THE DART-EUROPE PROJECT: TOWARDS DEVELOPING A EUROPEAN THESES PORTAL

Authors
Paul Ayris, Director of UCL Library Services and Copyright Officer, UCL
Chris Pressler, Director of Learning Resources, Dartington College of Arts
Martin Moyle, Team Leader (Science), UCL
Austin McLean, Director, Dissertations Publishing and Scholarly Communication,
        ProQuest Information and Learning

Abstract
        This paper will report on the new European theses project DART-Europe. The purpose of this project is to align institutional and national e-theses developments across Europe with the wider open archives movement by the construction of a European portal for research theses, thus enabling a global view of European institutional research assets. This project is driven through an innovative partnership between an information provider and an international body of university libraries and open access consortia. The project's goal is to explore the creation of a European model for the deposit, discovery, use and long-term care of research theses in an open access environment.
        The paper will outline the projected outcomes of DART-Europe, which is an active group of institutions in addition to a technical service. To this end, DART-Europe is engaged with disciplines and institutions that are widening the definition of research by redefining the formats of theses.
        For institutions and countries without a repository infrastructure, DART-Europe will enable the creation of a depository. Institutions and countries with a repository infrastructure can engage with DART-Europe to deliver their e-theses. DART-Europe acts as a technology bridge for researchers between those who have existing infrastructures and those who do not. The DART-Europe architecture assumes free at point of use access to full text theses, whether held on the DART-Europe server or by institutional repositories.
        This paper will provide session attendees with the current progress of this initiative, including a report on the 5 strands of the project, including: architecture; creation of a management tool kit; content acquisition; digital preservation and an investigation of business models.

Keywords
E-Theses, Open Access, Institutional Resources, Scholarly Communication, DART-Europe

The Open Access movement in the UK

        The movement to develop Open Access repositories in Europe has gathered considerable momentum over the last five years. In the United Kingdom, the SHERPA project has acted as the spearhead for the majority of research-intensive universities in the Russell Group to set up open access repositories. At the time of writing, SHERPA[1] has overseen the creation of 20 repositories in research-led institutions.
        A major driver for the establishment and population of the repositories is the RAE exercise in the UK. The Research Assessment Exercise is next due in 2008. It is of fundamental importance to universities, because it dictates which universities receive significant amounts of research funding from the Funding Councils. Each

---

[1] http://www.sherpa.ac.uk/.

submitting author has to submit four research outputs to the relevant subject panel, and these are scored by that panel. The present writer submits via UCL's History Department as part of UCL's RAE return.

Many universities want to link their in-house repository development to their RAE return. Copyright permissions can be cleared via the SHERPA-Romeo listing[2] and the full-text of papers mounted in repositories. The JISC (Joint Information Systems Committee) is also funding some studies and development work to facilitate this process.[3]

SHERPA has had a major impact on the speed and nature of the take-up of repositories in the UK and, as Chair of the SHERPA Management Committee, I applaud the work of the SHERPA team in their astonishing achievements.

Government Inquiry

A second major development in scholarly communication issues in the UK has been the impact of the House of Commons Select Committee's Report *Scientific Publications: Free for All?*[4] There has been enormous publicity for this Report in the media – television, radio, the broadsheets. The Government response to date has been disappointing and non-committal. The relevant Government Department, the Department of Trade and Industry, has committed only to maintaining a level playing field between commercial and Open Access initiatives and has set up a Research Publications Forum, on which the writer sits, to engender debate.

The Funders

A much more positive role has been played by the funders of research in the UK – particularly the Wellcome Trust and Research Councils UK. The Wellcome Trust has already mandated deposit for research outputs emanating from its funding grants. On 19 May 2005, the Trust issued a press release announcing changes to its conditions of grant. In relation to grants awarded after 1 October 2005, recipients will be **required** to deposit a copy of the final author manuscript in PubMed Central, or in the UK equivalent which the Trust is establishing, upon its acceptance for publication in a peer-reviewed journal. From 1 October 2006, this condition will apply to **all** recipients of Wellcome Trust grants, no matter when the grant was awarded.[5]

Research Councils UK (RCUK),[6] at the time of writing, are out to consultation on a statement on research outputs which is similarly minded to mandate deposit in Open Access repositories for research outputs (journal and conference papers) which are funded through its grants. It will be a momentous step if RCUK does indeed mandate deposit and the final statement is expected at some stage in the summer of 2005.

Open Access developments elsewhere in Europe

The United Kingdom is one small, but nevertheless influential, collection of countries in the European Union. All over Europe, countries are engaging in scholarly communication developments and advocating the use of Open Access repositories.

In the Netherlands, for example, a meeting in May in Amsterdam, sponsored by the JISC/SURF/CNI, celebrated the achievements of the DARE programme in creating Open Access repositories in all Dutch universities.[7] DARE also launched the Cream of Science, an Open Access network for scientific output by Dutch academics.

---

[2] http://www.sherpa.ac.uk/romeo.php.
[3] http://www.jisc.ac.uk/index.cfm?name=programme_digital_repositories.
[4] http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm.
[5] See http://www.wellcome.ac.uk/assets/wtx022821.pdf. The press release is at http://www.wellcome.ac.uk/doc_WTX025191.html.
[6] See http://www.rcuk.ac.uk/.
[7] See http://www.darenet.nl/en/page/language.view/home.

More than 200 top scientists in the Netherlands with 41,000 publications are now accessible for everyone. About 60% (25,000 papers) comprise full-text content, which is available.

Mirroring the Report of the House of Commons Select Committee in the UK, the EU is also funding a study of the scholarly publishing market. This Report is being undertaken by teams of library professionals and economists from Brussels and Toulouse, and will be finished by the summer of 2005. Early announcements of its findings, at the UNICA meeting in Vienna in 2004, indicate that their conclusions will be different from those of the House of Commons Report in the UK.

Examples of multi-national co-operations across Europe in Open Access initiatives are relatively few. Two exceptions are the new OpenDOAR listing of open archive repositories and the OAI Workshops held under the aegis of LIBER, the European Research Library Consortium. OpenDOAR[8] is a partnership between CURL, JISC, SPARC Europe and the Open Society Institute. The OAI meetings at Cern, the Organizing Committee for which I chair as Chair of the LIBER Access Division, are a major international player in disseminating technical and library advice and guidance on the use of the OAI protocol for metadata harvesting. The fourth in this highly successful series of Workshops is being held in October 2005.[9]

The role of E-Theses in Europe

There are a number of successful developments in Open Access in Europe. E-Theses, however, are not universally a driver for such initiatives. There *are* some exceptions. DiVA, the consortial Nordic E-Thesis initiative, has (at the time of writing) 3243 research dissertations from 12 institutions.[10] In the Republic of Ireland, there is great interest in e-thesis developments but no real activity in terms of the storage and dissemination of these materials. Until very recently, the same was true of the United Kingdom. There have been years of discussions over E-Theses, with no delivery of content. This has begun to change through the JISC-funded project called EThOS, run by Glasgow University in collaboration with the British Library.[11] The purpose of EThOS is to deliver a fully operational, easily scaleable and financially viable prototype UK e-theses online service, and supporting infrastructure, that will enable students, researchers and other end-users to search a UK Database of Theses (UKDoT) and to access, from the desktop, the full text, in secure format, of electronically stored theses, following selection from the UKDoT. It will also enable HE institutions, in partnership with the British Library, to promote and increase usage of their theses output. This is important work and EThOS deserves to be a successful exemplar project.

What is lacking in Europe is a co-ordinated multi-national approach to the discovery and retrieval of primary doctoral research in research theses. It is this lack that DART-Europe is designed to remedy.

DART-Europe: the partners

The conviction which unites current DART-Europe partners is the assurance that research dissertations are sources of information and knowledge which are little known and under-used and that this information should be made available to the global research and learning community in an Open Access environment, which is underpinned by the Berlin Declaration on Open Access.

UCL is a leading research-intensive University in the UK and is one of the two lead partners in the project. The second lead partner is Dartington College of Arts, which has an international reputation for the creative and performing arts. The

---

[8] See http://www.opendoar.org/.
[9] See http://oai4.web.cern.ch/OAI4/.
[10] See http://www.diva-portal.org/.
[11] See http://www.jisc.ac.uk/index.cfm?name=project_ethos.

University of Oxford is a world-renowned University which has a world class research record. Trinity College Dublin is likewise a major international University with impressive research outputs.

ProQuest Information and Learning[12] is a global information provider, based in Ann Arbor, Michigan, and Cambridge, England. ProQuest is a leading publisher of information solutions for Higher Education and has a very successful global track record of work with electronic dissertations.

UCL, Dartington College of Arts and ProQuest Information and Learning have established a research partnership with the University of Oxford and Trinity College Dublin to investigate the creation of a European model for the deposit, discovery and long-term use of research theses in an Open Access environment. The partnership, which will be aligned with national e-theses developments in each member country of the European Union, aims to create a unified search service and value-added services to gain much greater visibility for European research outputs which are embedded in research Masters and Ph.D. theses.

DART-Europe: objectives

DART-Europe has a project plan which is divided into eight interlinked workpackages:

1: creation of the DART-Europe Repository
2: creation of an E-Theses management toolkit
3: creation of the DART-Europe E-Theses portal (DEEP)
4: E-Theses support, primarily in terms of IPR and metadata
5: application of digital preservation within the project
6: identification of an e-theses business model
7: dissemination and advocacy
8: project management

The objectives of DART-Europe are:

- to specify and build the DART-Europe repository, enabling centrally-hosted repository services to be offered to European academic institutions and consortia.
- to create the DART-Europe E-theses portal, populated with e-theses metadata derived from local institutional/consortial repositories, and metadata from the DART-Europe repository.
- to create an e-theses management toolkit, to provide administrative support for the acquisition and deposit of e-theses at an institutional level.
- to investigate and publish "best practice" guidelines on aspects of e-theses management, including metadata standards and quality, copyright, and digital rights management.
- to examine the life-cycle of theses in non-traditional formats, in order to facilitate both their curation, and their exposure through the DART-Europe E-theses portal.
- to enable the long-term storage and preservation of e-theses in a trusted digital repository.
- to identify and address multi-lingual issues as they relate to the DART-Europe E-theses portal.
- to secure external funding for the medium-term development of DART-Europe.
- to construct a business model to underpin the long-term sustainability and development of DART-Europe's services.
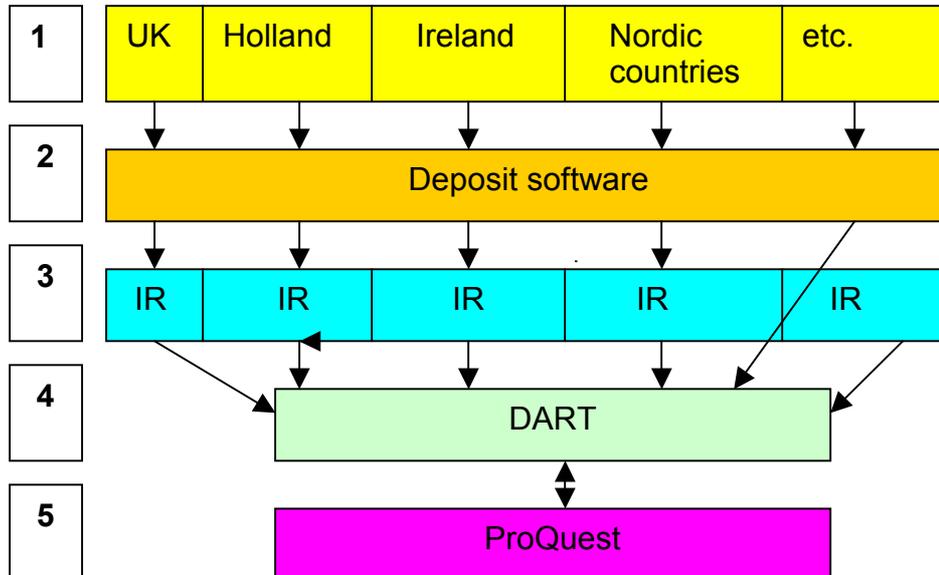
---

[12] See http://www.il.proquest.com/.

- to disseminate widely the work of the project and vigorously to advocate Open Access to e-theses, repository-oriented approaches to e-theses management, and the DART-Europe services.

DART-Europe architecture

The DART-Europe architecture works on four distinct, but inter-related, platforms:

1	UK	Holland	Ireland	Nordic countries	etc.
2	Deposit software				
3	IR	IR	IR	IR	IR
4	DART				
5	ProQuest				

These platforms correspond to:

1. Institution
2. Deposit
3. Local platform and local access
4. International platform
5. Preservation platform

Project timescales

The project will be carried out in at least 2 discrete phases. Phase 1 will begin in June 2005, and will last for 18 months. It will involve the evaluation of existing e-theses technologies, the establishment of the DART-Europe repository and the DART-Europe E-theses portal, the creation of an e-theses management toolkit, with supporting management guidelines, and the demonstration of preservation services. Exploratory work on multilingual aspects of the project will be undertaken. News of the project and its outputs will be disseminated during phase 1 through a project website and appropriate conferences and publications. Endorsement of DART-Europe in this phase will be sought from LIBER (Ligue des Bibliothèques Européennes de Recherche), as the pan-European library consortium.

Phase 2 will see the enlargement of the DART-Europe partnership, with special emphasis to be placed on new EU accession countries. Scoping studies, recommendations, and other outputs from phase 1 will be taken forward. Detailed work on multilingual aspects of the project will be carried out. External funding will be required for the implementation of phase 2 of DART-Europe: this will be sought from the European Commission.

The DART-Europe Repository

A central DART-Europe repository will be created. This will enable DART-Europe to offer hosted repository services, tailored to institutional requirements, to institutions without local repositories. There will also be scope for the creation of centrally-hosted consortial repositories, national repositories, and collaborative repositories in other configurations, within the DART-Europe repository.

Institutions who run local e-theses repositories, but who also wish to deposit an additional copy of locally-hosted e-theses in the DART-Europe repository[13], will be able to do so. This phase of work will investigate the mechanisms necessary to allow the most efficient population of the DART-Europe repository with supplementary copies of locally-managed e-theses.

The deliverables will be:
- Specification and creation of the DART-Europe repository.
- Investigation of mechanisms in support of the supplementary deposit of copies of locally-managed e-theses in the DART-Europe repository.

The DART-Europe E-theses Portal

The aim of this phase of activity will be to specify and build the DART-Europe E-theses portal. The portal architecture will be founded on interactions between existing institutional/consortial repositories, and the centrally-hosted DART-Europe repository. An end-user interface will be developed.

The portal will collate metadata about, and be a gateway to, the full text of European e-theses. These e-theses may be held in the DART-Europe repository, or may reside on servers held by institutional or consortial participants. Metadata which is held remotely will be harvested from the relevant repositories into the DART-Europe index. The portal itself will be OAI-compliant.

The end-user interface will support basic and advanced search in citations, abstracts and full-text (where indexed). It will be possible to limit searches to selected single or multiple contributing organisations. Additional features of the platform will include OAI-compliance, e-mail alerts to researchers, and XML export. The requirements for extended multi-lingual support, both in terms of metadata about e-theses, and of end-user interfaces, will be scoped during this phase, and a search interface in one European language in addition to English will be created.

Management information on user activity will be available to all sites contributing content to DART-Europe. As a minimum, this will include regularly updated reports on the number of searches and downloads, and the origin of download requests.

The deliverables will be:
- Creation of an index to European theses acquired through harvesting of metadata from contributing OAI-compliant repositories, combined with metadata relating to e-theses hosted by DART-Europe.
- End-user functionality, to include basic and advanced search, e-mail alerts and XML export.
- Creation of search interface in a second language.
- Examination of multi-lingual metadata issues.
- Availability of management information reports to contributing organisations.

E-theses Support

The purpose of this activity is to provide guidance to participating e-theses collection managers and administrators, especially in the areas of copyright, digital

---

[13] For instance, to facilitate retrieval through full-text indexing, or to take advantage of DART-Europe value-added services, such as digital preservation.

rights management, and metadata. The outputs will be web-based, and will include guidelines on EU and national copyright legislation, and on digital rights management – for instance, on securing author permissions for commercial redistribution, and on dealing with commercial embargoes. Guidance on best practice in e-theses metadata will also be made available: this will include a definition of the minimum metadata standards and quality thresholds required for participation in the DART-Europe portal, as well as general recommendations on appropriate metadata standards and specifications for the management of e-theses collections.

Any recommendations arising from Digital Preservation activity will be incorporated into the outputs of this activity. Guidance on supporting the full life-cycle of e-theses in non-traditional formats will also be included.

The deliverables will be:

- Web-based dissemination of advice and best practice guidelines in copyright, digital rights management, and metadata quality and standards, as related to e-theses.
- Assistance to DART-Europe participants in achieving compliance with EU Copyright directives and relevant national legislation.
- Definition of metadata standards required for participation in the DART-Europe E-theses portal.
- Support for the e-theses community in the acquisition and management of e-theses in non-traditional formats.

Future directions
The following areas have been identified for exploration in future development phases of DART-Europe:

- Many pure science and performing arts theses include large quantities of data. DART-Europe will investigate the movement and storage of these large files through discussion with the UK E-Science Programme, with a view to aligning large digital theses with other large research outputs.
- The opportunities for deeper interaction between the DART-Europe management toolkit and registry systems at the institutional level.
- Citation analysis and linking.
- Identification of and application to possible funders for mass retrodigitisation of European theses. Candidate bodies include the Mellon Foundation, Google, Hewlett-Packard, and the National Endowment for the Humanities.
- Uploading of ProQuest's existing European dissertation holdings to the DART-Europe E-theses Portal.