MediaBook

Summary

Source	Media	Headline	Date	Page
Irish Times Business	PRESS	Rebellion manuscripts go digital	06-Nov-2009	1
Irish Times Innovation	PRESS	MOVING THE LIBRARY ONLINE	06-Nov-2009	3

Rebellion manuscripts go digital

Behind the scenes of Dublin academia, a vast digitalisation of historical documents is under way. It may well raise the bar for archiving, writes **Karlin Lillington**

AT TRINITY College Dublin, 1641 is coming face to face with 2009, to the benefit of scholars but also the public.

A set of priceless manuscripts – aged, coffee-coloured and covered in the looping, spidery scrawl of many different 17th-century hands – is to emerge as a cutting-edge semantic web and documentation project accessible by all online.

The manuscripts are the famous – or, depending on your view, infamous – 1641 Depositions, eight volumes of 1,559 personal accounts of one of the most violent moments in Irish history, the 1641 Rebellion.

The depositions – witness statements taken mostly from Protestants – describe incidents of murder, rape and pillage but also many aspects of everyday life in the 1600s.

This rich material is valued by social, religious, historical and political researchers but is also the source of hundreds of years of bitter dispute over the truthfulness of the events described.

For many Northern Protestants, the depositions are part of a sense of national identity and they are as symbolic for the Protestant community in Ireland as the battles of the Boyne and Somme, according to the project's website, available at www.tcd.ie/history/1641.

"The 1641 Depositions are a major treasure, not just for Ireland but internationally," says TCD history professor Jane Ohlmeyer, one of the principal investigators on the project.

"So we wanted to make them available to a wider audience."

TCD has collaborated with Aberdeen University and Cambridge University for the project. IBM Ireland has been the main technology partner, with DJ McCloskey of the Language-Ware Group at IBM's Dublin Software Lab the contact point.

Funded by €1 million in grants from the UK's Arts and Humanities Research Council, the Irish Research Council for the Humanities and Social Sciences, and the TCD Library, the deposition project is scanning 19,000 pages of text from the depositions to produce a high-resolution image, which is then transcribed by hand, a stage which is close to completion, thanks to the work of three post-doctorates.

"These are very, very noisy texts, so they're being transcribed by three PhDs. No technology is going to be able to do that, and the level of historical expertise it takes to do these is quite extraordinary," he says.

Once transcribed using Text Encoding Initiative (TEI) guidelines, an agreed format for digitising documents so that they may be readily used and exchanged by scholars and libraries, the texts are marked up using IBM's Language-Ware software in XML (Extensible Markup Language), a set of rules for creating "metadata" by tagging data so that tagged objects are readily identifiable.

Such work would be impossibly time-consuming if tagging was done by hand, says McCloskey.

"There's a social network within these depositions themselves. We're building up a semantic network – a Web 3.0 view of that world," says McCloskey.

Ohlmeyer says this is a first for the humanities – LanguageWare has been used in the health sector and law enforcement, but not in this way.

IBM was interested in the effort because its researchers felt that if LanguageWare could tackle texts this complex and archaic, it could manage just about anything, says McCloskey.

"From a purely technical challenge view – which of course drives us – this is a perfect project. If we can analyse this, well, you can pretty much analyse anything."

Ohlmeyer agrees: "We have all this dirty data that makes a great sandpit for them to play in."

The fact that IBM's LanguageWare group is based out of the Dublin Software Lab made a project with such significant Irish historical ramifications especially meaningful for researchers, McCloskey says.

The digitisation has made the content of the depositions available to researchers in some startling new ways, says Ohlmeyer, revealing fresh, previously unseen patterns and relationships.

"Once data is marked up in TEI, the possibilities are just endless.

"For example, you can link an atrocity to a person to a place. We hope to do a sophisticated linguistic analysis on the depositions, and also are looking for grants to enable the data to be mapped."

The first tranche of data, from Ulster depositions, will be available next month, with the remainder to follow early in the new year.

Ohlmeyer sees the project, the first of its kind, as revolutionising research in the humanities because such large-scale views of complex documents have never been available before.

To that end, the research partners have applied for three further grants "to start creating a generic tool for humanities research and commercial applications", she says.

"By working with industry, we're creating tools that we could roll out to wider communities to work with texts. They could be of huge

interest for museums, for galleries, for libraries, social scientists, lawyers and many groups."

She notes the tools are open-source-code-based and therefore would be freely available, although there is also the possibility of creating IP (intellectual property) for Trinity and perhaps commercial possibilities that could flow from that.

"But in terms of what's driving it from our end, it's pure research," she says.

"What was clear to me is this kind of overlap between the humanities and technology has not really been happening.

"This project is truly interdisciplinary and multidisciplinary."

The depositions project could become a marker for future directions for digitising resources and doing research, she believes.

"If we get this right, there's a paradigm shift here."

Scenes from the 1641 Depositions, eight volumes of almost 1,800 personal accounts of one of the most violent moments in Irish history, the 1641 Rebellion. The depositions – witness statements taken mostly from Protestants – describe incidents of murder, rape and pillage, but also many aspects of everyday life in the 1600s. They are to be accessible online from next month

MOVING THE LIBRARY ONLINE

Irish universities are getting in on the idea of digitised, interactive libraries, with a new partnership between Trinity College Dublin and Microsoft setting the tone

THE MOVE towards digitised university libraries is gaining momentum worldwide with copyright concerns creating the only real doubts about how the transfer to digital will pan out.

As part of a partnership announced recently Trinity College Dublin is to work with Microsoft on the development of a user-friendly online environment for its digitised works, which includes some historical documents and work from the humanities department. Dubbed a "Virtual Research Environment" the infrastructure will attempt to draw all of this content in to one network, making access and collaboration between academics easier.

However, while the deal is ground-breaking by Irish university standards the move towards a digitised and interactive library is not new internationally. In countries such as the US and UK most large universities have long-established digitisation units that are working to make printed works available online – particularly those of historical significance.

"In the US there's been quite a long tradition of digitising what's called 'special collections', which would be rare and unique things," says MacKenzie Smith, associate director for technology in the MIT libraries. "Quite a few institutions in the UK have also done that . . . typically they're put on a website in public."

In many cases the main motivation for digitisation has been preservation of work. However, the most important driver has been increasing the accessibility of content to a wider audience.

"[Some works] are very difficult things to get to today unless you're a researcher with a lot of money to travel," says Smith. "I think access is the bigger problem although preservation is part of that."

Smith adds that the rate at which this move to digital is happening has increased significantly of late, particularly since Google undertook its aim to digitise millions of books for its Book Search service.

Digitisation is also happening at the other end of the timescale among modern books, although in this case at source. Many publishers now – particularly in the science and technology fields – accompany their printed works with "ebook" and "ejournal" editions which can be read on a standard computer or compatible eReaders.

"The trend is very much to go electronic in all areas," said John Kennedy, systems librarian at the University of Ulster. "There are a number of pluses in doing that; one of which is that you can have several people access something at the same time, another is that they can do so whenever and wherever they want.

"Another big benefit for us is that it cuts down on the kind of space we need to stock the physical copies in the first place."

Of course when publishers create digital copies of their own, the ownership rights around them are relatively clear – the same cannot be said for work produced long before the digital age.

This fact is the one thing creating real concern among universities, as there is legal uncertainty over whether they can digitise at all under existing agreements.

As yet, no one knows how far "fair use" stretches and libraries are having to seek agreement with rights-holders on a case-by-case basis in order to free up the work for users.

A failure to get such clearance could have a dramatic impact on the potential of the digitised library, which can be used to turn static work from a book into something adaptable and modular. In fact, a more integrated and layered communication platform is an integral part of what Microsoft is promising Trinity College, including what it calls web 2.0 features like tagging, instant messaging and RSS feeds.

"When something is digitised the computer has the power to process the information in a way that I don't have the time to do," says Prof Jane Ohlmeyer, the Erasmus Smith chair of modern history at Trinity College. "You want to be able to get the machine to do all that tedious cross-matching for you."

At its most basic level this social media edge allows people to edit and discuss documents in real time from opposite sides of the globe but the benefits can go far deeper than this. One example often cited is the new potential for data "mashups", which involves the combination of two independent sources of information to create something entirely new.

Where in the past this may have involved laborious grunt-work on the part of the academic as they siphoned isolated pieces of information from pages to compare and contrast, attaining such results is now as easy as the click of a mouse.

"The trajectory here is pretty clear – eventually all of our texts will be digital," says MacKenzie Smith. "I don't think there's any debate about whether we're going to create digital libraries, it's just a question of how quickly we can make it happen." ∎

Another big benefit for us is that it cuts down on the kind of space we need to stock the physical copies in the first place
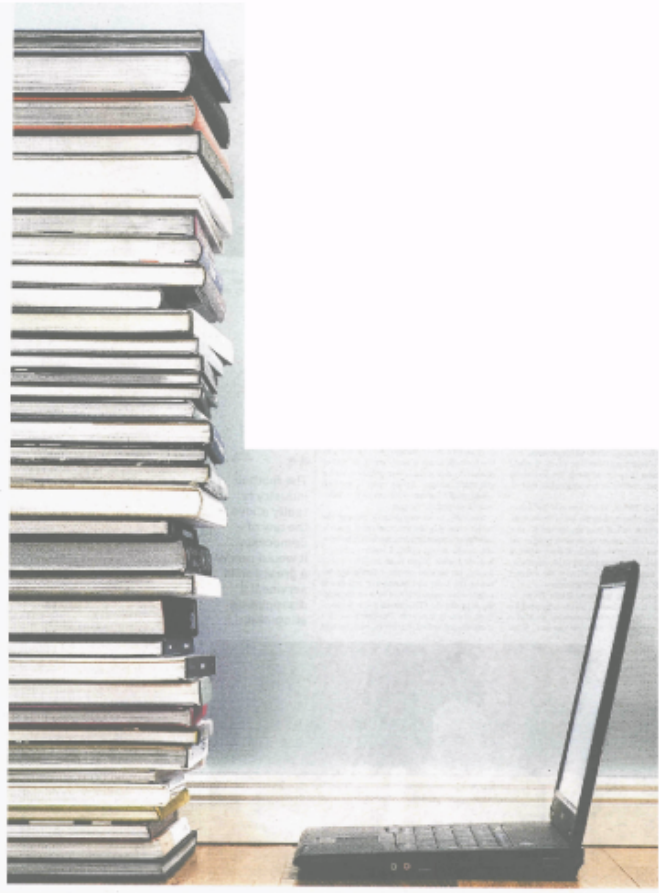
Fast and flexible Microsoft and Trinity College

TONY HEY, the man spearheading Microsoft's partnership with Trinity College Dublin, is "genuinely excited" by its potential despite admitting he had previously ignored the potential for innovation in the humanities.

Hey, who is corporate vice president at the external research wing of Microsoft Research, says that he has since seen the folly in this and understands just how important the area of study is to the research work Microsoft and others are doing at present.

"My work in the past has generally been around science which lends itself more to cutting-edge technology and new research techniques," says Hey while in Trinity College for the announcement on the partnership. "There is so much content there, though, and making all of it more accessible is very exciting."

With that increased accessibility comes the potential for greater collaboration according to Trinity College and Microsoft. However, Hey argues that the infrastructure now being created through the agreement has potential far beyond academia, stretching into the wider world of ed-

Tony Hey is at the forefront of the partnership between TCD and Microsoft

ucation and the country's economy.

"If you look at the example of the Book of Kells, people going to see it only ever get to see two pages at any one time," he says. "If it was all digitised and available to view at any time, anywhere it would have an enormous benefit to Ireland's culture and would vastly increase the interest in Ireland globally.

"Something like that would also be a valuable educational tool and I think the case needs to be made to the Irish Government that this could be very useful."

However, Hey makes no pretence

about Microsoft's motivation in striking the deal with Trinity, adding that it was going to benefit both sides.

"It would be dishonest to say that this was entirely altruistic – of course it wasn't," he says. "This partnership will be very useful to our research department and what comes out of it could eventually become features and services in future products."

Hey adds that his preferred approach is to build something tangible as quickly as possible, growing it from there. The plan is that as the software and infrastructure develops what works can be built on and what does not can be discarded.

The infrastructure is also being built heavily on open source code, something that Microsoft has begun to embrace after years of resistance. By making itself faster and more flexible – and by establishing partnerships like this one early – it helps Microsoft stay ahead of its competition.

Microsoft abandoned its short-lived plans to digitise books en-mass in the middle of 2008 opting instead to work with institutes on a one-by-one basis.