026500

BELIEF
Bringing Europe's eLectronic Infrastructures to Expanding Frontiers

WP3D9
Technical & Evaluation Report of
1st BELIEF Brainstorming Workshop - 11 & 12 April 2006 – Linz, Austria
"Future Knowledge Infrastructures - Scientific Repositories and Content Management over the GRID"

Version 0.8
September 25th, 2006

Start date: 1 November 2005 Duration: 24 months

The BELIEF Consortium consists of:
1. Metaware SpA, IT
2. CNR- ISTI, IT
3. University of Athens, GR
4. ERNET, IN
5. ENGITECH, IE
6. Escola Politécnica da Universidade de São Paulo, PUSP, BR
7. Wisconsin-Madison University, US

Responsible Author:
Alex Delis, Eleni Toli, Sokratis Pantazaras Univ. of Athens and Franco Zoppi - CNR-ISTI,
Contributors:
- Stephen Benians Metaware S.p.A., IT

Status:		**Confidentiality:**		
[ ]	Draft	[X]	- Public	- for public use
[ ]	To be reviewed	[ ]	- PP	- restricted to other programme participants (including the Commission Service)
[ ]	Proposal	[ ]	- RE	- restricted to a group specified by the consortium (including the Commission Service)
[X]	Final / Released to CEC	[ ]	- CO	- Confidential, only for members of the consortium (including the Commission Service)

Summary / Content:
WP3D7.1+9.1-T3.2 – Technical & Evaluation Report of the 1st BELIEF Brainstorming Meeting. 11 & 12 April 2006 – Linz, Austria. *"Future Knowledge Infrastructures - Scientific Repositories and Content Management over the Grid"*

Contents

Belief
Bringing Europe's eLectronic Infrastructures to Expanding Frontiers

1. Executive Summary

Despite the fact that institutional, regional, and community-based repositories exist and are dispersed throughout Europe, they generally lack in providing content services for dissemination information at a large scale. The scientific community now strongly believes that it is time to deal with the open problem of Content Management in an effective manner. As data, experiences and knowledge produced is on the increase, there is an urgent need for efficient Scientific Repositories that, together with Content Management will constitute the global Knowledge Infrastructures. It is anticipated that such Knowledge Infrastructures will further push the European scientific inquiry and spur business. The GRID infrastructure is expected to play a key supporting role in this development as it may offer abundant of computing and storage options.

These themes were discussed in a Brainstorming Workshop organized by the BELIEF Project, entitled "Future Knowledge Infrastructures: Scientific Repositories and Content Management over the Grid", that took place in Linz, Austria on April 11th-12th, 2006, directly after the open eInfrastructure Reflection Group workshop, which was organised on April 10th and 11th at the same location.

According to the mission of e-IRG, as it was drafted in Rome on December 10, 2003, e-IRG supports on the politically, advisory and monitoring level the creation of policy and administrative framework for the easy and cost-effective shared use of electronic resources in Europe (focusing on Grid-computing,data storage and networking resources) across technological, administrative and national domains. Common objectives and goals pursued by both, BELIEF and e-IRG, were the main reason for organising the workshop in Linz.

The Workshop was by-invitation only and engaged experts from the areas of scientific repositories, digital libraries, and the Grid in informal discussions and exchange of ideas on three core issues pertinent to the development of Knowledge Infrastructures, namely:

1. **Technical problems**: Knowledge Infrastructures call for the synergy of a number of technical areas and pertinent technologies including content structure and handling, digital libraries, data management, distributed computing and service provision all working atop the GRID. Large scale deployment of Knowledge Infrastructures on a continental and international level has not occurred to this point.
2. **Organizational problems**: The demarcation of actors, their role, and their interactions have to be clearly addressed. Moreover, vital issues regarding location of content, presence of storage elements, management of individuals and groups, access rights and models, ownership of content, and business models for ensuring sustainability need to be examined.
3. **User-oriented problems**: Content providers, owners of computing resources and of course end-users are expected to be shareholders of the European-wide Knowledge Infrastructure. Identifying constituent communities and offer for those appropriate services in the presence of diverse types of content are major issues to be tackled.

This document outlines the workshop organizational effort and provides a report on its activities that brought into foreground a large number of issues that have to be addressed in the context of European Knowledge Infrastructures.

Chapter 2 presents the workshop objectives while Chapter 3 offers an overview of the organization issues. Chapter 4 and 5 outlines the invited presentations and follow-up discussions that took place; the three abovementioned sets of issues were focussed upon with three groups of experts working in parallel to generate the technical outcome. Finally, Chapter 6 outlines the major outcomes, recommendations and follow-up steps of the Workshop.

2. Introduction

The main objective of **1st Brainstorming BELIEF Workshop** was to bring together experts in the areas of eInfrastructures, Grid computing, Digital Libraries, and Management of Data, along with selected IT professionals, researchers, and project managers to consider and debate the development of Scientific Repositories and Content Management in terms of technical, managerial, legal, financial, and user issues. The combination of Content Management and Scientific Repositories jointly termed as Knowledge Infrastructures (KIs) are expected to play a critical role in the continuity of European leadership in the area of e-Infrastructures.

The workshop was organized as a forum for experts to put forward key questions for the development of KIs and debate them in a collegiate manner. As such, the meeting proved to be very successful as it was very well attended and its informal setting provided the required environment for a critical and exciting exchange of diverse views.

the following section contains the position statements that helped the workshop participants focus on specific areas of concern. The objectives of the BELIEF project and its anticipated outcomes are succinctly presented at the end of this document .

2.1. The Position Statement of the Workshop

The idea of establishing Europe-wide, or even world-wide, networks of content repositories that will serve as infrastructures for scientific (and possibly other, even daily life) activities across the continent and the rest of the world is very exciting. Electrical power and phone services are available world-wide in a structured fashion, analogous computing services are forthcoming with all the current efforts on the Grid, but when it comes to information, the only existing counterpart is a chaotic, unstructured, offering-no-guarantees web. Institutional, regional, and community-based repositories do exist currently and are operational with great success, but their applicability and availability is rather limited. It is about time that reliable content-provision services are offered at a large scale, in the context of global Knowledge Infrastructures.

Any effort to establish such infrastructures has to address several problems, which can be broadly put into three main categories: technical, organizational, and user-oriented problems.

1. Technical problems: For the most part, the technological components required for large-scale Knowledge Infrastructures exist already, i.e., technology for organization, storage, search, retrieval, and preservation functionality is mature and available. Nevertheless, there are several additional issues that require attention. For example, existing technology has not been tested at an international scale, where content is truly distributed across a large number of locations. Also, large-scale Knowledge Infrastructures should be developed on top of the corresponding computing and network infrastructures, e.g., currently established Grids, but the latter's support for information processing has not yet matured. Furthermore, several desirable services take on a new form and raise a host of new challenges when applied at such a large scale, e.g., personalization, annotation, collaboration, multilinguality.
2. Organizational problems: Before any deployment, the operational model of European-wide Knowledge Infrastructures must be studied and decided upon. The failure probability of a venture of such magnitude is very high, so several issues need to be addressed: location of content storage (it should remain at its original creation site but this may not be enough),

actors and roles in the system, access rights of the individual and institutional participants, responsibilities of classification of incoming content, IPR management, legal foundations of the endeavor, potential billing strategies (e.g., "internal" vs. "external" users), centralized or distributed administration unit(s) supporting the system, and various organizational polices.

3. User-oriented problems: Any effort towards a European-wide Knowledge Infrastructure will fail if it does not serve the needs of its primary stakeholders, i.e., the content providers and the end users. The former may range from authorized individuals to large institutions (e.g., universities or public libraries) and media industry (e.g., broadcasting corporations), whose rights should be protected and whose benefits should be clearly identified. They would be depositing content that may range from very informal to very formal: from white papers and course projects, to technical reports and theses (undergraduate, graduate, PhD), conference and journal publications, manuscripts and books, images, videos, raw and curated scientific datasets, and others. Usefulness of this material will be most critical for successful adoption of any Knowledge Infrastructure by the end users. Hence, ensuring the quality of the material provided and offering an enriched set of services that capture users' needs that are truly missing from the web will be equally important. Identifying the appropriate services for different kinds of content and different communities of end users is a major issue.

3. Workshop Organizational Aspect

The Workshop was by-invitation only and experts from the areas of e-Infrastructures, digital libraries, data management and GRID-computing participated. The goal of the 1st BELIEF Workshop was to establish the state of the art and offer directions for drafting a research and development agenda on the technological, organizational, and user-oriented issues for Knowledge Infrastructures.

3.1. Inviting Attendees

Approximately 30 researchers, technologists and users were invited to participate in the workshop, exchange their ideas, discuss about challenges and propose future steps. The main criterion for the compilation of the invitees list was to select a diverse mix of people, with different technological and scientific backgrounds but either with pertinent scientific know-how or with long standing industrial experience. Thirtythree high-quality participants attended. Finally, it was a privilege to have in session three EU officers who helped with their points and perspective to better calibrate and focus the discussions. The provenance of the participants justifies the European character of the workshop, as participants from twelve different European countries (EU member states, New Member States and Non EU States) were present.

Regarding the participants profile, the major part (28%) came from the Technology Industry. The second largest group is formed by participants from Research / Scientific Institutions with 17 %. All the other groups (Library/Archive, Media Organization, European Commission and Others) were equally represented with 11 % each. The above figures illustrate not only the diverse backgrounds of the participants but also the diversity of approaches during the workshop

4. Views of Participants and Invited Talks

4.1. Position Papers

The consortium solicited short position papers from the invitees. The following offers an overview of individual statements that help form a context for the workshop sessions.

Peter Buneman, University of Edinburgh, UK

Scientific Repositories. The challenges of preserving knowledge.

Let me start with a deliberately contentious statement: institutional repositories are dangerous. Consider the huge loss to both our cultural and scientific heritage that was caused by placing it in "repositories" such as the libraries of Alexandria, Cotton, Louvain,. . . (the list is surprisingly long.) Most of our knowledge has been preserved by copying rather than by archival institutions. The analogy holds for our digital resources, as we increasingly entrust our scientific knowledge to centralized databases. Although we are making it more accessible, we are degrading its quality, making it less usable in the long term, and its survival is threatened.

This danger is particularly apparent in curated databases. These are databases that are constructed with a great deal of human effort. They are replacing the reference shelves of libraries – the dictionaries, gazetteers, indexes, encyclopedias. More interestingly, they are replacing – in some sense – scientific publication. In the field of molecular biology alone there are upwards of 700 on-line databases; moreover some of these databases represent an extraordinary amount of human effort. One of these databases has over 100 biologists involved in the creation, annotation and correction of the data. How do we ensure the longevity and accessibility of this costly data? How do we ensure that these databases can exchange data and cross-reference each other in a stable fashion? Most of all how do we assess the quality and validity of data in these databases? All these questions pose fundamental challenges for database research, and until we can make some progress with these issues, it is surely premature to design the "infrastructure" of scientific repositories.

Here are some of the technical challenges that confront us:

- **Data transformation**, exchange and integration: This is a relatively old topic in database research, which is crucial to curated databases. While low-level tools such as query languages that can talk to a variety of data formats and databases are now well-developed, declarative techniques for integration and transformation based on schema annotation and manipulation have been slow to come to market. The emergence of XML as a universal language for data exchange may have simplified some low-level problems, but has needlessly complicated some of the higher-level issues.
- **Data Citation**: How do we cite something in a curated database? One of the time-consuming tasks of a curator is to verify citations. As yet there are no standards for citing components of a database.
- **Annotation**: This is a growing area of activity in scientific databases. Some biological databases describe themselves as "annotation databases", and annotation is starting to become a form of scientific communication. However database technology provides little support for annotation.
- **Provenance**: Curated databases are often derived from other databases. Their value lies in the organization and annotation of data. When you see something in such a database, how

do you know where it has come from? This knowledge is a vital part of data quality, but in most databases systems it either lost or recorded in a variety of ad hoc forms.

- **Preservation**: How does one preserve a database? How does one keep it in a form that is independent of a specific data management system, and how does one preserve the history of something that changes frequently.

These are a few of the problems created by curated databases. There are many other areas such distributed computing and programming languages which can contribute. In addition, even if all the technical problems are solved, the social, legal and economic problems with web data are enormous.

Chris Chambers, BBC, UK

Content Management in Media Production

The current developments within news and entertainment media production are putting major pressures on infrastructure designs and management. Traditional approaches do not cater well for the proliferation of delivery platforms, the combination or audio, video and large amounts of descriptive data (often called "rich media") and the interactivity with users and delivery applications.

Media production structures based on IT networks are becoming commonplace, however there are a number of challenges that arise.

- The ability of such structures to **scale up** to interact with large audiences that can easily be achieved with media broadcast services.
- The vital need to maintain **security** against attacks and preserve personal and business privacy.
- The underlying need to **control costs** and allow for both public services and a fully cost recovered business model.
- The ability to expand the production of "**rich media**" content in a flexible and scalable way. Production processes often have to deal with linear (un-compressed) separate audio and video media along with handling live and near live content.

These issues place considerable challenges on performance of firewalls, storage and networks both within the production process and with the delivery to content audience.

Keith Jeffery, CLRC RAL, UK

Challenges for Content Management

Worldwide e-infrastructure provides many challenges. Significant ones are:

(a) **reducing the threshold barrier for the end-user** : if it is not available and easy to use anytime, anyhow, anywhere for variously-abled people in a multilingual / multimedia / multimodal fashion it will not be used;

(b) **ensuring sustainability**, availability, continuity and scalability: if the required services are not available 365/7/24, if the system does not provide scalable performance with increased demands as seen by the individual user it will not be used. An associated factor is cost; if the system costs too much to maintain / tune then it will fall into disuse;

(a) ensuring appropriate **trust including security, privacy and openness**: if the services and associated resources are not trusted by the end-user or an organisation then the system will not be used. This implies security systems and privacy protection but also transparency in the sense of freedom of information and in the sense of public scrutiny of auditable business.

These three aspects could be characterised as (a) semantics (b) autonomics (c) trust, although the concepts contain much more than these words imply.

Proposed Solution

The solution appears to lie with a service-oriented architecture where the services are 'intelligent' so allowing dynamic and flexible composition of quality-engineered software components with re-composition during runtime depending on non-functional requirements (such as performance, security, business continuity) in a 'genetic algorithm' sense but applied to modular services.

The underlying technological solution relies on metadata describing these services (for service discovery and composition) and also available resources (which may be wrapped by services). The metadata also describes – eg by constraints – the rights associated with the service or resource such as privacy, copyright, access rights, charges for access. The metadata must be formalised such that it is machine-understandable as well as machine-readable (else it is neither scalable nor interoperable).

A particular group of services will interact intelligently with the end-user to ensure understanding of the request to the system. Other services will wrap resources so allowing intelligent interaction and participation in composed service provision (e.g. in providing homogeneous access to heterogeneous repositories or facilitating interoperating business processes).

Acknowledgments

The ideas presented here, while the responsibility of the author, owe much to my colleagues in CCLRC, in ERCIM, in the W3C community and in the EC NGG Expert Group.

Peter Kunszt, Swiss National Supercomputing Center, Switzerland

Comments on the BELIEF Position Statement

Content provisioning and knowledge management is a very challenging topic, technically as well as legally and sociologically. As explained in the position statement provided for the workshop, it is essential that both the content providers and the consumers see clearly the benefits of such a service. By limiting the scope and consequently the content, we could address the problem at hand in a step-wise manner. Content that is already in the public domain, i.e. royalty-free and without charge is the first obvious candidate to start working on. These are public libraries and publications, websites, blogs, research databases, etc. Some projects have already started to involve this knowledge base in the searchable context of the web and the grid.

However, this is probably not the right thing to do as most of the technical and legal difficulties to be solved are for knowledge and content that is not free of charge. Excluding such content upfront with the promise to include it later will not work as experience has shown us in other contexts involving security and safety. Solving issues of security, accounting, billing, protection against piracy and fraud, etc, cannot be done 'later'. These concepts cannot be orthogonalized, they are pervasive and have to be designed into the system upfront, in a coherent and extensible way.

The most difficult problem here is to reach a consensus, a standard on how the content is to be protected and organized. The latest example of the BluRay disk's digital right management implementation shows that it was not the technology that delayed its release but standardization and reaching of agreements concerning content protection among the vendors of the BD and the providers of content (the movie and music industry). Considering that the interests of the consumers were largely neglected in the BD DRM (and therefore some suggest it will not be a market winner), a process that wants to standardize on a content protection format that suits both producers and consumers is a very ambitious task.

In addition to the security and protection aspects there are also data management issues to be addressed: The technology to serve up content is better explored. The model Google has introduced for the indexing of the web shows us the way to go, however, Google has the huge advantage that all the distributed data centers it owns are fully controlled by it and that it is therefore a homogeneous system, looking the same from every site of the world. And the web is, in today's standards, not such a big dataset to index over. Additional content coming from other sources would introduce a lot of heterogeneity and a large additional set of data. Scaling to that size may be easier than solving the data content heterogeneity, which is what the semantic web community has tried to address in the past decade.

Considering all this I suggest to select the problem to be solved very carefully, as the danger is large of choosing a well-known and already solved problem or a potentially too difficult problem to solve with today's methods.

Diego Lopez, RedIRIS, Spain
In our experience, one of the key issues in providing a successful service of this nature is the possibility of integrating different sources of data and identity through federated approaches.

In what relates to data, it is important to bear in mind that common content repositories must be coordinated with other (more local) sources of information. Researchers, as data producers, are keen to use local repositories to store the information they produce. The term local is not used here in a geographical sense, but rather with a wider meaning: local repositories may be based on common areas of interest, well structured (virtual) communities, professional reference sites, national associations, etc. Integrating these sources of information implies some risks:

- Data partition because of different syntax and semantics.
- Growth and maintenance problems because of incompatibility of protocols.
- Access limitations because of security measures.

To alleviate this problem, a federated model of data source integration can be applied. In this federated model, the aggregated repository collects data from the local sources that maintain their independence. As a common language, the Resource Description Framework (RDF) looks like the obvious choice. Several tools are already available for this using RDF in both push (UDS, OM) and pull (Jsearchy) modes. These tools could play a significant role in the development of KIs.

Solutions based on federated identity management allow users to access their local authentication system to identify themselves just once and then access different applications, located in different security domains, as their local authentication system sends the appropriate user attributes to each application's authorization points. There exist several of these federated identity management solutions for access control and personalization of Web applications. The most widely deployed in the academic community is Shibboleth. Such solutions could be used in the context of Knowledge Infrastructures and make them Shibboleth-enabled.

Malcolm Read, JISC, UK

Short Position Paper

We share the vision put forward in the Belief Position Statement and recognise that the aim of the first brainstorming meeting is a technology report. However, even at this early stage of your work there are some cultural, organisational and business issues that need to be considered.

Institutional policies and incentives

Why will researchers and academics wish to place their data resources in a repository? Will it be recognised and rewarded by their university or research institute? Will it add to their reputation or enhance their career? These factors need to be understood and institutional practice needs to be shared. Who owns these resources; the researcher, teacher or their employer?

Scope of repositories

I would argue that a repository should house, in a coherent manner, all information resources that an institution wishes to share. For a university this would include learning and teaching materials not just research data and outputs.

Sustainable business model

Short term capital funding to establish repositories is relatively easy to find. How will institutions meet the current costs? Will there be sufficient benefits, for example in administration costs of information management, to justify the on-going costs? Do we need to encourage shared repositories to reduce costs, particularly for smaller institutions?

Risk analysis

There are real uncertainties in achieving the vision set out in the Belief Position Statement. These risks need to be identified, their likely impact assessed and any mitigating actions identified.

4.2. Invited Presentations

This section gives an overview of the invited presentations . The slides of these, as well as those of the speech by Enric Mitjana, representing the European Commission, can be found on the BELIEF Project Website Digital Library (http://www.beliefproject.org/digital-library).

Scientific Repositories: challenges of preserving knowledge

by Peter Buneman (Laboratory for Foundations of Computer Science, School of Informatics, University of Edinburgh)

In the beginning of his presentation Peter Buneman focused on two key issues of scientific repositories: the presentation and the preservation of data. He compared the traditional repositories, e.g. libraries, to the digital ones. Building centralised repositories contains two risks, namely making the data less accessible and threatening its survival. Even traditional repositories, for example the 20th century libraries, did some things better than digital repositories, regarding the use of clear standards for citation, historical records and well understood ownership. The kind of data, either born digital or curated, is another issue that has an impact on the organization and annotation of databases, but also on its cost. Some other crucial topics are citation and archiving. A database should be an XML document if we want it to be "understood". The main conclusions of his presentation are:

- Useful knowledge repositories will support highly distributed, redundant copies.
- The architecture will have to be "conscious" of provenance and annotation.
- It is not at all clear that any current architecture is robust enough (or even better) than the "chaotic, unstructured" web.

The discussion that followed the presentation focussed around permanence, longevity and the distributed nature of digital repositories, but also on how to structure a European repository infrastructure. Especially regarding the European eInfrastructures there are several crucial questions: Would it be possible to see all these data under a single umbrella? Who is going to host this huge scientific repository? Which organization would you trust? Another important issue is the question of simplicity, e.g., we may not be able to reconstruct the different layers of infrastructure if we build layer upon layer and therefore we are taking a risk by creating too much fragility.

Grid Technology for Scientific Repositories - Current State, Perspectives and Issues

by Peter Kunszt (Swiss National Supercomputing Center)

This presentation offered first gave an overview of the state of art of grid technology and existing grid toolkits. Then he moved to challenges for the Grid Technology. He mentioned heterogeneity, data management, security, scalability, robustness and abstraction. In all these different areas, he focused on problematic or often forgotten aspects and on existing weaknesses. But he also gave some success stories, like the Sloan Digital Sky Survey or the Storage Resource Manager Interface. Very useful was his personal view on what the cornerstones for BELIEF could be: accessibility, quality and reliability. His final conclusions were:

- Knowledge and Content Management capabilities exist.
- Homogeneous systems already achieve good results.
- Some of the most difficult aspects, such as security, scalability, robustness still need work and time to mature.
- Choosing the right focus for content management and digital repositories in BELIEF will determine its technical feasibility in the short term.

The main discussion topics were: How can we define grid? Where does the issue of grid come in to scientific repository? How can we build a homogeneous system and actually do we need one? By building a homogeneous system are we maybe running the risk of causing the problems of monoculture? Do we need a generic or a community specific interface? To what extent has industry already picked up the Grid Technology?

Plenary discussion

Both presentations were followed by a lively and fruitful discussion and that was continued even after the scheduled slot. One thread of discussion that took place was on the differences between information and knowledge. Participants attempted to clarify and better understand the fundamental terms and concepts. The EU commission suggested: applying lessons-learned from the open-source community and the benefits one could derive in developing Knowledge Infrastructures. It was noted though that a roadmap is needed for the formation and organization of the European Knowledge Infrastructures.

5. Parallel Sessions

The Parallel Sessions took place on the morning of the second day. Three groups to discuss the main themes of the workshop were formed once preferences and scientific background of attendees were taking into account. In this section, the main issues and presentation points of each group are discussed.

5.1. Build Technology for Scientific Repositories

This session was facilitated and reported by Peter Buneman (University of Edinburgh). The attendees in the working group "Technology Issues" were:

- Patrick Aerts, NLF
- Angelos Bilas, FORTH
- Peter Buneman, University of Edinburgh
- Alex Delis, University of Athens
- Alberto Di Meglio, CERN
- Keith Jeffery, CLRC, RAL
- David Manset, MAAT-G
- Giorgio Occhioni, Engineering
- Heiko Schuldt, University of Basel

The working assumption is that the European and world GRID infrastructure offers virtually unlimited resources for data -and content at large- to be stored and be presented in numerous dynamic forms. Using contemporary e-Infrastructure and its emerging service-oriented mechanisms, it was agreed that new opportunities exist for a leap forward in the areas of data and content analysis, management, provenance, and dissemination on a global scale. Content and data here were envisaged as certainly more complex and "rich" than traditional counterparts in databases, information systems and digital libraries of today. European Scientific Knowledge Infrastructures (KIs) could and should facilitate the creation, management and the evolution of rich content at a global level. If this objective could be relaised in a well defined manner, unprecedented opportunities for business, SMEs, research and the education sectors are sure to follow. In this context, a number of key points were proposed for discussion and debated regarding vital KI build-technology.

Although the precise nature of a Scientific Repository is hard to define in black and white terms, there are some properties that such KIs should comply with. A repository should be:

- independent and self-contained
- easily transported over the GRID e-Infrastructure,
- ,exist for performance,

A repository should have a number of fundamental characteristics such as:

- the provision of persistent handles/identifiers for the objects,
- fundamental search and management services and preferably be monotone. This means data should be never overwritten as is the case in today's digital libraries, database engines and information systems; this characteristic would enable the trace of content evolution within the repositories which is deemed critical.

Repositories are expected to host huge number of digital objects that are of common interest to populous groups of users and/or help in carrying out analyses for R&D and business purposes. The type of such objects will undoubtly be "rich"; they are not only about structured/unstructured data but they will also feature "derived" elements including annotations, partial analyses done over time for segments of data, data history, ownership of sources, citations, and auxiliary information of multi-modal and multilingual nature.

Repositories are expected to be independent at all times and be self-contained. In this respect, it is anticipated that such building KIs blocks do not rely on other or existing services for the provision of data/content, metadata, and historical/administration records. **The notion of content permanence should be weaved into the design of scientific repositories**; once data become part of the storage infrastructure, they should remain available along with their provenance at all times. Moreover what is of key importance here is the self-describing nature of KI data so that they can readily attain independence and flexibility in use. One should be nevertheless aware of the problem of infinite regression as far as the self-describing aspect is concerned. Moreover, there is a core requirement that objects should provide persistent handles with which they could be referenced across the KI realm. Through their linking, handles would offer a viable mechanism to enable accurate provenance recording.

The size, frequency of access and the popularity of objects in repositories are vital issues as they are defining elements of such infrastructure. Repositories should impose no restriction in terms of size and frequency as both big and small entities as well as frequently and rarely accessed content may cause problems in the day-to-day operation of KIs.

In terms of their architecture, repositories could be organized in either a logical or a physical manner. In the former, there is a single organization and/or computing system that ascertains full responsibility for the content and its access while in the latter, multiple parties may contribute portions of data that make up the content in a cooperative way. By and large, modern databases follow the physical organization while digital libraries exploit more the logical style of organization. A related issue is whether replication and of what type should be allowed. In this respect, appropriate policies have to be adopted and the granularity levels of replicated content have to be identified.

Metadata will provide for timely access to very large volumes of content by being organized in multiple tiers, enhancing accessibility by providing multiple different "views" of content, and be ideally created in an automated manner. Scientific repository metadata should be treated in a uniform manner so that navigation and querying yield a feel-alike for users across the Knowledge Infrastructure. Standardization of metadata is expected to play a key role as it can assist in offering the sought "uniformity".

Diverse types of services and the way such offerings can be synthesized to form new ones constitute a major concern and technical challenge for the ultimate acceptance of repositories over the grid Services that will be required in addition to the rudimentary options currently offered by the current e-Infrastructure include:

- Creation, synthesis, addition, management-over-time, and multi-modal content-search and mining services. Although such services call for refinement at this stage, they will serve as the main "interface" to the repositories for both casual and expert users.

- Understanding-the-context services. This specialized breed of services will offer the capability to "view" specific portions by trying to better understand the context and/or needs of users. The context may change over time and can be different for the different types of content as well.
- Autonomous composition of services. Exploiting available and publicized services, new ones could be derived to enhance the access to content/data. Aggregation services could also help pull together resources from multiple repositories and present them to users.
- Administrative Services that deal with the transactional aspects of repositories. Procedures that include user registration for elements of the KIs, handling of trust credentials whenever this is needed, making content/credentials available to others, managing the monotonic aspects of the repositories and establishing security and integrity models.

The performance outlook of repositories is also a key issue. Should one follow the current grid approaches for data processing, it is certain that long responses and operational delays for the end-user will adversely affect the adoption of KIs? It was suggested that focusing on minimal sets of services that are frugal in their computational and resource demands will free the KI operation from delays and overheads. To this effect, QoS characteristics for grid services will have to be enabled.

Access control and security are likewise imporatant. Initially, it was not clear whether repositories should enforce their own access rights protocols and time embargoes or simply such functionality should be relegated to an external mechanism. There are pros and cons for both options as there seems that a trade-off exists between repository independence and operational overheads. The security policies and mechanisms should be analysed in detail as they play a key role in averting malicious attacks and defend content against illegitimate use. One way to achieve security would be to use firewalls among logical segments of grid servers. In doing so, the communications involved are of specific protocols that are permitted to go through and so, repositories and GRID farms that physically contain them could be safeguarded.

In the realization of Scientific Repositories key challenges that confront us are:

- **Data transformation, exchange and integration**: This is a relatively old topic in database research, which is crucial to curated databases. While low-level tools such as query languages that can talk to a variety of data formats and databases are now well-developed; declarative techniques for integration and transformation based on schema annotation and manipulations have been slow to come to market. The emergence of XML as a universal language for data exchange may have simplified some low-level problems, but has needlessly complicated some of the higher-level issues.
- **Data Management**: Storage provisioning is heterogeneous, and online and near-line storage have very different semantics. We have to cater for all those elements to function in a seamless manner.
- **Annotation**: This is a growing area of activity in scientific databases. Some biological databases describe themselves as "annotation databases", and annotation is starting to become a form of scientific communication. However database technology provides little support for annotation.

- **Heterogeneity**: Digital Content is served up by many and diverse sources. Inherently, one have to deal with various types of content, resources, and systems. This will require analysis and provisioning by different GRID servers whose individual worlds have to be unified.
- **Scalability and Robustness**: Large amounts of data are expected to be served and explored. To this effect, failures and disasters should be managed gracefully and in automated manner. Repositories should demonstrate reproducibility of results regardless of the status of the underline middleware.
- **Provenance**: Curated databases are often derived from other databases. Their value lies in the organisation and annotation of data. When you see something in such a database, how do you know where it has come from? This knowledge is a vital part of data quality, but in most databases systems it either lost or recorded in a variety of ad hoc forms.
- **Data Citation**: How can something be cited in curated databases? One of the time-consuming tasks of a curator is to verify citations. As yet there are no standards for citing components of a database.
- **Security**: access policies and their implementation have to be explored on the grid framework. How do we cope with contradictory policies and what are the best options for authorization, auditing and accounting are fundamental issues. Building centralized repositories may ultimately create less accessible content and may threaten its longevity and survival.
- **Preservation**: How does one preserve a database? How does one keep it in a form that is independent of a specific data management system, and how does one preserve the history of something changes frequently.
- **Abstraction and Personalization for users**: individuals are often not interested in the technology serving the content and knowledge and so repositories should provide for such abstractions. Dedicated portals are expected to play a significant role in this direction.

A number of major administrative issues were finally put forward: who does assume the initial repository costs by building this infrastructure? Who is the owner and responsible for its maintenance and continuous preservation? How does one deal with an "anarchic" system of mini-repositories and is this a viable avenue to follow in developing the European Knowledge Infrastructure?

In conclusion, the useful knowledge repositories will support highly distributed, redundant, and available copies of content and should be designed for performance and flexibility while being "conscious" of provenance and annotation. At this stage, it is not at all clear that any current architecture is robust enough (or even better) than the "chaotic, unstructured" web.

5.2. Organizational and Policy Issues

This session was facilitated and reported by Norbert Lossau (University of Bielefeld). The attendees in this working group were:

- Allan Foster, JISC
- Yannis Ioannidis, University of Athens (partially)
- Norbert Lossau, University of Bielefeld

- Andrea Manieri, Engineering
- Carlos Morais Pires, European Commission
- Malcolm Read, JISC
- Eleni Toli, University of Athens

In the early stage of the session, all participants noted the existing divergent trends in the European research arena regarding Knowledge Infrastructures (KIs). At the same time, ongoing efforts cut across multiple organization levels and affect a number of policy issues. More specifically, many institutions intend to or are about to build institutional Knowledge Infrastructures often in parallel to existing national initiatives. This occurs in a number of EU countries as the German GRID Initiative demonstrates. In addition, several new and ambitious GRID-based projects have recently started reinforcing the effort for KIs at a European level.

The working group came into a consensus that vital aspects which affect organizational and policy matters are: actors and stakeholders, the KI-structure, content, organizational and business models. All these have to function within a comprehensive framework. Below, each of these issues are briefly discussed as they are expected to greatly impact the forthcoming European KI-initiative.

The actors and stakeholders of Knowledge Infrastructures have to be clearly identified and their role should be sufficiently clarified early in the process. The working group recognized the following types of actors who are expected to play a significant role in the effort:

- Institutions (Administration and Management)
- Researchers /Academic Teachers
- SMEs and Commercial Players
- National Governments
- Research Funding agencies
- International Organizations
- Public Community

Evidently such a classification of stakeholders will better support research, development and archiving by researchers, teachers, and small medium enterprises that would not have otherwise access to a wide body of "rich" data and knowledge. As most funding is expected to come from individual institutions, consortia of universities and public organizations, the ultimate ownership should rest with the same actors. To this effect, the working group recognized that a more detailed listing of KI stakeholders should be put forward. This listing consists of:

- Institutions: higher education institutions and research organizations
- Researchers: the main peer review and quality assessment taskforce
- Academic teachers
- Content providers and SMEs
- Research and national libraries
- Repositories: existing scientific data e-Infrastructures based on digital objects
- Library communities
- Technical communities such as middleware, interconnectivity (i.e., EGEE) and applications (i.e., Diligent)
- GRID- and E-Research communities

- Commercial technology providers and software companies

All the above parties can contribute by means of technology, infrastructure, software, and of course rich data and knowledge. All attendees shared the opinion that the KI concept is being built upon addressing the greater benefit of the communities involved, and therefore, it should be a project of collaborative nature and certainly not self-absorbing.

The content of knowledge repositories can be designated as a set of various types that directly reflect the stakeholders' interests and/or operations. These types range from pure data to advanced services and sophisticated features which meet the needs of players. Scholarly publications along with their data, specialized analyses, and communications in the form of mailing lists, weblogs, etc., are a characteristic example of a baseline type. Library content and academic material from professional and academic associations are also examples of such basic types for the KI. Various forms of scientific, social, artistic and heritage nature are expected to have strong typing representation; the same will be the case with curated content from libraries, museums, galleries, archives, broadcasters, producers of moving images, and independent artists. Atop all these types, KI is expected to add information services, data feeds, and aggregations that jointly will help the querying and synthesis of information available in the infrastructure.

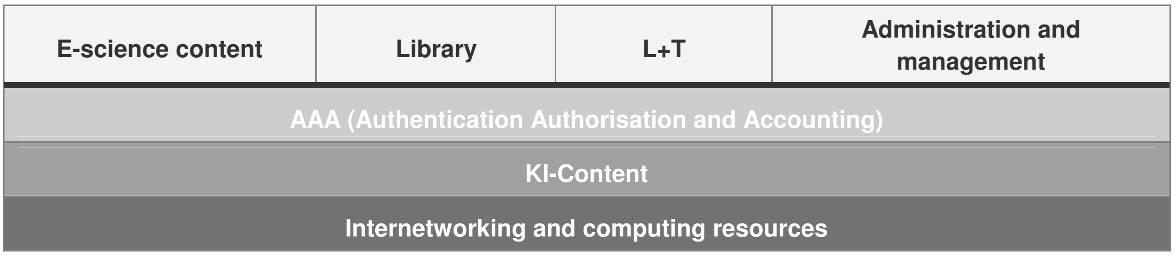
An alternative form of content could be datasets derived from data centres such as AHDS and DANS. Contributors of information could also be governmental and public bodies, research foundations, individual researchers or research communities (such as the "Human Genome Project") and any project whose data and results are to be publicly accessible. In addition, models and respective simulations of software artefacts as well as industrial processes and corresponding data are to be typed entities in KIs. Finally, KI content may entail available e-learning techniques and indeed any useful material originating from teachers, universities, institutions and VLEs and, of course, the already widely spread across the Internet TV streams, news boards and blogs.

A closely related issue is the format in which all of the above mentioned information and services will be available in KIs. This is by no means a question with a straightforward answer; simply considering the multitude of available popular options in small scale storage structures and servers can provide a glimpse of what is to be expected: information may be represented in pure text format, images and photographs, language recordings, multi and "rich" media such as audio and video, observation and experimental data, 3D artifacts computer generated images, maps, complex objects, living and dynamic documents, curated digital objects, and in general any form of software.

Next the question on how best to structure the elements of Knowledge Infrastructure was considered from the organizational and policy perspective. The e-Infrastructure consists of three main components which are:

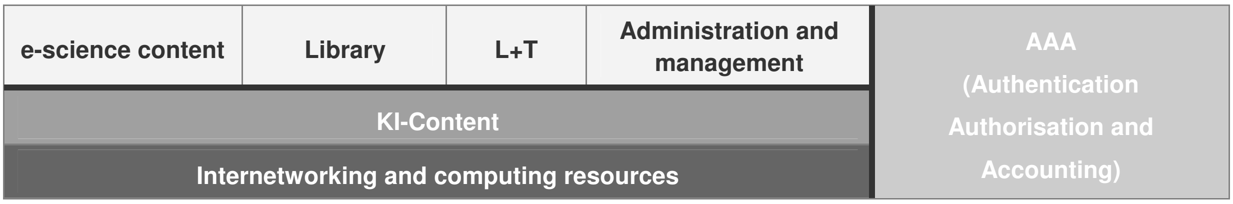
1. computing resources (Data Network, GEANT, Content/Repositories, Computational and Storage),
2. "middleware" (resource management, shared/enable technology) and the long-term archiving and access. This provides a unified view of the content/knowledge catered for also termed KI-Content.
3. authentication, authorization and accounting processes.

The following diagram reflects the above thinking; it also shows various access and management tools used by the various communities.



At present time, it appears that every application manages its own data/knowledge space within the top-level application layer. It was suggested that KI should offer a common and unified knowledge body for all user tools and access options.

It was also suggested that the authentication, authorization and accounting processes could be designed as a vertical element that lies across the KI-layers and are applicable to all layers. Taking this into account, the KI organizational figure changes to:



Before any deployment, the operational model of European-wide Knowledge Infrastructures has to be researched and evaluated carefully to avoid as much as possible the prospects of being unsuccessful. The failure probability of a venture of such magnitude is deemed high, so several questions need to be addressed.

The system implementation will take place in two phases: the setup and the operation. The latter largely refers to system sustainability and in turn points to the critical issue of funding that will ensure future successful operation. In this regard, it would be beneficial, if available budgets on national level could be shifted towards e-Infrastructures. Existing pertinent initiatives should be recalibrated and adjusted according to emerging needs and requirements. In general, the funding remains a complex issue that has to be decided before anyone embarks on the KI-initiative. Whether supported by individual governments or by interested organizations -such as JISC, SURF, DFG, BMBF- research and education funding bodies, and universities, the KI's survival must be ensured despite a number of factors that may present destabilising effects. Given that national funding can be unpredictable, the maintenance and sustainability of KI has to depend on different sources of income. Such support needs to be articulated in a well-laid contingency plan that can help deal with obstacles as they appear, to the extent that this is possible.

There are of course other responsibilities to be shared among the stakeholders: a critical decision is whether the KI structure will appear as centralised. Should KI be formed as a "virtual" organisation including a central office? What about knowledge exchanges among participant countries? What authority will each of the participants carry and according to which standards and considerations will decisions take effect? Who is to be responsible for development, deployment and maintenance of basic and advanced services? Among all these envisaged responsibilities there must exist some form of tight coordination.

Finally, the working group indicated that a very important point must be emphasised: by and large, the abovementioned institutions and government bodies can create knowledge-bases for themselves.

The main point here is to proceed correctly in such an endeavour so that the end result will not be merely a set of isolated repositories, but rather a set of infrastructures compatible with each other. This will ultimately offer the opportunity to create a Europe-wide or even global platform for knowledge exchange and service provision. To this effect, a research organization and/or foundation should lead the process, set standards and enforce their application. The UKOLN in the United Kingdom and the National Science Foundation in the U.S. could successfully fulfil this role. In the European arena, this has to be determined.

- In summary, the group maintained that such a significant endeavour will generate substantial new benefits and its investment will pay off in the future. To be successful, a number of policies have to be adopted by national governments, academic institutions, and EU research establishments. The Knowledge Infrastructure should provide open access to data, publications, information and knowledge at large.
- Its formation should be present in both regional and European-wide levels following a federated approach.
- Institutions, content providers, and SMEs should be presented with clear terms of involvement and derived benefits for contributing quality material.
- The intellectual property rights of contributing researchers, academic teachers, and knowledge workers should be safeguarded within the KI.

It was also perceived that addressing specific legal aspects of the initiative are rather premature at the time as funding and organization issues remain pending.

5.3. User Requirements

This session was facilitated and reported by Luigi Fusco (ESA). The attendees in this working group were:

- Stephen Benians, Metaware
- Donatella Castelli, CNR
- Chris Chambers, BBC
- Giovanni Giovanazzo, Engineering
- Luigi Fusco, ESA
- Enric Mitjana, European Commission

Successfully specifying user requirements is without a doubt a vital aspect in developing KI systems. Clearly, their operation, effectiveness and efficiency will ultimately be matched versus the needs and wishes of the end-users. Issues that this working group dealt with included:

- Definition of knowledge infrastructure from the use perspective
- The specific role of the different actors in the knowledge infrastructures
- Community-specific vs. Generic Infrastructure
- Grouping of User Needs
- Additional User-centric Issues

Definition of knowledge infrastructure from the use perspective

A number of key points regarding the KI definition from the user perspective include: How are users expected to look upon KI systems? What are their expectations? What do they wish to accomplish by participating in such a system?

KI is anticipated to be an ICT-based platform which is open to easily-designated members of communities according to agreed and established policies. In this context, actors (and this does not only refer to people) can contribute, share and access data, information, resources, etc., with the objective to generate and support science, promote community shared interests and encourage collaboration. KI provides and integrates in either structured dynamic way or in ad-hoc fashion resources to exploit data and information in order to create knowledge for use in its context. The term 'resources' refers not only to software and services, but also to physical and network components.

It was envisaged that the KI main functionalities should address ways with which we:

- Produce, control, share, preserve, retrieve, access and consume knowledge.
- Organise and enrich 'chaotic' environments that may appear in the user communities. This process may involve adding context with relevance to user communities.
- Support multidisciplinary applications where the need for across-communities needs for standards, common frameworks and approaches is pressing.
- Allow interoperability and cross-interaction among distributed KIs
- Respond to actors using services with defined restrictions as far as the time/quality/rate of the answer provided.

It becomes rather evident that the creation of KIs implies development and use of organisation elements, policies and standards to ensure well-defined interaction among participants. KIs are owned by "care-taker" communities and therefore will be operational for a long time.

The role of the different actors in the knowledge infrastructures

Several types of actors are to take part in Knowledge Infrastructures and hence, a number of key issues emanating from their interaction have to be clarified.

- The KI Community consists of all stakeholders in the system, including supporters and sponsors, even though they may not directly participate in the key functionalities mentioned above. The boundaries as well as the mode of interaction within KI and the different actor roles are defined by stakeholders with the use of policies.
- Producers stand mainly for generating new knowledge which is pervasive in KI. The term producer is not to be restricted only to physical persons. Producers of information can be entirely digital. For instance, an active software project can dynamically create data which is uploaded to the repository instantly according to specifications set by agreement of interested parties.
- Providers do undertake a distinct role. Although producers are providers, the reverse does not necessarily apply. For example, there can be providers of data, hardware, software, policies, tools, etc.
- Active consumers of information contribute to community specific resources (e.g. ontology, data handling tools, etc.). They also generate new applications.

- Administrators operate and maintain the KI.
- End users are simply people accessing the system and retrieving information or making use of available services to suit their purpose. Their interaction with the KI does not include providing back information.

The following diagram presents the different actors and their interaction within KI:

Community-specific vs. Generic infrastructure

KI must be built in accordance with a well-defined reference architecture which will maximize the benefits of the system and simultaneously will simplify maintenance and upgrading operations to the highest extent possible. Therefore, it would be advantageous if alternative propositions in terms of the reference architecture could be compared so that the best options are finally adopted.

The two extreme cases in system organisation are to implement either a community specific infrastructure or a generic one. To determine the best choice, a set of guidelines has to be put forward to reveal the pros and cons of each architectural option. Interoperability mechanisms and common standards across KIs must be identified, along with generic and specific functionalities. In order for the same component or service to be built across two differently organised platforms, we have to:

- Determine which legacy packages are to be migrated; the same applies to licenses of software packages.
- Possess the ability to manage different types of ontologies, mediators and metadata.
- Adopt a KI data model which relates to the typical flow of information (from data generation through processing to value adding to user, etc.)
- Study the organisation aspects of the communities involved as well as the policies for data access, sharing, etc.
- Describe and publish resources for the defining if KI boundaries (e.g., certification across the KI).

Another issue concerns the exposure of KI resources to different communities. For instance, one question that must be dealt with is how these resources can be used for creating new generic applications and how they use common standards (e.g. web services). Likewise, we must speculate on how new user applications can be made to employ generic and specific KIs for external communities, and what the cost implications of such a prospect will amount to. Also, engagement by new content providers at different levels must be supported to simplify initial entry and interaction with the community.

Grouping of User Needs

In establishing user needs, a process must be adopted to help derive such needs and their classification. This grouping will be the outcome of a categorisation of users into communities and taking into account similar requirements. Thus, it becomes imperative to concentrate on each community's internal usage of KI as well as the capability of knowledge exchange between such communities.

Grouping user needs, should not consider such requirements in a vertical manner. Needs are rather perceived as producers and consumers of information. This is a key notion in KIs and communities should not be placed in silos. Instead of treating communities differently, KIs should facilitate synergy and collaboration. **There is an outer ring of services, an inner ring of data and many different communities that can plug into both. In fact the KI could be perceived as an ecosystem of different users, as in the above diagramme.** The following diagram shows the interactions as well as the "plug-in" notion of user communities to available KI that feature both data and services.

Additional User-centric Issues

There is a set of other topics which call for contemplation in the user requirements area. One must reflect on both cost and complexity issues. **Costs must be supported by actors in the system**. The question arises concerning the actors and the costs that each has to bear. As far as complexity goes, one must determine the nature of the KI properties that are needed and implement them to optimal effect.

The issue of **ownership of knowledge repositories** is decisive on the basis of who will cover the costs including setting up, maintenance, performance-provision, etc. These in turn point to the subject of sustainability after the initial implementation and the manner that this can be ensured.

The system should be able to offer a **number of interfaces depending on the type of the users** and the nature of services available. The same stands regarding support for SMEs, incubators, spin-off, etc.

Interaction among different KIs must also be established through a set of well-laid and understood rules and policies. It is desirable that the system is flexible in terms of modification of policies and costs. The latter will assist in the evolution and further KI development in accordance with technological and sociological advancements.

Last but not least, it is vital to consider carefully issues and **requirements pertinent to archiving and preservation strategies.**

6. Outcomes, Recommendations and follow-up steps

In overview of the topics discussed during the meeting as well as the ideas expressed and possible future courses suggested, the following can be pointed out:

A set of desirable features of KI systems was established. These include the independence of repositories with respect to grid infrastructure, the provision of handles for the objects involved, as well as the preservation of content even after it has been modified in order to determine and trace its evolution. The system architecture, in terms of technology, can follow either a logical or a physical arrangement. Both of these approaches are widely used across database and digital library implementations.

Metadata are expected to be standardized and treated uniformly across the system. Apart from providing timely access to large amounts of data, metadata can also be employed to offer different views of content. The diversity of both knowledge and data that can be stored in KIs calls for careful planning in the areas of their accessibility and availability. Metadata can be used effectively for this purpose. By such means, the classification of users and the determination of the information path across the system will be plausible.

It is crucial that services are implemented initially on a low level scale (basic services), and that more sophisticated needs and requirements are satisfied by more complex services. Such services will stem as a result from a combination of lower-level applications. Service synthesis must take place dynamically while taking into account specific content delivery requirements. This hierarchical organization simplifies the service model and promotes system sustainability. Furthermore, it endorses the possibility of more advanced services being available to specific types of users, according to their status and role in the system.

All types of actors in the system, ranging from end-users to software corporations and non-profit organizations, can contribute content of their choice. This unavoidably results to extreme data quantity and complexity. Although quantity does not necessarily imply quality, this development can greatly enhance information availability across knowledge platforms, provided that content is being introduced following a set of a-priori agreed rules. The open source community has in the last years played a leading role in launching this practice with encouraging results. Their experience and expertise can be used to the benefit of the KI community as well. On the other hand, not all information and services are to be accessible to all users in the system, which in turn calls for the appropriate classification of data. Clearly, the adaptation of an authorization scheme will have to cater for the problem.

Grid technology is expected to serve as the backbone for implementing KIs along with other existing networking and technology resources. Since networking and grid infrastructures are already available across the continent, KI development should exploit these resources by making use of grid technology when constructing repositories. Grid platforms can also provide the basis for further evolution and implementation in Knowledge Infrastructures. Since fine-tuned performance is essential for the successful adaptation of KIs, QoS characteristics for grid services will have to be enabled.

Technical challenges that will have to be dealt with in the research implementation of Knowledge Infrastructures include: data transformation, exchange and integration, data management, annotation, heterogeneity, scalability and robustness, provenance, data citation, security, preservation, abstraction and personalization for users.

The adequate classification of actors and interactions in the KI environment is deemed critical, if the attempt is to be successful. Our pertinent working group identified a number of roles while taking into account that much of the content and funding will come from public organizations, consortia of universities (e.g. GRID- and E-Research communities, researchers, academic teachers), and individual institutions (e.g. national libraries, library and technical communities). The same actors should therefore claim the ultimate ownership of KIs. On the other hand, SMEs, commercial technology providers and software companies should be able to play a role commensurate to their involvement. The interaction of users with KIs designates the specific types of stakeholders: producers, providers, active consumers, administrators and end-users. It is essential to identify the needs and requirements of users, thus classifying a set of communities. Further development can be achieved by focusing on the internal usage of KI by each community, as well as the form of dealings and information exchange between communities. The content types available in KIs are to be determined on the basis of the stakeholders' interests and may vary from commonly used data types to sophisticated ones. KIs aim to the greater benefit of all participating communities, and thus, they should be of collaborative nature.

In terms of organisation and policy, the components of KI systems can be distinguished as computing resources, middleware and authentication, authorization and accounting (AAA) processes. Two organizational models were suggested; in these two models the AAA component is being differently laid out. KI implementations across (and not limited to) Europe should be designed along these models so that the end result is a set of compatible infrastructures. Implementation is to take place in two distinct phases namely, the setup and operation stages. The second stage is closely associated with the subject of sustainability. Ensuring continual funding and developing a contingency plan are a must in order to deal with possible complications along the way.

Especially on the issue of sustainability of e-Infrastructures it should be pointed out, that the e-IRG remarks considerably coincide with those of the BELIEF project. In particular the main points of the e-IRG recommendations are:

- governments and the Commission should develop policies and mechanisms to encourage increased investment in a more coherent and interoperable way across Europe
- the existing e-Infrastructure projects must be superseded by integrated sustainable services at national and European levels
- e-Infrastructures must be application-neutral and open to all user communities and resource providers. National funding agencies should be encouraged to fund multi-disciplinary and inclusive infrastructures rather than disciplinary-specific alternatives
- e-Infrastructures must inter-operate and adopt international standard services and protocols in order to qualify for funding
- the Commission should, within the seventh Framework Programme, develop a pan-European e-Infrastructure which explicitly encourages the further integration of national e-Infrastructure initiatives

Also, a number of issues that have to be addressed early on during the KI effort should include ownership of scientific repositories, cost coverage, interaction among KIs, as well as adoption of archiving and preservation strategies.

Meeting the expectations of the scientific and industrial community as well as those of the general public is a crucial point regarding the use of KI platforms and their ultimate wide-scale acceptance. Access to scientific repositories should be granted in conformity to agreed and established policies. Actors in the system should be able to contribute, share and access data following the overall spirit of the endeavour which is support the Sciences, promote European and global collaboration, and help economic growth in the continent.

7. Appendix

7.1. Workshop Agenda

The workshop was organized as a one-and-a-half day event and was co-located with the European the e-IRG meeting. The agenda was as follows:

Tuesday 11 April

15.15 – 15.45 Arrival and Registration
Opening Plenary
Moderator: Yannis Ioannidis, University of Athens
15.45 – 16.45 **Welcome and Introduction**
Welcome by the Austrian Grid Forum, Prof. Jens Volkert
Introduction of all participants, event specific objectives, process, and expectations
Presentation by the EU representative Enric Mitjana
Presentation by e-IRG Chair Dieter Kranzlmueller
16.45 – 17.00 Break
17.00 – 18.00 **Invited Presentations**
Peter Buneman, University of Edinburgh
Peter Kunszt, Swiss Supercomputing Center
18.00 – 19.00 **Open discussion**
Remarks by participants
Suggestions of 1-2 related topics per participant to be included in the discussion
19.00 Close of day and subsequent Dinner

Wednesday, 12 April

08.30 – 08.45 Formation of Working Groups and the corresponding topics
Parallel sessions / working groups
Technology issues
Moderator: Peter Buneman, University of Edinburgh
Organizational issues and policies
Moderator: Norbert Lossau, University of Bielefeld
User issues
Moderator: Luigi Fusco, European Space Agency
08.45 – 10.30 Parallel sessions / working groups
10.30 – 10.45 Break
10.45 – 12.30 Parallel sessions / working groups
12.30 – 13.30 Break
13.30 – 14.30 Parallel sessions / working groups
Discussions of summaries, action points and suggestions to the plenary
Reporting Plenary
Moderator: Donatella Castelli, ISTI – CNR

14.30 – 14.50 Reporting of "technology" group with follow-up discussion
14.50 – 15.10 Reporting of "policy" group with follow-up discussion
15.10 – 15.30 Reporting of "users" group with follow-up discussion
15.30 – 15.45 Break

Closing Plenary
Moderator: Yannis Ioannidis, University of Athens

15.45 – 17.00 Synopsis of group leaders and organizers
Looking to the future
Investigate synergies between BELIEF topics and e-IRG topics
Looking for collaborative opportunities and joint action points in FP7 and IST priorities
17.00 End of 1st Brainstorming meeting

7.2. The BELIEF project and its scope

BELIEF supports the goals of the Research Infrastructures activity area (and in particular in the area 3.2.3 "Communication Network Development – eInfrastructure – Consolidating Initiatives") of the FP6 Programme for "Structuring the European Research Area" by leveraging on European eInfrastructure-based initiatives and developing a denser network scheme through which further international outreach and industry engagement may be achieved.

The main objectives of BELIEF (2005 – 2007) are:

- To build and promote an effective Communication Network Platform;
- To develop, maintain and populate a multimedia Digital Library (DL) which makes accessible in a homogeneous way, through advanced services, the documentation produced by eInfrastructure projects and initiatives;
- To deliver four one-to-two-day brainstorming events on strategic multidisciplinary topics (20 – 30 participants), organized in strategic cities in Europe and in the US;
- To deliver two one-day strategic networking workshops with major potential user communities (40 – 60 participants) organized in strategic cities in Europe with a potential possibility in an NIS country;
- To deliver two two-to-three-day international conferences and exhibitions (up to 250 European & Indian and 250 European & Latin American participants) in India and Latin America, respectively, providing a platform to exchange experiences, showcase results and learn developments from emerging economies;
- To publish four Brainstorming reports covering technical research issues as well as legal, policy and accounting issues, and user needs, polarising the interests of the eInfrastructure Community and well as non; these will be made publicly available;
- To publish four (periodically revised and updated) BELIEF Research & Industry handbooks on specific business models tailored for different user communities, including discussions of exploitation and sustainability issues, case studies, and surveys of evolving and deploying eInfrastructures to user communities; these will be made publicly available.

BELIEF's expected results are:

- To develop synergies among communities of researchers and industries with ongoing eInfrastructure initiatives

- To form a worldwide accessible Digital Library of eInfrastructure multimedia documents serving research and industrial communities;
- To increase international co-operation by bringing into the European scene Indian and Latin American players;
- To provide an amalgamation of the current research communities' fragmentation and to sensitise the industry sector to eInfrastructure deployment
- To attract greater interest from Industries & SMEs on use & deployment European eInfrastructures

In addition, 860 to 1.000 organizations from industry, including SMEs, academia and research institutes, will pool experiences, knowledge and expertise during the brainstorming & networking workshops and international conferences and will establish a networking platform, which will serve as a foundation for further collaboration towards the ultimate goal of a consolidated international eInfrastructure initiative.

7.3. List of participants

	Name	Affiliation	Country
1	Atkinson Malcolm P	NESC	UK
2	Aerts Patrick	NLF	Netherlands
3	Balint Lajos	National Information Infrastructure Development Program NIIF	Hungary
4	Barbera Roberto	University of Catania	Italy
5	Baxevanidis Kyriakos	European Commission	EU
6	Benians Stephen	METAWARE	Italy
7	Bilas Angelos	FORTH	Greece
8	Buneman Peter	University of Edinburgh	UK
9	Castelli Donatella	ISTI - CNR, Pisa	Italy
10	Chambers Chris	BBC (Grid project)	UK
11	Delis Alex	University of Athens	Greece
12	Di Meglio Alberto	CERN	Switzerland
13	Foster Allan	JISC	UK
14	Fusco Luigi	European Space Agence (ESA)	Italy
15	Gagliardi Fabrizio	Microsoft	Switzerland
16	Giovinazzo Giovanni	Engineering	Ireland
17	Ioannidis Yannis	University of Athens	Greece
18	Jeffery Keith	CLRC, RAL	UK
19	Karayannis Fotis	GRNET	Greece
20	Kranzlmueller Dieter	University of Linz	Austria
21	Kunszt Peter	Swiss Supercomputing Center	Switzerland
22	Leenaars Michiel	NCF	Netherlands
23	Lopez Diego	RedIRIS	Spain
24	Lossau Norbert	University of Bielefeld	Germany
25	Manieri Andrea	Engineering	Italy
26	Manset David	MAAT-G	Spain
27	Mitjana Enric	European Commission	EU
28	Morais-Pires Carlos	European Commission	EU
29	Occhioni Giorgio	Engineering	Ireland
30	Read Malcolm	JISC	UK
31	Schuldt Heiko	University of Basel	Switzerland
32	Telksnys Laimutis	Institute of Mathematics and Informatics	Lithuania
33	Toli Eleni	University of Athens	Greece

7.4. Additional Information Resources

- Slides
- Evaluation Report