R: A Hitchhikers Guide to Reproducible Research

- Wise up

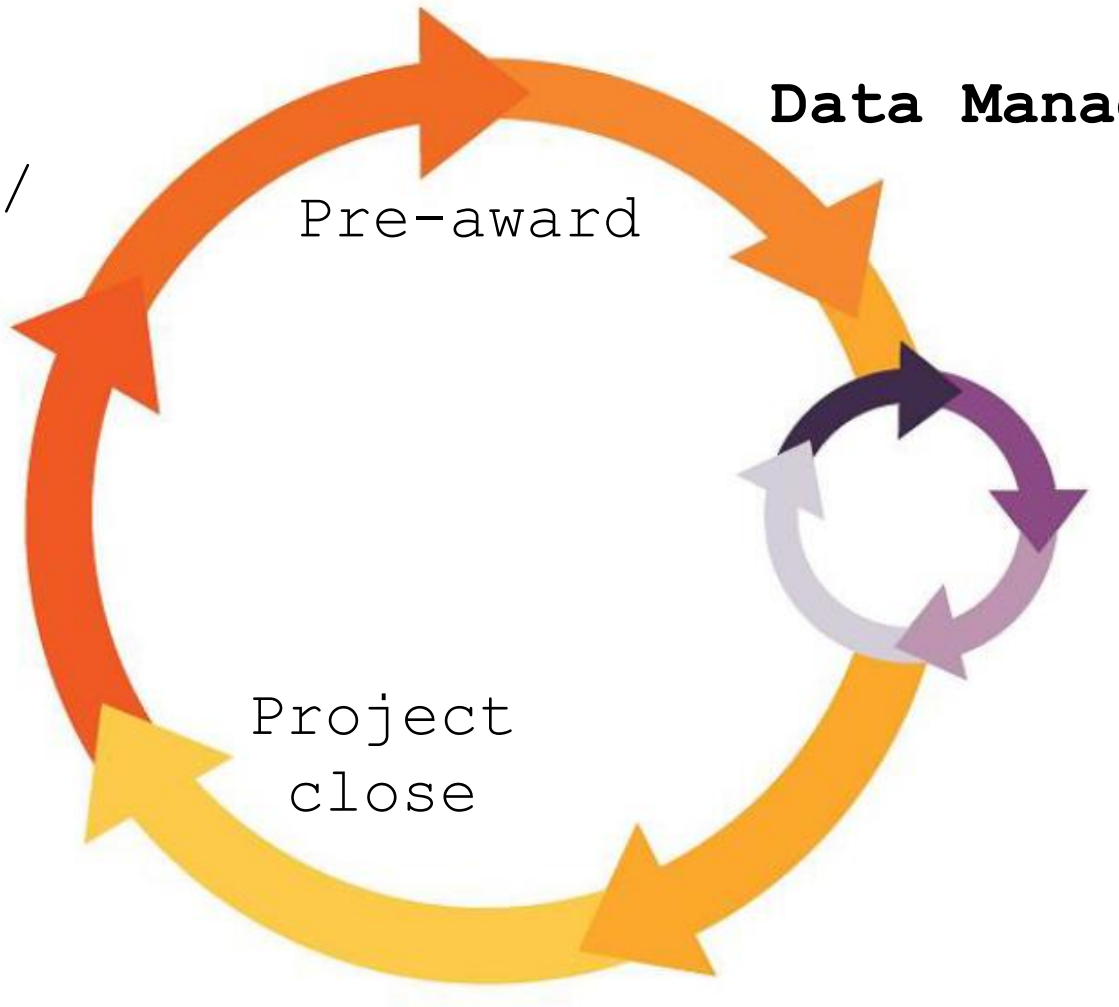
Brendan Palmer,
Clinical Research Facility - Cork &
School of Public Health
@B_A_Palmer

CRF-C
HRB Clinical Research Facility Cork

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

School of
Public Health

Plan your research



Data Ethics Canvas

2019-05

Data sources

Name/describe your project's key data sources, whether you're collecting data yourself or accessing via third parties.

Is any personal data involved, or data that is otherwise sensitive?

Limitations in data sources

Are there limitations that could influence your project's outcomes?

Consider:

- bias in data collection, inclusion/exclusion, analysis, algorithms
- gaps or omissions in data
- provenance and data quality
- other issues affecting decisions, such as team composition

Sharing data with others

Are you going to be sharing data with other organisations? If so, who?

Are you planning to publish any of the data? Under what conditions?

Ethical and legislative context

What existing ethical codes apply to your sector or project? What legislation, policies, or other regulation shape how you use data? What requirements do they introduce?

Consider: the rule of law; human rights; data protection; IP and database rights; anti-discrimination laws; and data sharing, policies, regulation and ethics codes/frameworks specific to sectors (eg health, employment, taxation).

Rights around data sources

Where did you get the data from? Is it produced by an organisation or collected directly from individuals?

Was the data collected for this project or for another purpose? Do you have permission to use this data, or another basis on which you're allowed to use it? What ongoing rights will the data source have?

Your reason for using data

What is your primary purpose for collecting and using data in this project?

What are your main use cases? What is your business model?

Are you making things better for society? How and for whom?

Are you replacing another product or service as a result of this project?

Communicating your purpose

Do people understand your purpose – especially people who the data is about or who are impacted by its use?

How have you been communicating your purpose? Has this communication been clear?

How are you ensuring more vulnerable individuals or groups understand?

Positive effects on people

Which individuals, groups, demographics or organisations will be positively affected by this project? How?

How are you measuring and communicating positive impact? How could you increase it?

Negative effects on people

Who could be negatively affected by this project?

Could the way that data is collected, used or shared cause harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)?

How are limitations and risks communicated to people? Consider: people who the data is about, people impacted by its use and organisations using the data.

Minimising negative impact

What steps can you take to minimise harm?

How could you reduce any limitations in your data sources? How are you keeping personal and other sensitive information secure?

How are you measuring, reporting and acting on potential negative impacts of your project?

What benefits will these actions bring to your project?

Engaging with people

How can people engage with you about the project?

How can people correct information, appeal or request changes to the product/service? To what extent?

Are appeal mechanisms reasonable and well understood?

Openness and transparency

How open can you be about this project? Could you publish your methodology, metadata, datasets, code or impact measurements?

Can you ask peers for feedback on the project? How will you communicate it internally?

Will you publish your actions and answers to this canvas openly?

Ongoing implementation

Are you routinely building in thoughts, ideas and considerations of people affected in your project? How?

What information or training might be needed to help people understand data issues?

Are systems, processes and resources available for responding to data issues that arise in the long-term?

Reviews and iterations

How will ongoing data ethics issues be measured, monitored, discussed and actioned?

How often will your responses to this canvas be reviewed or updated? When?

Your actions

What actions will you take before moving forward with this project? Which should take priority?

Who will be responsible for these actions, and who must be involved?

Will you openly publish your actions and answers to this canvas?

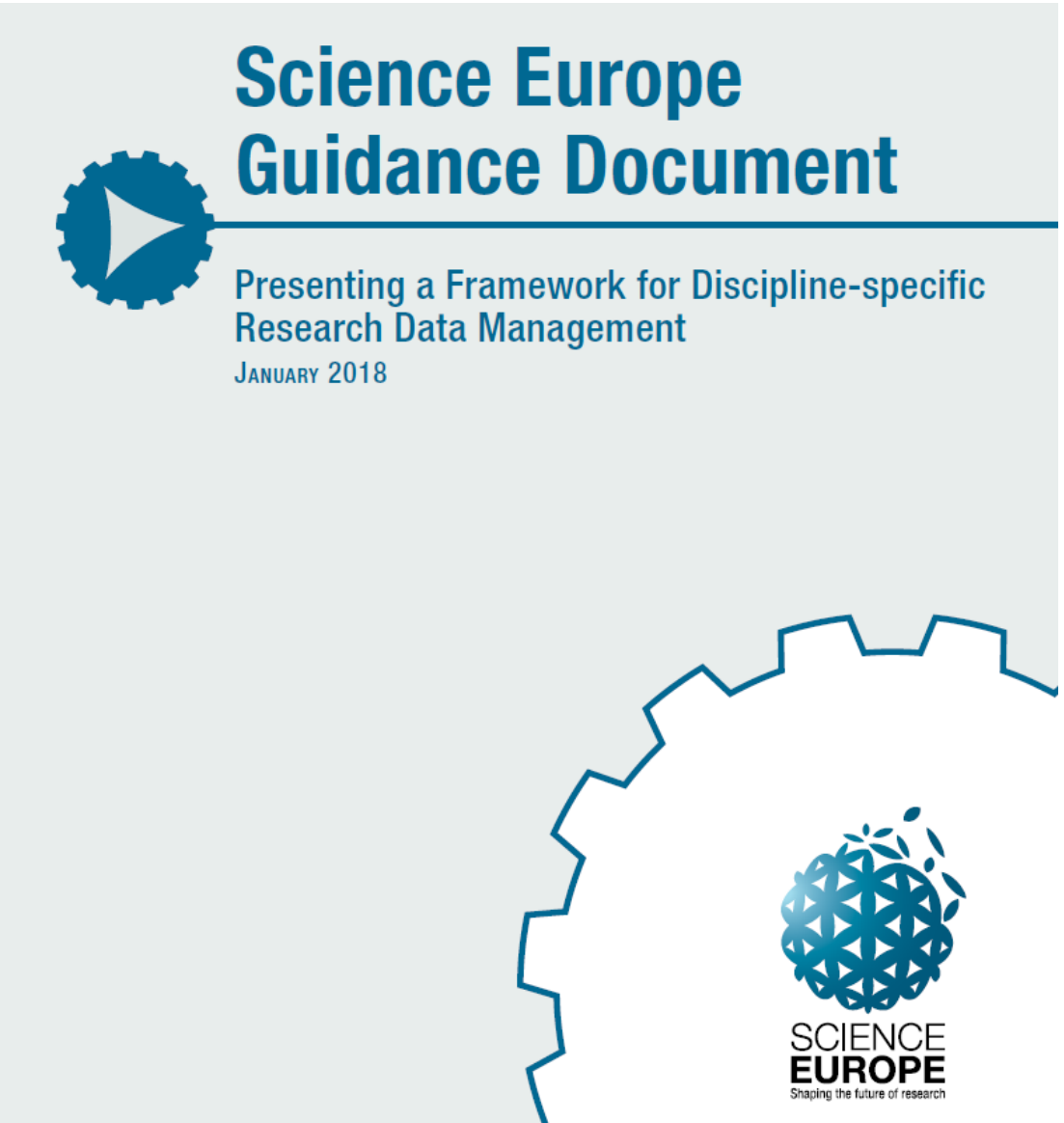
Plan your research

Health Research Board

HRB Policy on Management and Sharing of Research Data

For data gathered and generated in whole or in part from HRB-funded research, the following policy will be adhered to with effect from 1st of January 2020.

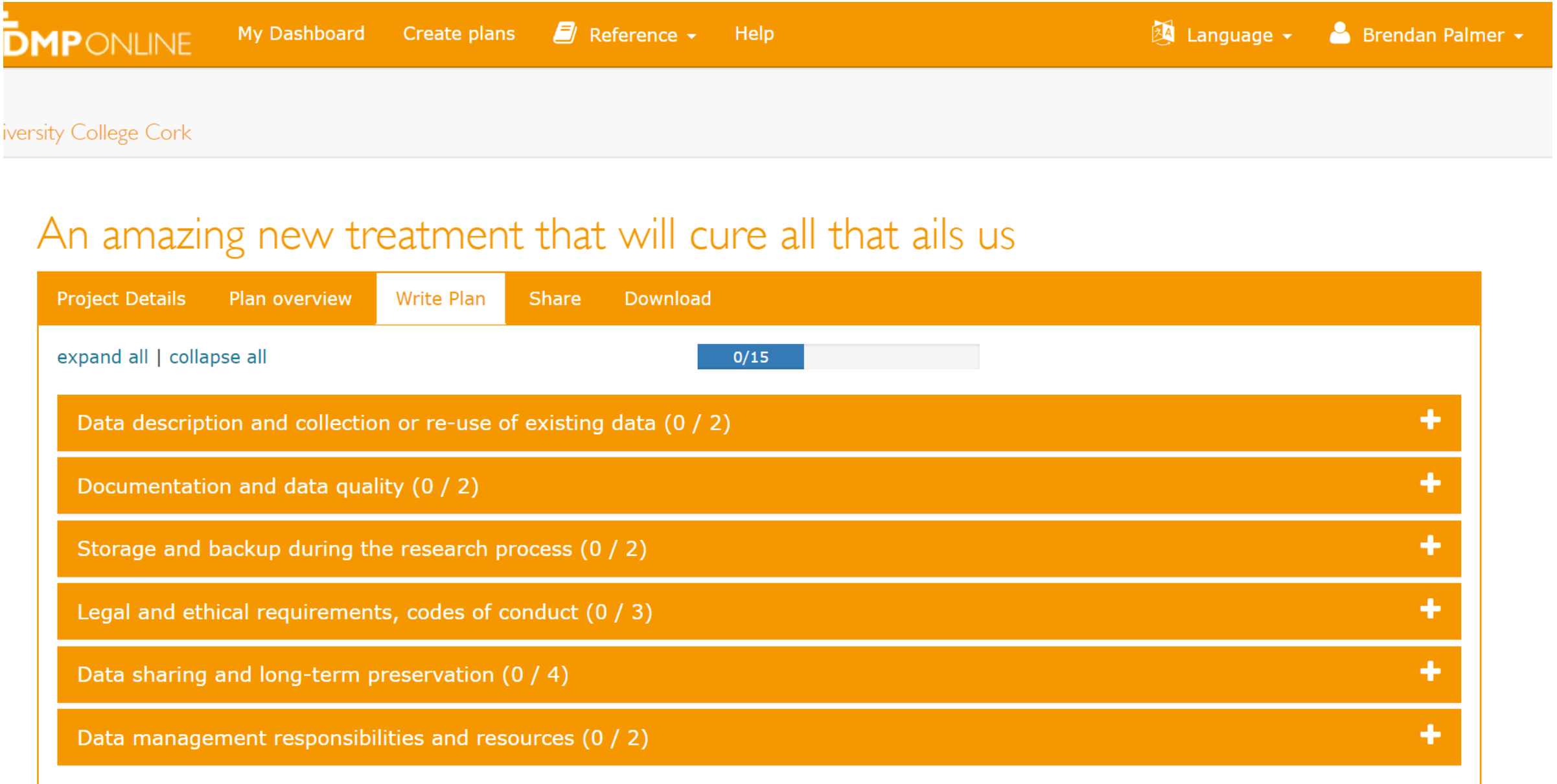
It all starts with a Data Management Plan



Seven main headings:

1. Data collection
2. Documentation and Meta-data
3. Ethics and Legal Compliance
4. Storage and Backup
5. Selection and Preservation
6. Data Sharing
7. Responsibilities and Resources

Taking those first steps



Taking those first steps

An amazing new treatment that will cure all that ails us

Project Details | Plan overview | Write Plan | Share | Download

expand all | collapse all

0/15

Data description and collection or re-use of existing data (0 / 2)

Within this section, we ask that you outline your plans for creating and managing data

How will new data be collected or produced and/or how will existing data be re-used?

B *I*

Save

Guidance | Comments

HRB Ireland | DCC

Explain which methodologies or software will be used if new data are collected or produced and specify which community standards (if any) will be used.

State any constraints on re-use of existing data if there are any.

Explain how data provenance will be documented.

Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.

Taking those first steps

An amazing new treatment that will cure all that ails us

Project Details | Plan overview | Write Plan | Share | Download

expand all | collapse all

0/15

Data description and collection or re-use of existing data (0 / 2)

Within this section, we ask that you outline your plans for creating and managing data

How will new data be collected or produced and/or how will existing data be re-used?

B *I*

Save

Guidance | Comments

HRB Ireland | DCC

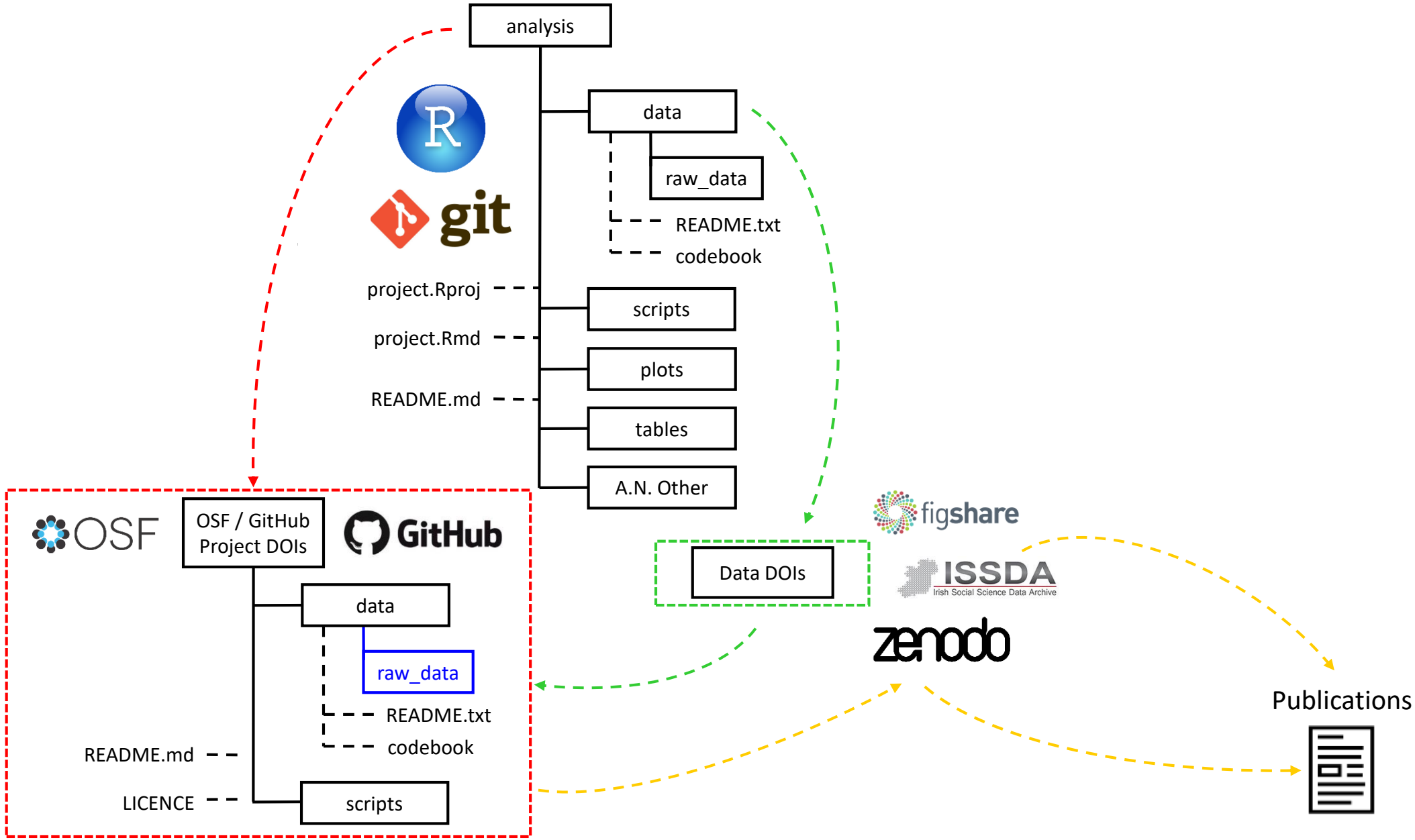
Explain which methodologies or software will be used if new data are collected or produced and specify which community standards (if any) will be used.

State any constraints on re-use of existing data if there are any.

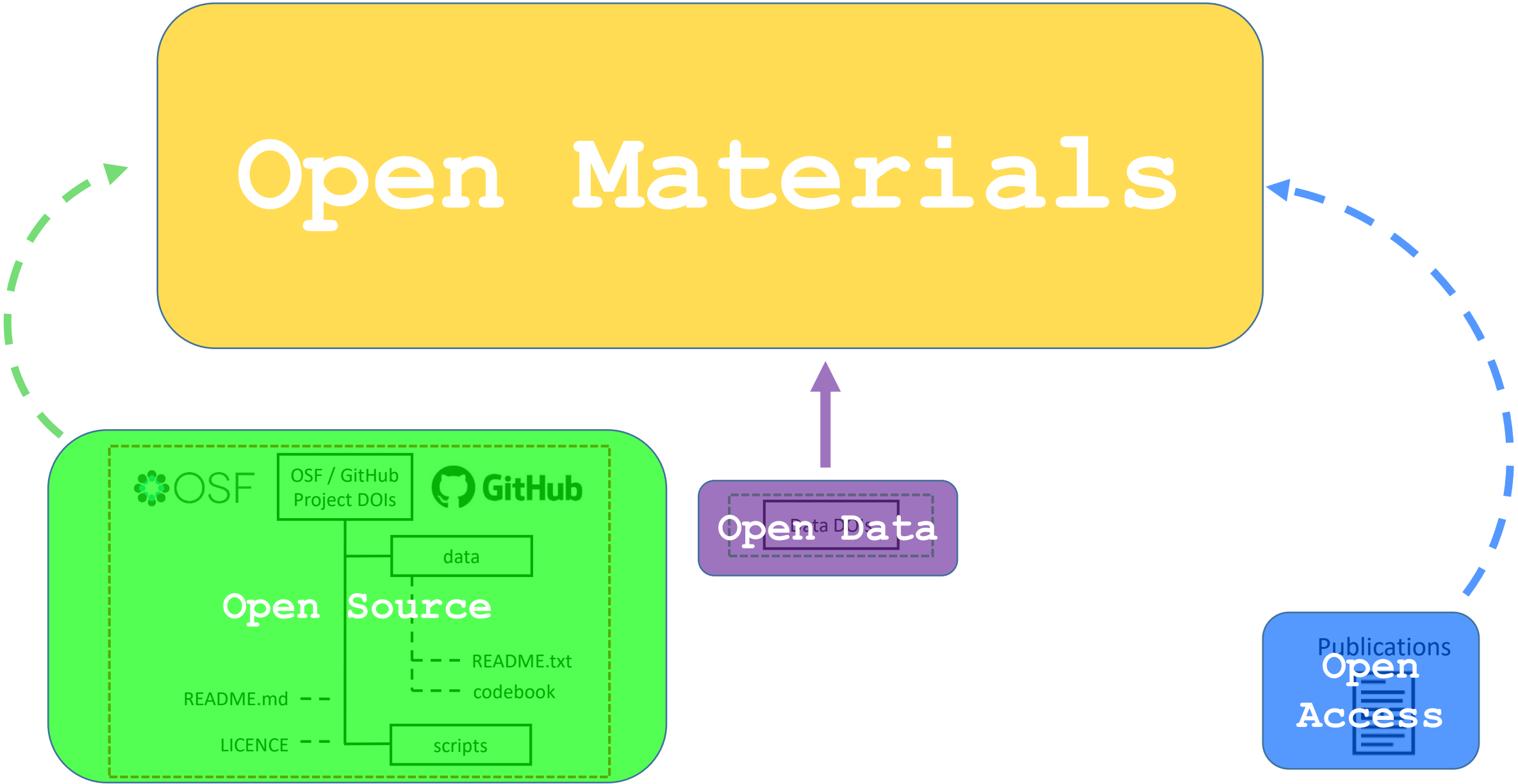
Explain how data provenance will be documented.

Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.

What does this allow us to do?



What does this allow us to do?



meetup

Cork (Ireland) R-Users Group