



F.A.I.R. Data Management

DELAD Workshop

Eoghan Ó Carragáin, Research & Digital Services, Boole Library

16th November, 2017

**A TRADITION OF
INDEPENDENT
THINKING**



UCC

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

@UCC – supports

IT Services:

- A wide range of cloud and on-premises storage & compute solutions
 - Research-friendly contracts and data agreements with major cloud providers, e.g. AWS, Azure, etc.
 - New “research NAS” available soon
- Advice & support on operational security, backups, data encryption
- Access to third-party services, e.g. HEAnet Filesender

UCC Library:

- Cork Open Research Archive (CORA)
- Service to mint DataCite DOIs for your datasets
- Working with IT Services on general purpose “Data Vault”
- Reviews of funder-requested F.A.I.R. Data Management Plans (via RSS)
- Working to standardise expectations of Irish funders through the National Open Research Forum

OVPRI:

- Research Integrity training

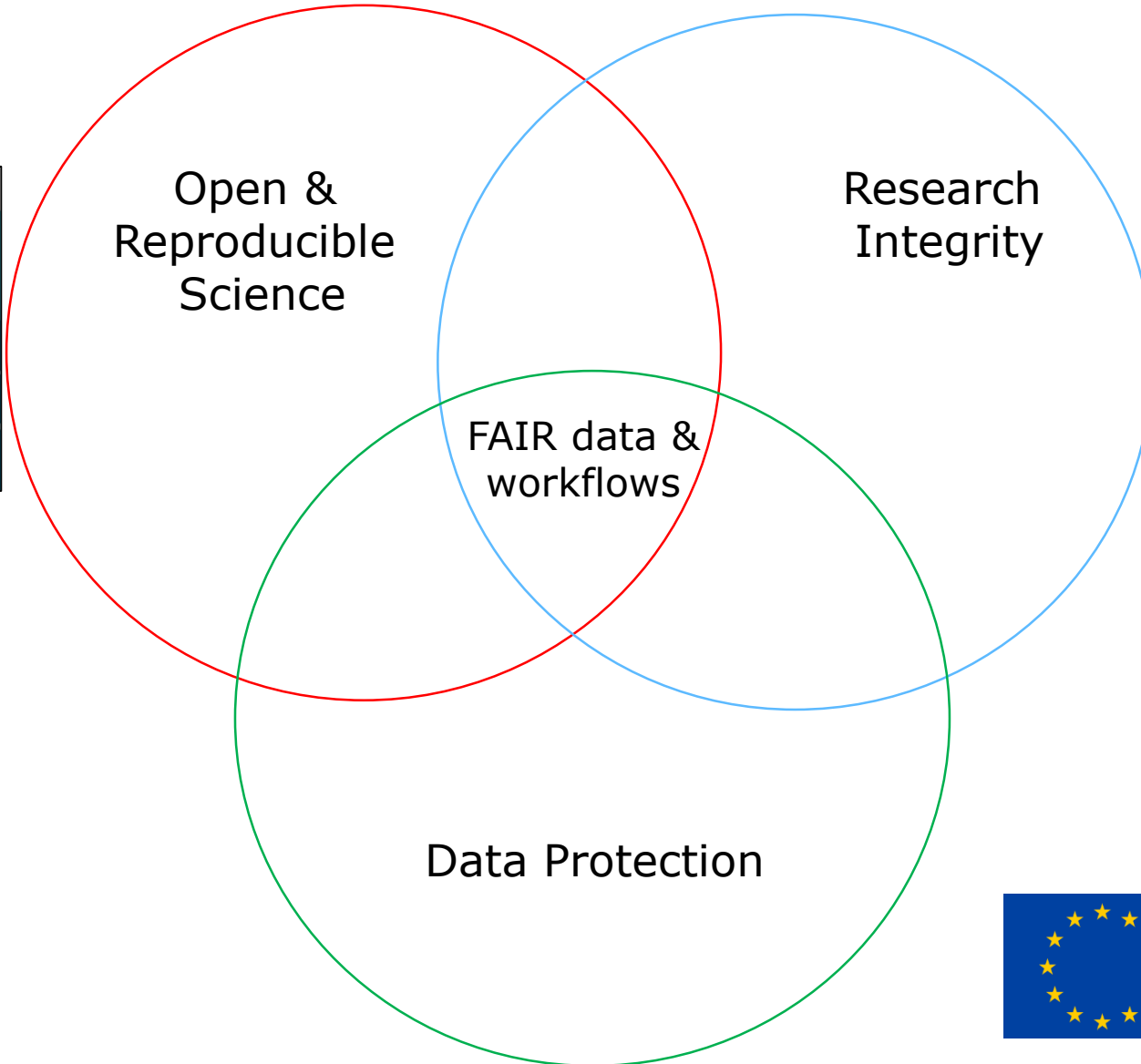
Other:

- R Workshops & symposia run by CRF-C (Darren Dahly, Brendan Palmer)

Overview

- Some context
- F.A.I.R. Data Guidelines
- Your Data

RDM trends



The need for Data Stewardship

- 90% of scientific data generated in the last 2 years
- Only 12% of NIH-funded data makes it to data repositories
 - Report of HLEG on European Open Science Cloud, 2016
- Requesting data from the author and journal supplementary data are insufficient

RESEARCH ARTICLE

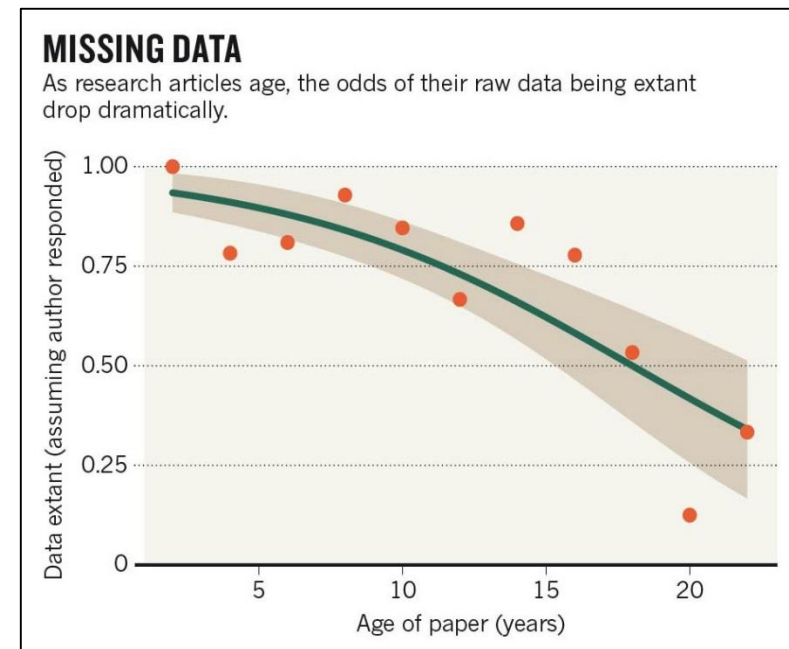
Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot

Martin Klein, Herbert Van de Sompel, Robert Sanderson, Harihar Shankar, Lyudmila Balakireva, Ke Zhou, Richard Tobin

RESEARCH ARTICLE

Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content

Shawn M. Jones, Herbert Van de Sompel, Harihar Shankar, Martin Klein, Richard Tobin, Claire Grover



Vines, T. H. et al. doi: 10.1016/j.cub.2013.11.014

Data/Code as 1st class research output

“Via the REF, we fully recognise **data as an equally valid form of research output**, and, through our open access policy for the next REF, we plan to reward research environments that deliver open access to a **wider set of outputs than just journal articles and conference papers.**”

<http://blog.hefce.ac.uk/2015/09/01/opening-up-research-data/>

COMING SOON

HRB Open Research

A platform for HRB-funded researchers to publish their research outputs in an open and accessible way

What research will this platform publish?

HRB researchers can use the platform to publish standard research articles, clinical trial findings, systematic reviews, study protocols, data sets, negative/null results, case reports and many other original submissions.

H2020: Open by Default

CHALLENGE

Wider access to scientific facts and knowledge helps researchers, innovators and the public find and re-use data, and check research results:

offers better value for EU research funds



a public benefit

encourages research across scientific fields



essential for solving today's complex societal challenges

SOLUTION

Horizon 2020 already mandates open access to all scientific publications

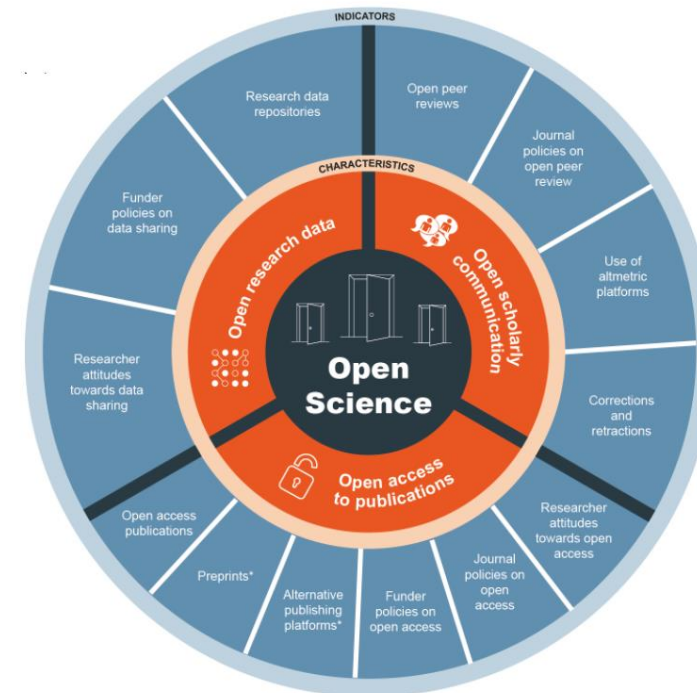


From 2017,
research data is
open by default,
with possibilities to opt out

Open Science – Policy platform

European Commission:

- All research publications Open Access by 2020
- All research data Open by Default (from Jan. 2017)
- €6.7 billion to be invested in “European Open Science Cloud”
 - INFRAEOSC H2020 calls from December 2017
- “Openness” expected to be a key pillar for FP9
- Recommendation to train 500K “core data experts”/“data stewards” in Europe over next 10 years
- High-level expert groups on:
 - European Open Science Cloud
 - Skills & Rewards for Open Science
 - New business models for scholarly communication
 - F.A.I.R. Data
 - Altmetrics



Go FAIR or go home



A set of Digital Object Compliance principles that describes the properties of digital objects that enables them to be **findable**, **accessible, interoperable and reproducible (FAIR)**.



A DMP is a **key element** of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a research project. As part of making research data **findable, accessible, interoperable and re-usable (FAIR)**, a **DMP should include information on:**



⁶ Data management should follow the **FAIR guiding principles** (Findability, Accessibility, Interoperability & Reusability). See, for example, Wilkinson, M. D. et al. (2016) *The FAIR Guiding Principles for Scientific Data Management and Stewardship*. Full text: <http://www.nature.com/articles/sdata201618>.



Investigator-led Projects Call, 2017:

7.10. Research Data Management
Describe the approach to data management that will be taken during and after the project, including who will be responsible for data management and data stewardship. Please consider the **FAIR Guiding Principles** for scientific data management and stewardship: Findability, Accessibility, Interoperability, and Reusability¹³.

F.A.I.R. Data Principles

What are the FAIR Data Principles?



Findable – *Assign persistent IDs, provide rich metadata, register in a searchable resource,...*



Accessible – *Retrievable by their ID using a standard protocol, allows for authentication/authorization, metadata remain accessible even if data aren't...*



Interoperable – *Use standard vocabularies, qualified references, shared and broadly applicable language for knowledge representation...*



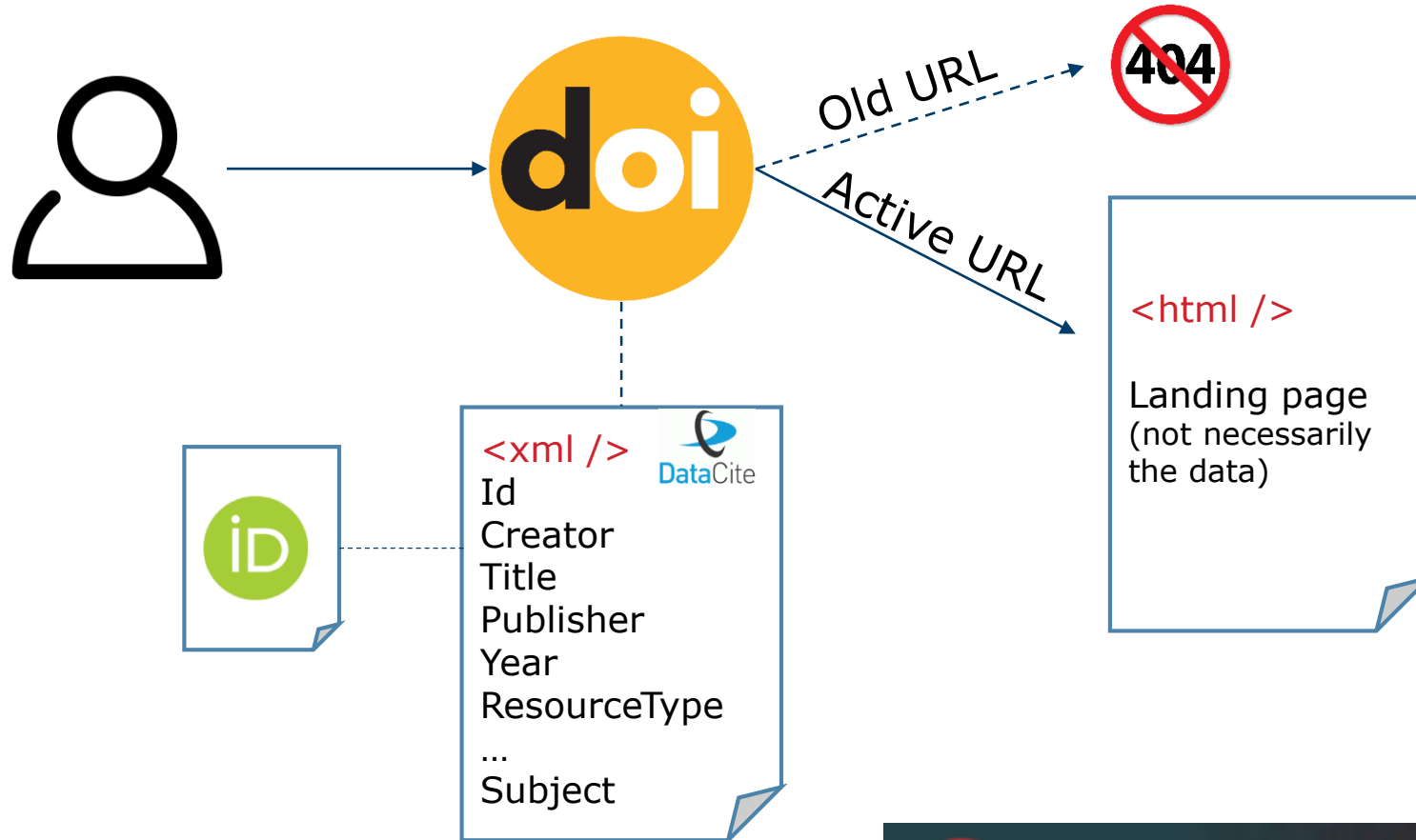
Reusable – *Rich, accurate metadata, clear licences, provenance, use of community standards...*

www.force11.org/group/fairgroup/fairprinciples

N.B.: generally, depositing data in a reputable data repository takes care of F. + A; it also helps with I. + R.



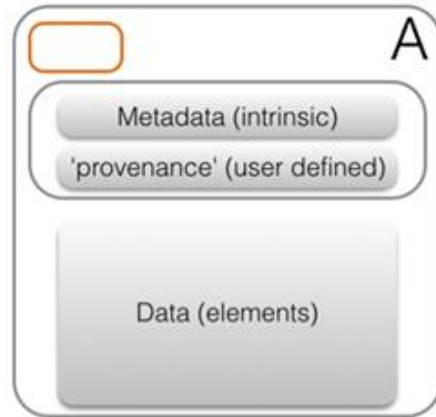
Findable (and citable)



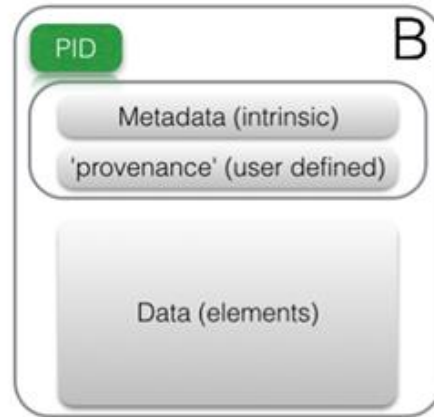


Accessible

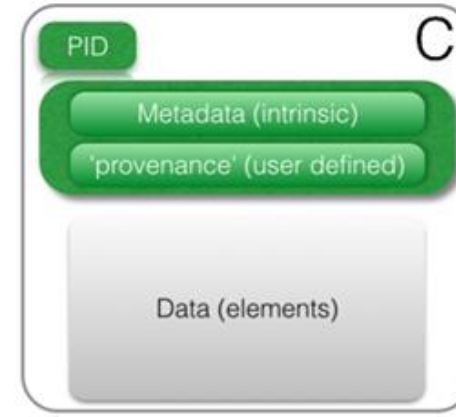
Re-useless data (80%)



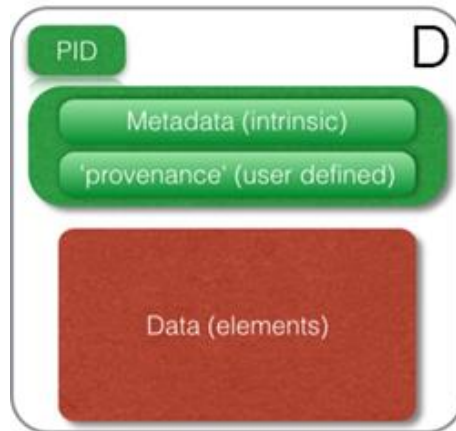
Findable



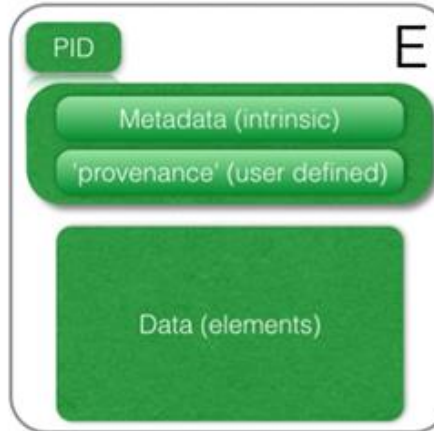
FAIR metadata



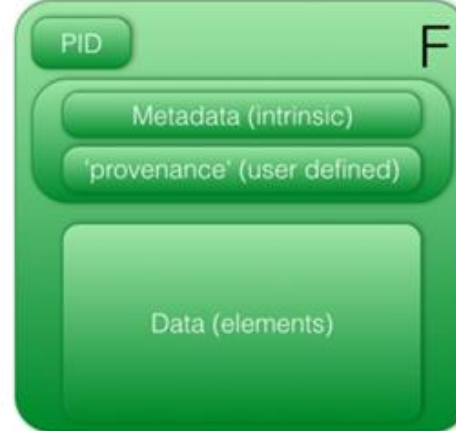
FAIR data-
restricted access



FAIR data-
Open Access



FAIR data-
Open Access/Functionally Linked



“As open as possible, as closed as necessary...”



Interoperability (the hard part!)

or people

The ability of two or more systems or components to exchange information and to use the information that has been exchanged

- IEEE



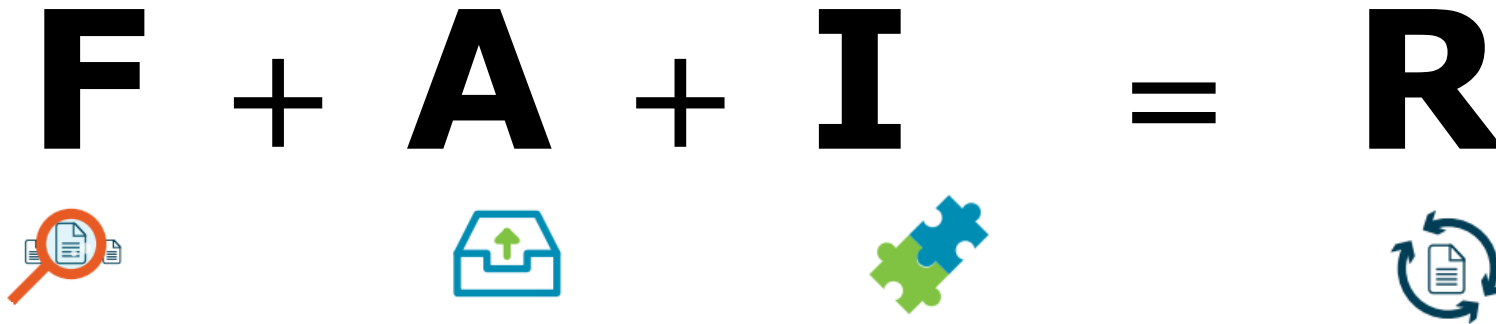
Level	Description	Example
Foundational/ Functional	Low-level data exchange via a standardized communications protocol, without any need to interpret the data	bit streams and data packets, TCP/IP, access protocols
Syntactic/Structural	Packaging of the data via encoding standards, file-formats, etc., so they can be parsed by the receiver	XLS, XML, JSON, CSV "Tidy data" convention for tabular data
Semantic	Receiver can interpret not only the structure but the formal meaning of data and take appropriate actions	Domain convention, e.g. semi-structured code-books; data-dictionaries; minimal reporting requirements W3C Linked Data specifications, e.g. RDF, LDP

Useful to think of it as a spectrum: where appropriate and practical, take steps to allow accurate (and automated) re-use of data by colleagues within and across disciplines



Reusable

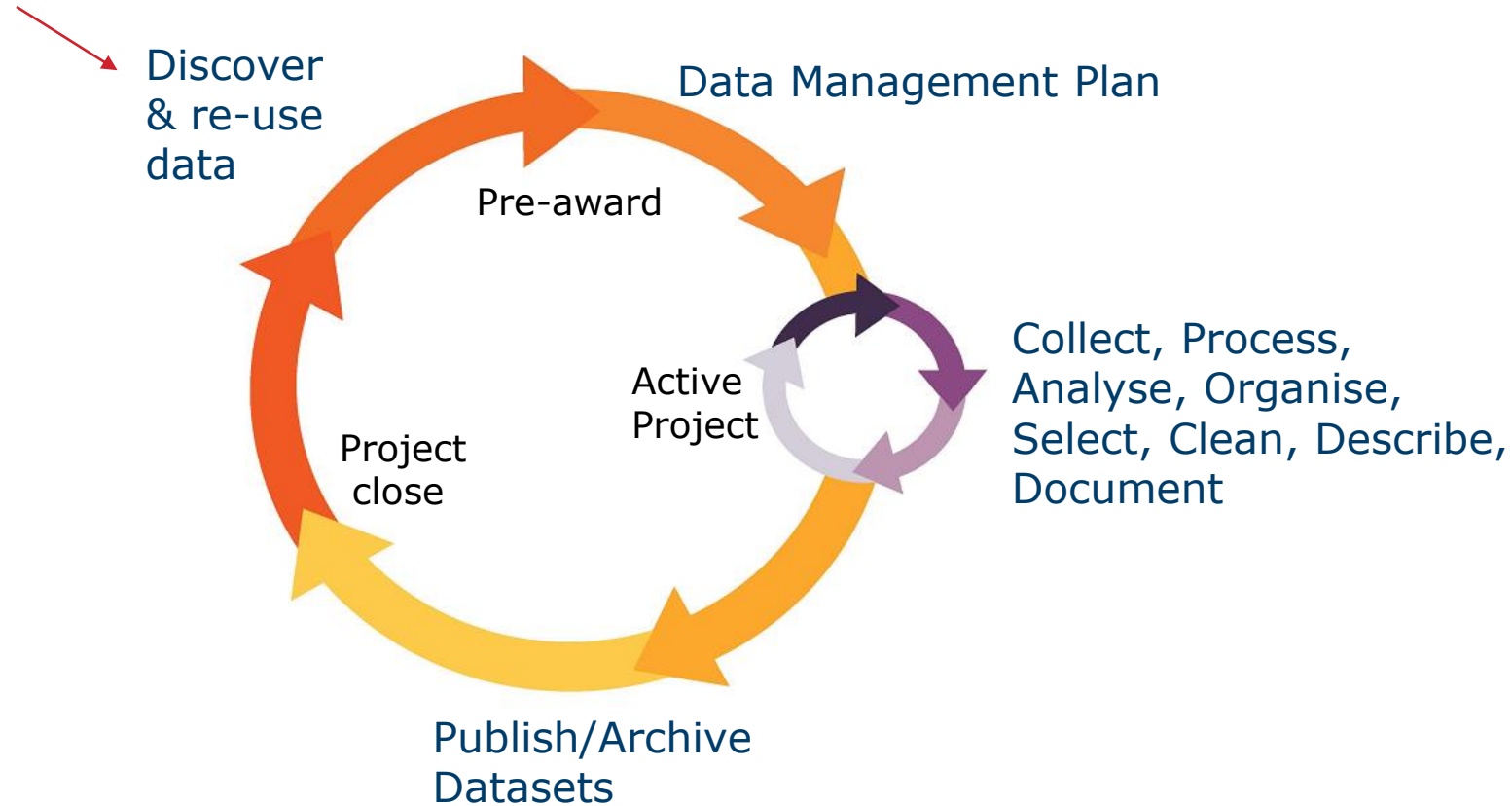
"Rich, accurate metadata, clear licences, provenance, use of community standards..."



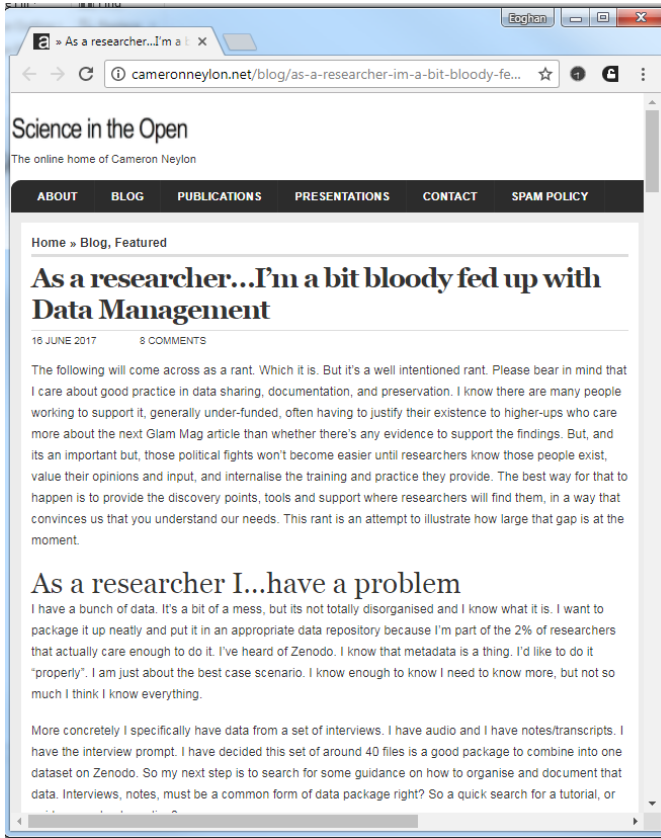
Some tips

Tip: put yourself in their (your future self's) shoes

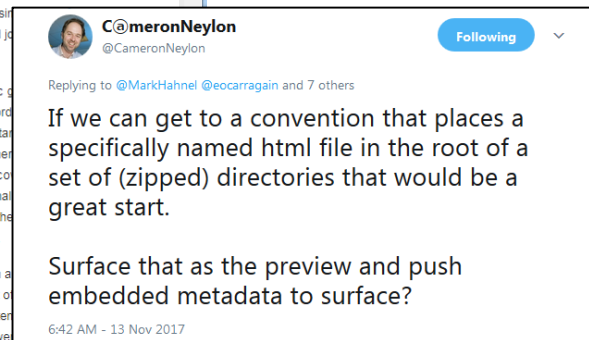
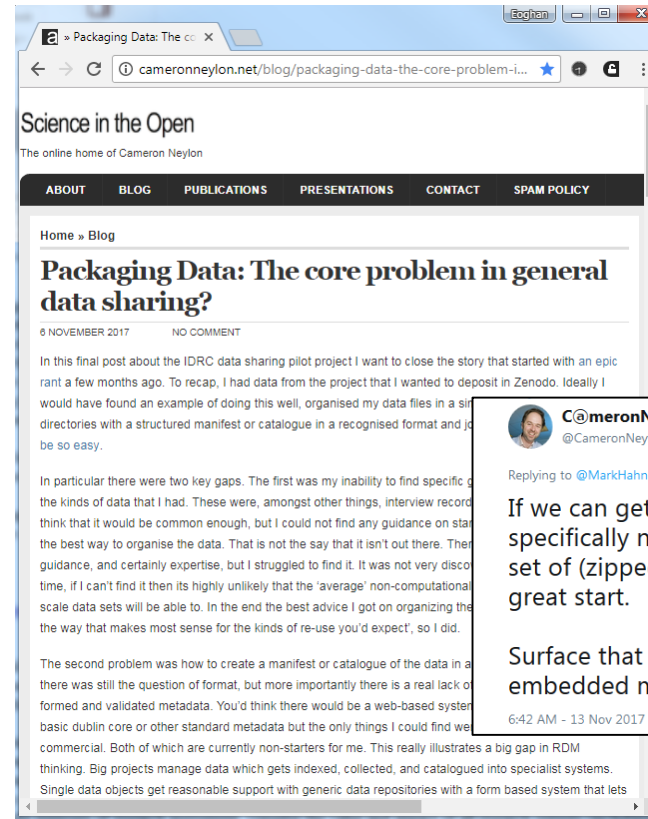
What formats, data organisation, documentation, licensing would **YOU** want/need to understand & re-use **YOUR** data/code/protocol?



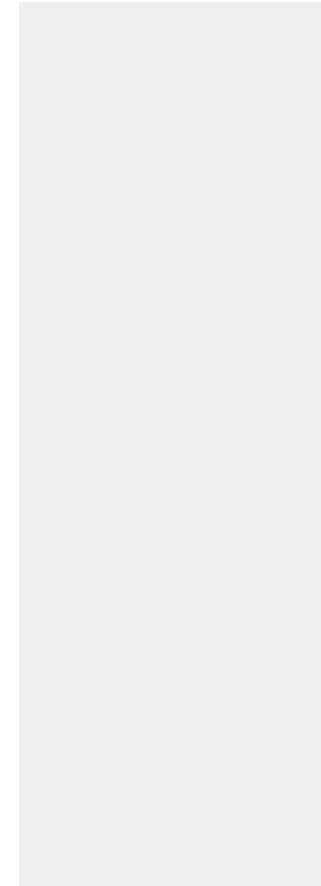
Tip: Don't forget the researcher!



<http://cameronneylon.net/blog/as-a-researcher-im-a-bit-bloody-fed-up-with-data-management/>



Re-use standards and consider a minimal approach that is workable on the ground





Questions?

eoghan.ocarragain@ucc.ie

**A TRADITION OF
INDEPENDENT
THINKING**



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh