Published on *Digital Agenda for Europe* (https://ec.europa.eu/digital-agenda)

Making culture digitally accessible

Published by Editor Connect on 05/07/2012

[1]

Mass digitisation requires national libraries, museums and archives to scan their collections and create electronic transcripts. An EU-funded project has now established a European centre where institutions can access experts, advice and cutting-edge technologies to support their digitisation programmes.

Share this

Date:

05/07/2012

Venue:

Speaker:

Europe has a rich cultural history, but until recently much of it has been almost impossible to access. Millions of books, documents and other printed materials have sat in boxes and dusty shelves, carefully preserved and guarded by museums, national libraries and archives. Only a few of these documents ever saw the light of day and only the rare scholar could visit institutions to blow off the dust to discover hidden gems.

But with mass digitisation, all this is changing. For more than a decade the European Commission has led the campaign to create a European-wide digital library. Cultural institutions have been encouraged (if not compelled) to digitise collections. For printed materials this can be done en masse. Millions of printed words have been scanned, automatically transcribed, and made fully searchable and accessible through Europeana, the central portal for the European digital library.

The FP7 project 'Improving access to text' (IMPACT [2]) has spent the last four and a half years working hard to support this effort. Its technological partners have developed a suite of post-scanning

software tools to improve the fidelity of digital transcriptions [3].

However, Hildelies Balk, the project's coordinator, argues that technological advances alone are not enough. 'Mass digitisation is such a large-scale task; we already have millions of pages now available electronically and online, but this represents just a tiny fraction, perhaps only 1 %, of the historical preserved material. Mass digitisation still needs support: institutions want guidance on the best technologies to use, support for implementing tools in their productive environment and help in setting up and managing digitisations programmes. It is a problem for the vast majority of libraries, museums and archives in Europe.'

So IMPACT has complemented its technology developments with efforts to support the strategic direction on mass digitisation and build the capacity of institutions to participate effectively in this area.

The project has provided training and support to staff involved in mass digitisation. IMPACT offered a help desk system which acted like a broker, matching end-user requests to project partners and to other digitisation experts. An established training programme dealing with large-scale digitisation issues and technologies was also made available on the project's website.

One of the major achievements of the project, however, has been its development of a technological framework, or architecture, which brings all the mass digitisation tools and technologies into a single place and ensures that technologies - those that are commercially available and those developed by Impact, for example - are all interoperable.

'The architecture we propose, and which has been adopted by all the project partners, is like the glue that binds everything together,' explains Clemens Neudecker, IMPACT's technical manager. 'It allows people to integrate different technologies and processing methods and offers a graphical interface, so projects are easy to manage. You can add whatever software or processing tools you have to the architecture and simply drag and drop files through a sequence of tools to refine and improve the electronic transcripts.' The IMPACT framework will provide libraries, museums and archives which are just starting a digitisation project with important business information. A suite of evaluation tools and resources will help them to decide upon the most effective combination of available tools for a collection.

'We wanted to allow libraries and archives to choose whatever software or systems they wanted and use them in whatever order they wanted,' Mr Neudecker continues. 'We don’t want people to worry about file formats, conversions or interoperability. The framework handles all this, as well as the challenge of scalability.'

The IMPACT project is due to end in June 2012, but the collective expertise of the partners and their experience of using and developing digitisation tools is now being opened up to the mass digitisation community through the IMPACT Centre of Competence. The everyday administration of the Centre and its help desk will be managed by the Miguel de Cervantes Virtual Library and the University of Alicante, Spain. The computing infrastructure and storage will be provided by the Poznan Supercomputing and Networking Centre, Poland. The main dataset for the IMPACT project is hosted by the PRIMA institute at the University of Salford, in the United Kingdom, and now has more than 500,000 digital images from the IMPACT partner libraries with more than 50,000 ground truth representations.

'The IMPACT partners are committed to keeping the momentum going,' remarks Dr Balk. 'We have developed so much expertise over the course of the project and want to make this available to

support institutions and enable their mass digitisation projects. The IMPACT Centre of Competence brings together the three main communities in this domain: content holders, researchers in the domain of image science, OCR and language technology, and the mass digitisation service providers, for example OCR software vendors, who are keen to engage with the digitisation community.' All three communities benefit from their interactions with the others.

The Centre is funded through subscriptions which cost EUR 10 000 for private bodies or companies and EUR 6 000 for public entities. When members contact the centre for advice, support or services they are directed to the most appropriate resources, tools and expert institutions from among the IMPACT partners.

'You cannot do mass digitisation on your own,' concludes Dr Balk. 'Cooperation is vital and the IMPACT partners now have years of experience working together in this field. Through the Centre of Competence we are now prepared to share our knowledge and experience with others and drive forward this exciting vision to really open up the wealth of historical resources we have in Europe.'

The IMPACT project received EUR 12.1 million (of total EUR 17.1 million project budget) in research funding from the EU's Seventh Framework Programme (FP7) under its ICT theme.

Useful Links:

- 'Improving access to text' project website [2]
- IMPACT project factsheet on CORDIS [4]
- ICT Challenge 4: Digital libraries and content [5]
- Impact Centre of Competence [6]
- Europeana [7]

Related Articles:

- Feature Stories - From the printed page to bits: new tools for mass digitisation [3]
- Feature Stories - Digitising our cultural heritage [8]

Country: NETHERLANDS

Information Source: Hildelies Balk, Head of European Projects at the National Library of the Netherlands

Hildelies Balk, Head of European Projects at the National Library of the Netherlands
Contact:

Newsroom Item Type:

- Projects news and results [9]

Around Europe & the World:

- Netherlands [10]
- Poland [11]
- Spain [12]

Source URL: https://ec.europa.eu/digital-agenda/en/news/making-culture-digitally-accessible

Links
[1] https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/newsroom/offer_id_8916_3040_0.jpg
[2] http://www.impact-project.eu/
[3] http://cordis.europa.eu/fetch?CALLER=OFFR_TM_EN&ACTION=D&DOC=3&CAT=OFFR&QUERY=01384cec02a5:3ba9:22d4319c&RCN=8874
[4] http://cordis.europa.eu/projects/rcn/85383_en.html
[5] http://cordis.europa.eu/fp7/ict/programme/challenge4_en.html
[6] http://www.digitisation.eu/
[7] http://www.europeana.eu/portal/
[8] http://cordis.europa.eu/fetch?CALLER=OFFR_TM_EN&ACTION=D&DOC=2&CAT=OFFR&QUERY=0137ffcea2f0:a4a4:20842f59&RCN=8696
[9] https://ec.europa.eu/digital-agenda/en/newsroom/all/projects-news-and-results
[10] https://ec.europa.eu/digital-agenda/en/country/netherlands
[11] https://ec.europa.eu/digital-agenda/en/country/poland
[12] https://ec.europa.eu/digital-agenda/en/country/spain