Building Community and Support for Open Science at Carnegie Mellon University – A Conference Report

Keith Webster
Dean of University Libraries and Director of Emerging and Integrative Media Initiatives

Huajin Wang, PhD
Biomedical Data Science Liaison
We do not do everything, but at what we do we aim to be the best in the world.
56 Members of NAE
17 Members of NAS
5 Members of NAM
44 Tony Awards
10 Academy Awards
116 Emmy Awards
12 Turing Awards
20 Nobel Laureates

▲ Won by alumni and current/former faculty

1 National Academy of Engineering
2 National Academy of Sciences
4 National Academy of Medicine
New Creative Industries

Concerts, Galleries, Daily Life, Aesthetics

Film

Digital Content
Games
Social Media
Urban Systems
Integrative Design
Maker Movement, Additive Manufacturing
Modular Content, Participatory Culture, Responsive Environments, Internet of Things

Civil and Manufacturing Infrastructure

Computing

Digital Signal Processing

Mobile Systems

Ubiquitous Computing
1ST INDUSTRIAL REVOLUTION
2ND INDUSTRIAL REVOLUTION
3RD INDUSTRIAL REVOLUTION
4TH INDUSTRIAL REVOLUTION
<table>
<thead>
<tr>
<th>Registry</th>
<th>Name</th>
<th>Abbreviation</th>
<th>Type</th>
<th>Domain</th>
<th>Taxonomy</th>
<th>Related Database</th>
<th>Related Standard</th>
<th>Related Policy</th>
<th>In Collection/Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Australian Antarctic Data Centre</td>
<td>AADC</td>
<td>Database</td>
<td>Environmental Sciences, Biology, Geology, Geographical</td>
<td>Geology</td>
<td>GBIF</td>
<td>None</td>
<td>None</td>
<td>PLOS Recommended Data Repositories, Scientific Data's Recommended Data Repositories</td>
</tr>
<tr>
<td></td>
<td>Archaeology Data Service</td>
<td>ADS</td>
<td>Database</td>
<td>Classical Archaeology, Natural History, Humanities</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Scientific Data's Recommended Data Repositories</td>
</tr>
<tr>
<td></td>
<td>TERN AEOCS</td>
<td>AEOCS</td>
<td>Database</td>
<td>Biodiversity, Data Storage, Data Dissemination, Scientific Data, Multidisciplinary, Geographical, Natural Sciences</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Scientific Data's Recommended Data Repositories, Scientific Data Policies</td>
</tr>
<tr>
<td></td>
<td>Australian Open Data Network Portal</td>
<td>ADDON</td>
<td>Database</td>
<td>Marine Environment, Natural Sciences</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Scientific Data's Recommended Data Repositories</td>
</tr>
<tr>
<td></td>
<td>ArrayExpress</td>
<td>ArrayExpress</td>
<td>Database</td>
<td>Chromatin Immunoprecipitation, DNA Methylation (CHIP-Seq), DNA Microarray, DNA Methylation, Data Expression, Data Dissemination</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>GigaScience - Standards of Reporting Checklist, American Association for the Advancement of Science - Science - Editorial Policies, Data Materials, Availability, Nature Publishing, Group - Nature Biotechnology - Availability of Data and Materials</td>
</tr>
</tbody>
</table>

**Record Status**
- Best Match

**Recommended Records**
- Recommended

**Associated Publication?**
- Has Publication

**Claimed?**
- Has Maintainer

**Record Status**
- Uncertain
- Deprecated
- In development
- Ready

**Standard Type**
- Reporting Guideline
  - 3
- Terminology Artifact
  - 1

**Registry**
- Database
  - 87
- Standard
  - 4
- Policy
  - 2

**Domains**
- Life Science
  - 62
- Biomedical Science
  - 25
Diederik Stapel, a professor of social psychology in the Netherlands, had been a rock-star scientist — regularly appearing on television and publishing in top journals. Among his striking discoveries was that people exposed to litter and abandoned objects are more likely to be bigoted.

And yet there was often something odd about Stapel's research. When students asked to see the data behind his work, he couldn't produce it readily. And colleagues would sometimes look at his data and think: It's beautiful. Too beautiful. Most scientists have messy data, contradictory data, incomplete data, ambiguous data. This data was too good to be true.
ALWAYS OPEN?

PHILOSOPHICAL TRANSACTIONS - 1665
ASSOCIATION OF IDEAS INDEX - 1960s
IMPACT FACTOR - 1960s
SALE OF JOURNAL TITLES BY SOCIETIES - 1960s-1980s
The Open Science Framework - cos.io
Open science by design

1. Provocation: connect and discover
2. Ideation: plan and design
3. Knowledge generation: observe and experiment
4. Validation: analyze and interpret
5. Dissemination: report and share
6. Preservation: store and maintain

https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century
Researchers in developing countries can see your work

More exposure for your work

Practitioners can apply your findings

Taxpayers get value for money

Higher citation rates

Compliant with grant rules

Your research can influence policy

The public can access your findings

CC-BY Danny Kinglsey & Sarah Brown
101 Innovative tools and sites in 6 research workflow phases
(< 2000 - 2015)
STRATEGIC RECOMMENDATION

Create a 21st century library that serves as a cornerstone of world-class research and scholarship.

In a world where digital is becoming the default format for information, the library will remain a vital presence on campus, sustaining serious scholarship and providing opportunities for interactive research and study environments. To support this important work for students, faculty, and staff, and to create 21st century library spaces for 21st century learners, the library will:

- Develop information specialists as partners in research, teaching, and learning.
- Collaborate with peer institutions to provide coordinated access to a global collection of information resources.
- Steward the evolving scholarly record, and champion new forms of scholarly communication.
- Be recognized globally as a leader in the development of the scholarly information ecosystem.
Open Science Framework: Start

A guide for how to use the open source project management platform, Open Science Framework

About Open Science Framework

OSF is a free, open source web application for project management from the non-profit, Center of Open Science. Researchers use OSF to collaborate, document, archive, share, and register research projects, materials, and data. This guide will teach you how to:

- create an account on OSF
- create a project in OSF and share with collaborators
- how to organize projects
- preregister a research project
- how to get a DOIs

Getting Started with OSF

Create an OSF Account

Get started by creating a free account with your Andrew ID. Find helpful info on how OSF can be used here.
<table>
<thead>
<tr>
<th>Service</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discover Your Research Topic</td>
<td>Discover and refine your research topic with resources that connect you with specialists and databases.</td>
</tr>
<tr>
<td>Find Funding &amp; Collaborators</td>
<td>Uncover potential collaborators and the funding to support your project.</td>
</tr>
<tr>
<td>Manage Information &amp; Data</td>
<td>Work with our specialists to evaluate, select, and navigate the tools to organize your data and keep your project on track.</td>
</tr>
<tr>
<td>Publish &amp; Share Your Research</td>
<td>Navigate the many options for publishing your research to optimize its accessibility and impact.</td>
</tr>
<tr>
<td>Measure Impact &amp; Manage Your Digital Identity</td>
<td>Identify when and where your research has been cited, as well as the scope of its reach.</td>
</tr>
<tr>
<td>Enhance Your Teaching</td>
<td>Advise your students about research methods and resources that enhance their learning inside and outside the classroom, unlocking your teaching potential.</td>
</tr>
</tbody>
</table>
Boy without one-sixth of his brain is a remarkably 'normal' 12-year-old

A boy who had one-sixth of his brain removed to stop daily seizures now acts like a typical 12-year-old, according to researchers and his parents.

Tanner Collins from New Stanton, Pennsylvania, was six years old when he had a tumor in the right hemisphere of his brain removed. At the time, his family was unsure if it was benign or malignant. Before the surgery, his parents, who are both nurses, felt they had exhausted options to rid him of frequent seizures. The anti-epileptic medications were also taking their toll on Tanner's personality, behavior, and performance in school, his parents Carl and Nicole Collins said.

When surgeons removed one sixth of a child's brain, here's what happened

By Susan Scotti, CNN
Updated 11:51 AM ET, Tue July 31, 2018

A.

Pre-operative (FLAIR)

Post-operative (TI)

B.

LH

RH

Surgeons removed part of a child's brain for epilepsy and it revealed itself.

Boy’s brain works just fine after a large piece was removed

Anterior

Posterior

LH

RH

LH

RH

NewScientist Live 2018
Tickets selling fast: book your place now!

News Technology Space Physics Health Environment Mind | Travel Live Jobs
The figure shows the left and right hemispheres of the patient's brain after the procedure. The dotted line indicates where the entire occipital lobe, which includes the brain's visual processing center, and most of his temporal lobe were removed. Despite the lobectomy and although the patient can't actually see the left half of his world, the preserved left hemisphere compensated for visual tasks such as recognizing faces and objects.

Related

Washington Post: A 12-year-old had one-sixth of his brain removed. He feels ‘perfectly normal.’

Newsweek: Lobectomy Study: Scientists Reveal Boy’s Incredible Recovery After Large Chunk of his Brain was Removed

PBS News Hour: A child lost a sixth of his brain, then made an amazing comeback

Cell Reports: Successful Reorganization of Category-Selective Visual Cortex following Occipito-temporal Lobectomy in Childhood

August 01, 2018

Case Study: Child's Lobectomy Reveals Brain's Ability To Reorganize Its Visual System

After three years, patient recognizes faces normally, despite removal of preeminent regions involved in facial recognition
Successful Reorganization of Category-Selective Visual Cortex following Occipito-temporal Lobectomy in Childhood

Liu TT, Nestor A, Vida MD, Pyles JA, Patterson C, Yang Y, Yang FN, Freud E, Behrmann M


Reporting date: 01 Jan 2018 Edit

Altmetrics: 633

Citation Metrics:

- Scopus: 0
- WoS: -

Journal Rankings:

- Cell Reports
  - SNIP: 1.65
  - SJR: 7.55
Title: Successful Reorganization of Category-Selective Visual Cortex Following Occipito-temporal Lobectomy in Childhood
Published in: Cell Reports, July 2018
DOI: 10.1016/j.celrep.2018.06.099
PubMed ID: 30067969
Authors: Liu, Tina T., Nestor, Adrian, Vida, Mark D., Pyles, John A., Patterson, Christina, Yang, Ying, Yang...
Abstract: Investigations of functional (re)organization in children who have undergone large cortical...
So far, Altmetric has seen 77 news stories from 71 outlets.

**deVolkskrant**
Hoeveel brein heb je eigenlijk echt nodig?
Volkskranti, 08 Sep 2018
Een 44-jarige man komt bij de dokter. Hij heeft de laatste tijd weinig kracht in zijn linkerbeen.

**MedPage Today**
Plasticity after Lobectomy: The Hunt for ALS Genes; Marijuana and MS Patients
MedPage Today, 07 Aug 2018
A child who lost more than one-third of the right side of his brain showed intact cognitive abilities and intermediate- and high...

**Tampa Bay Times**
A 12-year-old had one-sixth of his brain removed. He feels 'perfectly normal.'
Tampa Bay Times, 05 Aug 2018
It was a solution no parent wants to hear: To get rid of a brain tumor and stop their young son's seizures, surgeons would need...

**Brain's 'plasticity' amazes as boy recovers from drastic surgery**
Health24, 08 Aug 2018
The developing brain of a growing child has incredible ways of compensating for the loss of an essential brain region, a new...

**Epilepsie. Un enfant guéri par l'ablation d'une partie de son cerveau**
Free, 06 Aug 2018
Un garçon de six ans a subi une ablation partielle du cerveau afin de soigner son épilepsie.

**Après l'ablation d'une partie de son cerveau, ce garçon va bien**
Top Santé, 05 Aug 2018
Un peu avant l'âge de 7 ans, on lui a retiré un tiers de l'hémisphère droit de son cerveau pour soigner son épilepsie.

**Epilepsie : un enfant de 6 ans soigné grâce à l'ablation d'une partie de son cerveau**
Successful Reorganization of Category-Selective Visual Cortex following Occipito-temporal Lobectomy in Childhood

Tina T. Liu 1,2, Adrian Nestor 3, Mark D. Vida 1,2, John A. Pyles 1,2, Christina Patterson 4, Ying Yang 1, 5, Fan Nils Yang 6,7, Erez Freud 1,2, Marlene Behrmann 1,2, 8, 8, 9, 9

https://doi.org/10.1016/j.celrep.2018.06.099

Highlights

- A 3-year longitudinal investigation of visual plasticity post lobectomy in a child
- Stable left homonymous hemianopia and no remapping of the early visual cortex
- Competition between face and word selectivity within a single (left) hemisphere
- Intact cognitive abilities and intermediate- and high-level visual function

Summary
Bootstrapped regression slopes were calculated from the randomly picked 4 values (as a proxy for 4 sessions) after shuffling the condition labels in the upper (or equivalently the lower) RDM. This analysis yielded a distribution of the bootstrapped regression slopes (cyan histogram in Figure 4F), and the face and word dissimilarity slope (red dot in Figure 4F) and the house and object dissimilarity slope (yellow dot in Figure 4F) was each compared with this null distribution. To establish the statistical significance of the difference between bootstrapped slopes and the face and word dissimilarity slope or the house and object dissimilarity slope, we calculated the 95% CI of the obtained bootstrap distribution of the mean. For comparison, see RSA on a control anatomical region (LO2) in Figure S3.

Data and Software Availability

Behavioral and fMRI data and experiment scripts are available on KiltHub, which is a part of figshare (https://figshare.com/articles/Successful_Reorganization_of_Category-Selective_Visual_Cortex_Following_Occipito-Temporal_Lobectomy_in_Childhood/5919409/1).

Acknowledgments

This research was supported by NIH grant RO1 EY027018 (to M.B.) and a Presidential Fellowship from Carnegie Mellon University (CMU) (to T.T.L.). We thank Joel Greenhouse and Yuanning Li for statistical advice, Ev Fedorenko for providing the language localizer, and David Plaut for helpful comments. We also thank the patient, the controls, and their families for their time and cooperation; MRI technologists Scott Kudilla and Debbie Viszlay for help with imaging; and the VisCog group at CMU for fruitful discussion.

Author Contributions

Intact visuoperceptual function and category-selective organization in children with cortical resections

Dataset posted on 24.07.2018, 16:59 by Tina Liu, Erez Freud, Marlene Behrmann, Christina Patterson

This dataset includes fMRI raw data, behavioral data, experiment codes, and scripts related to the manuscript entitled: Intact visuoperceptual function and category-selective organization in children with cortical resections.

The fMRI folder includes raw fMRI data from 10 patients and 10 matched controls. In each participant, there was 1 anatomical scan and 3 functional runs of a category localizer, except in the case of one patient (NN) and the matched control, there were only two runs.

The behavioral folder includes raw data from 7 patients and 7 matched controls: each participated in 4 visuoperceptual tasks, focusing on global form perception (Glass pattern, contour integration) and pattern recognition (face recognition, object recognition).

All personally identifiable information has been removed. For more details, contact: behrmann@cmu.edu, tinaliutong@gmail.com.

FUNDING
NIH (RO1 EY027018)

DATE
24/07/2018
Intact visuoperceptual function and category-selective organization in children with cortical resections

Dataset created: Jul 24, 2018
Dataset updated: Jul 24, 2018
Dataset published: Jul 24, 2018

Dataset provided by figshare

Authors
Tina Liu; Erez Freud; Marlene Behrmann; Christina Patterson

License
https://creativecommons.org/licenses/by/4.0/

Description
This dataset includes fMRI raw data, behavioral data, experiment codes, and scripts related to the manuscript entitled: Intact visuoperceptual function and category-selective organization in children with cortical resections.
Organizers:
Huajin Wang
Ana Van Gulick
Melanie Gainey
Eric Yttri
The Inception

• “Open science is super important - let’s host an event to bring people together!”

• We need to engage researchers

• We need resources
Embedded liaison librarianship = collaboration with faculty

“Open science is super important - it brought so many collaborations to my research!!”

Eric Yttri
Assistant Professor of Biological Sciences
Carnegie Mellon University
Grant support and buy-in from college(s)

David Scaife Foundation:
“Here is some money - do some exciting things with it.”

Rebecca (Dean of MCS):
“This is the future of science!”

Rebecca Doerge
Dean of the Mellon College of Science
Carnegie Mellon University
City Hall of Gotham

Mellon Institute

The Venue
The Participants

Registrant Category

- Grad Student: 39%
- Staff: 11%
- Faculty: 17%
- Undergrad: 7%
- Postdoc: 19%

Registrant Department

- Biological Sciences: 15
- University Libraries: 9
- Computational Biology: 7
- Mechanical Engineering: 7
- Information Systems: 6
- Language Technologies Institute: 6
- Chemistry: 5
- Institute for Software Research: 5
- Psychology: 5
- Social and Decision Sciences: 4
- Heinz College: 4
- Pittsburgh Supercomputing Center: 4
- Mellon College of Science: 4
- Robotics: 4
- Psychiatry and Pediatrics: 3
- Public Policy and Management: 3
- Statistics: 3
- Health Sciences Library: 3
- Center for the Neural Basis of Cognition: 3
- Psychiatry: 3
- Physics: 2
- Human-Computer Interaction Institute: 2
- Electrical and Computer Engineering: 2
- Civil and Environmental Engineering: 2
- Machine Learning: 2
- Learning Research and Development Center: 2
- Figshare: 1
- Allen Institute: 1
- eScience Institute: 1
- Center for Machine Learning and Health: 1
Day 1:
Talks and Panels

- Open Science in Research
- Open Data and Reproducibility
- Open Tools and Platforms
- Open Access
The “Speed Dating”
Day 2: Hands-on Workshops

Track 1
- Preparing your data and code for reproducible publication (Code Ocean)
- Open Science on Bridges (Pittsburgh Super Computing Center)

Track 2
- Bioconductor (Sean Davis, NCI)
- ENCODE pipeline development framework: making your computational analysis scalable, reliable, portable and reproducible

Track 1+2
- Lightning talks: BenchSci, Protocols.io, Open Science Framework
- Lunch and KiltHub Deposit-a-thon (CMU Libraries)
Highlights
Open hardware and software

- open-ephys / plugin-GUI
- open-ephys / acquisition-board
- open-ephys / io-board
- open-ephys / headstage
Open hardware and software

Open-source Miniscopes

Wired

Wire-free

Cai et al., *Nature*, 2016
Shuman et al., *bioArxiv*, 2018

Denise Cai
Assistant Professor of Neuroscience
Mount Sinai
Large scale team science

The era of "brain observatories"

Allen Institute for Brain Science

Human Connectome Project (HCP),
(2012-2015), N = ~1,200

Healthy Brain Network (HBN),
(2016--), N = ~10,000

Adolescent Brain Cognitive Development,
(2016--), N = ~10,000

UK Biobank,
(2018--), N = ~500,000

Ariel Rokem
Senior Data Scientist
eScience Institute, University of Washington
Open framework for citizen science

Swipes for Science

https://github.com/SwipesForScience/SwipesForScience

braindrles.us  whaledr  appstract.pub  brainspot

Thanks eLife Innovation!
Different flavors of reproducibility

Parsing Reproducibility

“Empirical Reproducibility”

“Statistical Reproducibility”

“Computational Reproducibility”

V. Stodden, IMS Bulletin (2013)
Reproducible computing on HPC

Bridges-DL: Scalable AI for Open Research

Introducing NVIDIA Volta and DGX-2 to Bridges addresses the changing nature of research, building on Bridges’ strength in converged HPC, AI, and Big Data to provide research with an extraordinary platform for AI and AI-enabled simulation.

1 NVIDIA DGX-2: The most powerful AI system for the most complex AI challenges
- Couples 16 NVIDIA Volta 32GB SXM2 GPUs at 2.4 TB/s bisection bandwidth
- 81,920 CUDA cores and 10,240 tensor cores
- 2 Pf/s mixed-precision tensor performance
- 512 GB HBM2 memory
- 8 × 3.84 TB NVMe SSDs

9 HPE Apollo 6500 Gen10 servers: Balancing great AI capability and capacity
- Each couples 8 NVIDIA Volta GPUs with NVLink 2.0

Volta introduces Tensor Cores to accelerate neural networks, yielding extremely high peak performance for appropriate applications.

With 88 Voltas, Bridges-DL provides:
- 9.9 Pf/s mixed-precision tensor
- 251 Tf/s 32-bit
- 125 Tf/s 64-bit

Installation starting week!

Carnegie Mellon University
PSC
University of Pittsburgh

Paola Buitrago
Artificial Intelligence and Big Data Group Leader
Pittsburgh Supercomputing Center

Nick Nyström
Interim Director
Pittsburgh Supercomputing Center
National scale data collection and curation

NCI Cancer Research Data Commons (CRDC) - Concept

NCI Scope: “Create a data science infrastructure necessary to connect repositories, analytical tools, and knowledge bases”

Data commons co-locate data, storage and computing infrastructure with commonly used services, tools & apps for analyzing and sharing data to create an interoperable resource for the research community.*

National scale data collection and curation

Now all of SRA is taxonomically indexed!

[Image of a screen showing a web page with taxonomic analysis results.]

https://trace.ncbi.nlm.nih.gov/Traces/study/?stat_search=1561

Ben Busby
Lead, Bioinformatics Training
NCBI
National scale data collection and curation

ENCODE DCC

Submit

Production lab

Review

ENCODE DCC

Scientific community

Share

Genomic resources

Idan Gabdank
Senior Data Wrangler
ENCODE DCC
Evolving platforms by publishers

Bringing components together → Reproducible Document Stack

- Bring components together in a single narrative
  - Encapsulates usable code and data within the flow of a manuscript (for authors and readers)
  - Enhancement from static research article to full data and code interaction
  - Easy and accessible to everyone: connect Excel users with notebook coders

Supported by the publishing infrastructure
Pros and cons about pre-printing

A debate about preprinting in my lab

vaughn 8:53 AM
hi chris and caroline, so, now that the article is submitted, how do we feel about bioRxiv? You know I'm in favor, but I won't move forward unless we're unanimous

chris 11:25 AM
im in favor

caroline 12:32 PM
What do you see as the advantage of putting it on bioRxiv? Thus far I've put all my papers on bioRxiv and in my experience is that there are real drawbacks. 1. If there are any flaws that the reviewers catch, people have been reading the flawed version. 2. People pay less attention when the actual paper comes out. 3. (Less likely to be important) I've had difficulty where reporters wanted to write about paper, but the journal wouldn't let me talk to them.

If it gets closer to job application season, I would see the benefits as outweighing the disadvantages, but right now it seems to me that the disadvantages outweigh the advantages.

chris 12:52 PM
advantages: 1. community has a chance to comment/improve the manuscript. 2. public can read public-funded research without paywall 3. record of productivity for CV if reviews take too long

caroline 1:17 PM
I agree with #3, which is why I would support posting the paper later. Since Evolution Letters is entirely open access, #2 isn't currently a concern, so I would support posting to bioRxiv if we get a rejection from Evolution Letters and resubmit somewhere that isn't open access. #1 is valuable, but I think it's pretty rare. I am willing to be persuaded, but I'm not convinced.

vaughn 5:15 PM
On this conversation, which is really healthy. See: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005473
and: http://asapbio.org/preprint-info/preprint-faq
Recurring Themes
(and some favorite quotes)
Motivations and incentives

“There were microscopes that were too expensive for us ... why don’t we make it ourselves and make it open source?”

“Not just something to plug and play ... we want to constantly hack into our system and modify and make change for our specific research questions”

“...make tools more accessible for others, and others will contribute back”

“... but the open sources spirit caught on and bought a lot of the good will from the community”
Roles of funders and institutions

“Funders can encourage even force change in our behaviors”

“[NIH] should really make part of the [R01] renewal decisions on how much data is reused”

“Conversation [on reward system in terms of funding] has not formally happened.”

“Science always moves faster than evaluation processes, its going to change when some of us are on the other side.”
Challenges

“Data arriving at unprecedented volume, variety and velocity. Need new tools and approaches to process, analyze and interpret”

“Technology is ever evolving, need to be full time to learn all theses”

“People develop hard ware and code, but not necessarily have time to share”

“How to get grad students and postdocs the credit they need to get tenure track jobs”

“Till we get to a point how to measure the value /reuse of data, we won’t have incentives in place”
Reproducibility

“[A research project] is like a big iceberg, the tip is what’s shown, the whole thing is all the work that happens.”

“Grabbing data via point and click is not reproducible”

“With a few lines of code, get the data …reproducibly. If data changes, just have to re-run the code”
Value of Data

“Sharing data is not useful, reusing data is useful.”

“Open access ≠ useful resources… implementation is not FAIR”

“To make data valuable is work. It’s hard to do it well.”
“Is it worth spending the time? Yes, otherwise it’s not useful”

“Maintain balance of asking for enough metadata but not too much”

“Metadata standards…vary a lot by discipline. How discipline specific should we make it so it’s useful to genomics community but also useful for everyone else?”

“As a community we shouldn’t tolerate another data format. We need things as open, machine readable, and low level as we can possibly get”
Data ownership and privacy

“Who owns data, code, and research outputs? When, if ever, do they expire?”

“We know it's very valuable, but frankly we don’t know who owns it”

“The traditional Form of Consent comes with lots of challenges… because we don’t know how these data will be linked in the future to other data”
Moving Forward

• New collaborations among researchers

• Partnership among various stakeholders

• More open science events
Carnegie Mellon University

Nov 27-28, 2018

Day 1: 9:00 AM - 5:00 PM Day 2: 9:00 AM - 4:00 PM

Instructors: Dr. Rebecca Lowdon, Nathan Catlin

Helpers: Huajin Wang, Eric Kaltman
Innovation Roundtable 2018

Innovation Roundtable 2018: Collaboration Yields Powerful Research Results

- Monday, December 3, 2018
- 10 a.m. – 11 a.m.
- Cohon Campus Center, Rangos 3
- Register for this event: https://cmu-innovation-roundtable-2018.eventbrite.com

Presented by Carnegie Mellon University Libraries and the Software Engineering Institute, the Innovation Roundtable brings together CMU researchers to discuss how innovation and problem solving are enhanced through cross-disciplinary collaboration.
• Automation in data discovery
• Automation in data curation and generation
• Measuring and improving data quality
• Integrating datasets and enabling interoperability
• Biomedical data discovery and reuse
• The future of scientific data and how we work together

https://events.library.cmu.edu/aidr2019/
#AIDR2019
Carnegie Mellon University

kgw@cmu.edu
huajinw@cmu.edu

cmkeithw
HuajinBioLib