Building the

KNAW International Research Institute

on *e*-Science Studies in the Humanities and Social Sciences

(IRISS)

by
the Committee on a KNAW Research Institute for *e*-Science,

Prof.dr.ir. Wiebe E. Bijker (chair)
Prof. Kevin Schürer
Dr. Els Stronks
Prof. dr. Hans Uszkoreit
Ir. Peter Wittenburg
Prof. dr. Steve Woolgar
Dr. Gaspard de Jong (advisor)
Dr. Krijn van Beek (secretary)

15 September 2003

0. EXECUTIVE SUMMARY

This report analyses the current developments in the use of digital information processing and the related emergence of new forms of scientific work and new research paradigms of collaborative research — in other words, in *e*-science — in the humanities and social sciences. The report concludes that it is pertinent and timely for the KNAW to create a new research institute: the KNAW International Research Institute on *e*-Science Studies In the Humanities and Social Sciences (IRISS).

IRISS' mission should be to develop and stimulate the advanced and innovative use of ICTs in the humanities and social sciences, and to study the effects of the new ICTs on the practice, activity and quality of scientific research in the social sciences and humanities. This mission is to be pursued by an integrated programme of cooperative research between the humanities, social sciences, science studies, and computer sciences. The aim is to advance the use of ICTs in social science and humanities and at the same time to critically reflect upon their usage in the process. IRISS' mission is thus to help create a new form of humanities and social science scholarship, in addition to the existing traditions.

To realise this mission, IRISS should be built around a unique combination of substantive research in the humanities and social sciences, with the development of methodology and tools, and with critical technography (see 3.2 for an explanation) of the process of research and technology development. IRISS will thus draw on the following three constituencies:

- Humanities and social sciences (including such disciplines as history, philosophy, linguistics, literary theory, sociology, anthropology, Internet research, communication science)
- Science & technology studies (especially technography; see 3.2 below)
- Computer sciences (including information science, software development, systems engineering, library & information sciences)

These three constituencies will have equal intellectual importance in the research programming and practice of IRISS, although the actual interdisciplinary research should make the distinction between the three increasingly irrelevant and even invisible.

The basic questions of the research programme are: How can the intellectual mission of the humanities and social sciences be enriched by the use of ICT, Internet, new forms of digitised data, shared databases, formalized semantics, and collaborative and interactive research software? And what are the effects of the uptake of these new developments on the process, practices, quality and societal impact of scientific research in the social sciences and humanities?

Rather than differentiating these broad, basic questions into an elaborate research programme of smaller sub-programmes and projects, this report proposes a project-based research planning that allows for quick adaptations to new research questions and opportunities. The dynamic character of the current developments in *e*-science studies in the humanities and social sciences results in research problems being highly fluid and

priorities swiftly changing. A static research programme would not be able to adequately adapt to such changing circumstances.

IRISS should minimally have a critical mass of 25 researchers and 15 graduate students, and an annual budget of 4.3 M Euro, of which 2.3 M Euro structurally funded by the KNAW. An internationally recognized expert of high standing should be searched as Director, and this Director then should play a central role in the final design and implementation of IRISS, including the selection of staff.

In the process of drafting this report, the committee organised an expert meeting with 20 international guests. Reacting to an earlier draft of this report, these experts gave extremely valuable feedback. Partly, this feedback came in the form of concrete experiences and advice on how to carry out *e*-science studies in the humanities and social sciences, and how to create a research institute for that. Many of these concrete ideas have found their way into this report. Most importantly however, these international experts generally expressed support for the very idea of building IRISS and for building it around the integrated combination of humanities and social sciences, science and technology studies, and computer sciences.

Finally, in response to a specific question of the KNAW Board, the committee has concluded that much of the current research at NIWI (the KNAW Netherlands Institute for Scientific Information Services) would be relevant and appropriate within the proposed IRISS institute.

1. PROLOGUE[1]

In 1972, during his Turing Award acceptance speech, Edsgar W. Dijkstra declared that in their capacity as a tool, computers 'will be but a ripple on the surface of our culture', but as intellectual challenge 'without precedent in the cultural history of mankind'.[2] The central research imperative for IRISS is to see that the implications and consequences of Dijkstra's challenge are explored for the humanities and social sciences and that the resultant methodological wisdom is implemented.

This is not just about skills and tools — it addresses the intellectual mission of the humanities and social sciences. The promotional rhetoric of computing tells us that ICT provides solutions to problems. But research problems are not fulfilled by solutions: rather by giving rise to better, more difficult problems. The late classicist Don Fowler praised the great Eduard Norden 'for his championing of the notion that a good commentary does not solve problems but makes them worse'.[3] So does also any application of computing worthy of the humanities and social sciences. The *intellectual* challenge of which Dijkstra spoke is not to devise solutions to old problems, it is to problematise the apparently unproblematic.

What does this actually mean and entail for the humanities and social sciences? That is the central question with which IRISS will be engaged. The answer (or rather, the programme for better questions) comes out of doing *e*-science research and observing one's own and others' work from a methodological perspective.

This requires protected space in which to be pursued. Thus the social-intellectual role of a research institute for *e*-science in the humanities and social sciences. The best way to institutionalise is to provide an environment to support research and an infrastructure that will assure a steady pilgrimage to and migration from the new institute of serious scholars from around the world.

[1] Adapted from the contribution by Prof. Willard McCarty to the IRISS expert meeting, 25-26 May 2003.

[2] Dijkstra, Edsgar. 1972. The Humble Programmer. The ACM Turing Award Lecture. Communications of the ACM 15:859-866: 866.

[3] Fowler, Don. 1999. Criticism as commentary and commentary as criticism in the age of electronic media. In Commentaries – Kommentare. Aporemata 4, edited by G. W. Most. Berlin: Vandenhoeck und Ruprecht: 442.

2. INTRODUCTION

In October 2002 the Strategy NIWI Committee submitted its report to the Board of the KNAW, recommending to close the present Netherlands Institute for Scientific Information Services (NIWI) and to establish a new KNAW Institute for *e*-science research in the humanities and social sciences. In its position paper of October 2002 the KNAW Board has decided to establish a committee of national and international experts to study the desirability and feasibility of such an *e*-science research institute.

This committee consisted of: Prof.dr.ir. Wiebe E. Bijker (chair; Faculty of Arts and Culture, University Maastricht), Prof. Kevin Schürer (Director of both the Economic and Social Data Service (ESDS) and the UK Data Archive (UKDA), and Professor in History at the University of Essex), Dr. Els Stronks (Faculteit der Letteren, Universiteit Utrecht), Prof. dr. Hans Uszkoreit (Computerlinguistik, Universität des Saarlandes; and Language Technology Lab, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH), Ir. Peter Wittenburg (Max Planck Institute for Psycholinguistics, Nijmegen), Prof. dr. Steve Woolgar (Saïd Business School, University of Oxford). Dr. Gaspard de Jong (Director NIWI) participated as advisor to the committee, and Dr. Krijn van Beek (Raad voor Maatschappelijke Ontwikkeling) acted as secretary.

The committee's charge consisted of two phases.[4] This report concludes the first phase. In the first phase, the committee was to investigate the need and possibility for a KNAW *e*-science institute for the humanities and social sciences. For the second phase, which would only begin after a decision by the Board of the KNAW to create such an institute, the committee would be changed into a search committee to advise the Board of the KNAW about a Director. The committee met three times.

An important part in the process of this first phase was the organisation of an expert meeting in May.[5] International experts from all relevant scientific domains participated with a great amount of enthusiasm. This was indeed one of the most important findings from the expert meeting: a very significant commitment to the general idea of an *e*-science studies institute for the social science and humanities — almost all experts whom we approached, accepted the invitation; all participants sent the preparatory notes that had been asked for; the meeting started one minute after schedule, although on a bright Sunday afternoon in the heart of Amsterdam; and almost all participants sent additional comments and ideas after the meeting. This does not mean that there was an easy consensus — on the contrary. But our conclusion was that the engagement with which these debates were held should be taken as an indication of the pertinent, timely and intellectually challenging nature of the new institute. Also in more concrete forms — theoretical concepts, research projects, management advise — these experts have contributed greatly to the final result as presented in this report. When the research institute will be created eventually, it will be important to revitalize the relationship with this international group of researchers — among them, the institute already has some form of embryonic existence.

[4] See appendix 6.1 for more details.
[5] See appendix 6.3 for a list of participants.

During the process, the researchers at NIWI were periodically informed of progress and had a substantive input in the committee's discussions about the research programme. The committee's chair had two half-day workshops with the NIWI researchers, to discuss the various NIWI research activities.

The committee unanimously concludes that the establishment of a new International Research Institute on *e*-Science Studies in the Humanities and Social Sciences (IRISS) is timely and pertinent, and that the KNAW is a most suitable candidate for creating such institute. We will argue this conclusion by answering the following questions in the remainder of this report:

1. What is the relevant research area and what are the key scientific challenges? (section 3)
2. What could be the outline of a research programme to shape this research area and to contribute to answering these problems? (section 4)
3. What organisational requirements should inform the building of such a research institute — how would a blueprint for the institute look like? (section 5)

3. RESEARCH AREA AND KEY SCIENTIFIC CHALLENGES

3.1 e-Science in the humanities and social sciences

The introduction of personal computers, interconnectivity, networking, and Internet (briefly: information and communication technologies, ICT) in the humanities and social sciences research is having, and will continue to have, implications that range much farther than simply increasing the efficiency of scholarly work through, for example, the use of word processors and the use of digital libraries. These implications may have far more fundamental repercussions for the humanities and social sciences than they are having for the natural and technical sciences. For the latter, the coming of *e*-science probably meant primarily an increase in efficiency and effectivity, since the character of relevant scientific data did not change radically until now. In the case of the humanities and social sciences, a completely new style of scholarly work appears at the horizon: *e*-science in the humanities and social sciences can and should be more than a 'scholarly GRID.'

The concept of *e*-science comprises many forms of utilizing advanced networked information technology for boosting scientific progress. Just as with *e*-commerce, *e*-learning and *e*-government, the emphasis lies on exploiting the Internet for new levels of exchange and collaboration. Both as a concept and as a practiced form of collaborative research, *e*-science emerged in the largest scientific disciplines such as physics, biology and chemistry and then rapidly spread within the natural and engineering sciences.

In a narrow sense, the term '*e*-science' refers to GRID-computing, the powerful paradigm of scientific computing that turns the global network of participating computers into one single resource of data, programs and computing power. The proponents of this development view the GRID as the successor of the WWW. In their vision the GRID will be at least as powerful and influential for science and technology as the WWW has become for commerce and public information. Another ingredient of *e*-science is high performance visualization of complex data and computed results.

We view the term '*e*-science' in a wider sense. As we can clearly see that the globalised infrastructure of information technology will create unprecedented opportunities for all academic disciplines and enable new forms of collective research, we still cannot safely predict the particular developments in this mega trend beyond a time frame of a few years. *e*-Science shall therefore refer here to the multitude of new scientific methods and forms of collaborative research enabled by the globally networked information technology.

One important trend is the use of large data repositories. Currently, many digitisation projects, covering all sorts of data types and professional content, will lead to gigantic repositories of scholarly relevant data for all existing disciplines in the humanities and social sciences. Increasingly better methods to describe and characterize the objects in these repositories are developed with the help of basic descriptive metadata. In convergence with this growth of repositories, ICT engineers continue to develop and improve tools and methods to create and manipulate metadata in the broadest sense.

These tools operate at the semantic level and often require a shared structural framework of concepts.[6] The interoperability of computers on various levels, including the semantic level, is the key point. It is not the GRID itself, as a set of collaborating computers or services, that poses the real challenge, but the 'semantic GRID' (see table 1, below).

These developments, however, are still in an infant state and will need much fine-tuning to become useful for the humanities and social sciences. During this process researchers will try to make their basic knowledge explicit: they have to define the key terms of their research in a multilingual and even context-dependent form. Only in that way, this meta-knowledge can crystallise into formal frameworks that may help users to navigate complex spaces for identifying relevant knowledge. This explication work will be highly critical, since only little knowledge can be formalized, and formalization will always mean a severe simplification of the rich underlying data.

The availability of massive amounts of invaluable primary research data in 'virtual space' (i.e. digitally available from computer networks) will make it profitable for scholars from different backgrounds to collaborate in new ways and to develop new integrated methodologies. Even the notion of 'researcher' may change, since non-academics increasingly will be involved, for example by contributing to the process of data creation. Much more cross-disciplinary work can be expected. Also notions such reliability, replication, and even authorship may assume new meanings.

Not many researchers in the humanities and social sciences yet have a vision of these trends, their impacts, and the associated illusions, risks, and limitations. Consequently, there is no clear view among humanities and social science scholars as to where research should be heading. Technologists, on the other hand, merely see technical challenges and capabilities, and thus tend to move to an abstract level where they ignore important details of the research field and oversimplify the translation of a neat idea into a useable application.

We need to zoom in, to distinguish in more detail the kinds of work that are implied in the sketched trends of digitisation and networking. Table 1 gives an overview of a variety of research activities, ranging from the individual use of improved computing facilities in the top left corner to the collective development of new knowledge through the use of networks and cooperative architecture in the right bottom. Our committee's diagnosis is that the dominant trend in *e*-science in the humanities and social sciences is a move towards these latter styles of shared and cooperative styles of research. It is important to recognize that this does not make the individual scholar, working their way through archive boxes or reading in a library, obsolete; but it will change the scientific community to which they relate, and the employed criteria for adequacy, reliability, precision, and validity.

[6] Such a shared semantic framework or vocabulary of concepts and definitions is often called an 'ontology.'

	Computing	Data	Information	Scientific Content	Knowledge
Utilization	Scientific computing, Word processing	data bases	internal project WS	electronic authoring	individual digital memories (LifeBits, Haystack)
Exchange and Communication	exchange of computing resources (SETI)	individual data exchange	newsletters	discussion groups	exchange based on shared metadata standards and interlinking
Sharing	GRID computing	e.g. shared corpora	information services	e-journals e-print archives	semantic web
Collaboration	web-based applications	work on same data	distributed information services	joint authoring	Semantic GRID

***Table 1**: Trend in the use of information and communication technologies in the humanities and social sciences*

The diagnosed trend is pushed by efforts to combine recent developments in information science with the classic knowledge interests of the humanities and social sciences. This promises to create a potential for:

- Interoperability between disciplines;
- Operations that include resources from different domains;
- Flexible interpretation of texts from various sources;
- Creation of relational links by using natural language processing;
- Flexible inclusion of media (audio, video, images, 3d-objects); and
- Cooperative annotation of material.

These are possibilities that are especially important for the humanities and social sciences and that seem to surpass the challenges currently encountered by *e*-science developments in the natural sciences.

It will be necessary to develop new forms of 'middleware' — computer programs that serve to combine, or mediate between, two or more separate programs. Such middleware is a key element in any *e*-science, since it allows disparate and different systems to be joined into a common framework. The structure of data in the humanities and social sciences is intrinsically complex, unlike that of, for example, experimental physics where definitions are much stricter. Text and media form the natural domains of the humanities and social sciences, and knowledge technologies have now so far evolved that it is increasingly possible to handle even the complex, multi-interpretable, 'fuzzy' contents of the humanities and social sciences.

However, the tools for such advanced research cannot be obtained 'from the shelf.' In this respect the humanities and social sciences are no exception to other sciences: methodologies can only be developed by or in close cooperation with researchers themselves. Hence the necessity to combine methodology development, including software engineering and information science, with substantive humanities and social sciences research.

One of the challenges will be to tackle the multi-interpretable and fuzzy quality of data in the humanities and social sciences. The option, generally used in other domains (such as in the Semantic Web endeavour), to explicate meanings of concepts into ontologies as networks of well-defined words, does create the risk of creating too static an image of that particular knowledge domain. How can ontologies be made useful when the granularity of formal semantic descriptions is poor and static compared to what hermeneutic researchers are used to? How can we navigate in a knowledge domain where all contribute their own knowledge base? How can we construct inference engines that do not produce vacuous conclusions?

Taking stock of what the key challenges are in this research area of *e*-science in the humanities and social sciences, we can list:

- To inform researchers in the humanities and social sciences, and in knowledge engineering and information technology, about the needs, possibilities and limitations of ICT in humanities and social sciences research;
- To stimulate interaction between researchers and technologists;
- To develop optimal infrastructures to carry out this new type of research;
- To develop new ways of communication and interaction for researchers;
- To create a rich digital world of raw and processed data;
- To better understand the scientific requirements posed by the formalization of discipline semantics and their usage;
- To define useful standards and organize relevant discussions in the scientific community.

This requires close collaboration between humanities and social science researchers on the one hand, and computers scientists on the other. But that is not enough. A simple combination of 'scholars' and 'engineers', to label them briefly, would underestimate the fundamental challenges that are at stake. Such a combination would build on a simplistic view of 'scholars' using 'tools' developed by 'engineers.' The challenges for *e*-science in the humanities and social sciences are more fundamental, and require a more radical and built-in questioning of the role of science and technology in scholarly work.

Ideas about science and technology are at the heart of humanities and social science research in two main ways. Firstly, an understanding of science and technology is fundamental to questions about how knowledge is generated and distributed, what is the nature and source of expertise, and what are the social and organisational effects of the introduction of new technologies. Secondly, assumptions about the nature of science and technology underpin some of our deepest assumptions about the nature of research itself. What models of 'science' do we draw upon in attempting to persuade one

another? How do technologies of communication and representation inform the ways we organise ourselves?

To incorporate this monitoring and questioning of the underpinnings of the institute's projects, a third 'resource' is needed in addition to the computer science and the humanities and social sciences: science studies. The specific role for science studies within the proposed institute we will call 'technography.'

3.2 Technography for *e*-science in the humanities and social sciences

Science Studies is an important and often controversial interdisciplinary field comprising social studies of science and technology in all their aspects. It is provocative in challenging accepted views about science and technology, and in thereby raising profound questions about the very basis for social science and for scholarship more widely. Its analytic perspectives include relativism and social constructivism, social shaping, actor network theory and reflexivity. It covers substantive areas such as the genesis and reception of scientific knowledge and the sociology of technology, especially the social dynamics of information and communication (including Internet) technologies. This latter includes research into the likely effects of the new technologies on questions of identity, trust, privacy, representation and social organisation.

In all these guises Science Studies has enjoyed considerable success both within its own terms and, significantly, in influencing areas of scholarship well beyond science and technology. Of particular note, however, is a recent trend. Instead of merely reflecting upon and analysing scientific and technical developments, science studies scholars have increasingly begun working in collaboration with scientists and technologists. Information and analysis about the social and organisational dimensions of, for example, software development, has been fed back into the development process. Suggestions about the implicit assumptions which inform the design process are highlighted and discussed with the software developers. This development marks a significant contrast with early science studies which were, on the whole, simply reflective and analytic about the processes of scientific and technical development being observed. The more recent trend is of much greater involvement, with real time *in situ* feedback and collaboration between science studies researchers and the technical developers. Whereas early science studies attempted a form of detached analysis and comment, the more recent variants involve collaboration. To use Clifford Geertz' formulation of the distinction, this is much more a form of ethnographic study *in* the tribe rather than merely a study *of* the tribe.

More exactly, this new style of science studies research can be termed 'technography' (rather than ethnography) – meaning the intensive, reflexive, *in situ* and (most importantly) collaborative study of the genesis and development of new technical artefacts. This kind of research relation has proved attractive to various well-known high tech companies. For example, Xerox PARC and Intel have employed resident ethnographers-anthropologists (the term 'technographer' did not exist at the time). But we have yet to see its development in situations that are building information systems in the service of humanities, arts and social science research. Here is a unique opportunity to make a difference: to put in place a novel form of collaboration between science

studies and arts and humanities developments around the new technologies and, in the process, to advance scholarship in both science studies and the other areas.

A key feature of much humanities and social science research that relies upon technological systems, is that the claimed abilities and impacts of these technologies remain un-interrogated. For this reason it is important to include elements of research which retain a positively sceptical approach about the sometimes-exaggerated descriptions of technical capacity. This positive contribution can also be expected from the form of technography that is proposed here. Technography will guard against an uncritical belief that the increased adoption of ICT, the sharing of data, the mixing of methods, can do only well. Instead, a constant focus is needed on the 'essential tension' that exists around the introduction of *e*-science in the humanities and social sciences. The widespread adoption of new technologies can bring significant advantages. But historical examples also suggest that the implications, consequences and side effects are not always obviously positive, which provides extra reason for technographic feedback on the use and non-use of new technologies.[7]

One example of the possible application of a critical technographic perspective is provided by the idea of replication. It is frequently assumed that the adoption of new technologies such as the GRID and *e*-Science technologies, will require standardisation and then encourage replication. It will be part of the research agenda of the proposed institute to investigate the extent to which this is indeed the case, especially in the face of much existing historical and sociological evidence that (natural) scientists in practice go through a complex series of negotiation and interpretation processes such that they define what in each case might count as replication.

One especially interesting question is the extent to which the availability of new technologies of communication and interaction will radically affect existing working practices between social scientists and humanities scholars. Enhancements in the sheer speed of connection and bandwidth capacity are thought to be the possible impetus for greater sharing of data. In many social sciences, however, professional practices and identities are built around notions concerning the ownership of data. In anthropology, for example, reputation and identity are frequently built around ownership of and exclusive expertise in relation to a particular tribe. Also in the humanities the study of *unica* is important in much scholarly work. To what extent can new technologies modify currently entrenched disciplinary identities and structures of expertise?

3.3 Scientific challenges

Although there are some very innovative projects in the humanities using ICT, the overall picture is that ICT-use in the humanities is less developed than in most other disciplines. Many scholars in the humanities see ICT as a practical tool with limited scope and research relevance. Most researchers only use ICT in word processing,

[7] For example, the discovery of pulsars as an astrophysical phenomenon was made by the one radio astronomy group in the world, that did not adopt the new systems of automating recordings and their analysis. One implication (which enjoys almost no popularity in science policy circles!) is that new technologies can lead science to miss discoveries; or, conversely, that perhaps the more innovative research comes from groups which are least well equipped with the latest technology.

emailing, and Internet browsing. There is, consequently, not much insight into the effects of ICT on the research process in the humanities, nor into the effects on public access to and use of the knowledge thus produced. Similarly, the use of ICT is typically not seen as implying fundamental methodological issues.

Parts of the social sciences present a different picture. In economics for example, the creation, analysis and sharing of large databases is central to much research. Here, although it may not yet be possible to see many typical *e*-science research efforts – i.e. large groups of researchers cooperating via intelligent networks; though the creation of for instance the World Development Report might be seen as an *e*-science effort in itself - the evolution of economics as a discipline seems to naturally point towards using the possibilities of *e*-science rather earlier than later. A more or less similar argument can be given for demographers, social geographers, sociologists, health care scientists, etc. All these disciplines are increasingly making use of the new possibilities of ICT's (see below for examples).

This difference between the situations in the humanities and in the social sciences may suggest that it is only necessary to focus attention on *e*-science in the humanities. The committee, however, has concluded that it is important to focus on *e*-science in *the combination of* humanities and social sciences. There are two reasons for this. The first relates to the difference between social sciences and humanities, and the second regards the character of current *e*-science in the social sciences and humanities.

How should a line be drawn between humanities and social sciences; what should be included in, what excluded from an *e*-science stimulation endeavour? The simple distinction humanities—social sciences does not work: some humanities scholars do use advanced forms of ICT and shared databases; and many social scientists working in a qualitative or hermeneutic style do not use more ICT than for word processing. But more importantly than trying to establish the correct dividing line between users and non-users of ICT, is to understand what happens to these distinctions and identities when ICT is introduced: how does the uptake of ICT and the creation of shared databases change the character of scholarly practices and (disciplinary) identities? Thus, this committee proposes that a new institute will not be built on essentialist definitions of *e*-science, disciplines, and ICT-use; but that the new research institute should study the dynamic changes in these definitions and the identities.[8]

[8] One example is philology — a field in the humanities where the use of ICT methods may be expected to have far-reaching implications. The introduction of digital methods and tools will allow philology to acquire again the broad and challenging scope it had in the 19th century, of encompassing in an integrated way both language and literature studies— two fields that have become quite separated during the last century. Until recently, philology often implied detailed quantitative studies of words, images, and motives of an individual literary author. Consequently, it often 'fell between ship and shore': not belonging to mainstream language studies, where the focus is on language as a system; nor belonging to literary studies, since this philology was too much focused on language details and too little concerned with interpreting the literary content. The introduction of digital databases and methods may allow for more encompassing and sophisticated analyses — such, that the resulting increase in quantity leads to a quality jump: creating a new and fruitful combination of interpretative literary and quantitative language analysis.

Secondly, also in most projects that currently use ICT in an intensive way, the potentiality for interdisciplinary collaboration is not exploited to the full. Recent pioneering developments in the humanities and social sciences have shown that cooperative efforts under the heading of *e*-science do provide huge possibilities for doing interdisciplinary research. New questions are being asked, new ways of answering old questions are being developed, and innovative forms of cooperation across disciplinary borders are pursued. This seems to happen in all realms of the humanities and social sciences. At the same time this interdisciplinarity creates great challenges. These regard the technical aspects of large-scale cooperation, for example, interoperability, development of knowledge systems, and so on; these regard methodological questions and questions about the ontological underpinnings of different disciplines.

The best way to convey the key ideas, the flavour of the challenges, and the new potential of this type of research, is probably to give a few examples. Only concrete examples will exemplify the variety of ways in which new combinations of technical, methodological, and substantive issues may yield innovative projects. We will conclude this section with a few examples of projects that exemplify the new scientific challenges of *e*-science studies in the social sciences and humanities. More examples will follow in section 4.4.

Example: new use of corpora in literary studies

Traditionally, literary studies have a comparative and/or interpretative aim, focusing for instance on intertextuality, or on the influence of social developments on literary distribution and production, or on the appreciation of literary styles and genres within certain periods of time. Much of this traditional research will benefit from the creation of large, digitised literary corpora. The database structure of these digitised corpora — consisting of the original texts supplemented with mark-up in which knowledge of the structure, content and meaning of these texts is secured — allows researchers to search more fully and systematically, and enables them to verify impressions and results. Apart from these benefits, these corpora enable the research of new questions, since they permit scholars to look for detailed elements and aspects in a large corpus of texts.

Example: new data by combining different data banks

Ethnologists and anthropologists have gathered large repositories of artefacts and descriptive material about many cultures worldwide. Often these repositories have been digitised and stored according to some standard. In parallel, more and more language data from these cultures are also collected and stored according to another standard. There is much to be gained by combining these repositories into one virtual domain. Often artefacts, such as ritual masks, are just objects without much further clarification. Language resources could contain the myths that describe the purpose of these masks. To make such a combination of databases with a common search and access platform possible, interoperability of the two different standards that shaped the databases needs to be realised.

Example: cooperation among different disciplinary approaches

There is much debate about the evolution of languages and the factors that influence the processes of language change. Ethnobiologists, geologists and sociologists have gathered much material about different but related phenomena. Integration of these data to establish multi-layered maps over time

would allow new analyses of language modifications and shifts. The interoperability between these areas is still problematic however.

Example: new forms of scientific cooperation by shared access to sources
The genre of the Renaissance emblem was widespread throughout early modern Europe. It may be conceived as a web of citations, imitations and adaptations. Elements from classical, medieval and contemporary literature, painting, sculpture and music were used to compose emblem books on many different subjects and in many different styles.
The love emblem's uniquely Dutch character and the special position of the Dutch love emblem in Europe make this corpus attractive for a great number of scholars. A database with digitised love emblem books, based on a systematic and thorough encoding of the material, would be a great source also for non-Dutch researchers. Emblem studies mostly have a comparative aim, trying to find the origins and sources of pictorial and textual motifs and ideas. Before the digital era, emblem research was undertaken with scholarly reprints as its most important aid. This project changes that situation by offering a wide collection of flexible digital editions and related benefits (full texts search, but also indexes of motifs, sources and parallels).[9]

Example: more effective answering of old questions
Researchers at the KNAW's Meertens Institute successfully used a large database when studying the principles of contrafacta (new texts to extant melodies) in 16th and 17th-century Dutch songs. The relationship between text and music in contrafacta was investigated by the development of a strophic repertorium ('voetenbank') in which the metric schemes of songs were analysed. This allowed the identification of basic patterns of borrowing (of melodies, texts and literary aspects such as motifs).

These examples demonstrate that a move towards *e*-science in the humanities and social sciences has much more radical implications that a mere increase in efficiency or productivity. Old questions may be answered more quickly, but — more importantly — they also can be answered in totally new ways. These new ways may result from analysing date with new tools; from using larger databases and even from using hitherto unavailable data; and they may result from collaborations across disciplines that would be very difficult to accomplish outside an *e*-science infrastructure and research culture. The scientific challenge thus is, to realize the full potential of *e*-science in the humanities and social sciences by creating a truly new style of scholarly research. The next section will discuss how this ambitious and broadly formulated challenge can be translated into a research programme.

[9] The Emblem Project Utrecht intends to digitise twenty-five Dutch love emblem books, religious as well as profane. See http://www.let.uu.nl/emblems.

4. OUTLINE FOR A RESEARCH PROGRAMME

4.1 Central goal

The central goal for a KNAW research institute in *e*-science studies in humanities and social sciences can only be ambitious: to deliver top quality research in the social sciences and humanities, to apply and develop advanced ICT tools, to further related methodologies, and to understand the effects of these new developments for the research process and practices themselves. And as if this were not enough: all these aims are to be pursued in integrated rather than separate projects. In this section the outline for a research programme will be discussed.[10]

The basic questions of the research programme are relatively simple. *How can the intellectual mission of the humanities and social sciences be enriched by the use of new methods of ICT, Internet, new forms of digitised data, shared databases, formalized semantics, and collaborative and interactive research software? And what are the effects of the uptake of these new developments on the process, practices, quality and societal impact of scientific research in the social sciences and humanities?*

4.2 A project-based research planning

To breakdown the broad, basic questions into smaller sub-programmes or research projects would however be a mistake. To develop a research programme that specifies research lines for a long term in a static way, would defy the dynamic character of the current developments in *e*-science. Research problems in this field are highly fluid and priorities change swiftly. Therefore a project-based research programme that allows quick adaptations to new research questions and opportunities is advocated.[11] If a research programme like this is successful, it will have an impact on the scholarly debate and scientific agenda-setting, thus reshaping its own environment. The research programme needs to accommodate that changing environment in a bootstrapping operation by which it creates its own agenda in interaction with the scientific and societal worlds in which it operates. Such an agenda should not be defined in detail *a priori*.

More generally, a humanities and social sciences research programme of the fundamental and long-term character that is envisaged cannot be a deductively organised static programme. As Willard McCarty argues for humanities and computing: 'For the humanities, the analytical and heuristic potential of computing arises from its combination of manipulatory flexibility with a mechanically rigorous demand for consistent and explicit representation. Its principal importance for research lies in the ability of this combination to raise the question of epistemology in new and powerful forms. Research is obviously strengthened by what we come to know through it, but this

[10] One caveat in sketching this research agenda needs to be mentioned. The final shaping of the institute will be the responsibility of the new Director, who will have considerable leeway in choosing the most appropriate research and organisation models. This section is therefore primarily meant at yet another way of explicating the innovative and challenging character of the proposed research, and not as a recipe for running the institute.

[11] In chapter 5 we will also argue for a specific organisational structure of the institute and profile of its staff to enable such a flexible research programming.

knowledge, rigorously constrained by the nature of computing, also allows us to question by other means *how we know what we know*. The potential of computing for research, then, lies in the realization that the computer is more than a problem-solving machine or device for 'knowledge engineering'. For the humanities at least the success of a computing application derives its significance from the questions it raises. These may come from the residue of anomalies it illuminates by its inability to comprehend them – exceptions begging for a new rule. They may also follow simply from the expanding of horizons that computing enables.'[12]

A project-wise research planning is also necessary for other reasons. One important reason is to counter the forces of mono-disciplinarity, found in much current university scholarship. Innovative research in *e*-science studies in the humanities and social sciences needs to be interdisciplinary, but at the same time should maintain a relationship to the 'home' disciplines. There are always strong incentives to work within the boundaries of well-established disciplines — the easiest way forward is to stick to your own tribe. This mechanism can be countered by a project-wise organisation of the research. Other reasons for a project-based research planning are the usual advantages of project organizations: a more flexible agenda, more possibilities to invite top-researchers for fixed terms, more possibilities to hire assistant researchers on a temporary basis.

Adopting a project-based research programming does not mean that 'anything goes.' Not all projects will fit the mission of the research institute. Not all projects will meet the special requirement that were formulated in the previous section: addressing the problems of *e*-science studies in humanities and social sciences requires well integrated, collaborative research between the humanities and social sciences, science studies, and computer science. The implications of this requirement for the research programming will now be further explored.

To facility the discussion it is appropriate to distinguish the three 'constituencies' that feed the research programme.[13]

- Social sciences and humanities research
- Science studies research
- Computer sciences research

It is crucial that these three have no hierarchical relationship to each other. In some projects one may lead the agenda while the others offer support; in others the roles may be the other way around. Research excellence in all three domains must be the overall goal.. (This is of course also necessary to attract top quality researchers.) Researchers from all three constituencies are expected to publish in the premier international journals of their respective fields.[14] The symmetry of the three constituencies also

[12] W.McCarty, 'Field of Vision: Foundations for humanities computing', note to the committee, July 2003.

[13] Terminology is tricky here: to think of the proposed institute as consisting of three 'departments', or being built around three 'pillars' would reify the distinctions between these three intellectual and disciplinary 'starting points.' Because that is what they ideally only should be: starting points from where the thinking takes off, to be left behind when true collaboration develops.

[14] Taking the computer science constituency serious implies that also published or distributed prototypes of software packages or interfaces will count as scientific products.

implies that in addition to scholarly publications (the typical output of social science and humanities research), new ICT tools for science studies and technography will be developed, with demonstration applications in the humanities or social sciences being designed to test new ICT prototypes (an acceptable output of computer science research) and new enhanced research resources will be generated.

The committee has concluded that much of the current research at NIWI (the KNAW Netherlands Institute for Scientific Information Services) would be relevant, appropriate, and well fitting within the proposed research programme. The NIWI research programme has recently been shaping up, and still is quite flexible. Elements of the humanities and social science research, as well of science and technology studies are present in NIWI; computer science is less well represented. The challenging integration of the three constituencies that this report calls for is thus not part of NIWI practice.

4.3 Build on constructive tensions

The research programme that is conceived for IRISS is meant to stimulate, shape and respond to the radical changes in humanities and social science research that were sketched out in the previous section. Some of these changes will challenge long-established assumptions and distinctions. Thus, for example, is it quite likely that shifts will occur in the boundary between social sciences and humanities, the divide between qualitative and quantitative studies, the role of *unica* in the humanities, the difference between individual scholarly work and collaborative research, and in the identity of researchers, scientific communities, and indeed disciplines. Thus, new tensions will arise between such poles.

Rather than trying to define a research programme by choosing a position in all of these distinctions, the committee would argue strongly for viewing these tensions more constructively, as issues of research and reflection. Rather than choosing to focus the programme on either humanities or social sciences, it seems more appropriate to *do e*-science in these fields and then *study* how traditional criteria and distinctions take on new meanings.

The radical implication of this strategy is that such a research programme should not be interpreted as 'serving many masters', as 'not being able to choose.' Rather, the key element in the research programme is to study tensions that either already existeed but are increased by the introduction of ICTs, or that result as new effects from the development of *e*-science. The discussions during the May expert meeting fuelled this idea: what initially, in the beginning of the meeting, appeared as a conflict, often could later turn into an interesting issue that might indeed be effected by the emergence of *e*-science. Examples are the role of tools, the distinction between the 'hard' and 'soft' sciences, the tension between 'practical' and 'intellectual' aspects, the role of ontologies and standardisation, the importance of sharing databases, and the distinction between science and meta-science.

4.4 Examples of research projects

To conclude this section, some projects that could fit the broad research aims of this programme will be sketched out. In selecting these illustrative projects the committee benefited from the contributions by the participants in the May expert meeting and from the input of senior researchers within NIWI. These examples are not principally different from the ones given in section 3.3, but where the previous set was used to illustrate the general scholarly challenge with a rather broad brush, the next set of examples is elaborated in more detail so as to illustrate concrete forms of possible research.

4.4.1 Metadata infrastructures for social science survey data

The goal of this project is to develop standards and practices for metadata usage in the social sciences in order to facility greater interchange between researchers and greater interoperability between data resources. What kinds of metadata are used by social researchers when re-using data? What difference does the (non-) use of metadata make, and what is a minimum specification of relevant metadata that researchers need to take into account? What standards for social survey metadata follow from this? And what are the requirements for a system for social survey metadata and metadata management, which enables data-creators to routinely produce the required metadata? What ICT support can be given for the integration of related datasets? Does a system of metadata help to understand the value and meaning of integrated datasets? How can researchers be supported to generate (semi-)automatically metadata on surveys? These questions are relevant for both qualitative and quantitative research data. Since the problems are much larger for qualitative research, a first step would be to solve them for quantitative work.

The survey is the most frequently used data gathering method in quantitative social science research. Much empirical social research is based on primary data — data gathered within the context of the same research project. In these cases, the researcher knows very well the ins and outs of the data: details about the questionnaire (wording of questions and answer categories) and their coding; details about the sample, the population universe and non-response; details about the way the respondents were interviewed or completed the questionnaire; the theoretical background behind the questionnaire; the moment of doing the survey, and important events that may have influenced the answers respondents give; and so on. Often this information is not recorded in a way that is different from other projects. To make the data usable for other researchers, and certainly understandable to middleware within a GRID, these metadata (data about the data) would need to be captured in a comprehensive and standardized format.

The availability and the (potential) use of secondary data — data gathered within other projects — is increasing: for comparative studies in various countries, for example. For many important research questions different survey studies are available, covering different regions and different periods. These surveys differ, of course, in questions and answer categories, as well in the theoretical background. However, some are similar enough to try to combine them for cross-national (or cross-regional) and longitudinal comparison. For such a comparative and combinatory project researchers need tools to combine and analyse the datasets. These tools would build on metadata systems allowing an integration of the various surveys into one large database by homogenizing variables and answer categories. This homogenizing is at some cost, as

variation is removed from the data. This removed variation is also included in the metadata.

4.4.2 Co-constructing the GRID

The goal of this project is to identify the problems that can be expected in large-scale implementation of a GRID network (distributed computational resources) in the social sciences and humanities; to analyse the hidden assumptions in present thinking about the GRID that may hinder its adoption across the spectrum of sciences and humanities; and to develop ICT technologies that may help to solve these problems.

The GRID is seen as an important candidate for the succession of the present Internet. Its promise is the large-scale availability of computational resources for scientific and scholarly research. Computational resources should be seen in the broad sense: they may comprise computational power, simulations, shared databases, virtual access to scientific instrumentation, and complex collaboration environments. The proponents of the GRID wish to enable these new research environments for researchers in all fields. However, the present GRID paradigm is not capable of making this happen. The focus is on a subset of physics, computer science, and life sciences. It can therefore be expected that this GRID will only be relevant for a small subset of scientific research, unless the construction of the GRID is opened for other paradigms of thinking about technology development. Instead of only relying on small pilot projects created with GRID enthusiasts, and on the paradigm of systems building in high-energy physics and related computer sciences, a more inclusive and complex approach is needed. This can only be achieved by combining tool building with ethnographic research of the making of the GRID on the one hand, and with an immanent critique of the hidden assumptions in the GRID engineering paradigm on the other hand.

An ethnographic analysis of a GRID project in the making will form the core of this project. Interviews with users involved in the project are part of that. A comparative study of a GRID in a test lab situation with the GRID 'in the wild' will yield additional insights that can only be won through close collaboration between computer scientists, technographers, and scholars. A critical conceptual analysis of all relevant policy documents, technical reports and internal communications will support the analysis. An important part of the methodology will be the provision of feedback on the basis of the ethnographic research to the other researchers constructing the GRID.

4.4.3 Determinants of online collaboration among academic researchers

The goal of this project is to build a simulation of the influence of Internet-based collaborative software on collaboration among academic researchers; to analyse which factors determine the adoption/non-adoption of these tools by academic researchers; to discover the influence of the position in the networks of research groups and of individual researchers on collaborative behaviour; to identify emergent properties that are unpredictable from the perspective of an individual research group.

Funding agencies and government policies try to stimulate international collaboration, also in fields that have traditionally been individualistic in their research styles. This makes the question relevant of how robust collaborative behaviour of academic researchers actually is. Does it make sense to try to stimulate collaboration by investing in tailor-made collaborative software? Can increased collaboration be expected in all fields? By building a simulation that is able to encompass the real life

tensions between competition and collaboration that all academics experience, the configuration of factors shaping collaborative behaviour may be studied. Rather than just repeating the mantra that 'collaboration is a good thing', this project will yield precise insights to researchers and funding agencies about the advantages and problems of scholarly collaboration. It will involve ICT researchers for the simulation building, social science and humanities scholars for the substance of what is simulated, and technographers to trace the evolving relations and identities.

4.4.4 The future of Internet-based peer review in qualitative social science and humanities

The goal of this project is to analyse the practice of peer review as shaped by an increased use of the Internet in the qualitative social sciences and humanities. The project will explore the effects of two current trends: (1) formalising quality control processes and (2) the increased use of ICTs in these processes. A prototype ICT environment could be built for peer review, tailor-made for an improved quality control in qualitatively oriented research fields. Special attention will be paid to the peer review of book proposals, which are so much more important in the humanities than in the natural sciences.

An important difference with more quantitatively oriented research is that usually there is no community-wide consensus about the criteria for scientific quality itself, let alone about the quality of specific research. As a result, the landscape is determined by a number of independent schools that either compete for funding, or simply do not communicate at all. This intellectual landscape is embodied in scientific journals, often one or more per school. Peer review is not so much an authoritative judgement that can be backed up by reference to a set of formal criteria, but more an ongoing scholarly discussion. Two trends seem to affect the present peer review practices in these fields: a drive to more formalised quality control processes on the one hand, and an increased use of ICTs in quality control processes on the other hand. It is unknown how these trends will affect the peer review processes. Special attention to the qualitatively oriented fields is justified because they seem more vulnerable to adverse effects of formalisation. A good prototype ICT environment might help lessen the burden of peer review in these fields.

4.4.5 e-Voting and e-polling: systems, use, implications for research and democracy

The goal of this project is to develop and test different prototypes for polling and voting via the Internet. What are the effects of the voting and polling technologies on the participation in and outcome of the vote/poll?

Surveys are still the most used method for data-gathering in the social sciences, especially for opinion *polling*. (For commercial surveys, most companies now use the telephone.) Much research has been done into social factors that influence the reliability of questionnaires as an instrument to measure opinions and attitudes. One of the clear results is that the social context of answering a questionnaire influences the outcomes of the poll: the behaviour of the interviewer, the way questions and answers are formulated, the presence of others while completing the questionnaire — these are all relevant. Therefore, election ballots are organized so that everybody votes more or less in the same situation (the voting booth) without interference form others.

Increasingly, the Internet is used for *opinion polls* and the expectation (and aim of politicians) is to use the Internet also for voting. Using the Internet for opinion polls and voting may increase the *variety* of social situations in which people express their opinions: at home, even in bed, or while having lunch in the garden, in public places, at work or school, in bars. Several theoretical expectations are relevant here. Social psychology predicts that group processes and group pressure may influence the voting decisions; social identity theory suggests that place influences social identity, and therefore the opinion expressed by someone. The trust in the secrecy, security, and accountability of the system may influence the willingness to use the technology, or may cause voters to avoid expressing radical opinions. Does increased computer literacy reduce or increase these possible effects?

4.4.6 Policies, Programs, Projects, Papers — ICT based tools for data analysis

The goal of this project is to study the translations that take place between the different levels policy making. As policy making involves an extensive use of a variety of documents in all phases, the comparison of large sets of documents may show where continuities and discontinuities in the policy and implementation processes occur. This requires ICT tools for computer based text analysis, and frameworks to understand and interpret lexical and semantic change.

Policies are formulated by political decision makers, and then translated into specific *programs* by civil servants. Within the programs, relevant social actors formulate *projects*, and these projects may or may not be funded by the programs. When funded, the projects are executed and have or do not have *results* as intended by policies and the derived programs. Of course, every step is contested and negotiated, and there is not a formal, one-way deduction between these steps. On all levels, various actors are involved with their own perspectives, interests, resources, and aims. Thus the results of the policies may be very different from the original intentions.

A possible case-study is the EU policy on biotechnology. How does this policy shape the research activities? This question can be answered applying computer assisted content analysis to large sets of relevant documents such as general science and technology policy statements (policy level), biotechnology work plans in the various framework programs (implementation level), project proposals (response of researchers to the implementation of the policies), and research papers in journals (the outcome of the work done within the programs).

4.4.7 Analysing the society through the Internet

The goal of this project is to develop tools for using Internet interactions to study social phenomena by exploiting the new 'data flood' that is generated by the Internet: log files, web links, bookmarks, databases, and various types of website content. What tools and techniques can help to handle and reduce the enormous amounts of data, to analyze them, and to visualize them for purposes of social science research?

Society increasingly is getting an *electronic overlay*: people and social processes leave many electronic traces, which can be used to study these processes and behaviours. This may solve a fundamental dilemma in social research: direct observation is only possible in small samples, whereas for large samples surveys are the only method.

A good case-study would be a field of natural science, where the use of new media is relatively well developed. Do the use of Internet and email influence the knowledge production and use? How do social networks and collaborations develop, as studied via the new media? How is the presence on the web of research groups a means of communication with peers and/or with users of the knowledge?

4.4.8 Developing and assessing tools for e-science in humanities and social sciences

The goal of this project is to give separate attention to some of the implementation and evaluation issues related to ICT tool development, that will play a more integrated role in all other projects. Sub-projects would be:

Requirements: Analysis of specific requirements for e-science in the humanities and social sciences, for instance in respect to the nature of data sets used, analysis methods and research questions, taking (international) experience from previous e-science projects into account.

Implementation: Design and implementation of (and possibly also connecting to) an infrastructure of tools and methodologies for e-science in the humanities and social sciences, using readily available technology as well as adapting this to any specific requirements (see above).

Application: In order to assess the functionality of the implemented *e*-science infrastructure, a number of case studies should be carried out. These studies will be more traditional in content (e.g. 'lexical analysis of medieval Frisian literature', 'impact of blogging on news reporting', 'cross-cultural effects of the Hansa league'), but novel in their execution (virtual cooperation with shared data and tools). They will need to be done in close cooperation with external research groups in the humanities and social sciences that became part of the implemented e-science infrastructure.

Evaluation: Meta-level analysis of the impact and effects of the implemented *e*-science infrastructure on those humanities and social sciences research groups that are actively using it for their daily work, for instance in terms of general effectiveness, change in research questions, change in data handling, etc.

4.4.9 Changing relations between expert and lay knowledge

The goal of this project is to study the impact of Internet and electronic communication on the status of experts and the relations between expert knowledge and so-called lay knowledge.

Collaborative methods are changing the nature of the profession of the scientist. Experiments could be organised with collaborative systems of research, while at the same time investigating the changing role of the scientific professional *vis a vis* the non-scientists or lay-person. The binary distinction between expert and lay comes into question In medicine, for example, we see that lay knowledge gathered and exchanged by patients and patient organizations leads to new forms of (scientific) knowledge development.

One case-study could be the involvement of non-scientists in ordering the archives of museums. Museum archives contain immense amounts of not exhibited items. The Internet would — in principle — provides the possibility to put them all on display, except that for most of these archives, the items have not been described properly to the standards of the professional archivist. Archivists will need centuries to do this work properly themselves, but in a smartly designed collaborative network much

lay-knowledge about the archive items could be enrolled. Inevitably, and well in line with the central idea of IRISS, this would require a close involvement of computer scientists too.

Typical results from this project would include the software architecture to involve lay knowledge without abandoning scientific standards of quality and verifiability; at least one finished pilot project; scientific publications presenting new findings due to this new method in the field at hand (e.g. art history, if that is what the pilot museum was about); scientific publications about this new type of knowledge creation; publications about related issues such as intellectual property; insights into questions of inclusion and exclusion, privacy and trust, identity and social cohesion.

4.4.10 Semantic-based Interdisciplinary Search

The goal of this project is to develop a common semantic framework for cross-disciplinary collaboration and to study the implications of such collaboration for the character of scientific knowledge.

A possible case-study can be drawn from the combined fields of language and ethnology studies. A linguistic sub-project could be aimed at developing a language typology and a characterization of the relations between different languages. This would result in a database of language elements. The ethnological sub-project in the same region would produce databases on the myths in the various cultures, and on the artefacts used in rituals. Combining and comparing these two very different databases could yield further insights into the linguistic relations between languages and the cultural-historical relations between people — diachronically and systematically.

To be able to use such databases in this comparative and integrating way, both need to be supported by sets of metadata. These metadata sets will be built on different disciplinary ontologies, and thus additional knowledge about the relations between these ontologies need to be developed.

4.4.11 A variety of project ideas

In addition to the projects outlined above, a supplement list of ideas for potential projects is provided, arising from the participants in the expert meeting in May. These ideas may serve as further indication of the wide scope of the research programme, but also as testimony of the commitment and creativity of the international experts who were involved in discussing the plans for an institute for *e*-science studies in the humanities and social sciences.

Remote conservation/analysis of paintings over broadband/GRID networks, enabling conservators in Amsterdam to carry out research, perform conservation, and do theft/forgery of art investigations on paintings which are physically located anywhere in the world. They would use remote tele-microscopy, remote surgical instruments, and distributed art databases (provenance records) and crime records.

Wireless tools for anthropologists/linguists in the field, enabling researchers to capture film/video/audio/images, create metadata, and download other related information from remote databases during their fieldwork. They could then continue to perform information analysis and data aggregation in the field to guide further empirical fieldwork.

Tools for indigenous communities to enable these to access and assimilate relevant cultural and historical information and knowledge within archives, cultural institutions and universities. This might help them to claim land rights, repatriate stolen artifacts, and revitalize the maintenance of cultural heritage. It would support the development of tools for indigenous genealogies, and for language, song, dance, and ritual preservation and protection.

Single-point access to multiple heterogeneous data collections, contributing to solving a problem common to almost all modern computer-based projects. Usually, data collections have been created by various people at various times, and no single standard exists. But researchers want easy, direct, and integrated access to all data at once. They also want to be able to define their own perspectives over the data collections, and they want the access system to accommodate this new perspectives by organizing and presenting the data accordingly. ICT groups in various countries are working on solutions; some try to standardize the data, some create 'wrappers' to encapsulate the databases, some use machine learning techniques to map the data to a centralized metadata schema automatically. There is no clearly accepted best solution today, but there is hope that effective methods will be developed within the next 5-10 years.

Automated text clustering and summarisation, helping to solve a problem that exists for all researchers processing large amounts of literary data. Despite the existence of four or five small companies, no current summarisation system is very good. Ongoing ICT research focuses on multi-document summarization, headline creation, text clustering, opinion-based summarization, and summarization according to the user's specified topic(s). Research since 1996 has shown encouraging results.

The creation of a large and standard collection of resources that support cross-disciplinary humanities and social sciences study would facilitate interesting studies in various hitherto unrelated disciplines, and thereby promote subsequent cross-disciplinary studies. For example, the British National Corpus, a genre-balanced collection of several millions of words of text, has proven invaluable for lexicography/dictionary creation, computational linguistics, language and literature studies, and other endeavours. One could imagine a project that collects and unifies a variety of data and information about the Netherlands, including geological, geographical, ethnological, linguistic, cultural and social data throughout history. The next step would be to order this in a set of cross-linked datasets. The material must be easily available, for instance by using a set of integrated access and display engines via the web.

Epistemology of *e*-science will study issues of objectivity, validation, replication, and truth in this emerging field. The epistemic culture of this field is sure to differ considerably from the natural sciences.

Technography of *e*-objects will study the development of a particular ICT solution or tool, including the user side that is presumably external to the institute but can be included via interviews, utilization studies and so on.

The imagined user can be studied by focusing on the various understandings, definitions and theories of imagined users (and imagined uses) that are implied by different ICT and *e*-science approaches.

Transparency and sustainability: What is the impact of global transparency on sustainability and economic growth? The global reporting initiative by a centre of the UN environment program is in itself a major *e*-science operation.

Transparency and criminality: What is the impact of transparency of criminal records on criminal investigation and punishment? The recent surge in the distribution of child pornography is strongly Internet-based, but so are the searching and reporting efforts.

Collaborative peer review may be one valuable development of the system of peer review. Classic peer review is end of line quality control; Internet provides examples and possibilities to create quality development in earlier stages of the research process. In literary science a gradually changing scientific standard can be seen: due to computerization it seems more and more necessary that scientific claims be underpinned with quantitative research. In literary science an eloquent essay used to pass the scientific standard. IRISS might set out to analyse the change in measures of being scientific in literary science.

Classification and typology can develop into wholly new directions. For centuries archivists have created thesauri to classify documents. In a digital age it seems more than necessary to search for smarter ways of classification, if only because attaching keywords to archive material continues to be an immense effort. An interesting hypothesis in this respect is to stop working with metadata at all: smarter search engines and greater number crunching power may do the same job much more swiftly.

Evidence: The emergence of new knowledge technologies seems to change our notions of what counts as evidence and this seems especially acute in the humanities and social sciences. The increasing possibilities of quantitative methods seem to produce a strong focus on the quantitative underpinnings of any scientific claim, leading to new discussions about the replication of results. Also outside the core scholarly practices a growing emphasis on ‘evidence’ can be observed: as in ‘evidence based policy’ or ‘evidence based medicine.’ How does the very meaning of the concept of evidence develop in these various contexts, influenced by the development of *e*-science?

5. BLUEPRINT FOR A RESEARCH INSTITUTE

This final section aims to translate the analysis of the scientific challenges and the related research programme for *e*-science studies in the humanities and social sciences into a blueprint for a research institute. The proposed name for the institute is the *International Research Institute on e-Science Studies in the Humanities and Social Sciences* (IRISS).[15] In the sub-section 5.3, below, on critical size, it is argued why the previously sketched research programme can only be carried out by a full-sized research institute.

5.1 Mission

IRISS's mission is to develop and stimulate the advanced and innovative use of ICTs in the humanities and social sciences, and to study the effects of the new ICTs on the practice, activity and quality of scientific research in the social sciences and humanities. This mission is to be pursued by an integrated programme of cooperative research between the humanities, social sciences, science studies, and computer sciences. The aim is to advance the use of ICTs in social science and humanities and at the same time to critically reflect upon their usage in the process. IRISS's mission is thus to help create a new form of humanities and social science scholarship, in addition to the existing traditions.

IRISS is a research institute and seeks to be evaluated on its scientific production. The products may vary in nature: from scientific articles in reviewed journals, to IT network designs, and to new methodologies for doing advanced research in the humanities and social sciences. IRISS should be the (co)producer of new developments in ICT-use; not a mere consumer of new ICT possibilities, nor a helpdesk to provide ICT tools and expertise to others.

IRISS should aim at playing a central role in the development of the humanities and social sciences in the Netherlands, and at becoming a leading international centre of excellence in *e*-science studies in the humanities and social sciences.

5.2 Profile

To realise this mission, IRISS will be built around a unique combination of substantive research in the humanities and social sciences, with the development of methodology and tools, and with critical technography of the process of research technology development. IRISS will thus draw on the following three constituencies:

- Humanities and social sciences (including such disciplines as history, philosophy, linguistics, literary theory, sociology, anthropology, Internet research, communication science);
- Science and technology studies (especially for technography);
- Computer sciences (including information science, software development, systems engineering, library & information sciences).

[15] Much debate was spent on choosing this name, and a long list of alternatives was discussed. See appendix 6.2 for a discussion of the pro's and con's of the name IRISS.

These three constituencies will have equal intellectual importance in the research programming and practice of IRISS, although the practices of interdisciplinary research should make the distinction between the three increasingly irrelevant and even invisible.

IRISS will perform a combination of fundamental and applied research, and of academic agenda setting and policy relevant studies.

IRISS will engage in theory-guided empirical research (beyond mere classificatory exercises), and in empirically based theory development. It will be developing new technological environments and tools, as a direct result from close cooperation with humanities and social science scholars and technographers. As a result, innovative combinations of quantitative and qualitative research will be developed.

The IRISS profile can be further specified by stressing that in its research style and approach, it will be conceptually informed by and build on:

- Positive scepticism;[16]
- Technography;
- Recognising knowledge as being situated;
- Analysing: Technology as text ⇔ Text as technology;
- Constructivist analysis of epistemic objects
- Key role for mediation and representation;
- Co-evolution of object and knowledge;
- Infometrics and scientometrics;
- Internet expertise (technical/design);
- Computer and network sciences.

5.3 Critical size

Would it be possible to realise the outlined research programme with other organisational solutions than a full-fledged research institute? At least two alternative models can be considered — (1) a short-term KNAW project, drawing researchers from other institutes and working as a dedicated task force; and (2) a network of university institutes.

A task force on temporary basis we do not deem capable of creating the long-term intellectual investments that are needed to establish the fundamental innovation of research practice that is called for. The challenge of infusing research in the humanities

[16] A few examples may illustrate this positive scepticism. Improved access to and use of various knowledge bases are widely expected; but this may also lead to increasing capital intensity and higher research costs. Higher productivity and new creativity of the research system can be hoped for; but unrealistic promises may backfire and erode the legitimacy and authority of science. More interdisciplinarity is promised as a result of introducing digital interconnectivity and sharing of databases; but the development of such research tools and methods may also have the opposite effect of inward-looking investments leading to more mono-disciplinarity. Knowledge creation will be stimulated with the use of new ontologies; but static ontologies may also yield a 'tunnel vision' that blinds creativity and innovation. The use of cooperative networks will introduce new and challenging quantitative methods into the humanities and social sciences; but without a true integration of—or at least a careful balance between—quantitative and qualitative approaches, this may result in a decreased variation of epistemological cultures and the dominance of a one-sided model of research.

and social sciences with ICT in a more sophisticated way than dropping some new tools must not be underestimated. In this respect lessons can be learned from the experience in the Netherlands with 'alpha-informatica.'

In due time, IRISS will certainly become a partner in a network of institutes. To build such a network, however, strong institutes are crucial as hub and nodes. Especially in establishing new research practices the crucial role of personal interaction among researchers cannot be underestimated — around a whiteboard, a coffee machine, a reading table, or after a seminar. A network of institutes will serve the purpose of making databases more widely available and sharing tools and methods. The development of new scientific approaches — empirically and theoretically — and, indeed, the creation of a new field of scholarly work in the form of *e*-science studies does need the intellectual and social solidity of a physically existing institute. For similar reasons, IRISS will need a strong international fellowship programme. Intense collaboration with visiting scholars will help to give the institute a flying start, and a continuous exchange of researchers will help to establish IRISS internationally. It is also anticipated that past visiting fellows to IRISS will often continue to remain what might be termed 'corresponding fellows', thus enabling the institute to be linked to a varied programme of research, both nationally and internationally, that may be developed and implemented in higher education institutions other than IRISS, yet which may have originated from within IRISS.

What is the critical mass needed by such an institute to function as the scholarly gravity point that can create the new *e*-science studies in the humanities and social sciences that can be argued for? Comparison with other research institutes in the humanities and social sciences suggests that a minimum of 25 researchers and 15 graduate students (MA and PhD) is needed.

It is crucial to incorporate graduate students in the basic set-up of IRISS. One reason is that they contribute to the intellectual critical mass and the creation of an intellectually vibrant institute. Another reason is that such students form the driving force for the future. It has indeed been argued that exactly the uptake of ICT by young scientists has given rise to such astounding results in the natural and life sciences in a relatively short period of time.

5.4 Structure and budget of the institute

The structure of the institute can be kept quite simple so as to maintain maximum flexibility. One possible model will be discussed here.

The director is primarily responsible for the research programme and the research staff. The research programme will, in a stable situation after five years (see below), consist of 3-4 research lines; research groups to carry out projects can be organized around these lines, with a senior staff member (who preferably has a chair in this field at a Dutch university) as coordinator. Typically such a project group will consist of:
2-3 senior staff members (one of them being the coordinator)
0-1 guest researchers for longer or shorter periods
1-3 postdocs

2-5 PhD students
2-5 Master students

Several boundary conditions can be specified that need to inform the set-up and management of research groups:

- To have the intended impact on Dutch research, IRISS needs to build up and maintain a close working relationship with all relevant Dutch university institutes, faculties, and research groups, as well as with other KNAW institutes in the humanities and social sciences[17];
- To be able to play a serious role in the international research community, IRISS needs to maintain intensive relations with foreign institutes and researchers; the international fellowship program is one measure to help realise this;
- To acquire and maintain a position at the research front, IRISS' research programme needs to be flexible and adaptable to new research priorities and international funding possibilities, and will thus be organised in terms of 3-5 year projects;
- To generate extra funding, IRISS needs to apply for and cooperate in EU funded projects;
- To generate original research questions and to engage in research that is also relevant to non-academic audiences, IRISS will cooperate with non-university and industrial partners in contract research.

A rough indication of the annual budget for the institute is the following:

	Structurally financed by KNAW (M Euro)	External, temporary funds (M Euro)
Research staff	1.7	1.0
Support staff	0.6	0.2
Graduate students		0.8
Subtotals	2.3	2.0
Total	4.3	

A relatively large part of also the structurally financed positions will be filled on a temporary basis to guarantee flexibility in the research planning. The international fellowship programme will also be part of the structural funding, including grants and office and housing facilities.

The Director needs to be an internationally recognized expert in one of the relevant constituencies; a charismatic leader who can inspire staff to embark on the challenging and sometimes unsettling interdisciplinary cooperation that IRISS requires. The Director should have experience in research management and project acquisition. Although the Director need not be Dutch, political and diplomatic skills are important.

[17] Collaboration with the recently proposed KNAW data services centre will be crucial, for example. See the report "Networked Data Services — Towards a Future Data Infrastructure for the Social Sciences in the Netherlands" by the SWR-KNAW (draft August 2003).

In addition to the Director, the institute will have an Executive Administrator responsible for the support staff, housing, finance, personnel policy, and IT-infrastructure. One of the Executive Administrator's important roles will be to assist the research staff in the process of preparing grant proposals and in managing grants once they have been obtained. The Executive Administrator does need to be fluent in Dutch, as well as skilled in research management, contract research, and human resources and financial management.

Many senior researchers of IRISS will be part-time professors at Dutch universities. It is also conceivable that full professors at Dutch universities will be 'delegated' to IRISS for a period of some years.

An international Advisory Board should be in place from the beginning. Its function would not only be to advise the Director and to provide some of the necessary internal peer review, but also to help establish international relations of IRISS. Evidently, the KNAW can build for this on the commitment to the IRISS endeavor that already exists among the participants of the May expert meeting and the members of this committee.

5.5 Establishing IRISS: Growth and consolidation path

The KNAW, when deciding to create IRISS, should specify a five-year trajectory for the gradual growth of IRISS. A careful strategy, including the building up of national collaborations and international relations, should aim at bringing IRISS to full size only after five years. An intermediate evaluation after three years may help to check this process.

The evaluation criteria to be applied after the first five year period should be specified in advance. This is one of the issues that our committee discussed extensively with the experts in the May meeting. Drawing on their suggestions, the following evaluation criteria are proposed:

- International recognition (in the form of international publications; editorships of journals and special issues of journals; invitations for lectures and teaching; co-organisation of international conferences; requests to referee papers etc.);
- A sizeable part of the budget is externally funded (between 30% and 50%);
- A thriving network of international and national collaborations and a productive visiting fellowship programme exist;
- All senior researchers have university positions (the type may vary depending on the field);
- The institute must have produced some technology/techniques/data resources that are regularly used by social science and humanities researchers who are not themselves members of the institute;
- The institute has established its own publication series;
- The following medium output should be expected from every academic staff: one book length publication every fifth year, three journal length articles every year when no book is published.[18]

[18] A maximum of 50 % of the publications could be published in the in-house series. We explicitly advice against stressing the notion of 'peer reviewed publications' and/or numeric impact factors, and/or specific

The success of IRISS will depend crucially on its adopting a proactive outreach strategy. Excellent research itself is a necessary but insufficient determinant of research success. As many studies of the success of scientific and technological ventures have shown, IRISS will need to put in place an energetic set of activities for 'configuring its users.' This will entail the development and maintenance of a policy of interactive social science in publishing, conferencing, building networks of supporters and sponsors, etc. Here, the experiences and insights from science studies are highly relevant.

Another way for IRISS to help establish itself and to shape the scientific community and international research agenda, is to establish its own publication series. While its quality will in many ways be more difficult to judge than a 'peer reviewed journal', there are simply too few places for publication that will reach the complete targeted community with studies of interdisciplinary research. The institute needs a medium to address such communities with purpose written texts.

numbers of pages expected. The relevant rules in humanities, social sciences, science studies, and computer sciences are so different, that at least one of the communities would be disadvantaged.

6. APPENDICES

6.1 Committee's assignment: letter from the KNAW Board

Dir/443
March 2003

COMMITTEE ON RESEARCH INSTITUTE FOR *E*-SCIENCE
ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES (KNAW)

Background

In October 2002 the Strategy NIWI Committee submitted its report to the Board of the Academy. It recommended to close the present Netherlands Institute for Scientific Information Services (NIWI) and to establish a new KNAW Institute for *e*-science research in the humanities and social sciences. In its position paper of October 2002 the KNAW Board has decided to establish a committee of national and international experts to study the desirability and feasibility of such an *e*-science research institute.

Although the report of the Strategy NIWI Committee already provided indications on the potential, global research domain for such an institute, it was felt that the positioning of such an institute in the context of the Dutch research network required some further, open discussion before an expert committee could be established. On December 17, 2002, a workshop took place where selected ICT users and experts from the humanities, social sciences and life sciences met to discuss how an *e*-scientific institute could boost intra- and interdisciplinary sophistication in the humanities and social sciences. Among the participants were representatives from the NIWI. A broad spectrum of research topics was discussed, roughly in two wide domains: the tools and methods domain (with topics such as data coupling, webometrics and infometrics) and the science dynamics domain (with topics like accessibility, reliability, new communicative styles, intellectual property). Noticeable convergence arose when the discussion came to focus on metadata. A research strategy for the new institute, it was argued, could be primarily case-based in the following sense. Each case would consist in the coupling of two or more data domains (for which there should be an obvious need), such as survey and text data, or historical and archeological data. Such coupling requires the domain specialists to explicitly define the 'ontology' of their field, i.e. the core data types, concepts and their relations. Only then a coupling or integration can be established on the meta-data level. This approach would, moreover, contribute to a much-needed self-reflection in some of the humanities and social sciences, breaking through self-imposed, tradition-based isolation in some of their subdisciplines. Such an approach could be dynamic because cases would be completed within a few years, and replaced by new ones that are proposed by the research community. Dissemination of such sophisticated data architectures would be among the tasks of the institute. The metadata approach would, in addition, find a natural embedding in already existing large-scale European initiatives.

It should, however, be obvious that these discussion results should in no way delimit the freedom of the to be established committee; they rather demonstrate the remarkable potential of an *e*-science approach in the humanities and social sciences (without excluding 'spill-overs' to the life or natural sciences) and invite to a deeper analysis.

Assignment

The committee's task will be twofold:

1. In a first step, the committee should consider whether a research profile for an *e*-science institute can be defined, which is viable scientifically and within the national and international scientific network. The profile should be one for *e*-science in the humanities and social sciences, without excluding connections

of these sciences to the life or natural sciences. The design task of the committee will require interaction with NIWI's active researchers.
The committee should in particular consider:

- the optimal balance between basic and applied research in the institute's profile. It will be essential in both domains for the institute to publish in the world's top journals;
- whether, to what extent and how the present *e*-science research activities of the NIWI can be integrated in the institute's research programme;
- how a promising research programme is best organized as an institute, among other things in terms of departments, tenure-track and variable positions, type of infrastructure and in general: flexibility and dynamism in the organization;
- how such an institute should be embedded in the national and international *e*-science networks; the Committee should specifically consider the optimal mode of cooperation with related institutions and research groups within the Dutch universities;
- whether the institute is best served with a single or dual headed scientific directorate and what is the desired personal profile of the director (or directors);
- organizing an expert symposium as a sounding board during the conceptualization of the research profile.

The indicative number of KNAW-funded researchers will be approximately 12 to 14 researchers (excl. support staff).
The committee will present its recommendations to the Board of the Academy. The Board will then, at the shortest possible notice, come to a decision whether and in which terms an institute for *e*-science will be established.

2. In the positive case, the Board will turn the Committee into a search committee. For that purpose the composition of the committee will be extended with other members with relevant expertise and experience. The Committee's new task will be to consider suitable and potentially available candidates for the function of scientific director. Primary criteria should be internationally recognized scientific leadership in the field of *e*-science and recognized skills as scientific team leader.
It would be most welcome if the Committee could complete its work by summer 2003.

The Committee

The Committee on Research Institute for *e*-science has the following composition:

Prof. dr.ir. Wiebe Bijker (chair)
Dr. Els Stronks
Prof. dr. Hans Uszkoreit
Ir. Peter Wittenburg
Prof. dr. Steve Woolgar
vacancy (historian)[19]

Dr. Gaspard de Jong (NIWI) will function as advisory member of the committee.
Dr. Krijn van Beek will be the committee's executive secretary.

Board of KNAW

[19] In April Prof. Kevin Schürer joined the committee.

6.2 'What's in a Name?'

The proposed name of the institute — International Research Institute on *e*-Science Studies in the Humanities and Social Sciences (IRISS) — highlights some of the problems one encounters when formulating such an institute's research area and mission.

First, for some the prefix '*e*-' has become synonymous with hype and failure. To others, particularly to native English speakers, the label 'science' suggests an emphasis on natural and technical sciences. And finally, the phrase 'science studies' is often read as denoting the social studies of science.

IRISS is meant to be none of these exclusively: it is not naively built on wildly optimistic expectations about the application of ICT and digitisation in the humanities and social sciences, but on a broadly recognised need to help Dutch humanities and social sciences to maintain and further strengthen their forward position in a quickly changing world; the institute is aimed at the humanities and social sciences, but recognizes that innovative research cannot be realised with off-the-shelf tools and consequently needs an integrated approach to technological tool development and substantive humanities and social sciences research; and the institute's mission is not exclusively in the sociology of science, but rather seeks to integrate the production of scientific knowledge with a critical analysis of this production process and the use of the resulting humanities' and social sciences' insights in various practices.

The committee has considered a variety of names. We briefly summarize the arguments that support each of the elements in IRISS:

- '*e*-': for most people, this prefix does generate the right associations of computer and Internet supported scientific and scholarly work. When, in another five or ten years, this label may loose its meaning, it can be downplayed by giving more emphasis to the acronym IRISS as such;
- 'Science': the narrowness of the natural science connotation is countered by explicitly mentioning the social science and humanities at the end of the name; the combination '*e*-science' helps to recognize the important role of the computer sciences in the institute;
- 'Science studies': the combination of 'studies' and 'science' means to highlight the constant monitoring and technography of the on-going research projects in the institute;
- 'Humanities and social sciences': although not clearly visible in the acronym, these elements in the full name stress the primary relevance for the humanities and social sciences; it is also important to highlight that the institute does indeed regard *both* the humanities and the social sciences;
- IRISS: the acronym can stand alone, and does not seem to have problematic other meanings (see below).

One serious alternative would be a name that foregrounds a long term commitment to methodological innovation, of which the current focus on *e*-science is but one instantiation. The name 'Humanities And Social Science Centre for Innovative Technologies' (HASSCIT, pronounced *Hass-it*) would do this.

Whatever the name, a determined strategy of 'name branding' needs to be implemented once the institute is created. A website, annual report, (electronic) publication series, and active outreach to existing other institutions will help to establish the name of the institute.

The acronym IRISS already has several meanings, some of which we list below. None of these seems to pose any legal problems, but this needs to be checked more carefully.

- In Greek mythology, Iris is the personified goddess of the rainbow. She is regarded as the messenger of the gods to mankind, and particularly of the goddess Hera whose orders she brought to humans. Iris is the daughter of Titan Thaumas and the nymph Electra. She is portrayed as a young woman with wings and her attributes are a herald's staff and a water pitcher.
- The IRISS project of the European Union for an 'Integrated Research Infrastructure in the Socio-Economic Sciences' aims to foster access to information and mobility of European researchers in the socio-economic sciences by offering access to the local research facilities and archive of data. Since 1998, the IRISS fellowships offer European researchers (both junior and senior) the opportunity to spend time carrying out their own research in CEPS/INSTEAD. CEPS/INSTEAD is a socio-economic research institute in Luxembourg (*Centre d'Études de Populations, de Pauvreté et de Politiques Socio-Économiques / International Networks for Studies in Technology, Environment, Alternatives, Development*). It originated from a private, non-profit organisation founded in 1978 for comparative, trans-national research on persistent poverty in industrialised countries (within the framework of the first EC programme to combat poverty). The Centre was given, by Luxembourg law in 1989, the status of a Public Establishment with scientific, administrative and financial autonomy. (http://www.ceps.lu/iriss/iriss.htm)
- The Interdisciplinary Research Institute for Survey Science (IRISS) at Iowa State University is (according to their own website) one of the largest academic research organizations in the world focusing on the collection, analysis, and dissemination of information. Its faculty represent many of the disciplines on the Iowa State University campus, including education, family and consumer sciences, psychology, sociology, and statistics. (http://www.iriss.iastate.edu/)
- The Integrated Research Information Support System (IRISS) of the Research Center for Minority Institution on Puerto Rico provides a computer communications network environment for the faculty involved in research. The system provides faculty with connection and access to worldwide information resources to facilitate the expansion, strengthening, exchange, and dissemination of their biomedical research work and findings. (http://rcmiucc.org/iriss.htm)
- The IRISS-98 international conference on Internet Research and Information for Social Scientists, held in 1998 in Bristol, UK, seems to have been a unique event. (http://www.sosig.ac.uk/iriss/papers/proceed.html)
- The Institute for Research in the Social Sciences (IRISS) of the University of York, UK, is — according to their own website — one of the largest multi-disciplinary centres for social science research in the United Kingdom. Its units contain approximately 150 research and support staff and can call on the services of staff in the academic departments of Economics and Related Studies, Health Sciences, Health Studies and Social Policy and Social Work. (http://www.york.ac.uk/inst/iriss/)
- Iris, in most languages, refers to members of the flowering plant family *Iridaceae*. (http://biodiversity.uno.edu/delta/angio/www/iridacea.htm)

6.3 Participants in the May Experts' Meeting

	Bennis, Hans (Prof.dr. H.)	Meertens Instituut Postbus 94264 1090 GG Amsterdam	Hans.Bennis@meertens.knaw.nl
	Besselaar, Peter van den (Dr. P.A.A.)	NIWI Postbus 95110 1090 HC Amsterdam	peter.van.den.besselaar@niwi.knaw.nl
	Buitelaar, Paul (Dr. P.)	DFKI (German Research Center for Artificial Intelligence) GmbH Language Technology Department Stuhlsatzenhausweg 3 D-66123 Saarbrücken Germany http://www.dfki.de/~paulb/home.html	paulb@dfki.de
	Hagendijk, Rob (Dr. R.)	Universiteit van Amsterdam Faculteit der Natuurwetenschappen, Wiskunde en Informatica Afdeling Scheikunde Vakgroep Wetenschaps- en Technologiedynamica Postbus 20242 1000 HE Amsterdam	hagendyk@ishss.uva.nl
	Harmelen, Frank van (Prof. dr. F.)	Department of AI, Faculty of Sciences, Vrije Universiteit Amsterdam de Boelelaan 1081a, 1081HV Amsterdam, The Netherlands tel (+31)-20-444 7731/7700 fax (+31)-84-221 4294 http://www.cs.vu.nl/~frankh	Frank.van.Harmelen@cs.vu.nl
	Hine, Christine (Dr. C.)	Department of Human Sciences Brunel University Uxbridge UB8 3PH United Kingdom http://www.brunel.ac.uk/depts/hs/SOCSTAFF/CHNew.html	christine.hine@brunel.ac.uk
	Hovy, Eduard (Dr. E.)	Information Sciences Institute of the University of Southern California 4676 Admiralty Way Marina del Rey CA 90292-6695 http://www.isi.edu/natural-language/people/hovy/bio.html	hovy@isi.edu
	Hunter, Jane (Dr. J.)	The University of Queensland Level 7 GP South Staff House Rd Queensland 4072	jane@dstc.edu.au

		Australia http://archive.dstc.edu.au/RDU/staff/jane-hunter.html	
	Knorr-Cetina, Karin (Prof.dr. K.)	University of Konstanz Fachbereich Soziologie 78457 Konstanz Germany University of Bielefeld Fakultät für Soziologie und Institut für Wissenschafts- und Technikforschung Postfach 100 131 D-33501 Bielefeld Germany http://www.uni-bielefeld.de/iwt/gk/profs/knorr.htm	karin.knorr@uni-konstanz.de knorr@uni-bielefeld.de
	Mackenzie Owen, John (Prof.dr. J.S.)	University of Amsterdam Information Science Oude Turfmarkt 147 1012 GC Amsterdam The Netherlands http://cf.hum.uva.nl/bai/home/jmackenzie/home.htm	owen@hum.uva.nl
	McCarty, Willard (Prof.dr. W.)	King's College London Centre for Computing in the Humanities Room 11BB Main Building Strand London WC2R 2LS United Kingdom http://www.kcl.ac.uk/humanities/cch/wlm/	Willard.McCarty@kcl.ac.uk
	Nerbonne, John (Prof.dr.ir.)	University of Groningen Faculteit der Godgeleerdheid en Godsdienstwetenschap Opleiding Alfa-informatica Nieuwe Kijk in 't Jatstraat 104 9712 SL Groningen http://odur.let.rug.nl/~nerbonne/	nerbonne@let.rug.nl
	Renn, Jürgen (Prof.dr. J.)	Max Planck Institut für Wissenschaftsgeschichte Wilhelmstra_e 44 10117 Berlin Germany http://www.bu.edu/philo/faculty/renn.html	renn@mpiwg-berlin.mpg.de
	Ross, Seamus (Dr. S.)	Director of Humanities Computing & Information Management Glasgow University Humanities Advanced Technology and Information Institute (HATII) George Service House	S.Ross@hatii.arts.gla.ac.uk

		11 University Gardens Glasgow G12 8QQ Scotland http://europa.eu.int/ISPO/dlm/program/bio_sr_en.html	
	Sörensen, Knut (Prof. K.H.)	University of Trondheim Department of Interdisciplinary Cultural Studies 7491 Trondheim Norway http://www.hf.ntnu.no/itk/sts_sorensen/cv.htm	knut.sorensen@hf.ntnu.no
	Thaller, Manfred (Prof.dr. M.)	*Historisch-Kulturwissenschaftliche Informationsverarbeitung* *Universität zu Köln* *Albertus-Magnus-Platz* *50923 Köln* Germany http://www.hki.uni-koeln.de/people/thaller/mt.html	manfred.thaller@uni-koeln.de
	Vanhoutte, Edward (Dr. E.)	Belgische Koninklijke Academie voor Wetenschappen CTB - Centrum voor Teksteditie en Bronnenstudie Koningstraat 18 B-9000 Gent Belgium http://ger-ww.uia.ac.be/webger/ger/people/vanhoutte/default.htm	evanhout@uia.ua.ac.be evanhoutte@kantl.be
	Voorbij, Hans (Dr. J.B.)	Utrecht University Department of Information Science Faculty of Mathematics and Computer Science Achter de Dom 22-24 3512 JP Utrecht The Netherlands http://www.cs.uu.nl/groups/IK/archives/voorbij.htm	Hans.Voorbij@let.uu.nl
	Wouters, Paul (Dr. P.F.)	NIWI Postbus 95110 1090 HC Amsterdam	paul.wouters@niwi.knaw.nl
	Wyatt, Sally (Dr. S.)	Universiteit van Amsterdam Communicatiewetenschap Oude Hoogstraat 24 1012 CE Amsterdam The Netherlands http://virtualsociety.sbs.ox.ac.uk/people/wyatt.htm	wyatt@pscw.uva.nl
Secretary to the committee	Beek, Krijn van (Dr. K.W.H.)	Raad voor Maatschappelijke Ontwikkeling P/a Veronesestraat 2 NL 1077 RB Amsterdam	vanbeek@xs4all.nl
Chairman of the	Bijker, Wiebe (Prof.dr.ir. W.E.)	Universiteit Maastricht Faculteit der Cultuurwetenschappen	w.bijker@tss.unimaas.nl

committe		Postbus 616 6200 MD MAASTRICHT The Netherlands	
Advisor to the Committee	Jong, Gaspard de (Dr. G.A.)	NIWI Postbus 95110 1090 HC Amsterdam	gaspard.de.jong@niwi.knaw.nl
Committee member	Schürer, Kevin (Prof.dr. K.)	UK Data Archive University of Essex Wivenhoe Park Colchester CO4 3SQ United Kingdom http://www.essex.ac.uk/history/staff/schurer.htm http://www.esds.ac.uk	schurer@essex.ac.uk
Committee member	Stronks, Els (Mw. Dr. E.)	Universiteit Utrecht Inst. Nederlands/ Nederlands/Renaissance Letterkunde Trans 10 3512 JK Utrecht The Netherlands	Els.Stronks@let.uu.nl
Committee member	Uszkoreit, Hans (Prof.dr. H.)	Language Technology Lab Deutsches Forschungszentrum für Künstliche Intelligenz GmbH Stuhlsatzenhausweg 3 D-66123 Saarbrücken Universität des Saarlandes Computerlinguistik D-66041 Saarbrücken	uszkoreit@dfki.de uszkoreit@coli.uni-sb.de
Committee member	Wittenburg, Peter (Ir. P.)	Max Planck Institute for Psycholinguistics Wundtlaan 1 (Postbus 310) 6500 AH Nijmegen	Peter.Wittenburg@mpi.nl
Committee member	Woolgar, Steve (Prof. S.)	Saïd Business School University of Oxford Park End Street Oxford OX1 1HP England	steve.woolgar@sbs.ox.ac.uk