Jürgen Renn (ed.)

**ECHO – an infrastructure to bring
European Cultural Heritage Online**

The foundation papers of a European initiative

ECHO — AN INFRASTRUCTURE TO BRING EUROPEAN CULTURAL HERITAGE ONLINE

THE FOUNDATION PAPERS OF A EUROPEAN INITIATIVE

EDITED BY

JÜRGEN RENN

MAX PLANCK INSTITUTE FOR THE HISTORY OF SCIENCE, BERLIN, GERMANY

CONTENTS

INTRODUCTION

Three Max Planck Institutes in three European countries; for the History of Art in Rome; for Psycholinguistics in Nijmegen; and for the History of Science in Berlin, each with their international partners, have founded an initiative to bring the European cultural heritage online and to foster, at the same time, a new infrastructure for the humanities adequate to the Internet Age and competitive with similar ventures in the US. Under the label ECHO (European Cultural Heritage Online) the Max Planck Society has proposed this initiative for support by the European Commission. It has been extensively discussed at a workshop in October 2001 on "Humanities, Research and Cultural Heritage in Europe," organized by the European Commission in Brussels. This collection comprises the texts of the three keynote speeches held at this workshop, in addition to a selection of the foundation papers documenting the emergence of the ECHO initiative, all presented in chronological order.
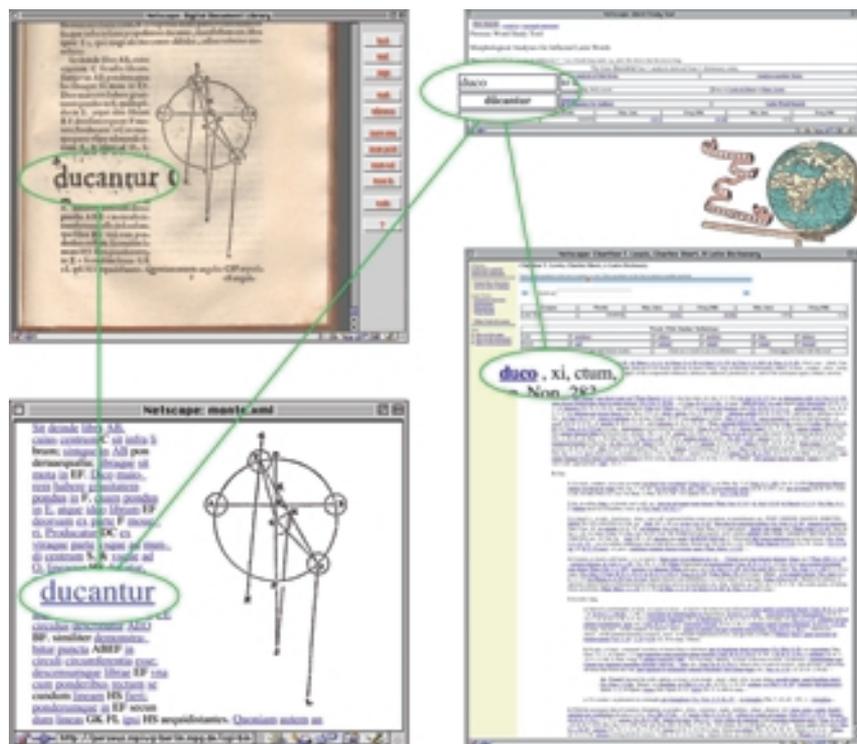
Example: The Max Planck Institute for the History of Science and its partners have developed an electronic environment for scholarly work with historical sources across various languages. A digital facsimile can be physically located anywhere in the world and linked with various commentaries that may be stored elsewhere, thus realizing the possibility of an Internet-based collaborative network. The example shows the link with an electronic transcription. The transcription in turn is linked to a morphological analyser allowing the identification of the given grammatical form and the root of the word. From here a further link leads to a dictionary entry providing an English translation and further explanations. (Illustration by Markus Schnöpf)

.

Jürgen Renn

Max Planck Institute for the History of Science, Berlin, Germany

Two Aspects of a Crisis
of European Culture

Two Aspects of
a Crisis of European Culture
Jürgen Renn

A Speech Given at the Workshop

*Convergence and Diversity of European Societies —
Legal and Economic, Social and Cultural Aspects
of the Research Framework*

Brussels, November 7, 2000

This is an occasion to review, to discuss, and also to enjoy perspectives of future research and prospective European collaborations, and I am grateful for the opportunity to join this discussion with a few remarks. I would like to begin, however, by pointing you not to a flourishing development but to what I consider to be nothing less than a crisis of European culture. This culture, the common European heritage that binds us together even more strongly than the European institutions, is presently in danger of being left behind, of missing the train, so to speak, of the rapid technological developments carrying us into a new information age. Indeed, the medium of today and tomorrow, the Internet, might leave behind a culture which is the heritage of our past but urgently needed to meet the challenges of the future.

The bulk of information which forms the core of European cultural heritage, the great works of literature and art, as well the treasures of scientific, scholarly, and philosophical writings going back to the dawn of the European civilization are in fact largely excluded from that information system which already now constitutes the backbone of an ever-more knowledge-based world. And the little culture that is included in the world wide web due to the efforts of a few pioneers is almost drowned by the tides of information garbage. This is one aspect of the crisis I mentioned, and it concerns the European society at large as well as the humanities which vitally depend on access to the information representing the European heritage. In fact, contrary to all disciplines in science and technology, which are improving their effectiveness by innovations exploiting the tremendous potential of electronic data processing, the humanities so far have been widely incapable of initiating a similar dynamics. But perhaps, their advancement towards computational humanities will simply be an innovation that Europe has to leave to others.

This is but one aspect of the crisis of European culture that I have in mind. In fact, the deficit in the extent to which cultural information is available on the net is accompanied by another deficit with perhaps even more pernicious consequences, the underdevelopment of cultural techniques adequate to the new information technologies. In the fields of language technology, image analysis, and the implementation of mathematics on the net, that is in fields of high economic and technological impact, bottlenecks become visible that are related to the negligence of an adequate transfer of the traditional cultural techniques of writing, depicting, and calculating to the new media. The problems of language technology, for instance, have for a long time been considered merely as an engineering challenge and not as a field to which the humanities can bring their century-long expertise in the linguistic representation of meaning. Meaning is, after all, not only in the text but also in the cultural context so strikingly absent from the new media. In short, the lack of implementation in the new media of the cultural information *and* techniques which are the domain of the humanities represents a major stumbling block to what, in my view, might otherwise become a second Internet revolution.

.

The Second Internet Revolution

This revolution will, however, not take place automatically, merely as a consequence of technical developments, but it requires the creation of new, content-laden information structures which can only result from an effort to overcome the present marriage of ignorance between the humanities and technology. What is needed is in fact not just hardware and software, but "thoughtware" such as open digital archives hosting, for instance, our cumulative knowledge on the history of architecture, the information available on the *ius commune* or language technology based on semantics. Thoughtware aims at capturing meaning and not bytes. Precisely because the creation of thoughtware depends on scholarly as well as technological competence, the new infrastructure I have sketched will not come true without the active participation of the scientific community and without the massive support of science policy.

There is a gradually growing awareness that the humanities will have to play a different role in the future. In fact, in several European countries projects have emerged that address the new challenges and that attempt to bring, albeit with great difficulty, various fields of the humanities to the Internet age, such as the Athens-based project on the history of science at the European periphery, the Rome-based project on the history of architecture, the Nijmegen-based project on language technology, or the Heidelberg-based project on international law. But such projects still suffer, as a rule, massively from a lack of technical support tuned to their scholarly needs. In particular, support on a European scale for such ventures is not yet available to the extent that is necessary.

A Proposal

This brings me to make, in conclusion, a proposal of how to address the challenges posed by the critical situation of European culture in the age of the Internet. I propose to found and fund an initiative to support the transfer of the European Cultural Heritage Online (ECHO). The initiative should take an organizational form which is appropriate to the distributed expertise involved in its realization. It should therefore pursue a double strategy. It should, on the one hand, establish a European network of projects engaged in making the cultural heritage accessible online. And it should, on the other hand, establish a European centre charged with supporting these projects and with accumulating the competencies necessary to boost the technologies needed by the single projects. These technologies will also most likely find applications well beyond the humanities.

The time for the creation of such an initiative seems right. Just now the U.S. National Science Foundation is reaching out to collaborate with European projects on digital libraries. But in my view, it is precisely the rich diversity of European culture that could prove to be a competitive

advantage and unique asset in facing the challenges of the second Internet revolution. I am also convinced that the momentum of the Max Planck Society's new Centre for Information Management, which has just now been created and which is specifically set up in anticipation of future international endeavours, could be brought to bear on a comparable European venture dedicated to the humanities.

The integration of Europe is evidently not just a matter of economic and administrative unification but also presupposes that European citizens can define themselves as belonging to a joint culture that can only be the result of the integration of our diverse historical and cultural pathways. It is to this integration that a European program dedicated to bringing the common cultural heritage online may generate a modest but very concrete contribution.

INTERNATIONAL WORKING GROUP

ECHO
EUROPEAN CULTURAL HERITAGE ONLINE

FOUNDATION NOTE

ECHO

EUROPEAN CULTURAL HERITAGE ONLINE

FOUNDATION NOTE

THE

FOUNDATION NOTE

FOR THE PROJECT

ECHO
EUROPEAN CULTURAL HERITAGE ONLINE

WAS PREPARED AND CIRCULATED BY THE

INTERNATIONAL WORKING GROUP

CONSISTING OF

P.A. COPPEN, UNIVERSITY OF NIJMEGEN, THE NETHERLANDS
E. KIEVEN, BIBLIOTHECA HERTZIANA, ROME, ITALY
S. LEVINSON, MPI FOR PSYCHOLINGUISTICS, NIJMEGEN, THE NETHERLANDS
J. RENN, MPI FOR THE HISTORY OF SCIENCE, BERLIN, GERMANY
URS SCHOEPFLIN, MPI FOR THE HISTORY OF SCIENCE, BERLIN, GERMANY
PETER WITTENBURG, MPI FOR PSYCHOLINGUISTICS, NIJMEGEN, THE NETHERLANDS

FOLLOWING UP DISCUSSIONS AT A JOINT WORKSHOP OF THE MAX PLANCK SOCIETY AND THE
EUROPEAN COMMISSION

CONVERGENCE AND DIVERSITY OF EUROPEAN SOCIETIES -
LEGAL AND ECONOMIC, SOCIAL AND CULTURAL ASPECTS OF TE RESEARCH FRAMEWORK

IN BRUSSELS, NOVEMBER 7, 2000

ABSTRACT

Europe is facing the large challenge to exploit its rich cultural heritage with the help of advanced resources from information technology and to make it freely accessible via the Web. Gigantic efforts and new organizational forms are necessary to preserve the existing knowledge and to transform this rich treasure to formats accessible to future generations. Only multi-national initiatives will provide the collective potential to meet this challenge and Europe should play a leading and competitive role here.

A strong interaction between cutting edge information technology and the disciplines of the humanities will be beneficial for both. New technical frameworks and archiving standards will emerge and influence the disciplines world-wide. Many existing technologies such as natural language processing will be driven forward by the demanding requirements of the humanities.

In order to meet these goals, the establishment of the ECHO project (European Cultural Heritage Online) is proposed. A Virtual ECHO Innovation Centre will house the technological expertise, develop reusable technical frameworks, establish archiving and presentation guidelines and standards, and organize the knowledge transfer. This centre will cooperate with many projects in various European countries, thus bringing together and coordinating the scholarly expertise available. As a start, a small ECHO scenario is envisaged under the 5th Framework, followed by a continuation of these efforts on a larger scale under the 6th Framework, once the chosen construction turns out to be successful. For the 5th Framework three exemplary projects were selected (Language Area, Knowledge History Area, and History of Arts Area) to serve demonstration and evaluation purposes.

Three Max-Planck-Institutes playing a leading role in the respective areas and the University of Nijmegen which has a long history in natural language and speech processing will act as the nucleus of the suggested ECHO project. ECHO will establish broad networks of experts in the humanities disciplines and look for an integration with the national programs. Currently the network comprises already more than 20 institutions distributed across Europe. Of especial interest are firm links to institutions from East European countries to integrate their rich archive material into a common European distributed on-line archive. In doing so ECHO will deliver an important contribution to an integrated European Research Area.

.

1. INTRODUCTION

1.1 Problem Description

Europe has a rich and varied cultural heritage of many dimensions. A rough estimation by the UNESCO credits the creation of a large percentage (>90%) of all innovations and cultural achievements world wide to a European source. In contrast with this, there is a surprising lack of European activity to use new electronic media to make this cultural heritage available to the general public, future generations, and even interdisciplinary science. Currently used methods of documentation, archiving and transferring the knowledge hidden in this great treasure, are far from being sufficient, let alone being competitive on an international scale.

Current programs for promoting European cultural heritage are to a great extent run by individual member countries. However, the amount of archiving involved is gigantic. Complex but standardized formats for the representation of online information have to be developed. New methods have to be invented to present information via the World Wide Web. These vast technological challenges make national programs no longer sufficient. Making European cultural heritage available on a new level is a European task and only a multi-national initiative will provide the collective ability to guarantee the transfer and presentation of information in larger relevant quantities.

Because language variation is strongly related to cultural development, Europe's rich cultural history has resulted in a great variety of languages and dialects. The cultural development has been driven in particular by enormous technological changes and innovations, and by correspondingly significant socio-economic changes. New language variants have continuously emerged, while others have died or are on the verge of becoming extinct. The WWW and modern media pose an unknown and potentially dramatic threat to traditional languages. It is time to multiply current efforts to transfer existing knowledge of endangered and extinct languages to the Web platform and to document and archive their acoustic/phonetic characteristics in multimedia dimensions.

Current projects are mostly nationally funded and therefore do not have the means available to archive and exploit the material using modern IT methods to the same extent that multi-national funding would allow for. Even pioneering ventures still suffer massively from a lack of IT support tuned to their scholarly needs. Receiving such a support would enable them to join forces to arrive at a critical mass for contributing to an innovative infrastructure for research, education, and the interested public. The implications of such a contribution exceed the boundaries of the humanities. In particular, "Open Digital Libraries" can be expected to break traditional

disciplinary boundaries and offer a new platform for interdisciplinary and comparative views on European culture, views that are nowadays still largely divided by barriers of language and national traditions.

With respect to application-oriented Information Technology, Europe evidently lags behind the United States. There, IT is seen as a first order strategic technology, i.e. the pervasive character of IT is well-understood, and in many areas, programs are defined to promote IT and to apply IT in more and more disciplines. The gap between IT in Europe and the US will become greater if Europe does not compete in many disciplines by funding influential consortia with a high potential impact factor. This factor should not be measured only in terms of short-term revenue. An important consideration in this respect is the 'image' of the IT field in Europe, which is of great importance to attract the best young people and to convince specialists world-wide that Europe has the potential of defining standards in IT. Currently, with some exceptions, this image is not present.

The new leading role defined by the eEurope initiative can only be achieved if a small number of excellent institutions collaborate on innovative projects to create the potential for "competitive dominance". Only such groups will be accepted as equal partners.

Europe's natural advantage of a rich treasure of cultural history offers a good chance of success to the ambitious eEurope idea. However, it is apparent that Europe has to transfer as soon as possible the leading domain specific knowledge in areas such as cultural heritage into a leading domain specific IT knowledge.

With respect to the humanities in Europe we can summarize that

- there is a considerable gap in applying advanced IT methods to the humanities, and there is almost no mutual fertilization between the two,
- national projects alone will not be sufficient to meet the great challenges of transferring the material and the knowledge about our cultural heritage to the new on-line media in an adequate and non adhoc way,
- Europe currently does not play a significant role in defining international guidelines and standards in the humanities, although the discipline knowledge is basically housed in Europe.

The challenges Europe is faced with to manage the information revolution in the area of the humanities and to play a leading role require new ways of thinking and new organizational solutions. Three Max-Planck-Institutes, each playing a leading role in their respective scientific

.

discipline, and the University of Nijmegen are willing to join forces, to merge their outstanding domain knowledge, to integrate their proven domain-oriented technological expertise, to bring together their wide-ranging networks of experts, and to look for close collaborations with the national programs. They propose to achieve this by starting the ECHO project.

1.2 Recent Developments

The development of ECHO relevant technologies is very dynamic. Trends in some of these areas will be sketched briefly in the following sections.

Web-Technologies

The continuously emerging Web is the innovation in IT with the greatest impact in the last decade. Continuously increasing bandwidths even for wireless transmissions will make the dream of ubiquitous computing via intelligent mobile devices a realistic option for the coming decade. The Web, based on the concept of individually defined semantic hyperlinks, has emerged as a gigantic "melting pot" of information of all kinds. However, due to the Web's rapid development, the shear volume of information makes it cumbersome to use. New technologies are developed, based on well-established concepts of information encoding (XML, XML-Schema, XSL, XLINK, etc.) They offer standardized methods for a proper definition of the structural properties of documents, separating layout details from content definition. Moreover, these technologies facilitate structural transformations and powerful links within and between documents. Only structurally well-defined documents adhering to these standards will be accessible to automatic information management in the future and are thus a pre-condition for any long-term archival solutions.

New initiatives have been started to better organize and structure the information offered on the web – at least with respect to information used by professionals. The term "semantic web" illustrates this commitment to a structure based on a proper semantics that is sensible and intuitively understandable. These structures can be created automatically, but in many cases they are still made manually. Metadata initiatives such as DC, IMDI, MPEG7 and others, as well as the definition of RDF schema and the Open Archives Initiative are intended to produce solutions towards this end. Advanced and more intelligent linguistic technologies will have to be employed to investigate the searchable and browsable spaces in order to locate easily and exactly the professional resources we are aiming at.

Multimedia Technologies

Kernel technologies such as CPU-speed and network bandwidth develop so rapidly that it will soon be possible to support video streaming based on standard encoding methods such as MPEG1/2/4 on mobile devices, or even to directly transmit MPEG-encoded camera recordings.

Methods for improved compression, media indexing and caching which have been introduced or are in development will allow for efficient data management such that the transmission of several streams in parallel will no longer be utopic. It is clear that future user interfaces will support multimedia data, making the information being offered far more accessible and understandable.

In the future, multimedia content will play a larger role in humanities projects (sound and movies in particular). Currently, a number of initiatives exist to automatically index such material with the help of sound and image processing. These will allow a goal directed access to individual media fragments, which will dramatically improve the efficiency for the user as well as for the transmission channels. The humanities need to catch up in this respect as well. At present, the methods used for major languages have to be adapted to minor languages for which less material and knowledge is available.

A side aspect, nevertheless utterly important, of multimedia technology is the automatic operation on sound and images of various sort. Only advancements in this area will allow humans to exploit all the archive material currently being digitized. Advancements of processing material taken from limited domains have been achieved. Nevertheless, these techniques are still in their infancy. The humanities have to actively test the status of these technologies by applying them to limited areas. In collaborations with experts projects have to be defined to push these technologies ahead.

Multimodal Technologies

Speech technology has made quantum leaps since the introduction of statistical methods, making its application to limited domains an interesting option. This holds for the recognition of speech as well as for its synthesis. However, speech technology in the humanities is presently almost non-existent. Speech synthesis offers new possibilities, especially for the presentation of speech phenomena and historic cultural events. Speech recognition will become a standard interface in modern interactive devices – particularly in the next generation of mobile devices. It will allow a broader public to profit from the wealth of information available in the Web.

Also on the horizon will be the automatic recognition and synthesis of parameters such as gaze, facial expressions and gesture, as part of intelligent user interfaces. Because the success of working with these parameters will be domain and culture dependent, it seems of great importance to gather knowledge and experience about their application in the humanities.

On the output side, animations and virtual reality will become state-of-the-art. They will be projected on life-like creatures performing synthesized lip movements, gestures, facial expressions, and eye movements. From psycholinguistics it is well-known that in many

communicative situations multimodality will facilitate the transfer of information. Multimodal user interfaces are more intuitive to humans, particularly when explaining complex spatial patterns such as route descriptions. They will naturally encourage users to explore the interface capabilities. For fast intent caption, which is so important for the acceptance of dialogues with machines, multimodality is obligatory.

Linguistic Technologies

Computational linguistics has also developed considerably during the last decade although the discipline can still be said to be in its infancy. Many problems are still to be solved. The exploitation of a gigantic amount of textual material for high precision retrieval is dependent on the success with which linguistic content can be captured into a compressed and structured form. Topic detection and summarization methods are continuously improving as is the knowledge of how to construct and combine language models, ontologies and lexica, which are vital for modern computational methods in text technology. Recently, the repertory of tools and components for the various main languages has grown very rapidly.

Within computational linguistics, a domain-specific approach and the availability of a proper knowledge is essential. Since the humanities mainly deal with more obscure languages and languages from other eras, it will require some major effort to port known methods and tools to new domains.

A special challenge in the web-era is to provide multilingual services. Currently, many projects world-wide are creating multilingual terminology registries and dictionaries to allow improved automatic processing. Machine translation however is still in its infancy and deals mainly with the main languages. To what extent such techniques can be applied to the textual sources typical for the humanities, needs to be investigated in detail.

Driven by the dramatic increase of the relevance of the web as an international and general source of information and knowledge, computational linguistics will develop quickly, achieving new milestones in the coming years. Initially, only the main languages will probably profit from this development. Within the humanities, efforts have to be made to keep pace with these developments. The application of these new methods to the humanities will also undoubtedly reveal their weaknesses and strengths. This will in turn lead to their improvement.

Others

Although the term knowledge management is ill-defined and heavily debated, we will use it in a broad sense for the bundling of methods and technologies for efficiently extracting knowledge from complex information, or, conversely, transform knowledge into multimedia documents. It is clear that knowledge management in this sense will improve as the ways of bundling these

technologies become smarter. For example, using the idle CPU time of the increasingly fast processors for intelligent background processing offers the potential to support the user in a much better way than is experienced today.

Another area where great improvements have already been achieved is the development of tools allowing the collaborative work of people working at various remote locations on distributed multimedia documents. In the future these techniques will become commonplace, allowing the users a continuous and smooth exchange of ideas and facts.

An increasing problem the humanities is faced with is the need to set-up archives which will store multimedia content for an unlimited time. Storage and access technologies have developed to the extent that nowadays major institutions can store multimedia contents in great quantities and offer it in a near-line mode. For typical bandwidths between 3 and 10 Mbps for MPEG2 encoded video streams, a well designed archive has several layers of storage which in total offers much more than 10 Terabyte capacity. It requires experienced system managers, well-designed organizational measures, and frequent media exchanges to maintain such archives over many years.

2. LONG TERM CONCEPTION (6TH FW)

2.1 Goals

Given the state of Information Technology in the humanities on the one hand and the development of the key technologies on the other hand, we can derive the major goals for the ECHO (European Cultural Heritage Online) initiative. The ECHO initiative aims at offering the framework for a coordinated integration of the efforts of pioneering projects from various European nations, providing them with the necessary IT support, and building, on the basis of such a joint effort, an innovative IT platform for making available the historical, cultural and language sources. Such a platform will comprise two levels:

The first level offers a set of standard tools to facilitate making large resources freely available to the Internet audience. These tools will especially serve humanities projects lacking high level IT support and competence. This level will not only provide the empirical basis for innovative research on the emergence of science in Europe, but also help to fill the cultural vacuum in the net and, in doing so, make eContent of high cultural value available to a wide public.

.

The second level aims at a novel kind of access system combining research and dissemination serving as a model for the organization of electronic knowledge in other domains as well. In such an access system, based on the cognitive organization of the sources, the original data are continuously improved by a direct application of research results and annotating efforts, thus immediately turning these results into instruments for further research. In this way an Open Digital Library would emerge, serving as a framework both for cooperative work on the sources and for the presentation of its results.

In order to provide a content-based access to the sources of the humanities, it will be necessary to develop suitable ontologies reflecting their cognitive organization. As these ontologies can only emerge in the context of scholarly work on the sources, one needs efficient working tools in order to associate the structure and content of a document with an emerging ontology. A digital library of the sources documenting European culture thus poses a generic problem of knowledge management whose solution may have paradigmatic character: the "co-exploration" of ontologies and empirical data-structures.

In short, ECHO aims at bundling and increasing the efforts in Europe to exploit the rich European cultural heritage with the help of advanced resources from information technology and to make the results available via the web. This may in fact have a double effect: First, the sources of the European cultural heritage which are, as a rule, dispersed over various public and private collections by the contingencies of history or of disciplinary specialization, can now, for the first time, again be reconstituted to their integrity - at least in digitized form and made freely accessible in the Internet. Second, the availability of a considerable body of cultural information in digitized form will contribute to transforming the Internet into a semantic web allowing to exchange and process information not just via technical protocols but in the languages of human culture.

To achieve this, we suggest the close collaboration between a number of local experts and an exemplary steering team, operating as a virtual centre of expertise for suitable technological solutions. This collaboration has to lead to the development of technological frameworks with potentially reusable components, and the establishment of a broad knowledge base on methods. The central team will take over the role as a centre for advanced information management in the humanities.

Both levels of the new platform can only be realized if minimal standards for rendering historical documents electronically are defined and implemented. Under the auspices of ECHO a number of pioneering projects from various European countries should agree on the definition and implementation of such joint standards, comprising not only standardized document struc-

tures in XML but also a core set of working tools and production lines to generate documents structured according to these standards, as well as a suitable environment to present the documents on the Internet.

ECHO also has to become a central resource for those institutions which are undertaking projects in the humanities requiring a considerable involvement of technology, but neither have the human nor the technical resources to put their work on the web or to store their material on a long-term basis. ECHO intends to look for partner institutions, such as existing museums and archives, to find long-term platforms for providing the kind of eContent discussed in this document. Also publishers have to be involved to work out new models of commercial eploitation in the web era. This implies that ECHO itself should only house such material during a transitional phase, or if no suitable ways can be found which are financially self-supporting.

With respect to the application of domain-specific information technology ECHO will claim to be one of the leading institutions world wide. In close collaboration with partners from research and archiving institutions and interested industry ECHO will:

- Create a virtual space containing of our rich cultural heritage
- Set and document milestones in the application of IT in the humanities
- Develop guidelines and standards with an international perspective
- Fertilize the development of IT due to leading edge operations and due to the potential of a synthesis between culture and technology
- Attract young people in particular also from the emerging industrial societies of Eastern Europe to become specialists in this area and to transfer knowledge and the technology to be accumulates by ECHO to their respective countries.

In the main phase ECHO will establish a stable collaborative network which can act as a virtual competence centre as well. ECHO will look for a collaboration with interested industry as well to extend current technological solutions, to promote a commercial exploitation of the found solutions and to define suitable frameworks for providing eContent. In doing so ECHO will be a major contribution on the way to an integrated European Research Area.

With these goals in mind ECHO will contribute to the goals of the EC's eEurope initiative to increase the efforts of providing high-quality and attractive eContent. In parallel, it will make sure that the cultural heritage of Europe will be present also in the medium of the future and serve as the backbone of a European identity.

.

2.2 Concept

The organizational concept of ECHO can best be described with the following picture. The Virtual ECHO Innovation Centre, constituted mainly by the technological experts of the major partners, will participate in actual projects in the humanities and collaborate closely with the local domain experts participating in ECHO. Furthermore, where necessary the virtual ECHO innovation centre will maintain contacts with technological partners regarding the actual projects it carries out.
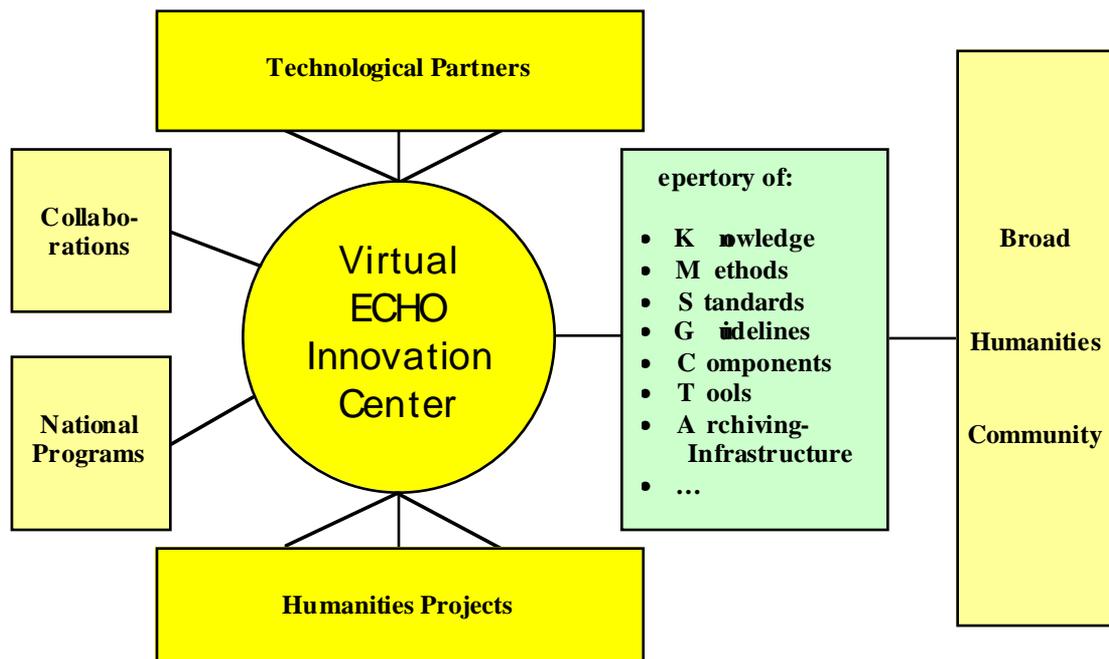
The selection and implementation of the projects is done in close collaboration with the national programs and related activities of the EC-member states. ECHO will look for international cooperation to benefit from the existing knowledge, and also to push forward emerging guidelines and standards where necessary. Although this isn't illustrated in the picture, in this respect ECHO will operate in a highly competitive environment.

The Virtual ECHO Innovation Centre will produce a repertory of technical components and solutions in collaboration with their partners which will contribute to the humanities projects within the ECHO project, and which will also be made available to other interested institutions of the humanities community.

A wide spectrum of measures will be taken to efficiently organize the project work and to transfer especially the knowledge about the usage of key technologies. The Virtual ECHO Innovation Centre will be implemented by establishing intensive communication channels between the

leading technological groups and experts. They will organise regular meetings to identify the technological commonalities between the different areas and to define the building blocks of the technological frameworks to be applied.

Technology experts from the Virtual ECHO Innovation Centre will be temporally detached to the selected archiving projects and share their worktime in the corresponding locations. In doing so a bidirectional transfer of knowledge and skills is guaranteed. Summer schools will bring a number of experts from different disciplines together to talk about infrastructural and technological aspects which the disciplines have in common to define reusable frameworks, guidelines and standards. Finally, the innovation centre also should temporally house a number of guests from external projects and organize training courses to transfer technological knowledge.

To realize the role of ECHO as envisioned, funding has to be provided for at least 6 years under the 6th Framework Program. For steering and controlling the work ECHO will have boards including the involvement of outstanding members of the community.

2.3 Technological Keypoints

The key technologies which will be utilized in the ECHO project have already been mentioned in part 1.2. ECHO will be especially focused on those technologies which:

- Offer the user a motivating and flexible interaction with the material investigated
- Perform background tasks for the user to anticipate his or her actions
- Provides an attractive, multimedia and multilingual interface.

3. 5TH FW PROJECTS

3.1 Goals

For the 5th Framework Program of the EC we suggest an ECHO start-up project with a duration of about 30 months. This start up project will serve as a proving ground for the concept to evaluate the capabilities of the main partners, to work out stable and reliable organizational pillars, and to establish the necessary framework. Furthermore it will be used to prepare the broad network of contacts and to select promising projects for the main phase. During this start-up project, ECHO has to demonstrate its vast potential by producing internationally recognizable results.

.

Therefore, the projects for the pilot phase will have to be built on existing work and knowledge and also be limited in scope. At least three areas of the humanities have to be represented. Based on the background of the main partners, three projects are suggested: the first taken from the area of languages, the second from the area of knowledge history, and the third from the area of the history of arts.

3.2 Language Related Project

Within the pilot phase the creation of a European archive of Endangered and Minority Languages is suggested. The current situation with respect to the accessibility of such resources as part of the European heritage is terrible. Many resources that document languages or the usage of languages are only accessible for a very small group of people directly involved – and often only individuals. Even worse is that very few people know that these resources exist at all. The consequences are that highly valuable material about minority and endangered languages cannot be re-used and sometimes is lost forever. Much of this highly valuable material is also in an unacceptable state.

We propose the joint effort of a number of European institutions that have the kinds of valuable resources required to create so-called metadata descriptions for the material based on standardized formats, and to create appropriate browsable hierarchies. Novel methods using the information in the existing metadata descriptions and the annotations and contributing to the idea of "Semantic Web" have to be developed and tested to generate such hierarchies automatically as well. In doing so, structured subspaces can be created which lends itself not only to easy navigation with the help of appropriate browsers, but also to being the subject of queries operating on structured elements with well-defined semantics. Recent demonstrations such as the one in Lund at the Official Opening Ceremony of the European Year of the Language, have shown the great potential of the described approach. The potential of such technologies can also be seen by the foundation of the Open Archives Initiative. Fast decisions would help Europe to stay ahead and ECHO to play a leading role world-wide.

In the pilot-phase it is also recommended to fund some projects which will include the digitisation of culturally valuable audio and video material which will otherwise be lost, and to link this material with linguistic descriptions. The media components are of particularly great importance for the lively documentation of languages. However, only a few institutions have the necessary human and technical resources available. ECHO will help in this respect and set world-wide standards. To achieve this, it is necessary to adapt the existing multimedia annotation and analysis tools to the special requirements. This will call for the following global innovations:

collaborative working methods between scientists working at different locations in Europe, the extension of existing search techniques to include the speech wave, and the development of a formal framework to interactively define and test grammatical descriptions.

3.3 Knowledge History Related Project

Within the pilot phase corpora documenting historical forms of scientific and technological knowledge shall be made available in the context of an Open Digital Library. The aim is to create an infrastructure allowing to study the emergence of scientific knowledge in Europe. It is planned to digitize key sources representing major epochs in the development of this knowledge (Antiquity, Renaissance, Scientific Revolution) in an exemplary form. These sources comprise, in particular, artefacts documenting ancient technological and scientific knowledge preserved in Pompeii, papers from the Venetian Arsenal documenting the knowledge of Renaissance scientist-engineers, and sources on the emergence of classical mechanics as the pilot science of the Scientific Revolution.

Science as a uniquely powerful form of knowledge without which the modern world and its future are inconceivable, has emerged in Europe under special epistemic, social, technical, and cultural circumstances. Any attempt to understand the conditions under which science was born in antiquity and became a dominant form of knowledge in the early modern period has to rely on a broad variety of historical sources still widely studied in separation from each other by different disciplines such as archeology, classics, the history of art, the history of technology, and the history of science. The information revolution now offers the chance of providing an integrated access to these sources and to develop the tools necessary for studying knowledge development as a long-term process structured by conditions and patterns, a process that has so far remained largely intractable in the context of specialized, local, and mono-disciplinary studies.

In several European countries a number of projects in the humanities have emerged that address the new challenges of the information revolution, attempting to make accessible, albeit with great difficulty, various collections of historical sources on the Internet. The historical materials presently under consideration or already in part available on the net comprise e.g. collections of cuneiform tablets documenting the emergence of mathematical thinking, measurement instruments preserved in Pompeii which document ancient practical knowledge of physics, medieval manuscripts on mathematics, the administrative archives of the cathedral of Florence documenting the technical knowledge of Renaissance artisans and engineers, the papers of Gal-

ileo Galilei, Thomas Harriot, Christian Huygens and Isaac Newton, documenting the birth of classical physics, and a collection of documents illustrating the spreading of classical science in the European periphery.

But even pioneering ventures still suffer, as a rule, massively from a lack of IT support tuned to their scholarly needs. If they could, however, receive such a support and, in addition, join their forces, they would have the critical mass to contribute to an innovative infrastructure for research with far-going implications also beyond scholarship in the humanities. In particular, an "Open Digital Library" for studying the emergence of scientific knowledge in Europe could be created that would contribute to overcoming traditional disciplinary boundaries and offer a new platform for interdisciplinary and comparative research on European culture, a research that is still largely divided by barriers of language and national traditions.

The ECHO initiative could offer the framework for a coordinated integration of the efforts of pioneering projects from various European nations, for providing them with the necessary IT support, and for building, on the basis of such a joint effort, an innovative IT platform for making historical sources available. It is envisioned, in particular, to develop a new kind of multi-media environment allowing to combine data-mining technologies with the exploration of the intrinsic semantic structures of the sources. The pilot phase, aiming at the digitisation of exemplary corpora of historical materials, serves to define the requirements for this platform and initiate its realization.

3.4 History of Arts Related Project

Roman architecture has since antiquity set a standard for European culture and architecture. Especially between the 15th to the 19th century Rome has acted as a European centre which stimulated architectural skills throughout Europe. Foreign artists and architects were sent to Rome to study the newest achievements there and Italian architects and craftsmen were called into almost all European countries. The Roman example was reproduced all over Europe and developed into a kind of common artistic language which paved the way for a truly European cultural self understanding.

The Bibliotheca Hertziana (Max-Planck-Institute for Art History) is setting up a database of Roman architectural drawings of the 17th to the 18th century as a start-up project within ECHO, which finally shall develop into a comprehensive database for architectural drawings. We aim to unite drawings which are now dispersed in various collections all over Europe into a virtual collection. The new electronic media will change the attitude of scholarly work, because they make it possible for the first time to create images, which do reproduce the quality, techniques, structure of paper and colours of the original drawing to such an extent, that it will be possible

to work with them as it was formerly only possible with the original drawings themselves. The database will give broad access to documents which were formerly difficult to consult and can help at the same time to protect the delicate original drawings. Our database will not only set a standard for the use of new technologies in the humanities but will also contribute to a profound knowledge of our common European cultural heritage.

Three components are essential for creating this database: (1) Each drawing will be described by metadata descriptions based on a set of well-defined elements. These metadata descriptions are part of a searchable and browsable universe which allow the user to input complex queries to find exactly those documents he is interested in. (2) With the help of digital cameras high resolution images of the drawings and pictures will be created. Even small details of the drawings can be analysed with the resolution envisaged. (3) Based on the proposed database of architectural drawings, combined and supplemented with archival sources such as historical descriptions and account books three dimensional reconstructions of design and building campaigns will be produced by applying Computer-Aided Design (CAD) technology. The possibility of actively manipulating the architectural objects will give completely new insights which was hitherto almost impossible. New ways of representation and visualization have to be found to support these possibilities.

In future image processing technology might be used to include features of the images themselves in the querying process. In how far such technology can already be applied in the start-up phase has to be checked.

3.5 Preparation of the 6th Framework Phase

Not only in the three areas of the humanities mentioned ECHO will establish a network of European experts and develop re-usable cutting edge technological frameworks and components. To give ECHO the central role in the humanities as intended the 5th Framework project has to be used to broaden this network in a number of dimensions such as domains within the humanities, number of European countries, number of projects, and spectrum of technologies. Echo will also establish the necessary links with industry during the 5th FW phase to ensure a knowledge and technology transfer in both directions.

Under the 5th FW ECHO will spend a considerable amount of time to discuss and test various collaboration models, set-up suitable management structures to handle distributed projects, define measurements which may serve to evaluate projects, and to discuss the ethical and legal issues generally involved in such humanities projects as intended.

.

4. EXISTING TECHNOLOGY AND EXPERIENCE

ECHO will rely on the experience, knowledge, and existing solutions of the partners. At present we can only give the profiles of the core partners.

4.1 MPI for Psycholinguistics

The MPI for Psycholinguistics has a great deal knowledge in many areas of linguistics which would be beneficial for ECHO. The Technical Department of the institute is one of the largest and most professional in all fields supporting psycholinguistics. Particularly noteworthy are: (1) Web-technologies and specifically those which support the Semantic Web idea, (2) multi-media technologies, (3) many years of experience in handling multimodal signals such as speech, gesture and gaze, (4) valuable knowledge in creating 3D animations including moving human-like creatures, and (5) experience in archiving large quantities of multimedia data.

In addition the institute has cooperated for years with other institutes and companies in the area of automatic speech recognition (U Nijmegen), computational linguistics (U Nijmegen, DFKI, U Sheffield, CNR Pisa, ESTEAM), and Artificial Intelligence (U Nijmegen, DFKI).

The institute would contribute to the ECHO project with part of its development staff, its Browsable Corpus tool (Metadata Editor and Browser), the EUDICO tool set, and its computer, network, and storage infrastructure. Dependent on the ECHO context, the MPI resources could be extended and adapted. The institute would take the lead with respect to speech and language technologies and establish a close collaboration with the specialists of the university of Nijmegen in particular.

4.2 MPI for History of Science

The Max Planck Institute for the History is among the pioneering institutions in Europe engaged in making sources of the history of science accessible in electronic form. The sources range from cuneiform texts of ancient Babylonia to the sources of 19th century physiology. A general aim is to reassemble in digitized form sources which have been dispersed over various public and private collections all over the world and thus as a rule are not accessible in their entirety for scholarly research. The Institute has also considerable experience in using cutting edge information technologies in order to create working environments supporting innovative scholarly work on such historical sources.

The Institute is a key partner in several international endeavours such as the Cuneiform Digital Library Initiative, funded by the American National Science Foundation, the Archimedes Project, funded jointly by the National Science Foundation and the German Science Foundation, and the Virtual Laboratory Project on the Experimentalization of Life, funded by the Volkswagen foundation. Within the Max Planck Society the Institute is also involved in joint initiatives such as the project to develop a Model Electronic Library.

4.3 Bibliotheca Hertziana (Max-Planck-Institute)

The Institute, founded in 1912, is one of the leading institutions in art history focusing its research on Italian art and architecture. The institute holds one of the largest and best libraries in art history. Since 1995 the Institute has started to build a multilingual electronic library catalogue in collaboration with the Institute for the History of Arts in Florence and the Centre for the History of Arts in Munich. Within the next two years this joint venture will provide the largest catalogue in the field of art history worldwide accessible via Internet. In order to make rare books available the institute will create a database containing our unique collection of Roman guide books from the 15th to the 19th century. The institute has also begun to digitize Roman city plans from the 15th to the 18th century (ORBIS URBIS) and will set up a Roman vedute-database in connection with it.

The institute intends to digitize its large collection of photographs (more than 500 000) during the next years and to set up a database for architectural drawings. During the last two years a CAD workstation was being set-up to create layered vector-based representations of famous architectural concepts and processes.

4.4 University Nijmegen

The University of Nijmegen has a long standing tradition in computational linguistics and corpus linguistics. Since the early seventies of the twentieth century, linguists of the English and former Computational Linguistics department have worked on the automatic grammatical analysis of English and Dutch. Nijmegen was the first European University offering a full undergraduate curriculum for language and speech technology.

The Nijmegen researchers participate in major projects within the fields of Natural Language Processing and Automatic Speech Processing. Together with the Nijmegen Max Planck Institute, they play a leading role in the Spoken Dutch Corpus, an international project aiming at building a large annotated corpus (6 million words) of spoken Dutch. The main resource for the grammatical description of the Dutch language, the "Algemene Nederlandse Spraakkunst" (or

ANS, the General Dutch Grammar), funded by the Dutch-Belgian Taalunie, is coordinated by Nijmegen researchers. They also carry out a project aiming at encoding the ANS material according to XML standards in order to make it available to a web audience.

Three decades of research on automatic grammatical description have led to a wealth of knowledge on language models, mainly for Dutch and English, but also in a general sense. Since the software developed has always been syntax directed, using separate components developed by the department of computer science, the language specific components of the Nijmegen language models can easily be replaced. Large parts are reusable for other languages.

In addition the Nijmegen researchers have been involved in a number of national and European projects to develop and apply speech recognition and speech synthesis technology.

Apart from the knowledge gathered by their own research, the current Language and Speech department (and future Linguistics department), have established a vast network within the field of computational linguistics. Nijmegen researchers cooperate with many national and international researchers working in the same or adjacent areas.

5. EXISTING CO-OPERATIONS AND NETWORKS

Both projects suggested can build on networks which have already been established. We will only briefly indicate them here.

5.1 Language Related Project

It is estimated that the participation of 15 European institutions from different countries would be sufficient to create the critical mass that we are aiming at for project viability and for the ECHO project to play a major role in international competition.

The following institutions have already expressed great interest in participating in an attempt to create a European virtual language space:

1. MPI for Psycholingistics, Netherlands/Germany
2. University of Lund, Sweden
3. University of Helsinki, Finland
4. MPI for Evolutionäre Anthropology, Germany
5. ELRA, France
6. CNRS/LACITO, France

7. LORIA Nancy, France

8. University of Tilburg, Netherlands

9. University of Lancaster, UK

10. University of Munich, Germany

11. ILSP, Greece

Various discussions have revealed that other institutions from for example Italy, Switzerland, Estonia, Slovakia, and Armenia are also highly interested in joining. Of especial interest is here to extend the collaborations to even more countries from East Europe.

5.2 Knowledge History Related Project

Four "core institutions" plan to launch a pilot initiative, preparing the ground for involving a wider network of projects and institutions all over Europe in the knowledge-related project. The four core institutions are the Max Planck Institute for the History of Science in Berlin, the Institute and Museum for the History of Science in Florence, the Department of History and Philosophy of Science of the University of Athens, and the Institute for Philosophy of the University of Bern. All four institutions are already engaged in larger projects making available on the Internet sources relevant to the emergence of scientific knowledge in Europe. In this context, all four institutions are also involved in developing innovative technologies for the scholarly exploration and public dissemination of these sources. The ECHO initiative would offer a framework not only for coordinating these efforts in order to achieve a joint platform as outlined above, but also boost European history of science as a whole towards an active participation in the information revolution by fostering a broad-based integration of technical and scholarly competencies.

In particular, further institutions are already part of a collaborative network to be involved in the ECHO initiative and strengthened by its outcome. Among them are the Sopraintendenza Archeologica di Pompei, Italy, the State Archive of Venice, Italy, the Opera di Santa Maria del Fiore, Florence, Italy, the Ludwig Maximilian University of Munich, Germany, the Cuneiform Digital Library Initiative involving such institutions as the Ermitage in St. Petersburg, the Pergamon Museum in Berlin, and University of Calfornia in Los Angeles; and the Perseus Project of Tufts University.

5.3 History of Arts Related Project

The Bibliotheca Hertziana is already actively collaborating with the Max-Planck-Institute for the History of Science, the Getty Foundation and other research institutes for art history worldwide to build up the database of architectural designs as described. Further collaborations are

.

intended with various European collections such as from Austria, Germany, Great Britain, and Sweden. The institute has great interest to also establish form collaborations with collections from East European countries, as access to their collections had been hitherto very limited. Contacts have already been established with collections in the Tchech Republic and Poland.

INTERNATIONAL WORKING GROUP

ECHO
EUROPEAN CULTURAL HERITAGE ONLINE

DISCUSSION PAPER

ECHO

EUROPEAN CULTURAL HERITAGE ONLINE

DISCUSSION PAPER

> "Was du ererbt von deinen Vätern hast,
> Erwirb es, um es zu besitzen."
>
> "All that you have, bequeathed you by your father,
> Earn it in order to possess it."
>
> (J.W. Goethe, Faust, der Tragödie - erster Teil, 682-683)

THE

DISCUSSION PAPER

EXPLAINING THE FOUNDATION NOTE OF THE PROJECT

ECHO
EUROPEAN CULTURAL HERITAGE ONLINE

WAS PREPARED BY AN

INTERNATIONAL WORKING GROUP

CONSISTING OF

L. BOVES, UNIVERSITY OF NIJMEGEN, THE NETHERLANDS
E. KIEVEN, BIBLIOTHECA HERTZIANA, ROME, ITALY
S. LEVINSON, MPI FOR PSYCHOLINGUISTICS, NIJMEGEN, THE NETHERLANDS
E. DEN OS, MPI FOR PSYCHOLINGUISTICS, NIJMEGEN, THE NETHERLANDS
J. RENN, MPI FOR THE HISTORY OF SCIENCE, BERLIN, GERMANY
URS SCHOEPFLIN, MPI FOR THE HISTORY OF SCIENCE, BERLIN, GERMANY
PETER WITTENBURG, MPI FOR PSYCHOLINGUISTICS, NIJMEGEN, THE NETHERLANDS

AND WAS PRESENTED AT THE WORKSHOP

HUMANITIES, RESEARCH, AND CULTURAL HERITAGE IN EUROPE

ORGANIZED BY THE EU
IN BRUSSELS, OCTOBER 24, 2001

SUMMARY

In this paper, we present the general concept of a large-scale action aiming at the establishment of the European Cultural Heritage On-line (ECHO). ECHO is an initiative of three Max-Planck Institutes in the field of the Humanities to create a European virtual infrastructure for the digital representation, storage, and conservation of European Cultural Heritage that can be accessed via the Internet without the need for specialist knowledge of the organisation of the collections, and without specialist ICT skills. At the same time this infrastructure will provide the basis for the emergence of novel and powerful research methods and environments for all sub-disciplines of the Humanities, and it will help to create Centres of Excellence and virtual research networks. ECHO will build upon, unify and strengthen the multitude of ongoing attempts to create a digital infrastructure for the Humanities and the preservation of Cultural Heritage.

ECHO intends to reach its goals by adapting emerging ICT technology to the needs of the Humanities and Cultural Heritage, and by developing new ICT tools and procedures specifically aimed at the requirements and needs of the Humanities. The new tools will build upon emerging standards in many of the sub-disciplines of the Humanities, and extend these standards by developing additional layers of semantic intelligence that will allow scientists, professionals in museums and the general public to access and query the growing digital collections in intuitive ways, using their preferred natural language in combination with advanced 3D graphical interfaces and virtual reality representations.

Cultural Heritage –or perhaps more appropriately: Cultural Memory, and the Humanities are intimately intertwined, but also distinguishable. The Humanities is the set of scientific disciplines that has the Cultural Memory as its major object of research. Therefore, it appears that support for the preservation and accessibility of Cultural Heritage also supports the progress of the Humanities. The other way round progress in the Humanities will enhance our understanding of the material and immaterial objects of Cultural Heritage, and their presentation to scientists and the general public.

The first chapter of this Discussion Paper sets the scene, defines the major goals of ECHO, and outlines in some detail the ways in which we intend to approach these goals. The second chapter of this Discussion Paper provides a sketch of the ways in which the ECHO initiative contributes to the economic, political and social objectives of the European Union. The third chapter of this Paper briefly describes visions of the impact that ECHO will have on three important sub-disciplines of the Humanities.

.

1. ECHO

In this chapter we introduce the ideas and considerations that form the basis of the ECHO initiative. We approach the subject from the point of view of the needs of Cultural Heritage and Humanities research, building upon recent and ongoing developments. To set the scene we first circumscribe what we mean by the terms Cultural Heritage and Humanities. Then we proceed to show how present and emerging ICT tools can and should be used to carry Cultural Heritage and the Humanities into a new era, which promises novel and highly interesting opportunities. We define the major objectives of the ECHO initiative, and we finish with an outline of the ways in which we think the objectives can be reached.

1.1 What is Cultural Heritage?

The seemingly clear concept 'Cultural Heritage' is quite difficult to define. Several somewhat different definitions have been proposed in the past. It is not the intention of this Discussion Paper to propose yet another definition. Therefore, we start from the 'working definition' proposed by the World Bank in a document summarising the results of a meeting in Washington, 26-27 January 1988.

> 'Cultural Heritage encompasses material culture, in the form of objects, structures, sites and landscapes, as well as living (or expressive) culture as evidenced in forms such as music, crafts, performing arts, literature, oral tradition and language. The emphasis is on cultural continuity from the past, through the present and into the future, with the recognition that culture is organic and evolving.'

This definition covers two aspects of Cultural Heritage that are essential elements of the basic idea behind ECHO. Firstly, it is said that Cultural Heritage encompasses both material objects and intangible 'objects' such as oral tradition and language. Other definitions tend to restrict the definition to material objects only. Secondly, the definition emphasises the fact that Cultural Heritage implies continuity as well as development and change.

There are two important aspects of Cultural Heritage that are probably implied by the definition of the World Bank, but are not explicitly expressed. The quote from Goethe's Faust emphasises that Cultural Heritage is something that has to be actively acquired, rather than being a passive asset. In addition, Cultural Heritage comes with an intricate set of values, which are subject to evolution and change. Many of the objects in our museums originated as simple artefacts, produced for daily usage. There are a large number of painters and writers who were considered as major pillars of the culture in their time, but who are now completely forgotten. The perception of Cultural Heritage as an active and evolving concept has given rise to the term Cultural Memory as a more accurate expression of its role and workings.

1.2 What are the Humanities?

The National Endowment for the Humanities uses the following definition of 'Humanities':

> '...humanities (include) the study of the following: language, both modern and classical; linguistics; literature; history; jurisprudence; philosophy; archaeology; comparative religion; ethics; the history, criticism and theory of the arts; those aspects of the social sciences that have a humanistic content and employ humanistic methods.'

This definition presents the Humanities as a set of academic and scientific disciplines. Moreover, the enumeration shows the enormous breadth of the ensemble of disciplines that entertain complex mutual relationships. Finally, it shows the almost complete overlap between Humanities and Cultural Heritage. In fact, one might say that the Humanities are the scientific disciplines that study Cultural Heritage. In this context it is interesting to point out that the definition of the Humanities suggests that the World Bank's definition of Cultural Heritage should be extended to cover such fields as ethics, religion, philosophy and law, which may be subsumed under 'culture', but which clearly deserve a mention on their own.

1.3 The State-of-the-Art

The fact that an organisation such as the World Bank gets involved with the value and the preservation of Cultural Heritage proves that these are important issues with major social economic significance and impact. This is indeed the case, as testified by the many and lively –but also quite diverse and dispersed- activities on regional, national, European and global scales attempting to harness modern Information and Communication Technologies to support the preservation, accessibility and valorisation of Cultural Heritage.

At first sight ICT and Cultural Heritage may seem fundamentally incompatible, as the former is conceived as cold and objective, while the latter is conventionally perceived as warm and subjective. However, there appear to be numerous aspects of Cultural Heritage, such as the catalogues of museums and libraries, which are eminently suited to ICT support. This explains the multitude of 'Digital Libraries' projects, and related projects to make the catalogues of large and small museums accessible via electronic means. However, it appears that most Digital Libraries and Digital Cultural Heritage projects are generally local and highly personal initiatives, resulting from coincidental collaboration between librarians, museum personnel, and Humanities scientists on the one hand, and 'local' computer scientists on the other. The natural consequence of this type of fragmentary collaboration is a multitude of incompatible and non-interoperable approaches. This begins with different methods used for digitisation and representation of the data, and continues through the level of annotation and classification systems

.

to disparate search and access procedures. To the defence of the visionary scientists it must be added that independent designs and development enterprises are not the only, and perhaps not the most important, cause of the lack of homogeneity of the representations and access methods. The different requirements imposed by individual collections and specific user populations have also contributed to the extant lack of interoperability.

In that context of emerging non-interoperable developments the awareness of the economic and social importance and potential of (digital) Cultural Heritage has given rise to the Action Line 'Cultural Heritage Applications' as part of Key Action III (Multimedia Content and Tools) of the 5th Framework Programme. In that Action Line a large number of projects and support actions have been launched, many of which address specific technological or organisational problems. Also, much effort has been spent on tracing developments in Europe and abroad, and building information systems that facilitate the discovery of existing activities, agencies, and resources.

Despite the results obtained in these and previous RTD projects and support actions, there is general agreement that the basic problems of the lack of interoperability and incompatibility remain to be solved. Consequently, policy papers such as the 'Florence Agenda' and the 'Lund Principles' emphasise the need for the creation of Competence Centres that are able to show appealing and convincing examples of what can be accomplished through using advanced technology on the basis of de facto (and eventually formal) standards and best practice procedures. Also, the need for better access procedures (based on advanced annotation, but also thanks to novel interaction and user interface techniques) is still recognised as extremely urgent, not least because of the relatively large 'cultural' distance between the Humanities and Cultural Heritage on the one hand, and ICT on the other.

The situation outlined above for the field of Cultural Heritage is almost identical to the situation in the Social Sciences, and especially in the Humanities, a set of disciplines that is characterised by a large degree of fragmentation caused by previously insurmountable differences between the object and the goals of the individual sub-disciplines. This is reflected in deep gaps, not only between the individual sub-disciplines themselves, but also between the ICT support, which is highly local and specialised in nature, resulting in non-interoperable solutions. Here too, different requirements imposed by the research traditions and the raw data have contributed to the emergence of non-interoperable resources and tools. Therefore, there is an urgent need for the establishment of an infrastructure to make existing tools interoperable, and to improve accessibility of digital resources across language borders for cross-discipline and comparative research. Recent developments in several ICT disciplines, for example image processing, data base technology and language technology, hold the promise to remove the borders that have

separated the Humanities disciplines and to enable the restoration of the Renaissance Homo Universalis who can access, use and integrate the latest research results of all disciplines in the Humanities.

1.4 The Future Role of the Internet in Cultural Heritage and the Humanities

It is easy to understand why information technology and the cultural disciplines have long been considered as fundamentally incompatible, an opinion that is still held in some circles. Until recently, there was effectively no useful way to digitise material objects and to make it possible to process the virtual representations by means of digital computers. However, there are notable exceptions to this 'rule'. Texts, as a form of information for which digital representations have been developed at the inception of the digital computer, are a case in point. Computer tools have been widely applied to the study of ancient Greek and Roman texts. These texts are alphanumeric and thus allow for digitisation by means of standard procedures. Moreover, they form a finite amount that is small enough to be processed by the computers from the 1970s. It should be noted, however, that the ASCII representation of the original manuscripts form only the tip of the iceberg, since all information encoded in the handwriting and layout of the manuscripts has essentially disappeared.

Only very recently the technology has become available to construct 'true' virtual representations of material objects, ranging from essentially two-dimensional manuscripts, via quasi-3D objects such as paintings, to full 3D objects such as sculptures and even landscapes. Harnessing the prospective full capabilities of these emerging virtual representations offers completely new opportunities, not only to do research in the humanities, but also to preserve full representations of material and immaterial manifestations of 'culture', even if they are inherently transient, such as is the case with classical and folk music, dance and theatre performance.

The Internet holds additional prospects for the Humanities and Cultural Heritage, in that it allows virtual representations of material objects physically located on different continents to appear on a single workstation. In this way medieval diptychs and triptychs that have been separated in the past can be re-united, in a representation that is sufficiently real for most research in the History of Art. Not only diptychs and triptychs have become separated. Many large paintings have been cut into two or three pieces to raise the total market price. The emerging virtual reality representations, combined with the global access through Internet, will allow us to virtually restore these paintings. Also immaterial manifestations of culture, such as dance, music, folklore and language will profit from the emerging digitisation methods and Internet. Small communities that have been living in diaspora for centuries will be able to virtually re-

unite, celebrate, and speak in their native language. These are some examples taken somewhat arbitrarily from a much larger set of opportunities that modern and existing ICT technology can offer to the Humanities and Cultural Heritage.

1.5 How to Realise the Promises?

The picture painted in the previous paragraph is –unfortunately- somewhat futuristic. There is no doubt that the compute power, storage capacity and transmission bandwidth needed for the type of applications sketched above will become available in the next decade. However, the hardware and communication requirements are only part of the puzzle. To make virtual representations of all forms of Cultural Heritage available in such a way that they can be accessed and used by all scientists in the Humanities, and by the people who form communities and cultures, much more has to be done. One obvious bottleneck is the digitisation (and as a corollary the preservation) of the objects and manifestations. That process is generally recognised to be extremely expensive, and also very difficult, if only because no generally agreed procedures exist. In the worst case this may result in 'digital obsolescence', due to the use of digital representation formats for which processing software is no longer available. The investments in the digitisation proper will never be recovered, in whatever terms or measures (economic, social, etc.), if the digital representations cannot easily be accessed and used for a long time to come. Access may be through Internet, but even then (or perhaps especially in the case of Internet access, where people have come to expect interoperability) the information has practically no value if it cannot be used in the same easy and standard way on all computers. Therefore, it is crucial to have appropriate standards, for the digitisation of Cultural Heritage, for the annotation of the digitised (material and immaterial) objects, for their representation and their use.

It is clear that digital resources alone are not sufficient; suitable software to access, analyse and exploit these resources is equally essential. Existing tools for accessing and processing digital representations of the basic material differ widely between the sub-disciplines, not in the last place because of different domain-specific requirements. The full promise of an interoperable digital infrastructure can only be realised if technology, software and procedures are available for the exploitation of the basic resources across the extant borders set by the research traditions and the different types of objects investigated by the Humanities disciplines. We will call the framework of technologies, procedures, best practice guidelines, and formal and de facto standards the digital infrastructure for the Humanities and Cultural Heritage. In creating that digital infrastructure ECHO will, of course, build upon and add value to the growing set of existing and emerging tools and standards in the individual sub-disciplines of the Humanities and the individual fields of Cultural Heritage.

1.6 The Objectives of ECHO

The ulterior objective of ECHO is to create the digital infrastructure for the Humanities and Cultural Heritage that will enable cross-disciplinary use of resources and consequent ground-breaking developments in the methodology and research traditions. One of the most important vehicles to reach that goal ECHO will show appealing examples of the added value of the use of advanced technology in the form of standards for data representations, annotations, and processing tools that support collaborative research that cuts across national, language and disciplinary borders. These examples will lead to a modular, self-organising infrastructure in which more and more data will be present that can easily be related. In that way ECHO intends to contribute to long-term policy goals such as the creation of a European Research Area.

The problems involved in making all manifestations of Cultural Heritage accessible on-line in digital representations that are standard, yet support creative uses for groundbreaking research, are so large and manifold that we cannot even begin to imagine the eventual ideal solution. This situation is familiar in the ICT industry, where completely new hardware and software were used to develop completely new services. All successful developments were based on, and gave rise to a kind of de facto standard, some of which were later adopted as formal ISO standards. ECHO intends to achieve a similar feat. By setting highly visible and appealing examples we expect to be able to lead the way towards a growing set of best practice guidelines and de facto standards that will allow the Humanities community to enter a completely new era.

A new era for the Humanities and Cultural Heritage can only be reached if the barriers of the past and present can be overcome. It has already been pointed out that the different requirements of the individual specialisms and their raw data and resources are a major contributor to these barriers. ECHO intends to break down the barriers by developing technology and tools that support access to the primary and secondary data in highly intuitive ways, based on the use of the combination of natural language queries and drawings or pictures. At the same time technology will be developed that allows to combine (semi-)formal annotations of independently created collections (which imply a human-made classification of the data) with information and relations derived from the data in those collections by means digital processing. This will allow scientists to detect new links between events and phenomena that have hitherto been investigated in isolation. Eventually, newly discovered relations can be added to existing annotations, and thus made available for future research. This will lead to a rapid growth of the much-desired incremental progress of the Humanities.

By making resources and processing tools accessible via Internet ECHO will open an era in which collaborative research in the Humanities will replace the solitary practice of the past. Digital access to full representations of the material and immaterial manifestations of Cultural

Heritage will enable top-level scientists working in smaller universities, without direct physical access to manuscripts, old prints, paintings, etc., to contribute in full to the advancement of their fields.

In concrete terms, ECHO intends to deliver a comprehensive set of tools to support the research in the Humanities, by developing software and procedures that will allow high level conceptual processing of virtual representations of material and immaterial objects of Cultural Heritage irrespective of the details of those representations, and across the borders of the national or disciplinary environments in which the basic representations were developed. It goes without saying that data and tools will be accessible via Internet. To obtain optimal results, it is mandatory to establish intensive interactions between data providers, technology developers and users. In addition, ECHO will create a Centre of Excellence for research in the Humanities. The most appropriate form and organisation of that centre remains to be determined. This Centre will be the hub in a Network of Excellence in Humanities Research, which in its turn will make a major contribution to the establishment of a European Research Area.

For the general population, the results of ECHO will become apparent in the form of significant support for emerging interactive virtual museums, where visitors can determine their own view of the objects, in the setting and combination of their choice. The results of ECHO will also support and strengthen ongoing developments to make the manifestations of small dispersed cultural communities accessible on-line. By doing so, the activities of ECHO will contribute to the preservation of these cultures in an increasingly global environment.

1.7 How are the Objectives of ECHO Pursued?

It is clear that the full vision of a digital infrastructure for the Humanities and Cultural Heritage requires an enormous effort, and developments in many fields as diverse as image processing, distributed databases, intelligent communication protocols, language technology, human-computer interaction, and the development of new research paradigms in the Humanities disciplines proper. ECHO does not intend to actively pursue developments in all these fields. On the contrary, the basic philosophy behind ECHO is to adopt, adapt and integrate developments in all the related fields of science and technology. Yet, it will be inevitable to also undertake R&D in some fields, especially those that relate most closely to the Humanities. In those fields ECHO intends to bring about new and much more productive tools and research paradigms.

The Max-Planck Gesellschaft is willing to take the lead and to coordinate the ECHO initiative. Together, the three institutes that have taken the initiative to propose ECHO cover a considerable part of the broad field of the Humanities. Moreover, these institutes play leading and directive roles in their own sub-disciplines and all three have close relations to R&D in the broad field of ICT.

1.7.1 Common Technology Framework

ECHO intends to build a comprehensive set of tools for the deployment of advanced ICT in the Humanities. These tools will integrate the latest developments in all relevant fields (image processing, data base technology, language technology, etc.). The tools will be documented, and made available to all scientists who carry out research in the Humanities, and to all organisations and agencies active in the preservation and presentation of Cultural Heritage. The Centre of Excellence will have sufficient resources to support scientists and other workers in the field in the use of these tools.

ECHO intends to make its technology support function bi-directional: it does not just want to help other scientists to use the tools and procedures it developed, it also intends to work with scientists to adapt the tools to specific needs of specific research projects. In exceptional cases ECHO may undertake to develop completely new tools if an interesting new research project in the Humanities or a new activity in Cultural Heritage cannot effectively be performed with the existing tools. Of course, ECHO will only actively pursue the development of new tools in the fields of its own core expertise. If additional developments are needed in related disciplines, ECHO will act as an intermediary between the party demanding new developments and the parties who might be able to supply the new tools.

1.7.2 Standardisation and Interoperability

As mentioned above, 'standards' are absolutely essential to achieve the objectives of ECHO. Only if data representations, Application Programming Interfaces and procedures are sufficiently standardised will it be possible to provide the hardware and software modules that are required for advanced applications in the field of the Humanities. A sufficient degree of standardisation is also a necessary requirement for making tools and resources interoperable. ECHO does not intend to develop and impose its own set of standards. On the contrary, ECHO will work with existing public and private organisations that have been active in the development of standards for the digital representation and preservation of Cultural Heritage. In this collaborative enterprise ECHO intends to play an active and guiding role, by acting as an intermediary between the groups and organisations that are already active in the field (such as Digital Libraries initiatives) on the one hand and ICT fields that do not yet have intensive contacts with the Humanities (such as Image Processing). By doing so, ECHO will push the emerging standards

.

in the fields that are not in the core of the Humanities disciplines in the directions that hold most promises for the specific needs of the Humanities and Cultural Heritage. To achieve this goal the partners in ECHO must maintain their highly visible and leading position in the European and international communities in all the fields that relate to Cultural Heritage.

ECHO will promote the emerging standards, tools and procedures by showing how their use results in superior performance in Humanities research and in Cultural Heritage preservation. It is expected that convincing examples of the opportunities offered by the use of standard tools and procedures will attract attention and interest of other scientists, who will then assimilate these tools and procedures in their own practice. Interoperability of existing and emerging resources and tools will be pursued by developing a new generation of knowledge and technology that is able to abstract away from the implementation and annotation details of existing resources. This will be done by adding increasingly powerful semantics to the formal syntactic devices that are currently available to process digital resources on the basis of annotations. ECHO will support the development of tools that can process the raw data that is indexed by the annotations, in order to automatically complete the annotations, and to automatically fill the gaps that appear when scientists ask novel questions. The same approach will be used to enable queries that address independent resources that may contain relevant information, but that could not be combined previously.

1.7.3 Language Tools

Despite their formality annotations of material or digital resources in Cultural Heritage and the Humanities are basically linguistic in nature. For one thing, annotations are usually created in the language of the country where the resources are being built, or in the Lingua Franca of the discipline (that is not always English). The fact that metadata tend to be linguistic is just one reason why 'Language' has a special role in the Humanities and Cultural Heritage. Large –and essential- parts of Cultural Heritage exist in the form of texts or stories, and therefore in the form of language. The 'great books' of the big religions are probably the most obvious examples; for cultures that did not use to have a tradition of writing, oral stories play a similar role. In addition, language is one of the most prominent features that define and unite societies that have their own culture. Last but not least, language has traditionally been the most important 'tool' of the Humanities: most of the scientific accomplishments of the Humanities disciplines consist of texts. This holds even for disciplines such as the History of Art, where the basic objects are material.

The many ways in which 'language' plays a role in the Humanities disciplines and in the preservation and accessibility of Cultural Heritage naturally gives rise to special attention to language in all the ECHO initiatives."

2. CONTRIBUTION TO EU POLICY OBJECTIVES

The objectives of ECHO fit seamlessly with several goals that the European Commission has set itself and the governments of the member states. The e-Culture and e-Europe programmes are clear examples.

ECHO is also explicitly conceived as a major contribution to the establishment of a European Research Area. In the past and present scientists in the Humanities formed small virtual, inherently international, but at the same time rather closed communities of specialists. These communities were –and still are- by necessity international because no single country can afford to support research groups in the field of –for example- Sumerian language and culture that have sufficient critical mass to be able to advance their science independently. This holds even for the classical European cultures, such as ancient Greek and Roman. These specialist communities were rather closed because only a small number of specialists had access to the basic resources in the form of material objects preserved in museums and libraries.

By making virtual representations of these basic resources accessible via the Internet ECHO will remove some of the most fundamental impediments that have affected the development of the Humanities research. Also by making available completely new kinds of tools and developing a completely new generation of research paradigms ECHO will lay the basis for a new kind of Humanities research preserving the flavour of international collaboration, but also offering opportunities for scientists of neighbouring disciplines to make contributions, as the original objects can be accessed and studied from new and different perspectives. For example, due to the accessibility of information about other (older, contemporary and subsequent) cultures in the Middle East and Mediterranean areas, scientists specialising in Mycenae culture will be able to discover links and influences that were hitherto impossible to detect. Accessibility will not be limited to 3D digital representations of the original objects; the fact that these objects, and especially the scholarly literature relating to them, will be searchable for scientists from related disciplines will be even more important. By creating bridges between the disciplines in the Humanities ECHO will not only contribute to breaking down the borders between the member states (and between Europe and the rest of the world) but also to removing the obstacles that have forced the specialists in the Humanities into unfortunate isolation.

2.1 Community Added Value and Contribution to EU Policies

It is evident that ECHO can only exist at the European level. No single organisation, nor any single member state is able to achieve the goals set by ECHO, viz. to create the digital infrastructure for the active preservation of European Cultural Heritage and the support of Human-

ities research in the 21st century. No single organisation or state can provide the critical mass that is needed to set the standards that are needed for the digital infrastructure, nor could any single organisation develop and promote all the necessary tools and procedures. At the same time, by being fundamentally European in nature, ECHO will also serve to strengthen and foster collaboration between the member states. After all, ECHO is designed to give much more than it takes, in that it makes available tools and infrastructures that no single organisation could ever have provided.

ECHO is conceived as a major contribution to several of the EU policies. We envisage ECHO as an essential contribution to the e-Culture programme. In fact, we are convinced that a Centre of Excellence and the toolbox that this Centre intends to provide are absolutely essential pre-requisites for the achievement of the eventual goals of the e-Culture programme.

ECHO will generate multiple spin-offs in the form of digital Cultural Heritage products. Virtual museums are just one example. We foresee a wide range of multimedia products targeted first at the professional market, but also soon at the very large market of interested laypersons wishing to prepare their travel and visits to the physical museums and monument sites. Therefore, we also foresee major contributions to the e-Europe initiative.

2.2 Contribution to Community Social Objectives

By far the most important contribution to the Community social objectives that ECHO has set out to make relates to the opportunities for the preservation of the cultures of small communities that have become dispersed due to unfavourable economic conditions in some geographical areas, or to unfavourable political conditions in the previous centuries. Today, small cultures, especially in the more remote areas of the EU are threatened by extinction, because the communities that support the cultures are shrinking in size. Also larger cultures –often connected to larger languages that dominate political discourse and economic activities- are acquiring increasing importance for those who stay in those regions. By creating a digital infrastructure for Cultural Heritage those who leave the region can effectively keep in touch with their native cultures, and even make their own contributions to its development and preservation. At the same time, those who stay in the region will feel strengthened in their attempts to keep their own culture alive, because they can perceive it as equally important and valuable as the intruding larger culture, thanks to the fact that similar facilities are available.

The tools and procedures to be developed by ECHO are initially targeted at a professional environment, consisting of scientists in the Humanities and museum staff. However, we have every reason to expect that these tools and procedures will also be made available in forms that are

suitable for use by interested laypersons. This will result in a range of products that will enhance the well being of many citizens, because they will be able to enjoy the rich Cultural Heritage of Europe (and of the rest of the world) from their armchairs.

Very soon these tools will also enter secondary and primary school education, where they will have an enormous impact on the possibility of arousing interest in Cultural Heritage among the pupils. This too will contribute to the preservation of small, potentially or actually endangered cultures.

2.3 Economic Development

We expect that ECHO will be at the basis of a completely new, and very large new business sector, which may be termed 'virtual cultural tourism'. We foresee the emergence of a wide range of products that will complement (and perhaps eventually substitute) the travel and museum guides that are presently on the market. By judicial deployment of emerging ICT tools, in combination with clever transmission protocols, distributed databases and image processing it will be possible to develop affordable museum guides that can be used to prepare for physical visits to museums, cities, and monument sites. The use of virtual reality will make very significant additions to what is presently possible with printed guides, even if they contain pop-ups that form 3D representations of buildings. Virtual reality technology will allow customers to move through the cities, buildings, and monument sites at their own will and discretion, and to look at the objects from all angles (even from below, a viewpoint that is usually never available, not even in real museums). We expect a large demand for these products, especially from the older generation, who tend to remain active after retirement.

Commercial exploitation of the results of ECHO is, unfortunately, not without problems. The most important problem that must be tackled and solved relates to Intellectual Property Rights. Museums own the objects in their collection in every way imaginable way, including the digital representations of those objects. Therefore, a legal and commercial infrastructure must be developed that allows publishers and multimedia companies to license the rights to use digital representations in their products.

Another potential bottleneck is the creation of the digital representations themselves. ECHO will make an active attempt to stimulate museums and scientists who create digital representations of material objects of Cultural Heritage to do so in a standardised manner, but also to ensure that those representations become available for other scientists, and eventually also for commercial use. We are confident that the most important objects in all major museums will become accessible within less than a decade. However, the full potential of the ECHO vision can only be realised if also the less well-known objects from the big museums, and the collec-

.

tions of the less famous museums are added to the on-line virtual collection. Many organisations, among which the World Bank, have already pointed out that the digitisation of the existing global Cultural Heritage will require a very large budget, in addition to very powerful and comprehensive agencies.

The practical, financial and organisational problems involved in the capturing of manifestations of living and oral cultures are probably orders of magnitude larger than the already gigantic problems connected to the digitisation of the material Cultural Heritage. ECHO will make every effort possible to convince state, regional and local authorities of the need to make funds available for the digital preservation of the oral cultures. Perhaps the most important argument that ECHO can contribute is the warranty that the investments in the digitisation will not be lost because of the lack of appropriate tools and software to use the data. Provided that the best practice guidelines and standards of ECHO are adhered to, the digital products can be used on all popular hardware and software platforms.

2.4 Science and Technology Prospects

The eventual goal of ECHO is little less than a revolution in the paradigms of Humanities research. The manifestation of this revolution will differ in different sub-disciplines. For example, in the History of Art scientists will profit from the opportunity to view (and eventually perhaps even to touch) virtual representations of objects in ways that were previously not possible. Moreover, these novel views of individual objects can be combined with similar views of other objects, at remote geographical locations, and that may never have been in the same location as the object under investigation. For linguists and literature scientists it will become possible to have simultaneous access to digital representations of manuscripts and old prints and to the full literature about the text under analysis.

The emerging ICT tools that ECHO intends to foster and propagate will fundamentally change the concepts of erudition. ECHO does not intend to make these concepts obsolete. On the contrary, we believe that they will always remain pivotal for high quality research in the Humanities. However, ECHO will bring about a qualitative change in the way in which these assets are used in daily work. Direct access to texts, citations, but also to knowledge, through multiple representations and annotations will make research in the Humanities much more productive, and also much more 'objective'. As the confidence that all existing evidence is brought to bear in a new interpretation of old (and new) facts, the risk that efforts get wasted in futile discussions (the single most troublesome aspect of the Humanities in the eyes of some observers) will be reduced.

Proper and judicious deployment of ICT tools in the Humanities will also contribute to a re-newed esteem for this extremely important branch of the academic disciplines. During the last two centuries the Humanities have suffered from a loss of credibility, due to the fact that the results of the research seemed to be less objective and falsifiable than in the natural (and to some extent also the social) sciences. The novel research paradigms that ECHO will develop, enable and propagate will help to restore that credibility.

3. VISIONS FOR THREE DOMAINS

In this section we briefly describe examples of the ways in which ICT tools have contributed to advanced research in the Humanities, how these tools have contributed to the preservation and dissemination of Cultural Heritage, and especially of the ways in which the kind of tools, best practice guidelines and standards that ECHO intends to develop can foster future research and development in the fields that we intend to address.

3.1 The History of Science

3.1.1 Introduction

Science as we know it, as a specific form of knowledge on which our modern civilization is based, has emerged in Europe. Hence the history of science deals with a vital part of the Euro-pean cultural heritage. Nevertheless, among the historical disciplines the his-tory of science is still relatively young and has long been neglected because it seemed that the unfailing progress of science needed no historical memory. As a consequence, history of science in Europe still tends to be split up into numerous sub disciplines such as the history of physics, the history of biology etc., auxiliary to the classical disciplines of the natural sciences.

The history of science has only recently been taken seriously as providing a resource for reflect-ing on the challenges of modern science and technology, ranging from ethical issues in gene technology via issues of knowledge organization in the information revolution to the role of cul-tural contexts for scientific innovations and their acceptance. The questions to the history of sci-ence emerging from such a perspective are by their nature interdisciplinary and can hardly be addressed by a discipline fragmented into narrow areas of specialization.

The information revolution has now created the potential for new interdisciplinary approaches to the history of science that might be adequate to this challenge. As a matter of fact, for the first time overarching issues such as the role of the ancient urban civilizations for the emergence of science, the relation between technical and scientific knowledge in the Renaissance, or the

.

nature of the scientific revolutions giving birth to 20th-century science can be systematically pursued on a broad empirical basis with the help of digital libraries of all relevant documents. A history of science exploiting this new potential may provide, at the same time, an ideal test-bed for all disciplines dealing with historical documents, being concerned with documents in virtually all European languages and from practically all historical periods.

3.1.2. ICT Potential

The development of information technology has fundamentally changed the potential of research in the history of science. The electronic storage of historical sources improves dramatically their accessibility and allows for new and powerful methods of retrieval. Electronic archives and software tools can be used to create advanced working environments making it possible to integrate historical details into coherent models of historical developments.

The realization of this potential depends on the availability of a wide range of sources accessible to the scientific community as a whole and on interdisciplinary scholarly co-operations on this material. Such co-operations will be characterized by a novel unity of research and dissemination and will draw on the potential of the Internet to cut across the traditional distinctions of research institutions, universities, and libraries.

A history of science brought to the digital age may, however, not only profit from the potential of the Internet to give rise to a new organization of knowledge but also substantially contribute to its realization. As a matter of fact, the history of science is directly concerned with some of the great challenges associated with the transition of the scientific culture into the age of the Internet, ranging from the problem of long-term electronic archiving of scientific documents to that of studying the relationship between the structures of scientific knowledge and its various forms of external representation.

3.1.3 Present Situation

At present, the necessity of exploiting the new technological potential for purposes of research, teaching, and dissemination in the history of science is widely acknowledged. Nevertheless, the potential described above remains largely unused. In fact, without an adequate infrastructure the gap between scholarly work and technological innovation remains, on the whole, insurmountable.

Still only a few exceptional pilot projects demonstrate what is possible in principle along the lines outlined above, such as an initiative aiming at reconstructing in the electronic medium the Babylonian archives of the third millennium B.C., the publication of an electronic representation of Galileo's manuscripts on the Internet, or the creation of a digital collection documenting the spreading of classical science to the European periphery.

But even such projects run the risk of remaining isolated endeavours if no adequate infrastructure is created turning them from exceptions into model ventures and spreading the know-how they have accumulated. In fact, numerous data cemeteries, web projects with dead links, outdated technology, and unkept promises point to the substantial difficulty of initiating a sustainable development leading to standard solutions for standard problems.

3.1.4 Vision

The vision of the ECHO initiative is to create an encompassing infrastructure for the historical heritage of European science in the Internet, constituted by a self-organizing network of open digital libraries. This infrastructure will, at the same time, represent a test bed for exploring strategic problems of science in the Internet age, such as archiving and knowledge organization. Its realization will generate new technology but also involve harvesting the results of advanced pilot projects. The initiative will create and implement living standards and tools capable of sustaining a widespread development of digital libraries, and establish a culture of connectivity between them.

A first step is the identification of standard problems for which standard solutions can be provided in close cooperation with the other branches of the ECHO initiative. Among the standard needs of digital libraries in the history of science is a universal navigation, display, and working environment allowing to navigate across a distributed collection of digital libraries, to display sources as images, texts, and translations, and to interactively work with these sources on the web. In addition, specific working environments for standard scholarly tasks such as transcription, translation, indexing, and commenting should be developed and implemented. Tools based on language technology such as automatic linking from texts to dictionaries in various languages and from various periods should become a standard module for all digital libraries generated under the ECHO framework. It is expected that the implementation of this base technology alone will substantially change the situation of the history of science by enabling a growing number of projects and institutions – whether involved in research, education, or archiving – to make their sources available on the net.

A further step is the creation of specific new tools for exploring digital collections of scientific documents, integrating the scholarly competencies of historians of science with the innovative potential of the new technologies. In particular, tools are needed which transform the results of scholarly work, e.g. the analysis of "ontologies" such as the organizational structure of an institution or the deductive structure of mathematical treatises, into instruments of navigation within a large collection of digital documents. As a matter of fact, the "co-exploration" of on-

tologies and data constitutes a major challenge of knowledge organization in the Internet age, a challenge that the history of science may help to take up by providing model solutions within its domain of competence.

The envisaged developments should not address any project or collection of historical documents in isolation but rather lay the ground for integrating them into a common framework characterized by connectivity and a guarantee for long-term access. The long-term access to a growing array of distributed digital collections, in particular, has to be ensured by an infrastructure that helps to implement standardized metadata, stability of reference, as well as to maintain the unrestricted availability and longevity of the data. The outcome would be digital laboratories with content-based ontologies allowing intellectual access to the milestone of the European scientific heritage, such as the creation of science in the ancient urban civilizations, the Scientific Revolution of the Renaissance, or the dawn of classical science around 1900.

3.2 The History of Art

3.2.1 Introduction

The importance of our knowledge of literature and history in order to recapture the sense of our common culture has been underlined elsewhere in this document. Here we would like to emphasize another field of the humanities as a vital factor of identification: the works of art and architecture. They represent the visual reminder of our European Cultural Heritage. They demonstrate the great variety and colourful diversity of artistic creativity and, at the same time, they testify of our common culture to an extent far beyond all other disciplines. It is these works of art – the Fontana di Trevi or paintings by Rubens and Rembrandt – which have formed our impression of a truly European culture and civilization.

3.2.2 ICT Potential

New technologies have changed the methods and possibilities of research in art history dramatically during the last years and will do so in the future. Databases that include images are a new and revolutionizing approach to all disciplines dealing with art and architecture. The digitisation of images can open up all the hidden treasures in public and private collections, and can give unlimited access to works of art that for conservation reasons were until now only available to very few specialists. Europe was slow in using the new technologies compared to the US, and especially to Canada. But now all over Europe collections are starting to set up inventories and databases. The digitisation of images is still a problem, however, due to the lack of user-friendly digitisation and access tools (one of the main topics of ECHO), and the financial resources that are necessary to digitise these complex data.

3.2.3 Present Situation

Roman architecture of the 16th -18th centuries set a standard for European art and architecture. Foreign artists and architects were sent to Rome to study the newest artistic achievements and Italian architects and artists were invited to almost all European countries. The Roman example was reproduced all over Europe and developed into a kind of common European artistic language that paved the way for a European cultural self-understanding.

The Bibliotheca Hertziana (Max-Planck-Institute for Art History) is setting up a database of Roman architectural drawings of the 17th-18th century. The aim is to unite drawings relevant to the subject and time given which are now dispersed in various collections all over Europe into a comprehensive virtual collection. The new electronic media will change the attitude of scholarly work, because they enable us for the first time to create images which reproduce the quality, techniques, structure of the paper and colours of the original drawings to such an extent, that it will be possible to work with them as it was formerly only possible with the actual original drawings. We will be able to reconstruct projects, which were never executed, but also to reconstruct the process of inventing an architectural work, using the possibility to introduce three-dimensional virtual reconstructions as well. The database will give broad access to documents that were previously difficult to consult and can help to protect the original drawings.

The Roman public monuments have set a striking example of a database with their "Roma Tridente" project, as well as the Univerità Roma 3 with "CROMA – Una struttura di ricerca per Roma."

The database will not only set a standard for the use of new technologies in the humanities but will also contribute to a profound knowledge of our common European cultural heritage. The database will be set up in collaboration with various European collections. Besides Roman institutions we are especially interested in collaboration with collections from east European countries, as access to them had been hitherto very limited. We are in contact with collections in The Czech Republic and Poland. Collaboration has also begun with collections in Sweden, Austria, Germany, United Kingdom, and France.

3.2.4 The Vision

A database like the one described above is only the starting point for a much more experimental approach to reconstructing design processes, analysing architectural structures and virtual reconstructions of projects never executed or lost – all of which has not been possible until now to such an extent.

The use of computer-aided design (CAD) and computer-based building-survey in recent times gives new impetus to classical research- and representation methods. The dramatic advancements in this field require an ongoing discourse on how to best utilize the scholarly and scientific potential of this fast-growing computer-technology without becoming an end in itself. Time saving and easy tools are absolutely necessary to diminish the burden of keeping abreast of the technology.

CAD-based presentations prepare accurate three-dimensional virtual copies of buildings but also of projects never executed. We are about to do so with Filippo Juvarra's famous Roman Academy design for a Royal palace of 1705 (which was copied by young architects from all over Europe). We will convert the drawings into three-dimensional models, make them "move," to explore their spatial quality, and will enter the interior of a structure never built.

Even with existing buildings, the virtual reconstruction provides new insight into the structure of the architecture, e.g. about the plan geometries or the geometrical layout of the vault. These geometries must be known precisely, because imprecise approximations will cause gaps, holes, offsets and incongruences in the CAD-model. The obligation to be precise is one main methodological advantage of CAD. Still more interesting and hitherto unknown advantages will appear, when more and more projects will be explored and "come to life" or lost and altered states of a building can be reconstructed.

In the case of the Roman basilica San Giovanni in Laterano, for example, where Francesco Borromini's projects concerning vault and facade of about 1650 remained unbuilt, Christoph Frommel and Augusto Rocca De Amicis made reconstructions on the basis of original drawings that were visualized by CAD –animation. An exact, made-to-scale and three-dimensional reconstruction is thus not only possible but also desirable: only by doing so we can visualize the space of the two-dimensional-drawing since Borromini's use of space is a key element in his architectural designs.

Comparing three-dimensional reconstructions, building surveys, original drawings or historical surveys by overlaying them in scale, is another scientific procedure, which is strongly facilitated and enabled through CAD. For the widely acclaimed Borromini exhibition in Rome and Vienna (1999/2000) members of the Bibliotheca Hertziana (C:L.Frommel/H.Schlimme) made for the first time a virtual reconstruction of the design process of the facade of San Carlino alle Quattro Fontane, which yield unexpected insight not only into the design process itself but into the actual building techniques.

56

New software enables us to explore the spatial qualities of a building. The new electronic measurement of Borromini's Roman church S. Carlino by Alessandro Sartor (University of Rome) gave striking results about the techniques implied, but also proved the reliability of the 17th century survey drawings.

New building survey methods, such as Single Image Photogrammetries as CAD-plug-in are now in use, and we also use a tachymeter with laser-telemeter in online mode.

Only international, large scale collaboration will be able to explore the exciting discoveries of space and time, uncover the hitherto hidden technical skills and achievements of the past, which formed our cultural heritage. The new techniques will lead towards a more profound appreciation of the great artistic and technical qualities.

3.3 Language as Object, Medium and Code

3.3.1 Introduction

The pivotal role of language in Cultural Heritage and the Humanities research motivates the need for special attention to 'language' and especially to language processing. Language is an object of the Humanities in the form of all branches of Linguistics, including Historical and Socio-linguistics. It has long been known that languages change over time, and that these changes index other changes in culture and society. There is an urgent need to record the present language situation in Europe, and to relate this record with older, and often incomplete records of historical situations. Today unprecedented technical facilities exist to capture 'language', not only in its written forms, but also in the form of audio-visual recordings.

Language has always been the single most important medium of the Humanities. It is not by coincidence that the Humanities are first and foremost associated with books and libraries. In the course of the history many natural languages have been used in the Humanities Discourse. This multilinguality constitutes one of the major stumbling blocks in the research, not only for human scientists who must devote large parts of their time to learn the relevant languages, but even more so for automatic tools that are used to support the human scientists in processing the primary and secondary sources.

Libraries, museums and collections have traditionally been the natural resources of Humanities research. These resources used to be accessible through catalogues, which can be viewed as the first forms of annotations, or as a usage of language as a kind of code to label and classify 'objects'. All scientists in the field know the experience of missing potentially relevant links and references only because they are given in unexpected wordings, one of the inevitable results of the fact that annotations used to be created by hand (and in many different languages as well).

.

Finally, research in the Humanities would become much easier if existing annotations could by completed by additional information that can be derived from the raw material that is indexed by the annotations.

3.3.2 ICT Potential

The potential of ICT for the preservation of languages as objects and for the processing of language as medium and code has long been clear. For example, there are 'antique' wax role recordings of Inuit stories and songs, and there are vast archives of radio and TV broadcasts and movies. These archives are still growing, but for the time being they are extremely difficult to access, because they are only indexed by means of small sets of metadata. The first initiatives to turn library catalogues into digital form, and to make them accessible through global computer networks date back to the early eighties. Recent research in the field of Digital Libraries has suggested that enormous amounts of information could be unlocked if it were possible to link the digital catalogues of different libraries. Emerging Language Technology holds many promises in terms of support of scientists in the Humanities in their analysis of both primary and secondary sources. Here, the border between language as medium and language as code is being blurred by the increasing capability of Language Technology to break the barrier between purely syntactic processing and semantic interpretation. The latter will revolutionise almost all disciplines in the Humanities (and all other sciences too), by providing direct access to virtually all relevant references to a topic under study, not just in the conventional (primary and secondary) literature, but also in other databases that relate to the same geographical region or the same age.

3.3.3 Present Situation

Today the official languages of the EU are all being recorded, if only in the form of radio and TV broadcasts and movies. As said above, the accessibility of the recordings leaves much to be desired. The smaller languages are less well documented, and it is unlikely that specific actions will be taken to make the fragmentary recordings that are being made accessible.

Current developments in Language Technology are very much focused on processing of business information that is produced electronically. Most of the linguistic records of the research in the Humanities only exist in the form of printed material of manuscripts. Efficient and cost-effective procedures are needed to digitise the materials, in such a way that scientists get access to the original visual appearance of the sources, but at the same time to its contents in the form of 'text'. Tools to process the texts with the objectives of the Humanities in mind remain to be specified and developed. Multilingual and multi-domain processing are mandatory.

Language as the code to index material and immaterial objects of Cultural Heritage and other resources in the Humanities has already received much attention in the past. Much work has been done to specify annotation schemes that should help to access the data, both linguistic and multimedia. Dublin Core is probably the most advanced example. At the same time it is generally recognised that annotation schemes such as Dublin Core only scratch the surface of what the Humanities really need, as is testified by the following quote:

> In the diverse world of the Internet, Dublin Core can be seen as a "metadata pidgin for digital tourists": easily grasped, but not necessarily up to the task of expressing complex relationships or concepts.

Obviously, more powerful tools are needed to support scientists and interested laypersons when they are out to search for specific information.

3.3.4 The Vision

ECHO will guide the development of technology that will enable completely new ways of doing research in Historical Linguistics, Socio-linguistics and Language Typology. For example, recent and historical recordings of regional variants of languages can be coupled to digitised representations of older phases of those languages, which are only documented in the form of text, dialect atlases, dictionaries, grammars, etc. At the same time links can be made with Geographic Information Systems and also with databases with information about political and economic history of the region. The combination and visualisation of these data, in the form of patterns that dynamically change over time, and perhaps differently in different regions, will provide new insights in the development of the languages, and in the interrelations of linguistic and non-linguistic factors that have been active.

ECHO will be involved in the development of advanced speech and image processing tools to automatically transcribe and annotate multimedia recordings of majority and minority languages, so as to make the recordings searchable on topics, and in the future even on words, sounds, and gestures. Automatic support for transcription, annotation and indexing will make it possible to document the major official languages, but also the lesser-used ones, as well as regional variants and dialects. This will strengthen small cultural communities by creating electronic resources similar to those of the dominant cultures and languages.

To make the enormous amounts of 'texts' in the Humanities available for on-line research, tools must be developed to digitise printed and written material. ECHO will contribute to the design and implementation of tools that link older phases of the languages to modern ones, so as to bootstrap the language technology tools that are needed to process all texts automatically, irrespective of their age and language. Also, tools will be developed that allow scientists to manually add annotations that cannot yet be provided automatically. For example, information that

.

is only present in the form of details in the handwriting in a manuscript, specialist knowledge about the origin or the meaning of words, etc. that previously used be present in the form of notes must be linked to the text proper. Processing will grow from Concordances and basic annotations to more detailed labels attached to individual sections and paragraphs. ECHO will work to provide full support for multilingual processing. In this context Language Technology will cross the border between automatic processing of texts as the medium in which the information is represented and transferred, and language as the code to index the information. ECHO will work to transform commercial multilingual Information Extraction technology into increasingly powerful tools that support scientists in the Humanities by making semantic interpretations of the contents of texts, drawings, pictures and paintings. At the same time growing ontologies will be developed that will help to bridge the gaps between currently incompatible annotation schemes.

Finally, ECHO will develop human-computer interaction technology based on a combination of natural language and image processing. This technology will help users –both professionals and lay persons- to browse through all relevant resources. Semantic web technology will be used to discover new relations that emerge from the coupling of hitherto isolated information sources. New virtual and enhanced reality techniques will be developed for the presentation of the results of 'intuitive' queries, so as to support the users in understanding newly discovered relations.

JÜRGEN RENN

MAX PLANCK INSTITUTE FOR THE HISTORY OF SCIENCE, BERLIN, GERMANY

TWO ASPECTS OF A CRISIS
OF EUROPEAN CULTURE

ECHO
AN INFRASTRUCTURE TO BRING
EUROPEAN CULTURAL HERITAGE ONLINE
JÜRGEN RENN

A KEYNOTE SPEECH GIVEN AT THE WORKSHOP

HUMANITIES, RESEARCH, AND CULTURAL HERITAGE IN EUROPE

BRUSSELS, OCTOBER 24, 2001

A Vision for European Culture

I would first like to thank the European commission for organizing this workshop and for the occasion to present an idea that has been the result of working experiences and discussions among scholars and institutions of about ten European countries, many of which are represented here today. The informal network that has emerged from these discussions is centred round three Max Planck Institutes in three European countries; for the History of Art in Rome; for Psycholinguistics in Nijmegen; and for the History of Science in Berlin, but it comprises not only research institutions but also museums and universities. It is not so easy to speak for this group here. In order to give you an impression of the wide-ranging interests covered by our group but hopefully also of the potential it constitutes, I will, also in the name of the other Max Planck Institutes, first give you a general introduction to our proposal for bringing Europe's Cultural Heritage Online, that is, to the ECHO initiative. Prof. Gerd Grasshoff from the University of Bern will then present you some ideas of how the vision of ECHO can be put to work, and finally Prof. Paolo Galluzzi from the Institute and Museum for the History of Science in Florence will stress the urgency of realizing the ECHO vision from his perspective as a pioneer of employing information technology in the public culture of science. In order to save time, we have not engaged in extensive demonstrations of what the participants of this initiative have already achieved or are proposing to realize. But in order to offer you at least an impression of the seeds that we would like to spread on a European scale, you will have the opportunity, in the sequel of the three keynotes, to take a glimpse at an array of exhibits that we have prepared for you.

This is a time in which technology and the values of European humanist culture seem to be as decoupled from each other as they have ever been in the recent past. Technological visions of progress, in particular, have lost their appeal of being guarantors of the progress of culture into the bargain. While catchwords such as "information society" or "postgenomic society", not to speak of "traffic of the future," have lost glamour and credibility as promises of a better civil society, scepticism if not hostility with regard to science and technology are spreading in Europe. The European culture, jointly created by the *homo faber* and the *homme des lettres*, faces a crisis: while it provided the foundation for magnificent technological achievements in a long-range development reaching back to antiquity, the cultural heritage of Europe and its values are dramatically losing ground in the techno-scientific world that has emerged from it. European cultural heritage is, in particular, strikingly absent from the medium of the future, the Internet. It is precisely the few shining examples of culture on the Web that make evident the potential of the bulk of information constituting our cultural memory, which is still not represented within the new medium. What we need is, in my view, a vision exploiting the new technological possibilities for the creation of a public culture of science, a vision that includes the humanities

.

and thus keeps alive the roots of our techno-scientific world in European cultural heritage. Such a vision guides the ECHO-initiative, aiming at a massive effort to bring European cultural heritage online.

On closer inspection ECHO must address a double challenge presenting itself to European culture in the age of the Internet, a quantitative and a qualitative one: the need to make a substantial amount of the sources constituting the European cultural memory electronically available, and the need to create an adequate intellectual, technological, and social infrastructure rendering this cultural memory accessible as a resource for addressing the questions of today, be they scholarly or from an orientation-seeking public.

Just imagine you could address the crisis of science education in Europe by creating the possibility of exploring the historical and cultural contexts of scientific and technological knowledge in a digital repository offering pathways from the questions of a student to the original historical sources documenting the emergence of this knowledge.

Just imagine you could understand how human gestures are shaped by nature and nurture, how they have developed historically, and how they are embedded in language and cultural contexts, and you can discover connections within in a digital repository comprising a browsable corpus of films documenting human behaviour as well as a digital archive of the great European works of art.

Just imagine you could understand how administrative structures have evolved in our civilization and you are able follow your enquiry from digital representations of the ancient Babylonian state archives to databases in which the administrative records of the great engineering ventures of the Renaissance are stored.

These are just some examples of concrete perspectives that would open up if the vision of ECHO were pursued.

Implementing the Vision

What needs to be done in order to realize this vision? A possible answer that merits careful examination is: … nothing, or at least nothing beyond what is already being done. There are indeed good arguments for claiming that the existing approaches suffice. Let us therefore take a closer look at what presently are the two principal ways of dealing with the challenge of the information revolution for the cultural and scientific heritage of Europe:

• the big player solution

- and the scout solution.

The Big Player Solution

Let me first turn to the "big player solution." Its basic assumption is that the dominating forces of the market, in particular the big players represented by large publishing houses and software firms, will take care of bringing cultural heritage to the Internet. This solution is most familiar from the present debates on electronic journals. While the few publishers who hold a near monopoly in certain areas of scientific publishing are indeed offering more and more material on the Internet, their approach has been rightly characterized as a "Faustian deal" in which a fatal price has ultimately to be paid by the scholarly community. In fact, although electronic dissemination is considerably cheaper than print dissemination, journal prices — in general still coupled to those of print subscriptions — not only continue to increase but, what is worse, the revenue accumulated by the publishers is in general not reinvested in a future-bound infrastructure for scientific information on the Web. On the contrary, the great challenges for such an infrastructure, for example the archiving problem or the problem of an integrated retrieval environment remain, for the time being, largely unsolved — menacing the longevity and interoperability of scientific and scholarly information in the electronic medium. This is the fatal price of the Faustian deal. There will be no escape from it as long as the scholarly community has to repurchase from the big players the information it produced in the first place, at the same time being left responsible for its infrastructure on the Web.

The situation of the digital availability of the primary sources of cultural memory, that is, of historical documents, of the great works of art and literature, or of language corpora is even more problematic. While sceptics are still debating the compatibility between culture and the Web, the big players have long since begun to secure exclusive rights on the reproduction of cultural artefacts and even to purchase important documents and collections with the intention of commercialising their digital images. From codices of Leonardo da Vinci to the photographs of Ansel Adams, every piece of cultural heritage is a potential asset in this new market. In the hope of spectacular gains, new firms have been founded, claims staked out, and portals opened up, as if in a gold rush. And indeed, looking back at its first phase, one already recognizes the typical ruins documenting the transiency of every gold rush: portals promising to become gateways to unimaginable cultural treasures which actually lead nowhere; key documents of European history are, on the other hand, confined to CD ROMs which are condemned to dust in stacks until they become outdated with the next generation of soft or hardware. Meanwhile they are banished from the World Wide Web which enlivens and enhances every significant piece of information exposed to it by its self-organizing connectivity, at least as long as this connectivity is not smothered by passwords or pay-per-view access.

.

It has become particularly evident that the big players have failed, in spite of their eagerness to control large domains of cultural heritage, to create an infrastructure that guarantees a steady and reliable flow of this heritage from the old medium into the new. On the contrary, they have contributed to an increasing inaccessibility of cultural heritage — not only because of the restrictive copyright laws they seek to impose but also because sources are now often held back by museums, archives, and libraries in the dim hope of future commercialisation. This hope can, however, hardly be sustained by a practice that amounts to a ruinous exploitation of limited resources rather than representing a concerted effort to augment them.

The Scout Solution

Let me turn to the "scout solution." It is based on the assumption that the transfer of cultural heritage to the new medium can essentially be achieved by pilot ventures in combination with an establishment of standards for production and dissemination. In contrast to the "big player solution," it amounts to the realization that bringing culture to the Internet actually means settling a new continent rather than just exploiting its resources in a gold rush. But it also amounts to the assumption that this can be done by merely sending out a few scouts to survey the new territory and set up a model farm here and there. However, this approach is not being criticized brashly. As a matter of fact, almost everything we have in terms of digital libraries demonstrating the potential of the new media for cultural heritage we owe to the breakthroughs achieved along this strategy. But it must be legitimate to ask whether this strategy is adequate to meet the principal challenge of the future, the creation of a self-sustaining development of European cultural heritage in the new medium.

Looking back at the successes and failures of the projects funded by national agencies as well as by the European community, we find indeed that so many of the feasibility studies, pilot projects, test beds, and proofs of concept, however impressive they are if taken by themselves, have actually failed to launch such a self-sustaining dynamics. The dead links, blind alleys, and empty databases characterizing some of the most ambitious homepages of such projects signal that they did not succeed in making a difference for the humanities at large, let alone for the role of cultural memory in an Internet society. To me, such projects are like chip factories in the jungle, incapable of leading off productive development because even the most basic infrastructure is lacking.

Admittedly, the humanities, responsible for preserving, exploring, and keeping cultural heritage alive, is a difficult environment for technical innovations. Scholars in the humanities have hardly even begun to realize that the new information technologies not only confront them with a competence problem that is unparalleled in the natural sciences but also that they are faced

with entirely new possibilities to overcome the deeply entrenched boundaries of narrow specialization. The assumption that the humanities can be catapulted into the Internet age by enticing them with exemplary pilot projects reminds a historian of the overly astute attempts of Jesuits in the 17th century to convert the Chinese mandarins to Christianity: they offered in fact a few extraordinarily beautiful clocks as a gift to the Chinese emperor in the futile hope that he would ask for more European technology and religion once the donated clocks needed rewinding.

The Agora Solution

In order to initiate the far-reaching upheaval that a comprehensive digitisation of cultural heritage would amount to for the humanities neither brute force nor missionary efforts will do. What is needed is rather an infrastructure that enables each single participant in this process to pursue his or her specific interests while contributing, at the same time, to a shared body of electronically represented knowledge. This brings me to a third approach that has so far not been seriously pursued and which we have labelled the "agora solution." It aims at launching a dynamics that combines the development of the whole with the benefit of the individual, a combination that has actually been the hallmark of all great civilisatory enterprises in Europe, beginning with the foundation of the Greek polis which achieved such a synthesis of interests in its agora.

Imagine that every scholarly project in Europe, every archive, museum, or library could join a network of digital libraries by making resources available on the Web with only a minimum effort, using a set of standard formats and the corresponding tools allowing their implementation. Imagine a universal electronic working environment for these resources and imagine that it would provide you, first of all, with an overview of what is presently available within the distributed network; that it would automatically combine all available texts with language tools comprising grammatical analysis and dictionaries; that it would offer interactive tools for studying texts as well as images of art and architecture; that it would allow us to show video scenes of human behaviour or scanned images of historical sources in combination with transcriptions and commentaries. Imagine that such a growing network of digital libraries would also include an interactive collaborative environment for multimedia annotation, that it would automatically establish as many meaningful links as possible between the heterogeneous bodies of knowledge stored within the network, and that it would offer options for converting the results of scholarly work on these bodies of knowledge into new ways of accessing the primary sources of European cultural heritage, opening up new vistas not only for specialists but also for the public at large. Then you get an idea of the new quality this cultural heritage may acquire if a concerted effort along the lines of the agora solution succeeds.

.

The Implementation of the Agora Solution

But how can this solution be implemented? Clearly a self-accelerating dynamics leading to an ever-more comprehensive electronic representation of the European cultural heritage can only emerge if certain minimal conditions are fulfilled. Among them are the requirements of open access, interoperability, modularity, and interactivity. Only if digital sources are made freely available on the Web, only if the same tools can be applied because they share compatible structures, only if diverse digital collections can be integrated to yield an interconnected whole, and only if it is possible to combine the power of computing with the power of the human mind in the analysis of sources, a set of data will turn into a digital athenaeum of European culture.

The essential elements for a successful pursuit of this approach do exist. Three Max Planck Institutes, representing — each together with its international partners — three different areas of the humanities, have accumulated considerable experience not only in making sources of European cultural heritage available online and in developing essential building blocks of the electronic working environment sketched above. The three groups have also engaged, within the limits of their possibilities, in efforts to share their experience with other scholars and institutions and to spread the usage of computer-assisted methods in the humanities. It is precisely on the background of this experience that the ECHO-initiative has been formulated. It has turned out, as a matter of fact, that even the most convincing standards, models, or tools will remain island solutions as long as those scholars and institutions in the humanities who are still lacking a comparable expertise in electronic information management are unable to join in.

It would, however, be an error to consider the implementation of the agora solution simply as a matter of technological developments which, once completed, have to trickle down from the initiated to the laymen. Even the realization of the universal working environment sketched above can hardly be completed without addressing a combination of technical, scholarly, and social issues. It is our experience that it makes just as little sense to develop standards without the tools to implement them, as it makes sense to develop tools without an understanding of the scholarly questions they should help to answer. The real challenge of the agora solution is thus to achieve an integration of scholarly and technical work. Its realization therefore presupposes an environment in which not only technology is spreading but also the knowledge about its innovative application to the intellectual problems of the humanities.

In order to realize such an environment we propose to create a somewhat unusual structure that combines support for a network of projects and institutions that are capable of substantially enhancing the presence of European cultural heritage on the Internet with the creation of an innovation centre. The centre would have the task of defining standards, of providing the network with tools developed on the basis of the accumulated experience of its participants, and of help-

ing them to implement the standards as well as the tools. The innovation centre would have the responsibility for developing the central building blocks of the infrastructure envisioned by the ECHO initiative, for instance the universal working environment mentioned above. It should be capable, at the same time, of boosting the technological competence of projects and institutions willing to contribute to a digital athenaeum of European culture, but which are lacking the infrastructure to do so on their own. Support by the innovation centre could include, for instance, the option to temporarily delegate collaborators to institutions within the network where they would participate in concrete projects and then bring their experience back to the centre. The innovation centre would thus constitute, almost literally, the agora of the network, that is, the central market place where information is being exchanged and experience accumulated.

The Potential Impact

In conclusion, let me briefly come back to what I said in the beginning about the growing distance between the development of technology and European cultural heritage. In my presentation I have mainly concentrated on the information revolution as a challenge for humanistic culture. My account would miss, however, an essential facet of the present situation without acknowledging that cultural heritage also represents in turn both a challenge and a resource for the further development of the Internet as well as of the society whose backbone it has become. It is, in fact, hardly imaginable that the present efforts to create a semantic web, to improve the interfaces between humans and computers, or to establish ever-more sophisticated metadata standards will really succeed without taking into account the knowledge accumulated over centuries by the humanities about the relation between signs and meaning, about the functioning of language, about the way meaning is stored by images, and about structures organizing human knowledge. And what is even more important: when the development of technology and its social implementation confronts us with moral dilemmas or at least with the necessity to decide between alternatives that cannot be distinguished just by criteria of efficiency, we have to be able to reflect upon the past experiences of our societies as they are stored in European cultural heritage. But even as a source of reflection on our present situation, it had better be available online.

.

PAOLO GALLUZZI

ISTITUTO E MUSEO DI STORIA DELLA SCIENZA, FLORENCE, ITALY

HUMANITIES, RESEARCH, AND CULTURAL HERITAGE IN EUROPE

HUMANITIES, RESEARCH, AND
CULTURAL HERITAGE
IN EUROPE
PAOLO GALLUZZI

A KEYNOTE SPEECH GIVEN AT THE WORKSHOP

HUMANITIES, RESEARCH, AND CULTURAL HERITAGE IN EUROPE

BRUSSELS, OCTOBER 24, 2001

Previous speakers have touched upon the many crucial topics relating to the initiatives to be undertaken and the methods to be followed in order to exploit the added value of cumulative work in the process of the transition of the humanities and cultural heritage from the analogic world into the digital domain.

This transition is producing a great revolution both in the way in which scholarly research has been traditionally conducted and in the established methods of production of tools to access original documentation in museums, libraries and archives.

As my colleagues have already stressed, a lot of work has been done in this perspective by a variety of different subjects. A large part of the European scholarly community has become fully aware of the importance of sharing common standards in order to be able to build up a significant mass of information easily accessible to the user with the minimum of cost and of investment in human resources.

Previous speakers, as well as the Discussion paper, have provided us with more than sound evidence regarding the need for a European network of centres of excellence to deal with the strategic topics presented by the transition into the digital world. This network will have to perform original research; to monitor the development of applications; to create software for the specific requirements of research and user applications in the humanities; to play a guiding role as to the delicate problems of how to preserve digital records; to define and distribute standards and tools for improving the interoperability of digital archives and systems – like multilingual thesauri and ontologies - in order to favour an easy and rapid user access to the information available in the different digital repositories; and, finally, to provide authoritative certification of the information available through the web. This is a big challenge which implies the establishment of a programmatic and permanent international cooperation between several competent European cultural institutions, which will stimulate and assist smaller institutions working in the humanities to enter what is, for the greater part of them, the still unexplored world of ICT.

In this perspective it is important to pay due attention, not only to the technical problems and to the methods to develop effective cooperation, but also to the fact that the transition of humanities and cultural heritage into the digital domain implies an extensive re-structuring of traditional methods and styles of research and, above all, of the way of producing and organizing documentation of all kinds.

Let me point out – as a telling example – the transformation which ICT is introducing into the world of documentation and of cultural heritage. Museums, libraries and archives – as we know them today – are the result of a long evolutionary process which began many centuries ago. Museums, for instance, started out in 16th century Europe as promiscuous bodies where art objects,

.

instruments, artifacts of various nature, natural items (both organic and inorganic), books and documents were intrinsically integrated and displaced one aside the other. The conceptual glue of what we would today rate as a heterogeneous ensemble was the exceptional character and the uniqueness of all the collected entities and the curiosity that they aroused in the visitor. Later evolution of museums, libraries and archives resulted in a process of progressive separation of objects from books and from documents. The evolution of museums might be very schematics recounted as a story of ever stronger specialization. First, the separation of art collections from scientific collections (during the Enlightenment). Then artistic collections underwent a further process of specialization with the separation of paintings (this marks the origins of Pinacotheca) from sculpture (sculpture galleries), from drawings (print and drawing rooms) and from minor arts (decorative arts departments). The same process took place with regard to scientific collections. After their separation from art collections specialization progressed through the division of mathematical and physical objects from natural history items; later on, in the mid-19th century, technological and industrial collections gave rise, under the impulse of the Universal Exhibitions, to new independent museums; finally, in more recent times, new institutions devoted to the preservation and presentation of bio-medical and ethno-anthropological collections were founded.

This evolution, which arrived at completion at the beginning of the 20th century, has produced a fragmentation of cultural documentation and cultural heritage according not only to disciplines, but also to the various categories of objects (books, archival records, drawings, etc.) and to their different material supports (canvas, marble or stone, paper, fabrics, etc.).

What has to be stressed from the point of view of our discussion today is that the specialization process has placed museums, all museums, in a negative condition, when it comes to giving full cultural value to the collection they preserve. The same could be said for libraries and archives.

These limitations derive from the fact that objects are presented in museum atomized and decontextualized. This makes it almost impossible for visitors to grasp their real meaning and to perceive their importance. In order to achieve this goal it would be necessary to re-connect the object to the context from which it has been extracted. If we take as example the situation of a painting in a pinacotheca, its context is not simply made up of other related paintings preserved in the same or other museums, but it consists of a variety of different documents: books, documents, preparatory drawings, biographical information about the painter and or the commissioner; information about the technical process adopted by the author to produce his work, about his personal motivations, religious and political beliefs etc. And this bulk of information

has also to be considered in a vertical perspective, that is to say against the background of the chronological evolution of iconology, of the techniques to render three-dimensional space on a flat surface like that of painting, of the contemporary cultural and social debates, and so on.

I cannot develop here this argument more fully. But I think that this schematic history of the process of specialization of museums (and the same holds for libraries and archives) is relevant to our discussion. Let us see in what way.

Once we have decided on dedicating our energies to the digital conversion of the immense cultural heritage preserved in Europe, it is essential to avoid the mistake of reproducing – as it might instinctively seem logical to do – not only the fragmentation of objects and records in different places, but, above all, their compartmentalization into specialized structures (like museums, archives and libraries), structures which are conceived as totally independent and de facto extraneous one from the other.

The new frontiers and formats of the humanities in the digital Age will have to coincide with the re-composition of what the material history of cultural patrimony has fragmented. The fragmentation process has been determined, to a large extent, by preservation concerns (it is easier to preserve paintings gathering them together in the same building, and the same holds true for the preservation of books, archival records, drawings, natural history objects, the video tapes, sound tracks and so on). But, once the information has been processed and transformed into a digital entity, it has no more to follow the fate of its physical support. There is thus no reason to store it according to the same systems used to preserve the object it emulates. In the digital domain there are no buildings or walls, and we are not obliged to reproduce distinctions based on the topologies of material objects or on the various nature of their physical shells. It is possible – and indeed necessary – to re-organize digital records into new cognitive architecture, where the strict constraints of the physical world no longer apply.

The logical consequence of what I have been saying is that we can finally re-arrange the records which digitally mirror the real objects in a totally new orderly architecture. In this way we will introduce a radical alteration of the traditional –that is, pre-digital – system of managing information pertaining to the humanities and to the cultural heritage. We will thus build new repositories – intrinsically federal repositories – heterogeneously populated by a variety of different entities, which a long tradition has accustomed us to consider as totally independent and unrelated items.

This scenario suggests to establish an important precept and a first guideline:

.

The precept: in the transition of humanities and cultural heritage into the digital world we must not only to take the greatest care in optimizing technical procedures and in creating and using common tools. It is equally essential to develop a critical analysis of the implications of this transition, in order to pursue the appropriate goals and not to miss the opportunity of introducing more effective practices and styles of processing, storing and accessing cultural information.

The guide-line: even if, in terms of practical action, we might decide to undertake the immense task of converting cultural heritage and documentation into digital form by respecting the specialization of traditional repositories (that is, museums, archives and libraries) and treating them one by one, in different and separate phases, we have to be aware from the outset that the product we are aiming at is a totally different construct from all traditional repositories. That it is a federal, net-distributed repository where digital records are stored within an open and integrated architecture; where information is not preserved into a rigid spatial dimension, but is like a knot in a cobweb - a knot which maintains its own specific identity, but which, at the same time, is linked to all the other knots with which it shares relations, common properties and dynamic analogies. This implies that we have to strongly contrast the idea of the Digital Museum, of the Digital Library and of the Digital Archive, conceived as digital clones of existing museums, libraries and archives. Our aim will be that of creating a unique repository in which conceptual, material, contextual and historical interrelations are built into the stored data.

A repository of this conception will have enormous impact upon research and upon education. It will also offer important job opportunities, openings for innovative services and for the development of a very promising economy of culture, from which private companies will benefit.

It will thus become finally possible to overcome the limits of fragmentation of cultural heritage and primary sources, a result which, until now, we have been able to achieve only occasionally through purely textual production and, above all, through temporary exhibitions. The latter – at least the good ones – have been so popular in the last few decades just because – this is my guess - their formula is programmatically based upon the presentation in contest of items which are preserved in different institutions scattered in different places. For a limited period of time good exhibitions provide a simulation of how much more effective would be a fruition of cultural heritage based on the reconstruction of the intellectual, symbolic, political and social interweaving which connect a variety of dispersed and isolated items. In the digital domain the effective promotion of cultural heritage guaranteed through re-contextualization by temporary exhibitions will have to become a customary and permanent practice.

This process of restructuring information after digital conversion will imply the development of permanent cooperation, much closer than in the past, between scholars (who will have to trace the multiple threads which interconnect digital data and to provide authoritative quality and authenticity certification) and the professionals traditionally in charge of document management (librarians, museum curators and archivists).

Cooperation among European excellence centres must be grounded upon the definition of common standards, upon the practice of cumulative work and upon the shared awareness that the stronger institutions must assume responsibility for helping bodies with less experience and/or with minor critical mass to enter this new dimension. Cooperation will be fundamental also in defining the long-term goals to be achieved, in trying to foresee the major implications as to the organization of research and its fall-out upon cultural heritage, and in envisaging the new scenarios to which we look forward. For these reasons, I believe that it would be fruitful to launch a specific European program to support cooperation between authoritative and expert partners engaged in building exhaustive and heterogeneous or hybrid thematic repositories of the kind described above, using common standards. This will help towards creating those de facto standards upon which the Discussion paper so rightly insisted.

After an initial phase of spontaneous and individualistic approaches to these new problems on the part of several bodies, it seems to me that in Italy the awareness of the necessity of working both on technical issues and on the conceptual, cognitive and organizational implications of the transition into the digital Age, is gaining ground.

The presence of central bodies with national competence (the Ministero per i Beni e le Attività Culturali and the Ministero dell'Istruzione, dell'Università e della Ricerca) should, in principle, make coordination easier. Recently a project of a National Digital Library was launched during a Conference in Rome, in the presence and with the intervention of an authoritative representative of the European Commission, in the person of Dr. Bernard Smith.

The museum world is still hesitant in Italy to proceed along the new roads. On the contrary, the national Archives have recently shown, not simply interest, but also noteworthy capacity in developing integrated projects, in proposing original reflections concerning key problems like record authenticity, privacy, certification, and so on. Finally, during the last years, Italian institutions active in the humanities and in cultural heritage, have established, thanks to their participation in European projects, close relations with European partners.

In general, awareness of what I consider the inevitable integration, at least in the long term, of information stored separately in the analogic domain, into a hybrid and heterogeneous digital repository is still limited.

.

The Istituto e Museo di Storia della Scienza, located in Florence, has been very active during the last ten years in stimulating reflection on this new scenario, in attempting to foresee its implications, and in testing digital languages and tools for innovative applications.

Drawing inspiration from the motivations and goals that I have been talking about, we have developed a series of innovative projects: integration and interoperability between heterogeneous archives; creation of virtual museums not cloned upon the existing ones; promotion of the cultural value of objects which, if not re-connected to other intellectual data and to other objects, cannot express the extraordinary cultural value embedded within them; multimedia representation of objects, ideas, historical situations of which we have neither visual records, nor tools for understanding their functions, their dynamic workings etc.; exploitation of the new integrated architecture of the data repository to generate research tools and/or educational services to be accessed by school teachers and students.

Inasmuch as we share the analyses and the preoccupations contained in the Discussion paper, we also promoted, a few months ago, the creation of a national centre of excellence, to be located in Florence, with the task of acting as a national orientation and coordination agency at the service of public bodies and cultural institutions facing the problems of the transition to the digital Age. The provisional title of this Agency (a joint public/private effort) is Istituto Nazionale per le memorie digitali (INMED). The pre-feasibility study of INMED has met with the approval of some of the most authoritative national experts. A team of professionals is now preparing a feasibility study. We hope that this project will soon make a concrete start. The fields of activity of INMED, which will act as consulting body for cultural institutions, will be three-fold: Research (new software, ontologies, standards, best practices, e-hand books), Conservation (of digital records) and e-learning. The latter is a crucial and urgent task. It involves explaining to the large number of individuals and institutions active in the humanities and cultural heritage affected by the transition to the digital domain that, in order to avoid exclusion, they must start a new phase of learning (in fact, a lifelong learning process). It also involves convincing them that this is a totally new game with new rules, no longer based on individual work, but on programmatic cooperation on an international level, on the creation of products and tools entirely different from traditional ones; products to be continuously maintained and upgraded and to be made accessible to ever more demanding and active users; finally, that is necessary a systemic approach to the variety of new problems and tools to be managed.

The contribution of a new institutional body of this kind is absolutely essential on our national territory. But its action would be even more productive if joining efforts on a European level with other competent private and public partners, in order to trace a more direct path to the con-

struction of a new frontier for the development of research and education in the humanities, for the promotion of cultural heritage and for granting universal access to these fundamental resources.

GERD GRASSHOFF

UNIVERSITY OF BERN, SWITZERLAND

DIGITAL OBJECTS OF EUROPEAN CULTURAL HERITAGE: GETTING THE VISION TO WORK

DIGITAL OBJECTS OF
EUROPEAN CULTURAL HERITAGE:
GETTING THE VISION TO WORK
GERD GRASSHOFF

A KEYNOTE SPEECH GIVEN AT THE WORKSHOP

HUMANITIES, RESEARCH, AND CULTURAL HERITAGE IN EUROPE

BRUSSELS, OCTOBER 24, 2001

A Reader's Tour through Digital Libraries

The digitisation of objects from Europe's cultural heritage is not a new issue. Over the last few years many digitisation initiatives have been started to develop a suitable technological platform and to make the digital content of sources available on the Internet. Individual institutions support many of these projects, however some national and international digitisation projects should now coordinate their efforts.

Despite the hefty sums meanwhile invested, the search for published results—the corresponding offers of digital sources on the net—ends up all too frequently in the same frustrating pattern. As a rule one is confronted with an elaborately designed and very promising entry portal (for an example see fig. 1, a project of the renowned Library of Congress).

Figure 1. Entry portal to a digital library

The reader's interest is now awakened and, encouraged by this homepage, he follows several links to further project information, at least tries to, but is soon confronted with a less stimulating page announcing the following:

Figure 2. Error of link reference

So far, so good—technical errors can and do happen, even if they don't exactly convey the best advertising for an Internet project. So instead the reader chooses a different link to get an idea of the digital library's inventory and to possibly see one or the other document. This is in accordance with the basic interest of every library user, who, upon entering the digital library hall wants to find out which books are available and where they are kept.

As usual the initial entry leads to a catalogue, in this case a list of the collections available so far. The current list is shown in fig. 3. Looking at this catalogue one is struck by the fact that only highly specialised collections are featured, for example "Pamphlets from the D. A. P. Murray Collection." In fact these collections can be useful in a compact and specialised way, but they do not comprise a central inventory of sources, certainly not one which could be described as representing the main inventory of the sources of our cultural heritage. And so, seen from the perspective of my subject, the history of science, which researches one of the pillars of European cultural heritage, it is remarkable, astounding even, that in the example given, there is not one single collection from any period whatsoever which relates to the history of science. This is the case despite the Library of Congress having one of the largest collections ever and being in a position to offer an incredible range of sources for this subject. This shortcoming is not only characteristic to the Library of Congress but rather reflects the general state of most national and international digitisation efforts. It is sobering to realise that despite the numerous and varying ongoing projects there are still large parts of Europe's cultural heritage that are not digitally available.

Figure 3. List of Collections

Whilst lamenting these limitations of content the reader may nevertheless press on to the digital sources currently available. On finding a catalogue entry linked to an interesting document the following image is shown: The link from the collection's list leads to an electronic catalogue card similar to those found in a normal library index. The required link is found upon reaching the very end of this card—to the modern book one had hoped to find in the electronic library. If one was hoping to download the contents of the book on the computer to peruse a certain document at leisure— a document previously only available in the special reading rooms of certain libraries — then one is about to be disappointed again. The links don't offer access to the whole document but instead to samples, which are obviously supposed to sufficiently represent the content of this source (see fig. 4).

Digital reproductions

Total file-space for digital reproductions: 3 Gbytes

Types of material: [Documents][Images][Motion pictures][Sound recordings]

Documents	Archival page images (printing quality)
Marker	Filename repeats identifier followed by two digits for page number wi extension ".tif"
Formats	TIFF Group IV
Tonal resolution	1 bit
Spatial resolution (approx)	300 dpi
Quantity (approx)	15,554 pages
File size	20-50 Kbytes
Samples	Sample printing

Figure 4

This experience can only be compared to that of the frustrated library user who enters the luxuriously equipped entrance of a gigantic library building merely to find a humble catalogue containing lists of the special collections available. If the user should then happen to be interested in one of these special collections and feels he would like to look through the pertinent documents, he is fobbed off with information about only certain parts being available for study but not the complete work.

Admittedly, a few regrettable coincidences might have taken place on this exemplary tour of the Internet's digital inventory. Unfortunately this experience is rather the rule than the exception and confronts three basic problem areas:

- Comprehensive digital collections of cultural heritage are rare.
- Access interfaces are highly individualistic and often of low functionality.
- Sustainability of collections is poor.

One could take the view that the future tasks of national and international research programmes lie in the elimination of their existing problems; namely the digitisation of extensive collections of significant objects from Europe's cultural heritage; the optimization of user interfaces; and safeguarding a permanent availability of the (complete) documents. However, I see more fundamental problems and believe that these support programmes should not been continued using current targets and strategies. Using a few theses I will outline in the following some of these programmes' more fundamental failings and point to an alternative approach.

A New Echo

In previous concepts two groups of people played a major role: one of these groups is the provider of digital content— often also the source's proprietor. Universities, museums, libraries or archives, whose sources are to be published in a digitised form on the Internet, are all examples of providers operating in this way. The second group are the technical developers: experts who as a rule cooperate with the providers and prepare the technical platforms for digitisation and with whom online-presence is achieved. These providers and technical developers are, in former concepts, those responsible for the conception and realisation of digital projects and those to whom the support programmes are directed. I would like to call this configuration the *Provider paradigm* and at the same time offer an alternative approach, which seems to me much more advantageous, the *Utility paradigm.*
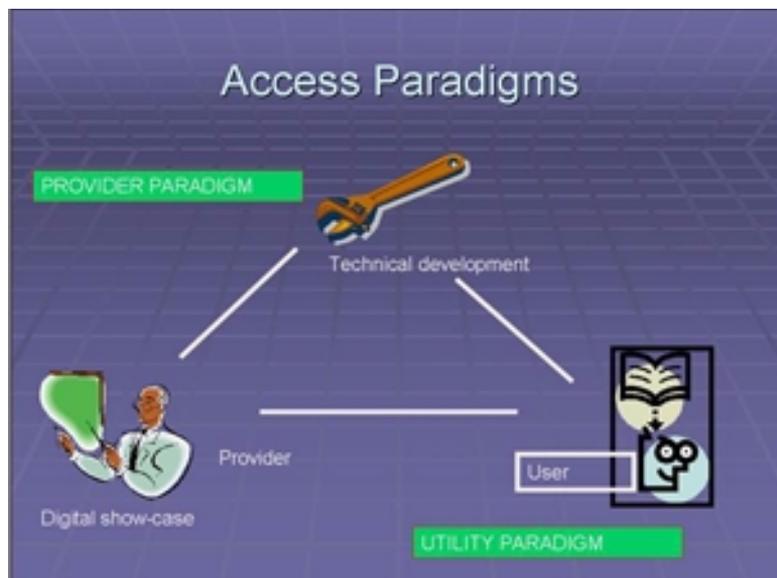

Figure 5. Provider and Utility Paradigm

I would like to use an example to illustrate how essential it is that the user be included in the concept for the digital publication of objects from Europe's cultural heritage. The requirements of the user should be the central issue in all decisions made by providers with regards the technical development of the form the source presentation will take and the possibilities for digital access. Solutions should therefore be found which allow the user (i) to obtain a large quantity of digital sources rapidly by using simple modes of access and limited resources and (ii) to be able to evaluate these sources optimally and relate them to own interests and interpretations. Neither of these requirements is easy to achieve—it is illusionary to assume that the simple provision of digital sources on the Internet is sufficient to meet user requirements.

.

The results of digital projects that originated under the influence of the *Provider paradigm* often resemble the exhibition of objects from Europe's cultural heritage as coined by the museum concepts of the 19th century or earlier. Curiosity cabinets, galleries, and presentations of treasure were aimed mostly at putting as many of Europe's precious objects as possible on display to the public. The observer was presented with individual showpieces whose historical meaning as objects from Europe's cultural heritage had necessarily to be revealed by the object itself. The contextualization of such objects, today taken for granted, is on the other hand a comparatively new attainment. In a similar way, the web pages of many providers can be perceived of as being antiquated curiosity cabinets, wherein particularly valuable pieces can be presented—preferably in an aesthetically superior manner. Parts of Europe's cultural heritage, namely the objects used in daily life, are missing from these collections although they are also part of this heritage. There is absolutely no textual embedding and description of the objects, which in this way become isolated from their historical context of origin and use. Ultimately it is impossible for the user to break away from this isolation with his contributions and to include the digital objects in a new context.

One should not believe that the user of digital sources is satisfied using super intelligent search engines or clever catalogues to track down curiosity cabinets of this type to then view their content with profound admiration. As the word itself says, we are dealing here mostly with source *users* and not source *viewers*. I would like to argue that the user has to be offered the chance not only to optimally utilise digital sources for their own interpretations, but also to contribute further information about the digital source and thus place it more closely within its historical context. The traditional perception of a digital object from Europe's cultural heritage must be challenged in the same way as the hitherto concepts for exhibiting such objects on a website. The following example may clarify this abstract thesis: the use of a digitally available source that can be attributed irrefutably to Europe's cultural heritage but whose most interesting attributes are first revealed when presented in a digital programme specially designed according to the *utility paradigm*.

Objects of Cultural Heritage: the example of the *Flora Graeca*

My example concerns a work on flora from the botanical literature of the early 19th century, a botanical manual comprehensively listing an inventory of the plants of a region, type for type. In this case we are not speaking about any old flora but about the most magnificent and precious flora ever created: The *Flora Graeca Sibthorpiana*.[1]

---

[1] John Sibthorp, James Edward Smith: *Flora Graeca, sive plantarum rariorum historia quas in provinciis aut insulis Graeciae legit, investigavit, et depingi curavit*, Johannes Sibthorp, M.D. [...]. London: Payne & Voss, 1806-40. 10 vols. with 966 col. copper etchings.

The *Flora Graeca* represents the botanical yield of two field trips to the Levant that John Sibthorp, at that time professor of Botany at Oxford, carried out during the years 1786-7 and 1794-5. This work is famous mostly due to its unusually beautiful and detailed illustrations. An example can be seen in fig. 6, the portrayal of a grapevine, Latin: *Vitis vinifera*. Ferdinand Bauer, the Austrian illustrator and one of the best botanical artists ever, illustrated the work and the famous and no less talented James Sowerby did the engravings. Bauer and Sowerby coloured the first issues of the work themselves by hand. A ten-volume opus on the flora of Greece was created by the joint efforts of these first class experts. It was printed on best quality paper and illustrated with 966 hand coloured copper engravings. Due to cost factors, only 70 copies were produced worldwide—a true masterpiece of the history of European science.
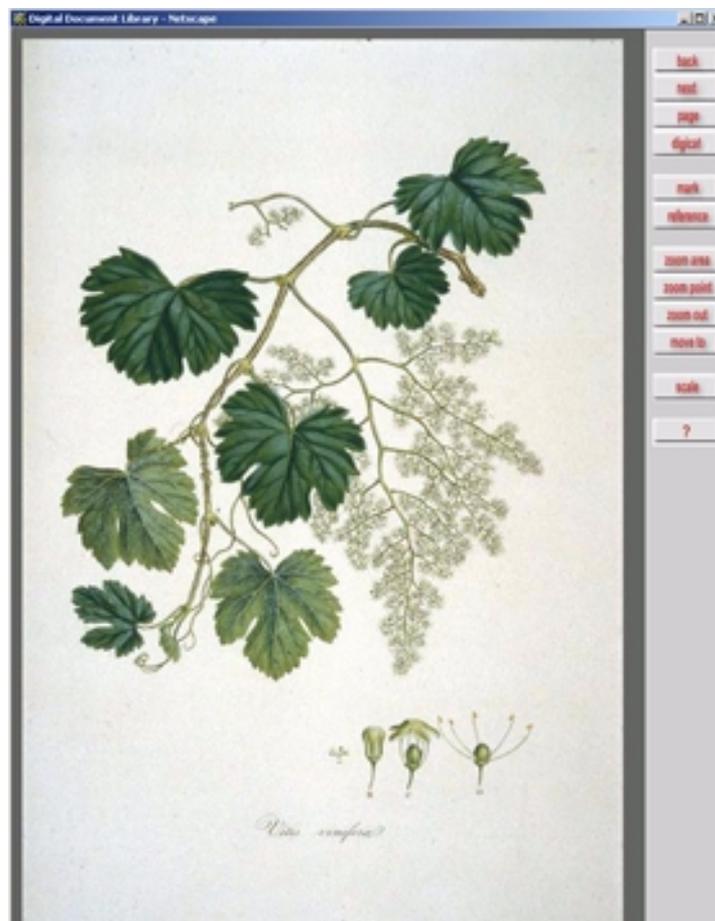
*Figure 6. Grapevine in Flora Graeca – complete view*
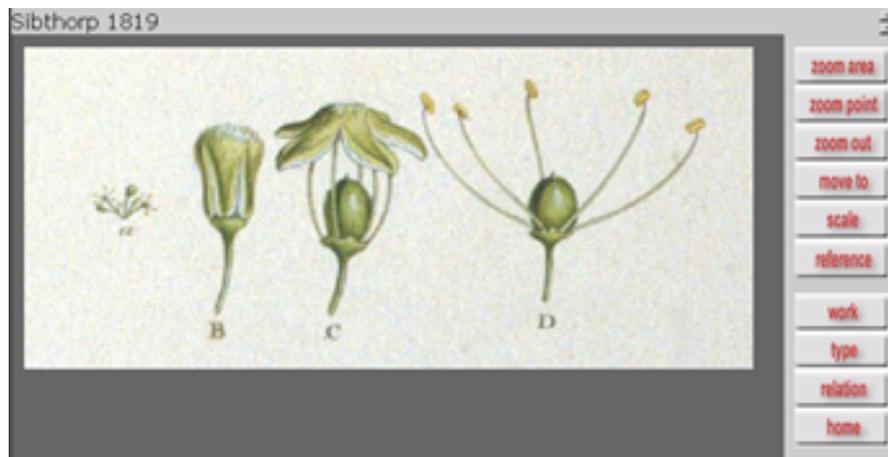
Figure 7. Details from the illustration of the grapevine

The significance of a digital publication of this treasure becomes apparent when one considers the few remaining copies of this work and their usually complicated accessibility in libraries. How should this digital publication be undertaken to best gratify the wishes of the user?

The user would naturally like to obtain an initial overall impression of the textual pages and illustrations. Accordingly, the digital publication must be able to show at a glance the complete image, as is the case in fig. 6. An inspection of the image seldom ends with the sole wish to see a complete view. A closer inspection of its individual sections is desired, to zoom in and study details of this carefully elaborated image—as if using a magnifying glass on the original to trace its technical refinements.

The conceptual and technical difficulties of the digital publications available today become apparent when the aspired combination of these two initial and obvious user requirements are taken into account. The overall impression is still accomplished as a rule even when the frequently long download times make for a very sluggish digital document. But the user is seldom given the choice to look at details of individual sections. This user requirement cannot be satisfied with a given number of selected image areas made available by the provider. No provider can foresee which particular details the future user of the sources will find interesting—therefore he can provide at best an exemplary choice, similar to the sample pages of the digitised textual sources of the Library of Congress. This standard selection would really only satisfy a few user requirements and leave many interesting questions open. This solution is as inadequate as offering only the single pages of a book.

The technical demands on the digital document would rise considerably were it technically possible for the user to zoom in on any part of the image for detailed study using a specially prepared interface and possibly even further software-tools of his choice. It does not suffice to publish the overall view on the Internet in a standard resolution and provide the user with tools to zoom in. The resolution of this general view would be insufficient to support a detailed study.

If on the other hand high-resolution images were put on the Internet, these data files would be so large that they would easily surpass the average user's capacity for transferral and storage of data. A solution for these basic user requirements can only be achieved when the provider combines high-resolution data with the possibility to generate the required sections from digital images and then to send them via the Internet.


Figure 8. Grapevine by Friedrich Gottlob Hayne

Nevertheless even these detailed views of the image only make certain showpieces visible, like the tip of an iceberg whose real potential lies hidden under the surface of the water. The perspective potential is what defines the real European cultural heritage and will only be visible when the user interprets, processes, and interacts with its objects.

Compare the illustration of the grapevine from *Flora Graeca* with the second, more recent rendition of the same kind of plant in fig. 8. This image originated in Berlin as an illustration for a pharmaceutical compendium. This opus appeared in print in a smaller format than the *Flora Graeca* and was also less expensive both to produce and acquire. With its 14 volumes and 648 hand-coloured copper engravings, however, this work is still an extraordinarily impressive product of the history of science. Friedrich Gottlob Hayne edited this book under the title *Getreue Darstellung und Beschreibung der in der Arzneykunde gebräuchlichen Gewächse, wie auch solcher, welche mit ihnen verwechselt werden können*.[2]

For the reproduction of the grapevine in this work, the overall perfunctory impression is, that very different means were employed in comparison to those used for the *Flora Graeca*. The image is much more schematic, the colours are more striking, the details much rougher, and finally various details can be found which are not present in Ferdinand Bauer's picture, like for example the lusciously displayed grapes. Still a second glance reveals this presentation as a copy of the first work. Look at the number of leaves, their shape, the way the vine tendrils curl, and the unfurling clusters of blossom! In comparing the two images directly, it becomes apparent that the manufacturers of the second book used Ferdinand Bauer's illustration as a template (fig. 9).

Meanwhile we should not disqualify Hayne or his illustrators when comparing them to Sibthorp and Bauer! Ferdinand Bauer was no more acting as a human camera than Hayne was, creating exact photographic replicas of a living grapevine from Greece or elsewhere. Even an intuitive guess would reveal that scientifically used botanical illustrations are not exact copies of nature. On the contrary, if one analyses these or similar plant representations in detail, one can see that different stages of development are combined so that colours, form, and scale are not true to nature, but rather chosen in a way to present as much information about the plant as simply and with as many characteristics as possible. The conveyance of relevant and correct scientific information played a decisive role in the manufacturing process of such images. The use, by Hayne for instance, of earlier illustrations as templates for their own drawings was a vital practice that ensured adherence to the knowledge established so far.

---

[2] Friedrich Gottlob Hayne: *Getreue Darstellung und Beschreibung der in der Arzneykunde gebräuchlichen Gewächse, wie auch solcher, welche mit ihnen verwechselt werden können*. Berlin: at the writer's expense, 1805-37. 1805-37[-46]. 14 vols. with 648 col. copper engravings.
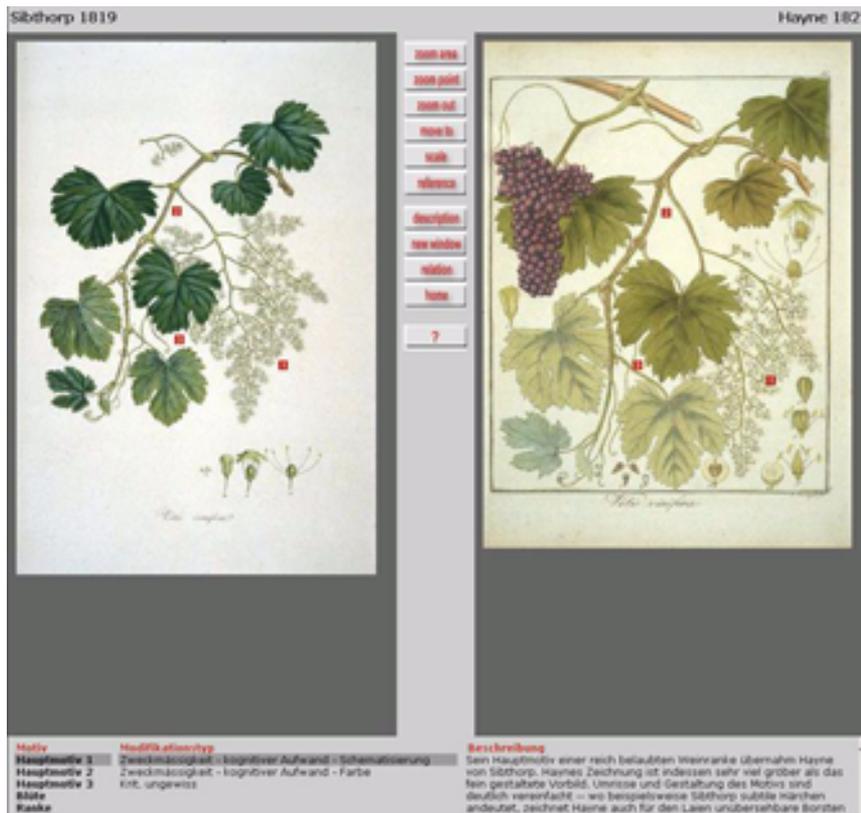
Figure 9. Comparative view of the two grapevine illustrations

This copying process should in no way to be seen as a blind portrayal—on the contrary the motives are systematically modified and adapted into other illustrations according to changing requirements. In fig. 10 we can see for example the comparison of two corresponding sections of the two grapevine illustrations. Whilst the copyist duplicates nearly all the twists and turns taken by the tendrils, as shown in *Flora Graeca*, he deviates from his model in one small part. He omits the (incorrect!) detail of one tendril which bifurcates *twice*—because grapevine tendrils, as is in shown correctly in later versions, only bifurcate *once*. Therefore we see here in the frame of a later version an example of independent correction of the template, the copy turns out not to be a degeneration of the original motive but rather an improved version!

.

Korrektur:

Links sehen wir einen Ausschnitt aus der
Abbildung des Weinstocks von John
Sibthorp, rechts den entsprechenden
Ausschnitt bei Friedrich Gottlob Hayne
(Kopierbeziehung). Hayne kopierte fast alle
Wickelungen und Windungen der Weinranken
bei Sibthorp minutiös – allein an der hier
vergrössert abgebildeten Stelle verzichtet er
auf ein Detail.

Eine der Ranken bei Sibthorp gabelt sich
nicht nur einmal, wie es viele Abbildungen der
Stichprobe zeigen, sondern zweimal
(markiert jeweils durch *). Diese doppelte
Gabelung fehlt bei Hayne: Wie alle anderen
teilt sich auch diese Ranke auf seiner
Abbildung nur ein einziges Mal (vgl. auch hier

Figure 10. Comparison of two corresponding detail views

This kind of intelligent and critical use of templates in copying single elements from earlier scientific portrayals is, in the case of botanical illustrations, quite common to see. It has been shown in a systematic and extensive examination of these and other botanical illustrations from the 18th and early 19th centuries that during this time natural plants were in no way the only direct naturalistic templates used for scientific illustrations. Earlier well-received illustrations were seen as a kind of basic scientific knowledge, whose single elements could, without disregarding scientific diligence, by all means be used and integrated into other illustrations. When utilising this copy method, mistakes could be identified and corrected on one hand and on the other, additions, modification and reductions carried out according to changing requirements and the scientific goals of a new project.[3]

The object seen in isolation—the digitised image of the grapevine from the most magnificent flora of all times—is transformed in relation to the copyist's findings to something completely new. If we look at the whole image, we see nothing but a more or less pleasing and detailed image of a plant. If, however, we see it in its historical context we suddenly have an extremely interesting insight into the working techniques of the botanists and botanical illustrators of that time. Without a direct comparison of the pictures in digitised form one could hardly have found their contextual relations. As a rule these volumes are not kept at the same library, at least this is a very rare occurrence, and even then the user can usually only look at them one by one in a special reading room. The subtle differences and similarities seen in botanical illustrations cannot be reconstructed from memory—at least this fact has been recognised in recent times as digital images have become available for comparison.

---

[3] Kärin Nickelsen: *Wissenschaftliche Pflanzenzeichnungen – Spiegelbilder der Natur? Botanische Abbildungen aus dem 18. und frühen 19. Jahrhundert.* Bern Studies in the History and Philosophy of Science (Bern Studies, 1), 2000.

Only a user having the necessary technical appliances and options is able in this way to put such digital sources in relation to each other. It will only be possible to carry out such comparative studies when the technical requirements are given and the libraries' inventories consolidated within the same working environment. The working environments shown in the illustrations provide the means to do this. It is not only possible to view the images alone, it is also feasible to place any two botanical images or other representations beside each other in order to compare them both in their entirety and in detail. These aids should not, however, be conceived of solely as instruments to provide the user with the opportunity for a detailed study of digital objects. The user should also be able to suitably publish his findings—again on the Internet with the ability to refer to his own comparisons and aids from the source analysis—this can also be achieved with visual instruments as we can see in more detail below.

The discovery of copy relations with the help of digital aids tells many unexpected things about what is actually being presented in these images and what the authors and illustrators actually depended on in their work. In this respect a suitable *demonstration* of a cultural object also fundamentally changes its *interpretation*— for example our perspectives on its cultural environment—in this case on the scientific practice of contemporary botany. The original object, namely the single pages of the *Flora Graeca*, are transformed in the process of these new interpretations to a new object of Europe's cultural heritage, in that they are now seen from a completely new angle—no longer as single, naturalistic image wonderful to behold but rather as an object of scientific design achievement in the frame of a complex network of botanical knowledge from Europe at the beginning of the 19th century.

The users and viewers of the original sources were able, through their active personal contribution of interpretation and contextualization, to create something new. What is more, this was achieved without having to or even being able to supply the provider or original technical developer with a paper outlining these constructive modifications. To be able to do this for other cases, the current and future digital projects should make it their aim to provide suitable technical aids. This should not encourage the heartfelt admiration of Europe's cultural heritage but rather provide and encourage active research and interpretations of these digitised objects and its free subsequent publication.

The Digital Library of the Future: a brief sketch

How can the technical aims set by the utility paradigm best be realised for the optimal use of sources? What if we wanted to publish, for example, the results of the work done comparing different copies? The textual descriptions alone of the differences and similarities between the images will remain too abstract to communicate the required content. To make the results really

.

accessible, the presentation of the image is vital. It soon becomes apparent that the presentation of single, complete images next to blocks of text, as is the case with traditional book publications, is not sufficient. Their communication demands, as was the case with the original discovery of relations in copies, that: (i) any chosen detail of the image be easily magnified for closer inspection and (ii) the images be viewed side by side for optimal comparison, and (iii) it can be annotated. This is no longer possible with printed medium—the results of digital research are best published in digital form. This form of publication necessitates the fulfilment of the same basic requirements as previously outlined for researching digital sources.

At present, to render them accessible to the public it is necessary to save all botanical images, on one's own responsibility, on local servers. In contrast, the digital publications of the future will have access to numerous, externally-stored sources in the Internet—for example, to images portrayed in single, comparative or detailed views, but also to the additional letters of the authors and botanical illustrators, or to the relevant fundamental works of contemporary science, etc. The sources would be stored, using uniform standards, on clusters of high-performance machines, whose existence the public need not realize. For standard use, all that is needed is an average computer fitted with a common browser which supports the simple technology used for digitisation. In the future network of digital libraries, source documents stored worldwide in standard formats and requiring only standard software will be accessible with very little effort. These documents will be accessed via a reference system comparable to file references or bibliographical details. The references will be documented by a network of computers and relocated to the correct data servers, without the user having to be informed of these details. The centrally-stored data volume itself will be made available by the provider, who, in this configuration, does not have to concern himself with data management.
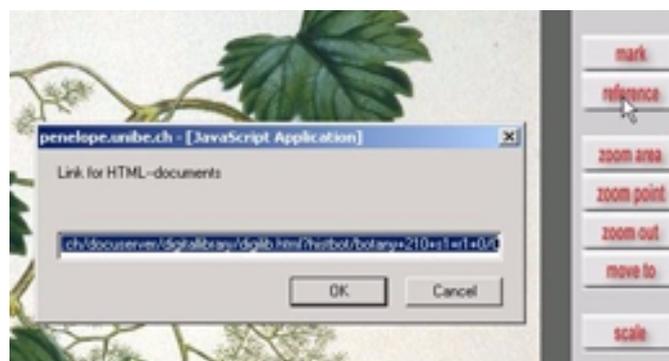
Figure 11. Creating a link referring to a specific digital source

The virtual library allows digital documents to be adapted in a constructive and previously unknown manner. It does not just offer an unimagined network of previously isolated content, with the ensuing creation of completely new objects of cultural heritage. Above all it allows the user to do something that will leave every librarian and protector of Europe's cultural heritage cold. The user has the possibility, not only to refer to and access the object of his study in an optimal way, but can even mark these documents in parts relevant to his needs. It is also possible to define these marks and annotations for reference addresses in such a way that not only the object of study, but particular passages of text or certain images can also be referred to. The form this will take in practice can be seen in the images previously shown from the much-cited Bern research project (fig. 9). On the left side of the user panel, the image in question can be seen. The user has marked certain parts of this image in red and the numeration of these marks featured in the right-hand text window—or in any other text—may be referred to. The technology is set up so that the markings as well as any further work done on the digital sources appear solely on the user's screen and are then only stored locally. The original provider of the digital sources has nothing more to do with this additional information. He can publish the original sources without using extra information—the networking, processing, and interpretation are carried out by the user himself.

http://penelope.unibe.ch/docuserver/compago/
botspezial/arten_bot.html

Der Link *Arten* führt zu einer Auflistung und knappen Skizze der für das Projekt ausgewählten Pflanzenarten (vgl. Abb. 4.34). Über die reine Benennung hinaus wird kurz begründet, warum gerade diese Arten für die Fragestellungen des Projektes interessant sind. Über Klick auf die Artnamen oder auf die verkleinerten Bilder sind von hier aus die betreffenden Artbeschreibungen zu erreichen.

4.8.3 Quellen

Digital abrufbare Quellen zum Botanikprojekt: http://
penelope.unibe.ch/docuserver/compago/botspezial/
quellen_bot.html

Neben den Bildern sind über COMPAGO auch digitalisierte Textquellen zu dem Botanikprojekt zugänglich. Eine Übersicht über die derzeit abrufbare Literatur findet sich über Klick auf den Link *Quellen*. Aufgenommen wurden ausgewählte Texte der Sekundärliteratur sowie Auszüge aus den im Projekt vertretenen Werken der botanischen Fachliteratur. Durch Klick auf die Titeleinträge können die Texte im DIGILIB-Bildschirm aufgerufen werden. Abbildung 4.36 zeigt ein Beispiel für einen solchen Aufruf. Da es sich sämtlich um bitonal digitalisierte Dokumente handelt, werden sie in festem Massstab präsentiert (vgl. dazu die Ausführungen in Kapitel 2; dort findet sich auch einer Erklärung der zusätzlich verfügbaren Button-Funktionen).

Neben einer Auswahl direkt einsehbarer Primär- und Sekundärquellen ist eine vollständige Liste mit Nachweis der für das Projekt verwendeten Literatur abzurufen (vgl. Abb.

Figure 12. Standard text with references to a virtual digital library

Some colleagues involved in the Bern project are already using this new system of publication to refer to digital sources. As can be seen in the following image (fig. 12), it is possible to link traditional texts with distributively-stored digital sources in order, for example, to use these as evidence for argumentation. In this future form of publication, the origin of bibliographic details for source texts, citations, images and other documents used to study an object will no longer be featured in the traditional form of footnotes in scientific texts. Should a reader wish to question the reasoning behind a publication in full detail, then he must physically obtain the source documents cited in the footnotes. The invention of the scientific footnote system in the 19th century was an incredible progression for transparent, evidentially-drawn argumentation, and theoretically, given the exact indication of the sources used for one's own argumentation, the source could be consulted in a library for public criticism. In practice, this method is of limited use only. Who verifies the vast amounts of citations stored in archives all over the world? In these times of the Internet we have entirely new technological opportunities, which seem to completely overshadow the traditional footnote system. In only a few years, the user of scien-

tific texts could find classic footnotes extended with functions to automatically generate digital sources which in turn can offer a critical revision with absolutely no time loss involved. The introduction of a system of digital sources tailored to users' needs does not just provide a generally easier-to-access propagation of Europe's cultural heritage. It changes the fundamental way in which scientific knowledge can be reviewed. The critical debate of scientific theses is no longer tied to a privileged entry on the sites of libraries and archives, where traditionally-cited references are accessible only if one is prepared to make a long research trip. No system of political decisions or declaration of intentions could so rapidly accomplish the integration of third world countries into the scientific community of Europe and North America.

Technical Platform for a Digital Library

Which technological platform can fulfil the aims of a worldwide digital library? To achieve this goal, a few technical and organisational requirements must be given. A few main features are mentioned here in brief, which have been outlined in more detail elsewhere.[4]

The digital library of the future has three basic requirements:

1. Access to digital sources must be permanent. A link to other sources must, as with normal publications, produce sources from decades or possibly centuries gone by, which the author of the original document has put together from his own texts and documents, or from references to other documents. Should this permanence not be guaranteed then there would be no motivation for the author to compose his document using these widely-distributed documents.

2. According to the *agora* principle, it is not possible to coordinate the digitisation and permanent publication for a distributively-stored digital library, which itself incorporates sources from a complicated and centralised administration. The participation in the library requires the following:

- A self-organised participation whereby document providers are able to register and authorise their own inventories.
- The provider should be able to either digitally prepare and publish their sources themselves or authorise participating libraries to manage them.
- Security systems should generate back-up inventories so that each piece of data is preserved with at least one lasting duplication of the sources to make sure that no hardware failures, changes of address or other mishaps can have any adverse effects on the availability of dig-

---

[4] Gerd Graßhoff, Hans-Christoph Liess, Kärin Nickelsen: COMPAGO - der systematische Bildvergleich. Handbuch. Bern Studies in the History and Philosophy of Science (Bern Studies, Educational Materials; 3), 2001.

.

ital sources.

- Offer free access to the software used by the basic technology of open source libraries that is
  - low-priced
  - maintained and further developed by a development community. The history of computing shows how stable and efficient this strategy is. Today's operating system for the Internet—Unix—is based on this approach.
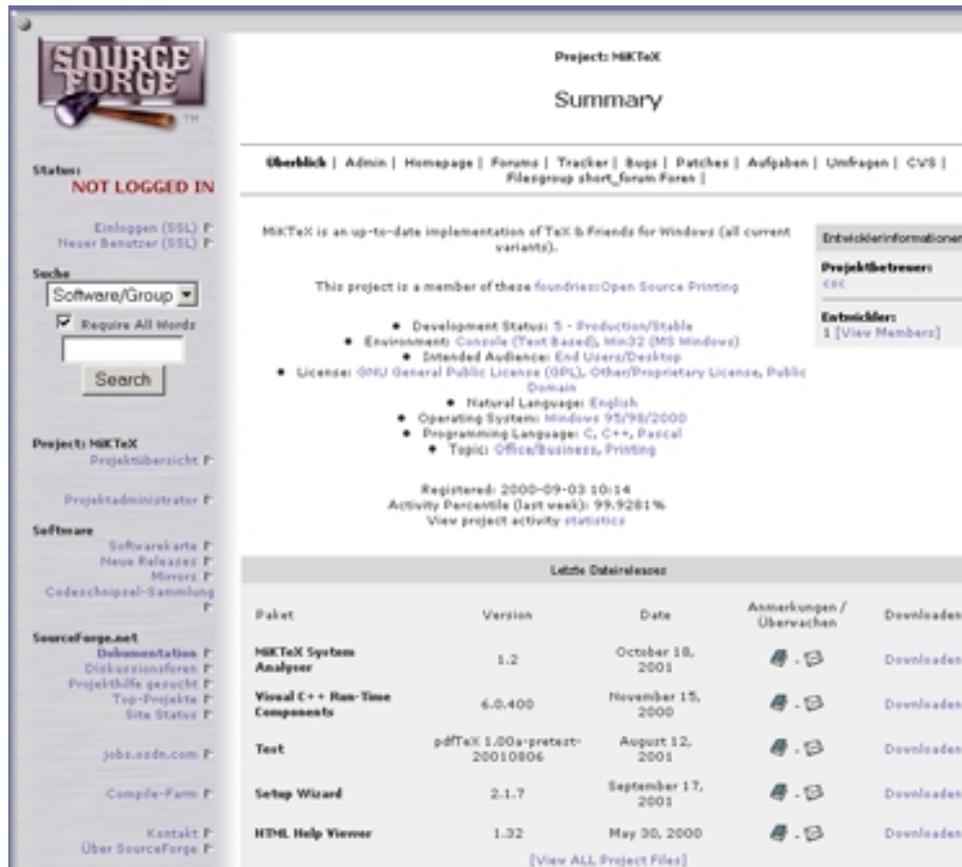
Figure 13. Access page to an open software repository

- The provider institutions allow free access to the objects from Europe's cultural heritage— there should be no licensing or fee-paying.

Nowadays all the basic techniques are already available, they simply need developing into a completely coordinated, technical system and, with the help of leading user institutions and providers fully develop into an attractive and exemplary programme. Inevitably, the revolution will come to pass and our times of the digital showcase will progress to an integrated, virtual, cultural space for all.

STEPHEN C. LEVINSON AND PETER WITTENBURG

MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS, NIJMEGEN, THE NETHERLANDS

LANGUAGE AS CULTURAL HERITAGE

Language as Cultural Heritage
Stephen C. Levinson and Peter Wittenburg

ABSTRACT

The ECHO proposal aims to bring to life the cultural heritage of Europe, through internet technology that encourages collaboration across the Humanities disciplines which interpret it – at the same time making all this scholarship accessible to the citizens of Europe. An essential part of the cultural heritage of Europe is the diverse set of languages used on the continent, in their historical, literary and spoken forms. Amongst these are the 'hidden languages' used by minorities but of wide interest to the general public. We take the 18 Sign Languages of the EEC – the natural languages of the deaf - as an example. Little comparative information about these is available, despite their special scientific importance, the widespread public interest and the policy implications. We propose a research project on these languages based on placing fully annotated digitized moving images of each of these languages on the internet. This requires significant development of multi-media technology which would allow distributed annotation of a central corpus, together with the development of special search techniques. The technology would have widespread application to all cultural performances recorded as sound plus moving images. Such a project captures in microcosm the essence of the ECHO proposal: cultural heritage is nothing without the humanities research which contextualizes and gives it comparative assessment; by marrying information technology to humanities research, we can bring these materials to a wider public while simultaneously boosting Europe as a research area.

The Hidden Languages of Europe – the Example of Sign Languages

ECHO has the ambition to simultaneously advance access to the cultural heritage of Europe and to promote and network Humanities research across the continent, through development of the technology to make Humanities research on cultural heritage available on the internet. One of the richest elements of the cultural heritage of Europe is, obviously enough, the major languages of the member states, and their accumulated national literatures. This is well enough appreciated that on-line access to such resources may seem hardly important. Nevertheless, which of us can read Beowulf or the Norse Sagas or Petrarch in the original? Here the provision of images of the early manuscripts, line by line glosses, on-line dictionaries, and cultural footnotes could in fact provide a wonderful resource, allowing wide public access to the historical roots of all the languages we share here in Europe, with obvious educational applications. The technology required here would essentially be the same as that required for displaying and interpreting materials from the History of Science (see the relevant section of this booklet).

There is another side to the linguistic heritage of Europe, a hidden side. There are disappearing dialects, and the languages of ancient minorities, from the 13 Romany dialects to Breton to Fries, on which information could be provided. Many of these ethnic minorities are now dis-

persed around Europe and beyond, and the possibility of reconstructing a "virtual community" – to which people could contribute spoken language materials – would be very attractive to the public. One could for example have a Celtic site, inside which one would find Irish, and therein one could find recordings and films of spoken Irish from different generations, collected not only from Ireland but from all the destinations of the Irish diaspora.

There are also the many languages of immigrant communities which now play such an important role in our economic and political life, and indeed in our demographic future. The provision of materials on Turkish, the Caribbean Creoles, Tamil, Chinese, Vietnamese and all the other languages of our immigrant communities would be much appreciated, and would help to network the wealth of scholarly knowledge about such languages in the Universities of Europe.

But let us give the most striking example of a hidden linguistic richness right around the continent, about which the proverbial man on the street knows next to nothing. In Europe there are many languages that are all around you but which you have never heard. They are the Sign Languages, languages used by the profoundly deaf and their relatives. In every deaf community, Sign Languages are the natural, preferred means of communication. Much of the public imagines that such Sign Languages are just gestures, some primitive universal iconic medium, not realizing that each of the Sign Languages of Europe have evolved over centuries into real languages with the full expressive power of any of the major spoken languages. They have their own distinctive words, and their own syntax – their own ways of combining words into sentences. Each of them is the focus of a rich cultural system, and they have unexpected cultural connections across and beyond the member states. Thus Austrian Sign is related to Russian sign, French Sign to American sign, but British Sign is unrelated to American Sign, and so forth. Some member states have more than one Sign Language, as in Catalan vs. Spanish Sign, or Belgian French vs. Flemish Sign (the latter not closely related to Dutch Sign incidentally). Altogether, there are at least 25 Sign languages of Europe, and at least 18 indigenous Sign Languages within the existing EEC.

There is an important ethical dimension to the appreciation of Sign languages. The history of Sign languages in Europe has been a roller-coaster ride, from high points in the Enlightenment to the lowest depths in Nazi-occupied Europe. Much of the history is tragic, with repeated acts of suppression of Sign in favour of the dominant spoken languages taught through finger-spelling and attempts at vocalization. New justifications for the suppression of Sign have been based on medical advances (cochlear implantations), but Sign remains the favoured expressive medium for all deaf people. Member states still take different attitudes to their Sign languages, and

only three (Greece, Finland, Portugal) give them constitutional recognition despite two resolutions of the European Parliament (1988, 1998) encouraging this. There are thus important public policy implications here.

There is tremendous public interest in Sign Languages, which is currently unsatisfied. There are frequent enquiries from the parents of deaf children, or from those who have a deaf colleague or in-law, betraying both a curiosity and concern and a deep public ignorance about the essential facts.

Sign languages present a major technological challenge. Needless to say, remote communication by Sign is currently problematic. In Texas there is a pilot experiment with video-phones, and spoken-phone to video-phone links provided by bilingual operators – this has proved to be a great success. In Britain there is a project to create a 'virtual signer' – a computer driven animation which will take English as input and produce British Sign Language as output, in order to facilitate communication between the hearing and the deaf. Research on Sign Language is equally technologically demanding. We must have ways to annotate visual recordings (e.g. with translations into written language), and search and retrieve segments of such visual episodes. This annotation should be carried out collaboratively between partners working at remote sites and other researchers should be able to add comments in a flexible way. Preliminary solutions are sketched below.

Sign languages are scientifically important to our understanding of the very foundations of human language. No other animal can displace its communication system from one modality to another in the way that human language can be displaced from the vocal-auditory channel to the manual-visual one. Language has a deep biological and psychological basis – but the very existence of Sign Languages challenges our current understanding of these foundations. For example, language is largely in the left-hemisphere of the brain, but spatial cognition is largely processed in the right hemisphere – so what happens when a spatial language is processed by the brain? It is thus scientifically imperative to understand the nature of Sign Languages at a greater depth than we currently do. The first thing to do is to deepen our understanding of each of these languages. There are some strong research traditions, for example in the national institutes of the deaf of the various member states – but there are also many member states where no such scholarly tradition has yet been established. But the comparative study of the Sign Languages of Europe has hardly begun. So here is a natural network waiting to be established across the continent – a network which would bind together scholars who currently work largely within national boundaries, and without the benefit of the technology that we could bring to them.

For the study of Sign is technologically demanding – it requires from its very nature a multi-media approach, a method of handling video documents with rich annotations and full retrievability. Most current work uses ad hoc means and relatively primitive home-grown technology.

So here would be an immensely valuable project – to introduce to the citizens of Europe these unheard languages in their midst, to show them how conventionalized and expressive they are, and how they have the status of real languages, to explain historical relatedness and cultural differences. We would do this first in the following way. Each of the 18 languages of current member states – from Portuguese Sign, to Greek Sign, to Belgian Sign, to British Sign – would be represented in the following way:

- Historical pages, detailing e.g. the eighteenth century pioneers in the various countries
- A video clip retelling the same story from a wordless picture book in each languages, allowing visual comparisons between each and every language
- A large set of words in each language – e.g. the word for fire, man, tomorrow – showing how different the different Sign systems are, and relatedness where it occurs
- Conversational exchanges in each language, to show that these are living languages used in ordinary face-to-face communication
- Folklore and poetry in Sign, to illustrate that there accumulated cultural riches – a 'manual literature' paralleling the 'oral literature' of traditional spoken tongues
- More detailed linguistic material, arranged on a comparative basis, for each of the Sign Languages.
- Using the tools described below, internet users could find the video sequences of interest to them, search for specific features and compare e.g. the same sentence in three different Sign languages on the same screen.

This kind of information could be assembled from across Europe from the national researchers already at work on the local Sign systems.

Such a program would open up a hidden part of European heritage to the wider public. It would help us all to appreciate the achievements, the culture, of our silent fellow citizens. But it would have real scientific value too – the fact is that no such comparison has ever been done, and we have no clear idea about what the results would look like. The results would be of great interest to students of language and psychology. It would be a first – yet another example of Europe taking the lead in an area rich with cultural appreciation, policy implication and ethical dimensions.

Tools for Multi-Media Handling of Spoken Language, Sign and Other Cultural Performances

Cultural objects – scanned or digitized – are of little intrinsic interest without the information and scholarship that sets them in context, explains their importance and compares them to others. All this information is linguistic – and linguistic tools are therefore fundamental to the entire ambitions of the ECHO project: they play a crucial role in retrieval of information, making it accessible across the national languages, and so on. In addition, language can also be the object of attention, as when the object is an important historical text. Here, we focus on just one kind of language material, recordings of linguistic performance unfolding over time. The tools required here are specifically related to the dynamics of events, and we will describe these below.

First, however, imagine users of the internet wishing to find out something about the languages of Europe. They can click on a map of Europe and then successively narrow down the area they are interested in, and the language they want to hear and see – finally a list of different kinds of recorded speech events are offered to them, and through successive choices they arrive at a specific film, of say a person describing the local city in Dutch Sign Language or the local dialect of Romany. They can get the simultaneous translation in different major languages, or they can search for a specific feature, like a greeting or the word "Hello". (Maps of course can be more than a way to find the data – they can also offer, for example, historical overviews of the extent of different language areas over time.) Alternatively, instead of using the map, they can browse in conceptual spaces and/or search with keywords 'description in Dutch Sign' or the like, and rapidly they find the relevant excerpts.

This multidimensional mode of access requires underlying structure of course. We have developed a "Browsable Corpus" system which delivers all this, while at the same time facilitating the task for the provider of the information through a specific Browsable Corpus Editor which structures the input information and guarantees adherence to the constraints of the keywords required for efficient retrieval (the set of 'metadata' attributes). The metadata set used was developed within a European initiative and supports description elements allowing researchers and also the interested public to enter detailed queries. For the general user metadata harvesting with the help of OAI based protocols and Dublin Core elements is supported. The Browsable Corpus not only navigates the user to the objects of interest (assembled as a single coherent set of documents and data, though they may in fact be distributed around the internet), but also automatically launches the tools appropriate to the kind of data. This kind of structure in fact is

essential to the entire ECHO project, and we plan to make the Browsable Corpus cover all the major types of data covered by research in the humanities, whether pictures, music, ancient text, living language or for that matter any kind of performance.

So, the Internet user navigates to the linguistic performance of interest. There we launch a viewer to see the digitized film of the event, by streaming of media fragments and synchronized annotations. We have developed a specific tool set, EUDICO, which is designed to input, view and analyse annotated multi-media records. Here we have had to do the underlying conceptual work: any event unfolds over time, and annotations need to be locked to a time frame or to already existing annotations. But some annotations are dependent annotations, and inherit the time frame of the higher order annotation, while others are overlapping in time. Again others may refer to annotations on the same or other tiers. All this requires an underlying computational structure which can also be exported to open XML-based formats. EUDICO is built on such a structure, and consequently you can search video-records for specific attributes (say the Dutch Sign for 'girl'), and immediately see just the relevant sequence, and all the time-locked annotations (translation, linguistic annotations, etc.). EUDICO has general potential application to a video-record of any cultural event or performance of any kind – for example, the user could find and compare the styles of football playing across the member states, or study techniques of traditional boat building (if they have been documented by video recordings). But in fact EUDICO has been developed so far specifically to handle linguistic and gestural material, and this provides some insight into the way in which a specific research goal in the Humanities can drive technological development. Language-special features include the display and analysis of the speech signal, the specific annotation tools, and the kind of time-bound manipulations available from the interface. For example, for the study of gesture and Sign, one needs to be able to lock into two or more very small time windows enumerated in milliseconds, so that overlapping events can be characterized both independently and in relation to one another (say speech and gesture, or manual Signs and facial Signs).

The kind of tools sketched here need development in various directions. The Browsable Corpus needs to be developed to handle different kinds of object material, to support flexible user-definable hierarchies and to have general procedures for finding and launching analysis tools in a distributed environment. The EUDICO event analysis tool needs to be generalized so that annotations can be made onto the same general database over the net, so making it possible for remote research partners to collaborate on the annotation of specific events and to add comments to a particular time-locked series. These are quite demanding technological developments.

Conclusion

This short contribution sketches an approach to language as cultural heritage, within the framework of ECHO. We have suggested that the many distinct Sign Languages of Europe are part of the hidden cultural heritage of our continent, which can be unlocked by provision of the technology that would make the systematic comparison and collaborative research on these languages possible – while at the same making them available to the general public on the internet. The required technological tools are quite clear and circumscribed, but would have wide application. The Sign Language project can stand as a microcosm for the ECHO initiative: let us use the technology to simultaneously satisfy public interest in our cultural heritage and boost research and collaboration in the Humanities – in doing this we will drive the technology in demanding and profitable directions, and help to construct a European research area in the Humanities.