MEMORANDUM OF UNDERSTANDING
For the implementation of a European Concerted Research Action
Title: An interoperable supranational infrastructure for digital editions (Interedition)

A
Abstract
There is a great need for international cooperation in the development of tools for scholarly digital editing and analysis of literary material. Research groups from individual institutes exchange research results through conferences and journals and report on their methodological advances, but these results and advances often cannot be scientifically verified or tested because the used hardware and software is not available to or easy to use by other researchers. A shared technical infrastructure for the preparation, editing, publishing, analysis and visualization of literary material does not exist. This Action aims to form an international Management Committee of researchers that have a thorough experience in electronic editing and digital text analysis for scholarly purposes in a national context. A series of meetings will be called for researchers in the field of literary research and information technology to meet on the topic of a shared supranational networked infrastructure for digital scholarly editing and analysis. The Action will deliver a roadmap for the implementation of such an infrastructure.

Keywords
tools development, interoperability, usability, scholarly cooperation and collaboration, scholarly digital infrastructure

B. Background
B.1. General background
The area the Action concentrates on is literary research. This is the scholarly study of the form and function of literature and texts in their cultural context. (The closely related discipline of linguistics takes (spoken) language as its subject matter). Literary phenomena are studied in a perspective of international reception and context as a long standing tradition. Comparative literary studies research literary developments as supranational phenomena. A plethora of international bodies exists bound by a common well defined literary research subject.
With the advent of the internet the possibilities of putting literary research materials on the web became apparent and the added value of digitized and digitally shared material for literary research and researchers was quickly acknowledged. Early explorers of the new possibilities for scholarly editing and research in the digital era, like Peter Robinson and Jerome McGann, delivered groundbreaking work by theorizing and showing how digital tools could be used to transform literary research sources and research output into scholarly reliable digital representations for publication and analysis.
However, the preparation and production of digital editions of literary material in a scholarly adequate fashion as well as the development of technical infrastructure and software to support scholarly digital editing has been foremost a concern on an institutional or at most a national level. The tools and practices developed, like Collate, Anastasia, TUSTEP, Juxta, ARCHway's EPT etc., were often conceived to function on an international scale, but are actually primarily operated within an institutional context - contrary to the international setting of literary research in general. Arguably interoperability (which is the possibility to have analysis tools using each other's functions and/or output as automated services) and portability of tools (which is the possibility to put a tool to use for automation within another technical context) have not been a focal point within the development of tools.
This situation does not apply solely to the European region. Scholars and research groups

from other regions are also confronted with a lack of an interoperable networked infrastructure for digital literary research. Canada's TAPOR-project, for example, is addressing similar interoperability and infrastructural issues.
As literary research and scholarly editing become ever more international and collaborative undertakings, the need arises for digital editing and publishing tools that are maintainable, interoperable and usable on a supranational level. The tools in existence lack in support for interoperability, grid readiness, usability and maintainability needed to use them as the international 'hard research infrastructure' for digital literary research.
This Action will combine the expertise of a number of researchers of different countries in the larger EU-region to study the current state of affairs in scholarly digital editing. This Action will identify the common causes and reasons for the perceived lack of an international digital infrastructure for digital scholarly editing.
COST by far offers the most appropriate framework for this Action, as it aims to bring together the existing expertise from different national initiatives in the field of digital scholarly editing. The Action will foremost provide a platform to combine this existing knowledge and will identify common issues and problems. On that common basis the Action will then produce a roadmap for the development of an international shared digital infrastructure for digital scholarly editing. The Action explicitly tries not to identify a new research area. Instead it will point out the common issues and problems underlying the current state of affairs in digital scholarly editing. The Action then will suggest a model to solve this issues in an international cooperation between the researchers and/or research groups.
A common cause for the apparent non-existence of a shared international infrastructure for digital scholarly editing may be found in the lack of an international organizational body geared specifically towards this problem. This Action is precisely aimed at instantiating such a body or platform, built upon existing expertise. As COST is specifically aimed at combining research efforts from different countries in the larger EU-region by initiating platforms for international discourse, it is the appropriate framework for this Action.

B.2 Current state of knowledge

Momentarily Canada's TAPOR-project comes closest to - but is still far from - what Interedition would want to achieve. TAPOR's aim is to provide an integrated workbench of text analysis tools by way of a portal. To achieve this TAPOR provides an online interface to a number of tools that in essence have been around for quite a number of years but did not share formats, programming languages or technical infrastructure. Although relying on somewhat older computing techniques, TAPOR cannot be overestimated in its intention and success of integrating a variety of text analysis tools from different technical and organizational origins.
As for standardization in format, the initiative that has made a considerable and – from a methodological viewpoint – important advance within the community of literary scholars is the Text Encoding Initiative (TEI). This initiative laid down the fundamental methodology and practices for the standardized XML-description of literary works. The TEI-consortium carries it's mission to disseminate good practices in encoding of digital editions out to this very day.
Considering tools for creating digital editions quite a number can be identified, to name but a few: TUSTEP - TUebingen System of TExt processing Programs (http://www.zdv.uni-tuebingen.de/tustep/tustep_eng.html); The ARCHway Project's EPT – Edition Production Technology (http://beowulf.engl.uky.edu/~kiernan/ARCHway/entrance.htm); University of Virgina's Juxta (http://www.patacriticism.org/juxta/?page_id=2); Classical Text Editor by the Commission for Editing the Corpus of the Latin Church Fathers (CSEL, http://www.oeaw.ac.at/kvk/). The majority of these tools is technically oriented and

preoccupied with the collation and representation of variant texts. Though this is a very valid focus (as different representants of the same work do reveal much about the origins and intentions of a work), in most cases it brings along a quite technology oriented approach of the interface of the tools. This technology oriented window on text obstructs the main body of literary text researchers from using the tools that could considerably advance their understanding of text and literature.

Over the last 30 years a vast amount of cd-roms, websites, pdf's etc. etc. have seen the light and could be enumerated under the label 'digital edition'. It is impossible to list all in this proposal, as examples the results of the Canterbury Tales Project (http://www.canterburytalesproject.org/) may be indicated as well as the Biblioteca Virtual Miguel de Cervantes (http://www.cervantesvirtual.com/index.jsp). Almost every digital edition in the corpus that could be compiled, relies on its own methodology, techniques, architecture and information carrier.

As far as can been evaluated at this point there has been no state of the art usability research into the use of all the tools, portals, digital editions etc. that the fore mentioned research has sprouted.

Interedition is innovative in several ways. All tools, formats and applications considered in the preparation of this proposal were originally developed on local infrastructure, and were dependent on local organizational facilities, knowledge and coding practices. Local practices prevented wide spread support and usage of the tools. As indicated TAPOR appears to be the first to provide an integrated toolset on line, but still does this on a centralized local architecture. Interedition wants to initiate an international collaborative for an investigation of the possibilities of decentralized development and hosting of tools for the production of digital editions. Interedition's challenge is to govern the decentralized development of tools – which means that participating researchers and developing organizations may use locally preferred computing technique and architecture – by a set of recommendations and guidelines support by the research community so that the tools developed may be hosted, operated and further developed on any other distinct part of the shared infrastructure. In essence this could be summarized as supporting decentralizing technical approaches to build solutions by finding methodological consensus on an organizational level.

Interedition favors a decentralized approach: decentralized as to where the tools may be built, decentralized as to where they may be operated. To make this possible, Interedition seeks to formulate recommendations for protocols for interoperability of tools. In this way the specifics of implementation are free to local builders, but the general interoperability over a shared network can still be guaranteed.

To reach these goals Interedition opts for a radical different approach to initiating the development of solutions. Interedition does not want to prescribe how solutions should be built and by whom. Rather Interedition will identify key (or core) functionalities and will indicate which functionalities should interoperate over a shared infrastructure, without prior stating means and method. Interedition then asks its associated partners for a best solution trajectory. As developing progresses on local levels, Interedition only monitors and guards functional interoperability. It does not impose technical rules or requirements, which are best left to local cooperating developers.

As usability is a general issue to the existing solution, a major role for Interedition monitoring the cooperation between the associated partners will be in consideration to unbiased and independent usability tests. In this respect Interedition will be the first initiative that lets development of tools interact with intended end users (meaning non-technical oriented literary scholars) right from the outset. This direct involvement of end users monitored by Interedition in the form of concrete usability studies conducted by an independent associated partner will guard usability for the indented community of end users at a maximum level.

B.3 Reasons for the Action

The Action is mainly aimed at European scientific technological advance. Its immediate benefits are information exchange between scholars and programmers working in different countries at digital scholarly editing, a substantial increase in scientific discussion about methods and techniques, and the possibility to discuss how to solve present problems. Future benefits will be an expected increase of scientists making use of digital tools for their scholarly textual editing and literary analyses, more reliable digital tools through development cooperation; better supportability and maintenability of the produced programming code and tools through shared knowledge and methodological approach, a general increase in (international) cooperation on creating digital editions; an increase in cooperation on the production of edition technology; and a general advancement in numbers of editions published. The advancement in the number of editions is expected not only because of the expected growth of users, but also because the tools will be open to use for any researcher. Scholars working in e.g. 'smaller languages' who would otherwise be limited in publishing their editions in print, can then put their scholarly work on the web, unhinderd by economical criteria where scientific criteria should reign.

The means that are needed for the immediate benefits are the possibility to organize and attend international expert meetings to discuss the problem. The means needed for the future benefits are the roadmap or manual that will be drawn up during the Action.

B.4 Complementarity with other research programmes

The Action has complementarity with the planned projects CLARIN and DARIAH. CLARIN (Common Language Resources and Technology Infrastructure, http://www.mpi.nl/clarin/) aims to establish an integrated and interoperable research infrastructure of language resources and its technology. It focuses on linguistic reseach, while the Action focuses on textual scholarship and literary research. The linguistic field is related, but needs a different approach where tools and infrastructure are concerned. Some tools will be necessary for all fields. It will be useful to contact CLARIN in an early stage of the Action to check where the projects could benefit from each other. DARIAH (Digital Research Infrastructure for the Arts and Humanities, http://www.dariah.eu/) will provide a coordinated infrastructure for supporting preservation of cultural heritage in Europe and providing access to research material from the humanities. The topic on which to contact DARIAH in an early stage of the Action may be dissemination, as one of DARIAH's key processes is to enhance digital scholarship in the humanities and arts across Europe. Dissemination of the results of Interedition could well be easily included in integrated dissemination actions.

The HyperNietzsche Project (http://www.hypernietzsche.org/base.html) is a relevant linked project because it creates a digital infrastructure for the collection of materials concerning one author. Interedition will make use of the experiences and expertise of the project. HyperNietzsche users may be found willing to be end users for the Interedition example tools and in this way provide the Action with input on which to develop the roadmap.

C. Objectives and benefits

C.1 Main/primary objectives

The primary objective of the proposed Action is to produce a 'roadmap' or 'manual' conceptualizing the development of a technical infrastructure for collaborative digital preparing, editing, publishing, analyzing and visualizing of literary research materials.

C.2 Secondary objectives

The secondary objectives are

(1) to accompany the roadmap with a set of (at least 5) proof of concept web services functioning as an example implementation of the technical infrastructure model, this being the best way to show end users in what ways they will benefit from adhesion to the Roadmap;
(2) the development of concept proposals for additional funding from other funding bodies, to guarantee development of the tools referred to in the Roadmap and to meet with new parallel requests by end users from different countries;
(3) attain a wider European involvement with partners from (esp. eastern) European countries which are up till now not represented in Interedition.

C.3 How will the objective be achieved?
The paramount condition to attain the primary objective, producing a manual that will be endorsed by a larger community of scholarly researchers in the field of (digital) literary research, is the possibility for a sufficiently large number of experts from that field to meet and discuss the concepts and requirements of the infrastructure. Therefore one objective of the Action should be to create enough occasions for such researchers, as members of the Management Committee or a Working Group, to meet on this subject. These meetings should also include a number of IT-researchers, well-informed on the particular needs of the literary research community, to guide and support the literary domain experts in inferring viable technical recommendations. The secondary objectives will be the prime goals of three of the work packages, cf. the description under E.

C.4 Benefits of the Action
The Roadmap or Manual that will be delivered should describe an interoperability protocol for on line services that support collaborative scholarly editing and publishing. The manual will also describe a development and maintenance strategy for tools delivered to the infrastructure and will include research results addressing the usability issues specific to a digital literary research environment. The benefits of this Roadmap will be:
(1) general availability of a guide for further interoperable technical research and development
(2) which will help to solve the problem of sparse knowledge sharing between the research groups relating to digital editing and text analysis in Europe
(3) and which will solve the lack of coordination between these groups
(4) and will help to avoid duplication of tools;
The set of (at least 5) proof of concept web services will help Interedition to:
(5) reach out to scholars in order to improve their possibilities for doing innovative research
(6) and to give them the opportunity to share and combine their separate research, leading to new collaboration
(7) ideas for new collaboration will also be coordinated by Interedition, resulting in new European projects
(8) and an increase of end users of digital editing and text analysis tools all through the European region.

C.5 Target groups/end users
Target groups for the Action are both scholars of literature and those researchers/programmers who are working or planning to work on the development of tools for those scholars, from as broad a geographic area (the wider European context) as possible.
(1) Researchers of literature: scholars who are already using digital tools, to stimulate them to share their experiences and tools and have them test the solutions Interedition is working towards,

(2) Researchers of literature: scholars who are not using digital tools yet, to have them share their experiences and frustrations and also test the solutions Interedition has in mind. Usability (userfriendliness) here is of main importance.
(3) Researchers and/or programmers who are currently building digital tools for scholarly editing: to focus them implicitly on interoperability (technical guarantee for reusability of the tools they are building) as well as on usability (userfriendliness of the tools)
(4) From the wider European context: reach out to all scholars and programmers in Europe who may benefit from the Interedition COST Action.
The likely end users belong to the first target group. Interedition aims to enlarge the group with end users from the second target group, users of digital editing and text analysis tools on a European scale with scholars who up till now did not make use of them

D. Scientific programme

D.1 Scientific focus

The most important research tasks of the Action are:
(1) to describe the current state of the art in tools development for digital scholarly editing and text analysis,
(2) to identify the technical aspects for which agreement is needed to guarantee that the tools to be developed are usable in any other technical, institutional or national context and to arrive at technical agreement on those points (e.g. import and export format of the files of the text to which researchers would like to apply certain tools),
(3) to identify usability aspects which have to be taken into account for the targeted end users, who cannot be expected to have the same ease in using digital tools as representatives of the 'hard' sciences, and
(4) to develop an example implementation of the envisioned technical infrastructural model, consisting of a small set of web services in a prototypical form (necessary for dissemination to esp. end users without much experience in using digital tools, and reaching out to new partners).

The expected scientific impact is a considerable advancement of international collaborative literary research. It will be possible to revive a number of research endeavors that are dormant by lack of technical infrastructure. The use of delivered tools within the research community will have a significant methodological impact, resulting in promoting reproducibility of measurements and research results, also resulting in a more algorithmic modeling approach to literary research.

The human and technical means to achieve the objectives of the Action are the availability of key researchers and programmers in the field for meetings at different locations in Europe; it would enhance the productivity of those meetings were it possible for the attendants to access work stations and developers' machines to actually test technical solutions that are being discussed; access to the internet is of course necessary as well.

D. 2 Scientific work plan – methods and means

The work plan for the Action will contain the following items: (i) the identification of shared and reusable functionality, as part of a larger plugin and web service architecture; (ii) the selection of an adequate stack of web services and grid protocols, together with (iii) the creation of guidelines for the use of these protocols; (iv) the definition of interfaces for the reusable functions; (v) the selection of existing software components suitable to be ported or 're-wrapped' for delivery in a web services based architecture; (vi) identification of promising

services for proof-of-concept implementation, which will be extended as the needs arises. The items will be divided among four (or if needed five or six) Working Groups.

To some extent the problems to be tackled are technical. The main potential problem is formed by the rapid changes in grid and web services infrastructure. Architectural choices should therefore strike a right balance between experiment and stability. Non-technical problems include different national and institutional traditions and methods in literary studies. The editing environment to be created should not enforce a single editorial method but rather inspire exchange and cooperation of different methods.
In successfully dealing with these problems, a platform for edition and literary studies will be created that is innovative in being open to extension, by partner institutions and non-partners alike. This program will thus create the foundation for a component-based, flexible platform that will facilitate the use of electronic resources in literary studies. It will truly be a shared architecture for a shared heritage.

In deciding on which technical approach is best, the input of end users is of the highest importance. As to method and means, this implies that experiments play a key role in the gathering of information by the Working Groups. Feed back from end users, especially about usability, will be taken into account in making decisions for the roadmap.

The Working Group 'Prototyping' will provide at least five example tools which will be provided to the end users found willing to test and therefore are key to the experimental setting of the Action. The Working Group 'Strategic IT Recommendations' will closely watch the experiments and will input their evaluations in recommendations for the Working Group 'Roadmap for Interoperability', whose objective is to generate the roadmap for building the digital infrastructure for scholarly editing and textual analysis in a way that will guarantee international collaboration and prevent duplication of tool building in different countries and institutions.

The fourth Working Group, 'European Dimension', will actively reach out to other European countries, disseminating the aims and results of the Action to as broad an audience as possible. The aim of this Working Group is to invite at least one new partner for each half-yearly meeting, thus enlarging the group of European partners steadily throughout the Action. The Working Group 'European Dimension' will also coordinate new international grant proposals.

The Action will coordinate the scientific discussion on all the topics mentioned. The discussions will take place by e-mail and on the Interedition discussion forum, but can only lead to concrete decisions in regular meetings of the Project Manager, members of the Management Committee and coordinators of the Working Groups.

E. Organisation

E.1 Coordination and organisation

The COST Action is implemented through a concerted action, which means that the research is carried out in and financed by the participating countries, while COST provides the necessary coordination. This means that the preparational work and the actual programming and testing of the example tools will be done by all scholars and programmers who take part in their institutional roles, and that the COST Action will take care of the organization of the meetings needed to evaluate the research and experiments done, discuss and plan ahead, and

generate the roadmap with recommendations for building an internationally shared digital infrastructure for scholarly editing and textual analysis.
A Management Committee will be installed with the required representatives from each signatory country and from the non-COST institutions. The Management Committee will supervise and coordinate the implementation of the Action. A Project Manager will be appointed to take care of daily business. The Management Committee will establish the necessary Working Groups (more about them in E.2) and appoint coordinators to them. The different types of coordinative (research)work will be distributed over the Working Groups. MC and coordinators together will draw up detailed plans, arrangements for the distribution of tasks, and defining methods for the different phases of execution of the Action, including the planning and follow up of the Actions budget.
The Project Manager will be working closely with the coordinators of the Working Groups. They will convene on a half year basis to evaluate current developments, to plan ahead, and, if necessary, to adjust plans. They will also be responsible for the staging and continuous updating of the Interedition website. The website is meant to inform all partners and scholars and programmers from the target groups of the status of the project and will also enable discussion, the yeast of which will be taken into account at the meetings. At each of the half year meetings a new interested partner should be invited, preferably from a non-Western-Europe organisation.
At the kick-off meeting it will be decided whether it is useful to establish an extra Working Group, functioning as a quarter master for the other working groups taking planning and initiating the other working groups as its core tasks. The Management Committee initiating this Action calls for open Working Groups and wants to welcome any humanities researchers or research groups that have an interest in exploring the possibilities to develop the research infrastructure.
There will be four Milestones, the first a year after the start of the project and the next each a year later. The Milestones will consist of the presentation of a concept-version of the yearly expanded Roadmap and of the release of one or two example tools (cf. F, Timetable, for a more exact planning). At the end of the Action, the tools will be in beta-form and the test results of end users have been evaluated and taken into account for the guidelines presented in the Roadmap. The Roadmap will be in its final form.

Milestones

12 months: Roadmap: Concept version, 1, with outlined chapters and topics to be described
Release of Example Tool 1, version alpha

24 months: Roadmap: Version 1.1
Release of Example Tool 1, version beta
Release of Example Tool 2, version alpha
Release of Example Tool 3, version alpha

36 months: Roadmap: Version 1.2
Release of Example Tool 2, version beta
Release of Example Tool 3, version beta
Release of Example Tool 4, version alpha
Release of Example Tool 5, version alpha

48 months, end of Action: Roadmap: Version 2
Release of Example Tool 4, version beta
Release of Example Tool 5, version beta

The members of the Management Committee will play a pivotal role in the coordination of national research. They will discuss and implement plans for possible common research teams, conferences and workshops, short-term scientific missions or exchanges between developers' groups. The signatories have already come up with the following ideas for workshops:

- (Collaborative) Digital Text Editing
- Digital Text and Image Annotation
- Electronic transcription of notebooks and modern manuscripts
- Scholarly Annotation Exchange (SANE)
- Automatic comparison of versions of the same text (Text Collation)
- Electronic edition APIs
- TEI P5 XML for scholarly editions
- Schema Design and Documentation (e.g. with TEI ODD)
- XML Transformation, Querying and Processing (XSLT/XQuery)
- Subversion repository management
- Text Analysis Tools
- Tools for Vocabulary Analysis in Literary Texts

E.2 Working Groups

We intend to form four working groups: 'European Dimension', 'Prototyping', 'Strategic IT Recommendations', and 'Roadmap for Interoperability'.

The Working Group 'European Dimension' has three tasks.
(1) To reach out to ongoing research that might be usefull to the objectives of the other Working Groups (e.g. inventory of existing digital tools for Scholarly practices in the wider European area);
(2) to disseminate the results of the other Working Groups to an audience as large as possible;
(3) and to initiate new grant proposals on an international scale.

The Working Group 'Prototyping' has as its tasks
(1) to do applied research into interoperable tools, and
(2) to feed its results into the Work Group 'Strategic IT Recommendations'.

The Working Group 'Strategic IT Recommendations'
(1) will evaluate the information of the Working Group 'Prototyping'
(2) will make use of the inventory of research practices and wishes by the Working Group 'European Dimension' and
(3) will combine the evaluations of theory and practice in actual recommendations on IT which will then be fed into the Working Group 'Roadmap for Interoperability'.

The Working Group 'Roadmap for Interoperability' will use the input of the Working Groups 'Prototyping' and 'Strategic IT Recommendations' to generate the actual roadmap that will hold recommendations for building an internationally shared digital infrastructure for scholarly editing and textual analysis.

If necessary, other Working Groups will be established (cf. E.1).

E.3 Liaison and interaction with other research programmes
At the start of the Action the Management Committee and the Project Manager of the HyperNietschze Project, (also a ESF/COST funded project) will see whether it is efficient to cooperate as to practical things such as sharing address databases, dissemination, journals, conference, logistics etc. The Working Group 'European Dimension' should continuously investigate the existence of projects which could be 'natural' partners for the Interedition initiative. Interedition wants to be an open platform for cooperation and will actively reach out to other projects. One of the objectives of this Working Group is to invite a new (Eastern-European) partner to each half-yearly meeting.

E.4 Gender balance and involvement of early-stage researchers
This COST Action will respect an appropriate gender balance in all its activities and the Management Committee will place this as a standard item on all its MC agendas. The Action will alco be committed to considerately involve early-stage researchers. This item will also be placed as a standard item on all MC agendas.
It is commonly known that young researchers play a key role in changing traditional approaches in science. This is especially so in the case of new media and the use of digital tools for research purposes. For Interedition to succeed it is necessary to involve young researchers and providing them with the knowledge and the tools to help them to establish a digital turn in textual scholarship and literary research. In several countries the majority of doctoral and post-doctoral researchers in textual scholarship have been women over the last ten years. It is possible that Interedition will have to take special care to guarantee enough involvement of male researchers.

F. Timetable
The Action will have a duration of four years. The Milestones of the Action in its totality are described under E.1. For each of the four Working Groups, the following timetable is suggested.

Working Group 'European Dimension'
Year 1: inventory of related projects on a European and global level
Year 2: start of (coordination) of dissemination of first results (e.g. arranging that concrete research results are presented at suitable conferences)
Year 3: further coordinative tasks of dissemination and reaching out to new partners
Year 4: actively start spreading recommendations and results to interested policy makers and organisations, disseminate results to interested partners and projects not directly associated with/to Interedition, publish advise on the continuation of dissemination after Interedition finishes.

Working Group 'Prototyping'
Year 1: in the kick-off meeting , take a decision on the first three tools to prototype; release the first version of one of these at the first Milestone
Year 2: release a revised version of the tool(s) delivered in the year before, based on evaluation of end users and release the first version of two more tools
Year 3: as year 2

Year 4: release a revised version of the tools delivered in the year before; evaluate the results of tests by end users, and keep the Working Group ‘Strategic IT Recommendations’ informed and on track.

Working Group ‘Strategic IT Recommendations’
Year 1: provide input for the Working Group ‘Prototyping’ on the example tools to develop
Year 2: theoretical research
Year 3 and 4: documents and recommendations

Working Group ‘Roadmap for Interoperability’
Year 1 and 2: keep the other Working Groups on track through discussions
Year 3: start writing the Roadmap
Year 4: finishing the Roadmap

The COST Action will cover the organization of the meetings in which the Milestones will be presented. The preparational work and the actual programming and testing of the example tools will be done by all scholars and programmers in their institutional roles.

G. Economic Dimension
The following COST countries have actively participated in the preparation of the Action or otherwise indicated their interest: <BE,HR,FR,DE,NL,GB>. On the basis of national estimates, the economic dimension of the activities to be carried out under the Action has been estimated at 9 Million € for the total duration of the Action. This estimate is valid under the assumption that all the countries mentioned above but no other countries will participate in the Action. Any departure from this will change the total cost accordingly.

H. Dissemination plan
H. 1 Who?
The target audiences for the dissemination of the results of the Action are other researchers working in the field; other research frameworks; research Institutes and Academia; Standards Bodies (e.g. the TEI-consortium). Of specific importance for textual scholarship are the subscribers to mailing lists such as Humanist, Digital Medievalist, Digital Classicist, Antiquist, Corpus Linguistics lists and related lists. Consortia working towards integration of language and text corpora such as CLARIN, DARIAH, Digital Historical Corpora, Poetics and Linguistics Association etc.
Further target audiences are European level policy makers; Government policy makers, and regional planners and policy makers.

Belgium (Flanders):
-FWO (‘Fonds voor Wetenschappelijk Onderzoek’ – Flanders) The central national body in Flanders (Belgium) for funding of research across a.o. the humanities.

Croatia:
Croatian Ministry of Science

France:
Centre National de la Recherche Scientifique (CNRS)
Ministère de l'Enseignement supérieur et de la Recherche

Germany:

Deutsche Forschungsgemeinschaft (DFG)
The German Federal Ministry of Education and Research (BMBF)

Great-Britain:
-AHRC: the Arts and Humanities Research Council. The central national body for funding of research across the humanities, particularly (but not exclusively) within universities. Within the AHRC, a strategic initiative for ICT focusses on the impact of digital methods for research.
-JISC: The Joint Information Systems Committee. They coordinate a wide range of digital initiatives across universities, notably in e-Science.

The Netherlands:
-NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek – Netherlands Organisation for Scientific Research). The central national body in the Netherlands for funding of research across a.o. the humanities.

H. 2 What?
The dissemination methods we intend to use are
- postings of general information on relevant public websites, directing visitors to the Interedition website
- posting of working documents on the Interedition website, partly freely available and partly on a password protected subpage;
- in relation to the Interedition website: an electronic communication network (e.g. an internet discussion forum);
- publications: state of the art reports, interim reports, case study reports, proceedings, guidelines, manuals, final reports;
- events: workshops, seminars and conferences organised by the MC, contributions to other national and international conferences and symposia;
- articles in peer-reviewed scientific and technical Journals;
- non-technical publications.

H. 3 How?
The dissemination of the results will be tuned to the different target groups and end users (cf. C.5). Although publications are very important, events and workshops are equally or even more important. For the dissemination of results to target group (1), scholars who are already using digital tools, we will organize presentations on big international humanities computing conferences (e.g. the yearly Digital Humanities conference). In the presentations the focus will lie on technical, theoretical and methodological innovations in research.

Target group (2), scholars who are not using digital tools yet, will be approached with presentations on more general international conferences (e.g. the yearly International Medieval Congress in Leeds). In their presentations for this target group, scholars participating in the Interedition project will focus on presenting the advantages of the new digital tools that are being developed. Examples of their research and examples of the use of the tools have proved to be the best instruction of the problems. Dissemination will also take place on a national level, in any subdiscipline in the field of literary research and textual scholarship. Especially young researchers will play an important role in this method of dissemination. These presentations lead to discussion with researchers in the field which may yield useful information for the progress of the Action and may lead to new partners.

Presenters will therefore report back after these events and the results from discussions will be taken into account for the dissemination plan of the Action.

Target group (3), researchers and/or programmers who are currently building digital tools for scholarly editing, will in most cases already be included in the Interedition network. If projects and persons unknown to the Action before are met at conferences or elsewhere, they will be informed about Interedition and, if relevant, invited to join the Action. The same approach goes for target group (4), scholars and programmers in Europe who may benefit from the Interedition COST Action.

The dissemination to institutions, standards bodies, etc., will concentrate either on the technical side of the tools and the problems and solutions, and/or on the research process, showing how scholars approach tools, why up till now use of the tools was problematic and how Interedition wants to solve these problems.

Special opportunities for dissemination may be:
-The 150th anniversary of the publication of Darwin's Origin of Species will be marked by a new digital variorum, which we are preparing. This anniversary will attract massive publicity.
-2008: Royal Netherlands Academy of Arts and Sciences (Amsterdam, the Netherlands), 200th anniversary
-2010: Centre for Scholarly Editing and Document Studies of the Royal Academy of Dutch Language and Literature (Gent, Belgium), 10th anniversary
-2011: Royal Academy of Dutch Language and Literature (Gent, Belgium), 125th anniversary

Part II –Additional Information NOT PART OF THE MoU

A. List of experts

-consulted during the drafting of the proposal and who have already expressed interest in participating in the Action (highlighted: experts that might be part of the Management Committee).

Dr. Karina van Dalen-Oskam
Huygens Institute (KNAW)
PO Box 90754, NL-2509 LT Den Haag, The Netherlands
karina.van.dalen@huygensinstituut.knaw.nl
+31-70-3315875

"Joris van Zundert" <joris.van.zundert@gmail.com>
"Ronald Dekker" <ronald.dekker@huygensinstituut.knaw.nl>,
"Peter Boot" Peter.Boot@huygensinstituut.knaw.nl

Dr. Barbara Bordalejo
Institute for Textual Scholarship and Electronic Editin, University of Birmingham
Graduate Institute of Theology and Religion, University of Birmingham, Elmfield House, Selly Oak Campus, Birmingham B29 6LQ United Kingdom
B.Bordalejo@bham.ac.uk
+44-121-4158441

Dr. Peter Robinson
Institute for Textual Scholarship and Electronic Editing, University of Birmingham
Graduate Institute of Theology and Religion, University of Birmingham, Elmfield House, Selly Oak Campus, Birmingham B29 6LQ United Kingdom
P.M.Robinson@bham.ac.uk
+44-121-4158441

Prof. Dr. Fotis Jannidis
Technical University of Darmstadt
Institut für Sprach- und Literaturwissenschaft, Hochschulstrasse 1 (S 1 03 / 182) , D-64289 Darmstadt, Germany
jannidis@linglit.tu-darmstadt.de
+49 - 6151 - 16 2597

"Thorsten Vitt" <vitt@linglit.tu-darmstadt.de>,

Mr. Edward Vanhoutte
Centre for Scholarly Editing and Document Studies (KANTL)
Koningstraat 18, B-9000 Gent, Belgium
edward.vanhoutte@kantl.be
+32 (0)9 265 93 50

"Ron Van den Branden" <ron.vandenbranden@kantl.be>,

Dr. Dirk van Hulle
University of Antwerp, dept. of literature studies
Departement Letterkunde, Letteren en Wijsbegeerte, Universiteit Antwerpen, Stadscampus R120, Rodestraat 14, B-2000 Antwerpen, Belgium
dirk.vanhulle@ua.ac.be
+32- 3 220 4550

Dr. Paolo D'Iorio
l'Institut des Textes et Manuscrits Modernes, ENS,
45, rue d'ULM, Paris F-75005, France
diorio@ens.fr
+.33-1-44323233

Mr. Martin Wynne
Oxford Text Archive,
Oxford University Computing Services, 13 Banbury Road, Oxford OX2 6NN, UK,
martin.wynne@oucs.ox.ac.uk
+44-1865-610600

State and University Library Goettingen
Dr. Heike Neuroth
Niedersächsische Staats- und Universitätsbibliothek Göttingen, D-37070 Göttingen, Germany
neuroth@sub.uni-goettingen.de
+49-551-39-3866

"Andreas Aschenbrenner" <aschenbrenner@sub.uni-goettingen.de>,
"Mittelbach, Jens" <mittelbach@sub.uni-goettingen.de>,

Dr René van Horik
DANS
P.O.-Box 93067, NL-2509 AB, Den Haag, The Netherlands
Rene.van.Horik@dans.knaw.nl
+31-06-23297389

Mr. Neven Jovanoviÿ
Department of Classical Philology at the Faculty of Humanities and Social Sciences, University of Zagreb
neven.jovanovic@ffzg.hr

James.Cummings@computing-services.oxford.ac.uk
"Adrian Packel" <packel@perseus.tufts.edu>,
"Marilyn Deegan" <marilyn.deegan@wanadoo.fr>,
"Robert Kummer" <rokummer@gmail.com>,

Not yet contacted:

Prof Marko Tadic
Institute of Linguistics, Faculty of Humanities and Social Sciences, University of Zagreb;
Prof Damir Cavar / Dunja Brozovic Roncevic, Institute of Croatian Language and Linguistics, University of Zagreb;

B. Additional information

History of the proposal

In the Fall of 2006 the department of e-Reseach of the Huygens Institute KNAW took the initiative to invite a small group of experts on digital text editing for a two-day meeting in The Hague, the Netherlands, to discuss the possibilities of sharing the technical workload for the development of tools for this specific scholarly activity. The invited experts accepted the invitation and immediately proposed still other experts to join the group. During the meeting, which took place on 25 – 26 January 2007, it was agreed to continue the discussion on a more formal and continuous level and the Interedition Expert Group was formed. It was decided to actively search for possibilities for funding more meetings, which resulted in the COST-proposal for the implementation of a European Concerted Research Action. More interested partners were asked to join and did join after the meeting. The topic that was discussed most in the first 10 months of the initiative was the collation of texts, in relation to the planned migration of Peter Robinson's programme Collate.

Recent publications

For the topic the Action wants to address: Peter Robinson's article 'Current issues in making digital editions of medieval texts—or, do electronic scholarly editions have a future?' In Digital Medievalist 1.1 (2005) , http://www.digitalmedievalist.org/article.cfm?RecID=6

For an overview of reseach in humanities computing: A Companion to Digital Humanities, ed. Susan Schreibman, Ray Siemens, John Unsworth. Oxford: Blackwell, 2004, http://www.digitalhumanities.org/companion/. (e.g. chapter 9: Literary studies, chapter 22: Electronic Scholarly Editing, etc.)

Overview of current and/or planned research of partners and their view on how Interedition will be of use in the future.

Peter Robinson, Institute for Textual Scholarship and Electronic Editing, University of Birmingham

In the next years, the central aim of our research is the creation, in partnership with others, of systems to allow many different scholars to collaborate in dispersed cooperative editions, in an efficent and fruitful manner. This requires creation of software to enable this collaborative work, and agreement on standards and methodologies to underpin this software. Our fundamental needs are for systems for web-based transcription, collation, analysis and commentary, and for 'live' updating of dynamic editions as different scholars make different contributions. Especially, we see it as essential that the scholars do not, as at present, need to be tightly bound into a single controlling project, with an elaborate and dedicated infrastructure. Rather, we imagine a scheme where scholars may collaborate based only on their knowledge and abilities, and a shared vision of what scholarly editing is. The common structures of the standards and methodologies underlying the collaboration will remove the need for the costly and exacting project-based control schemes now in place.

We are well placed to test and deploy these developing schemes over the next years. We are working with editors in fields ranging from ancient Egyptian, through biblical and medieval texts to modern authors. We have already commenced work on some of these tools, in the form of the CollateX application, and are collaborating with others making other of these tools. We plan at least one workshop a year, to train other scholars in the use of these tools, and we plan also to offer support on our own server to give the editions made by these tools a home for the foreseeable future.

Dirk van Hulle, University of Antwerp – Centre for Manuscript Genetics

The current research project is entitled 'Manuscript Genetics: Combining image-based and text-based approaches to electronic editing'. On the one hand, digitization projects like HyperNietzsche are centred around images: scans of documents or digital facsimiles of manuscripts are the starting point. The great advantage of this *image-based approach* is that it works with stable and unique sigla that identify a particular zone on a particular document. On the other hand, a *text-based approach* starts from transcriptions encoded in XML. The research at the Centre for Manuscript Genetics at the University of Antwerp investigates whether both approaches mutually exclude each other or whether they can be combined. Different authors have different writing habits. Some jottings never make it into the final version, other jottings do not even belong to a version. These are usually referred to as paralipomena. The current research aims at a solution to make the translation between document-oriented and text-oriented approaches as smooth as possible, to fully integrate the paralipomena in the edition, to link them to external sources and to indicate their place within the temporal dimension of the writing process.

Planned research: 'Reading Notes: From marginalia to the published text'. The aim of this research is to devise a flexible electonic infrastructure that can accommodate and visualize complex writing processes for different kinds of writing – both scientific and literary – from the authors' reading (including their personal library and reading notes) to the publication of their works. Both the current and the planned research aim at interoperability of tools.

James Cummings, Oxford Text Archive

The Oxford Text Archive (OTA), as part of the Oxford University Computing Services (OUCS) is directly involved with numerous projects and services hosted at OUCS as well as some research projects which the OTA itself hosts. As an archive we have been involved with the Text Encoding Initiative (TEI) since its inception. Not only is the University of Oxford one of the 4 institutional hosts of the TEI, but Dr James Cummings, the OTA's Research Officer, has been elected to the TEI Technical Council multiple times. He is currently working hard on the next major revision of these internationally-used Guidelines for text encoding. In addition James is also involved in other projects such as: a pan-European project for making manuscript description metadata interoperable, the Arts and Humanities Data Service: Literature, Languages and Linguistics, an Oxford project digitising the diaries of William Godwin, a scholarly edition of The Conversion of St Paul, as well as research on legacy data migration, and repository management. He has provided schema for a number of projects such as an EU-funded project 'System Aided Compilation and Open Distribution of European Youth Language' (SACODEYL). The OTA is also the main UK partner in the pan-European 'Common Language Resources and Technology Infrastructure' project (CLARIN). The OTA helps to disseminate the British National Corpus, another project located in OUCS alongside projects such as the JISC-funded Open Source Software Advisory Service (OSS Watch) and Institute: Arts and Humanities.

Edward Vanhoutte, Centre for Scholarly Editing and Document Studies

The Centre for Scholarly Editing and Document Studies (CTB) of the Royal Academy of Dutch Language and Literature (KANTL, Ghent, Belgium) focuses on producing electronic scholarly editions of valuable literary source materials. Like their non-electronic counterparts, electronic scholarly editions are the result of text critical research, aiming to investigate the origin and evolution of a literary work through comparison and evaluation of different textual versions of the work. Meaningful electronic editions make use of standardised representation formats like XML (eXtensible Markup Language) and markup languages like the TEI (Text Encoding Initiative) guidelines for representing and transcribing the source materials. A collation phase, which compares and records the differences between the text versions under consideration, provides the base for an interface to the electronic edition, that –by virtue of the versatility and manipulability of XML–can highly enrich the users' engagement with the textual tradition.

Although transcription and annotation of literary source materials will remain the realm of a human editor, we believe the collation phase is susceptible to automation. However, until now, automatic text collation for humanities research purposes has remained an under-investigated research area (notable exceptions like the Collate software aside). Automatic version control systems for source files in software development processes could provide an intuitive starting point, but currently are too unsophisticated to deal with XML structures or complex textual variation at word level. Text critical research is in need of a generic open source tool that enables comparison of an unlimited number of XML texts, can represent these results in an XML format, and preferably can be flexibly used in a number of use scenarios (standalone, built-in in electronic editions, web-based,…). The CTB wants to play an active role in the development of such a tool and believes the Interedition project an ideal basis for a coordinated effort.

Neven Jovanovic, University of Zagreb

The project Croatiae auctores Latini: A digital corpus of Croatian Latin writers deals with designing and building a corpus of texts written in Latin by Croatian authors. This corpus

will be freely accessible over the internet, offering materials for several types of study, research, and scholarship. The question connected with Interedition project is: how to present a large number of little known texts in a language which (though once an international medium of expression) today belongs to the category of 'cultural heritage'.
How should a database of Neo-Latin literature look like? How will these texts be read? By whom? To what purpose? How can we connect the people reading the texts, help them to build a community of researchers? How can these texts help people learn more Latin, learn more about forgotten aspects of European culture?
First steps in building a corpus are simple: the texts have to be selected, obtained, and made computer-readable (digitized). But here the task of editors is just beginning. A database is worth to people who know what is inside; so secondary information about the texts, authors, genres, periods, places also has to be encoded, stored, made accessible; the texts need to be ordered, put in some kind of context.
The Interedition project will be crucial in supporting these tasks of the editors; it will also help bring the texts and the whole corpus to the maximum 'user-friendly' level possible. It would also improve the database prototypes according to opinions of users --- both experts (in Latin and Neo-latin literature), learners (i. e. students), and interested laypersons (e.g. somebody who wants to know more about Croatia, or about Renaissance theology).

Fotis Jannidis, Technical University of Darmstadt
The aim of Interedition partly resembles the German TextGrid project.. TextGrid aims to create a community grid for the collaborative editing, annotation, analysis and publication of specialist texts. It thus forms a cornerstone in the emerging e-Humanities. Text sciences research the complex processes in the genesis of literature and their contexts. Despite modern information technology and a clear thrust towards collaboration, text scientists still mostly work in local systems and project-oriented applications. Current initiatives lack integration with already existing text corpora, and they remain unconnected to resources such as dictionaries, lexica, secondary literature and tools. However, this integration and interconnection bears a wealth of opportunities.
Integrated tools that satisfy the specific requirements of text sciences could transform the way scholars process, analyse, annotate, edit and publish text data. Working towards this vision, TextGrid aims at building a virtual workbench based on e-Science methods.
The installation of a grid-enabled architecture is obvious for two reasons. On the one hand, past and current initiatives for digitising and accessioning texts already accrued a considerable data volume, which exceeds multiple terabytes. Grids are capable of handling these data volumes. Also the dispersal of the community as well as the scattering of resources and tools call for establishing a Community Grid. This establishes a platform for connecting the experts and integrating the initiatives worldwide. The TextGrid community is equipped with a set of powerful software tools based on existing solutions and embracing the grid paradigm.
It is expected that the network of technical collabrators that Interedition wants to establish will help TextGrid to accomplish tools that would otherwise not have been possible to realize.

Karina van Dalen-Oskam, Huygens Instituut KNAW
The Huygens Institute carries out research into both the history of literature and the history of science. The rapid development of information technology plays an important role in this research, in that it advances the digital techniques that are used for deciphering as well as analyzing texts. The department of e-Research specializes in humanities computing research (e.g. stylistics and authorship verification) and in the development of the tools needed for this research and for digital editing of (literary) texts. In this, usability of the developed tools proofs extremely important for the success of the tools (meaning they are actually used by the

intended scholars). Because guaranteeing a userfriendly programme takes a lot more work than just a working tool which can be used by technically more advanced researchers, the department of e-Research thought it wise to look for other institutions having the same problem, hoping that opportunities would arise to share the development of tools that are needed by scholars in different countries, working on different languages. Through Interedition the Huygens Institute hopes to contribute to enhancing the possibilities for digital editing and text analysis on a broader, European scale.