

Cardiff, Digitisation Conference 2007, Dr. Ralf Goebel, Panel Session: E-content selection and management, 19 July 2007

The “Digital Library” in the future

Ladies and gentlemen,
research and teaching are almost entirely digital nowadays. This also applies to information services and to conveying information. Allow me to give you an example: In a lecture on art history the lights are always switched off and the video projector switched on. This is because, for an art historian to be able to analyse, interpret and visualise his subjects, he always relies on an image. In the past, the lecture theatre lights were dimmed so that slides or microfiches could be shown. Now, of course, slides have been replaced by digital images. These digital images – and the added value digital features that come with them – play a vital role in research and teaching in every discipline, not only in art history.

That’s the reason why in Germany the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), as the largest non governmental funding agency, is providing as much as €50 million in funding for the establishment of digital libraries. The DFG is the central funding organisation for research in Germany. Its main purpose is to fund research projects by scientists and academics in universities and research institutes. In other words, the mission of the DFG is to promote science and the humanities — and therefore, by ex-tension, to promote academic libraries.

The digitisation of valuable historical sources and books accounts for approximately €15 million per annum. Experience in recent years, however, has shown that drawing up stringent selection criteria for digitisation is not successful. All attempts at selecting or creating a canon of works for inclusion in a digital collection have failed because of the heterogeneity of research interests and approaches – and rightly so, in my opinion. The DFG is now of the opinion that mass digitisation is the most promising route to take. Anything that is not available online will soon be irrelevant to the fast-moving global world in any case. Of course that does not mean that we do not support the digitalisation of special collections. But if we do that, scientists are involved in the project ideas from the beginning.

DFG also is particularly active in licensing for existing electronic journals. However, the costs involved are so high that we only can afford a selection. This selection is based on criteria such as the scientific relevance, the existing level of availability in Germany as well as the price. The opportunity to make the contents accessible for the entire German population also plays a major part in the decision. To date the DFG is funding national licences for Germany to the tune of around 15 to 20 million euros a year. In comparison to the total amount of research funding in Germany, 15 to 20 million euros are, of course, a paltry sum. There are, however, initial signs of a political move towards boosting this area.

Finally, there is also the question of German literature printed in the 20th Century to consider, especially books that are out of print. For digitising this literature we will provide up to €10 million per annum as of next year. Unfortunately, I cannot yet go into much detail about the selection criteria – one of the key factors to take priority will necessarily be the consideration of the legal preconditions for this project. As an initial step in this process stands the creation of a joint working group with representatives from the DFG and German publishers. In the end we hope to see the first proposals for that digital library of the 20th century at the end of the year. A key idea is to use a DFG funded structure that we call Virtual Subject Libraries. The purpose of DFG-funded Virtual Subject Libraries is to build subject-specific information systems which researchers can access directly via the Internet. Currently we support over 30 Virtual Subject Libraries, e.g. for economics , physics , political science , history , and for regional categories such as Middle East and Eastern Europe. Over the last few decades, the participating special subject collection libraries have accumulated practically exhaustive subject collections of nationally and internationally published knowledge. The digitisation and nationwide electronic provision of these subject-specific holdings will include the digitisation of historic periodicals and monographs from 1800 onward, the digitisation of non-public-domain periodicals and monographs and the digitisation of subject-relevant product lines from certain publishers, if their acquisition is of a high priority for special subject collection libraries.

So the DFG provides a significant amount of funding for the creation of digital libraries of both books and other print media. The digitised content will then be made accessible via centralised portals such as the portal for digitised imprints ZVDD, which will also act as long-term archives. The digitisation of books, however, is set to become less significant over the coming years for a number of reasons. This is partly because, in a few years' time, most literature available in the public domain worldwide will be available online in any case. This is not only due to the efforts of the European and American research funding bodies, although they play a significant role, but is also due to public-private partnerships with investors such as Google and Microsoft. This means that, in future, the DFG will focus more attention on the digitisation of images, television and radio broadcasts as well as scientific primary data. Finally, I personally view the term "collection" as an anachronism, no longer appropriate in the digital age of the 21st Century. In my opinion, the philosophy of collecting is losing the crucial role it once had in modern information society. Allow me to explain:

From a historical point of view, a collection is seen as a cohesive set of information, which can either be defined thematically, in terms of the medium, the driving interest of the collector, or the acquisition profile of the library. For the way research was done in the early 18th Century, collections undoubtedly played a fundamental role as the key to understanding nature. This was where the ordering, classification and description of minerals, plants and animals occurred, this was effectively where the knowledge of natural history was produced. At the same time, the act of collecting, particularly that of art objects, served as a means for developing and establishing the growing middle classes.

How do we treat the concept of a collection nowadays? Allow me to take a couple of "modern day" examples to illustrate my point: A few years ago we were all still buying records and CDs, in other words complete compilations of individual songs packaged for us by the publishers or producers. Today, young people compile their own personalised "best of" compilations of the latest songs using file sharing networks on the Inter-net. You can find Harry Potter in the internet, and you can't be sure whether this is the original manuscript or a manipulated copy or a text that was written by someone else. If the old truth that a book is a book is a book is not true any longer,

the idea of an authorised collection or the concept of a neatly rounded ensemble also is obsolete. Another example: In the past, we all had encyclopaedias as books on our shelves, whose quality and timelessness were guaranteed by their publishers. These encyclopaedias had a claim to legitimacy and completeness for decades to come. Indeed, you will still find the 24-volume “Brockhaus” encyclopaedia – the equivalent of the Encyclopædia Britannica – in my parents’ living room. Nice, but never used, because now, more and more people are using Wikipedia, the freely accessible encyclopaedia where readers can also contribute to and edit the articles they read, and where both the knowledge base and scope of knowledge are dynamic and subject to constant change. These new forms of active and specifically non-canonised knowledge formation constitute part of what is known as “Web 2.0”.

Web 2.0 is based on the principle of collaboration on dynamic content. In the scientific community we refer to this principle as the “virtual research and learning environment”. Unfortunately, at the moment the virtual research environment is still more of a dream than a reality. The few attempts that have been made have predominantly been based on proprietary environments or conventional portal technology and are restricted to individual disciplines or communities. In the not too distant future, however, there will be tools and web services that will work in any environment and on any platform, which will break the technical limitations and the boundaries between disciplines that have existed up to now. Users of these new interdisciplinary environments will always be not only users, but also co-authors.

Accepting the loss of significance of the concept of “librarian authorised collections” in the 21st Century also makes the concept of comprehensive digital collections obsolete. The same applies for the presentation of these collections in electronic format using portals. After all, the issue of selecting the material for digitisation is then irrelevant. As far as books are concerned, this can already be observed today. Google is currently in the process of digitising entire libraries – in Germany and around the world – or at least all of the material in those libraries that is in the public domain. So in future, most books that are out of copyright will be available online. The question of what gets digitised and added to a collection and what does not is then no longer an issue. In a few years’ time, these books from around the world that have been digitised will give us an initial idea of what this new type of digital

knowledge-based society is like. The “mass digitisation” of books that is taking place now will tomorrow be commonplace in many other areas such as art, music, historical documents, animal calls, mediaeval manuscripts, voice recordings of dialects, adverts, radio news reports etc. In the near future the question won't be what to select but how to navigate.

In addition to the important task of digitising analogue material there is, of course, also the huge task of handling the generated digital data that is created each day. How many digital photos are taken each day, worldwide? How many digital documents are created each day on how many computers and servers? How much data is produced by automated calculations and measurement processes? Just think of all the satellite and geographical data! This data is the raw material of the future. Not oil, but digital information will be the resource of the future.

In order to be able to deal with the digital data of tomorrow, we need to build strong supporting structures and tools today, which will enable us to collaborate in handling this enormous data volume and still hold on quality assurance principles. This applies just as much to the analogue materials that are digitised as it does to the genuine digital data. Establishing these new structures and tools needs to be an international effort. Representatives from the scientific and educational sector need to give high priority to this task and promote it for the sake of the scientific and educational community. So instead of talking about the selection criteria that we should use for creating digital collections in the next few years, I would far rather discuss the tools that will make it possible to manage the data we can anticipate.

On the JISC website I found some points for debate concerning the JISC digitisation strategy. The first question was about the role of specific collections and archives. Should JISC focus on specific areas or should JISC act more responsive, funding proposals from different areas without restriction. The DFG does both: in specific action lines we ask for ideas and proposals in fields we believe they are important. That lines are designed for mass digitalisation proposals. And at the same time we allow people to think free without restrictions. About 50% of the digitalisation budget goes into the funding of defined specific areas like the book digitalisation of German imprints of the 16th and 17th century – which is without any doubt a mass

digitalisation effort. And 50% of the budget we reserve for proposals that reflect the needs of the scientific community – without any content restrictions.

The second question was to what extent JISC should open its strategy to international needs. I believe that in the future we will cooperate much closer than we do that now, especially in building standards or in the development of tools and services. As far as the content side and the issue of how to select is concerned I think that it will be easier for each nation to concentrate on the national resources at first. That also is the idea of the European Digital Library which should be filled with national content of the member states. For that reason Germany focuses on German materials in the moment. That policy can change – but unless we are not equipped with easy processes and tools to prevent double digitalisation the limitation on national resources is more simple.

The third question was to what extent JISC should open its programmes for digitalisation ideas from a user perspective. DFG offers very new a line of action which is called “Libraries and Archives in Association with Research Institutions”. This funding line aims to enable researchers and information service centres to submit joint proposals that meet their common needs. Only joint proposals, in which the concerns of researchers and of information service centres are closely linked, will be supported.

Ladies and gentlemen, allow me to conclude with three summary points:

1. The concept of the “complete collection” and of the “portal” is losing relevance in the digital world of the 21st Century.
2. Collections and portals are being replaced by dynamic platforms such as virtual research and learning environments. These need to be developed and run by users from science and education independently. The users should organize themselves.
3. The development of tools to facilitate collaborative use of mass data will only be possible with international cooperation. The major research funding bodies from Europe and the USA need to cooperate more closely in this field.