly launched on the platform. A revenue will be provided to the eNMR consortium as a share of each license sold by the manufacturer. It is foreseen that this activity by itself could sustain one FTE to work on the e-NMR platform
2. Organize training workshops on the tools of the e-NMR platform, access to which will be granted for free to academic partners, and charge a participation fee
3. License use of the e-NMR platform to pharmaceutical industries as a fee-for-access or allowing them to deploy a stand-alone version of the platform that can run locally for a flat fee. Additional revenues can be generated through contracts for training activities.
4. Explore the interest of SMEs in the pharmaceutical and biotech fields to make use of the e-NMR platform and identify a suitable corresponding payment scheme (e.g., on the basis of CPU usage)
5. Seek national funding, also through the NGIs.

Points 2-4 will be implemented after surveying the interest and demands of the user community as well as potential industrial stakeholders; the survey will be carried out as part of the service activities (DL2.8). The results of survey will be analysed to help direct future strategies for the sustainability of eNMR (DL2.9).

Monitoring, Standardisation and Outreach (WP1) – WP1 has a two-fold purpose: the first one is to build a global overview of computational approaches in bio-NMR to assess the state-of-the-art of the protocols currently used, also to identify commonalities; the second one is to build connections between the bio-NMR and e-science communities, which will be instrumental for the development of the e-NMR platform. This will require the definition of widely used standards for both software computational protocols and data storage. The combination of these actions will foster the adoption and use of the e-infrastructure by the European bio-NMR community, with a world-wide outreach.

This will be prompted through spreading and training initiatives. The latter will increase the awareness of the e-NMR initiative throughout the relevant scientific communities, as well as encourage the acceptance of the proposed standards.

Assessment of the state-of-the-art, and establishment of software protocols and data storage standards (**T1.1**)

The large amount of expertise that is already available within the three EU-NMR infrastructures represents only partly the worldwide community of NMR software users and developers. Each member of EU-NMR has implemented various computational protocols, some of which are being developed within EU-NMR. However, many computational developments do occur in other laboratories, often associated to EU-NMR through EC-funded projects or via the NMR-Life coordination action. To achieve the best assessment of what can be considered state-of-the-art in NMR computational approaches and monitor new developments in that area, it is best to directly involve the laboratories contributing to software development and their respective expert researchers, by setting up expert groups. These will be gathered to present, compare and discuss their approaches. Workshops will be organized to test innovative and/or optimised approaches on defined NMR computational problems (**D1.3** and **D1.4**). Already identified groups contributing NMR software will be contacted; in addition an open call will be made. The list of computational aspects that will be assessed will include (not exhaustive):

- automated peak-picking and signal recognition methods
- automated backbone and side-chain resonance assignment
- automated assignment and structure calculation approaches
- structure refinement approaches

…

In order to perform these assessments, we will create a growing database of both raw and interpreted NMR data (e-NMR database, **D1.1**). For this purpose, we will establish a common archiving format. A well organized and unified data structure and storage is a key to ensure transparency across all core facilities and will allow a unified scheme for data access and exchange within the e-NMR platform. A well defined structure is also required to define pathways within the e-NMR platform. Further, the model should allow gathering and storage of all relevant information from data collection to the final validated structures, enabling subsequent quality checks and facilitating data deposition in international public repositories. A well defined structure is also required for GRID computing both within the NMR large scale infrastructures (NMR-GRID) and, at a later stage, at the European level (e.g. EGEE-II) The three I3 NMR infrastructures will closely interact with the EMBL-EBIMSD to define the requirements of such a database. The e-NMR database will contain the past and current projects of the I3 core centres; the users of the EU-NMR infrastructures will be encouraged to make their data available as well. The e-NMR database will comprise raw NMR data (time domain data), processed spectra, resonance assignments and various restraint sets for different biomolecular systems including proteins, DNA/RNA and complexes thereof. A selection of the data in the e-NMR database will be accurately examined and annotated in order to serve as reference data sets for the aforementioned assessment activities, as well as to help NMR software developers in their work by providing a validated ensemble of test data sets (**D1.2**)

The e-NMR database will be provided freely to the NMR community via the NMR-Life web site (www.postgenomicnmr.net). Considering the biological relevance of most systems under study and the efforts that go into sample production, NMR measurements and data analysis, it will not be possible to provide data for testing in "blind" manner as is done in other critical assessment experiments such as CASP for protein structure prediction and CAPRI for the prediction of the structure of protein complexes. We will benefit here also from developments in data standardization

and book keeping from other related initiatives such as CCPN and the FP6 STREP project Extend-NMR.

Establish tight connections between the e-Science and bio-NMR communities (**T1.2**)

e-NMR is a project that complements the EGEE II I3 project. For a better synergy between these two initiatives, WP1 will specifically address their coordination, by transferring knowledge from one project to the other. Initially EGEE results need to be disseminated to the NMR community in order to have a precise view of what EGEE Grid platform can provide. In particular, NMR experts need to know what a Grid-platform can provide to them in terms of resources and support, to identify new possible scenarios and applications. ICT and software experts at the NMR infrastructures need to know the EGEE middleware development stage to evaluate the need of future enhancements specific for the NMR domain. Throughout the project a continuous interaction is useful both for guiding NMR activities in the EGEE view and to report preliminary results of NMR to the EGEE designers. INFN will have a leading role in the task. It will promote the project results among EGEE and e-Science community, supporting the project participation to the main conferences and workshops with posters, lectures and live demonstrations of the achieved results (Milestone M1.2: "Demonstration of e-NMR at e-Science event"). In particular, tight connections will be established with the EGEE Life Science Cluster, which periodically organises workshops with analogous e-Infrastructure projects, and is currently leading the process of defining a smooth transition from EGEE to EGI sustainability model with respect to the Life Sciences grid community, cooperating in strict contact with EGI_DS management, the EGI Policy Board, the NGIs and the future EGI Council.The results of e-NMR will be reported to the EGEE community. Through cross-dissemination WP1 will establish a tight link between NMR teams and EGEE members, enabling also the development of suitable policies for the integration of e-NMR into EGEE and for access to e-NMR services. This is an essential issue to increase the potential impact of the present project. A report on these activities will be made available (D1.6).

Within the above-mentioned framework, we will also work towards establishing policies for access, resource sharing and repositories within e-NMR. This objective will be tackled in close contact with the scientists and management of other large international e-Science projects through participation at policy forums, and in particular through interactions with the e-Infrastructure Reflection Group (e-IRG). The implementation of these policies will be based on tight contacts with national, regional and European e-Infrastructure agencies. The development of policies related to data storage and deposition will be facilitated by the EMBL-European Bioinformatics Institute.

Spreading and training (**T1.3**)

Training activities will take place at two levels: i) within the partnership, to train the bio-NMR partners in the implementation and management of the e-NMR Grid; ii) within the bio-NMR user community, to disseminate the use of the e-NMR platform. The first level of training activity will contribute to linking the life science community and the e-Science community, while the second level of training activity will increase the awareness of the European scientific community about e-infrastructures and contribute also to spreading best computational practices. The present training activities will result also in establishing a virtual NMR research community, thereby fostering collaboration among different research centres and making scientific processes more efficient.

The state of advancement of the e-NMR infrastructure will be presented at the yearly meetings of the EU-NMR users. Training events for users will also be organized on these occasions. Separate training events focusing on the use of the e-NMR platform and protocols for specific bio-NMR computing themes will also be held. The topics covered will include basic concepts of spectral assignment, automated backbone and side chain assignment routines, derivation of experimental constraints, biomolecular structure calculation, and structure model refinement. In a step-by-step procedure the participants in the course will be made familiar with a variety of techniques and will

apply these to standardised data sets, providing insight into the strengths, weaknesses and differences in the various approaches. A strong focus will be on the interpretation of the outcome of structure determination procedures in terms of validation of derived restraints and the accompanying structure models.

To complement the above activities and allow e-training of users, on-line tutorials and standardized data sets will be made freely available, describing the workflow for bio-NMR data analysis and interpretation and its implementation within e-NMR. The state of advancement and impact of spreading and training activities will be evaluated periodically by the MC (at least yearly). Evaluation results will be included in the annual project reports to the EC. A global overview of the achievements of the efforts deployed in the present task will be produced and made publicly available towards the end of the project (**D1.5**)

Finally, there will be a significant synergy of the activities within this task and within task 1.2 to disseminate the achievements of the e-NMR project also within the e-Science community. In this respect, partner 4 (INFN) will be proactive to present the results and involve the partnership of e-NMR on the occasion of large public events of the e-Science community, such as user forums, conferences on grid computing, schools and workshops, etc.

B.1.3.2 Timing of work packages and their components

	3	6	9	12	15	18	21	24	27	30	33	36
WP0: Project coordination	M0.1											M0.2
Task 0.1				D0.1				D0.2				D0.3
Task 0.2												
Task 0.3												
WP1: Monitoring, Standardisation and Outreach						M1.1 M1.2						
Task 1.1			D1.1		D1.2	D1.3				D1.4		
Task 1.2											D1.6	
Task 1.3								D1.7			D1.5	D1.8

M0.1: Organization of kick-off meeting
M0.2: Organization of concluding general meeting
M1.1: Report on the results of the first assessment workshop based on the standardized test data sets constituting D1.2
M1.2: Demonstration of e-NMR at e-Science event
D0.1, D0.2, D0.3: Organization of annual meeting and preparation of report
D1.1 e-NMR database model
D1.2 Test data sets for software development and testing
D1.3 Assessment of state-of-the-art: first workshop for assessing the state-of-the art in computational aspects of bio-NMR
D1.4 Assessment of state-of-the-art: second workshop for assessing the state-of-the art in computational aspects of bio-NMR
D1.5 Training of bio-NMR users: training sessions for bio-NMR users at the yearly meeting of the EU-NMR large facility and global evaluation of all training activities performed
D1.6 Report on connections established within Task 1.2
D1.7 and D1.8: annual reports on the dissemination of e-NMR results in the bio-NMR and e-Science communities

B.1.3.3 Work package list /overview

Work package No	Work package title	Type of activity	Lead participant No	Lead participant short name	Person months	Start month	End month
0	**Management**	**MGT**	**1**	**BMRZ**	**24**	**0**	**36**
1	**Monitoring, standardization & outreach**	**COORD**	**2**	**CIRMMP**	**64**	**0**	**36**
	TOTAL				**88**		

B.1.3.4 Work package descriptions

Workpackage number	0	**Start date or starting event:**		Month 0		
Workpackage title	Management					
Activity type	MGT					
Participant number	**1**	2	3	4		
Participant short name	**BMRZ**	CIRMMP	BCBR	INFN		
Person-months per participant	**9**	9	3	3		

Objectives

O0.1 Setup the structure of the collaboration network with an adequate reflection at the project management level

Description of work

T0.1 Produce and submit to EC periodic progress reports (Partner 1). Prepare and update regularly Project Presentations.

T0.2 Overview financial matters and financial reporting to EC (Partners 1, 2)

T0.3 Initiate contacts with national funding bodies, potential industrial partners, national e-infrastructure operators and other possible stakeholders in order to guarantee medium to long term self sustainability of e-NMR (Partners 1-4)

Deliverables

D0.1 First year project report

D0.2 Second year project report

D0.3 Third year project report

Workpackage number	1	**Start date or starting event:**			Month 0	
Workpackage title	Monitoring, standardization and outreach					
Activity type	COORD					
Participant number	1	**2**	3	4	5	SubA
Participant short name	BMRZ	**CIRMMP**	BCBR	INFN	EMBL	CNRS
Person-months per participant	12	**14**	12	18	6	1
Participant number	SUBB					
Participant short name	S-NMR					
Person-months per participant	2					

Objectives

O1.1 Critical assessment of NMR computational approaches

O1.2 Policy study and participation to standardization efforts

O1.3 Training and dissemination, also through participation in concertation meetings

Description of work

T1.1 Assessment of the state-of-the art in computational aspects of bio-NMR and establishment of software, protocols and data storage standards (Partner 1,2,3, 5 Sub-contractors A, B)

T1.2 Establish tight connections between e-science and bio-NMR communities (Partners 1-5)

T1.3 Spreading and training, both within the partnership to train the bio-NMR partners in the implementation and management of the e-NMR Grid and within the bio-NMR user community to disseminate the use of the e-NMR platform

Deliverables

D1.1 e-NMR database model (month 9)
D1.2 Test data sets: complete sets of 10 raw and interpreted NMR data made for software development and testing (month 15)
D1.3 Assessment of state-of-the-art: first workshop for assessing the state-of-the art in computational aspects of bio-NMR (month 18)
D1.4 Assessment of state-of-the-art: second workshop for assessing the state-of-the art in computational aspects of bio-NMR (month 30)
D1.5 Training of bio-NMR users: training sessions for bio-NMR users at the yearly meeting of the EU-NMR large facility and global evaluation of all training activities performed (month 33)
D1.6 Report on connections established within Task 1.2 (month 33)
D1.7 and D1.8: annual reports on the dissemination of e-NMR results in the bio-NMR and e-Science communities (month 24 and month 36)

B.1.3.5 Efforts for the full duration of the project

Workpackage[1]	WP0	WP1	TOTAL per Beneficiary
BMRZ	**9**	12	21
CIRMMP	9	**14**	23
BCBR	3	12	15
INFN	3	18	21
EMBL		6	6
CNSR (SubA)		1	1
S-NMR (SubB)		2	2
TOTAL	24	63	89

Efforts are given in person-months. The WP leader is in bold.

B.1.3.6 List of milestones

Milest. no	**Milestone name**	**WP no.**	**Expected date**	**Means of verification**
M0.1	Kick-off meeting	WP0	Month 3	Meeting report available
M1.1	First assessment of the state-of-the-art	WP1	Month 18	Workshop report available
M1.2	Demonstration of e-NMR at e-Science event	WP1	Month 18	Demonstration report available
M0.2	Concluding general meeting	All	Month 36	Final project report available

[1] Please indicate in the table the number of person months over the whole duration for the planned work , for each work package by each beneficiary

B.1.4 Service Activities and associated work plan

B.1.4.1 Overall strategy and general description

Our main objective with respect to Service is to deploy and support a new NMR e-Infrastructure that will be integrated with EGEE at a final stage (**D2.4** and **D2.5**). The e-NMR infrastructure will consist of shared resources both in terms of the infrastructure itself (network, computational resources, software) and of a new virtual research framework for the bio-NMR community. As a basic premise, the proposal will start from a well known grid framework, the one currently in operation in the EGEE project The INFN partner has already a very in depth knowledge of the deployed middleware services, a successful experience of supporting new user communities and a practical experience in the installation and day to day operation of the current EGEE middleware gLite. This framework will be implemented at the NMR infrastructures.

Regarding the e-Infrastructure utilization, the NMR collaboration is already well established, thanks to Trans-national access projects, which have been in place since 1994, as well as to a number of collaborative RTD initiatives (e.g., in FP6, NDDP, UPMAN, bioDNP, …). However, additional efforts are needed to benefit from the possibilities of the grid technology in order to provide the users with a powerful and efficient e-Infrastructure. Critical areas, such as data access, usage of commercial licences in the Grid, security and proper definition of NMR workflows have been identified.

A challenging part of the present service activity will be the setup and operation of the NMR Virtual Organization(s) supporting the activity of bio-NMR scientific community. The establishment of a VO dedicated to bio-NMR researchers will constitute an unprecedented step for this scientific community towards the adoption of common computational protocols and the establishment of a thorough interoperability among different laboratories.

The deployment of the infrastructure and of the framework enabling the use of the applications selected in the Networking and Joint Research activities will take place by exploiting as much as possible the middleware and tools already available. The necessity to introduce some adaptations in the middleware will probably arise and will be evaluated when this occurs in order to guarantee that the NMR communities will be able to fully profit of the EGEE e-Infrastructure.

The e–NMR grid will be initially open to the researchers operating at the facilities of the three NMR partners (internal staff and visiting users). After the successful installation of the Grid operation center (**T2.2**) and of (significant parts of) the e-NMR platform (WP3), access will be granted also to EU-NMR users from their own premises and then to other scientists. We therefore foresee to be able to open the infrastructure to the broader scientific community extending beyond the users of EU-NMR before the last year of this project.

Deployment of the e-NMR grid infrastructure (**T2.1**)

The deployment of the e-NMR grid infrastructure will be the first step enabling the provision of service (**D2.1**). In order to implement from the very beginning full compatibility with the EGEE middleware, an agreement will be reached with the NMR sites to adopt the Grid Security Infrastructure (GSI) based on the PKI mechanism and X509 certificates already in use within EGEE. Every user must obtain a personal International Grid Trust Federation compliant certificate and register with enmr.eu VO. Access to the services of the e-NMR portal will be granted only to users having a valid certificate.This will involve the establishment of Registration Authorities (RAs) trusted by the corresponding national Certification Authorities (CAs) for those sites and research institutes which are not yet enabled to issue EUGridPMA trusted digital certificates. This Service Activity task will provide guidelines and support to establish the RAs where needed.

Elaborating on the outcomes of Networking and Joint Research activities, we will identify the middleware and grid services to be deployed at the NMR sites, according to the requirements of the selected applications. The grid services will comprise the Computing Element (CE) and Storage Element (SE). In parallel, we will establish the NMR Virtual Organization. This will require the deployment and maintenance of a set of central grid services such as: Virtual Organization Membership Service (VOMS); Workload Management System (WMS); LCG File Catalogue (LFC).

These activities will continue throughout the whole lifetime of the project in order to keep the e-NMR grid infrastructure aligned with the most recent middleware releases, to upgrade it with the new tools and components produced by the Joint Research activities, and to possibly include new NMR sites willing to join the e-NMR infrastructure.

Within this task, after a suitable time since the e-NMR Grid is fully operational, we will survey the user community as well as identify and address potential users outside the academic bio-NMR community, in order to evaluate their interest or demands in terms of fee-based services. The interest in having training course held upon payment of a registration fee will also be explored. The results of the survey will be available by month 24 (D2.8) and will be analysed in order to provide input to WP0 (D2.9) to help direct subsequent actions aimed at guaranteeing the sustainability of the e-NMR platform after the end of the present project.

e-NMR Grid Operation Centre (GOC) (**T2.2**)

This task aims to set up an operational structure for the management, control and support of the e-NMR infrastructure (**D2.2** and **D2.6**). This will be constituted by a set of tools to proactively monitoring the e-Infrastructure, checking the status and accounting the use of the central grid services and of the remote computing and storage resources. These tools will be chosen among the ones already in use within EGEE (e.g. GridICE, DGAS, SAM, GStat, RTM), in order to make the integration of the operations of the e-NMR infrastructure into EGEE smoother. In particular, GridICE and DGAS have been developed by the INFN partners of the present project, and thus constitute a preferable choice, if suitable for their intended use in this project. An important parameter to measure the success of the project via the GOC will be the number of members enrolled in the eNMR VO, including both researchers who port their application(s) to the e-NMR GRID as well as new users of the e-NMR platform. A reasonable target for the project is to attain 120 members at its end, including also non-European members.

Of particular importance is the support system for both grid end-users and resource centres administrators. This system typically consists in a complete portal containing a large number of features, such as: documentation, knowledge base, wiki, trouble ticketing system, mailing lists, SVN repository, middleware repository, bug tracking system, links to the monitoring tools, etc. The ticketing system will be eventually interfaced with the Global Grid User Support (GGUS), the official support system used by EGEE and OSG grid infrastructures. A mechanism will be implemented to handle operational problems and answer support requests.

A certain level of coordination with the relevant EGEE Services Activities is needed in order to leverage as much as possible on already existing operational procedures and tools, and in view of the full integration of the e-NMR infrastructure inside EGEE.

Integration of JRA components (**T2.3**)

Since the beginning the grid middleware deployed on the infrastructure will be based on the gLite distribution, produced by the EGEE-II project. It is expected that new software components

developed to enable the e-NRM platform will be delivered by the Joint Research Activities during the lifetime of the project.
The aim of this task is to test and certify the components developed in the Joint Research Activities (**D2.3** and **D2.7**). In particular, their integration with the current tools and middleware deployed on the e-NMR infrastructure will be taken care of. The scalability and the robustness of the new components have in fact to be proved before integrating them in the production service, in order to avoid unexpected behaviour leading to a degradation of the overall performance or in the worst case to the malfunctioning of the service. Furthermore, automatic deployment tools and procedures have to be put in place in order to make easier the work at the resource centres when upgrading their site. This activity is typically carried out with the help of a pre-production test-bed, i.e. a set of dedicated computing resources distributed around a few sites among the project's partners. After successfully passing the quality tests a middleware release including the new components is ready to be deployed smoothly on the production grid infrastructure, without interrupting the service provisioning. The whole activity may require a dedicated software repository and a bug tracking system to be setup at the GOC, and is of course strongly linked with the JRA.

B.1.4.2 Timing of work packages and their components

Timing of the workpackages and of their components

	3	6	9	12	15	18	21	24	27	30	33		36
WP2: e-NMR Grid deployment and operation				M2.1				M2.2	D2.4				D2.5
Task 2.1				D2.1				D2.8	D2.9				
Task 2.2						D2.2					D2.6		
Task 2.3							D2.3						D2.7

M2.1: Grid infrastructure operational (Month 12)
M2.2: First integration of JRA developed components (Month 24)

D2.1 e-NMR grid infrastructure operational
D2.2 Assessment of the e-NMR infrastructure and GOC report
D2.3 New components' integration and deployment report
D2.4 Report on e-NMR infrastructure state-of-the-art toward integration with EGEE
D2.5 Final report on e-NMR infrastructure usage and integration with EGEE (update of D2.4)
D2.6 Second Assessment of the e-NMR infrastructure and GOC report (iteration of D2.2)
D2.7 New components' integration and deployment updated report (iteration of D2.3)
D2.8 Survey of the interest and demands of the user community as well as of potential industrial stakeholders with respect to the provision of fee-based services and/or training activities
D2.9 Evaluation of the results of the survey and implementation of consequent actions

B.1.4.3 Work package list /overview

Work package list

Work package No	Work package title	Type of activity	Lead participant No	Lead participant short name	Person months	Start month	End month
2	**e-NMR Grid deployment & operations**	**SVC**	**4**	**INFN**	**52**	**0**	**36**
	TOTAL				**52**		

B.1.4.4 Work package descriptions

Workpackage description

Workpackage number	2	**Start date or starting event:**		Month 0		
Workpackage title	e-NMR Grid deployment and operation					
Activity type	SVC					
Participant number	1	2	3	**4**		
Participant short name	BMRZ	CIRMMP	BCBR	**INFN**		
Person-months per participant	10	10	10	22		

Objectives

O2.1 Deployment of the e-NMR infrastructure.

O2.2 Support of the e-NMR infrastructure

Description of work

T2.1 Deployment at the three NMR sites of the e-NMR Grid services including security and registration services, ensuring full compatibility with the EGEE middleware (Partners 1,2,3, 4)

T2.2 Establish the e-NMR Grid Operation Center (GOC) for management, control and support of the e-NMR infrastructure (Partner 4)

T2.3 Certify and integrate the software components developed in the Joint Research Activities with the middleware deployed in the e-NMR infrastructure (Partners 1,2,3, 4)

Deliverables

D2.1 e-NMR grid infrastructure operational (document describing the hardware, the middleware and the grid services deployed at the NMR sites, and the GOC supporting the infrastructure) (Month 12)

D2.2 Assessment of the e-NMR infrastructure and GOC report (document evaluating the use of the infrastructure in terms of number of users, number of jobs run, CPU time used, storage capacity used, integration level with EGEE, effectiveness of GOC procedures and users' support) (Month 18)

D2.3 New components' integration and deployment report (document listing the programming libraries, APIs, tools, applications released by the JR activities and describing the procedure adopted for their integration and deployment on the infrastructure) (Month 21)

D2.4 Report on e-NMR infrastructure state-of-the-art toward integration with EGEE (Month 27)
D2.5 Final report on e-NMR infrastructure usage and integration with EGEE (Month 36)

D2.6 Second Assessment of the e-NMR infrastructure and GOC report (Month 33, iteration of D2.2)

D2.7 New components' integration and deployment updated report (Month 36, iteration of D2.3)

D2.8 Survey of the interest and demands of the user community as well as of potential industrial stakeholders with respect to the provision of fee-based services and/or training activities (Month 24)
D2.9 Evaluation of the results of the survey and implementation of consequent actions (Month 27)

B.1.4.5 Efforts for the full duration of the project

Workpackage	WP2	TOTAL per Beneficiary
BMRZ	10	10
CIRMMP	10	10
BCBR	10	10
INFN	**22**	**22**
EMBL		
CNRS (SubA)		
S-NMR (SubB)		
TOTAL	52	52

Efforts are given in person-months. The WP leader is in bold.

B.1.4.6 List of milestones

List of milestones

Milest. No.	**Milestone name**	**WP no.**	**Expected date**	**Means of verification**
M2.1	e-NMR grid	WP2	Month 12	Demonstration of e-NMR grid
M2.2	First integration of JRA developed components	WP2	Month 24	Demonstration of NMR grid computing over the three Large Scale infrastructure sites

B.1.5 Joint Research Activities and associated work plan

B.1.5.1 Overall strategy and general description

The joint research activities of this project are combined in a single workpackage aimed at developing the uniformed and integrated e-NMR platform that forms the core of the e-NMR infrastructure. We will also establish a long-term sustainable technical framework to support the research of the scientific communities directly or indirectly involved in bio-NMR within the general European e-infrastructure. These communities range from the scientists, typically chemists or physicists, developing new experimental protocols and schemes for the application of high resolution NMR to biological systems to the biologists exploiting NMR spectroscopy to obtain further information into their systems of choice.

In this activity therefore we will address the entire range of issues from those associated to the implementation of individual applications to those related to the development of a technical general framework and its long term support. This effort will go along with effective dissemination (carried out as part of the Networking activities) throughout both the NMR community and the e-science community of the problems encountered, of the ways out found, of the aspects still awaiting resolution. By providing an integrated and unified NMR computational platform at all sites, geographical and infrastructural differences will disappear, reinforcing the supra European character of the EU-NMR I3.

The foreseen objectives of the joint research activities and the corresponding indicators are:

- *Objective 3.1:* identify research applications to be ported to the e-NMR Grid, which will be enhanced thanks to the use of a shared e-Infrastructure. The identification of research applications will determine the exact needs in terms of middleware deployment and development (see also next objective). Progress will be measured by the list of enhanced applications, and their impact and relevance in the various fields of bio-NMR research.
- *Objective 3.2:* develop, integrate and support the enhancement to the gLite stack allowing the identified NMR applications to exploit the e-Infrastructures across Europe and possibly in the world. Progress will be measured by the list of needs for enhancements identified and by the corresponding implementation list.
- *Objective 3.3:* Develop an integrated, harmonized biomolecular e-NMR platform covering all computational aspects of biomolecular NMR. Adapt/optimize the e-NMR platform by implementing state-of-the-art (from WP1). The initial version of the e-NMR platform will support a limited number of methods and protocols that have been well-tested and are in current use at the various centres in order to allow its early deployment. However, as the project evolves, the platform will be extended to provide additional tools and protocols selected from the workshops addressing the critical assessment of NMR computational methods (within WP2). Progress will be measured by the number of applications and protocols released.
- *Objective 3.4:* Identify missing parts in the e-NMR platform that would facilitate the overall integration and develop them if required. While NMR structural computational tools are well developed for the study of proteins in particular, developments are needed when it comes to nucleic acids (DNA/RNA). Our objective is therefore to identify and develop when needed missing components.

<u>Identification of NMR research applications suitable for porting to the Grid (**T3.1**)</u>

The first step will be to identify the research applications most relevant for the e-NMR infrastructure and prioritise them for inclusion in the platform for service provision. This activity will thus address the evaluation and the optimisation of the performance of the selected applications

in the context of the e-NMR Infrastructure. In parallel, we will also address the evaluation and adoption of the best technological resources for the specific needs of the applications selected. In this way, an optimised framework for the exploitation of the computational resources of the e-Infrastructure will be defined and adopted. Particular attention will be devoted to the problem of data security, as this has both organisational and technical concerns which will impact both on the Networking and Service activities. From this activity we will obtain a list of applications that are suitable for porting to the Grid (**M3.1**) and indications on the computational infrastructure needed to fully sustain them.

The identification of significant use-cases is essential for the correct definition of system requirements. The very first part of the present task will thus focus on the identification of a defined set of use-cases to allow the evaluation of the applications mentioned in the preceding paragraph (**D3.1**). Use-cases should be significant in terms of usage scenario, involvement of organizations (research centres, data providers, etc.), interaction of human and technological resources. They should cover a wide set of possible applications of Grid technologies to bio-NMR applications. Moreover, some of them should be existing scenarios that can be enhanced by the use of advanced technologies, while others should describe new scenarios made possible by the existence of a Grid infrastructure. Taking into account all these elements, use-cases can drive the study of the research strategies for the design and development of the infrastructure. Indeed based on these preliminary analyses, the ICT experts may define the system requirements for the research infrastructure supporting NMR applications.

In the first half of the second year, the present task will address the extension of the list of applications and of the set of use-cases. The latter will allow not only evaluation of the applications added to the initial list, but also more thorough evaluation of the applications selected first. Part of these efforts will be useful also for task 1.1.

Enhancement of Grid middleware for e-NMR support (**T3.2**)

In addition to the aforementioned issues, part of the joint research activities will be aimed to identify the enhancements required by the selected NMR applications that need to be integrated and shared in the European NMR Research Infrastructure and in the general European e-Infrastructure. As a basic premise, we will start from the well known grid framework, the gLite stack currently in operation in the EGEE project, where the INFN partner has already a very in depth knowledge.

Based on NMR system requirements and the EGEE existing and planned features, we will define the research developments required to enhance EGEE for fully enabling NMR applications. The EGEE middleware deployed and configured on partner sites as part of the Service activity will be useful to test and evaluate the current release of EGEE in a bio-NMR setting. Using the e-NMR infrastructure, a study for the NMR framework integration with the EGEE middleware will be performed. This will allow us to evaluate which parts of the middleware need to be enhanced. Some minor software development to help the gridification of e-NMR applications can also be performed within this task, e.g. code to enable GSI access to e-NMR portals for grid job submission, adaptation of existing tools to manage huge grid workloads, etc. However, no new grid elements nor enhanced versions of existing gLite grid elements are planned to be released by the project.

The first results will be summarized at Month 18 in a Request for Enhancement (RFE) to the EGEE community, reporting the main features that need to be improved and the fields where greatest research efforts should be planned (**D3.3**). This deliverable will explicitly take into consideration also the security issues that can arise from the interaction with users and industrial stakeholders. In particular, the relevant communities will be surveyed for their requirements with respect to data security. At the same time, gLite developers will be asked to provide feedback on all the technical aspects addressed in D3.3, as well as to point out relevant new solutions that might have become

available during the course of the project. Consequently, an updated version of the RFE document (**D3.7**) taking into account the feedback provided from gLite middleware developers will be delivered at the end of the task planned at Month 30. Security issues will addressed in a separate deliverable, before the end of the task, in order to start exploring potential solutions before te end of the project (D3.9), as these aspects may be relevant for the involvement of industrial stakeolders in the e-NMR community. The INFN partner will ensure that the gLite will be available to the e-NMR consortium / user community after the end of the present project. In this frame, INFN will guarantee that the specific requirements of the e-NMR consortium / user community will be either included in the standard gLite distribution or will make a suitable patch available for each gLite release.

Development of the e-NMR platform and implementation of state-of-the-art protocols (**T3.3**)

Within the European bio-NMR infrastructures as a whole, a large amount of expertise is available covering both most computational aspects of NMR as well as a very extensive range of different types of systems studied (proteins, nucleic acids and their complexes, paramagnetic systems, ligand-receptor adducts, …). Each bio-NMR infrastructure, however, has its own expertise area. We will bring together all of this knowledge for the EU-NMR user community in a uniform computational platform, which will be deployed at all sites, implementing state-of-the-art protocols and tools (**D3.6**). In this way, the specific expertise of the partners will be spread over all EU-NMR research infrastructures. In addition, expertise from other leading groups contributing to NMR software development will be considered in the development of the e-NMR platform. This part of the work will be tightly connected to the Networking activities.

The platform will allow easy communication between various software programs as well as interfacing to databases. As shown in Figure 2, the e-NMR platform will cover essentially all

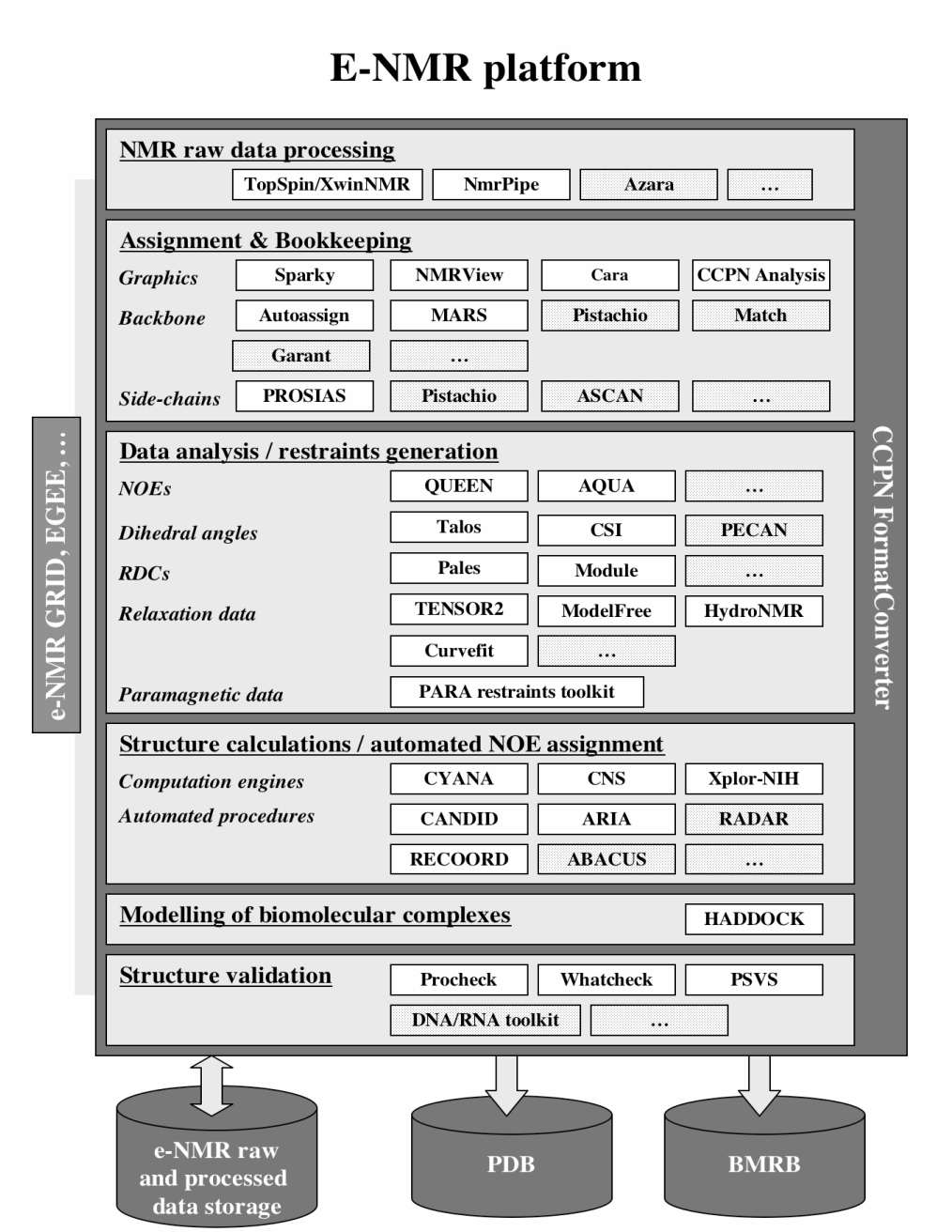


Figure 2: Schematic representation of the e-NMR platform to be implemented in this project. The initial version of the platform will support only the software indicated in white boxes. As the project evolves and depending on the results of the assessment of the state-of-the-art in NMR computational approaches (WP2), new software will be added to the platform (grey shaded boxes).

computational aspects of biomolecular NMR, from data processing to structure validation. Next to providing transparent path and data exchange between the various software, the platform will include well tested and documented tutorial examples describing the entire process from raw NMR data to final structures (see also Networking activities). In first instance, the platform will be limited to software commonly used at the three NMR facilities for which ample expertise is available (white boxes in Figure 2). As the project evolves and as a result of the assessment of NMR computational methods in the Networking activities (WP1), new tools and procedures will be implemented.

An important step will be to define a common software directory structure in which all required NMR programs will be installed. Using a well-defined structure will allow easy deployment and maintenance of the platform at all sites and possibly at other interested sites or at user sites in the future. The software structure will also facilitate the development of streamlined protocols.

More in detail, we will install all programs identified in the course of Task 3.1 in the e-NMR grid. Experts will then run the use-cases of Task 3.1 to define the optimal parameters to individually run the software tools installed also as a function of parameters such as number of available constraints (for structure calculations) or number of relaxation rates measured (for quantification of protein flexibility). After this first phase, we will aim at implementing protocols that allow seamless communication between different programs. These protocols will cover all aspects involved in typical bio-NMR applications, from raw data to interpreted results. The protocols will be designed to required minimal or no human intervention. Validation of the protocols will be performed against the standardized data sets of Task 1.1. The protocols will be progressively expanded to include different options for software selections at each individual step while keeping best performance. Evaluation of individual software tools within different protocols may be performed if needed. Users will be offered the possibility to run different protocols in parallel and compare their results. New software tools highlighted within networking activities or appearing in the scientific literature will be included as needed.

The e-NMR platform will interface with the e-NMR database (D1.1) in order to store the data generated, which will allow users to re-evaluate their data should new procedures and/or better analysis tools become available.

The e-NMR platform will be integrated by on-line tutorials and demos, which will allow users to learn easily the scientific, experimental and technological concepts that are behind the tools integrated in the platform. In addition, the tutorials and demos will greatly facilitate the use of the e-NMR platform for users that have received little or no training within the context of the present project, as well as remind trained user of the various possibilities offered by the platform. These training materials will be produced and made available all along the process of construction of the e-NMR platform (**D3.5**). A through review of all materials will be carried out close to the end of the project.

Identification and development of missing NMR applications within the e-NMR platform (**T3.4**)

An important part of the study of biomolecular interaction and function involves DNA and RNA structures and their complexes with proteins. However, most of the focus and research in biomolecular structure determination has been on proteins, resulting in more than 39,000 protein structures at the Protein Data Bank. In contrast, only ~1,700 structures of protein-nucleic acids complexes, ~1000 DNA structures and ~550 RNA structures are available from the Protein Data Bank. The reasons for this huge discrepancy are found in a variety of experimental difficulties, from obtaining samples to structure determination techniques. Consequently, so far little focus has been on the investigation and validation of DNA/RNA structures, whereas this becomes an increasingly important aspect of understanding biological structure and function. The currently available software for investigating nucleic acid structures provides mainly descriptive tools, whereas true structure validation tools are largely lacking (**D3.2**). Within the e-NMR consortium knowledge is available on both the experimental side of nucleic acid structure determination, as well as on the protein structure validation. We will therefore investigate the feasibility of translating well-established concepts from protein structure validation to DNA/RNA structure validation (**D3.4**). The current growth of the number DNA/RNA structures at the Protein Data Bank should allow for building a reference database of high quality including well-validated nucleic acid structures, which in turn can be used to compare with and validate newly determined DNA/RNA structures and complexes thereof with proteins.

In addition, we will evaluate the needs of users for new methods of analysis of NMR data, and will

also monitor the appearance of new methodologies and applications of bio-NMR that (may) require development of software tools to be then included in the e-NMR platform.

B.1.5.2 Timing of work packages and their components

	3	6	9	12	15	18	21	24	27	30	33	36
WP3: Design and development of the e-NMR Grid platform												D3.6
Task 3.1		D3.1		M3.1								
Task 3.2				M3.2		D3.3		D3.9		D3.7		
Task 3.3			M3.3			D3.5						D3.8
Task 3.4					D3.2					D3.4		

D3.1: Release of first set of use-cases
D3.2: Overview of existing validation software able to deal with nucleic acids
D3.3: Request for Enhancements of gLite to support bio-NMR applications
D3.4: Structure validation suite for DNA/RNA
D3.5: First set of online tutorials (3) describing in details the optimal use of the e-NMR platform for solution NMR
D3.6: Integrated e-NMR platform
D3.7: Updated RFE of gLite to support bio-NMR applications
D3.8: Final set of online tutorials describing in details the optimal use of the e-NMR platform for solution NMR (iteration of D3.5)
D3.9: Analysis of security issues brought about by the user community and by industrial stakeholders

M3.1: List of NMR applications suitable for GRID porting
M3.2: Recommendations (software requirements) for NMR software developers to allow easy porting to the e-NMR GRID
M3.3: Description of e-NMR software installation structure

B.1.5.3 Work package list /overview

Work package list

Work package No	Work package title	Type of activity	Lead participant No	Lead participant short name	Person months	Start month	End month
3	**Design & development of the e-NMR Grid platform**	**RTD**	**3**	**BCBR**	**184**	**0**	**36**
	TOTAL				**184**		

B.1.5.4 Work package descriptions

Workpackage description

Workpackage number	3	**Start date or starting event:**			Month 0	
Workpackage title	Design & development of the e-NMR Grid platform					
Activity type	RTD					
Participant number	1	2	**3**	4	5	SubA
Participant short name	BMRZ	CIRMMP	**BCBR**	INFN	EMBL	CNRS
Person-months per participant	46	46	46	30	12	1
Participant number	SubB					
Participant short name	S-NMR					
Person-months per participant	4					

Objectives

O3.1 Identify research applications to be ported onto the e-NMR Grid

O3.2 Enhance grid middleware (gLite) for NMR applications

O3.3 Develop a unified e-NMR computational platform implementing state-of-the-art

O3.4 Identify and develop missing components of the e-NMR platform

Description of work

T3.1 Identification of NMR applications suitable for porting to the Grid (Partners 1-5)

T3.2 Enhancement of Grid middleware for e-NMR support (Partners 1-4)

T3.3 Development of the e-NMR computational platform and implementation of state-of-the-art protocols (Partners 1,2,3,5, Sub-contractor A)

T3.4 Identification and development of missing NMR applications with the e-NMR platform with emphasis given to smooth communication between the components and structure validation

Deliverables

D3.1: Release of first set of use-cases (month 6)

D3.2: Overview of existing validation software able to deal with nucleic acids (month 15)

D3.3: Request for Enhancements of gLite to support bio-NMR applications (month 18)

D3.4: Structure validation suite for DNA/RNA (month 30)

D3.5: First set of online tutorials (3) describing in details the optimal use of the e-NMR platform for solution NMR (month 18). The material will be continuously improved/updated.

D3.6: Integrated e-NMR platform (month 36)

D3.7: Updated RFE of gLite to support bio-NMR applications (month 30)

D3.8: Final set of online tutorials describing in details the optimal use of the e-NMR platform for solution NMR (iteration of D3.5) (month 36)

D3.9: Analysis of security issues brought about by the user community and by industrial stakeholders (month 24)

B.1.5.5 Efforts for the full duration of the project

Workpackage	WP3	TOTAL per Beneficiary
BMRZ	46	46
CIRMMP	46	46
BCBR	**46**	**46**
INFN	30	30
EMBL	12	12
CNRS (SubA)	1	1
S-NMR (SubB)	4	4
TOTAL	185	185

Efforts are given in person-months. The WP leader is in bold.

B.1.5.6 List of milestones

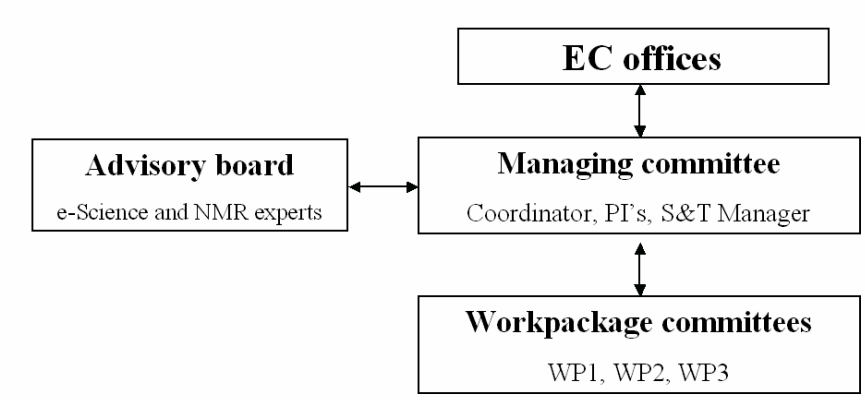
List of milestones

Milest. no.	**Milestone name**	**WP no**	**Expected date**	**Means of verification**
M3.1	GRIDable NMR applications	WP3	Month 12	List of suited applications
M3.2	e-NMR software recommendation	WP3	Month 12	Report with e-NMR recommendations
M3.3	e-NMR software directory structure	WP3	Month 9	e-NMR software directory structure on CD/DVD

B2. Implementation

B.2.1 Management structure and procedures

The project will be managed in a two-tiered hierarchical way. The lower level is the workpackage one. The workpackages involve networking activities (WP1), service activities (WP2) and the joint research activities (WP3). Each workpackage is coordinated by the PI of one of the project partners. The workpackage coordinator chairs the workpackage committee in which all partners responsible for the activity are represented. The workpackage committee is responsible in the broadest sense for performing the work in the workpackage, such that the contracted deliverables and milestones are achieved in time. The top-level management body is the Management Council (MC), which is chaired by the Project Coordination, Prof. H. Schwalbe (Partner 1 PI). The MC will be responsible for the overall scientific and strategic management, as well as the definition of the policies of the present I3 also in relation to the other I3, networking, and coordinative initiatives that are already on-going or will be launched at the time of start of the present project and that have relevance for the activities proposed here. The MC is also in charge of the setup, organization and operation of the bio-NMR Virtual Organization that will support the activity of the bio-NMR scientific community within e-NMR. The MC will be assisted in its decision-making process by an Advisory Board, which will comprise also members not directly involved in the present partnership. The Board will be chaired by Dr. Robert Jones, Project Director of EGEE-II. Other personalities, both from the e-Infrastructure and the bio-NMR communities, have been contacted to join the Board.
All partners are represented in the MC by their PI's. To cope with the scientific and technological complexity of the present project, which also requires efforts and visits to ensure effective harmonization at all the partner sites, it appears useful to extend the composition of the MC in order to include a Scientific and Technological Manager of the project. Therefore, the MC will comprise a fifth member (Prof. Antonio Rosato, Partner 2) who will act as S&T Manager. The S&T Manager will be responsible for executing the decisions of the MC and will take care of day-by-day operation of the project. MC meetings are organized at least semi-annually; videoconferencing may substitute for physical meetings. A minimum of one physical meeting of all MC members per year will take place.
The Advisory Board will be composed by a maximum of five members, including its Chair. The exact composition of the Board will be defined during the negotiation phase of the project. The Advisory Board will meet on a yearly basis or electronically whenever needed on occasions when the MC deems necessary to receive indications on specific matters. The Advisory Board will be particularly involved in providing indications to the MC for the aspects concerning e-Science policies and strategies to addressing the computing aspects of the present e-NMR project.
A scheme of the management structure is provided below.



As mentioned, the PI of Partner 1 is the Project Coordinator and Chair of the MC. Partner 2 has been appointed with the responsibility for all administrative and financial duties, e.g. with respect to the following central functions:

a. Organisation and preparation of minutes of the MC;
b. Transmission of any information connected with e-NMR to all parties concerned;
c. Signalling any contingencies in the execution of the project to the MC;
d. Financial administration of the project.

To fulfil these duties, a part-time project management specialist and a part-time secretary will be hired. They will take care of the general management, of assisting the S&T Manager in the preparation of the overall project reports on the basis of the contributions from the various workpackages, and of the I3 administration, including the preparation of financial reports to be submitted for approval to the MC.

The Coordinator and the S&T manager ensure that all relevant information (progress reports and financial data) is distributed up and down the chain from the individual scientist executing part of a task that should lead to the achievement of deliverables or milestones, through the workpackage committees and the MC, to the EU appointed scientific and administrative project officers seconded to the project.
Communication in the e-NMR project will proceed first of all through the project WEB site, which contains all relevant information. This information is accessible to the partners on a need to know basis. The second mode of communication consists of workshops, which cater to a more or less specialized audience, progress meetings, annual user meetings and symposia. At the annual meeting, to which representatives of external institutes and industries may be invited, the annual report is presented and decisions are made with respect to the plans for the next period. In order to make the information exchange process as efficient as possible, it is our intention to organize the various meetings for the TA, NA and JRA subprojects as much as possible on the occasion of meetings for the whole e-NMR project. Internal progress reports for e-NMR will be made available semi-annually. The partners are obliged to submit timely their contributions for this report. Milestones and deliverables are released immediately after they have been reached.

At the start of the project and at the end of each project year the partners submit through the various management levels their detailed plans for the execution of the work. These plans should be based on the state-of-the-art at the start of the project and on reports of concept progress for the preceding phase, the results obtained, the input from any external expert(s) that might have been contacted, and the results obtained at locations outside the partnership. The proposals should contain verifiable deliverables and milestones, and a good planning, inclusive of the alternate routes that will be taken in case of contingencies in the execution, the planned manpower etc. The MC will decide on the realization of these plans and allocate the funding for their execution. The activity of the Advisory Board will be particularly useful under several aspects: i) to make the MC aware of technological progress that may have an impact on the present project; ii) to help the MC evaluate different solutions that exploit newly available results (from within or outside the partnership); iii) to improve planning. In this way, optimal use can be made from the available resources.
The control of the various tasks is based on the informal tracking of the project through communications by the researchers in charge with the workpackage coordinators and the screening of material submitted to the project WEB site. The management support team will prepare a semi-annual evaluation of the project based on material that has been submitted to the project database and short progress reports submitted to the coordinator. Because of the close coupling of planning and control in an annual cycle, the project can be controlled closely and adjusted to developing needs and can deal efficiently with contingencies.
The present management structure is closely matched to the scale and the complexity of this project. The size of the MC enables thorough discussion among all the members and at the same time

guarantees a quick decision-making process. The MC involves the PI's, who also act as coordinators of the various workpackages. This ensures that the MC becomes promptly aware of any deviation from the planned project development, starting with any relevant problem that should occur already at the level of individual tasks. All members have a very long tradition in the management of scientific collaborations of the size of, or even larger than, the present one. The activity of the S&T manager will allow smaller problems, even involving administrative aspects, to be solved promptly, minimizing the involvement of the entire MC only to those issues (scientific, technological, administrative, financial) that require deeper analysis and a coordinated response by the partnership. Obstacles or deviations from the plans are thus addressed at three different levels: by the workpackage coordinator at the level of individual tasks, with the help of the S&T manager at a higher level or when some administrative aspects may need attention, and finally at the level of the MC when elaborate actions and/or strategic decisions are needed. Periodic reviews of the state of the project carried out by the S&T manager and submitted to the MC will ensure that no potential bottlenecks are overlooked.

B.2.2 Beneficiaries

Participant N. 1

Center for Biomolecular Magnetic Resonance (BMRZ), Goethe-University Frankfurt, Germany
Principal investigator: Harald Schwalbe

Expertise: The BMRZ currently has the greatest number of NMR spectrometers of any EU NMR Infrastructure. Cryoprobe technology, available on four different instruments ranging from 600 MHz to 950 MHz. BMRZ is coordinating the I3 European network of NMR infrastructures (EU-NMR). BMRZ operates an internationally recognized Large Scale NMR Facility supporting biomolecular NMR research in the EU, being funded since 1995 under HCM, TMR and IHP. The research staff provide access, training and expertise to users who have determined many protein structures of ligand binding and catalytic domains including GPCRs, kinases and phosphatases. The institutes operate facilities for chemical synthesis, molecular biological and biochemical studies of the investigated systems. In particular, fermenters, synthesisers and instruments for recombinant techniques and chemical synthesis are available for the production of stable isotope enriched proteins including cell-free production of protein samples. This site has developed a group of researchers entirely devoted to NMR in ligand design. Currently eight senior researchers are directly involved in related protein preparation, NMR screening and NMR structure. These researchers develop their own pulse programs, screening protocols and software and will focus on overcoming the technical problems of structural biological studies on membrane proteins.

Role: The main tasks of the BMRZ partner will be: *a*) coordinating the scientific program of the e-NMR infrastructure *b*) providing protocols and workflow for protein structure determaintion from NMR data, organizing the dialogue between scientists on establishing best standads. *c*) deployment and management of the e-NMR Grid hardware and software at the Frankfurt site (T2.1).

Human resources: BMRZ includes three groups involved in this project, the groups headed by Prof. Harald Schwalbe, by Prof. Volker Dötsch and by Prof. Peter Güntert. The research groups of BMRZ have demonstrated internationally recognized excellence in NMR based structural biology and include key players such as Prof. Schwalbe, recipient of the Gerhard Hess prize of the German Research Foundation (DFG) and Karl Winnacker prize, a Pew Scholar in Biomedical Sciences and Alfred P. Sloan fellow. Prof. Schwalbe is one of the three directors of the recently funded Cluster of Excellence: macromolecular complexes (DFG). Prof. Volker Dötsch has been appointed to Frankfurt from the University of San Francisco. He has pioneered the development of NMR methods to study proteins in their native environment (in-cell NMR). Prof. Peter Güntert has recently been appointed from RIKEN Genomic Sciences Center (Japan) to the University of Frankfurt. This was possible because Peter Güntert has been awarded the Lichtenberg professorship, one of the most prestigious German prizes, by the Volkswagen Foundation. He has developed the suite of NMR software packages DIANA-DYANA-CYANA, one of the two principal computational methods that are used world-wide by the NMR-community to calculate NMR structures.

The NMR groups are also involved in an international Graduate Research School "Structure and Function of Biological Membranes" initiated by the Max Planck Institutes of Biophysics and Brain Research. Members of the BMRZ are involved in three "Sonderforschungsbereiche" of the Deutsche Forschungsgemeinschaft: "Molecular Bioenergetics", "RNA-Ligand Interactions" and "Functional Membrane Proteomics".

Staff who may be engaged in the transfer of knowledge include Drs. M. Betz, F. Bernhard, F. Löhr, C. Richter, J. Wirmer, V.Jaravine, A.Bagaria and H. Jonker.

Participant N. 2

Interuniversity Consortium for Magnetic Resonance on Metalloproteins (CIRMMP)
Principal investigator: Prof. Ivano Bertini

Expertise: CIRMMP is the managing institution of the Magnetic Resonance Center of the University of Florence (CERM). CERM has a long-term, excellent knowledge and tradition in the characterization of structure and dynamics of proteins in solution through NMR. It has continuously contributed to the improvement of NMR-based techniques for macromolecular structure determination, and is currently involved in European Structural Proteomics projects. It is equipped with the most advanced instruments for spectroscopic and biophysical characterization: 11 NMR nuclear magnetic resonance spectrometers, which span the largest range of existing magnetic fields, from very low (relaxometers) to very high ones (900 MHz). Most instruments are equipped with cryoprobes and several have optimized probes for paramagnetic systems, or automatic sample changers for large numbers of samples. Also the highest wide bore magnet spectrometer for solid state NMR (850 MHz) is available. The group also has access to EPR spectrometers and mass spectrometers (ESI- and MALDI-TOF), and apparatus for structural crystallography. More than 100 protein structures were solved in the laboratory using NMR and/or X-ray spectroscopies. The research group has considerable experience in structural biology, particularly in structural and dynamic protein characterisation in solution through NMR and unique expertise in paramagnetic systems, having characterised several metalloproteins and developed software tools to include paramagnetism-based restraints in solution structure calculations. CIRMMP also has extensive experience in the development of new computational and experimental high-throughput screening methods and drug candidate optimisation, as well as in the application of developed methodologies to selected disease targets. Finally, CIRMMP has a long-standing expertise in the application of bioinformatics to the analysis of genome sequences and of metalloproteins.
CIRMMP has participated in several EC-funded projects. At present CIRMMP is coordinating the Coordination Action “Focusing NMR on the machinery of Life” (NMR-Life) and is involved in the I3 EU-NMR as one of the research infrastructures providing transnational access to high-field instrumentation. CIRMMP is also involved in several RTD projects in Italy and in Europe, through STREPs.

Role: The main tasks of the CIRMMP partner will be: *a*) collaborating with the coordinator for the scientific and the day-to-day management of the project; *b*) taking care of all financial matters relevant for the project and collaborating to the preparation of periodic reports; *c*) assessment of the state-of-the-art in computational aspects of bio-NMR, establishment of standards and contribution to training (tasks T1.1, T1.3); *d*) establishment of connections with the e-Science community (T1.2); *e*) deployment and management of the e-NMR Grid hardware and software at the Florence site (T2.1); *f*) identification of applications of applications suitable for porting to the GRID (T3.1) and development of an unified e-NMR computational platform (T3.3). CIRMMP will be in charge of WP1.

Human resources: **Ivano Bertini**, Professor and Director of CIRMMP, has pursued technological advancements in solution structure determination, particularly for paramagnetic metalloproteins, and established a molecular biology department for protein expression. He has published more than 600 articles and solved more than 100 protein structures. **Claudio Luchinat**, Professor, is the author of more than 400 publications in scientific journals of international renown and of four books on NMR and relaxation. His research interests are in the structural and dynamic characterization of metalloproteins and metalloenzymes, and drug discovery. **Antonio Rosato**, Associate Professor, is the author of more than 60 publications. His research work focuses on structural biology, on the development of software tools to assist the analysis of NMR data and on the bioinformatic analysis of genomes and proteins. **Antonio Rosato** will act as scientific

and technological manager of the project.**Andrea Giachetti**, PhD, is an expert in informatics at CIRMMP and has worked on software development for NMR-based studies of protein-ligand interactions. Two expert postdocs will be hired for the project.

Participant N. 3

Bijvoet Center for Biomolecular Research, Universiteit Utrecht (BCBR)
Principal investigator: Dr. Alexandre M.J.J. Bonvin

Expertise. The Bijvoet Center for Biomolecular Research is a joint research institute of Utrecht University and the Chemical Sciences Research Council (NWO-CW) whose research focuses on mechanisms of molecular recognition and interactions. The research of the NMR group aims at obtaining fundamental insight in biological processes using the methods of high-resolution NMR spectroscopy. A major aim is to understand biomolecular recognition processes in terms of detailed 3D structures and dynamics. A further aim is the development of new methodology supporting the structural biology research. Novel methods have been and are being developed to make the structure determination process more efficient, to validate protein structures derived by NMR, to study dynamical processes of biomolecules, and to study transient states in photoreceptor molecules. The NMR group possesses a wide range of expertise in fields such as molecular biology, protein expression, development of new NMR methodology and its application to biomolecular systems, and development and application of computational and molecular modelling approaches. In the past, relaxation matrix-based structure refinement methods and validation software have been developed (IRMA, DINOSAUR, AQUA). A recent major achievement is the development of the protein-protein docking approach HADDOCK (High Ambiguity Driven protein-protein DOCKing) that can make use of biochemical and or biophysical information such as NMR titration data or mutagenesis data to drive the docking process. More than 350 research groups are now using this approach worldwide. The NMR group has and is participating to several EU-funded programs and networks contributing to the standardization of NMR computational approaches (FP5 RTD projects NMRQUAL and FIND, FP6 STREP projects Extend-NMR and NDDP). It also belongs to the large EU-funded Structural Genomics consortium SPINE-II complexes.

Role. The main tasks of the BCBR partner will be: *a*) assessment of the state-of-the-art in computational aspects of bio-NMR, establishment of standards and contribution to training (tasks T1.1, T1.3), see Workpackage-Description tables); *b*) deployment and management of the e-NMR Grid hardware and software at the Utrecht site (T2.1); *c*) development of a unified e-NMR computational platform (T3.3); *d*) identification of applications of applications suitable for porting to the GRID (T3.1) and of missing applications within the platform (T3.4). The BCBR partner will be in charge of WP3.

Human resources. **Alexandre M.J.J. Bonvin** (PI) is Associate Professor within the NMR Spectroscopy research group of the Bijvoet Center for Biomolecular Research, UU. His expertise resides in computational methodology development and application, both for NMR and molecular modelling projects. **Rolf Boelens**, Professor of Chemistry is leading the NMR group and is an expert on both experimental and computational aspects of NMR. Johan van der Zwan, scientific programmer, is in charge of computational infrastructure within the NMR group. Sjoerd de Vries is a PhD student working on the development of tools for data-driven docking. To be appointed: two postdoctoral fellows for each two years.

Participant N. 4

National Institute of Nuclear Physics, Padua (INFN)
Principal investigator: Prof. Mirco Mazzucato

Expertise: INFN (http://www.infn.it/), is a public governmental research organization, which promotes, co-ordinates and funds nuclear and high-energy physics related researches. It is organized in 4 National Laboratories, 19 Departments (called Sections located in major Universities) and 11 Local Groups (see *http://www.infn.it/mappa.php*). INFN staff research personnel amounts to more than 1500 peoples with an equivalent number of associates from University and other Scientific National Institutes. INFN has a

considerable experience on high performance distributed computing. Already in 1998 INFN deployed a Wide Area CONDOR Pool distributed all over Italy (see *http://www.infn.it/condor*); at the end of 1999 INFN launched the INFN-GRID project (*http://www.infn.it/grid*), to evaluate/develop the use of GRID technologies in facing the stringent computing requirements of the High Energy incoming LHC experiments at CERN. Since 2001 INFN has played a major role in the EU DataGrid and DataTAG projects, the CERN based LCG and WLCG projects, the National Grid Projects GRID.IT (*http://server11.infn.it/firb-grid*) and LIBI, and more recently the EGEE and EGEE II and the grid infrastructure extension projects like EUmedGRID, EUChinaGRID, EU-IndiaGrid and EELA, GridCC and BioinfoGRID The INFN contribution to these projects comprises the setup of the INFN Production GRID, with more then 3000 CPU's deployed in more than 20 sites, the development and reengineering of the grid Middleware, in particular of the Workload Management service, the Virtual Organization Membership Service (VOMS), the Glue Schema, the new Web Service CE implementation with CREAM and CEMON, the Grid Accounting service DGAS, the GRID Monitoring service GridICE, some new components related to grid policies GPBOX and the SRM interface to parallel file systems (Storm) together with the dissemination ad training activities.

Role: INFN main task in e-NMR will be the participation and coordination of the SA activities that consists in supporting the deployment and the operations of the grid infrastructure for the NMR community shared with EGEE. INFN will also take part to the NA and JRA activities. Experience to carry out this task comes from the role played in the various worldwide Grid projects quoted above and from the role that INFN has in the EGEE projects. Moreover the experience coming from the coordination of the INFN Production Grid (the INFNGRID) will be of utmost importance in running and coordinating the infrastructure needed by e-NMR. Furthermore INFN has an important role in the development of Grid software and has extensive knowledge of the problems related to testing and certification activities.

Human resources. Prof. Mirco Mazzucato: INFN Director of Research. CNAF (INFN Advanced Computing Centre) Director, since 2004. INFN Grid Project Project Manager since 2000. INFN delegate and member of the Management Board in the (W)LCG, World Computing Grid for LHC Computing Grid Project at CERN since 2002 (~ 6000 physics in Europe, USA and Asia) Member of the Project Management Board and coordinator of the Italian Federation in the FP6 European project EGEE (and its successor EGEE II) MIUR Italian delegate in the e-Infrastructure Reflection Group. Dr. Marco Verlato, INFN Senior Technologist. Ph.D. in Physics in 1998, experience in HEP data analysis and farm computing, involved in EU grid projects since 2001 (EU-Datagrid, EU-DataTAG and EGEE) contributing on several tasks: pure middleware development, middleware deployment on the local computing farm, development of interfaces between the LHC experiments application software and the grid middleware, interoperability between European grid and Globus-based US grid.

Participant N. 5

European Molecular Biology Laboratory - European Bioinformatics Institute, Hinxton, UK
Principal investigator: Dr. Kim Henrick

Expertise The European Bioinformatics Institute (EMBL-EBI) is the flagship European public science institute in the field of bioinformatics. It is an outstation of the European Molecular Biology Laboratory (EMBL) headquartered in Heidelberg. The Macromolecular Structure Database group (MSD) was established to ensure that all aspects of 3D structure data are placed in the public domain and served to the scientific community. The MSD is heavily involved in a biological structural network that spans wet and dry science, generates standards and collects, orders and disseminates information. The MSD macromolecular structure relational database (http://www.ebi.ac.uk/msd) is designed to be a single access point for protein and nucleic acid structures and related information. The database is derived from Protein Data Bank (PDB) entries. Relational database technologies are used in a comprehensive cleaning procedure to ensure data uniformity across the whole archive. The expertise provided by this subcontractor with respect to the design of biological databases and their integration within software platforms that are made available to broad scientific communities is unique in Europe. Through our PDB activities we are a partner with the

BioMagResBank (BMRB) and the NMR toolkit produced at the EBI is in production use by the BMRB to process depositions of NMR related data.

EMBL-EBI also has extensive experience in handling NMR data on a large scale, both with respect to derived data (shifts, constraints) and raw NMR data (as part of the Databank for Experimental NMR spectra, DEN). It has collaborated closely with the CCPN project since its inception, and uses the CCPN framework to store and handle NMR data. The MSD has been a partner is a number of other international collaboration projects including previous EU projects, NMRQUAL and ExtendNMR and plays a significant role in the NMR world wide community. We have been involved in the RECOORD project, a database that contains recalculated structures for protein entries from the Protein Data Bank (PDB) where the restraint lists used for the recalculation originate from the Filtered REstraints Database (FRED) at the BioMagResBank. We are currently funded by the BBSRC (UK) to improve NMR aspects of depositions to the PDB.

Role: The main tasks of EMBL-EBI will be: a) to establish a common data storage structure at the various partner sites and create a central storage repository for the e-NMR generated data; b) assist in establishing test data sets for software development and testing and c) provide a link to the CCPN project.

Human resources: **Dr K. Henrick** is the head of the European Macromolecular Structure Database group with 20 years experience in X-ray diffraction for structural biology and data base technologies. **Dr Wim Vranken** is responsible within the MSD group for all NMR aspects of protein structure and database models. EMBL-EBI has established connections with CCP4 (X-ray), CCPN (NMR) and the RCSB (Protein Data Bank), and has accumulated wide experience of developing database applications for use in protein structure analysis, and in handling NMR data on a large scale.

B.2.3 Consortium as a whole

The partners of the e-NMR project represent the three core members of the existing European NMR Integrated Infrastructure Initiative EU-NR (partners 1, 2 and 3) plus an expert group in e-Science (partner 4). Together they provide a large amount of expertise necessary to achieve the objectives of the e-NMR project. For some specific tasks, however, the expertise of outside groups will be used. These will be sub-contracted to the project. The choice of S-NMR and CNSR-CRMN as a sub-contractors is motivated by the fact that they will only provide expertise for some specific aspects of the networking and research activities of the project. In contrast, the INFN and EMBL (EBI) groups (partners 4 and 5), which are not current partners in EU-NMR, play a crucial role in the project by providing the necessary e-Science and research/networking expertise, being involved in all aspects of the project including services and providing the link to the EGEE project.
The assignment of the coordination of workpackages to the partners reflects their scientific as well as managerial skills. Indeed, the three NMR partners are differentiated by their scientific expertise in the various facets of bio-NMR as well as by their previous experience in the coordination and management of European projects. Partner 1 has a very solid and successful experience in the coordination of large EC-funded projects, and is currently coordinating the EU-NMR initiative. Partner 1 has also a long tradition in the development of innovative bio-NMR experiments, and has tight links to pharmaceutical and biotechnological companies thanks to his expertise in the investigation of protein-ligand interactions. Partner 2 has been involved since the mid-nineties in the managing of European coordination and networking projects, such as Concerted Actions (framework programme 5), Infrastructure Cooperation Networks (framework programme 6) and is presently managing the Coordination Action NMR-Life. Partner 2 is active in the development of innovative methods for the application of bio-NMR to study the structure and intermolecular interactions of proteins in solution and in the solid state. The latter is a rapidly developing aspect of bio-NMR, which has gained visibility in the last five/ten years. Partner 3 is a world-wide leader in the design and development of computational tools for the various aspects of bio-NMR, ranging from the calculation of the structure of proteins, nucleic acids and macromolecular adducts based on

NMR data to the assessment of the quality of the results obtained from NMR data. Finally, Partner 4 is highly skilled in the development of Grid technologies and their application to the various fields of science, and has a long tradition in the implementation and management of Grid-based services.

The European Molecular Biology Laboratory (EMBL) acting through its outstation the European Bioinformatic Institute (EMBL-EBI) will help in the establishment of common database standards (**T1.1**) and in ensuring a smooth communication between the software components of the e-NMR platform (**T3.4**). CNRS-CRMN-Lyon in the person of Dr. Torsten Herrmann will be subcontracted by partner 3 to contribute to the development of the eNMR platform (WP3) by porting his unique automated assignment and structure calculation software suite to the grid and to provide expertise for the software evaluation rounds under WP1. Spronk NMR Consultancy (SubB) will be sub-contracted by partner 2 to provide specific training courses (**T1.3**) and contribute to the development of DNA/RNA validation tools (**T3.4**). Note that by involving Spronk NMR Consultancy, a new SME based in Lithuania, the e-NMR project is contributing to the enhancement of the industrial base within Europe thereby strengthening the industrial expertise in computational aspects of NMR and structural biology in Europe.

i) Subcontracting:

Sub-contractor A: CNRS-CRMN Lyon
Principal investigator: Dr. Torsten Herrmann

Task	*Person Mo.*	*Role*
WP1	1	Monitoring, standardization & outreach (e-NMR database model.)
WP3	1	Design & development of the e-NMR Platform

Role: The subgrantee will contribute to the e-NMR objective of implementing state-of-the-art computational methods by making the software ATNOS/CANDID (joint property of Torsten Herrmann and ETH Zurich, Switzerland) available for large-scale grid and cluster-computing calculations. To achieve this goal, it will be required to adapt the software layer needed to run ATNOS/CANDID on a Grid resource and to verify the software compatibility with given e-NMR grid standards (WP3). Additionally, the subgrantee will perform a series of test calculations on the available computer resources to ensure proper performance of ATNOS/CANDID.
Furthermore, the subgrantee will contribute to the assessment of state-of-the-art computational methods by providing the best ATNOS/CANDID protocols available (WP1).

Reason for subcontracting: Dr. Torsten Herrmann has established a strong record in biocomputational NMR and became one among a handful of leading experts in the field of automated NMR data analysis and protein structure determination. His research introduced novel computational-theoretical concepts for NMR data analysis leading to decisively improved accuracy and efficiency of NMR structure determination of proteins. In particular, the programs ATNOS and/or CANDID for automated collection of conformational restraints have evolved into standard processing tools for NMR data analysis worldwide. Very recently, with the presentation of two novel mathematical approaches for automated backbone and side-chain assignment, MATCH and ASCAN, a comprehensive data analysis pipeline for fully automated protein NMR structure determination without any human intervention got implemented into the novel UNIO protocol.

UNIO NMR data analysis enables to determine protein structures including NMR measurement time within one week.
In 2008, he joined the European Centre for High-Field NMR (CRMN) in Lyon as Senior Research Associate (CDD Chercheur haut niveau CNRS) and received the qualification for professorship in theoretical chemistry by the French Ministry of National Education. The European Center for High Field NMR in Lyon, a research unit of the University of Lyon, is affiliated with the CNRS (French National Center for Scientific Research) and the Ecole Normale Supérieure de Lyon, one of the leading Grandes Ecoles in France, and as such a centre of excellence for teaching and research in all fields of science. Notably the CRMN hosts state-of-the-art equipment (the first 1 GHz NMR spectrometer is expected to be installed in summer 2009) and is the home to research groups actively involved in developing NMR spectroscopy in structural biology, metabolomics research, and materials applications. The Center also operates the Rhône-Alpes Large-scale Facility for high-field NMR (RALFNMR) which is a Lyon-based European I3 (Integrated Infrastructure Initiative) NMR facility funded by the EU, the CNRS, the French Ministry for Education and Research, the Atomic Energy Commission (CEA), and the Région Rhône-Alpes (www.ralf-nmr.fr). RALFNMR provides a high-level expertise in both solution and solid-state NMR techniques for the investigation of molecular structure, dynamics, and function. The NMR Center provides an ideal platform for gathering a wide-range of data sets and interacting with its broad and diverse user base. This network of some of Europe's best scientists will definitively impact on the progress of the work package.

Estimation of costs: We will allocate 10,000 € to Sub-contractor A taken from the budget of Partner 2.

Sub-contractor B: Spronk NMR Consultancy
Principal investigator: Dr. Chris Spronk

Task	*Person Months*	*Role*
WP1	2	Provide training and develop tutorials
WP3	4	Contribute to the development of structure validation tools, in particular for DNA/RNA

Role: The main tasks of the Spronk NMR Consultancy sub-contractor (SubB) are: *a)* to provide expertise in the field of validation of NMR structures (Task 3.4). *b)* to provide high level theoretical and practical training programmes covering various aspects from NMR data interpretation to structure calculation, validation and interpretation (Task 1.3).

Reason for subcontracting: Spronk NMR Consultancy is a small enterprise that offers a variety of services and expertise in biomolecular NMR structure determination and validation to commercial and academic institutions worldwide. The company is the only one in the world with a core specialization in computational biomolecular NMR services. As such, the company like no other can assess and address the needs and concerns of commercial parties regarding the use of the e-NMR technology, infrastructure and grid-computations. A major advantage of the company is that it combines high scientific standards with a practical hands-on approach. Therefore it is an ideal partner to test the technology in the development phase, address problems from the users' perspective and develop training programs not only from a theoretical point of view but, more importantly, aimed at practical use of the e-NMR technology by non-experts.
The knowledge and renown scientific experience in the field that is present in the company is unique in its kind, with a scientific track record that spans the whole trajectory from biological experiments, NMR structure determination and refinement up to structure validation. In particular,

the company has a unique expertise on the validation and the development of automated methods for the refinement of NMR structures of proteins (>10 publications with a total of >250 citations) that is not available in research groups but is of critical importance and has a high added value for the results obtained from e-NMR applications. This specific scientific expertise, especially in the field of structure validation, makes the company the best choice available worldwide for aiding in the development of validation technology specifically for NMR structures of RNA and DNA. Currently this technology is still rudimentary and there is great need for its development to allow future progress in the field.
In addition to performing the main tasks T1.3 and T3.4, the companies' hands-on experience in handling data sets, and identifying and solving associated problems, provides indispensable knowledge for the process of curating and validating test data sets (T1.1: D1.2). Reliable test data are crucial for the success of developing state of the art and best practice procedures (T1.1: D1.3 and D1.4). Also here Spronk NMR Consultancy will contribute its experience, since it not only applies state of the art NMR technology, but also continuously develops state of the art itself and through its international network of close *collaborations* with companies and leading academic scientists.

Finally, Spronk NMR Consultancy is based in Lithuania, and is thus uniquely placed to further the results of the present project in Eastern Europe.

Estimation of costs: We plan to allocate 62,000 € to Sub-contractor B, corresponding to 9.3% of the total expenses of Partner 2.

ii) Other countries:
Not applicable.

iii) Additional partners:
Not applicable.

iv) Third parties
Not applicable.

B.2.4 Resources to be committed

Resources that will complement the EC contribution
Partner 1 will contribute to the project with its currently available computational resources comprising a Linux cluster and a storage RAID 5 array of several Terabytes. One additional Linux cluster will be made available at the project start; the current technical staff for computer will care for the maintenance of the whole computational infrastructure.

Partner 2 will commit to the project its currently available computational resources, consisting of around 50 workstations and two clusters of Linux machines. Storage space will also be made available, consisting of RAID 5 arrays with a total capacity of several Tbyte. Technical workforce for the maintenance of the computational infrastructure will also be provided.

Current computational resources of Partner 3 include 30 Linux PC workstations and two Beowulf clusters (36 32bits processor system and 48 64bits opteron cluster). The group distributes the docking program HADDOCK (http://www.nmr.chem.uu.nl/haddock) and provides web services for the prediction of biomolecular interactions (WHISCY server http://www.nmr.chem.uu.nl/whiscy). The indirect costs of Utrecht University contain all costs for the general scientific infrastructure. On this basis, Utrecht University will contribute to the required computational GRID infrastructure.

INFN will made available to the project the GILDA t-infrastructure (https://gilda.ct.infn.it/) for the training about gLite middleware. It consists of more than 20 computing and storage resource sites in 3 continents, of a fake Certification Authority and a VO for demonstrations and tutorial purposes, and its own set of Grid services. Furthermore, for the Grid Operation Centre activities, part of the servers and tools of the INFN Production Grid (http://grid-it.cnaf.infn.it/) will be shared and made available to the e-NMR Infrastructure.

Description of resources and budget

The above-mentioned resources will be used to constitute the initial pilot e-NMR Grid. The pilot e-NMR Grid will be based on the cluster calculation facilities of the NMR infrastructures, over which the middleware will be deployed. No costs for the computer hardware are charged to the present project; expenses for maintenance and update of the computing facilities are already in the current budget of the NMR infrastructures. The EC contribution will cover the costs of the personnel involved in the set-up and operation of the e-NMR Grid and of the services associated with it, as well as the costs of the personnel involved in the JRA activities. Overall, the planned budget will allow each partner to employ two units of personnel, at different levels of expertise, for the project. In the case of Partners 1-3, one of these two units will be exclusively involved in the JRA activity (Type of activity: RTD), while the second unit will be partly dedicated to JRA's and partly to support the Service activity (Type of activity: Other). The JRA activity of Partners 1-3 will be dedicated to the identification, extension, optimization of software tools and protocols to be implemented in the e-NMR platform. Analogously, Partner 4 will dedicate part of his personnel resources to JRA activities, focusing mainly on the development of middleware for e-NMR, and partly to support the deployment of the middleware at the three NMR sites for Service provision. In total, we plan a total budget for personnel costs (excluding subcontracting) of 1,010,000 € and 110,000 € for JRA and Service activities, respectively. The partitioning of personnel costs between these two types of activities reflects the relative effort in terms of person-months, as described in section 1 for the various workpackages.

In more detail, for the JRA we have planned four tasks. Task 3.1 **(**Identification of NMR applications suitable for porting to the Grid) is carried out mainly by the NMR partners, with help from the INFN partner. Personnel costs of 300,000 € will be devoted to this task. Task 3.2 (Enhancement and development of Grid middleware for e-NMR support) will be aimed to identify, develop and support the enhancements required by the identified NMR applications that need to be integrated and shared in the e-Infrastructure. An enhancement of gLite for NMR will be produced. We foresee an expense for personnel, mainly provided by the INFN partner, of 180,000 €. In task 3.3 (Enhancement and development of Grid middleware for e-NMR support), we will develop, test, install, maintain the e-NMR platform, which will cover essentially all computational aspects of biomolecular NMR, from data processing to structure validation. Related training materials and demos will also be created. This task, which is the main responsibility of partners 1-3 will involve 410,000 € of personnel costs. Finally, 120,000 € of personnel costs will be devoted to task 3.4 (Identification and development of missing NMR applications), which will focus on the development and integration into e-NMR of new software tools that cover aspects of bio-NMR for which there are no satisfactory computational tools, as well as novel aspects of bio-NMR that might emerge during the lifetime of the present project.

This activity will address the evaluation and the optimisation of the performance of the selected applications in the context of the e-NMR Infrastructure. In parallel, it will address the evaluation and adoption of the best technological resources for the needs of the applications selected. Finally, the identification of significant use-cases will be tackled.

The resources devoted to the networking activities (Type of activity: Coordination) will be used to cover travel and subsistence expenses for the people attending the training events and the workshops devoted to the assessment of the state-of-the-art, as well as all other costs associated with organizational aspects. Networking activities will also involve the active participation of the senior scientists involved in the project to spreading the results and participation in policy-making events. Travel and subsistence expenses associated with these activities will also be charged to the project. More in detail, we have reserved 5,000 € for the kick-off meeting of the project, and 10,000 € for each of the three annual meetings of the project (MC meeting and Advisory Board meeting, with the possibility of inviting other experts or representatives of user organizations, other networks or other European projects relevant to e-NMR), for a total over the entire project of 35,000 €. As described in section 1, we plan to organize two workshops for the assessment of the state-of-the-art in bio-NMR computational approaches involving the major expert groups all across Europe and possibly also overseas. A cost of 46,000 € per workshop is planned (total per two workshops: 92,000 €). Participation in events relevant for policy-making, setting of e-Science standards and other similar activities of strategic relevance for the project is assumed to involve costs of 18,000 € for the whole project. 51,000 € per year will be devoted to spreading and training activities (total 102,000 € over the entire project). The latter activities may involve training workshops held at the NMR infrastructures or at meetings organized within the frame of EU-NMR or NMR-Life, or sessions aimed at spreading knowledge of e-NMR held at the users' sites or at events organized by users (also potential) of the bio-NMR infrastructures.
A total cost of 248,120 €, corresponding to an EC contribution of 186,000 € is planned for subcontracting. The role and importance of the contribution of subcontractors has been described in section 2.3.Finally, WP0 is devoted to management. The overall EC contribution for management activities, including indirect costs is 122,000 €. This is partly allocated to the partners in order to cover their costs for contributing to project reporting.

Sustainability
The present project will create a computational infrastructure that will be of outstanding importance to users of the European NMR research infrastructures, as it will provide them with a state-of-the-art platform for the analysis and interpretation of experimental data. e-NMR will thus become central to complement the service provided in terms of measurement time as offered by EU-NMR. Thus, the NMR research infrastructures will devote part of their budget that is currently used for maintenance and refurbishing of the infrastructure itself to sustain the future operation of e-NMR project. The e-NMR consortium will also strive to establish a collaboration with the manufacturers of NMR machines. If the NMR machine operation software interfaces with the e-NMR platform to perform data processing it will be of mutual benefit. On one hand, the long-term sustainability of the e-NMR grid infrastructure is ensured and on the other, the customers' demands are met that the interoperability and streamlining difficulties in the current computational protocols for the data analysis are solved.
It is already foreseeable that the user-base and therefore the demand for the e-NMR platform will enlarge in the future beyond the community of scientists who only apply NMR spectroscopy in the context of structural biology. Indeed, it should be possible to involve other scientific communities, which also make use of NMR spectroscopy, for example in the field of bio-banks. An European infrastructure for bio-banking is being planned (www.biobanks.eu). Another relevant scientific field or community, respectively, is that of metabolomics, where researchers routinely use NMR techniques and have links with bio-banks as well.
In summary, the prospects for medium-term sustainability of the project are good. This is due both to the role that the e-NMR platform will play in the context of the services that are and will be provided by the NMR infrastructures and to the possibility of creating close links with other scientific communities that are more biologically or biomedically oriented.

B3. Potential impact

B.3.1 Strategic impact

The e-NMR project builds on long-standing experience and collaborations at a European level. The three NMR partners (BMRZ, CIRMMP, BCBR) are the three core members of the FP6-I3 NMR research infrastructure (EU-NMR); as such they have been providing access to European users since FP3. They also have been and are collaborating within several European projects and networks (e.g. FP6-STREP-NDDP, FP6-STREP-UPMAN, FP6-CA-NMRLife, FP6-DesignStudy-BioDNP, FP6-SPINE2, …). The e-Science partner (INFN) has been and is involved in several European e-Infrastructure projects (e.g. EGEE/EGEE II, EUmedGRID, EUChinaGRID, EU-IndiaGrid, EELA, GridCC and BioinfoGRID). As such, they already have collaborative arrangements with a large number of both European and worldwide research groups. Through various EU projects, they are also linked to both large (manufactures of NMR instrumentation, pharmaceutical) companies and small and medium enterprises (SME) within Europe. The e-NMR project will bring together the bio-NMR and e-Science European communities. Both communities already operate at an international level. We therefore expect a large international impact, which could not be realized at a national level.

The present project will provide new opportunities for users accessing to the NMR research infrastructures, by enabling a more thorough analysis of the experimental data collected. This will result in a better use of the data and in a higher reliability and visibility of the results (in terms of biological implications) derived from the raw NMR data. We also foresee that the whole process of the analysis of bio-NMR data will become more straightforward and more rapid thanks to the availability of the e-NMR platform of optimised protocols (at least for the more common needs of the users). In parallel, the e-NMR platform will also allow an increase of the quality of the results obtained through bio-NMR, thanks to the provision of standards and to enhanced interoperability among different laboratories, which in turn will encourage data cross-validation as well as the development of better assessment methodologies. The joint research activities planned in this project will additionally bring about an increase in the quality of information extracted from NMR data (in particular for what concerns structural information), which is a long-standing open issue in the field. The availability of selected and well-documented test-cases will permit benchmarking exercises and will foster more rapid innovation of computational applications for bio-NMR.
Access to e-NMR will provide users with all of the above benefits in an integrated manner and requiring only minimal efforts for the set-up of calculations. Therefore users will be able to exploit bio-NMR infrastructures more thoroughly and the usefulness of these infrastructure for the European scientific community will be enhanced. The e-NMR service activities will not only result in an increase of the quality and quantity of access to bio-NMR infrastructures by the current pool of users, but will also encourage access by new user groups, from disciplines more far from molecular spectroscopy/biophysics (i.e. closer to “pure” biology). This is because at present bio-NMR is still perceived as a technique that can be applied only by highly trained experts due to the complexity of the procedures for data analysis. The e-NMR services will significantly reduce this complexity and thus will make the technique accessible also to non-experts.

The service activities of the e-NMR project are expected to have a broad international impact. Indeed, the project will reach from the start a broad European bio-NMR user community via the links provided by the EU-NMR and NMR-Life projects as well as a broad European e-science community via the EGEE/EGEE-II link. The partnership itself is spread over three different European countries, which will enhance right from the beginning of the project its pan-European impact. Indeed, this aspect constitutes one of the main reasons why the present endeavour cannot be tackled at a national or local level. Reaching very soon a critical mass of scientific expertise and,

simultaneously, of the user community, with a correspondingly large geographical distribution (such as that ensured by EU-NMR) is a vital requisite for the attainment of the impacts described in the preceding paragraph. Another important outcome of the project will be the increased awareness and usage of GRID computing in the European Life sciences community, and, at the same time, an increased interest of researchers working in e-Science towards Life science problems. The e-NMR services will thus stimulate new European-level collaborations bringing together different disciplines from computer sciences and biological/biomedical sciences.

The fact that the resources deployed by e-NMR partners, as well as their research communities, are already spread across different European countries is not the only reason why the present project has significantly higher chances than any national or regional level project to attain a significant effect on European science. In fact, the integration of bio-NMR resources with the EGEE Grid infrastructure requires the adoption of methodologies and policies that are agreed at the European level and are compliant with the EGEE ones. Similar requirements apply to the new middleware components which will developed within the Joint Research activities but will be certified and integrated in the frame of the Service activities. Discussions with national bodies cannot substitute for the more extensive indications and impacts that are brought about by having and working at an European-level dimension. Furthermore, the portfolio of applications of bio-NMR is so vast that it cannot be reasonably covered by the expertise present in a single site. Indeed, the involvement of different research infrastructures across Europe allows all disciplines within the field of bio-NMR to be adequately represented and thus addressed in the planned development activities.

The required steps to achieve the expected impact include:

- deployment of a e-NMR grid at the research infrastructures and link to the EGEE grid
- identification of an ensemble of applications suitable for porting to e-NMR covering the various facets of the application of bio-NMR
- development of test cases for the implementation of optimised protocols and to benchmark new software tools and calculation protocols
- enhancement of GRID management software and adaptation of NMR applications for e-NMR GRID computing
- establishment of support and dissemination mechanisms and training of our user community
- opening the e-NMR platform to the target scientific communities

With respect to assumptions and external factors, it must be noted that both NMR in Life sciences and GRID computing are well established fields. In addition, the management structure is flexible enough and the network of relations of the partner is broad enough to guarantee that the occurrence of any unprecedented external event that can be potentially harmful or limiting for the impact of the project will be readily identified and adequately taken care of. Indeed, any significant technological advancement that should occur during the lifetime of the present project will be carefully evaluated by the project's Management Committee, also with the help of the Advisory Board of e-NMR, in order to make a decision if and how to incorporate it into the reach of our activities. If necessary because of specific scientific and/or technological implications, other experts may also be contacted to obtain advice. The present project does not rely on specific assumptions, and the majority of the individual tools and expertise needed for its successful undertaking are available to the present partnership. Therefore, the onset of technological problems preventing completion of the planned activities is quite unlikely. We also do not expect problems in successfully linking the fields of e-Infrastructures and bio-NMR: the latter has a long history in both software and hardware development and middleware GRID computing is now a mature field.

B.3.2 Plan for the use and dissemination of foreground

Results of the research activities of the e-NMR project will be immediately made available to the scientific community mainly through the NMR research infrastructures, thanks to their continuous activity in hosting visitors on the basis of the transnational access project EU-NMR, as well as through the web. We will make sure to take all the due precautions for protecting the potential for commercial exploitation of the results when relevant. New results will be the object of dissemination and exploitation activities involving all the partners through the meetings, workshops and training events organized in the context of the Networking activities of the present project as soon as practically feasible.
The activities of e-NMR are expected to give rise to a number of publications in scientific journals of the highest international standing. At present, there are no limitations concerning prior commitments with respect to the exploitation of knowledge. In general, the results of the project will be made freely available to the scientific community. As part of the activities of this project, we will devote a specific section of the e-NMR web site to distribute the advancements achieved within e-NMR.
The web site will also allow us to advertise the events organized and promote the results obtained in the networking activities, as well as to advertise any other initiative taken within e-NMR or involving e-NMR in collaboration with other existing external projects, networks or scientific organizations. In addition, the scientific results achieved within e-NMR will be advertised at international conferences through lectures and posters. Posters describing the activities within e-NMR will also be posted at international meetings, particularly those reaching an audience outside the bio-NMR and e-Science communities, in order to reach scientists who can profit in their research from the activities being pursued in the present consortium, thereby increasing the awareness of bio-NMR as well as e-Science throughout the Life sciences community and fostering the future organization of other interdisciplinary initiatives.
Newsbreaks regarding the achievements of the e-NMR members will be collected, and made available to the EC. An electronic newsletter may be distributed periodically, within and outside e-NMR.

More in detail, for the dissemination of the knowledge produced in this project, we will mainly follow the actions below:

- Among the activities of e-NMR, the maintenance of a dedicated web site to distribute the advancements achieved within the activities of the present project is explicitly planned. The web site will allow to promote the results obtained in the various research and networking activities, as well as to advertise other initiatives taken within e-NMR.
- Results of the research activities of the e-NMR consortium will be immediately made available to the scientific community at the partner sites providing access through the web, taking due precautions for protecting the potential for exploiting the results.
- For reaching to our European user community, we will make use of the user database and web services already in place with the EU-NMR infrastructure (see http://www.eunmr.eu and http://www.postgenomicnmr.net).
- New developments and services will be actively advertised at the annual user meeting of the European NMR Large Scale Infrastructures. Our networking activities are also aimed at reaching the worldwide NMR community and establishing links to other communities with both Life- and e-Sciences.

- The scientific results achieved within e-NMR will be advertised at international conferences through lectures and posters. Posters describing the activities within e-NMR will also be posted at international meetings, particularly those reaching an audience outside the bio-NMR community, which can profit in their research from the activities being pursued in the e-NMR consortium.
- The results of the research activities of e-NMR will be communicated in a number of publications in scientific journals of the highest international standing.
- Wherever possible and worthwhile workshops and training courses will be organised for broad target groups in the Life sciences, which further the education in bio-NMR computing and promote the use of advanced NMR computational techniques to solve research questions in these areas.

The formal issues with respect to the access rights to knowledge, and the dissemination and exploitation of knowledge generated in the project are regulated in the Consortium Agreement of the project. For the exploitation and management of intellectual property the following rules will be followed

- Whenever applicable, patenting of the results obtained within the e-NMR activities will be encouraged.
- In order to reach the largest community possible, we will implement mainly open source software or software freely available for non-profit organizations if possible. This is a clear requirement if the e-NMR software is expected to run at an international level on a GRID.
- Software requiring licensing will only be made available at the partner sites within the consortium. Access to commercial users might be granted provided these can prove that they own valid commercial licenses. Access to the e-NMR platform will be controlled by a portal that will be developed with the service and access activities.

B4. Ethical issues

Not applicable

B5. Consideration of gender aspects

The European Research Frameworks have been particularly important for promoting the participation of women in cooperative scientific programmes and the Co-ordinator and all partners engaged in this proposal are fully committed to this philosophy and implementation of gender equality in science.

There are no specific gender issues involved with any of the workpackages of the present proposal. Note also that the research planned in this project will not involve studies of living organisms promoted by the consortium.

Access to the e-NMR grid will be open to any research team, irrespective of race, religion and or gender. Because women constitute only a small fraction of senior scientists active in bio-NMR as well as in e-Science in Europe, the percentage of women involved in the various tasks related to the Service Activities planned in the project will reflect this paucity.

The networking activity are open to scientists from all disciplines in the Life sciences. Participation mainly from researchers active in the fields of bio-NMR, e-Science, and, possibly, bioinformatics is foreseen. The gender balance for the networking activities of the project will thus reflect that within these communities.

With respect to the joint research projects, the consortium members and the Management Committee will be responsible for actively recruiting female scientists, for providing them with projects that are as rewarding and demanding as those of their male colleagues. They will also be expected to promote their female post-docs as prominently as the men. Likewise, they must give female group members the same responsibilities (students to supervise, administrative jobs, etc) as their male counterparts, and eventually help them to establish independent research groups. This approach is already well established in our partner laboratories, but is worth emphasizing formally.