Support action (SA) proposal
ICT Call 9
FP7-ICT-2011-9

PRESERVATION OF DATA for the DIGITAL SOCIETY [PODDS]

Support action
Date of preparation: 16 April 2012
Version number *(optional)*: 1.10

Work programme objective addressed
ICT-2011.4.3 Digital Preservation

Name of the coordinating person: Jamie SHIERS
E-mail: Jamie.Shiers@cern.ch
Fax: +41 22 766 9954

Participant no. *	Participant organisation name	Part. short name	Country
1 (Coordinator)	EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH	CERN	CH
2	MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.	MPG	DE
3	ISTITUTO NAZIONALE DI FISICA NUCLEARE	INFN	IT

Proposal abstract

The preservation of data for future re-use is a topic that is becoming increasingly important across a wide range of disciplines and is expected to become essential in the coming years as a cornerstone of the digital society. The strategy of this project is to combine the forces of several large-scale organizations to ensure that research work in this area is well coordinated and that the results are made available to a wide range of disciplines. The project aims to both document and verify techniques for the preservation of digital data with a strong focus on the continued ability to fully re-use the data, to raise awareness of the importance of this issue as widely as possible, to act as a forum for the interchange of information and best practices, and to develop a roadmap for the coming decade in order that digital preservation moves from being a challenge to a solution. Disciplines targeted range from Life Sciences to High Energy Physics – fields that have diverse needs in terms of data volume and lifetime as well as in access patterns and associated authentication and authorization requirements. Specific use cases that we wish to address include: the preservation of data for a small number of decades for educational outreach, preservation of data and the associated environment needed to maintain the ability to reanalyse it, as well as e-records of all kinds. The durations that we target are a small number of decades for the first two cases and up to a century or more for the latter. The goals can thus be summarised as: to consolidate and disseminate the current state of the art in this domain, including existing techniques and known challenges; to use a range of projects, networks and fora to further promote the fundamental necessity of providing solutions to an agreed set of use cases; and to develop a community-endorsed multi-disciplinary roadmap that clearly identifies the steps necessary to establish pilot services in this area.

TABLE OF CONTENTS

1 SCIENTIFIC AND TECHNICAL QUALITY

1.1 Concept and Objectives

Digital preservation is a complex issue that must be solved as a cornerstone of the digital society. Two often quoted use cases are the ability to redo or perform new analyses on existing data[1] and/or for educational outreach. These bring with them specific challenges that go much beyond the "simple" preservation of the bits, as we describe in more detail in section 1.1.2. Moreover, in the digital society it is essential that data be preserved at least for many decades, if not for much longer (consider electronic patient records, scans and the plethora of e-documents that are now managed in somewhat ad hoc ways and for which long-term preservation is not guaranteed).

Whilst much research has been done in this area, the results are not always widely disseminated – particularly between different communities – and a ***proven end-to-end and sustainable service*** that could handle even modest requirements[2] – is not yet available. However, there is today a wide spectrum of communities and applications that requires data preservation at a much larger scale and for significantly longer periods. This includes sciences (High Energy Physics, Astronomy and Astrophysics, Life and Earth Sciences and so forth), arts and humanities, industry and service sectors as well as libraries and archives of all forms. To avoid the loss of huge amounts of precious data it is imperative that long term data preservation and re-use are seen as an integral part of the data lifecycle. A trusted data preservation infrastructure needs to be put in place to ensure community uptake and the communities must be engaged to ensure that data re-usability, both within and without the community, is guaranteed.

What makes this proposal special is the combination of skills and experience that the partners bring:

- *A set of well established international and multi-disciplinary networks that will be essential for all phases of the project;*
- *A proven track record of resuscitating long-dead data with clear documentation of the issues involved;*
- *Having to address preservation of data for several decades as an integral part of the supported scientific programmes;*
- *In handling massive data volumes and rates (tens to hundreds of PB stored, sustained data rates of several GB/s) in fully distributed environments;*
- *Having direct involvement in a wide range of primarily scientific disciplines with complex software and hardware environments.*

Further details of these skills and experience as well as the complementarity of the partners is provided in section 2.3.

The purpose of the proposed project is therefore:

[1] For example – in the case of scientific data – when called for by new theoretical insights, new or improved analysis and/or simulation techniques, or to search for previously unseen signals observed at a newer facility.
[2] For example, the preservation of 1PB of data for 1 decade, ***including the ability to continue to use the data in a meaningful manner***. This includes maintaining and migrating the needed software and metadata.

- To consolidate and disseminate through **well-established channels** the current state of the art in terms of research into digital preservation – with a strong focus on reuse – as well as existing techniques and known challenges;
- To use previous, current and planned projects, networks and fora to further **promote** the fundamental necessity of providing solutions to a documented and agreed set of use cases;
- To develop a multi-disciplinary medium- to long-term **roadmap** for the digital preservation domain, clearly identifying the steps necessary to establish pilot services in this area and mechanisms whereby they could be validated from a technical and sustainability viewpoint, leading to a pan-European / truly international set of service offerings in the coming five to ten years. This roadmap should be endorsed as widely as possible by recognised institutes, bodies and communities.

The challenge involved is clearly international as is explained in the discussion of the key use cases below.

In other words, this proposal is strongly aligned with the objectives under point d) of the ICT-2011.4.3 Digital Preservation objective:

"d) Promotion schemes for the uptake of digital preservation research outcomes including outreach to new stakeholders and road mapping activities."

The proposal aims are also well aligned with the recommendation of the e-IRG and the High Level Expert Group (HLEG). Indeed the need for data preservation and its ensured re-usability was highlighted by Neelie Kroes, the Vice-President of the European Commission responsible for the Digital Agenda for Europe, in her foreword to the HLEG "Riding the Wave" report:

"My vision is a scientific community that does not waste resources on recreating data that have already been produced..."

The cost of repeating large-scale scientific experiments, such as re-launching a satellite or re-building a particle accelerator is prohibitive – tens or hundreds of millions of Euros. Some data – from Supernova events, social surveys, and historical documents – are simply irreplaceable, once lost they are gone forever. This highlights the need to ensure that data is preserved in an open and re-usable manner.

The project would be centred on a small number of funded partners and would rely heavily on the input and guidance from a much larger set of communities, including those beyond the domains represented by these core institutes. These disciplines and institutes would also give formal direction to the project through a International Advisory Committee (see section 2) that would provide early feedback on the evolving roadmap for long-term digital preservation for continued use.

In the work package tables below we count only the funded effort that would be more than matched by the unfunded effort from the partners (see sections 2.2 and 2.4).

1.1.1 Related Disciplines and Projects

The disciplines and projects with which we will interact are described in detail in section 2.4 "Resources to be Committed".

1.1.2 PODDS Use Cases

We list below the main use cases that we wish to address with some simple but non-exhaustive examples. Further details would be given in the roadmap document as well as via the outreach and dissemination activities.

1. *The preservation and continued ability to use e-records of all kinds for at least the lifespan of an individual (say 100 years: for some disciplines even longer retention periods are desirable);*

Whilst this area is broad, encompassing the service sector through scientific disciplines and indeed all digital information, a key example relates to health records. The Partner [PARTNER] project, for example, is studying the use of hadron therapy as a means of providing better patient treatment, something that is of key concern both to individuals – some 80% of whom will develop some sort of cancer during their lives – as well as to society as a whole. The preservation of scans and indeed all patient records, as well as the ability to continue to use them at least during the lifetime of an individual (and indeed beyond to allow for better statistics on treatment types and evolution of conditions) is of extreme value. Numerous other examples exist and will be elaborated on in the roadmap. By definition, this problem needs to be solved at the international level: at least European and preferably, by the use of the appropriate standards and if necessary gateways, globally. The motivation for an international response can be seen in the example, which involves initial diagnosis and treatment at a first hospital, follow-up and referral at hadron-therapy centres across Europe and the involvement of further sites in research, discovery and training at an additional site.

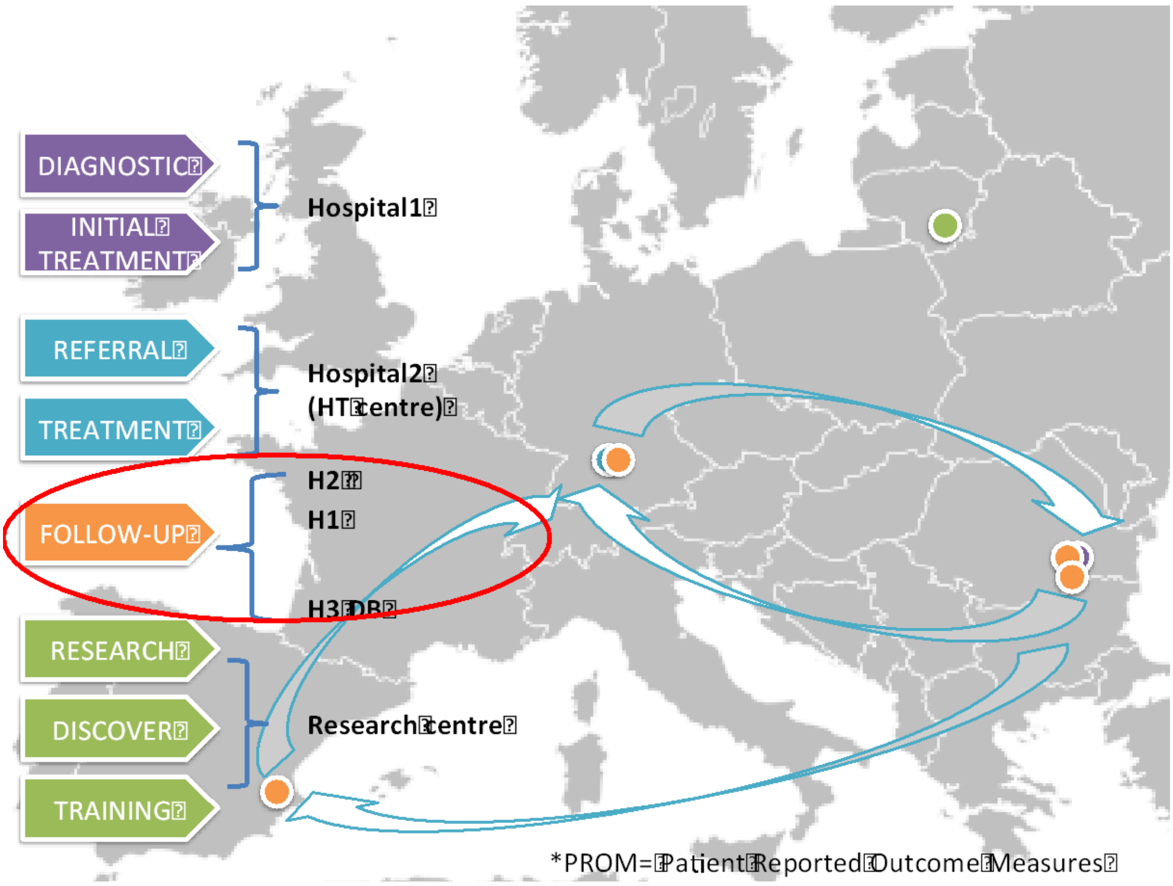

Figure 1 Example of data flow from the initial diagnosis and treatment at a first hospital, follow-up and referral at hadron-therapy centres across Europe and the involvement of further sites in research, discovery and training at an additional site.

2. *The preservation and continued ability to re-use (e.g. re-analyse) scientific and library data for a small number of decades;*

Each generation of space telescope brings with it new discoveries, many of which motivate additional searches in the data from previous devices. These may provide additional information, such as a more complete understanding of an object, e.g. from different wavelengths and / or the time evolution of such an object. This is shown in the diagram below, where information combined from X-ray, optical, radio and infrared telescopes can provide a much more complete picture than from a single device.

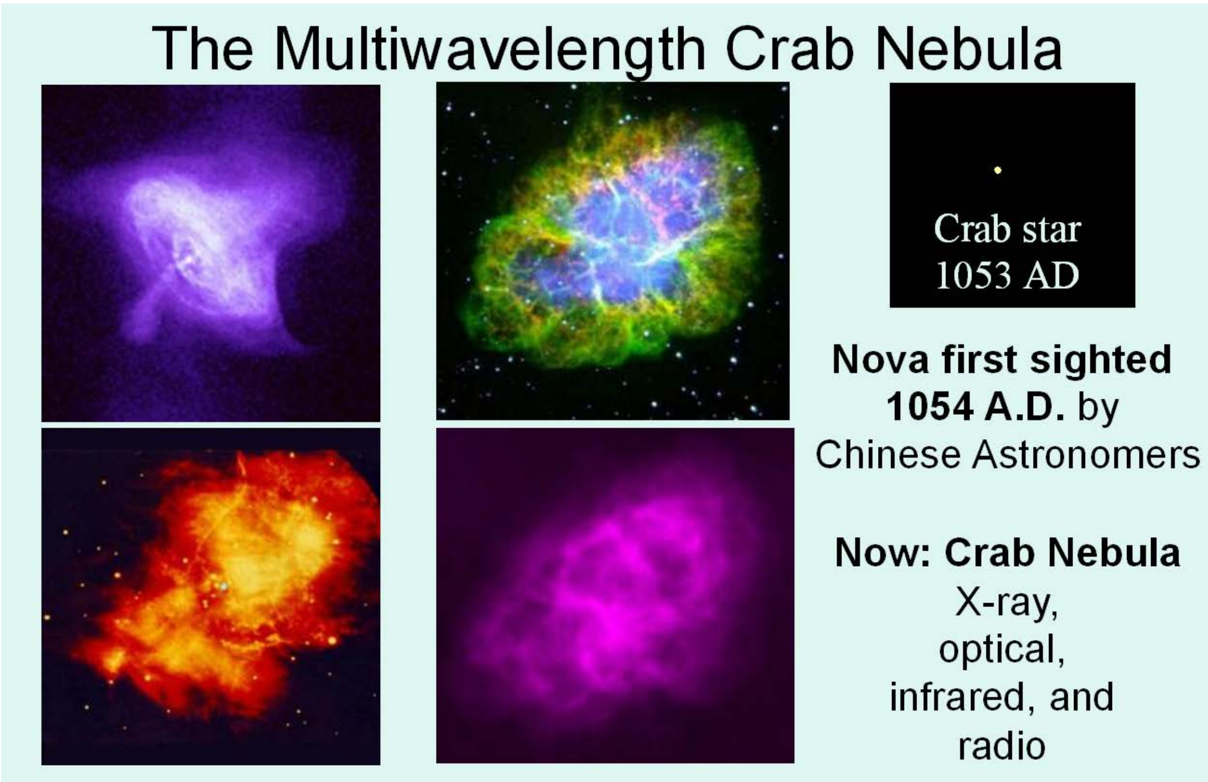

Figure 2 Slide courtesy of Robert Brunner @ CalTech

Numerous other examples exist and may be motivated by better theoretical understanding or modelling and again cover multiple scientific domains. For example, in High Energy Physics, searches for evidence for the standard model Higgs boson in data from the Large Hadron Collider (LHC) at CERN have recently motivated re-analysis of data from the previous collider, the Large Electron Positron collider (LEP).

In performing these re-analyses, it is essential that the software and "knowledge" environment is preserved – more realistically, often adapted, re-written but most importantly verified for a potentially quite different computing environment. This critical area is one where we believe that we can bring highly pertinent skills.

Many other examples of re-analysis of old data exist, along with much experience on the difficulties and challenges that are associated with this work.

Once again the necessity of an international response can be seen from the nature of these disciplines: the devices involved (space telescopes, particle accelerators) are today operated and exploited by international collaborations and thus any solution must at a minimum allow for continued access from the collaborations and institutes involved in acquiring the data and, where appropriate, made available for wider educational and even public access.

3. *The preservation of scientific data for educational outreach for a small number of decades.*

As our knowledge of the universe expands and new data are collected, we find it useful to return not only to our past conclusions but also to the old data themselves and check whether or not it all survives in a consistent interpretation. Having access to data from experiments all over the world can raise outreach efforts to the public to another level by letting non-experts interact with the scientific experience in a way not previously possible. The outreach tools developed for these efforts can also be used for undergraduate college courses and to train graduate students who will be the next generation of scientists and/or take their skills and training from such an education into other fields.

By improving our educational tools for the general public, we will also develop better techniques for teaching new graduate students and collaborators which will allow them to contribute more quickly to the experiments. (For example, by repeating key analyses performed in the past – the "historical base" that they need if they are expected to make the next big discoveries).

Well understood datasets can address the aforementioned outreach and education efforts. There are four main groups who can learn from and benefit from these data:

1. The general public;
2. High school science students;
3. College students studying courses in particle physics or computing;
4. Graduate students in particle physics.

Each of the above groups brings its own challenges and may interact with the data in different ways.

Thus this work will raise the overall awareness and appreciation of scientific work so that it becomes an integral part of future educational models, addressing the needs identified above.

1.1.3 Strategy

The key elements of our strategy are summarised below.

a. *The consolidation of existing requirements, knowledge and research results into a single knowledge repository that is as broad as possible in terms of the disciplines that it represents. This includes not only sciences (High Energy Physics, Astronomy and Astrophysics, Life and Earth Sciences and so forth), but also arts and humanities as well as industry and service sectors;*

Whilst there is on-going work in a number of domains, such as the DPHEP [DPHEP] initiative in High Energy Physics, these efforts are somewhat fragmented even within single disciplines. Through the extensive international networks described in more detail in section 2, but which include the EIROforum [EIROforum] , the Max Planck Society [MPG] as a whole, and the extensive range of disciplines and institutes that are thus represented, we would consolidate all existing knowledge, research results as well as requirements into a single knowledge base. This would be used as input to further activities, such as broader outreach and dissemination and roadmap preparation.

b. *The motivation for and benefits from digital preservation would be widely presented to as many disciplines as possible, both to increase awareness of this matter and to seek further input and requirements for future activities. Existing networks and contacts would be used for this purpose and further ones established as necessary;*

Using the partnerships and networks of the members of the consortium, that include a variety of projects, meetings and conferences in which these institutes are involved, we would actively campaign to raise awareness of the needs and benefits from digital preservation in all domains in which we have knowledge and contacts, simultaneously seeking additional input on future directions. We foresee a variety of levels at which participation would be possible, all of which would be beneficial to the development of the roadmap and the overall goals of the project but would encourage as active and as wide participation as possible, with a strong preference for pro-active rather than re-active response. We firmly believe that the members of the consortium, both at the level of institutes as well as individuals, have the reputation and influence to make this happen, for example through the EIROforum.

c. *Building on this work, areas of key concern and urgent future work would be identified and prioritised with the explicit approval of the disciplines concerned;*

As part of the iterative process that we see, which includes regular presentation and publication of our results, we would identify and prioritise the most pressing areas for future work. Additional checkpoints would be made via the International Advisory Committee described in section 2.

d. *A roadmap document would be drafted, widely disseminated and revised, seeking the approval of as many representative bodies and communities, such as the EIROforum members.*

Through the regular presentations of our work together with the proposed annual roadmap documents, we would actively seek the input from and approval of a wide range of disciplines and institutes of international renown. They in turn would be encouraged to use their networks to cast even wider, as well as to explicitly endorse the documents produced. Thus, the final roadmap document would come with the official support from a wide range of disciplines and associated institutes.

1.1.4 Work Packages and Deliverables

The work packages that we propose follow closely the call itself, covering Digital Preservation Technologies, Digital Preservation Outreach and Roadmap Preparation, plus a small management work package.

Each work package sees the participation of all the three partners, to facilitate the transfer of knowledge and workload. The leading partner of each package contributes about 80% of the needed effort, evaluated in person months.

The milestones and deliverables identified for each work package are detailed in section 1.3 and a summary of the task breakdown, together with the proposed means of verification, is given in the table below.

Task	WP	Description	Means of Verification
1.1	1	Report to EU – the various milestones, deliverables and annual project review	D1.2 (M12), D1.3 (M24)
1.2	1	Organise and chair the Project Management and Technical Management Boards	Agenda & Minutes via Project Website (D1.1)
2.1	2	Storage solutions for preservation of data	D2.1 (M9)
2.2	2	Solutions for preservation of software libraries and metadata	D2.1 (M9)
2.3	2	Build a small test-bed to evaluate proposed solutions	D2.2 (M22)
2.4	2	Implement an automatic test-suite to verify such solutions	D2.2 (M22)
3.1	3	Information capture and consolidation	D3.1 (M12)
3.2	3	Preparation of supporting material	D3.2 (M18)
3.3	3	Increase Awareness and Promote Collaboration	D3.3 (M22)
4.1	4	Establish an interim roadmap for Digital Preservation	D4.1 (M12)
4.2	4	Establish an endorsed roadmap for Digital Preservation	D4.2 (M24)

The timing of the various work packages, milestones and deliverables is shown in the following diagram.

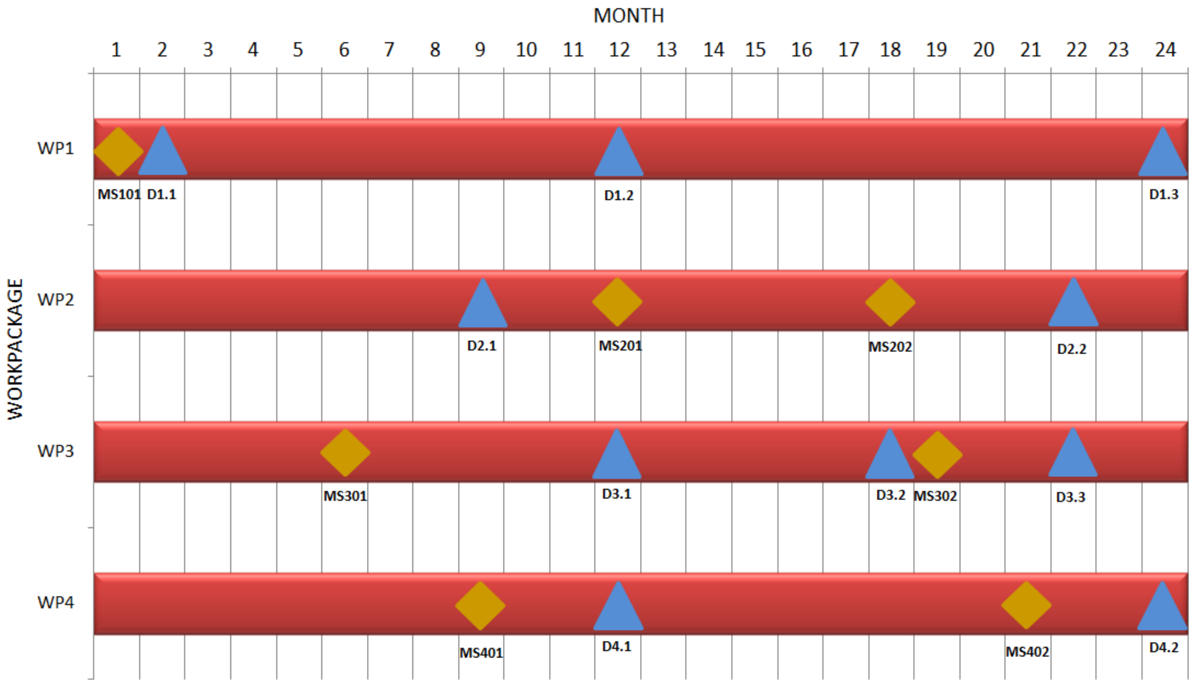

Figure 3 Timing of the various work packages

Key:


Milestone

Deliverable

List of Milestones

MS101 – Project mailing lists
MS201 – Test-bed installed
MS202 – Test suite implemented
MS301 – Requirements capture
MS302 – Workshop Announcement
MS401 – Draft Interim Roadmap
MS402 – Draft Roadmap

List of Deliverables

D1.1 – Project website established
D1.2 – Annual Report
D1.3 – Final Report
D2.1 – Report available technologies
D2.2 – Test-bed results report
D3.1 – Outreach report
D3.2 – Public web site on Digital Preservation
D3.3 – Workshop to increase awareness
D4.1 – Interim Roadmap for Digital Preservation
D4.2 – Final Roadmap for Digital Preservation

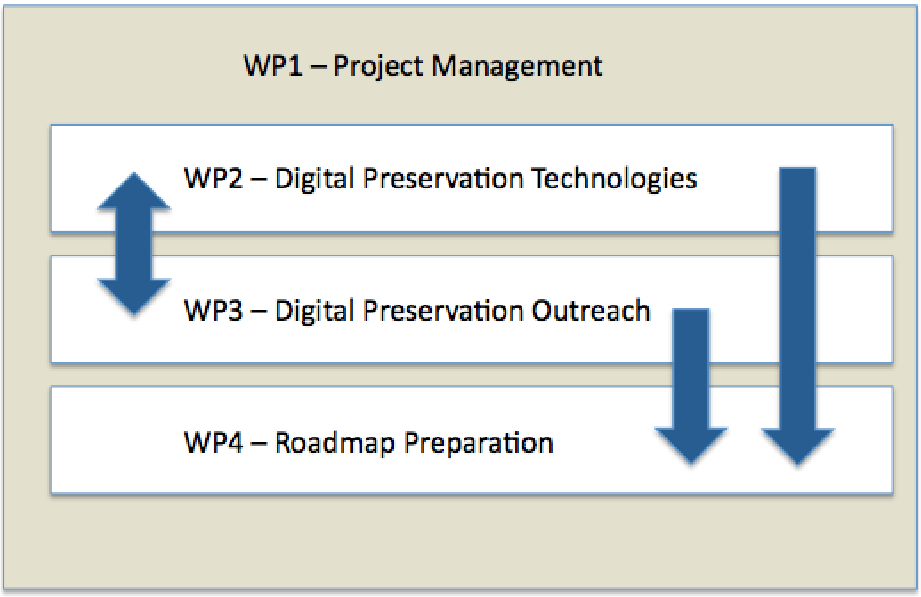

Figure 4 Interdependencies between work packages

1.1.5 Summary of Expected Impact

As described in further detail in section 3.1, the main impacts relevant to this call that we expect to achieve (in order of decreasing priority) are:

- Wider Adoption of Research Results;
- Sustainable Access to Information;
- More Efficient Selection of Resources to be preserved.

1.2 (Section not required for Support Actions)

1.3 PODDS Work plan

Motivation

The three main use cases that we wish to address are:

1. The preservation and continued ability to use e-records of all kinds for at least the lifespan of an individual (say 100 years: for some disciplines even longer retention periods are desirable);
2. The preservation and continued ability to re-use (e.g. re-analyse) scientific and (software) library data for a small number of decades;
3. The preservation of scientific data for educational outreach for a similar duration.

These use cases have different needs and challenges associated with them and it is expected that additional use cases will need to be addressed, such as the preservation of arts and humanities data – assumed to be digitised analogue data which can be compressed and/or tolerate minor losses of fidelity which would not be acceptable for many other cases. Given these timescales it is essential to consider not only the migration of the data across various media but also to handle both minor and major changes in storage systems and data formats. Furthermore, the continued ability to use the data brings with it additional challenges, as do issues including authentication, authorisation and accounting.

The impact of this work is described in detail in section 3.

Strategy

The main elements of our strategy would be:

a. The consolidation of existing requirements, knowledge and research results into a single knowledge repository that is as broad as possible in terms of the disciplines that it represents. This includes not only sciences (High Energy Physics, Astronomy and Astrophysics, Life and Earth Sciences and so forth), but also arts and humanities as well as industry and service sectors;
b. The motivation for and benefits from digital preservation would be widely presented to as many disciplines as possible, both to increase awareness of this matter and to seek further input and requirements. Existing networks and contacts would be used for this purpose and further ones established as necessary;
c. Building on this work, areas of key concern and further development would be identified and prioritised with the explicit approval of the disciplines concerned;
d. A roadmap document would be produced, widely disseminated and revised, seeking the approval of as many representative bodies and communities, starting from the EIROforum members.

Initial informal discussions regarding multi-disciplinary digital preservation have been held at events including the IEEE Massive Storage and Technologies symposium in 2009 and 2010 and the EGI Technical Fora in 2010 / 2011. These discussions have attempted to capture multi-disciplinary requirements in terms of a matrix and lead us to believe that there is commonality across a wide range of disciplines even if absolute priorities in terms of functionality differ. By capturing these requirements in a single document we believe that a common roadmap highlighting the various steps and timelines involved in establishing first prototypes, then demonstrators of technology, functionality and sustainability leading to persistent service offerings is both timely and necessary.

It is clearly expected that this project lead into further activities that would be outlined in the roadmap. Although the partners involved cover a wide range of disciplines, they do not cover all. A further important activity would be the development of a future proto-consortium that could be formed by merging with other activities covering disciplines that are less represented by the PODDS partnership and/or extending the existing collaboration in the necessary areas. This falls naturally into the dissemination, outreach and road mapping activities that are planned.

1.3.1 Work Package List

Work package No	Work package title	Type of activity	Lead partic no.	Lead partic. short name	Person-months	Start month	End month
WP1	Project Management	MGT	1	CERN	2	1	24
WP2	Digital Preservation Technologies	SUPP	3	INFN	24	1	24
WP3	Digital Preservation Outreach	SUPP	2	MPG	22	1	24
WP4	Roadmap Preparation	SUPP	1	CERN	24	1	24
	TOTAL				72		

Table 1.3a: Work package list. Only EU funded person months are reported in the table. About the same amount of effort from senior personnel will be added from the institutes involved, mainly through the named experts listed in section 2.2. The coordination of WP1 will be funded directly by CERN at a cost of 20% of a senior member of personnel.

1.3.2 List of Deliverables

Del. no.	**Deliverable name**	**WP no.**	**Nature**	**Dissemi-nation level**	**Delivery date (proj. month)**
D1.1	Project website established	WP1	Web Document	PU	M2
D1.2	Annual Report	WP1	Document	PU	M12
D1.3	Final Report	WP1	Document	PU	M24
D2.1	Report available technologies	WP2	Document	PU	M9
D2.2	Test-bed results report	WP2	Document	PU	M22
D3.1	Outreach report	WP3	Document	PU	M12
D3.2	Public web site on Digital Preservation	WP3	Web document	PU	M18
D3.3	Workshop to increase awareness	WP3	Workshop	PU	M22
D4.1	Interim Roadmap for Digital Preservation	WP4	Document	PU	M12
D4.2	Final Roadmap for Digital Preservation	WP4	Document	PU	M24

Table 1.3b: Deliverables List. Dissemination level code: PU=Public

1.3.3 List of Milestones

Given the goals of the project, the primary milestones that we propose concern the evolving road map. They are intended as a means of verifying that the work is proceeding in the correct direction as are as both the EU and concerned disciplines and institutes are concerned. The plan is to present the draft roadmap at least twice per year – additional presentations will be scheduled as part of the outreach and dissemination activity but at a schedule that is hard to predict – and will be complemented by an interim roadmap (D4.1) as well as a final version (D4.2).

Milestone number	**Milestone name**	**Work package(s) involved**	**Expected date**	**Means of verification**
MS101	Project mailing lists	WP1	M1	Mailing lists established and in use
MS201	Test-bed installed	WP2	M12	The test-bed is installed and ready to be used in order to test the data preservation technologies
MS202	Test suite implemented	WP2	M18	The test suite for running automatically data preservation test
MS301	Requirements capture	WP3	M6	Participation at conferences and fora to collect requirements
MS302	Workshop Announcement	WP3	M19	Public announcement of the workshop on data preservation
MS401	Draft Interim Roadmap	WP4	M9	Draft of Interim Roadmap document (D4.1) presented to IAC
MS402	Draft Roadmap	WP4	M21	Draft of Roadmap document (D4.2) presented to IAC

Table 13c List of milestones

Work package number	WP1	**Start date or starting event:**	M1			
Work package title	Project Management					
Activity Type	MGT					
Participant number	1	2	3			
Participant short name	CERN	MPG	INFN			
Person-months per participant:	0[3]	2	0			

Objectives

- Manage and monitor progress towards stated goals;
- Coordinate interactions with the European Commission;
- Ensure effective communication between project participants and related projects;
- Provide administrative support to ensure timely, high-quality technical and financial reporting.

Description of work

This work package will be responsible for:

Task 1.1: All reporting to EU – the various milestones, deliverables and annual project review;

Task 1.2: Organising and chairing the two bodies foreseen within the project, the Project Management Board and the Technical Management Board. The Project Management Board (PMB) consists of a representative of each partner and is chaired by the Project Coordinator. Its purpose is to ensure that the project is on track with respect to its objectives and deliverables. The Technical Management Board will consist of the Project Coordinator and representatives from the work packages and will be responsible for following the progress of the project with respect to the defined work plan.

Deliverables

D1.1: Project website established. (M2).

D1.2: Annual Report describing the progress made by the project during the first year. (M12)

D1.3: Final Report describing the progress made by the project during the second year. (M24)

[3] The coordination of the WP1 will be funded directly by CERN at a cost of 20% of a senior member of personnel

Work package number	WP2	**Start date or starting event:**		M1		
Work package title	Digital Preservation Technologies					
Activity Type	SUPP					
Participant number	3	2	1			
Participant short name	INFN	MPG	CERN			
Person-months per participant:	20	2	2			

Objectives

- Analyse pros and cons of the available “data preservation techniques” given the requirements gathered from the involved scientific communities
- Deploy the most promising techniques analysed on a small test-bed to carry on standard test about functionalities, scalability and reliability.
- Use the use cases provided by involved scientific communities in order to stress test the deployed data preservation solutions
- Provide recommendations to user communities on metadata management for improving data preservation capabilities
- Provide recommendations to user communities on the software stack used to analyse data in order to improve data preservation capabilities

Description of work

This work package will be responsible for finding and evaluating solutions for providing data preservation techniques to different scientific and other communities. It is split into the following tasks:

Task 2.1: Search for available storage technologies that could be relevant in this context. This search will be realized also by means of participation to conferences and fora covering these topics, including the ones organized by WP3, as well as searching the available literature (publications, web etc.).

Task 2.2: Search for available solutions to preserve the software libraries and the metadata needed to understand the data semantics and their characterization.

Task 2.3: Build a small test-bed where the most promising solutions will be installed for testing, after a general evaluation of the capabilities of each proposed solution. Deploy applications provided by user communities (WP3) on the above mentioned test-bed and evaluate their strong and weak points and if they fulfil the user requirements (WP3).

Task 2.4: Implement an automatic test-suite that will check if the whole chain (from the user application to the data read) is correctly working for every change in the environment and infrastructure, such as: changing storage media, changing storage management software, changing operating system, etc.

The results will be disseminated by WP3 participating to conferences on the field of data technologies. The same results will be provided as input to WP4, in order to write the roadmap to drive future development and services.

Deliverables

D2.1: Report on the available technologies on data preservation and their characteristics (M9)

D2.2: Report on the test-bed installation and test executed in order to guarantee data preservation for the users communities involved into the project. (M22)

Work package number	WP3	**Start date or starting event:**		M1		
Work package title	Digital Preservation Outreach					
Activity Type	SUPP					
Participant number	2	3	1			
Participant short name	MPG	INFN	CERN			
Person-months per participant:	18	2	2			

Objectives

- Consult widely with organisations, user communities and projects on Digital Preservation needs and activities. Encourage and open dialog with and between the communities and foster trust;
- Prepare document summarising the use (and usefulness) of preserved data for outreach activities like international master classes and science courses, as one of the major motivations for data preservation;
- Prepare and organise presentations about data preservation at a large variety of conferences and workshops, in order to increase awareness of the problem and to stimulate activities in a wide range of fields;
- Organise workshops dedicated to data preservation policies and techniques;
- Prepare and disseminate information material about data preservation: motivation, techniques and use cases.

Description of work

This work package is responsible for generating and increasing the awareness of the necessity of data preservation. Information about all aspects of data preservation is to be gathered from a wide variety of fields and projects and is to be used to optimise models for data preservation and define best practices and policies. Moreover a series of workshops and seminars will be used to encourage collaboration and information exchange amongst the diverse research communities and projects.

The work is to be organized into three distinct tasks.

Task 3.1: Information capture and consolidation: Create and distribute surveys for requirements capture. Gather information from a wide variety of fields. Correlate and identify synergies.

Task 3.2: Preparation of supporting material: Public web pages, flyers, posters and white papers on data preservation best practices and policies.

Task 3.3: Increase Awareness and Promote Collaboration: Increase the awareness of the need for data preservation. Engage communities, presentations at conferences and workshops, directly approach research communities, projects, policy makers and funding bodies. Organise seminars and workshops to encourage collaboration and information sharing among a diverse range of research communities and projects.

Deliverables

D3.1: Report on outreach as motivation and use case for preserved data (M12)

D3.2: Public web pages as information tool for data preservation (M18)

D3.3: Organisation of dedicated workshop to increase awareness of the project results. (M22)

Work package number	WP4	**Start date or starting event:**		M1		
Work package title	Roadmap Preparation					
Activity Type	SUPP					
Participant number	1	2	3			
Participant short name	CERN	MPG	INFN			
Person-months per participant:	20	2	2			

Objectives

- Prepare an interim roadmap for Digital Preservation as a basis for further feedback;
- Regularly present the evolving roadmap at technical meetings / workshops / conferences to solicit further input;
- Based on this feedback and further iteration with key stakeholders, service providers and user communities prepare a final report that has the maximum endorsement (EIROforum members, leading International Organisations, user communities and others).

Description of work

The purpose of this work package is to establish a roadmap that is as inclusive as possible in terms of disciplines in terms of their requirements, techniques and where appropriate prototypes and/or solutions in the arena of digital preservation. The intent is to build on existing work and make use of all available networks to ensure that the roadmap not only captures the needs of as many disciplines as possible but also has the endorsement of the communities and institutes involved. The international dimension would be addressed both via the disciplines – many of which are international in nature – and by direct contact with related projects and institutes in other regions, such as the US and Asia-Pacific. The roadmap would state clearly the global requirements, highlight the priorities of specific communities and make concrete proposals as to how these needs can be addressed in the short, medium and long-term. It is foreseen that it would propose a small number of potentially sustainable models that would need to be demonstrated leading into a long-term service handling multiple disciplines at the International and Exabyte scale.

Task 4.1: Establish an interim roadmap for Digital Preservation and present soliciting further feedback;

Task 4.2: Establish a roadmap for Digital Preservation with endorsement from EIROforum members, leading International Organisations, user communities and others.

Deliverables

D4.1: Interim Roadmap for Digital Preservation (M9)

D4.2: Final Roadmap for Digital Preservation, endorsed by EIROforum members, data preservation projects and a wide spectrum of application domains and associated institutes (M21)

1.3.4 Summary of effort

In the table below the number of EU funded person months is shown for each work package and for each partner. *About the same amount of effort from senior personnel will be added from the institutes involved, mainly through the named experts listed in section 2.2*. In particular, the coordination of the WP1 will be funded directly by CERN at a cost of 20% of a senior member of personnel.

Partic. no.	**Partic. short name**	**WP1**	**WP2**	**WP3**	**WP4**	**Total person months**
1	**CERN**	**0**	2	2	**20**	24
2	**MPG**	2	2	**18**	2	24
3	**INFN**	0	**20**	2	2	24
Total		2	24	22	24	72

Table 1.3e Summary of effort in terms of EU funded person months. The work-package leader for each WP is indicated by the corresponding person-month figure ***in bold****. The coordination of the WP1 will be funded directly by CERN at a cost of 20% of a senior member of personnel.*

1.3.5 Risks and Contingency Plans

The main risks and associated contingency plans are described by work package in the table below.

Table 1 – Risks and Contingency Plans

Work Package	Risk	Impact	Probability	Mitigation
WP1	Inability to manage consortium	Possible failure to meet targets, e.g. milestones and deliverables.	LOW	The consortium is small and the partners have a good track record of working together successfully over many decades.
WP2	Inability to find example of feasible data preservation technologies	Possible failure in meet the target, e.g. providing input for writing the roadmap	LOW	In the project are involved people with great experience in data preservation and storage technologies. The WP2 will participate to dedicated conference and workshops to be informed on emerging works and technologies in the field of data preservation
WP3	Lack of success in soliciting talks (public conferences) and organising workshops	Insufficient degree of awareness and of triggering projects of data preservation across any fields	LOW	Persistence; Distribution of work load among interested persons (MPG personnel)
WP4	Lack of success in obtaining consensus on roadmap	Failure to achieve key impact of project	LOW	Persistence; Existing multi-disciplinary discussions suggest that convergence is likely
WP4	Lack of Trust in the roadmap	Insufficient uptake from research communities	LOW	Active engagement of communities from project outset. Ensure community needs are respected and foster trust by ensuring a high level of openness and knowledge of the proposed roadmap details.

2 IMPLEMENTATION

2.1 Management structure and procedures

The PODDS consortium consists of 3 partners that have a long history of working together on international scientific programmes as well as projects and services related to distributed computing infrastructures. All have strong involvement in multi-disciplinary activities as well as extensive networks of international partners. Given this, together with the relatively small size of the project, a simple management structure is considered appropriate. It is therefore proposed to establish only two boards within the project – the Project Management Board (PMB) consisting of one representative of each organisation and chaired by the Project Coordinator (PC) and a Technical Management Board (TMB), consisting of the leaders of each work package and again chaired by the PC.

The Project Coordinator will be assisted by the Coordinating Partner and an Administrative Assistant in daily administrative and financial management.

These bodies would hold regular phone or video conferences in addition to technical meetings organised within the work packages themselves. It will be a fundamental principle that all meetings will permit remote participation, given the distributed nature of the consortium and the importance of collaborating with complementary projects and institutes worldwide.

2.1.1 Project Coordinator

The Project Coordinator will ensure that the project meets all its contractual obligations (including all reports and deliverables), that the participants execute the defined work plans, and that the project achieves its goals. The Project Coordinator interacts with the following bodies:

- European Commission: The Project Coordinator will be the sole liaison with the European Commission for the project.
- Project and Technical Management Boards: The Project Coordinator will chair the Project (PMB) and Technical Management Board (TMB).

Dr. Jamie Shiers (CERN) is proposed as the PODDS Project Coordinator.

2.1.2 Coordinating Partner

The Coordinating Partner is responsible for the scientific coordination, administration, and financial management of the project. The Coordinating Partner will be responsible for the distribution of the EC financial contribution to the project's partners.

CERN is the Coordinating Partner for PODDS. Having led the European DataGrid (EDG) project as well as the EGEE series of projects, it has extensive experience in European Framework Programme projects and in performing the administrative, legal and financial services necessary to ensure the effective management of the project and the coordination of the consortium.

2.1.3 Administrative Assistant

The Administrative Assistant (WP1) will help the Project Coordinator by dealing with everyday tasks related to the management, administration, and financial reporting aspects of coordinating the project. Specifically, these tasks include arranging meetings, taking minutes, and disseminating information to the project participants and partners.

2.1.4 Project Management Board

The Project Management Board (PMB) consists of a representative of each partner and is chaired by the Project Coordinator. Its purpose is to ensure that the project is on track with respect to its objectives and deliverables and guarantee the involvement of the partners during the course of the project.

Meetings will take place at least once every quarter and may be held physically or via telephone or video conferencing. Additional meetings may be called if necessary. The agenda must be provided at least one week prior to the meeting and must include a project status report from the Project Coordinator.

All Parties shall agree to abide by all decisions of the PMB. All disputes shall be submitted in accordance with the provisions of the Grant Agreement and Consortium Agreement.

Although not formally members of the PMB, we plan to use the bi-annual milestones foreseen for WP4 (road mapping) to solicit feedback from the wide network of communities with which we will be working to ensure that the work is proceeding in the correct direction. Not only will these reports be publically available but they will also be presented at appropriate conferences and workshops and to the key institutes concerned.

2.1.5 Technical Management Board

The Technical Management Board (TMB) will consist of the Project Coordinator and representatives from the work packages. The Project Coordinator will be the Chair of the TMB.

Meetings will be held fortnightly and may either take place physically or via telephone or video conferencing. The Chair of the TMB will prepare the agenda of the meeting in consultation with the members of the TMB. Minutes and actions from the meeting will be made available to participants before the next meeting.

The TMB is responsible for following the progress of the project with respect to the defined work plan, raising any issues (internal and external) encountered, and ensuring that other members are aware of significant events in each activity. The TMB is also responsible for the approval of deliverables and milestones.

Decisions will generally be made by consensus. Where no consensus can be reached the issue will be forwarded to the PMB for discussion and a decision.

2.1.6 International Advisory Committee

In order to strengthen the interaction with the targeted communities, an international advisory committee will be setup. This committee will receive early drafts of the evolving roadmap and will be involved in ensuring that the direction of the project is well aligned with their needs. Key representatives from the various communities and projects described in sections 2.3 and 2.4 will be asked to join this committee, i.e. OpenAIRE/OpenAIREPlus, DPHEP and EUDAT projects, plus the main scientific domains with whom we plan to work.

2.2 Individual Participants

2.2.1 CERN

Brief description of legal entity

CERN is an International Organisation with its headquarters in Switzerland and is the largest particle physics laboratory in the world. It is currently exploiting the Large Hadron Collider (LHC) that is now entering its third year of data taking. The LHC is the world's most powerful accelerator and provides research facilities for several thousand researchers from all over the globe. The LHC experiments are designed and constructed by large international collaborations and will collect data over a period of 10-15 years. Up to 1 million computing tasks are run per day with some 15 petabytes of data generated per year. Through a sequence of upgrades and enhancements, the LHC project is expected to continue for a number of decades and data will need to be preserved not only during its lifetime but also far beyond for the well established scientific reasons explained in sections 1 & 3. In addition, there are strong arguments to preserve data as well as the capability to use it for educational outreach: not only to teach science but also attract further generations into higher education and research (see section 3).

Main tasks in project and relevant experience

CERN will lead the proposed project as well as work package 4 (roadmap preparation). It will participate in all other work packages. Given the long lifetime of the LHC experiments and the large volume of data involved, preservation has to be addressed during the active data-taking phase. Moreover, CERN plays an active role in numerous other domains, including Life and Earth Sciences, as well as providing technical guidance to numerous International Organisations and is thus well placed to help prepare a multi-disciplinary roadmap as well as assist in outreach and dissemination activities. CERN has prominently contributed to a number of EGEE-related grid projects and currently leads the Services for Heavy User Communities work package of EGI-InSPIRE. Under FP6 and FP7, the IT department has been involved in some 20 European Commission-funded projects covering a wide range of topics. The department has been at the forefront of computing for many years in all aspects including storage and data management.

Profiles of individuals undertaking the work

Dr Jamie Shiers leads the Experiment Support group in CERN's IT department with a strong focus on the data processing analysis activities of the LHC experiments using worldwide distributed computing environments. He has participated in several European grid projects—EGEE, EGI_DS, EGI-InSPIRE (Heavy User Communities), EnviroGRIDS (Earth Science), PARTNER and ULICE (both Life Science). He has also been involved in the database, data and storage management and preservation fields, being the Vice-Chair for European Activities of the IEEE Computer Society Committee on Mass Storage Systems.

Professor Rolf-Dieter Heuer

Rolf-Dieter Heuer is the Director General of CERN, as well as chair of the EIROforum, a partnership between eight of Europe's largest inter-governmental scientific research organisations: CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL and ILL. He has held numerous senior positions in particle physics and has been a leading advocate for Data Preservation activities for many years. In his role as EIROforum chair, he is well-placed to push for coordinated Data Preservation activities across all its members as well as between the numerous communities that these organisations represent. His position as Director General of CERN also allows him to provide influence at the highest level across CERN member states and beyond as well as to other International Organisations and to the numerous disciplines with which CERN has active partnerships.

In addition, CERN plans to hire on project funds one staff with experience in data preservation for its involvement in the road-mapping (WP4) and other work packages as well as the necessary manpower to cover project administration and financial reporting (kept deliberately light-weight).

2.2.2 MPG

Brief description of legal entity

The Max Planck Society is Germany's most prominent research organization. Since its establishment in 1948, no fewer than 17 Nobel laureates have emerged from the ranks of its scientists, putting it on a par with the best and most prestigious research institutions worldwide. The more than 13,000 publications each year in internationally renowned scientific journals are proof of the outstanding research work conducted at Max Planck Institutes – and many of those articles are among the most-cited publications in the relevant field.

The currently 80 Max Planck Institutes conduct basic research in the service of the general public in the natural sciences, life sciences, social sciences, and the humanities. Max Planck Institutes focus on research fields that are particularly innovative, or that are especially demanding in terms of funding or time requirements. Their research spectrum is continually evolving: new institutes are established to find answers to seminal, forward-looking scientific questions, while others are closed when, for example, their research field has been widely established at universities. This continuous renewal preserves the scope the Max Planck Society needs to react quickly to pioneering scientific developments.

Main tasks in project and relevant experience

WP3 – Digital Preservation Outreach

MPG leads work package 3 (Digital Preservation Outreach) and will also participate in all other work packages. The Max Planck Society is particularly well suited to this task given the very wide range of disciplines that are supported by its many institutes plus its international reputation. This is complemented by the specific experience of the individuals involved in the active re-use of preserved data over a period of some decades, as detailed below.

Profiles of individuals undertaking the work

Prof. Dr. Siegfried Bethke is currently a Director of the Max-Planck-Institute of Physics (Werner Heisenberg Institute) and co-chair of the International Advisory Committee of the Study Group for Data Preservation and Long-Term Re-use in High Energy Physics (DPHEP). He is a leading expert in the scientific motivation for data preservation with hands-on experience of re-vitalising data. Given his background and positions, he is well placed to play a leading role in the outreach activities that are foreseen, pushing the case as widely as possible through multiple disciplines and institutes of the Max Planck Society and the public, as well as influencing high level funding and strategic bodies.

Dr. Stefan Kluth is a member of the ATLAS collaboration at CERN and responsible for the ATLAS Tier2 centre hosted at the Rechenzentrum Garching (RZG). Another of his research activities concerns the analysis of data from past experiments running at electron-positron colliders PETRA at DESY in Hamburg and LEP at CERN, Geneva. Both experiments have closed down many years ago, but there is still a lot of good science obtained from re-analysing data of past experiments.

Dr. John Kennedy is a Data Services manager at the RZG Computing Centre of the Max Planck Society and Max Planck Institute for Plasma Physics. Dr. Kennedy has over a decade of experience in distributed computing with his current focus being data storage, access, and long term preservation. He acts as deputy work package leader in the EUDAT operations work package and has experience in large scale data access and processing as well as Federated AAI and grid computing. Through his work at RZG, Dr. Kennedy supports a diverse set of research communities including Plasma Physics, Astrophysics, Psycholinguistics and High Energy Physics and has a deep insight into their varied data service needs.

In addition, MPG plans to hire on project funds one staff with experience in data preservation for its involvement in WP3.

2.2.3 INFN

Brief description of legal entity

INFN (http://www.infn.it/) – the National Institute of Nuclear Physics – is an organization dedicated to the study of the fundamental constituents of matter and conducts theoretical and experimental research in the fields of sub-nuclear, nuclear, and astro-particle physics. This requires the use of cutting-edge technologies and instrumentation, which the INFN develops both in its own laboratories and in collaboration with the world of industry. INFN has over 1500 research staff with an equivalent number of associates from Universities and other National Scientific Institutes.

Main tasks in project and relevant experience

INFN leads work package 2 (Digital Preservation Technologies) and will also participate in all other work packages. Since 2001 INFN has played a major role in the EU DataGrid and DataTAG projects, the WLCG project, and more recently the EGEE and EGI projects. The contribution to these projects includes the development and reengineering of many grid Middleware components as well as the operation of the INFN Production Grid, with more than 30,000 CPUs deployed at 27 sites. Of particular relevance for the WP2 is also the role played by INFN in projects promoting the use of the grid technologies in other research communities such as LIBI, BioinfoGRID, e-NMR, CYCLOPS, BioVel.

INFN is the leading partner of the Italian Grid Infrastructure (IGI) JRU, established to boost the construction of the national e-infrastructure for e-Science. IGI puts together the public research institutions and the public consortia, active in the field of grid computing, either as resource provider but also representing large research and academic user communities.

Profiles of individuals undertaking the work

Marcello Maggi has unique and extensive experience in data preservation issues, being the person responsible for the archived data and libraries of the ALEPH experiment dismounted in 2000. These data and libraries are still preserved, functioning and used. He has coordinated many activities in data analysis and has a broad experience of many different techniques in the field.

Giacinto Donvito is actively involved in the fields of Distributed Storage and Distributed Computing for several scientific communities. He has significant experience in the installation, configuration and optimization of various storage systems (Lustre, GPFS, Xrootd, Hadoop, dCache, etc.) to provide data access to scientific applications. He is the system administrator of the CMS and Alice Tier2 at INFN-Bari. He has been the technical coordinator of the LIBI project for bioinformatics and has participated to several other projects in bioinformatics (BioinfoGRID, BioVel). Since 2004 he has participated in all the EGEE and EGI projects.

Claudio Vuerli participated in the "Grid.it" multidisciplinary project funded through the FIRB and in the co-funded "DRACO" project to set up an Astrophysical Italian Grid; he was engaged in a proposal to port in EGI the simulation pipelines of the ESA Planck mission and coordinated the astrophysical community in EGEE-II and EGEE-III projects; currently he coordinates the grid-related activities in the context of the INAF Information System and since 2010 the Astrophysical community in EGI; he participates in various projects funded by the European Commission, including EGI-InSPIRE.

Giorgio Pietro Maggi is the Project Manager of the INFN Accounting Systems and the Director of the INFN "Direzione Sistema Informativo". He is a member of the coordination and executive committees of IGI (Italian Grid Infrastructure) that he has contributed to develop and integrate into the European Grid. He has coordinated the activities of the realization of the CMS Tier2 and in a series of national and European projects (EGEE, LIBI, BioinfoGRID, BioVel, ReCAS) for the establishment of distributed computing infrastructure and the promotion of the grid technologies in many different research communities. He is the local leader of the INFN-GRID project and is a member of the Executive Board of this project.

In addition, INFN plans to hire on project funds one staff with experience in data preservation for its involvement in WP2.

2.3 Consortium as a whole

The PODDS consortium consists of partners that have a long history of working together in a variety of domains, including in the development and exploitation of e-infrastructures, as part of scientific collaborations and through a variety of EU-funded and other projects. Together they not only represent a wide range of disciplines but have proven skills in the area of data preservation for re-use, including resurrection of old data for re-analysis. In terms of storage technologies and the provision of related services, they can all be considered world leaders. The Max Planck Society has been a leader in terms of innovative storage technology for many decades and has regularly presented its work at the IEEE Massive Storage Systems and Technologies conferences. Both CERN and INFN are leading providers of storage and data transfer systems that are exploited in the Worldwide LHC Computing Grid: all are committed to the provision of such services for at least the duration of the LHC programme and most likely much longer.

With respect to research communities, they represent a wide range of scientific disciplines, that includes High Energy Physics, Life Sciences, Astronomy and Astrophysics, Fusion, Molecular Biology and so forth but also, through the Max Planck Society as a whole also numerous other fields including Arts and Humanities and many others.

<u>High Energy Physics:</u> all partners are strongly involved in support for High Energy Physics (HEP). CERN is the host laboratory for the Large Hadron Collider (LHC) and is one of the main HEP laboratories in the world today. The Max Planck Institute for Physics, one of the largest Max Planck institutes is a key HEP research centre, actively participating in all phases of HEP experiments. From the initial conception, design and construction of large detectors through to their operation and subsequent analysis and publication of the resulting data, the Institute plays an active role. INFN, the Italian National Institute of Nuclear Physics, plays a leading role in all aspects of HEP and is also strongly involved in distributed computing initiatives and infrastructures.

<u>Astronomy and Astrophysics:</u> the “Istituto Nazionale di Astrofisica” (INAF) is represented in this and other projects via INFN. INAF is actively involved in all activities of astronomy and astrophysics using a variety of instruments worldwide. Such activities encompass the porting of applications on DCIs, the provision of astronomical requirements and use-cases, the creation of bridges between DCIs and astronomical data e-infrastructures (with particular reference to the Astronomical Virtual Observatory) and the long-term maintenance of legacy software trying to make it compatible with new emerging DCIs. CERN also has a strong link to astronomy and astro-particle physics, where data from CERN experiments are used to study the ordinary stellar evolution as well as supernovae. The Max Planck Society supports several large scale experiments including the MAGIC and HESS gamma ray telescopes as well as numerous satellite projects and radio and optical astronomy, including the Max Planck’s Effelsberg 100m radio telescope located near Bonn, and the international Very Large Telescope project. Moreover, large-scale simulations of the universe are undertaken by the Max Planck institute for Astrophysics to allow comparisons between theory and observation

<u>Fusion:</u> The Max Planck Society is strongly represented in the field of fusion research with two fusion experiments of its own, the ASDEX Upgrade Tokamak experiment located in Garching near Munich and the Wendelstein 7-x Stellarator located in Greifswald. In addition the Max Planck Society participates in the international JET and ITER projects.

<u>Life Sciences:</u> both through the use of grid and other distributed computing infrastructure technology CERN is actively involved in a variety of life science projects. These include hadron-therapy projects, such as PARTNER and ULICE, as well as EGI-InSPIRE and its explicit support for the Life Science community (Life Science Grid Community – LSGC). The Max Planck Society is also strongly involved in Life Science activities, through its numerous institutes. INFN supports numerous Life Science Initiatives through the Italian Grid Infrastructure, through the Laboratorio Internazionale di BioInformatica (LIBI) project, through the tps.infn.it hadron therapy treatment planning centre, through its involvement in the EU-FP7 project eNMR and so on.

Molecular Biology: The Max Planck Society is active in a diverse range of projects in the field of Biological Research. Projects in the field of large scale genome sequencing including Evolutionary Anthropology (Leipzig), Plant Breeding Research (Koln) and Molecular Genetics (Berlin) are complemented by projects in areas such as Proteomics and Molecular Structural biology (Martinsried).

Other Disciplines: The Max Planck Society prides itself on performing a broad range of basic research in the fields of life sciences, natural sciences as well as social and human sciences. In addition to the fields already mentioned the Max Planck Society conducts research in areas such as art history, psycholinguistics, ornithology, meteorology and medicine. The diversity of research undertaken within the Max Planck Society as well as its participation in numerous national and international projects places it in an ideal position to perform outreach activities for long-term data management and preservation.

2.3.1 Complementarity and Completeness

Each partner has a well-defined role within the project that matches well their interests and experience. Together, these cover the full spectrum of the proposed project plan as described in the work plan (section 1). Furthermore, there is complementarity in the networks that each partner is able to mobilize that will allow us to achieve the desired impacts (section 3). In summary, INFN will be responsible for the work related to Digital Preservation Technologies (WP2), matching well their knowledge and experience; the Max Planck Society will be responsible for Digital Preservation Outreach (WP3), making the most of the very broad spectrum of activities of the Organisation as a whole (close to 80 institutes worldwide); CERN will be responsible for coordinating the development of the Roadmap (WP4).

2.3.2 Other Related Initiatives

We list below some of the main projects related to digital preservation for re-use and/or those that will be important for gathering of requirements and outreach. The table is not intended to be exhaustive but highlight some of the main related projects and activities.

Project	PODDS partners involved	Description
APARSEN	CERN	This Network of Excellence gathers a very diverse set of practitioner organisations and researchers in order to bring coherence, cohesion and continuity to research into barriers to the long-term accessibility and usability of digital information and data, and to build a long-lived Virtual Centre of Digital Preservation Excellence. Joint research activities will cover technical methods for preservation, access and re-use of data holdings over the whole lifecycle; legal and economic issues including costs and governance as well as digital rights; and outreach within and outside the consortium to help to create a discipline of data curators with appropriate qualifications.
EUDAT	CERN, MPG	The “EUropean DATa” project is a three-year project that will deliver a Collaborative Data Infrastructure (CDI) with the capacity and capability for meeting future researchers’ needs in a sustainable way. It aims at providing a variety of community-oriented services

		that include long-term preservation and thus there is a clear link between the two activities. PODDS is complementary activity (as with OpenAIRE/OpenAIREPlus below) with a strong focus on addressing the software and metadata aspects required for continued use and re-use of data for the three primary use cases outlined in section 1.
DPHEP	CERN, INFN, MPG	An ICFA-endorsed study group on Data Preservation and Long-Term Analysis in High Energy Physics. This international working group, which involves numerous collaborations and major laboratories, is preparing a "Blueprint document" which defines the key requirements and use cases for HEP. It also has initiated discussions with other communities, including Astronomy and Astrophysics.
DASPOS	(CERN)	The Data And Software Preservation for Open Science (DASPOS) proposal has as a primary goal a "curation challenge", whereby a scientist from one of the LHC collaborations would be able to perform an analysis on the data of another LHC experiment. As such, it builds on the work done by DPHEP in defining the requirements in the area of Data Preservation for HEP. (CERN is not formally a member of this collaboration but is the host laboratory for the LHC experiments and is thus indirectly involved).
OpenAIRE/ OpenAIREPlus	CERN	OpenAIRE and OpenAIREPlus are projects oriented at providing persistent Open Access for publications and linking such publications with the scientific data involved. They are therefore complementary projects: PODDS will focus on the initial data and the possibility to re-use it whereas OpenAIRE/OpenAIREPlus focus on the publication phase.
EGI-InSPIRE and other EU-funded projects	CERN, INFN, MPG (various)	All PODDS partners are part of the EGI-InSPIRE project and all are involved in a number of other EU-funded projects. EGI-InSPIRE has dedicated support tasks for High Energy Physics, Life Sciences, Earth Sciences and Astronomy and Astrophysics. These contacts allow us to obtain requirements from the relevant communities as well as to disseminate the results of the project.

2.4 Resources to be committed

PODDS is proposed as a 24 month project with 3 funded partners, each of whom is requesting 24 person-months each, i.e. a total of 72 person months. The EU funded resources will be assigned to the partners based on the work package breakdown in section 1, with each partner allocating 80% of the funded resources to the work package that they lead with the remaining 20% spread over the other two technical work packages.

To this will be added about the same number of person months from senior personnel from the institutes involved mainly through the named experts listed in section 2.2. In particular, the coordination of the project will be funded directly by CERN at a cost of 20% of a senior member of personnel. As the costs of these personnel are higher than those that we plan to hire, the consortium members would be contributing significantly more than 50% to the personnel costs of the project.

The partners will make available free of charge the necessary office space and infrastructure to support the funded personnel, together with the necessary IT resources needed by any test-beds (INFN), outreach and dissemination or other such costs.

Taking the personnel and infrastructure resources together, it is estimated that the partners will be contributing two thirds of the total cost of the project.

In addition to these resources, that can be considered an integral part of the project, numerous other associated partners have expressed their interest and commitment to provide input regarding the requirements and feedback on the evolving roadmap. These include:

- The EIROforum members;
- The Science and Technology Facilities Council (STFC) in the UK;
- Representatives of the Life and Earth Sciences, Astronomy and Astrophysics communities, as indicated in section 2.3 above;
- The DPHEP and DASPOS projects.

It is expected that, as the project progresses, the efforts of WP3 in particular will attract more unfunded resources to the project.

Finally, the extensive networks that the PODSS project will mobilise to achieve the desired impact are described in detail in section 3.2.

3 IMPACT

3.1 Expected impacts listed in the work programme

We list below the expected impacts of the project in order of decreasing priority (most important first). Although all are considered important, the largest impact that is expected is in terms of wider adoption of research results. In addition, through the existing techniques that we will document, it is expected that the preserved resources will remain meaningful and usable for longer periods than would otherwise be the case. However, given the lifetime of the proposed project, this will clearly extend into the period when the project is over but can still be considered an important outcome. Finally, by highlighting the appropriate techniques and the use cases that they address, we would attempt to improve the selection of which resources are candidates for preservation.

The international nature of modern research means that all of the use cases that we intend to address need to be solved at the European if not international level: this includes the explicit examples that we have given in the area of e-health records (use case 1), scientific and associated data (such as metadata and software libraries) for eventual re-analysis (use case 2) and finally the preservation of such data for educational outreach purposes. A true solution to the problems faced by the proposal can only be found if we address the international research communities as a whole and ensure that data preservation and re-use become a fixed part of their data lifecycles.

Although not explicitly listed as an expected impact, we would also like to emphasise the intended value of the foreseen roadmap in terms of making concrete and measurable steps towards sustainable services for long-term preservation in the full spectrum of disciplines that are foreseen.

We explain in more detail below how each of these impacts will be achieved, including the various steps that are involved.

3.1.1 Roadmap

The roadmap – one of the key deliverables of the project – will encapsulate the key requirements and use cases of the various user communities, summarise the state of the art in terms of digital preservation for future re-use and make concrete proposals for the next steps in achieving long-term service offerings in this area. An expected outcome would be to prototype such services at a convincing scale demonstrating not only that the key technical requirements were met but also confirming (or otherwise) a small number of realistic funding and operational models. The roadmap, as stated before, should not only have the input from a wide range of communities but also their explicit endorsement, including the EIROforum members and other representative institutes and bodies. The roadmap is to be used to align the data preservation plans of the varied research communities and to provide direction for the momentum which has been generated during the project lifetime. The true impact of the roadmap will be experienced in the years following the project as it paves the way to projects and e-infrastructures where long term data preservation is guaranteed and the data that we generate today is still available for tomorrow's research.

3.1.2 Wider Adoption of Research Results

This topic is strongly related to the activities of WP2 (Digital Preservation Technologies) and WP3 (Digital Preservation Outreach). However, it is clearly expected that an outcome of WP4 (Road map Preparation) would be yet further adoption of these results in a sustainable and long-term fashion. Both aspects of wider adoption are essential: by the "supply industry" as well as by end-users.

With our strong networks within many user communities, we believe that we are very well placed to motivate and achieve much wider adoption of research results – this will be done through the many channels explained in section 3.2.

Equally challenging is that of the supply industry. Here, our contacts through technology related events, such as the IEEE MSST (see section 3.2) will high-light the requirements and pressure will be put on leading vendors to address appropriate parts of the overall problem domain, some of which are addressed by other topics of the current call.

The supply industry in this context could include not only commercial suppliers but also projects such as EUDAT, which include data preservation as one of the use cases that they wish to address.

On the other hand, neither commercial providers nor other projects are currently addressing the specific issues that we intend to stress, that include the retention of meta-data and software libraries and databases that are an integral and essential part of digital preservation for long-term re-use.

3.1.3 Sustainable Access to Information

One of the key issues that the proposed project addresses is that of keeping resources not only available but also meaningful and usable. Whilst there are many technical issues in retaining the bits over long periods of time, the ability to be able to re-use the data is central to our proposal. Depending on the use case(s) involved, the degree of complexity varies (see section 3.1.4 below). To take the most complex use case – that is to be able to exploit the full potential of the (experimental, in this example) data has significant implications. If full flexibility is to be retained, i.e. to cover the use cases mentioned in section 1 (*when called for by new theoretical insights, new or improved analysis and/or simulation techniques, or to search for previously unseen signals observed at a newer facility*), decisions must preferably be taken early on in the lifecycle of the experiment involved and significant effort expended in software quality and documentation. This is the foundation for the preservation of the analysis and simulation chains, together with verification tools, for migrations to new hardware and software environments. These are issues that cannot be addressed by industry suppliers and / or generic storage providers but require careful design and consideration by the communities involved. As described in section 1, particular attention to these details is a differentiating factor of our proposal. Given the focus that we plan to place on these issues, we expect our project to have a unique impact in this area.

3.1.4 More Efficient Selection of Resources to be preserved

Through better definition of requirements and the associated use cases, we intend to provide guidance related to the selection of which data and associated information / software need to be preserved and hence the appropriate preservation processes, methods and technologies. For example, for educational outreach a highly simplified representation of the data is appropriate, whereas for full re-analysis a much more complex environment needs to be preserved. A clear summary of the needs of the HEP community in this domain has been made by the DPHEP consortium and is summarised below. This needs both to be confirmed as well as extended for all other disciplines that we intend to cover.

For example, the DPHEP blueprint document (in preparation) defines the following levels:

Preservation Model	**Use Case**
Provide additional documentation	Publication-related information search
Preserve the data in a simplified format	Outreach, simple training analyses
Preserve the analysis level software and data format	Full scientific analysis based on existing Reconstruction

Preserve the reconstruction and simulation software and basic level data	Full potential of the experimental data

Depending on the use case that needs to be addressed, different data needs to be preserved and different levels of complexity are involved. Thus, if the need is for Outreach, the data need only be preserved in a simplified format, typically not only in complexity but also in data volume as well.

3.2 Spreading excellence, exploiting results, disseminating knowledge

In order to increase the impact of the project, we will use the channels described below. Specifically, we intend to target key disciplines such as High Energy Physics, Life Sciences, Astronomy and Astrophysics, Fusion, Molecular Biology and also numerous other fields including Arts and Humanities and many others, raising the awareness of the issues and of the results that we have obtained, whilst seeking further input on requirements and future directions. This would be done via the direct contacts that we have with those communities, as described in section 2.3 above. Furthermore, by working through the EIROforum, Max Planck and other institutes we would aim to increase awareness further still and at the highest possible level. Via publications, presentations and press releases we would additionally target the media as well as the public at large.

3.2.1 EIROforum

As stated on the EIROforum web page:

"EIROforum is 'a partnership between eight of Europe's largest inter-governmental scientific research organisations' that are responsible for infrastructures and laboratories: CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL and ILL.

It is the mission of EIROforum to combine the resources, facilities and expertise of its member organisations to support European science in reaching its full potential."

In addition to working through the EIROforum members and thus reaching the key communities that they represent, the EIROforum represents an excellent channel to address issues such as Science Policy, Education and Outreach, all of which are essential to our mission.

The Director General of CERN is currently the chair of the EIROforum (until June 2012) and is thus well placed to raise awareness of this issue in a broad range of research fields through the EIROForum and its members.

3.2.2 MPG

The Max Planck Society consists of some 80 institutes worldwide that cover a very wide range of disciplines. The institutes are specialised in disciplines ranging from natural sciences, life sciences, social sciences to the arts and humanities. Working through the Organisation we will thus be able to reach a very large number of disciplines and communities. In addition, the renown of the organisation will similarly enable us to reach policy makers as well as media and the public at large.

3.2.3 INFN

INFN research activity is carried out at two complementary types of facilities: the 19 Divisions (Sezioni) each located at a university physics department, a national computing facility (CNAF), and the four National Laboratories that provide major facilities available to the national and international scientific community. The Divisions thus provide a direct connection between the Institute and the academic world. INFN is also a precursor and incubator of interdisciplinary activities thanks to the developed techniques that have found broad applications in informatics, health, defence, cultural heritage and environment preservation. INFN has a long tradition of dissemination and training, dedicated both to the scientific community and to the general public.

Thus, in this project, INFN will not only provide additional channels to reach a wide reach of communities but also with its leading role in various international projects will allow us to disseminate and exploit the project's results wider still.

3.2.4 CERN

With the third year of LHC operation just commencing, and with the promise of exciting scientific results, CERN is very much in the public eye. In addition to its leading role in the domain of High Energy Physics, CERN is also an active partner in numerous other projects, including Earth Science, Life Science, network, grid and cloud technology and so forth. It also has the fundamental need to find and provide solutions for long-term data preservation for re-use that match at least the timescales foreseen for LHC exploitation. Including possible upgrades of the accelerator complex, this could amount to a few decades.

Through targeted press releases, via presentations to visiting officials, through publication channels, conferences as well as CERN's worldwide multi-disciplinary network, the results of the proposed project would be further disseminated.

3.2.5 Other Data Preservation Activities / Projects

The International Committee for Future Accelerators (ICFA) has endorsed a study group on Data Preservation and Long-Term analysis in High Energy Physics (DPHEP).

The objectives of the Study Group are to:

- Review and document the physics objectives of the data persistency in HEP;
- Exchange information concerning the analysis model: abstraction, software, documentation etc. and identify coherence points;
- Address the hardware and software persistency status;
- Review possible funding programs and other related international initiatives;
- Converge to a common set of specifications in a document that will constitute the basis for future collaborations.

All institutes involved in the PODDS proposal are active members of the DPHEP consortium. Through the DPHEP "blueprint" document, its various outreach and dissemination activities, which reach the highest levels in European and International funding organisations, we will not only raise awareness of the importance of these activities but also seek active guidance on the direction of the project. Whilst its mandate primarily concerns High Energy Physics, DPHEP also has active participation from other disciplines, including Astronomy and Astrophysics.

3.2.6 International Conferences, Symposia and Workshops

Members of the PODDS consortium and the communities that they represent are active participants in a wide range of international events. These provide an excellent means of further disseminating the results of the project and seeking further input and direction. We list below some of the key events that are currently targeted as well as concrete actions in each case.

- The EUDAT vision is to support a Collaborative Data Infrastructure which will allow researchers to share data within and between communities. During the project's 3 year lifetime EUDAT will organize numerous conferences including user forums, collaboration wide conferences and smaller, more focused, workshops. Through these EUDAT will bring together a broad range of communities and infrastructure providers and so will provide an excellent opportunity for the PODDS project to further the awareness of the need for data preservation and more specifically data re-usage. The conferences will take place throughout 2012,2013 and 2014 and are as such well aligned with the time-frame of the PODDS project.

- The IEEE Massive Storage Systems and Technology (MSST) symposium takes place every year and is widely attended by industry, technology researchers as well as service providers. Panels and presentations on data preservation issues, ranging from hardware technology to discipline-focussed requirements, have been held at the 2009, 2010 and 2011 symposia. These events represent not only an excellent opportunity to reach out to international researchers but also leading service and technology providers.

- The IEEE Nuclear Science Symposium (NSS) & Medical Imaging Conference (MIC) is an annual event involving some 1500 delegates from the High Energy and Astrophysics and Medical Communities worldwide. In contrast to the IEEE MSST, its focus is more on the science rather than the technology providers. It offers an example opportunity to reach multiple communities and share information with them: focussed workshops are an integral part of the events and multi-disciplinary data preservation would be a welcome topic. The deadline for abstract submissions for workshops to be co-located with this event is May 2012 and such a proposal is in preparation.

- The European Grid Infrastructure (EGI) delivers integrated computing services to European researchers. EGI's mission is to allow researchers of all fields to make the most out of the latest computing technologies for the benefit of their research. EGI is a federation of over 350 resource centres and coordinated by EGI.eu, a not-for-profit foundation created to manage the infrastructure on behalf of its participants: National Grid Initiatives (NGIs) and European Intergovernmental Research Organisations (EIROs). EGI-InSPIRE (EGI-Integrated Sustainable Pan-European Infrastructure for Research in Europe) is a four-year project helping to establish EGI as an e-infrastructure to support research and innovation in Europe. EGI-InSPIRE is a collaborative effort involving dozens of partners in Europe and beyond. The project is co-funded by the European Commission's 7th Framework Programme (contract number: RI-261323) to help lay down the EGI operational and support processes, as well as to build a sustainable e-infrastructure, independent from project cycles. EGI-InSPIRE organises and runs two major events per year, the EGI Community Forum (CF) and Technical Forum (TF). All of the PODSS partners are actively involved in EGI-InSPIRE, the 2012 CF being held on the Garching campus that houses also several Max Planck Institutes, the Garching Computing Centre (RZG) and several EIROforum members. At all EGI events so far, workshops and sessions have been held on gathering requirements regarding data preservation. So far, this has led to a draft requirements matrix across the disciplines that have explicit support tasks in EGI-InSPIRE WP6 "Services for Heavy User Communities" (led by CERN), namely High Energy Physics, Life Sciences, Earth Sciences and Astronomy and Astrophysics. Future EGI fora would be used similarly to broadcast the results of the project, present interim roadmaps, gather further requirements and disseminate widely also to other communities and service providers involved in EGI.

- The International Conference on Computing in High Energy and Nuclear Physics (CHEP) is held every 18 months and rotates between Europe, the Americas and the rest of the world. In addition to delegates from these fields it also attracts people from industry as well as other key scientific disciplines. It typically involves some 600-700 delegates. At the forthcoming CHEP 2012 event, to be held in May 2012 in New York, there will be a plenary presentation on the results of the DPHEP organisation as well as a half-day parallel session on data preservation activities. For the following event, to be held in October 2013 in Amsterdam, a half-day plenary session is foreseen to cover not only the activities in HEP but also those in other disciplines as well as the response from industry. Members of the PODDS consortium are actively involved in the organisation of these events.

- The High Energy Physics Unix Information Exchange (HEPiX) forum brings together worldwide Information Technology staff, including system administrators, system engineers, and managers from the High Energy Physics and Nuclear Physics laboratories and institutes, to foster a learning and sharing experience between sites facing scientific computing and data challenges. Participating sites include BNL, CERN, DESY, FNAL, IN2P3, INFN, JLAB, NIKHEF, RAL, SLAC, TRIUMF and many others. The HEPiX organization was formed in 1991, and its semi-annual meetings are an excellent source of information and sharing for IT experts in scientific computing. HEPiX will be a natural place to focus on particularly the technology-oriented aspects of our work whilst also addressing key institutes and that are involved.

3.2.7 Media

- International Science Grid This Week (iSGTW) is an international weekly online publication that covers distributed computing and the research it enables. It reports on all aspects of distributed computing technology, such as grids and clouds. It also regularly features articles on distributed computing-enabled research in a large variety of disciplines, including physics, biology, sociology, earth sciences, archaeology, medicine, disaster management, crime, and art. It is a widely read e-publication that is well suited to dissemination and outreach regarding the progress and results of the activities of the proposed project. It reaches both technical staff as well as senior managers in research institutes worldwide.

- The CERN Courier is an International Journal of High Energy Physics that is published both on paper and also online via the Institutes of Physics. It is widely distributed to many major institutes and universities worldwide. Key articles published in the Courier are picked up by the online and printed press. Given the level of this publication, we would expect to publish an article on an approximately annual basis, coinciding with a key event or publication.

3.2.8 Summary

We have listed above some of the main channels that we intend to use for the dissemination and exploitation of project results.

Through these measures, we believe that we can effectively address policy-makers and funding agencies, a very wide range of disciplines as well as the media and the public at large.

4 ETHICAL ISSUES

Ethical Issues Table

	YES	PAGE
Informed Consent		
• Does the proposal involve children?		
• Does the proposal involve patients or persons not able to give consent?		
• Does the proposal involve adult healthy volunteers?		
• Does the proposal involve Human Genetic Material?		
• Does the proposal involve Human biological samples?		
• Does the proposal involve Human data collection?		
Research on Human embryo/foetus		
• Does the proposal involve Human Embryos?		
• Does the proposal involve Human Foetal Tissue / Cells?		
• Does the proposal involve Human Embryonic Stem Cells?		
Privacy		
• Does the proposal involve processing of genetic information or personal data? (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)		
• Does the proposal involve tracking the location or observation of people?		
Research on Animals		
• Does the proposal involve research on animals?		
• Are those animals transgenic small laboratory animals?		
• Are those animals transgenic farm animals?		
• Are those animals cloned farm animals?		
• Are those animals non-human primates?		
Research Involving Developing Countries		
• Use of local resources (genetic, animal, plant etc)		
• Impact on local community		
Dual Use		
• Research having direct military application		
• Research having the potential for terrorist abuse		
ICT Implants		
• Does the proposal involve clinical trials of ICT implants?		
I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	**YES**	

5 ANNEX: GLOSSARY

ALEPH	Apparatus for LEP Physics at CERN
ALICE	A Large Ion Collider Experiment
APARSEN	“Alliance Permanent Access to the Records of Science in Europe Network”
ARGUS	The Argus Authorization Service development started as part of the European Enabling Grids for E-sciencE (EGEE) collaboration. Currently the Argus project is part of the European Middleware Initiative (EMI).
BioinfoGRID	This project aims to connect many European computer centres in order to carry out Bioinformatics research and to develop new applications in the sector using a network of services based on futuristic Grid networking technology that represents the natural evolution of the Web.
BioVel	Biodiversity Virtual e-Laboratory
CEmon	Provides a common interface for publishing information about a computing element in your network.
CERN	European Organization for Nuclear Research
CH	Switzerland
CHEP	Computing in High Energy Physics conference series
CMS	Compact Muon Solenoid
CPU	Central Processing Unit
CREAM-CE	Computing Resource Execution And Management – Computing Element
CYCLOPS	Surveillance and Security tool.
D*X.X*	Deliverable *X.X*
DataTAG	Project aimed at creating a large-scale intercontinental Grid testbed
DE	Germany
dCache	A mass storage solution
DGAS	DGAS (Distributed Grid Accounting System) is the accounting system deployed in the Italian Grid Infrastructure
DPHEP	Study Group for Data Preservation in High Energy Physics
DRACO	Datagrid for Italian Research in Astrophysics and Coordination with the virtual Observatory
EDG	European DataGrid
EGEE	Enabling Grids for E-sciencE
ESA Planck	European space observatory studying Cosmic Microwave Background
e-health	Healthcare practice supported by electronic processes and communication
e-Infrastructure	Infrastructure that allows researchers to easily make use of the technology that underlies this complex and heterogeneous ecosystem.
e-NMR	e-NMR is deploying and unifying the **NMR computational infrastructure** in system biology, a project funded under the 7th framework programme of the European Union (Contract no. 213010 - e-NMR). NMR plays an important role in life sciences (biomolecular NMR), and structural biology in particular, at both European and

	international levels
e-Science	Healthcare practice supported by electronic processes and communication
EGI	European Grid Infrastructure
EGI CF	EGI Community Forum
EGI TF	EGI Technical Forum
EGI_DS	EGI Design Study
EGI-InSPIRE	EGI Integrated Sustainable Pan-European Infrastructure for Researchers in Europe
EFDA-JET	European Fusion Development Agreement – Joint European Torus
EELA1&2	E-Infrastructure shared between Europe and Latin America
EIROforum	A partnership between eight of Europe's largest inter-governmental scientific research organisations.
EMBL	European Molecular Biology Laboratory
EnviroGRIDS	Building Capacity for a Black Sea Catchment Observation and Assessment System supporting Sustainable Development.
ESA	European Space Agency
ESFRI	European Strategy Forum on Research Infrastructures
ESRF	European Synchrotron Radiation Facility
ESO	European Organisation for Astronomical Research in the Southern Hemisphere
ETICS1&2	e-Infrastructure for Testing, Integration and Configuration of Software
EU	European Union
EU DataGrid	Aims to build the next generation computing infrastructure providing intensive computation and analysis of shared large-scale databases
EUDAT	European Data Infrastructure
EUAsiaGrid	Sustaining and enabling interoperability between Europe and Asia
EUChinaGRID	Sustaining and enabling interoperability between Europe and China
EUIndiaGrid1&2	Sustaining and enabling interoperability between Europe and India
EUmedGRID	Sustainability of e-Infrastructures across the Mediterranean
European XFEL	A research facility currently under construction in the Hamburg area, Germany
FIRB	Foreign Investment Review Board
FNAL	Fermi National Accelerator Laboratory
FP*X*	Framework Programme *X*
Glue Schema	Abstract modelling for Grid resources and mapping to concrete schemas that can be used in Grid Information Services
GPFS	The General Parallel File System from IBM
GridICE	A distributed monitoring tool designed for Grid systems
GRID.IT	A National Italian Project on Enabling Platforms for High-performance Computational Grids
GridCC	Providing a real-time Grid for distributed instrumentation
Hadoop	Aims to provide an open-source implementation of frameworks for reliable, scalable, distributed computing and data storage

HEPiX	High Energy Physics Unix Information Exchange forum
IEEE	Institute of Electrical and Electronics Engineers
IEEE MIC	IEEE Medical Imaging Conference
IEEE MSST	IEEE Mass Storage Systems Technical committee
IEEE NSS	IEEE Nuclear Science Symposium
ICT	Information Communication Technology
IGI	Italian Grid Infrastructure
ILL	The *Institut Laue-Langevin* is an international research centre based in Grenoble, France that operates one of the most intense neutron sources in the world
INAF	National Institute for Astrophysics
INFN	National Institute of Nuclear Physics
ISGC	International Symposium of Grid Computing
IT	Information Technology
IT	Italy
JRU	Joint Research Unit
LEP	Large Electron Positron Collider
LHC	Large Hadron Collider
LIBI	The LIBI Grid Platform: an International Laboratory for Bioinformatics
Lustre	Parallel distributed file system
(M*X*)	Month *X*
MGT	Management of the consortium
MPG	Max Planck Society
MS*XXX*	Milestone *XXX*
OGF-EU	http://ogfeurope.eu/
OMII-EU	http://omii-europe.omii.ac.uk
PARTNER	Particle Training Network for European Radiotherapy
PB	Peta Byte
PC	Project Coordinator
PMB	Project Management Board
PODDS	Preservation Of Data for the Digital Society
ReCAS	
SRM	Storage Request Manager – a protocol widely used in HEP to access storage elements
SToRM	A storage resource manager for generic disk based storage systems separating the data management layer from the underlying storage systems.
SUPP	Support Action
TMB	Technical Management Board
ULICE	Union of Light Ion Centres in Europe
VOMS	Virtual Organisation Membership Service

WLCG	Worldwide LHC Computing Grid
WMS	Workload Management Service
WP*X*	Work Package *X*
Xrootd	Project which aims at giving high performance, scalable fault tolerant access to data repositories of many kinds