
FAIR Workshop Series:

Integrating Box and RStudio

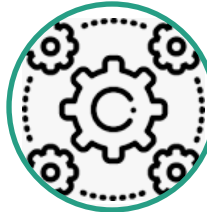
April 20, 2022



Findable



Accessible



Interoperable



Re-usable

**Data scientist,
epidemiologist &
biostatisticians
working together**

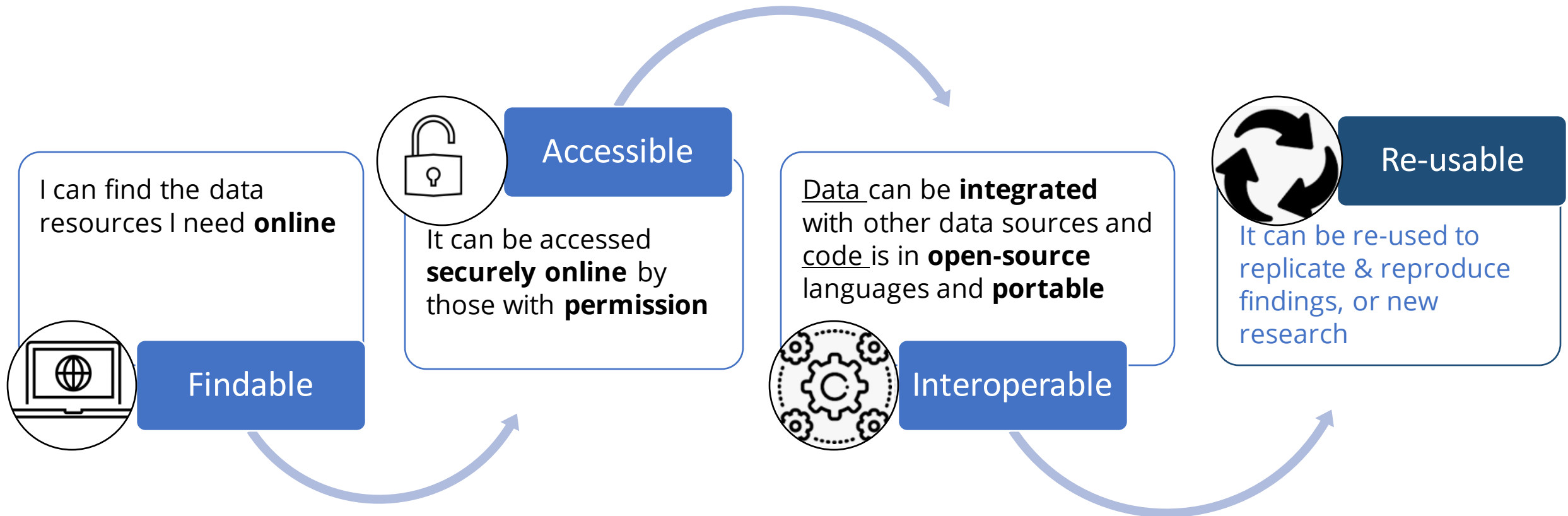


Why did you register for today's workshop?

Why should you bother with FAIR principles?

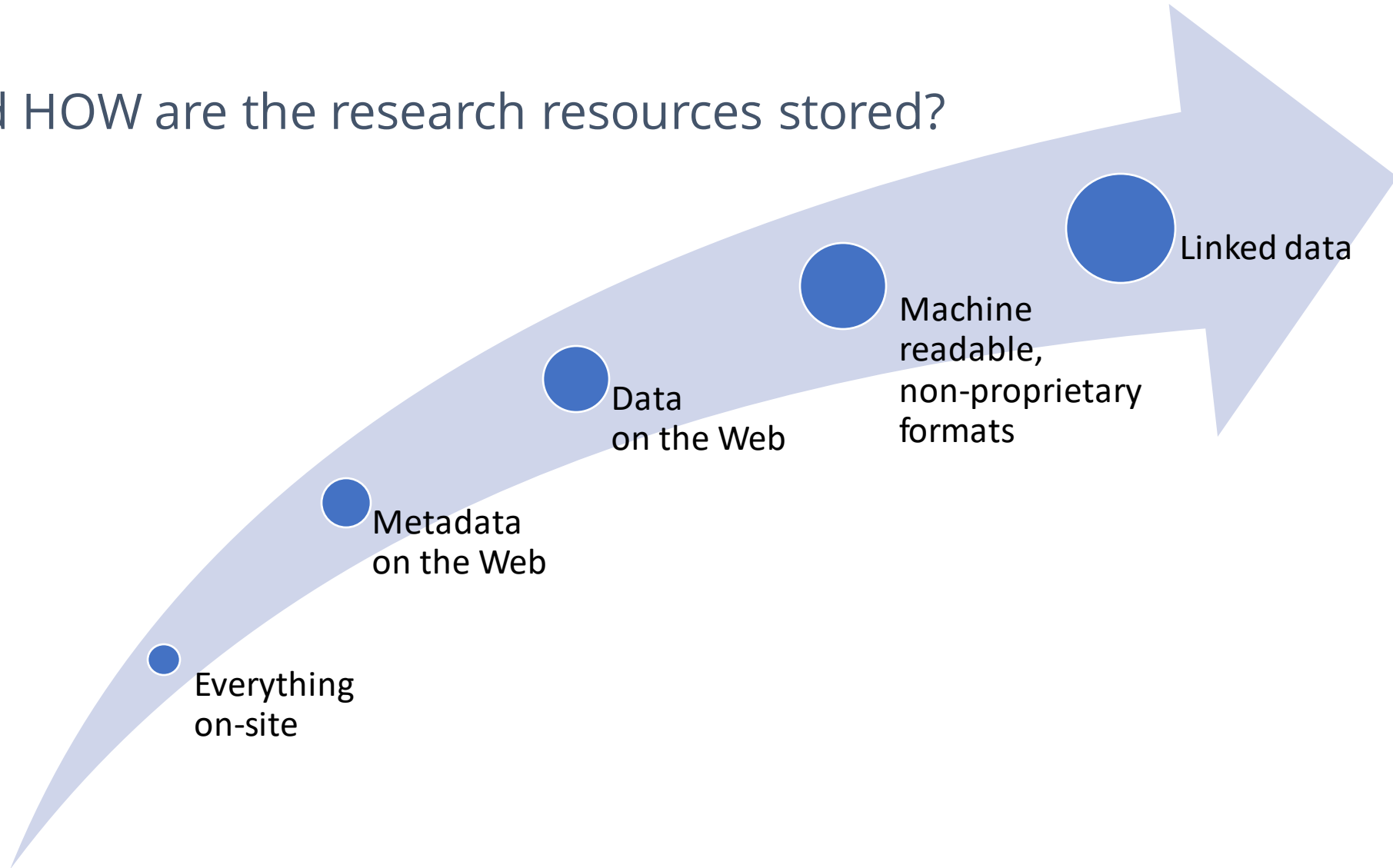
- **Regular, common challenges addressed by FAIR practices**
 - You're not sure where which data/code version you need is stored
 - Barriers to accessing data/code
 - Not having license to proprietary software (e.g. SAS, STATA, Matlab)

Are my data and code FAIR ?



Key steps towards increasing degrees of FAIRness

WHERE and HOW are the research resources stored?



The simplest way to start being FAIR (we could think of)

File/Folder

Access management
Version control
Track changes/usage



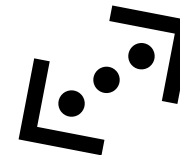
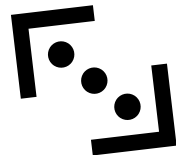
Put your
Data in Box

Code

Access management
Version control
Track changes/usage



Push/Pull
code in
GitHub



Run analyses in
your computer/server

Why putting my dataset in Box makes it more FAIR?

- Box assigns files a unique URL
- Manage access/permissions to data
- Versioning control of data
- Integrate data from different sources/governance without moving the data
- Remote analyze data

Rstudio: Use 'boxr' to access your data in Box

Package 'boxr'

January 19, 2021

Type Package

Title Interface for the 'Box.com API'

Version 0.3.6

URL <https://github.com/r-box/boxr/>

BugReports <https://github.com/r-box/boxr/issues>

Description An R interface for the remote file hosting service 'Box' (<<https://www.box.com/>>). In addition to uploading and downloading files, this package includes functions which mirror base R operations for local files, (e.g. `box_load()`, `box_save()`, `box_read()`, `box_setwd()`, etc.), as well as 'git' style functions for entire directories (e.g. `box_fetch()`, `box_push()`).

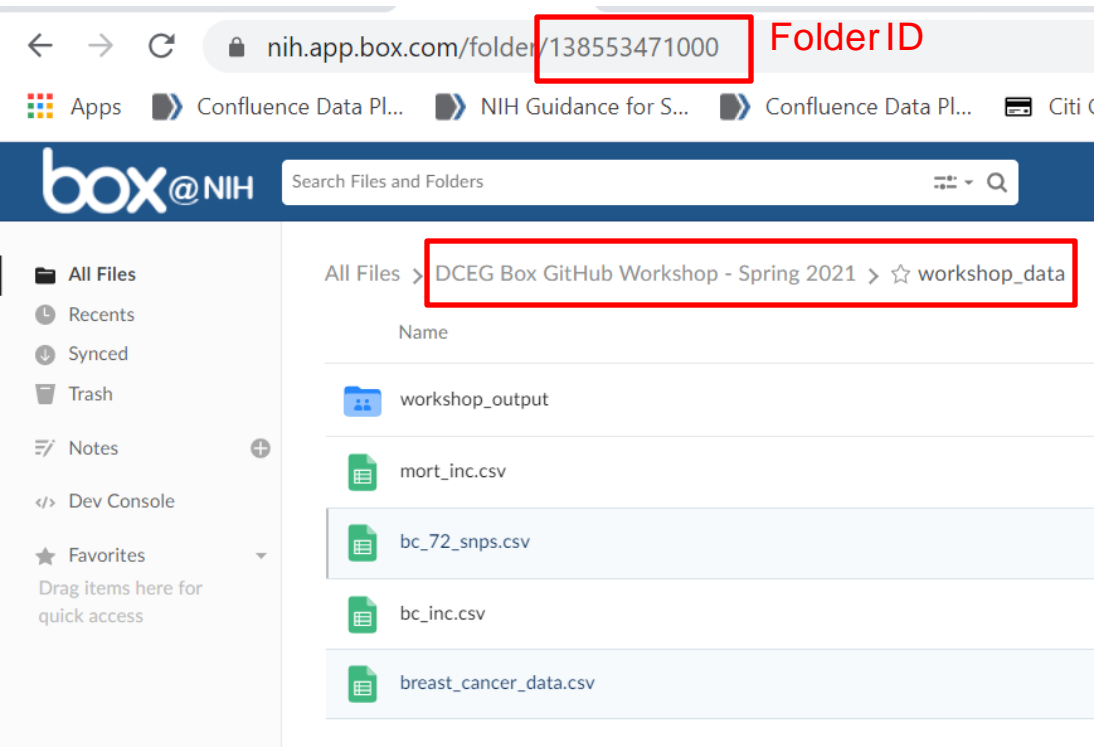
Workshop Objectives: Boxr example script

1) **Rstudio and Box**

- Read data from Box and run your analyses
- Save data/output to Box from Rstudio
- Understand versioning control of data
- Manage permissions to access data

Before we get to Rstudio

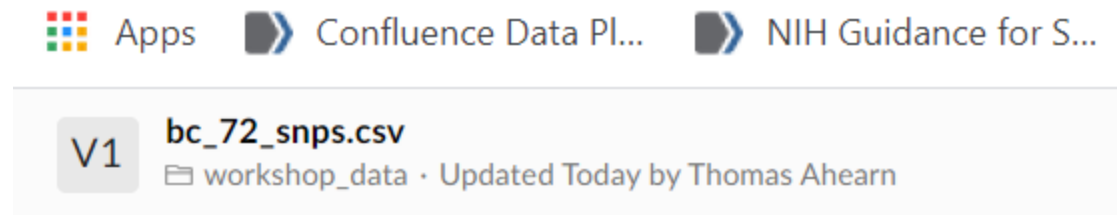
Anatomy of Box.com: Folder IDs, File IDs, and Version IDs



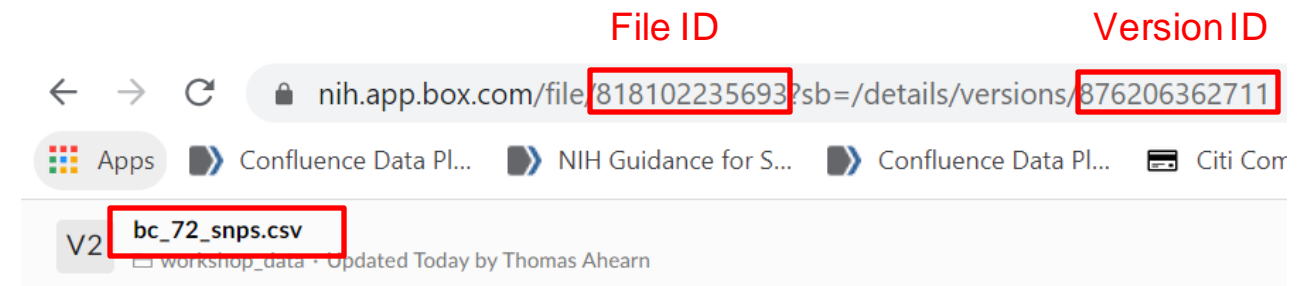
This screenshot shows the Box.com web interface for a user named 'nih.app.box.com'. The address bar displays the URL 'nih.app.box.com/folder/138553471000', where the folder ID '138553471000' is highlighted with a red box and labeled 'FolderID'. The left sidebar shows a navigation menu with options like 'All Files', 'Recents', 'Synced', 'Trash', 'Notes', 'Dev Console', and 'Favorites'. The main content area shows the breadcrumb path 'All Files > DCEG Box GitHub Workshop - Spring 2021 > ☆ workshop_data', with the final folder name 'workshop_data' highlighted by a red box. Below the breadcrumb, a list of files is shown, including 'workshop_output' (folder), 'mort_inc.csv', 'bc_72_snps.csv' (selected), 'bc_inc.csv', and 'breast_cancer_data.csv'.



This screenshot shows the Box.com web interface for a file. The address bar displays the URL 'nih.app.box.com/file/818102235693', where the file ID '818102235693' is highlighted with a red box and labeled 'File ID'.



This screenshot shows the details for the first version (V1) of the file 'bc_72_snps.csv'. The file is located in the 'workshop_data' folder and was updated today by Thomas Ahearn.



This screenshot shows the details for the second version (V2) of the file 'bc_72_snps.csv'. The address bar displays the URL 'nih.app.box.com/file/818102235693?sb=/details/versions/876206362711', where the file ID '818102235693' is highlighted with a red box and labeled 'File ID', and the version ID '876206362711' is highlighted with a red box and labeled 'VersionID'. The file name 'bc_72_snps.csv' is also highlighted with a red box.

Data governance in Box - Folders

The screenshot shows the Box@NIH web interface. On the left is a sidebar with navigation options: All Files, Recents, Synced, Trash, Notes, Dev Console, and Favorites. The main area displays a folder named 'Example_data' containing several files: logit.Rdata, breast_cancer_data (1).csv, Example_R_code.R (V2), breast_cancer_data.csv (V2), logistic_model.Rdata, density_plot.Rdata, upated_data.Rdata, Data_dictionary.docx (V3), and iCARE_PLOS-ONE_2020.pdf. A 'Share' dialog box is open over the folder, titled 'Share 'Example_data''. It shows 'Invite People' with a search bar and a list of users (MG, DR, JA, +8). Below this is a list of permission levels: Editor (checked), Viewer Uploader, Previewer Uploader, Viewer, Previewer, and Uploader.

Name	Updated	Size
logit.Rdata	May 14, 2021 by Grace Hong	1 MB
breast_cancer_data (1).csv		2.2 MB
Example_R_code.R (V2)		892 B
breast_cancer_data.csv (V2)		2.2 MB
logistic_model.Rdata		1.1 MB
density_plot.Rdata		7.8 KB
upated_data.Rdata		512.5 KB
Data_dictionary.docx (V3)		15.5 KB
iCARE_PLOS-ONE_2020.pdf		1.2 MB

Share 'Example_data'

Invite People

Shared with MG DR JA +8

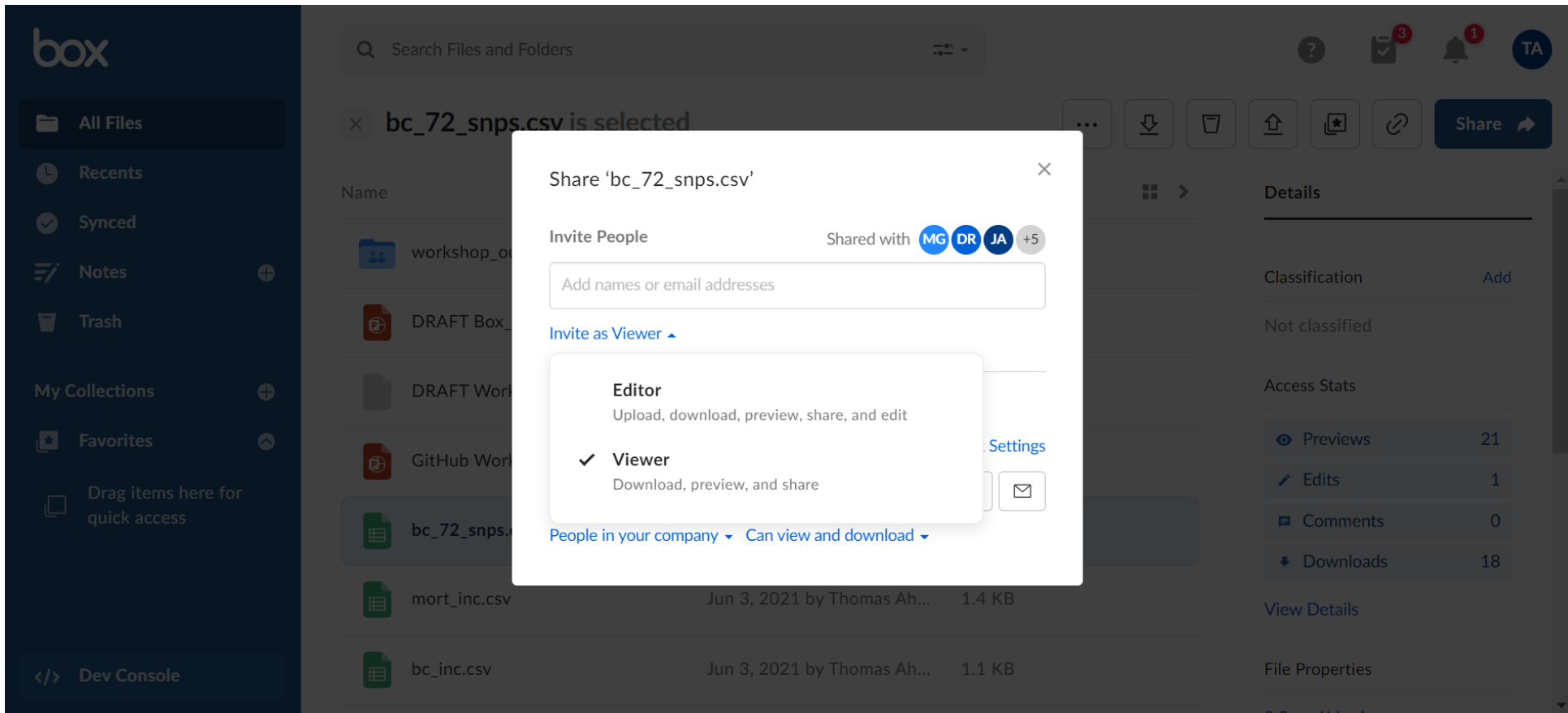
Add names or email addresses

Invite as Editor

- ✓ **Editor**
Upload, download, preview, share, edit, and delete
- Viewer Uploader**
Upload, download, preview, share, and edit
- Previewer Uploader**
Upload and preview
- Viewer**
Download, preview, and share
- Previewer**
Preview only
- Uploader**
Upload only

Share Box folders with investigators, giving specific levels of permission

Data governance in Box - Files



Share Box Files with investigators, giving specific levels of permission.

Data governance in Box

V2 **breast_cancer_data.csv**
Example_data · Updated Apr 5, 2021 by Parichoy Pal Choudhury

	famhist	menarche_dec	parity	birth_dec	agemeno_dec	height_dec	bmi_dec	rs
	0	8	0				10	1
	0	10	0				4	1
	0	1	0				10	1
	0	8	2				5	2
	0	8	2				9	2
	0	1	3				8	2
	1	9	2				8	1
	1	9	1				9	1
	0	5	2				6	1
	0	1	1				10	1
	0	3	0				5	2
	1	8	4				10	1
	0	10	2				10	1
	0	10	2				9	2
	1	8	2				1	1
	0	9	0				10	2
	0	5	2				3	2
	0	9	4				7	1
	0	8	3				2	2
	0	10	2				4	2

Access Stats for 'breast_cancer_data.csv'

65
PREVIEWS

1
EDITS

0
COMMENTS

45
DOWNLOADS

Today

Thomas Ahearn

Previewed today at 4:09 PM

Last Week

Clara Bodelon

Downloaded May 22, 2021 at 2:40 PM via Box Drive

Daniel Russ

Previewed May 21, 2021 at 11:21 AM

Daniel Russ

Previewed May 21, 2021 at 11:18 AM

Ayodele Ajayi

Downloaded May 21, 2021 at 9:18 AM via Box Drive

Close

Export

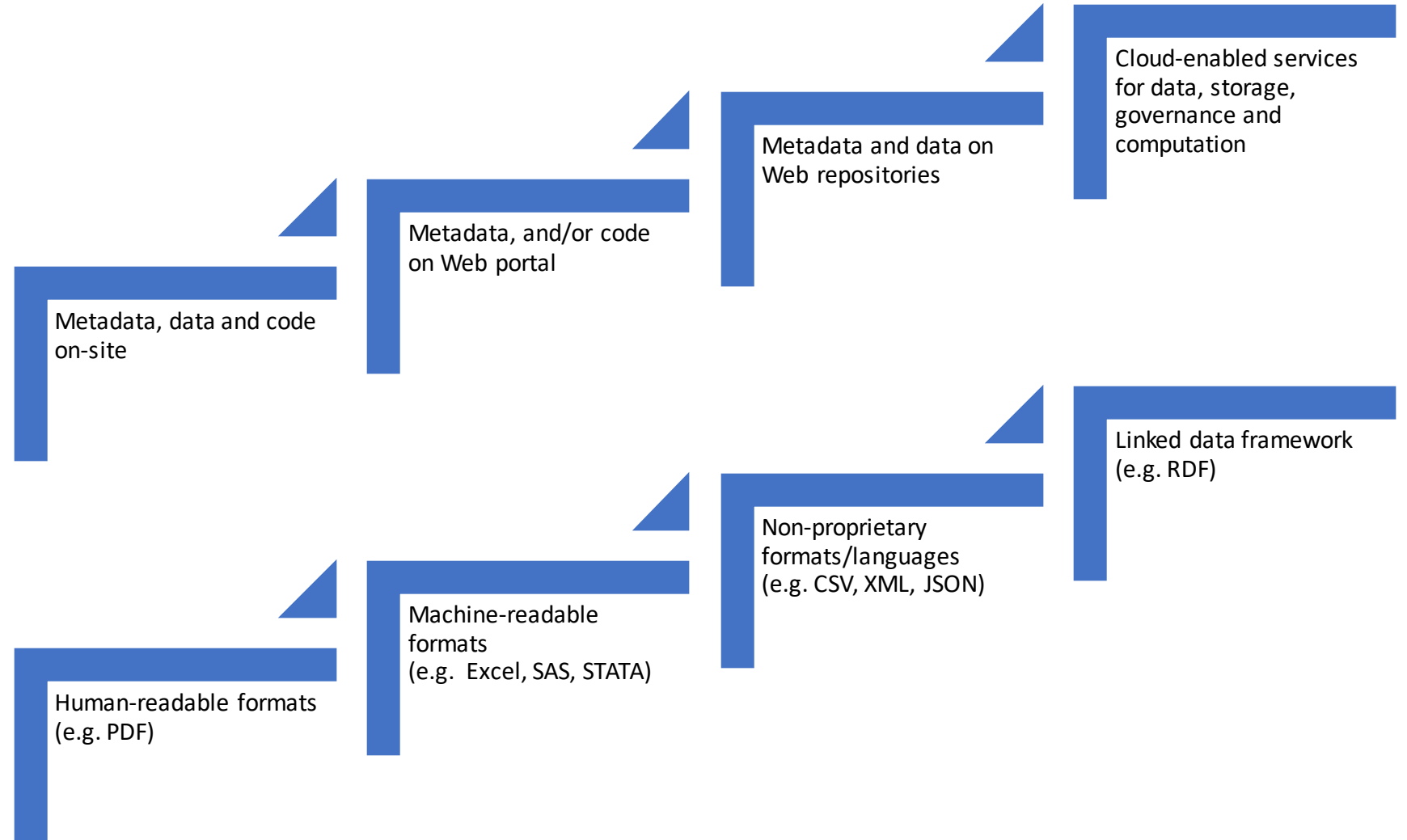
Monitor who has accessed the data

After the Workshop

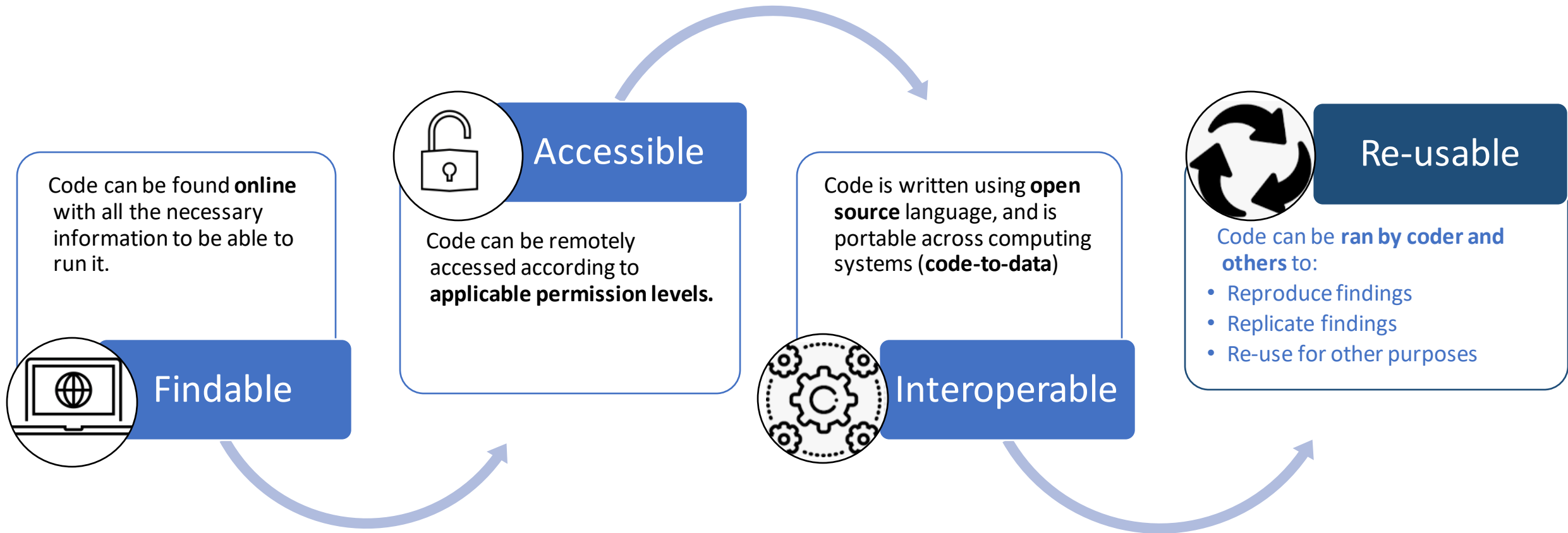
- Post-workshop survey
- Helpful links that will be shared:
 - Workshop webpage
 - FAIR Fridays

Increasing degrees of FAIRness in two dimensions

**Where are the
research
resources?**



Is my code FAIR ?



Why should I bother with FAIR principles?



Replicability



Reproducibility



Re-usability