

Main Articles

DRIVER: Building the Network for Accessing Digital Repositories across Europe

[Martin Feijen](#), [Wolfram Horstmann](#), [Paolo Manghi](#), [Mary Robinson](#) and [Rosemary Russell](#) present an outline of the DRIVER Project and its achievements so far in supporting and enhancing digital repository development in Europe.

[Main Contents](#)[Section Menu](#)[Email Ariadne](#)[Search Ariadne](#)

Introduction: Why DRIVER Is Needed

OpenDOAR [1] lists over 900 Open Access repositories worldwide. Approximately half of them are based in Europe, most of which are institutional repositories. Across Europe many more repositories are being set up and supported by national and regional initiatives such as the Repositories Support Project [2] in the UK and IREL-Open [3] in Ireland.

A recurring challenge for repositories is that of engaging researchers in Open Access and motivating them to deposit their work in OA repositories. Local efforts to embed the repository within the research processes of an institution are not always successful. What is needed is the embedding of repository use in research and research publication processes on a large scale, across Europe.

Researchers and indeed other users of digital information systems have high expectations for the provision of digital content. Retrieval should be fast, direct, and versatile. Ideally, retrieval of full text should be just one click away. The current state of institutional repositories does not fully support these expectations. While many valuable services such as OAIster [4] and BASE [5] have been established to search and retrieve bibliographic records (metadata), the resource itself is sometimes hidden behind several intermediate pages, obscured by authorisation procedures, not fully presented, or not retrievable at all.

What is needed is a unified approach to managing this challenging and evolving repository landscape. This approach must ensure a high level of interoperability across repositories and allow the development and improvement of retrieval services providing fast and efficient retrieval of content. DRIVER (Digital Repository Infrastructure Vision for European Research) [6] is an EU-funded project which provides such a unified approach. It is the largest initiative of its kind and is leading the way in supporting and enhancing repository development in Europe and indeed worldwide.

DRIVER makes possible the development of high-quality search and associated services for the

research community, which will enable effective retrieval and use of content held in repositories. A unified European approach will also ensure that repositories and their use become an accepted part of research and research publication processes across Europe.

This article outlines the aims of the DRIVER Project, its achievements so far, their implications for the European repository and research community and the future aims of DRIVER.

DRIVER Aims

DRIVER is a multi-phase effort, the initial phase lasting 18 months, and is funded under the EU Sixth Framework Programme, Research Networking Testbeds. It sets out to build a testbed for a future knowledge infrastructure of the European Research Area.

The DRIVER partnership has ten partners from eight EU countries and benefits from a wide range of experience. Many partners are already well known to the European repository community through their involvement in national and international repository initiatives (CNRS, SHERPA, SURF, UKOLN, University of Ghent) while other partners are experienced in the development of technical infrastructure and services (University of Athens, University of Bielefeld, CNR, ICM, University of Goettingen) to support such a community.

The DRIVER Project aims to deliver any form of textual scientific output, including scientific/technical reports, working papers, pre-prints, articles and original research data, to the various user groups.

Work on the current testbed began in June 2006 and has five main objectives:

- To organise and build a virtual, European scale network of existing institutional repositories from the Netherlands, the United Kingdom, Germany, France, and Belgium.
- To assess and implement state-of-the-art technical infrastructure, which manages the physically distributed repositories as one large-scale virtual content resource.
- To assess and implement a number of fundamental user services.
- To identify, implement and promote a relevant set of standards.
- To prepare the future expansion and upgrade of the digital repository infrastructure across Europe and to ensure widest possible involvement and exploitation by users.

Below we present an outline of the DRIVER studies, guidelines and technical infrastructure services, as well as the benefits and implications of DRIVER for research in Europe and elaborate on the future development of the project.

Studies

Within the DRIVER Project a number of strategic and co-ordinated studies on digital repositories and related topics have been carried out.

Maurits van der Graaf and Kwame van Eijndhoven (SURF) conducted an inventory study of OAI-compliant repository activities in the EU [7].

The *DRIVER's Guide to Repositories* edited by Kasja Weenink, Leo Waaijers and Karen van Godtsenhoven (SURF and University of Ghent), is due to be published by Amsterdam University Press in December 2007. It aims to motivate and promote the further creation, development and networking of repositories. It contains comprehensive and current information on digital repository-related issues particularly relevant to repository managers, decision makers, funding agencies and infrastructure services as stakeholders. DRIVER has identified five specific, complex and long-term issues which are essential to the establishment, development or sustainability of a digital repository: the business of digital repositories; stimuli for depositing materials into repositories; intellectual property rights; data curation; and long-term preservation.

The success of a repository is dependent on having addressed these five issues sufficiently. Good practice and lessons learned as part of this report will assist stakeholders in both their day-to-day and long-term challenges, and can help them avoid reinventing the wheel. The *DRIVER's Guide* focuses on international and trans-national approaches which go beyond local interests.

The *Investigative Study of Standards for Digital Repositories and Related Services* [8] by Muriel Foulonneau and Francis André (CNRS) reviews the current standards, protocols and applications in the domain of digital repositories. Special attention is paid to the interoperability of repositories to enhance the exchange of repository data. The study is aimed at institutional repository managers, service providers, repository software developers and all players taking an active part in the creation of the digital repository infrastructure for e-research. It aims to raise discussion on these topics and to support initiatives for the integration and in some cases the development of new standards, in addition to the current interoperability mechanisms that have been implemented in digital repositories. The study not only looks at the current situation, but also at the near future: what steps should be taken now in order to support future demands? This study is also due to be published in December 2007.

Repositories as Content Networks

The landscape of digital repositories is multi-faceted with respect to different countries, different resources such as text, data or multimedia, different technological platforms, different metadata policies, etc. However there is also a considerable degree of homogeneity across parts of this landscape: the main resource type provided by digital repositories is text (see Figure 1) and the common approach for offering textual resources is via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Therefore the current test-bed phase of DRIVER is focusing on textual resources that can be aggregated across many repositories and made directly accessible via OAI-PMH. These resources are mainly articles, but also include lectures, theses, reports etc. Direct access means that the user can download and use the full text of these resources with only a few clicks, anytime, anywhere, and without payment.

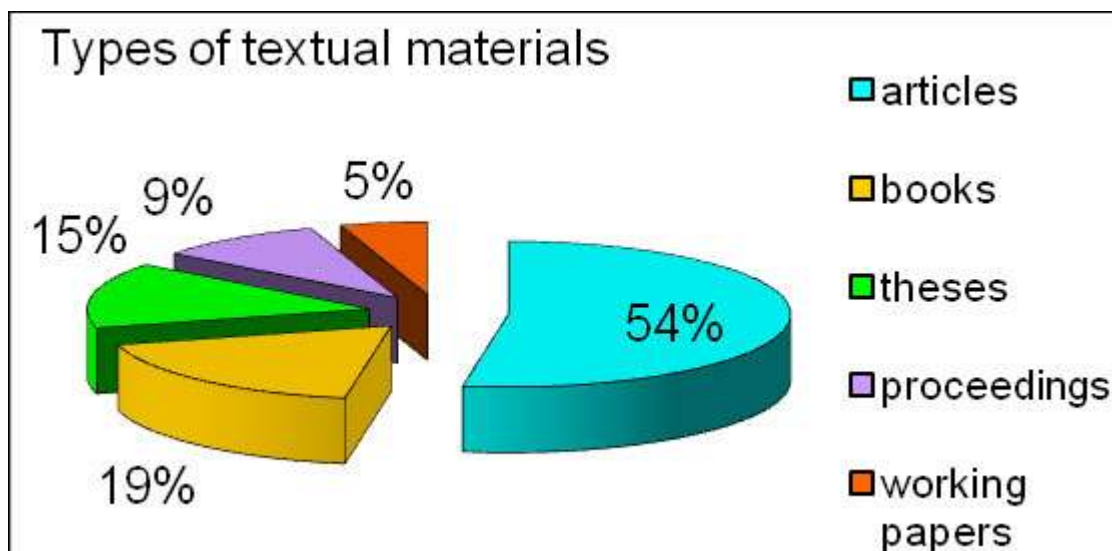


Figure 1: Types of textual materials held by digital repositories

Why Have Guidelines?

As distributed systems, repository networks critically depend on interoperability in terms of technology and content provision. Practical experience with repository networks using OAI-PMH (e.g. DAREnet [9], HAL [10]) reveals firstly that homogenous use of the protocol considerably strengthens the quality of services for the end-user. A number of additional issues around the use of the protocol have been defined and listed in *Investigative Study of Standards for Digital*

Repositories and Related Services [8]. They can be grouped around three themes:

- the harmonisation of the metadata content (e.g. use of the inverted name form in DC element i.e. Author: Smith, John)
- the uniform use of the OAI-PMH protocol (e.g. use of the transient parameter for deleted records)
- the lack of a mechanism for the transport of metadata that relate to a resource that consists of multiple digital files (e.g. a thesis that has 15 separate PDF files)

These themes have been addressed in DRIVER by highlighting the implications for local repository managers. He/she can support DRIVER locally by offering content in a specific manner. For this purpose the project has drawn up the *DRIVER Guidelines for Content Providers: Exposing textual resources with OAI-PMH* [11]. The main goals are to provide guidance for managers of new repositories in defining their local data management policies for textual resources; for managers of existing repositories to take steps towards improved services; and for developers of repository platforms to add supportive functionalities in future versions. Aspects of content provision other than exposing textual resources - e.g. managing versions or exposing scientific data - may be addressed in future versions of the guidelines.

Compliance with the guidelines is needed to enable the full range of services available through the DRIVER infrastructure. A search service, for example, that promises to list only records that provide a full-text link, cannot process all contents of a repository that offers metadata-only records or obscures full-text by authorisation procedures. Compliance with the guidelines will help to differentiate between those records. However repositories that do not conform to all the mandatory or recommended guidelines will still be harvested, but depending on the degree of conformance, content may not be retrievable. The guidelines do not, of course, prescribe which records should be held in a local repository. DRIVER offers support to local repositories to implement the guidelines on an individual basis. Support can be delivered through the Internet [12] or via the DRIVER helpdesk [13]. DRIVER is committed to any possible solution that can be realised by central data processing. However the sustainable, transparent and scalable road to improved services is via the local repositories.

Retrieval of full text with bibliographic data is a basic but necessary step forward to approach rich information services based on digital repositories. Future guidelines will elaborate on further steps with respect to other information types such as primary data or multimedia and on more complex information objects that are composed of several resources.

Infrastructure

The DRIVER guidelines take a 'top-down' approach, in trying to ensure that repository data is exposed in a standard way. At the same time the DRIVER infrastructure takes a 'bottom-up' approach and provides the technology to harvest content from multiple repositories and manage its transformation into a common and uniform 'shared information space'.

The relevant aspects of this information space are that:

- the services needed to maintain it (stores, indexes, aggregators) are distributed on computers under the jurisdiction of several organisations, thereby reducing their individual effort and cost;
- such services can be added to the infrastructure at any time, in order to provide additional content and functionality;
- special *enabling services* automatically administrate the resources of the available services in order to maximise quality of service;
- content in the information space, i.e. DRIVER records, is that harvested from repositories, after cleaning and enrichment with provenance information;
- all records in the system can be reused via standard interfaces and protocols (SRW/CQL for querying, OAI-PMH for harvesting, etc.) by DRIVER services, such as a search

service and OAI-Publisher service, but they can also be used by external applications.

DRIVER's software infrastructure is operated as a service-oriented architecture based on open Web standards such as SOAP [14] and provides 'core' functionalities for administering distributed services and content. Data from distributed repositories are harvested and indexed based on extensive experience with content aggregation in BASE [5]. Interfaces for a wide range of functionality services such as 'search', 'recommend' or the management of digital collections and communities support the use and integration of software services. Deployment of these services in a robust hardware environment as well as authentication and authorisation procedures complement this scheme.

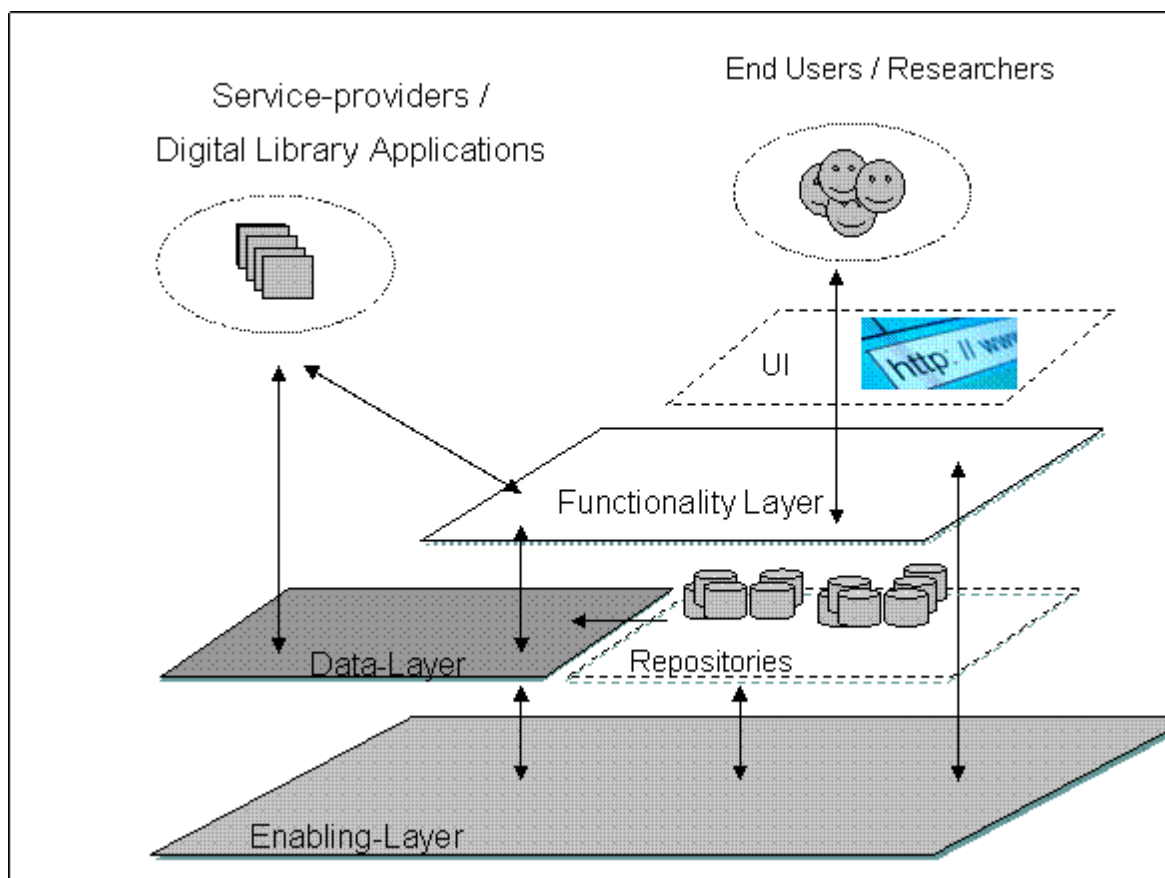


Figure 2: DRIVER infrastructure layers

The infrastructure will offer tools that enable repository managers to register their repositories within the DRIVER information space and obtain immediate feedback on the conformance of their OAI-PMH interfaces and their degree of alignment with the guidelines.

Repositories from any country will be able to register with the infrastructure and expect their content to be extracted, 'cleaned', and aggregated within an information space for integrated use. The information space describes all documents according to a rich and uniform metadata format, which extends typical digital resource information (e.g. author, title, year) with *provenance* information (e.g. country, language, institution) and *technical* information (e.g. metadata formats available for the resource, repository platform).

Reuse is a key feature of DRIVER. Traditional approaches have involved each organisation deploying a new system with high maintenance costs (possibly installing, customising services and constructing from scratch, i.e. harvesting/cleaning, a new information space). However in DRIVER organisations can reduce costs and effort by relying on the DRIVER infrastructure and

co-operating in order to populate and reuse the global information space as needed. Services developed by external organisations can also be shared.

DRIVER records can therefore be accessed through different standard interfaces and protocols, thereby opening the information space to the external world (SRW/CQL for querying, OAI-PMH for harvesting).

Benefits for DRIVER Users

DRIVER has wide-ranging benefits for both end-users and users of the DRIVER infrastructure.

Benefits for End-users

- *High quality of content.* Only high-quality resources that have been selected by participating institutions will form part of the DRIVER services.
- *High quality of search results.* Search results are relevant because DRIVER is using validated metadata supplied by its partners to build its search indexes.
- *High quality of service.* As a result of DRIVER's innovative distributed technical infrastructure, using services distributed over different nodes in Europe, the services are available at any time, anywhere.
- *Easy to use.* All that is needed is a browser and Internet access. Access to the full text is only one click away.
- *Full-text access.* All references link to the full text (usually a pdf file) that can be read online or printed for later use.
- *No fee, no restrictions.* All documents are open access, for everyone, worldwide.
- *Broad spectrum.* Covers all scientific domains and all of Europe. In the initial testbed version (2007) there will be 60 partners in Belgium, France, Germany, the Netherlands and the United Kingdom. This number will grow rapidly in 2008 from 60 to over 200 and partners will come from almost all European countries.

Benefits for Infrastructure Users

- DRIVER provides a shared information space that can be reused by service providers.
- The repository community can use the DRIVER infrastructure to provide national or regional repository search services which draw only from repositories in that country or region. Such customised searches can be made available from the national or regional Web sites.
- Individual repositories and their institutions can enjoy the increased visibility afforded by being harvested by DRIVER. Any document returned in a search will display the logo and reference of the host repository and institution.
- DRIVER also provides assistance and support for external service providers to enhance the search and other services they provide.

The testbed phase of DRIVER is a first step. The services, content, partners and quality will continue to be expanded in DRIVER-II.

DRIVER and the Research and Repository Community in Europe

As a European project, DRIVER provides a clear voice for the repository community in Europe. DRIVER is working to increase awareness of Open Access and repositories among those directly involved in research such as researchers and research funders as well as raising the awareness of the issues among the general public. DRIVER can also lobby policy makers on behalf of the community.

Through the experience of its partners in the development of national networks e.g. DAREnet [9], DINI [15], Archives Ouvertes /HAL [10] and SHERPA [16], DRIVER can provide advice to national initiatives and groups in the development of such networks as well as providing information and support through the DRIVER Support Web site [12] and Wiki [17].

The DRIVER Support Web site provides links to national groups, projects and services and acts as a source of news and information to the research and repository community in Europe. DRIVER is working with national groups e.g. Belgian repositories [18], to facilitate the development of national Web sites in association with the DRIVER Support Web site.

The DRIVER Wiki provides a space for contributions from the general repository community. Information on individual repositories, local projects, and events or news may be added to the Wiki to inform the wider community. Where national co-ordinating groups or projects do not exist, the Wiki will provide a platform for their initial development.

Existing networks can benefit from an increased visibility in the European repository community through linking with the DRIVER Support Web site and through participation in the Wiki.

DRIVER Services

Since the focus of DRIVER has been on developing infrastructure, it has not aimed to provide a pre-defined set of services. The infrastructure includes open, defined interfaces which allow any service providers working at a local, national or subject-based level, to build services on top. They will be able to reuse the data infrastructure (the Information Space) and the software infrastructure to build or enhance their systems. Services can therefore be developed according to the needs of users.

However in order to demonstrate exemplary functionality, and the potential of repository infrastructures, several services are being developed by the project. These include end user services (e.g. search) and services for repository managers (e.g. the validator tool), as outlined below. Further services will be developed in the future.

Search

The search service is particularly important in demonstrating the functionalities and capabilities of the DRIVER infrastructure. It is a generic search tool for querying the DRIVER information space, designed both for end-users (to gain an idea of DRIVER functionality) and also for service providers (to demonstrate the potential of the infrastructure). DRIVER search is intended only to provide access to full-text records.

The current version of the search interface offers an 'advanced search' which allows searching by selected field, as well as refinement by document type, language, date of publication etc. In addition it uses the concept of 'collections' and 'communities'. A sample range of collections is offered for searching at a broad subject level e.g. medical science, biology, history. Users will be able to subscribe to communities (containing a set of collections) and will be informed of changes to relevant collections. Customisation for local or subject needs could be implemented through the use of collections, i.e. search services could be built around a specific subset of the available records. Subsets could be chosen based on geographical location, or document type e.g. e-theses. Browsing and navigating will also be available at a later date.

A list of all the repositories currently included in the search is available via the interface, together with the number of documents held within each repository.

The public search final release will contain two search options: 'search clean' and 'search all'. By narrowing searches to 'search clean', users will query only those repositories in the information space whose content is strictly conformant to the DRIVER guidelines.

The DRIVER search test release is due to be made publicly available on the Web in the latter part of October 2007.

Validator

The Validator tool will validate repositories for conformance to 1) the DRIVER guidelines, 2) standard OAI-PMH functionality, and 3) OAI-PMH functionalities specific to DRIVER issues. It is currently in a test phase but will allow repository managers to check their repositories for conformance.

Once the validator service is fully integrated into the DRIVER infrastructure, it will automatically provide feedback to the harvester in real-time and will assist in informing the decision whether to harvest a specific record or not.

The Validator will be on public release (i.e. beyond testbed countries) towards the end of 2007.

Mentor Service

A mentor service is being developed to assist developers and managers of institutional repositories across Europe. The purpose of this service is to introduce those who are developing and managing institutional repositories to their peers on a one-to-one basis to enable the sharing of experience and the development of a supportive and active repository community.

This service is not an alternative to materials or advisory services already available online or to email discussion lists, but instead will provide advice on issues not typically available through formal sources of information. The service operates on the goodwill of the mentors and as such, is free of charge.

Requests for mentors can be submitted via the DRIVER Support Web site. The mentor team consider each request individually. Where mentoring is considered suitable, the mentor team will identify and contact several possible mentors from the database of mentors. This service will be expanded and developed further as DRIVER progresses.

The Future: DRIVER-II

As a result of a successful bid for Framework 7 funding, DRIVER-II will commence at the end of 2007. A further three partners will join the core DRIVER partnership and a further six countries have been identified as likely future partners.

It is important to remember that any repository or network of repositories can benefit from the DRIVER infrastructure and services. DRIVER has thus far concentrated its effort on the support of institutional repositories and the management of textual content in repositories. Through DRIVER-II, it is acknowledged that subject repositories and subject-specific services are key services needed by the research community. Therefore, in DRIVER-II, subject communities will be invited to become involved in the project and DRIVER technical developments will focus on the development and enhancement of services for specific communities. Moreover, subject-based communities introduce other forms of information management: scientific data and other non-textual content play an important role. Textual publications together with supplementary materials can form new aggregated types of content, sometimes referred to as 'enhanced publications'. In DRIVER-II the technical focus will therefore expand from the management of textual content in repositories also to include the handling of such complex objects.

References

1. OpenDOAR <http://www.opendoar.org/>
2. Repositories Support Project <http://www.rsp.ac.uk/>

3. IREL-Open <http://www.irel-open.ie/>
4. OAister <http://www.oaister.org/>
5. Bielefeld Academic Search Engine (BASE) http://base.ub.uni-bielefeld.de/index_english.html
6. DRIVER <http://www.driver-community.eu/>
7. van der Graaf, M., "DRIVER: Seven Items on a European Agenda for Digital Repositories", *Ariadne* Issue 52, July 2007. <http://www.ariadne.ac.uk/issue52/vandergraf/>
8. A preprint is available: "DRIVER review of technical standards" <http://www.driver-support.eu/en/about.html>
9. DAREnet <http://www.darenet.nl>
10. HAL <http://hal.archives-ouvertes.fr/>
11. DRIVER Guidelines <http://www.driver-support.eu/en/guidelines.html>
12. DRIVER Support Web site <http://www.driver-support.eu>
13. DRIVER helpdesk helpdesk@driver-support.eu
14. SOAP <http://www.w3.org/TR/soap12-part1/>
15. DINI <http://www.dini.de>
16. SHERPA <http://www.sherpa.ac.uk/>
17. DRIVER Wiki <http://www.driver-support.eu/pmwiki/>
18. DRIVER Belgium <http://www.driver-repository.be/>

Author Details

Martin Feijen
SURFfoundation

Email: Feijen@surf.nl
Web site: <http://www.surffoundation.nl/>

Wolfram Horstmann
Goettingen State and University Library

Email: whorstmann@sub.uni-goettingen.de
Web site: <http://www.sub.uni-goettingen.de/>

Paolo Manghi
CNR-ISTI

Email: paolo.manghi@isti.cnr.it
Web site: <http://www.isti.cnr.it/>

Mary Robinson
SHERPA European Development Officer
University of Nottingham

Email: Mary.Robinson@nottingham.ac.uk
Web site: <http://www.sherpa.ac.uk/>

Rosemary Russell
Research Officer
UKOLN
University of Bath

Email: r.russell@ukoln.ac.uk
Web site: <http://www.ukoln.ac.uk/>

[Return to top](#)

Article Title: "DRIVER: Building the Network for Accessing Digital Repositories across Europe"
Author: Martin Feijen, Wolfram Horstmann, Paolo Manghi, Mary Robinson and Rosemary Russell

Publication Date: 30-October-2007 Publication: Ariadne Issue 53

Originating URL: <http://www.ariadne.ac.uk/issue53/fejjen-et-al/>

[Copyright and citation information](#) File last modified: Saturday, 03-Nov-2007 19:38:05 UTC

[Main Contents](#)

[Section Menu](#)

[Email Ariadne](#)

[Search Ariadne](#)

Ariadne is published every three months by [UKOLN](#). UKOLN is funded by [MLA](#) the Museums, Libraries and Archives Council, the [Joint Information Systems Committee \(JISC\)](#) of the Higher Education Funding Councils, as well as by project funding from the JISC and the [European Union](#). UKOLN also receives support from the [University of Bath](#) where it is based. Material referred to on this page is [copyright Ariadne \(University of Bath\) and original authors](#).