# Online Library Reference Tools:
# A Proportional Learning of Open Research Data Repositories

## Dr. Bhalachandra S. Deshpande

Assistant Professor, Dept. Library and Information Science, Bangalore North University,

Kolar - 563103

balusnd@gmail.com

## ABSTRACT

*Open repositories for Research Data Management (RDM) are a project to organise, store, cite, preserve, and distribute the data generated during research. There are several interdisciplinary and subject-specific open repositories for RDM that provide excellent capabilities for long-term data management. The goal of this research is to compare and contrast the features of three major open data repositories: Zenodo, FigShare, and Mendeley Data.*

**KEYWORDS:** Open Access, Research Data Management, Open Repositories

## 1. INTRODUCTION

Learned societies provide equal opportunities to all the citizens without any kind of bias. In prospective environment like this, people enthusiastically engage in learning, responsibly make use of resources around them and play a vital role in nation building. The first and foremost need here is to make learning resources like journals easily available. Providing open access to resources is the on- point solution to this scenario. Open Access is a boon to the scholars when journal list prices have been rising faster than inflation [1]. Especially, the scientific resources are to be made available freely to the research community. With this objective, a number of open-access journals have increased during the recent years. While, the funding agencies have also strongly voiced to provide free access to research which are based on public funds. Currently, the open repositories for Research Data Management (RDM) are one such initiative to organize, store, cite, preserve, and share the collected data derived from the research. There are many multidisciplinary and subject specific open repositories for RDM offering exquisite features to manage research data for long term. The current article aims to presents a comparative study of Open Data Repositories- Zenodo, FigShare and Mendeley Data.

## 2. OPEN ACCESS CONCEPTION

"Open Access" as defined by the Budapest Open Access Initiative (BOAI), is making information resources open on the public Internet to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the Internet itself. [1]

A study by Antelman, K. (2006) indicated that freely available articles do have a greater research impact [2]. Metrics indicate that, as of December 2019, **14,119** open access journals and **4,516,860** Articles from **130** Countries are listed in the Directory of Open Access Journals. While, currently 4,725 institutional and cross-institutional repositories have been registered in ROAR. FAIR stands for 'Findable, Accessible, Interoperable and Reuseable', which make the term 'open access' easier to discuss [3]. The interlibrary loan depends on the loaning library in terms of days or weeks to provide the requested article/information. But think of an Open access to the same article available online will be very much fast-paced service for a researcher. The Figure 2 is a direct answer to why we need to support open access.



Credit: Australian Open Access Support Group

**Figure:   Benefits of Open Access**

## 3. RESEARCH DATA MANAGEMENT

The intended audience of research publishing is generally other researchers. Open access benefits researchers as readers by providing access to papers that their libraries do not purchase. Researchers are the direct benefactors from open access, as no library can subscribe to all journals– this is termed as the "serials crisis" [4].

As proposed by open access advocates, open access promotes research progress, efficiency, and transfer of information [5]. Faster discoveries benefit every researcher, not only those whose library can afford to subscribe to the particular journal.

In the view of above scenario, sharing research and its data will promote R&D activities around the world. So Research Data Management (RDM) aims to organize, store, cite, preserve, and share the collected data derived from the research.

Sharing research data can help to promote your research, increase citation rate and raise researcher. By providing research data, the possibilities that open are:

- published work can be verified by peers
- share private links with colleagues and reviewers.

- control access to data
- Showcase the data with slides and posters
- backup data and metadata using steadfast cloud based infrastructure
-  Securely store research data for long term

RDM platforms help scientists get credit for making their research available. There are many other benefits when research data is made freely available. It also increases transparency and confidence in research findings. Furthermore, data can be reused and speed up the discovery.

## 3.1 Figshare

Figshare is a web-based interface which is designed for scholarly research data management and dissemination. Figshare supports storing, sharing, discovering research and receive citations for all the research outputs with over 5000 citations of Figshare content till date.

Figshare originally was created as a solution to store research outputs in one single place, simultaneously allowing it to be discovered by the academic community. Figshare allows academics upload, cite, share and discover all sorts of research outputs with the secure hosting options and long term preservation of data.

## 3.2 Zenodo

Zenodo is derived from- *Zenodotus*, father of the first recorded use of metadata and the first librarian of the Ancient Library of Alexandria. Zenodo supports sharing, cititng, storing and discovering research for all the research outputs. Zenodo hosted by CERN and is funded by EU.

## 3.3 Mendeley Data

Mendeley Data is an open research data repository, where scholars can upload and share their data. Datasets can be shared privately among peers, and also published to share with academic community around the world. Sharing data is significant for science, since it facilitate data reuse and encourage reproducibility of research. Also it helps gaining exposure for research outputs, as Mendeley Data provides DOI for every dataset and can be cited.

## 4. FINDINGS

| Features | zenodo | figshare | MENDELEY DATA Put your research data online today |
|---|---|---|---|
| **Upload limit** | 50 GB per dataset | 5 GB, 20 GB of free private space | 10 GB per dataset. |
| **Provide DOI** | Yes | Yes | Yes |
| **File Types** | Publications, posters, presentations, datasets, images, video/audio files, software and lessons. | Any file format | Images (.tiff/tif), JPG, PNG; Video; Sound; Word (.docx); Selected programming languages and scripts (e.g.: .json, .java); PDF; Text; CSV; Excel files and 3D models (e.g.: .cif, .mol, .mol2, |

| | | | |
|---|---|---|---|
| | | | .nii) |
| **Versions** | • Supports versioning on the dataset level.<br>• Once the record has been published, you can no longer change the files in the record, nonetheless, new version can be created (DOI versioning) | • Supports version control of all publicly available data.<br>• Any privately stored data can also be altered or deleted as you wish. | Supports versioning of the dataset |
| **Software** | Zenodo is run with Invenio (an open source software framework), wrapped by a small extra layer of code that is also called *Zenodo*. | No Info | No Info |
| **Citation Support** | Zenodo integrates with GitHub to make software citeable. | No Info | Data will include a Force11 compliant citation so that other researchers can effortlessly cite the research. |
| **Servers Support** | • Zenodo servers are managed via OpenStack and Puppet configuration management system.<br>• Zenodo frontend servers are responsible for running the Invenio repository platform application which is based on Python and the Flask web development framework. | Figshare is hosted on Amazon Web Services to ensure the highest level of security and stability for the research data. | • Data is stored on Amazon S3 (a storage and hosting service) servers, which assures the integrity and security of data.<br>• In addition, published datasets are archived with Data Archiving and Network Services(DANS), to preserve your data over the long term. |
| **Metadata & Search** | • Metadata is indexed in an Elasticsearch cluster for fast and powerful searching.<br>• Metadata is stored in JSON format in | •Figshare supports OAI-PMH.<br>•Figshare items with the type dataset are included in Google Dataset Search results. This is achieved by including schema.org JSON- | Published dataset metadata is aggregated to DataCite's metadata index (a comprehensive research datasets metadata index) and to the OpenAIRE portal. |

| | PostgreSQL | LD markup in the public item's metadata HTML pages. | |
|---|---|---|---|
| **Reporting** | After uploading data, Zenodo will take care of the reporting. | Supports reporting | No Info |
| **Licenses** | Allow uploading under a variety of different licenses and access levels | By making data public, it is under the Creative Commons 4.0 licences | Can choose a licence to publish it under, from a range of Creative Commons and open software and hardware licences. |
| **Support for Publishers** | No Info | Provide cloud solutions for publishers | Provides a platform to showcase your authors' and journals' datasets, making them discoverable alongside datasets from over 200 journals. |
| **API** | REST API | open API | open RESTful API |
| **Admin Rights** | Provide communities to build a hub of curated information with a group of users | • Provide control access to private files and folders with trusted colleagues. <br> • Provide Private link sharing | Supports sharing unpublished data with the colleagues and funding bodies. |
| **Certification** | No Info | No Info | Has received the industry-recognised CoreTrustSeal certification. |
| **Institution Support** | No Info | Helps academic institutions store, share and manage all of their research outputs. | Offers modular research data management and collaboration solutions for the Institutions with a range of institutional packages that can be tailored to best suit individual research data requirements. |
| **Delete or Edit** | No Info | Supports version control of all publicly available data. Any privately stored data can also be altered or deleted. | Draft datasets can be deleted within the web interface or API, but published datasets can be deleted by contacting Mendeley Data. |
| **Peer Review** | Promote peer-reviewed openly accessible research, and curate the uploads posted on the front-page | No Info | Datasets posted to Mendeley Data are currently moderated. |

## CONCLUSION

Research data management is a necessary academic activity for long-term data storage and sharing. As a result, RDM must be regarded a significant aspect of scholarly publication. Researchers may use open research data repositories to help them manage their research data more efficiently. These repositories make data freely available to the global academic community for use in their research projects. At this point, an individual researcher must be aware of the different open data repositories that are available.

The current study evaluated the properties of three major open research data repositories, namely Zenodo, Figshare, and Mendeley Data, in this regard. The importance of open access and research data management has been highlighted in this article. There has also been an attempt to give broad information about the repositories under investigation. The study's findings lead to the conclusion that Zenodo has the highest data upload limit of the three repositories. All three repositories include the key elements required by a researcher, such as DOI, File Types, citation support, licencing, and search (metadata harvesting). Zenodo was the only company that mentioned the reporting feature. Whereas, peer-review is conducted in Zenodo and Mendeley Data.

## REFERENCES

[1] Van Noorden, R. Open access: The true cost of science publishing. *Nature* 495**,** 426–429 (2013) doi:10.1038/495426a

[2] ANTELMAN, K. (2004). Do open-access articles have a greater research impact?. *College & research libraries*, *65*(5), 372-382.

[3] Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; et al. (15 March 2016). "The FAIR Guiding Principles for scientific data management and stewardship". Scientific Data. 3: doi:10.1038/sdata.2016.18

[4] Laakso, Mikael; Björk, Bo-Christer (2013). "Delayed open access: An overlooked high-impact category of openly available scientific literature". Journal of the American Society for Information Science and Technology. 64 (7): 1323–1329. doi:10.1002/asi.22856

[5] ALA Scholarly Communication Toolkit Archived September 8, 2005, at the Wayback Machine

[6] Anon, (2020). [Blog] Available at: https://blogs.bournemouth.ac.uk/research/researcher-toolbox/research-outputs/bu-open-access-mini-guide/ [Accessed 12 Jan. 2020].

[7] Zenodo - Research. Shared. (2020). Retrieved 13 January 2020, from https://zenodo.org/

[8] figshare - credit for all your research. (2020). Retrieved 13 January 2020, from https://figshare.com/

[9] Mendeley Data. (2020). Retrieved 13 January 2020, from https://data.mendeley.com/

_____