

CRIS2018 – Extended Abstract Submission

DuraSpace for FAIRness and Data Protection: DSpace and Fedora

Authors

Tim Donohue, tdonohue@duraspace, USA, DuraSpace

Michele Mennielli, mmennielli@duraspace.org, Italy, DuraSpace

David Wilcox, dwilcox@duraspace.org, Canada, DuraSpace

Keywords

FAIR principles, EOSC, GDPR, open source, open access, DSpace, Fedora, repository

Topics

FAIR communication, sharing and profiling of research

Current trends in Open Access and Open Data Management

Dataset description and metadata

Other topics

Extended Abstract

With the European Open Science Cloud initiative the concept of FAIR data has become an official recommendation for the European Member States, but the principles are now widely recognized and accepted as guiding principles at a global level. Repositories are a crucial component in the research ecosystem to make sure the FAIR data principles are respected and implemented.

With this paper we want to share with the euroCRIS community that two of the most widely adopted open source repositories, DSpace and Fedora, allow almost 3000 institutions around the world (the users) to be able to comply with FAIR data principles. We will analyze the 4 different dimensions of FAIRness (Findability, Accessibility, Interoperability, Reusability) to highlight the features that already comply with the principles and to identify any gaps.

Findable data must be described by rich metadata, assigned a unique persistent identifier, and be indexed for search.

Fedora provides support for identifiers, an agnostic approach to metadata standards, and patterns for indexing resources to different external indexes.

DSpace provides support for identifiers, extendable metadata schemas, and patterns for indexing resources to internal and external indexes (including SEO).

Accessible data must be retrievable by their identifier, using a standardized, open protocol with support for security, and metadata must remain accessible even if the data are no longer available.

Fedora implements a standardized, well-documented REST-API and includes support for secure access. Metadata can remain available even if the data are removed or otherwise made inaccessible.

DSpace implements a well-documented REST-API and support for fine-grained, configurable access controls. Metadata can remain available even if the data are removed or otherwise made inaccessible.

Interoperable data must use an accessible, shared language, along with vocabularies and references to other data and metadata.

Fedora supports multiple languages and can be integrated with external vocabularies for metadata. Linking to other data and metadata is supported and encouraged.

DSpace supports multiple languages and importing / configuring controlled vocabularies or authority control. Linking to other data and metadata is supported via URIs or identifiers.

Reusable data must conform to community standards and have a clear usage license along with provenance information.

Fedora provides rich metadata support with capabilities to link to usage licenses and maintain complete provenance records.

DSpace provides extendable metadata schemas, usage licenses and captures basic provenance within metadata.

Another important European Commission initiative that will have an important impact on the research ecosystem, and on research data management in particular, it is the EU General Data Protection Regulation (GDPR), which will be enforced on May 25th, 2018. GDPR compliance has a lot to do with distinguishing between levels of sensitive data and applying different rules for access, up to and including non-reversible anonymization of highly sensitive personal data. Some of these requirements could be supported by the repository, while others would need to be handled at a higher level in the software stack, with CRIS systems playing a crucial role. In our presentation we will focus on the repository's contributions, and we will engage with the audience in a discussion on how to deal with the new requirements from a broader (ecosystem) perspective.