

No Claims for Universal Solutions - Possible Lessons from Current e-Humanities Practices in Germany and the UK

Tobias Blanke
Arts and Humanities e-Science Support Centre
King's College London
Drury Lane
tobias.blanke@kcl.ac.uk

Andreas Aschenbrenner
State and University Library Goettingen
Germany
aschenbrenner@sub.uni-goettingen.de

Marc Küster, Christoph Ludwig
University of Applied Sciences Worms
Germany
kuester@fh-worms.de
ludwig@fh-worms.de

Abstract

The aim of this paper is to review some emerging successful practices in e-Humanities from the perspective of Germany and the UK. All the reviewed projects work on the collaboration of science and computing practices with arts and humanities. This paper argues that claims to universality are the wrong way of promoting future research in e-Humanities. We need to open local spaces in which arts and humanities researchers can engage with e-Science tools and methodologies in a way they are used to engage with other research in their domain. This suggests to 'embed' e-Science in arts and humanities research practices.

1. Introduction: e-Science in the arts and humanities

The cradle of e-Science has been up to now clearly in sciences disciplines such as high energy physics or astronomy. However, the last years have seen an expansion of e-Science into other disciplines like social sciences and humanities. The attempt to export e-Science to these new disciplines, in order to build novel cross-disciplinary, international research infrastructures, has shown a remarkable will to experimentations within the e-Science community. The paper will cover the expansion into arts and humanities. It discusses this from the perspectives of UK and German humanities e-Science work, pointing to early successes but also problems.

It can be seen at the first success of the attempt to adopt e-Science technologies to arts and humanities research that a new discipline has been emerging internationally. In many countries in Europe this new discipline is called e-Humanities, researchers in the US typically prefer the term 'cyberinfrastructure for the Humanities' [1]. The UK programme is called Arts and Humanities e-Science with a recent tendency to re-brand itself as Arts and Humanities e-Research. The authors of this article are undecided on the matter how to call the new emerging discipline. For the purpose of this article, we will use the term e-Humanities in order to fit into the more international context, while recognizing that the term e-Science offers some benefits, as it points to a specific form of collaboration. We are going to show that many of the successful grass-roots activities in e-Humanities originated from newly formed productive partnerships with sciences. This meant not only to work together on a common problem but also the adoption of a new different viewpoint in humanities research regarding old questions. For example, looking at the way computer science tries to simulate the understanding of meaning in music and texts helped the research practice of textual studies and musicology. Discovering the statistics behind mining technologies means understanding what a more mathematically oriented methodology can deliver and what it will fail at. This enhanced perspective on research practices and methodologies is also part of the theme of e-Science which will hopefully not get lost in the re-branding as e-Humanities.

The aim of this paper is to review some emerging successful practices in e-Humanities from the perspec-

tive of Germany and the UK. We focus on those projects which work on the collaboration of established science and advanced computing practices with arts and humanities. This means ignoring many successful projects in the wider domain of Humanities Computing, which is centrally concerned with the common methods and techniques for analysing source materials, tackling problems and communicating results in the humanities disciplines. We concentrate on those projects emerging from e-Humanities which use novel e-Science tools and methodologies.

This paper argues that claims to universality are the wrong way of promoting future research in e-Humanities. We need to open local spaces in which arts and humanities researchers can engage with e-Science tools and methodologies in a way they are used to engage with other research in their domain. The UK Arts and Humanities e-Science Scoping Study suggests to ‘embed’ e-Science in arts and humanities research practices.¹ We will then see that, for example, methodology work and the understanding the computational methods by arts and humanities researchers are a high priority. So are lean and interoperable middleware solutions, which really seriously consider how to enable researchers as ‘peers’.

This is our experience working in two well-established e-Humanities communities. The UK Arts and Humanities e-Science initiative² has been running for three years and produced several grass-roots projects by arts and humanities researchers working together with computing and e-Science specialists. In Germany, TextGrid³ was from the beginning the attempt to integrate existing community tools and services into one platform. The project included a broad range of researchers from textual studies and computing.

It is often emphasized in the e-Science community that there cannot be one solution that fits all problem sizes but often enough large-scale projects still dominate. Too often these projects raise unachievable expectations and claim to bring about nothing else but a ‘regime change’ for (humanities) research. There is a place for large-scale infrastructure also in e-Humanities - mainly in the provision of data and computational resources - but this article would like to focus on local solutions without such universal claims. From experience, they work best for the community we are working with.

For the purpose of this paper, we have structured the experimentations for such local e-Humanities solutions into three distinct domains. Firstly, there is the attempt to consider seriously the challenges of the data deluge that is driving e-Science uptake in all disciplines. These projects define on a local and domain specific level what it means for humanities to work with large amounts of digital data. Sec-

ondly, we look at the experimentations with novel methodologies where arts and humanities researchers seek to understand better the impact of computational methods onto their domains. Researchers have not been simply interested in finding the right tools but also in understanding how these tools work. Much work has gone into methodological research, which should be encouraged by future e-Humanities activities. This methodological research reflects a need for arts and humanities researchers to present their particular research processes to other domains like computing, as barriers of understanding seem to be particularly widespread. Thirdly, we would like to present some more promising e-Infrastructure initiatives we have been involved in. We would like to call them ‘lean grids’, though some of them have got nothing to do with grid technologies. They are built for local solutions — helping with concrete research problems. Such local solutions could then be the building blocks for larger research infrastructure work. For infrastructures in particular, universal solutions seem to be rejected by researchers.

In the next section, we will look at how there are no universal solutions for arts and humanities researchers when it comes to working with their specific data sets.

2. Working with Data

It has often been emphasized that humanities data is particularly complex and difficult to deal with [9]. Reviewing some existing e-Humanities projects, there are generally three problems when it comes to humanities data from a computational perspective.

1. There are few standard formats or interfaces that make systems interoperable and data not isolated. In the next subsection, we will look at attempts to build such standards for a specific discipline.
2. As we deal with human reasoning and discourse realised in the data, there is an increased semantic barrier. Research material is highly context dependent. For instance historical data is only valid for a particular location in a particular time. In subsection 2.2, we see how researchers look at AI technologies in order to make sense of their data.
3. Access to qualitative human based data needs novel methods of selection. We will see in this section how projects attempt to define this access by using new kinds of interfaces.

As a result of these problems, the general assumption is often that humanities data is particularly difficult to deal with because most of it is created by humans and not by automatic simulations. The data is therefore error prone

¹<http://www.ahds.ac.uk/e-science/e-science-scoping-study.htm>

²<http://www.ahessc.ac.uk>

³<http://www.textgrid.de>

and inconsistent. But, recordings for linguistic analysis are not necessarily any smaller nor less complex than recordings of physical, environmental or medical phenomena. At first sight, what seems to be missing in the humanities is a language that would help with data integration and comparison. Such a language has to be domain specific. As texts are currently the best understood data items in the humanities community, it seems appropriate to look at textual studies standards first.

2.1. Data Interoperability

There have been several efforts in the wider domain of digital humanities to work on interoperability languages for specific subdomains. Text Encoding is a good example. It has been at the heart of the disciplines of Digital Humanities for a long time. Text Encoding is widely used and at the same time of high theoretical impact. Contrary to some other more experimental Humanities Computing activities, it is (already) recognized as a valid addition to humanities research in general by employing computational methods. Text (en-)coding in the humanities, however, has mainly been centered around markup. The Text Encoding Initiative (TEI) standard has been very successful as a way for humans to annotate texts and make them understandable to other humans. It remains to be seen whether TEI with its many degrees of freedom in choosing a concrete encoding strategy can have a similar impact for text exchanges between computational agents. This, however, would be its main use case in more computationally oriented e-Science for humanities. TextGrid experienced this as the challenge to maintain the ability of TEI to express semantic depth while at the same time using it as an abstraction language to ensure interoperability. Universal claims towards standard interoperability languages even in the particular subdomain of textual studies have proven to be difficult to realise.

TEI has been used in various projects for both, as a standard format to further homogeneity as well as a technical means for enabling semantic depth. Due to its openness and powerful annotation mechanisms, however, interoperability across different projects proves to be difficult. Creating homogeneity and interoperability in arts and humanities data seems sometimes like an impossible task. Available data varies as to semantic depth and completeness. In general, it can be said that semantic depth needs to be reduced, as the data from various sources is combined in striving for completeness. Interoperability often means to stick to the lowest common denominator between heterogeneous sources (cf. the experiences of the Dublin Core Metadata Initiative).

The TextGrid project in Germany aims to enable both dimensions at the same time: semantic depth and completeness. To achieve this, it developed the core encoding approach, which follows a simple principle: one can always

go from a higher semantic degree to a lower semantic degree; and in possession of a suitable transformation script, this mapping can be done automatically. TextGrid encourages all its participating projects to describe their data in an XML-based markup that is suitable for their specific research questions. At the same time projects can register a mapping from their specific, semantically deep data to the respective TextGrid-wide core encoding that is a reasonably expressive TEI-subset. Users can then do both, cross-analyse all data available in the TextGrid repository based on the TextGrid core encoding (completeness), as well as retrieve the project-specific encodings (semantic depth).

The problems to define standards for the interchange of human created data while at the same time providing the necessary semantic depth to be suited for arts and humanities research, led to attempts in the community to work on other ways of generating standard compliant ways to understand the deeper semantics and meaning in humanities data. At the centre of these attempts are decision support systems to help researchers deal with the uncertain nature of humanities data and other AI technologies such as text and data mining to help with sense-making.

2.2. Making Sense of Humanities Data

Data in arts and humanities is most of the times created manually with enormous effort. Hence, the per-unit value of humanities data is often greater than from scientific sources (despite the often high financial investments in scientific projects). The analysis of these data items takes sometimes years when degraded inscriptions and manuscripts need to be deciphered. In Oxford, a project called *Image, Text, Interpretation: e-Science, Technology and Documents* works on such a decision support system to decipher and interpret degraded manuscripts, which are fragmentary, stained and damaged. Advanced imaging technology helps to develop and deliver an image-processing tool, which can be applied by researchers on a range of documents [13]. An essential part of the tool will be to incorporate expert knowledge on how to reassemble damaged manuscripts, using a combination of advanced imaging technologies together with data mining and AI approaches. These kinds of technologies are never universal generic solutions but need to be trained according to specific targets. Decision support systems to minimise uncertainties have successfully served medical or financial analysis for a number of years. They offer some promising prospects for doing the same for humanities data. Another tested technology to help with access to unstructured information by unveiling the hidden structures is text mining. Again, mining solutions only work for specific subdomains as in this case archaeological records.

The York-based Archaeological Data Service (ADS) is responsible for developing *Archaeotools: Data mining,*

facetted classification and e-Archaeology.⁴ Over 40000 reports of grey literature in archaeological excavations lie potentially idle, as they are hard to access. Archaeotools is an attempt to provide access to these records using automated metadata generation techniques. These will index datasets for new links in the records in terms of When, What and Where. The underlying datasets combine over one million records from the National Monuments Records of Scotland, Wales and England as well as Historic Environment Records from numerous local authorities and the ADS's own archive holdings.⁵ The formation of the facets is supported by existing thesauri and the University of Edinburgh's geoXwalk service,⁶ which will support geospatial information access to the data. A 3D-space will visualise facets and their links and provide access to deeper unpublished archaeological literature, whereas users will be able to ask for their own specific research interests to be represented in the indexing of these research records. This will create flexible access to resources up to now neglected in research.

Both the Oxford and the Archaeotools project show ways to prepare humanities data for the consumption of computers. As they are based on machine learning techniques, they only work in specific subdomains. It will be at least as important for researchers to offer new interfaces to this data that reflects the way they regularly work with the data.

2.3. Working with data: Interfaces that reflect the thinking of researchers

It comes to no surprise that much of e-Science in arts research is concentrated on defining new interfaces. In [3], the Newcastle based project *Data Services for Associated Motion Capture User Categories (AMUC)*⁷ is analysed for the way it integrates the 'thinking' of performance arts research into novel interfaces to live performance data. Its prototype retrieval tool impresses with its idea to build a sketch-based interface in order to best mimic the thinking and understanding of this very particular user group. Another performance arts project *E-Dance*⁸ aims to investigate the impact of the Access Grid interface on distributed performance. The involved computer scientists expect the specific dance-led research perspective help improve the Access Grid interface and related ones. Using the experiences of dance research, the Access Grid might become closer to its original promise of an entire room alternative to the classic desktop — as envisioned in 'Group-Oriented Collaboration: The Access Grid Collaboration System' [6].

⁴<http://ads.ahds.ac.uk/project/archaeotools/>

⁵<http://www.nesc.ac.uk/action/esi/download.cfm?index=3714>

⁶<http://www.geoxwalk.ac.uk/>

⁷<http://ahessc.ac.uk/files/active/0/AMUC-report.pdf>

⁸<http://www.ahessc.ac.uk/e-dance>

Central to building this whole room interface, which would support arts research with the Access Grid, is the exact recording and definition of spatio-temporal aspects of live performances. *E-Dance* will advance the development of building a non-textual and visual interface to computing infrastructures by locating and relocating dances in a virtual space using live and mediated performances. The non-linear character of performance shall help develop a pre-linguistic, event-based interface that should benefit other spatio-temporal determined research practices such as archaeological excavations, too.

All projects presented in this section work with data that is already available but poses challenges to the research communities as new computing based methodologies need to be 'embedded' into established research practices. All of these computing based methodologies from text mining to semantic annotation are proven to work in other sciences (especially in medical sciences). In the following section, we look at projects that look for more experimental methodologies and their use in arts and humanities research. They work again on the local level by investigating concrete research problems and how e-Science tools and methodologies might help.

3. Methodological experiments

Deep methodological differences between computing sciences and humanities are often best expressed in seemingly minor differences in language, which we experience again and again when arts and humanities research meets computing. Here, the e-Humanities experience is not different from the one Humanities Computing has had for several decades. Language is often experienced as a subtle indicator of cultural differences with a major impact on technological approaches. For example, *authenticity* cannot be solved with checksums from a humanities perspective, authenticity is ensured by recording context information about the creation of a document and its purpose. After all, a JPEG2000 derivate of a TIFF image may be intellectually pretty much the same thing, the checksum will deny this, however.

Another example for two different languages in computing and humanities is the usage of the word *transparent*: Computer scientists would generally say transparent to any technical procedure that is invisible to the user. Humanities scholars on the other hand mostly consider any technology transparent, which transparently displays its full complexity to the user without obscuring it. Innovation emerges exactly at those intersections of different languages and perspectives. For example, the *Purcell Plus* project at Goldsmith College⁹ works on the better understanding of the new availability of huge amounts of digital music resources

⁹<http://doc.gold.ac.uk/mas01tc/PP/>

on the Internet. Over night so to say, musicology has transformed itself from a data-poor discipline into a data-rich one. With these new resources, new empirical approaches to musicology seem to be feasible for the first time. *Purcell Plus* focuses on the comparison of computational methods to describe similarities of music pieces with the more traditional theoretical musicology based ones. This shift in methodologies has only become possible with e-Science means to integrate music resources and provide useful access to them.

If arts and humanities researchers therefore demand transparency and authenticity, they talk about requirements of their actual research methodologies rather than requirements for computing support. In the next subsection, we look at more new methodologies that have only become possible due to the emergence of e-Science paradigms in arts and humanities. We first consider new modelling methodologies for the curation of artifacts, before in section 3.2 we investigate how computational methods can be used to cleanse and enrich humanities data. 3-D models are bound to revolutionise museum studies and curatorship.

3.1. Advanced visualisation

University College London (UCL) has a long-standing expertise in curatorship and e-Science. Typically, curators work in the physical world using hands and advanced sensors to assess an object's identity and its current condition. The information gained from this assessment is then published in a report consisting of a combination of photographs and texts. Both forms of representations are necessarily selective and interpretative. Texts require exact descriptions, and images reflect only 2-dimensional aspects of the 3-dimensional world. *E-Curator - 3D colour scans for remote object identification and assessment*¹⁰ works on a true revolution of the methods in the field of curation and museum and archive studies. Technology today allows to have an exact 3D-model of the whole object with great detail in terms of surface structure and other significant properties of an object like colour and shape. UCL's Arius3D scanner is the first of its kind in Europe providing high resolution 3D-geometry through the use of a laser triangulation system at a 100 microns point spacing.

Such 3D-models would be highly beneficial to any arts and humanities research domain but in particular museum studies. They could help identify degraded surface structures and allow comparisons without the need for the researcher to be physically co-located with the object. However, it is only now and with programmes such as the e-Science initiative that the technology and the expertise can come together to make use of such benefits. As a data item that could be shared and manipulated, exact 3D models are

relatively unknown to large user communities - particularly in the arts and humanities. *E-Curator* has targeted these unknown methodologies and developed recommendations and policies how to validate and share 3D data for curatorial practices.

High resolution colour scans of one object require hundreds of megabytes of storage space, and can only realistically be shared using advanced grid technologies such as Storage Resource Broker (SRB). But using grid technology is only part of this project's challenge. The real question is rather how to provide the meaningful link-up of data when it comes to 3D-models of complex historical artifacts. How can 3D models be purposefully linked, securely distributed and shared between museums and academic institutions? How can then the integrity and quality of the data be ensured to allow, for instance, the verification of that condition by insurers? The challenge *E-Curator* therefore addresses is not so much a technical but a policy and methodology one. *E-Curator* discusses how virtual objects - even if they are 3D models and therefore the highest possible recreation of the original object - can be accepted as a means of research and other curatorial practices.

3.2. Simulating the past

We have just seen that particularly museum studies and musicology are not data-poor disciplines. Again, it is wrong to make any universal claim that humanities in general lack data. Consequently, methodology research is dominating the e-Humanities agenda in museum studies and musicology. For both, it is important to find out how computational methods compare to existing non-computational ones. Sometimes, however, there is just not enough data to work with. This is particularly true for historical data where even if there is some data available it is not enough or not good enough to draw any meaningful conclusions. In earlier work [2], we reviewed the efforts at UCL to work on available digitisations of historical census data held at the private company Ancestry.co.uk. We explored possibilities using probabilistic indexing and matching in order to (semi-)automatically create a longitudinal database of individuals across the census. To do so, systems have to be trained to incorporate heuristic modelling procedures as currently applied by historians. They need to be trained to realise humanities methodologies how to match historical records. This is an extremely difficult task for two reasons: (1) the records are very inconsistent due to human errors while recording them in the 19th century and (2) the overall size of US census holdings alone from 1790 onwards exceeds the size of digital weather recordings by the US National Climate Data Centre and youtube video data.¹¹

¹⁰<http://www.museums.ucl.ac.uk/research/ecurator/>

¹¹http://www.wired.com/science/discoveries/magazine/16-07/pb_intro

Probabilistic methods are one proven way in computing to work with insufficient and inconsistent data recordings in the humanities. However, they can only function if there is a certain amount of existing data and not if there are simply not enough and no recordings at all. Archaeology, for instance, often deals with time periods of which we have little or no information at all. That's why in archaeology agent-based modelling (ABM) as a way to simulate the past is already established. In agent-based modeling (ABM), a system is modelled as a collection of autonomous decision-making entities called agents [4].¹² The interaction between agents reinforces these decisions or changes existing behaviour. ABM is interesting to researchers of the past as it provides a description of a system of human interaction and can highlight emerging and therefore unknown behaviour. Both characteristics are vital to cleanse and recreate inconsistent historical data sets. Of course, not everything in societal behaviour can be described in mathematical equations, which is why ABM is still regarded with suspicion in many parts of the community. Individual and group behaviour is just too complex and the complexity of differential equations increases exponentially as the complexity of behavior increases [4]. But grids can at least help deal with this complexity and showcase current limitations of using ABM in humanities research. On the other hand, ABM may provide the only chance to create at least some more meaningful data by running experiments that explore the conditions of certain historical events. It is in its current state nowhere close to a universal solution of a realistic modelling of human societies but it might provide at least some insights where all other methods have to fail due to necessarily insufficient data.

A Birmingham based project, entitled *Medieval Warfare on the Grid: The Case of Manzikert*,¹³ will use computational clusters to recreate the military logistical context of the Battle of Manzikert in 1071 using ABM. In Manzikert, the Byzantine army was defeated by the Seljuk Turks. This defeat has been one of the milestone events in the demise of Byzantine power and plays even nowadays a key role in explaining the relationship of East and West. As there is only limited data available, there is no consensus between military historians on basic facts like the route or the sizes of the involved armies. The project will explore methodologies to use the relationship of modelled data and historical data to develop a better understanding of the battle. With e-Science technologies and methodologies, it is possible to run such complex historical simulations on data from different sources such as transport infrastructure, agriculture and military organisation.

This section has described new methodologies for arts and humanities research following the emergence of e-

Humanities as a new discipline. All of the presented projects ask specific questions relevant to particular research domains in arts and humanities. ABM, for instance, will at least currently help only in particular cases like the battle of Manzikert. The next section will shed light on arts and humanities demands towards infrastructures that would really work for them. It is our experience, that it is most important to keep such infrastructure projects focussed and simple. Again, a generic solution covering all research domains is likely to fail. In the next section, we discuss briefly two ways of building 'lean grids' reflecting two specifics of arts and humanities research. It is relatively poor in terms of funding and spread out among many disciplines and sub-disciplines.

4. Lean grids

There are already some large e-Infrastructure initiatives for arts and humanities such as the European ESFRI project DARIAH.¹⁴ DARIAH stands for 'Digital Research Infrastructure for the Arts and Humanities' and is funded to conceptualize and afterwards build a virtual bridge between different humanities and cultural heritage data resources in Europe. However, such projects as useful as they might be are likely to be the exception. The TextGrid experience and the UK Arts and Humanities initiative have clearly shown that the need for access to national and international large scale grid services is currently relatively limited for humanities and arts. This has several reasons which are external to the community. The difficulty of the GSI-based grid security systems is often quoted here. Also, arts and humanities research often takes place in smaller institutions, which do not have the necessary IT support not to mention grid support to connect to national grid services. However, the most important reason seems to be that most of digital research in the arts and humanities is done in an interactive manner as a way of humans requesting resources, annotating data or running small supporting tools. The model of 'running jobs' on the grid is alien to such research practices. At the moment at least, large and in particular grid-based infrastructures do not support interactive behaviour well. In this section, we will look at two alternatives. Firstly, TextGrid is modelled less on the example of grid systems but as a digital ecosystem. Secondly, we will look at 'volunteer thinking' approaches and how they might help in a distributed library environment for humanities data. We see both as examples of what we would like to call 'lean grids' — approaches which incorporate the idea behind grids to share resources while at the same time not relying on building heavy infrastructures.

¹²<http://www.pnas.org/content/99/suppl.3/7280.full>

¹³<http://www.cs.bham.ac.uk/research/projects/mwgrid/>

¹⁴<http://www.dariah.eu>

4.1. ROAs, SOAs and Digital Ecosystems

The ‘grid way’ — at least not in the strict sense as a specific technology stack — is by no means the only approach for e-Humanities networks. E-Humanities and cultural heritage federations such as ‘TAPoR (Text Analysis Portal for Research)’¹⁵ or ‘Bricks (Building Resources for Integrated Cultural Knowledge Services)’¹⁶ [12] build on more light-weight paradigms. In particular, they position themselves as portals and / or networks of services and data (‘resources’), though especially TAPoR seems to have complemented its originally pure Service Oriented Architecture (SOA) + SOAP Web Services approach¹⁷ with collections of Desktop tools for specific types of services.

This in itself is hardly surprising. As we have seen, much of the work with e-Humanities data is in its nature highly interactive and can only be partially automated (this, incidentally, is the main reason that TextGrid consciously opted for a fat-client approach towards the grid). This is quite different from ‘traditional’ e-Science approaches that concentrate more on batch mode processing and the testing of mathematical models in a sequence of (potentially parallel) program runs. Human users and the strong emphasis on interacting with data are much less of a priority there.

Both Service Oriented Architectures (SOAs) and Digital Ecosystems (DEs) share the grid’s and e-Science’s focus on peer interaction between software components, be they interacting capabilities or self-organizing agents. For example, DEs are defined as ‘agents-based, loosely coupled, domain-specific and demand driven interactive communities which offer cost-effective digital services and value-creating activities that attract agents to participate and benefit from it’¹⁸ and the OASIS SOA Reference Model [10] defines SOA as ‘a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains’. According to this, grids, SOAs and DEs tend to minimize if not outright overlook both the importance of content as the key element (not only) in the e-Humanities and the role human users play in the network.

It is for this that we emphasize the role of data and human agents in e-Humanities DEs [7] [8] and position in particular TextGrid squarely in a mesh of human and machine agents working on data. Fortunately, Resource-Oriented Architectures (ROAs), based on the RESTful model [11] [5], fit in very well with this approach and, for that matter, with the renewed emphasis on repositories and registries, treating data also technically as key factors in any e-Humanities DE. We see TextGrid as a key subsystem of such lean, open and highly interoperable ‘Ecosystems

of Ecosystems’ of humans, services and data in the e-Humanities domain. RESTful interfaces have also been implemented in several projects of the Arts and Humanities e-Science initiative in the UK - among them *e-Dance*, *AMUC* in Newcastle and the archaeology e-Research portal of the VERA project in Reading.¹⁹

TextGrid is a relatively well funded national German project. Most e-Infrastructure work in arts and humanities, however, should start with the economic argument that arts and humanities are ‘poor’ compared to the research funding sciences are often used to. This means we need to adjust efforts to an economic environment that does not allow to build large transnational e-Infrastructures. This is a especially important argument for international collaborations, which are not supported by collaborations between national or supra-national funding bodies. King’s College London is suggesting to build a ‘poor person’s grid’ to support arts and humanities data exchange between Taiwan and UK.

4.2. Poor person’s grids

The primary aim of the project is to address the issues involved in developing more resource-efficient ways of providing researchers with integrated discovery of and access to arts and humanities research assets held in digital libraries that are physically distributed on a global scale. The model for the potential solutions that we will address is a collaborative one, a ‘poor person’s grid’ based on the concept of ‘volunteer thinking’. This is a recent evolution of distributed computing, in which volunteers on the web contribute to scientific data analysis — examples include projects such as *GalaxyZoo*,²⁰ and *Herbaria@home*.²¹ *GalaxyZoo*, for example, used the knowledge of thousands of participants to solve the simple (in individual cases) task of classifying galaxies into spiral and elliptical galaxies; such ‘armchair astronomers’ have identified over 500 overlapping galaxies, whereas astronomers had previously known of only 20 such systems. Volunteer thinking has proved highly reliable because the process incorporates highly redundant verification of results from many volunteers. This approach could be adapted to producing detailed and fine-grained metadata for arts and humanities material, thus building a ‘Citizen’s Digital Library’ for arts and humanities.

As a case study, we intend to investigate the approach in the context of the digital library of cultural heritage material developed by the digital archive programme in Taiwan; this will allow us to address the challenges of creating a functional and sustainable digital asset management system across two greatly different cultural contexts. The

¹⁵<http://tapor.humanities.mcmaster.ca/home.html>

¹⁶<http://www.brickscommunity.org/>

¹⁷<http://portal.tapor.ca>

¹⁸<http://www.ieee-dest.curtin.edu.au/2007/06060720-20Call20for20Papers.pdf>

¹⁹<http://vera.rdg.ac.uk/>

²⁰<http://www.galaxyzoo.org/>

²¹<http://herbariaunited.org/atHome/>

unified catalogue of TELDAP (Taiwan e-Learning and Digital Archive Program) contains over two million items, including images and multimedia files, in thematic categories ranging over geology, anthropology, Chinese rare and ancient books, maps, and archaeology. The case study will involve making this archive more accessible to UK researchers (and to the wider international community) by using the ‘volunteer thinking’ methodology outlined above. The metadata records will be ‘translated’ into English and adopted to international standards. New records will be added, while old ones will be clustered into collections in order to add meaningful context to individual records. We will use this work as a case study, to advertise the work in universities, schools and other places and to enlist volunteers.

These e-Infrastructure initiatives have more or less directly emerged from the grass-roots activities of e-Humanities in Germany and the UK. They close gaps and attempt to bring together different tools and methods into one framework rather than creating a grand new solution that fits all sizes. They can be seen as a major emerging trend for future e-Humanities.

5. Conclusion and Future Trends

Both TextGrid and the e-Science initiative in the UK, have created a lasting collaboration between arts and humanities with scientists and computing researchers to solve challenges posed by the new digital resources available. In order to address the specific needs of the arts and humanities communities, it is important that such collaborations start from concrete research questions in arts and humanities. Then, the dialogue will become a productive one, from which both sides will be able to gain new insights and perspectives. The enthusiasm to work together is there — now, it is a question of keeping this momentum going and remain focused on the concrete interests and needs.

We argued in this paper that our experience in these existing projects suggests the way forward is to start with ‘embedded’ research activities. According to our review of current trends in e-Humanities, it is important for e-Humanities to refuse claims to generic, universal solutions. E-Science in arts and humanities has neither in Germany nor the UK limited itself to more traditional ideas of e-Science, which link e-Science mostly to the application of certain advanced network technologies in supporting sciences. These technologies like the grid have their place but they might generate the impression that the solution is already there and only needs to find the right application and those willing to give the money. From our experience also in the wider world of e-Science, e-Science will fail if it deems itself as just an application of technologies. Rightfully, it will then be perceived as an invasion of some technology know-it-

alls, which know the solution without knowing the problem. This perception might be even worse in arts and humanities. Their development and success was at least in the past seldomly linked to advances in computing. Men and women are still vastly superior to machines when it comes to discussing history, analysing concepts or revolutionizing arts.

References

- [1] American Council of Learned Societies. Our cultural commonwealth: The final report of the ACLS Commission on Cyberinfrastructure for the Humanities and Social Sciences, 2006.
- [2] T. Blanke and S. Dunn. The arts and humanities e-science initiative in the uk. In *E-SCIENCE '06: Proceedings of the Second IEEE International Conference on e-Science and Grid Computing*, page 136, Washington, DC, USA, 2006. IEEE Computer Society.
- [3] T. Blanke, M. Hedges, and S. Dunn. E-science in the arts and humanities - from early experimentation to systematic investigation. In *E-SCIENCE '07: Proceedings of the Third IEEE International Conference on e-Science and Grid Computing*, Washington, DC, USA, 2006. IEEE Computer Society.
- [4] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *PNAS*, (99), May 2002.
- [5] T. Dillon, C. Wu, and E. Chang. Reference architectural styles for service-oriented computing. *Network and Parallel Computing*, pages 543–555, 2007.
- [6] I. Foster and C. Kesselman. *The Grid 2: Blueprint for a New Computing Infrastructure*. Morgan-Kaufmann, second edition, 2004.
- [7] M. W. Küster and C. Ludwig. Digital ecosystems of ehumanities resources and services. In *Proceeding of the Second IEEE International Conference on Digital Ecosystems and Technologies*, 2008.
- [8] M. W. Küster, C. Ludwig, and A. Aschenbrenner. TextGrid as a digital ecosystem. In E. Chang, editor, *DEST 2007*, 2007.
- [9] M. Nentwich. *Cyberscience. Research in the Age of the Internet*. Austrian Academy of Science Press, Vienna, 2003.
- [10] OASIS. Reference model for service oriented architecture. Oasis committee specification, OASIS, 08 2006.
- [11] L. Richardson and S. Ruby. *RESTful web services*. O’Reilly, Farnham, 2007.
- [12] T. Risse, P. Knežević, C. Meghini, R. Hecht, and F. Basile. The BRICKS infrastructure — and overview. In *8th International Conference EVA 2005 Moscow*. EVA, 2005.
- [13] M. Terras. *Image to Interpretation. An Intelligent System to Aid Historians in Reading the Vindolanda Texts*. OUP, 2006.