

# Achieving Reproducibility: Perspectives and New Tools

Professor Brian Wandell

Director, Stanford Center for Cognitive and  
Neurobiological Imaging (CNI)

Deputy Director, Wu Tsai Neurosciences Institute

QUANTITATIVE MEASUREMENTS

∞

COMPUTATIONAL MODELS

∞

CHECK AND SHARE

Thanks to the  
**Simons Foundation** and the  
**Arnold Foundation**  
for their support

# Commercial disclosure

I am an active participant in Flywheel Exchange LLC, a commercial venture related to the work described here.

**FLYWHEEL**

[HOME](#) [PRODUCT](#) [SOLUTIONS](#) [CUSTOMERS](#) [PARTNERS](#) [COMPANY](#)

[SCHEDULE A DEMO](#)

## Cloud-Scale Collaborative Science

Unleash discovery with comprehensive data management, computational analysis and secure collaboration.

For: [Core Facilities](#) [Imaging Research](#) [Clinical Research](#) [Multi-Center Studies](#)



FLYWHEEL ON THE ROAD

Catch Flywheel at ASFNR Oct. 15th-17th & SFN Nov. 3rd-6th

# Early days of MRI - hardware oriented, little ability to do much with data management

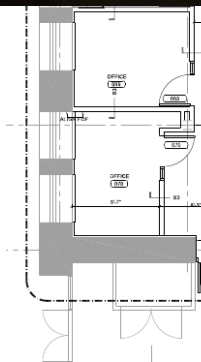
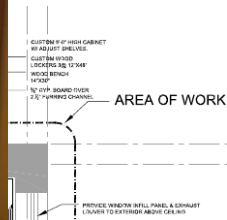
1992

- MRI acquisition methods have become more complex
- MRI computations (reconstruction, data analysis) have become more complex
- Networking and computer technology have advanced



# Implementing our MRI center made us think about data management

2008



## **Motivation**

Help people check and share their work to  
advance scientific understanding together

# Replication vs. Computational reproducibility

## **Replication**

Obtaining the data again, usually by independent investigators using similar methods, equipment and protocols

We can't help people do this

## **Computational reproducibility**

Confirm the calculations (e.g., statistics, images, numerical relationships) calculated from existing data

We can help people do this



Dave  
Donoho

Computational reproducibility ... must be designed into a project from the beginning. **One does need to develop a whole set of programming and research disciplines** with the end result in mind and stick with them.

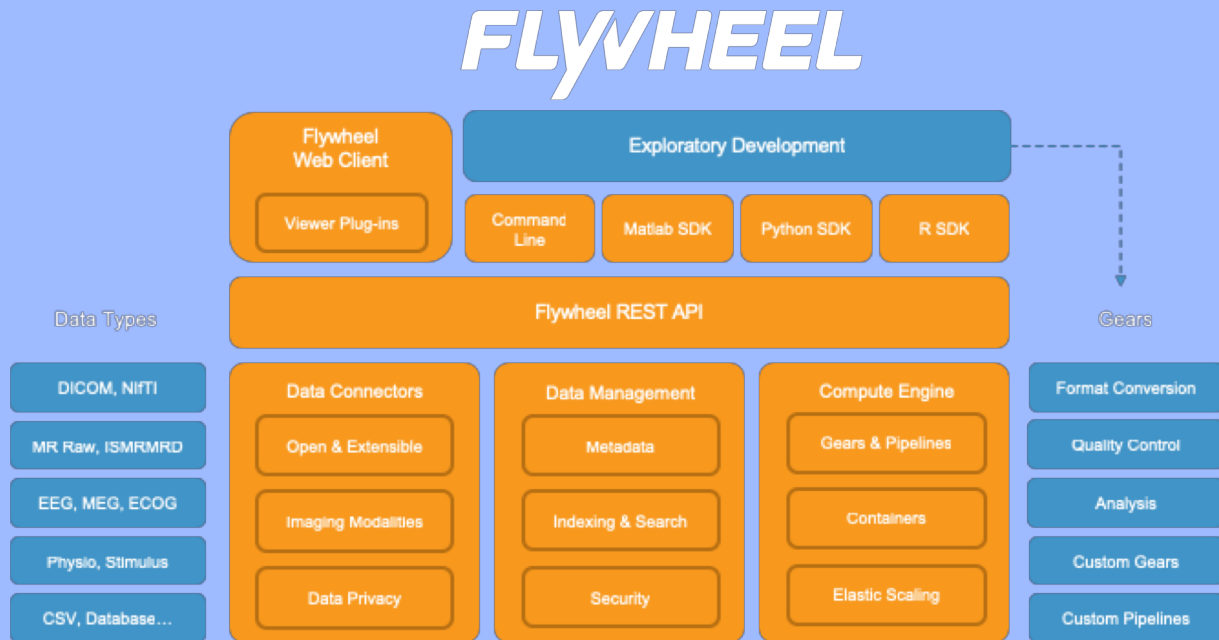
# Computational reproducibility: The MRI Center perspective

- **User needs**

- Reproducible data and computation
- Tools, not rules

- **What Flywheel built**

- Acquisition, metadata, visualization
- User-rights
- Search and data reuse
- Embedded Gears
- Collaborating



# Data management is a pre-requisite to computational management are both important

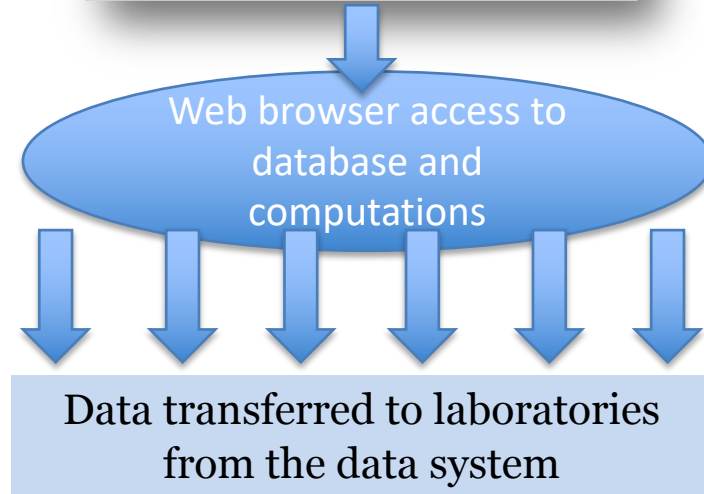
- Most MRI Centers provide one of these data retrieval options from the MRI scanner
  - Copy data to CD, DVDs, USB hard drive
  - Copy to a server and remote login
- Data and metadata are transferred to a system controlled by a student or post-doc
- **Limitations** – reuse and sharing become burdensome; metadata and pre-processing information are frequently lost



Data transferred to laboratories  
from the scanner

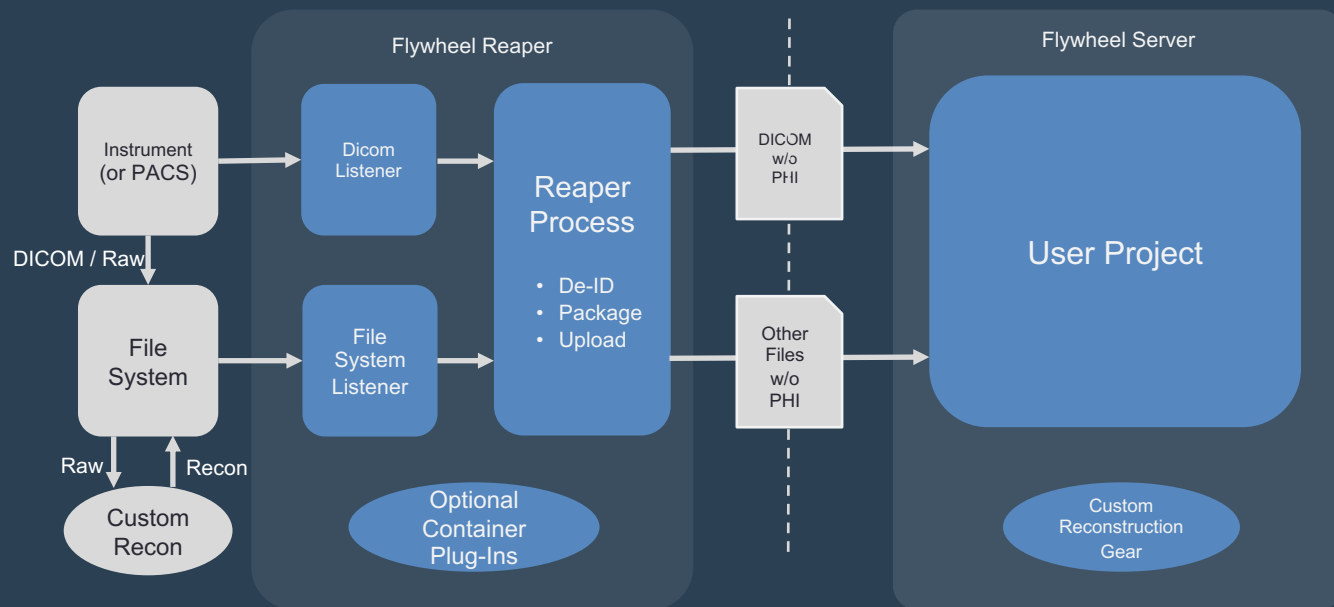
# Data management is a pre-requisite to computational management are both important

- Archiving MRI Center data eliminates the need for users to gather the data again for publication
- The data should be available through a platform-independent web browser to simplify access
- Basic tools, such as search, visualization, and pre-processing can be available through the browser
- The data are ready for sharing and reuse; metadata can be stored; pre-processing methods shared can be shared



# Data Capture Architecture

- Data capture must classify and interpret many different MR data types
- Users ask to add legacy data, or to include PACS (medical) data
- Adding behavioral and physiological data is valuable for reproducibility
- Permitting custom MRI reconstruction methods has been important at several MRI Centers



# Tools are better than rules

- Users have different preferences about to prepare their data (pre-process) for detailed analysis
- Project members change over time, so user rights must be easy to manage
- Many users seek help with pre-processing, managing project notes, and data visualization
- The Center and users have a shared interest in Quality Assurance (QA)
- Tools to help upload to an archival site (e.g., NDA, Journal) simplifies grant and publication compliance

The image displays three overlapping screenshots of the FLYWHEEL web application interface. The top screenshot shows the 'Rules' management page, where users can configure 'Gear Rules' for file classification (e.g., 'P-File Classification', 'DICOM Classification', 'MUX Recon') and view an 'About Gear Rules' section. The middle screenshot shows the 'Users' management page, featuring a table of users with columns for Name, Email, and Role, and a form to 'Add Existing User'. The bottom screenshot shows a 'Visualization' window with three MRI brain scan images: Axial, Coronal, and Sagittal views, each with technical metadata like 'img (128/256)', '176 x 240', and 'Zoom: 229%'.

Rules

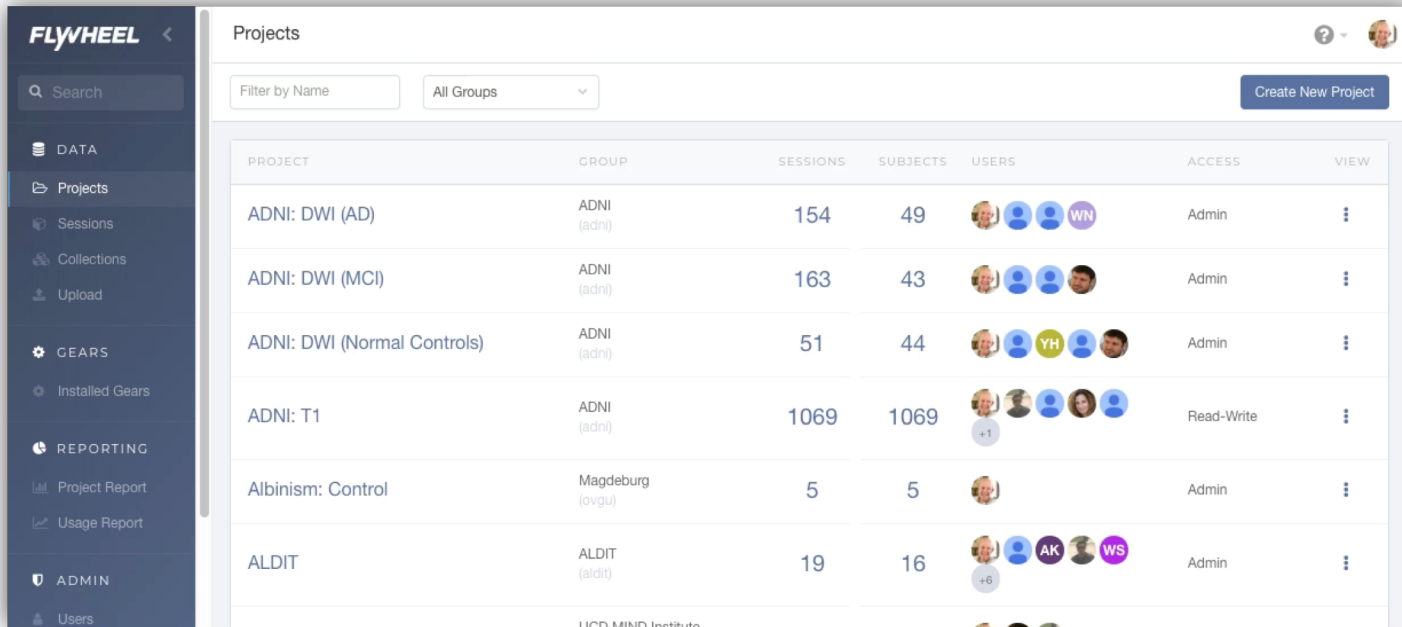
Users

Visualization

# Project overview – Background, Users, Pre-processing rules



- Information about each project is organized for sharing and editing
- Users can be added or removed from the project with different permission levels
- Notes, papers, rules for processing the data can be specified for each project



The screenshot shows the FLYWHEEL web application interface. On the left is a dark blue sidebar with navigation options: DATA (with sub-items: Projects, Sessions, Collections, Upload), GEARS (with sub-item: Installed Gears), REPORTING (with sub-items: Project Report, Usage Report), and ADMIN (with sub-item: Users). The main content area is titled "Projects" and features a search bar, a "Filter by Name" input, a "All Groups" dropdown menu, and a "Create New Project" button. Below this is a table listing various projects with columns for Project Name, Group, Sessions, Subjects, Users, Access, and View options.

PROJECT	GROUP	SESSIONS	SUBJECTS	USERS	ACCESS	VIEW
ADNI: DWI (AD)	ADNI (adni)	154	49	WN	Admin	
ADNI: DWI (MCI)	ADNI (adni)	163	43		Admin	
ADNI: DWI (Normal Controls)	ADNI (adni)	51	44	YH	Admin	
ADNI: T1	ADNI (adni)	1069	1069	+1	Read-Write	
Albinism: Control	Magdeburg (ovgu)	5	5		Admin	
ALDIT	ALDIT (aldit)	19	16	AK WS +6	Admin	

# Providing users with tools



- Many types of calculations can be installed and shared between users
- Here is a list of the 'Gears' on my lab's site that I share with Grill-Spector
- There are dozens, including, FreeSurfer, FSL, HCP, tractography, file format (DCM), directory format (BIDS), and ...

The screenshot shows the FLYWHEEL web application interface. The left sidebar contains navigation menus for DATA, GEARs, REPORTING, and ADMIN. The main content area displays a table of projects with columns for PROJECT, GROUP, SESSIONS, SUBJECTS, USERS, ACCESS, and VIEW. The table lists various projects such as ADNI: DWI (AD), ADNI: DWI (MCI), and Albinism: Control, along with their respective session counts, subject counts, user avatars, and access levels.

PROJECT	GROUP	SESSIONS	SUBJECTS	USERS	ACCESS	VIEW
ADNI: DWI (AD)	ADNI (adni)	154	49		Admin	
ADNI: DWI (MCI)	ADNI (adni)	163	43		Admin	
ADNI: DWI (Normal Controls)	ADNI (adni)	51	44		Admin	
ADNI: T1	ADNI (adni)	1069	1069		Read-Write	
Albinism: Control	Magdeburg (ovgu)	5	5		Admin	
ALDIT	ALDIT (aldit)	19	16		Admin	
Autism Phenome	UCD MIND Institute (mind)	625	625		Admin	
BCBL_ILLITERATES	Wandell Lab (wandell)	92	73		Read-Write	
Brain Beats	Wandell Lab (wandell)	8	8		Admin	

# Computational record keeping - Gears and their parameters



- The Gears you used and their parameters are stored in the Provenance tab
- Here is a FreeSurfer run

FLYWHEEL	
Search	
DATA	
Projects	
Sessions	
Collections	
Upload	
GEARS	
Installed Gears	
REPORTING	
Project Report	
Usage Report	
ADMIN	
Users	
Groups	
Access Log	
Gear Rule Templates	
SOC ECoG (Hermes)	Wandell Lab (wandell) 5 5  Admin
SVIP (Philips)	Simons Foundation (simons) 139 139  Admin
SVIP Released Data (SIEMENS)	Simons Foundation (simons) 283 281  Admin
SVIP: Unreleased Data	Simons Foundation (simons) 81 81  Admin
TBI: NeuroCor	Palo Alto VA (vapa) 5 5  Admin
Templates Adult	Templates (template) 4 4  Admin
Templates Macaque	Templates (template) 1 1  Admin
Templates Neonatal	Templates (template) 12 12  Admin
Templates Pediatric	Templates (template) 6 6  Admin
UMN	John Day Lab (jwday) 164 158  Admin
Velscope	Oral Eye (oraleye) 3 3  Admin
	Wandell Lab

# Taking advantage of cloud-scale computing



- **Cloud-scale** means you can select and run many Gears on different data sets at the same time
- The cloud providers sets up the machines upon request; they have a lot

A screenshot of the FLYWHEEL web application interface. The left sidebar is dark blue with the "FLYWHEEL" logo at the top. It contains a search bar and several menu categories: DATA (with sub-items: Projects, Sessions, Collections, Upload), GEARS (with sub-item: Installed Gears), REPORTING (with sub-items: Project Report, Usage Report), and ADMIN (with sub-items: Users, Groups, Access Log). The main content area is white and shows the details for a project named "ADNI: T1". At the top, it displays statistics: 1069 Sessions, 0 Flagged, 1069 Subjects, and 1 Attachments. Below this is a navigation bar with tabs for Information, Permissions, Attachments, Template, Rules, and Analyses. The "Information" tab is active. The main content area is divided into sections: "Description" (with a "Markdown Supported" label and an "Edit" button), "Custom Information" (with "Save", "Cancel", and "+" buttons, and a message "No custom fields added yet."), and "Notes". On the right side, there is a "Quick Tip" box with text explaining project descriptions and notes. At the top right of the main content area, there are buttons for "View Data" and "Download Project".

# Taking advantage of cloud-scale computing



Video

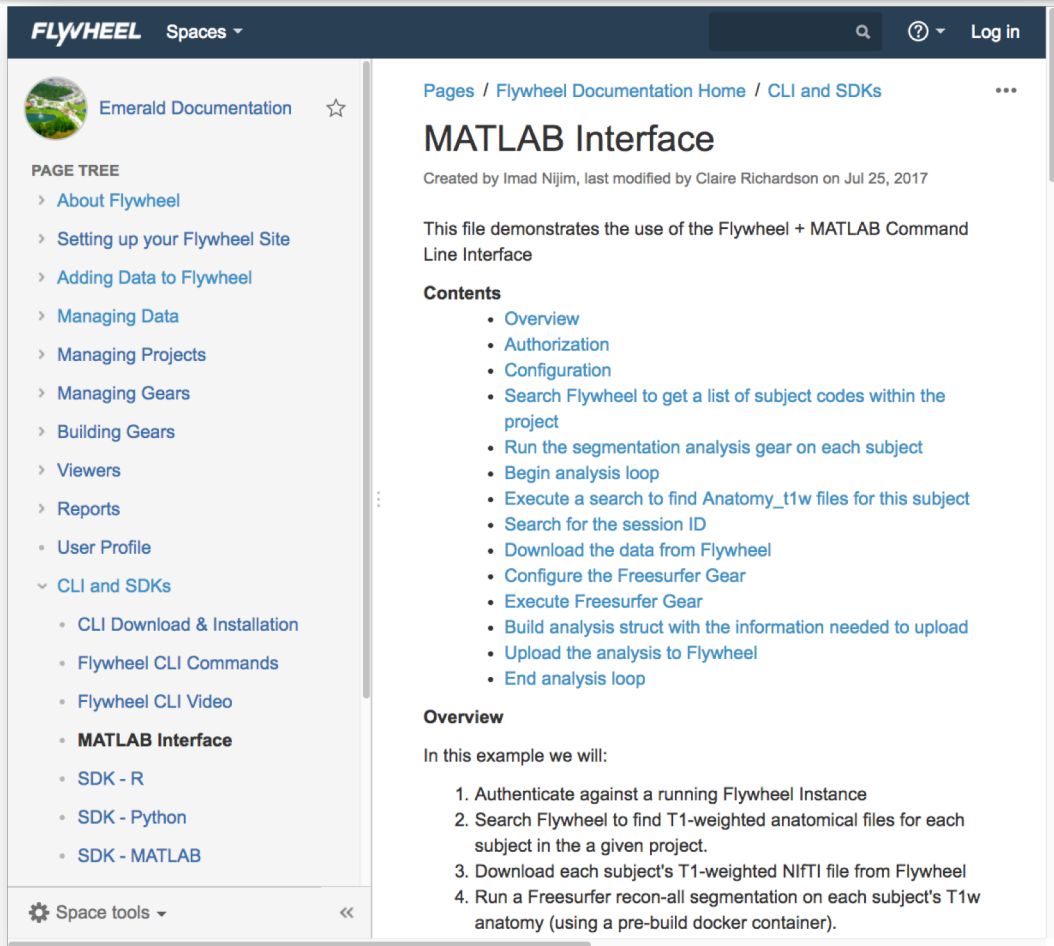
- **Cloud-scale** means you can select and run many Gears on different data sets at the same time
- The cloud providers sets up the machines upon request; they have a lot

The screenshot shows the FLYWHEEL interface. On the left is a dark sidebar with navigation options: DATA, Projects, Sessions, Collections, Upload, GEARS, Installed Gears, REPORTING, Project Report, Usage Report, ADMIN, Users, Groups, and Access Log. The main content area is titled "Projects · ADNI: T1 · Data" and shows "Displaying 1-50 of 1069" items. It features a table with columns for Actions, Advanced Filters, Acquisitions, Session, Subject, Project, and Analyses. The table lists various data points with timestamps and session IDs. Below the table is a pagination control showing page 1 of 7. On the right, there is a "Filter By Name" section with a "Run Analysis Gear" button. Two analysis results are visible: "freesurfer-recon-all-v600 04/11/2017 21:35:26 Session Analysis" and "fsl-bet 11/01/2018 16:01:49 Session Analysis".

Actions	Advanced Filters	Acquisitions	Session	Subject	Project	Analyses	Provenance
<input type="checkbox"/>	Timestamp			Subject			
<input type="checkbox"/>							
<input type="checkbox"/>							
<input type="checkbox"/>							
<input type="checkbox"/>							
<input type="checkbox"/>							
<input type="checkbox"/>							
<input type="checkbox"/>							
<input type="checkbox"/>							
<input checked="" type="checkbox"/>							
<input checked="" type="checkbox"/>							
<input type="checkbox"/>							

# Computation: Exploration and methods development

- Language bindings (SDK)
  - Python
  - Matlab
  - R
- REST database endpoints
- Computational sharing



The screenshot shows the Flywheel documentation interface. The top navigation bar includes the Flywheel logo, a 'Spaces' dropdown, a search icon, a help icon, and a 'Log in' button. The main content area is titled 'MATLAB Interface' and is part of the 'CLI and SDKs' section. The page was created by Imad Nijim and last modified by Claire Richardson on July 25, 2017. The main text states: 'This file demonstrates the use of the Flywheel + MATLAB Command Line Interface'. A 'Contents' section lists the following topics: Overview, Authorization, Configuration, Search Flywheel to get a list of subject codes within the project, Run the segmentation analysis gear on each subject, Begin analysis loop, Execute a search to find Anatomy\_t1w files for this subject, Search for the session ID, Download the data from Flywheel, Configure the Freesurfer Gear, Execute Freesurfer Gear, Build analysis struct with the information needed to upload, Upload the analysis to Flywheel, and End analysis loop. An 'Overview' section follows, stating 'In this example we will:' and listing four steps: 1. Authenticate against a running Flywheel Instance, 2. Search Flywheel to find T1-weighted anatomical files for each subject in the a given project, 3. Download each subject's T1-weighted NIFTI file from Flywheel, and 4. Run a Freesurfer recon-all segmentation on each subject's T1w anatomy (using a pre-build docker container).

**FLYWHEEL** Spaces ▾

Emerald Documentation ☆

**PAGE TREE**

- [About Flywheel](#)
- [Setting up your Flywheel Site](#)
- [Adding Data to Flywheel](#)
- [Managing Data](#)
- [Managing Projects](#)
- [Managing Gears](#)
- [Building Gears](#)
- [Viewers](#)
- [Reports](#)
- [User Profile](#)
- ▾ **CLI and SDKs**
  - [CLI Download & Installation](#)
  - [Flywheel CLI Commands](#)
  - [Flywheel CLI Video](#)
  - **MATLAB Interface**
  - [SDK - R](#)
  - [SDK - Python](#)
  - [SDK - MATLAB](#)

Space tools ▾ <<

Pages / [Flywheel Documentation Home](#) / [CLI and SDKs](#) ⋮

## MATLAB Interface

Created by Imad Nijim, last modified by Claire Richardson on Jul 25, 2017

This file demonstrates the use of the Flywheel + MATLAB Command Line Interface

**Contents**

- [Overview](#)
- [Authorization](#)
- [Configuration](#)
- [Search Flywheel to get a list of subject codes within the project](#)
- [Run the segmentation analysis gear on each subject](#)
- [Begin analysis loop](#)
- [Execute a search to find Anatomy\\_t1w files for this subject](#)
- [Search for the session ID](#)
- [Download the data from Flywheel](#)
- [Configure the Freesurfer Gear](#)
- [Execute Freesurfer Gear](#)
- [Build analysis struct with the information needed to upload](#)
- [Upload the analysis to Flywheel](#)
- [End analysis loop](#)

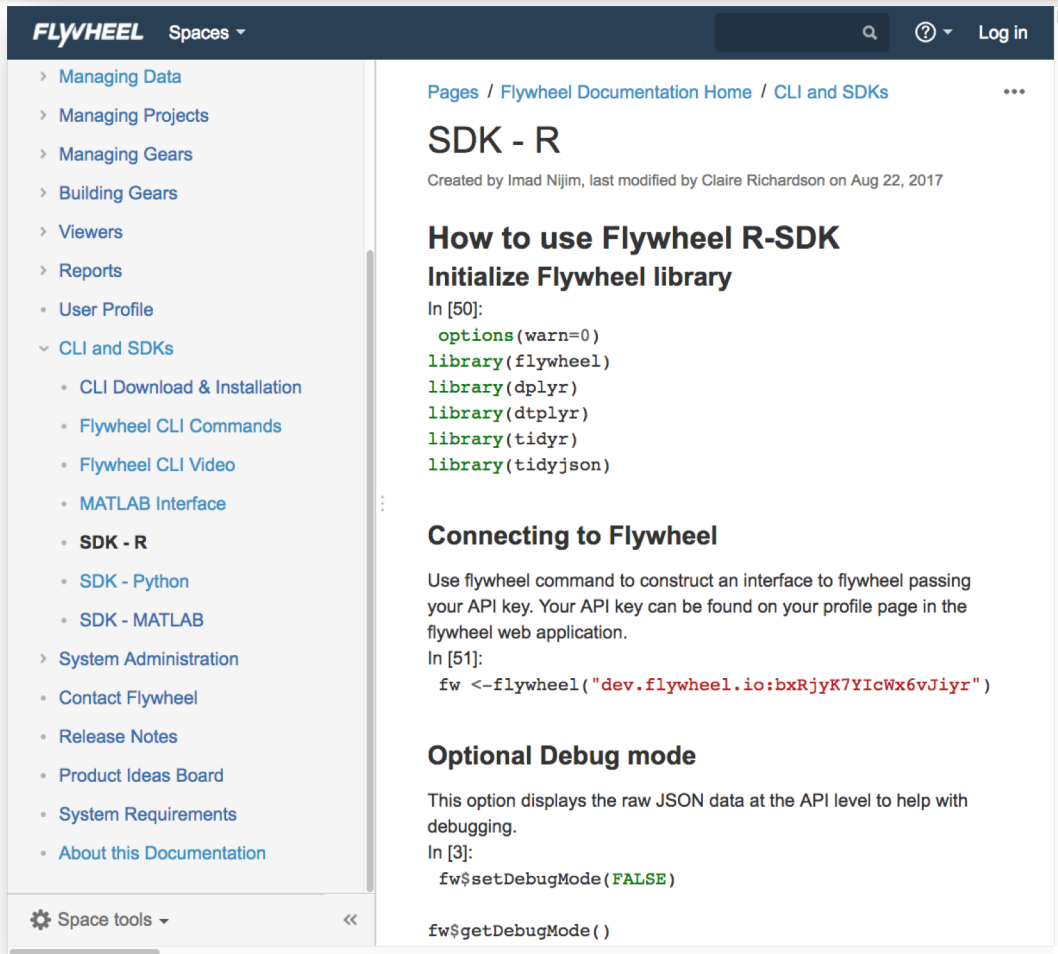
**Overview**

In this example we will:

1. Authenticate against a running Flywheel Instance
2. Search Flywheel to find T1-weighted anatomical files for each subject in the a given project.
3. Download each subject's T1-weighted NIFTI file from Flywheel
4. Run a Freesurfer recon-all segmentation on each subject's T1w anatomy (using a pre-build docker container).

# Computation: Exploration and methods development

- Language bindings (SDK)
  - Python
  - Matlab
  - R
- REST database endpoints
- Computational sharing



The screenshot shows the Flywheel documentation website. The top navigation bar includes the Flywheel logo, a 'Spaces' dropdown menu, a search icon, a help icon, and a 'Log in' button. The left sidebar contains a navigation menu with categories like 'Managing Data', 'Managing Projects', 'Managing Gears', 'Building Gears', 'Viewers', 'Reports', 'User Profile', and 'CLI and SDKs'. The 'CLI and SDKs' category is expanded, showing sub-items such as 'CLI Download & Installation', 'Flywheel CLI Commands', 'Flywheel CLI Video', 'MATLAB Interface', 'SDK - R', 'SDK - Python', and 'SDK - MATLAB'. The main content area displays the 'SDK - R' page, which includes a breadcrumb trail 'Pages / Flywheel Documentation Home / CLI and SDKs', the title 'SDK - R', and a subtitle 'How to use Flywheel R-SDK'. Below this, there is a section titled 'Initialize Flywheel library' with a code block showing the initialization of the Flywheel library for various languages: 

```
In [50]:
options(warn=0)
library(flywheel)
library(dplyr)
library(dtplyr)
library(tidyr)
library(tidyjson)
```

 Another section titled 'Connecting to Flywheel' provides instructions on using the flywheel command with an API key and includes a code block: 

```
In [51]:
fw <- flywheel("dev.flywheel.io:bxRjyK7YIcWx6vJiyr")
```

 The final section, 'Optional Debug mode', explains that this option displays raw JSON data at the API level for debugging and includes a code block: 

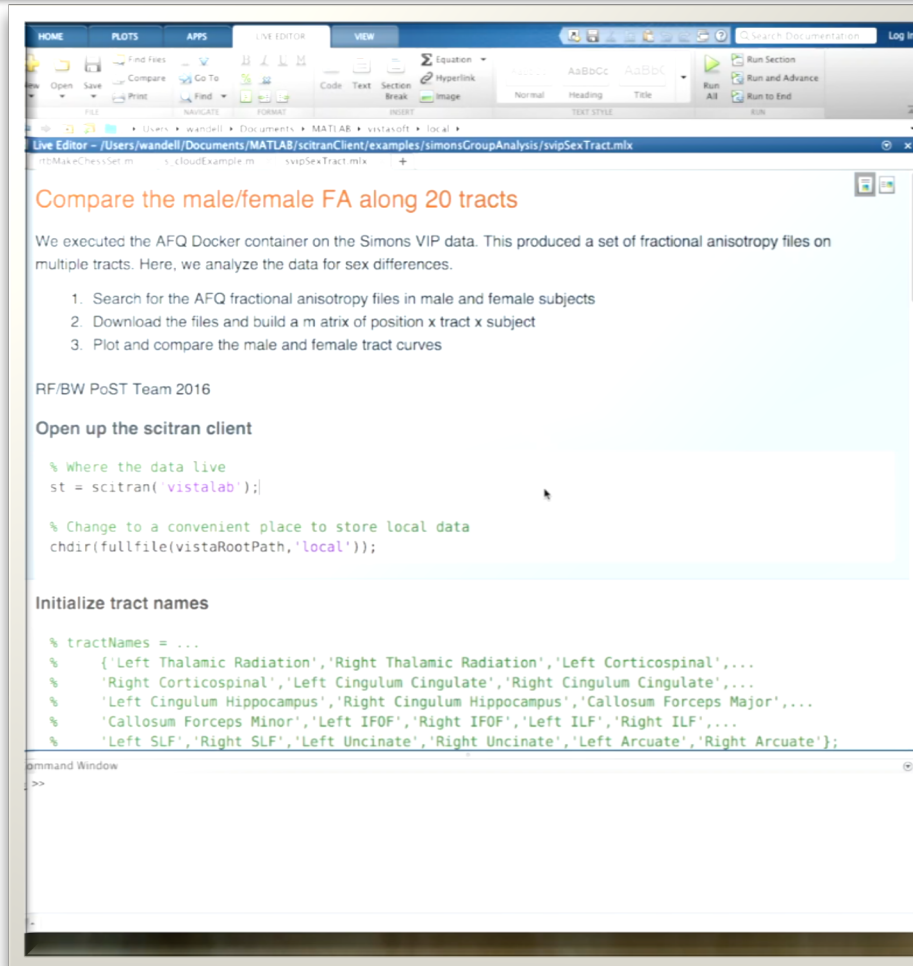
```
In [3]:
fw$setDebugMode(FALSE)

fw$getDebugMode()
```

# Computation: Software development kit (SDK) interface



- The SDK can be used with Jupyter Notebooks or Live Script to make
- This shows the Live Script interface to search, read data, perform analyses, store the analyses in the system

A screenshot of the MATLAB Live Editor interface. The window title is "Live Editor - /Users/wandell/Documents/MATLAB/scitransClient/examples/simonsGroupAnalysis/svipSexTract.mlx". The main content area shows a Live Script with the following text:

Compare the male/female FA along 20 tracts

We executed the AFQ Docker container on the Simons VIP data. This produced a set of fractional anisotropy files on multiple tracts. Here, we analyze the data for sex differences.

1. Search for the AFQ fractional anisotropy files in male and female subjects
2. Download the files and build a matrix of position x tract x subject
3. Plot and compare the male and female tract curves

RF/BW PoST Team 2016

Open up the scitrans client

```
% Where the data live
st = scitrans('vistalab');

% Change to a convenient place to store local data
chdir(fullfile(vistaRootPath, 'local'));
```

Initialize tract names

```
% tractNames = ...
% {'Left Thalamic Radiation', 'Right Thalamic Radiation', 'Left Corticospinal', ...
%  'Right Corticospinal', 'Left Cingulum Cingulate', 'Right Cingulum Cingulate', ...
%  'Left Cingulum Hippocampus', 'Right Cingulum Hippocampus', 'Callosum Forceps Major', ...
%  'Callosum Forceps Minor', 'Left IFOF', 'Right IFOF', 'Left ILF', 'Right ILF', ...
%  'Left SLF', 'Right SLF', 'Left Uncinate', 'Right Uncinate', 'Left Arcuate', 'Right Arcuate'};
```

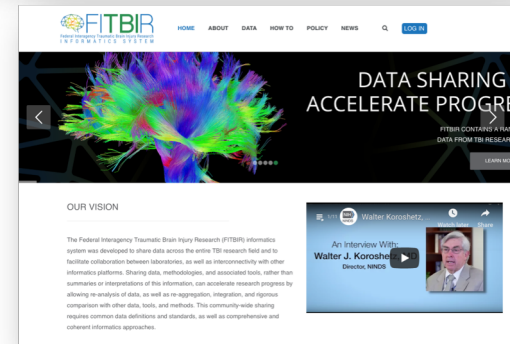
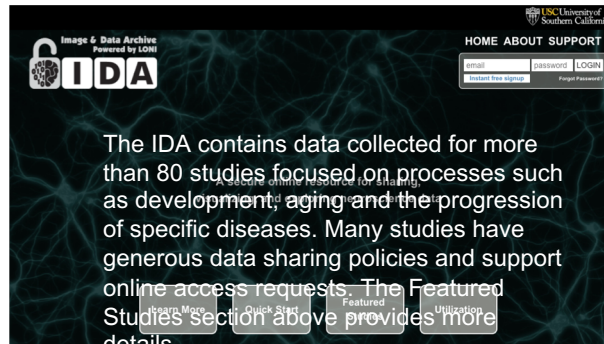
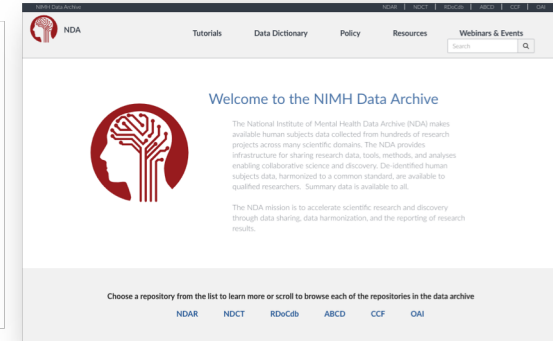
Command Window

```
>>
```



# Databases are designed for different purposes

- At time of publication, curated data may be placed on a shared site
- These sites add value by checking for quality, providing some pipelines
- They differ in many ways from our MRI Center goal – mainly they are not designed to manage ongoing projects, support a broad range of cloud computations
- These are valuable tools, but with another purpose



LORIS, COINS, XNAT, OTHERS ...

# Statistical reproducibility differs from computational reproducibility

- Null hypothesis statistical testing and group comparisons are deeply ingrained in the neuroscience publication culture, particularly cognitive and clinical neuroscience
- Our tools are for many types of computations, not the final statistical hypothesis testing

The screenshot shows the NIH website's 'RIGOR AND REPRODUCIBILITY' page. The header includes the NIH logo and navigation links. The main content area features a sidebar with a list of links: Reporting Guidelines, Application Instructions, Training, Funding Opportunities, Meetings and Workshops, Announcements, Publications, and Resources. The main article is titled 'Principles and Guidelines for Reporting Preclinical Research' and includes social media sharing icons, a paragraph about a 2014 workshop, and a list of principles. The first principle is 'Rigorous statistical analysis', which includes a sub-section on journal policies. The second principle is 'Transparency in reporting'.

NIH National Institutes of Health  
Turning Discovery Into Health

Search NIH

NIH Employee Intranet | Staff Directory | En Español

Health Information | Grants & Funding | News & Events | Research & Training | Institutes at NIH | About NIH

Home » Research & Training » Rigor and Reproducibility

## RIGOR AND REPRODUCIBILITY

### Rigor and Reproducibility

- Reporting Guidelines
- Application Instructions
- Training
- Funding Opportunities
- Meetings and Workshops
- Announcements
- Publications
- Resources

## Principles and Guidelines for Reporting Preclinical Research

[f](#) [t](#) [e](#) [m](#) [+](#)

NIH held a joint workshop in June 2014 with the Nature Publishing Group and Science on the issue of reproducibility and rigor of research findings, with journal editors representing over 30 basic/preclinical science journals in which NIH-funded investigators have most often published. The workshop focused on identifying the common opportunities in the scientific publishing arena to enhance rigor and further support research that is reproducible, robust, and transparent.

The journal editors came to consensus on a set of principles to facilitate these goals, which a considerable number of journals have agreed to endorse. These principles are shown below.

[Open all](#) | [Close all](#)

- Rigorous statistical analysis**  
A section outlining the journal's policies for statistical analysis should be included in the Information for Authors, and the journal should have a mechanism to check the statistical accuracy of submissions.
- Transparency in reporting**

### Related Links

[Rigor and Reproducibility FAQs](#)

The **FAIR** principles are a set of community-developed guidelines to ensure that data or any digital object are

- **Findable**
- **Accessible**
- **Interoperable**
- **Reusable**

We are trying to support computational reproducibility following the **FAIR** principles

A screenshot of a web page from the journal Scientific Data. The page title is "The FAIR Guiding Principles for scientific data management and stewardship" by Mark D. Wilkinson, Michel Dumontier, and Barend Mons. The page includes a navigation bar with "nature > scientific data > comment > article", a "MENU" dropdown, and the journal logo. Below the title, it says "Comment | OPEN | Published: 15 March 2016". The article is identified as "Scientific Data 3, Article number: 160018 (2016)" with a "Download Citation" link. The "Abstract" section begins with the text: "There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis".

nature > scientific data > comment > article

MENU ▾

SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

## The FAIR Guiding Principles for scientific data management and stewardship

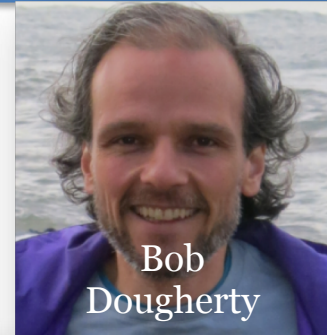
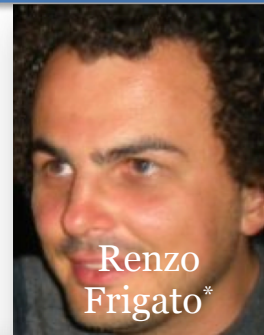
Mark D. Wilkinson, Michel Dumontier [...] Barend Mons ✉

Scientific Data 3, Article number: 160018 (2016) | Download Citation ↓

### Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis

# Achieving Reproducibility: Perspectives and New Tools for MRI Centers



The Researcher's Data &  
Analysis Platform

learn more at [flywheel.io](https://flywheel.io)

**FLYWHEEL**

 Sign in with Google