

Communication and data curation

V1. Communication issues in data curation

As we have seen: communication is central to almost every area of data curation . . .

And is itself an area of data curation

Scientific communication of data and the results of data analysis
is an essential part of science, technology, and scholarship,
with effects in two directions

Of course more data, and more data analysis, is a good thing,
but the increase is causing a crisis in scientific communication

Communication and the data curation

As we have seen *communication* is central to almost every area of data curation

[And it itself an independent area]

Areas of curatorial activities

Collection:	Support the collection and acquisition of data
Organization:	Employ an appropriate data model and use appropriate standards
Storage:	Support reliable and effective storage
Preservation:	Ensure that data will be understandable and useable in the future
Discoverability:	Support the ability to search for and locate relevant data
Access:	Support the ability to retrieve and distribute data
Workflow:	Support the ability to systematize data workflows
Identification:	Support the ability to identify, authenticate, and validate data
Integration:	Support integration of data from different sources using different data models
Reformatting:	Support reformatting for use by different tools or to match new format standards
Reproducibility:	Support ability to reproduce results, ensuring scientific validity
Sharing:	Support sharing data between researchers, teams, and institutions.
Communication:	Support representation, publishing, and visualizations that provide insight
Provenance:	Support identifying what inputs and calculations are responsible for data values
Modification:	Support management of corrections and updates
Compliance:	Ensure compliance to legal, regulatory, and local policy requirements
Security:	Ensure that data is secure from tampering or inappropriate access and distribution

Our definition of data science (again)

Data science is concerned with all aspects of
the **creation, management, analysis**, and **communication** of data
focusing in particular on
the application of *computational methods* to *digital data*

The data science objective: *extracting useful knowledge from data*

Methods of curatorial action

There are many methods and techniques employed to achieve the objectives just listed, but five categories stand out as particularly important:

Analysis

To determine needs, and develop relevant data models and *metadata*, and reformat, correct, or update data.

Documentation

To record essential information (typically via *metadata*)

System design and implementation

To support all data curatorial activities

To support the generation and use of data documentation and processing documentation

Policy

To specify objectives, procedures, practices, and formats.

Process

To ensure success and efficiency by managing the development of appropriate organizational units and roles, providing training, advocating for change, and managing curatorial activities.

What we will take up next

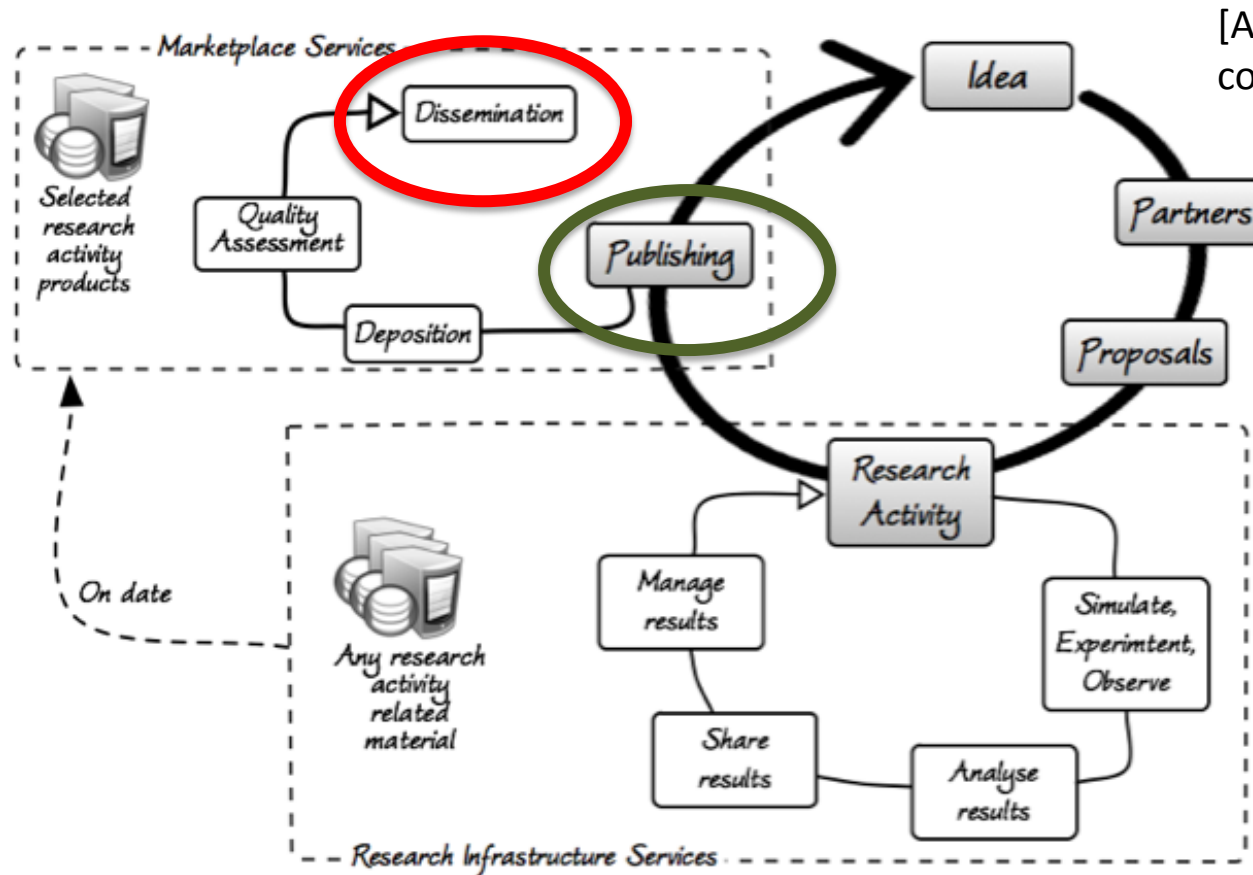
Ok, communication is relevant to many aspects of data curation

But we have a particular focus this week:

Scientific and technical communication

particularly journal literature, where scientific results are reported

Scientific communication is how data gets noticed



[After all, if the results of analyzing data are not communicated, then what's the point of it all?]

Scientific and technical communication is a critical part of the data lifecycle, with effects flowing both ways:

- from the research process,
- and back into the research process.

. . . the crisis

But scientific and technical publishing is in crisis

as we'll see in the next video