DSpace Product Visioning Working Group Report

2021-12-06

Tal Ayalon
Kimberly Chapman
Scott Hanrath
Barbara Hirschmann
Agustina Martinez-Garcia
Beate Rajski

Introduction

The DSpace Product Visioning Working Group was charged by the DSpace Steering Group in January 2021 to help develop a vision for DSpace beyond the DSpace 7 release. The Working Group was asked to "engage stakeholder groups … to outline a strategic vision and high-level priorities for the DSpace platform." The Work Group focused on providing high-level recommendations for ways in which DSpace, as both a technical and community-led project, could evolve to enhance its status as a preferred, trusted, and widely-used repository solution.

In the course of its work the Working Group gathered information from a variety of sources. Activities included:

- participating in and reviewing January 2021 DSpace Leadership Group breakout sessions on opportunities for DSpace integration and interoperability
- reviewing previous DSpace strategic planning exercises
- conducting an environmental scan of repository platforms and related initiatives
- interviewing DSpace Tier 1 Service Providers
- surveying Tier 2 and 3 Service Providers
- surveying contacts for national DSpace User Groups

Sections 1 through 3 of this report contain the Working Group's primary recommendations for DSpace's governance groups to consider:

1. Modularize the DSpace architecture
2. Empower users
3. Explicitly conform to selected standards and international initiatives

These recommendations do not generally include specific feature specifications (though examples are offered), and certainly omit many important features that DSpace should offer to be successful. The recommendations are intended rather to gesture toward a framework the community might agree on for defining and prioritizing future work. Because the Working Group did note requests for specific features in its information gathering, section 4 of this report

includes an overview of primary DSpace use cases and features that could be considered to support them.

1. Modularize the DSpace Architecture

DSpace is used by a wide variety of organizations to meet a range of needs. These include, but are not limited to, "traditional" institutional repositories, open access and data repositories, digital collections management, publishing, and research information systems. DSpace has strengths both as a stable, out-of-the-box solution for these needs and as a platform that can be adapted to new needs. These two strengths can play out as tensions, risking scenarios where meeting new use cases requires diverging from the core platform, which in turn makes it difficult for organizations to maintain and upgrade as DSpace releases are launched. This is both a technical challenge (e.g., how best to incorporate new features into the code base) and community challenge (e.g., community engagement, community skills-building and contributions, how best to prioritize features for new releases).

DSpace could benefit from a modular architecture that could assume a smaller, well-defined core repository application with intentional hooks and interfaces for self-contained plugins to override existing functionality or provide extended functionality. Other open source applications may provide inspiration, e.g., Drupal, WordPress, ArchivesSpace. This architecture could support both plugins officially supported and released as integral parts of DSpace releases and plugins supported by third parties. It may also help focus on opportunities for integration with other platforms. While the shift to an extensible, modular architecture can help supporting a wider range of use cases for the DSpace platform, there are instances where other platforms exist and are more appropriate to fulfil some of those use cases. Examples include dedicated solutions for digital preservation, or stand-alone CRIS systems. In these instances, the ability to integrate the DSpace platform with such systems can be more appropriate. Future DSpace developments should be designed in ways that allow for an easy integration with new and future tools or services in those areas most relevant to the community. In this respect, the development direction that DSpace is taking with its new API, together with incorporating features that support next-generation repository behaviours, can be a big step forward in supporting a wider range of integrations.

Such a structure could open up new possibilities for feature prioritization and release management processes. It may also offer new incentives and opportunities for community member engagement. This focus would push the community to further refine its governance structures and processes to provide the best environment for evolving the core DSpace platform. Clear participatory opportunities and a shared understanding

of roles for voices from a range of expertise – technical, functional, and beyond—would be required.

2. Empower Users

DSpace is a powerful and flexible application. In order to be of greatest value to its wide user base, DSpace should provide a user experience that directly empowers repository administrators and other users to make full use of its customizable and configurable features. Replies to our distributed surveys mentioned several areas in which the various populations of DSpace users would like to be more empowered.

Plugins / Customizations

As a highly customizable system designed to cater to a wide range of repositories within a wide range of institutions, it is common to find institutional repositories and other systems that are built on DSpace, but incorporate specific customizations required by the particular institutions. Survey responses included several different aspects of DSpace customization, among them discovery, development, distribution and installation.

Institutions would like to empower users (institutional users / administrators in this case) to easily locate existing customizations that might fit a need not addressed by the core DSpace application, or even browse those customizations for ideas on how to improve their repository. Absent an existing customization, they would like to be empowered to create their own customizations and use a simpler, more streamlined process for incorporating local customizations into a DSpace repository. Attention to the distribution and installation of customizations could benefit both software vendors and their clients.

Enable greater control through the admin UI

Institutional users who are not IT professionals but act as repository managers, curators, administrators, etc. should to the greatest extent possible be empowered to fully configure repository functionality using admin UI tools, without the need of vendor or IT department intervention or server access.

DSpace 7 has already changed the UI / UX design landscape with the introduction of an Angular 2 unified DSpace UI, customizable via Bootstrap, but institutional users would still like to make some UI alterations via the admin UI, for example making menu changes; adding, removing or editing facets; static page editing; theming; query result weighing, and more look & feel parameters.

More automated features for content enrichment

When using or maintaining a DSpace system, users require and expect more than a "file store" they can browse or search. As information professionals, we know that accurate and varied metadata enriches a repository, not only due to the improved discoverability it provides, but also by expanding the scope of information that is provided to users.

Institution-created metadata is only one aspect of content enrichment available to today's researcher. The DSpace community has indicated that integrating DSpace with external services which offer such content enrichment, which could include translated interfaces, media engagement-gauging via academic citation data integration, inclusion of errata and retraction information, and more. This is an area in which DSpace could have a more robust support system, eliminating some of the need to collect such peripheral data manually outside of the repository system.

Note

For more background about user empowerment and what it means for DSpace, see a longer version of this chapter which will be made available separately.

3. Explicitly conform to selected standards and international initiatives

Repository software plays an important role in applying and enforcing community standards and norms. Potential users look for software tools that comply with certain standards out-of-the-box. It is therefore essential that DSpace as a community strategically interacts with existing and new initiatives that define principles, standards, and other best practices for digital repositories and scholarly communication infrastructures. This includes monitoring of existing and new standardization initiatives, defining a process for deciding which standards DSpace should apply and implement, as well publicly documenting the extent to which DSpace complies with a certain standard.

In the long-term, DSpace should aim to become "certification-ready," with "validation-ready" software solution for endorsed initiatives (e.g. certificates such as the CoreTrustSeal, metadata standards such as openAIRE and DataCite, or guiding principles such as the FAIR Data Principles and the TRUST Principles for digital repositories).

We recommend that the DSpace community collaborate with broader repository visioning activities, for example COAR's Next Generation Repositories initiative. This will help position DSpace not only as a standards-compliant software solution but also

as a stakeholder active in shaping the future role of repositories in the scholarly infrastructure ecosystem.

Example Standards and Initiatives

Principles

- The TRUST Principles for digital repositories: https://doi.org/10.1038/s41597-020-0486-7
- FAIR Data Principles: https://www.go-fair.org/fair-principles/

Metadata Standards

- openAIRE Guidelines: https://guidelines.openaire.eu/en/latest/
- DataCite Metadata Schema: https://schema.datacite.org/
- Common European Research Information Framework (CERIF): https://cordis.europa.eu/article/id/8260-cerif-common-european-research-information-format

Best Practices

- COAR Community Framework for Good Practices in Repositories: https://www.coar-repositories.org/coar-community-framework-for-good-practices-in-repositories/
- COAR Repository Toolkit: https://coartraining.gitbook.io/coar-repository-toolkit/

Certificates

- CoreTrustSeal Requirements: https://www.coretrustseal.org/wp-content/uploads/2019/02/CoreTrustSeal-Draft_Requirements_2020-2022_v00_01.pdf
- DINI Certificate: https://edoc.hu-berlin.de/handle/18452/22465

Other requirements and initiatives

- PlanS Requirements for Open Access Repositories: https://www.coalition-s.org/plan-s-practical-advice/
- Next Generation Repositories (behaviours, protocols, technologies that will enable new and improved technologies)
  - https://www.coar-repositories.org/news-updates/what-we-do/next-generation-repositories/
  - https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf
- Notify project: https://www.coar-repositories.org/notify/

- A data citation roadmap for scholarly data repositories:
  https://www.nature.com/articles/s41597-019-0031-8
- Data Repository Selection: Criteria that Matter (see COAR Comment)

4. Features and important development areas

DSpace can be used for a wide range of use cases, which is one of its strengths. Both the focus on a single use case and the mixture of different scenarios are possible. All use cases require better integration with preservation tools, smart storage of large files and compliance with accessibility and General Data Protection Regulation best practices. The FAIR principles of making content findable, accessible, interoperable and reusable apply to all use cases. DSpace should support persistent identifiers like DOI, ROR, and ORCID for best possible interoperability

Institutional Repository

DSpace adheres closely to repository standards, such as the COAR recommendations and PlanS requirements for Open Access repositories. Automations for data import, content enrichments and better support of persistent identifiers make it easy to use. For extended use scenarios such as publication platforms, external services should be able to integrate.

Data Repository

As a data repository, DSpace is aligned with FAIR principles. It supports basic features such as large file handling, external storage connectivity, dedicated rights management, and file-based licensing. For special use cases, integration options are supported, e.g. with data processing systems such as CKAN, support for viewers for different digital objects, Jupyter Notebooks, and more specialized data repositories.

Digital Collection Management

DSpace enables dynamic collection creation mechanisms and focuses on representation and interaction with Digital Objects through media-specific extensions such as embedded viewers, video streaming, video transcription, OCR, and use of the International Image Interoperability Framework (IIIF).

Current Research Information System (CRIS)

DSpace enables the capture of content, particularly publications, according to the [Common European Research Information Framework (CERIF)](#) standard for the aggregation and delivery of research information. The use of persistent identifiers and classifications guarantees an easy integration of the content into third-party CRIS systems.

Learning Object Repository (LOR) / Open Educational Resources (OER)

Learning objects can be created with a wide variety of tool chains. An automated upload to DSpace enables the publication, versioning and permanent provision of the artifacts.