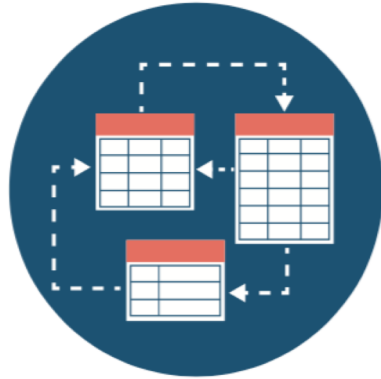# Workflow and Connection Diagrams

Trisha Cruse, Executive Director DataCite

Lynn Yarmey, Director of RDA/U.S. Community Development RDA Secretariat

TAG E: Workflow recommendations between data repositories and publishers

Patricia Cruse, Executive Director, DataCite
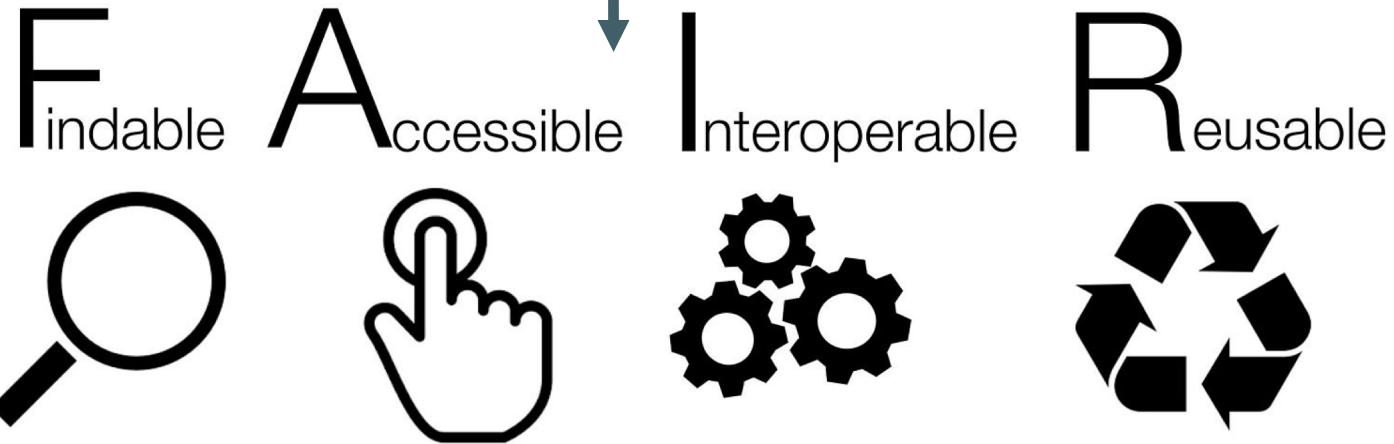Mark Servilla, Principal Investigator of EDI

AGU: Enabling FAIR Data Project

What

Our task: Workflow recommendations between data repositories and publishers.

In order to…

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}

What does it really mean?

- Investigating **workflows** associated with **data** and **article availability** across **publishers** and **repositories**.

- A key component of the investigation is the development of a **use case** that can apply to a **range of stakeholders**.

- However, TAG E recognizes that any one article and associated data can have **multiple use cases**.

Who are the actors?

- Data Producer
- Paper Producer (Author)
- Data Repository
- Publisher
- Persistent Identifier Service Provider
- Manuscript Peer Reviewer

Keep in mind

- **Data** = non-textual materials, including datasets, statistics, transcripts, survey results, and the metadata associated with these objects. The data are the factual information that is necessary to replicate and verify research results.

- **FAIR Compliant Repository (TAG A and B** = components in a FAIR repository: data, metadata, persistent identifier (PID), support for data citation, support for landing pages

- **Persistent identifier (PI or PID)** = a long-lasting reference to a document, file, web page, or other object. Digital Object Identifiers (DOIs) are one of the more widely used PIDs by the scholarly community.

- ?

Two variables drive the workflow

1. Time of data **deposition** in a FAIR compliant data repository

2. Status of data **availability**

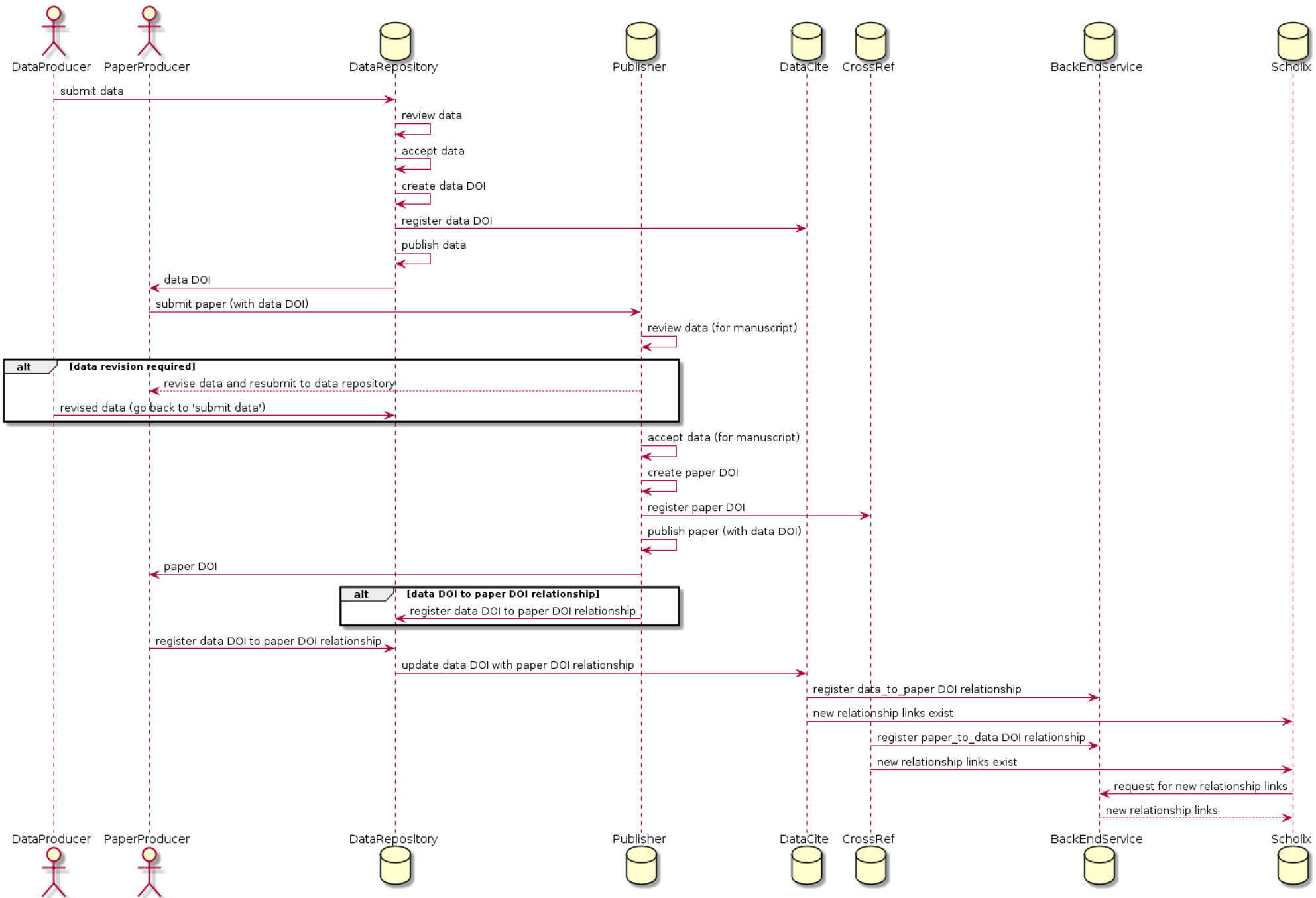| 1.Time of data deposition in a FAIR compliant data repository		
A	Data deposited prior to manuscript submission (Recommended)	The data that supports the research, to include the visualizations, has been deposited in a repository and published with a persistent identifier as part of a URL that resolves to a landing page at the time of manuscript submission.
B	Data deposited after manuscript	The data that supports the research, to include the visualizations, is not published in a repository. It is possible that the repository is actively working on the data, but it has not yet been published, and no citation is possible at this time.
C	Data published as supplementary information	This is no longer an option. Data must be in a repository and cited within the manuscript.
D	Data made available after manuscript publication	This is no longer an option. The manuscript will not be published if the data are not in a repository.

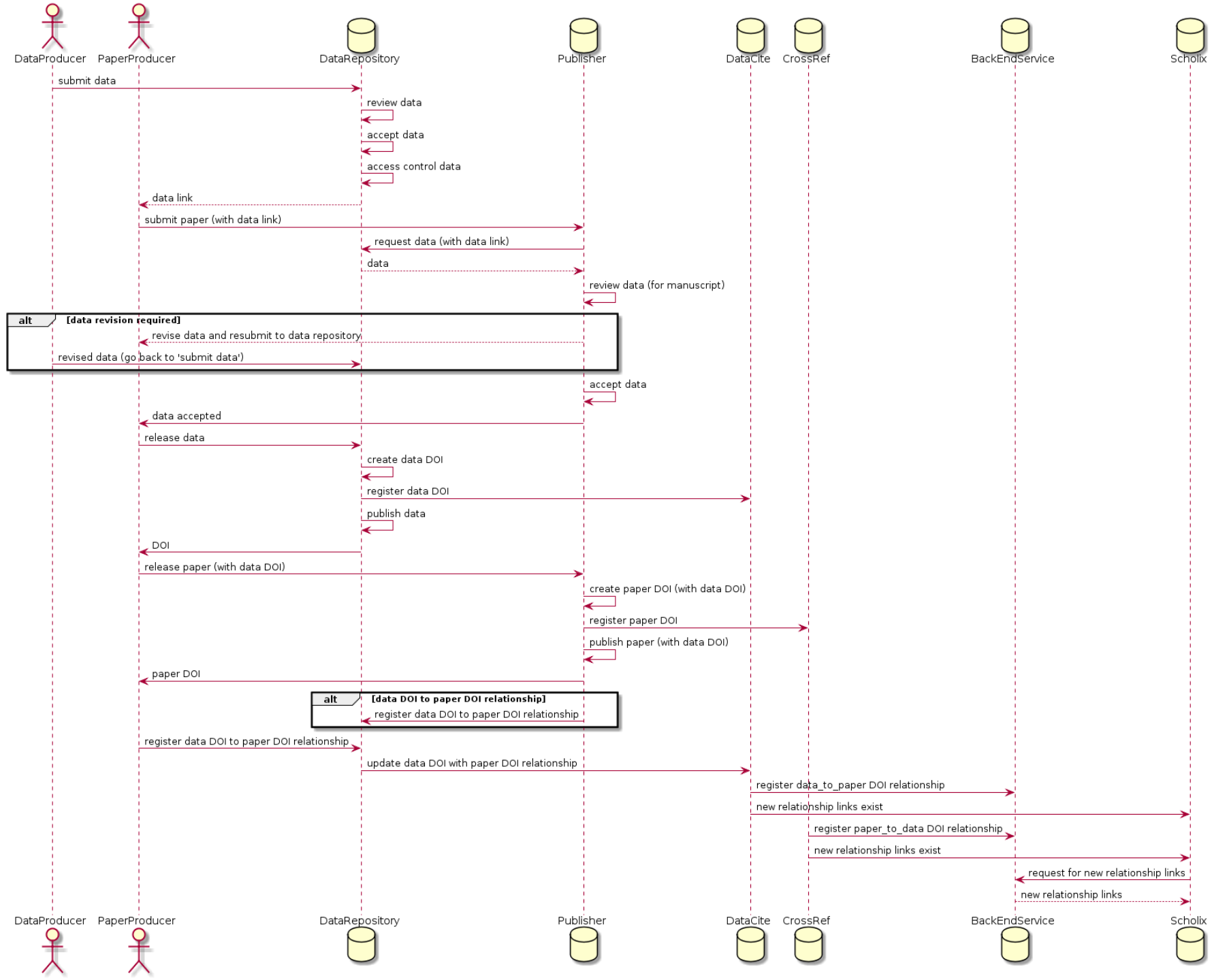| 2. Status of data availability		
1	Data publicly available	The repository has completed all curation work on the data and made it discoverable and accessible through a landing page. A persistent identifier has been assigned and registered. The repository supports data citation.
2	Data held until publication	The author, in conjunction with the repository, plans for the data to be made publically discoverable and accessible in conjunction with the publication of the manuscript.
3	Data are protected (access control); metadata available	Data that are sensitive and/or protected must be placed in a repository that can provide access control protections. Metadata should be made discoverable and accessible. The data availability statement in the manuscript as well as the repository metadata should make the steps for acquiring access clear.
4	Data reused	Data that have been published and cited previously follow the same process as #1 - Data publically available.

data deposition + data availability = use cases

	A Data deposited prior to manuscript submission	B Data deposited after manuscript submission	C Data published as supplementary information	D Data published after manuscript publication
1 Data publicly available	**A1** – Recommended 🙂 ⭐			**D1 - Not allowed**
2 Data held until after manuscript publication	**A2 -** Recommended 🙂	**B2**	**C2 – Not allowed**	
3 Data are protected (e.g. sensitive data, endangered species, etc.)	**A3**	**B3**		

A1 – Data deposited prior to manuscript and made public

A2 –
Data deposited prior to manuscript but not published until the manuscript is published