E-based Humanities and E-humanities on a SURF Platform
Digest of expert meeting 31 August 2004
Final version 13 September 2004
This report has been read and agreed upon by the participants

Attendees:

Kurt de Belder. Hoofd elektronische diensten UB. Univ.van Amsterdam.
Dr Karina van Dalen-Oskam. Hoofd afdeling neerlandistiek. NIWI-KNAW.
Dr Peter Doorn. Hoofd afdeling geschiedenis. NIWI-KNAW.
Prof. dr Jaap van den Herik. Computer Science Univ. Maastricht. Recht & Informatia Univ. Leiden.
Dr Hans Kamermans. Archeologie. Univ. Leiden.
Dr José de Kruif. Projectgroep cultuurgeschiedenis, Letteren. UU.
Prof. dr John Nerbonne. Alfa informatica, Letteren. RUG.
Dr Ben Peperkamp. Inst. Nederlands Letteren. UU.
Dr Henk Wals. Adj. Directeur IISG-KNAW, Directeur CHI-KNAW.

Dr Leo Waaijers. SURF.
Dr Joost Kircz. KRA.

Written comments prior to the meeting were given by:
Leen Breure (UU), Peter Doorn, José de Kruif, John Nerbonne, Eep Talstra (VU), Henk Wals.

==========

Chairman Leo Waaijers welcomes all attendees and expands on the goals SURF has in mind with this expert meeting. Leo is project leader of the *ICT and Research* platform. The aim is to widen our understanding of the notion of E-Science in the humanities and to reach an inventory of issues needed to allow SURF to fulfil its role as initiating, facilitating and stimulating ICT organisation in Dutch academic research and higher education. A report was commissioned to Joost Kircz. This report and its recommendations are now on the table for: comments, critiques and, in particular, a discussion on the priorities. The report, the outcome of this meeting and the 3rd Conference on ICT and Humanities to be held 30th September and 1st October, will be the basis of a policy discussion on SURF's tasks vis-à-vis the humanities.

Joost Kircz expressed his thanks to all people who provided input to the report and commented on the final result.
He shortly reviewed the main message of the report:
1. A clear distinction has been made between ICT as enabling technology that enhances actual research programmes and the next phase in which novel research emerges that intrinsically uses electronic tools and methods. The first phase is coined E-based Humanities, the later E-Humanities. In the first phase the electronic corpora are used as reference tool, in the second phase they are real research corpora that develop and change in the usage.

2. A crucial issue is the availability of electronic corpora, data sets and the relevant software needed to fully exploit the capabilities of methods based on electronic tools.
3. Electronic corpora need structuring, which demands on the one hand standardisation, and on the other hand sufficient domain dependent idiosyncratic annotations. This whole issue of descriptive languages, metadata, ontologies, thesauri, etc. is still in full development.
4. Typically for the field is the rift between researchers who work according to traditional methods and those that already exploit electronic methods and technologies. The need for electronic means is very unbalanced between the various (sub) disciplines.
5. Education and training, not only for students, but also for mature researchers are badly needed.
6. To allow E-Humanities to flourish it is needed to quote John Nerbonne "*to stimulate work that is convincing for de non-computationally oriented humanist and that makes essential use of the computer".*

First round of discussions aimed at a general assessment of the report.

The attendees praised the report, but it was felt that the tedious point of how traditional humanities' research could take advantage of the new possibilities needs more discussion and attention. The heterogeneity of the humanities was mentioned as a great obstacle to the creation of "methodological commons" as defended in the report. An interesting discussion emerged to what extent good examples and flexible methods and standards could assist in promoting ICT usage in traditional domains, such as literature studies. However, it was also argued that in all sciences we see various tiers of research programmes that have very different levels of ICT involvement.
A general agreement was expressed on the devastating low priority computer literacy has in the new BaMa curriculum of the universities. The issue of education was felt as one of the prime concerns.
Also the issue of financial support that diminished after the eighties of last century and the weak position of Humanities Computing (alfa-informatica) was mentioned. Only in Groningen the group is still alive and kicking. A stronger bond between informatics and humanities was deemed necessary. Nevertheless, in many faculties slowly genuine novel ICT applications can be found.
A special point of attention is that journal publications that want to expand on methodology, often face problems with editorial boards of scholarly journals. It is not unusual to be asked that methodology and/or data be tucked away in appendices. The specialised journals on computing and humanities are normally not read by main stream researchers, which is negative for the reach out of new research avenues.
Due to the greater ingrowth of ICT in civil life, the rift can be bridged if the advanced research endeavours keep close contact with its peers. It might be a good idea to concentrate the SURF activities to those that break down the

barriers and enhance accessibility, whilst leaving the front-end research to NWO. Part of such a project would include the promotion of re-using methods and data, next to the availability of well-structured research corpora.

The second round of the discussion was on the prioritising of the recommendations in the report.

The most enthusiastic discussion constantly revolved around the issues of methodological training and tools against the need for corpora per se as starting point for new works.
Also the issue of metadata (and languages) got a lot of attention as proper domain compliant structuring is a pre-condition for research. Not only the sheer size of the corpus, but the capability to manipulate it is of crucial importance. It was stressed that the sheer creation of electronic corpora is not enough. Research corpora need structuring, but it is obvious that not all essential elements can be defined prior to the actual research. The capability for dynamic enrichment of research corpora is an important precondition for success.

In order to reach some tentative conclusions all participants (excluding Waaijers and Kircz) were asked to rate the 10 recommendations of the report.
For reference sake, we mention them again in shorted form.

0- The creation of large digital corpora as a precondition for E-Humanities
1- Changing the emphasise of the curriculum from training in (commercial) office automation packages to a more methodological coupling of intrinsic research interests to modelling and architectural system demands.
2- The extension of the helpdesks from system gatekeepers to people who help to select software and make product analyses of (commercial) software.
3- To improve the role of Information science as an interdisciplinary field. As said above, this means strengthening the bond between informatics, information science, and the particular (sub) domain.
4- The need for data repositories that warrant re-use of data for extensive periods of time.
5- The building of a program library, where people can deposit their working own software, provided it has a clear description and manual.
6- The development of Webservices to enable Internet asses to a great variety of data and other repositories.
7- Furthering the work on Metadata structuring and ontology languages.
8- The development of ontology-based authoring tools and collaborative software.
9- The development of human-computer interfaces and visualisation techniques to enable multiple media rendering of data. In the discussion it was coined Multimodularisation.

Though no new recommendations were suggested, the following comments were made and agreed upon. The creation of corpora (#0) as well as data repositories (#4) must be considered as preconditions. The recommendation on information science (#3) must be made concrete by emphasising a closer tie between informatics, information science and the various humanities' domains. Recommendation #6 is a bit too technical and must be interpreted as using web services as threshold lowering techniques that enable peers to learn what is possible and beautiful. Recommendation #9 on human-computer interfaces was recoined *Multimodularisation*, to express the need to allow for a manifold of presentation schemes, that enable dedicated manipulation by the end-user. On the creation of corpora itself it was mentioned that although the fact that they must be research driven, it is important to emphasise also the general cultural importance, as is the case with the very large US repositories on, e.g., US history or ancient cultures.

The participants were asked to rate the 10 recommendations by assigning a percentage to each of them, in such a way that the total will be 100. The integrated results are presented in the table. Although this result is only the outcome of one vibrant afternoon, it is an important ingredient for the final commandments on a SURF E-humanities policy.

#	Recommendation subject	Score
0	Corpora	21
1	Curriculum	16
6	Web services	14
3	Bond Informatics-Humanities	13
7	Metadata structuring	9
4	Repository research data	6
5	Program library	6
9	Multimodularisation	6
2	Expand helpdesk function	5
8	Authoring tools	3