Introduction to Software Sustainability

DH & RSE Summer School, 27 July 2021
Day 2: Software Sustainability

Dr Mary Chester-Kadwell
Cambridge Digital Humanities
Cambridge University Library

What sort of thing is Code anyway?

methodology

steps to reproduce analysis

citable publication

tool to get something done

REF-able artefact

research output

software package

personal workings

experiment

…?

…?

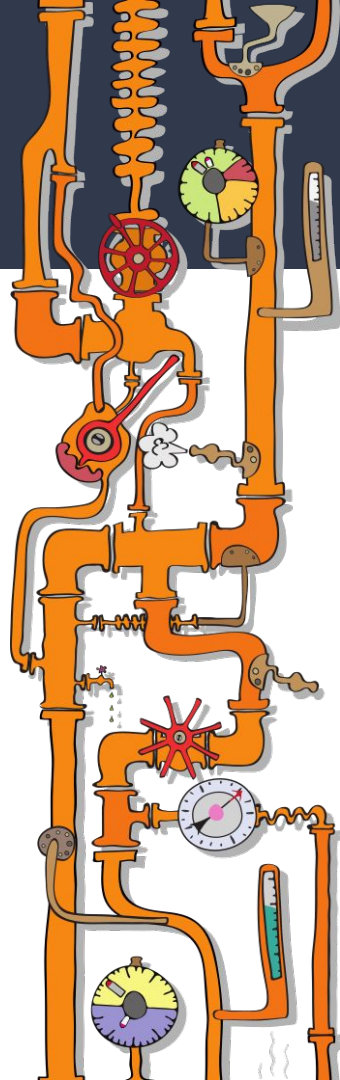
scripts hacked together

The Afterlife of Research Code

- What happens to your code when your project is finished?
- Will someone be able to:
  - Find your code?
  - Use your code?
- Why would you want them to?

What is ‘Good’ Research Code?

- **Open?**
  - Open access, standards, data
- **3 R’s?**
  - Repeatable, Reproducible, Reusable
- **Usable?**
- **Sustainable?**



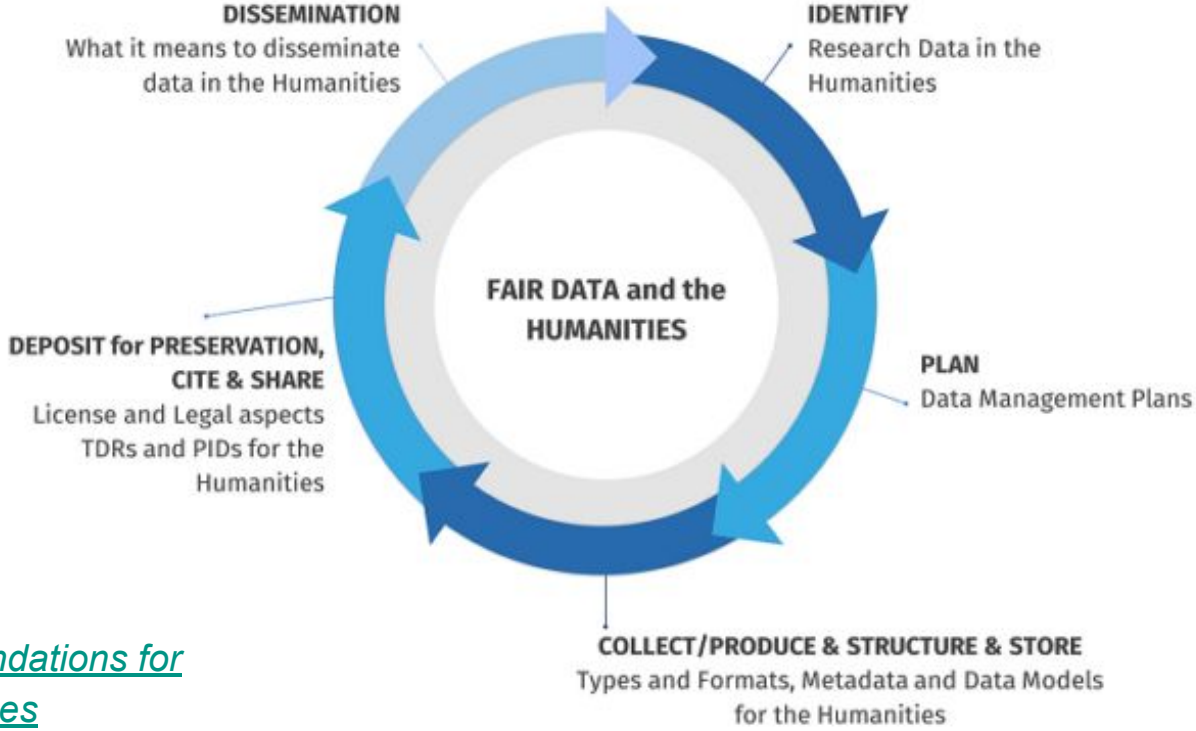
Usability & Sustainability

Criterion	Sub-criterion	Notes – to what extent is/does the software...
Usability	Understandability	Easily understood?
	Documentation	Comprehensive, appropriate, well-structured user documentation?
	Buildability	Straightforward to build on a supported system?
	Installability	Straightforward to install on a supported system?
	Learnability	Easy to learn how to use its functions?
Sustainability and maintainability	Identity	Project/software identity is clear and unique?
	Copyright	Easy to see who owns the project/software?
	Licencing	Adoption of appropriate licence?
	Governance	Easy to understand how the project is run and the development of the software managed?
	Community	Evidence of current/future community?
	Accessibility	Evidence of current/future ability to download?
	Testability	Easy to test correctness of source code?
	Portability	Usable on multiple platforms?
	Supportability	Evidence of current/future developer support?
	Analysability	Easy to understand at the source level?
	Changeability	Easy to modify and contribute changes to developers?
	Evolvability	Evidence of current/future development?
	Interoperability	Interoperable with other required/related software?

Source: https://software.ac.uk/sites/default/files/SSI-SoftwareEvaluationCriteria.pdf

What is FAIR?

- Findable
- Accessible
- Interoperable
- Reusable



ALLEA Report: *Let's be FAIR! Recommendations for Sustainable Data Sharing in the Humanities*

FAIR for Research Code: Simple Recommendations

Findable	Identifier & public registry
Accessible	Public repository
Interoperable	Metadata & standards
Reusable	License & documentation

- Five Recommendations for FAIR Software
- Society for RSE: SORSE event FAIR 4 Research Software

Some Helpful Links

- Making Your Code Citable
- How to cite and describe software
- Some Software Registries
- Choosing a repository for your software project
- ChooseALicense.com
- 5 Recommendations for FAIR Software
- FAIR 4 Research Software events (videos)
- University of Cambridge Office for Scholarly Communication (many resources on related topics)
- Society of Research Software Engineering (not just for ‘engineers’; friendly and helpful)

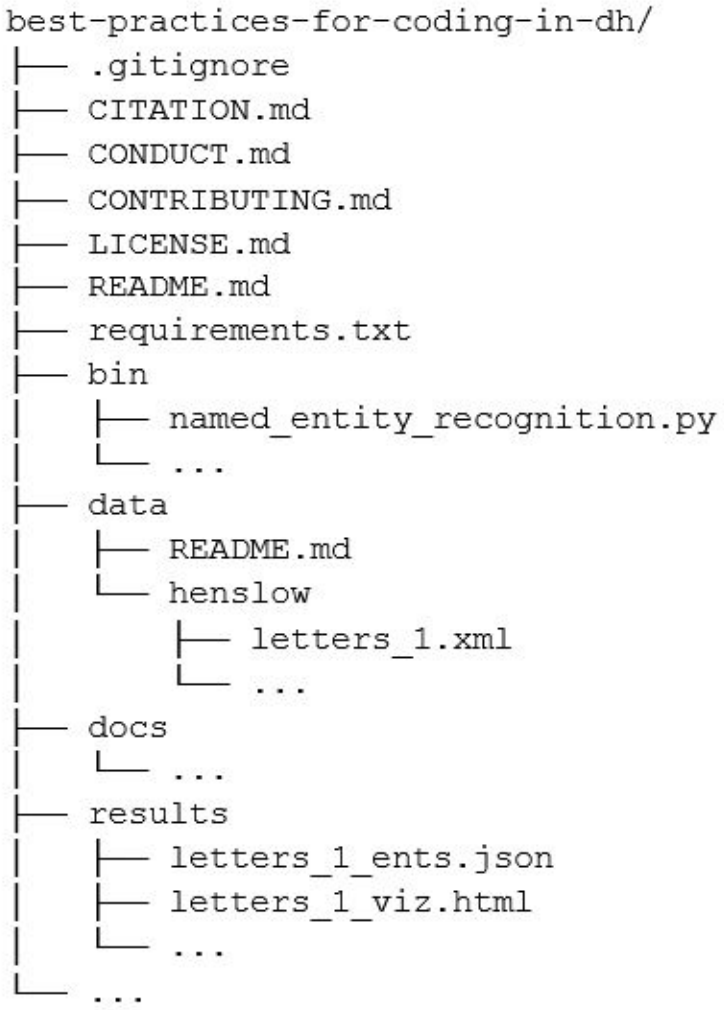
Version Control

- What is version control?
- What is it for?
- What sort of version control programmes are available/used?
- What is the difference between git and GitHub?

Project Structure

Noble, William Stafford. 2009. “A Quick Guide to Organizing Computational Biology Projects.” PLoS Computational Biology 5 (7): e1000424. https://doi.org/10.1371/journal.pcbi.1000424

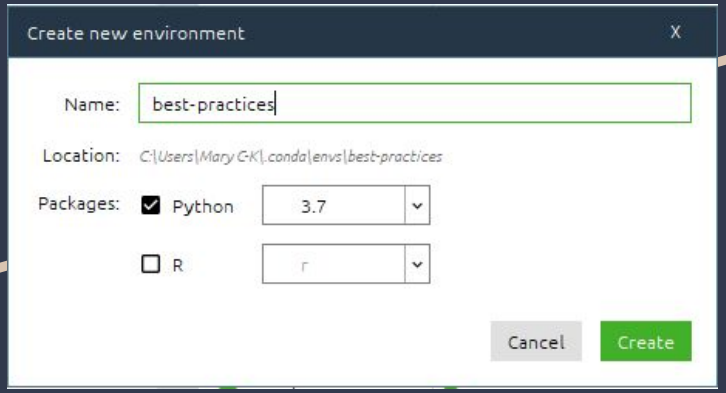
Research Software Engineering with Python
https://merely-useful.tech/py-rse/getting-started.html#getting-started-structure



Dependencies

- What are dependencies?
- Why should you care about recording project dependencies?
- How can you manage dependencies?
- All about `requirements.txt`

Virtual Environments



- What is a virtual environment?
- Why should you care about creating virtual environments to run your code?
- When should you use virtual environments? (Hint: always!)

Reusable Programs

- **Command-line programs**:
  - Automation
  - Reusability

```
(venv) $ python bin/named_entity_recognition_4.py -s
data/henslow/letters_152.xml

Irish: NORP
Arbutus: PRODUCT
Trichomanes: ORG
first: ORDINAL
Wilson: PERSON
Spring: DATE
Summer: DATE
Wilson: ORG
a few days ago: DATE
```

Usable Programs

- Graphical User Interface (**GUI**) programs:
  - Usability