

DSpace 2015-18 Strategic Plan - Technology

Technology Goals

Over the last few years, the Steering Group along with various strategic working groups have validated the following vision statements which describe the goals of the DSpace open source product:

1. DSpace will focus on the fundamentals of the modern "Institutional Repository" use case. We are striving to meet the IR needs of the next 5-10 years.
2. DSpace will be "lean", with agility and flexibility as primary goals.
3. DSpace will include a "core" set of functionality that can be "extended" (think plugins) or have "hooks" (integration points) to complementary services/tools
4. DSpace will be designed in such a way that it can be easily/quickly configured to integrate with new & future tools/services in the larger digital scholarship "ecosystem"
5. DSpace will support low-cost, hosted solutions and deployments (by featuring an easy, "just works" setup)

Assumptions

It is worth being aware that several assumptions are made in the drafting of this strategic plan for technology:

- We do NOT plan to rewrite DSpace from scratch, for the following reasons:
 - We have a highly active (and global) development community on the existing platform. We are averaging 50+ contributors in recent major releases. We also have a very active and healthy set of service providers.
 - A complete rewrite would require significant funding and centralized resources, neither of which are currently available. There also seem to be few (if any) grant opportunities to rebuild existing, established platforms.
 - A complete rewrite is very risky in the open source world. While in some cases it can succeed, it also can run the risk of fragmenting or fracturing a user community or developer community.
- We ARE aiming for a potentially substantial leap forward in user experience / web user interface.
 - We've heard the feedback that neither of the two UIs (JSPUI or XMLUI) provides an optimal user or administrative experience. So, a User Interface rewrite or major refactoring would be "on the table".
- The below actions and goals are ambitious, but we believe they are achievable provided that
 1. We can also achieve our [Sustainability goal of increasing project revenue to hire a Product Manager](#), and
 2. Institutional stakeholders are willing to commit developers to spend time working on organized development sprints under the direction of the DSpace Technical Lead to achieve the deliverables.

Based on this proposed value proposition and assumptions, the Steering Group recommends the following actions corresponding to each goal:

Goal 1: DSpace will focus on the fundamentals of the modern "Institutional Repository" use case.

In November 2002, DSpace was initially announced as an out-of-the-box "institutional repository software platform" (see [DSpace 1.0 release announcement](#)). While that basic goal has not changed, the common needs and use cases of an "institutional repository" have changed significantly in the last decade or so. Therefore, this goal is oriented towards striving to retain DSpace's niche while revitalizing it to meet current and future use cases associated with the modern repository platform.

- **Action 1A:** Verify and validate the needs of a "modern institutional repository". This is instrumental in formalizing the value proposition of DSpace.
 - This activity has already begun part of the 2014 Vision Survey and [Use Case](#) gathering by DCAT. However, the gathered [use cases are still being analyzed and prioritized](#).
- **Action 1B:** Survey the community on Technology Roadmap / drafted feature rankings
 - Ranking of features or validation of our Technology Roadmap
 - Also an opportunity to possibly gather volunteers for specific feature projects
- **Action 1C:** Identify minimum set of functionality/features for 'IR-core' and refactor codebase to provide this. This core may not be functional as-is, since it may require plugins that aren't extensions (e.g. authN)

Goal 2: DSpace will be "lean", with agility and flexibility as primary goals

Since its initial release in 2002, numerous features, configurations and options have been added to the DSpace codebase in an ongoing effort to keep up with the changing needs of its user base. While many of these changes have helped us to achieve new use cases, in some instances they have also complicated the codebase and made setup and upgrades more complex. Therefore, this goal is oriented towards cleaning up (and simplifying) the codebase and its configuration options, while also working towards avoiding duplication (of code and development efforts). We feel DSpace can be a "leaner" platform, which will allow the codebase to better adapt to the needs of the future and simplify its maintenance, setup and upgrade processes.

To be "lean", the DSpace technology platform should *avoid duplicative functionality* except where necessary to meet use cases or achieve "flexibility" goals. Where unnecessary duplicative functionality already exists, the technology team should choose a "best option" solution, or propose building a new solution when a "best option" does not exist.

- **Action 2A:** Converge on a single, out-of-the-box user interface (UI). DSpace will no longer be released with multiple User Interfaces (JSPUI vs XMLUI). A single user interface should be developed as DSpace's out-of-the-box UI. Early discussions on the requirements of this single UI (and some brainstormed candidates) are at [Brainstorms on a Future UI](#).
- **Action 2B:** Converge on a single, out-of-the-box search/browse system. DSpace will only support Apache Solr for search/browse, and the older, deprecated Lucene and DB search/browse system should be removed.
- **Action 2C:** Converge on a single, built-in statistical engine. DSpace will only support a single, built-in statistical engine (based on Apache Solr), and support for Elastic Search statistics should be removed or migrated to an optional module. Support for Google Analytics will be retained, as it's an optional integration with an external statistics engine.

- **Action 2D:** Develop a basic User Interface "style / layout guide". In order to ensure a consistent user experience across all pages/features within the User Interface, we should provide basic guidelines for layout and styling of common page elements, etc. Examples may include basic guidelines for how errors/warnings/notices should be displayed, what class(es) to use for types of buttons, etc.

Goal 3: DSpace can be "extended" (think plugins) or have "hooks" (integration points) to complementary services/tools

There will obviously be limitations to what DSpace can and should do, so we need to have ways to support plugins/addons/extensions to that core functionality. Not all users of DSpace will need to achieve the same set of Use Cases, so we will need to define which are "core" and which would be better implemented as plugins/extensions (either centrally supported or third-party supported).

- **Action 3A:** Define a family of specifications/interfaces/tooling for 'implementation' plugins (like authN, storage, persistent ID services, etc).
- **Action 3B:** Define a family of specifications/interfaces for functional extensions to 'core' DSpace (working title: 'modules'), and refactor existing bundled code to conform to new model (if appropriate/cost-effective).
- **Action 3C:** Provide infrastructure/tools for a module registry, where users can discover, and install modular extensions. Likely include both modules maintained by committers and community contributions.
- **Action 3D:** Devise a flexible but rigorous system of versioning all components (core, module, etc) where compatibility requirements can be checked/enforced by the build/deploy tools.
- **Action 3E:** Define and expose new interfaces (in 'core' DSpace and possibly modules) to allow local customized code to run: 'integration points'.
- **Action 3F:** (Highly dependent on UI architectural work) Provide a user-discoverable registry/library of user interface templates (working title: 'themes'), that can be installed and adapted for local use.

Goal 4: DSpace will be designed in such a way that it can be easily/quickly configured to integrate with new & future tools/services in the larger digital scholarship "ecosystem"

In order to continue to play a key role in the larger digital scholarship "ecosystem", DSpace must provide ways to both share and consume data/content from external services. We should strive to make all information in DSpace "shareable", and also ease the process of adding information to DSpace by providing Administrators with tools to consume data from other locations.

DSpace should provide easy and out of the box integration with external services in the following areas:

- **Action 4A:** Support ingest of complete metadata records (items) from external services, with or without files. External services may include: CrossRef, [DataCite](#), PubMed, ORCID works, [SHARE](#), etc.
- **Action 4B:** Provide the ability to consume external authority control sources to enrich specific metadata fields. External authority control services may include: ORCID or VIVO for author/contributor data, Fundref for funder metadata, Sherpa Romeo for Publisher OA policies information, etc.
- **Action 4C:** Expose DSpace metadata and content to external services. Allowing pushing metadata and content to external APIs, or allowing external services to harvest (pull) information from a DSpace repository. For example, DSpace content should be made available to [Europeana](#), [OpenAIRE](#), [RIOXX](#) compliant harvesters (UK), [SHARE](#)
- **Action 4D:** Expose DSpace usage data to external services.
- **Action 4E:** Integrate with external Authentication and Single Sign on services. Examples may include: [UK Federation](#), [OpenAthens](#), [OpenID](#), Google/Facebook/LinkedIn/ORCID authentication
- **Action 4F:** Integrate with external services providing identifiers ([Handle](#), [DOI](#), [DataCite](#), ...)
- **Action 4G:** Integrate with external storage and backup services ([DuraCloud](#), Amazon Glacier/S3, [Arkivum](#), [Archivematica](#), ...)

To integrate with parallel projects and initiatives (fedora, hydra, islandora) we first need to pin down the use cases of what those integrations will bring to DSpace, or what these will bring to the other platforms. They currently do not fit immediately in any of these five areas.

Goal 5: DSpace will support low-cost, hosted solutions and deployments (by featuring an easy, "just works" setup)

DSpace should be easy to install without requiring Java development expertise, to configure without requiring server access, and to monitor from within the application. Basic configuration options, including the look and feel and selecting themes should be accessible from within the DSpace online administration area.

- **Action 5A:** Improve and simplify the installation experience for DSpace. This may include, but is not limited to...
 - Investigate download, packaging and installation tools for Java web applications to make it easier to build a working system. (What do similar systems use?)
 - Examine options for lightweight installation, with most configuration taking place from the web interface upon first use (see for example [WordPress](#), etc)
- **Action 5B:** Improve and assist with the upgrade experience for DSpace, especially in terms of simplifying the management of local customizations (branding, custom themes, etc). This may include, but is not limited to...
 - Investigate options to assist with upgrades (for example highlighting changes from core code or configurations)
- **Action 5C:** Make configuration and basic theming an easier experience for hosted or low-cost deployments by migrating most options to the administrative interface. Some examples include..
 - Move configuration of basic theme configuration options (colours, logo) into administrative interface;
 - Allow most configurations to be edited (and refreshed) from the administrative interface
- **Action 5D:** Ensure data is never solely stored in "transient" technologies (e.g. Solr indexes or other such indexes) where it could be accidentally corrupted or lost. All DSpace data should be stored in a stable, persistent data storage system (e.g. database, filesystem), and then indexed from that location into tools like Solr, etc.
- **Action 5E:** Provide recommendations around scaling and load-balancing large DSpace instances.
- **Action 5F:** Provide administrators with additional system reporting features within the UI. Example use cases may include..
 - Alert administrators when new upgrades are available
 - Alert administrators when common system issues or misconfigurations are encountered (e.g. "Solr is not accessible / working", "assetstore is unavailable / unattached", "space in assetstore is low");
 - See [Admin UI - System Alerts via Admin UI](#) for more examples.

