

# IDS Metadata and Data Quality

Amy O'Hara  
AISP Network Meeting  
June 20, 2019



---

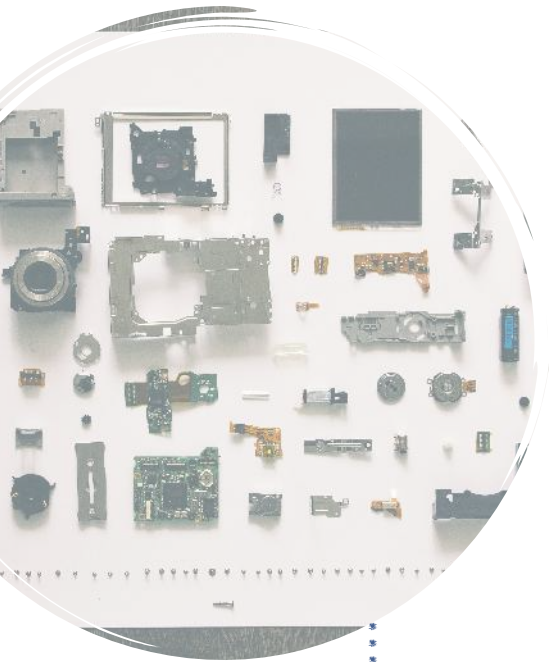
McCourt School *of* Public Policy



# Overview

Metadata  
Data quality

Explain concepts  
Describe examples



# Metadata

Design metadata system for users  
of the collection

Types of users?

Strategic, analytical, operational  
Program administration, compliance  
Research, evaluation



## Types of metadata

Descriptive  
Provenance  
Technical  
Rights  
Preservation  
Citation



# Technical

Check-in to files and layouts  
Item level fill rate and ranges  
Joinable as planned  
Automate production  
Change detection

# Guides

[https://www.ands.org.au/\\_data/assets/pdf\\_file/0004/728041/Metadata-Workinglevel.pdf](https://www.ands.org.au/_data/assets/pdf_file/0004/728041/Metadata-Workinglevel.pdf)



ANDS Guide



## Metadata

Level: Working

Last updated: December 2016

Web link: [www.ands.org.au/guides/metadata-working](http://www.ands.org.au/guides/metadata-working)

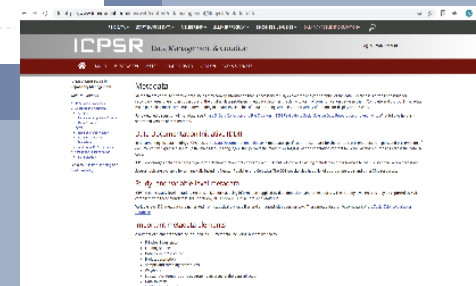
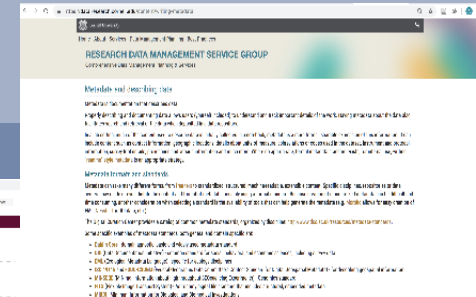
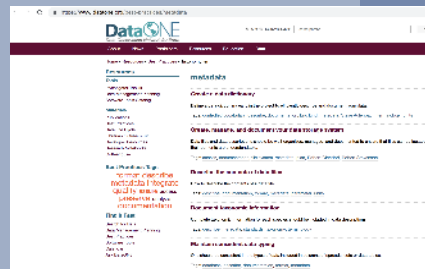
This Guide is intended to provide a simple generic working level view of the needs, issues, and processes around metadata collection and creation as it relates to research data. For those needing information and advice about the RDCS metadata schema used for Research Data Australia, please see the [Content Providers Guide](#).

### Table of Contents

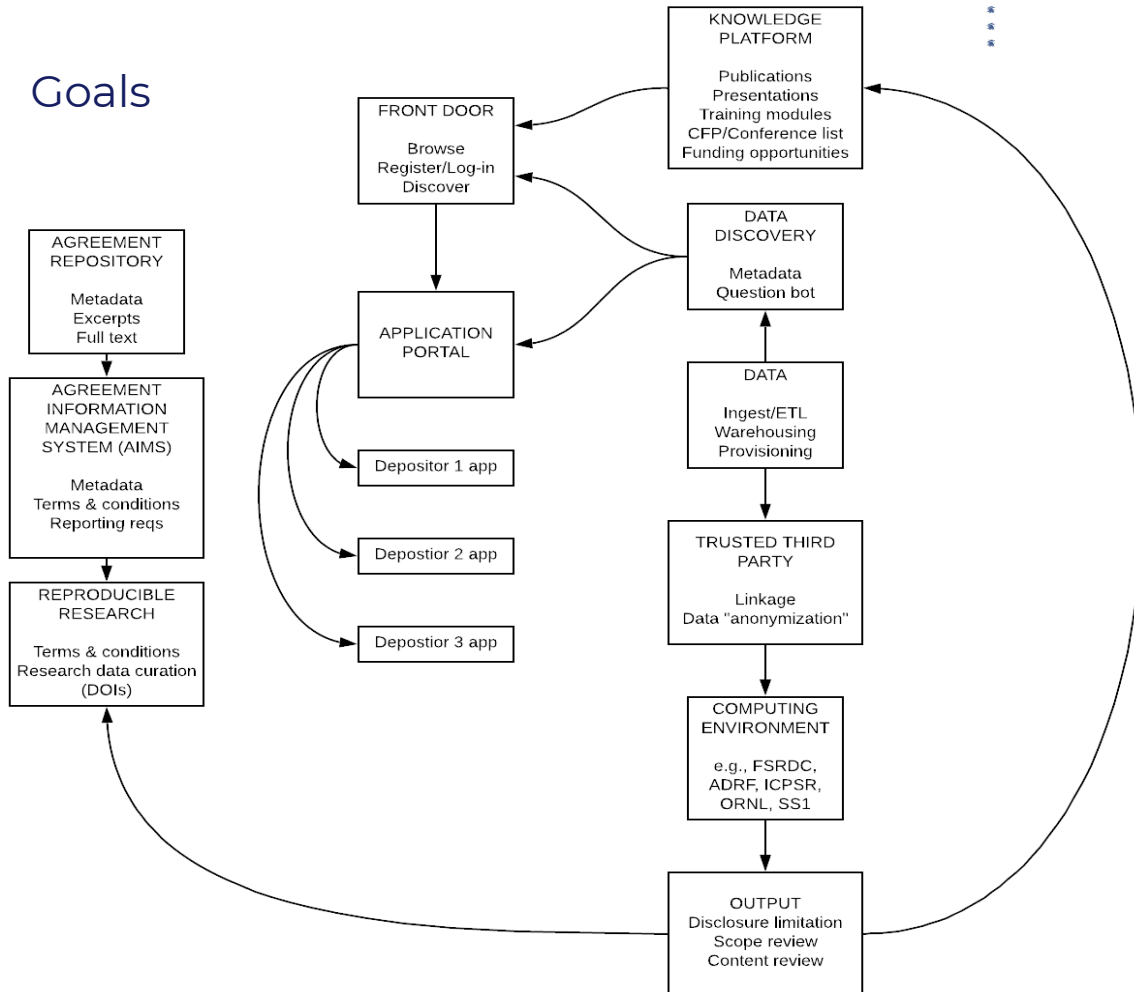
Metadata	1
Key points	2
What is metadata?	2
Metadata for research data	3
Why create metadata?	3
Finding data	3
Determining the value of data	4
Accessing data	4
Using and reusing data	4
Levels of metadata	4
Types of metadata	5
The lingua franca of metadata	6

ands.org.au

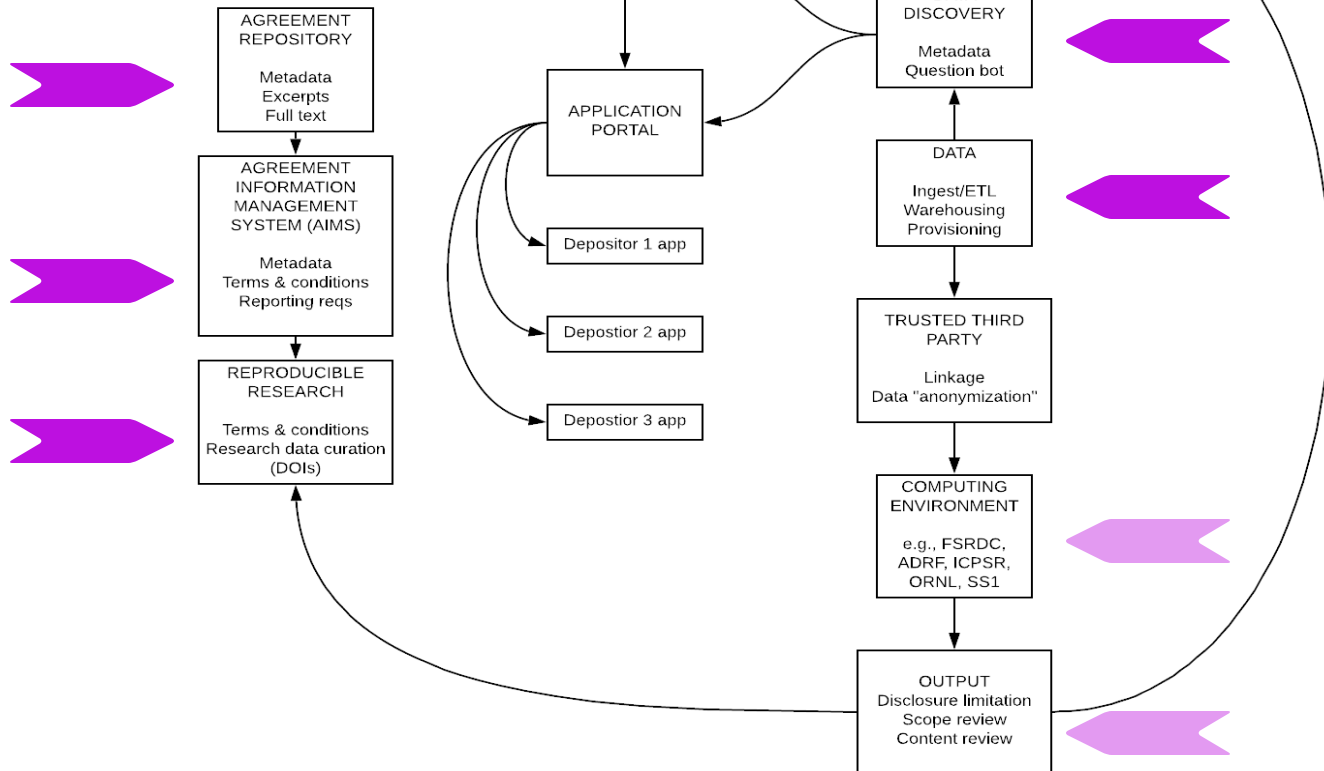
1



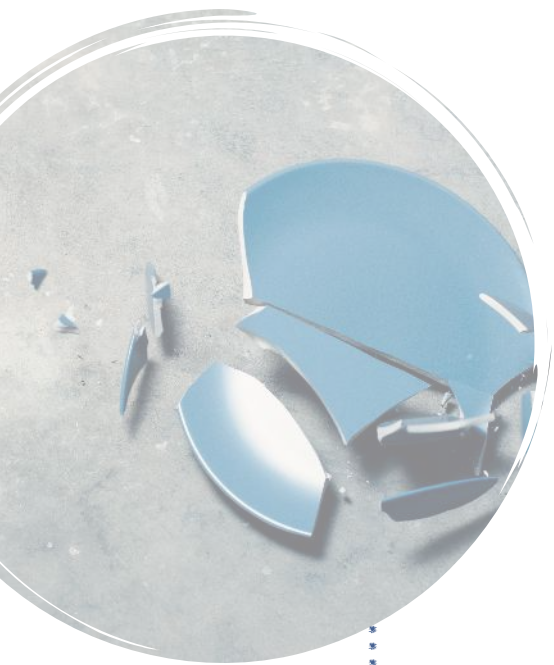
## Goals



## Goals



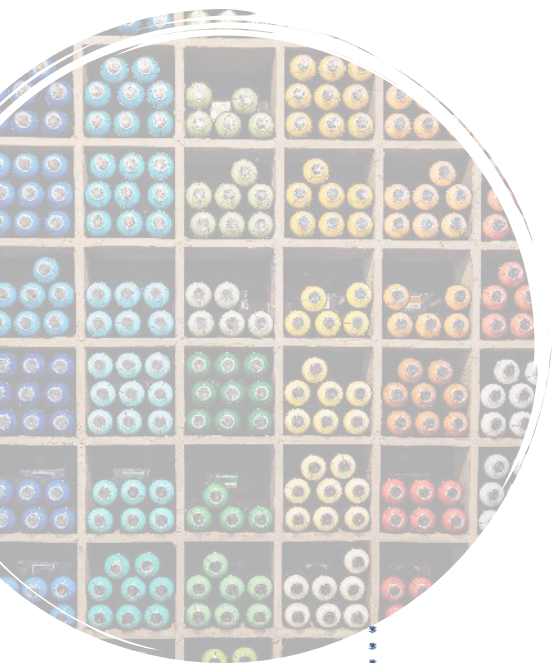




## When things go wrong

WIC - interoperability  
Public housing - documentation  
Commercial data - quality control  
2010 census - operational decisions

React  
Remedy  
Resolve  
Remember



# Data quality

What you care about

Most important fields for  
Linkage/de-duplication

Who and where

How many and for how long

Outcomes and costs

Reliability

Availability

Representativeness

※  
※  
※  
※

※  
※  
※  
※

# Dimensions of quality

Relevance

Accuracy and reliability

Timeliness and punctuality

Coherence and comparability

Accessibility and clarity

<https://www.mathematica-mpr.com/our-publications-and-findings/publications/transparency-in-the-reporting-of-quality-for-integrated-data-a-review-of-international-standards>

[archive.stats.govt.nz/methods/data.../guide-to-reporting-on-admin-data-quality.aspx](https://archive.stats.govt.nz/methods/data.../guide-to-reporting-on-admin-data-quality.aspx)

※  
※  
※  
※

※  
※  
※  
※

# When things go wrong (2)

WHOOPS | MAY 24, 2019

## Here's an Actual Nightmare: Naomi Wolf Learning On-Air That Her Book Is Wrong

By Yelena Dzhuravova



Photo: Roger Askenw/REX/Shutterstock

In the pantheon of nightmares, somewhere between "falling into an endless pit" and "back at high school but naked" is "going on national radio and learning, on-air, that the book you wrote and is to be published in two weeks is premised on a misunderstanding." Naomi Wolf, unfortunately, is living that nightmare.

When she went on BBC radio on Thursday, Wolf, the author of *Virginia and the forthcoming *Outrages: Sex, Censorship, and the Criminalization of Love*, probably expected to discuss the historical revelations she'd uncovered her book. But during the interview, broadcaster Matthew Sweet read to Wolf the definition of "death recorded," a 19th-century English legal term. "Death recorded" means that a convict was pardoned for his crimes rather than given the death sentence.*

Wolf thought the term meant execution.

<http://nymag.com/intelligencer/2019/05/naomi-wolfs-book-corrected-by-host-in-bbc-interview.html>

MONEYBOX

## A Widely Cited Statistic That Supposedly Proved Student Debt Was a Rich Person Problem Was the Result of a Coding Error

By JORDAN WEISSMANN

APRIL 26, 2019 • 4:25 PM



Dear Reader: Reuters hasn't taken any good news photos for stories about student debt since 2012, so I'm stuck reusing this one, from an Occupy Wall Street protest, where borrowers hung signs with their balances from their necks.

Andrew Burton/Reuters

Back in January, the Urban Institute—a widely respected Washington think tank—published a statistic about student debt that pretty soon caught fire among economics journalists. Based on an analysis of federal data, its researchers concluded that 49 percent of all outstanding education loans belonged to highest-earning quarter of American households.

1 True

<https://slate.com/business/2019/04/elizabeth-warren-college-debt-plan-bad-stat-urban-institute-code-error.html>

## A new book says married women are miserable. Don't believe it.

Many books aren't fact-checked, and we're increasingly realizing they're full of errors.

By Kelsey Piper | Jun 4, 2019, 12:40pm EDT

f t SHARE



Happiness researcher Paul Dolan made a splash with the claim that married women admit they're miserable once their spouses leave the room. It was based on a misreading of survey data. | Public Domain Pictures

<https://www.vox.com/future-perfect/2019/6/4/18650969/married-women-miserable-fake-paul-dolan-happiness>

---

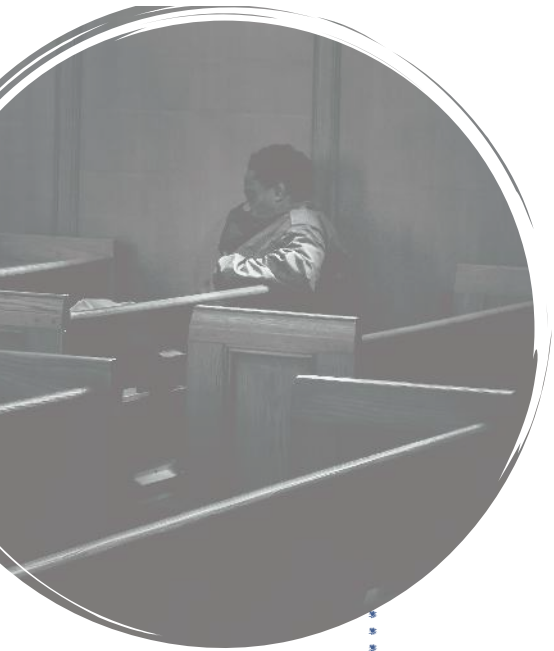
---

Getting help

Do-it-yourself  
Guides  
Training  
Tools

Have it done  
Use info scientists  
Contractors  
Crowdsource





## Example: civil justice

Court records, service providers,  
commercial data, open data

Users are researchers/journalists  
Users are legal services providers

Integrated or federated  
Common data model



# Administrative Data Research Initiative

Supports the secure and responsible use  
of administrative data for research and  
policy evaluation.

Domains and disciplines  
Directories and inventories  
Templates and guides  
Community building  
Intermediaries  
Products



Amy O'Hara, Ph.D.  
Massive Data Institute  
Georgetown University  
[amy.ohara@georgetown.edu](mailto:amy.ohara@georgetown.edu)

