Digitus Dei ist hic!

S. Berner
sberner@ecognus.com

One does not usually associate 'humanities' with 'information technology', yet such an association not only exists, but is thriving on state endowments in a number of countries, creating an interesting mongrel known as 'digital humanities' or 'humanities computing' that goes both ways to enrich the two fields in ways that were unthought of just a few decades ago.

Digital humanities (DH) is, very broadly speaking, the interdisciplinary point at which computing meets culture. It is also what happens to these cultural artifacts as a result of being worked on by computers (DH as tool), and what humanities have to offer computational science (DH as lens). The balance is a little skewed, especially in this age where technology is king and all the important underlying principles are considered quaint – few of the young these days are concerned with learning philosophy, religion or history, and the big thrust is towards a belief that 'technology will solve all'.

DH did not suddenly appear onstage in the 1990s, with the advent of the Internet. Indeed, computers started to be used in humanities in the mid 1940s, with Roberto Busa's complete lemmatization of the concordance with critical edition to the works of St. Thomas Aquinas under the title *Index Thomisticus*. Busa, an Italian Jesuit and professor emeritus of professor of Ontology, Theodicy and Scientific Methodology at the Gregorian Pontifical University in Roma, who pioneered computational linguistics, writes of his project in his Foreword to the very informative *A Companion To Digital Humanities* (Schreiberman, Siemens and Unsworth 2004):

'Between 1941 and 1946, I began to look for machines for the automation of the linguistic analysis of written texts. I found them, in 1949, at IBM in New York City' (Schreiberman *et al*. 2004:xvi).

In the half-century since DH began, its activities have grown in number and sophistication, developing from a loose collection of techniques to a coherent academic field, while at the same time remaining very broad in its scope. Its main interest is the 'text' encompassing the full range of 'multimedia', and its goal – especially following the advent of the World-Wide Web – is to use IT in understanding cultural artifacts and using that understanding to affect the way IT is developed and used (Schreiberman *et al*. 2004:xxiii).

DH 'probes how standard software may be used to make new knowledge from our cultural inheritance and from the contemporary world' (Kings College 2006). Epistemologically, two issues are focal to the DH: (a) how we know what we know, and (b) how we imagine what we do not know (McGann 1997). The primary effect of computing on humanities is not to accelerate the pace of research, or making its performance more efficient or accurate, although all those improvements can occur, but to furnish a new means of approach to the long-term problems in the study of cultural artifacts (McCarty 2007).

Humanistic informatics, a term used mainly in Europe, teaches humanities-based approaches to the rapidly developing computerized society. The field is quite broadly defined, and includes courses in humanities computing, basic introductions to how computers work,

historical developments of technology, technology and learning, digital art and literature and digital culture. It focuses on the effects of linguistic and visual symbols and face-to-face communication, culture and media. It also incorporates sociology science on values, conduct and organization in human socialization, and is also concerned with technical science with particular focus on how information technology is designed and developed in accordance with the people using it. As such, it is therefore interdisciplinary: a field, which has grown concurrently with the development of communication media.

A number of DH initiatives have emerged in recent years, of which a few should be mentioned:

- The **Text Encoding Initiative** provides standards that enable libraries, museums, publishers, and individual scholars to represent a variety of literary and linguistic texts for online research, teaching and preservation. It is a consortium of institutions and research projects which collectively maintains and develops a standard for the representation of texts in digital form. Originally sponsored by the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing, the TEI is now an independent membership consortium, hosted by academic institutions in the US and in Europe.
- The British **Arts and Humanities Data Service (AHDS)** is a national UK service aiding the discovery, creation and preservation of digital resources in and for research, teaching and learning in the arts and humanities, and currently covers archaeology, history, visual arts, literature, language and linguistics, and the performing arts.
- The University of Michigan **Humanities Text Initiative** is an umbrella organization for the creation, delivery, and maintenance of electronic texts, as well as a mechanism for furthering the library community's capabilities in the area of on-line text. It has been providing on-line access to full text resources since 1994. It is part of the university's Digital Library Production Service (DLPS). Created in 1994, its origins are in the University Library's 1988 efforts to create an Internet based 'textual analysis' capability through a service then known as UMLibText. Since its creation, the HTI has amassed perhaps the Internet's largest and certainly the richest collection of materials in SGML. As well as creating text collections available to the Internet community and working with scholars on the creation of new electronic editions, the HTI supports the delivery of externally created SGML collections and has collaborated with publishers and with other academic operations to design and build local access mechanisms for their titles (Powell and Kerr 1997).
- The **Perseus Digital Library** at TUFTS University is an evolving digital library whose primary goal is to bring a wide range of source materials in the field of humanities to as large an audience as possible with the aim of strengthening the quality of questions, leading to new avenues of research and connecting more people through the connection of ideas. It is a heterogeneous collection of texts and images pertaining to the Archaic and Classical Greek world, late Republican and early Imperial Rome, the English Renaissance and 19th Century London. The texts are integrated with morphological analysis tools, student and advanced lexica, and sophisticated searching tools that allow users to find all of the inflected instantiations of a particular lexical form. The current corpus of Greek texts contains approximately four million words by 33 different authors. Most of the texts were written in the fifth and fourth centuries B.C.E., with some written as late as the second century C.E. The corpus of Latin texts contains approximately one million five hundred thousand words mostly written by authors from the republican and early imperial periods. The digital library also contains more than 30000 images, 1000 maps, and a comprehensive catalogue of sculpture.
- University of Chicago **ARTFL Project**: The Project for American and French Research on the Treasury of the French Language (ARTFL) is a cooperative enterprise of the Laboratoire ATILF (Analyse et Traitement Informatique de la Langue

Française) of the Centre National de la Recherche Scientifique (CNRS), the Division of the Humanities, the Division of the Social Sciences and Electronic Text Services (ETS) of the University of Chicago. At present the corpus consists of nearly 2000 texts, ranging from classic works of French literature to various kinds of non-fiction prose and technical writing. The eighteenth, nineteenth and twentieth centuries are about equally represented, with a smaller selection of seventeenth century texts as well as some medieval and Renaissance texts. A Provençal database that includes 38 texts in their original spellings was recently added. Genres include novels, verse, theater, journalism, essays, correspondence and treatises. Subjects include literary criticism, biology, history, economics and philosophy. In most cases standard scholarly editions were used in converting the text into machine-readable form and the data contain page references to these editions.

There are a few issues, of course, with traditional academia 'accepting' the DH and recognizing it at par with other computing disciplines. This has to do with a number of issues – social, cultural, intellectual and even curatorial. Multimedia, documentaries, interactive games and other visual media tend to be considered 'entertainment' and 'popular culture' and as such are not being given the same status as, say, ancient Greek plays, despite the fact that Aristophanes' *The Birds* were played at festivals in ancient Greece and that the notion of 'high culture' is apparently now long dead. There is a need to train senior practitioners in the field of humanities on how to use the available software, develop packages that 'deliver the goods' and provide continuous IT support to faculty – something not yet available as a standard at academic institutions. Non-verbal knowledge-bearing objects are not yet as well understood as 'texts' in their classical definition and curators face issues with how such digital knowledge-bearing objects are to be preserved for the long term. There are not enough academic appointments in the field of DH. In the UK there are the King's College London and Glasgow, both of which have made appointments at the lecturer and senior lecturer levels and teach academic courses in humanities computing, and at Oxford, whose Humanities Computing Unit shares an appointment with the English Faculty. In Canada there is the Humanities Computing Centre at McMaster, which is in the process of making a second appointment at the assistant professor level, and the Canadian Institute for Research Computing in Arts, University of Alberta at Edmonton, which shares an appointment at the full professor level with English. In the USA there is only the Institute for Advanced Technology in the Humanities, University of Virginia, which shares a tenured appointment with English and is about to share another with computer science (McCarty 2007).

There is also the issue that quite a number of academics regard the computer out of ignorance as an object essentially alien to the concerns of the arts and humanities and so think it either a threat or a salvation from beyond their turf. Academic institutions change relatively slowly, academics even slower. At the same time, developments in IT and computing change at 'the speed of light'. People accustomed to unchanging texts find it intimidating to deal with programmes that need perpetual upgrading, and so-called 'short courses' in computing do not do anyone justice. And yet, research in cognitive sciences, education and artificial intelligence, not to mention natural linguistics, show that the relationship between man and machine, culture and technology, society and science is closer today than it ever was before.

Apparently, constraints in available technology and resources often inhibit the data capture of materials related to the humanities, and the materials and research methods of humanists present fundamental challenges for information technology research. The hardware and software used for the data capture of 'standard' documents is often inadequate for the ingestion of humanities materials. Humanists are an underrepresented group in IT research. Yet there are three major reasons why it is important to have more efficient data capture of materials relating to humanities to ensure the:

- Full representation of human knowledge in digital libraries
- Greater availability of humanities digital content could enable new modes of inquiry for humanists
- Extraction of knowledge from humanities content to pose challenges that will foster innovative information technology research (NSF 2002).

We are back to a neat definition of what DH aims to achieve. Humanists are on the move, albeit slowly and cautiously – time will show if the achievements merit the effort. No doubt, however, it will be interesting to watch.

References

Kings College. 2006. Introduction to the digital humanities [Online]. Available WWW: **http://www.kcl.ac.uk/schools/humanities/cch/digihum/**.

McCarty, W. 2007. What is humanities computing? [Online]. Available WWW: **http://staff.cch.kcl.ac.uk/%7Ewmccarty/essays/McCarty,%20What%20is%20humanities%20computing.pdf**.

McGann, J. 1997. Imagining what you don't know: the theoretical goals of the Rosetti Archive . [Online]. Available WWW: **http://jefferson.village.virginia.edu/%7Ejjm2f/old/chum.html**.

National Science Foundation (Centre National de la Recherche Scientifique). 2002. A data capture framework and testbed for cultural heritage materials: NSF Proposal. [Online]. Available WWW: **http://ldp.library.jhu.edu/vhost-base/ldp/documents/ITR_NSF_2002.pdf**.

Powell, C.K. and Kerr, N. [1997]. SGML Creation and delivery: the humanities text initiative. *D-Lib Magazine* July/Agust. [Online]. Available WWW: **http://www.dlib.org/dlib/july97/humanities/07powell.html**.

Schreiberman S., Siemens, R. and Unsworth,. J (eds). 2004. *A companion to digital humanities*. Blackwell: Malden MA.

About the author
Sam Berner (B.Ed., Dipl. LIS, Postgraduate Diploma in Information Management) is a principal of the company ECognus (Brisbane, Australia). She is a knowledge management consultant, assisting small to medium enterprises to benefit the most from their intellectual assets. ECognus also provides services in the area of tailored software applications and the digitization of business processes.