www.ichim.org

Les institutions culturelles et le numérique
Cultural institutions and digital technology

École du Louvre
8 - 12 septembre 2003

THE DIGITAL PRODUCTION LINE: A DUTCH IT RESEARCH PROGRAMME FOR THE BENEFIT OF THE CULTURAL HERITAGE SECTOR

Paul M. Doorenbosch M.Sc.
Koninklijke Bibliotheek, national library of the Netherlands,The Hague

Prof. dr. H. Jaap van den Herik
Institute for Knowledge and Agent Technology (IKAT), Maastricht

« Acte publié avec le soutien de la Mission Recherche et Technologies du Ministère de la Culture et de la Communication »

Abstract

A consortium, consisting of seven IT-research institutions, the five leading cultural heritage institutions and a number of commercial parties in the Netherlands, has submitted an IT research programme for the digitised cultural heritage, called the Digital Production Line (DPL). Although the partners in the consortium have different points of view on the cultural heritage domain, their common motivation is to develop "new methods and techniques in the domain of multimedia and agent technology to support the cultural heritage sector effectively in their interaction with (among others) researchers and the general public."

The three main objectives of the DPL programme are:
To increase the efficacy and efficiency of digital access to Dutch cultural heritage for research, education, and the general public (including tourism).
To reinforce the knowledge infrastructure of IT researchers, the cultural heritage sector, and relevant companies.
To improve the international position of Dutch researchers within the interdisciplinary research field of IT and humanities.

An additional benefit of the DPL will be to open up new possibilities for the presentation and study of the cultural heritage. The DPL will take the digitised cultural heritage as its input. Five IT research lines (fundamental and applied research): Metadata, Interoperability, Semantic-based Access, Navigation and Presentation, and Knowledge Enhancement, form the roadmap to a futuristic Holodeck. The five IT research lines are flanked by two society-oriented research lines: Digital Durability and Intellectual Property Rights.
The results of the programme will not only be of interest to the Dutch cultural heritage and IT world, but will be sufficiently generic for the cultural heritage sector world-wide to benefit. The consortium thinks it important to concertate and co-operate with other parties, nationally and internationally, that are involved in related research.

Keywords: Cultural Heritage, IT Research, Digitisation, Semantic-based Access, Knowledge Enhancement, Interoperability, Navigation and Presentation

1. Introduction

The collective memory of the Netherlands is stored in our cultural heritage. Huge numbers of archives, books and magazines, paintings and other art objects, audio-visual resources, folkloristic and archaeological objects are kept in various places, often in buildings that are themselves part of our cultural heritage. All these witnesses of our past and present are indispensable components of our national identity and our society, and contribute to the quality of life. Nowadays many people, be they teachers, researchers, politicians or people with a general interest in culture, would like to have cultural collections and information about them at their disposal, independent of time or location: at home or in the street, on their mobile or on their computer. Therefore, museums, archives and libraries try to achieve effective access to digitised objects and digital durability of our entire cultural heritage. These are also the main focus points underlying the recently submitted Digital Production Line (henceforth: DPL) research programme. Figure 1 illustrates the DPL metaphor, visualising the interconnected components of IT-based access to the digitised cultural heritage of the Netherlands.

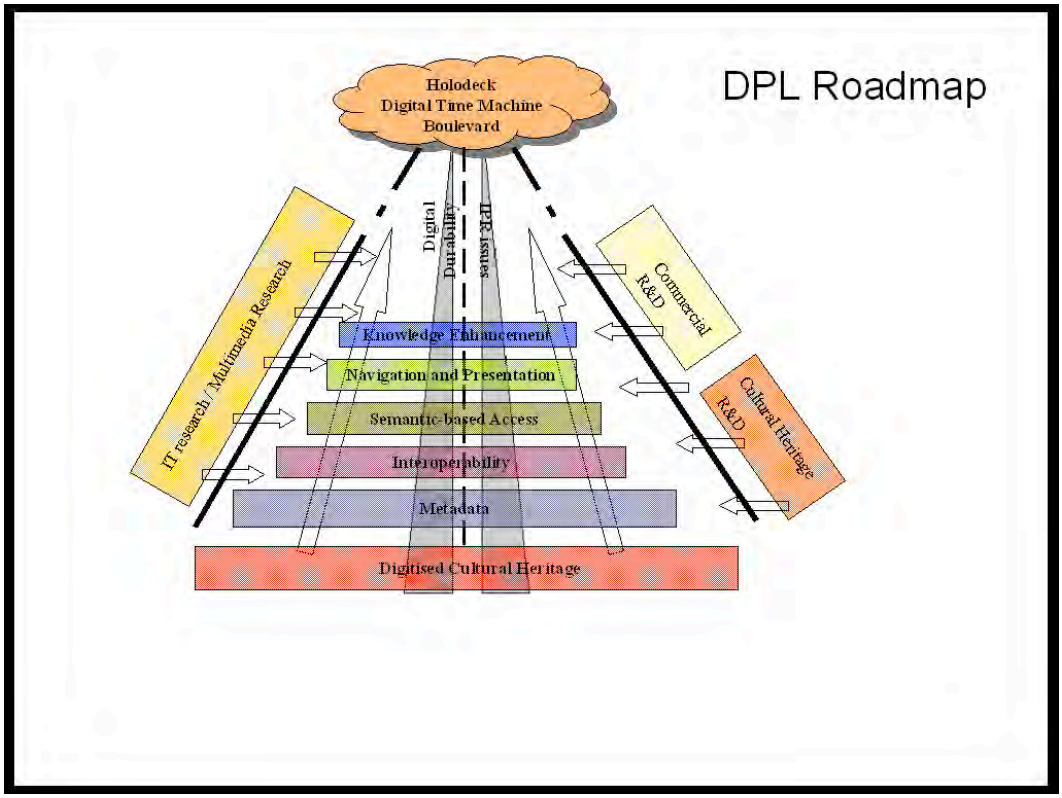

Fig. 1: A graphical illustration of the Digital Production Line.

The volume of the Dutch digitised cultural heritage is huge and increasing every day. The funds and time required to present all our cultural materials in a traditional way are lacking. The use of advanced information and communication technology (IT) enables sustainable accessibility of the digitised cultural heritage.

The main objective of the DPL programme is to carry out fundamental and applied research, resulting in an overall knowledge infrastructure providing the public with direct, time and location independent access to the major part of the digitised Dutch cultural heritage, and answers that are relevant and meaningful in the context of their queries. This is a very ambitious goal. Therefore it is split into seven distinct research lines, five IT-related topics (metadata, interoperability, semantic-based access, navigation and presentation, and knowledge enhancement) and two society-related ones (digital durability and IPR issues). Each research line consists of three to four different projects. The research lines each address issues of major importance to the cultural heritage sector. They are discussed in more detail in section 5.3.

The submitted proposal is in the final stage of the auditing process. In November of this year the governmental commission will decide whether the proposal will be granted 30 million Euros, half of its total costs of 60 million. Even if the funding is denied, the proposal has enabled the IT and cultural heritage sectors to explore their common interests and develop joint research plans and to build a strong network to the mutual benefit of both. This has been a positive outcome, even before the project has actually started.

2. The organisational structure of the project

The programme has been prepared by a broadly-based consortium, that will also direct its execution. Due to the grant conditions the consortium is composed exclusively of Dutch partners. During the realisation of the programme, concertation and co-operation with other parties, nationally and internationally, who are interested in co-operation or involved in related research, will be a major point of attention.

Below we introduce the consortium members, with a description of their activities and their relation to the project.

The consortium accommodates four different kinds of participants:
The NWO Research Councils for the Humanities (GW) and the Physical Sciences (EW);
National representatives of the cultural heritage domain;
IT research groups of five Dutch universities, two research institutes, and one law section;
IT companies (Cap Gemini Ernst & Young, IBM, Oracle, SUN).

2.1 The Netherlands Organisation for Scientific Research (NWO)

NWO (http://www.nwo.nl) plays an important co-ordinating role in the Netherlands with respect to science management. As regards the DPL programme, NWO takes care of the submission of the proposal, will monitor and co-ordinate the progress of the project, and will apply its quality assessment procedures at agreed intervals during the project. Thus, NWO has a facilitating role in bringing together the IT and humanities sectors and will offer manpower, contacts, and organisational structures to enable the participants to perform research, to organise meetings and conferences and to publish articles, books, and brochures. Moreover, NWO will be responsible for the administration, for monitoring the continuity, for the application of the research results, and for the interaction with the general public.

2.2 The cultural heritage sector

In the research programme we distinguish three main sections in the cultural heritage sector, viz. museums, libraries, and archives. The monument and archaeology sections are not actively participating in the project, although they may be involved at a later stage. Obviously, the aims and intentions of these three cultural heritage sections are closely connected and rather concurrent, especially in the way they try to attract and 'entertain' visitors. However, with respect to IT developments, and in particular in the effective and efficient application of IT tools, their goals, organisational methods, and stages of development diverge considerably.

From the cultural heritage sector, the following organisations are core members of the consortium and initiators of the programme: the National Library of the Netherlands, the Rijksmuseum, the National Archives, the Netherlands Institute for Cultural Heritage, the Netherlands Institute for Sound and Vision, and the Netherlands Digital Heritage Association. There are other cultural heritage institutions involved in the individual projects and more and more organisations are expressing their interest in participation.

The organisations mentioned above started during the last decade to digitise (parts of) their collections. At first these digitised collections were relatively small and simple in structure. This resulted in a motley collection of digitised objects, with different kinds of gateways to the Internet and with a variety of access paths from the perspective of the user. Moreover, the distributed character of the digitising process has led to scattered storage and varying accessibility procedures. The past few years have seen increasing investments in cross-sectoral projects, such as Memory of the Netherlands (http://www.geheugenvannederland.nl), carried out by the national library of the Netherlands (http://www.kb.nl) [see the presentation by the same author at ICHIM 2003] and the Cultuurwijzer/Cultuurwijs projects by the Netherlands Digital Heritage Association, DEN (http://www.den.nl). 'Cultuurwijzer' (http://www.cultuurwijzer.nl) is a selective portal to a large number of digital heritage information resources for the general public. 'Cultuurwijs', built in a similar way, is directed at school-age children. The proposed multimedia-based approach of the DPL fully covers the existing problems and developments in the cultural heritage sector. It will provide a generic structure and solutions for improved access to digitised cultural heritage collections, exceeding the limits of the individual collections.

2.3 The universities and IT institutes

The five universities which will be involved in the programme's IT research are: Technical University Eindhoven, University of Maastricht, University of Groningen, University of Tilburg, and Vrije Universiteit Amsterdam. The IT institutes involved are: the Telematica Instituut, and the Max Planck Institute (MPI). The University of Leiden (Faculty of Law, Department of Law and Computer Science) participates in the research

line on intellectual property rights. Three partners (University of Groningen, University of Maastricht, and Max Planck Institute) will involve humanities scholars in their DPL Projects. Many of the IT research groups involved have connections with the cultural heritage sector in the context of one of the individual projects mentioned above. Many of the researchers involved participated in earlier joint projects of the IT and cultural heritage sector, such as ToKeN 2000 (http://www.token2000.nl). The huge amount of data as well as its diversity make cultural heritage applications a challenging research domain. The research approach and directions chosen in the projects such as ToKeN 2000 form a good starting point for the DPL programme.

2.4 The IT companies

A well-known example of an existing partnership in this research domain is that between IBM and the National Library of the Netherlands. There are many more of such fruitful partnerships, for instance between CGEY (Cap Gemini, Ernst and Young) and the digital durability project executed by ICTU, as well as the database development projects carried out jointly by Oracle and several small-scale university groups. Finally, it should be mentioned that SUN (partner of the Netherlands Institute for Sound and Vision) has agreed to provide relevant IT knowledge originating from their multimedia development centre in the United States.

The companies' interest in the DPL and their willingness to contribute to it is illustrated by the various research plans incorporated in the projects. These companies expect that the research carried out as part of the project will enable them to develop products, tools and services which will serve the community at many levels of education and in many functional applications. Their commitment is a sign in itself that they believe that the research as undertaken in the DPL programme will contribute substantially to future IT developments in research, education and entertainment or tourism.

3. General strategy, context, and aims

For the DPL to achieve its goals, a general strategy is needed, as well as a well-formulated context and clearly-set aims.

3.1 General strategy

The general strategy is characterised by interdisciplinarity, co-operation and a pragmatic attitude.

Interdisciplinarity

One of the focus points of NWO's strategy for the coming years is interdisciplinarity. Interdisciplinarity is inherent in the programme's overall design, since it is defined as an IT programme for the cultural heritage sector. Interdisciplinarity plays a part in all the subjects addressed, in the various backgrounds of the researchers involved, and in the way the programme will be run. We consider it to be an essential feature of the programme that a large part of each of the various projects will be carried out at the locations where the cultural objects are actually stored (or exhibited), i.e. in the museums, libraries or archives. This implies that a Ph.D. student or postgraduate researcher will have his/her working place in a cultural heritage environment and will be supervised and advised by a cultural heritage expert on a daily basis.

Co-operation

Co-operation is achieved at the level of the research lines. The five IT research lines each involve at least two distinct IT research departments (university or institute), at least two cultural heritage institutes and at least one IT company. The projects within each research line are focused primarily on that specific research line but will also be strongly connected to other research lines. In each project, all three types of participants are involved: IT researchers, cultural heritage representatives, and IT companies. A schedule of meetings and a protocol of how to exchange information will reinforce the co-operation.

Pragmatic attitude

Each IT research project group relies on two types of input: (1) digitised cultural heritage objects (accompanied by clearly stated problems) and (2) database management systems, multimedia machines, and advanced IT tools. The objects and tools to be used in the project are provided by the cultural heritage sector and the IT companies to the researchers. It may happen that conditions are imposed (due to institutional rules or to company regulations) in certain stages of the project. For instance, when a tool is new and the specifications are not yet publicly available. In such situations it is agreed that all parties should show a pragmatic attitude since this will benefit effective co-operation necessary to achieve their joint objectives.

3.2 The context

In the description of the general strategy, the context was already implied. Research is carried out in two places: in the cultural heritage institutions and in the IT institutions (universities). Although the DPL programme has a national orientation, it is at the same time embedded in an international context since (1) it has an international board of advisors, (2) all researchers involved have many international contacts in their field, (3) all the cultural heritage institutes are involved in international projects, and (4) most companies have an international basis and will provide researchers with knowledge of the latest developments in multimedia and agent technology. International developments will continue to be followed closely and we are interested to meet parties that are willing to exchange information or explore possibilities for co-operation.

3.3 The aims

The overall objective of the DPL programme, as described in the introduction can be divided into three distinct goals:

1. To increase the efficacy and efficiency of digital access to the Dutch cultural heritage for research, education and the general public;
2. To reinforce the knowledge infrastructure of IT researchers, the cultural heritage sector, and relevant companies;

3. To improve the international position of Dutch researchers within the interdisciplinary research field of IT and the humanities.

1. To increase the efficacy and efficiency of digital access to the Dutch cultural heritage for research, education and the general public

Currently, the accessibility of the rapidly growing amount of digitised cultural heritage materials is unsatisfactory. The main problem is that the digital materials are stored in many different databases, each with their own format and structure. Obviously, effective and broad accessibility of the cultural heritage is important in particular for scholars in the humanities. Due to the current lack of tools (e.g., standardised metadata, poor navigation, bad interoperability, etc.) they are prevented from finding and using all the available information. The IT research undertaken in the DPL programme will open up new opportunities for research in the humanities. The development and application of IT techniques as foreseen by the DPL is necessary to increase the efficacy and efficiency of digital access.

2. To reinforce the knowledge infrastructure of IT researchers, the cultural heritage sector, and relevant companies

One of the main goals of the Council for the Physical Sciences is to enhance the knowledge structure with respect to computer science in the Netherlands. The Council for the Humanities has as an important aim to provide their researchers with a digitised, durable and accessible data infrastructure of relevant cultural heritage objects. Although there are examples of interesting cross-sectoral programmes, the development of a knowledge infrastructure for the intersection area of IT and cultural heritage is still in its infancy. The DPL programme will support this infrastructure along three lines: (1) the research lines as described in section 5.3 need junior and senior researchers from both sectors, (2) relevant IT companies supporting the development of specialised products and the infrastructure for the cultural heritage sector will be involved; and (3) at several locations new senior positions will be created on the intersection of IT and cultural heritage.

Bringing together front-rank researchers from the universities, the two IT institutes, and the Research and Development departments of innovative IT companies and cultural

heritage institutions will give a major thrust to reinforcing the knowledge infrastructure of the intersection of the IT and cultural heritage sectors. On the one hand, many IT researchers have only a basic understanding of the specific problems and challenges in the cultural sector. On the other hand, the cultural heritage sector is largely unaware of the potential solutions that the IT sector may offer for their problems.

3. To improve the international position of Dutch researchers within the interdisciplinary research field of IT and the humanities.

An enhanced infrastructure for the digital cultural heritage will clearly demonstrate the success of the development of novel IT and multimedia techniques tailored to the domain of research in the humanities. The use of novel IT techniques and multimedia tools will facilitate the creation of new combinations and new viewpoints on facts in the cultural heritage domain. As a result, the DPL will strengthen the international research positions of both the Dutch IT and cultural heritage sectors.

4. Scientific relevance

4.1 Scientific and technological innovations

The scientific relevance of the DPL programme lies primarily in its innovativeness that is both practically applicable and interdisciplinary. The scientific innovations are guided by the structure of the DPL programme. As described earlier, we distinguish five IT-oriented research lines and two society-oriented research lines. Each of these research lines is expected to result in several innovative methods and ideas. All products will contribute to achieving our three aims, outlined in section 3.3. Moreover, we expect the products to play leading roles in the development of the futuristic concept of the Holodeck or the more mundane concept of the Boulevard programme. The latter is a national programme to build a 'boulevard' that is both physical and virtual, where current political and social events are explained in their historical context, from various perspectives and points of view. To achieve this the Boulevard will use advanced IT technologies.

The major breakthroughs expected from the DPL programme are (1) ubiquitous accessibility, (2) effective personalised presentation, (3) unlimited knowledge enhancement, and (4) ”perennial” durability combined with a balanced intellectual property rights study. Breakthrough (1) coincides with our first aim regarding accessibility; breakthroughs (2) and (3) with our research aim regarding the infrastructure and the benefits for researchers in the humanities; breakthrough (4) with our aim to improve the international position of the Netherlands in the domain of digital durability and intellectual property rights.

4.2 Ubiquitous accessibility

The first aim of the DPL programme (see section 3.3) is: increasing the efficacy and efficiency of digital access to the Dutch cultural heritage. Achieving this aim is considered to be a major advance, for which many smaller scientific and technological improvements have to be accomplished. In the research programme for each research line future scenarios are outlined illustrating ubiquitous accessibility to the cultural heritage and its environment. As a case in point, we take the metadata research line. To get access to an art object, an agent acting on behalf of the user should be able to navigate to the correct location. Metadata serve as signposts in the navigation process. The better the metadata, the better the results in terms of precision, recall, speed, and reliability. Moreover, metadata can pave the way with respect to interoperability obstacles. Furthermore, metadata can support presentation, semantic-based access, and even knowledge enhancement. Hence, improvements in the definition and use of metadata are needed for breakthroughs towards ubiquitous accessibility. A comparable contribution will be made by the other six research lines.

4.3 Effective personalised presentation

The second aim of the DPL programme is to reinforce the knowledge infrastructure. One major step towards achieving this aim is an effective personalised presentation. The user wishes the answers to his/her questions to be tailored to his/her information needs. The underlying idea is that a computer should know what these needs are. Therefore, information on the user’s interests should be gathered. There are many methods which

ultimately fall into one of two categories: (1) those that feature obtrusive interaction with the user, and (2) those that feature unobtrusive observation. The main drawback regarding obtrusive methods is that they burden the user with additional actions. For the unobtrusive observation the design of the user interface is a major issue. Gathering knowledge on the user consists of three essential questions: (1) what user behaviour to observe? (2) where and when to observe the user? and (3) how to interpret the observed behaviour? All three questions are highly relevant. Technically there are various possibilities, in particular with the help of metadata. The main problem, however, is the ability to identify which user is interacting with, for instance, a museum's system, otherwise it will not be possible to update and use the right user profile. From a legal point of view, privacy issues are at stake here as well. We need to ensure that the user's information is kept safe. For instance, the information on the server should be anonymous or encrypted. Moreover, there should be a transparent privacy policy. From the above it is clear that there is a tension between a perfectly personalised performance by a museum agent and the protection of the user's privacy, although this will diminish when in future optimal conditions for privacy protection can be realised. Moreover, it is well known that people tend to accept minor breaches of privacy, when this is outweighed by the benefits. It is to be expected that this will be the case with effective personalised presentation.

4.4 Unlimited knowledge enhancement

Knowledge enhancement means making clear how objects, information, or links are interpreted. It may deal with a certain type of object (e.g., a certain painting belongs to the impressionist school of 1880-1890), but also with types of material, contemporaries, and other, not immediately obvious relations. Thus, new structures encompassing diverse objects from different collections may become apparent. Automatic knowledge processing techniques may even take over parts of the research task, in particular when very large quantities of object data are involved. Knowledge processing techniques will also find new structures in non-analysed object data collections and object databases. This development brings us to the domain of data mining, which will account for many smaller breakthroughs in the various projects. In summary, the third breakthrough will be unlimited, qualitatively improved access to data. The desire for this breakthrough is voiced emphatically by cultural heritage organisations (Krikken & Piel, 2002). For them it

is essential to have techniques that are able (1) to scale up to very large databases, (2) to produce overviews, analyses, and indices that can be inspected and understood on a human scale, and (3) to normalise, enrich, and clean up very large databases that could not be normalised by humans on a human time-scale. Therefore, the knowledge enhancement breakthroughs will be in data mining and knowledge discovery techniques that scale up to very large databases beyond the human scale. To this purpose the research line relies on state of the art work in machine learning, a subfield of artificial intelligence. Moreover, considerable scientific innovation can be expected to arise from investigating the very large scale of cultural heritage data in combination with the rich multi-modality (sound, image, text) and relational structures (thesauri, history) in these data.

A schematic example, illustrating the various aspects that are relevant when offering access to users with different needs, as well as the creation of relations between the objects themselves and between objects and the knowledge about them is shown in figure 2.

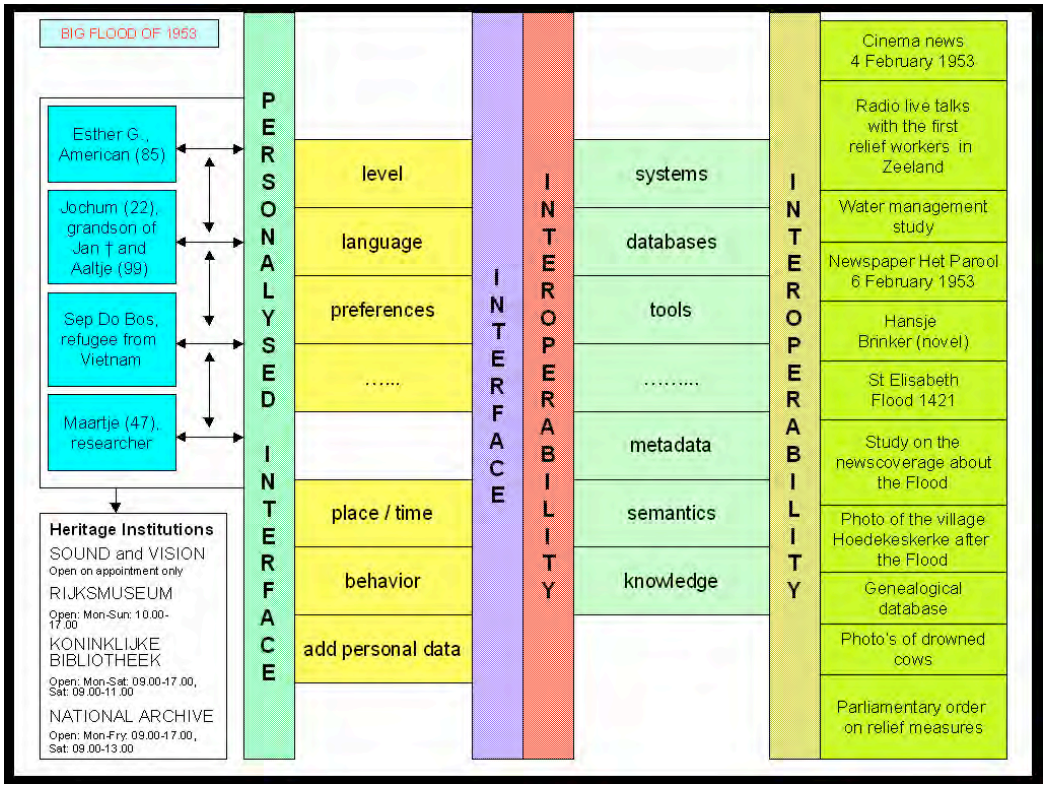

Fig. 2: A schematic example based on the subject of the Big Flood of 1953, the last large-scale flood disaster to hit the Netherlands.

The right-hand column of figure 2 contains the resources (objects, databases etc.). These resources, that are distributed geographically, should appear to the user as a coherent collection. This requires technical interoperability (top section second column from the right) as well as interoperability with respect to content. On the left a number of different types of users is listed, with different backgrounds and queries, whereby users may change roles over time (a scholar may at some point in time be a tourist). With regard to presentation, feedback and interaction interfaces must take into account these various aspects of users' roles (second column from the left). The final objective is to realise communication between users and the combined resources and the knowledge about them, as efficiently and effectively as possible.

4.5 "Perennial" durability and a balanced IPR study

The third aim is to improve the international position of Dutch researchers within the interdisciplinary research field of IT and the humanities. The fourth breakthrough will result from the co-operation between IT, cultural heritage and law researchers.
Digital durability obviously is a key research issue world-wide and will therefore be investigated thoroughly in other projects too. A proposal for an integrated network project called PATCH was submitted to the European Commission's 6th framework programme (Steenbakkers, 2003). Unfortunately it was not granted funding, but the intention is to find other means to carry it out. The Koninklijke Bibliotheek in the Netherlands has also planned a research project regarding some of its aspects. IBM and the National Library have build an e-archiving system (DIAS), and currently the development of modules to facilitate digital durability of, amongst others, digitised cultural heritage materials, is being investigated. In the DPL programme we will investigate durability in relation to metadata, interoperability, and intellectual property rights. We expect this combination to lead to new insights into the intricacies of the digital durability issues.

When objects need to be preserved digitally, we also need to preserve the metadata that accompany these objects. With respect to topical metadata one inherent problem is clear: the meaning of terminology changes. Hence, in the future, concepts with the same meaning can be referred to by different terms. Moreover, the use of a classification system will change in the course of time, due to its growth and the development of scientific

insights. For a long time cultural heritage researchers believed that clarifying the connection between older records and newer ones and indicating the association between historically connected objects could only be performed manually (at least for the greater part). The DPL programme shows that the use of supporting tools is possible, helping the user to make historical connections at the topical level. In the case of retro-digitisation (a scenario in the submitted research proposal) there is also a lack of interconnectedness at the level of the descriptive metadata. Standards for the description of objects in older catalogues and inventory systems deviate from the current standards or are completely lacking, resulting in poor connections. Nowadays, when we digitise the objects we must document their descriptions carefully. It is of vital importance that the researchers know the standards according to which the catalogues were compiled, as well as the changes made in the course of time. The need for a relatively new kind of metadata is also felt: administrative and technical metadata, describing the procedural and technical aspects of the digitisation process itself. (see for instance the working group on preservation issues of metadata, http://www.rlg.org/preserv/presmeta.html) . Much work on metadata and the preservation of digital objects has already been done in the library world (OCLC, 2001; and Koninklijke Bibliotheek: http://www.kb.nl/kb/menu/ken-arch-en.html)

Within the next five to seven years many digitisation projects will be performed by cultural heritage institutions in co-operation with universities and IT companies. In the submitted research proposal the complexity of intellectual property rights is discussed together with an appropriate problem statement and five research questions. The breakthrough expected to result from all the precise and detailed investigations into the legal aspects is: the building of a legal software agent that translates the complex legislation on intellectual property rights, which differs world-wide, in such a way that it is still legally correct and comprehensible to non-lawyers, such as IT researchers and cultural heritage investigators and organisers.

5. The Digital Production Line: situation analysis

The DPL programme focuses on the Dutch cultural heritage. Below, some figures and characteristics of the various collections are given in order to enable an effective situation

analysis. In section 5.1 we provide some qualitative and quantitative data regarding the cultural heritage, then in section 5.2 we formulate the general problem statement of the programme. Section 5.3 presents the division in research lines of the DPL programme.

5.1 Data on the cultural heritage

Throughout history, cultures have preserved objects for their cultural value. These objects allow future generations to recognise their own culture and to become aware of their common identity. As such, cultural heritage acts as a 'glue' for cultures. Moreover, it is the subject of education and scientific research. In addition, cultural heritage is a main provider of tourist attractions. Every year, our national cultural heritage institutions are visited by many Dutch citizens and foreign visitors.

Figures and values

The total size of the Dutch cultural heritage is difficult to estimate but is certain to be huge. In the Netherlands there are at least 80 large collections that together contain more than several millions of objects (Quick scan, 2002). The economic value of this heritage is even more difficult to estimate since the real value is symbolic rather than economic. Nevertheless, the estimated monetary value (22 billion euro's in 1998, art collections only; according to a survey by the Dutch National Bureau of Statistics) underscores the enormous value of our cultural heritage. This is accentuated by the fact that the government is spending around 200 to 250 million euro's on an annual basis on the management of the cultural heritage sector. Revenues and secondary economic effects are probably much larger.

Four characteristics of the cultural heritage are of major relevance to the DPL programme.
The volume of the cultural heritage is huge.
The cultural heritage objects are distributed all over the country. They are exhibited or stored in 900 museums, 400 archives, and 1100 libraries.
The collection of cultural heritage objects is heterogeneous, ranging from buildings to books and pictures.

Cultural heritage is generated in a largely unpredictable autonomous process. Material and immaterial products of human activity and creativity enter the domain of cultural heritage in a continuous and perennial stream.

Digitising the cultural objects

As a result of the characteristics mentioned above, managing the cultural heritage has always been a major challenge to governments and institutions. In the past decades, as a first step in dealing with this challenge, many cultural heritage institutions started to digitise their cultural objects and to make inventories using computer technology. A substantial but still small part of the collective heritage is now digitally represented in documents, images, or even three-dimensional models. However, the digitally "represented" cultural heritage objects have inherited the same four characteristics mentioned above.

The digitised cultural heritage is large, encompassing many digital documents.

The digitised cultural heritage objects are stored on many different computers and databases, at many different locations.

The collection is heterogeneous as it is stored in a wide variety of digital formats and systems.

New elements of digitised cultural heritage are continuously being produced.

An effective and efficient accessibility and opening up of the digitised cultural heritage is hampered considerably by these four aspects.

A period of five years reviewed

The following observations were reported five years ago in the SURF study entitled Alles uit de Kast (WTR-SURF, 1998).

Only a small part of our cultural heritage has been digitised as yet.

The use of IT in cultural heritage institutions is limited to the storage and presentation of raw data.

The cultural heritage institutions lack the knowledge and experience to exploit the full potential of IT.

The accessibility of the digitised cultural heritage is poor.

The number of digital visitors to the cultural heritage institutions could be increased considerably.

Today, it is rather sad to note, these observations still apply. A very recent report of the Telematica Instituut (Bruinsma & Velthausz, 2002) concludes that:
Many cultural heritage institutions digitise their collections for collection management purposes only.
An overall vision on IT and the cultural heritage is lacking.
Over the past years many individual initiatives were funded, but without an eye for coherence.

The Telematica Instituut advises to develop a nationally co-ordinated approach of the use of IT in the cultural heritage sector. This should involve co-operation between researchers, institutions and companies. The DPL is such a co-operative approach, comprising science, culture and industry. It applies innovative IT techniques to the cultural heritage.

5.2 General problem statement

The main problem addressed by the DPL is how to deal with effective accessibility and the opening up of the digital cultural heritage despite its size, (geographic) distribution, and heterogeneity. In addition, the DPL explores new possibilities for the preservation, presentation, and study of the cultural heritage.

The working area of the DPL consists of the digitised objects of the cultural heritage domain. However, what is a digital working area without virtual roads, without lighting and without signposts? The DPL programme will build these roads, with lighting and signposts, with buildings and facilities.

Researchers and users will co-operate in new ways in the Digital Production Line. The first obstacle is the disagreement on definitions between the IT and cultural heritage domains. Therefore, right at the outset of the programme we will have to agree on standard definitions of the key concepts.

5.3 The structure of the programme

The five main IT research lines of the DPL, i.e., metadata, interoperability, semantic based-access, navigation and presentation, and knowledge enhancement, in combination with the supporting research lines of digital durability and intellectual property rights, have been identified by scientists and domain experts as representing the main components for achieving the goals set. Here we will briefly discuss the various research lines.

Metadata

Of all the research lines, metadata is most closely related to the established practice of curators and other cultural heritage workers. As metadata is in some respects a digital analogue of the addition of labels and explanatory texts to the preserved objects, its incorporation as a research line is fairly obvious.

Interoperability

The interoperability research line follows from the need to connect the databases of different cultural heritage institutions. The heterogeneity of the formats and the semantic variance of objects employed by the various institutions necessitates a focus on interoperability.

Semantic-based access

The DPL examines the vast knowledge bases of the digital heritage. Automatic identification and classification of the various knowledge elements are of major importance. Due to the heterogeneity of formats and conventions, indexing and classifying relevant knowledge is a difficult task, technically as well as semantically.

Navigation and presentation

When navigating the cultural heritage knowledge bases, users are invited to translate their personal needs and desires into digital actions. Presenting the - intermediate or final - results in a personalised way is therefore considered to be an essential component. Once the three previous stages have been completed successfully, in other words, when

metadata has been added to the cultural heritage objects, when the databases are interoperable and the knowledge contained in them can be accessed in a semantically meaningful way, efficient navigation and presentation techniques provide the final stage of making the cultural heritage accessible to various groups in society (researchers, experts and the general public).

Knowledge enhancement

The main goal of cultural heritage research is to discover new knowledge, to add to the existing body of knowledge about the objects and their creation processes. Knowledge enhancement is a rather new research topic that originally focused on generating new knowledge from existing knowledge. Nowadays, knowledge enhancement is supported by a plethora of techniques, such as neural networks, pattern classifiers, data mining, etc. It is not without reason that knowledge enhancement is at the top of the hierarchy of the production line. It is the most advanced research line of the DPL programme.

Digital durability and intellectual property rights

These more society-oriented research lines are vital to the programme's success. Research in both areas is carried out exclusively in relation to the IT research in the five basic research lines, not independently.

5.4 Reference architecture

The DPL will be a dynamic, distributed, service-based environment, comprising autonomous cultural heritage entities and service providers that make available enhanced content. The cultural heritage institutions will utilise content enhancement services to obtain additional descriptive metadata to improve the accessibility of their content. The use of open standards will ensure that the DPL is able to implement solutions from multiple vendors and can easily be extended with new partners. Effort put into digital durability issues will support longevity of the cultural heritage content, and will improve accessibility of content through the adoption of uniform content formats.

For a programme like this to achieve optimal results, it needs a reference architecture in which content, services and algorithms and software - both delivered or under

development - can be managed and accessed efficiently. The DPL shall adhere to the following general architectural principles:

Utilisation of a service-based architecture

This is necessary to ensure that the development of user services and the underlying IT-services (e.g. web-enabling, web services) are not mutually dependent. Thus, the architecture will be dynamic, supplying a framework in which new technology can easily replace earlier solutions. In addition, new and unexpected user applications can be realised quickly and easily.

Use of open standards

It is expected that through the use of open standards (such as J2EE, UDDI etc), specific products that carry out standardised functions can be readily positioned within the framework of the overall architecture, making use of well-defined and clear interfaces. Moreover, it makes the specific selection of products less relevant, thereby increasing the flexibility of the DPL participants with regard to product choices both now and in the future.

Modularity that positions the functions of the architecture clearly and unambiguously

A high degree of modularity is necessary because many parts are developed independently, but simultaneously. Implementation of the applications should also be possible on a modular level. Careful agreements should prevent loss of coherence or work being duplicated during development.

Administratability/controllability as prime concerns

With many different parties contributing to the processes of both meta and primary information collection, control tasks should be embedded within the architecture.

Take into account the interests of all stakeholders

The DPL programme has many different stakeholder perspectives which need to be clearly accommodated by the reference architecture.

Content assets and enhancement services will be decentralised

This means that the cultural heritage institutions will remain autonomous and maintain their own assets, and enhancement services will be developed and operated by many of the project partners. Distributed solutions will accommodate this approach.

Co-existence with existing legacy digitised content and metadata

Investments made in the past in access and availability of resources and knowledge must be re-usable in the newly developed environment. For the ongoing generation of cultural knowledge it is necessary that older knowledge stays available and re-usable.

5.5 Exploitation

No government will fund a project unless it can state clearly what follow-up is foreseen for the results of the programme. Evidently the cultural heritage sector itself will take care of the content, the digitised objects. But smaller institutions may face problems coping with it over time. Due to, for instance, the lack of durability of formats it will be difficult to keep resources accessible. The same goes for the services and tools. The consortium will take responsibility for drawing up a business plan for the period after the project's end. As the programme benefits the cultural heritage sector, it is clear that the consortium partners from this sector will have to fund future exploitation or at least must try to raise the necessary funds. The dissemination and continued development of the scientific results will be the responsibility of the universities. Companies will be given the right to develop commercial products based on them (detailed agreements regulating this still have to be drawn up). Participants in the programme on the other hand, will claim the right to licence these products at discount fees or develop new services bases on them themselves. One of the challenging dilemmas that needs to be solved during the next months and years.

References

Krikken, J. and Piel, W.H. (2002). Mining museum riches: an introduction to digital methods in the discovery of changes in the diversity and distribution of life on earth. In J. Meij (Ed.) Dealing with the data flood: Mining data, text and multimedia. The Hague, The Netherlands: Stichting Toekomstbeeld der Techniek, 140-165.

OCLC (2001). Preservation metadata for digital objects: a review of the state of the art: a white paper by the OCLC/RLG Working Group on Preservation Metadata. January 31, 2001 (http://www.oclc.org/research/pmwg/presmeta_wp.pdf).

Quick scan digitalisering cultureel erfgoed in Nederlandse collecties. Reekx Advies, April 2002. (= Quick scan of the digitisation of cultural heritage collections in The Netherlands)

Steenbakkers, Johan (2003). The PATCH project: integrating research & standardisation of digital preservation. The Hague, The Netherlands: KB, June 2003 (http://www.kb.nl/kb/ict/dea/patch/patch.pdf)

Velthausz, D. & Bruinsma, E. (2002). Inventarisatie infrastructuur digitaal erfgoed: een onderzoek naar visies, belemmeringen en oplossingen .Telematica instituut (= Survey digital heritage infrastructure: an investigation of views, obstacles and solutions) (http://www.minocw.nl/cultuurict/doc/2002/rapportIDE.PDF)

WTR-SURF (1998). Alles uit de kast: op weg naar een nationaal investeringsprogramma digitale infrastructuur cultureel erfgoed . Utrecht: Wetenschappelijk Technische Raad SURF. (= 'Bring it all out': towards a national investment programme for a digital cultural heritage infrastructure) (http://www.surf.nl/publicaties/index2.php?oid=47)

Appendix

The following is a list of potential services that could be developed and utilised within the DPL.
Optical Character Recognition (OCR)
Teletext (subtitles for the deaf and hard of hearing)
Identification of silence, music and speech
Conversion from speech to text (Dutch)
Speaker Identification
Shot Detection
Key frame-extraction: most representative frame per shot
Transcoding to different audio and video content formats
Handwriting: recognition and annotation
The spoken cultural heritage: speech technology for search and retrieval
Automatic recognition of visual cultural heritage objects
Knowledge enhancement through combining images and text
etc.