

# Making FAIR Fair to the Researchers

Bhavesh Patel, Ph.D.

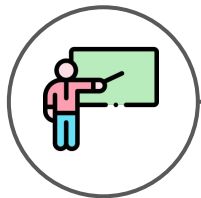
Research Professor



# About This Presentation



20 min + Q&A



Share some lessons I have  
learned for making FAIR more fair

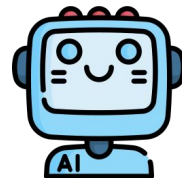
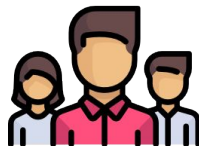


# **FAIR Principles**

# The FAIR Principles

## Overview

The FAIR Principles (2016) provide instructions for **making research outcomes reusable** by humans and machines



# The FAIR Principles

## Adoption

The FAIR Principles have been **supported and adopted worldwide** by all stakeholders including researchers, publishers, and funders



Example: The new Data Sharing Policy of the National Institutes of Health (NIH) effective January 2023 requires all grant proposals to include a plan for making data FAIR

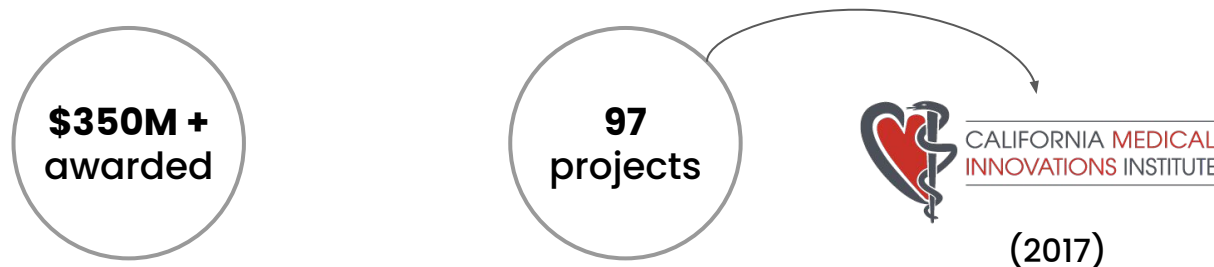


# **My Encounter with FAIR**

# My Encounter with FAIR

## The NIH SPARC Program

**SPARC is an NIH Common Fund Program** aimed at accelerating the development of neuromodulation therapies (2015-2025)



# My Encounter with FAIR

## The NIH SPARC Program

SPARC imposes **strict guidelines to awardees for making data FAIR**

- Organize data into a standard folder structure
- Follow file and folder naming conventions
- Include metadata using standard file formats and vocabularies
- Upload dataset on the SPARC data platform Pennsieve
- And more!

I was put in charge of doing this for the SPARC data at our Institute



# My Encounter with FAIR

## Challenges



**Difficult** – I was not trained for this



**Time-consuming** – The guidelines are very extensive



**Boring** – I wanted to do research, publish papers!

# My Encounter with FAIR

## Solution: Automation

Idea: Automate the process of preparing and sharing SPARC data

2017:  
Started developing  
Python scripts for our  
group's data



2018:  
Presented the idea  
at the SPARC  
Hackathon



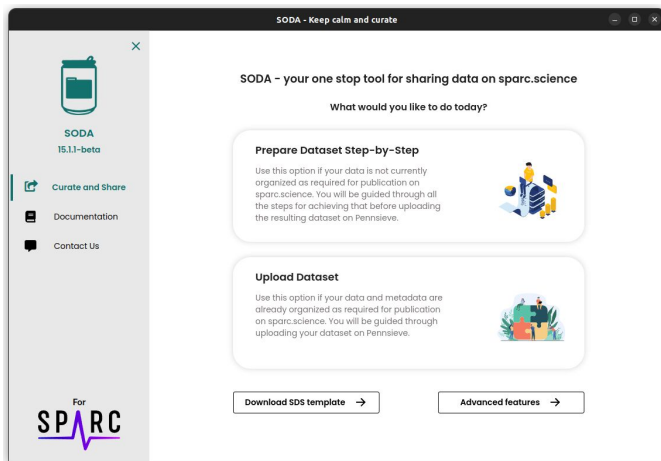
2019:  
Received funding  
from SPARC for  
further development



# My Encounter with FAIR

## Software to Organize Data Automatically (SODA)

SODA is an open source and free desktop software that **simplifies the preparation and sharing** of datasets according to the SPARC guidelines



***Like a tax filing tool but for filing FAIR SPARC datasets***

# My Encounter with FAIR

## SODA Performance



Reduces time, errors, and difficulty



Has helped process 27+ TB of data (400k+ files) from 1000+ users (since 2021)

# My Encounter with FAIR

## SODA-inspired Tools

SODA has inspired many tools aimed at supporting researchers in making their data FAIR



Making immunology  
related data FAIR  
(NIAID)



Making clinical  
research data FAIR  
(NIH Bridge2AI)



A team from the NIH  
Brain Initiative has  
forked SODA to help  
make neurophysiology  
data FAIR

# My Encounter with FAIR

## Lesson Learned #1

**User-friendly tools** can help reduce researchers time and effort,  
and are needed to make FAIR more fair to them





# **FAIR Research Software**

# FAIR Research Software

## Definition of Research Software

Any software created **during the research process**  
**or for a research purpose**



Data analysis



Computational  
model



AI/ML  
model



# FAIR Research Software

## FAIR4RS Principles

**FAIR Principles for Research Software (FAIR4RS Principles, 2022):**  
**17 principles tailored for research software**

F1. Software is assigned a globally unique and persistent identifier.

F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.

F1.2. Different versions of the software are assigned distinct identifiers.

F2. Software is described with rich metadata.

F3. Metadata clearly and explicitly include the identifier of the software they describe.

F4. Metadata are FAIR, searchable and indexable.

(and more)

# FAIR Research Software

## Problem

The FAIR4RS Principles by design  
**do not provide actionable instructions**

How do I assign a unique identifier?

How do I provide rich metadata?

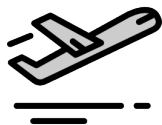


# FAIR Research Software

## FAIR-BioRS Guidelines

### FAIR Biomedical Research Software (FAIR-BioRS) Guidelines

Minimal, actionable, step-by-step guidelines for  
complying with each of the FAIR4RS principles



**December 2021**  
Beginning of this effort



**August 2023**  
Manuscript published

# FAIR Research Software

## FAIR-BioRS Guidelines Overview



Full guidelines: [fair-biors.org](https://fair-biors.org)



Step-by-step process



Clear & actionable instructions

The screenshot shows a web browser window with the address bar displaying 'fair-biors.org'. The page title is 'Guidelines'. The main heading is 'Guidelines', followed by 'FAIR-BioRS Guidelines v2.0.0'. The content is organized into sections and numbered steps:

- 1. Prepare prior to the development of the software**
  - 1.1. Select a version control system platform to work from (GitHub, Bitbucket, or GitLab are suggested) and create a repository there for your software.
  - 1.2. Select a license and include the license terms in a file called "LICENSE" using plain text or markdown syntax. Locate it in the root directory of the software. While the FAIR4RS principles do not require research software to be open-source, it is highly recommended to use a license approved by the Open Source Initiative (OSI). Amongst those licenses, it is encouraged to use the permissive MIT or Apache 2.0 licenses. Use [choosealicense.com](https://choosealicense.com) and/or the [SPDX License List](https://spdx.org/licenses/) for help. During development, ensure that the software's license is compatible with the software's dependencies.
- 2. Follow coding standards and best practices during development**
  - 2.1. Have code-level documentation (e.g., in code comments, description in the file headers) when deemed necessary for code reuse.
  - 2.2. Record dependencies as per standard practices for the coding language, e.g. in a requirements.txt file for Python code, in a package.json file for Node projects, or in a DESCRIPTION file for R packages.
  - 2.3. Follow language-specific standards and best practices (e.g. [PEP 8 Style Guide for Python Code](https://pep8.org/), [Google's R Style Guide for R code](https://google.github.io/styleguide/rguide/), etc.) and document them (c.f. 3.2).
  - 2.4. Ensure that inputs/outputs of the software follow any applicable community standards (e.g., General Feature Format (GFF) for genomic annotation files). Use [fairsharing.org](https://fairsharing.org) for finding relevant standards.

# FAIR Research Software

## Lesson Learned #2

**Actionable guidelines** for making research outcomes FAIR can reduce burden on researchers and are needed to make FAIR more fair to them



# FAIR Research Software

## Codefair



Codefair is a free and open source GitHub app that acts as  
**your personal assistant for making software FAIR**

1

**Install** Codefair (codefair.io) on the GitHub repository of your software

2

**Develop** your software as usual from GitHub

3

**Track** and address FAIR compliance issues through the Codefair GitHub issue

# FAIR Research Software

## Codefair Issue Dashboard

The screenshot shows a web browser window displaying the 'FAIR Compliance Dashboard #1' issue page. The interface includes a top navigation bar with tabs for Code, Issues (1), Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. The issue title is 'FAIR Compliance Dashboard #1' with 'Edit' and 'New issue' buttons. Below the title, it says 'codefair-app [bot] opened this issue 14 minutes ago · 0 comments'. The issue content is a comment from 'codefair-app [bot]' with a 'Contributor' button. The comment text reads: 'Check the FAIRness of your software'. It explains that this issue is the repository's dashboard for all things FAIR and provides a link to the documentation. It then lists two sections: 'LICENSE' with a green checkmark and an 'Edit License' button, and 'Metadata' with a red X and an 'Add Metadata' button. The 'Metadata' section explains that CITATION.cff and codemetada.json files are expected at the root level and provides a link to the FAIR-BioRS Guidelines. On the right side, there are settings for Assignees (No one—assign yourself), Labels (None yet), Projects (None yet), Milestone (No milestone), Development (Create a branch for this issue or link a pull request.), Notifications (Unsubscribe button), and 0 participants.

FAIR Compliance Dashboard #1

codefair-app [bot] opened this issue 14 minutes ago · 0 comments

codefair-app [bot] commented 14 minutes ago • edited

Contributor

### Check the FAIRness of your software

This issue is your repository's dashboard for all things FAIR. Keep it open as making and keeping software FAIR is a continuous process that evolves along with the software. You can read the [documentation](#) to learn more.

**LICENSE** ✓

A LICENSE file is found at the root level of the repository.

[Edit License](#)

**Metadata** ✗

To make your software FAIR, a CITATION.cff and codemetada.json are expected at the root level of your repository, as recommended in the [FAIR-BioRS Guidelines](#). These files are not found in the repository. If you would like Codefair to add these files, click the "Add metadata" button below to go to our interface for providing metadata and generating these files.

[Add Metadata](#)

Assignees

No one—[assign yourself](#)

Labels

None yet

Projects

None yet

Milestone

No milestone

Development

[Create a branch](#) for this issue or link a pull request.

Notifications

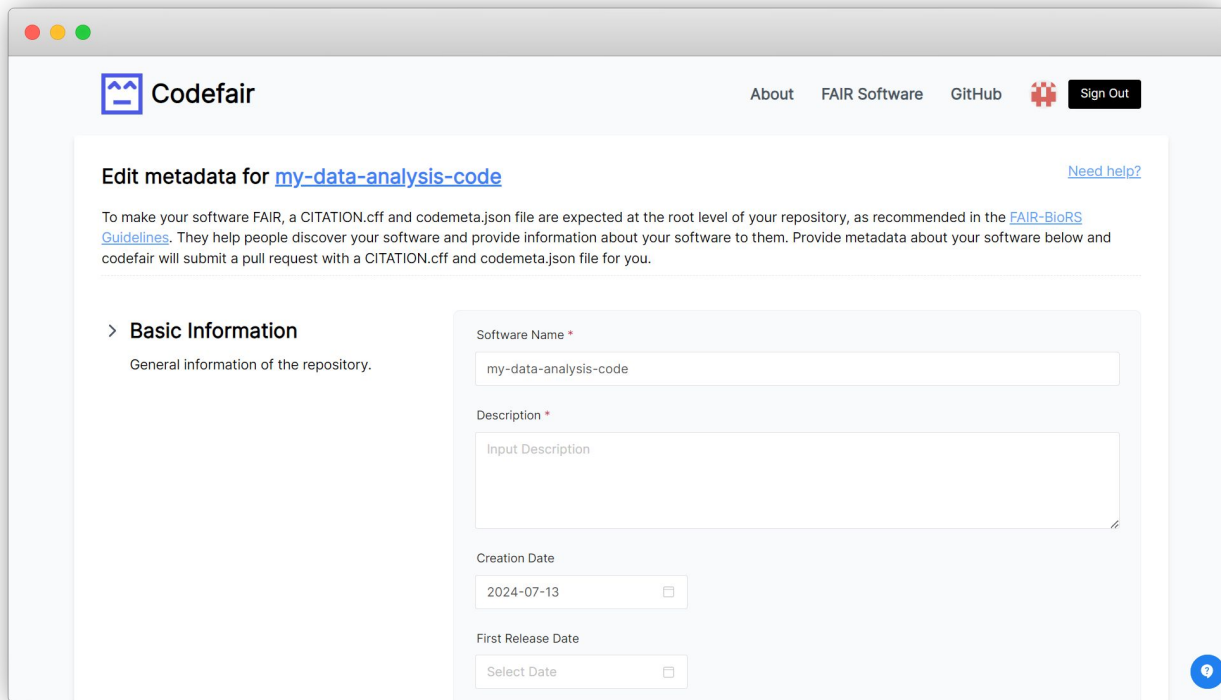
[Unsubscribe](#)

You're receiving notifications because you're watching this repository.

0 participants

# FAIR Research Software

## Codefair UI



The screenshot shows a web browser window with the Codefair logo and navigation links (About, FAIR Software, GitHub, Sign Out). The main content area is titled "Edit metadata for my-data-analysis-code" with a "Need help?" link. Below this is a paragraph explaining the purpose of the metadata files. A section titled "Basic Information" is expanded, showing a form with fields for Software Name, Description, Creation Date, and First Release Date.

**Codefair** About FAIR Software GitHub Sign Out

Edit metadata for [my-data-analysis-code](#) [Need help?](#)

To make your software FAIR, a CITATION.cff and codemeta.json file are expected at the root level of your repository, as recommended in the [FAIR-BioRS Guidelines](#). They help people discover your software and provide information about your software to them. Provide metadata about your software below and codefair will submit a pull request with a CITATION.cff and codemeta.json file for you.

> **Basic Information**  
General information of the repository.

Software Name \*  
my-data-analysis-code

Description \*  
Input Description

Creation Date  
2024-07-13

First Release Date  
Select Date



# FAIR Research Software

## Codefair UI

The screenshot shows a web form for FAIR Research Software. At the top, there's a URL field with the value "https://example.com/download/1.0.0". Below it is a "Current Version Release Notes" section with a text area containing "Initial stable release.". The main section is titled "> Additional Information" and contains a description "Additional information about the software.". To the right of this section is a "Development Status" dropdown menu with the value "Select Category". Below this is a "Is Source Code Of" section with a text area containing "Bigger Application". Below that is a "Is Part Of" section with a text area containing "Bigger Suite". At the bottom left, there is a "Save draft" button. At the bottom right, there is a "Save and push to repository" button, which is highlighted with a red rectangle. A blue circular help icon is located at the bottom right corner of the form.

https://example.com/download/1.0.0

Current Version Release Notes

Initial stable release.

> Additional Information

Additional information about the software.

Development Status

Select Category

Is Source Code Of

Bigger Application

Is Part Of

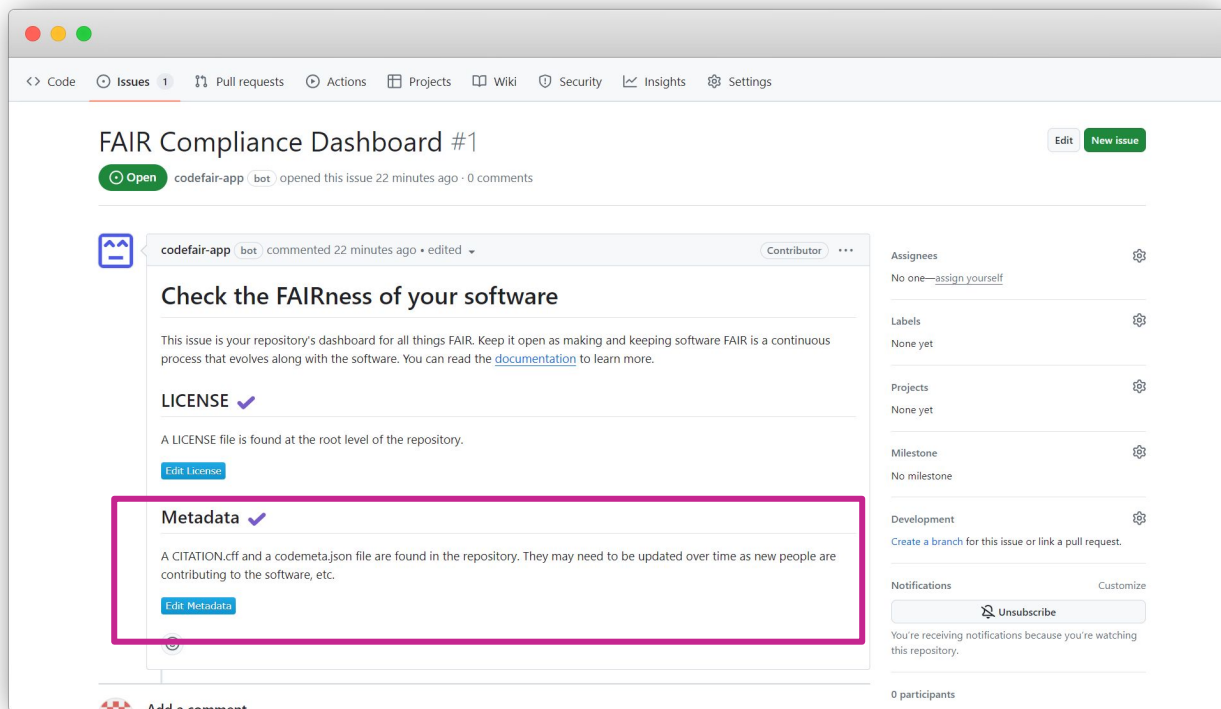
Bigger Suite

Save draft

Save and push to repository

# FAIR Research Software

## Codefair Issue Dashboard



The screenshot shows a GitHub issue titled "FAIR Compliance Dashboard #1" created by the "codefair-app" bot. The issue is currently open and has no comments. The main content of the issue is a checklist for ensuring software is FAIR:

- Check the FAIRness of your software**

This issue is your repository's dashboard for all things FAIR. Keep it open as making and keeping software FAIR is a continuous process that evolves along with the software. You can read the [documentation](#) to learn more.
- LICENSE ✓**

A LICENSE file is found at the root level of the repository.  
[Edit License](#)
- Metadata ✓**

A CITATION.cff and a codemeta.json file are found in the repository. They may need to be updated over time as new people are contributing to the software, etc.  
[Edit Metadata](#)

The "Metadata" section is highlighted with a red rectangle. On the right side of the issue, there are sections for Assignees (None yet), Labels (None yet), Projects (None yet), Milestone (No milestone), Development (Create a branch for this issue or link a pull request.), Notifications (Unsubscribe), and 0 participants.



# Closing Comments

# Closing Comments



Requiring researchers to make their research outcomes FAIR can rapidly become **unfair** to them



FAIR can be made more **fair** to the researchers through actionable guidelines and user-friendly tools for implementing them

# Thank You!



bpatel@calmi2.org



fairdataihub.org

*Find these slides and  
all resources here*



[bit.ly/fairfair24](https://bit.ly/fairfair24)