

# DRIVER: Building a Sustainable Infrastructure of European Scientific Repositories

**Norbert Lossau**

DRIVER Scientific Coordinator; Director,  
Göttingen State and University Library,  
Platz der Göttinger Sieben 1, 37077, Göttingen, Germany,  
[Lossau@sub.uni-goettingen.de](mailto:Lossau@sub.uni-goettingen.de)

**Dale Peters**

DRIVER Scientific Technical Manager,  
Göttingen State and University Library,  
Platz der Göttinger Sieben 1, 37077, Göttingen, Germany,  
[peters@sub.uni-goettingen.de](mailto:peters@sub.uni-goettingen.de)

## **Abstract**

DRIVER has a clear vision: *All research institutions in Europe and worldwide make all their research publications openly accessible through institutional repositories.* The vision follows the Berlin Declaration, which called in October 2003 for ‘free and unrestricted access to sciences and human knowledge representation worldwide’. Initiated by the internationally renowned German research organisation the Max-Planck-Society, and signed by many international research organisations and institutes, the Berlin Declaration has set a political statement. In building a sustainable infrastructure for scientific repositories, DRIVER brings to this statement the reality of scholarly communication in the future.

**Key Words:** open access; institutional repositories; Berlin Declaration

## **Introduction**

DRIVER is a multinational initiative, co-financed by the European Commission. In response to the Berlin Declaration, the DRIVER initiative started initially as a test-bed project in June 2006, designed to explore the development of a distributed infrastructure to enable enhanced interoperability of data. Currently in its second phase, DRIVER has achieved a network of digital repositories containing research and other scholarly publications from twelve partner institutions or national clusters across ten European countries. Partners have come together in a consortium that comprises the National and Kapodistrian University of Athens; the University of Bielefeld; the Institute of Information Science and Technologies of the Italian National Research Council; the SURF Foundation; The University of Nottingham; the University of Bath; the Interdisciplinary Centre for Mathematical and Computational Computer Modelling at the University of Warsaw; the University Library of Ghent; the Göttingen State and University Library; the Technical Knowledge Centre of Denmark; the University of Minho; and the National and University Library of Slovenia.

With substantial funding amounting to M€2.7, the main deliverables are currently to provide a production-quality digital repository infrastructure and to establish a European digital repository confederation. In conjunction, these objectives will achieve a cross-country community of content providers and an infrastructure for the international networking of open access repositories.

A nucleus and early inspiration for DRIVER was the national DARE<sup>1</sup> programme in The Netherlands, which released a network and services for institutional repositories in 2004. One of the identified success factors of the DARE programme was community building. The affiliation in the DRIVER Confederation is based, similarly, on a willingness to comply with a set of principles and some agreements related to organisation, data and technology.

## **DRIVER Activity Areas and Outcomes**

From the outset, DRIVER has adopted a result-driven approach to structure the consortium activities. In considering the experiences learned from the DARE programme, DRIVER identified four main areas of activity:

1. organise digital repository content providers and build a cross-country community as a Confederation;
2. develop an open source software suite to network physically distributed repositories;
3. implement this software as operational instance of a digital repository infrastructure in Europe;
4. provide an online DRIVER Portal.

## 1. Community and Confederation

DRIVER views the building of an informed repository community as a prerequisite for a functioning infrastructure. A key lesson of the DARE programme in The Netherlands, and confirmed by other advanced national communities like [SHERPA](#)<sup>2</sup> in the UK, has been the relevance of active community building. The nature of such communities is varied; DRIVER sees communities in various contexts: national, subject-based, technical (e.g., repository platform providers like ePrints, D-Space or Fedora).

The rationale for building an informed repository community is based on multiple factors. For example, communication is more efficient between twenty national communities than between five hundred local institutional repositories. Technical and data quality requirements aimed at local repositories are more effectively discussed and disseminated in the constraints of the local environment. A significant problem for OAI service providers like OAIster, BASE or DARENet lies in the effort required to manually correct or 'clean' records submitted by individual repositories. Jackson et al.<sup>3</sup> stress that despite the importance of disseminating descriptive metadata capable of supporting interoperability, the manner in which institutions are implementing Dublin Core in practice is seldom discussed. Improving the quality of metadata (e.g., by standardising usage of fields like DC:author [Name], [Surname]) remains an enormous challenge to interoperability and harmonising the application profiles for local OAI-PMH implementations would significantly support automated harvesting processes.

Community building also involves both the strategic level (mainly library directors) and the operating level (repository manager and team) to support the network mission and the necessary implementations. This is especially important where additional effort is required to improve data

retrospectively. Communities are further characterised by a certain level of commitment from their members. Again, this commitment is a prerequisite for a reliable, large-scale infrastructure in the use, for example, of persistent identifiers and the guarantee for long-term access to online publications in repositories.

A central objective of the current DRIVER project is to establish a Confederation of digital repository communities. As organisational backbone of the DRIVER infrastructure it is envisaged beyond the duration of project funding, to ensure the sustainability of both the infrastructure and the organisational network of repository communities. The meeting of the first DRIVER Advisory Board, held in conjunction with the 37th Annual LIBER Conference in Istanbul in July 2008, unanimously recommended the Confederation as an essentially European organisation and endorsed its subsequent development internationally. Currently under evaluation is the business model of a Confederation and its legal status, either as an independent legal entity or constituted within an appropriate existing organisation such as SPARC Europe or LIBER.

## **2. DNET Software**

In June 2008 the DRIVER Consortium released version 1.0 of the DRIVER Network Evolution Toolkit (DNET)<sup>4</sup> under an Open Source Apache License. The license type was chosen after discussions that included commercial software companies, in order to support widest possible collaboration.

DNET Version 1.0 offers 23 services in total, and supports three groups of users:

- The repository network administrator is supported to harvest, clean, enrich and aggregate data from local repositories (e.g., by the *Repository Network Manager, Resource Monitoring*)
- The local repository manager can check compatibility of local data with the DRIVER index format (*DRIVER Validator*)
- End users and, in particular, researchers can use the virtual knowledge base with some fundamental features (search, browse, profiling).

From a software architecture point of view, DNET recognises three main layers:

- service management
- data management
- end-user services.

Service management enables the interaction of all service layers, the dynamic addition and grouping of single services and service monitoring. The services can be used in combination or stand-alone. Increasingly, new services are being added by DRIVER network partners or third-party service providers. The D-NET v. 1.0 software release is also made available to any organisation willing to run independent installations of the DRIVER software and able to assume responsibility for the maintenance of such infrastructure. With this release, DRIVER demonstrates its support for the development of further open infrastructures and its willingness to promote other service providers and relative repository communities.

### **3. Infrastructure Implementation**

A possible deployment scenario of the DNET software is envisaged in the implementation by a repository community of a national repository network. Such an example is already evident in [RECOLECTA](#)<sup>5</sup>, the Spanish national service for the search, access and retrieval of full text open access scholarly publications of Spanish academic and research institutions. A joint collaboration project between REBIUN (Network of Academic and Research Libraries of Spain) and FECYT (Spanish Foundation for Science and Technology), the aims of RECOLECTA are twofold:

- coordinate and stimulate the development of a network of interoperable digital scientific repositories as a means of building a national knowledge infrastructure based on the open access principles and open and international standards and protocols;
- to develop and allow third parties to develop, a set of services on top of the aggregated content, targeting different needs: search, re-use, collaboratives, citation, subject portals, national web portals, etc.

The DRIVER Infrastructure offers technology that can be customised to the specific needs of such organisations willing to build a uniform information space from an arbitrary number of heterogeneous OAI-PMH data sources. Communities may thus define applications operating over the infrastructure that are specific to their needs and that can be implemented as a customisation and extension of the DRIVER Infrastructure.

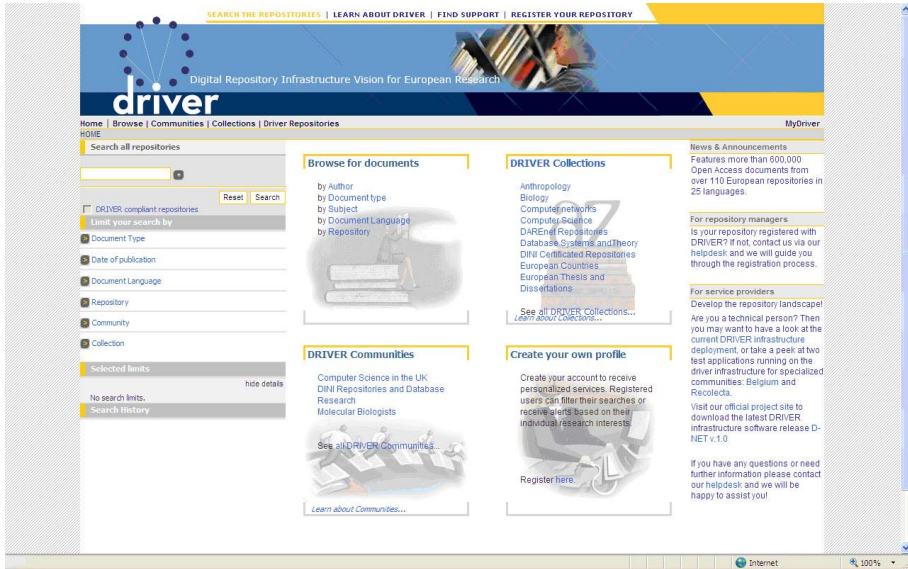
D-NET has been developed as a technically advanced, open source toolkit for re-use by repository networks. Deployment thereof guarantees interoperability, saves costs and time, but is not mandatory. More critical to the functionality of a common infrastructure is the use of DRIVER services and data models to support interoperability. The DRIVER Guidelines<sup>6</sup> aim to achieve formal syntax and declared semantics in improving metadata usage and practice.

Based on common standards and accepted practice, the DRIVER Infrastructure is readily extendable to become a global repository infrastructure. Negotiations are underway in collaboration with eIFL<sup>7</sup>, with representatives of national repository communities in Eastern Europe, the Baltic countries, China, India, Africa and Latin America. International relations are also maintained between the DRIVER Infrastructure and relevant repository initiatives in the United States, such as the Object Reuse and Exchange (ORE)<sup>8</sup> project and SPARC<sup>9</sup>.

#### **4. The DRIVER Portal**

In the interim it has become clear that networking repositories means more than harvesting metadata through the OAI protocol. Existing service providers report many problems arising from heterogeneous implementation of this protocol. Interoperability at the price of manual correction of data is to be avoided, as this is not scalable. At the same time the workload for local repository providers needs to be contained within acceptable limits.

Fig. 1: DRIVER Portal.



The online portal service (Figure 1) offers basic usability in search, browse, personal profiling and support for research communities. The service functionality relies on repository community support by DRIVER, allowing repository managers to share and exchange knowledge and experience around initial setup of institutional repositories (IR's) and the organisational aspects of IR management. In addition, the portal service relies on raising professional awareness in areas of repository data management:

- high quality of data is a priority, as it is the basis for automated services like profiling, statistics, metrics, and also for links to research evaluation systems (CRIS);
- delivery of open access full text;
- interoperability is achieved through the common set of metadata policies, as described in the DRIVER guidelines, and offers maximum exposure of research data.

The DRIVER Guidelines<sup>10</sup>, now approaching a second version, are maintained by a volunteer metadata expert group comprising DRIVER partners

and other European experts to improve the quality of available metadata and to assist managers of new repositories in defining their local data management policies. The harmonisation of OAI implementations and application profiles serves DRIVER too, in avoiding harvesting errors, ultimately ensuring a degree of uniformity across all repositories in the DRIVER platform.

## **Joining the DRIVER Community**

The inclusion of individual repositories into the DRIVER platform is possible, but national communities are preferred as natural partners in the European repository network. One institution or a group of institutions usually takes responsibility for building a national repository community, such as DARENet<sup>11</sup> in the Netherlands, SHERPA<sup>12</sup> in the UK, OA.Net<sup>13</sup> in Germany, RECOLECTA<sup>14</sup> in Spain, and HAL<sup>15</sup> in France. National communities are represented by country correspondents, designated to liaise with the DRIVER community, reporting on the national status of repository development, organising local events, and translating repository guidelines into local languages. This effective and dynamic interaction is maintained in a wiki format, available online<sup>16</sup>, to identify local contact details to encourage membership and support of an existing national community. National nodes may in time build up national data aggregators, cleaning data and providing additional services to the local community as required.

Where no such national nodes exist, dedicated repository managers are encouraged to indicate on the DRIVER wiki their interest in taking up this role, and thus joining the DRIVER community. The benefits are many:

- By joining the DRIVER community individual repositories contribute to the successful implementation of open access principles in a powerful, international community.
- Certain research areas, such as health, climatology and geomorphology can no longer be adequately served regionally. Local researchers require an international infrastructure in which their research output becomes visible as part of a common international knowledge base.
- Local funding and research organisations can build their own interface on the DRIVER data index to demonstrate national efforts.
- National communities share in developments contributed to the DRIVER Infrastructure by service providers.



The DRIVER community is sustainable in the form of a Confederation, in which the members assume responsibility for the DRIVER objectives, contribute data, offer services, share expertise and suggest strategic direction to the DRIVER community. These objectives form strong natural incentives, beyond project funding, to build an international repository community that is firmly embedded in national communities and supported by significant alliances with LIBER, SPARC Europe and eIFL. The DRIVER *Confederation* is envisaged as the permanent organisational backbone of the DRIVER Community.

## Conclusions

DRIVER's mission is to expand its content base with high-quality open access research output, as well as provide support for repository managers and state-of-the-art services for the end user. By building a robust network of content providers, enhanced with the complex set of services DRIVER offers, the DRIVER Infrastructure also enables service providers to develop new applications on top of the basic services. It acts as a showcase for repository development, as a networking tool for the DRIVER community and demonstrates a range of end-user services. The community is defined in the DRIVER Confederation, comprising data providers, service providers, the major repository software platforms, representatives of emerging standards, protocols and data models, and NGO's serving the repository community.

However, much remains to be done. The lobby for open access mandates from all funding organisations is high on the list of priorities for the repository community and DRIVER has signed a Memorandum of Understanding with SPARC Europe to join forces. The coordinated approach presented in the DRIVER Confederation will undoubtedly hasten and facilitate the implementation of open access. Related effort is required in building open access consortia to revise contracts with journal publishers, to convert current subscription payments to parallel publishing models, i.e., 'subscription + institution-wide "green road" IR deposit permission'. The pilot Springer Open Choice agreements are indicative of potential future developments in this area. Another example is seen in the role of DRIVER in coordinating repository interaction with publishers in the PEER<sup>17</sup> project. A pioneering collaboration between publishers, repositories and the research community, PEER (Publishing and the Ecology of European Research) will investigate the effects

of large-scale, systematic self-archiving on reader access, author visibility and journal viability, as well as on the broader ecology of European research.

Such coordinated action is required to support both open access journal publishing and open access book publishing, as seen in the European OAPEN project.<sup>18</sup> New bibliometrics need to be devised in support of the principles of peer review, including citation analysis and impact measuring services. Digital curation and long-term preservation services are urgently required by the repository community.

DRIVER provides a voice to national repository communities and makes them visible in the European and international context. The development of local institutional repositories is accelerated, with strong economies of scale achieved in DRIVER's shared services, which can be built on top of local repositories. Such a network replicates and innovates traditional modes of scholarly communication to allow comparative searching within subject domains — and can help to fill repositories, as researchers, research institutes and universities want to have maximum visibility—best realised in a network of content repositories.

Five years ago, the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* acknowledged the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access. DRIVER 'Digital Repository Infrastructure Vision for European Research', has come to represent this vision of worldwide networks of content repositories, offering a robust infrastructure supporting scholarly communications of the future.

## Notes

---

<sup>1</sup> Waaijers, L. (2005), 'Towards the self-filling repository [The DARE Programme. Successes and lessons learned: from libraries to libratories]', Presentation delivered at the CERN workshop on Innovations in Scholarly Communication (OAI4), Geneva (Switzerland), <http://eprints.rclis.org/archive/00007248/>

<sup>2</sup> SHERPA/RoMEO, Publisher copyright policies & self-archiving. <http://www.sherpa.ac.uk/projects/sherparomeo.html>

- <sup>3</sup> Jackson, A.S., M.-J. Han, K. Groetsch, M. Mustafoff and T. Cole. (2008), Dublin Core Metadata harvested through OAI-PMH. *Journal of Library Metadata*. 8(1) p.5–21
- <sup>4</sup> D-Net: release of the DRIVER Software, <http://www.driver-repository.eu/D-Net.html>
- <sup>5</sup> RECOLECTA, <http://www.recolecta.net/buscador/>
- <sup>6</sup> DRIVER Guidelines, [http://www.driver-support.eu/documents/DRIVER\\_guidelines\\_1%200.pdf](http://www.driver-support.eu/documents/DRIVER_guidelines_1%200.pdf)
- <sup>7</sup> eIFL.net is an independent foundation that strives to lead, negotiate, support and advocate for the wide availability of electronic resources by library users in transitional and developing countries, <http://www.eifl.net/cps/sections/home>
- <sup>8</sup> Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources, <http://www.openarchives.org/ore/>
- <sup>9</sup> The Scholarly Publishing and Academic Resources Coalition (SPARC) is an international alliance of academic and research libraries working to correct imbalances in the scholarly publishing system, <http://www.arl.org/sparc/index.shtml>
- <sup>10</sup> DRIVER Guidelines, [http://www.driver-support.eu/documents/DRIVER\\_guidelines\\_1%200.pdf](http://www.driver-support.eu/documents/DRIVER_guidelines_1%200.pdf)
- <sup>11</sup> DARENet is the network of Digital Academic Repositories in the Netherlands. It is a result of the national DARE programme that aimed to coordinate and stimulate the development of repositories containing scientific output in the Netherlands. Since April 2008 DAREnet has been integrated in the scientific portal NARCIS, <http://www.narcis.info/index/tab/darenet/>
- <sup>12</sup> The SHERPA partnership has within its membership a range of examples of repository environments and institutional structures within research-led institutions, offering the ideal environment for exploring and testing ideas for repository development, which can be evaluated and disseminated to the wider community, <http://www.sherpa.ac.uk/index.html>
- <sup>13</sup> Supported by the DFG, the project OA-Network is a joint collaboration of the Humboldt Universität zu Berlin and the Universities of Göttingen and Osnabrück. It aims to virtually integrate all document and publication services with a DINI certificate and to increase the number of DINI certified repositories, <http://www.dini.de/projekte/oa-netzwerk/>
- <sup>14</sup> RECOLECTA, <http://www.recolecta.net/buscador/>
- <sup>15</sup> The HAL archive is a national aggregator which offers a common platform to multiple archives comprising French universities, major higher education schools and major research institutes which have signed a common protocol to make the

HAL archive system their common platform to host the national research output, <http://www.driver-support.eu/national/france.html>

<sup>16</sup> DRIVER wiki information on national communities: <http://www.driver-support.eu/national/index.html>

<sup>17</sup> PEER Project to Examine Large-Scale 'Green' Open Access Impacts, <http://digital-scholarship.org/digitalkoans/2008/10/13/peer-project-to-examine-large-scale-green-open-access-impacts/>

<sup>18</sup> Open Access Publishing in European Networks (OAPEN) is a project in open access publishing for humanities and social sciences monographs. The open access movement has developed rapidly in the sciences and in journal publishing. The consortium of university-based academic publishers who make up OAPEN believe that the time is ripe to fully explore the possibilities of open access for the humanities and social sciences, <http://www.oapen.org/>