

How to make your data repository successful and sustainable: learnings from figshare-powered repositories

Mark Hahnel, Founder & CEO for Japan Open Science Summit 2021



データリポジトリの成功と持続可能性のために:

figshare機関版からの知見

Mark Hahnel, Founder & CEO for Japan Open Science Summit 2021

This webinar is about:

- What is Figshare and what we do
- What we think is important
- How different organizations are working with us
- And how we implemented their feedback into their repositories

How and when Figshare started? Why did you start it?

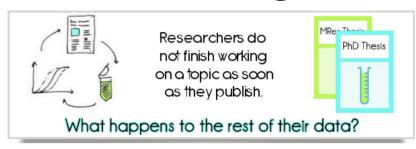
I owe my business to my frustration as a Scientist – Figshare Founder Mark Hahnel: Of Schemes and Memes Blog. (2014, June 20). Nature.com. http://blogs.nature.com/ofschemesandmemes/2014/06/20/i-owe-my-business-to-my-frustration-as-a-scientist-figshare-founder-mark-hahnel



FigShare | beta PUBLISH ALL YOUR DATA



feedback, visibility, etc.



figshare.com



A **cloud based** research data management system where you can:







Manage your research outputs privately and securely

Make your research outputs citable, sharable, discoverable



"I want my research to have more impact than anyone else's."



"Investigators are expected to **share with other researchers**, at no more than incremental cost and within a reasonable time, **the primary data**, samples, physical collections and **other supporting materials** created or gathered in the course of work under NSF grants"



http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4

"NIH expects the **timely release and sharing of data** to be no later than the acceptance for publication of the main findings from the final dataset."

http://grants.nih.gov/grants/policy/data sharing/data sharing guidance.htm#time



"NEH is committed to timely and rapid data distribution."

http://www.neh.gov/files/grants/data_management_plans_2012.pdf



Who are using figshare.com and why?

figshare proved to be a good idea. Since 2012 we have offered...

26 million +

page views

7.5 million +

downloads

800,000 +

user uploads

2 million +

articles

500,000 +

collections

5,000 +

projects

Remdesivir					Vehicle					
0	0	0	0	0	0	0	0	0	0	0
3.638322	7.944003	7.037755	6.421312	7.097592	6.328537	5.003218	6.205009	8.243229	6.159435	7.360754
3.520792	7.805665	6.591881	6.432046	7.270482	5.639842	7.115913	6.952361	7.26932	7.358531	7.72567
0	6.789424	7.544269	6.308445	5.908887	4.022503	5.840717	0	6.90447	5.465937	6.724548
0	6.103718	4.15875	5.282032	6.799568	5.032269	4.757699	3.861664	6.111172	6.418617	6.307458
6.832035	5.735359	5.767165	5.899246	5.581897	3.337214	1.587728	3.908346	2.894674	4.36701	5.18362
3.524714	6.560891	3.964147	5.722443	4.586469	1.982908	3.243703	4.177106	5.529449	3.913707	4.74489
6.474144	4.711339	3.561905	4.433067	4.937532	2.344143	0	4.820715	4.240739	0	2.142409

(0.4.07.1.0)

LungBodyWeight Gross Lung Lesions

Nose Swabs RNA Th

Throat Swabs RNA Rectal Swabs RNA

Nose Swabs virus

sx (34.97 kB) 1

Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2

Cite

Download (34.97 kB)

Share I

nbed + Collect

2582 views

355 downloads

2 citations r₹

Version 3 ➤ Dataset posted on 21.05.2020, 20:04 by Brandi Williamson, Friederike Feldmann, Benjamin Schwarz, Kimberly Meade-White, Danielle Porter, Jonathan Schulz, Neettje van Doremalen, Ian Leighton, Claude Kwe Yinda, Lizzette Perez-Perez, Atsushi Okumura, Jamie Lovaglio, Patrick Hanley, Greg Saturday, Catharine Bosio, Sarah Anzick, Kent Barbian, Tomas Cihlar, Craig Martens, Dana Scott, Vincent J. Munster, Emmie de Wit





Browse Search on The NIH Figshare Ar... Q



citations

CATEGORIES

KEYWORDS

Perforin-2

MACPF

LICENCE

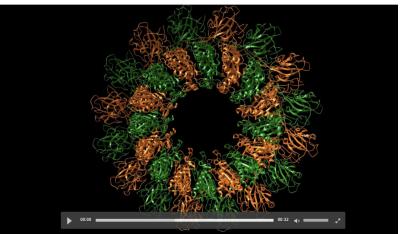
Immunology

Infectious Diseases

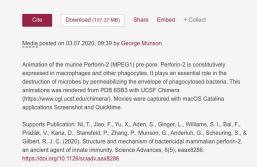
© 0 CC BY 4.0

MPEG1

· Innate Immunity



Animation of the Perforin-2 (MPEG1) pre-pore which forms at neutral pH and likely precedes the acid-dependent transition to the bactericidal pore in phagosomes.



Publication Abstract:

Researcher incentives

- Any "non-traditional" scholarly outputs
- Any file types (1,000+ formats to preview)
- Free storage (5GB per file)
- Free DOIs
- Metrics (views, downloads, citations, altmetrics)
- Easy to use, no training needed

What do you think is important for data repositories?

Some of Figshare's core beliefs

- Academic research outputs should be as open as possible, as closed as necessary
- Academic research outputs should never be behind a paywall
- Academic research outputs should be human and machine readable/query-able
- Academic infrastructure should be interchangeable
- Academic researchers should never have to put the same information into multiple systems at the same institution
- Identifiers for everything
- The impact of research is independent of where it is published and what type of output it is

What was the market response?

What we heard from:

Individuals

- Researchers wanted to publish different types of outputs
- And get credit for them
- But they didn't have a digital storage space to publish those 'non-traditional' outputs

Institutions

- Institutional or subject-specific repositories preferred
- Generalist repositories are okay but more control needed
- Keeping up with new policies and the advances in technology is difficult

How did you respond to those gaps?

An infrastructure that you have to build and maintain simply configure and let run

The Figshare platform is ready to run and constantly <u>maintained</u>, enabling your researchers to store and publish their data with ease. <u>Integrations</u> are available, allowing to use institution IDs and Single Sign On (SSO).

The workflows are made for research data, and customisations and extra services such as curation can be added on.



How does it work for institutions?

This is your repository

To-do lists Add a to-do list
Getting Started
☐ Read and complete the implementation guide and form: https://goo.gl/for
☐ Arrange next meeting (calendar link in description) 1 comment
☐ Read the admin and end user guides 1 comment Mark Hahnel
Add a to-do
Implementation Tasks
☐ Setting up your stage (test) portal 1 comment
☐ Storage 1 comment
☐ User Authentication 1 comment
☐ Creating User Accounts 1 comment
☐ Branding 1 comment
☐ Groups (i.e. journal pages) and their URLs 1 comment
☐ Quota 1 comment
□ DOIs (1 comment)
☐ Other Customisations 1 comment
☐ Setting up your production portal 1 comment
☐ Support and Engagement 1 comment
Add a to-do

What are the considerations when building a repository?

Stamping—FAIR Data Badge? OA label in Google Scholar?

Feedback—peer-review? Comments? Usage stats?

Marketing—strategy? Engagement? SNS sharing?

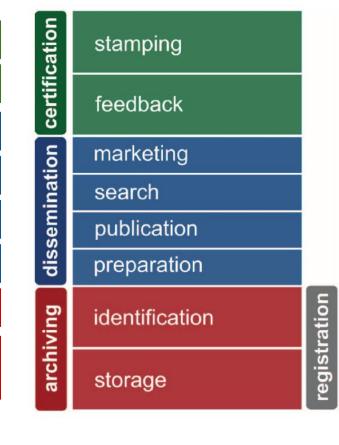
Search—discoverability? Indexing services? Harvesting?

Publish—License? Embargo? Versioning?

Curation—who? Metadata schema? Workflow?

Persistent identifiers—DataCite? Crossref? Handles?

Data sovereignty—which country? Your own storage? Storage—data size?







<u>Vision & Objectives</u> <u>Principles</u> <u>Design Assumptions</u> <u>User Stories</u> <u>Behaviours</u> <u>Technologies</u> <u>About</u>

About the Technologies

This section presents a snapshot of the current status of technology, standards and protocols available to support each behaviour.

The Next Generation Repositories Working Group has explicitly focused on the generic technologies required by all repositories to support the adoption of common behaviours. However, we also recognize that there are other technologies and standards that may be useful for certain content types or disciplinary communities.

In some cases, the technologies required to support a specific behaviour are not yet sufficiently mature, or it is not yet clear what technology will prevail. In other cases, where there are not currently no appropriate technologies to support the specific behaviour. In these cases, the Working Group was not able to recommend specific technologies, however we will continue to monitor developments and make recommendations as new or better technologies become available.

List of technologies

- · Activity Streams 2.0
- COUNTER
- Creative Commons Licenses
- ETag
- HTTP Signatures
- International Image Interoperability
 Framework
- IPFS
- Linked Data Notifications
- OpenID Connect
- ORCID
- ResourceSync
- Signposting
- Sitemaps

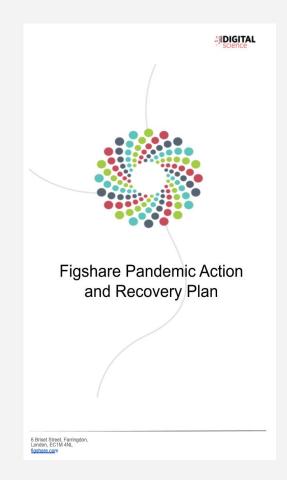
Policy Compliant Infrastructure

ISO27001

ISO 27001 is a specification for an information security management system (ISMS). An ISMS is a framework of policies and procedures that includes all legal, physical and technical controls involved in an organization's information risk management processes.

DataCite Node
Region Specific Storage - AWS GovCloud
Single Sign on Integrations
Customizable Terms and Conditions
Accessibility: WCAG 2.1 AA, Section 508, and EN 301 549

Google Scholar
Google Dataset Search Alpha Partner
ORCID Founding Member
Make Data Count



How are different organizations using figshare to build their repositories?

Repositories we power are used in many wonderful ways

Research Centers Portfolios Museums Special Collections University Press

Open Educational Resources (OER) Theses Repository Institutional and Data Repository

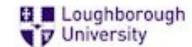
Conference Materials Archives Private Collaboration

Institutional and Data Repository

Several institutions use their Figshare portal as all-in-one institutional and data repository.



- University of Leicester
- Loughborough University
- <u>KiltHub</u> @ Carnegie Mellon University





Theses Repository

Some institutions use their Figshare instance as a full institutional repository or as a theses repository.

- Monash University Theses
- University of Leicester Theses
- Purdue University Graduate School







Portfolios

Whether it's to display the outputs from a particular project or for assessment, several institutions use collections to group portfolios together.

- Artist Boss @ Bath Spa University (REF submission)
- Art Pop Intersections @ University of Salford





Open Educational Resources (OER)

Some institutions use Figshare exclusively for learning objects.

- <u>Utrecht University</u>
- Stockholm University
- <u>La Trobe University</u>



Museums

Museums can use Figshare for showcasing their exhibits.

- Smithsonian Research Data
- The British Museum
- Museum of Domestic Design & Architecture (MoDA)
 - @ Middlesex University London







Special Collections, Archives

Creating a subgroup for special collections is a great way of displaying the contents of the collection in a controlled, but public way.

- The Planet Bethlehem Archive
 @ University of Sussex
- Zamani Project @ ZivaHub (University of Cape Town)
- Music Archive of Monash University





Private Collaboration

Researchers can invite their collaborators to a private project to share data privately. If and when the project is ready to be shared, it can be made publicly available, or selectively.

see how restricted publishing works

 Murdoch Children's Research Institute (case study)



Conference Materials

Display outputs from the conference held at your institution—presentations, posters, conference proceedings, recordings from the event, and more.

 <u>List of conferences</u> that used Figshare to share their materials







Are there any organizations using figshare for their repository in Japan?



- J-STAGE Data is a research data repository service provided by Japan Science and Technology Agency (JST).
- Datasets associated with articles published on J-STAGE main platform, where 3,000+ publications are hosted.
- Implementation started in late 2019; full launch in March 2021 after
 1-year pilot period.

論文とデータの相互参照





ALL CATEGORIES JOURNALS SEARCH Q

6 Journals in total







日本リモートセンシング学会 誌/Journal of The Remote Sensing Society of Japan







:)

mark@figshare.com @MarkHahnel