

# DSpace 7 Benefits

## Is It Worth Upgrading?

*Matus Formanek*

---

### ABSTRACT

This study discusses the importance of the DSpace open-source software that supports numerous digital libraries and repositories around the world. With the release of DSpace version 7, a natural question that arises is whether the new version offers enough new functionalities to motivate system administrators to upgrade. This paper briefly describes the most important changes, including new features and bug fixes, included in DSpace 7.4 and prior minor versions. The next parts of this paper explore our estimate that there are several thousand DSpace-based systems globally that will likely have to be upgraded in the near future. The main reason for this need is that older versions of DSpace (including 5.x) have reached the end of their developer support period or are reaching it in mid-2023. Based on our own upgrade experience, we propose suggestions and recommendations on migrating from the previous DSpace 6.3-based environment to the new one in a case study that concludes this article.

### INSTITUTIONAL REPOSITORIES AND DSPACE

Countless new research papers are being produced around the world daily thanks to the hard work of scientists and research teams. The significance of their research is fully realized when it is shared with other scientific communities. Scientists need to disseminate the outputs of their scientific work in an appropriate way. Khan and Sheikh emphasized in their work that higher education institutions are confronted with the intricate task of preserving and disseminating their intellectual outputs. To effectively address these challenges and ensure open access (OA) to research works, institutional repositories (IRs) as information systems emerge as indispensable and substantial tools for safeguarding these scholarly documents in various digital formats.<sup>1</sup>

In the context of this study, the term institutional repositories refers to digital archives established to collect, preserve, and distribute scholarly and research-related digital scientific outputs within an organization, particularly in academic and research institutions.<sup>2</sup> As Zervas et al. further stated, institutional repositories were originally developed to showcase their academic work.<sup>3</sup> However, valuable digital content must not only be disseminated but also preserved in a reliable way. As Khan, Khousa, and Thelwall noted, digital repositories—institutional repositories included—assume a significant role in facilitating these procedures by furnishing the necessary policy and infrastructure to uphold efficient curation and enduring preservation in contemporary times.<sup>4</sup>

An appropriate application and other support software is required to create a digital (institutional) repository or digital library. Either proprietary or open-source software is used. In our work, we will focus on a particular type of open-source software for supporting digital libraries and repositories. As Mapulanga pointed out, the utilization of open-source software has empowered developers globally, granting them the capability to craft tailor-made features and functionalities for the systems.<sup>5</sup> This type of software is usually obtained and used for free, which

**Matus Formanek** ([matus.formanek@fhv.uniza.sk](mailto:matus.formanek@fhv.uniza.sk)) is Assistant Professor, Department of Mediamatics and Cultural Heritage, Faculty of Humanities, University of Zilina, Slovakia. © 2023.

is one of its indisputable advantages. To maintain an objective perspective, it should be acknowledged that implementing and maintaining an open-source operating system and software can entail significant additional costs, despite the potentially low or nonexistent initial acquisition cost of the software.

## **DSPACE**

The focus of this paper is an advanced, open-source software called DSpace. It is developed to run modern open digital libraries and repositories in academic or other organizations and is supported by a wide community of developers and enthusiasts around the world. DSpace provides a robust software framework for durable access to research materials and other valuable digital content, assets, and digital resources, available online, all with an eye on their long-term preservation. For example, DSpace is applied to preserve and provide access to academic faculty and students' papers, preprints, projects, audio/video visual materials, datasets, or any other important materials.<sup>6</sup> In the words of the software developers, the long-term goal of open-source software, which also includes DSpace, is, among other things, to ensure that "current and future generations have access to our collective digital heritage."<sup>7</sup>

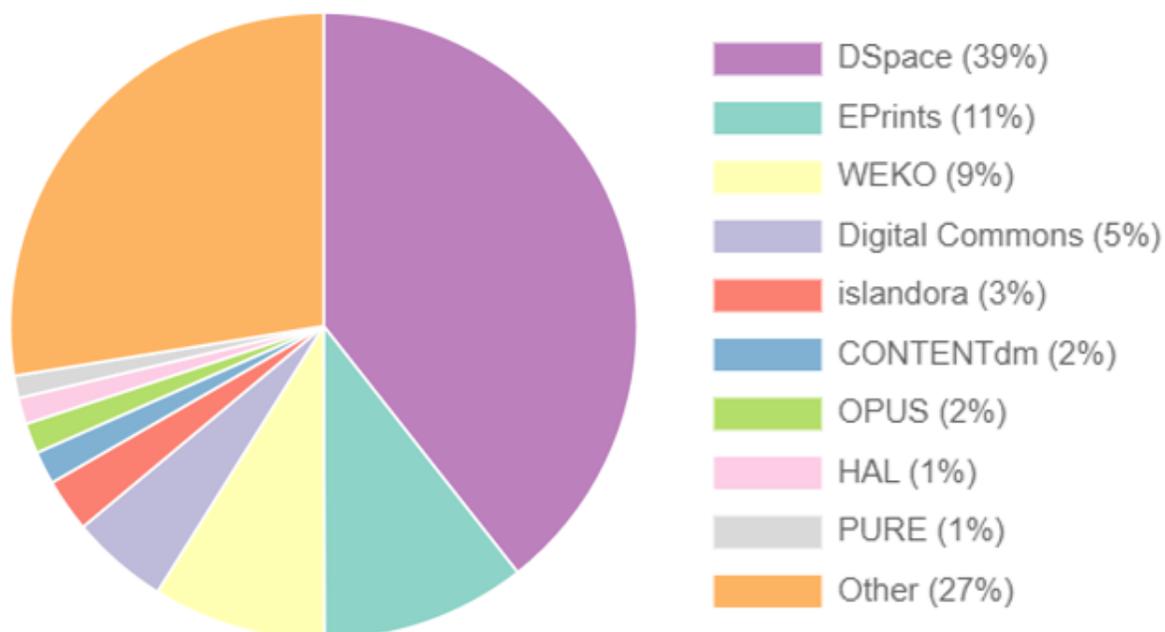
The DSpace project was founded in 2002 by Hewlett Packard and MIT. As a result of their close and intensive collaboration, DSpace 1.0 was released on November 4, 2002.<sup>8</sup> As an advanced software tool for supporting digital libraries and repositories, it thus celebrated its 20th anniversary last year.

Since its inception, development has progressed rapidly. As the number of active users grew and these users demanded more and better support, it was necessary to establish the DSpace Foundation—a nonprofit organization that provided technical leadership and administration to support the increasing number of DSpace users. The dynamic speed of development has become a characteristic sign of DSpace even today. Currently, the development of the DSpace project is managed by Lyris and directly supported with in-kind contributions of development resources and financial donations through a community sponsorship program. Lyris provides leadership and innovation for open technologies that promote durable, persistent access to digital data. This organization collaborates with many academic, scientific, cultural, and technology communities.<sup>9</sup>

DSpace, as software, is available to download for free from the official Lyris website (<https://DSpace.lyris.org/download/>). It may also be freely installed and used, as a low-cost, on-premise software solution. In that case the repository instance is installed on a local server at an institution. Alternatively, if preferred, it is possible to host the entire digital repository solution in the cloud as a subscription service for archiving, sharing, and managing content. In our study, we focus primarily on the important steps and principles of setting up and operating a locally run instance of DSpace as well as understanding these steps and principles in more depth.

Currently, DSpace is the world's most widely adopted software system for running open archives, libraries, and repositories in general. This claim is supported by numerical data obtained by the globally respected, authoritative registry called Open Directory of Open Access Repositories (OpenDOAR), which regularly collects and updates relevant metrics each month. DSpace has held the largest share for adopted software for open access institutional repositories, approximately 39 percent, among other similar software solutions for several years in a row (see figure 1).

**Figure 1.** DSpace’s dominant position among other similar software solutions.<sup>10</sup>



We employed shown data to demonstrate the extensive utilization and broad acceptance of DSpace in comparison to other alternative systems. Additionally, the data retrieved from OpenDOAR (see fig. 1) are supported by equivalent up-to-date data from other authoritative registries such as Registry of Open Access Repositories (ROAR) and Re3data.<sup>11</sup> We will consider data from the DuraSpace Registry as supplementary because this registry was created directly by Lyrasis (which is the organization that manages DSpace’s development).<sup>12</sup> In view of this fact, it is reasonable to assume that this registry mainly includes the systems supported by them (except DSpace, Fedora, and VIVO are also systems maintained by Lyrasis). Thus, in this case, we do not consider the high percentage of DSpace-based systems in the registry by Lyrasis to be representative.

All retrieved data is shown in table 1 below and up to date as of January 10, 2023.

**Table 1.** DSpace market share.

	OpenDOAR	ROAR	Re3data	DuraSpace registry
Total number of entries	6,002	4,725	3,072	3,703
Sum of DSpace-based repositories and libraries	2,363 (DSpace) + 16 (DSpace-CRIS) <b>(39.64%)</b>	2,377 <b>(50.3%)</b>	120 <b>(3.9%)</b>	3,138 <b>(84.7%)</b>

As is shown in table 1, except in the Re3data Registry, DSpace is the most popular software. These findings are consistent with the findings of the following research papers and surveys focusing on the topic of open repositories and the widespread popularity of DSpace. Surveys conducted by

several studies found that institutional repositories are more likely to use DSpace, Fedora, Eprints, Greenstone, and Invenio in this order.<sup>13</sup> Previous research by Ivanović et al. has also found very similar results for DSpace adoption by institutional repositories.<sup>14</sup> Scientific papers written by Khan and Sheikh and Ahammad also highlight the important position of DSpace as a world leader in open-source software for digital repositories, or open-source digital library software (OSDLS). These studies, which only focused on universities in Pakistan, showed that DSpace is used by 47.6 percent of the digital library or repository systems.<sup>15</sup> Gurikar and Hadagali also investigated the adoption of open-source software and the role of underlying software for institutional repositories within India. Based on the data collected, they confirmed that DSpace and Eprints are the two most popular tools used.<sup>16</sup>

### **END OF SUPPORT FOR OLDER DSPACE RELEASES**

Released in December 2013, DSpace 4.x is a release that remains currently unsupported along with all previous DSpace versions since 1.4.0. Its support ended in July 2021. The newer major DSpace release (5.x) is officially unsupported as of January 1, 2023. This date represents the release end of important security updates. The current major release 6.x (or 6.4) will also see end-of-life as of July 1, 2023.<sup>17</sup> After this date, bug or other security fixes will no longer be published for these releases. System administrators may worry with this announcement, as explained in the next paragraph. In addition, Tim Donohue, technical lead for DSpace at Lyrasis, posted a brief explanation for end of support on the Lyrasis wiki pages. The primary reason for DSpace 5.x as well as 6.x is as follows:

[T]he support period for any major release was usually 3–4 years. The support period for 5.x and 6.x was necessarily extended to allow for the development of 7.x (which is a significant update to the DSpace platform). As you may be aware, in the last year, several major open-source libraries have discovered major security vulnerabilities. While DSpace 5.x and 6.x were luckily not impacted, those events forced us to reanalyze the ongoing support promised for these releases. Because both versions 5.x and 6.x were designed and built on technologies from 6–7 years ago, they have become difficult to maintain. The age of these releases impacts our ability to promise that all future security vulnerabilities can be addressed. These old codebases are showing their age and cannot be made as secure as the 7.x platform. Therefore, we highly recommend all institutions begin planning an upgrade to DSpace 7.x as soon as reasonably possible.<sup>18</sup>

### **DSPACE MIGRATION AND UPGRADE: CASE STUDIES IN CURRENT LITERATURE**

Worldwide, there are relatively few scientific or conference papers that are in any way related to the issue of migration, modification, or upgrade of a repository/library based directly on DSpace. Zervas et al. focused on the transformation of a locally running repository into a Current Research Information System (CRIS), which we consider as an upgrade process.<sup>19</sup> Bimbe et al. dealt with challenges related to implementation and upgrading a DSpace-based institutional repository at the University of Zambia. Their paper explained the steps taken, the challenges faced, and the process taken to install the software and perform an upgrade to the latest version.<sup>20</sup> Casados et al. described in their paper an entire workflow of transformation from a legacy repository that was developed in house to a solution based on DSpace-CRIS.<sup>21</sup> Kawale and Vandana published their conference paper relating to the process of installation and customization of a DSpace-based institutional repository.<sup>22</sup>

As these studies show, the process of migration or upgrade is not entirely simple, even in the case of DSpace. This paper can help the expert community to raise and partially resolve key issues and assure system administrators that upgrading is a step forward. However, it needs to be carefully considered first. The steps need to be individually explored and tested in the context of each organization’s capabilities. Zervas et al. observed that a shared objective is to ensure the continuous alignment of digital repositories with the latest technological advancements. Institutional repositories must consistently adjust their mission, objectives, and services to align with those of the university and the community they cater to.<sup>23</sup> Iowa State University successfully migrated its old digital repository to an open-source DSpace 7 instance as early as January 2022. This migration was accomplished following extensive years of research and meticulous planning, with the aim of presenting users with a more streamlined interface, enhanced search functionalities, and a more adaptable repository structure, while preserving all the services and statistical data that users have grown to value.<sup>24</sup>

Given the fact that DSpace is currently the most popular software for running digital libraries and repositories, the need or necessity for early upgrades may become an issue or challenge for many systems worldwide. In the following section, we will attempt to sketch the distribution of each DSpace version in a production environment.

**NUMBER OF DSPACE-BASED REPOSITORIES RUNNING WORLDWIDE**

The aim of this section is to approximate the number of repositories worldwide that require upgrading, in accord with the already confirmed information about ending support of the current version.

To be able to approximately describe the current state of DSpace repositories in the world, we need to at least estimate the relative distribution between the different DSpace versions within the repositories deployed in a production environment. We cannot state this information with complete accuracy, but we can rely on the available data online in the repository registries.

Our survey was conducted to collect data from the DuraSpace Registry maintained by Lyris. It is the unique registry that also provides information about the version of DSpace used for each registered system—other registries (such as OpenDOAR/ROAR/Re3data) do not include this information in their entries. Figure 2 is a screenshot with a view of our local departmental repository entry, with the highlighted information about the used major version of DSpace. The specific minor version (e.g., 6.3) is not shown.

**Figure 2.** The digital repository of the Department of Mediamatics and Cultural Heritage in the DuraSpace Registry.

Organization Name	Site Name	Site URL	Type of Organization	Country	Technology	Version
Department of Mediamatics and Cultural Heritage, Faculty of Humanities, University of Zilina	Digital repository of The Department of Mediamatics and Cultural Heritage	uniza.sk	academic	Slovakia	DSpace	6.x

On January 10, 2023, in accordance with the methodology, we accessed the open registry DuraSpace<sup>25</sup> and filtered the list of all DSpace-based systems from that registry. The only filtering criterion was used technology (DSpace) independent of the geolocation of the system, its content,

or the access method. As of that date, the DuraSpace Registry displayed and exactly matched 3,138 entries that met the specified conditions. The registry arranged them row by row in a table format. Unfortunately, it was not possible to set up a more sophisticated filter so that we only retrieved entries for a particular major version of DSpace. Furthermore, the entries could be browsed on the site by groups containing only 25 entries. Thus, it was not possible to batch export all appropriate entries at once into some structured format like XLSX, CSV, etc. We contacted Lyris technical support by email with a request to export all entries as a single file, but support for this was not provided and the request was unmet. So, we laboriously copied all 3,138 entries from the web into a local MS Excel spreadsheet, page by page—126 pages in total. Then, we formatted and analyzed the structured data in MS Excel. We focused on the last column of the spreadsheet which contained, in most cases but not always, information about the version of DSpace that the system or repository was supposedly using.

The results are summarized as follows: Of the 3,138 registered DSpace repositories, only 2,415 entries contained in their metadata the information about the specific DSpace version, which represents 76.9 percent of the total number of registered systems. To avoid unwanted bias in the results as much as possible, we continued to work only with this set of repositories that had the used DSpace version properly listed in their metadata entries. Results are provided in table 2.

**Table 2.** DSpace’s versions share in DuraSpace Registry.

<b>DSpace version</b>	<b>Number of registered systems</b>	<b>Share in %</b>
1.x	709	29.36%
3.x	198	8.20%
4.x	313	12.96%
5.x	565	23.40%
6.x	613	25.38%
7.x	17	0.70%
<b>Total</b>	<b>2,415</b>	<b>100.00%</b>

Based on the provided data and as of January 2023, we observed that only 17 systems declared the usage of the latest major version of DSpace 7, labeled as “DSpace 7.x.” We suppose that there are certainly many outdated entries in the DuraSpace Registry and elsewhere because this information is only submitted by administrators. Lyris, as a service provider, does not proactively prompt administrators to continuously update the metadata. Moreover, it is possible that many systems already exist but are not registered anywhere. Further, many systems may have been—and certainly have been—upgraded to a higher version of DSpace but the information on the upgrade has not been corrected in the registry. It is therefore probable that the provided data does not fully reflect the actual real state. In support of this claim, the disproportionately high share (almost 30 percent) of the oldest DSpace version (i.e., 1.x) is noted in the total representation. If we completely ignore the probably highly overstated share of older DSpace versions (and consider only versions 5.x and newer), the adjusted values are shown in table 3.

**Table 3.** Estimated share of specific DSpace versions in DuraSpace Registry.

Known DSpace version	Number of registered systems	Share in %
5.x	565	47%
6.x	613	51%
7.x	17	2%
<b>Total</b>	<b>1,195</b>	<b>100%</b>

Based on the data mentioned and evaluated above, we can assume that DSpace 7.x is used globally in only approximately 0.70 percent to 2 percent of cases, within the community of organizations using DSpace. The most popular versions of DSpace are 6.x or even the older 5.x by a considerable margin. In conclusion, we estimate that up to 47 percent of global digital repositories and libraries running DSpace are using an outdated version (5.x or older) of this software.

If we apply this percentage to data retrieved from the authoritative and more recent OpenDOAR repository, we can reasonably assume that of the 2,379 registered repositories and digital libraries using DSpace (see table 1), only around 16 to 48 systems are running on some of the latest major versions of DSpace 7.x as of January 2023. Most DSpace systems—on the order of thousands—are already using the aging version 6.x or even the outdated version 5.x, which is alarming. Numerically speaking, we estimate that there are over 2,000 open repositories and libraries based on DSpace worldwide waiting to migrate to DSpace 7, which is a significant number. We are aware that our results may not be completely accurate and reliable due to the highly probable outdated data in the repositories, but we consider them in kind as indicators or approximations when describing the state of DSpace systems in the world at the beginning of 2023. A more rigorous and extensive survey in obtaining accurate numbers would be needed. From an information security perspective, the upgrade/migration to DSpace 7.x will help to protect digital objects. The subsequent sections present an introduction to DSpace 7 and discuss its principal advantages.

### **DSPACE 7 DEVELOPMENT**

The DSpace 7 project began after the Open Repositories 2015 conference. As a part of the DSpace technology roadmap and the strategic plan of Lyrasis, the developers, as well as the working groups formed during 2015–2017, focused on the new frontend user interface prototype based on Angular as well as on the backend based on the REST API. These key components became the basis for the upcoming DSpace 7 version.<sup>26</sup>

The release of the new major version of DSpace 7 (i.e., 7.0) was announced on August 2, 2021, via a press statement also published on the Lyrasis website. The statement indicated that DSpace version 7.0 brings a number of reasonably radical changes (which will be discussed later in this paper) as reflected in the change to over 1 million lines of code inside the software.<sup>27</sup> These changes are so significant in comparison to previous versions of DSpace that it is not just only another evolutionary step in its development, but a change in the fundamental paradigms on which DSpace is built.

Atmire, a company that co-develops DSpace with Lyrasis, has made the first preview release of DSpace 7.0 publicly available on its website for testing and feedback purposes since May 20, 2019. As Atmire states, DSpace 7 has been designed using test-driven development practices, aiming to detect and resolve potential issues at earlier stages of the development process, thereby ensuring more stable releases right from the outset.<sup>28</sup>

In a Lyrasis press release announcing the completion of DSpace 7.0, Kristi Park, DSpace governance chair and Texas Digital Library executive director, is quoted saying that the release of the final version of DSpace 7.0 was preceded by five beta releases, as well as the subsequent participation of the developer community in a public event called DSpace 7 Testathon in April 2021. Its goal was to report and mend bugs within the latest beta.<sup>29</sup> Consequently, the collaborative effort involving over 60 community developers, led by LYRISIS, Atmire, and 4Science, led to the development of DSpace 7.0. The project received funding support from over 25 institutions worldwide.<sup>30</sup>

A more detailed explanation of DSpace 7 benefits is offered in the following section.

#### **DSPACE 7.4 BENEFITS**

The major version of DSpace 7, according to the wide developer community, declares many radical changes as well as incorporating completely new features that not only improve the overall level of working with the system, but also add new functionalities and known bug fixes. There are several wiki-style articles published by Atmire and Lyrasis available on the internet, which draw the attention of system administrators to the most important changes made within DSpace 7 and its internal software components.<sup>31</sup> However, a detailed, scientifically prepared study focusing on a more comprehensive analysis and deployment of DSpace 7, which could appeal to system administrators, has been lacking in the expert community so far. We believe that summarizing and reviewing the new functionalities and the most significant changes in DSpace 7, as well as their impacts, will help administrators answer the main questions regarding the potential upgrade of their digital libraries and repositories to the newer major version of DSpace 7.

In the following paragraphs, we provide an overview analysis of selected major innovations made within DSpace 7. All information in our study is based primarily on a synthesis of official information available within the Lyrasis.org website. Other important sources are technical articles; technically oriented blogs and wikis maintained by experts in the field, especially those written by Tim Donohue, technical lead for DSpace at Lyrasis; educational YouTube channels by Furrukh Hussain Zai; official DSpace tutorials; and webinars published by institutions collaborating with Lyrasis.<sup>32</sup> We integrated, summarized, and synthesized the obtained information enriching it with our own DSpace 7 testing experience. We consider the experience of installing DSpace in your own environment (i.e., on site) to be nontransferable. We have actively used DSpace since version 4. We also built from the ground up the Digital Repository of Mediamatics and Cultural Heritage, which we have been maintaining since 2017 and is based on DSpace 6.3. For the purposes of this paper, we downloaded and installed the then most up-to-date version of DSpace 7.x, which was version 7.4, from October to December 2022, in parallel to the existing legacy repository still based on DSpace 6.3. Both systems run on the same environmental setup with two virtual machines running Ubuntu Server 22.04 LTS, four cores, 4GB of RAM. We were able to compare the two systems on the fly—systemically and architecturally in detail.

Lyrasis reports that the planned main benefits of DSpace 7 have been divided from the very beginning of their phased roll-out into tiers. These tiers are priority lines that contained the

separated tasks and milestones according to their level of importance/necessity as identified in a survey conducted among the “DSpace Steering & Leadership members and institutions they represent.” Tier 1 represents the highest priority and tier 5 is reserved for low priority. Eight out of nine features with the highest priority from tier 1 were implemented into versions 7.1, 7.2, and 7.3. The last available version of DSpace, 7.4, adds new functionality from tier 2 and tier 3, so we consider the introduced changes as significant.<sup>33</sup>

Based on all these new features, it is possible to conclude that DSpace 7.4 is already a sufficiently robust, stable, and feature-rich version from a system point of view, which we will describe in more detail. Since the first release of DSpace 7 (i.e., 7.0), several serious bugs have been fixed. We carry out a detailed comparative analysis and testing against the previous major version of DSpace (6.x) based on the numerous implemented changes, and in the following paragraphs, we will present those that are most significant. For clarity of this study, we have divided the list of changes into separate groups:

- Changes at software system level
- Advanced integration with other external services
- Working with digital content
- New security features and important bug fixes

The release of DSpace 7.5, which is a part of the roadmap, occurred in February 2023. It brought the next set of additional functionalities and enhancements within each tier. For example, the adding of item counts for communities and collections was planned. The next minor versions of DSpace 7.x (7.6, 7.7, etc.) will be released as needed to ensure that all important milestones within tiers 1 through 4 will be met.<sup>34</sup>

## **CHANGES AT THE SOFTWARE SYSTEM LEVEL**

This section describes the most significant and important changes that have taken place directly inside DSpace 7.x regarding system administration, its internal components, and the installation process. These changes are relevant to each system administrator who is involved in system installation, maintenance, or overall system administration.

### ***Splitting the System into Frontend and Backend Interfaces***

The most significant change made within DSpace 7.x, and the longest in preparation, is the splitting of the system into a frontend part, based on Angular technology, and a full-featured server-side, so called backend, with a completely new redesigned REST API.<sup>35</sup> Both these components are now installed separately, and as a result, they are linked to each other via references in configuration files.<sup>36</sup> According to Morris, it is possible to install and run each part (frontend and backend) on a separate server.<sup>37</sup> This installation allows DSpace to adapt more dynamically to the specific requirements and needs of a given environmental setup, or to scale the required performance more efficiently.

The completely new user interface, or frontend, combines the advantages of both now nearly obsolete web interfaces known from DSpace 6.x—XMLUI and JSPUI. They no longer meet the required technical criteria for modern dynamic and responsive websites, especially XMLUI, which relies on an old, unmaintained Apache Cocoon Framework. The new Angular technology used in DSpace 7 extends the capabilities and helps improve the user experience: the web interface is fast, natively responsive, and easily customizable and themed via HTML and CSS.<sup>38</sup> It better supports search engine optimization (SEO), which means that inserted digital objects are “easily

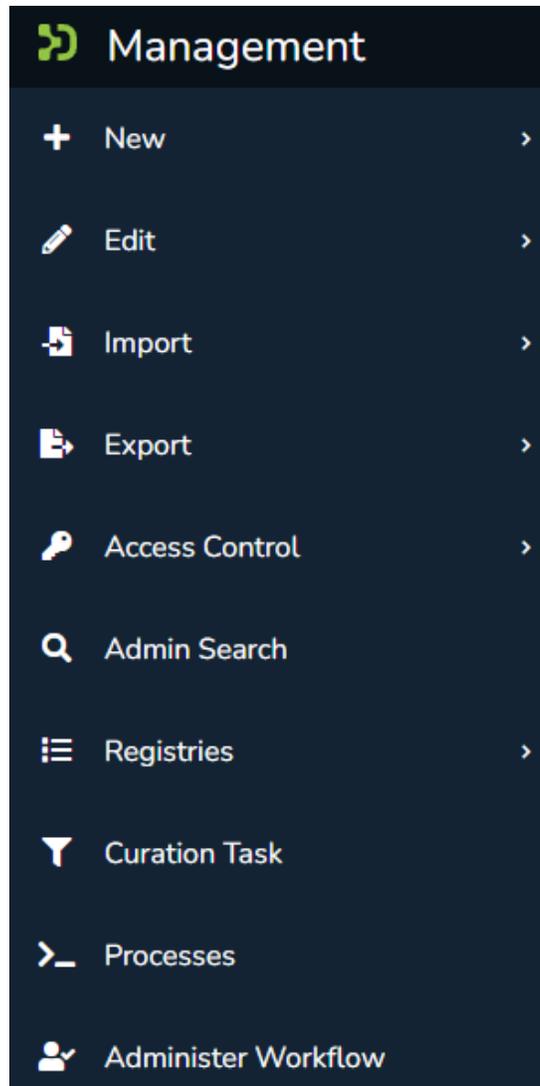
indexed/searched by Google or Google Scholar ... we verified this by running a series of tests with Google Scholar team.”<sup>39</sup>

Other significant features of the Angular-based frontend interface include improved accessibility and support for web archiving techniques (for example, Internet Archive harvesting). The web archiving process harvests, collects, and preserves digital content and artefacts from the internet (various web pages, media, blogs, forums, and other online sources) into a maintained archive. DSpace 7 integrates with other services, which will be further discussed. Angular as a technology, paradoxically, does not even require Javascript support to run such applications. As Donohue noted, in this case, each user gets “precompiled, static HTML pages each time they click links or buttons.”<sup>40</sup> Applications are written in TypeScript, an enhanced version of Javascript supported by Microsoft and Google.<sup>41</sup> Due to the advantages, “all users should immediately migrate to and utilize the new Angular User Interface.” The disadvantage is that there is no migration path from either the XMLUI or JSUI to the new user interface.<sup>42</sup> In other words, upgrading to DSpace 7 requires creating and modifying the completely new web user interface, but this can be done by any administrator with slightly advanced knowledge of HTML and CSS.

The Angular-based frontend serves as an interactive graphical extension over the systemically more important backend, which represents the server-side of the DSpace architecture. The backend’s most important part is the REST API. All functionalities available through the User Interface are now equally achievable through this interface. The complex backend includes tools to support OAI-PMH, SWORD, and RDF in addition to the REST API.<sup>43</sup> The original REST API, used since DSpace 4.0, was very limited in functionality due to read-only features. The updated and completely modernized REST API allows both read and write functionalities and provides “an easier path towards future integrations with DSpace (by other third-party platforms or plugins).”<sup>44</sup> In reviewing the REST API, we note that these radical changes have enabled integration with modern and widely deployed services for managing and making scientific content accessible. The old, outdated REST API did not support those features and it is, in short, necessary to update it to allow for further digital library system evolution, “following modern best practices, supporting modern technologies, and easier overall maintenance (maintenance).”<sup>45</sup>

### ***New Admin User Interface and Its Features***

Many changes were implemented in the graphical web interface during the redesign of DSpace’s frontend. The main feature noted by users upon login is the new modern admin panel for system administrators, which was added to the interface in DSpace 7.0, as shown in figure 3. Administrators can run and monitor all workflow processes as well as self-created backend scripts and other actions related to system administration, including various curation tasks, through this so-called Admin UI.<sup>46</sup> The outputs of these actions are also visible through this graphic interface. Administrators can now execute the filter-media script from the Admin UI in order to immediately update thumbnails and full-text indexing. In DSpace 6.x, this could only be done via the console or command line.<sup>47</sup>

**Figure 3.** New admin panel.

Based on our own experience, it is also possible to edit or search the digital objects more intuitively through the Admin UI than in any previous DSpace versions. Authors placed more emphasis on controlling the environment with the mouse. This is evident in the addition of well-known drag-and-drop support or in the newly added so-called quick actions buttons. The authors noted that it is possible to withdraw items or make items private. The process of metadata editing has also been simplified. As Donohue pointed out, it “introduces suggest-as-you-type for field name selection of new metadata.”<sup>48</sup> In the case of problems that may occur during possible workflows, system administrators can log in as a specific person (type EPerson in DSpace), allowing them to “debug issues that a specific user is seeing, or do some work on behalf of that user.”<sup>49</sup> From the administrator’s and user’s point of view, the control and management of the system is clearly more comfortable and convenient than it was compared to older versions of DSpace.

#### ***Batch Import and Export from Admin User Interface***

As of version 7.4, DSpace supports direct batch import and export of collections and items to and from a ZIP file through the Admin UI. This workflow could only be done via the console in the past.

As a result, the migration process of collections or partial items between DSpace systems has become faster. However, it is still not possible to migrate an entire site, or repository, through the web interface in this way.

### ***New Content Detection and Analysis Framework***

This feature is not completely new. Older versions of DSpace have been able to successfully extract textual content from items and index it for fast searching. Since DSpace 7.3, a new text and metadata extraction filter for full-text indexing has been implemented by using a software tool written in Java—Apache Tika. The new filter is enabled by default and supports all the most popular text formats containing digital media, including Adobe PDF, Microsoft formats (Word, Excel, PowerPoint), CSV, HTML, RTF, TXT, as well as all open document formats.<sup>50</sup>

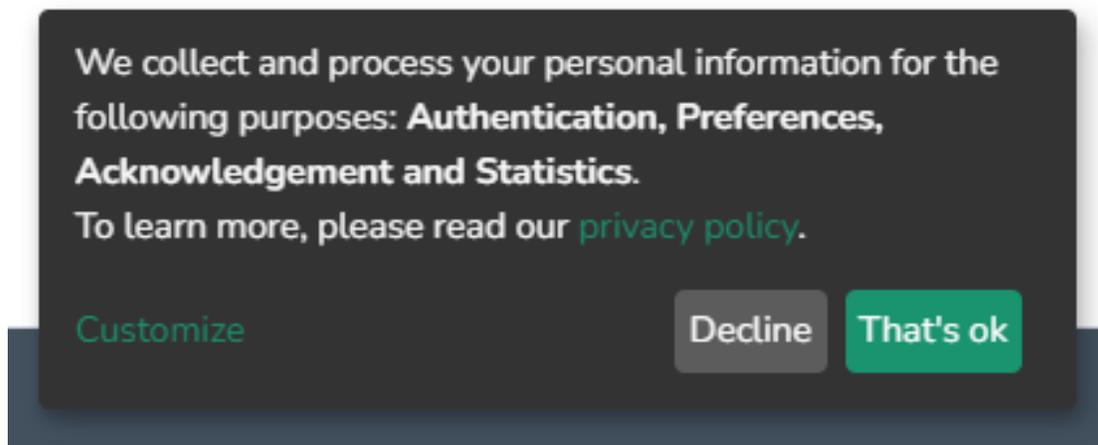
### ***New Feedback Form***

A new feedback form is now linked in the footer of every page. The system administrator defines “feedback.recipient” conveniently already in local DSpace’s config file.<sup>51</sup>

### ***Improved GDPR Alignment***

Many system administrators remember the coming of the General Data Protection Regulation (GDPR), especially in the European Union, when they had to create and define their own user agreement on all web pages with some cookie preference popup window or banner. It should describe how their data is used through a privacy statement policy. DSpace 7 natively integrates this functionality and, therefore, there is no need to create a banner from scratch. Now, a cookie preference popup, as is shown in figure 4 below, immediately appears when first accessing the site. By default, it displays on the bottom bar, near the right corner. The system administrator can modify the wording and behavior of the popup.

**Figure 4.** A cookie preference popup window.



### ***Welcome Email***

The introduction email represents an additional advancement in the system. This elective feature has been accessible since version 7.3, allowing the transmission of a welcome email to all recently registered users. This capability can be activated by utilizing the recently introduced mail.welcome.enabled backend configuration in your local.cfg file.<sup>52</sup> While a simple functionality, users may be pleased with it from the list of the new DSpace 7 features.

### ***Easier Installation Process***

A more significant change within DSpace 7.x is the simplification of its installation process in comparison to DSpace 6.x. The ability to quickly install DSpace 7 via container technology, such as Docker, has been added, without requiring much technical knowledge and skills. However, Donohue recommends this method only if you just want to try it out quickly for development or test environments.<sup>53</sup> From our point of view, for production use cases, a traditional step-by-step installation using a manual is better.

As we observed, the DSpace installation process requires extra attention and experience with advanced operating system administration, such as Windows or Unix-like. We recommend following the installation guidelines, which are available thanks to a wide community of experienced users. The community has made available a number of valuable tutorials in text form as well as in video formats.<sup>54</sup> We especially recommend the expert YouTube channel called *Ready Access to Free Library and Information Management Technologies* by Furrukh Hussain Zai.<sup>55</sup>

Vimal Kumar, a digital libraries researcher, has contributed valuable work on the topic of digital libraries through his scientific papers as well as the DSpace Geek blog, which contains tutorials on this topic.<sup>56</sup> We have also successfully used these resources to install a testing instance of DSpace 7.4, so we consider these steps as helpful.

Installing DSpace 7.x also requires a recent version of the database system (DBMS). This is PostgreSQL 14, at the time of writing.<sup>57</sup> Donohue provided a cautionary notice to readers regarding the deprecation of Oracle Database support, with its complete removal scheduled for mid-2023. Consequently, all websites should diligently plan for a migration to PostgreSQL.<sup>58</sup>

In addition to the database system, where the data is stored, a newer version of the Tomcat web server (9.x) is also required because it fixes several critical security vulnerabilities. Among the software prerequisites, the Open JDK v11 and other software components (Apache Ant, Maven, etc.—all are described in DSpace's manual) must be installed.<sup>59</sup>

Compared to older DSpace versions, there were no problems with the integration of the various software prerequisites mentioned above when installing DSpace 7.x. Based on our own experience, we state that webapp deployment is completely seamless within the Tomcat 9.x configuration, even for less experienced users, in contrast to DSpace 6.x, where we had to use the older Tomcat 8 instead of the version 9 due to many persistent app deployment issues.

On the other hand, DSpace 7.x installation includes a few extra steps. Donohue warns that Apache Solr is now excluded from the DSpace installer and must be installed independently as a separate dependency in conjunction with the DSpace backend.<sup>60</sup> In DSpace 6.x's installer, it was already embedded.

Solr is one of many important software dependencies used within DSpace. It is used in high-traffic websites, which enables features such as full-text searching and real-time indexing. To perform the testing, we installed Solr 8.11.1 according to the available tutorials by Kumar and System Librarian and ran the service on default port nr. 8983 with no problems or errors, as shown in figure 5 below.<sup>61</sup> We consider the Solr installation simple, and it can be accomplished by even a moderately advanced DSpace system administrator.

**Figure 5.** SOLR service running on port 8983.

### ***Advanced Integration with Other External Services***

Significant improvement had been made with the release of DSpace 7.3 in its integration with other services. For this reason, we have decided to dedicate a separate section to describe these changes. The support for importing metadata via external APIs, such as ORCID, PubMed, Sherpa Journals or Sherpa Publishers among others, is considered the most significant and improved features.

Many integration features included in DSpace 7.x are ported from DSpace-CRIS—a separated project maintained by 4Science. CRIS (Current Research Information System) is an extension to DSpace that provides a dedicated set of functionalities that give added value to researchers' work and the repository. It is

a picture of the current research activities undergoing in a research organization. The scope of a CRIS system is very broad as it covers people (researcher profiles, CVs, supervisions), research outcomes (publications, patents, research data), funds, projects, professional activities (invited presentations, peer or editorial reviews, etc.), or management information (metrics, bibliometrics, statistics). The content of a CRIS is analyzable and reportable. Content within a CRIS system can be easily connected and linked with other content.<sup>62</sup>

As we see, DSpace-CRIS aims to collect data about research activities and does so through its close integration with other services, which is also its biggest advantage. The DSpace-CRIS is not covered in this paper, although it is also DSpace-related. It represents a separate thread of DSpace software development, so our focus remains on the relevant new features implemented in DSpace 7.x.

### ***ORCID Authentication and Synchronization***

As of DSpace 7.3, enhanced ORCID authentication and synchronization is implemented. Each user with a DSpace account, or EPerson, can have the DSpace Researcher Profile assigned in the system. It can be synchronized, or initially created, via ORCID Integration. This feature is turned off by default and must be enabled manually, if needed.<sup>63</sup>

Although the integration with the ORCID service has been supported by DSpace for a long time, all advanced ORCID-related features are fully available in DSpace-CRIS.

### ***External Lookup of Digital Object's Metadata***

While this interesting feature is planned, the functionality is not yet available. It is based on an integration with services providing persistent identifiers (e.g., DOI) to enrich an in-progress submission.<sup>64</sup> After the implementation, digital content can be directly imported from nine new external services including CrossRef, Scopus, Web of Science, PubMed Europe, CiNii, NASA

Astrophysics Data System (ADS), VuFind.org, SciELO.org, and the European Patent Office (EPO). DSpace's official YouTube channel also mentioned the support of arXiv in this way.<sup>65</sup>

### ***Support for OpenAIRE API***

The OAI-PMH interface has undergone enhancements, allowing for the import of comprehensive funding information (including Funder, Funder Identifier, Funding Stream, and Funding ID) from the OpenAIRE API when importing Project Entities (Research Projects). This functionality has been implemented through a novel external source utilizing the Live Import from external sources feature.<sup>66</sup>

### ***Single-Sign-On (SSO) using OpenID Connect (OIDC) Authentication***

This entirely new authentication plugin has also been adapted from the DSpace-CRIS project. It enables the support of authentication through various providers like Google, Microsoft, Amazon, and others.<sup>67</sup>

### ***Support for Remote Handle Resolver***

Since DSpace 7.4 it is possible to run Handle Server remotely on a dedicated system. This way may be preferred by some administrators.

## **WORKING WITH DIGITAL CONTENT**

Another large group of improvements are changes related to working with digital content, specifically individual files and items.

### ***Drag and Drop Support***

The main submission and workflow features have been redesigned into “a one-page submission process with a drag-and-drop interface & a searchable MyDSpace.”<sup>68</sup> It is shown in figure 6. Manual submission of digital objects is therefore slightly easier and faster for all users.

**Figure 6.** An item submission.

Moreover, starting from DSpace version 7.4, there is now added support for Markdown, HTML, and MathJax within item abstracts. Additionally, all metadata fields now include support for line wrapping.<sup>69</sup>

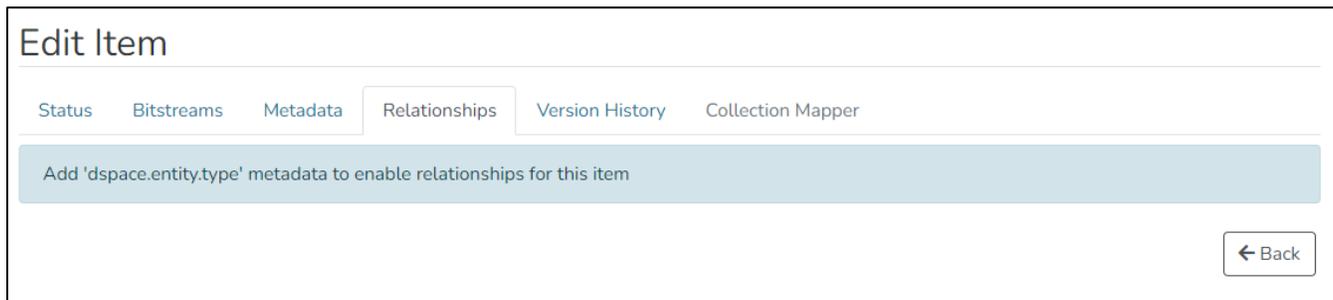
### ***Extended Support for Item-Level Versioning***

Since DSpace 7.1, it has been possible to create a new version of an item. Versioning is a new functionality to build the history of a digital object. Users have the ability to create new versions of an item whenever they make any modifications to it. Item versioning is enabled by default.<sup>70</sup>

### ***Support for New Configurable Entities with Relationships***

DSpace 7 supports an optional feature through which relationships between different DSpace “configurable entities” (e.g., items) can be captured and expressed, as shown in figure 7. These entities can have configurable relationships to other entities, for example, a set of journal hierarchy entities (Journal, Volume, Issue, Publication) and a set of research entities (Publication, Project, Person, OrgUnit).<sup>71</sup>

**Figure 7.** Item relationships.



Since DSpace 7.3 there has also been support for Versioning of Configurable Entities. It is now possible to create versions of Entities that automatically preserve all relationships.

When Configurable Entities are enabled, in the “Edit Collection” page you can select an Entity Type (e.g., Person, Project, Journal, etc.) that Collection will accept. Once configured, this Collection will only accept new Submissions of that Entity Type, and will be one of the recommended Collections to Submitters whenever they start a new Submission of that Entity Type.<sup>72</sup>

Entities can also be imported, including their relationships from another system via the Simple Archive Format.

### ***Item Access Status***

This is an optional feature that must be manually enabled via the new “showAccessStatuses” setting in the config file. Subsequently, all item lists can display the status of the item (e.g., “open access,” “restricted,” “metadata only,” “embargoed”).

### ***Redesigned Search Box***

Currently, users possess the ability to conduct searches within the header of any web page by activating the magnifying glass icon. Subsequently, the search results page has been augmented to incorporate automatic search highlighting, user-friendly drop-down and search filters, as well as the option of thumbnail-based results.<sup>73</sup> These advancements are designed to optimize the search process, enabling users to promptly access pertinent information and resources. In comparison to earlier versions of DSpace, DSpace 7 now offers a more sophisticated and comprehensive search functionality, significantly enhancing its overall usability and accessibility for users.

### Support for Video Content Streaming

When enabled, DSpace can now stream videos and view images in full screen, using a basic out-of-the-box Media Viewer.<sup>74</sup> Downloading the video is no longer required and users can directly stream from the website, as shown in figure 8.

**Figure 8.** Video streaming plugin activated in DSpace 7.



The screenshot shows the DSpace 7 user interface. At the top, there is a navigation bar with the DSPACE logo and links for "Communities & Collections", "Statistics", and "All of DSpace". Below this is a breadcrumb trail: "Home • Katedra mediamatiky a ku... • Seminárne a semestrálne ... • Audiovizuálna tvorba • Pre". The main content area features a large heading "Prezentačné video FHV UNIZA" and a video player. The video player shows a large green letter 'M' over a night city scene, with a progress bar at 0:06 / 2:22. To the right of the video player, there is a "URI" field with the value "https://repositor.fhv.uniza.sk/handle/1234/281", a "Collections" field with the value "Audiovizuálna tvorba", and a button labeled "Full item page". Below the video player, there is a "Files" section listing "Lapšanský\_oprava.mp4(300.24 MB)".

### NEW SECURITY FEATURES AND IMPORTANT BUG FIXES

The last, but no less important, section reviews security aspects: fixing known bugs in the systems involved in running DSpace 7 as well as adding some security features.

#### *Apache Log4j/Log4shell Security Vulnerabilities*

DSpace version 7.1.1 fixes a critical vulnerability in the Apache Log4j Library (v2) that allows potential hackers to remotely execute commands and take full control over a system. The Log4j library is used to create logs in many Java frameworks and applications. It is also used within DSpace 7.x. This vulnerability is known as CVE-2021-44228.<sup>75</sup>

According to the National Institute of Standards and Technology (NIST), a part of the US Department of Commerce and one of the oldest physical science laboratories, this vulnerability received the highest possible base score of 10.0 (level: critical). This is therefore an extremely serious problem in terms of systems security.<sup>76</sup> We must emphasize that this is a compelling

reason to update from older minor versions of DSpace 7.x. It is necessary to verify that the server is not running the unpatched version DSpace 7.1 in the backend, only if you have tried or worked with DSpace 7.0. It is equally important that the server is running Apache Solr v8.11.1 (or above) as only this version of Solr is patched for this vulnerability.

Fortunately, as Donohue stated, DSpace 6.x, 5.x, or 4.x (or below) are not vulnerable, as they all use log4j v1 exclusively with a default configuration which is not impacted.<sup>77</sup> Our recommendation is to exclude all DSpace dependencies which relied on log4j v1 as well since log4j v1 is currently at end of life. The ideal solution is a complete migration to DSpace 7.4 or above.

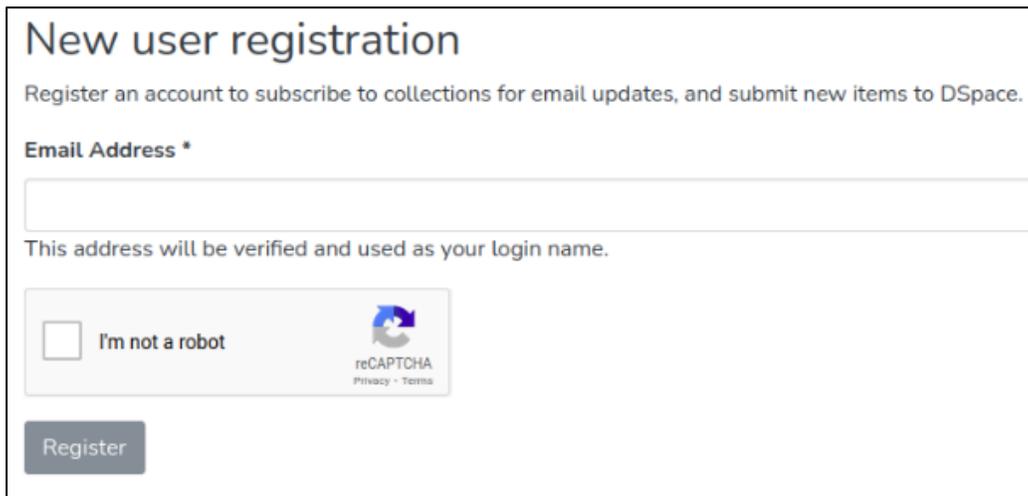
### ***Password Authentication Security Enhancements***

Since DSpace 7.4, all users must provide a current password when changing their password. Password minimum requirements are now configurable, allowing administrators to require users to create more secure passwords.<sup>78</sup>

### ***Support for Google reCaptcha***

As of DSpace 7.4, the new user registration form requires completing the Google reCaptcha feature, as show in figure 9. When activated, this feature provides a robust supplementary safeguard against potential undesired activities, such as large-scale registrations carried out by automated bots utilizing randomly generated or stolen email addresses.<sup>79</sup> This feature confirms that the user is indeed a living being.

**Figure 9.** Google reCAPTCHA integration in DSpace 7 web interface.



The image shows a web form titled "New user registration". Below the title is a subtitle: "Register an account to subscribe to collections for email updates, and submit new items to DSpace." There is an input field for "Email Address \*". Below the input field is a note: "This address will be verified and used as your login name." Below the note is a reCAPTCHA widget with a checkbox labeled "I'm not a robot" and the reCAPTCHA logo. Below the reCAPTCHA widget is a "Register" button.

### ***Other Internal Bug Fixes and Improvements***

It is highly recommended to always upgrade to the latest minor version (at the time of writing this was DSpace 7.6), because the previous versions may contain various vulnerabilities. Important improvements and fixed issues are highlighted:

- DSpace version 7.0 was vulnerable to CVE-2021-41189. A Community or Collection Admin could escalate their permissions to become a full Administrator.
- DSpace 7.2 was vulnerable to CVE-2022-22965. The vulnerability affects Apache Spring Libraries which are used in Java/JDK 9 or above. These libraries are again vulnerable to remote command execution. This vulnerability was scored 9.8 of 10.<sup>80</sup> Donohue recommends upgrading the DSpace backend to version 7.2.1 immediately.<sup>81</sup> Optionally,

upgrade the Apache Tomcat to version 9.0.62, which provides extra guards against this vulnerability.

- Since DSpace backend version 7.2, indexing and file download performance has been improved, which has contributed to an overall increase in the system experience. Numerous other minor bug fixes or accessibility improvements have been implemented.

The following section reviews key points of this process, which we faced during the experimental upgrade. The most significant improvements will be described briefly with details in the references.

## **THE UPGRADE OF DIGITAL REPOSITORY OF MEDIAMATICS AND CULTURAL HERITAGE—THE CASE STUDY**

### ***A Short History of the System Upgrade***

Proof of concept of the Digital Repository of Mediamatics and Cultural Heritage was already tested during 2017, when it was built on the older version of DSpace 6.1. Later, in the first half of 2018, we completed the customization of the XMLUI environment based on user and expert UX/UI testing. The redesigned web interface was completed and based on DSpace 6.2. The system was subsequently launched into production. The repository already contained several dozens of digital objects at that time. It was upgraded to the version 6.3 until the end of 2018. All minor upgrades (6.1, 6.2, 6.3) were successfully carried out using the following procedure framework:

1. A new virtual server with the latest version of the underlying operating system available at the time was installed. We preferred to use Ubuntu Server with long-term support as a good choice.
2. All necessary software prerequisites and dependencies were installed in the new system environment as well as a default installation of DSpace using the appropriate version available at the time according to the instructions in the manual.
3. All system settings, from local.cfg/dspace.cfg and many other configuration files, as well as the entire customized XMLUI web interface were copied to the new system.
4. The digital objects were transferred to the empty repository by simply copying the original assetstore<sup>82</sup> from the previous system to the new. The next necessary step in the migration process of the repository content was still the transfer of the PostgreSQL database and its content. On the old system we performed a so-called database dump used for DSpace; this dump was copied as a single file to the new environment and then restored via the entire restore procedure on the target server.
5. After restarting all system services, the repository started under the same domain name, or IP address, with the same URL and the same digital content as it was in the previous repository.

The above-mentioned procedure has always been our effort to run the system on the latest (updated) software components (such as Apache Tomcat, PostgreSQL, Java JDK, etc.) as well as on the latest operating system—although DSpace, according to the manual, also allows upgrading the version on the remaining system. However, we have preferred a fresh install and then migration of the repository content and settings every time.

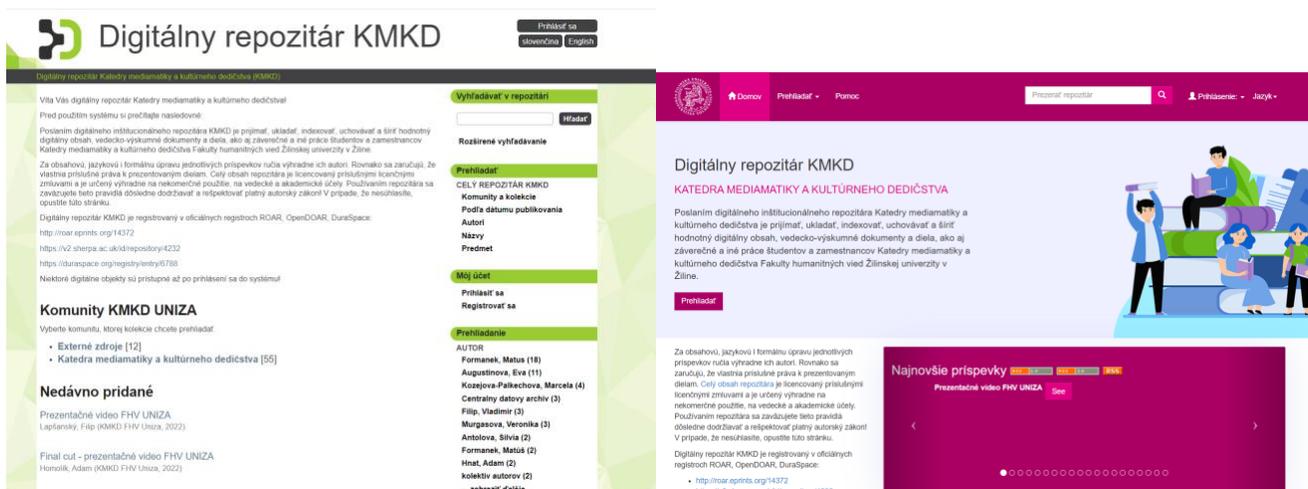
The installation of new versions of DSpace has been carried out without any major difficulties each time using the available manuals and procedures recommended by the experienced professionals, such as Vimal Kumar, known as System Librarian, or Furrukh Hussain Zai. However, the critical point in the mentioned procedure is the migration of content—entire digital objects—between the

two DSpace instances. When it comes to digital content, it's not just about the digital files (e.g., PDFs, etc.) located in the assetstore folder. An essential part of a digital object, as defined by the University of California, is also metadata (structural, descriptive, administrative). These metadata are stored in a SQL database system (e.g., PostgreSQL).

### Need to Upgrade to DSpace 7.4

We have considered DSpace 6.3 to be fully compliant and to meet our requirements for a long period of time. Moreover, in the departmental environment, the web interfaces of the repository (XMLUI as well as JSPUI, as shown in fig. 10) have been optimized several times towards search engines, so a lot of work has been completed (translation to the Slovak language, etc.).

**Figure 10.** Translated XMLUI web interface (on the left) and JSPUI (on the right).



However, with the arrival and dynamic development of the new major version of DSpace (7.x) during 2022, we have started to focus intensively on the idea of a necessary upgrade.

The currently used underlying operating system (Ubuntu Server 20.04.5 LTS) is becoming obsolete and will also need to be replaced soon. After a thorough information research and considering all the benefits, we decided to try to upgrade the repository to DSpace 7.4 in November 2022, using the well-known procedure described above. We installed the latest Ubuntu Server 22.04.1 LTS as well as DSpace 7.4: the default backend (REST API) and the Angular-based frontend. Both components worked correctly in default settings but without any content. The customized settings as well as the whole assetstore were copied/migrated to DSpace 7.4 without any issues. We also tried the procedure described in a related post by Kumar (2022b). However, a critical problem occurred during the database migration.

We assume that the migration from DSpace 6.x to DSpace 7.4 could not be easily performed using the previously described procedure via a dump and restore of the database, since the structure and content of the database are probably not fully compatible between these versions. Or, we have made a mistake of which we were not aware. This issue (incompatibility) resulted in an internal server error message after the database restoration. The migrated digital objects remained inaccessible. Thus, the DSpace 7.4 instance remained unusable because of the database/system error. To resolve this problem and following the recommendations of the manual (Donohue 2022e), we decided to run a migrate command over the imported database as well as the "index-discovery -b" command to upgrade the structure of the DSpace database and subsequently reindex

the entire content of the site, so that the database became compatible with newer versions of DSpace. However, this step did not solve the problem. After further unsuccessful attempts, we decided to look intensively for other ways on how to migrate the digital objects. We found that exporting and then importing the entire site structure of the DSpace repository via Archival Information Packages (AIP) seemed to be potentially suitable and universally applicable. Tim Donohue recommended this as an alternative approach where you can use the AIP export tools to export AIPs from your old site and then import them into the new site.<sup>83</sup> The nature of AIPs and content migration via AIPs will be discussed in a following section.

### ***AIP Backup (Export)***

AIP is defined in the OAIS (Open Archival Information System) reference model as

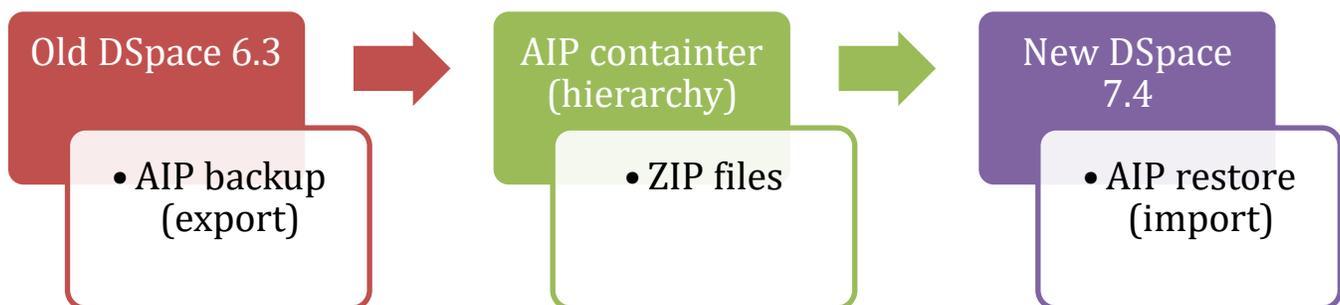
an information package that is used to transmit archival objects into a digital archival system, store the objects within the system, and transmit objects from the system. An AIP contains both metadata that describes the structure and content of an archived essence and the actual essence itself. It consists of multiple data files that hold either a logically or physically packaged entity.<sup>84</sup>

According to this definition, AIP is used to archive or export digital objects within the DSpace system. The ability to work with AIP packages has been available in DSpace since its inception (version 1.0), so this is not a new feature. Therefore, we will utilize it in an appropriate way on an appropriate occasion.

In the context of the DSpace system, AIP may include a single item, collection, community, or an entire site (site AIPs contain sitewide information). The files with content (such as PDFs), if they exist, are always part of an item's AIP. So, each AIP is logically self-contained. This fact implies that it is possible to restore it without the rest of the archive. We can export or restore a single item, collection, or community. AIPs containing a larger logical object (site, collection, or community) do not include any physical child objects directly in itself (e.g., all items in collection or community) but contain only references (links) to all "child objects" in the metadata structure. If we want to successfully restore a collection, we need to do it only with the corresponding items' AIPs.

We propose to use the above-mentioned principles in this paper and export the entire site from the original system (DSpace 6.3) and transfer this object (AIP packages hierarchy) to the new environment (DSpace 7.4) via the AIP container(s), as shown in figure 11.

**Figure 11.** AIP import and export process.



Each AIP is represented fundamentally as a ZIP file containing metadata (expressed via METS manifest) and, in the case of an item, all related digital content (bitstreams, licenses, and any other associated computer files) too.

When exporting an AIP, only the latest version of the items is taken from the repository, which is considered as a slight disadvantage of using AIP for export. Thus, past versions of an item are always lost. Unfinished submissions are not described in the AIP either.

The export command executed over the site root will also invoke the export of all integrated communities and collections. This gives a complete hierarchical export of the entire content of the digital repository including the access permissions to certain collections/communities, or information about the system users (called EPersons with their DSpace logins and passwords) and user groups. The single item export does not contain any mentioned administrative metadata.

In the original system (DSpace 6.3), we initiated the recursive command to export the entire hierarchical tree structure of the entire site, starting from its root defined in the command itself:

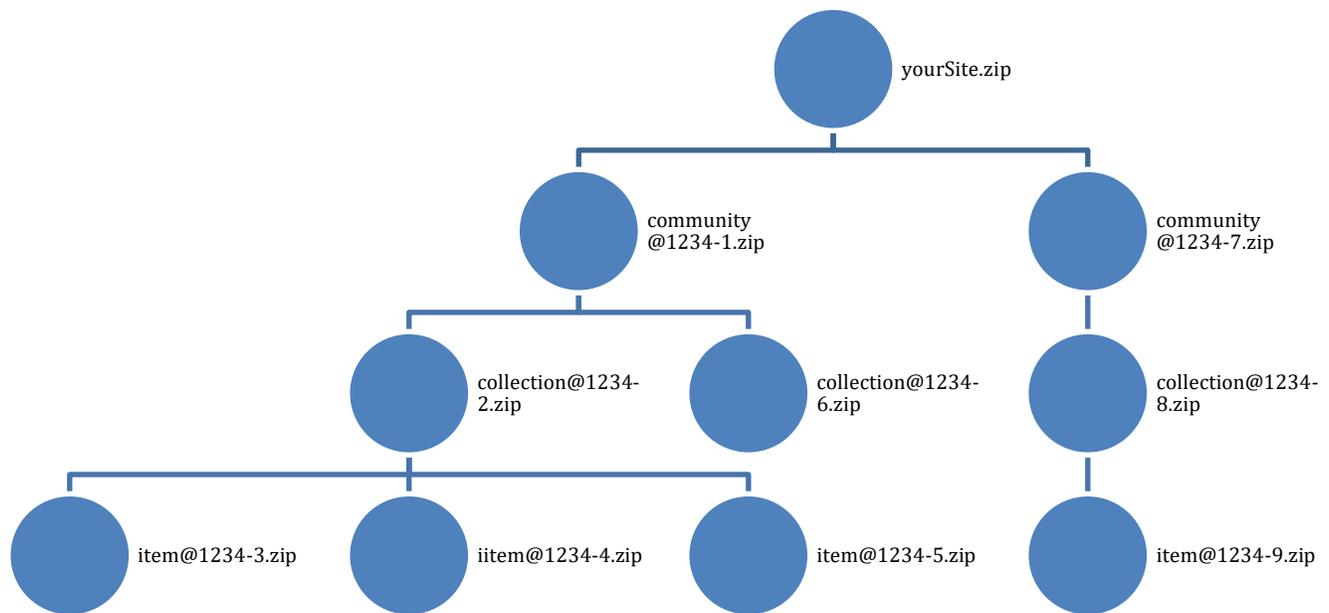
```
sudo /dspace/bin/dspace packager -d -a -t AIP -e
  spravca.repozitara@fhv.uniza.sk -i 1234/0 /tmp/kmkd/kmkd.zip
```

where:

- The prefix “**sudo**” escalates the privileges when running the command to the administrator level.
- The string “**/dspace/bin/dspace**” represents the absolute path to the executable file located in the specific “DSpace-home” directory.
- The parameter “**packager**” triggers export or import of packages.
- The “**-d**” switch modifies the packager settings to default export mode.
- The “**-a**” switch extends the export to all child objects in the structure. This is an important setting if we want to export an entire site, collection, or community.
- The “**-t AIP**” switch defines the type of package to be used in the export. We need the AIP type.
- The “**-e spravca.repozitara@fhv.uniza.sk**” switch defines the specific user (DSpace administrator) who has the privileges to run the export.
- The “**-i 1234/0**” switch defines the top-level root entity from which the export starts. We want to export the entire site, so the root will always be in the form “your-handle-prefix/0”. In our case, we use “1234” as a handle prefix.
- The string “**/tmp/kmkd/kmkd.zip**” represents the absolute path to the folder with the root ZIP file of the entire site (in our case). This ZIP file does not exist before the export. It is created when the command was performed.

After the command is executed, according to the above-mentioned structure, the export process starts. Depending on the number and size of (AIP) packages, this action takes an estimation of minutes to tens of minutes. Consequently, a hierarchical structure of internally related objects, i.e., ZIP files (AIP archives), is created. As a result, the separate ZIP files are created side by side and arranged (e.g., by name) in the server’s file system (e.g., in the folder /tmp/kmkd/) after export. The relationships between the AIP archives are stored at the metadata level in the associated XML file that is part of each ZIP archive.

The structure of AIP/ZIP packages is shown in figure 12.

**Figure 12.** AIP hierarchy example.

The root package called “yourSite.zip” (“kmkd.zip” in our case) does not physically contain all other ZIP files of the entire site inside. It only references the AIP packages of communities that are located one level lower in the hierarchy (see fig. 12) through XML metadata records. Similarly, each community’s package refers to the package (or packages) of the collections that are placed in the hierarchy below it. The physical content files are only part of the item packages which also take up the most disk space. In our case, the export of all objects—284 ZIP files were created—required a disk space of about 13 GB and took approximately 10 minutes.

### ***AIP Backup (Import)***

All ZIP files created in the /tmp/kmkd folder were then copied to a new virtual server based on Ubuntu 22.04 LTS, which already had been running a default installation of DSpace 7.4. The new DSpace instance already had a precreated administrator (i.e., EPerson), under whose account we could start importing the transferred objects.

For importing AIP packages, we recommend using a command similar to the following:

```
sudo /dSPACE/bin/dSPACE packager -r -a -f -t AIP -e
spravca.repozitara@fhv.uniza.sk -i 1234/0 -o skipIfParentMissing=true
/home/dkar/kmkd/kmkd.zip
```

where:

- “**sudo /dSPACE/bin/dSPACE packager**” has already been explained when exporting.
- The “**-r**” switch initiates the restore process.

- The “-f” switch forces so-called force restore. Any conflicts with potentially existing data in the target site (if it previously existed) are resolved by overwriting.
- The “-t AIP” switch defines the package type as AIP.
- The “-e spravca.repozitara@fhv.uniza.sk” switch defines the administrator account that has the permission to perform the import.
- The “-i 1234/0” switch says that we want to import the entire site (therefore that 0) characterized by a particular handle-prefix. We used an unregistered handle prefix “1234” in our case.
- The “-o skipIfParentMissing=true” switch defines additional import options, telling the packager not to deal with the parent object relationship. Since we are already starting the import process from the site’s root (and the root object has no parent), this switch is mainly used when importing the entire site in one step.

The above-mentioned command starts the import process over the site root. A hierarchical restore of the entire site will be performed by linking the AIP packages at the metadata level. Depending on the site size, this process takes tens of minutes to hours of time. In our case, the import process took about 25 minutes. Thus, the import process is more time-consuming than the export.

When the import process is completed, we highly recommend performing a comprehensive reindexing of the entire repository content. This crucial step is essential for ensuring the accurate visibility of all inserted metadata in the new system. Reindexing plays a vital role in maintaining data accuracy, optimizing search engine performance, and ensuring the overall functionality of the system. By accurately reflecting all metadata and content in the new system, reindexing guarantees data consistency and integrity. Additionally, it ensures that all data fields are appropriately indexed and searchable, effectively preventing any discrepancies between the migrated data and the search index.

Reindex the DSpace repository content by executing the following command:

```
sudo /dspace/bin/dspace index-discovery
```

Consequently, all digital objects should be accessible through the frontend (web) interface.

When using an unregistered handle prefix (in our case only the string “1234”), based on our own experience, we must emphasize that the new DSpace 7 configuration file (local.cfg) must already contain and define the same handle prefix. It must be identical with the prefix from the original system because the exported AIP packages also contain the handle prefix from the original system. This prefix must be recognized by the new system. Otherwise, the entire import will fail due to the inability of the system to translate the unknown handle prefix.

## **SHORT DISCUSSION ON DSPACE SYSTEM MIGRATION**

The default web interface (frontend) of our testing instance of DSpace 7.4 is shown in figure 13. It also contains items (recent submissions), communities, and collections that we have imported through the AIP restore process.

**Figure 13.** Screenshot of a testing instance in DSpace 7.4.<sup>85</sup>

**DSpace 7**

DSpace is the world leading open source repository platform that enables organisations to:

- easily ingest documents, audio, video, datasets and their corresponding Dublin Core metadata
- open up this content to local and global audiences, thanks to the OAI-PMH interface and Google Scholar optimizations
- issue permanent urls and trustworthy identifiers, including optional integrations with handle.net and DataCite DOI

Join an international community of [leading institutions using DSpace](#).

The test user accounts below have their password set to the name of this software in lowercase.

- Demo Site Administrator = dspacedemo+admin@gmail.com
- Demo Community Administrator = dspacedemo+commadmin@gmail.com
- Demo Collection Administrator = dspacedemo+colladmin@gmail.com
- Demo Submitter = dspacedemo+submit@gmail.com

Search the repository ... Search

### Communities in DSpace

Select a community to browse its collections.

Now showing 1 - 2 of 2

Externé zdroje  
Voľne prístupné externé zdroje

Katedra mediamatiky a kultúrneho dedičstva

### Recent Submissions

No Thumbnail Available

**Item**  
Prezentačné video FHV UNIZA  
(KMKD FHV Uniza, 2022) Lapšanský, Filip

**Item**  
Final cut - prezentačné video FHV UNIZA

The complete migration from an older DSpace system to a newer or latest version is much more complex and includes far more steps than is possible to describe in this article. In addition to the transfer of the digital objects themselves from one environment to another, system administrators will have to deal with the customization and overall optimization of the web graphical user interface as well as with the start-up and detailed setup of all other DSpace components and services, in particular, enabling HTTPS in the production environment, setting up OAI-PMH, and integrating ORCID, which are not covered here due to the scope of the work.<sup>86</sup> All these important steps must be kept in mind and considered during migration because their implementation is quite time-consuming.

However, we still consider the transfer, i.e., the export and subsequent import of digital objects, to be the key moment of the whole migration. A digital repository without any content (even in the latest version and with advanced features enabled) is a worthless system.

## CONCLUSION

Taken together, we have made our own well-reasoned prediction that thousands of DSpace systems around the world are likely to soon face the necessity of migrating to the newer version of DSpace 7. This is a logical, but technically nontrivial, step that may confuse some system administrators, especially if an unexpected bug or state is detected and the administrator is unable to fix it themselves. This paper provides helpful tips to institutions deciding on performing a critical system upgrade and further describes key points and reasons for the migration or upgrade. We have described the most important benefits and functionalities of the new DSpace 7.4 version as well as briefly dealt with a case study, where we have given our proposals regarding the transfer of digital objects between two systems with different versions in an alternative, but more universal way, using a practical example.

DSpace 5 has been end-of-life since January 2023. The support for DSpace 6 officially ended in the summer of 2023 as well. We urge system administrators to consider the importance of early upgrading or migrating the systems which they are responsible for due to this end in support for previous versions. Administrators must consider these end-of-life statements and the question becomes: What happens if the repositories running DSpace 6.x will not be updated by mid-2023? We do not expect a massive failure of open DSpace-based repositories, but all systems built on outdated, obsolete, or unpatched software are at a significantly higher risk for potential exploitation of security vulnerabilities by attackers in the online environment.

For further detailed research, it would be appreciated if the metadata fields in authoritative registry records like OpenDOAR and others could be extended with a field tracking the specific version of the system used. We also suggest that authoritative registries send periodic communication (e.g., annually) to repository administrators as a prompt or reminder to update the relevant entries they contain. This would prevent outdated entries, links, and other discrepancies that may occur in registers.

The upgrading to DSpace 7 provides an undisputable number of new advanced and sophisticated features that simplify the overall work with digital entities. In general, DSpace 7 represents a significant step forward in the development of software for open digital libraries and repositories on worldwide. It may be considered as a revolution, not only as an evolution in this field. As the most important changes are highlighted and briefly discussed in our study, we recommend trying the latest version of DSpace 7, getting to know its new environment, and then developing your own upgrade plan.

## ENDNOTES

- <sup>1</sup> Kabir Khan and Arslan Sheikh, "Open Source Software Adoption for Development of Institutional Repositories in University Libraries of Islamabad," *Information Discovery and Delivery* 51, no. 1 (January 2023): 47–55, <https://doi.org/10.1108/IDD-10-2021-0113>.
- <sup>2</sup> Patrick Mapulanga, "Digitising Library Resources and Building Digital Repositories in the University of Malawi Libraries," *The Electronic Library* 31, no. 5 (September 2013), 635–47, <https://doi.org/10.1108/EL-02-2012-0019>.
- <sup>3</sup> Marios Zervas et al., "Next Generation Institutional Repositories: The Case of the CUT Institutional Repository KTISIS," *Procedia Computer Science* 146 (2019): 84–93.

- <sup>4</sup> Nushrat Khan, Mike Thelwall, and Kayvan Kousha, "Are Data Repositories Fettered? A Survey of Current Practices, Challenges and Future Technologies," *Online Information Review* 46, no. 3 (August 2021): 483–502, <https://doi.org/10.1108/OIR-04-2021-0204>.
- <sup>5</sup> Mapulanga, "Digitising Library Resources."
- <sup>6</sup> MacKenzie Smith et al., "DSpace: An Open Source Dynamic Digital Repository," *D-Lib Magazine* 9 no. 1 (January 2003), <http://www.mybestdocs.com/smith-m-et-al-DSpace.htm>.
- <sup>7</sup> "Research: Technologies and Services to Ensure Sustained Access," Lyrasis, accessed January 11, 2023, <https://www.lyrasis.org/membership/Pages/Research.aspx>.
- <sup>8</sup> Tim Donohue, "DSpaceResources," last modified November 20, 2020, <https://wiki.lyrasis.org/display/DSPACE/DSpaceResources#DSpaceResources-DSpaceSystemDocumentation>.
- <sup>9</sup> "Organizational Home," Lyrasis, accessed January 25, 2023, <https://dspace.lyrasis.org/organizational-home/>.
- <sup>10</sup> "OpenDOAR Statistics," JISC, accessed January 10, 2023, [https://v2.sherpa.ac.uk/view/repository\\_visualisations/1.html](https://v2.sherpa.ac.uk/view/repository_visualisations/1.html).
- <sup>11</sup> "Registry of Open Access Repositories," eprints, accessed February 1, 2023, [http://roar.eprints.org/cgi/roar\\_search/advanced?location\\_country=&software=DSpace&type=&order=-recordcount%2F-date](http://roar.eprints.org/cgi/roar_search/advanced?location_country=&software=DSpace&type=&order=-recordcount%2F-date); "Registry of Research Data Repositories," Re3data, accessed January 15, 2023, <https://doi.org/10.17616/R3D>.
- <sup>12</sup> "DuraSpace Registry," Lyrasis, accessed January 11, 2023, [https://registry.lyrasis.org/?gv\\_search&filter\\_10=DSpace&filter\\_4\\_6&filter\\_3&filter\\_20&filter\\_28&mode=all](https://registry.lyrasis.org/?gv_search&filter_10=DSpace&filter_4_6&filter_3&filter_20&filter_28&mode=all).
- <sup>13</sup> Khan, Thelwall and Kousha, "Are Data Repositories Fettered?"; George Pyrounakis, Mara Nikolaidou, and Michael Hatzopoulos, "Building Digital Collections Using Open Source Digital Repository Software: A Comparative Study," *International Journal of Digital Library Systems* 4, no. 1 (2014): 10–24, <https://doi.org/10.4018/ijdl.2014010102>.
- <sup>14</sup> Dragan Ivanović et al., "FAIRness of Repositories & Their Data: A Report from LIBER's Research Data Management Working Group" (2019), <https://doi.org/10.5281/zenodo.3251593>.
- <sup>15</sup> Khan and Sheikh, "Open Source Software Adoption"; Nur Ahammad, "Open Source Digital Library on Open Educational Resources," *The Electronic Library* 37, no. 6 (2019): 1022–39, <https://doi.org/10.1108/EL-11-2018-0225>.
- <sup>16</sup> Rushmanasab Gurikar and Gururaj S. Hadagali, "Use of Open Source Software in Indian Institutional Digital Repositories: A Study," *Library Philosophy and Practice* (2021): 4608, <https://digitalcommons.unl.edu/libphilprac/4608>.
- <sup>17</sup> Tim Donohue, "DSpace Releases," last modified July 5, 2023, <https://wiki.lyrasis.org/display/DSPACE/Releases>.

- <sup>18</sup> Tim Donohue, "Support for DSpace 5 and 6 is ending in 2023," last modified March 24, 2023, <https://wiki.lyrasis.org/display/DSPACE/Support+for+DSpace+5+and+6+is+ending+in+2023>.
- <sup>19</sup> Zervas et al., "Next Generation Institutional Repositories."
- <sup>20</sup> Nason B. Bimbe et al., "Challenges in Reinvigorating and Upgrading DSpace-Based Institutional Repositories: A University of Zambia (UNZA) Library Case Study," (IDS Working Paper 483, March 2017), <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/12858>.
- <sup>21</sup> Teblos Casados et al., "A Quest to Upgrade from a Legacy to a Modern Open Source Repository," (poster presented at 2019 Texas Conference on Digital Libraries, May 22, 2019), <https://hdl.handle.net/2249.1/156391>.
- <sup>22</sup> Yogesh Kawal and Kumari Vandana, "Installation and Customization of DSpace: An Institutional Repository," (paper, International Virtual Conference on Contemporary Problems and Prospects of Libraries and Information Centres, Dindigul, Tamilnadu, November 2021), [https://www.researchgate.net/publication/360962968\\_Installation\\_and\\_Customization\\_of\\_DSpace\\_An\\_Institutional\\_Repository/citations](https://www.researchgate.net/publication/360962968_Installation_and_Customization_of_DSpace_An_Institutional_Repository/citations).
- <sup>23</sup> Zervas et al., "Next Generation Institutional Repositories."
- <sup>24</sup> Hope Craft, "Embracing Open: Institutional Repository Migrates to DSpace 7," Iowa State University of Science and Technology, accessed January 11, 2023, <https://www.lib.iastate.edu/news/embracing-open-institutional-repository-migrates-DSpace-7>.
- <sup>25</sup> "DuraSpace Registry," Lyrasis.
- <sup>26</sup> Kristi Searle, "The DSpace 7 Project—A Simple Summary," Lyrasis (April 2017), <https://dspace.lyrasis.org/the-dspace-7-project-a-simple-summary/>.
- <sup>27</sup> "DSpace 7 Press Release," Lyrasis (2021), <https://dspace.lyrasis.org/dspace-7-press-release/>.
- <sup>28</sup> "AVAILABLE NOW: DSpace 7 Preview Release," Atmire (May 2019), <http://www.atmire.com/articles/detail/available-now-DSpace-7-preview-release>.
- <sup>29</sup> "DSpace 7 Press Release," Lyrasis.
- <sup>30</sup> "DSpace 7 Press Release," Lyrasis.
- <sup>31</sup> Tim Donohue, "Release Notes," last modified Jun 29, 2023, <https://wiki.lyrasis.org/display/DSDOC7x/Release+Notes>; "AVAILABLE NOW: DSpace 7 Preview Release," Atmire.
- <sup>32</sup> "DSpace 7.4 Upgrading to DSpace 7.4," DSpace videos (2022), <https://www.youtube.com/watch?v=2ECza9Zd5PM>; "Connect DSpace 7.4 Frontend with Backend," RAFLIMTS (January 2023), <https://www.youtube.com/watch?v=gCYHhD2Bf5Y&list=PLjVaNnuLQgZzt4iGYj03lrcl-3bSmFvUu&index=2>; "DSpace 7.4 Backend on Ubuntu 22.04 with Installation Guide," RAFLIMTS (December 2022), <https://www.youtube.com/watch?v=N0wfsV4v1WM>; "DSpace 7.4 Frontend on Ubuntu 22.04 with Installation Guide," RAFLIMTS (December 2022),

<https://www.youtube.com/watch?v=fh4KXck4TWo>; “Video Recording from the Webinar: DSpace 7 and DSpace 7 CRIS: How to Bring Your Institutional Repository to the Next Level?,” PCG Academia (July 2022), <https://pcgacademia.pl/en/news/video-recording-from-the-webinar-DSpace-7-and-DSpace-7-cris-how-to-bring-your-institutional-repository-to-the-next-level/>.

- <sup>33</sup> Ivan Masár, “DSpace Release 7.0 Status” (September 2022), <https://wiki.lyrasis.org/display/DSPACE/DSpace+Release+7.0+Status#DSpaceRelease7.0Status-Whatfeaturesarecominginalater7.xrelease?7.x>.
- <sup>34</sup> Masár, “DSpace Release 7.0 Status.”
- <sup>35</sup> “DSpace 7.4 Upgrading to DSpace 7.4,” DSpace videos.
- <sup>36</sup> “Connect DSpace 7.4 Frontend with Backend,” RAFLIMTS.
- <sup>37</sup> Carol Minton Morris, “DSpace 7 UI Project Plain Language Summary” (July 2017), <https://wiki.lyrasis.org/display/DSPACE/DSpace+7+UI+Project+Plain+Language+Summary>.
- <sup>38</sup> Donohue, “Release Notes.”
- <sup>39</sup> Morris, “DSpace 7 UI Project Plain Language Summary.”
- <sup>40</sup> Donohue, “Release Notes.”
- <sup>41</sup> Morris, “DSpace 7 UI Project Plain Language Summary.”
- <sup>42</sup> Donohue, “Release Notes.”
- <sup>43</sup> Donohue, “Release Notes.”
- <sup>44</sup> Morris, “DSpace 7 UI Project Plain Language Summary.”
- <sup>45</sup> Morris, “DSpace 7 UI Project Plain Language Summary.”
- <sup>46</sup> “DSpace 7.4 Upgrading to DSpace 7.4,” DSpace videos.
- <sup>47</sup> Donohue, “Release Notes.”
- <sup>48</sup> Donohue, “Release Notes.”
- <sup>49</sup> Donohue, “Release Notes.”
- <sup>50</sup> Donohue, “Release Notes.”
- <sup>51</sup> Donohue, “Release Notes.”
- <sup>52</sup> Donohue, “Release Notes.”
- <sup>53</sup> Tim Donohue, “Try out DSpace 7,” last modified April 2023, <https://wiki.lyrasis.org/display/DSPACE/Try+out+DSpace+7>.

- <sup>54</sup> “Install DSpace 7.2 on Ubuntu 22.04,” System Librarian, accessed November 28, 2022, <https://hyperlink.co.ke/2022/06/17/install-DSpace-7-2-on-ubuntu-22-04/>; Vimal Kumar, *DSpace Geek* (blog), 2022, <http://DSpacegeek.blogspot.com/>; Jeremiah Kellogg, “Install DSpace 7 Backend and Frontend on a VirtualBox Debian 11 Server,” February 2022, <https://www.youtube.com/watch?v=Ta-BpqBFJsg>; Furrukh Hussain Zai, “DSpace,” accessed October 28, 2022, <https://www.youtube.com/playlist?list=PLJVaNnuLQgZxbJKXbBH3kvTGK6SwveaxB>.
- <sup>55</sup> “DSpace 7.4 Backend on Ubuntu 22.04 with Installation Guide,” RAFLIMTS; “DSpace 7.4 Frontend on Ubuntu 22.04 with Installation Guide,” RAFLIMTS.
- <sup>56</sup> Kumar, *DSpace Geek*.
- <sup>57</sup> Kellogg, “Install DSpace 7 Backend and Frontend on a VirtualBox Debian 11 Server.”
- <sup>58</sup> Donohue, “Release Notes.”
- <sup>59</sup> “DSpace 7.4 Backend on Ubuntu 22.04 with Installation Guide,” RAFLIMTS.
- <sup>60</sup> Donohue, “Release Notes.”
- <sup>61</sup> Kumar, *DSpace Geek*; Vimal Kumar, “Upgrading DSpace 6 to 7” (blog post), December 2022, <http://dspacegeek.blogspot.com/2022/12/upgrading-dspace-6-to-7.html>; “Install DSpace 7.2 on Ubuntu 22.04,” System Librarian.
- <sup>62</sup> Zervas et al., “Next Generation Institutional Repositories.”
- <sup>63</sup> “DSpace 7.4 Upgrading to DSpace 7.4,” DSpace videos.
- <sup>64</sup> Masár, “DSpace Release 7.0 Status.”
- <sup>65</sup> Donohue, “Release Notes.”
- <sup>66</sup> Donohue, “Release Notes.”
- <sup>67</sup> Donohue, “Release Notes.”
- <sup>68</sup> Masár, “DSpace Release 7.0 Status.”
- <sup>69</sup> Donohue, “Release Notes.”
- <sup>70</sup> Donohue, “Release Notes.”
- <sup>71</sup> Donohue, “Release Notes.”
- <sup>72</sup> Donohue, “Release Notes.”
- <sup>73</sup> Donohue, “Release Notes.”
- <sup>74</sup> Tamás Kiss and Nagy Zoltán Kanász, “Rekordoldali médiamegjelenítés a DSpace 7-Ben,” *Digitális Bölcsészet* 17, no. 3 (December 2020), <https://doi.org/10.31400/dh-hun.2020.3.1475>.

- <sup>75</sup> This vulnerability is described in more detail at <https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2021-44228> and <https://logging.apache.org/log4j/2.x/security.html#Fixed in Log4j 2.15.0>.
- <sup>76</sup> “CVE-2021-44228 Detail,” NIST, accessed December 19, 2022, <https://nvd.nist.gov/vuln/detail/CVE-2021-44228>.
- <sup>77</sup> Tim Donohue, “DSpace & log4j Critical Vulnerabilities (CVE-2021-44228 and CVE-2019-17571),” accessed November 29, 2022, <https://groups.google.com/g/DSpace-tech/c/QR59bS4nIT0>.
- <sup>78</sup> Donohue, “Release Notes.”
- <sup>79</sup> Donohue, “Release Notes.”
- <sup>80</sup> “CVE-2022-22965 Detail,” NIST, accessed December 19, 2022, <https://nvd.nist.gov/vuln/detail/cve-2022-22965>.
- <sup>81</sup> Donohue, “Release Notes.”
- <sup>82</sup> The assetstore folder is a part of the DSpace storage layer—a specific folder on the filesystem that contains the uploaded, ingested, or generated files (documents, images, audio, video, datasets, etc.) according to Tim Donohue, “Storage Layer,” last modified January 4, 2019, <https://wiki.lyrasis.org/display/DSDOC6x/Storage+Layer>.
- <sup>83</sup> Tim Donohue, “Migrating DSpace to a New Server,” accessed December 16, 2022, <https://wiki.lyrasis.org/display/DSDOC7x/Migrating+DSpace+to+a+new+server>.
- <sup>84</sup> “Archival Information Package (AIP),” IASA, accessed January 25, 2023, <https://www.iasa-web.org/tc04/archival-information-package-aip>.
- <sup>85</sup> Available online at: <https://repozitar.fhv.uniza.sk>.
- <sup>86</sup> “DSpace 7.4 Upgrading to DSpace 7.4,” DSpace videos.