Project Acronym:	INDICATE
Project Title:	International Network for a Digital Cultural Heritage e-Infrastructure
Contract Number:	261324
Starting date:	1 September 2010 **Ending date:** 31 August 2012

Deliverable Number:	D 5.1
Title of the Deliverable:	Case study report – Long Term Preservation
Task/WP related to the Deliverable:	Task 5.1 Long Term Preservation
Type (Internal or Restricted or Public):	Public
Author(s):	Bahadir Aydinonat, Roberto Barbera, Sergi Fernandez, Antonella Fresa, Ayman Khoury, Stefanos Kollias, Hakan Koray Ozluk, Mercè López, Maurizio Messina, Marzia Piccininno, Sabrina Tomassini, Franc Zakrajsek
Partner(s) Contributing:	All

Contractual Date of Delivery to the CEC:	month 12
Actual Date of Delivery to the CEC:	October 2011 – updated on January 2012

Project Co-ordinator

Company name : Istituto Centrale per il Catalogo Unico (ICCU)
Name of representative : Rosa Caffo
Address : Viale Castro Pretorio 105, I-00185 Roma
Phone number : +39.06.49210427
Fax number : +39.06. 06 4959302
E-mail : rcaffo@beniculturali.it
Project WEB site address : http://www.dc-net.org

Context

WP 5	Research Case Studies
WPLeader	I2CAT
Task 5.1	Task 5.1 Long Term Preservation
Task Leader	MCT
Dependencies	All WPs
Starting date	September 2010
Release date	October 2011 – updated on January 2012

Author(s)	Bahadir Aydinonat, Roberto Barbera, Sergi Fernandez, Antonella Fresa, Ayman Khoury, Stefanos Kollias, Hakan Koray Ozluk, Mercè López, Maurizio Messina, Sama Moustafa, Marzia Piccininno, Sabrina Tomassini, Franc Zakrajsek
Contributor(s)	All
Reviewers	Roberto Barbera, Franc Zakrajsek
Approved by:	Rossella Caffo (ICCU)

HISTORY

Version	Date	Author	Comments
01	2 August 2011	Sergi Fernandez	First version, index and structure
02	2 August 2011	Antonella Fresa	First version chapter 3 and chapter 4
03	3 August 2011	Sergi Fernandez	Re-structuration and responsabilities
04	4 August 2011	Mercè Lopez, Sergi Fernandez	Updated executive summary and chapter 6.
05	8 August 2011	Bahadir Aydinonat	Added 2.1, 2.4, 5.3, 6.4.7
06	10 August 2011	Bahadir Aydinonat	Change in TOC, 2.4, 2.4.1, 2.4.2, 2.4.3 and annex.
07	11 August 2011	Bahadir Aydinonat	Added 10 Conclusion
08	17 August 2011	Hakan Koray Ozluk	Added 5. Magazzini Digitali

09	6 September 2011	Antonella Fresa	Added section 2.2 Added section 4.1 Added chapter 7 Added Annex 1 (integrated and edited into the document content from Bahadir file) Added Annex 2
10	7 September 2011	Sabrina Tomassini Roberto Barbera Mercè López Sergi Fernandez	Updated executive summary Updated chapter 6 General review Change in TOC, 6.3, 6.4
11	8 September 2011	Sergi Fernandez	General review
12	13 September 2011	Frank Zakrajsek Sabrina Tomassini Mercè López	Updated chapter 6.3 General review
13	14 September 2011	Ayman Khoury Mercè López Frank Zakrajsek	Added chapter 8 Updated chapter 6.3.6.
14	19 September 2011	Stefanos Kollias	Added chapter 6.3.3
15	20 September 2011	Marzia Piccininno Sama Moustafa Mercè López	Updated chapter 4.3 Added 6.3.1 General review Removed chapter 9 Removed 8.2.5
16	21 September 2011	Mercè López Bahadir Aydinonat	Updated 8.2 and conclusions General review
17	29 September 2011	Roberto Barbera	Internal review
20	10 October 2011	Antonella Fresa	Final versions submitted to EC
21	27 January 2012	Antonella Fresa	Added bibliography

Table of Contents

1 Executive summary

This document is the result of the research done by the INDICATE partners through the Long Term Preservation Case Study. The main objective of this use case is to review the current situation of digital preservation processes in Europe, the state of the art of the technology used and the relation between preservation institutions and e-Infrastructure providers.

This document is composed by 9 sections:

Section 2 gives an overview of the whole use case, its relationship with other projects and the actions that have been taken to gather all the information required (surveys, workshop, etc.).

Section 3 explains in deep the digital preservation scenario, its threats, strategies, standards and ICT services.

Section 4 gives an overview of the current state of the art and reviews the majority of European projects related to digital preservation and associated technologies, with special emphasis in persistent identification for preservation.

Section 5 introduces two examples of digital preservation systems: Magazzini Digitali from Italy and the National Library of Turkey.

Section 6 explains the relationship between preservation institutions and e-Infrastructures providers, gives a complete overview of the current status of the e-Infrastructures at global level and details preservation institutions related to e-Infrastructure providers in each of the INDICATE partner countries.

Section 7 presents a paper about the problems of IPR in the preservation field.

Section 8 points out the importance of staff training on digital preservation.

Section 9 presents the conclusions of the use case.

The document is completed by two annexes, detailing the survey results country by country and the conclusions of the DC-NET Conference on Preservation.

2 Introduction

2.1 Scope of the case study

The INDICATE project mainly focuses on the use of e-Infrastructures for digital cultural heritage in countries all around the Mediterranean. Establishing a network made up of experts, researchers in the relevant fields in Mediterranean, the project helps to share experience, promote standards and guidelines, and seek harmonization of best practices and policies[1].

This Long Term Preservation Case Study examines the current situation of preservation among the established working groups; analyses the current status of e-Infrastructures in the partner countries and how e-Infrastructures can be used by the cultural sector.

The general aim of the study is to gather information and communication technology requirements, to identify capabilities of e-Infrastructures, to match requirements to capabilities, to identify gaps, issues, and problems at technology and other levels, to investigate how these issues may be addressed also by looking at examples of best practices and to contribute results to the dissemination work-package.

To investigate these, a workshop about Long Term Presentation was held in Ankara, Turkey in July 2011. A survey was then prepared to review preservation experiences in various countries for this workshop. National and International participants discussed practice examples, uses of e-infrastructure in digital preservation and shared policies and strategies.

2.2 The liaison with DC-NET

INDICATE is a sister project of DC-NET (www.dc-net.org). DC-NET is an ERA-NET, started in December 2009, whose main objective is to agree on a joint action plan among the participating countries in the area of the digital cultural heritage and the development of priority services enabled by e-infrastructures. Eleven European countries participate to DC-NET, eight being the original partners of the project and three having joined the network more recently.

DC-NET has identified the digital preservation as the first priority and a conference has been devoted to the discussion about the needs of the cultural sector in terms of digital preservation and how the e-infrastructures can contribute to this priority. The conference took place in Budapest on 23-24 June 2011 under the aegis of the Hungarian Presidency of the European Union. As a basis for the follow-up of the conference, a text of conclusions was approved and endorsed by all the participants. All the presentations delivered at the conference and the texts of the conclusions are available on the DC-NET project website.

There is a natural liaison between the DC-NET ERA-NET and the INDICATE project. For this reason, it has been considered relevant and complementary with the work done in INDICATE the results of the Second DC-NET Conference, held in Budapest, especially dedicated to the theme of preservation. The conference ran for two days with the participation of experts from all over Europe and representatives of the European Commission, both from the e-infrastructure Unit and the Competitiveness and Innovation Programme ICT Policy Support Programme (CIP ICT PSP) Digital Libraries.

It is worth mentioning here that the work done by DC-NET and INDICATE cannot be disconnected from the work done in the frame of the projects supporting Europeana (www.europeana.eu). Europeana is the flagship initiative referred by the Digital Agenda for Europe to create the European portal for culture.

To this regard, it is worth mentioning that the DC-NET conference followed the Linked Heritage conference which was held the day before, on the 23th of June 2011. Linked Heritage is a Best Practice Network about the coordination of standards and technologies for the enrichment of Europeana.

This European background is very important to contextualise the INDICATE action that is indeed not an isolated initiative, but it is instead part of a strategy of the European Member States, which in INDICATE starts to share it with the neighbouring Mediterranean countries.

The first day of the DC-NET conference was devoted to present the various aspect of the main theme. The working day started with the keynote speeches that introduced the issue of the digital preservation from different perspectives: the cultural authorities (i. e., Ministries of national cultural resources) who are investing in the digitisation and who want to preserve their investments; the European Commission, namely the e-infrastructure Unit, who is supporting the creation of digital infrastructures and wants to make it available and exploitable also by the researches in the humanities; the national research and educational networks that are looking for new domains to enlarge the potential of their e-infrastructures. Then, the DC-NET ERA-NET achievements were presented and eventually the actual needs of the memory institutions were discussed more in depth with some specific experience from Hungarian archives, national library and museums.

The second day was devoted to the analysis of three best practices from Austria, France and Switzerland. Then, technical and legal aspects of the long term preservation were discussed. The conference terminated with an animated round table were all the subjects discussed along the conference were revisited in the light of a future possible joint plan of activities to promote coordinated efforts for more tailored e-Infrastructure services made available to cultural heritage organisations. The conference terminated with the approval and the endorsement of the conclusions by the participants. Full text of the DC-NET conference conclusions is provided in Annex 2 for information.

The INDICATE study on digital preservation perfectly fits in the scope described by the conference conclusions. The INDICATE project is strictly related to DC-NET because it represents a practical experimentation of the policies defined by DC-NET and its enlargement to the Mediterranean region. This is particularly true for the topic of preservation.

In fact, while discussing at the DC-NET conference, examples of best practices, limitations and advantages of the e-infrastructures, with a particular attention to clouds, the INDICATE project looked more closely at what the cultural institutions are doing (through the INDICATE survey and the workshop held in Ankara).

It is worth also to mention here that two partners of INDICATE are also members of the DC-NET ERA-NET, namely: the Italian Ministry of Culture (MiBAC-ICCU) and the French Ministry of Culture (MCC). This will contribute to a better alignment and harmonisation of the approaches and understanding between the two projects, which in fact address the same topic from different perspectives.

2.3 The Workshop in Ankara

The study how e-infrastructure can enabl digital preservation is considered as a core topic in the work program of the INDICATE project and for this reason a specific workshop has been planned and held in Ankara (Turkey) on the 7 – 8 July 2011.

The scope of the workshop was to present the best practice examples of digital preservation, to discuss how e-Infrastructures can be used to solve digital preservation problems, to analyze the resources which e-Infrastructures offer, and how they can be deployed to deal with digital preservation.

For the preparation of the workshop, a survey was carried out to analyse the preservation studies of the partner countries and in other various volunteering countries.

At the preliminary stage, Turkish partners contacted the project manager of INDICATE and the Turkish Ministry of Culture and Tourism for possible dates, venues, participants, details of organization. Later, appropriate options were assessed and necessary formal procedures like simultaneous translation, audiovisual systems, catering services were completed. National Library was selected among the other candidate venues. The partners contacted with Mr. Tuncel Acar who is the head of the Turkish National Library for the venue and their presentation. Also Mr. Bulent Yilmaz from Hacettepe University and Mrs. Nilufer Onal from Atilim University were contacted and invited to give a presentation at the workshop. National and foreign experts, academicians and librarians etc. were invited to share their ideas, discuss practices and experiences.

The first day of the workshop was planned as open for all participants. After the welcome speech, the presentation of INDICATE was given by the project coordinator Mrs. Antonella Fresa and Turkish partner Mr. Bahadir Aydinonat to the Turkish community. Turkish NREN and Grid provider Tubitak explained the NREN and Grid Infrastructure in Turkey, what they did offer to institutions and their users and some technical information about the network structure. After that, Prof. Dr. Bulent Yilmaz gave detailed information on the digitisation case studies in Turkey and the relevant project Access-it. International practices on preservation were shared to participants like Magazini Digitali from Italy and Cultnat from Egypt. Lastly, Ankara City Archive project by Atilim University and Turkish National Library digitization and preservation experiences were shared.

The second day was planned as a group activity, where partners discussed deliverables and survey results. Turkish partner Bahadir Aydinonat summarized the survey results per country and gave information about preservation strategies / policies, NREN Grid Connections etc. Tubitak gave a report about a TERENA study on the connection of cultural institutions. NREN and Grid possibilities for each country were discussed. Lastly, the deliverables about the review, handbook and a new shortened preservation survey was planned for the next project meeting.

2.4 The Survey about preservation of digital cultural content

As mentioned above, Long Term preservation using e-infrastructures is one of the main topics of the work plan of the INDICATE project. According to this, a survey was prepared by the Turkish Ministry of Culture and Tourism to be analyzed at the workshop held in Ankara. This survey mainly aims to figure out the ways of a preservation process of digitized records in different countries and different institutions. It analyzes how a preservation system is built, maintained and sustained. It includes the structure of the system, preservation methods, the standards, the tools, the software used, etc. The partners of the INDICATE project have considered to make a new quick survey in Ankara because the previous one was deemed too complex. A shorter survey with 10-15 basic questions will get more feedbacks, results; will be easier to analyze. The new survey is currently under construction and it will reviewed at the next project meeting.

2.4.1 Methodology

In the Introductory part of the survey, what we aim to is to get basic information about the date, the country, name and website of the institution. The second part mainly focuses on policies and strategies and questions are related to global preservation strategies, legal deposit service, copyright and access policies. The third part of the survey ask questions about what do they preserve, how do they preserve, where do they preserve. The fourth part focuses on the international standards, administrative, technical, preservation, repository metadata records. The fifth part is mainly on back-up strategies and file integrity, and the final part is about future planning and conclusions.

2.4.2 Overview of the results of the survey

The volunteered participant countries are Italy, Spain, Greece, Turkey, Slovenia, France, Ireland, Netherlands, Lithuania and Sweden. The institutions interviewed are mostly National Libraries and National Archives. There are also specialized institutions like Council Libraries, Audiovisual archives, Data centres and museums.

After all the interviews completed and results analyzed, what we have seen is mostly the lack of global policies and strategies about preservation. In most countries, these works are done through individual efforts. Many institutions have started to digitize and preserve their items in their own ways. Because there is no effective training strategy in preservation, most of the staff finds the right way after many useless trials. This can cause overprint of the records, waste of space and time, financial sustainability issues, etc. Awareness about the preservation in professional manner is important; this points the importance of staff training. A global/national standardization about policies, training, copyrights, related laws, descriptive standards, storage, accession etc is a must[2].

When the results were analyzed, the best examples of preservation have been found in Biblioteca Nazionale Firenze, Roma, Catalunya National Library (BC) and French National Library and Ministry which all implemented a strategy for preservation (e.g., preferred file formats, conversion options, best storage options and how to secure system), handled most of the issues about copyrights, legal deposit services and financial sustainability. Their staff is well-trained, works professional, aware of the special standards, metadata, secure storage systems that they work on. Most of these institutions are connected to their countries' NREN and Grid and use its capabilities. Secure storage in different servers and different places/cities and dark archive with a restricted access can be the best ways of backup and recovery. Interoperability in national and international manners should be provided.

To conclude, the results of the survey are important and should be well-examined. The detailed review is in Annex I.

3 Concept of Preservation

3.1 What is digital preservation

Digital preservation is the active management of digital information over time to ensure its accessibility and understandability. A huge amount of precious digital information created and stored all over the world becomes inaccessible every few years at a very fast pace. Moreover, paper documentation is more and more converted in electronic format, and that characterises the transaction to the e-only. Those electronic resources need to be preserved and maintained since their availability has to be guaranteed for the future.

Preservation of digital information is widely considered to require more constant and ongoing attention than preservation of other media because of its vulnerability. The unique characteristic of digital forms makes it easy to create content and keep it up-to-date, but at the same time brings many difficulties in the preservation of this content.

One aspect is the vunerability of the media on which digital contents are stored. This issue however is not covered by the present project.

Another challenge is the issue of long-term access to data. Digital technology is developing quickly and retrieval and playback technologies can become obsolete in a matter of years. When faster, more capable and less expensive storage and processing devices are developed, older versions may be quickly replaced. When a software or decoding technology is abandoned, or a hardware device is no longer in production, records created with such technologies are at great risk of loss, simply because they are no longer accessible. This process is known as digital obsolescence.

Those reasons justify the need for the acquisition, preservation and maintenance of digital resources, in order to assure that the contained information may be always accessible and usable. This constant input of effort, time, and money to handle rapid technological and organizational advance is considered a major stumbling block for preserving digital information. Indeed, while we are still able to read our written heritage from several thousand years ago, the digital information created merely a decade ago is in serious danger of being lost.

Digital preservation is the set of processes and activities that ensure continued access to information and all kinds of records, scientific and cultural heritage existing in digital formats. This includes the preservation of materials resulting from digital reformatting, but particularly information that is born-digital and has no analog counterpart. In the language of digital imaging and electronic resources, preservation is no longer just the product of a program but an ongoing process. In this regard the way digital information is stored is important in ensuring its longevity. The long-term storage of digital information is assisted by the inclusion of preservation metadata.

Digital preservation is defined as: long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required for. “Long-term” is defined as long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely[3]. “Retrieval” means obtaining needed digital files from the long-term, error-free digital storage, without possibility of corrupting the continued error-free storage of the digital files. “Interpretation” means that the retrieved digital files, files that, for example, are of texts, charts, images or sounds, are decoded and transformed into usable representations. This is often interpreted as “rendering”, i.e. making it available for a human to access. However, in many cases it will mean able to be processed by computational means.

There are a lot of available solutions, methods and software for the registration and storage of the electronic resources. But, though, there is a more and more growing interest and attention for those

concerns, it is difficult task to guarantee the preservation because of the continue evolution of formats and the growing amount of information data to be stored.

Think of losing official records, a museum archive, irreplaceable scientific data, or even a collection of family photos, and we realize digital preservation is affecting us all.

3.2 Digital preservation VS digital curation

Digital curation is the selection[4], preservation, maintenance, collection and archiving of digital assets. It is generally referred to the process of establishing and developing long term repositories of digital assets[56] for current and future reference by researchers, scientists, historians, and scholars.

The digital curation lifecycle comprises the following steps[7]:

- **Conceptualise:** conceive and plan the creation of digital objects, including data capture methods and storage options.
- **Create:** produce digital objects and assign administrative, descriptive, structural and technical archival metadata.
- **Access and use:** ensure that designated users can easily access digital objects on a day-to-day basis. Some digital objects may be publicly available, whilst others may be password protected.
- **Appraise and select:** evaluate digital objects and select those requiring long-term curation and preservation. Adhere to documented guidance, policies and legal requirements.
- **Dispose:** rid systems of digital objects not selected for long-term curation and preservation. Documented guidance, policies and legal requirements may require the secure destruction of these objects.
- **Ingest**: transfer digital objects to an archive, trusted digital repository, data centre or similar, again adhering to documented guidance, policies and legal requirements.
- **Preservation action:** undertake actions to ensure the long-term preservation and retention of the authoritative nature of digital objects.
- **Reappraise:** return digital objects that fail validation procedures for further appraisal and reselection.
- **Store:** keep the data in a secure manner as outlined by relevant standards.
- **Access and reuse:** ensure that data are accessible to designated users for first time use and reuse. Some material may be publicly available, whilst other data may be password protected.
- **Transform:** create new digital objects from the original, for example, by migration into a different form.

3.3 Digital preservation threats

The main threats of digital preservation are here below outlined[8]:

- failure of any chain of preservation may be imagined as involving changes in, or non-maintainability of essential hardware, software or support environment;
- the human methodology established for preservation may not be followed (sudden changes of a whole team of people, etc.);
- changes in the user community's Knowledge Base may mine the intelligibility and the understandability of the data;
- one may have a loss in the chain of evidence and a lack of certainty of provenance or authenticity;
- encodings, currently considered uncrackable, used to establish lack of tampering may eventually be broken using increasingly powerful processors or sophistication of attack;

- the custodian of the data, an organisation or project, no matter how well established, may at some point in the future cease to exist;
- even if the organisation exists, the mechanisms to identify the location of data, for example a DNS entry pointing to a host machine, may no longer be resolvable.

Mandating the continued use of specific systems or formats is one possible way to try to ensure preservation. For example, we might try to mandate all images to be JPEG, all documents to be PDF/A, and all science data to be kept as XML files, or demand that a specific ontology be adopted. Even if we were to be successful for a limited time, one thing we can be sure of is that things would change and the mandates would fail.

3.4 Digital preservation strategies

In 2006, the Online Computer Library Center[9] developed a four-point strategy for the long-term preservation of digital objects that consisted of:

- Assessing the risks for loss of content posed by technology variables such as commonly used proprietary file formats and software applications.
- Evaluating the digital content objects to determine what type and degree of format conversion or other preservation actions should be applied.
- Determining the appropriate metadata needed for each object type and how it is associated with the objects.
- Providing access to the content[10].

There are several additional strategies that individuals and organizations may use to actively combat the loss of digital information.

- **Refreshing**. Refreshing is the transfer of data between two types of the same storage medium so there are no bitrate changes or alteration of data[11]. For example, transferring census data from an old preservation CD to a new one. This strategy may need to be combined with migration when the software or hardware required to read the data is no longer available or is unable to understand the format of the data. Refreshing will likely always be necessary due to the deterioration of physical media.
- **Migration**. Migration is the transferring of data to newer system environments. This may include conversion of resources from one file format to another (e.g., conversion of Microsoft Word to PDF or OpenDocument), from one operating system to another (e.g., Windows to Linux) or from one programming language to another (e.g., C to Java) so the resource remains fully accessible and functional. Resources that are migrated run the risk of losing some type of functionality since newer formats may be incapable of capturing all the functionality of the original format, or the converter itself may be unable to interpret all the nuances of the original format. The latter is often a concern with proprietary data formats.
- **Replication**. Creating duplicate copies of data on one or more systems is called replication. Data that exists as a single copy in only one location is highly vulnerable to software or hardware failure, intentional or accidental alteration, and environmental catastrophes like fire, flooding, etc. Digital data is more likely to survive if it is replicated in several locations. Replicated data may introduce difficulties in refreshing, migration, versioning, and access control since the data is located in multiple places.
- **Emulation**. Emulation is the replicating of functionality of an obsolete system[12]. Examples include emulating an Atari 2600 on a Windows system or emulating WordPerfect 1.0 on a Macintosh. Emulators may be built for applications, operating systems, or hardware platforms.

- **Metadata attachment**. Metadata is data on a digital file that includes information on creation, access rights, restrictions, preservation history, and rights management[13]. Metadata attached to digital files may be affected by file format obsolescence. ASCII is considered to be the most durable format for metadata[14] because it is widespread, backwards compatible when used with Unicode, and utilizes human-readable characters, not numeric codes. It retains information, but not the structure information it is presented in. For higher functionality, SGML or XML should be used. Both markup languages are stored in ASCII format, but contain tags that denote structure and format.

3.5 Digital preservation standards

To standardise digital preservation practice and provide a set of recommendations for preservation program implementation, the Reference Model for an Open Archival Information System (OAIS) was developed.

An OAIS is an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community.

The term OAIS also refers, by extension, to the OAIS Reference Model for an OAIS. The OAIS Reference Model is an ISO standard (ISO:14721:2003)[15][16] and defines a common framework in order to analyse and describe concepts and terminology for Digital Archives. The major aim of OAIS Reference Model is to facilitate a much wider understanding of what is required to preserve information for the long term. Long term means long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community.

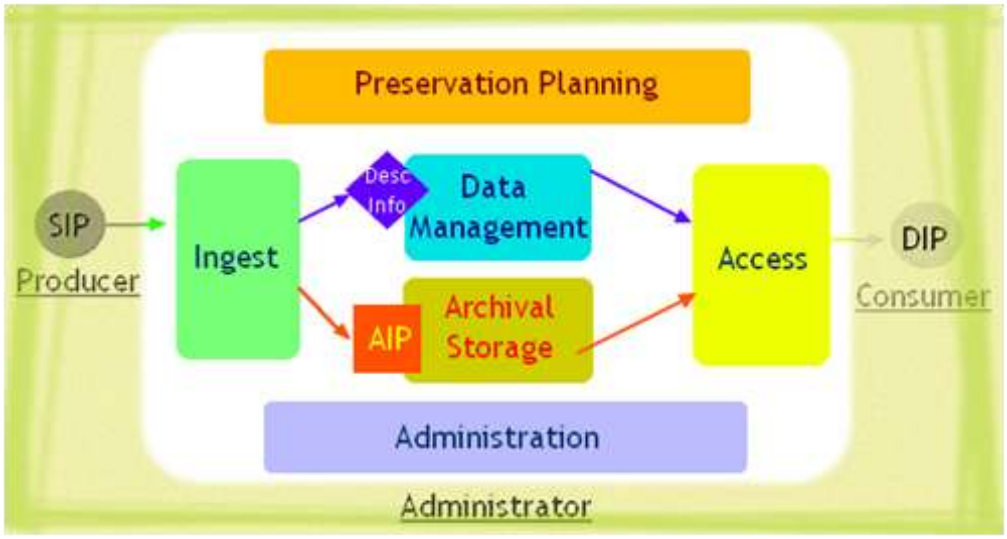
The key concepts of the OAIS Reference Model are:

1. The Content Information and its Data Object which need to be preserved
2. The Representation Information which allows to interpret the Data Object
3. The Preservation Description Information which describes the Content Information and is needed for long-term preservation
4. The Information Package which represents the element managed by the OAIS digital archive (i.e. ingested, stored, preserved and accessed).
    a. The Producer submits Submission Information Package (SIP) to the OAIS Archive through the Ingest which produces the Archival Information Package (AIP);
    b. The Archive stores and guarantees the AIP long term preservation;
    c. The Consumer accesses the Archive through the Access, by submitting queries and orders. The Access returns result sets and the preserved information as Dissemination Archival Information (DIP);
5. The Designated Community[17] and its Knowledge Base which describe the group of persons interested to the preservation of the content information and their knowledge background required for understanding the content information itself (i.e. what a community is assumed to know).

In order to deal with those key concepts and manage the information packages, OAIS defines a Functional Model by identifying 6 macro functional components:

1. **Ingest** – to accept and validate input (SIP) and prepare for storage and management (AIP);
2. **Data Management** – to populate, maintain and query wide variety of Descriptive Information;
3. **Archival Storage** – for storage and retrieval of AIP and component data objects;
4. **Access** – to support consumer in determining existence, description, location and availability of information of interest;

5. **Preservation Planning** – to monitor for changes that could impact ability to preserve and maintain archival information. It maps out preservation strategy, as well as recommends appropriate revisions to this strategy in response to evolving conditions;
6. **Administration** – manages all the Archive activities.



3.6 Services that can be constructed to preserve objects

Given the constantly changing world, a system which aims at addressing the main digital preservation issues does not have to force a specific way of doing things but instead it should be able to allow anything to be accommodated. For example, we cannot mandate a particular way of producing metadata or, using the OAIS terminology, Representation Information. While it might have some advantages in terms of interoperability in the short term, in the long term we would be locked into a dead-end. However, this should not prevent us from advising on best practice.

- One of the recognised techniques of isolating dependencies on hardware, software and environment is virtualisation. This refers to the technique of identifying abstract, important, interfaces/processes which can be implemented on top of concrete implementations.
- Changes in Knowledge Base can only be truly solved by the community itself, but procedures can be proposed which help to ensure that gaps in understandability are at least recognised and the information requested from the community is gained before it is entirely lost.
- Provenance and authenticity is, in part at least, dependent on social and information policy concerns, process documentation, and other aspects which cannot have a purely technical solution. However, some tools can be made available to ameliorate the risks of security breaches. Systems security and data integrity are only two aspects of provenance and authenticity, and we should be careful not to assume that tools for these problems will provide solutions to larger problems.
- Custodianship should always be regarded as a temporary trust and techniques are needed to allow a smooth handing over of holdings from one link in the chain of preservation to the next.
- The provision of a definitive system of persistent actionable identifiers which spreads the risk of the deterioration of identifier systems can be proposed[8].

4 State of the art of preservation

4.1 The EC Workshop on digital preservation

On the 4-5 May 2011, the European Commission organized in Luxembourg the workshop: "Future of the Past". The workshop was organized by the unit Cultural Heritage and Technology Enhanced Learning of the Information Society and Media Directorate-General. Experts in digital preservation participated to the workshop coming from all over Europe, with the participation of 61 delegated from 17 countries.

The scope of the workshop was to discuss about the future of the research on digital preservation with a particular focus on the new Horizon 2020 programme that is going to start in 2013. The starting point of the workshop has been to observe that "rate of growth in the creation of digital data is rapidly outstripping the rate of growth in data storage technologies". This gap needs actions to be taken in order to be fulfilled. The participants agreed that tools and methodologies have been developed, also thanks to the support of the European Programmes.

However, even if skills capabilities have improved in the last years, there is still much to be done in terms of training and awareness and development of "a corpus of Best Practice to refer to". Also, the dialogue between preservation specialists and cultural institutions (those "who produce and curate data") needs to be improved.

Preservation of knowledge, as an evolution of the preservation of data and the ability to demonstrate the authenticity of the preserved data are two more challenges for the future research in this field. Integration of preservation into the standard information lifecycle and sound economic models are important factor for the success. The "temporal nature of data" has been discussed, including the integration of the whole business process and the user experience of interaction with data into the staff to be preserved.

The full text of the report of the workshop is available online at:

http://cordis.europa.eu/fp7/ict/telearn-digicult/future-of-the-past_en.pdf

The results of this workshop have been reviewed by the INDICATE partners as a valuable starting point for the use case study.

4.2 EC Projects

Some of the most recent and successful EC funded projects on digital preservation are presented in this section because of their relevance for the scope of the INDICATE study.

Some of the INDICATE partners and/or experts participating to the INDICATE working groups have participate to the following projects and their experience is therefore transferred into INDICATE.

4.2.1 DPE

The Coordinator of INDICATE was involved as partner in DPE.

Digital Preservation Europe (DPE)[18] is a Coordination Action funded by the Information Society Technologies (IST) under the Sixth Framework Programme of European Commission. The project focuses on dissemination of the Digital Preservation issues to professionals as well as non professionals. Building on the earlier successful work of ERPANET, DPE aims at facilitating pooling of the complementary expertise

that exists across the academic research, cultural, public administration and industry sectors in Europe and at securing a shared knowledge base of the processes, synergy of activity, systems and techniques needed for the long-term management of digital material.

DPE fosters collaboration and synergies between many existing national and international initiatives across the European Research Area and addresses the need to improve coordination, cooperation and consistency in current activities to secure effective preservation of digital materials.

The main objectives of the project are:

- raise the profile of digital preservation;
- promote the ability of Member States acting together to add value to digital preservation activities across Europe;
- use cross-sector cooperation to avoid redundancy and duplication of effort;
- ensure auditable and certificated standards for digital preservation processes are selected and introduced;
- facilitate skills development through training packages;
- enable relevant research coordination and exchange;
- develop and promote a research agenda roadmap;
- help both citizens and specialists recognise the central role that digital preservation plays in their lives and work.

4.2.2 SHAMANN

SHAMAN – Sustaining Heritage Access through Multivalent ArchiviNg[19] - is a Large Integrated Project co-funded by the European Union within the Seventh Framework Programme.

SHAMAN aims to create a technology environment which may be used to manage the storage, access, presentation, and manipulation of potentially any digital object over time. In this sense, the solutions developed by SHAMAN aim a wide mix of target groups to minimize negative effects, but also to seize opportunities related to securing long term access and usability of digital content.

Moreover, SHAMAN aims to contribute in mitigating losses of socially valuable digital assets and minimize costs of poor digital content management, as well as to enable productivity gains in records management, generate new value-added services by re-use of preserved content and engender new services upon DP productions process components.

Under a mid-term vision, SHAMAN designs and progressive implements large-scale European-wide collections with innovative access services that support communities of practice in the creation, interpretation and use of cultural and scientific content, including multi-format and multi-source digital objects. They are combined with robust and scalable environments which include semantic-based search capabilities and essential digital preservation features.

For the longer term, SHAMAN develops radically new approaches to Digital Preservation, such as those inspired by human capacity to deal with information and knowledge, providing a sound basis and instruments for unleashing the potential of advanced ICT to automatically act on high volumes and dynamic and volatile digital content, guaranteeing its preservation, keeping track of its evolving semantics and usage context and safeguarding its integrity, authenticity and long term accessibility over time.

The main Scientific & Technical Goals are:

- To establish an Open Distributed Resource Management Infrastructure Framework enabling Grid-Based Resource Integration, reflecting, refining and extending the OAIS model and taking advantage of the latest state of the art in virtualisation and distribution technologies from the fields of GRID computing, Federated Digital Libraries, and Persistent Archives;

- To develop and integrate technologies to support Contextual And Multivalent Archival And Preservation Processes which are adapted and significantly extended from the fields of content and document Management and Information Systems;
- To develop and integrate technologies to support Semantic Constraint-Based Collection Management to target one of the key challenges in automating one class of digital preservation core functions;
- To support the Managing Of Future Requirements by securing Interoperability With Future Environments and Maintaining Essential Properties of the preserved content.
- To foster the take up of DP technologies by facilitating suites of products and services able to attack problems and needs from an integrated point of view.

SHAMAN's main output is the implementation, testing and delivery of a next generation, long term digital preservation framework including systems and tools for analysing, ingesting, managing, accessing and reusing information objects and data across libraries and archives. It comprises the definition of the SHAMAN Theory of Preservation integrating the analysis, ingestion, management, access to and reuse of information objects across distributed repositories. The data preservation capabilities offered secure the authenticity and integrity of data objects through time.

4.2.3 PLANETS

Planets, Preservation and Long-term Access through Networked Services[20], is a four-year project co-funded by the European Union under the Sixth Framework Programme to address core digital preservation challenges. The primary goal for Planets is to build practical services and tools to help ensure long-term access to our digital cultural and scientific assets.

The Planets Project delivers a sustainable framework to enable long-term preservation of digital content, increasing Europe's ability to ensure access in perpetuity to its digital information.

In particular Planets delivers:

- Preservation Planning services that empower organisations to define, evaluate, and execute preservation
- Methodologies, tools and services for the Characterisation of digital objects
- Innovative solutions for Preservation Actions tools which will transform and emulate obsolete digital assets
- An Interoperability Framework to seamlessly integrate tools and services in a distributed service network
- A Testbed to provide a consistent and coherent evidence-base for the objective evaluation of different protocols, tools, services and complete preservation plans
- A comprehensive Dissemination and Take-up program to ensure vendor adoption and effective user training.

The project enables organisations to improve decision-making about long term preservation, ensure long-term access to their valued digital content and control the costs of preservation actions through increased automation and scalable infrastructure. Intensive Dissemination and Take-up activities ensures the widest possible adoption of results in the user community and enable commercial tool and service providers to compete in a new market place for differentiated preservation services and tools.

A key objective of PLANETS is to create an environment within which commercial services to support different aspects of the preservation lifecycle can flourish. The PLANETS interoperability framework, through the publication of open, common interfaces, provides the foundations for a competitive market place to which third party vendors can bring preservation tool and service products.

The Planets project ended on 31 May 2010. Planets results will be maintained and developed by a follow-on organisation called the Open Planets Foundation (OPF)[21]. OPF is a not-for-profit company, registered in the UK. To find out more about the OPF and how to join, please visit: www.openplanetsfoundation.org.

The Open Planets Foundation (OPF) has been established to provide practical solutions and expertise in digital preservation, building on the research and development outputs of the Planets project.

4.2.4 CASPAR

CASPAR - Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval[22] - is an Integrated Project co-funded by the European Union within the Sixth Framework Programme (Priority IST-2005-2.5.10, "Access to and preservation of cultural and scientific resources").

The core objectives of the CASPAR project are to design and implement a common digital preservation framework for heterogeneous data based on the OAIS Reference Model along with the research, development and integration of advanced components, for use in a wide range of preservation activities. These components are integrated in a common software platform that enables the building of services and applications for several digital preservation purposes. Starting from the common CASPAR technology and organisational framework, the ambitious challenge for CASPAR is to derive a variety of innovative applications allowing OAIS-compliant digital preservation strategies to be undertaken in different domains, including Science, Cultural Heritage and Contemporary Arts.

In particular, CASPAR intends to:

- Implement, extend, and validate the OAIS reference model (ISO:14721:2003)
- Enhance the techniques for capturing Representation Information and other preservation related information for content objects
- Design virtualisation services supporting long term digital resource preservation, despite changes in the underlying computing (hardware and software) and storage systems, and the Designated Communities.
- Integrate digital rights management, authentication, and accreditation as standard features of CASPAR.
- Research more sophisticated access to and use of preserved digital resources including intuitive query and browsing mechanisms
- Develop case studies to validate the CASPAR approach to digital resource preservation across different user communities and assess the conditions for a successful replication.
- Actively contribute to the relevant standardisation activities in areas addressed by the project.
- Raise awareness about the critical importance of digital preservation among the relevant user-communities and facilitate the emergence of a more diverse offer of systems and services for preservation of digital resources.

4.2.5 NESTOR

NESTOR - Network of Expertise in long-term STOrage of digital Resources in Germany[23] - is a cooperation association including partners from different fields, but all connected in some way with the subject of Digital Preservation. The following cooperation partners are currently active in Nestor:

- Bayerische Staatsbibliothek (Bavarian State Library)
- Deutsche Nationalbibliothek (German National Library)
- Fernuniversität Hagen (Hagen Open University)

- Georg-August-Universität Göttingen / Niedersächsische Staats- und Universitätsbibliothek Göttingen (Georg-August University, Göttingen / Lower Saxony State and University Library, Göttingen)
- Humboldt-Universität zu Berlin (Humboldt University in Berlin)
- Landesarchiv Baden-Württemberg (Baden-Württemberg State Archive)
- Stiftung Preußischer Kulturbesitz / SMB - Institut für Museumsforschung (Prussian Cultural Heritage Foundation / SMB - Institute for Museum Research)
- Bibliotheksservice-Zentrum Baden-Württemberg (Baden-Württemberg Library Services Centre)
- Institut für Deutsche Sprache (German Language Institute) ´
- Computerspiele Museum Berlin (Computer Games Museum)
- Goportis
- PDF/A Competence Center

NESTOR was set up as a BMBF (German Federal Ministry for Education and Research) sponsored project (2003-2009) and has been continued on an independent basis by the former project partners together with other organisations since July 2009.

As a network it brings together a disparate array of institutions affected by digital preservation, experts and active project participants, all committed to the exchange of information, the distribution of tasks, the development of standards and the exploitation of synergy effects etc.

NESTOR's work is not, however, restricted to Germany-based activities. The partners maintain close links with corresponding initiatives in other countries and are actively involved in European and international initiatives and projects.

4.2.6 APARSEN

APARSEN - Alliance Permanent Access to the Records of Science in Europe Network[24] - is a Network of Excellence co-funded by the European Union within the Seventh Framework Programme that aims to bring together an extremely 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, exploiting our diversity by building a long-lived Virtual Centre of Digital Preservation Excellence.

The links of APARSEN may be summarized in the following figure.

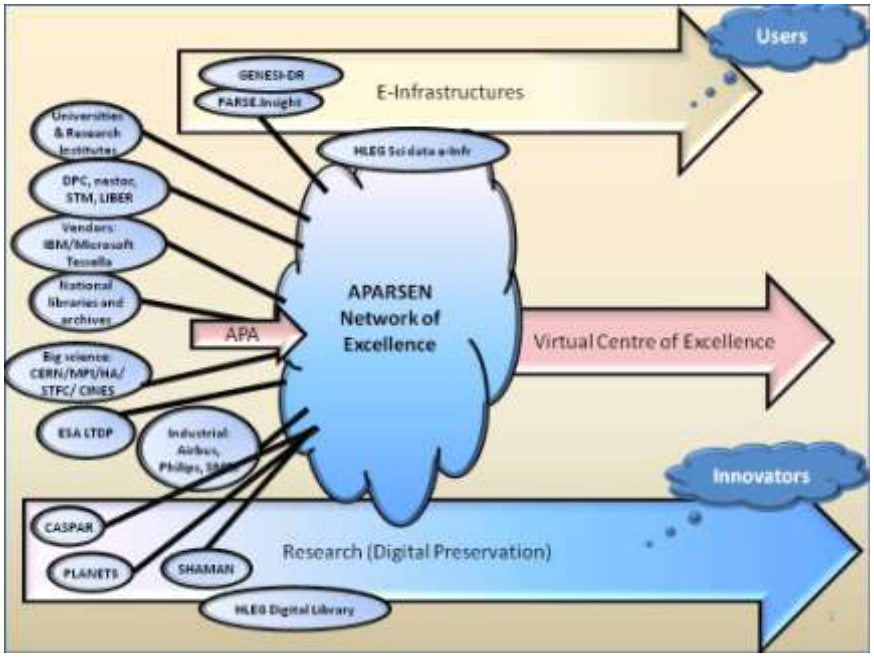


Figure 1. APARSEN Network of Excellence

The Alliance aims to develop a shared vision and framework for a sustainable organisational infrastructure for permanent access to scientific information. In particular it aims to:

- Support the development of a sustainable European Digital Information infrastructure that guarantees the permanent access to the digital records of science, whether documents or data, across all fields of research, scholarship and technology;
- Be a strategic partner for the Commission of the European Union and national governments in order to strengthen European and national strategies and policies and their implementation in the area of long-term preservation of and access to the digital records of science, thereby contributing towards Europe as an Information Society;
- Be a platform that will assist key stakeholders in the world of science and scientific information to cooperate amongst themselves and with other organisations on digital repositories for science;
- Strengthen the role of European parties in worldwide efforts to secure long-term preservation and access of the digital records of science;
- Be a global centre of excellence for digital preservation.

4.2.7 ENSURE

ENSURE - - Enabling kNowledge Sustainability Usability and Recovery for Economic value[25] - is a Collaborative Project co-funded by the European Union within the Seventh Framework Programme. Drawing on motivation from use cases in health care, finance and clinical trials, ENSURE aims to significantly extend the state-of-the-art in digital preservation which to-date has focused on relatively homogeneous cultural heritage data.

ENSURE use cases bring up a large number of issues which have yet to be fully addressed:

- safely leveraging scalable pay-as-you-go infrastructure such as clouds;
- having businesses understand the economic implications of preservation;
- conforming to regulatory, contractual and legal requirements as part of a whole workflow;

- managing long term integrity and authenticity significant intellectual property or highly personal data;
- using off-the-shelf IT technologies for preservation to support different types of digital resources.

Building on prior work, ENSURE addresses the above mentioned issues with novel approaches and tools:

- Cost and Value. Evaluate the cost and benefit of different quality solutions, enabling the selection of the most cost-effective solution.
- Preservation Lifecycle Management. Build on lifecycle management approaches to manage the preservation lifecycle, ensuring regulatory compliance, allowing changes in the preservation approach to reflect environmental changes, addressing evolution of ontologies, and managing the quality of the digital objects over time.
- Content-Aware Long-Term Data Protection. Ensure long-term data protection, addressing changes in personally identifiable information, new and evolving regulations, managing user identities over decades, etc.
- Leveraging Wider ICT. Evaluate the costs, risks and benefits, and demonstrate how to use emerging commonly available IT to enable scalable solutions for long term digital preservation, considering in particular cloud storage and virtual application image capture.

4.2.8 SCAPE

The SCAPE - SCAlable Preservation Environments[26] – project, co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137), develops scalable services for planning and execution of institutional preservation strategies on an open source platform that orchestrates semi-automated workflows for large-scale, heterogeneous collections of complex digital objects. These services enables to:

- Identify the need to act to preserve all or parts of a repository through characterisation and trend analysis;
- Define responses to those needs using formal descriptions of preservation policies and preservation plans;
- Allow a high degree of automation, virtualisation of tools, and scalable processing;
- Monitor the quality of preservation processes.

SCAPE aims to enhance the state of the art of digital preservation in three ways:

- By developing infrastructure and tools for scalable preservation actions;
- By providing a framework for automated, quality-assured preservation workflows;
- By integrating these components with a policy-based preservation planning and watch system.

These concrete project results are validated within three large-scale testbeds from diverse application areas.

4.2.9 PROTAGE

One of the DC-NET Partner (the Swedish Archives) has been the coordinator of the PROTAGE project.

PROTAGE[27] is a research project (STREP) in the seventh framework programme. The name PROTAGE is an acronym for PReservation Organizations using Tools in AGent Environments.

The mission of the PROTAGE project is to investigate and initiate complementary new approaches to digital preservation that make long-term digital preservation easy enough for users to be able to help preserving

their own content, while reducing the cost and increasing the capacity of memory institutions to preserve digital information.

The PROTAGE approach to digital preservation is based on pro-active autonomous software agents that are independent of hardware and software technologies. This represents a shift of focus in digital preservation from information systems to preservation-friendly digital objects. The idea is to link these digital objects to long-term digital preservation processes by using agent-based software technology. The PROTAGE project, based on the latest research on digital preservation strategies and on autonomous systems, builds and validates flexible and extensible software agents for long-term digital preservation and access that can cooperate with and be integrated in existing and new preservation systems.

The main objectives of the project are:

- Research the potential of software agent ecosystems to support the automation of digital preservation tasks;
- Demonstrate the technical feasibility of such a system;
- Analyse implementation in different organisational environments;
- Explore possible integration with other digital preservation environments;
- Explore synergies with other RTD activities in digital preservation.

The target end users of the PROTAGE project are digital curators and content creators, including individuals creating and managing their own digital information. The developed solutions will be flexible and extensible, so that they can be utilised by archives, libraries, museums, private and public sector organisations and individuals

The PROTAGE point of view is to consider challenges in digital preservation by using concepts of Agent Ecosystems. An agent ecosystem is a bio-inspired open-ended environment populated by autonomous agents which interact in a flexible way. Autonomous agents can be regarded as similar to biological entities that are provided with some initial energy, abilities and capabilities of learning from their experience.

Self-organization and self-adaptation are two facets in digital preservation introduced by the PROTAGE approach. Self-adaptive systems work in a top-down manner. They evaluate their own global behaviour and change it when the evaluation indicates that they are not accomplishing what they were intended to do, or when better functionality or performance is possible. Such systems typically operate with an explicit internal representation of themselves and their global goals. Self-organizing systems work in the bottom-up principle. They are composed of a large number of components that interact according to simple and local rules. The global behaviour of the system emerges from these local interactions, and it is difficult to deduce properties of the global system by studying only the local properties of its parts. Such systems do not use internal representations of global properties or goals; they are often inspired by biological or sociological phenomena.

Intended application areas of PROTAGE agent technology in digital preservation include:

- Submission and ingest of digital material - Digital resources that are to be submitted to a repository are usually checked for consistency in terms of file formats and metadata. Currently, this puts substantial workload on the personnel responsible for handling the submissions, either at the submitting side or at the receiving side. Agent based digital preservation solution will lighten this burden, since ingest can be heavily automated, both in metadata collection and in the quality assurance of the delivered resources. Another way of lightening this process is that the agent-based systems are also open for automatically shortening the period of transfer of inactive records to an archive. Agents could also be configured to exercise certain control over the objects intended for archiving, within the system in which they are created and maintained.
- Monitoring preservation - Globally, the volume of digital information to be preserved for the medium to longer term is colossal. Encapsulation of repository data within agents reduces their

complexity and increases their manageability. Monitoring the preservation system with the help of analysis agents will reduce complexity, making preservation activities easier and will support dynamic and flexible organization of personal and institutional information repositories, distributed over the web. These in turn enable sharing of information between users at the knowledge-level, and automated discovery of new relevant information through collaborative information exchange between software agents.

- Transfer between repositories - Instead of having to manually schedule for large deliveries of digital material at both the transferring and receiving repository, software agents can negotiate between themselves when the delivery should be made in order to minimize impact on both network traffic as well as on the systems and storage solutions in respective repositories.

Expected results are:

- Allow content producers to create and publish in a preservation-compatible manner,
- Provide digital repositories with means of further automating the preservation processes,
- Facilitate seamless interoperation between content providers, libraries and archives, and end-users throughout Europe.

4.2.10 DIGICURV

DigCurV - Digital Curator Vocational Education Europe Project[28] - is a project funded by the European Commission's Leonardo da Vinci programme to establish a curriculum framework for vocational training in digital curation.

DigCurV brings together a network of partners and organizations from Europe, Canada and the USA with a strong track record of international work in the field of digital libraries and digital preservation, to address the availability of vocational training for digital curators in the library, archive, museum and cultural heritage sectors needed to develop new skills that are essential for the long-term management of digital collections.

DigCurV addresses the availability of vocational training for digital curators in the library, archive, museum and cultural heritage sectors needed to develop new skills that are essential for the long-term management of digital collections.

The thirty-month project aims to identify, analyse and profile existing training opportunities and methodologies, survey training needs in the sector, while identifying the key skills and competences required of digital curators. It establishes a curriculum framework from which training programmes can be developed in future.

4.2.11 KEEP

KEEP Project[29] is co-funded by the European Union's Seventh Framework Programme for research and technological development (FP7) and focuses on the development of emulation services (KEEP Emulation Services) to enable accurate rendering of both static and dynamic digital objects: text, sound, and image files; multimedia documents, websites, databases, videogames etc.

The overall aim of the project is to facilitate universal access to the cultural heritage by developing flexible tools for accessing and storing a wide range of digital objects. KEEP also considers legal issues concerning the implementation of emulation-based systems and proposes solutions which comply with European and national copyright laws.

The big success of computing technology, where machines are rapidly superseded, has created a serious and growing challenge of how to preserve access to digital material produced on obsolete machines. Cultural heritage organisations are particularly sensitive to the threat of major data loss resulting from

technical obsolescence. KEEP develops the KEEP Emulation Services to enable the accurate rendering of these objects, designed for a wide variety of computer systems, so that they can be securely accessed in the long term.

KEEP addresses the problems of transferring digital objects stored on outdated computer media such as floppy discs onto current storage devices. This involves the specification of file formats and the production of transfer tools exploited within a framework, and taking into account possible legal and technical issues. KEEP addresses all aspects ranging from safeguarding the original bits from the carrier to offering online services to end-users via a highly portable emulation framework running on any possible device. In addition to producing a software package, the project delivers understanding about how to integrate emulation-based solutions with an operational electronic deposit system. Existing metadata models are researched and guidelines are developed for mapping digital objects to emulated manifestations. KEEP seeks ways to integrate its work with the outputs of other digital preservation projects and software. Overall, KEEP contributes to the next generation of permanent access strategies based on emulation.

Although primarily aimed at those involved in Cultural Heritage, such as memory institutions and games museums, the KEEP Emulation Services can also serve the needs of a wide range of organisations and individuals because of its universal approach.

4.2.12 National Digital Stewardship Alliance

The Infrastructure Working Group works to identify and share emerging practices around the development and maintenance of tools and systems for the curation, preservation, storage, hosting, migration, and similar activities supporting the long term preservation of digital content.

The group is currently surveying emerging practices around the use of large-scale and cloud storage platforms for digital preservation. Future projects may include:

- Promoting the collaborative development of free and open source tools, using sustainable software development processes and to create a structure for the support of the software over time
- Exploring the use of computer forensics tools for the appraisal, processing, and preservation of born-digital collections
- Encouraging communities with highly specialized needs (e.g., geospatial, datasets, observational data) to develop storage networks or access services that can serve the entire community

The National Digital Stewardship Alliance[30] was launched in July 2010 as an initiative of the National Digital Information Infrastructure and Preservation Program, which Congress established in 2000 and the Library of Congress administers.

Members of the NDSA collaborate to preserve access to our national digital heritage by:

- Broadening access to our nation's expanding digital resources;
- Developing and coordinating sustainable infrastructures for the preservation of digital content;
- Advocating standards for the stewardship of digital objects;
- Building a community of practice around the management of distributed digital collections;
- Promoting innovation;
- Facilitating cooperation between government agencies, educational institutions, non-profit organizations, and commercial entities;
- Fostering the participation of diverse communities and relationships across boundaries;
- Raising public awareness of the enduring value of digital resources and the need for active stewardship of these national resources.

4.2.13 We-preserve website initiative

WePreserve [31] is a collaborative platform which highlights the synergistic activities of DigitalPreservationEurope (DPE), Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval (CASPAR), and Preservation and Long-term Access through NETworked Services (PLANETS). This platform hosts many initiatives like conferences, forums and joint training modules in order to introduce the Principles of Digital Objects Preservation.

The three partners work together to:

- Develop and deliver a collaborative web platform shared by the projects to serve as a common entry point to digital preservation and curation projects and provide common services, a calendar of events, an information board and resources;
- Collaborate on the development of training and educational activities, events and programmes in Europe;
- Support the dissemination of publications and the mechanisms to ensure their visibility (e.g., by automatic means such as OAI-PMH); and,
- Consider on a quarterly basis the possible ways we might extend the collaborative agreement.

4.3 Work on persistent identifiers: DOI, NBN-Italy

The persistent identification of resources on the Internet is one of the fundamental problems of the cultural world, especially within 'digital library' and 'digital preservation' applications which seek to use the Internet as a platform for searching and disseminating contents.

Current initiatives like the European Digital Library (EDL)[32] and Europeana[33], clearly show the need for a certified and stable digital resource reference mechanism in the cultural and scientific domains; also projects aimed at supplying Europeana with digital content like ATHENA worked on the identification on the state of the art of persistent identifiers within the museum community in order to demonstrate, in general terms, how important identifiers are to the delivery an organisation's information services[34]. The lack of confidence in digital resource reliability hinders the use of the Digital Library for preservation, research, citation and dissemination of digital contents.

A trustworthy solution to realize the "long-term availability" of the cultural contents is to associate to any digital resource a Persistent Identifier (PI) that certifies its authenticity and ensures its accessibility.

Currently, some technological proposals are available, but the scenario shows that we can't expect/impose a unique PI technology or only one central registry for the entire world. Moreover, different user communities do not commonly agree about the granularity of what an identifier should point to.

Another component essential to implement a PI service is the reliability and credibility of the institution that stands security for the maintenance of the PI-URL association register.

4.3.1 Persistent identifier standards

The association of a Persistent Identifier (PI) to a digital resource can be used to certify its content authenticity, provenance, managing rights, and to provide an actual locator. The only guarantee of the actual persistence of identifier systems is the commitment shown by the organizations that assign, manage, and resolve the identifiers.

The following is a brief description of the most widely diffused technological platforms. Only the NBN standard will be described in details in the next section.

- The Document Object Identifier system (DOI [35]) is a business-oriented solution widely adopted by the publishing industry, which provides administrative tools and a Digital Right Management System (DRM).
- Archival Resource Key (ARK [36]) is an URL-based persistent identification standard which provides peculiar functionalities that are not featured by the other PI schemas, e.g., the capability of separating the univocal identifier assigned to a resource from the potentially multiple addresses that may act as a proxy to the final resource.
- The Handle System[37] is a technology specification for assigning, managing and resolving persistent identifiers for digital objects and other resources on the Internet. The protocols specified enable a distributed computer system to store identifiers (names, or handles) of digital resources and resolve those handles into the information necessary to locate, access, and otherwise make use of the resources. That information can be changed as needed to reflect the current state and/or location of the identified resource without changing the handle.
- Finally, the Persistent URL (PURL[38]) is simply a redirect-table of URLs and it's up to the system manager to implement policies for authenticity, rights, trustability, while the Library of Congress Control Number (LCCN [39]) is the persistent identifier system with an associated permanent URL service (the LCCN permanent service), which is similar to PURL but with a reliable policy regarding identifier trustability and stability.

In the Library domain the **National Bibliography Number** (NBN – RFC3188[40]) has been defined and is currently promoted by the Conference of European National Librarians (CENL)[41]. This standard identifier format assumes that the national libraries are responsible for the national name registers. The first implementations of NBN registers in Europe are available at the German and Swedish national libraries; NBN has been registered and adopted also by the Nordic Metadata Projects[42]

In Italy, the National Central Library of Florence, supported by the *Fondazione Rinascimento Digitale,* and in cooperation with the National Central Library of Rome and the Marciana National Library (Venice) is now actively working on a new NBN architecture. The decision to utilise the NBN is due to the fact that it is a namespace of the exclusive pertinence of national libraries (every country has registered a sub-namespace at the Library of Congress: for Italy: NBN:IT ISO 3166); this guarantees the presence of the requisites of stability and permanence necessary for an institution that intends to manage a PI service.

The project aims to set up a national register of persistent identifiers for the digital cultural objects on the Internet, and experiment a service of resolution and access to these resources. The technology is based on the Uniform Resource Name (URN[43]), utilising the National Bibliography Number (NBN) as the reference namespace, with the aim of facilitating citizens' access to resources, arousing sensitivity and encouraging cultural institutions to implement digital preservation programmes for their own digital resources.

Unlike URLs, URNs are not directly actionable (browsers generally do not know what to do with a URN), because they have no associated global infrastructure that enables resolution (such as the DNS supporting URL). The NBN solution is conceived especially for those resources that do not possess any type of identification (i.e. doctoral theses, digitalisations of antique books, etc), but in perspective, it can also be extended to unifying all digital cultural resources under a single code, even those already identified by codes like ISBN, ISSN, SICI. The resources identified will thus be able to reside on the systems of the cultural institutions that have the rights to manage their sub-domain, and/or in legal deposit systems.

4.3.2 NBN Technical Architetcture

- At the highest level there is a **Root node or Central node** (the National Central Library of Florence, in the prototype currently being developed), which is responsible for the top-level domain (NBN:IT in our case). The Central Node maintains the central register, where all NBN names generated by any NBN leaf nodes are stored. The central node can assign sub-domains to institutions that accomplish a registration procedure; it can resolve a user-query directly or redirect it to the appropriate lower level agency; it checks the NBN records harvested from sub-domain registers for policy compliance and uniqueness. The Central Node supports the management of a Central Agency, which is responsible for the quality and reliability of the answers provided by the whole infrastructure to the user requests.
- The second-level domains (e.g.: NBN:IT:UR, NBN:IT:FRD, etc.), authorized by the Central node, are defined as **Inner nodes (**second level naming authorities). They manage specific sub-domains which control other lower level domains. The Inner nodes can define their own policies for generating NBN names or registering institutions in relation to the specific domains (e.g.. Cultural Heritage, Scientific, Broadcasting, etc.). The inner node harvests the NBN records from the leaf and/or inner nodes under its responsibility and performs checks that are similar to those described for the central node but for a smaller set of resources. The inner node cannot generate NBN names but can resolve them directly or by redirecting requests to the appropriate nodes.
- At the bottom of this hierarchy there are the **Leaf nodes** (e.g. NBN:IT:UR:CNR), that are responsible for the bottom-level sub-domains, assigned to the agencies that manage the actual digital libraries. The Leaf nodes are the only ones that harvest publication metadata from the actual repositories and assign unique identifiers to digital objects, according to the policies defined from the Inner nodes. Also leaf nodes can resolve all the NBN names.

It is easy to see that this hierarchical multi-level distributed approach implies that the responsibility of PI generation and resolution can be recursively delegated to lower level sub-naming authorities, each managing a portion of the domain name space. Given the similarity of the addressed problems, some ideas have been borrowed from the DNS service.

Within this architecture each node harvests PI information from its child nodes and is able to directly resolve all identifiers belonging to its domain and sub-domains. Besides, it can query other nodes to resolve NBN identifiers not belonging to its domain. This implies that every node can resolve every NBN item generated within the NBN:IT sub-namespace, either by looking up its own tables or by querying other nodes. In the latter case the query result is cached locally in order to speed up subsequent interrogations regarding the same identifier.

This redundancy of service access points and information storage locations increases the reliability of the whole infrastructure by eliminating single points of failure. Besides, reliability increases as the number of joining institutions grows up. A distributed architecture also increases scalability and performance, while maintaining unaltered the publishing workflows defined for the different repositories.

The NBN Persistent Identifiers system will be implemented by the *Magazzini Digitali* project (see next section).

5 Examples of preservation

5.1 Magazzini Digitali - the Italian Digital Preservation Service[44]

Magazzini Digitali aims to set up a long term digital preservation system for electronic documents published in Italy, distributed via a digital communication network and deposited according to the provisions of the legal deposit law (L. 106/2004, DPR 252/2006). The project is funded by the Italian Ministry of Cultural Heritage and supported by the *Fondazione Rinascimento Digitale*, a private institution that supports the use of digital technologies in the cultural domain. The two National Central Libraries of Florence and Rome and the National Marciana Library of Venice are the project partners.

For the purposes of the project, Digital Preservation is defined as a public service to be provided by trusted digital repositories in order to ensure - for deposited digital resources – viability (permanence over the long time of a bit sequence), renderability (property of the bit sequence to be readed from a device in order to be displayed to a user), authenticity (identity and integrity) and availability for designated communities[45].

5.1.1 Technical Architecture

The technical infrastructure is based on:

- data replication (different machines located in different sites);
- simple and widespread hardware components, non vendor-dependent, that can easily be replaced (e.g. personal computers with four 2000 GB hard disks, using widespread and inexpensive Serial Advanced Technology Attachment (SATA)[46] technology);
- open source operating system and utilities (widespread acceptance means less dependencies);
- data replication based on open source disk synchronization utility (rsync[47] for UNIX); Redundant Array of Independent Disks (RAID)[48] is not used, in order to avoid hardware dependencies (e.g. disk controllers);
- each site is composed by a set of autonomous and independent nodes, each node on a given site has a mirror node on the other site, so each site will contain – in a symmetrical way - both master nodes and mirror nodes (*see Figure 2*);
- Each physical file is replicated twice on different computers within the same node.

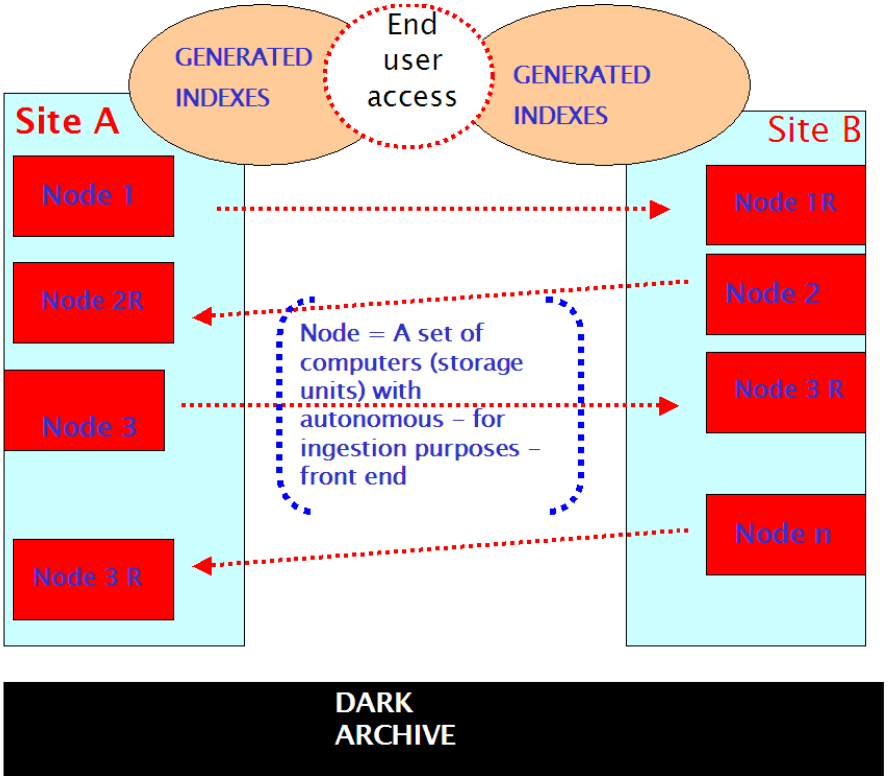


Figure 2. Magazzini digitali technical architecture overview

The two main deposit sites, managed by the National Central Library of Florence and the National Central Library of Rome, are intended for final user access and long term preservation. The third site, managed by the National Marciana Library, is a dark archive for long term preservation and disaster recovery services only. Its architecture is the same as the other two sites, but it doesn't supply access services to the public.

As regards the 3 repositories, from a physical point of view, and in order to avoid any threats to the security of the overall service, they are all located on external ISO 27001[49] certified data centers (or collocation centers[50]), owned and managed by 3 different companies (to reduce the commercial risk of "domino" effects). The 3 collocation centers have to be at least 200 km far away from each other (to reduce the risk of natural threats). The compliance with the ISO 27001 international security standard is a requirement for a domain specific certification of *Magazzini Digitali* as a trusted digital repository (e.g. according to DRAMBORA[51] or TRAC[52] provisions).

The repositories are also ISO 14721-2003 OAIS[53] compliant, as regards their logical and functional model.

5.1.2 Metadata

The *Magazzini Digitali* system can ingest two kinds of files:

- data wrapped in WARC containers: WARC (ISO 28500) container aggregates digital objects for ease of storage in a conventional file system[54].
- metadata wrapped in MPEG21-DIDL containers[55]: MPEG21-DIDL (ISO 21000) is a simple and agnostic container suitable for the representation of digital resources (sets of metadata compliant to different *Schemas*)

Nevertheless managing different metadata *schemas*[56] will be a problem:

A *long term archive*, particularly a *legal deposit long term archive*, cannot rely on stores of metadata based on few *schemas* and fed by a few principal sources; a *long term archive* has to face stores of metadata based on *schemas* that can origin from – using the PREMIS language - different *agents* (ex. g. OAI-PMH metadata harvesters, metadata extractors like JHOVE, Librarians, etc). Each *schema* can change in the course of time and semantic overlap among elements belonging to different schemas (ex. g. PREMIS, MIX) will be probably the norm rather than the exception.

Magazzini Digitali is taking into account this problem and the fact that no tools seem to be available. Perhaps the most promising research activities are about "linked data"[57] .The term Linked Data (Tim Berners Lee, 2006) refers to a set of best practices for publishing and connecting structured data on the Web. Key technologies that support Linked Data are URIs (a generic means to identify entities or concepts in the world), HTTP (a simple yet universal mechanism for retrieving resources, or descriptions of resources), and RDF (a generic graph-based data model with which to structure and link data that describes things in the world)

5.1.3 Legal Framework and service model

Magazzini Digitali realizes what the most recent Italian law on legal deposit provides for: a trial period for legal deposit on a voluntary basis of electronic documents. This legislation can be regarded to as a strong commitment for national libraries (The "commitment" is one of the requirements of a trusted digital repository[58]) to set up the foundations of a Digital Preservation Network:

In order to achieve the following three main goals:

- To implement an *organizational model* suitable for creating the national and regional archives of electronic publishing production, and for being extended also to electronic resources of other domains, different from the libraries ones;
- To implement a *service model* suitable for balancing the right-holders interests in contents protection with the final users ones in contents access;
- To implement a *system* suitable for ensuring long term preservation and access to digital contents, as well as their authenticity (identity and integrity), a legal and agreements framework is strongly needed.

The *Magazzini Digitali* working group, under the coordination of the Italian Ministry of Culture and with the participation of all the involved stakeholders has produced:

An agreement between the three National libraries and the FRD, already signed, in order to set specific roles and responsibilities of each institution and to define an organizational and financial sustainability plan

An agreement schema between the three National Libraries and each electronic publisher (or electronic content provider) will join the project, already signed with the main publishers associations, in order to fulfil law provisions. The main points of this agreement are:

- Periodical harvesting of the agreed publisher's electronic documents by the National Libraries of Florence and Rome;
- Documents storing in ISO 27001 certified external data centers, and in ISO 14721-2003 OAIS compliant trusted repositories;
- Libraries (Florence, Rome, Venice) commitment to ensure long term preservation and access of the deposited documents, and to track any changes in the same documents;
- Authorization for Libraries to perform any necessary actions (refreshing, duplication, migration etc) in order to achieve long term preservation and access of the deposited documents;

- Access allowed to registered users, on the libraries LANs, with a tracking system;
- Files printing and/or downloading subjected to a specific license;
- Access allowed also in the regional deposit libraries LANs, with some restrictions.

A license model for legal deposit documents printing and/or downloading, with different service levels, already signed with the main publishers associations, with the following main points:

- three different paper printing possibilities: 15%, 50%, or the whole document;
- three different possibilities to send printed documents or files: only to Italian legal deposit libraries; only to Italian libraries; to all libraries everywhere;
- two different ways of sending documents to libraries: postal service or fax for printed documents, professional services (e.g. ARIEL, NILDE, …) for files, with file destruction after the download;
- two different downloading possibilities: only for registered users inside the legal deposit library premise; for registered users inside and outside the legal deposit library premise, in the last case via professional services (NILDE, ARIEL, …), with file destruction after the download;
- choices are up to publishers;
- no commercial utilization allowed;
- no economic compensation for publishers for two years, to be defined after.

To the purpose of extending the test-basis, the project is taking into account also Ph. D. digital thesis, resulting from specific agreements with universities, and digital resources resulting from digitisation projects funded by the Italian Digital Library initiative.

5.2 Long term preservation at the Turkish National Library

5.2.1 Library and the Collection

Beyond an ordinary library; Turkish National Library contributes the Turkish culture, science, literature and the arts as a national archive, museum and research centre.

The national library of Turkey has a collection of 2.931.095 works that include books, printed works, rare manuscripts of art, printed work with the Arabic alphabet, daily newspapers, magazines, bulletins, annuals, posters, maps, musical notes, audio-records (CD and tape), pictures, etc[59].

5.2.2 Digitization and Preservation

The digitization and the preservation works of the Turkish National Library has started in 2004 which aim to preserve maintain and hand on the Turkish Cultural Heritage in a healthy manner for domestic and foreign researchers. Turkish National Library mainly digitizes and preserves manuscripts, old / rare books, serials and gramophone records, paintings, classic movie posters.

National Library is one of the institutions that start to work in this area in Turkey. 27.309 manuscripts digitalized, preserved. These manuscripts are being served in 5 different languages (Turkish, English, Arabic, Persian and Russian) online through the http://www.yazmalar.gov.tr. 4028 volumes of Ancient Ottoman Turkish Periodicals have been preserved since 2008. This Periodical Information System is online through http://sureli.mk.gov.tr. 5500 old gramophone records are now online to researchers through http://mkutup.gov.tr/sesbankasi. 513 paintings from various Turkish Artists are digitized and preserved. Some of them are online through the web. 36.807 classic Turkish Movie posters are also digitized and

preserved. In accordance with the Encounters project, paintings, figures, movie posters will be all digitized and preserved. All the preserved materials in the library are approximately 20 TB.

5.2.3 NREN and GRID Connections

Tubitak Ulakbim provides the network infrastructure for Turkish Institutions and researchers. Turkish National Library is a member of the Turkish NREN (Ulaknet) and it is connected with a 8 Mb/s bandwidth.

5.2.4 Policies and Strategies and Technical Information

Preservation works and processes in Turkey have done mostly individually. There is a lack of collaboration with the legal and the public institutions about digitization and preservation. Most of the institutions including Turkish National Library do not check for overprints. A global policy has to be developed about preservation.

Turkish National Library mostly preserves manuscripts and rare Books, periodicals, gramophone records, paintings, figures and classic Turkish movie posters. Main formats of these preserved items are tiff, jpeg, mp3, pdf. Master copies of the digitized materials are kept for archival purposes. End-user copies are the same as the master copies. Those preserved items are stored in CD-DVD, Internet, RAID Array, Network Attached Storage, institutional repositories. The library takes daily incremental backups, for the integrity of the materials; there is a MDA and SHA1 algorithm check. National Library uses Red Hat Linux operating system which is working on Mysql database. Classic standards are used like Marc, Dublin core, AACR, Turkish Electronic Record Management System for administrative, technical, preservation and international metadata and standards. Turkish National Library has a membership system. Users have to be a member to use the library materials. There is a fee based usage system for copyrighted materials[60].

5.2.5 Conclusion

To sum up, Turkish National Library is one of the leader institutions that digitizes and preserves Turkish Cultural Heritage. Libraries' collection includes heritage items like manuscripts, old rare books, figures, paintings and gramophone records which were digitized and preserved since 2004. Librarians are mostly in charge of digitization and preservation but staff still needs special training.

Lastly, a global policy and strategy about preservation has to be developed nationally. In order to integrate Turkish Cultural Heritage to the world successfully, Common standards, strategies must be implemented and NREN, Grid connection possibilities must be evaluated. The next steps of the Turkish National Library should be undertaken in this direction.

6 Relationship with e-Infrastructures

6.1 Current status of the European e-Infrastructures

In this section we mention the European e-infrastructures network **GÉANT**[61] and the on-going FP7 projects GN3 which is responsible of its maintenance and evolution. We mention also the European Grid Initiative **EGI** and the partnership for a high-performance computing infrastructure **PRACE**.

6.1.1 GÉANT

Known simply as GÉANT, GN3 is the third generation of the successful GÉANT network and project that lies at the heart of the EU's e-Infrastructure strategy. Co-funded by the European National Research and Education Networks (NRENs) and the European Commission, the project operates the pan-European GÉANT network and a portfolio of advanced services for the research and education community.

The GN3 project started in April 2009 and it has duration of 48 months. Following extensive consultation between partners, the project established so far the general plan for the long-term development of the GÉANT backbone network. Along with the completion of the transport equipment RFI, the project is well placed to proceed with the long-term equipment and connectivity procurement plans, which will run in parallel, with the former starting first. The project started the extensive procurement of transmission and switching capability, capable of supporting 100Gbps transmission speeds.

GÉANT upgraded the western ring capacity by implementing 40Gbps wavelengths on the Frankfurt–Amsterdam route in addition to the existing Frankfurt–Geneva route, and 20Gbps wavelengths on the Amsterdam–London–Paris–Geneva, Geneva–Madrid and Geneva–Milan routes. GÉANT upgraded the eastern ring capacity to 20Gbps in the Frankfurt–Prague–Vienna–Milan route. Access capacity for DFN[62], SWITCH[63] and GARR[64] was increased to 20Gbps.

Significant improvements to off-fibre connected NRENs included increased connectivity to Israel with a second STM-16 (2.5Gbps) link to Frankfurt, and an upgrade of the University of Malta access circuit (partially provided by GARR) to 1Gbps. ULAKBIM[65] access was upgraded from 2 x 2.5Gbps to 2 x 10Gbps and the SANET[66] access circuit was upgraded to 10Gbps.

A total of 9 new point-to-point circuits (6 GÉANT Plus and 3 GÉANT Lambda), were delivered on the GÉANT infrastructure.

Following the launch of ACE (America Connects Europe), an NSF (National Science Foundation) project to fund research connectivity between US researcher and those connected to GÉANT, GN3 project has been leading much of the work with US partners to provide a common US-Europe connectivity framework that will support multi-domain connectivity services in a coherent managed way.

GÉANT, during GN2 and GN3 phases, has developed and extended the portfolio of its services and of the connected NRENs. Issues of service availability and service adoption were recognized,, as well as the need to understand user requirements and to develop a viable model for service development. Seven services from the GÉANT portfolio (out of a total of 13) were in production and three in strategy/design phase by end of the second year. Two services were launched:

- eduGAIN – April 2011
- eduPKI – March 2011

Both of these services have already acquired users.

With new service development, there has also been a strengthening of the project interaction with organized scientific and humanities disciplines. The processes of identifying, developing and delivering the connectivity and functionality are being integrated into the service development and delivery cycle. Relationships are being enhanced with user groups across disciplines (biotechnology, arts & culture, health & medicine) and complementary e-Infrastructures and programmes (Grids, Future Internet, ESFRI[67]).

As a result of the more formalized interaction with large pan-European user groups, a more consistent and professional approach is now being implemented based on a liaison process between GÉANT, the NRENs and the Large User Groups. This process was prototyped during interaction with the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT[68]) and COLUMBUS[69] projects.

The GÉANT IPR Policy, consisting of a broad set of guidelines, supplemented by standard licences and a set of Frequently Asked Questions, was adopted by the Consortium.

Figure 3. GÉANT Global Connectivity 2011

The GÉANT project is focused on understanding its users' requirements and offering services that meet their specific needs. In collaboration with the NRENs, GÉANT is developing user-focused, multi-domain services aimed at delivering seamless network performance across borders and domains and to roll these out at national level to institutions, projects and researchers through the GÉANT Service Area.

The range of services currently offered, and those in development for introduction during the term of the project, are grouped into three types:

- Connectivity Services

  Access to the GÉANT network provides the standard, high bandwidth IP connectivity (GÉANT IP). In addition, GÉANT offers the benefit of a "virtual" private network created by reserving capacity on the network backbone (GÉANT Plus or GÉANT Lambda). These specialised point-to-point connections provide dedicated bandwidth without the cost and difficulty of building and managing a private network.

- Network Performance & Operations Services

Delivering consistently high performance is key to a successful user experience. Users can benefit from the range of GÉANT network monitoring, security and support services exploited by NRENs to assure optimum performance for projects and institutions. Connectivity is supported by a comprehensive range of network monitoring and management services. These optimise network performance by providing 24x7 monitoring across the GÉANT Service Area infrastructure, enabling fast identification and remedy of any faults on the network as well as providing powerful security to prevent and detect malicious attacks. The areas of tools and services in this group include: ***Performance measuring and monitoring, Performance enhancement, Network provisioning, Security.***

- End User Applications

The development of end user application services are an important focus of the project. Mobility and roaming will allow researchers, academics and others to move freely across network boundaries, enabling safe and secure research.

- **eduroam**[70] is a global service that provides secure roaming connectivity to users at hundreds of participating institutions across Europe, including universities, libraries and research institutes.
- the **eduGAIN** [71] service, being developed in the GÉANT project, aims to establish a confederation of identity providers, enabling member organisations associated with different federations to seamlessly and securely exchange information as though they were part of the same national identity provider
- the **eduPKI**[72] service being developed within the GÉANT project aims to ease the adoption of digital certificates within the project in a cost-effective way. It aims to create a service able to support other of the project's services in defining their security requirements, and to provide them with digital certificates.
- **eduCONF**[73] - pan-European Videoconferencing service, currently under development. At the moment most NRENs offer the multivideoconferencing service, and eduCONF aims at harmonizing and optimizing these services.
- **GIdP**[74] is a simple eduGAIN-compatible Identity Provider system, where user identities and attributes are registered. GÉANT services across the pan-European backbone and the national networks can then redirect users to GIdP for authentication and authorisation.

6.1.2 EGI

The associated **European Grid Infrastructure (EGI)**[75] includes in excess of 300 sites across 50 countries, offering around 240,000 processor corse, and more than 100 petabytes of tape and disk storage. The infrastructure is available to users around the clock achieving a sustained workload of half a million computer tasks, or jobs, every day. EGI-InSPIRE (European Grid Initiative – Integrated Sustainable Pan-European Infrastructure for Research in Europe) in a collaborative effort involving more than 50 institutions in over 40 countries. EGI-InSPIRE is coordinating the transition from the previous project-finded system to a sustainable pan-European e-Infrastructure, by supporting grids of high-performance computing (HPC) and high throughput computing (HTC) resources. EGI users are represented by virtual research communities, covering bioinformatics, chemistry, high energy physics, fusion physics, health science and medicine, life science, astrophysics, earth science, earth observation, humanities and computational science.

EGI-InSPIRE aims at supporting and developing the European DCI community in three important ways:

- Integrate resource providers more closely, both politically and technologically, within the National Grid Initiatives (NGIs) and European International Research Organisations (EIROs), represented by the project consortium and beyond.
- Support the development of policies by the community of European resource providers, in collaboration with other bodies in Europe and worldwide, to ensure effective technical management, integration and operation of the EGI.
- Coordinate the development and support of Virtual Research Communities currently using the production infrastructure within the European Research Area and encourage other structured user communities onto the infrastructure.

EGI-InSPIRE will also support the work of EGI.eu within the European E-Infrastructures Forum, a collaboration that includes DEISA, PRACE, GÉANT and TERENA. EGI.eu will work with other DCI technology projects such as EMI, IGE, StartusLab, VenusC and infrastructures such as EDGI for desktop grids, to bring the technologies underpinning the computational and data infrastructures needed by our user communities into production. Strong links will also be established with the SIENA project to ensure that the needs of the production infrastructure are incorporated into the standards roadmap being developed within the wider community.

6.1.3 PRACE

The mission of the **PRACE**[76] (Partnership for Advanced Computing in Europe) Research Infrastructure is to enable high impact European scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. The PRACE RI seeks to realize this mission through world class computing and data management resources and services open to all European public research through a peer review process. Thanks to the broad participation of European governments through representative organizations, a diversity of resources can be provided by the PRACE RI including expertise throughout Europe in effective use of the resources. To aid users and potential users as well as preparing the next generation scientists and engineers PRACE has an extensive pan-European education and training effort. PRACE also encourages collaboration with industry and industrial use and conducts annual Industrial Seminars at locations throughout Europe. PRACE also seeks to strengthen the European HPC industry through various initiatives and has a strong interest in improving energy efficiency of computing systems and reducing their environmental impact. The PRACE RI is established as an international non-profit association with seat in Brussels and is named 'Partnership for Advanced Computing in Europe AISBL'. It has 21 member countries whose representative organizations are creating a pan-European supercomputing infrastructure, providing access to computing and data management resources and services for large-scale scientific and engineering applications at the highest performance level.

6.2 The European and the "global" research areas

At the onset of the 21st Century, the way scientific research is carried out in many parts of the world is rapidly evolving to what is nowadays referred to as e-Science, i.e. a "scientific method" which foresees the adoption of cutting-edge digital platforms known as e-Infrastructures throughout the process from the idea to the production of the scientific result. The e-Science vision is depicted in the figure below.

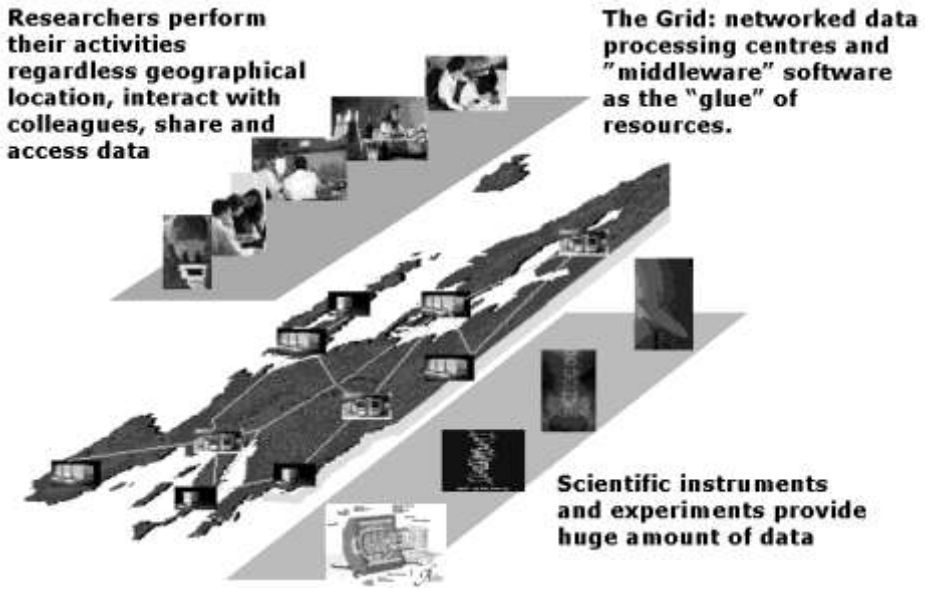


Figure 4. The "vision" of e-Science

Scientific instruments are becoming increasingly complex and produce huge amounts of data which are in the order of a large fraction of the whole quantity of "information" produced by all human beings by all means. These data are often relative to inter/multi-disciplinary analyses and have to be analyzed by ever-increasing communities of scientists and researchers, called Virtual Organisations (VOs), whose members are distributed all over the world and belong to different geographical, administrative, scientific, and cultural domains. The emerging computing model which is being developed since a decade or so is what is called the "Grid", i.e. a large number of computing and storage devices, linked among them by high-bandwidth networks, on which a special software called middleware (intermediate between the hardware and the operating system and the codes of the applications) is installed, allowing the resources to behave as a single huge "distributed" computer which "dissolves" in the fabric of the Internet and can be accessed ubiquitously through virtual services and high-level user interfaces.

The Grid and the underlying network constitute the e-Infrastructure (see Figure 5).

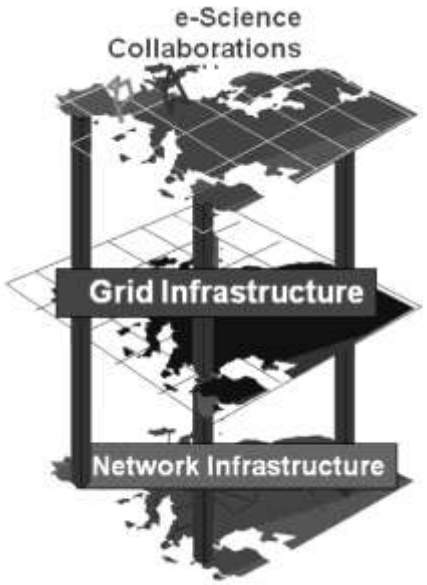


Figure 5. e-Infrastructure model of the European Research Area

The European Commission (EC) is heavily investing through its Framework Programmes in e-Infrastructures and this platform is by now considered as one of the key enablers of the European Research Area (ERA). In fact, at the top of the three-layers model of an e-Infrastructure there is the most important "network": the human collaboration among scientific communities of researchers that work together on unprecedented complex multi-disciplinary problems whose solutions are highly beneficial for the society and the progress at large.

The European Research Education Network, which connects about 3900 Institutions in more than 40 countries in the continent, and support the work of more than 30 millions of students, teachers, and researchers, is realized in the context of the GÉANT, GÉANT2[77], and GN3[78] projects, coordinated by DANTE[79].

The pan-European Grid has been realized by flagship projects like the EGEE[80] series, for High Throughput Computing (HTC) applications, and DEISA[81], for High Performance Computing (HPC) ones. Nowadays, long term sustainability of Grid and HPC infrastructures is being ensured by the EGI[75] and PRACE[82] initiatives.

In order to bridge the digital divide between Europe and other less developed regions of the world, over the past 6 years the European network and the European Grid have been expanded well outside the borders of the "old continent" in the context of several successful EC co-funded projects that have been complemented by other national/regional initiatives. The current "landscape" consists of: ALICE[83] and ALICE2[84] (network projects for Latin America), EUMEDCONNECT[85] and EUMEDCONNECT2[86] (network projects for the Mediterranean and the Middle-Eastern region), GÉANT2-ERNET[87] (network collaboration for India), ORIENT[88] (network project for China), SANREN[89] (the South African National Research and Education Network), SEEREN, SEEREN2[90], SEE-FIRE[91], and SEE-LIGHT (network projects for the South-Eastern European region), TEIN2[92] and TEIN3[93] (network projects for the Asia-Pacific region), the Ubuntunet Alliance[94] (an international initiative aiming to create a Regional Research and Education Network in Sub-Saharan Africa), EELA[95] and EELA-2[96] (Grid projects for Latin America), EUAsiaGrid[97] (Grid project for the Asia-Pacific region), EUChinaGRID[98] (Grid project for China), EU-IndiaGrid and EU-IndiaGrid2[99] (Grid projects for India), EUMEDGRID[100] and EUMEDGRID-Support[101] (Grid projects for the Mediterranean and the Middle-Eastern region), SAGRID[102] (the South African National Grid Initiative), and SEE-GRID[103], SEE-GRID2[104], and SEE-GRID-SCI[105] (Grid projects for the South-Eastern European region).

All together, the above mentioned projects/initiatives have created the "global" network and the "global" Grid depicted in Figures 6 and 7.

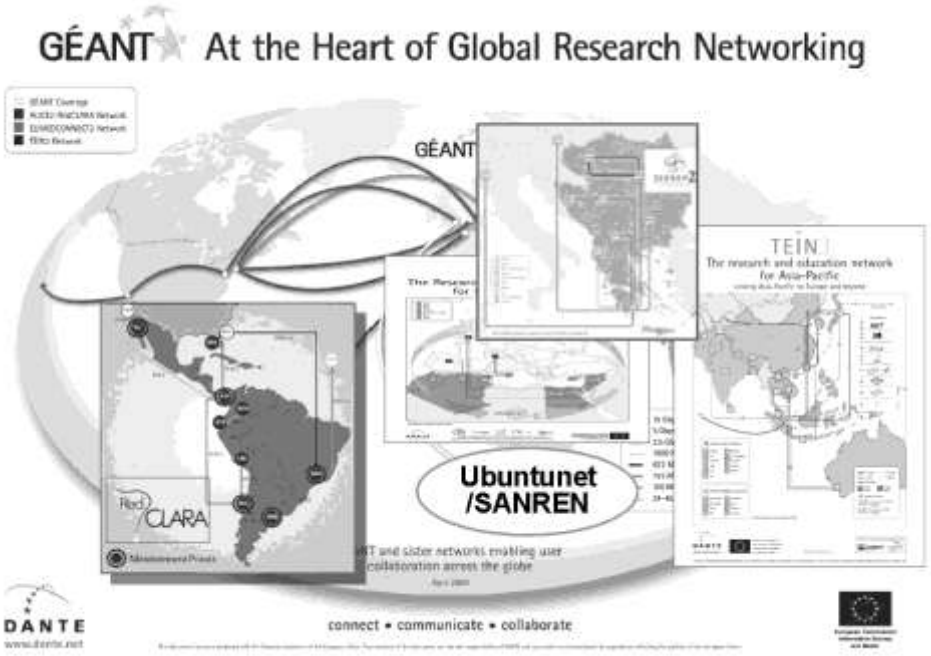


Figure 6. The “global” network

Figure 7. The “global” Grid

All projects share the same work plan whose “virtuous cycle” is depicted, in a graphical way, in Figure 8.

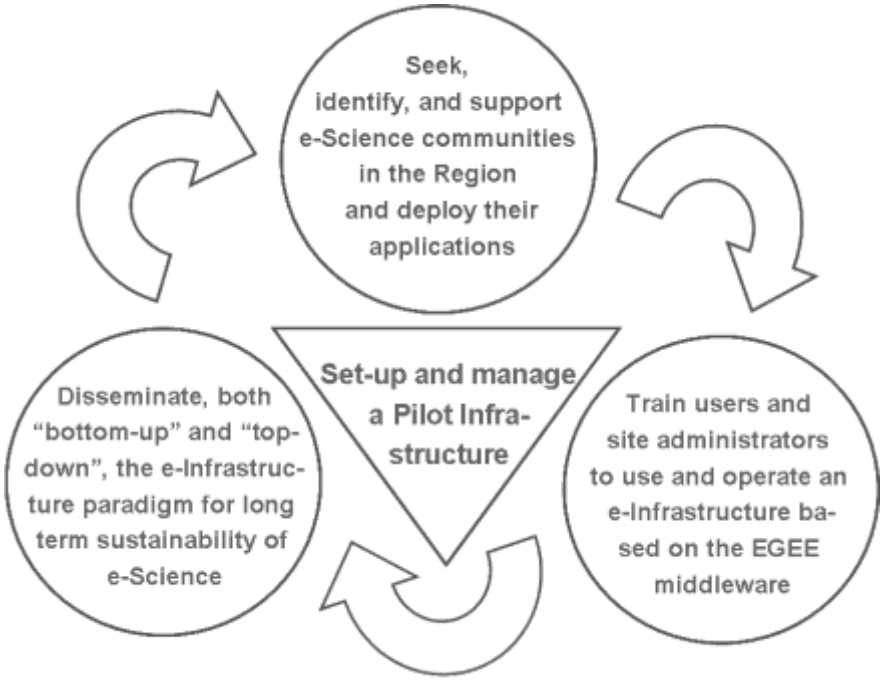


Figure 8. The "virtuous cycle" of regional Grid projects

In the following sub-sections we will describe all the "regional" e-Infrastructures cited above and, for each of them, we will show the network layer, the Grid layer and the applications deployed and running underlining how the three "actors" enable the virtuous cycle depicted in Figure 8.

6.2.1 E-Infrastructures in the Asia-Pacific Region

Network

Network provision in the Asia-Pacific region has long been an issue for e-Infrastructure development as it has been characterised by extreme heterogeneity, lack of connectivity and single points of failure in the links to other regions such as the US and Europe. Through the efforts of the Asia-Pacific Advanced Network[106] organisation and the TEIN3 project, these issues are presently being addressed and the countries in the region are now benefitting from an increasingly mature high performance network for research and education, similar to those in other regions. The third generation of the Trans-Eurasia Information Network (TEIN3) provides today a dedicated high-capacity Internet network for research and education communities across Asia-Pacific. It currently connects 11 countries in the region (Australia, China, Indonesia, Japan, Korea, Laos, Malaysia, Philippines, Singapore, Thailand and Vietnam) and provides direct connectivity to Europe's GÉANT2 network (see Figure 9).

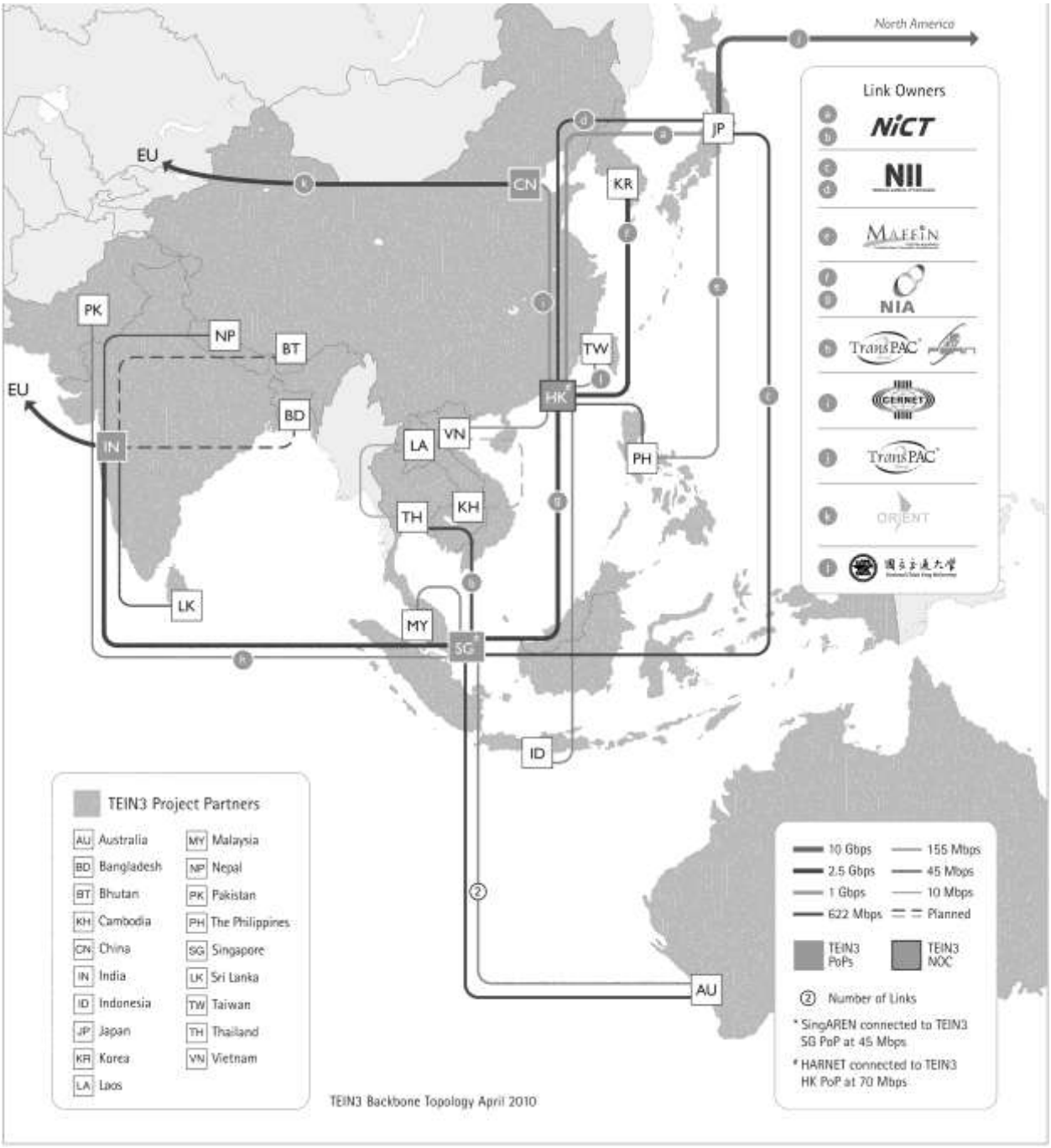


Figure 9. Topology map of the TEIN3 network

International cooperation coordinated through APAN is beginning to address network weaknesses such as bottlenecks and network disruptions such as those caused by the typhoon Morak. In that case, ASGCnet (the Taiwanese Research and Education Network) provided TEIN and APAN with a backup route to Europe during the network outage affecting them.

The recent incorporation of APAN as a legal entity based in Hong Kong will further help to strengthen the organisational arrangements and structures at the policy level to improve network provision in the region. The APAN organisation is based on the principle of national representation, with only one Primary Member from each country. Currently, there are 15 Primary Members forming a Council, the highest governing body of APAN. The governance structure of APAN follows the principles of subsidiarity as it allows individual

members to operate largely under their own rules according to the policies and funding principles of their countries but benefitting from the international collaboration and coordination provided through APAN. This arrangement aims to overcome the significant differences in the socio-economic and political characteristics of countries in the region. For the sake of completeness, Figure 10 shows the Asia-Pacific backbone topology by funding source.

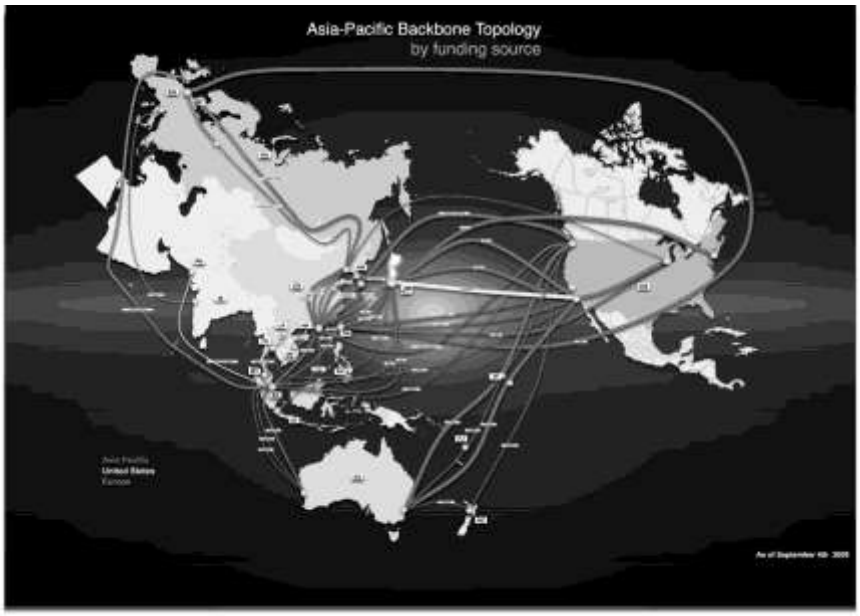


Figure 10. Asia-Pacific backbone topology by funding source

Grid

The development of grid infrastructures in the Asia-Pacific region has been driven by the participation in the CERN Large Hadron Collider[107] experiments as well as a number of applications of specific interest such as biomedical research, engineering applications and disaster mitigation. The EUAsiaGrid project, co-funded by the European Commission in the context of its Seventh Framework Program, has played a crucial role in building the capacity currently available in a number of countries such as Malaysia, Indonesia, the Philippines, Thailand and Vietnam. This infrastructure ensures that local support and key services such as user interface nodes, compute and storage elements can be taken for granted by researchers and that the core of a sustainable e-Infrastructure in the region is put in place.

While a number of countries in the Asia-Pacific region have significant investments in e-Infrastructures for research, the level of funding is still very heterogeneous and only a few countries have National Grid Initiatives (NGIs) that provide the necessary coordination at the national level to leverage the capability of Grids to provide persistent and sustainable e-Infrastructures that can be taken for granted by researchers and that enable them to focus on their substantive research. Until recently, most Grid-related initiatives were based at individual institutions that sought to build up capacity to support specific research projects and application areas. As a consequence, many resource providers ended up trying to support installations with different middleware stacks, stretching their resources. Clearly, a coordinated approach to the development of a persistent and sustainable e-Infrastructure would not only maximise the return on investment by enabling a wider range of researchers to benefit from the resources but would also help resource providers cope with the heterogeneity and continuous evolution of grid technologies.

Through the coordination and support provided by EUAsiaGrid, much needed local capacity has been developed, both in terms of resources available as part of the world-wide EGEE infrastructure and in terms of the supporting human infrastructure that is needed to carry on their ongoing operation and effective

exploitation by researchers. To minimize barriers to access the Grid infrastructure, the EUAsiaGrid project also created and maintained a catch-all, application neutral, Virtual Organisation called EUAsia. Furthermore, several Certification Authorities (CAs) approved by the International Grid Trust Federation[108] already operate in the region, with the Academia Sinica Grid Computing one[109] serving as a catch-all CA and taking care of users of those countries that do not yet have their national CA. Any researcher from the region interested in trying the Grid for his/her research can get a certificate through a nearby Registration Authority and immediately subscribe to the EUAsia VO. Although application neutral, nodes serving this VO have installed many application packages to be easily available. Also, each partner has set up a user interface to provide local access to the Grid.

The EUAsiaGrid project will come to an end in June 2010 but the EUAsiaGrid Consortium has agreed to keep open, on a best effort basis, the existing infrastructure (see Figure 11) and a Memorandum of Understanding is being approved and signed by the following partners:

- Academia Sinica, Taipei, Taiwan;
- Advanced Science and Technology Institute, Quezon Cirty, Phylippines;
- CESNET, Czech Republic;
- Hydro and Agro-Informatics Institute, Bangkok, Thailand;
- Institut de la Francophonie pour l'Informatique, Hanoi, Vietnam;
- Istituto Nazionale di Fisica Nucleare, Italy;
- Institut Teknologi Bandung, Bandung, Indonesia;
- National University of Singapore, Singapore;
- Universiti Putra Malaysia, Selangor, Malaysia.

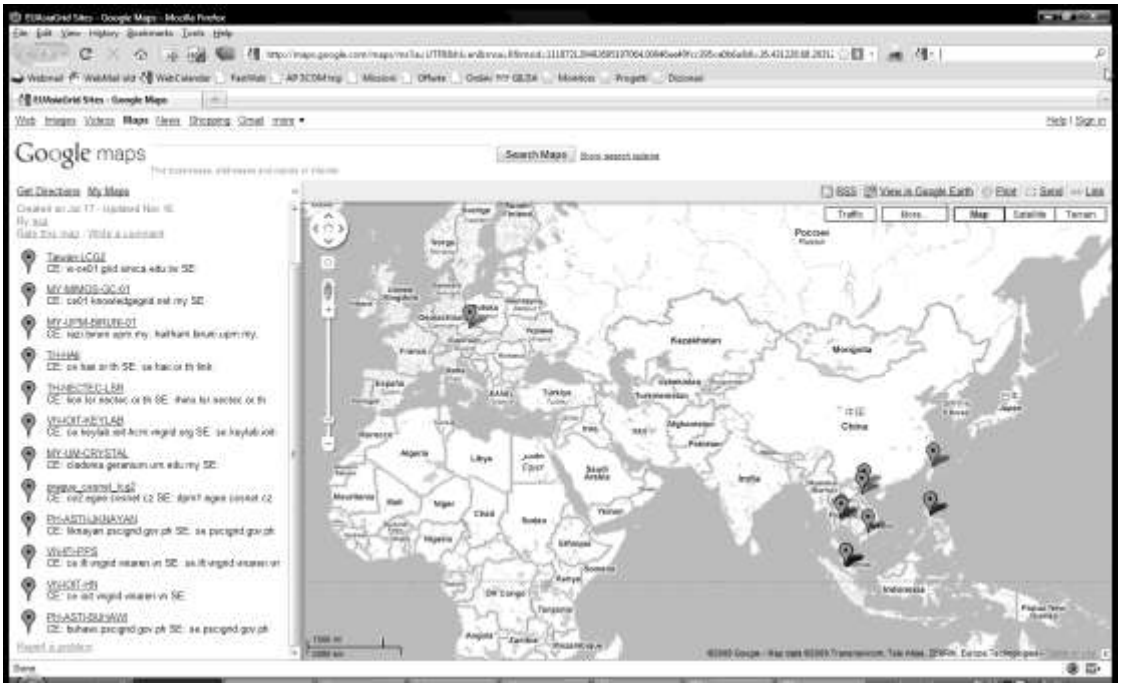


Figure 11. Map of the EUAsiaGrid project Grid infrastructure

6.2.2 E-Infrastructures in China

Network

The Chinese network for education and research is made of two infrastructures: CSTNET[110] and CERNET[111].

CSTNET, the China Science and Technology Network (see Figure 12), is the final phase of almost eleven years long activities in developing network connections among Chinese Academy of Science's (CAS) institutes from all over China.

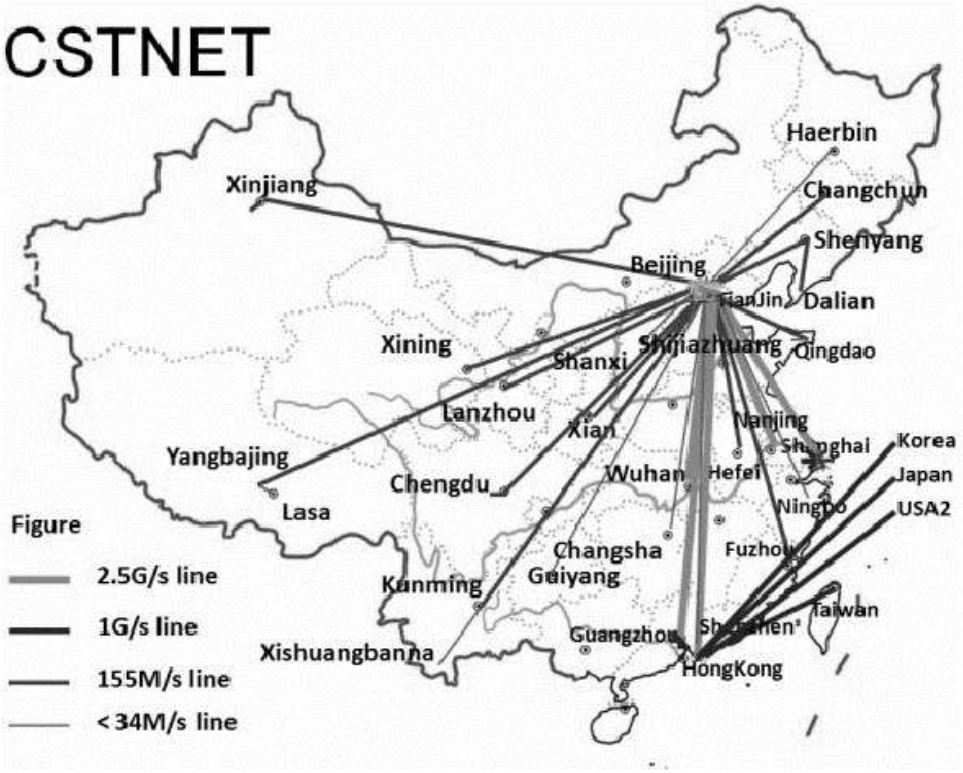


Figure 12. Topology map of the CSTNET network

CSTNET is a nationwide network for the scientific and technical communities, relevant government departments and hi-tech enterprises, providing services such as network access, host trusteeship, virtual host and domain name registration. CSTNET is becoming one of the top large-scale networks in China and is playing a key role in the development of China Internet Industry.

It is made of a backbone at 10 Gb/s, Metropolitan Area Network (MAN) links at 1 Gb/s and Wide Area Network links at 155 Mb/s – 2.5 Gb/s. The network interconnects 12 sub centres that cover more than 20 Chinese provinces, linking more than 100 research institutes of CAS and many other scientific and technical communities in the country for a total of more than 1 million end users.

CERNET, the China Education and Research Network, is a network connecting hundreds of Universities in more than 20 Chinese towns (see Figure 13).

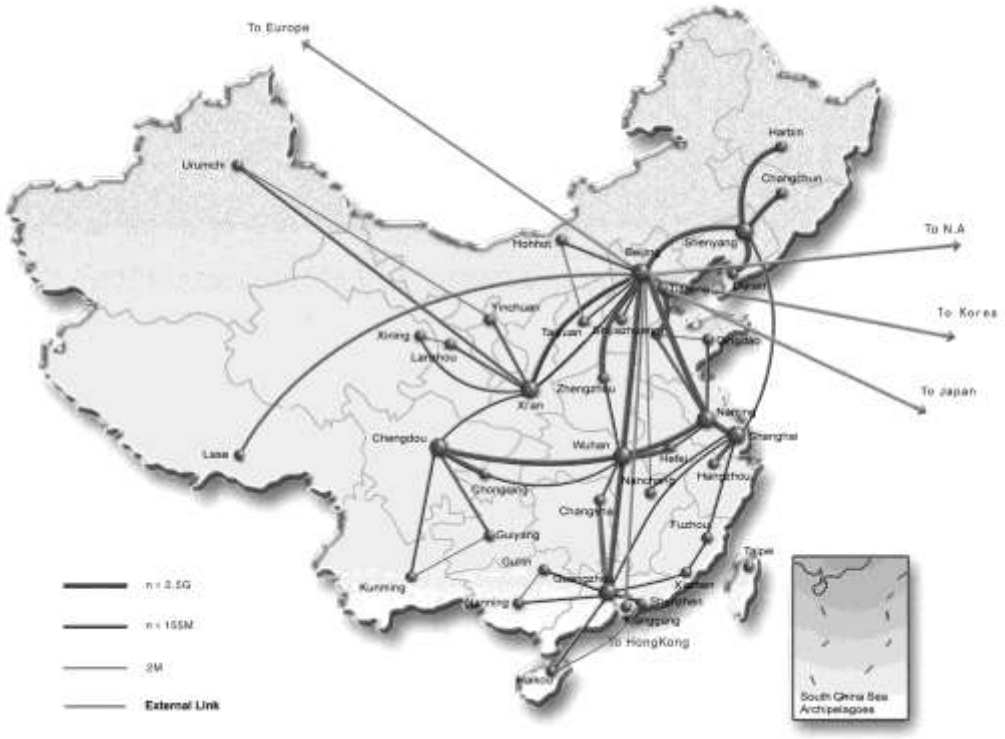


Figure 13. Topology map of the CERNET network

The development of CERNET started in 1994 as the first IPv4 nation-wide Internet backbone. Since 2003 CERNET is evolving in to CERNET2, the largest next-generation Internet backbone which is the core network of the China Next-Generation Internet (CNGI) demonstration project, a nation-wide, and world's largest, academic network based on a native IPv6 backbone. When fully completed, CERNET2 will connect all key research universities distributed in 20 cities around China at speeds comprised between 2.5 and 10 Gb/s and it will provide IPv6 connectivity to more than 200 universities, other research institutions and R&D organizations.

The basic connection between China and the rest of the world is ensured by means of three international connections: ORIENT, TEIN3, and GLORIAD.

ORIENT (see Figure 14) is a collaborative ICT project aiming to connect the research and education networks of China and Europe. Jointly funded by China and the European Commission, the project has procured and currently operates a high capacity data-communication link between the pan-European GÉANT2 backbone network and Chinese research and education networks.

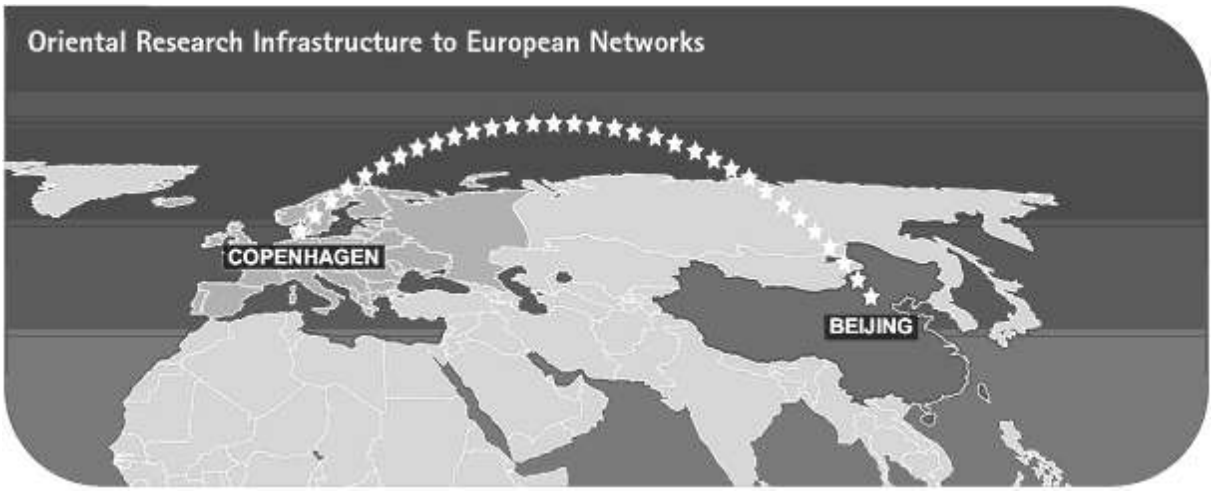


Figure 14. Pictorial map of the ORIENT network

The ORIENT project, working with TEIN2 has successfully procured a 2.5 Gb/s link on the shortest-possible trans-Siberian route. In fact, ORIENT is the single largest circuit using this important east-west communications route. After a period of testing and optimisation, the circuit was brought into full production service in January 2007.

TEIN3 (see Figure 9 above)**,** the third generation of the Trans-Eurasia Information Network, provides a dedicated high-capacity Internet network for research and education communities across Asia-Pacific. TEIN3 already connects researchers and academics in China, India, Indonesia, Japan, Korea, Laos, Malaysia, Nepal, Pakistan, the Philippines, Singapore, Sri Lanka, Taiwan, Thailand, Vietnam and Australia, Bangladesh, Bhutan and Cambodia are in the process of getting connected, bringing to 19 the total number of partners involved in the project.

GLORIAD (see Figure 15) is built on a fibre-optic ring of networks around the northern hemisphere of Earth, providing scientists, educators and students with advanced networking tools that improve communications and data exchange, enabling active, daily collaboration on common problems.

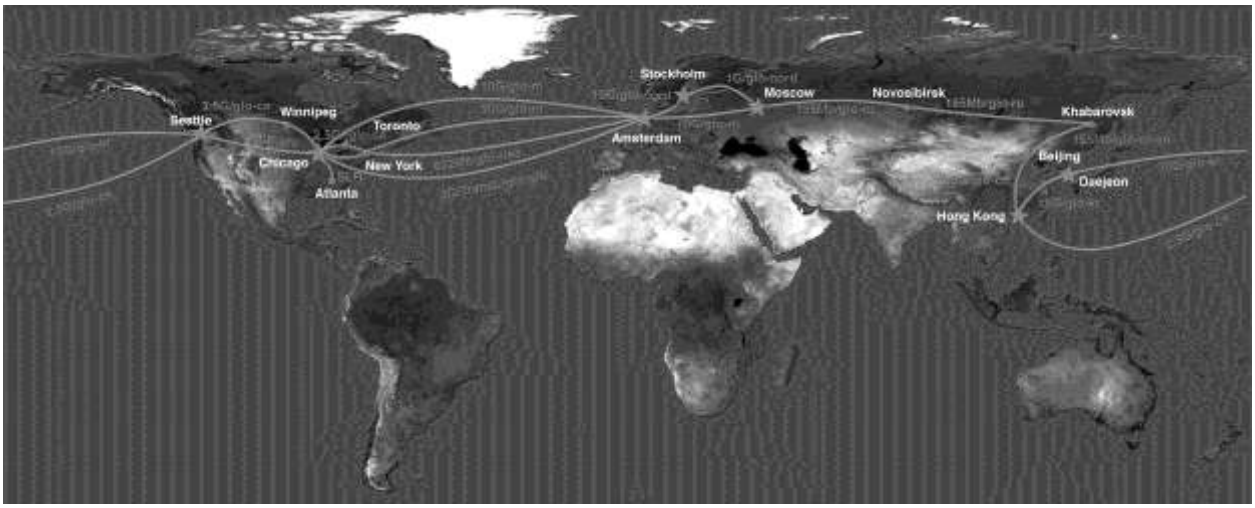


Figure 15. Topology map of the GLORIAD network

With GLORIAD, the scientific community can move unprecedented volumes of valuable data effortlessly, stream video and communicate through quality audio- and video-conferencing. GLORIAD exists today due to the shared commitment of the US, Russia, China, Korea, Canada, the Netherlands and the five Nordic countries of Denmark, Finland, Iceland, Norway and Sweden, to promote increased engagement and cooperation between their countries, beginning with their scientists, educators and young people. The benefits of this advanced network are shared with Science & Education (S&E) communities throughout Europe, Asia and the Americas.

Grid

China has developed both grid middleware and nation-wide e-Infrastructures. The most relevant are:

- CNGrid (see Figure 16), the China National Grid Project, is supported by the "High Performance Computer and its Kernel Software" project which, in turn, is a key project belonging to the National High-Tech R&D Program. The CNGrid is a test bed for the new generation of information infrastructure by integrating high performance computing and process transaction capacity. It efficiently supports various applications including scientific research, resource and environment research, advanced manufacturing and information service by sharing resources, collaborating and service mechanism. It also propels the progress of national e-Infrastructure and related industry through technology innovation. It is based on a middleware developed in China named GOS;
- ChinaGrid (see Figure 16), the China Education and Research Grid, is an important project funded by Chinese Ministry of Education and aims at constructing a public service system. It is also supported by the National High Technology Research and Development Program of China in the context of the “863 Program”. The goal of ChinaGrid is to integrate heterogeneous mass resources distributed in the China Education and Research Network (CERNET), share those resources in the CERNET environment effectively, avoiding the “resource islands”, provide useful services, and finally form the public platform for research and education in China;
- CROWN Grid is a test-bed to facilitate scientific activities in different disciplines, based on the Globus Toolkit middleware. It was formerly developed at the Beihang University and then became matter of cooperation between UK and China.

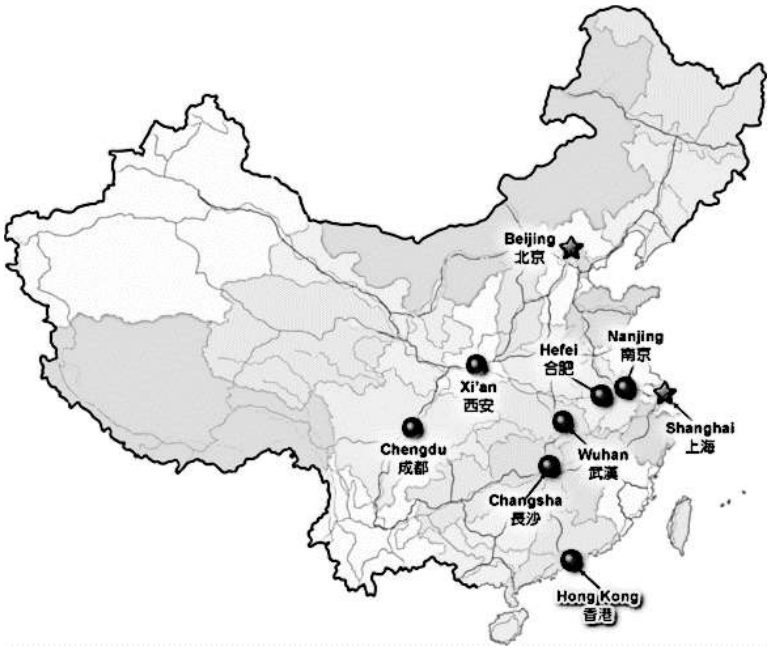


Figure 16. Map of the CNGrid and ChinaGrid infrastructures

A lot of other smaller grid infrastructures deployed just for few or, in some cases, just for one application are also present in the wide Chinese scenario.

In 2005, the European Commission, in the framework of its Sixth Framework Programme, funded the EUChinaGRID Project that ran from the 1st of January 2006 to the 31st of March 2008. The project aimed to foster a wider cooperation between Europe and China in the field of e-Infrastructures.

In detail, the main goals of the project were:

- to build a common Grid test-bed between China and Europe (see Figure 17);
- to support a set of applications which were selected as demonstrators;
- to study the middleware interoperability between gLite and GOS, as a basis for the real interconnection of the European Grid operated by the EGEE project and that managed by CNGrid;
- to study IPv6 compatibility of involved middleware.

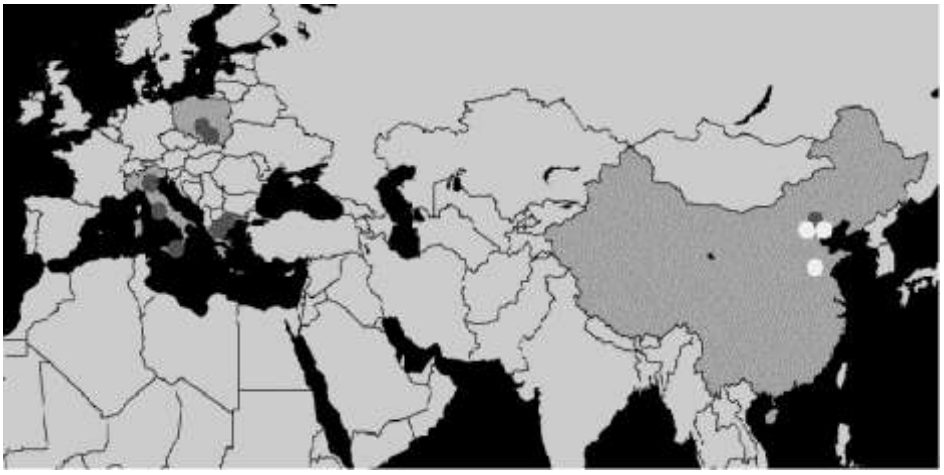


Figure 17. Map of the EUChinaGRID infrastructure

An intense activity of dissemination of and training on Grid computing paradigm was also part of the EUChinaGRID work plan and several hundreds of people were inducted to install, operate, access and use the Grid services deployed in the context of the project.

The project showed the feasibility and interest of such a common infrastructure. A "gateway" to connect gLite- and GOS-based infrastructures was developed and the IPv6 study produced a simple code checker to verify basic IPv6 compliance and a report was prepared on components of several middleware. This activity opened the way for a serious route to interoperation between European and Chinese e-Infrastructures.

6.2.3 E-Infrastructures in India

Network

In the last few years, the connectivity developments in India have been enormous, both at National and International level. In October 2006, as part of the EU-India co-operation program in ICT, for the first time the GÉANT-ERNET link was established to promote collaborative research between European and India with a link at 45 Mb/s. At present, the TEIN3 link interconnects the GÉANT network with India at 2.5 Gb/s.

The most prominent landmarks in the connectivity area since 2006 have been:

The establishment of the 45 Mb/s ERNET-GÉANT link and routing of regional WLCG data to CERN and subsequently the EU-IndiaGrid traffic to EGEE in 2006;

The upgrade of the GÉANT-ERNET link from 45 Mb/s to 100 Mb/s in 2008 and then to 175 Mb/s in 2009 and the upgrade of domestic bandwidth for the Indian organisations participating to WLCG;

The establishment of a dedicated 1 Gb/s TIFR-CERN link for LHC research in 2008 and peering with GÉANT in 2009;

The establishment of the National Knowledge Network (NKN) in April 2009;

The TEIN3 link at 2.5 Gb/s connecting India to GÉANT since February 2010;

The approval, by the Government of India, of the full National Knowledge Network Plan, in March 2010, with a total budget of about 1 billion euro.

The office of Principal Scientific Adviser to the Government of India and the National Knowledge Commission[112] have recently recommended the creation of the National Knowledge Network (NKN) as absolutely necessary for India's development. The objective of the National Knowledge Network is to bring together all the stakeholders in Science, Technology, Higher Education, Research & Development, and Governance with speeds in the order of 10's of Gb/s coupled with extremely low latencies. NKN will interconnect all institutions engaged in research, higher education and scientific development in the country, over a period of time. It would enable use of specialized applications, which allow sharing of high-performance computing facilities, e-libraries, virtual classrooms, and very large databases.

In the initial phase of NKN, 15 core locations and about 57 institutes covering leading national R&D labs and educational institutes, have been connected at varying bandwidths of 100 to 1000 Mb/s (see Figure 18).

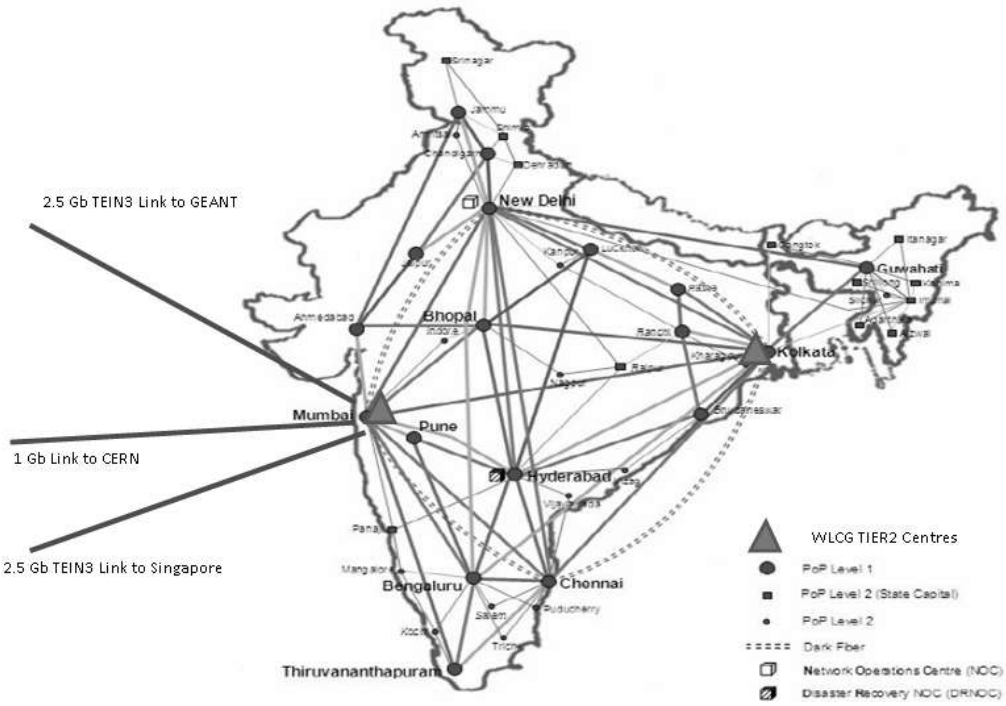


Figure 18. Topology map of the NKN network

NKN has been proposed as national program and the network will be sustained through continuous government funding. In its final phase, around 5,000 leading national academic and research institutes are going to be connected by NKN. On March 2010, the Government of India approved the full National Knowledge Network Plan with a total budget of about 1 billion euro.

NKN with its multi-gigabit, low-latency, optical fibre based backbone is acting as national transport for all existing networks. The Indian National Grid Initiative GARUDA[113] is based on NKN. The regional WLCG in India is going to be migrated to NKN. NKN will provide transport to ERNET[114] (the Indian National Research and Education Network) replacing its existing backbone. The main design consideration for NKN is to create an infrastructure that can scale and adapt to future requirements.

The project's ultimate aim is to unite stakeholders in science, technology, higher education, R&D and e-governance. The NKN is expected to foster collaboration and the creation of new national intellectual assets, enabling the sharing of high-performance computing facilities, e-libraries, virtual classrooms, and more. The NKN will also provide access to global content on emerging technologies, thus allowing close coordination among different institutions across nations.

NKN will be used for the following major applications.

Education: Education is going to be a major application to be deployed over NKN. E-Learning services, Digital Libraries, Data Centres, Compute Servers, Secure monitoring systems, Information search services, voice and video conferencing across educational institutions are waiting for large scale deployment. Besides, high-degree molecule decomposition, polymer synthesis simulations, aerodynamic and thermo-dynamic modeling are all waiting for the appropriate environment. All these applications not only require very high bandwidth, but also require real-time guarantees that the required bandwidth will indeed be available on-demand. In fact, the applications envisaged are more appropriate for India, more than any other country in the world. All of them have a direct impact on the quality of life and the quality of education in our country. Countrywide virtual classrooms alone will justify any amount of bandwidth and any amount of investment.

Health care and related sciences have experienced exponential development of knowledge and it is almost impossible for any medical library to store such huge knowledge database, which is constantly updated. The NKN will be able to address this concern as well. Also the proposed network would facilitate in reducing the differences among different medical institutions in infrastructure, teaching material, and quality of knowledge, skill and teachers. NKN will thus foster knowledge sharing and collaborative research. Besides, NKN will enable applications in the domain of telemedicine leading to better quality of life. Cardiac care, eye care, cancer care are a few applications that touch human lives.

Agriculture would be a major thrust area in the field of content creation. The content would include research on horticulture, livestock, fisheries, biochemistry, agronomy, environmental science, microbiology, seed research and genetics. The content shared on the proposed network would also include the agricultural statistical research, remote sensing, GIS5, production & area study and estimation.

Global R&D Applications: Today, unlike in the past, the research and development activities are not limited to few developed countries. India has been playing an increasingly important role in the global scenario and has actively contributed in development of new technologies. NKN would provide access to global content on emerging technologies, would allow close coordination among different institutions across nations. Therefore, NKN will help promote further research in specialized areas like bio-informatics, Grid computing, genetics, etc.

In the design philosophy of NKN, high speed connectivity to global research networks have also been envisaged. To provide high-speed connectivity to users of NKN, a TEIN3 Point of Presence (PoP) has been co-located at the ERNET PoP in Mumbai and it is acting as hub for connecting research networks in South

Asia, except Pakistan. From Mumbai, two high speed links at 2.5 Gb/s each have been commissioned to Europe and Singapore and are now operational providing direct connectivity both to GEANT and the TEIN3 PoP in Singapore. India is now acting as a hub for connectivity between Europe and the Asia-Pacific countries. The European Commission is partly funding the connectivity under TEIN3. At present, the TEIN3 PoP is located inside the C-DAC campus in Mumbai. In the long run, it will be relocated at the NKN PoP, still in Mumbai.

Grid

In India, two main Grid Initiatives have been taken at governmental level: Regional WLCG set up by the Department of Atomic Energy (DAE), in coordination with the Department of Science & Technology (DST), and the GARUDA National Grid Initiative. The EU-IndiaGrid project, operating within the Sixth Framework Program of the European Commission, has played a bridging role between European and Indian grid infrastructures and its successor, EU-IndiaGrid2 (EU-IndiaGrid2, 2010), aims at increasing the cooperation between European and Indian e-Infrastructures capitalizing on the EU-IndiaGrid achievements.

The Large Hadron Collider, built at CERN near Geneva, is the largest scientific instrument on the planet and it just started its data-taking phase. In full operation, it will produce roughly 15 million gigabytes of data annually, which thousands of scientists around the world will access and analyse. The mission of the Worldwide LHC Computing Grid project is to build and maintain a data storage and analysis infrastructure for the entire high-energy physics community that will use the LHC. The Indian Department of Atomic Energy (DAE) is actively participating to the scientific program taking active part in CMS and ALICE experiments, devoted to find answers to the most fundamental questions at the foundations of matter constituents. The data from the LHC experiments will be distributed around the globe, according to a four-tiered model. Within this model, to support researchers with required infrastructure, India has also setup regional Tier-2 centres connected to CERN. In India there are two Tier2 centres: one for CMS at TIFR in Mumbai and one for ALICE at Saha-VECC in Kolkata. These centres provide access to CMS and ALICE users working from Tier-3 centres at Universities and national labs and LCG Data Grid services for analysis. TIFR is presently connected to CERN at 1 Gb/s and very soon it will exploit the 2.5 Gb/s TEIN3 link. Now TIFR and VECC are also being connected through NKN at 1 Gb/s.

Specific activities are also ongoing in the area of Grid Middleware Software development, devoted to ensuring grid enabling of IT systems. These activities cover the area of Grid Fabric management, Grid Data management, Data Security, Grid workload scheduling and monitoring services, fault tolerant systems, etc. DAE developed number of Grid based Tools in the area of Fabric management, AFS file system, Grid View and Data Management, which are being deployed by CERN in their LHC Grid operations since September 2002. So far, the number of software tools and packages such as a correlation engine, Grid operations monitoring, problem-tracking system, Pool Database Backend Prototype, Scientific Library Evaluation and Development of Routines, AliEn Storage System and Andrews File System, were developed by DAE team members under a Computing Software agreement. Currently, BARC engineers are working on the enhancement of the Grid View software tool.

GARUDA is a collaboration of scientific and technological researchers for a nation-wide Grid comprising computational nodes, mass storage systems and scientific instruments. It aims to provide the technological advances required to enable data and compute intensive science for the 21st century.

C-DAC, one of EU-IndiaGrid's main partners, ensures progressive evolution and durable integration as manager of the Indian National Grid Initiative and, from the start of its activity, the EU-IndiaGrid project established an excellent collaboration with GARUDA. The map of GARUDA sites is shown in Figure 19.

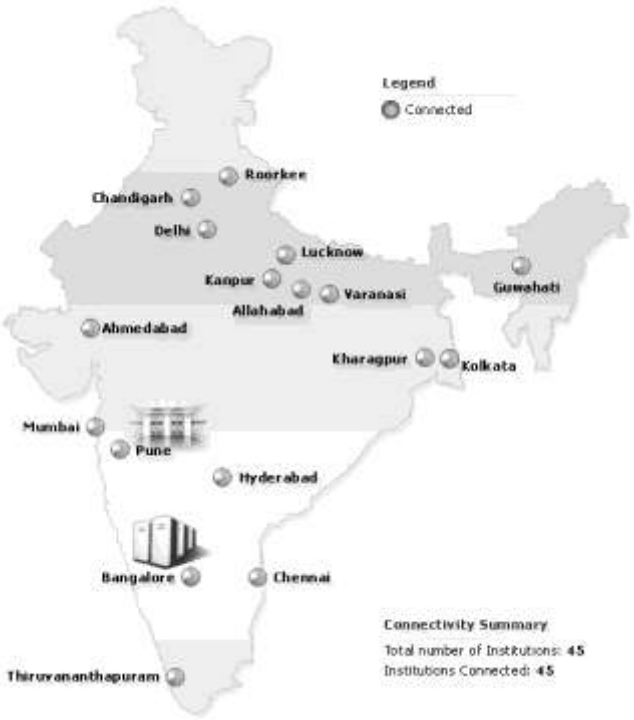


Figure 19. Map of the GARUDA infrastructure

GARUDA has transitioned form the Proof of Concept phase to the Foundation Phase in April 2008 and currently is in its third phase: Grid Technology Services for Operational Garuda. This phase has been approved and funded for three years until July 2012.

Some of the envisaged deliverables of this phase include:

Delivering Service based Grid with tools to support ease of use;

On-demand provisioning of resources;

Ensure QOS and end-to-end reliability for applications;

Open and standards based implementation;

Supports Inter-operability across Grids;

Deploy select identified application as service for end user consumption.

The GARUDA project coordinator, CDAC, established in November 2008 an IGTF recognized Certification Authority[115] which allows access to worldwide Grids for Indian Researchers.

GARUDA aims at strengthening and advancing scientific and technological excellence in the area of Grid and Peer-to-Peer technologies. It will also create the foundation for the next generation Grids by addressing long term research issues in the strategic areas of: knowledge and data management, programming models, architectures, grid management and monitoring, problem solving environments, tools and grid services.

EU-IndiaGrid and EU-IndiaGrid2 are part of a group of projects, funded within the Sixth and Seventh Framework Programs for Research and Scientific Development of the European Commission, which aim at integrating the European grid infrastructure with other regions in order to create one broad resource for scientists working on existing or future collaboration.

The EU-IndiaGrid project ran from 2006 to 2009. The leading responsibilities of the EU-IndiaGrid Indian partners and the project bridging role between European and Indian e-Infrastructures gave to EU-IndiaGrid project the opportunity to be at the core of the impressive developments in India in the e-Infrastructures domain and to effectively contribute at improving cooperation between Europe and India in this area. In all these activities, the role and the contribution of EU-IndiaGrid partners, as well as the bridging role of the EU-IndiaGrid project, was particularly relevant and obtained full recognition at the highest level by representatives of the Indian Government and of the European Commission so contributing at supporting the improvement of the e-Infrastructures capabilities. Its successor EU-IndiaGrid2 will run from January 2010 to December 2011 and capitalizes on the EU-IndiaGrid achievements by acting as a bridge across European and Indian e-infrastructure to ensure sustainable scientific, educational and technological collaboration. The project launch, on January 11th 2010, occurred in the same week of the EU-India Thematic Workshop on Research Infrastructures, one of the agreed bilateral actions under the overall ambit of India – European Union Science & Technology Cooperation. The Workshop, where the EU-IndiaGrid2 project actively contributed, underlined the role of e-Infrastructure to favor Euro-India Science and Technology cooperation and the role of agreement where the project actively contributed. The EU-IndiaGrid contribution in the Euro-India e-Infrastructures cooperation's scenario and the perspectives for EU-IndiaGrid2 can be well resumed in the words of Dr. Chidambaram, Principal Scientific Advisor to the Govt. of India, who gave the opening speech both at EU-India Thematic Workshop on Research Infrastructures and at the EU-IndiaGrid2 project launch: "*I am happy to learn about the second phase EU-IndiaGrid2 project – Sustainable e-Infrastructures across Europe and India. The first phase has benefited immensely a variety of scientific disciplines including biology, earth science and the Indian collaboration for the Large Hadron Collider (LHC). The successful working of the initial phase of multi-gigabit National Knowledge Network, Indian Certification Authority, and participation in Trans-Eurasia Information Network (TEIN3) phase 3 are some of the important building blocks for supporting virtual research communities in India and their collaboration work with other countries.*"

6.2.4 E-Infrastructures in Latin America

Network

CLARA[116] (Cooperación Latino Americana de Redes Avanzadas) is the legal entity responsible for the implementation and management of the network infrastructure that interconnects the Latin American NRENs: the RedCLARA. The RedCLARA backbone (see Figure 20), composed of 9 nodes, interconnects 12 Latin American NRENs: RNP (Brazil), InnovaRed (Argentina), REUNA (Chile), RENATA (Colombia), CEDIA (Ecuador), RAICES (El Salvador), RAGIE (Guatemala), CUDI (Mexico), RedCyT (Panama), RAAP (Perú), RAU2 (Uruguay) and REACCIUN (Venezuela).

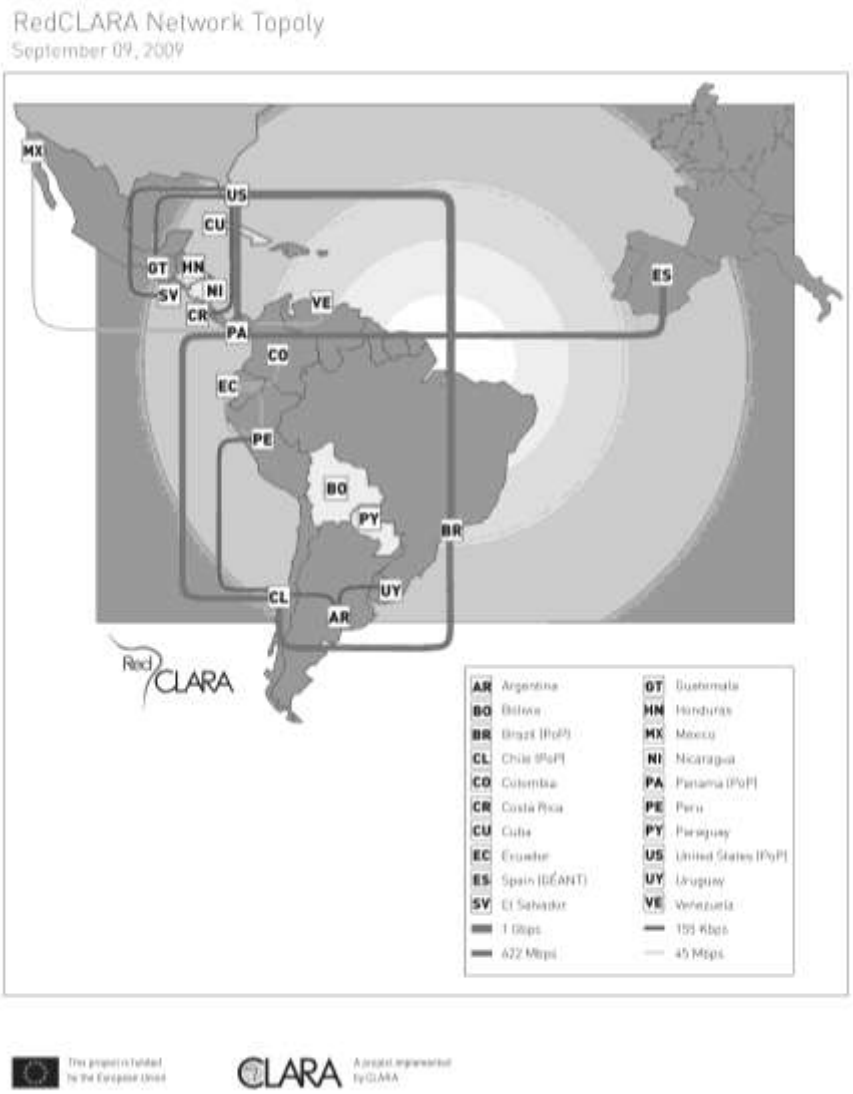


Figure 20. Topology map of the RedCLARA backbone network

The RedCLARA backbone is also connected to the pan-European network GÉANT, as well as to other international research networks.

CLARA was created in June 2003, as a membership association for national NRENs in Latin America. The European Commission EuropeAid project ALICE, within the @LIS programme, was jointly executed between 2003 and 2008 by 4 European NRENs (from France, Italy, Portugal and Spain) and the Latin American NRENs members of CLARA, with the coordination of DANTE, afterwards assisted by CLARA itself. The principal deliverable of ALICE was the RedCLARA network, inaugurated in September 2004. The ALICE project was in large part financed by the European Commission, with contributions from the Latin American NRENs. After ALICE termination, in March 2008, the network continued to be maintained by Latin American contributions.

In 2008, a new project, ALICE2, was approved by the European Commission and will receive funding until 2012. It has as principal aim to build a robust and modern regional network, which will be financially sustainable after its end. The project, which began in November 2008, is coordinated by CLARA which is seeking to acquire long-term access to telecommunications infrastructure, such as optical fiber and wavelengths, which can be used to provide scalable network capacity with low maintenance cost. The resulting network is expected to display large increases in bandwidth, compared with the present capacity. The first results of this new approach became available in the "Southern Cone" countries in 2009, where international links of 10 Gb/s were made available. Additionally, the ALICE2 roadmap includes the improvement of network connectivity of the individual Latin American NRENs through sharing the new telecommunications infrastructure.

In the second quarter of 2009, CLARA, InnovaRed, RNP, the AugerAccess[117] project and Silica Networks enabled provisioning of RedCLARA's first 10 Gb/s "lambda", between Buenos Aires and Santiago of Chile. The joint infrastructure provided access to the Pierre Auger Southern Cosmic Ray Observatory[118] in Malargüe, Argentina.

Complementary to the investments in ALICE2, Brazil will also invest up to US$ 10,000,000 in three years to fund connections between Mercosur[119] countries, which will form part of RedCLARA. This investment is complementary to ALICE2 and will be used as a contribution to counterpart funding of this project. In this context, a joint Brazil-Argentina project is under study with engineering and management under CLARA responsibility, to be operational by 2010 and connecting Buenos Aires, Rosario, Uruguaiana, Porto Alegre, and Sao Paulo. For the connection Brazil–Uruguay opportunities of acquiring optical fibre are under investigation. There are also discussions under way for a Brazil–Paraguay connection, using fibre on energy transmission lines from the Itaipu Binational hydroelectric scheme.

Plans in the remainder of the region are less well-defined, but follow the same strategy of seeking strategic partners with their own optical fibre. In many countries, the most likely candidates are electrical companies which can install optical cables along their high-voltage transmission lines. Many of these countries have agreements with their neighbours which have led to cross-border integration of electrical transmission networks, and also, as a consequence, of optical fibre networks.

In México, the local NREN CUDI is currently negotiating the creation of a new high capacity network based on optical fibres belonging to electrical companies. This network will provide 10 Gb/s links to 9 cities as well as cross border links to the neighbour countries of the US (three 10 Gb/s links) and Guatemala (a single 2.5 Gb/s link).

In Central America, since 2001, an ambitious plan for regional development, known as the Proyecto Mesoamérica, formerly Plan Puebla-Panamá[120], is being carried out by the contiguous set of countries from México to Colombia with financial support from the Inter-American Development Bank (IADB) and the Central American Bank for Economic Integration (CABEI). Amongst the projects being carried out are SIEPAC (System for Electrical Interconnection of Central American countries) and AMI (Mesoamerican Information Highway), and these are expected to be completed during the lifetime of the ALICE2 project, permitting a terrestrial fibre link between Mexico and Panama (to the Colombian border).

Similar integration is going on in the northern Andean region of South America, where the Colombian electrical company ISA (Interconexiones Elétricas S.A.) is a leading player in initiatives to build transmission lines interconnecting the power grids of several countries (Colombia, Venezuela, Ecuador, Peru, Bolivia). Internexa, a telecommunications company of the ISA group, now acts as a player in the international telecommunications market and has provided connectivity for RedCLARA between Colombia, Ecuador and Peru since 2008.

Thus, there are reasonable expectations within the near future of continuous high-capacity terrestrial connections between Latin American countries from Mexico as far south as Bolivia, and also within the Southern Cone countries (Argentina, Brazil, Chile, Paraguay and Uruguay). In order to complete the North-

South connectivity, there remains the gap along the Pacific coast between southern Peru and central Chile where only conventional telecommunications companies are present. This is the area of the Atacama Desert, home to several large-scale astrophysics observatories, such as Atacama Large Millimeter/submillimeter Array (ALMA), European Southern Observatory (ESO) and Cerro Tololo Inter-American Observatory (CTIO). These sites are of great interest to the international scientific community and it is expected that this will provide the necessary impetus to establish high-capacity connectivity to the international research networks, both northward, towards the terrestrial links to Central and North America, and also southward, towards to the "Southern Cone" interconnections and their submarine cable links to the US and Europe, through the provision of fibres to Peru and central Chile, in partnership with RedCLARA.

Finally, it should be emphasized that the upgrades planned in RedCLARA through the introduction of high-capacity terrestrial fibre links will also benefit the NRENs, who will be able to share the same fibre links in order to improve internal connectivity in the countries traversed by the international links. This process has already begun in Argentina, where InnovaRed is building out a new 10 Gb/s backbone network in partnership with RedCLARA (and others). This is expected to serve as the new paradigm for NREN deployment in the foreseeable future.

Grid

The EELA-2 Project (E-science grid facility for Europe and Latin America) is by far the most inclusive initiative that has gone on in Latin America in the area of distributed computing infrastructures. EELA-2, ended on the 31st of March 2010, aimed at building a high capacity, production-quality, scalable Grid to answer the needs of a wide spectrum of applications from European-Latin American scientific collaborations.

Its focus was on:

- Offering a complete set of versatile services fulfilling Applications requirements;
- Ensuring the long-term sustainability of the e-Infrastructure beyond the term of the project.

Such an ambitious project would not have been possible without the prior existence of a consolidated e-Infrastructure, set up with the early intention to build a sustainable Grid platform. This was the objective of the EELA first-phase project that provided its users with a stable and well supported Grid which proved, over 2006-2007, that the deployment of an European-Latin American e-Infrastructure was not only viable but also responding to the real needs of a significant part of the Scientific Community.

By the end of the project, the EELA-2 Consortium encompassed 78 Institutions from 16 countries, 5 from Europe (France, Ireland, Italy, Portugal and Spain) and 11 from Latin America (Argentina, Brazil, Chile, Colombia, Cuba, Ecuador, Mexico, Panama, Peru, Uruguay and Venezuela).

The EELA-2 infrastructure, shown in Figure 21, consists of 29 Resource Centres for a total of 8,000 CPU cores and 200 TB of storage.

Figure 21. Map of the EELA-2 infrastructure

6.2.5 E-Infrastructures in the Mediterranean and the Middle-East

Network

The EUMEDCONNECT project, co-funded by the European Commission in the context of its Sixth Framework Program, has played a pioneeristic role in the promotion of Communication Networks, as fundamental components of e-Infrastructures in the Mediterranean. This activity is currently being coordinated with two main initiatives: the EUMEDCONNECT2 project and the recently launched Arab Scientific Research and Education Network[121] initiative.

EUMEDCONNECT2, a follow-up of EUMEDCONNECT, is co-funded by the European Commission in the context of its Seventh Framework Program and aims to sustain and upgrade the high-capacity IP-based data-communications network serving the research and education communities in seven countries across the southern Mediterranean, enabling them to participate in collaborative projects. Offering a direct link to GÉANT, its pan-European counterpart, EUMEDCONNECT2 allows approximately 2 million users in around 700 institutions across North Africa and the Middle East to collaborate with their peers at more than 3,000 research and education establishments in Europe. EUMEDCONNECT2, whose topology map is shown in figure 22, acts as a real gateway to global research collaboration.

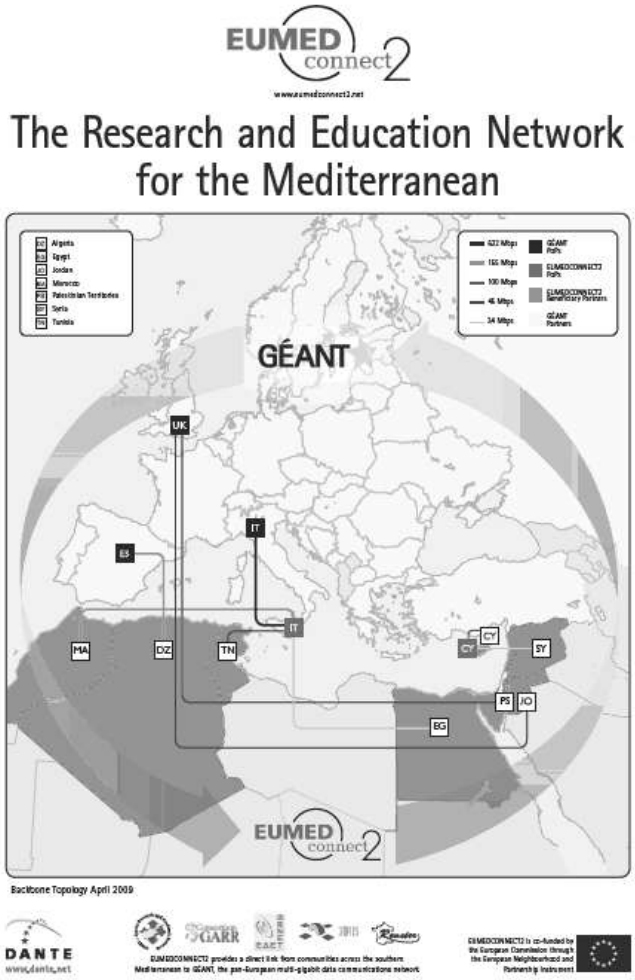


Figure 22. Topology map of the EUMEDCONNECT2 network

The Arab Mediterranean countries participating in the EUMEDCONNECT project series (Morocco, Algeria, Egypt, Palestine, Jordan and Syria) have signed the "Rome Declaration" in September 2006. These countries stated that they will support the establishment of National Research & Education Networks (NREN) in their countries, lead the efforts to further develop this regional network, and also conduct the proper promotion among these countries to use this research infrastructure. These countries are currently working to establish a legal organisation called ASREN which will focus on raising the necessary funds to build the regional network and to promote the use of networks and e-Infrastructures in research and education.

Grid

Co-funded by the European Commission within the Sixth Framework Programme, the EUMEDGRID has run in parallel but in conjunction with EUMEDCONNECT project and has supported the development of a Grid infrastructure in the Mediterranean area. EUMEDGRID also promoted the porting of new applications on the Grid platform, thus allowing Mediterranean scientists to collaborate more closely with their European colleagues. EUMEDGRID has disseminated Grid awareness and competences across the Mediterranean and, in parallel, identified new research groups to be involved in the project, helping them to exploit Grids' enormous potential to improve their own research activities.

The implementation and coordination of a Grid infrastructure at a national (or wider) level can be regarded as an opportunity to optimize the usage of existing, limited storage and computing resources and to

enhance their accessibility by all research groups. This is particularly relevant for the non-EU countries involved in the project.

Many research fields have indeed very demanding needs in terms of computing power and storage capacity, which are normally provided by large computing systems or supercomputing centres. Furthermore, sophisticated instruments may be needed to perform specific studies. Such resources pose different challenges to developing economies: they are expensive, they need to be geographically located in a specific place and they cannot attract a critical mass of users because they are usually very specific and are relevant only for small communities of researchers scattered across the country/region. This is the case even in some strategic domains such as water management, climate change, biodiversity and biomedical activities on neglected or emerging diseases. Thus, a significant part of researchers is forced to emigrate to more developed countries to be able to continue their scientific careers. However, thanks to the creation of global virtual research communities and distributed e-Infrastructure environments, all these drawbacks can be overcome: through an appropriate access policy, different user groups can use resources wherever dispersed, according to their availability. Furthermore, geographically distributed communities working on the same problem can collaborate in real time thus optimizing not only hardware and software resources but also human effort and "brainware".

The EUMEDGRID project was conceived in this perspective and has set up a pilot Grid infrastructure for research in the Mediterranean region which is interoperable and compatible with that of the EGEE project and related initiatives. The EUMEDGRID's vision focused on improving both the technological level and the know-how of networking and computing professionals across the Mediterranean thus fostering the introduction of an effective Mediterranean Grid infrastructure for the benefits of e-Science.

The EUMEDGRID achievements can be categorised into two main areas:

- The creation of a *human network* in e-Science across the Mediterranean;
- The implementation of a pilot Grid infrastructure, with *gridified* applications, in the area.

Cooperation among all the participants has been demonstrated by the enthusiastic participation to common workshops and meetings organized during the lifetime of EUMEDGRID and the success obtained fostering the creation of National Grid Initiatives and national Certification Authorities (CAs) officially recognized by IGTF. Impressive results were also obtained in the events of knowledge dissemination on Grid technology and services. A large community, including system administrators, researchers, and final users, was involved with good results in terms of number of participants (more than 700 individuals) and feedback obtained through dedicated questionnaires.

The promotion of National Grid Initiatives carried out in all non-EGEE Partner Countries registered a good level of success with programmes already operational in Algeria, Egypt, Morocco and Tunisia and well advanced plans in Jordan and Syria. The national impact and policy level awareness in some of these countries has led to an initial financial support of the initiatives.

The project has been very active in promoting the creation of national Certification Authorities which issue digital certificates necessary to allow secure Grid access to the users. The process is completed in Morocco, the first African Country to become member of EUGridPMA[122], the international organisation to coordinate the trust fabric for e-Science Grid authentication in Europe, and well advanced in the other countries. In the meanwhile, a temporary catch-all CA was set-up at INFN in order to fulfil the needs of EUMEDGRID users not having a Certification Authority in their countries.

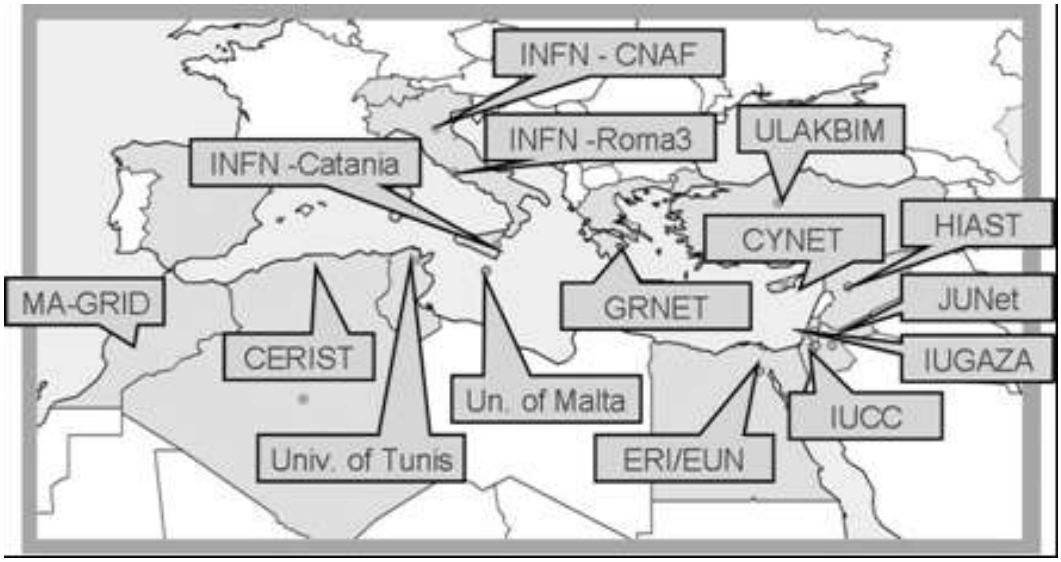


Figure 23. EUMEDGRID partners operating Grid sites

A pilot Grid infrastructure, composed to date of 25 sites in 13 countries, was set up during the time span of EUMEDGRID.

Besides its scientific mission, EUMEDGRID had also a significant socio-economic impact in the beneficiary countries. Fostering Grid awareness and the growth of new competences in EU Neighbours' scientific communities is a concrete initiative towards bridging the digital divide and the development of a peaceful and effective collaboration among all partners.

e-Infrastructures also contribute to mitigate the so-called "brain-drain" allowing brilliant minds in the area to stay in their regions and contribute significantly to cutting edge scientific activities, concretely enlarging the European Research Area (ERA). Research and Education Networks and Grids are fundamental infrastructures that will allow non-EU researchers to carry out high quality work in their home laboratories without the need to migrate in more advanced countries.

An extended Mediterranean Research Area could thus be seen as a first step towards the realisation of more politically ambitious plans of open market, open transportation infrastructures, free circulation of citizens, etc.

Finally, the EUMEDGRID Consortium has agreed before coming to an end to keep open, on a best effort basis, the existing Infrastructure and a formal agreement has been approved and signed by the following partners:

- Centre de Recherche sul l'Information Scientifique et Technique (CERIST), Algeria;
- Centre de Calcul Khawarezmi (CKK), Tunisia;
- Centre National pour la Recherche Scientifique et Technique (CNRST), Morocco;
- Consortium GARR (GARR), Italy;
- Cyprus Research and Academic Network (CYNET), Cyprus;
- Egyptian Universities Network (EUN), Egypt;
- Electronic Research Institute (ERI), Egypt;
- Greek Research and Technology Networks S.A. (GRNET), Greece;
- Higher Institute of Applied Sciences and Technology (HIAST), Syrian Arab Republic;
- Jordanian University Network (JUNET), Jordan;
- Istituto Nazionale di Fisica Nucleare (INFN), Italy;
- Tubitak Ulusal Akademik ag ve Bilgi Merkezi (TUBITAK ULAKBIM), Turkey;
- Universita ta Malta (UoM), Malta.

EUMEDGRID finished in 2008 but the new project EUMEDGRID-Support, co-funded by the European Commission, has recently started on the 1st of January 2010 and will continue until the 31st of December 2001. EUMEDGRID-Support obviously builds on EUMEDGRID outcomes and aims at:

- Pushing for a consolidation of the existing EUMEDGRID infrastructure and for the development of sustainable e-Infrastructures in the Mediterranean region in a broad, general, meaning;
- Promoting the completion of the process of creation of Certification Authorities in the Mediterranean Countries;
- Exploiting the maximum level of synergy with other initiatives and projects and specifically cooperate with the EPIKH[123] project for advanced knowledge dissemination actions.

6.2.6 E-Infrastructures in South-East Europe

Network

In the past 6 years, a number of targeted initiatives funded by the European Commission via its RTD programmes, as well as national and regional funding sources, have contributed to bridging the digital divide in the South-Eastern European (SEE) region.

The SEEREN and SEEREN2 (South-East European Research and Education Networking initiatives) projects have established the SEE segment of the pan-European GÉANT network and successfully connected the research and scientific communities in the region. Most of the countries in the region are now part of GÉANT. Currently, the SEE-LIGHT project is working towards establishing a dark-fibre backbone that will interconnect most national Research and Education networks in the region. The dark fibre backbone is funded by Hellenic Plan for the Economic Reconstruction of the Balkans (HiPERB). The topology of available fibres at the study/analysis stage is show in Figure 24.

Figure 24. Topology map of the SEE-LIGHT network

Grid

The SEE-GRID and SEE-GRID2 (South-East European GRid e-Infrastructure Development) projects have established a strong human network in the area of scientific computing and have set up a powerful regional Grid infrastructure, and attracted a number of applications from diverse fields from countries throughout South-East Europe. The current SEE-GRID-SCI project, ending in April 2010, empowers the regional user communities from fields of meteorology, seismology and environmental protection in common use and sharing of the regional e-Infrastructure. Current dedicated resources for these 3 major VOs are in the order of 2000 CPU cores and 300 TB of storage, spread over more than 40 Grid sites (see Figure 25).

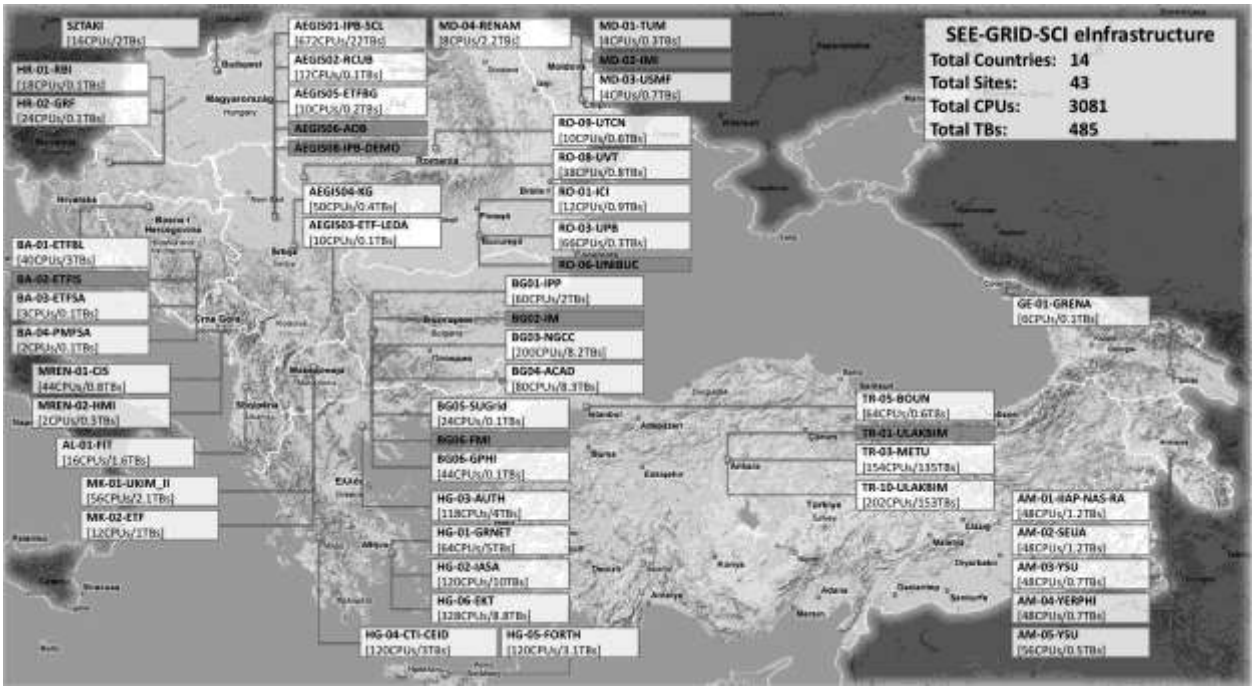


Figure 25. Map of the SEE-GRID-SCI infrastructure

The joint regional operations consist of maintaining deployed core services for SEEGRID Virtual Organisation and three discipline-specific VOs, as well as core services for the ops.vo.egee-see.org VO used for testing of the infrastructure. A set of operational and monitoring tools is maintained and used to manage and assess the status of the infrastructure. In this way, operations are distributed and countries which are not part of pan-European EGEE infrastructure are effectively supported.

The Grid initiatives are coordinated by the Greek Research & Technology Network[124] and the wider consortium, which is in long-term formalised via a Memorandum of Understanding for a multi-national Joint Research Unit, consists of representatives from National Grid Initiatives of Bulgaria (IPP), Romania (ICI), Turkey (ULAKBIM), Albania (UoPT), Bosnia-Herzegovina (UoBL), FYR of Macedonia (UKIM), Serbia (UOB), Montenegro (UOM), Moldova (RENAM), Armenia (IIAP NAS RA), Georgia (GRENA) and Azerbaijan (Institute of Physics).

The wider SEE region can be considered as a model region that has achieved European e-Infrastructures full integration, apart from network aspects where the non-GÉANT countries such as Albania, Bosnia-Herzegovina and Caucasus still require specific funding actions for network links.

The High Performance Computing initiatives in SEE are starting in different countries separately, with an upcoming HP-SEE project to coordinate them at a regional level.

6.2.7 E-Infrastructures in Sub-Saharan Africa

Network

With few exceptions, African universities and research centres lack access to dedicated global research and education resources because they are not connected to the global infrastructure consisting of dedicated high capacity regional networks. The consequence is that research and higher education requiring such access can currently not be conducted in Africa and the continent is not well represented in the global

research community. This is witnessed by the world map of scientific divide (see Figure 26) where territory size shows the proportion of all scientific papers (published in 2001) written by authors living there[125].

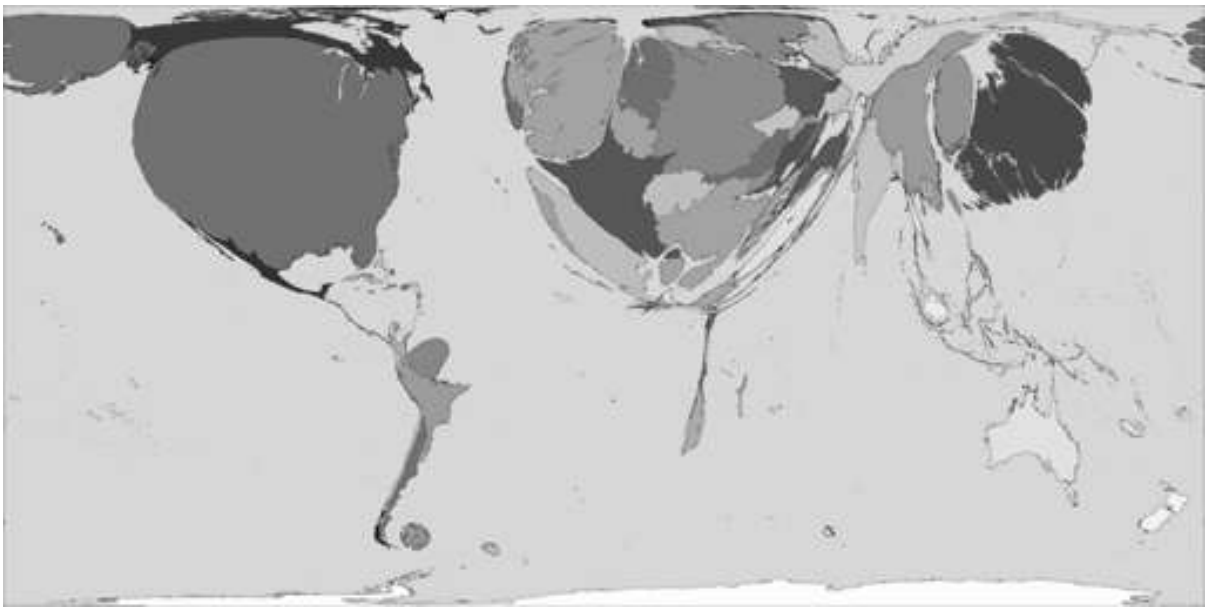


Figure 26. World map of the scientific divide

An important bottleneck is the lack of direct peering with other research and higher education networks. This bottleneck can be removed only by creating dedicated National Research and Education Networks (NRENs) connecting research and tertiary education institutions in each African country to a Regional Research and Education Network (RREN) interconnected to the peer infrastructures on other continents. In this context, a pioneering and very important role has been played by the Ubuntunet Alliance. Incorporated in 2006, Ubuntunet gathers the following 12 NRENs in Eastern and Southern Africa: Eb@le (Democratic Republic of Congo), EthERNet (Ethiopia), KENET (Kenya), MAREN (Malawi), MoRENet (Mozambique), RwEdNet (Rwanda), SomaliREN (Somalia), SUIN (Sudan), TENET (South Africa), TERNET (Tanzania), RENU (Uganda), and ZAMREN (Zambia) and it is fostering the creation of new ones in Botswana, Burundi, Lesotho, Namibia, Mauritius, Swaziland, and Zimbabwe.

The mission of the Alliance is to secure affordable high speed international connectivity and efficient ICT access and usage for African NRENs. In this respect, Ubuntunet has been one of the stakeholders of the FEAST project [126] (Feasibility Study for African – European Research and Education Network Interconnection) that, between December 2008 and December 2009, has studied the feasibility of connecting African NRENs to the GÉANT network and has documented the relevant issues in the region inhibiting these enabling technologies. In its final study (FEAST final report, 2010), FEAST has identified the opportunities available in Sub-Saharan Africa in terms of new intercontinental submarine cables with abundant capacity (see Figure 27) and emerging regional and national terrestrial fibre optic backbones.

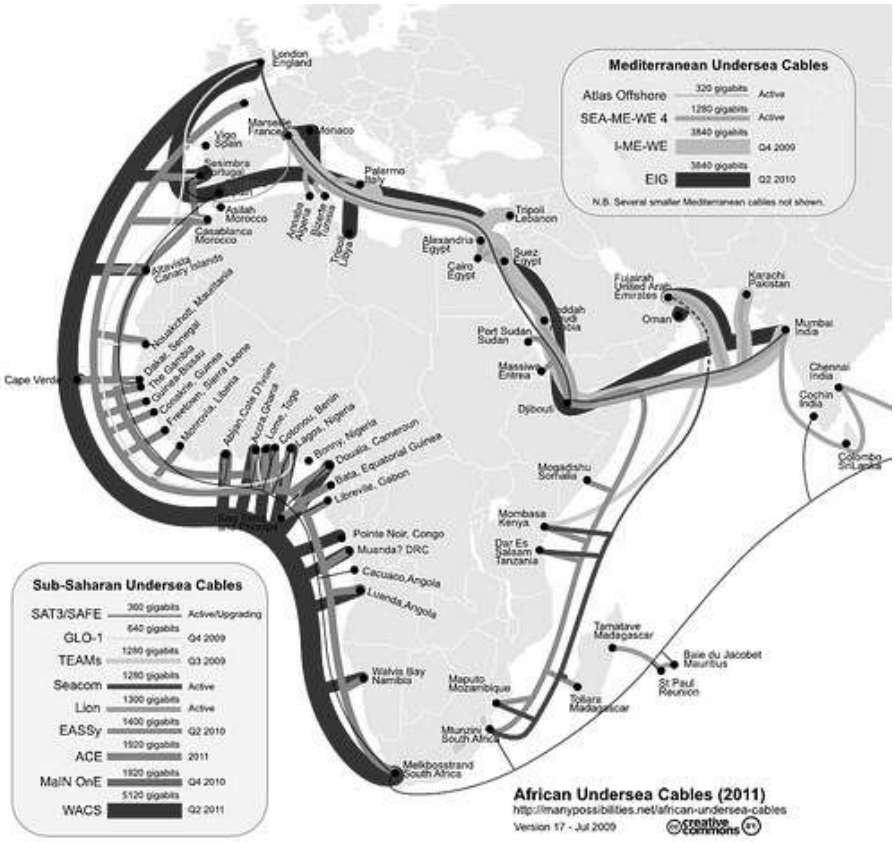


Figure 27. Map of the submarine cables currently available around Africa

FEAST has also paved the way for the creation of the AfricaConnect project[127] that will take care, under the coordination of DANTE, of the creation, in the next 4 years, of a RREN in Sub-Saharan Africa at a total cost of 15 M€, 80% funded by the European Commission and the rest co-funded by the beneficiary countries.

Grid

Notwithstanding the large dissemination activities of strategic projects, such as IST-Africa[128], EuroAfrica-ICT[129], and eI-Africa[130], co-funded by the European Commission in the context of its Sixth and Seventh Framework Programs, the Sub-Saharan region of Africa has seen the least amount of activity in distributed computing initiatives. However, the recent advent of affordable international bandwidth, the reform of national telcoms policies and the subsequent construction of high-bandwidth national research networks in the early part of the first decade of the century has had a catalytic effect on interest in deploying e-Infrastructures in the region. These naturally have a scope well-beyond that of Grid computing projects for scientific research, but have been identified by researchers, higher-learning institutions, and governments in the region as enablers of collaboration and tools to reduce the effect of the digital divide discussed above.

As in other cases discussed in this chapter, scientific projects requiring significant infrastructure – in particular the Southern African Large Telescope (SALT, 2010) and the Karoo Array Telescope[131] – were great stimuli of the interest in deploying networks and Grids in the region. The remote location of the scientific equipment and the wide geographic separation of the members of the collaborations using it were prime motivators, for example, for the development of the South African NREN. Data sharing considerations were long a concern, too, for the South African participation to two experiments of the Large Hadron Collider.

Two groups of research centres participate to the ALICE and ATLAS experiment, respectively, and the hub of medical and fundamental nuclear physics research undertaken at the iThemba Laboratories was one of the original drivers for experimenting with a national data and compute grid.

South Africa is the only country in the Sub-Saharan region with a dedicated activity to coordinate distributed computing, which started with two projects centrally funded by Department of Science and Technology. These were the national research and education network (SANREN) and the Centre for High-Performance Computing[132], which was inaugurated in 2006. The plan for a high-speed network connecting the country's universities and national laboratories generated interest in the creation of a federated distributed computing infrastructure based on the grid paradigm. The creation of a Joint Research Unit in mid-2008 was the start of this project, which aimed to integrate existing computing clusters and storage distributed in the institutes into a national grid computing platform.

The South African National Grid[102] by the start of 2010 consisted of a federation of seven institutes taking part in Grid operations and belonging to the SAGrid JRU, with open activities under way for further inclusion of other universities in the country:

- Meraka Institute (Cyberinfrastructure Programme, Pretoria);
- University of Cape Town , including the UCT-CERN Research Centre;
- University of the Free State (Bloemfontein);
- University of Pretoria (Pretoria);
- North-West University (Potchefstroom);
- University of Johannesburg (Johannesburg);
- University of the Witwatersrand (Johannesburg);
- iThemba Laboratory for the Accelerator-Based Sciences (Faure);

The development of the national Grid was based in many ways on the experience acquired in Europe, starting with the model of EGEE-III and regional activities. The gLite[133] middleware stack was adopted as standard at all sites, ensuring that the infrastructure would be easily used by Virtual Organisations operating on the EGEE resources. Integration into operational tools such as the Global Grid User Support (GGUS), Grid Operations Database (GOCDB) and monitoring tools such as the Real Time Monitor (RTM) and GSTAT ensure that the operations in South Africa are performed in a compatible manner to that of the other international infrastructures.

Grid computing services and identity management are most often secured and managed with X.509-standard digital certificates issued from a trusted Certificate Authority. A major obstacle in the Sub-Saharan region was the lack of a CA accredited by IGTF. Since there is indeed no region of the IGTF responsible for Sub-Saharan Africa, the nearest Policy Management Authority (PMA) is that responsible for Europe and the Near East: EUGridPMA. A proposal to accredit a new CA for South Africa, the SAGrid CA, was accepted by EUGridPMA in 2009 and full accreditation is expected by the end of 2011. To avoid delays, the INFN CA has assigned Registration Authorities in several South African institutes which are able to issue digital certificates for individuals and services locally.

The Grid infrastructure in South Africa makes of course use of the high-bandwidth SANReN network and aims to integrate the distributed computing resources attached to it providing their users with a powerful platform for collaboration and scientific research. This platform, due to its interoperability and operation as a single unit, can be considered as an extension of international infrastructures elsewhere, and access and usage of it is to a large degree location-independent. Coordinated training and development events both in South Africa and the broader region, undertaken in collaboration with the GILDA t-Infrastructure[134] have expanded the base of competent site administrators and users, in concert with similar activities undertaken by the EUMEDGRID-Support project (see above). This foundation work is essential in developing the base of applications, technical experts and eventually (and most importantly) users in the region.

6.3 Status of the Cultural Community in each partner country and relations with e-Infrastructure providers

6.3.1 Egypt

Research and Education Networks and Grids such as the Egyptian Universities Network (EUN) are fundamental infrastructures that will allow non-EU researchers to carry out high quality work in their home laboratories without the need to migrate in more advanced countries. However, the focus of these research and education networks has emphasized the development of strategic domains in the natural sciences, and/or socioeconomic fields.

Unfortunately, at the moment, there are no official collaborations between the cultural communities and e-infrastructure providers in Egypt, concerned specifically with cultural preservation or documentation. Any collaboration that have taken place, have done so in an unofficial and/or spontaneous capacity, with no direct goal of cultural documentation.

6.3.2 France

Two national priorities: Digitisation and Research. Digitization of cultural heritage has been defined as a priority by the French Ministry of Culture and Communication. Digitisation of heritage collections and contemporary art aims at making them accessible to everyone on the Internet and to multiply the uses for Research, for education, for cultural tourism, for regional development, and more broadly for the sake of all public. It is a means of promoting cultural diversity, cultural democratisation and promoting cultural transmission of knowledge.

The national digitization plan is a pillar of the French strategy to contribute to the European digital library Europeana, the other two axes being the mass digitization of books held by the Bibliothèque nationale de France (BNF), and the Audiovisual Archives of Institut National de l'Audiovisuel (INA).

The major objectives set by the digitization plan French are the same ones that have been defined at European level: providing content in all domains (archives, libraries, monuments, art, archeology ...), offer a structured developing of networks of institutions and collections, achieve critical mass, promote free and open access, promote the uses for all.

The national digitization plan is implemented through an annual call for projects. This call is open to public cultural actors (national or territorial) as well as non-profit private organizations. The plan encourages the development of new cross-collaborations between institutions (archives, libraries, museums, heritage ...) and is organised along six "thematic programs" to structure the projects of digitisation and thereby achieve coherent and critical mass.

All heritage collections that are digitized in France are recorded in the national inventory of digitised collections Patrimoine numérique[135], French contribution the MICHAEL portal.[136]

The thematic approach has allowed the creation of large corpus of coherent content and encouraged the creation of cultural portals, especially around contemporary art, reinforcing and networking digital culture. The ministry provides recommendations and standard specifications for each type of document / format[137]. All of these recommendations are in coherence with the national General Interoperability Reference Framework (RGI)[138].

Two other priorities are on the agenda of the Ministry of la Culture et de la Communication are the **development of Cultural research and of digital innovation as well as the integration in the European framework.**

MCC is also in charge of scientific research in the fields of culture and heritage. The same department (DREST) of the secretariat general is in charge of developing this research, and of the co-ordination of Cultural higher education (arts, architecture, music and dance…) and of the implementation of the national digitisation plan and the support to cultural technologies. MCC has for this developed a wide range of partnerships with research institutions, like the national Centre for Scientific Research (CNRS) or Universities.

In 2010, the Ministry of Culture and Communication has embarked on a new initiative to support digital innovation and to develop new digital cultural services for the general public based on digitised collections : 3D, augmented reality, mobile devices, collaborative tools, social ... 60 projects in all cultural sectors have been identified and are available on the platform "Culture Labs[139]".

In this area, MCC is also strongly taking part in the new Joint Programming Initiative (JPI) "Cultural heritage".

On all these matters, there is a strong cooperation between the ministry of Culture and Communication and the ministry of Higher education and Research (MESR). MCC is part of the steering committee put in place by the MESR in order to update the French national roadmap for Research infrastructures (see below). The Michael French instance integrates two dimensions with two interoperable databases: Patrimoine numérique (Culture) and NUMES (research and higher education). A general cooperation agreement has been signed for this by the two ministries.

A new impetus to digitisation and research through the "investing for the future plan". In the framework of the national plan "investing for the future", 750 M€ have been allowed to digitisation of cultural heritage, in order to foster collaboration between public and private sector for projects of new services in the field. Projects such as digitisation of out of print books (BnF and publishers) or digitisation of heritage movies (Centre national de la Cinématogrphie et de l'image animée - CnC and rights-holder) are underway. Those projects have a business model where the State supports the investments and the final user will pay for reimbursement.

A part of the budget is also allocated to the support of Research and Development activities in the field of digitization, on-line access or preservation of digitized heritage.

6.3.2.1 ON-LINE ACCESS: the national aggregator for digitised cultural heritage "Collections" on Culture.fr

The promotion of open technologies and standards has allowed the development of the national aggregator "Collections" on www.culture.fr.

Collections is the "single access point" to archives, libraries and museums resources is a flagship service, available on htpp://www.culture.fr in order to be easily identified by the general public. It is integrated in the public IT strategy to develop access to culture on the internet and broaden the audience of cultural heritage. Collections offers simple access to resources developed by archives, libraries, museums and other heritage services (archaeology, monuments) in France. More than 4 million records and 3 million digitised objects are available.

"Collections" is the French national aggregator contributing to Europeana.

6.3.2.2 Digital preservation

6.3.2.2.1 A national reference framework

In the framework of the national "digitisation" steering committee, a specific effort has been undertaken since early 2007 to co-ordinate the actions of the ministry of Culture and Communication concerning the long term preservation of heritage. It aims to establish a referential that transversally brings together the existing recommendations of the different sectors and actors. This national reference guide[140] takes the form of a reference web site, easily updated, and accessible on the "digitisation" site of the Ministry of Culture and Communication.

6.3.2.2.2 Digital preservation in the French law – digital legal deposit

The Fench law "Copyright and neighbouring rights in the Information Society" (« Droit d'Auteur et Droits Voisins dans la Société de l'Information » loi n° 2006-961 du 1er août 2006 dite DADVSI) transposes into French law the European directive 2001/29/CE on "the harmonisation of certain aspects of copyright law and related rights" and amends the code of intellectual property (CPI).

In matters of long term preservation of digital heritage, the law now permits: "the reproduction of a work, executed to the ends of preservation destined to preserve the conditions of its onsite consultation by libraries accessible to the public, by museums or archive services, on reserve that these seek no economic or commercial advantage."

The same DADVSI law introduced in the Code of heritage an exception for reproduction rights favouring organisations in charge of legal deposit. In virtue of this text the author cannot forbid the depositing organisations from:

1°Consultating the work on premise by duly accredited researchers on individual consultation strictly reserved for research ends,

It is the same for neighbouring rights holders:

« The artist, singer songwriter, producer of phonograms or videograms the audiovisual communication company cannot forbid the reproduction and the communication to the public of documents cited in article L.131-2 under the conditions foreseen in article L 132-4

The text also states that the producer of a data base cannot forbid the extraction or reuse through making available the whole or part of the data under the same conditions.

The dispositions of the heritage code concerning the legal deposit (article L131-2) include digital documents: “printed, graphic, photographic, sound, audiovisual, multimedia documents, irrespective of the production technique, publishing and distribution, will be subject to a compulsory legal deposit, from the moment they are put at the disposal of a public via material media, irrespective of the nature of the media.

At the BnF it is the Audiovisual department that is in charge of receiving the legal deposit of all multimedia documents or data-bases distributed on a media. INA is in charge of legal deposit for radio and television, the CNC (Centre National de la Cinematographie) for film. The decree 2006-696 of July 13th, 2006 makes provision for the BnF to be able to “demand deposit of a digital file, in substitution of the deposit of a printed, graphic or photographic document”.

At last, the title IV of the DADVSI law cited above (law n° 2006-961 of August 1st 2006) opened the legal framework for archival of internet sites of the French domain “Are also submitted to legal deposit the signs, signals, writing, sounds or messages of any nature that are the object of a public communication in electronic form.”

This disposition is also cited in the heritage code (article L.131-2). The law foresees the “collection through automatic means” (Internet site harvesting), but also covers the possibility to determine other means in agreement with website publishers. The organisations in charge of this legal deposit are the BnF and the INA. The law foresees that a code or access restriction cannot hinder the collection by the organisation in charge of the legal deposit.

6.3.2.3 Infrastructures for Digital preservation in the Cultural sector

Each institution is in charge of its own policy in accordance to national reference framework mentioned earlier. The national libary (BnF), the national audio-visual archive (Ina), the national archives service (SIAF) have developed specific programs (BnF and SIAF detailed below).

6.3.2.3.1 The "Distributed preservation and archiving system" (SPAR) of the BnF

The recent acceleration and growth of digital collections (over 100 teraoctets per year today) and the diversity of their formats, confronts the BnF with the challenge of their conservation: it is essential to securely and lastingly store the digital objects on a solid and effective foundation.

After one year of study the BnF launched the SPAR project[141], a full-fledged digital warehouse. Its conception is based on international norms that are recognised in the world of preservation of digital information (OAIS model)

SPAR is more than a simple storage space for secure data. It makes multiple copies of digital objects that allow one to anticipate further copies before definitive loss. But it also allows, through the precise and complete recognition of input data formats, to guarantee the continuity of access by making necessary transformations in case of technological obsolescence of computer restitution tools. Providing this guarantee implies a constant process of tech-watch on the formats, prototyping and tool testing. The overall arsenal is inherently foreseen in the conception of SPAR. Furthermore, SPAR allows to step back at any given time to restore objects to their original formats.

SPAR is a system put at the service of the community. It must guarantee that restored documents have not been altered. To do this, SPAR tags each stored object with a digital marker. In order to guarantee the rights of "communicability" of the used digital objects, SPAR relies on expert system of rights management that allows it to calculate the user licences of the digital objects and to apply necessary restrictions according to the user profile (researcher at the BnF, web user…)

SPAR does not have the vocation to only by an internal tool at the BnF. With a will to mutually share expertise and costs, the BnF will open its system to other partners and institutions, thus providing the service of "third party archivist" of digital heritage.

6.3.2.3.2 The development of electronic archiving : the pilot project of the PILAE platform steered by the directorate of French Archives (SIAF)

The SIAF, with support of the General directorate of state modernisation (DGME) is presently producing a pilot platform that will function during a transitory period from 2008-20012 for the National archives in order to receive, preserve and communicate the born digital archives produced by the central services of the state (material from data bases, bureaucratic documents or produced from electronic documents management systems, electronic message services, immaterial data-flow such as public tenders…). The pilot implements the standard for archival data exchange; allows one to receive the transferred archives in this format and control a part automatically according to the conditions set down by the agreements established between the different actors of the archive process, partly in manual form by the archivists of the national archives; to automatically convert input data to lasting formats, documents that are not in the archival target format; to securely conserve the data by replication on two distant sites and, finally, to transmit them via download. In the first stage the pilot will be operational for the producers of the archived documents in the ministries, for the archivists in the different ministries and the archivists at the national archives.

The pilot could also be reused as a model for other administrations that wish to implement an archiving platform in their own environment.

6.3.2.4 Participation of Cultural institutions to national research infrastructures

Cultural institutions are interested in two national infrastructures: RENATER and TGE ADONIS. RENATER and TGE ADONIS are part of the French Research Infrastructures roadmap. RENATER is a generic service infrastructure. TGE ADONIS is dedicated to the Social Sciences and Humanities.

6.3.2.4.1 RENATER

The National Telecommunications Network for Technology, Education and Research **RENATER**[142] (Réseau National de Télécommunications pour la Technologie, l'Enseignement et la Recherche) is the French NREN It provides network services (access capacities start from 10 Mb/s and go up to 10 Gb/s) to all research institutions in France.

RENATER was deployed in the early 1990s to combine telecommunications infrastructures for research and education. The GIP RENATER (Public Interest Group) was created in 1993 to act as the project manager for this network. The infrastructure of the RENATER network consists of national links connecting points of presence in French regions and overseas territories together with the international links, and the exchange nodes between Internet service providers, known as SFINX (Service for French Internet eXchange) in Paris and REUNIX in Reunion Island.

All national cultural institutions having research activities (the French national library BnF, the Centre Pompidou, the Ircam, the Louvre, the arts schools and the architecture schools etc) are connected to RENATER, mainly for broadband connectivity, not for storage.

6.3.2.4.2 TGE ADONIS

The **TGE ADONIS** (Integrated Access to Digital Data and Documents in the Humanities and Social Sciences) is a Major Facility launched by CNRS to connect Humanities and Social Sciences. It proposes a range of services to those research communities. Cultural institutions are involved mainly through their co-operation with the CNRS for research programmes.

The contribution of ADONIS to the National Research Strategy has several aspects. A major concern is the need to archive and to promote integrated access to digital data and documents in the humanities and Social Sciences; this is regarded as of national strategic importance, given the high level of French investment in this domain. Such activities are also a major component of the European Research Infrastructure "roadmap", and the TGE thus represents French participation in that strategy at the international as well as national levels.

Co-ordination of a coherent archiving and access service for research datasets in the Humanities and Social Sciences is something which cannot easily be done at the level of individual institutions or laboratories. The TGE thus has a major role in promoting the sharing of expertise amongst laboratories, in encouraging the definition and adoption of best practice in such areas, and in co-ordinating the provision of relevant technical services. It should be emphasized that ADONIS itself is a facilitator and not a funder of projects

The four main activities of Adonis are:

- Co-ordination of a network of national Digital Data Centres (centres de ressources numériques). Each of these Centres provides advice and support to a particular research community within the SHS; each of them is attached to an existing laboratory with specialist knowledge.
- Promotion of interoperability amongst resources by means of Guides to Good Practice, training, and other facilities.
- Provision of the ADONIS Grid: a national service infrastructure providing massive storage, networking, and long-term archival facilities. This set of services is provided in partnership with the Computing Centre of the National Institute of Nuclear and Particle Physics (CC-IN2P3) and with the National Computing Centre for Higher Education (CINES). The Adonis Grid offers research teams the following possibilities:
    - web hosting for scholarly and documentary research projects, using a wide variety of solutions based on open technologies;
    - secure offsite storage for data produced by scholarly research projects;
    - high performance mass-computing capabilities using the facilities of the CC-IN2P3, for example for data modeling and 3D imaging;
    - long-term archival storage for research data using the international model of the Open Archival Information System (OAIS),in collaboration with the CINES.

  A range of utilities to improve interoperability and harvesting of stored data is also planned, including tools for format conversion, publication and verification of metadata, etc.
- Development and deployment of the ISIDORE[143] platform (Integration of Data Interconnection Services on Research and Education): a state of the art intelligent query engine, using advanced techniques to permit semantic searching of the "deep web".

ISIDORE is a new software system designed to enhance searching for information in the human and social sciences. It is the outcome of a major effort on the part of the TGE Adonis and its community. ISIDORE provides a single point of access to the full range of digital outputs from the human and social sciences, taking full advantage of the specialist knowledge of the research professionals and technical experts who classify, select, digitize, organize, structure and archive digital data.

Version 1.0 of ISIDORE, released on 4 April 2011, facilitates navigation across more than a million documents from 860 sources, organized into 42 collections: journals, works, bibliographies, scientific datasets, blogs, scholarly archives, research news, university and municipal library catalogues... the number of sources is expected to increase regularly.

ISIDORE brings together the following components:

- a collector of information, harvesting data according to international standards and protocols;
- a workflow in which the information is enriched and enhanced by means of such scholarly resources as authority files;
- a search engine which makes use of this enriched information;
- a public web interface at the site www.rechercheisidore.fr;
- a public repository which can be accessed independently, using data linking methods.

As of today, ISIDORE is the most significant scholarly "open data" project in France to make use of linked data technologies in this innovative way.

ISIDORE is a way of promoting the use of digitized cultural content for Research purposes. The participation of Cultural institutions is illustrated for example by the integration of the national digital library GALLICA in the project.

6.3.3 Greece

The major part of cultural heritage digitisation in Greece, involving more than 200 cultural institutions, was based on structural funds of about 120 MEuro in the period 2007-2009. New funding is expected, starting in 2011, based on structural funds - through the Digital Convergence Programme (running until 2013), in collaboration with the Hellenic Ministry of Culture and Tourism (HMCT) and the Hellenic Ministry of Education, Lifelong Learning and Religious Affairs (HMELLR).

The Hellenic Ministry of Culture and Tourism is responsible for the digitisation and cultural developments related to the fields of museums, ancient monuments, contemporary arts, theatre, cinematography and audiovisual archives. 56 branches of the Ministry (Efories of Ancient Monuments) are possessing and have started digitising and annotating their content during the last few years. In parallel, HMCT is aggregating content descriptions - metadata from other public and private museums and archives, so as to make their digital content (further) accessible to the Greek and European citizens. Specific attention in the current and following years will be paid through the HMCT Directorate of Monuments and Archives to extending digitisation and annotation of objects to include large 3-D ancient monuments.

The Hellenic Ministry of Education, Lifelong Learning and Religious Affairs is responsible for the digitisation of content from libraries, especially public, and the state archives and for creation of educational content. HMELLR is currently setting up a Panhellenic Strategy for organising the fields of libraries and archives, including the aggregation, access and preservation issues.

Next digitisation activities and projects are expected to take place within the "Digital Convergence" Information Society Operational Programme and the newly developed and initiated National Strategic Reference Framework (ESPA) 2007-2013. Referring to content, all types of content will be digitised, including books, photographs and 2-Dobjects, paintings, videos, films, sound documents, music, 3-D (archaeological) objects and monuments.

In the current phase both Ministries, as well as Greek Cultural Organisations have been providing cultural content to Europeana, at a ratio of about 1% of the whole Europeana aggregated (through respective metadata) and offered content.

This content has been mainly and will continue to be provided through the following projects:

1) ATHENA (2008-2011) in the museum and archival sector. HMCT participates in Athena.
2) Europeana Local (2008-2011) in the library and small museum/archival sector.
3) EUScreen (2010-2013), following the Videoactive (2008-2010), in the audiovisual sector. The Hellenic Audiovisual Archive of HMCT participates in the projects.
4) European Film Gateway (2008-2011) in the film sector. The Greek Film Centre participates in EFG.
5) APENET (2009-2011) in the archival sector. General State Archives of HMELLR participate in Apenet.
6) CARARE (2010-2013) in the archaeological and architectural content sector. HMCT participates in Carrare.
7) Judaica (2010-2012) in the Jewish library/museum/archival sector. The Greek Jewish Museum participates in Judaica.
8) ECLAP (2010-2012) in the theatrical content sector.
9) Linked Heritage (2011-2013) in the museum and archival sector. HMCT participates in the Linked Heritage project.
10) Digital Contemporary Art (2011-2013) in the Art sector. The Hellenic National Gallery articipates in DCA.

The National Technical University of Athens has been providing technical assistance in most of the above projects, having created a Europeana Ingestion Tool that has been adapted and used in them.

HMCT is developing plans to play the role of national aggregator for content related to museums, monuments, arts, music, films, and audiovisual archives. In 2011 and further HMCT will devote much effort on the digitisation and annotation of archaeological content and monuments. In this framework, HMCT participates in the CARARE (archaeological content sector aggregator) project started in March 2010, which is aggregating archaeological objects for Europeana. HMCT has already started planning aggregation of such content from its branches, i.e., 56 Eforeies of Archaeological Content. There will be further investigation for aggregating archaeological content the field of Museums, Archaeological Institutes and Foreign Archaeological Schools.

The HMCT is also a partner in the DC-Net project, which aims at examining and relating the digitisation activities in European countries with their e-infrastructures and, particularly, with the research ones. In the framework of this activity, NTUA has been assisting HMCT in generating a national group, including representatives from HMCT and HMELLR, representatives of the NRENs, in particular of the GRNET and the SYZEFXIS network, providing results that will be useful for the advancement of technological developments in the cultural sector, in collaboration with e-infrastructures, for its sustainability and for digital preservation.

The Hellenic Ministry of Education, Lifelong Learning and Religious Affairs develops plans to play the role of National Aggegator for content related to libraries and archives as well as of educational content. In this framework, they are planning to offer access, through their web site, to some 18.00.000 pages digitised in their public library network, and to about 5.000.000 pages of the greek state archives.

The resulting target, as a whole, is to double the content that has been successfully digitised in the former Operational Period, while, at the same time, managing to preserve and make all already digitised content accessible to the general public.

6.3.4 Italy

GARR-G, the R&E network backbone currently in operation is based on high-bandwidth circuits up to 10Gbps, using various technologies such as DWDM, SDH and MPLS. The meshed topology, interconnecting

45 network Points of Presence (PoP), ensures a high level of resilience and reliability of the network; thanks also to its wide coverage of the country, it interconnects more than 400 user organizations all over the national territory.

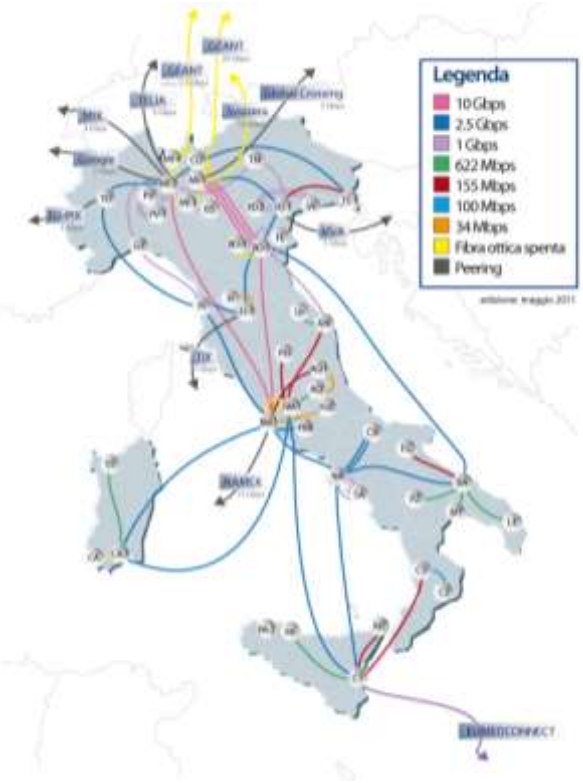


Figure 28. GARR-G Network Topology

GARR-G is interconnected with the rest of Internet through General Internet connections, Research connections and National peerings. GARR-G network ensures its community the interconnection with General Internet using three of the main international commercial upstream providers (Global Crossing, Level3 and TELIA). Interconnections are located in three different PoPs for resiliency. The total aggregated capacity of the links is 7.5 Gbps. The GARR-G network is part of the worldwide system of Research and Education Networks (NRENs). It connects to other NRENs in Europe and worldwide through a 10 Gbps link (plus 2.5 Gbps backup link) to the GÉANT pan-European backbone. Other interconnections, such as LHCOPN, DEISA, EXPRes, FEDERICA, have been set up within specific projects and to provide connectivity to specific groups of users.

The GARR-G network is interconnected with all major Italian commercial ISPs (Telecom Italia, Wind, Fastweb, etc.) at the main public Internet Exchange Points in the country, namely MIX (in Milan) [R 17], NAMEX (in Rome), TOP-IX (in Turin), TIX (in Florence) with an aggregated capacity of about 20 Gbps. Furthermore, GARR-G peers with Local Public Administration Network Infrastructures. Currently, peering agreements are operational with the Regional Area Networks of Basilicata, Emilia Romagna, Marche, Toscana and Piemonte. Direct peerings have been also established with the networks of Presidency of the Council of Ministers and Vatican. GARR recently activated a direct connection with the Central Administration National Network (SPC, Public Connectivity System), for those users that need connectivity with specific applications provided in this environment.

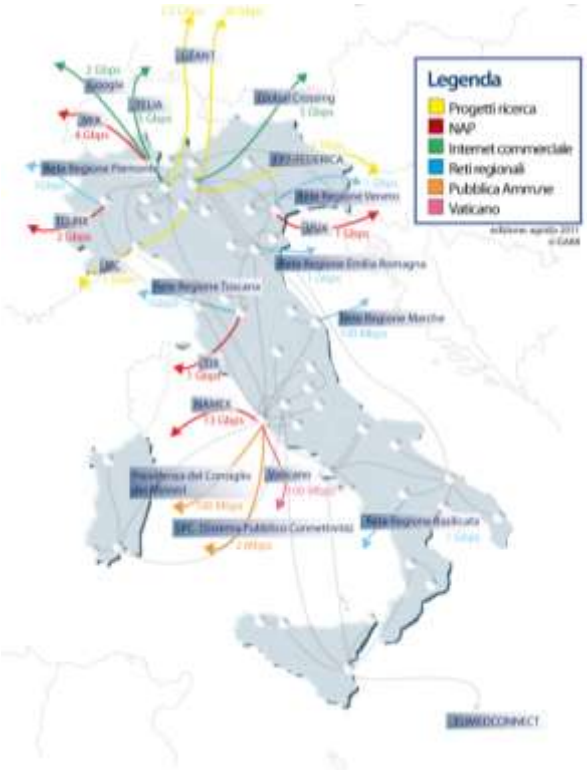


Figure 29. GARR Peering topology

GARR supports many institutions belonging to the world of Cultural Heritage, in particular offering technical support in their research, preservation, promotion and exploitation activities. Many excellences of this sector are members of GARR user community. These cultural institutions work daily with the wider and multi-disciplinary University Research community. They had established collaborations at a local level, but also at national and international level. Among the institutions that actively use the GARR network services through its broadband transmission, there are Libraries, Memory Institutions, Museums, and Central Departments of the Italian Minister of Culture.

The main e-infrastructure application areas for this community are:

- Sharing and transmission of large amounts of data between institutions distributed throughout the country that collect, store and process the digital objects.
- Use of web-applications, allowing individual institutions to promote more effectively its activities on the preservation and promotion of Cultural Heritage.

The GARR services portfolio includes a number of technical activities. GARR offers support for the configuration and management of network devices, the fault management and information security incidents, but also for the domain names registration (.it and .eu) and IPv4 and IPv6 address allocation.

GARR provides to the University and Research community advanced services that can meet the new demands of study and development, especially in terms of high value-added services such as authentication and authorization, mobility support, security and data protection, collaboration, information sharing and multimedia applications (such as multi-conferencing and VoIP, but also webTV and radio). GARR also

provides support for research projects with dedicated networks both within its own network in the international collaborations and participates directly in European research projects.

The GARR network and its services are dedicated to the Italian Research, Academic and Education communities. Currently the GARR network connects approximately 450 end sites, including research centers, laboratories and other facilities, universities, Institutes for Research in Health Care (IRCCS), Music Conservatories and Academies of Art (AFAM), Libraries, Archives, Schools, Museums and other R&E institutions of national relevance, for overall 2.000.000 end users:

- 129 Laboratories and Research centers of major Scientific organizations in Italy, including CNR, ENEA, INGV, INFN and ASI,
- 88 Universities,
- 31 Music Conservatories and Academies of Art (AFAM),
- 38 Institutes for Research in Health Care (IRCCS),
- 26 National and University libraries ,
- 17 Astronomical and astrophysical observatories,
- 74 other Research and Education facilities of national and international interest,
- 77 Schools and Hospital Special Schools for children who are patients (HSH), funded by a specific action of the Ministry of Education until 2009.

The following tables show a list of sites linked to GARR network that are of interest for the Cultural Heritage community. In addition, all University departments connected to the study, research and preservation of Cultural Heritage are also connected to the network.

N.	Name	Key CH-related activities
1	CNR IBAM - Istituto per i Beni Archeologici e Monumentali	Restoration and preservation of CH
2	CNR ICCOM - Istituto di Chimica dei Composti Organometallici e Laboratorio di Spettroscopia Laser.	Imaging tecniques applied to CH
3	CNR ICEVO - Istituto di studi sulle Civiltà dell'Egeo e del Vicino Oriente	Cartography and digitalization of photo and maps
4	CNR ICIS - Istituto di Chimica Inorganica e delle Superfici	Restoration and preservation of CH
5	CNR ICRM - Istituto di Chimica del Riconoscimento Molecolare	Restoration and preservation of CH
6	CNR ICVBC - Istituto per la Conservazione e la Valorizzazione dei Beni Culturali	Presevation, promotion and exploitation of CH. Part of the CHARISMA FP7 project
7	CNR IDASC - Istituto di Acustica e Sensoristica "O. M. Corbino"	Monitoring and preservation of CH
8	CNR IFN - Istituto di Fotonica e Nanotecnologie	Monitoring, preservation and restoration of CH

N.	Name	Key CH-related activities
9	CNR IIT - Istituto di Informatica e Telematica	ICT applied to CH. Developed the "geomemories" platform (www.geomemories.org)
10	CNR IMM - Istituto di Microelettronica e Microsistemi	Monitoring and preservation of CH
11	CNR ISAC - Istituto di Scienze dell'Atmosfera e del Clima	Monitoring and preservation of CH. Several projects ongoing: TEACH, MESSIB, HI KNOW e Climate for Culture
12	CNR ISCIMA - Istituto di Studi sulle Civiltà Italiche e del Mediterraneo Antico	Study of ancient Mediterranean, ICT applied to CH. Project "Archeologia e Calcolatori"
13	CNR ISMN - Istituto per lo Studio dei Materiali Nanostrutturati	Monitoring and preservation of CH
14	CNR ISTEC - Istituto di Scienza e Tecnologia dei Materiali Ceramici	Monitoring, preservation and restoration of CH
15	CNR ISTI - Istituto Scienza e Tecnologie dell'Informazione	Visual Computing and Networking Multimedia Information System applied to CH; and ICT applied to preservation and monitoring.
16	CNR ISTM - Istituto di Scienze e Tecnologie Molecolari	Monitoring, preservation and restoration of CH
17	CNR ITABC - Istituto per le Tecnologie Applicate ai Beni Culturali	Study of CH, virtual reality reconstructions, etc
18	ENEA	8 Research centers ENEA are involved in: - CH research and preservation activities - Dissemination, exploitation and sharing of CH materials - Promotion and dissemination of knowledge, e-learning etc
19	Sincrotone Trieste	CH research, monitoring and preservation activities, anti-forgery technologies

Table 1: Large Italian Research institutes involved in CH research, preservation and restoration and key activities carried out

N.	Research institutes related to Ministries (other than Cultural Ministry)
1	Accademia dei Lincei – Roma
2	Accademia della Crusca – Firenze
3	Biblioteca Hertziana - Roma
4	CINSA – Consorzio Interuniversitario Nazionale per le Scienze Ambientali di Venezia
5	CORILA – Consorzio per la Gest. Del Centro di Coord. Delle Attività di Ricerca Inerenti il Sistema Lagunare di Venezia
6	CRS4 - Cagliari
7	CRUI – Conferenza dei Rettori delle Università Italiane – Roma
8	Fondazione Ettore Majorana – Erice
9	Soprintendenza Speciale Polo Museale Fiorentino
10	INFM (Ist. Naz. Per la Fisica della Materia) - Laboratorio nazionale TASC – Trieste
11	ISS – Istituto Superiore di Sanità - Roma
12	ISTAT – Istituto Nazionale di Statistica – Roma
13	IUE – Istituto Universitario Europeo – Firenze
14	IVSLA – Istituto Veneto di Scienze, Lettere ed Arti - Venezia

Table 2: Research institutes related to Ministries other than Cultural Ministry

N.	International institutes relevant to CH in Italy
1	Ecole Française – Roma
2	ESA-ESRIN Roma
3	European Commission - Joint Research Centre - Ispra
4	Holy See – Vaticano
5	KHI – Kunsthisotrisches Institut in Florenz
6	New York University in Firenze

Table 3: International institutes in Italy

N.	Central, National and Public Libraries
1	Biblioteca Angelica - Roma
2	Biblioteca Casanatense - Roma
3	Biblioteca Estense e Universitaria - Modena
4	Biblioteca Marucelliana – Firenze
5	Biblioteca Medica Statale - Roma
6	Biblioteca Medicea Laurenziana - Firenze
7	Biblioteca Nazionale - Bari
8	Biblioteca Nazionale Braidense - Milano
9	Biblioteca Nazionale Centrale di Firenze
10	Biblioteca Nazionale Centrale di Roma
11	Biblioteca Nazionale Marciana di Venezia
12	Biblioteca Palatina - Parma
13	Biblioteca Provinciale - Bari
14	Biblioteca Riccardiana - Firenze
15	Biblioteca Statale Antonio Baldini - Roma
16	Biblioteca Universitaria - Bologna
17	Biblioteca Universitaria - Genova
18	Biblioteca Universitaria - Napoli
19	Biblioteca Universitaria - Padova
20	Biblioteca Universitaria - Pavia
21	Biblioteca Universitaria - Pisa
22	Biblioteca Universitaria Alessandrina - Roma
23	ICCU – Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche
24	Istituto Centrale per i Beni Sonori e Audiovisivi

Table 4: Central, National and Public Libraries

N.	National Archives
1	Archivio di Stato di Catania
2	Archivio di Stato di Firenze
3	Archivio di Stato di Milano
4	Archivio di Stato di Napoli
5	Archivio di Stato di Palermo
6	Archivio di Stato di Roma
7	Archivio di Stato di Torino
8	Archivio di Stato di Venezia
9	Archivio Centrale dello Stato - Roma

Table 5: National Archives

6.3.5 Slovenia

ARNES[144] is the Slovenian NREN and provides services with strong e-Infrastructure requirements. ARNES provides a powerful resilient backbone network infrastructure, based on Gigabit and 10 Gigabit Ethernet technology. Network PoPs (Point-of-Presence), to which organizations eligible for ARNES services connect, are present in major Slovenian towns.

Figure 30. ARNES. Source: http://www.arnes.si

ARNES provides advanced web and network services to eligible organizations and web hosting, electronic mail and dialup Internet access to individuals:

- web, email, servers (web hosting (html, PHP/MySQL, LAMP), domain-name registration and redirection),
- network services (IP-level connectivity, dedicated point-to-point connections, Eduroam and other services provided by the ARNES network),
- multimedia services (audio, video and computer communication in real time using many different methods – web and H.323 videoconferencing, archiving of videoconference events),
- ARNES AAI (infrastructure enabling simple and secure use of web and network services within a particular community of users and service providers),
- security (protecting the ARNES network and local-area networks of organizations, advanced security services for individuals, and coordination and intervention in Internet incidents),
- registry services and
- services for individuals.

The institutions connected to ARNES should meet the criteria laid down by Slovenian government and could be classified into the following groups:

1) Research and development.

Research organizations, technological parks, developments parks:	40 institutions
Other research groups:	18 institutions
2) Education.	
Universities	24 institutions
High schools	160 institutions
Primary schools	521 institutions
Other education institutions	128 institutions
3) Cultural institutions.	
Libraries, museums, archives	179 institutions
Other cultural organizations	25 institutions
4) Government.	
Ministries	12 institutions
Other administration	7 institutions

Preservation institutions connected to ARNES are included among cultural institutions.

Museums:

- Narodni muzej
- Narodna galerija

Libraries:

- Narodna in univerzitetna knjižnica (NUK)
- Mariborska knjižnica
- Mestna knjižnica Ljubljana
- Koroška osrednja knjižnica
- Osrednja knjižnica Celje

Archives:

- Pokrajinski arhiv Koper
- Zgodovinski arhiv Ljubljana
- Pokrajinski arhiv Maribor

Other cultural institutions

- Slovensko narodno gledališče Drama
- Slovensko mladinsko gledališče Ljubljana
- Slovenska kinoteka
- Zgodovinski arhiv v Celju
- Lutkovno gledališče Ljubljana
- Slovenski gledališki muzej

6.3.5.1 Grid computing in Slovenia

ARNES took leading role to organize grid computing in Slovenia. ARNES represent Slovenian part in the European Grid Initiative and supports common providing European service for the European research area. Main services are located at ARNES and Jožef Štefan Institute. Information services on ARNES enables Nagios, odkrivanje prostih kapacitet, VOMS, job manager – glite-CREAM and Nordugrid ARC, beleženje v dnevniške datoteke, information system (BDII), Accounting – APEL) etc.

6.3.5.2 Preservation by means of e-Infrastructures. Slovenian projects and initiatives.

As part of its activities, ARNES is involved in the operation of several international organizations:

- TERENA (Trans European Research and Education Networking Association)
- CEENet (Central and Eastern European Networking Association)
- RIPE (Regional Internet Registry)
- EURid (The European Registry of Internet Domain Names)
- FIRST (Forum of Incident Response and Security Teams)
- Euro-IX (European Internet Exchange Association)

ARNES is a founding member of DANTE (Delivery of Advanced Network Technology to Europe) and a member of NREN PC (National Research and Education Programme Committee).

GÉANT. ARNES in involved in the pan-European data network dedicated to the research and education community. Together with Europe's national research networks, GÉANT connects 40 million users in over 8,000 institutions across 40 countries. GÉANT includes two strands of work. It provides support to the development of research and education networking in the less advanced countries in Europe. Annual reporting on European NRENs' network and service operations due to identify and understand trends between countries and developments over time - TERENA Compendium of national research and education networks.

EGI: European Grid Infrastructure. ARNES is one of the founders of the initiative is a foundation European Grid Infrastructure. The project aims to guarantee the long-term availability of a generic e-infrastructure for all European research communities and their international collaborators. Its mission is to enable access to computing resources for European researchers from all fields of science, from High Energy Physics to Humanities.

6.3.5.3 Repositories and aggregator services

GISKD[145]

Repository for the immovable heritage of Slovenia.

ATHENA[146]

Harvesting for museums than were contributing their content to Europeana.

NUK[147]

National and university library is a repository for libraries.

KeyToNature[148]

Slovenian project partner Natural museum of Slovenia prepares a repository of biodiversity for educational purposes.

Trubar[149]

National central repository for learning and education with integrated national catalogue and a repository for educational e-contents.

6.3.6 Spain

RedIRIS[150] is the Spanish NREN and integrates what could be called RRENs (Regional RENs) that are the final providers of services in specific regions with strong e-Infrastructure requirements.

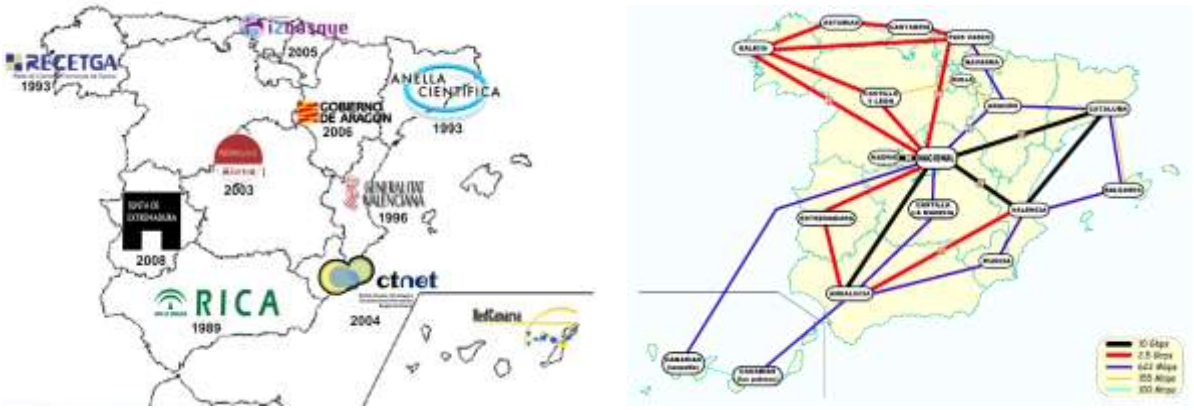


Figure 31. Red IRIS, Spain

The following graphic depictures the main services that the spanish NREN offer to the connected institutions: email, digital identity, mobility, storage, eScience services, security, connectivity and private networks.

Figure 32. Services offered by Red IRIS to the connected institutions

RedIRIS = 430 institutions + 33 centres connected

RedIRIS has a clear policy to accept or deny the inclusion of new institutions in their network and provision of services. The institutions connected to RedIRIS are classified into the following groups or categories:

A.1) Universities and centres for higher education. This group includes research units in university hospitals and attached centres.

A.2) Public research entities. Including the major science centres with independent legal capacity.

A.3) Other research institutions. Legal, non-profit institutions that participate in National R&D&i Plan projects or equivalent at regional or international level.

B.1) Entities that manage R&D&i programmes. Entities that receive public finance for research

B.2) Entities with relevant content. Non-profit institutions, with independent legal capacity, which have significant digital content that is relevant for the scientific-technical community and which place this content at the said community's disposal.

B.3) Entities that participate in R&D programmes. Entities that participate in National R&D&i Plan projects or equivalent at regional or international level, during and for the development of these projects.

B.4) Entities of special interest. Other entities of special interest for the Spanish science and technology system. In these cases, RedIRIS may provide the technical resources necessary so that the traffic exchanged with these institutions is limited to the goals established.

Specifically, preservation institutions connected to RedIRIS are included in the B.2 category as relevant content owners. The main institutions on that category are:

Libraries:

- Biblioteca Abadía de Montserrat (ABMONTSERRAT)[151]
- Biblioteca Nacional (BNE)[152]
- Biblioteca Pública Central de La Rioja (bcr)[153]
- Biblioteca de Castilla y León (BCL)[154]

- Biblioteca de Cataluña (BNC)[155]
- Consorci de Biblioteques Universitàries de Catalunya (CBUC)[156]
- Consorcio de Bibliotecas Universitarias de Galicia (CBUG)[157]
- Real Biblioteca (Patrimonio Nal.) (PATRIMONIONACIONAL)[158]

Museums:

- Museo Arqueológico Nacional (man)[159]
- Museu d´Història de la Medicina de Catalunya (MHMC)[160]
- Consorci de les Drassanes Reials i Museu Marítim de Barcelona (MMB)[161]
- Museu d´Història de la Medicina de Catalunya (MHMC)[162]

Other (Royal academies, institutes...):

- Real Academia Española (RAE)[163]
- Real Academia de Ciencias Exactas, Físicas y Naturales (RAC)[164]
- Real Academia de Ciencias y Artes de Barcelona (racab)[165]
- Institut Cartogràfic de Catalunya (ICC)[166]
- Institut d'Estudis Catalans (IEC)[167]
- Instituto Arqueológico Alemán (DAINST-MADRID)[168]
- Instituto Cervantes (CERVANTES)[169]

RedIRIS has 18 points of presence across Spain, which we have introduced as RRENs (Regional RENs). They are the final providers of services in specific regions. In Catalonia, for example, the Regional RENs is called **Anella Científica**, a flexible and reliable infrastructure that connects a wide range of institutions involved in the world of R&D and innovation. In addition to connectivity with different speeds and modes of connection, it offers many associated services.

On this context, supercomputing centers and communication networks offer a wide range of possibilities not yet enough exploited for the cultural sector.

6.3.6.1 List of Spanish Regional Research and Educational Networks

- RedIRIS170
- Red Informática Científica de Andalucía171
- Rede de Ciencia e Tecnoloxía de Galicia172
- Comunidad Autónoma de Madrid173
- Red Canaria Académica de Recursos de Información Avanzados (RedCanaria)174
- Red de Ciencia, Tecnología y Sociedad de la Información de la Región de Murcia (Red CTnet)175

- Red de Ciencia y Tecnología del País Vasco (Red i2BASQUE)176

6.3.6.2 Most relevant Supercomputing Centers in Spain

- Barcelona Supercomputing Center–Centro Nacional de Supercomputación (BSC–CNS)177
- Centro Extremeño de Investigación, Innovación Tecnológica y Supercomputación (CénitS)178
- Centro de Supercomputación y Visualización de Madrid (CeSViMa)179
- Centro de Supercomputación de Galicia (CESGA)180
- Centro de Supercomputación de Murcia (CESMU)181
- Centro Informático Científico de Andalucía (CICA)182
- Fundación Centro de Supercomputación de Castilla y León (FCSC)183

6.3.6.3 RES (Spanish Supercomputing Network)

The Spanish Supercomputing Network is a distributed supercomputing grid created in 2007 to give support to research groups in Spain. Barcelona Supercomputing Center is the coordinator of the RES, which includes a total of eight supercomputers distributed in Spain in different institutions/universities:

- Barcelona Supercomputing Center
- CeSViMa, Universidad Politécnica de Madrid, UPM
- Instituto Astrofísico de Canarias, IAC
- Universidad de Málaga
- Universidad de Santander
- Universidad de Valencia
- Universidad de Zaragoza
- ITC (Instituto Tecnológico de Canarias)

The total number of processors in the Spanish Supercomputing Network is 15170 PowerPC 970, with a Peak Performance of 138,5 TFlops.

6.3.6.4 Preservation by means of e-Infrastructures. Spanish projects and initiatives.

6.3.6.4.1 Web archiving initiatives. The Padicat project

Preserving Internet content for future generations is one of the new challenges on long-preservation issues. National and initiatives are growing worldwide[184] (ex. Internet Archive[185]), as well as the need of global

exchange and international relations have been detected and structured through the international internet preservation consortium (netpreserve.org[186]).

In order to acquire, preserve and make accessible knowledge and information from the Internet for future generations, the Biblioteca de Catalunya lead the first project on web archiving in Spain: PADICAT. During the last years, other organisations like the Spanish National Library (BNE) had also created web archiving projects. Another relevant project on web archive is ONDARENET[187] (Basque digital heritage), responsible of preservation and access to the Basque web.

PADICAT[188] is the Web Archive of Catalonia, a repository for the heritage born digital: the web pages. The aim of PADICAT is to preserve and give access to old versions of web pages published on the Internet, since 2005.

Project created by the Biblioteca de Catalunya[155], the public institution responsible for compiling, processing and distributing the bibliographic heritage of Catalonia, and by extension the digital heritage, with the technological collaboration of the Centre de Serveis Científics i Acadèmics de Catalunya[189] (Center for Scientific and Academic Services of Catalonia, CESCA).

6.3.6.4.2 Repositories and aggregation services. Hispana

Access to **3,079,432** digital objects of **155** repositories.

Hispana[190] brings together the digital collections of archives, libraries and museums according to the Open Archives Initiative being promoted by the European Union. With respect to Spanish digital repositories, it performs functions that are analogous to those of Europeana regarding European repositories, acting as a content aggregator of digital collection databases.

Highlights of these collections include the institutional repositories of Spanish universities and the digital libraries of Autonomous Communities, which offer access to growing sets of all types of materials (manuscripts, printed books, photographs, maps, etc.) of the Spanish bibliographic heritage.

Through Hispana, the digital libraries of the Autonomous Communities and other local digital libraries contribute their content to the EuropeanaLocal project, in which the Ministry of Culture participates as a national coordinator, together with 32 other institutions from 26 countries.

Hispana also incorporates the content of CER.es[191], the collective catalogue of the Digital Network of Spanish Museum Collections.

This network brings together museums that share a unified documentation and management system for their collections (DOMUS), which has been developed by the Ministry of Culture and is shared by museums and collections throughout Spain. CER.ES offers unified access to the collections of 61 of these museums, which belong to different administrations and specialities, in addition to 100,000 pieces of cultural asset and 130,000 images.

It is the first service of this type developed in Spain, and it allows establishing a common strategy between the various administrations, private entities and the Ministry of Culture for participation in Europeana and in the different initiatives of the 7th Framework Programme by applying the European Council's recommendations on digitisation, on-line accessibility to cultural material and digital conservation, published in the Official Journal of the European Union on 24 August 2006. It moreover allows complying with the points that provide details on the initiative, particularly point 6, which invites Member States to reinforce national strategies and objectives for digitisation and digital conservation; to contribute to Europeana, a common, multilingual access point to European cultural heritage; to improve the framework conditions for digitisation and on-line accessibility; to reinforce coordination within Member States; and to contribute to an effective overall view of progress at a European level.

Hispana includes a directory of the digitisation projects that are being carried out in Spain, that constitutes an instrument for coordination of the digitisation projects, preventing, for example, the digitisation of the same work two or more times[192].

6.3.6.4.3 Cooperative repositories. MDC: Digital Memory on Catalonia

The MDC[193] is a cooperative project that offers open access to digitized collections related to the Catalan heritage from scientific, cultural and research institutions. Each institution is responsible of the digitization process, that must be done according the "Digitisation standards: minimum requirements"[194] published and updated regularly by the CBUC. The repository is opened to libraries and archives collaborate, integrating and disseminating their selected digitised collections about Catalonia and/or heritage.

The MDC is a project lead by the Biblioteca de Catalunya (BC) and the Consorci de Biblioteques Universitàries de Catalunya (CBUC). Created on 2006, offers open access to 2 milion documents from 18 institutions, references to be increasingly the next years.

The repository is managed with CONTENTdm[195] (Digital Collection Management Software), that allows image description with international standards and ensure interoperability. As a result, dissemination and visibility of the documents are achieved, which are also accessible through other international repositories as Europeana[33] or OAIster[196].

6.3.6.4.4 Institutional repositories. The Calaix project

Calaix[197] is an institutional repository created to store, preserve and make accessible the digital material (such as documents in text, images, maps, and multimedia) from the Departament de Cultura de la Generalitat de Catalunya (Culture deparment of the autonomus government of Catalonia) related to the managment of Cultural Heritage (archeological and restauration activities, UNESCO Word Heritage sites, etc...)

Calaix uses the protocol (OAI-PMH) and is explicitly designed to enable Internet search engines to locate and identify its contents, and maximize their impact beyond conventional web pages. The website, metadata and digital files are stored in CESCA. The various units involved in the project provide metadata databases that are converted to their own metadata model Calaix. This conversion does not involve any loss even through information is displayed differently. Indexing is done according to the AAT, Art and Architectural Thesaurus (Catalan version) and documents consultation is guaranteed by the fact that they all have their own URL.

The project started on 2010 and was tecnhnically developed by Centre de Serveis Científics i Acadèmics de Catalunya[189] (Center for Scientific and Academic Services of Catalonia, CESCA).

6.3.7 Turkey

Turkish NREN is the ULAKNET[198] which was founded as a R&D Facility Institute of TÜBİTAK(*) in 1996. The Turkish Academic Network and Information Centre ULAKBIM consists of National Academic Network (ULAKNET) Unit, which undertakes the task of formation and operation of research and education network infrastructure in Turkey, and Cahit Arf Information Center, which provides information and document supply services nationwide.

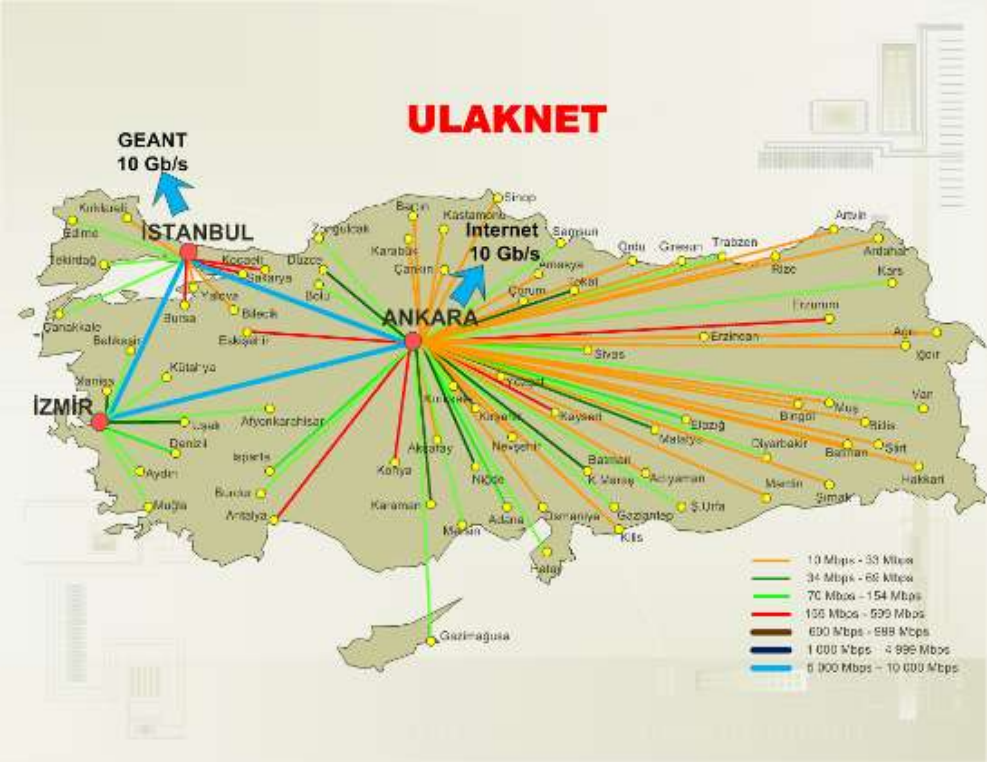


Figure 33. ULAKNET network

Ulaknet provides network services always one step ahead of the expectations of its users, made up of 100.000 lecturers and research assistants, and more than 2.500,000 higher education students. As well as operating the national research network infrastructure, ULAKBIM also coordinates nation-wide grid activities and operates national grid-based high performance computing (TR-Grid-HPC) e-infrastructure.

The following organizations which are the institutional elements of the national innovation system, are among the institutional users of ULAKBİM[199]:

- Universities,
- National science and technology organizations,
- National information and documentation centers,
- The organizations which offer support, make evaluations, and determine policy and strategy, for research and development,
- The national metrology, accreditation, and documentation centers,
- Technology observation and evaluation centers,
- Test laboratories,
- New technology display centers,
- Technology support units,
- Technoparks, science parks and technocities,
- Patent offices,

- Organizations, other than those cited above, which have a considerable interactive effect on the innovation system, from the point of view of the information they can offer and the decisions taken by them.

The objectives of ULAKBİM are[200]:

- To establish and operate a computer network enabling interaction within the institutional elements of national innovation system, and to provide information technologies support to help information production,
- To maintain the network interactive, high speed, open to new technologies and in global standards, in national and international context,
- To provide connections of the national network with similar networks abroad, to develop cooperation, and to represent Turkey in the level of academic and research networks on international platforms,
- To improve national network according to current needs and developments, to establish new networks and to give information services over the network,
- To carry out research and development work in the field of Information technologies,
- To offer information and document services to the end users most of whom are academicians using traditional methods and new information technologies,
- To improve and diversify the information and document services offered over ULAKNET according to current needs and developments,
- To carry out studies on making the information and document services become widespread at national level,
- To carry out work on the indexing of national periodicals and to provide access to these through the network, in order to contribute to national academic information exchange,
- To develop information exchange-based cooperations with the instituions abroad giving similar information services, and to represent TUBITAK in the level of national information and document services level on international platforms.

Ulakbim fields of activity are[201]:

- Undertaking the establishment and operation of research and education network infrastructure in Turkey (ULAKNET). ULAKNET provides Internet and network connections of universities, military and police academies, libraries, research and development organizations and some governmental organizations.
- The regular operation of ULAKNET, the continuity of services provided over ULAKNET and opening up of new services, network planning and network security studies are carried out.
- To take into consideration the security needs of the ULAKNET backbone and its end nodes, a Computer Emergency Response Team (Ulak-CSIRT), forming cooperation among ULAKNET's users to defend against and respond to cyber attacks.
- Wide spreading IPv6 applications over the ULAKNET backbone and to establish its connectivity with the rest of the IPv6 backbone in the world.

- Leading and coordinating the activities of national research communities in high performance computing towards a Turkish Grid, representing the TR-Grid National Initiative in regional and global grid projects, deploying the Turkish National Grid e-infrastructure.
- eduroam which stands for Education Roaming, is a RADIUS-based infrastructure that uses 802.1X security technology to allow for inter-institutional roaming.
- To carry out information gathering and sorting studies that will contribute to the information production of universities and research institutes, and cataloging and organization of this information, to develop software using only the open source codes in order to realize the information and documentation delivery services provided by Cahit Arf Information Center in the electronic medium.
- Participation national and global network infrastructure and grid projects.

6.4 E-Infrastructures for Long Term Preservation

The huge efforts on digitization require a management of all the digital material to guarantee for the future the preservation of such important data and their availability. Furthermore, digital preservation will become more and more strategic as the digital production increases.

Challenges on capacity storage, retrieval and long-term access to data can be easily faced thanks to the distributed computing infrastructures. Cooperation and coordination will be necessary to handle the rapid technological and organizational advance and solve the continue evolution of formats and the growing amount of information data to be stored.

Challenges:

1. Preservation level. Determinate the preservation level required for the different content types. The last years each institution has decided individually;
2. User needs. Collect and evaluate regularly the user needs and adapt, if necessary, the access policies and programs;
3. Metadata and documentation standards. Discuss and update regularly the metadata used. Challenges on interoperability, multilingualism, mapping, ontologies and linked data;
4. Formats. Standardisation of formats and adaptability to changes;
5. Capacity storage. The grid and the potential of distributed content. No single repository but several, challenge is how to interconnect them;
6. Personal training. Due to the nature itself of digital documents, the scenario will change continuously, and they will be necessary new skills, knowledge and an open mind as a model and way of working.;
7. Economic sustainability. It would be necessary to define the better solutions to assure long-term preservation with reasonable costs. Key: Cooperation and global policies. To solve financial issues: organise global policies and cooperation strategy in order to reduce preservation costs. Distributed repositories, use of national infrastructures to store and preserve items from different institutions;

8. National and international global policies. Key aspects need to be agreed at a national and international level. Urge of preservation policies and guidelines;
9. Interoperability. Tools and services in a distributed service network;
10. Persistent identification of digital objects. Detailed references on section 4.3 of the deliverable.

E-Infrastructures described in the previous sub-sections have been so far mostly used for exact sciences and a little adoption has been observed so far by the Digital Cultural Heritage community.

Indeed, e-Infrastructures, with their high performance computing and storage systems, can be very beneficial platforms for Digital Cultural Heritage for the following reasons:

- High performance/throughput computing systems:
  - Large scale, computationally intensive, problems can be tackled, e.g. “ab initio” sound creation, document/image digital restoration, etc.;
- High performance storage systems:
  - Geographically distributed replicas of files:
    - Fault-tolerant digital preservation;
- Simplified authentication systems:
  - Single Sign-On (SSO); support of Federations of Identity Providers;
- Fine-grained authorization systems:
  - Allow to precisely define “who”:
    - Individual user(s), group(s), organization(s), the whole world;
  - And “what”:
    - Read/edit/delete data and metadata;
    - Search, browse;
    - Creation of new repositories;
    - Role assignments, etc.

7 IPR and preservation

The issue of the protection of rights associated with the digital content that are preserved for a long term usage is a key factor for the correct implementation of the digital preservation initiatives. For this reason, a specific chapter is devoted to this matter in the present deliverable.

Since the matter is very much for specialists, it has been agreed to ask to Dr. Peter Mezei of the University of Szeged to introduce the matter for the scope of this deliverable. Dr. Mezei agreed to participate on a voluntary basis with the attached paper.

This article was presented at the DC-NET conference on Digitisation of Cultural Heritage and Long Term Preservation – the Role of eInfrastructures, 23-24 June 2011, Budapest. The present article was written with the help of the project named „TÁMOP-4.2.1/B-09/1/KONV-2010-0005 – Creating the Center of Excellence at the University of Szeged", which is supported by the European Union and co-funded by the European Regional Fund.

Dr. Péter Mezei, PhD, is an associate professor of law at the Institute of Comparative Law, University of Szeged, Hungary. His major field of interest is the comparative copyright law, with a special focus on digital copyright law and internet law. He is an elected member of the Hungarian Copyright Expert Council since 2010. He has delivered almost 30 university lectures and conference presentations in Finland, France, the United States, Canada, Germany, Austria and Russia. He is the author of almost 40 publications. His English language materials are available on http://ssrn.com/author=1697918. (mezei@juris.u-szeged.hu)

Future Legal Aspects of Long Term Preservation
Dr. Péter Mezei, Ph.D.

> *„Why, Mars Tom, if you knowed what chuckle-heads dem painters is (…) I see one of 'em a-paintin' away, one day, down in ole Hank Wilson's back lot, en I went down to see, en he was paintin' dat old brindle cow wid de near horn gone—you knows de one I means. En I ast him what he's paintin' her for, en he say when he git her painted, de picture's wuth a hundred dollars. Mars Tom, he could a got de cow fer fifteen, en I tole him so. Well, sah, if you'll b'lieve me, he jes' shuck his head, dat painter did, en went on a-dobbin'.* ***Bless you, Mars Tom, DEY don't know nothin'.****"*[202]

The words of Jim, the black slave friend of Tom Sawyer and Huckleberry Finn, serve as a great starting point to discuss whether *the European Union* (the painter) *can deal with the copyright problematic of the use of the cultural heritage* (paint down the one horn cow) *in the landscape of digital technologies and the Internet* (in ole Hank Wilson's back lot) well.

It is a well-known fact that copyright law and technologies have developed hand in hand in the course of time.[203] As soon as new technologies emerged, copyright law and the rights holders reacted on them, and they usually tried to force them back into the shadows. However, the rights holders always understood it within a short period of time that the new technologies could be used to create new works, new types of works, new type of data carriers and new business models as well.

The clash between the rights holders' and the society's interests was speeded up by the emergence of digital technologies. Just think of the Internet, and mainly filesharing.[204] Downloading a sound recording, a movie, a software or an e-book by the users doesn't seem to be a socially necessary use of copyrighted works; however, the digital preservation of culturally important creations – that has also led to harsh debates – seems to be a socially important activity. Indeed, Mtima and Jamar argued that "*[m]ass*

digitization of the world's books is the answer to many copyright social utility dilemmas."[205] I share their opinion in this respect, and I believe that the European Union has to take the lead in cultural preservation for the sake of the society.

Although there have been several programs before the Google Books Project whose aim was the (digital) preservation and making available to the public of culturally valuable creations (especially the Gutenberg Project), Google was the first to conduct book digitization on a mass scale *and at the same time* without the purpose to ask for the permission of the rights holders.[206] This is why the Google Books Project has necessarily led to court proceedings.[207] The parties in the actual cases have decided, however, to settle the dispute out of court. They published the original Settlement Agreement in 2008[208] and the Amended Settlement Agreement (ASA) in 2009.[209] Judge Chin has, however, declared the ASA as not fair, adequate and reasonable in March 2011 – mainly due to the copyright problematic of Google's opt-out policy, the class action settlement (Rule 23) objections and the constructive monopoly of Google over the content of the digital archive (and the foremost over orphan works).[210]

Google's ambitions clearly opened the eyes of the European regulators that the *private digitization projects are unstoppable*, and therefore the European Union has to react on them correctly.[211] Commissioner Viviane Reding noted in 2009 that *"[w]e should create a modern set of European rules that encourage the digitisation of books".*[212] Just a bit more than a year after this, the Comité des Sages published its *New Renaissance report* that included many straightforward but naturally non-compulsory recommendations about the topic of digital preservation, including the issue of orphan works and the role of Europeana in this process.[213] In March 2011 Commissioner Neelie Kroes declared that "*[t]here is a serious risk that there will be a '20th century black hole' on the internet. It is a duty of our time not to let this happen."*[214] She was actually referring to the orphan works question in the previous sentences, and she also emphasized that the Commission was planning to finalize the text of the proposal for an orphan works directive soon. The proposal was finally published in May 2011.[215]

In the light of the recommendations of the Comité des Sages and the draft regulations of the Proposal, we should discuss two major topics that closely relate to long term preservation: the issue of orphan works at all, and the participation of public partners in the digitalization processes.

Special attention need to be paid to the first issue: the *scope* of the Proposal. The Comité des Sages recommended to include all types of works (subject matters) into the future European norm.[216] The Proposal covers, however, only books and several audio and audiovisual works.[217] Bearing in mind that the digitization of all culturally important works would be difficult and expensive[218], it seems to be reasonable at first glimpse to cover only the above subject matters. Furthermore, the Proposal requires the Commission to check on a yearly basis whether the enlargement of the directive's scope is necessary in the future or not. Nevertheless the disregard of several types of works seems to be illogical in the light of other provisions of the Proposal.

Whilst the New Renaissance distinguished between commercial and non-commercial uses, and required payment of a reasonable remuneration to the rights holders for every commercial use of the orphan works, and "expected" payment for every non-commercial use,[219] the Proposal differentiates between permitted uses (the reproduction of the works and the making available of the digital copy) and authorized uses.[220] The latter uses may surpass the boundaries of the permitted uses, if they fulfill the strict requirements imposed by the Proposal on them. One of these requirements is the payment of remuneration to the rights holders.[221] It follows *a contrario* from this provision that the European Commission did not envisage a payment of remuneration in case of the permitted uses. It consequently means, however, that the failure to include every subject matter into the scope of the future directive shall be explained by other than financial reasons.

The New Renaissance further recommended supporting the new European model of the protection and use of orphan works by extensive rights information databases, such as ARROW.[222] The Proposal seems to bring this recommendation to perfection since it requires the users to record all the necessary information in databases in respect of the diligent search for an orphan work, the result of the diligent search, and the uses of the orphan works.[223]

Another notable element of the New Renaissance is the Comité's opinion on the public-private partnerships. The Comité understands it well that the private partners have the funding, technology and expertise in the field of digital preservation. This is why the New Renaissance said that *"[t]he key question is not whether public-private partnerships for digitisation should be encouraged, but 'how' and 'under which conditions".*[224] The European Commission also stressed that *"[t]he Directive is without prejudice to the freedom of contract of such organizations in the pursuit of their public interest missions."*[225] The wording of the text is, however, notable. The proposal subordinates the participation of private partners to the public (social) interests, rather than giving free hands to the private partners.

After all, the article shall be finished with the words of Jim in the following transformed way: "Bless you, Mars Tom, ***DEY know somethin'***" – but only something. Remember the motto I used at the beginning. Even if Mark Twain died before the First World War, and therefore his works are already in the public domain, that quote was available on the website of Google Books and the Project Gutenberg, and not on the Europeana. I truly believe that the European Union is on the right track, both in respecting the interests of the rights holders and the society, but we still have to do a lot to help the digital preservation of and granting access to culturally important creations by the tools of the legal system.

8 Training

8.1 Importance of staff training

The importance of training in digital archives relies on the creation of a better awareness of the facilities offered by the repositories for both the owner of the collection and the user looking for the available information, as well as of the different techniques of creation and preservation of contents. There are a number of aspects to consider in implementing the training process.

The first aspect in this process is to understand the benefits of a digital archive as a sustainable tool for content and collection management and preservation. The best methods to start with to achieve this goal is to get familiarized with the different models of digital archives.

The second aspect in the training process is assessing the technology and resources (hardware and software) needed for the creation of the archive. Networking is also beneficial for the purpose of using open source software and the possibility to adapt to a unique working environment while keeping a standard.

The third aspect focuses on the organization, management and accessibility of contents, as well as on the determination of users' needs and on how to satisfy those needs in the digital archive. All data needed to describe and identify the digital content for better retrieval through different search engines, basically by the use of metadata, are to be assessed.

The forth aspect is the maintenance of the archive though backup strategies and continuous evolution of the contents based on hardware and software changes.

Finally, the last focus is on the intellectual property and copyright of digital contents and the concept of Creative Commons.

8.2 Common resources for training on digital preservation

As mentioned before, staff training is one of the most important aspects of the digital preservation when preserving the cultural heritage. Many European countries were interviewed about their preservation policies / strategies for the workshop in Ankara and these survey results indicate that lack of expertise is one of the problems witnessed when preserving cultural heritage. In order to establish a well built preservation system, training about digital preservation must include common aspects like metadata usage, interoperability, standardizing digital archives, conservation in multiple places etc.

8.2.1 Metadata

Metadata are the data providing information about certain contents. It is important to understand the types of metadata necessary to maintain digital repository. There should be emphasis on:

- the metadata standards;
- mapping and applying metadata standards;
- understanding the relationship between data and metadata;
- metadata for better search on the internet;
- metadata creation for specific culture digital content (photo, video, sound recordings, etc.);
- the quality of metadata.

8.2.2 Interoperability

Interoperability is the capability of two or more systems of communicating and sharing information. It is one of the main rules in standardizing the information and in facilitating the efficient dissemination of contents.

One of the main goals of digital archives is to provide the user with the best service of searching with the minimal effort, and to facilitate the link between the data provider (the one who maintains one or more repositories) and the service provider (the one who requests from data providers and uses the metadata as a basis for building value-added services).

To reach this step, the learning of the different protocols for Metadata Harvesting is an important issue.

8.2.3 Migration of non-standard digital archives to standard digital archives

Most institutions working in the field of conserving culture heritage have gone through a process of digitisation and conservation of the contents they own. Those attempts are mostly non-standard. It is very important to integrate those attempts in archiving to a better and more standard form.

This learning process should be done through practice examples of non-standard archives, with an emphasis on how to migrate those archives to a more standard one based on:

- The standardization of the content itself;
- The use of metadata.

8.2.4 Conservation in multiple locations

It is also important to replicate data in multiple locations. Backing up is making copies of the digital repository which may be used to *restore* the contents in case of data losses. Loss of digital data is more frequent than that of the originals or hardcopies; the backup process is then of great importance.

The learning process should include the integration of different storage media as magnetic tape, optical storage, as well as remote backup service (internet). The important formats of location and data repository include:

- On-line backup storage;
- Near-online storage (ex: magnetic library);
- Off-line storage (storage not directly connected to the computer);
- Off-site backup,

9 Conclusions

This document points out the importance of the long term preservation of digital cultural content, to safe the large investments that the cultural institutions are doing in the digitations of cultural heritage. Heritage items have to be preserved without any loss of data. They have to be stored, maintained and managed in appropriate ways. Digital objects are much more 'fragile' than traditional analogue documents such as books or other hard copy mediums. Digital objects are also much more venerable to physical damage. One scratch on CD-ROM containing 100 e-books can make the content inaccessible, whereas to damage 100 hard copy books by one scratching move is - fortunately - impossible. A flash memory stick can drop into glass of water or get magnetised, portable hard drive or laptop can slip from your hands and get irreparably damaged in a second[226].

E-infrastructures aim at empowering researchers with an easy and controlled online access to facilities, resources and collaboration tools, bringing to them the power of ICT for computing, connectivity, storage and instrumentation. Usage of e-infrastructures may assist to address challenges on storage, data security and access for digital preservation systems. Therefore, integrating digital preservation systems with large capacity networks or international NREN or Grids will rapidly become prevalent worldwide[227].

In order to provide this, policies and strategies have to be developed nationally and globally. What we have seen in the workshop in Turkey and in the survey results is that:

1. Institutions implement their own methods and there is a need for common policies and strategies to make possible the sharing of best practice.
2. Institutions and their staff are mostly unaware of the e-infrastructure possibilities; in some cases they are not even connected to the e-infrastructure providers.
3. Training of the staff is a must.
4. There are issues about copyright, accession, sustainability, storage, back-up and data integrity that need to be addressed at the different levels, from the policy domain to the technological research and development.
5. There is a lack of common approaches and solutions to the problem of the persistent identifiers.

Some of the actions we suggest to perform about preservation of cultural heritage are the following:

- Interoperability and a global policy about preservation in each country should be provided;
- In order to overcome the legal issues about copyright, accession and data sharing, necessary changes in laws should be undertaken;
- Staff training should be provided in order to put people in a position to factually contribute to the development of operational preservation system;
- Publications and guidelines about how to build preservation systems including selection, acquisition, conversion, storage, backup, preferred file formats, etc., should be provided to institutions;
- Institutions should contact their e-infrastructure providers and to discuss and plan the use of e-infrastructure possibilities in the Digital Preservation systems;
- Interoperability and cooperation possibilities about NREN and Grid connections can be shared widely with similar projects like DC-NET, EPIKH, SUMEDGRID Support, etc.;

- After the determination of national policies and strategies, international activities about harmonisation of policies and strategies at international level should be undertaken.

10 Annex 1. Survey Results country by country

Template of the Survey

PART ONE : INTRODUCTION
Name / Type of Institution:
Country:
Project name:
Realization date:
Website:

PART TWO : POLICIES
Do you have a digitization / preservation policy?
Is there any legal deposit service policy of your digitized object; Is it a legal policy through the institutions or your own specific policy?
Does your institution have its own annual budget for preservation process? (migration, refreshing, digitization etc.)
Does a National Research & Education Network (NREN) exist in your country? (Please find a comprehensive list of existing NRENs at the bottom of this page http://www.terena.org/activities/compendium/) - If yes, are any sites of your organization connected to it? Through how many links? What is the total bandwidth of your links?
Are there any plans in your country to build a National Grid Initiative (NGI)? (Respondents based in European countries can find here http://knowledge.eu-egi.eu/ a map of NGI' s in preparation) If yes, is the NGI already operational? If yes, could you provide its name and website?
Is there any specific preservation service used like Archive-It, Digital Archive?
What are the Quality Control Criteria's used during digitization process?
What's the strategy about the records which have copyright?
Do you provide an open/free access or do you have a membership system?
Is the master digitized copy being served to users? If not - Is the master digitized copy – End user copy records differ?
Do users have to pay to use/view/download the digitized record?
Have you used OCR software for texts?

PART THREE : TECHNICAL INFORMATION
Types of items that'll be digitized and preserved? (Books, manuscripts, cultural heritage, audio-video, GIS etc.)
Do you check other institution's preserved records in order to prevent overprints?

Formats of digitized / preserved materials? (tiff, jpeg, mp3,aac, wma, doc, etc.)
Total preserved items, total percentage with the collection, Stored data size?
How do you preserve your digital records? (CD-DVD, Internet, RAID Array, Network Attached Storage, institutional repositories, web archiving etc.)
Is it possible to view the preserved record in various display types? (PDA-Mobile-Other browsers)
Which functions are included in the browsing/viewing software? (Hypertext links highlight etc.)
Any special software, tool used for viewing and downloading the records?

PART FOUR : STANDARDS, METADATA, FORMATS
Which standards are used for descriptive metadata? Is there any kind of Persistent Identifier for digital objects (e.g. DOI)?
Which standards are used for administrative/technical metadata?
Which standards (if any) are used for preservation metadata (es. PREMIS)?
Which model (if any) is used for the digital repository (e.g. OAIS – ISO 14721 – 2003)?
Which standard is used for letting the repository be harvested (es.OAI-PMH)?
Is there any International Standards used at the preservation process? (ISAD(G)), ISO 15489-1,ISO 14721: 2003 etc.)

PART FIVE : STORAGE- BACK UP
What's the back up strategy of your preserved data?
How often do you back up your preserved data?
Do you have a recording history of your preserved records?
How do you keep the integrity of your digitized records? (hacker attacks, viruses, hardware failure problems, environmental issues)
What's the renewal timeline of your preserved records?

PART SIX : PLANNING AND CONCLUSİON
Have you implemented a Preservation Plan?
Is a trustworthiness certification (e.g. TRAC checklist, DRAMBORA) part of your Preservation Plan?
Have you provided for a financial sustainability plan?
Do you have a policy for migrating data to more recent technological platforms?
What are the main problems, difficulties you have witnessed during the preservation process?
Do you think any other additions, comments, issues missing in this survey?

The summary of the replies is synthetised in the following sections.

10.1 ITALY

Istituto Beni Culturali

POLICY

- Preserves digital born records.
- Has an annual budget
- No answer given to NREN and Ngi
- No answer given to Preservation service, Quality control Criteria, copyrighted material, access type, master-user copy questions, usage policy, OCR
- LESS INFORMATION ABOUT THE POLICY PART OF THE SURVEY*

TECH INFORMATION

- no answer given to overprint question
- Preserves mostly PDF
- 80.000 records in institutional repositories
- No answers given to functions and tools, software used.

STANDARDS

- Encoded Archival Description (EAD) and Encoded Archival Context (EAC) for metadata, Uniform Resource Name (URN) for identifiers, MAG for administrative metadata, Premis for preservation metadata, OAIS for repository, ISAD(G) for international standard

BACKUP

- Back up strategy: 3 tape copies, synchronous copy on main and disaster serves. Progressive daily back up, full weekly back up. Keeps recording history
- Integrity of the records: Special data center with Information Security Management System (ISMS) standard 127001 and periodic integrity check
- Renewal time-line is 5 years.

CONCLUSION

- A preservation plan is in implementation process. No certification for the plan, There is a sustainability plan, policy for more recent techology transfer, digital born records is not suitable for LTP.

Biblioteca Nazionale Firenze, Roma

POLICY

- Mibac for digitization through state and other libraries. Magazzini Digitali for preservation aims to be national infrastructure for Italian libraries.
- Magazine Digitali is the legal deposit platform by the law. Has 3 year annual budget for an experimental implementation
- Garr is the NREN, Magazzini Digitali connected to 2-8 Mb. There is an Italian Grid (IGI)
- There is no specific preservation service; quality control is embedded into digitization process work-flow. Copyright strategy is based on deposit of electronic materials and license for accessing and using with different levels.
- Free access for the test period, user copies are served with lower resolution, some of them ocr'ed.

TECH INFORMATION

- Types of items preserved are digital born items, book, manuscripts, maps, music sheets, There is overprint check, Main formats are Tiff and Jpegs. Total 198 TB.
- Data storage is based on replication: different machines, different data centers (200 km.) all 27001 verified.
- Functions and tools software used are to be defined.

STANDARDS

- Unimarc and dublincore are used for descriptive metadata, National Bibliographic Number (NBN) for is used for identifiers, MAG (Metdati Amministratavi Gestionali) mapping to Mets/Mix for administrative metadata, Premis embedded into MAG for preservation metadata, OAIS, ISO 14721 – 2003 for repositories, OAI_PMH is for harvesting.

BACKUP

- Has own strategy: Magazzine Digitali Infrastructure, Keeps recording history of any file created or changed, Back up strategies and Renewal timeline are to be defined, MAG has a checksum.

CONCLUSION

- A Preservation plan is not yet implemented, there are plans for certification. Sustainability plan is not yet provided. Migrating to recent technologies will be defined. Problems are defining a plan, sustainability, training, copyrighted materials.

10.2 SPAIN

Girona City Council

POLICY

- There is no global policy, has their own policy, there is no legal deposit service, There is a budget for preserving and restoring but not migrating.
- There are no answers given to NREN and NGI and quality control questions.
- Uses specific service called Padicat and iArxiu in future.
- Provides open access, serves user copy, uses OCR.

TECH INFORMATION

- Preserves mostly manuscripts. Also books, audio-video and photographs. Checks for overprints, formats for maps and preservations: Tiff, Consultation: mp3, jpeg, Manuscripts: Pdf
- 18 TB, 2.5 million digitized, 90% in Repositories, planning a mobile access and new web.

STANDARDS

- EAD, ISAD, IPTC for photos, Sepiades (Sepia Data Element Set) for descriptive metada, There is no identifiers used, there is no answer given to administrative metadata, Premis used for preservation Metadata, OAI used for repositories, OAI_PMH, SRT used for harvesting, ISAD, ISO 15489-1 are the international standards.

BACKUP

- For active documentation: weekly, monthly and yearly. For non-active: monthly, yearly. Doesn’t keep a recording history, Uses Hash check for integrity. There is no renewal.

CONCLUSION

- There is no preservation plan but needed. No certification planned, there is sustainability plan, and there are no plans for migrating data to recent technologies.
- Issues: Infrastructure linking, several repositories interconnecting, storage capacity, metadata simplification. Selecting what to keep and who should decide.

Catalunya National Library

POLICY

- There is a preservation policy. Collaborating with several projects and collecting digital resources from Catalan domain, implementing repository for both natural born and digitized materials.
- Legal deposit service for digital preservation is to set a new one. There is annual budget depending to library's budget
- Rediris is the Spanish NREN, Anella Cientifica is the Catalan one. There is no answer for NGI.
- Padicat used for preservation service. There is no answer given to Quality Control. Library digitizes public domain materials, agreements with right holders for born web archive. Open access is provided. Master copy is for preservation, User copies are low resolution copies. For Commercial purposes users have to pay tax to library if they want high resolution copy. OCR is used.

TECH. INFORMATION

- Preserves almost anything from books to maps. Checks for overprints. Tiff and jpeg for printed, Wav and Bwf for sound, mpeg1 and mpeg2 audiovisuals. 68Tb Tiff, 7.7 tb html, 3.3 jpeg, 2.000 hours of audio-video stored. Preservation in institutional repositories. No various displays, offers hypertext links. There is a new preservation system is being developed.

STANDARDS

- No answer given to descriptive metadata, Dublin core and DOI used for identifiers. DC, own standard and MIX for administrative metadata. METS/BC's standard based on premis used for preservation metadata, OAIS used for repositories, repositories cannot be harvested. There is no international standard.

BACKUP

- There is no global strategy on backup. But they preserve records in different media types. No answer given to how often question, log file system for keeping history. Restricted serves access, cyclic virus detection, raid disk structure, MD5 check, climate and electric safe servers used for keeping the integrity of the records.

CONCLUSION

- A preservation plan is now being implemented, Certification, sustainability is not yet planned.

Arxiu Nacional de Catalunya

POLICY

- Digitizes documents related to Catalan Heritage, there is no legal deposit service, embedded in General budget.
- Not linked with NREN and NGI
- There is no specific preservation service, quality control criteria is based on image completeness (resolution, noise cleaning, brightness etc.)
- If ANC has the copyright owner, their rights are paid. If ANC is not the owner, users can't get digital copies for research purposes not commercial uses. Provides open access with low resolution copies. Generally there is no ocr use for the type of the material.

TECH. INFORMATION

- Preserves almost anything in archive. Anc doesn't check overprints because it has unique collection. Formats produced from digitized materials: tiff, avi, mpeg2, wav, mp3, aac, dpx, jpeg. 1.925.250 million items storing 25 TB. Stores in repositories since December 2010. No Various Display types, special functions, special software, tools.

STANDARDS

- No standards for descriptive metadata, persistent identifiers, preservation metadata, repositories or international standards. ISAD(G) and NODAC used for administrative and technical metadata.

BACKUP

- Computer experts are responsible from the records hosted on the server. Daily backups. Only track the file when processing. Department of Culture's policy is responsible for the integrity of the data.

CONCLUSION

- Preservation plan, certification, sustainability plan is not yet implemented.
- Future challenges are storage, format standardization, and viability of repository.

10.3 GREECE

Hellenic National Audiovisual Archive

POLICY

- Henaa updates their policies relevant to EU policies. There is no answer to legal deposit service. Basic needs are covered from the annual budget. Major works can be structural funds and other Eu programs.
- Greek NREN and Grid exist but Henaa is not connected, there are possible plans for linking.
- There is no special preservation service used, quality control criteria and copyright questions is not answered. No Ocr due to file type.
- Provides open access for research, demands payment for commercial uses. Downloading is permitted. Low resolution copies are served to users.

TECH. INFORMATION

- Preserved items are mostly audio-video. Future plans for cross check overprints. Formats of the preserved items are DPX, MPEG2 and Digital Beta. 3000 items storing 300 TB in LTO3 and Digital Betacam Tapes. Low resolution http/rtmp streaming enabled, functions included are annotation viewing, video streaming. Flash browser plug-in needed for viewing.

STANDARDS

- Simple Dublin core standards are used for descriptive and administrative metadata. No standards used for preservation metadata, digital repositories, harvesting, international standards.

BACKUP

- There is no global backup strategy. (cyclic timeline, renewal) Keeps recording history only.

CONCLUSION

- Preservation plan is not implemented. Certification is not planned. There is a short term sustainability plan. Also there are future plans for migrating data to recent technologies.

National Archive of Monuments

POLICY

- HMCT implements a regional cultural management scheme. But an operational digital

curation policy is not yet formulated. There is a legal deposit service for electronic records. Preservation activities used from the annual budget. EU funds are also used.

- No HMCT sites linked with Greek NREN. There is Hellas Grid. There is no special preservation service used. Quality Control Criteria implemented are DNAM's technical digitization standards.
- HMCT offers public and licensed content which is free for access but paid for commercial. Master copies differ from User Copies (image format, analysis and record info) User copies are served. No OCR used due to type of materials.

TECH. INFORMATION

- Types of preserved items are cultural heritage objects, monuments, archaeological sites, 3D and Virtual Reality. Their formats are raw, tiff and jpeg. 138.847 objects are stored in dvds. Total 10 % of the collection. Display Types, Functions, special tools software are not used.

STANDARDS

- For descriptive, administrative, preservation metadata general standard CIDOC CRM is used. No Digital repository has been set, Harvesting enabled through xml exports.

BACKUP

- A back up of the database and the backup of the photographic material is kept at DNAM. There are no regular backups. There is no recording history. Antivirus checks and environmental control mechanisms for the data integrity. There is no renewal timeline answered.

CONCLUSION

- A preservation plan and certification is not implemented. There is short term sustainability plan. Migrating through Europeana platform.
- A comprehensive policy and strategies are needed for national digital curation strategy.

10.4 SLOVENIA

Ministry of Cultrure - Indoc Centre

POLICY

- A preservation policy exists not explained. There is no legal deposit service, annual budget. They're not connected to the Sloven NREN and NREN is protyping the grid.
- There is no special preservation service used, international and national guidelines used for quality control, licensed texts cannot be given to 3rd party. Provides open access. Lower resolution copies for internet use. No OCR.

TECH. INFORMATION

- Preserved items are cultural heritage, photos and GIS. There is no check for overprints. Preserves 1 TB of data. Stores is Raid disks, Cd/Dvds. Various display is enabled. A dynamic GIS map allows searching by name, place, Doi, municipality and parcel.

STANDARDS

- EAD, ISAD, IPTC for photographs and SEPIADES, Dublincore. No answers given to preservation metadata, digital repository, harvesting standard and international standards.

BACKUP

- Different for active and non active documentation. For active: weekly, monthly, and yearly. For non active: monthly, yearly. Keeps no recording history. Physical and software protection for the files but no details given.

CONCLUSION

- There is no plan for preservation. No certification, No sustainability plan, Migrating is not systematically planned.

10.5 FRANCE

France National Library

POLICY

- There is a preservation policy including not only curation and binding also digitization, preservation, staff training, awareness-raising, research and technology monitoring. An official legal deposit service exists since 2006. Bnf has an annual budget for preservation processes.
- France NREN is RENATER, Bnf is connect to it through 1gbits bandwidth. There is no NGI answer given.
- There is no special service used for preserving.
- According to the French laws, Bnf has the right to digitize but access is only to library users. Open access for non copyrighted materials.
- Master copies are served to professional users with an agreement. User copies are low resolution copies. To view copyrighted materials users have to register as a researcher and pay the annual fee. OCR is used.

TECH. INFORMATION

- Preserves almost anything from books to maps. Collaborates with 70 other French libraries for selection of the digitized materials, so checks overprints. Preserve formats: Tiff for prints, Mpeg2 for video, wav for video, arc for web archiving. 1.500.000 items preserved, 1% total. 516 TB. Storage for 2 tape libraries (LTO5 and t10000) and one additional NAS copy.
- Various display types are planning. Epub format, a web application for mobiles and ipads. Hightligh, zooming, scrool fullscreen functions enabled.

STANDARDS

- Dublincore for descriptive metadata, DIDL on Gallica. ARK is used for identifier, For administrative/technical metadata, METS, PREMIS, MIX, TEXTMD, MPEG7 are used. Spar is used for digital repository. OAI_PMH is used for harvesting.

BACKUP

- Two copies are stored on different tape technologies located on different sites. Two copies are systematically done. The provenance metadas are stored in OAIS. For keeping the integrity Oais system is not connected to internet, access provided by special software. Administrative access is restricted. Disaster Recovery Prodecures are in place between 2 locations. MD5 checksums, tape conditions are checked with software.

CONCLUSION

- Full preservation plan is not yet in place but a storage level plan is in place. A certification from the French National Archives. There is a short term sustainability plan. Migrations to recent technologies are continuous.
- A new software has to be designed for the needs. Diversity of the library collections, changes in planning and policies are the issues faced.

Ministère de la Comunication et la Culture

POLICY

- Yes there is a policy by national digitization steering committee since 2007. There is a legal deposit service for audiovisual by Bnf, INA for radio and T.V and CNC for film. RENETAR is the NREN, all national cultural institutions are connected. France Grilles is the Grid.
- For quality control, the ministry provides recommendations and specifications for each type of document.
- The process is done with a total respect of intellectual property rights. Open access for public domains is provided. The master copy is for preservation, user copies are low resolution copies. OCR is used.

TECH. INFORMATION

- All types are taken according to the national digitization plan. There is national inventory of digitized collections for cross check overprints. The ministry provides standard specifications for each type of document / format.
- Functions and specific tools, software depend on the service and the institution, there is option for various viewing.

STANDARDS

- The ministry provides the standards for descriptive, administrative etc. metadata. Different standards are used according to the domain and the objectives. Minimum standard is Dublincore. Premis for preservation metadata, OAIS for digital repositories, OAI_PMH for harvesting, ISAF(G), ISAAR (CFP), ISDF and ISDIAH for international standards.

BACKUP

- Backups are done by each institution in charge. Bnf uses SPAR, SIAF also developed a programme.

10.6 TURKEY

National Library

POLICY

- There is no global policy. There is no legal deposit service for digital content. There is not a separate budget for the digitization and preservation. They use Annual budget.
- Turkish NREN is Tubitak Ulakbim's ULAKNET. National Library is connected with a 8 Mb bandwidth. There is also Tr-Grid.
- Metus Library software is used as a special service. Quality control criteria depend on the Electronic Record Management System. National Library provides membership system. Users have to pay for some digital content. End-user copies are the same as originals. OCR is used.

TECH. INFORMATION

- They preserve almost all types of items from Manuscripts / Rare Works to gramophone records. For the talkin library section there is a overprint check. Main formats are tiff, jpeg, mp3, pdf which stores 20 TB. Main Storage places are CD-Dvd, Raid Discs, and Institutional Repositories.
- There is an option for viewing various formats, Hypertext linking is functioned. There is no special software, tool used.

STANDARDS

- Marc and Dublincore for is used for descriptive metadata, AACR for administrative metadata, Electronic Record Management System for preservation metadata.

BACKUP

- Daily periodical back up in data storage units. Keeps recording history. MD5 and SHA1 algorithm checks for the integrity of the records.

CONCLUSION

- There is sustainability plan for the preservation processes. Continuous transferring to new technologies when the technology occurs. Financial issues are the main problem faced.

Suleymaniye Manuscripts Library

POLICY

- No global policy, strategy for digital preservation. There is no legal deposit service. There is an annual library budget. Preservation processes are being met from that budget.
- Ulaknet is the NREN but our library is not connected. Also there is a Tr-Grid.
- No special service used for preservation process. Quality control criterias are related with the success of scan (Resolution, Noise, Focus)
- Membership system is used for users to buy the digital copyrighted content. Master copies are raw copies, user copies are tiff or jpeg. No OCR software is used.

TECH. INFORMATION

- Preserved items are manuscripts and old rare works. There is no overprint check. Used formats are Tiff and Jpeg. Stored data size is 10 TB which is all of the manuscripts and 30% of the old rare works. Mostly Cd-Dvd storage and institutional repositories. There is no option for different types of viewing platforms. There are no special functions.

STANDARDS

- No information about that section.

BACKUP

- Daily and weekly backups are taken. Keeps recording history.

CONCLUSION

- There are plans for a new project for migrating old digital content to new technologies.

State Archives

POLICY

- There is no global digital preservation strategy or policy used also there is no legal deposit service. NREN Ngi questions are not answered.
- There is no special preservation service are used. Quality Control is related with the success of scanning. (Resolution, Noise, Brightening, Focus etc.)
- Archive materials in Turkey have no copyright. Institution uses membership system. Master copies are for archival purposes. User copies are low resolution copies which take low space.
- Future web usage will be charged system. No OCR software is used.

TECH. INFORMATION

- Preserves any kind documents in paper format. No overprint check due to type of the institution. Used format is mostly Tiff. 15 million items storing 45 TB. 5% of the total collection.
- Storage Units are DVDs, Internet and Raid Array Discs.
- Hypertext function is enabled, No choice for different platforms. There is no special software, tool used.

STANDARDS

- ISAD(G) and ISO 15489-1 used for metadata records. File types and creation date metadata's stored.

BACKUP

- Strategy of backing up is storing in different Medias in different places. Daily, weekly and monthly intervals. Regular checkups are done to check the integrity. There is no renewal unless there is a hardware or software failure.

CONCLUSION

- There is no comprehensive future preservation plan. Financial sustainability provided by the annual budget. Continuous Migration strategy to new formats, technologies. Main problem witnessed is technological issues.

10.7 IRELAND

An Comhairle Leabharlanna - The Library Council

POLICY

- The project acts as a repository for digitized material created by the public libraries. The project has previously provided training and facilitation for the digitisation process and to build capacity.
- There is no legal deposit service.
- There is no special budget for preservation.
- Heanet is the NREN but the organization is not directly connected. Grid-Ireland at www.grid.ie is the NGI.
- For quality control criteria, provided training and facilitations is used.
- Copyright policy: Copyright on the content held at askaboutireland.ie is held by the contributing public libraries, Open access is provided.

TECH INFORMATION

- Preserves Books, manuscripts, images, photos, cultural heritage materials, audio, and video.
- There is no overprint check because the project acts a repository for the content of various public libraries.
- Main formats are tiff, jpeg, mp3, pdf, doc etc. total 20.000 items which stored in repositories, network attached servers.

STANDARDS

- Basic Dewey system is used. Other questions about this subject are not answered.

BACKUP

- Redundant servers with backup external drives, as supplied under a SLA with our internet service provider. Backup external hard drives stored off-site for some content. Daily, weekly, monthly backups.

CONCLUSION

- Sustainability relies on ongoing (annually reviewed) national funding from the Irish government.

Note: While the survey will develop an awareness of the baseline for digital preservation of cultural heritage, some questions which explore the ability (or not) of respondents to interoperate and to share (or indeed remotely backup) one another's data would be interesting. If the INDICATE project has a "vision" of where digital preservation should be going, then questions which explore the viability of this vision might add some value.

10.8 THE NETHERLANDS

DANS (Data Archiving and Networked Services)

POLICY

- We have identified file formats and information properties which we preserve and have defined methods in order to do so. We have compiled a list of Preferred Formats.
- There is no legal deposit service, no special budget.
- No answer for NREN but gives the Grid: Big GRID www.biggrid.nl.
- Copyright policy: We enable the depositor to keep the data under embargo for two years. On request this can be prolonged. Another option is to limit downloads to persons who have requested permission to do so.
- Born digital records are preserved and accessed with free registration.

TECH INFORMATION

- Preserves pdf, formatted text, unformatted text, comma separated data, gis data, audio/video etc. There is a list of preferred formats for the digitized items. Items sored in NAS, with dual dual tape backup in separate locations. Online stored size is 2.6 T.B, offline size is 25 TB.

STANDARDS

- Urn is used for descriptive metadata, for administrative/technical metadata Qualified Dublin Core is used. OAIS is used for repository. OAI-PMH is used for harvesting. Aims to use ISO 16363 (in development) as an international standard.

BACKUP

- Backup strategy is to take tape backup on two separate locations. A third copy in the cloud is planned. Daily backup is taken. Keeps record history.
- IT-service provider guards our server systems against hackers. For hardware failures, we have our backups. We build our electronic archive system with security in mind. We intend to add virus checking to our ingest process.

CONCLUSION

- A preservation plan is not yet implemented. About Certification: We have recently performed a self-audit on the RAC (ISO 16363 draft), which was reviewed by the author team of the RAC. We intend to acquire formal RAC certification in 2015. Sustainability depends on the Royal Netherlands Academy of Arts and Sciences, they're not independent.
- Main issues are working by formal procedures, keeping records of all data handlings, keeping all preservation information, running checksum controls.

10.9 LITHUANIA

Lithuanian Art Museum

POLICY

- For the preservation policy, they claim to have but no details given.
- There is no legal deposit service.
- They're aware of NREN and Grid but only grid's details are given. www.litgrid.lt
- Quality control depends on image and metadata quality control.
- Copyright depends on the agreements with owners. Free access is provided.

TECH INFORMATION

- Preserves Cultural heritage, artworks, audio-video, Checks for overprints, Main formats are tiff, jpeg, pdf, mp3.avi. 2 TB items storing 20% of the total collection. Preserves in CD-DVD, RAID Array, web archiving.

STANDARDS

- CDWA and Dublincore for descriptive metadata standards, No answers to administrative, preservation, repository and harvest metadata. ISAD(G) and Minerva is used for as an international standard.

BACKUP

- Backup strategy is mainly on main repository: RAID Array; back-up copies: magnetic tapes, CD-DVD. Weekly backups, Keeps history log, For the integrity Firewall, password protection, virus protection

CONCLUSION

- Main issues are capacity planning, poor network connection.

10.10 SWEDEN

National Archives

POLICY

- No details are given about the preservation policy.
- There is a legal deposit service: According to the Swedish Constitution those public records (regardless of form) that are not specifically decided to be disposed, shall be preserved. In the very end, they will be transferred to the National Archives.
- The institution has an annual budget.
- In the process to connect the National Archives through the Swedish NREN, Grid: Swedish National Infrastructure for Computing , http://www.snic.vr.se
- Quality Control: The National Archives has a special production procedure named "the Digital Chain" which includes quality control.
- Copyright / Access : No specific strategy besides that digitized documents with known copyright is not provided on the Internet but can (if not covered by secrecy regulation) be available in the reading rooms of the National Archives. All public records that are not covered by secrecy regulation are open to the public.
- Master – end user copies are the same. OCR: most of our sources are handwritten and are not suited for OCR.

TECH INFORMATION

- Preserves Archival records, maps, audio-video and photographs. The records are unique, stored size is 1.3 Pbyte. The physical storage media is LTO tapes and SAN.
- Various Display, Special Functions: there are DjVu-readers for Iphone and Ipod, Easy panning and zooming of document images.

STANDARDS

- Descriptive, international metadata: National implementations of ISAD(G), ISAAR, EAD and EAC , Administrative and Preservation metadata : METS and PREMIS, Repository: OAIS and TRAC, Harvesting Metadata: OAI-PMH and EAD

BACKUP

- Strategy: Two identical copies stored and maintained in different part of the country. These two copies will never be in contact with each other.
- 5 years is the renewal timeline. Keeps record history. Integrity is provided with a A closed secure digital archive with very restricted access. (archival copies)

CONCLUSION

- There is preservation, financial sustainability plan but details missing. Planning to get certificated. Main issue : Documentation of old accessions items.

11 References

[1] INDICATE Project description: http://www.indicate-project.eu/index.php?en/94/about

[2] Digital Preservation Survey (Annex I)

[3] Consultative Committee for Space Data Systems. (2002). Reference Model for an Open Archival Information System (OAIS). Washington, DC: CCSDS Secretariat, p. 1-1

[4] Erin Scime (8 December 2009). "The Content Strategist as Digital Curator". A List Apart.

[5] "What is Digital Curation?". Digital Curation Centre. Retrieved 2008-04-01.

[6] Elizabeth Yakel (2007). "Digital curation". Emerald Group Publishing. Retrieved 2008-04-01.

[7] Digital Curation Centre: http://www.dcc.ac.uk/

[8] Giaretta, David (2011). Advanced Digital Preservation. 1st Edition. Springer. pp 62-64.

[9] Online Computer Library Center: http://www.oclc.org/

[10] Online Computer Library Center, Inc. (2006). OCLC Digital Archive Preservation Policy and Supporting Documentation, p. 5

[11] Cornell University Library. (2005) Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems.

[12] Rothenberg, Jeff (1998). Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Washington, DC, USA: Council on Library and Information Resources. ISBN 1-887334-63-7.

[13] NISO Framework Advisory Group. (2007). A Framework of Guidance for Building Good Digital Collections, 3rd edition, p. 57.

[14] National Initiative for a Networked Cultural Heritage. (2002). NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials

[15] CCSDS.org - CCSDS Recommendations and Reports - All Publications: http://public.ccsds.org/publications/AllPubs.aspx

[16] ISO 14721:2003 Spec: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683

[17] Consultative Committee for Space Data Systems. (2002). Reference Model for an Open Archival Information System (OAIS). Washington, DC: CCSDS Secretariat, p. 3-1

[18] DPE Project: www.digitalpreservationeurope.eu

[19] SHAMAN Project: shaman-ip.eu/shaman

[20] PLANETS Project: www.planets-project.eu

[21] Open Planets Foundation: www.openplanetsfoundation.org

[22] CASPAR Project: www.casparpreserves.eu

[23] NESTOR Project: www.langzeitarchivierung.de/eng

[24] APARSEN Project: www.alliancepermanentaccess.org

[25] ENSURE Project: ensure-fp7-plone.fe.up.pt/site

[26] SCAPE Project: www.scape-project.eu

[27] PROTAGE Project: www.protage.eu

[28] DIGICURV Project: http://www.digcur-education.org/

[29] KEEP Project: www.keep-project.eu

[30] National Digital Stewardship Alliance Project: www.digitalpreservation.gov/ndsa

[31] Wepreserve initiative: www.wepreserve.eu

[32] EDL Project: http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/

[33] Europeana Project: http://www.europeana.eu/portal/

[34] D3.4 'Assessment of requirements for persistent identification of objects, collections and institutions' and D3.5 'Technical and policy infrastructure to support persistent identifiers' are available at http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents. Both deliverables examine persistent identifiers by multiple points of view, with particular regards to the requirements for the successful implementation of persistent identifiers in the CHS, the technical infrastructure needed for PIDs, and the policy for the management of PIDs.

[35] Document Object Identifier system: http://www.doi.org

[36] Archival Resource Key: http://tools.ietf.org/html/draft-kunze-ark-14

[37] Handle System: http://www.handle.net/

[38] Persistent URL: http://purl.oclc.org

[39] Library of Congress Control Number: http://www.loc.gov/marc/lccn.html

[40] IETF RFC 3188 Using National Bibliography Numbers as Uniform Resource Names: http://tools.ietf.org/html/rfc3188

[41] Conference of European National Librarians: http://web3.nlib.ee/cenl/

[42] Nordic Metadata Projects: http://www.kansalliskirjasto.fi/extra/muut/meta/

[43] IETF RFC 2141 URN Syntax http://tools.ietf.org/html/rfc2141

[44] Section written thanks to the contributions of Maurizio Messina, expert from the Italian Ministry.

[45] This definition is based on:

a)*Trustworthy Repositories Audit & Certification (TRAC)* http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf (for the concept of "trusted digital repositories"*)*;

b) Luciana Duranti, *Un quadro teorico per le politiche, le strategie e gli standards di conservazione digitale: la prospettiva concettuale di InterPARES,* <<Bibliotime>>, 9(2006), 1

http://didattica.spbo.unibo.it/bibliotime/num-ix-1/duranti.htm (to assess the *authenticity* of a digital resource, the *public service* must be able to establish its *identity* and demonstrate its *integrity*)

c) PREMIS 2.0, 2008 , PREsevation Metadata: Implementation Strategies, http://www.loc.gov/standards/premis/ (for the concepts of "viability, renderability, understandability, authenticity, identity");

d) ***OAIS**. Reference model for an Open Archival Information System,* ISO 14721:2003 (for the concept of *archive* and *designated community:* "an organization that intends to preserve information for access and use by a designated community). See also the OAIS Bluebook: http://public.ccsds.org/publications/archive/650x0b1.PDF

[46] Serial Advanced Technology Attachment: http://it.wikipedia.org/wiki/Serial_ATA

[47] Rsync software: http://it.wikipedia.org/wiki/Rsync

[48] Redundant Array of Independent Disks: http://it.wikipedia.org/wiki/RAID

[49] ISO/IEC 27001:2005 "specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining and improving a documented Information Security Management System within the context of the organization's overall business risks"

[50] Colocation centre: http://en.wikipedia.org/wiki/Colocation_centre

[51] Digital Repository Audit Method Based on Risk Assessment: http://www.repositoryaudit.eu/

[52] *Trustworthy Repositories Audit & Certification (TRAC)* http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

[53] ISO 14721-2003 Spec: http://public.ccsds.org/publications/archive/650x0b1.PDF

[54] IISO 28500:2009 "specifies the WARC file format: to store both the payload content and control information from mainstream Internet application layer protocols, such as the Hypertext Transfer Protocol (HTTP), Domain Name System (DNS), and File Transfer Protocol (FTP); to store arbitrary metadata linked to other stored data"

[55] ISO/IEC 21000-2:2005: "The Digital Item Declaration Model describes a set of abstract terms and concepts to form a useful model for defining Digital Items [...], is based upon the terms and concepts defined in the above model. It contains the normative description of the syntax and semantics of each of the DIDL elements, as represented in XML".

[56] Schema is used here as http://www.w3.org/XML/Schema: "XML Schemas express shared vocabularies and allow machines to carry out rules made by people"

[57] Linked Data: http://www.w3.org/DesignIssues/LinkedData.html, http://en.wikipedia.org/wiki/Linked_data

[58] *Trustworthy Repositories Audit & Certification (TRAC)* http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

[59] Turkish National Library: http://www.mkutup.gov.tr/menu/11

[60] Digital Preservation Survey (Annex 1 - 10.6)

[61] GÉANT Project: http://www.geant.net/

[62] The german NREN – DFN (Deutsche Forschungsnetz, http://www.dfn.de/)

[63] The swiss NREN – SWITCH (Serving Swiss Universities: http://www.switch.ch/)

[64] Consortium GARR: http://www.garr.it/

[65] The Turkish NREN – ULAKBIM Turkish Academic Network and Information Center : http://www.ulakbim.gov.tr/eng/

[66] Slovak Research and Academic Network

[67] European Strategy Forum on Research Infrastructures: http://ec.europa.eu/research/esfri/

[68] Exploitation of Meteorological Satellites: http://www.eumetsat.int/Home/index.htm

[69] COLUMBUS Project: http://www.esa.int/esaHS/ESAAYI0VMOC_iss_0.html

[70] education roaming: http://www.eduroam.org/

[71] eduGAIN: http://www.edugain.org/

[72] eduPKI: http://www.edupki.org/

[73] eduCONF: http://educonf.geant2.net/

[74] GÉANT Identity Provider: http://www.geant.net/SERVICES/ENDUSERAPPLICATIONSERVICES/Pages/GIdP.aspx

[75] European Grid Infrastructure: http://www.egi.eu/

[76] Partnership for Advanced Computing in Europe: http://www.prace-ri.eu/

[77] GÉANT2 Project: www.geant2.net

[78] GN3 Project: www.geant.net

[79] Delivery of Advanced Network Technology to Europe: www.dante.net

[80] EGEE Project: www.eu-egee.org

[81] DEISA Project: www.deisa.org

[82] PRACE Project: www.prace-project.eu

[83] ALICE Project: alice.dante.net

[84] ALICE2 Project: alice2.redclara.net

[85] EUMEDCONNECT Project: www.eumedconnect.net

[86] EUMEDCONNECT2 Project: www.eumedconnect2.net

[87] GÉANT2-ERNET Project: global.dante.net/server/show/nav.1416

[88] ORIENT Project: global.dante.net/server/show/nav.1418

[89] South African National Research and Education Network: www.meraka.org.za/sanren.htm

[90] South-Eastern European Research and Education Network: www.seeren.org

[91] South-Eastern European-FIRE Project: www.seefire.org

[92] Trans-Eurasia Information Network 2: www.tein2.net

[93] Trans-Eurasia Information Network 3: www.tein3.net

[94] Ubuntunet Alliance Project: www.ubuntunet.net

[95] E-Infrastructure shared between Europe and Latin America: www.eu-eela.org/first-phase.php

[96] E-Infrastructure shared between Europe and Latin America 2: www.eu-eela.eu

[97] EUAsiaGrid Project: www.euasiagrid.eu

[98] EUChinaGRID Project: www.euchinagrid.eu

[99] EU-IndiaGrid, EU-IndiaGrid2 Projects: www.euindiagrid.eu

[100] EUMEDGRID Project: www2.eumedgrid.eu

[101] EUMEDGRID-Support Project: www.eumedgrid.eu

[102] South African National GRID: www.sagrid.ac.za

[103] South-Eastern European GRID: www.see-grid.org

[104] South-Eastern European GRID2: www.see-grid.eu

[105] South-Eastern European GRID-SCI: www.see-grid-sci.eu

[106] Asia-Pacific Advanced Network: www.apan.net

[107] CERN Large Hadron Collider: www.cern.ch/lhc

[108] International Grid Trust Federation: www.igtf.net

[109] Academia Sinica Grid Computing: www.twgrid.org/en/index.php?option=com_content&task=view&id=27&Itemid=153

[110] China Science and Technology Network: www.cstnet.net.cn/english/index.htm

[111] China Education and Research Network: www.edu.cn/english_1369/index.shtml

[112] National Knowledge Commission: www.knowledgecommission.gov.in

[113] Indian National Grid Initiative GARUDA: www.garudaindia.in

[114] Indian National Research and Education Network: www.eis.ernet.in

[115] Indian Grid Certification Authority (IGCA): ca.garudaindia.in

[116] Cooperación Latino Americana de Redes Avanzadas: www.redclara.net/index.php?lang=en

[117] AugerAccess Project: www.augeraccess.net

[118] Pierre Auger Southern Cosmic Ray Observatory: www.auger.org

[119] Mercosur: www.mercosur.int

[120] Proyecto Mesoamérica: www.planpuebla-panama.org

[121] Arab Scientific Research and Education Network: www.asrenorg.net

[122] EUGridPMA Project: www.eugridpma.org

[123] EPIKH Project: www.epikh.eu

[124] Greek Research & Technology Network: www.grnet.gr

[125] Worldmapper - Science Research: www.worldmapper.org/display.php?selected=205

[126] FEAST Project: www.feast-project.eu

[127] AfricaConnect Project: www.africaconnect.eu

[128] IST-Africa Project: www.ist-africa.org

[129] EuroAfrica-ICT Project: www.euroafrica-ict.org

[130] eI-Africa Project: www.ei-africa.eu

[131] Karoo Array Telescope: www.kat.ac.za

[132] Centre for High-Performance Computing: www.chpc.ac.za

[133] gLite Project: www.glite.org

[134] GILDA t-Infrastructure: https://gilda.ct.infn.it

135 Patrimoine numérique: www.numerique.culture.fr .

[136] MICHAEL Project: http://www.michael-culture.org/

[137] French ministry's recommendations and standard specifications for each type of document / format: http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_04.htm

[138] http://references.modernisation.gouv.fr/sites/default/files/RGI_Version1%200.pdf.

139 General Interoperability Reference Framework: http://culturelabs.culture.fr/

[140] MINISTERE DE LA CULTURE ET DE LA COMMUNICATION. CONSERVATION A LONG TERME DES DOCUMENTS NUMERISES. http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/conservation.pdf

[141] SPAR Project: http://www.bnf.fr/en/professionals/preservation_spar/s.preservation_SPAR_presentation.html

[142] National Telecommunications Network for Technology, Education and Research: http://www.renater.fr/

[143] Integration of Data Interconnection Services on Research and Education: http://www.rechercheisidore.fr

[144] Slovenian NREN (ARNES): http://www.arnes.si/

[145] GISKD Project: http://giskd.situla.org

[146] ATHENA Project: http://www.athenaeurope.org/

[147] NUK Project: http://www.nuk.uni-lj.si/

[148] KeyToNature Project: http://www.keytonature.eu

[149] Trubar Project: http://www.sio.si/

[150] Spanish NREN (RedIRIS): http://www.rediris.es/

[151] Biblioteca Abadía de Montserrat: http://www.abmontserrat.es/

[152] Biblioteca Nacional: http://www.rediris.es/rediris/instituciones/www.bne.es

[153] Biblioteca Pública Central de La Rioja: http://biblioteca.larioja.org/

[154] Biblioteca de Castilla y León: http://www.bcl.es/

[155] Biblioteca de Cataluña: http://www.bnc.es/

[156] Consorci de Biblioteques Universitàries de Catalunya: http://www.cbuc.es/

[157] Consorcio de Bibliotecas Universitarias de Galicia: http://www.bugalicia.org/

[158] Real Biblioteca (Patrimonio Nal.): http://www.patrimonionacional.es/

[159] Museo Arqueológico Nacional: http://man.mcu.es/

[160] Museu d´Història de la Medicina de Catalunya: http://www.comb.cat/

[161] Consorci de les Drassanes Reials i Museu Marítim de Barcelona: http://www.diba.cat/

[162] Museu d´Història de la Medicina de Catalunya: http://www.comb.cat/

[163] Real Academia Española: http://www.rae.es/

[164] Real Academia de Ciencias Exactas, Físicas y Naturales: http://rac.es/0/0_1.asp

[165] Real Academia de Ciencias y Artes de Barcelona: http://www.racab.es/

[166] Institut Cartogràfic de Catalunya: http://www.icc.cat/

[167] Institut d'Estudis Catalans: http://www.iec.es/

[168] Instituto Arqueológico Alemán: http://www.dainst.org/

[169] Instituto Cervantes: http://cervantes.es/

[170] RedIRIS: www.rediris.es

[171] Red Informática Científica de Andalucía: www.cica.es/comu/rica.es.html

[172] Rede de Ciencia e Tecnoloxía de Galicia: www.cesga.es/ca/defaultC.html

[173] Comunidad Autónoma de Madrid: www.madrimasd.org/RedTelematicaMadrid

[174] Red Canaria Académica de Recursos de Información Avanzados: www.rcanaria.es

[175] Red de Ciencia, Tecnología y Sociedad de la Información de la Región de Murcia: www.f-integra.org/

[176] Red de Ciencia y Tecnología del País Vasco: www.i2basque.es

[177] Barcelona Supercomputing Center–Centro Nacional de Supercomputación: www.bsc.es

[178] Centro Extremeño de Investigación, Innovación Tecnológica y Supercomputación: www.cenits.es

[179] Centro de Supercomputación y Visualización de Madrid: www.cesvima.es

[180] Centro de Supercomputación de Galicia: www.cesga.es

[181] Centro de Supercomputación de Murcia: www.cesmu.es

[182] Centro Informático Científico de Andalucía: www.cica.es

[183] Fundación Centro de Supercomputación de Castilla y León: www.fcsc.es

[184] List of Web Archiving Initiatives worldwide: http://en.wikipedia.org/wiki/List_of_Web_Archiving_Initiatives

[185] Internet Archive Project: http://www.archive.org/

[186] International Internet Preservation Consortium (IIPC): http://www.netpreserve.org/

[187] ONDARENET Project: http://www.euskadi.net/ondarenet

[188] PADICAT Project: http://www.padicat.cat/en

[189] Centre de Serveis Científics i Acadèmics de Catalunya (Center for Scientific and Academic Services of Catalonia): http://www.cesca.es/en

[190] Hispana Project: http://hispana.mcu.es/en/estaticos/contenido.cmd?pagina=estaticos/presentacion

[191] Red Digital de Colecciones de Museos de España (collective catalogue of the Digital Network of Spanish Museum Collections): http://ceres.mcu.es/pages/SimpleSearch?index=true

[192] List of digitisation projects carried out in Spain: http://hispana.mcu.es/en/comunidades/directorio.cmd

[193] MDC Project: http://mdc.cbuc.cat/

[194] "Digitisation standards: minimum requirements". CBUC. http://hdl.handle.net/2072/97495

[195] CONTENTdm Digital Collection Management Software: http://www.contentdm.org/

[196] OAIster Project: http://oaister.worldcat.org/

[197] Calaix Project: http://calaix.gencat.cat/

[198] Turkish Academic Network and Information Centre ULAKBIM - National Academic Network (ULAKNET) Unit (Turkish NREN): http://www.ulakbim.gov.tr/eng/

[199] ULAKBİM users: http://www.ulakbim.gov.tr/eng/users.uhtml.en

[200] ULAKBİM objectives: http://www.ulakbim.gov.tr/eng/missions.uhtml.en

[201] ULAKBİM fields of activity: http://www.ulakbim.gov.tr/eng/ulaknet/

[202] Mark Twain: *Tom Sawyer Abroad*, 1st World Library – Literary Society, Fairfield, 2004, p. 29-30. (Emphasis added by me: P.M.)

[203] See for example: Yoshiyuki Tamura: Rethinking Copyright Institution for the Digital Age, *WIPO Journal*, 2009/1: p. 66-68.

[204] See: Mezei, Péter: Digital Technologies - Digital Culture, *Nordic Journal of Commercial Law*, Issue 2010/1, p. 9-13. Available at http://www.njcl.utu.fi/1_2010/mezei_peter.pdf.

[205] Mtima, Lateef - Jamar, Steven D.: Fulfilling the Copyright Social Justice Promise: Digitizing Textual Information, *New York Law School Law Review*, 2010/11: p. 103.

[206] On the Google Books Project see especially: Jonathan Band: The Long and Winding Road to the Google Books Settlement, *John Marshall Review of Intellectual Property Law*, 2010: p. 227-329.

[207] Authors Guild v. Google, Inc., *No. 05 Civ 8136* (S.D.N.Y. Sept. 20, 2005) and McGraw-Hill Cos., Inc. v. Google, Inc., *No 05 Civ 8881* (S.D.N.Y. Oct. 19, 2005).

[208] Available at: http://www.googlebooksettlement.com/intl/en/Settlement-Agreement.pdf.
[209] Available at: http://books.google.com/booksrightsholders/agreement-contents.html.
[210] The Authors Guild, et al., v. Google, Inc., 05 *Civ*. 8136 (S.D.N.Y.), Opinion, 22 March, 2011.
[211] Katharina De La Durantaye: H is for Harmonization: The Google Book Search Settlement and Orphan Works Legislation in the European Union, *New York Law School Law Review*, 2010/11: p. 167-168.
[212] Viviane Reding: *Digital Europe – Europe's Fast Track to Economic Recovery*, The Ludwig Erhard Lecture 2009, Lisbon Council, Brussels, 9 July 2009: p. 9.
[213] *Report of the 'Comité des Sages': The New Renaissance*. Available at: http://ec.europa.eu/information_society/activities/digital_libraries/doc/reflection_group/final-report-cdS3.pdf. (Further: *New Renaissance*.)
[214] Neelie Kroes: *Addressing the orphan works challenge*, Speech 11/163, March 10, 2011. Available at: http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/163&format=HTML&aged=0&language=EN&guiLanguage=en.
[215] *Proposal for a Directive of the European Parliament and of the Council on certain permitted uses of orphan works*, Brussels, 24.5.2011, COM(2011) 289 final, 2011/0136 (COD). Available at: http://ec.europa.eu/internal_market/copyright/docs/orphan-works/proposal_en.pdf. (Further: *Proposal*.)
[216] *New Renaissance*, 5.3.2.2.ii).
[217] *Proposal*, Art. 1(2).
[218] Hendrik Kafsack: Für Zusammenarbeit mit Google Books, 11. Januar 2011, *FAZ.net*. Available at: http://www.faz.net/artikel/C31158/eu-fachleute-fuer-zusammenarbeit-mit-google-books-30324081.html.
[219] *New Renaissance*, 5.3.2.2.vi).
[220] Compare to *Proposal*, Art. 6. and 7.
[221] *Proposal*, Art. 7(1) 4-5.
[222] *New Renaissance*, 5.3.2.2.viii).
[223] *Proposal*, Art. 2(1), Art. 3(4), Art. 6(4), Art. 7(1) 2., and further recitals (13) and (15).
[224] The seven conditions of the participation of private partners in the cultural preservation imposed by the Comité des Sages are listed in New Renaissance, 9.3.5. Further, the Comité recommended that all of the future public-private partnership agreements shall last until a maximum of seven years. See: New Renaissance, 9.3.6.
[225] *Proposal*, Art. 6(3) and recital (18).
[226] DPE. Digital Preservation introduction: http://www.digitalpreservationeurope.eu/what-is-digital-preservation/
[227] European Commission - Research Infrastructures Programme - e-Infrastructures activity: http://cordis.europa.eu/fp7/ict/e-infrastructure/