



Supporting FAIR Data Principles with Fedora

David Wilcox

DuraSpace, Beaverton OR, Canada

dwilcox@duraspace.org, orcid.org/0000-0001-5411-9208

Abstract

Making data findable, accessible, interoperable, and re-usable is an important but challenging goal. From an infrastructure perspective, repository technologies play a key role in supporting FAIR data principles. Fedora is a flexible, extensible, open source repository platform for managing, preserving, and providing access to digital content. Fedora is used in a wide variety of institutions including libraries, museums, archives, and government organizations. Fedora provides native linked data capabilities and a modular architecture based on well-documented APIs and ease of integration with existing applications. As both a project and a community, Fedora has been increasingly focused on research data management, making it well-suited to supporting FAIR data principles as a repository platform.

Fedora provides strong support for persistent identifiers, both by minting HTTP URIs for each resource and by allowing any number of additional identifiers to be associated with resources as RDF properties. Fedora also supports rich metadata in any schema that can be indexed and disseminated using a variety of protocols and services. As a linked data server, Fedora allows resources to be semantically linked both within the repository and on the broader web. Along with these and other features supporting research data management, the Fedora community has been actively participating in related initiatives, most notably the Research Data Alliance. Fedora representatives participate in a number of interest and working groups focused on requirements and interoperability for research data repository platforms. This participation allows the Fedora project to both influence and be influenced by an international group of Research Data Alliance stakeholders.

This paper will describe how Fedora supports FAIR data principles, both in terms of relevant features and community participation in related initiatives.

Key Words: fedora; repository; fair data; open source; linked data

1. Introduction

Fedora, the Flexible, Extensible, Digital Object Repository Architecture, is open source repository software that stores, preserves, and provides access to digital content.¹ Fedora is built around the notion of flexibility; content can be modeled in a variety of different ways to support both simple and complex use cases. This flexibility is primarily based on linked data. Fedora uses Resource Description Framework (RDF) triples to create semantic links between resources, thereby allowing for data models unrestricted by traditional file and folder hierarchies. Along with this flexibility, Fedora supports millions of resources, both large and small, with configurable storage capabilities. But perhaps most importantly, Fedora is interoperable; it has been designed around a robust REST-API and an event-based messaging service that establish well-documented patterns for integrating Fedora with other applications and services to build a larger system (Technical Specifications, 2018).

Institutions adopt Fedora for a variety of reasons; as a flexible system Fedora can satisfy a number of use cases and requirements. However, most institutions turn to Fedora for its flexibility — while local use cases may start out relatively simple, they will inevitably grow more complex over time. Fedora supports this natural growth by accommodating more complex needs as they arrive. Just as importantly, Fedora is designed with durability in mind. Digital preservation is a complex topic, and Fedora does not seek to be an all-in-one digital preservation system, but it provides a number of features and integration patterns that support an overall digital preservation strategy (Duraspace, 2018). Fedora has also been successful — it is not enough for a project to be open source, it must also be sustainable and well-adopted. Fedora has been deployed in over 400 institutions around the world, which demonstrates its stability and success (Wilcox, 2018). Fedora also focuses on standards; as an API-driven application, Fedora implements a set of modern, well-adopted web standards to provide its services. These standards help ensure that data don't become trapped in a Fedora repository with application-specific

customizations, while also making it easier to integrate with other applications and services to share data. Finally, Fedora is backed by a thriving, global community that provides distributed support and control.

The FAIR Data principles (Force11, n.d.), introduced by Force11 in 2014 and first published in 2016, provide a set of guidelines for making data Findable, Accessible, Interoperable, and Reusable. Each principle has an associated list of criteria which can be aligned with the relevant Fedora features in order to demonstrate how Fedora can effectively support the FAIR Data principles.

2. FAIR Data Principles

This section will look at each FAIR Data principle and its criteria in turn and describe how Fedora meets these criteria to support each principle.

2.1. Findable

Findability is defined by the following criteria:

1. (meta)data are assigned a globally unique and eternally persistent identifier.
2. data are described with rich metadata.
3. (meta)data are registered or indexed in a searchable resource.
4. metadata specify the data identifier.

As a resource-centric repository, Fedora assigns each resource, whether it be a metadata record or a file, a Uniform Resource Identifier (URI) that serves as a persistent identifier for that resource. Additional persistent identifiers may also be used; for example, a public-facing DOI could be registered and mapped to a Fedora resource, and that DOI could be stored as a metadata property on the resource in Fedora. This also means that metadata resources can be linked to the data they describe by storing and linking to the data identifier. Strong support for metadata is another key Fedora feature; any type of metadata based on any schema (including custom fields) may be used. This flexibility allows Fedora to be used across research domains. Fedora also provides strong support for indexing — metadata and data (in the case of

text-based resources) can be indexed in any number of external indices; common use cases include Solr, Elasticsearch, and triples stores.

2.2. Accessible

Accessibility is defined by the following criteria:

1. (meta)data are retrievable by their identifier using a standard protocol.
 - a. the protocol is open, free, and universally implementable.
 - b. the protocol allows for authentication and authorization.
2. metadata are accessible, even when the data are no longer available

Fedora provides a well-documented REST-API (Woods, 2018). This API serves as an open protocol for access resources in the repository using their identifier. Any REST-based client can easily request repository resources, and standard authentication can be applied to such requests to prevent access by users or machines without the proper credentials. Once a user has been authenticated, Fedora uses the World Wide Web Consortium (W3C) Web Access Control standard to enforce authorization (WebAccessControl, 2018). This will ensure that authenticated users only receive the appropriate level of access based on their credentials. This same authorization scheme can be used to control access to data and metadata separately — a repository administrator could choose to lock down access to data while still providing access to the related metadata.

2.3. Interoperable

Interoperability is defined by the following criteria:

1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
2. (meta)data use vocabularies that follow FAIR principles.
3. (meta)data include qualified references to other (meta)data.

In terms of metadata, Fedora has strong multilingual support. Any language may be used, and metadata can be stored in whatever format is most relevant (e.g. RDF, XML). Additionally Fedora's native linked data support allows

vocabularies to be used to enhance knowledge representation by referencing well-known terms rather than using custom values. For example, a subject field could reference a term in the Library of Conference Subject Headings vocabulary (Library of Congress, 2011) and store the URI for that term. That way, a user can follow the URI to gather more information on the subject. This concept of following links to other resources is a fundamental principle of linked data and also supports the FAIR data notion of interoperability.

2.4. Reusable

Reusability is defined by the following criteria:

1. meta(data) have a plurality of accurate and relevant attributes.
 - a. (meta)data are released with a clear and accessible usage license.
 - b. (meta)data are associated with their provenance.
 - c. (meta)data meet domain-relevant community standards.

Fedora's rich metadata support allows for a wide variety of attributes, and RDF can be used to link resources to their licenses. These licenses could be stored as resources in the repository, or they could be external licenses such as those provided by Creative Commons.² Fedora also has an optional Audit module that can be enabled to track the provenance of resources in the repository (including metadata). Once enabled, the module will create PREMIS metadata (PREMIS, 2018) associated with events in the repository (e.g. when something is added, changed, or deleted) and this PREMIS metadata can be stored and queried for provenance reporting. Finally, another optional application can be used with Fedora to provide additional functionality on top of the standard set of Fedora services. This application, the API Extension Framework, can be used to build and share modules to do things such as metadata validation to ensure compliance with relevant community standards (API Extension, 2018).

3. Related Community Initiatives

Fedora is more than software; it is also a community. The Fedora community participates in a variety of international efforts aimed at making progress on key project priorities. One such effort is the Next Generation Repositories report that was published by the Confederation of Open Access Repositories

in 2017 (Confederation of Open Access Repositories, 2017). This report is based on the efforts of an international working group which included participation from the Fedora Product Manager. The report recommends a number of behaviours and supporting technologies that the next generation of repositories should implement, and these recommendations are very much in line with the FAIR data principles.

The Fedora community is also involved in the Research Data Alliance,³ an international group focused on enabling research data sharing across borders and around the world. This group is obviously well-aligned with the FAIR data principles, and there are a number of interest and working groups within the RDA that are making progress toward these shared goals. One such group, the Research Data Repository Interoperability working group, recently published recommendations on a data packaging standard for increased interoperability between repository platforms on a machine-to-machine level (RDA Research Data, 2018). This group was co-chaired by the Fedora Product Manager, and the recommendations are consistent with both the FAIR data principles and ongoing work in the Fedora community to support standardized data import and export.

4. Supporting and Sustaining Fedora

Fedora is stewarded by DuraSpace,⁴ a not-for-profit organization funded primarily through membership. Institutions join DuraSpace and direct annual funding to support the project(s) of their choice. In 2017, 74 DuraSpace member institutions supported Fedora with \$562,300 in funding (Fedora Community, 2018). This funding pays for 2 full time equivalent (FTE) staff members, as well as travel for conferences, workshops, and user groups, marketing and communication, and other priorities as determined by the project governance group. Fedora is designed, built, and maintained by the community; DuraSpace provides support but the majority of the development is done by members of the community.

5. Conclusion

The FAIR data principles represent an important community goal of making data Findable, Accessible, Interoperable, and Reusable. However, in order

to put these principles into practice they must be broken down into criteria that can be supported by infrastructure. For each principle, Fedora has a set of features that satisfy these criteria and support the overall implementation of the FAIR Data principles. This can be demonstrated not only at the level of the software, but also in the Fedora community's participation in related community efforts that further the goals of the FAIR data principles at a more strategic level. As community-supported, open source software, Fedora will continue to evolve to meet the needs of the research data management community as data is made more Findable, Accessible, Interoperable, and Reusable.

References

- API Extension. (2018). *Fedora API extension framework*. Retrieved from GitHub. <https://github.com/fcrepo4-labs/fcrepo-api-x>.
- Confederation of Open Access Repositories. (2017). *Next generation repositories*. Retrieved from <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>.
- Duraspace. (2018). *Fedora and digital preservation*. Retrieved from <https://duraspace.org/fedora/resources/publications/fedora-digital-preservation/>.
- Fedora Users. (2018). Retrieved from <https://duraspace.org/fedora/community/fedora-users/>.
- Force11. (n.d.). *The FAIR data principles*. Retrieved from <https://www.force11.org/group/fairgroup/fairprinciples>.
- Library of Congress. (2011). *Library of Congress subject headings*. Retrieved from <http://id.loc.gov/authorities/subjects.html>.
- PREMIS. (2018). *Preservation metadata maintenance activity*. Retrieved from Library of Congress website: <https://www.loc.gov/standards/premis/>.
- RDA Research Data Repository Interoperability Working Group. (2018). *Research Data Repository Interoperability WG final recommendations*. <http://dx.doi.org/10.15497/RDA00025>.
- Technical Specifications. (2018). Retrieved from Fedora website <https://duraspace.org/fedora/resources/technical-specifications/>.
- WebAccessControl. (2018). Retrieved from W3C wiki. <https://www.w3.org/wiki/WebAccessControl>.

Wilcox, D. (2018). *2017 Fedora annual report*. Retrieved from <https://wiki.duraspace.org/display/FF/Annual+Reports>.

Woods, A. (2018). *RESTful HTTP API*. Retrieved from Fedora 4.7.5 documentation: <https://wiki.duraspace.org/display/FEDORA475/RESTful+HTTP+API>.

Notes

¹ <https://duraspace.org/fedora/>.

² <https://creativecommons.org>.

³ <https://www.rd-alliance.org>.

⁴ <https://duraspace.org>.