Grid TALK GridBriefings
Grid computing in five minutes

Digitising culture: Grids and eHumanities

The era of eHumanities has arrived. No longer consigned to manuscript, memory or museums, digitisation is transferring the humanities to our computers. "Our original sources are scattered in many ways: historical documents in archives and libraries, works of art in museums, archaeological finds in ancient sites, disappearing languages on the tongues of native speakers," says Peter Doorn, director of the proposed humanities infrastructure, DARIAH[1]. "Once these are digital, we want to be able to find them easily, to process and analyse them with user-friendly yet innovative software tools, and to publish our findings, in order to share and discuss our new insights with our colleagues." This article takes a look at how grid computing, along with other technologies, is realising this aim.

A social grid

Today a significant amount of digital material is available to researchers, through digitisation programmes but also because more and more data is 'born digital'. However the volume of these new datasets can pose problems. In addition, while many resources for humanities research are available in digital form, this is not true for everything and many researchers find that what they need for their research is either not yet digital, not accessible, not in a form that is usable, or that the software tools to analyse that particular material are lacking. Questions of storage, access and analysis are major challenges for humanities researchers. In the same way that science researchers have turned to technologies such as grid computing to solve these problems, those in the humanities are also realising the advantages offered by such solutions.

Grid computing offers humanities researchers a number of benefits:

- Grids can more easily deal with the enormous and constantly increasing amounts of data that have come into existence through digitisation initiatives.
- Grid structures can enable collaborations between researchers working on similar projects who are separated by physical distance, by integrating them into virtual research communities and allowing them to securely share data between trusted sources.
- Grids can provide tools to researchers spread across the globe that were previously only available on a local level.

1 - Peter Doorn, http://www.dariah.eu/documents/DARIAH_Newsletter-3_June2009.pdf

A virtual museum

For such a small continent, Europe is packed full of culture. From films to paintings, photos to books there is a wealth of material waiting to be explored. And, since 2008, Europe's collections have been even easier to visit through Europeana.eu.

Europeana is a hybrid of digital library, museum and archive which allows visitors to search and browse through 4.6 million digital images, texts, sounds and videos held in collections across Europe.

The project is mainly funded by the EU and is run by a core team based at the national library of the Netherlands, the Koninklijke Bibliotheek.

By uploading digitised content from cultural institutions, the project's creators hope to provide access to 10 million objects by late 2010. Whether it's watching footage of the fall of the Berlin wall, viewing original scores by Mozart or admiring the Gutenburg Bible, Europeana is a simple way to discover Europe's cultural heritage.

Comentarios al Apocalipsis, Beato de Liébana, 1047
The National Library of Spain (in Europeana)

Ioli Kalavrezou, Harvard University - *"With access to the digital infrastructures the humanities are becoming almost a new discipline. A variety of resources are now at the disposal of the scholar right at his/her desk. Not only accessing the vast textual and visual records, but the approach to the material can become more interdisciplinary, which in return encourages more collaboration among researchers. The availability of the digital infrastructures have opened up information to anyone who is associated with universities and research institutes and has learned to 'search.'"*

Projects such as TextGrid in Germany are opening up the potential of grids for the humanities. TextGrid is developing a virtual research environment (VRE) for humanities researchers. A repository provides a long-term archive for humanities research data while a laboratory gives access to tools with which to edit, analyse or annotate data. The grid plays an essential role in both, and allows for collaborative work, independent of location. Now in its second phase, TextGrid's consortium includes academic communities in philology, linguistics, art history, and musicology who are working together to achieve sustainable operation and a wide set of users.

Heike Neuroth, TextGrid – *"Early on, D-Grid recognised the necessity and opportunity in inviting the humanities into a national grid initiative. D-Grid tasked the TextGrid project to unlock the promises of digital infrastructure for the humanities. TextGrid benefits from the expertise of the other D-Grid communities, yet it also contributes back its experiences with regard to data and metadata management, digital curation, handling unstructured data, and generally in establishing interactive virtual research environments. From this collaboration across communities a new digital infrastructure is emerging that extends the grid concept of virtualised hardware to semantically enriched data and methodologies, with the human user at its centre."*

Uniting the humanities

eHumanities research is currently fragmented, with islands of researchers each using their own tools and applications, tailored to answer specific research questions and methodologies. Much research does not yet agree on and use common standards, nor make raw data available in digital form.

This need for collaboration has not gone unnoticed. Many institutions, including the ESFRI-projects (European Strategy Forum in Research Infrastructures) CLARIN and DARIAH are working towards a more harmonious form of eHumanities. DARIAH (Digital Research Infrastructure for the Arts and Humanities), for example, aims to create one European data area in which scholars and students can access scattered resources, as well as giving expertise and sharing best practices.

In October 2009, eight networks, infrastructure projects, and planning initiatives set up CHAIN (Coalition of Humanities and Arts Infrastructures and Networks). CHAIN is committed to overcoming the barriers created by researchers working with incompatible technologies. By ensuring their present, proposed, and future activities are interdependent and complementary, CHAIN members want to create a shared environment where technology services can interoperate and be sustained, enabling new forms of research in the humanities.

Bringing music to life

Musical instruments long consigned to the history books have been brought back to life thanks to an innovative partnership between scientists and musicians.

The ASTRA project is reviving the sounds of instruments such as the epigonion, barbiton, syrinx, salpinx and aulos. By using grid computing techniques, researchers are able to model ancient instruments, and recreate timbres and sounds of notes not heard for centuries.

courtesy Luca Petralia

ASTRA uses a technique called physical modelling synthesis to reconstruct these lost sounds. Equations and algorithms describe the physical structure of the instrument, and sounds are generated by modelling it as a mechanical system with different configurations for each note. Hundreds of computers in the GÉANT, GILDA and EUMEDGRID infrastructures are used to process these calculations - without the power of grids it would take a powerful desktop computer four hours to produce just one sound lasting 30 seconds. And the story doesn't stop there. Once reconstructed, ASTRA's Lost Sounds Orchestra brings these instruments to life, giving performances of the reconstructed instruments for us all to enjoy.

John Byron, Australian Academy of the Humanities - *"The global humanities are undergoing a radical transformation through the clever use of grid architectures. New ways of working are bringing scholars together who may never have collaborated otherwise and are enabling entirely new ways of encountering the world, raising questions about what it is to be human that have simply not occurred to us before. Humanities research will repay many times over the investment in tailored approaches to generating tools, methods and technologies that meet their particular needs."*

Building a community

In order to ensure the growth of eHumanities, the humanities community needs to be closely involved when building e-infrastructures. Discussions at the Digital Humanities 2009 conference, held at the University of Maryland in June, highlighted this and proposed that initiatives should build on and enhance existing resources, rather than starting from scratch. For the field to progress, training in new computer aided techniques is also necessary. Universities have already started introducing this into the academic curriculum, so advanced computing technologies are understood and integrated into research from the start.

Sustainability

Funding remains an issue for eHumanities. Compared to the sciences, humanities have relatively less sophisticated IT infrastructure and support as they are often allocated smaller grants. Therefore sustainable infrastructures such as DARIAH are needed to support the permanent availability of digital resources, tools and infrastructures in the field. By their very nature grids could offer a solution to this issue. Grid technologies allow researchers to make better use of resources already available, providing a cheaper way to store and access data compared to setting up isolated data sources.

Daniel Röwenstrunk, Edirom, University of Paderborn – *"Grid technologies offer a good chance for smaller academic disciplines like musicology to 'stand on the shoulders of giants' and so benefit from developments in the (computer) sciences. The gains are not primarily in the use of distributed storage and processing power but in approved and standardised concepts for accessibility of content and methods, improved collaborations and co-operations and services like authentication and digital preservation."*

Martin Wynne, Oxford e-Research Centre - *"Humanities computing has a long history of creating tools, resources and services which don't work with each other, and which are not sustainable in the long term. We need to create a digital infrastructure to overcome the current barriers, making it possible to free the enormous innovative and creative potential in our disciplines."*

How to summarise *Le Monde*...in Polish

Researchers tackling questions such as 'find all video clips of Tony Blair on the BBC in 2007', 'list all uses of "enthusiasm" in 19th century English novels written by women', or 'summarise *Le Monde* of March 17 2008 - in Polish' are being given a helping hand by ESFRI project, CLARIN.

CLARIN (Common Language Resources and Technology Infrastructure) is working to establish a research infrastructure for languages. CLARIN's mission is threefold. Firstly it aims to unite digital archives containing language based material across Europe. Secondly it wants to provide access to the many language and speech processing tools developed by researchers over recent years. Lastly it wants to provide web-based services for researchers to gain access to the means by which they can answer the questions above. And grid technologies are helping it on its way.

"CLARIN is about connecting the islands that exist today within the humanities, about making available the vast amount of linguistic resources and tools to the research community in an easy-to-use way," says Dieter Van Uytvanck from CLARIN. "We hope today's grid infrastructure can be one of the bridges to interconnect the fragmented eHumanities landscape."

Availability and access

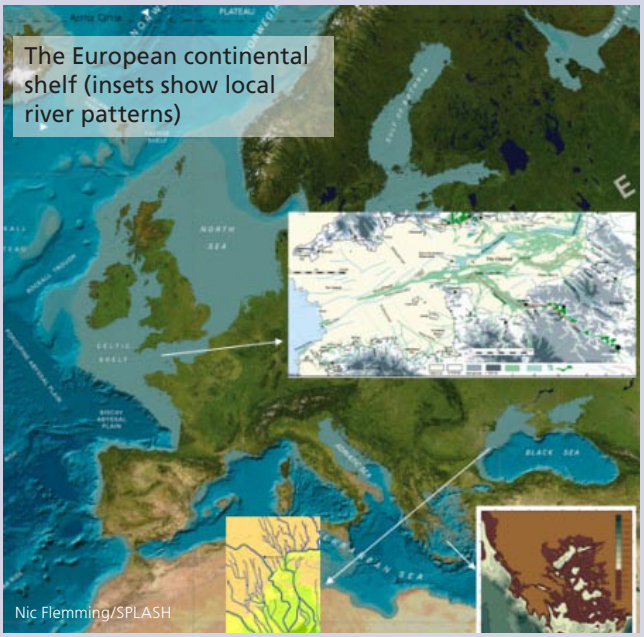
While more and more humanities data is being digitised, this is not yet true for everything. However as issues such as standards and training are addressed, and with infrastructures such as DARIAH being set up, the amount of digital data available will continue to grow.

Making raw data available in digital form can benefit researchers across disciplines allowing for research to become ever more interdisciplinary. The SPLASH project for example hopes to exploit physical and geochemical data for the use of archaeologists. This boom in eHumanities also gives the wider public access to objects and collections they may otherwise never encounter, through digitisation projects such as Europeana (see first page).

Andreas Aschenbrenner, DARIAH - *"Scholars in the humanities have been shaping the 'Digital Humanities' for several decades, and collaborative infrastructure such as grid technologies is one of the next steps forward. However, the tools and requirements differ from those experienced grid communities. To serve the humanities with grid technologies, we are currently establishing gateways to existing technologies used in the humanities, including repositories and digital curation, authentication based on Shibboleth and OpenID, and others. Once those gateways are established, large and diverse communities will be able to explore the emerging, grid-based research environments for the humanities."*

Diving back in time

Grids are poised to help researchers unearth civilisations and settlements submerged beneath our seas.



The SPLASH-COS (Submerged Prehistoric Archaeology and Landscapes of the Continental Shelf) network, an EU-funded COST Action, is bringing archaeologists, marine geophysicists, environmental scientists and commercial and industrial organisations together in a bid to delve into our watery past.

It will draw on data describing the prehistoric archaeology and palaeoenvironments of the European continental shelf, a 3.2 million square kilometre area of land which lay above sea level during the Ice Ages. Using the aid of grid technologies, large sets of seabed physical and geochemical data already collected and archived will be identified in an attempt to reconstruct these past landscapes.

SPLASH-COS will set the design requirements for researchers to access, browse and manipulate data, stored in archives across Europe. The network will aid in the collecting and sharing of information for the future, ultimately helping us learn more about our early history.

Paul Ell, Queen's University Belfast - *"The humanities is on the cusp of a methodological revolution. In the future the reliance on archives and libraries as the basic data sources for humanists will change to web-based e-sources and humanists require assistance in harnessing their potential. E-Science, and specifically the Data Grid, have the potential to markedly assist this change. The Data Grid will link disparate e-resources large and small, it will facilitate data harvesting and ingestion, and it will help to ensure that e-resources are sustainable."*

Paul Wouters, Virtual Knowledge Studio for the Humanities and Social Sciences - *"If we want the humanities to profit from investments in large informational infrastructures such as the grid, developing the right kinds of interfaces is the critical issue. This means interfaces between researchers and technology, but also interfaces between different humanities disciplines, as well as between the humanities, the social sciences, the sciences, and the public at large. Without those interfaces, the infrastructures will be virtually useless for the advancement of scholarship in the 21st century."*

Rutger Kramer, Data Archiving & Networked Services – *"Arts & Humanities and Social Sciences researchers should have the correct data at their disposal at any time. Our aim is to enable researchers to find the data they need, by giving them access to data across scientific disciplines and beyond national boundaries. Building data infrastructures on grid technology allows us to profit from the work already done on federating access, reliable storage, data exchange and virtual collaboration."*

For more information:

TextGrid: ***www.textgrid.de***
DARIAH: ***www.dariah.eu***
CLARIN: ***www.clarin.eu***
ASTRA: ***www.astraproject.org***
SPLASH: ***www.york.ac.uk/depts/arch/news/Splash.html***
Europeana: ***www.europeana.eu***
Edirom: ***www.edirom.de***
e-IRG: ***www.e-irg.eu***
EGEE (Enabling Grids for E-sciencE): ***www.eu-egee.org***
iSGTW: ***www.isgtw.org***
GridTalk : ***www.gridtalk-project.eu***

GridTalk is co-funded by the EC under FP7