

ETHOS: a national OAI and digitisation service for e-theses in the United Kingdom

Jill Russell, Chris Awre, Anthony Troman, and Paul Needham

It has long been the goal of the academic library sector in the UK to make more effective use of the many theses they receive every year from postgraduate and masters research at their institutions. Whilst holding the theses in the Library makes them available for all those willing to travel, or via microfilm through the British Library, their use is limited and does not reflect the depth of research and knowledge contained within them. As the use of Microsoft Word etc. increased the creation of theses in a digital form, it became more feasible to consider how these e-theses could be used for improved management and access. Many institutions have moved to incorporate e-theses within their institutional repository developments: this both supports institutional preservation practices and allows the e-theses to be exposed publicly for wider dissemination and access via the repository website and OAI-PMH.

The use of OAI-PMH has become almost ubiquitous now in commonly used repository software, allowing the metadata and contents of the repository to be exposed for harvesting by interested parties. This relative boom in the number of OAI data providers has made a huge impact on the amount of digital content available for access around the world. However, for that access to be beneficial to end-users there needs to be equivalent development of OAI service providers to carry out the harvesting of the repositories and serve the aggregated results to intended audiences. The number of OAI service providers has not kept pace with repository development, and there is a danger that the good intentions of repository development will not be fulfilled.

One of the main reasons for this is the lack of mature business models to support OAI service providers, especially where harvesting is part of an open access publication model. The ETHOS feasibility project in 2005-6 (<http://www.ethos.ac.uk>) was funded by the JISC to identify how a financially viable and sustainable national OAI service provider for e-theses could be established. The model produced received a positive independent evaluation and JISC has now funded the ETHOSnet project to take the prototype developed in the feasibility project and develop this into a full service, to be launched in August 2008.

The ETHOS model is a combination of established OAI service provider architecture and the addition of added value digitisation and preservation services that provides a complete package to underpin the sustainability of the service. For a national service to work effectively harvesting must be feasible from all interested educational institutions, regardless of which

repository software is in use. Therefore, clear guidelines are being developed to guide users of all currently used repository software systems: these include EPrints, DSpace, Fedora, BePress's Digital Commons, and Ex Libris's DigiTool. Others will be added as and when they are adopted within the UK. The guidelines are being encapsulated within the wiki-based EThOS toolkit (currently at <http://ethostoolkit.rgu.ac.uk>), which also provides guidance to institutions on the development of e-theses within their own repositories: such guidance will lay down the requirements EThOS has to operate as an effective harvester and service.

The EThOS feasibility study developed a qualified DC schema for UK e-theses, labelled UKETD_DC (http://ethostoolkit.rgu.ac.uk/?page_id=72), and the repository guidelines largely encompass how this can be implemented within each system and how the metadata, and content, should be exposed for harvesting: a key part of the EThOS model is that the content of the e-thesis, and not just the metadata, should be harvested. To make this possible, the dc:identifier field within the UKETD_DC record contains a link to the content binary file, with dc:relation used for the jump-off or splash page, an approach not commonly used in other DC metadata records. Plugins currently exist for EPrints and DSpace to assist with this, whilst work is ongoing for guidelines for the other three systems. The full text content of the e-thesis is being harvested to aid delivery and preservation, though it will only be used in accordance with the wishes of the providing institution.

The central hub harvester resides at the British Library, and has been developed around the EPrints 2 software. The prototype is now being scaled up to meet service demand, and will deliver a one-stop access point to all UK e-theses, plus metadata records of bound theses. The user interface links this discovery service with locate and request services. Full text e-theses discovered via EThOS will provide a link to the full text from the metadata record. This link can either be to the home repository, or to the copy held at the British Library. Following the link will also reveal a range of added value services available from the British Library, based around the sale of printed or CD versions of the thesis (whereas online electronic access will remain free to meet the principle of open access).

To meet JISC's requirements for a financially viable and sustainable service, the British Library has developed a cost-recovery model for EThOS. Whilst the sale of non-Internet versions of the e-thesis will provide a degree of income, it cannot cover the costs of the EThOS service in its entirety. As a way of extending EThOS beyond currently available e-theses, the British Library will be offering a digitisation service for older bound theses. Institutions wishing to take advantage of this service will be charged an annual subscription fee, which will both pay for the digitisation and maintain the OAI service running alongside: the two services will be operated and presented seamlessly as one. Digitisation priority will be by user demand, with institutions able to nominate additional theses for digitisation as can be encompassed within their subscription. Digitised copies of the theses will be

retained by the British Library for preservation, but can also be delivered back to the home institution for ingest into their local repository.

The EThOSnet project will deliver not just the EThOS service, but will act as an advocate of the service to the community and help to generate both interest and engagement in participating with the service. A Technological Development workpackage is also looking to the future; examining the development of e-thesis metadata, compound e-theses, plagiarism, discovery to delivery workflow enhancements and the architecture of the EThOS service.

The presentation at OR08 will describe the development of an OAI service provider to serve across open repositories at all universities in the UK, and the associated business model that will help this service to be sustainable into the future.

7th December 2007