

Kicking On:



Brian Flaherty

Growing Data Services at NeSI





Understanding the eResearch Ecosystem in New Zealand



Workshop Reflection Report
eResearch NZ 2018 Conference
Queenstown, New Zealand

Broadening computational capabilities & skills

Research

Partnering with Science

ID & Auth2
/ IAM,
Cyber
security

Service &
Resource
Governance

Data, Methods, Models, Code

Research Platforms, Virtual Labs, Scientific
Gateways

Research Artifact
Management

Analysis & Modelling
Tools

Storage Infrastructure

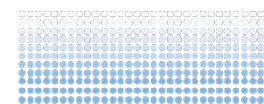
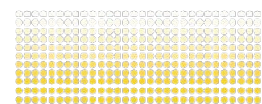
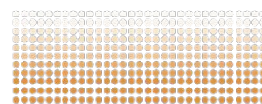
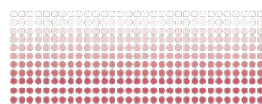
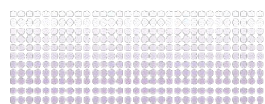
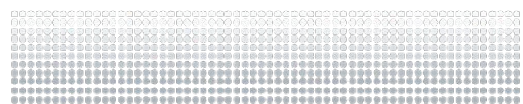
Compute Infrastructure

Advice,
Training,
Outreach

Research
Software
& Data
Engineeri
ng

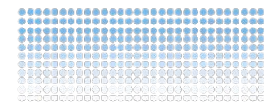
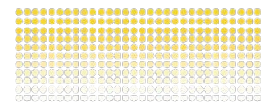
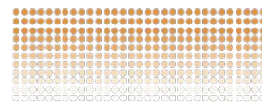
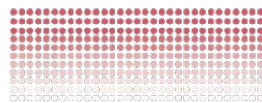
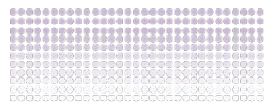
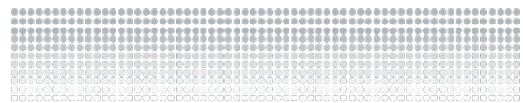
Network Infrastructure

Data Centre



Community Consultation Findings Report 7th September Update

New Zealand eScience Infrastructure



Data Transfer & Sharing



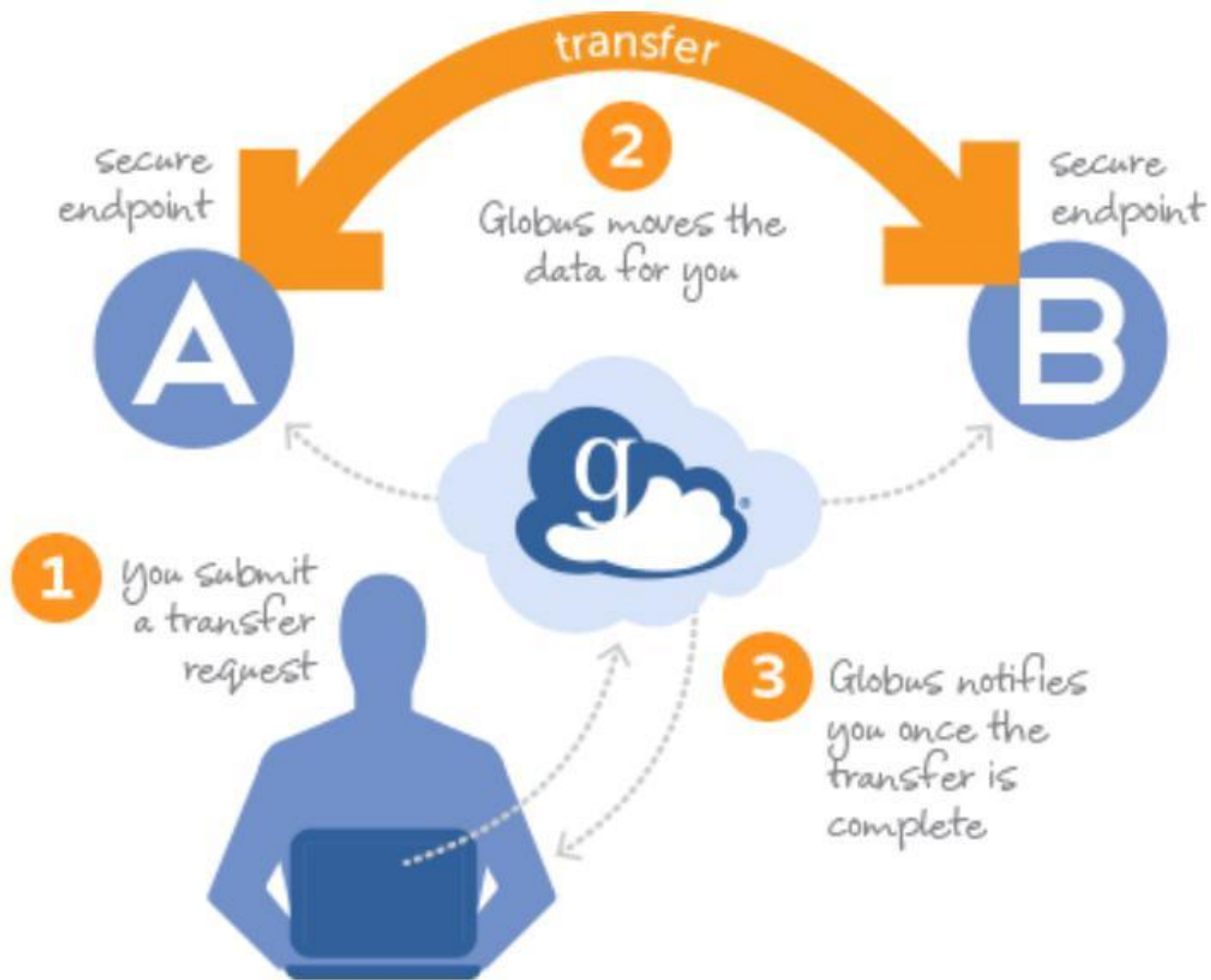
CLOSED



SHARED



OPEN



globus

≡

File Manager

File Manager

RECENTLY USED

Brian F NeSI laptop

NeSI Wellington DTN B...

RDCEP Publish

Published Globus Data

PINNED BOOKMARKS

You have no pinned bookmarks

Bookmark Manager

Activity

Endpoints

Publish

Groups

Console

Account
brianflaherty@globusid.org

Help

Globus Home

File Manager

NeSI Wellington DTN Benchmarks

Collection

Path

/

select none up one folder refresh list columns

10G.dat01/16/2019 11:36am

50GB-in-medium-files02/01/2019 12:40pm

5GB-in-small-files01/16/2019 11:36am

big_file01.out11/29/2018 12:50pm

big_file02.out11/29/2018 12:55pm

big_file03.out11/30/2018 11:45am

big_file04.out11/30/2018 12:40pm

big_file05.out01/11/2019 04:12pm

Brian test01/30/2019 01:09pm

DanS01/17/2019 03:00pm

Davemac02/01/2019 11:29am

Gareth02/01/2019 10:01am

Jun-Test01/30/2019 01:09pm

Martin_UoA02/01/2019 12:52pm

michael01/17/2019 10:37am

to_directory01/22/2019 12:38pm

Share

Transfer or Sync to...

New Folder

Rename

Delete Selected

Preview (limited)

Download (https)

Open (https)

Get Link

Show Hidden Items

Deactivate

big_file03.out01/22/2019 03:18pm10.48 GB

clhi_Amon_MPI-ESM-LR_piControl_r1i1p1_282...01/23/2019 05:00pm633.51 KB

data01/23/2019 02:59pm—

img0.jpg01/25/2019 09:34am53.19 KB

NeSI staff ID photo.JPG01/21/2019 02:00pm7.34 MB

NeSI-2017-Annual-Review-FULL.pdf01/29/2019 01:51pm35.66 MB

Start

Transfer & Sync Options

Start

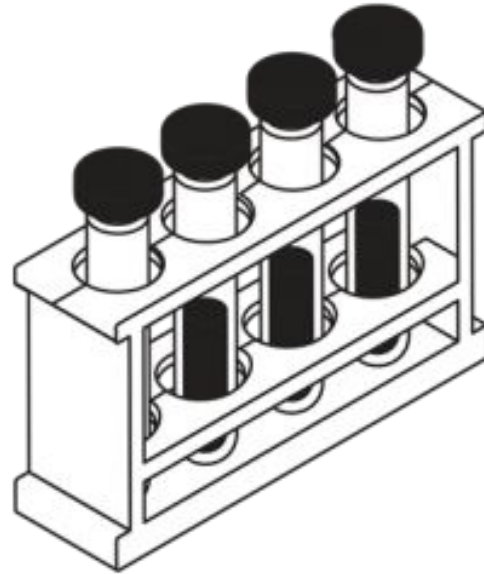
Panels

Bookmark Manager

Brian F NeSI laptop

/~/Data Test/

VIRTUAL LAB

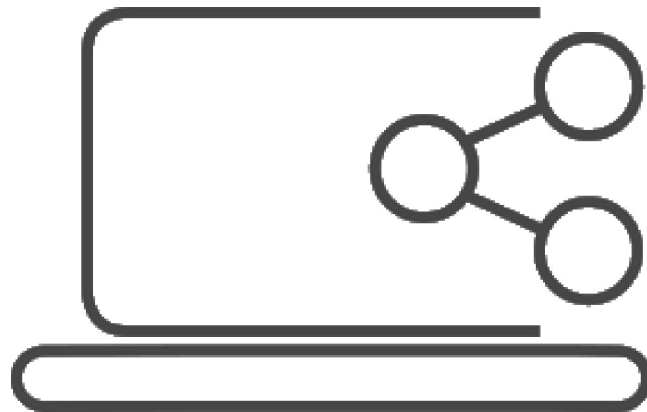




VIRTUAL LAB for Genomics training

cutdapt
samtools
bwa
fastqc
bedtools
Hisat2
MultiQC
subread (featureCounts)
STAR
IGV
SeqMonk

DATA REPOSITORIES, DATA PUBLISHING & DISCOVERY



*“NeSI is a national level organisation, there is a great opportunity for NeSI to push for higher data standards for researchers overall, by advocating **FAIR Data Principle** (Findable, accessible, interoperable and reusable)”*

NeSI Community Consultation Findings Report

RESEARCH DATA PORTAL



Harmonized Cancer Datasets

Genomic Data Commons Data Portal

Get Started by Exploring:



Projects



Exploration



Analysis



Repository

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary

[Data Release 14.0 - December 18, 2018](#)

PROJECTS

43

PRIMARY SITES

68

CASES

33,096

FILES

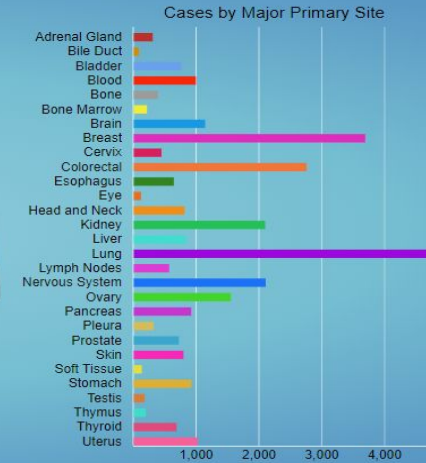
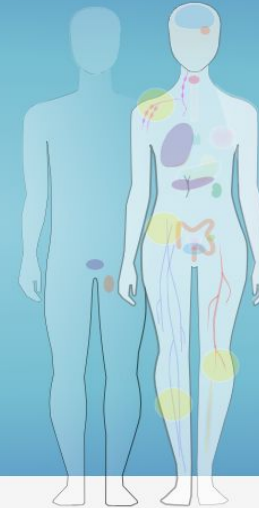
358,679

GENES

22,872

MUTATIONS

3,142,246



GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:



[Data Portal](#)



[Website](#)



[Data Transfer Tool](#)



[API](#)



[Data Submission Portal](#)




[Documentation](#)



[Legacy Archive](#)

International Cancer Genome Consortium

 **ICGC** Data Portal

[Cancer Projects](#) [Advanced Search](#) [Data Analysis](#) [DCC Data Releases](#) [Data Repositories](#)

Cancer genomics data sets visualization, analysis and download.

[Search](#)

e.g. BRAF, KRAS G12D, DO35100, MU7870, F1998, apoptosis, Cancer Gene Census, imatinib, GO:0016049


Advanced Search

[By donors](#) [By genes](#) [By mutations](#)

Data Release 27 April 30th, 2018



Cancer projects	84
Cancer primary sites	22
Donor with molecular data in DCC	20,487
Total Donors	24,077
Simple somatic mutations	77,462,290

[Download Release](#)

 **PCAWG**
PanCancer Analysis
OF WHOLE GENOMES

Browse more than 2800
harmonized whole genomes

The Pan-Cancer Analysis of Whole Genomes is an international collaboration from over 700 scientists to identify common mutation patterns across more than 60

Instant analysis with cloud
computing

Getting headaches spending days or weeks downloading data? Instead, upload your analytic software to our Collaboratory cloud or Amazon Web Services and get

Modern Research Data Portal

It's how research data management is done!

This simple web application demonstrates use of the Globus platform. Login/logout and authorization for all API calls is handled by the Globus Auth service. The application mimics how a portal may be used by researchers to browse, download, transfer, and analyze datasets. After logging in, click [TRANSFER](#) to view a list of sample datasets; click [GRAPH](#) to generate graphs for a dataset; or click [PROFILE](#) to update your application profile.

Modern Research Data Portal Resources

Source code, examples, papers, and other useful documentation.

Globus Transfer API

API reference for transfer and sharing functions.

Globus Auth API

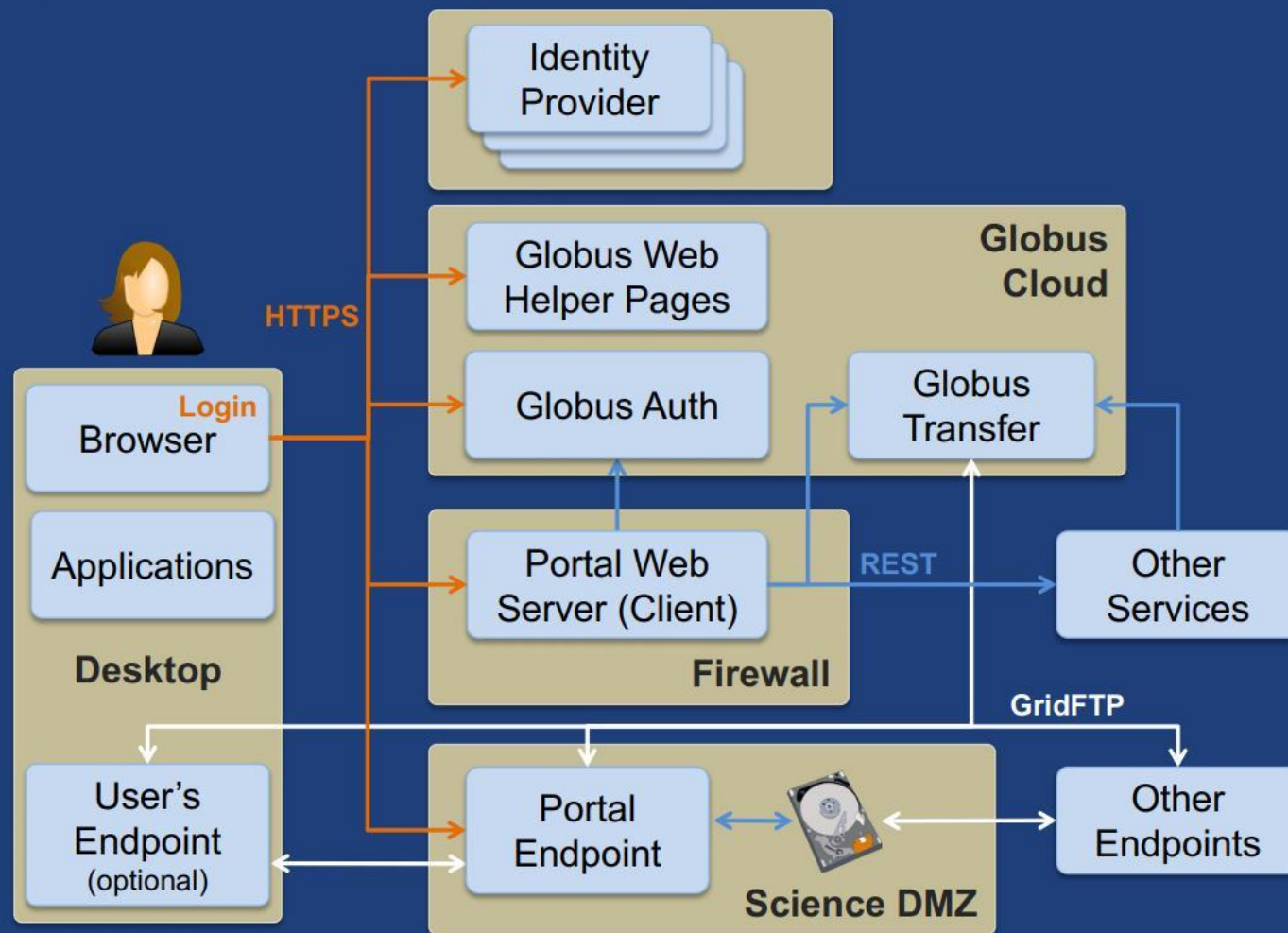
API reference for authentication and authorization.

[GLOBUS DOCUMENTATION →](#)





Prototypical research data portal



Data sovereignty

...to manage data storage and access within a Māori values context, something that is different from the standard 'public repository' or 'open access' philosophy.

...approved access by researchers, data curation and improved connectivity to tangata whenua will be hallmarks of practice.



Te Mata Ira

Guidelines for Genomic Research with Māori

Mr Maui Hudson
Dr Angela Beaton
Mrs Moe Milne
Dr Waiora Port
Prof Khyla Russell

Dr Barry Smith
Dr Valmaine Toki
Lynley Uerata
Dr Phil Wilcox

October 2016

Welcome to the DMPTool

Create data management plans that meet institutional and funder requirements.

[Get started](#)

DMPTool by the Numbers


35,203
Users


31,916
Plans
[More](#)


245
Participating Institutions
[More](#)

Top Templates

Digital Curation Centre
NSF-BIO: Biological Sciences
NIH-GEN: Generic
Arctic Data Center: NSF Polar Programs [DRAFT]
NSF-SBE: Social, Behavioral, Economic Sciences

[More](#)

DMPTool News

Prettier and mobile-ready(-er)

[Go to the blog](#)

 [RSS](#)

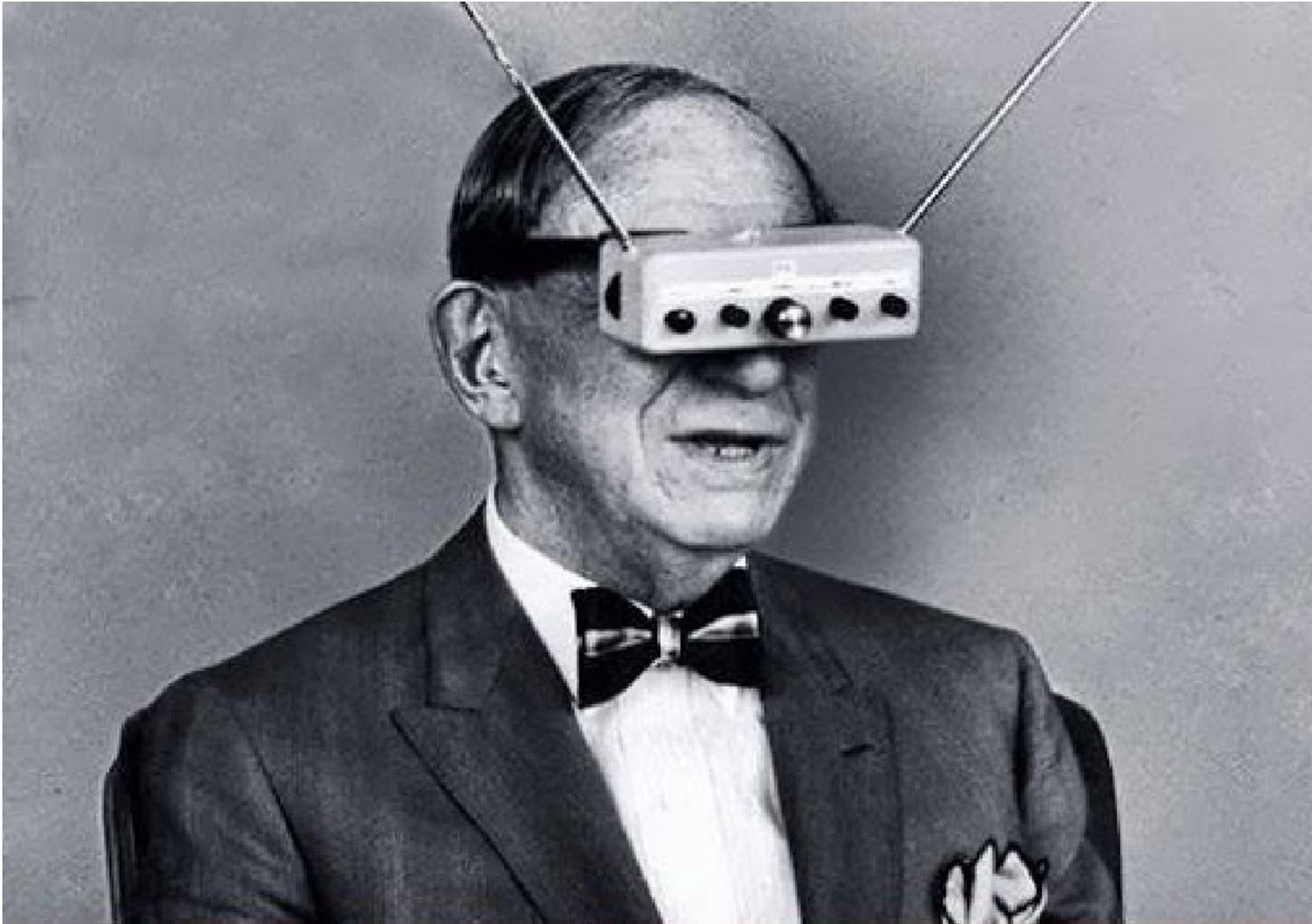
[About](#) [Terms of use](#) [Privacy statement](#) [Accessibility](#) [Github](#) [Contact us](#)

the big(ger) picture



Discovery



Like all the men of the Library, in my younger days
I traveled; I have journeyed in quest of a book,
perhaps **the catalog of catalogs**

Jorge Luis Borges, The Library of Babel

DATA

eResearch
2020



**National
Research
Data**
Programme

The Case for
Research Data

Find data for research

Find, access, and re-use data for research - from over one hundred Australian research organisations, government agencies, and cultural institutions

All Fields ▾ Search for Data

Q Search

☐ Publicly accessible online

Advanced Search Map Search

Browse By Subjects



Humanities and Social
Sciences



Business, Economics and
Law



Medical and Health Sciences



Engineering, Computing and
Technology



Built Environment and
Design



Biological Sciences



Agricultural and Veterinary
Sciences



Environmental Sciences



Earth Sciences



Physical, Chemical and
Mathematical Sciences

NZ RIS Data Principles

1. Provide a **system-wide view of research**, science and innovation information.
2. Ensure open data which is easily accessible and widely used.
3. Protect personal and commercially sensitive data.
4. Enable the re-use of data.
5. Reduce collection and reporting burden.
6. Ensure data is trusted, authoritative and well-managed.
7. Enable easy and automatic movement of data between systems.

(meta)data

capture the meaning of data,
not just data itself

Our researchers don't design their data to be findable, let alone searchable.

Our institutions, and indeed our research funders, do not have a consistent approach to data management or sharing meta-data.

Our research communities do not have agreed meta-data standards for creating, collecting or sharing new data

Data Repositories



- Subjects
- Content Types
- Countries
 - International (1)
 - Italy (1)
 - New Zealand (9)
 - Taiwan, Province of China (1)
 - United States (1)
- AID systems
- API
- Data access
- Data access restrictions
- Database access
- Database licenses
- Data licenses
- Data upload
- Data upload restrictions
- Enhanced publication
- Institution responsibility type
- Institution type
- Keywords
- Metadata standards
- PID systems
- Provider types
- Quality management
- Repository languages
- Software
- Repository types
- Versioning

← Previous 1 Next →

Sort by ▼

Found 9 result(s)

Landcare Research Data Repository



Datastore Landcare Research Manaaki Whenua

Subject(s)

Natural Sciences Agriculture, Forestry, Horticulture and Veterinary Medicine Life Sciences

Geosciences (including Geography) Agriculture, Forestry, Horticulture and Veterinary Medicine Biology

Content type(s)

other Plain text Archived data Images Raw data Scientific and statistical data formats

Country

New Zealand

The Landcare Research DataStore ('the DataStore') is the general data catalogue and repository for Environmental Research Data from Landcare Research. Much of Landcare Research's research data is available through specific web pages, but many datasets sit outside these areas. This new data repository provides a mechanism for our staff to deposit and document this wider range of datasets so that they may be discovered and potentially re-used.

Pacific Islands Families Study



PIF Study

Subject(s)

Empirical Social Research Public Health, Health Services Research, Social Medicine Pediatric and Adolescent Medicine

Non-European Languages and Cultures, Social and Cultural Anthropology, Jewish Studies and Religious Studies Humanities

Humanities and Social Sciences Social Sciences Social and Behavioural Sciences Medicine Medicine Life Sciences

Social and Cultural Anthropology and Ethnology/Folklore

Content type(s)

Standard office documents

Country

New Zealand

The Pacific Islands Families (PIF) Study is an ongoing longitudinal birth cohort study that has been tracking the health and development of 1,398



Browsing University of Waikato Research by Type

0-9 [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

Now showing items 1-17 of 17



Type

[Authored Book](#) [9]

[Book](#) [10]

[Chapter in Book](#) [278]

[Commissioned Report for External Body](#) [155]

[Composition](#) [19]

[Conference Contribution](#) [1335]

[Discussion Paper](#) [6]

[Internet Publication](#) [7]

[Journal Article](#) [5492]

[Oral Presentation](#) [4]

[Other](#) [32]

[Performance](#) [2]

[Report](#) [78]

[Software](#) [1]

[Technical Report](#) [62]

[Thesis](#) [23]

[Working Paper](#) [693]



- ☒ Search Research Commons
- ☐ Search this Community

USAGE STATISTICS

For this community

For all of Research Commons

KiltHub

Discover research from **Carnegie Mellon University** ▼

 **+ Follow**



help? ✕

sort **Relevant** ▼

type **ANY** ▲

licence **ANY** ▼

+ Follow this search

- ☐ Collections
- ☐ Items
- ☐ Projects
- Item type
- ☐ Figure
- ☐ Media
- ☐ Dataset
- ☐ Fileset
- ☐ Poster
- ☐ Paper
- ☐ Preprint
- ☐ Presentation
- ☐ Thesis
- ☐ Code
- ☐ Book



High-throughput mechanotransduction in Drosophila...
Ardon Shorr ▼ today



Robot Design for Everyone: Computational Tools that...
Ruta Desai



Session 1, Opening Remarks, Discussion
Serge ▼ 09/02/2019



Wet Lab to Web Lab
Anisha Keshavan 07/02/2019



Preprints and Open Data in the Life Sciences
Vaughn Cooper 07/02/2019



Open Ephys: Open-Source Tools for Electrophysiology
Josh Siegle 07/02/2019



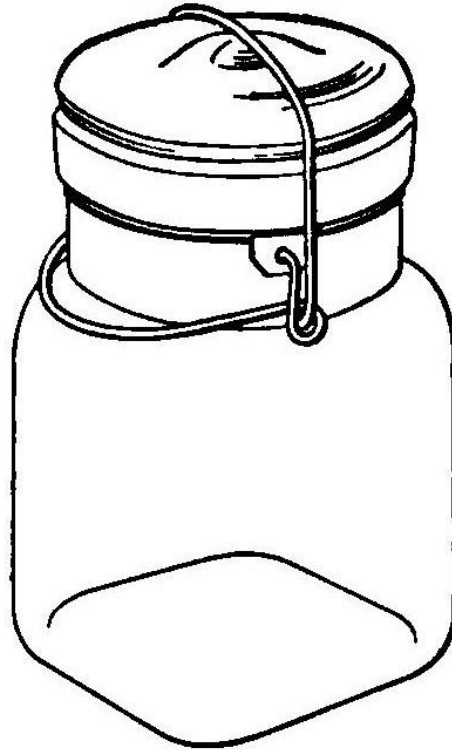
Miniscope Project: Developing an Open-Source Platform for Neurote...
Denise Cai 07/02/2019



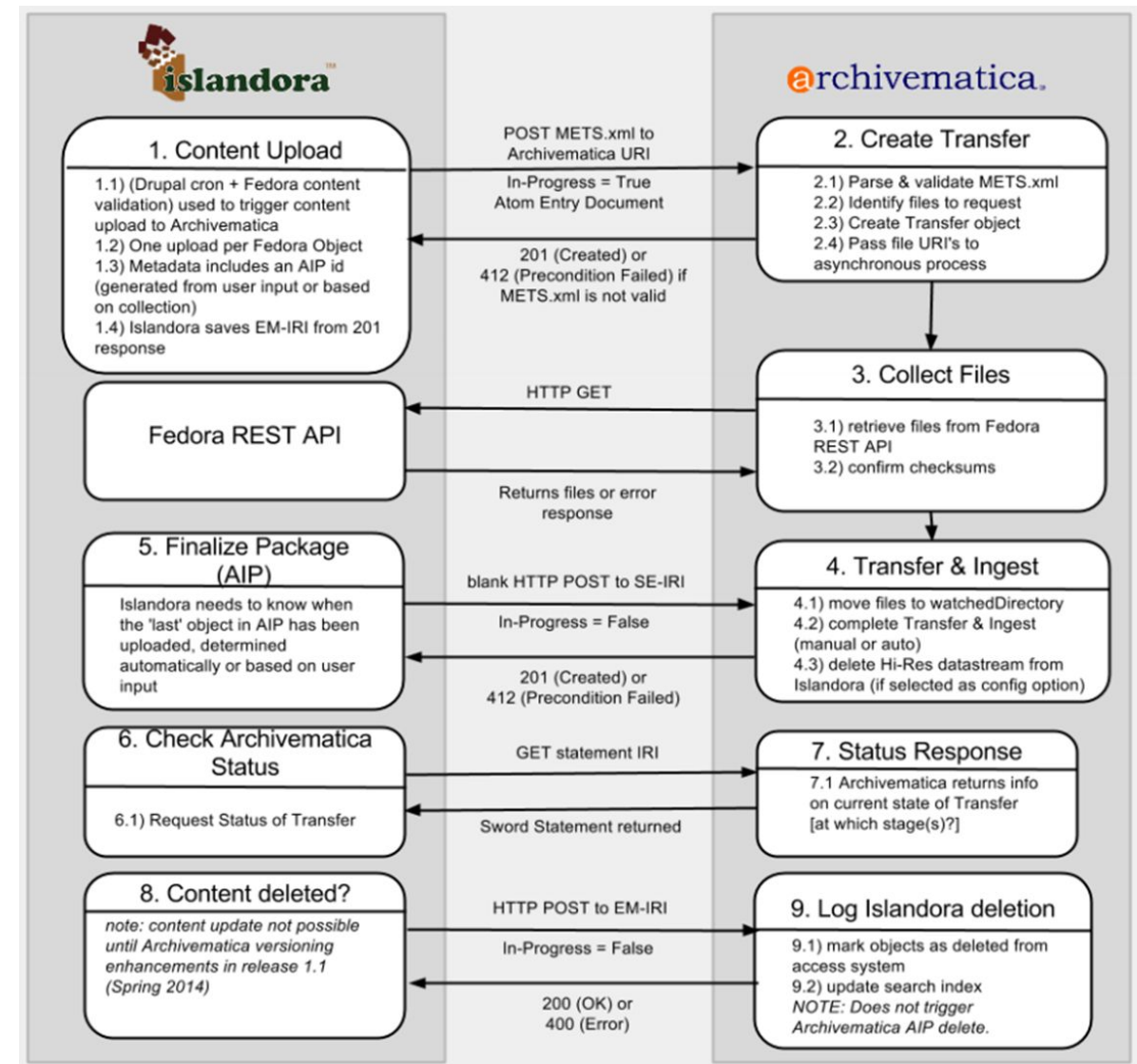
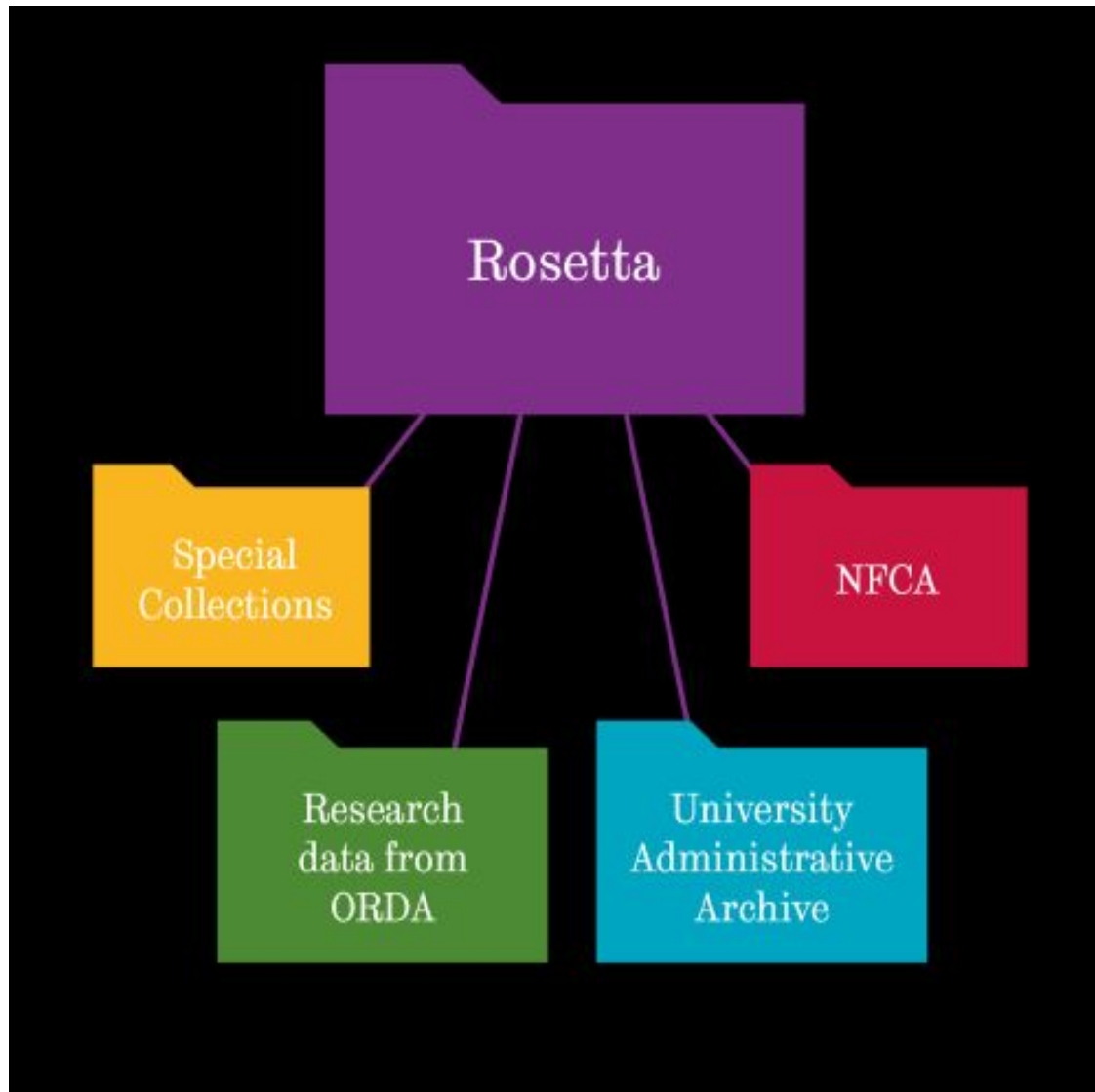
Open Science Symposium 2018 Program
Melanie Gainey ▼ 07/02/2019



preservation



Connecting Data Repositories to Preservation systems





Citing and Archiving Research.

A Baden-Württemberg state project to develop infrastructure to support computational science

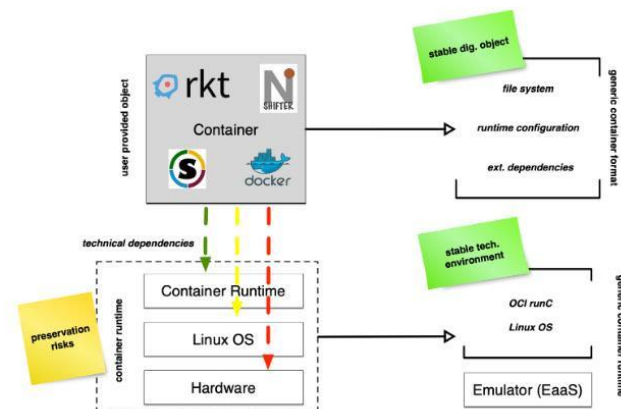
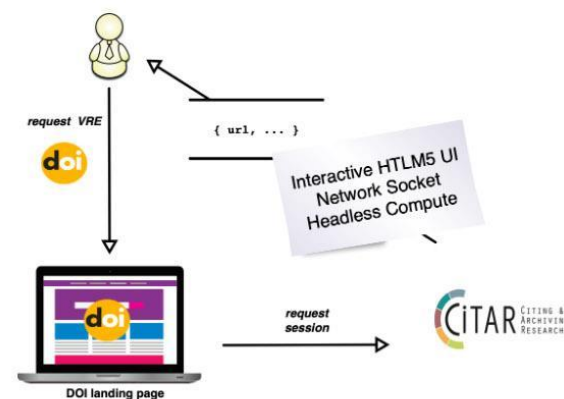
• repeat • replicate • reproduce • reuse

Manage, cite and preserve Virtual Research Environments (VRE)

Support software-based research, CiTAR provides workflows to

- automated import virtual machines and containers
- assign persistent identifiers
- re-run virtual machines and containers with external data
- archival and long-term preservation

[Use case and demos ...](#)



Long term preservation of containers

- ingest of a container from a productive environment (virtual research environment)
- conversion into an archivable container format (currently available for Docker and Singularity)
- test run in a virtualized environment (bwCloud)
- documentation of dependencies (hardware, network)
- documentation of input/output parameters

The focus of maintenance planning is on the container runtime environment.

[Use case and demos ...](#)

DPaaS

Digital Preservation as a Service



NeSI @ eResearch NZ - Talks & Workshops:



Monday 18 Feb

2:10 - 2:30 pm

Understanding research drivers for NZ's advanced research computing

2:30 - 2:50 pm

How NeSI helps Manaaki Whenua - Landcare Research monitor land cover changes

3:30 - 3:50 pm

NeSI Futures

4:30 - 5:30 pm

Training Community BoF

4:50 - 5:10 pm

Catering to domain (Genomics) specific eResearch needs

Tuesday 19 Feb

11:00 - 11:20 am

The NeSI HPC Computer and Data Analytics Service

11:00 am - 12:30 pm

Open Space Session - BYO topics!

1:30 - 1:50 pm

Visualization capabilities of NeSI's new high performance computers

1:30 - 1:50 pm

A day in the life of NeSI's Apps Support

1:50 - 2:10 pm

NeSI and your data: Scalable storage

1:50 - 2:10 pm

Research Software Engineering (RSE): What's in a name?

Tuesday 19 Feb (cont.)

2:10 - 2:30 pm

Scaling new data services at NeSI

2:30 - 2:50 pm

Insight into the new NeSI platforms

3:30 - 4:30 pm

(Inter)national collaborative research infrastructure strategies BoF

3:30 - 4:30 pm

Research Software Engineering BoF

4:30 - 5:30 pm

Research Cloud NZ BoF

Wednesday 20 Feb

11:10 am - 4:00 pm

Hacky Hour / Bring your own code