DIDs Supporting Research Data Ecosystems

Carly Huitema, Ph.D.
Manager – Agri-food Data Strategy

AGRI-FOOD DATA CANADA
AT THE UNIVERSITY *of* GUELPH

Semantic Engine Team Contributors

Past and present

Michelle Edwards
Director, ADC

Carly Huitema
Manager, ADC

Mathew Mozaffari
Developer, ADC

Steven Mugisha Mizero
Developer, ADC

Setayesh Sanavi
Developer, ADC

Niraj Patil
Developer, ADC

Ali Asjad
Developer, ADC

Adam Sadowski
Developer, ADC

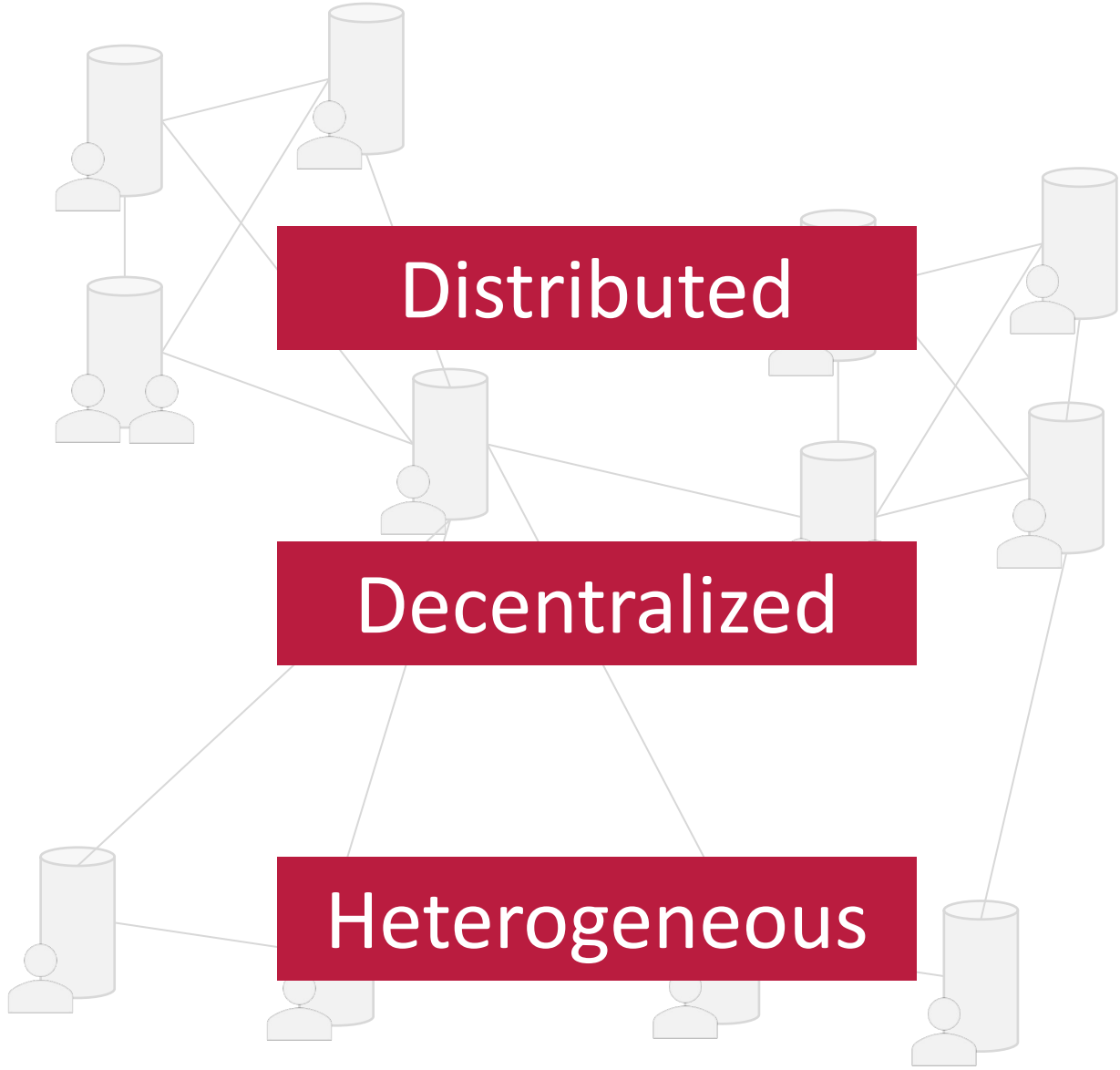
Xingjian Xu
Co-op developer, ADC

Lyndsy Acheson
Co-op developer, ADC

Elisha Matharu
Co-op developer, ADC

AGRI-FOOD DATA CANADA

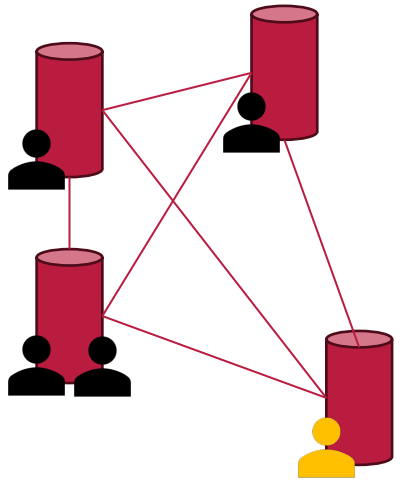
Research data infrastructure



ADC

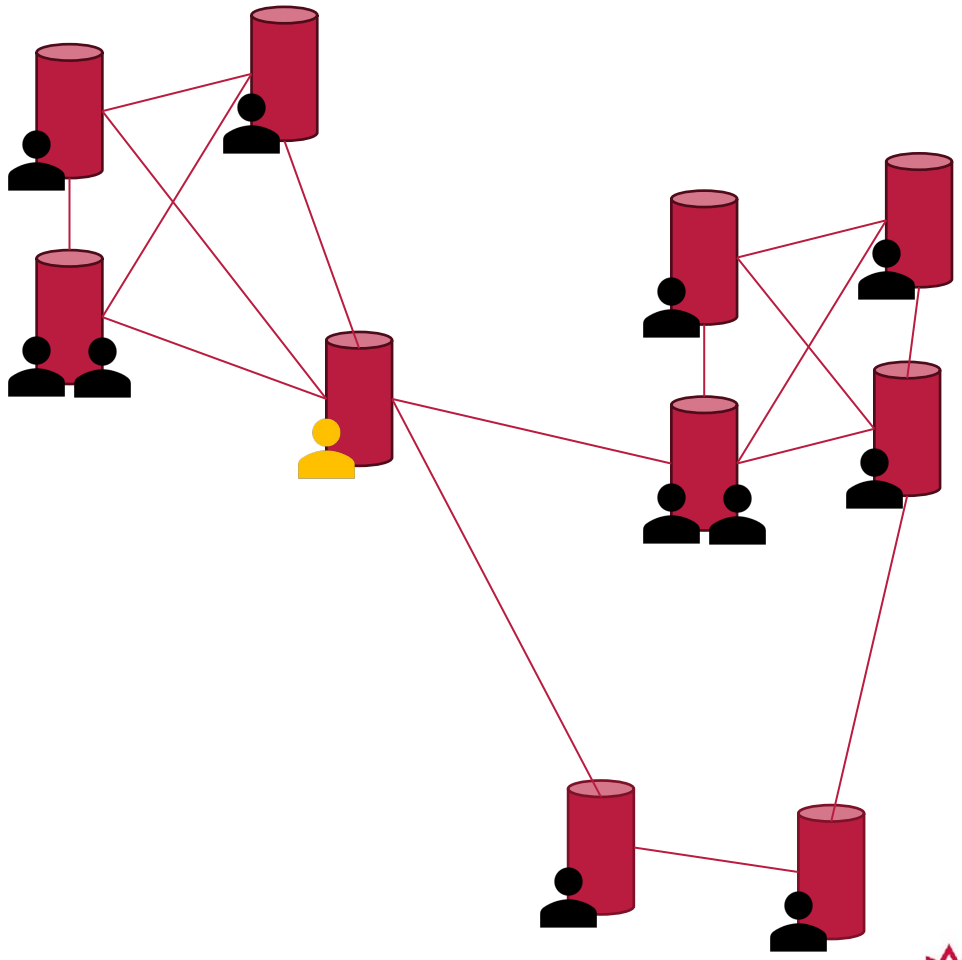
Research collaborations organized in projects

- A researcher is often part of a funded team to address a specific grant topic
- They may share resources, send data files, develop shared infrastructure



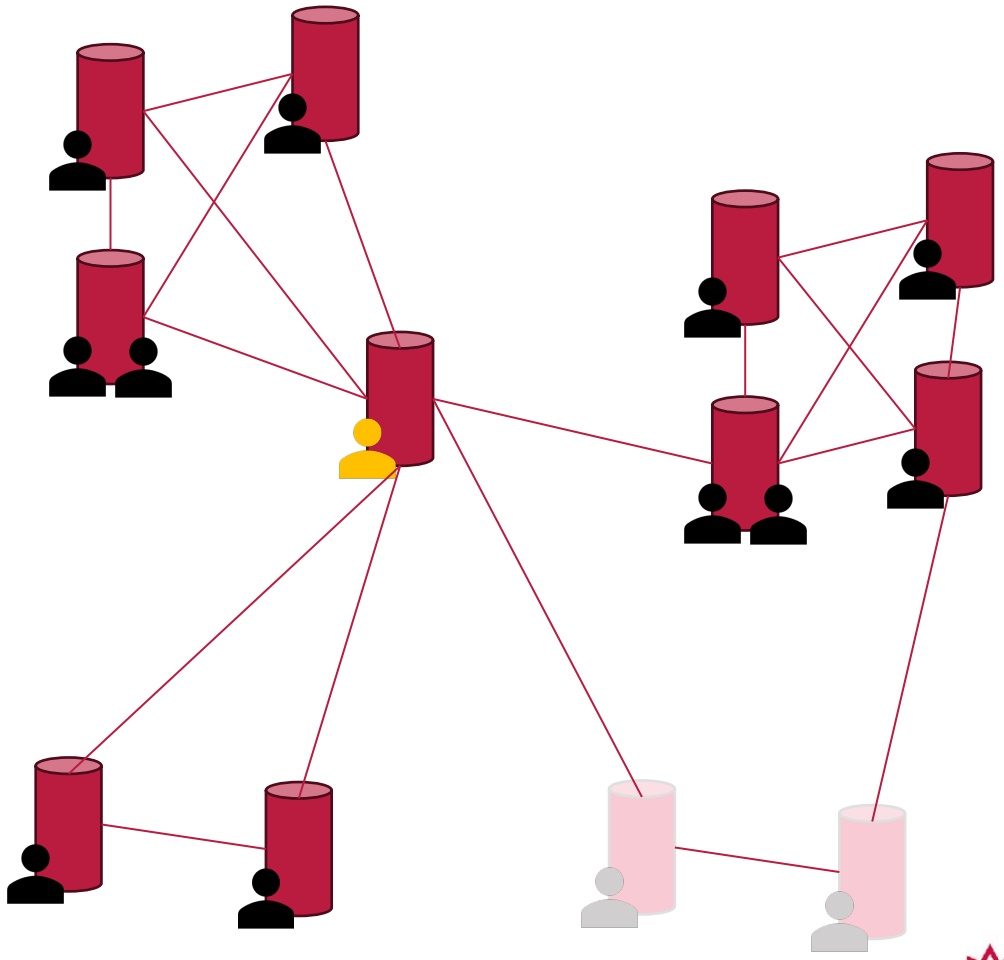
Research ecosystem complexity increases

- Researchers can be part of several projects at the same time



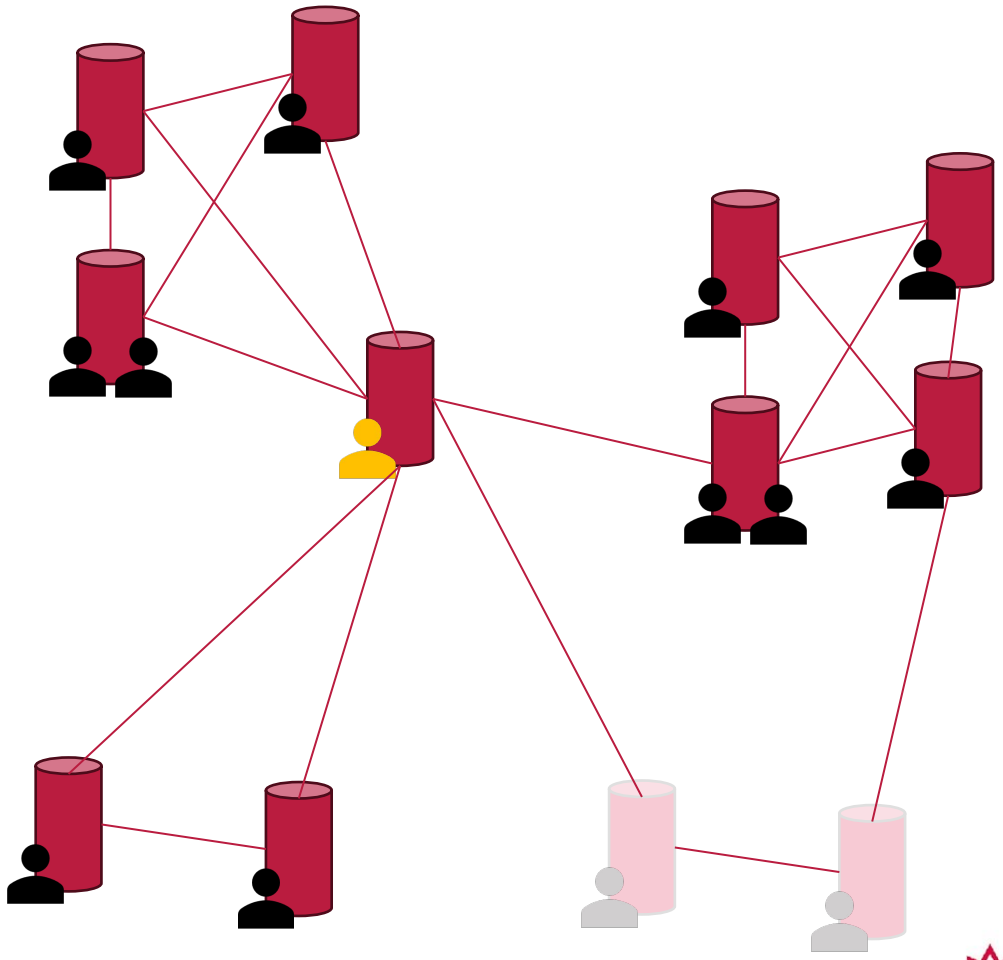
Research structure is dynamic

- Research projects are always in the process of forming and dissolving.



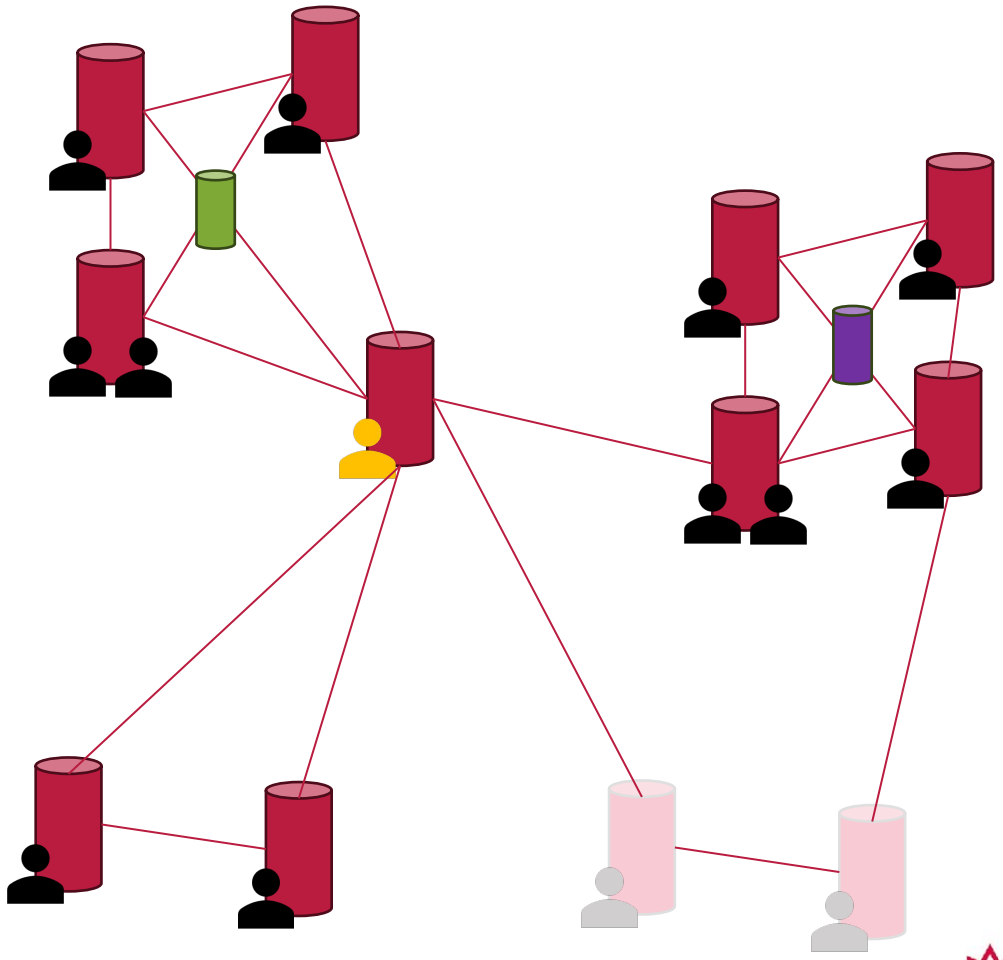
Data Complexity Considerations

- Some data is sensitive
  - E.g. health data, commercial data, indigenous data, personal data
- Some data is very large
- Diverse realities and opinions on data openness



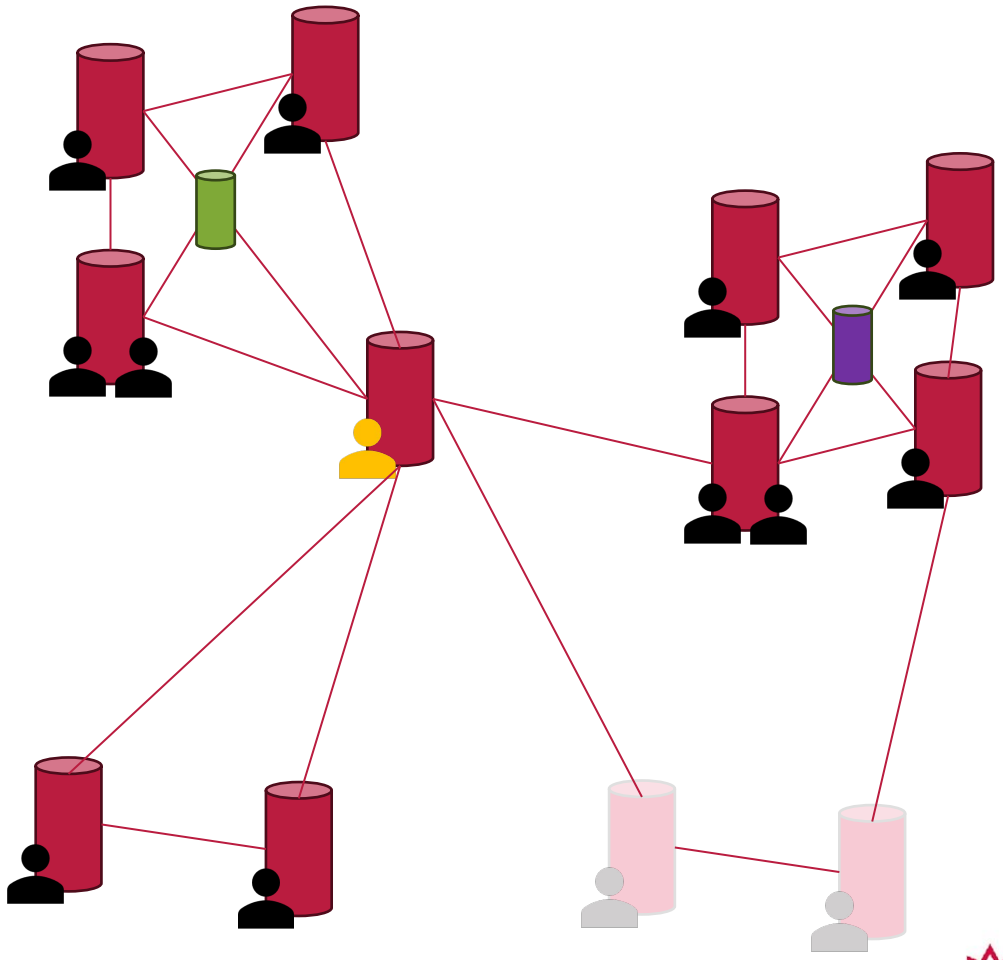
Challenge 1: Data Access

- Data access a mixture of solutions
  - Given a copy of the data (or subset, anonymized, etc.)
  - Given login credentials to multiple systems
  - Create new centralized infrastructure to serve a specific project.



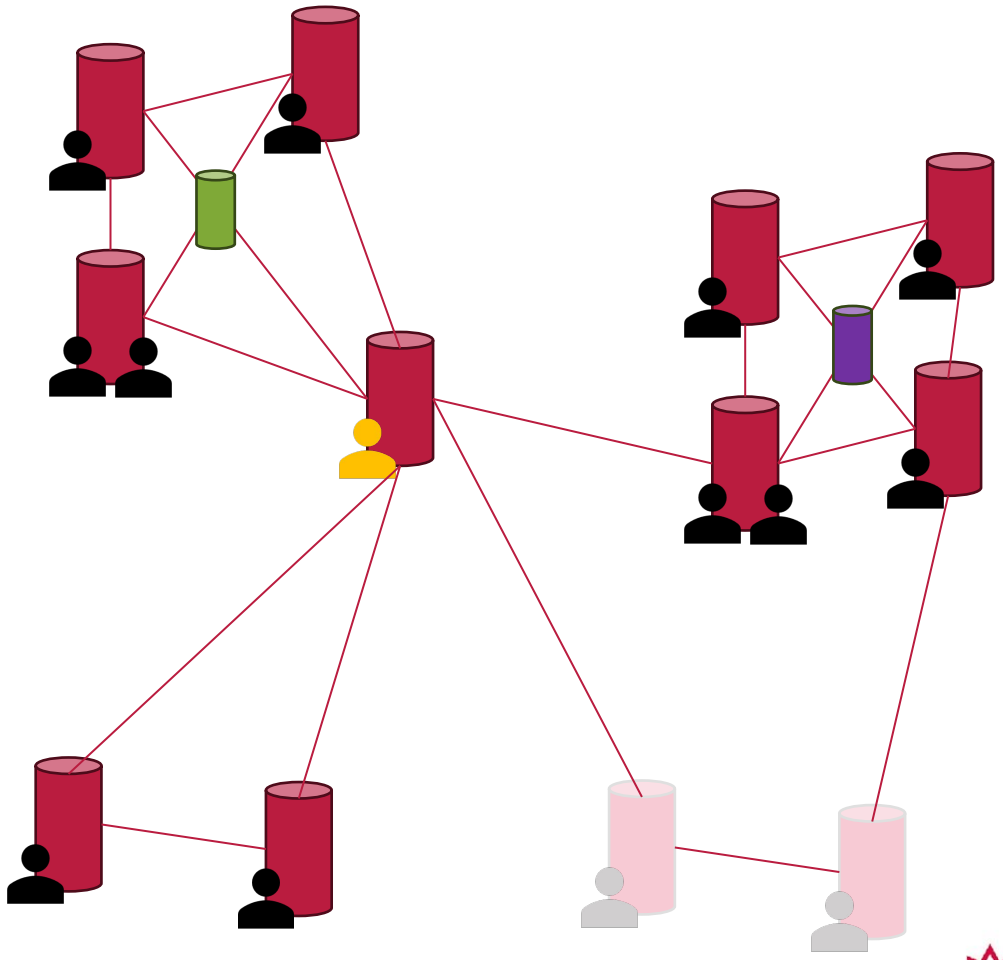
Challenge 2: Provenance Tracking

- Infrastructure to build global provenance tracking is lacking
- Currently built on citations and statements in publications
  - May use DOI identifiers of the datasets



Challenge 3: Secure Data Transfer

- Not widely used across all scientific disciplines
  - Discipline specific strength
  - Globus for secure file transfer in Canada
  - Other sharing includes Dropbox, OneDrive, Google Drive, email etc.

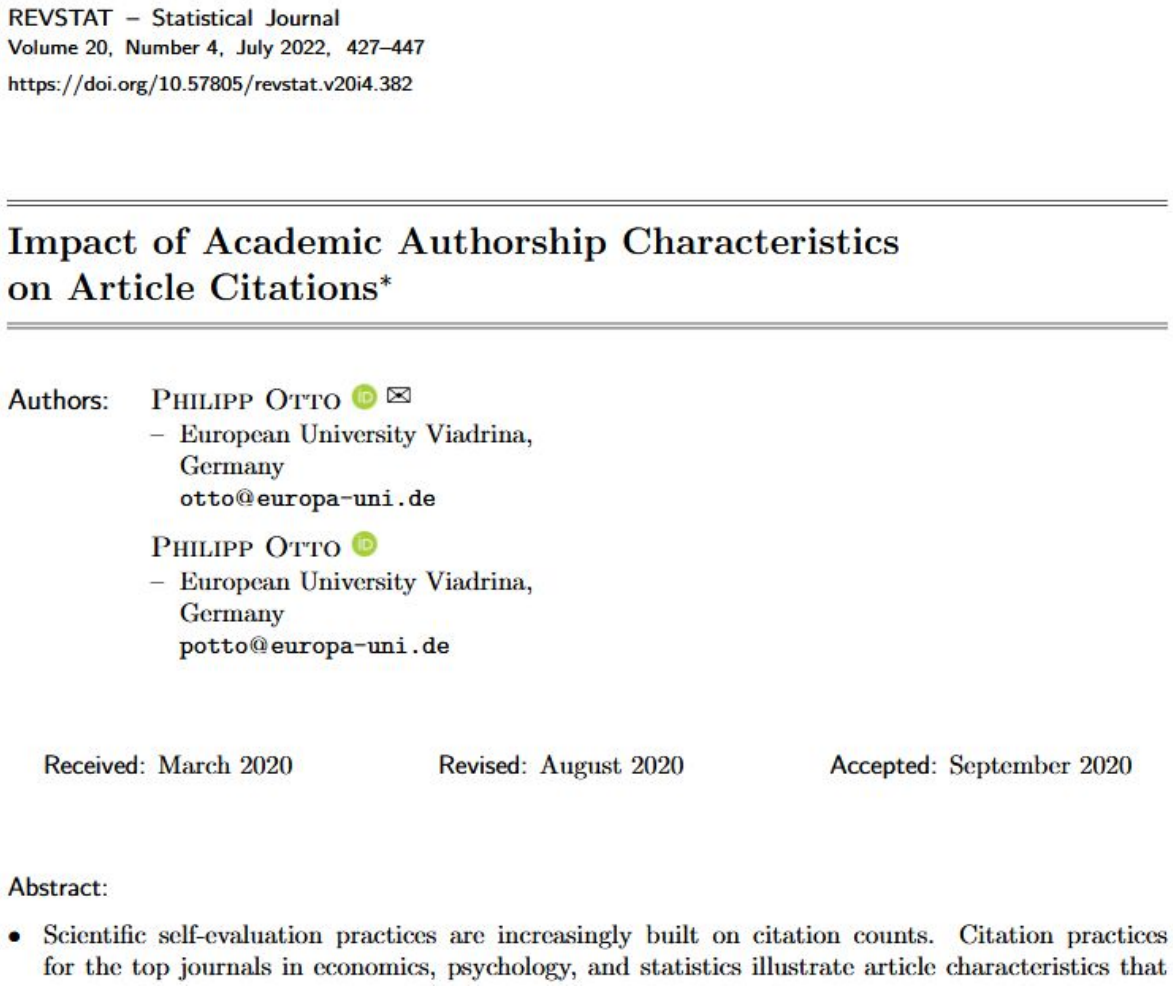


Researcher ecosystems use identifiers

- Persistent Identifiers: PIDs
  - Research infrastructure for unique global identifiers
  - ORCIDs, DOIs, ARKs, RORs, PURLs, RAiDs
    - Most popular are DOIs for research outputs including publications and datasets

ORCIDs for People

- Open Researcher and Contributor ID (ORCID)
  - Non-profit registered in USA
- ORCID is a persistent identifier for people in the research ecosystem
- Researchers own their ORCID records
- Global PID



REVSTAT – Statistical Journal
Volume 20, Number 4, July 2022, 427–447
https://doi.org/10.57805/revstat.v20i4.382

Impact of Academic Authorship Characteristics on Article Citations*

Authors: PHILIPP OTTO
– European University Viadrina, Germany
otto@europa-uni.de

PHILIPP OTTO
– European University Viadrina, Germany
potto@europa-uni.de

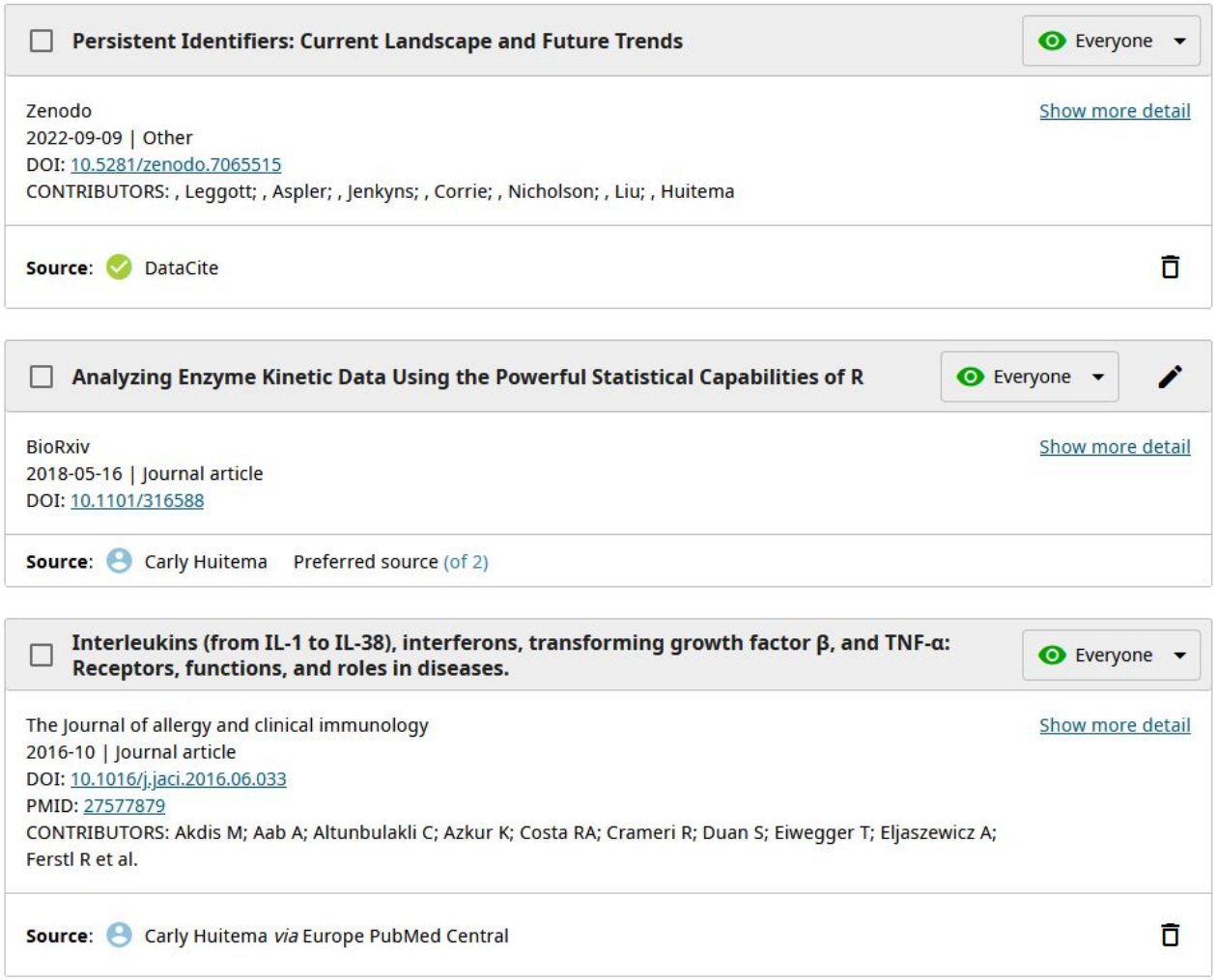
Received: March 2020 Revised: August 2020 Accepted: September 2020

Abstract:

- Scientific self-evaluation practices are increasingly built on citation counts. Citation practices for the top journals in economics, psychology, and statistics illustrate article characteristics that

ORCID record example

- ORCID data is either self-attested or updated via trusted organizations (researcher approved)
- OAuth and audit trails for verification of data

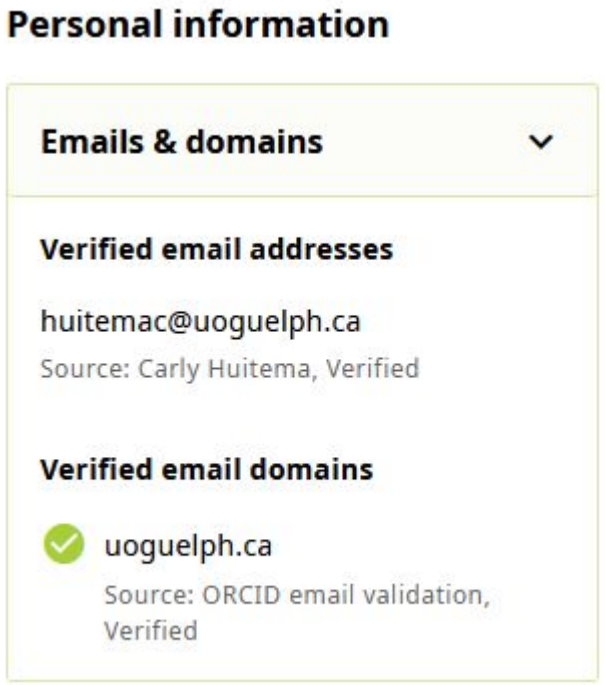


Decentralized Identifiers (DIDs)

- W3C standard Decentralized Identifiers (DID)
- The identifier (DID) resolves a DID Document
  - Resolution method depends on DID method (200+)
  - Storage and security of DID source data depends on DID method
    - Ledger, files, websites, history etc.
- DID document content contains:
  - Authentication keys – public keys for specified services
  - Assertion keys – public keys for signing statements
  - Key agreement keys – public keys for secure data channels
  - Endpoints – methods to contact DID subject

Proposal 1

- List researcher DIDs in ORCID record
- Could be used for:
  - Provenance tracking
    - Sign datasets and other outputs
  - Authentication
    - Key-based authentication
  - Secure data transfer
    - Key agreement keys and endpoints



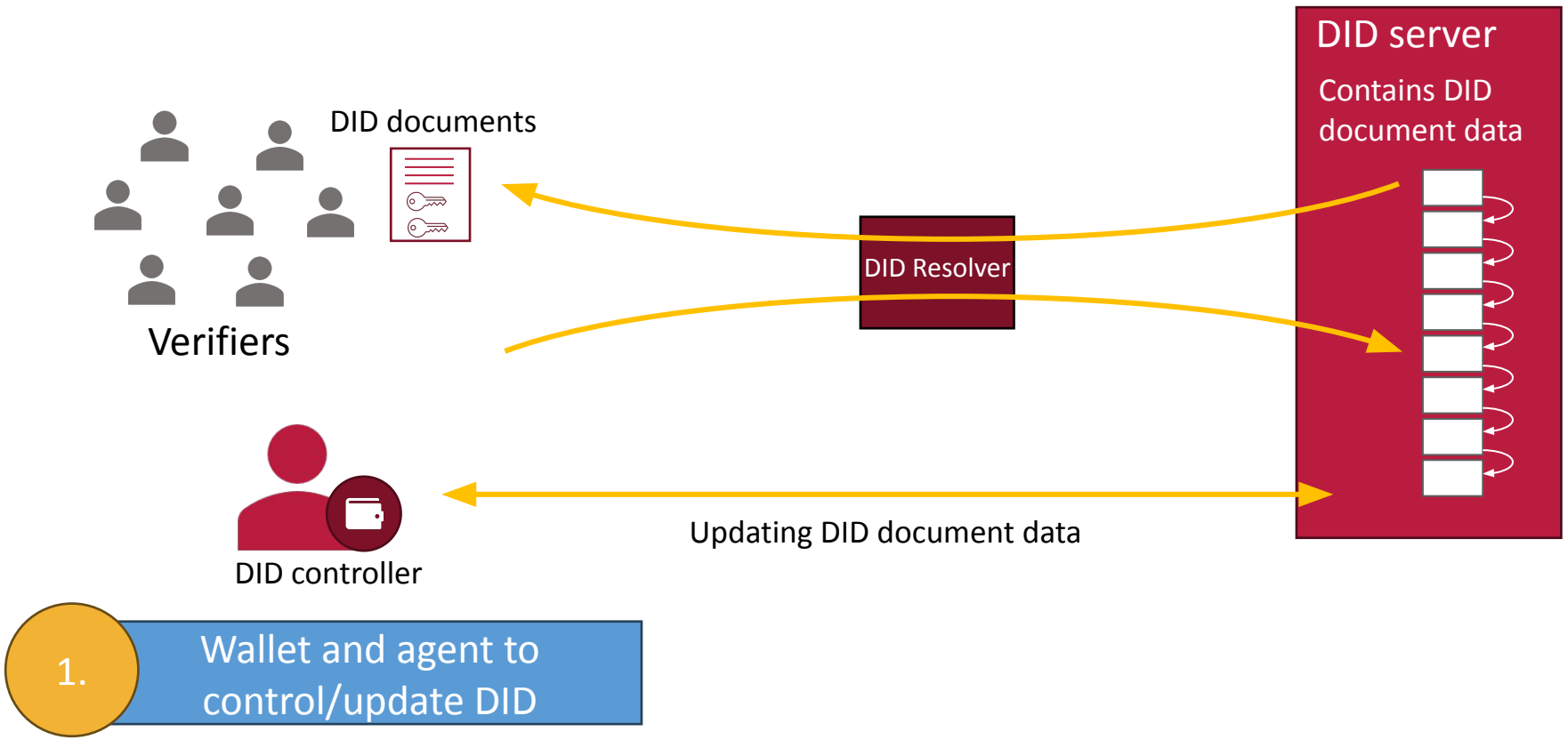
Verified DIDs

did:example:123456789abcdefghi

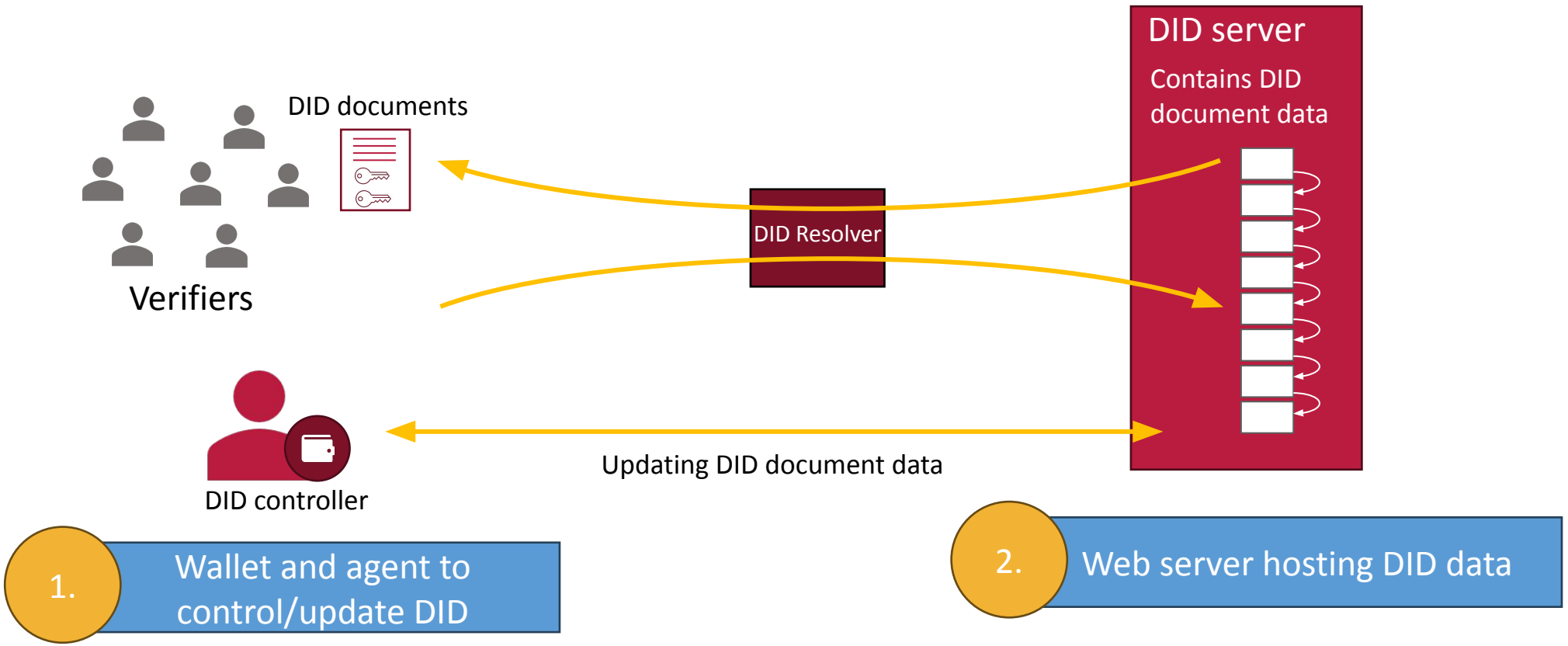
What can a DID do in a data ecosystem?

- Support authentication, assertions, secure communication/transfer
- DIDs available for building data spaces
  - Data Spaces can build on top of these available keys and endpoints
- ORCID with DIDs
  - Trust signals from ORCID record
  - Researcher PID integrated into existing workflows and infrastructure

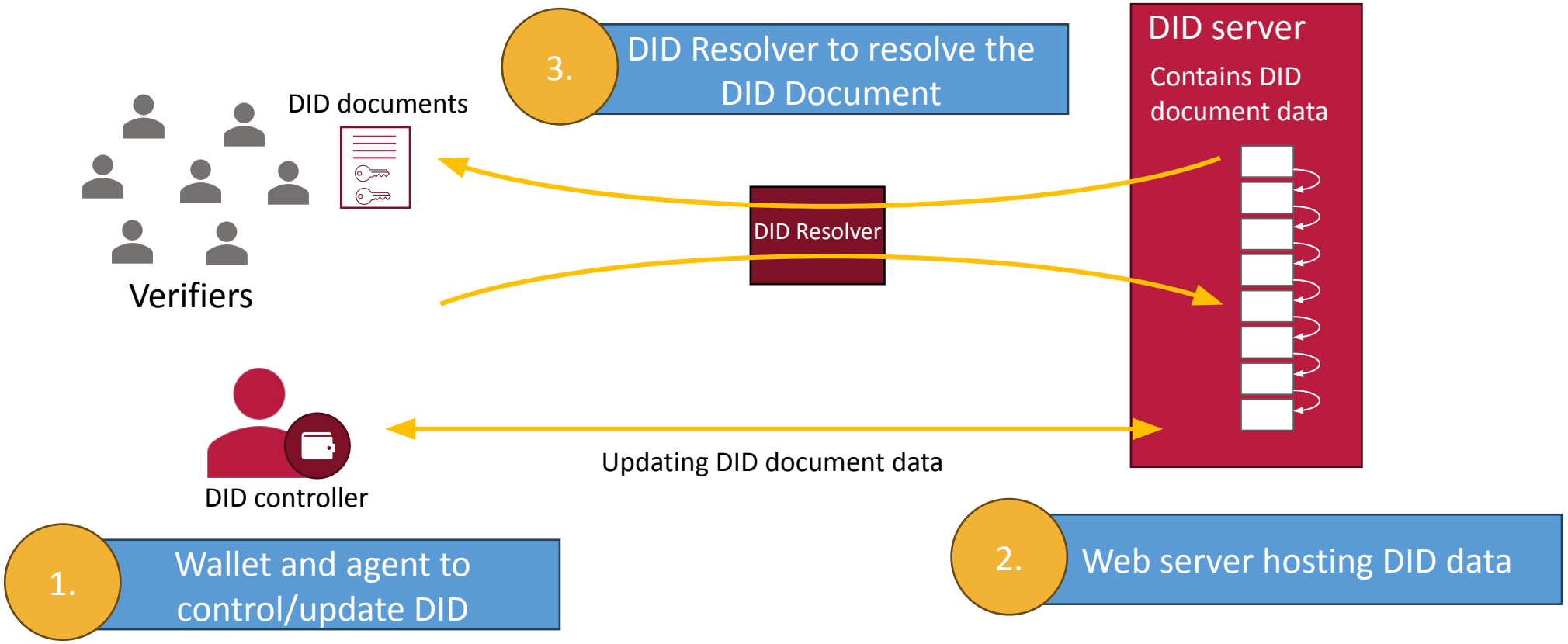
DID Ecosystem Architecture



DID Ecosystem Architecture



DID Ecosystem Architecture



Proposal 2

- ORCID host DID infrastructure
  - Let researchers create DIDs as well as list them
  - Bootstrap more secure research Data Spaces
  - Support others to build DID infrastructure

DID methods: did:webvh and did:webs

- Self-certifying identifiers
- Key pre-rotation
- Verifiable history, cryptographic evidence
- Witnesses and watchers supported
- Web-based publication
- Portable DIDs

https://didwebvh.info/

aca-py.org and Credo

Python implementation

https://trustoverip.github.io/tswg-did-method-webs-specification/

https://github.com/GLEIF-IT/did-webs-resolver

KERI-Py: https://github.com/WebOfTrust/keripy

How to follow this work

- **Agrifooddatacanada.ca**
- Trust over IP at Linux Foundation Decentralized Trust
- Educational content: Bite Size Trust on YouTube
  - https://www.youtube.com/playlist?list=PL0MZ85B_96CEoV-jvUbTovUAOA531bce_

THANK YOU

Agri-food Data Canada at the University of Guelph is an innovation platform for Canada's agriculture and food sectors.

AGRI-FOOD DATA CANADA
AT THE UNIVERSITY *of* GUELPH

50 Stone Road East, Guelph, ON N1G 2W
adc@uoguelph.ca

agrifooddatacanada.ca