

June 05, 2018

Balancing Multiple Roles of Repositories:

Developing a Comprehensive Institutional
Repository at Carnegie Mellon University

David Scherer, Scholarly Communications and Research
Curation Consultant, CMU

Lisa Zilinski, Research Data Consultant, CMU

Dan Valen, Strategic Partnerships, Figshare



Introduction

Carnegie Mellon University

Founded in 1900 by Andrew Carnegie

13,961 students from 109 countries

47%UG, 53% Grad

1,391 Faculty from 42 countries

BIG IDEAS THAT SHAPE YOUR WORLD START HERE



1,000+ companies greenlight
across the U.S. and internationally



Cognitive tutors, which improve
the math skills of 500,000+
students each year



The first U.S. school to award
a degree in drama in 1914



James Gosling (CS 1983)
invented the **Java computer
programming language**.



The first **exclusive higher
education partner of the Tony
Awards®**, co-creating the first,
national recognition program
honoring K-12 theatre educators



The **first smile in an email**, created
in 1982 by professor Scott Fahlman



Kevlar Fiber, invented by alumna
Stephanie Kwolek (MM 1946)



The Last Lecture, written
by alumni Randy Pausch (CS 1988)
and Jeffrey Zaslow (DC 1980)



In 1919, **CMU's first Ph.D.** was
awarded to Mao Yisheng, father of
Chinese bridge construction.



Duolingo, invented by professor
Luis von Ahn (CS 2003, 2005), had
10 million downloads in 12 months
and was named Apple's 2013 app
of the year.



The leading university center
for cybersecurity, providing
support to more than 110
centers around the world



The **world's first university
robotics department**,
founded in 1979



Alumnus **Andy Warhol**
(A 1949), pop artist pioneer
and cultural icon



One of only 25 universities
invited to be a member of the
**World Economic Forum's
Global University
Leaders Forum**

My Heart is in the Work



Source: CMU Fact Sheet

<https://www.cmu.edu/assets/pdfs/cmufactsheet.pdf>

**Carnegie
Mellon
University**

figshare.com to Figshare for Institutions

Some of team 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

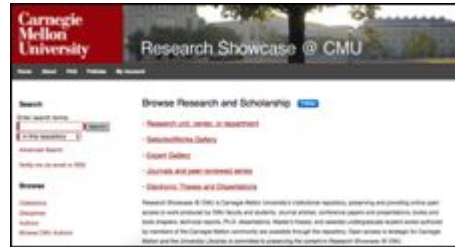
Source: <https://knowledge.figshare.com/articles/item/mission-statement-and-core-beliefs>



**Carnegie
Mellon
University**

Repository Landscape at CMU

Traditional IR



Archives



Data Repository



Evolution of Repository Services at CMU

October 2008 – Creation of Research Showcase Repository (Digital Commons)

February 2015 – First dataset accepted by CMU Libraries, added to archival repository

February 2015 – CMU evaluates repository solutions for new data repository and other repository needs

October 2015 – Formation of the Digital Repository Task Force to evaluate repository solutions

October 2015 - May 2016 – CMU Evaluates RIM Systems and related platforms and services

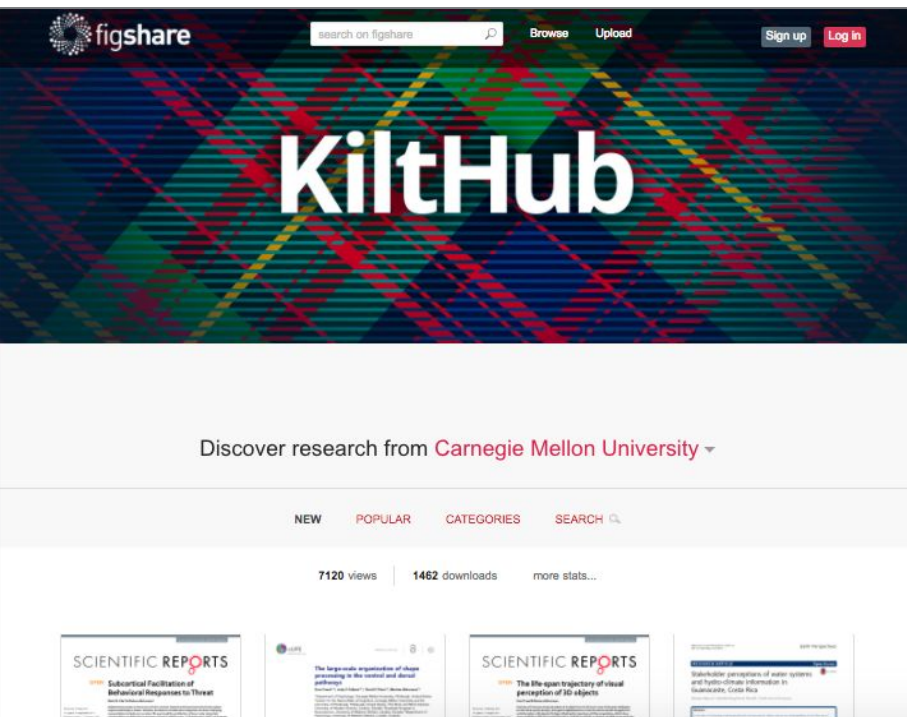
February 2017 – CMU and Digital Science announce strategic development partnership

February – March 2017 – CMU Libraries conducts campus-wide naming contest for new repository

April 2017 - KiltHub Repository becomes operational, deposits first dataset

KiltHub – Figshare for Institutions

kilthub.cmu.edu



figshare

search on figshare

Browse Upload

Sign up Log in

KiltHub

Discover research from Carnegie Mellon University

NEW POPULAR CATEGORIES SEARCH

7120 views | 1462 downloads | more stats...

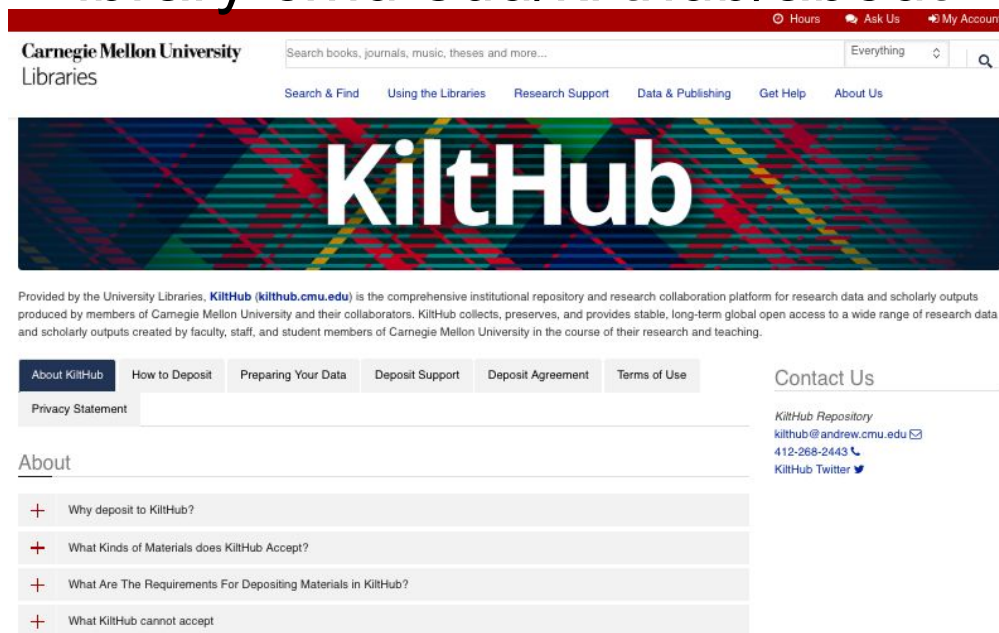
SCIENTIFIC REPORTS

The largest organization of shape processing in the ventral and dorsal pathways

SCIENTIFIC REPORTS

The life span trajectory of visual perception of 3D objects

library.cmu.edu/kilthub/about



Carnegie Mellon University Libraries

Search books, journals, music, theses and more...

Everything

Search & Find Using the Libraries Research Support Data & Publishing Get Help About Us

KiltHub

Provided by the University Libraries, **KiltHub** (kilthub.cmu.edu) is the comprehensive institutional repository and research collaboration platform for research data and scholarly outputs produced by members of Carnegie Mellon University and their collaborators. KiltHub collects, preserves, and provides stable, long-term global open access to a wide range of research data and scholarly outputs created by faculty, staff, and student members of Carnegie Mellon University in the course of their research and teaching.

About KiltHub How to Deposit Preparing Your Data Deposit Support Deposit Agreement Terms of Use Privacy Statement

About

- Why deposit to KiltHub?
- What Kinds of Materials does KiltHub Accept?
- What Are The Requirements For Depositing Materials in KiltHub?
- What KiltHub cannot accept

Contact Us

KiltHub Repository
kilthub@andrew.cmu.edu
412-268-2443
KiltHub Twitter

The KiltHub Repository Team

The Repository Services Team

David Scherer

Scholarly Communications and
Research Curation Consultant

Katie Behrmann

Repository Specialist

Ann Marie Mesco

Data Deposit Coordinator

The Data Services Team

Lisa Zilinski

Research Data Consultant

Eric Kaltman

CLIR Fellow for Data Curation in
the Sciences

Emma Slayton

CLIR Fellow for Data Visualization
and Curation

Liaison Librarian Team

Streamlining Workflows

Roles

Repository Administrator, Site Level Admin (RA)

Data Deposit Administrator, Site Level Admin (DA)

Data Services Team (DST)

Liaison Librarians (LL)

Research Data Consultant (RDC)

Deposit Workflow Amongst Admins Before Curation Workflow

1. Deposit notification comes to all site level administrators
2. RA conducts initial review from notification
3. RA assigns DA as curator to Datasets and Filesets

Streamlining Workflows

Data Deposit Workflow

1. DA is assigned as curator from RA
2. DA checks KiltHub deposit metadata for deposit consistency
3. DA checks README.txt for dataset for deposit consistency
 - a. If dataset meets deposit requirements, DA moves to Step 10
 - b. If dataset does not meet deposit requirements, DA moves to Step 4
4. DA sends email to researcher making contact, informing them that they will be in touch with questions regarding their deposit.
5. DA contacts LL regarding deposit*

*Ideally, LL would have worked with researcher advance of deposit. Additionally, LL should contact RST letting them know of their interactions with researchers for forthcoming deposits.

Streamlining Workflows

Data Deposit Workflow

6. When necessary, DA contacts DST via RDC Cc'ing LL
7. Based upon comments and feedback between DA, LL, and DST, the researcher is contacted by LL or DA (Based on LL preference), Cc'ing all
8. Researcher responds to LL/DA
9. Based on circumstances, AMM or researcher revises dataset
 - a. If AMM is to revise, she should revise the materials
 - b. If researcher is to revise, dataset should be declined. AMM should add notes to the Comments field what the researcher should revise

Streamlining Workflows

Data Deposit Workflow

10. AMM reviews dataset after revisions to dataset have been made

Steps 7 - 9 can be repeated as needed to ensure deposit is consistent with data deposit requirements

11. DA approves dataset for deposit

Balancing Requirements and Ease-of-Deposits

Requirements

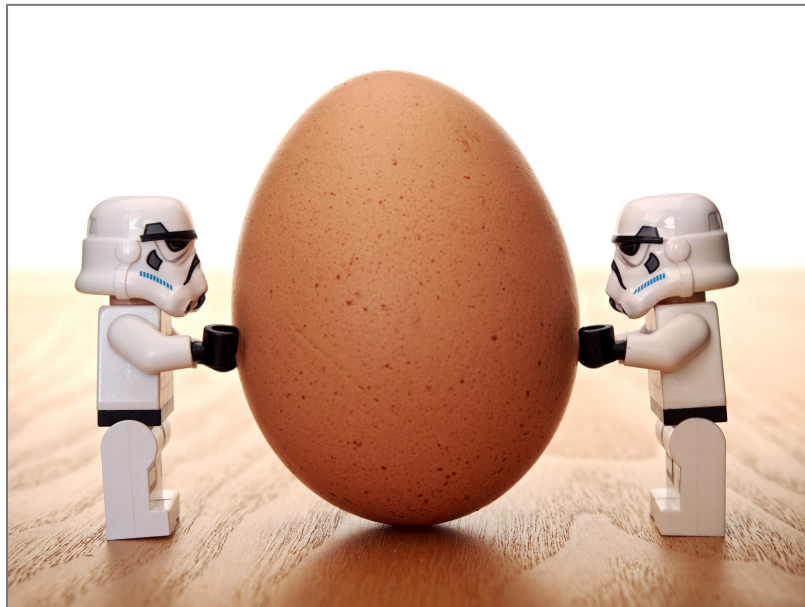
Submission Metadata

README.TXT

Data Dictionary

File Naming
Conventions

Proper File types for
access and preservation



<https://pxhere.com/en/photo/806073> CC0

Ease-of-Deposit

Quick and easy
submissions

Fast-turnaround

DOI for
publications/funders

Tied to
publication/funder
timeframes

IRs and irs

Institutional Repositories

- What does it mean for a repository to be an “Institutional Repository?”
- What is the role of the Institutional Repository? How is this affected by its capabilities and limitations?
- What does the IR collection and why?
- Is the mission of all IRs the same?

institutional repositories

- How is the repository administered?
- What level of control/administration does an institution need over its repositories?
- How do you provide a repository as a service?
- How should the repository be seen in the broader scholarly communication ecosystem of an institution?

Figshare IR Advisory Board and the community

The image is a composite of three screenshots related to Figshare's Institutional Repository (IR) project. On the left is the 'Public roadmap' for version 17.05.2018, featuring the Figshare logo and a list of development tasks such as 'Development process', 'Environment definition', and 'Current feature release schedule'. In the center is a Basecamp project page for the 'Figshare IR advisory group', showing discussions and project updates. On the right is a blog post titled 'Announcing the figshare Institutional Repository...and Data Repository...and Thesis Repository...really just an all-in-one next gen repository'. A large teal arrow points from the roadmap towards the Basecamp page, indicating a flow of information or a link between the two.

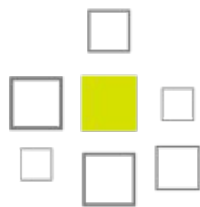
Public roadmap
Version - 17.05.2018

Figshare IR advisory group
14 Discussions 6 Files Add the first: To-do list Text document Event
Latest project updates
2:46am Patrick S. gave S.Chang@latrobe.edu.au access to the project: Figshare IR advisory group
1:54am Patrick S. posted a message: [Two docs to consider: Figshare roadmap vs Figshare as an IR](#)
May 9 Alan H. gave Florin A. access to the project: Figshare IR advisory group
See all updates
Discussions Post a new message Watch a quick video about Discussions
Patrick S. [Two docs to consider: Figshare roadmap vs...](#) - Hi all, For those of you who haven't seen both of these documents I've attached our Figshare roadmap from April 2018, and
Dan Valen

Announcing the figshare Institutional Repository...and Data Repository...and Thesis Repository...really just an all-in-one next gen repository
22/03/18 14:15 by Alan Hyndman
Figshare is building out our existing data (or non-traditional research outputs) repository functionality to cover all the functionality needed for a traditional Institutional Repository (IR). We have worked hard over the past 5 years to build what we believe to be the best data repository available. We have worked on distributing our infrastructure to comply with funder storage policies around the globe, whilst consistently adding functionality, from citation counts to customizable metadata. As such, we now have a customizable system that can live on institutional domains, with institutional branding and institutional DOIs. We accept any file format and are focussed on making all outputs

https://figshare.com/blog/Announcing_the_figshare_Institutional_Repository_and_Data_Repository_and_Thesis_Repository_really_just_an_all-in-one_next_gen_repository/389

The IR and Research Information Management



DSpace



eprints
repository software



Elements

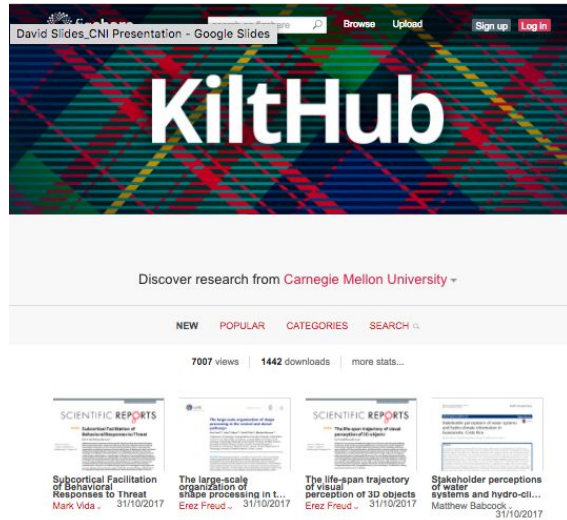


CONVERIS
Research Information System

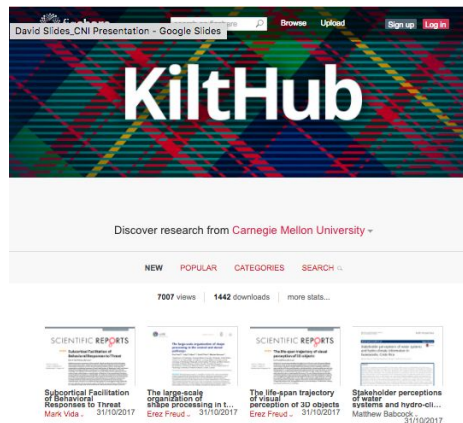
DigitalMeasures

**Carnegie
Mellon
University**

The IR and Research Information Management



IR to RIM

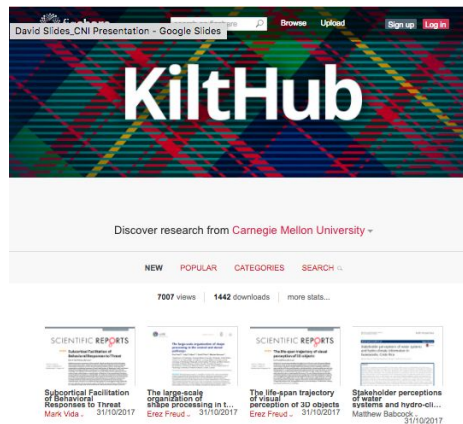


**Metadata
Harvesting
and Data Match**



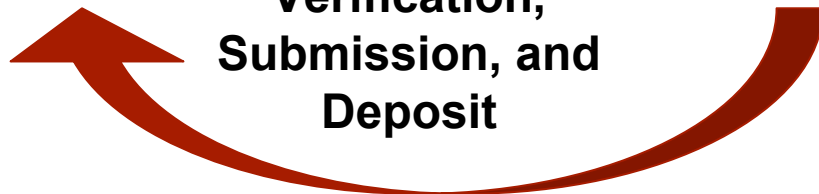
Elements

RIMs to IRs



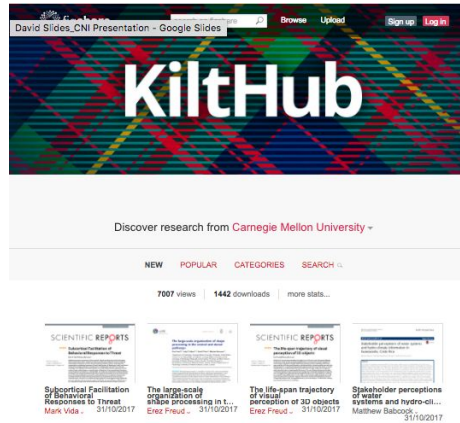
Elements

Article
Verification,
Submission, and
Deposit



Carnegie
Mellon
University

IRs and RIMs



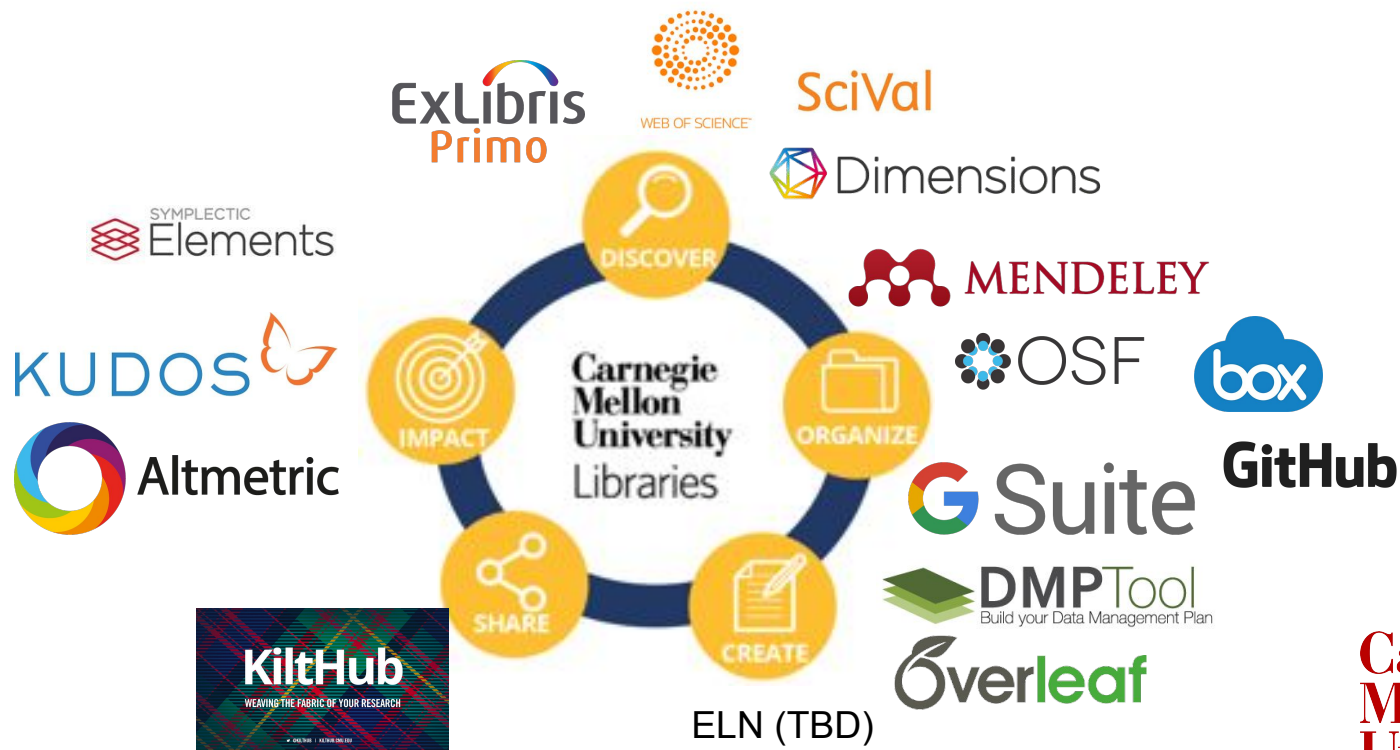
Monitoring OA
Status and
Engagement



Elements

Carnegie
Mellon
University

The Repository in the Scholarly Communication Ecosystem



KiltHub

WEAVING THE FABRIC OF YOUR RESEARCH

🐦 @KILTHUB | KILTHUB.CMU.EDU

Weaving the Fabric of Research: The Warp and Weft

Research Data

A dataset is typically a file or collection of files and its corresponding metadata. This metadata may include preservation and presentation information, as well as annotations and ancillary content that is associated with the data. KiltHub is limited in scope to the raw data, or data that supports the claims made in publications/presentations/figures.

Scholarly Outputs

When evaluating scholarly outputs for inclusion in KiltHub, *scholarly outputs* should be broadly understood to mean the published or otherwise publicly presented outputs, created by individuals or units within Carnegie Mellon University in the course of their scholarly or professional work that represents the intellectual life and academic endeavor of the university, and may have enduring value.

Thank you

Questions / Comments

David Scherer

daschere@andrew.cmu.edu
@davidascherer

Lisa Zilinski

ldz@andrew.cmu.edu
@l_zilinski

Dan Valen

dan@figshare.com
@dnvln

Pages

KiltHub Repository:
<https://kilthub.cmu.edu>

KiltHub About Pages:
<http://library.cmu.edu/kilthub/about>