

# In the interest of science: The Digital Information Priority Initiative by the Alliance Partner Organisations \*

Matthias Kleiner

*President of the Deutsche Forschungsgemeinschaft (German Research Foundation)*

The Digital Information Priority Initiative by the Alliance Organisations is a remarkable joint initiative by all German Science Organisations in the domain of digital information. It has already started to produce significant results and is most likely to enhance and transform the information infrastructure in a significant and sustainable manner at least in this country. But rather than presenting the operative details of the initiative, this article reflects on the reasons and motivations that prompted science organisations to take such an active role in addressing the topic of information infrastructure. To make it clear from the outset: the notion of “science” is not used in the narrower sense of denoting only the natural sciences but in the broader sense of including the social sciences and humanities as well. To bear this in mind will help to avoid misunderstandings and undue reductions.

The Digital Information Priority Initiative by the Alliance Organisations started in the summer of 2008, when the Digital Information Priority Initiative was adopted by all institutions that are represented in the so-called Alliance of German Science Organisations: Alexander von Humboldt-Foundation, Deutsche Akademie der Naturforscher Leopoldina, German Research Foundation, German Academic Exchange Service, Fraunhofer-Gesellschaft, Helmholtz Association of German Research Centres, German Rector’s Conference, Max Planck Society, Leibniz Association and German Research Council.

The Alliance Organisations agreed upon six priority areas: National licensing, Open access, National hosting strategy or better: a national strategy for safeguarding permanent and sustainable access to digital material, Primary research data, Virtual research environments and Legal frameworks.

So far the basic information. But what has prompted science organisations to take such an active role in addressing the topic of information infrastructure? As a result of the digital revolution, which has swept over the science generation of the late second and early third millennium, the significance of information infrastructure has changed considerably. Information infrastructure is no longer a background service that, as a researcher, one only took notice of when it failed to function – as was the case for years or even decades – for example, if the journal or the book that one needed was not available at the library, if one did not understand the library catalogue system, or if it took weeks before an item loaned from another library finally arrived.

In fact, in the two preceding decades, we have, through the digitisation of scientific research, experienced how the provision of information has changed as well. It has transformed from a background

---

\*Presented at APE Berlin, January 2010.

service structure to a parameter that not only supports scientific research, but rather one that has an increasing influence on it. This is reflected by two developments: First of all, the basic conditions for scientific research, has – as a result of digital information technology – changed in a way that was hardly foreseeable just two decades ago. Secondly, the digital revolution has not only considerably changed the basic conditions for scientific research, but rather that digital technology has impacted scientific research itself. In other words: It no longer seems appropriate to distinguish between the areas of “scientific research” and “information infrastructure”. In fact, these areas have become increasingly intertwined. And not only in the sense that science utilises infrastructures. But rather in the sense that input from science becomes increasingly important for the optimisation of information infrastructures and, vice versa, that scientific research is increasingly influenced by information infrastructures. Against this backdrop, it is imperative for the science organisations to play an active role in the transformation process and to address the associated challenges.

## **1. The basic conditions**

Digital information technology initially resulted in the disappearance of boundaries in the dimensions of space and time for scientific research. The space in which scientific communication and scientific research take place has, for quite some time already, known no national borders. Science is a global player. What started with the mobility revolution – the overcoming of space – was brought to an end by the digital revolution. While the mobility revolution served to overcome distances, the digital revolution virtually eliminated them. To communicate face-to-face with my colleagues around the world, I no longer need to leave my lab, my workshop or my office. And I am not limited to communicating and collaborating with those who are performing research at the same location as I am. Rather, distance no longer plays a role in deciding whom to talk to for a spontaneous reaction, an answer or immediate advice.

But the elimination of distances goes further. Access to knowledge is no longer limited by the physical availability of a book, journal or statistics that I need at a given moment. Rather, through digital technology, it is theoretically possible for all researchers to access any information they need for their scientific research from any place on Earth (provided the information exists in digital form).

In the dimension of time, there has been an immense acceleration of scientific research. The technology available to us today not only makes it possible to publish knowledge faster than in the analogue era, it also enables us to receive knowledge the instant it appears. And not only receive it, but also comment on it and expand it with new knowledge.

As a result of the elimination of the boundaries of space and time through digital technology, knowledge can be made available without time delay to anyone who wishes to partake. And this is the vision that we, as science organisations, pursue when we talk about what information infrastructures should look like in the future: All researchers should be able to access the information they need from any location and without delay.

Science organisations are working on implementing this objective. The technical requirements for achieving this goal already exist. However, there are still a number of legal, organisational and financial hurdles that need to be overcome before that which is theoretically possible today becomes daily practice in science and research.

## 2. The impact on scientific research itself

How does digital technology change the process of knowledge production? This will be illustrated by using four very different examples:

(a) In the humanities, initiatives have formed under the umbrella term "Digital Humanities" in recent years. These initiatives have the goal of establishing a sustainable infrastructure for the access to information as well as for the exchange and evaluation of information with the aid of modern information technology. The TextGrid project, which is funded as part of the D-Grid Initiative by the German Federal Ministry of Education and Research, is such an example. In the TextGrid project, an Internet-based platform is being created that offers the researchers tools and services for evaluating text-based data. This enables joint access to data stored in various locations in different archives, independent of the form of data and independent of software. Using the methodically advanced corpus linguistics as a model, the platform will allow entirely new research questions to be examined in the philologies as well.

For example, it is now possible for the first time to perform comparative analyses of hundreds of German-language novels from past centuries on a representative basis. In doing so, systematic comparisons of novels with other types of text can be performed; or existing annotations on the book history and the original formative context can be used to evaluate phenomena such as narrative structures and characters from an additional social-historical perspective. In short: A well-grounded analysis with a broad source base with numerous interlinked, interdisciplinary approaches will be possible. It is the merging of large quantities of text into a single body that can be searched and processed as a whole that makes such research even possible. Certainly, we are still in the beginning stages of this development, and what may sound exotic in many ears – "digital humanities" – will make the humanities an extremely innovative and revolutionary discipline.

(b) The second example is the way in which publishing occurs. Most of the current forms of publication are based on a more or less refined peer-review system. Be it a journal article or a monograph, it is always a small group of experts that decides whether a work sufficiently satisfies all scientific quality requirements for being published and whether it is "allowed" to contribute to the production of specialised scientific knowledge. The peer-review system and its quality control function remain unchallenged in their role of defining what bears enough merits to be published.

In addition to the classic peer-review system, another quality assurance process evolves. Here, I am speaking of the peer-review systems that are based on the collective or, one could say, swarm intelligence of a community. These review systems are open and transparent; all who can and who desire to make subject-specific contributions are able to do so. This publication type is also based on the principle that new scientific findings are subject to the criticism of the scientific community and must withstand this criticism. In this regard, there is no difference to the conventional, generally anonymous, peer-review system. The main differences are the time factor and the transparency. Knowledge that is published and reviewed in this way can immediately enter scientific discourse without the further delay associated with an organised peer-review process. But in this process, not only is the respective manuscript reviewed, but new knowledge is collectively generated through dialogue. In this way, publishing becomes a dynamic process in which collective intelligence in a virtual research environment yields new knowledge.

The current "test cases" of the open peer-review process nicely demonstrate that we are in a transition phase. In "Atmospheric Chemistry and Physics" and "Economics – The Open-Access, Open-Assessment E-Journal", two open access journals that practice the open peer-review process, a traditional review, overseen by an editorial staff, still takes place. It is, however, complemented by the open evaluation of the community. The traditional approach is apparently still perceived as being necessary in order to

guarantee that at least two competent assessments are presented to ensure the publication-worthiness of manuscripts. There is, after all, no guarantee that the community will respond to an article.

It will certainly take some time before the publication practice will transform from its current form to a process of collective knowledge generation. As *digital immigrants*, we will find it difficult to make the transition, as we feel secure in the old system, and the new system appears unripe, vulnerable and chaotic to us. But the upcoming generation of *digital natives* will not have such reservations. They – the *digital natives* – will not only use digital technology to simplify already established processes, they will have internalised that digital technologies give rise to new processes. Digital technology will no longer be a service – it will be an element that structures the research process.

(c) The third example to illustrate how digital technology influences the process of knowledge production applies to primary research data – that is, the data that serves as the basis for the process of scientific knowledge development or upon which the process of scientific knowledge development draws.

Everybody is familiar with the problem: In research projects, extensive resources are often used to acquire data that initially serves the function of shedding light on the scientific problem that was the motivation for collecting the data. But what happens when the project has ended? When the questions one had have been answered? For DFG-funded projects, this data must be preserved for ten years in order to be able to re-examine the results if necessary. These regulations for good scientific practice are aimed at being able to verify the results at a later point in time should it be necessary. It is not the intent of the regulations to store the data in order to reuse it in other research contexts.

I do not mean to imply that all data that is acquired during the course of scientific projects may also potentially be relevant in other research contexts. There may, in fact, be very many data sets that are so singular, so specific, that it would simply make no sense to place them in other contexts and interpret them.

But what about data that is collected in long, well-described measurement series and which is, above all, not reproducible, for example, climate or environmental research, of oceanography or geophysics? Meteorological measurements cannot be reproduced. Measurement series for documenting the pollution of ecosystems or for tracing the spread of parasites or viruses are likewise not reproducible with respect to time.

If this data is not stored in a transparent and sustainable structure, the result is an irresponsible economic loss. Some disciplines have addressed this responsibility and, during the Geophysical Year of 1957, started to establish the so-called World Data Centres for the earth and environmental sciences. Under the umbrella of the ICSU World Data System and financed by the country in which each facility is located, 52 data centres in 12 countries are currently operated. These centres serve to make available scientific data on time scales ranging from seconds to millennia, particularly with regard to changes to the geo- and biospheres.

This is a development that all disciplines must address. All disciplines face the challenge of collecting, storing, and providing data that is of general importance to the respective area in such a way that it may also be used in other research projects. At first glance, it may appear to be a simple task. Why – one could ask – are not existing institutions, libraries for example, charged with establishing such a structure? And precisely here lies the danger and the challenge. Creating infrastructure is not enough. It is not the technical aspects of the task that are problematic. Rather, the problems lie in the acceptance of such new structures by those who are to use them, the researchers themselves.

National infrastructure for data storage can only be successful if it is accepted by the members of the respective scientific subject area. And it will only be accepted if the basic standards are developed from within the subject area in the first place:

- These are standards that must be adhered to, beginning with data collection. To ensure that it can be used by others, an instruction note of sorts needs to be included with the data right from the start. You experience the different ways in which temperature, for example, can be interpreted every day in the weather forecast, where both the measured temperature and the wind-chill are reported. The measured value is not scientifically usable until the measurement conditions that influence the value, the devices that were used, and the calibration methods for those devices are linked to the value.
- The respective disciplines must also define standards for providing data with meta-information in such a way that this information can again be identified in automated processes.
- Further standards relate to aspects of intellectual property and data privacy.
- And, of course, all of these standards must be designed to accommodate rapid progress in information technology over the long term.

In order for such essential data infrastructures to succeed, it is critical that the scientific disciplines themselves structure the process and define the rules and standards which a given discipline views as necessary and reasonable. Experts from the area of information infrastructure can accompany and support this process, but they cannot replace the opinion-forming process within science and the humanities. In addition to the ICSU World Data System mentioned above, the European Social Survey also serves as an example. The longitudinal studies of the social sciences are particularly dependent on a sustainable and supra-regional data infrastructure. The European Social Survey project has been successful in establishing a system on a broad European basis that makes available in aggregate form social science data from more than 30 countries.

Using primary research data as an example, it is clear how important contributions from science and the humanities are in establishing infrastructure which, in turn, stimulates new scientific queries.

(d) A fourth example, illustrating the merging of information infrastructures and scientific research, is in the area of bibliometrics. This is a field that is becoming increasingly independent and which has a growing influence on science, and has to be seen with great concern. Oftentimes the summation of impact factors replaces the discussion on the content of a publication. Furthermore, as a result of the long-stated dissatisfaction with the impact factor, other, new processes and standards will be developed – be it the Hirsch index, which has been in use for some time, or processes based on the electronic measurement of requests, either alone or in some combination thereof. The point of such processes is apparently not to assess research performance with respect to content and quality, but instead to do so by means of a pseudo-objective index. In a 2008 *Genome Biology* contribution, Gregory Petsko ironically demonstrated just how absurd it is to reduce scientific (life) work to an index in a story in which Saint Peter requires an impact factor greater than ten for entry into heaven [2].

Moreover, given the strong fixation on the impact factor attempts are being made to apply processes developed for specific disciplines to areas for which these methods simply are not suited. The subject areas are so diverse that they cannot be measured using the same evaluation criteria. This becomes a problem as soon as impact factors become a parameter that influences how research funding is, for example, allocated to universities.

It is of course legitimate for university administrators to be concerned about knowing how the individual department in their universities compare within their disciplines on a national and international level. But this can only mean that, for example, Environmental studies at one university are compared with other departments from the same area. It cannot mean that Environmental studies should be compared with production engineering or biochemistry at the same university. That is as obvious as it is straightforward. And, nevertheless, a great deal of damage is being and can be done here. There are

already instruments that are leading universities and state ministries to believe that they can measure the research performance of their departments by evaluating the publication output database.

But the current obsession with bibliometrics has another effect which, from the perspective of science must also be alarming. The strong fixation on any type of bibliometric factor reduces flexibility in publishing. The position of journals with high bibliometric factors becomes firmly entrenched. Researchers who know that their future resources are dependent on how well their publications perform in bibliometric processes may opt to submit an article to a journal not based primarily on whether the journal reaches the scientific community that one wishes to reach. They may choose the journal where they can reach the highest scores. It becomes almost preposterous when it is more important to publish reviews than original research papers. As a result, the scientific content becomes less relevant while the packaging becomes more important. Certainly, this is not in the best interest of science.

The scientific subject areas cannot and must not leave the matter of evaluating scientific performance up to automated processes, regardless of how sophisticated they may be. Bibliometrics does not replace the intellectual debate with scientific knowledge.

Here, too, all scientific disciplines are again asked to develop evaluation criteria by which the subject areas are willing to be measured and can be measured. This means appropriately taking into consideration the entire spectrum of publication forms available in a given subject area if indicators for the evaluation of research performance are to be defined and determined. The European Educational Research Quality Indicators project, funded within the 7th Framework Programme by the EU Commission, is, therefore, justified in including not only journal articles, but also books and contributions to edited volumes, which are considerably more important for the educational sciences [3]. At the same time, the necessity to take into consideration all publication forms when developing evaluative bibliometric processes shows that such processes will not be ready tomorrow. This was made apparent to us by the report "Towards a Bibliometric Database for the Social Sciences and the Humanities", commissioned by the ESF, the DFG and other European funding organisations [1]. Among other things, this report points out the need to also take non-English literature into consideration in the databases that are used for bibliometric purposes and to define criteria for this purpose that state to what extent publications that are clearly national in focus may be considered to be of high quality.

By way of these four examples – digital humanities, open peer-review processes, primary research data and leaning towards bibliometrics – one can clearly see how areas of scientific activity merge with information infrastructure.

What conclusions must we, as science organisations, draw from what has been said so far? What are the consequences of the fact that the digital revolution does not only affect the basic conditions of scientific research, but also – to a certain extent – scientific research itself?

As science organisations, which we consider to be trustees of the interests of science and research, we feel that we have a responsibility to play a greater role than in the past in helping to shape information infrastructure, conveying the interests of individual disciplines as well. In other words: to clearly articulate the interests of science and to represent those interests.

For the future composition of information infrastructures this means that the previously accepted and well-functioning division of roles and work in the production of knowledge, knowledge dissemination and knowledge preservation is reaching its limits and must be reconsidered and restructured. For a long time, the roles were clearly and precisely defined: researchers produced new knowledge; publishers ensured that this knowledge was aggregated and disseminated; and, lastly, libraries had the task of organising the knowledge and making it accessible for research. This distribution of tasks was the result of

a long differentiation process in which the professional fields of researcher–publisher–librarian became specialised and distinguished.

This division is no longer functioning well. The tasks – as I have tried to show – can no longer be clearly divided into production, dissemination and access. And this brings us back to the question: what has prompted science organisations to take such an active role in addressing the topic of information infrastructure?

The mutations that scientific research has undergone as a result of digital information technologies make it imperative that we as researchers recognise the responsibility with which we will undoubtedly be faced when it comes to organising the information infrastructures of the digital scientific era. We are doing that in fields that we defined as the six central priority areas for cooperation in the Alliance's Digital Information Initiative. The roles and tasks for science and science organisations within the Alliance's Digital Information Initiative are the follows:

- In the licensing of content, we work in the interest of science to achieve the best-possible provision of information under reasonable conditions; another important goal is to ensure that the information is not only visible, but that it is also usable for the purpose of the *digital humanities*;
- for open access publishing, we would like to establish the basic conditions that this form of publishing needs – this includes, of course, financial frameworks, but also organisational and legal frameworks;
- with the national hosting strategy the question is, above all, how we can organize and guarantee permanent, secure access to digital resources;
- with primary research data, we want to ensure that this data is collected, archived and made openly accessible for further scientific use according to accepted standards;
- virtual research environments support researchers in structuring virtual workspace and research space;
- with respect to the structure of the legal frameworks, it must be a matter of concern to us as science organisations to bring the perspectives and interests of science into the legislative process. We aim to convey the decision makers that it is a responsibility for society as a whole to make the knowledge generated by researchers accessible to the research system under reasonable conditions.

For conclusion a short interim assessment of what has been achieved so far through the Alliance Initiative and what the lasting effects can be will be given. The Alliance Initiative was started in the summer of 2008. The initial term was for four years until 2012. But even at this point in time, less than half way through the period, it is already becoming obvious that this initiative will have longer-term impact – not only through the progress and achievements in the working groups as such but also in the capacity of how the German Science Organisations are able to collaborate effectively beyond their individual institutional boundaries in identifying challenges, defining goals and strategies and implementing the new ways together. By way of this improved networking and collaborating the German Science Organisations will become a more focused and organised partner for publishers, for political and funding bodies and for any other player relevant for the development of modern information infrastructures.

## References

- [1] D. Hicks and J. Wang, Towards a bibliometric database for the social sciences and humanities, 2009, available at: [http://works.bepress.com/diana\\_hicks/18](http://works.bepress.com/diana_hicks/18).
- [2] G.A. Petsko, Having an impact (factor), *Genome Biology* 9 (2008), 107, available at: <http://genomebiology.com/2008/9/7/107>.
- [3] Project website under <http://www.eerqi.eu/>.