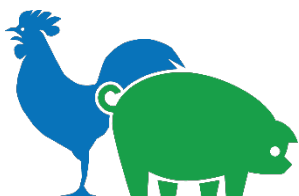


# Ag Data Commons

## Infrastructure for data publication and access

DCMI Virtual  
September 23, 2020

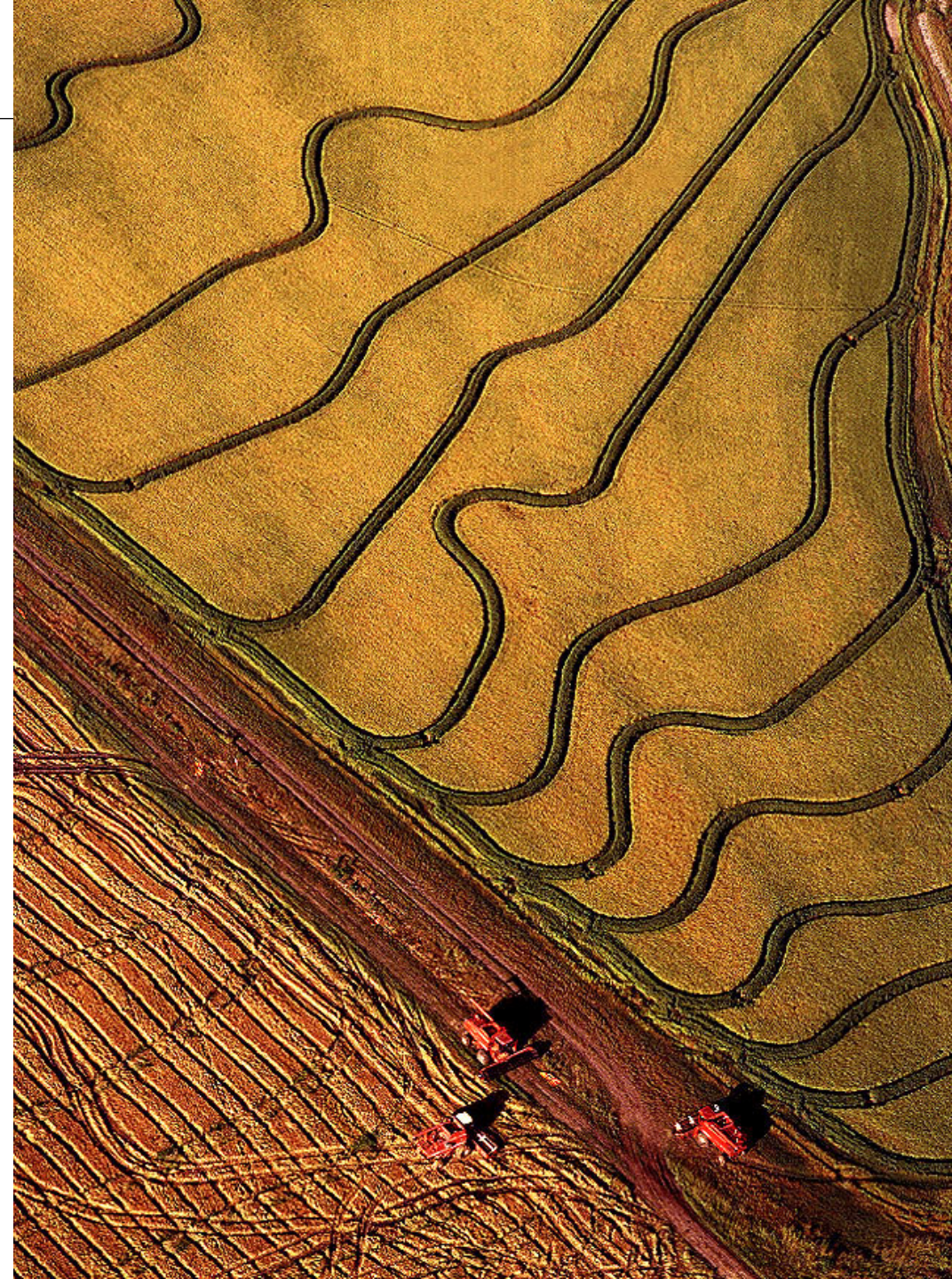
Erin Antognoli  
Data Curation Team Lead & Metadata Librarian  
National Agricultural Library, United States Department of Agriculture



# USDA National Agricultural Library

The National Agricultural Library (NAL) is one of five national libraries of the United States. It houses one of the world's largest collections devoted to agriculture and its related sciences.

<https://www.nal.usda.gov>



# NAL provides a variety of online resources

## **AGRICOLA**

Search citations from two agricultural science databases: indexed journal articles and the NAL catalog.

## **PubAg**

Find full-text articles and citations to peer-reviewed journal articles in the agricultural sciences.

## **Ag Data Commons**

Access open data relevant to agricultural research.

## **Digital Collections**

Explore NAL collection materials available in digital format.

## **NAL Thesaurus**

Browse an online vocabulary tool of agricultural terms in English and Spanish.

## **Invasive Species**

Learn about invasive species that threaten the United States.

## **Nutrition.gov**

Uncover reliable resources on food, healthy eating, physical activity, and food safety.

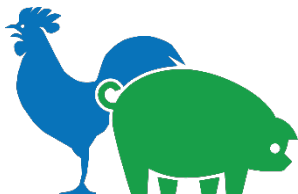
## **LCA Commons**

Find Federal life cycle assessment data, research, and information systems.

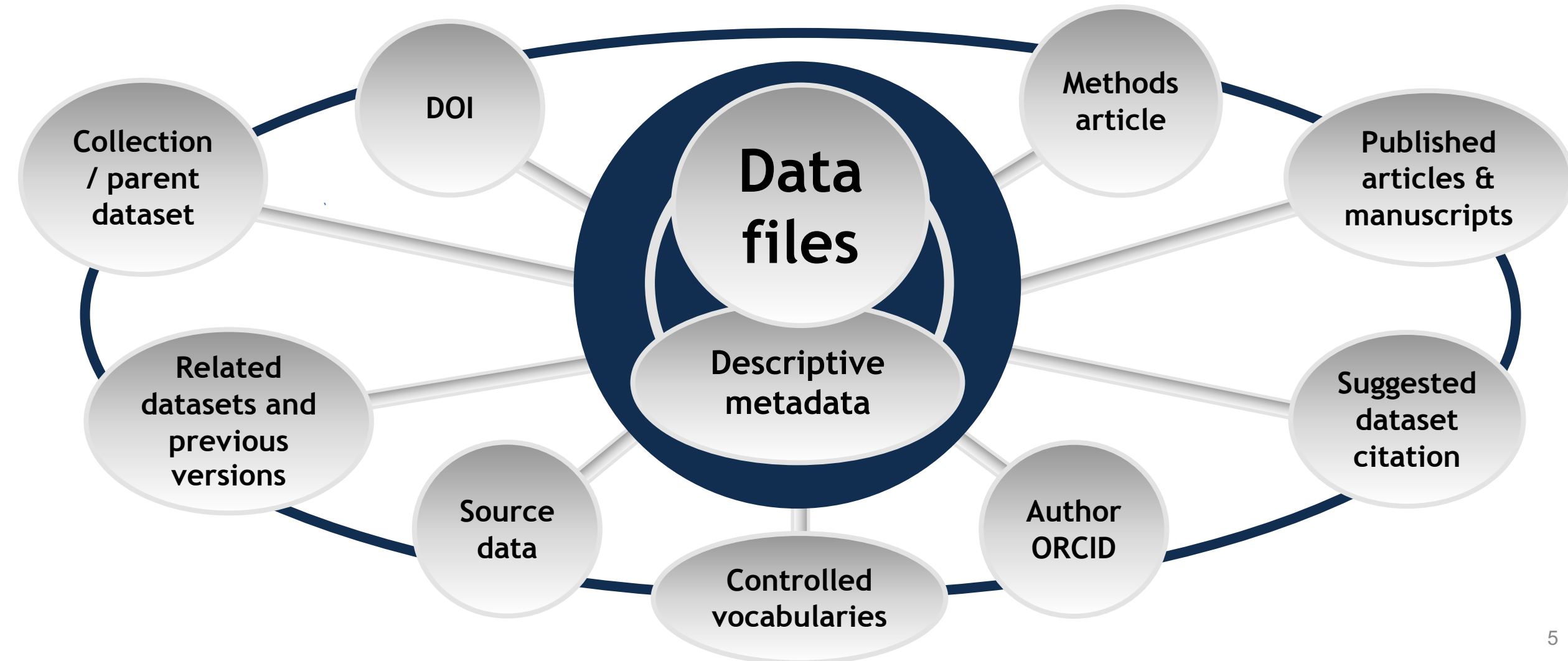
# The Ag Data Commons...

- Is a catalog and data repository for USDA-funded research data
- Provides expert services for creating, curating, and enabling access to complete and machine-readable scientific metadata (FAIR data)
- Creates infrastructure for linking information, data, publications, people,...
- Helps the USDA-funded research community meet public access requirements

<https://data.nal.usda.gov/>



# Ag Data Commons dataset



## Findable

- Rich metadata
- Persistent identifiers

## Accessible

- Fixity
- Data & metadata available to target audience

# FAIR Principles

## Interoperable

- Open formats
- Common metadata standards
- Controlled vocabularies

## Reusable

- Usage license
- Provenance
- Community standards

# Ag Data Commons provides discovery and access



## Discovery

- Search and discovery tools allow researchers to find data
- Linked information promotes discoverability

## Access

- Public access to data through a catalog of USDA funded research and public APIs
- Open licenses allow for reuse of data
- Data repository for long-term archiving (as required to support DOI service)

# Ag Data Commons synthesizes metadata

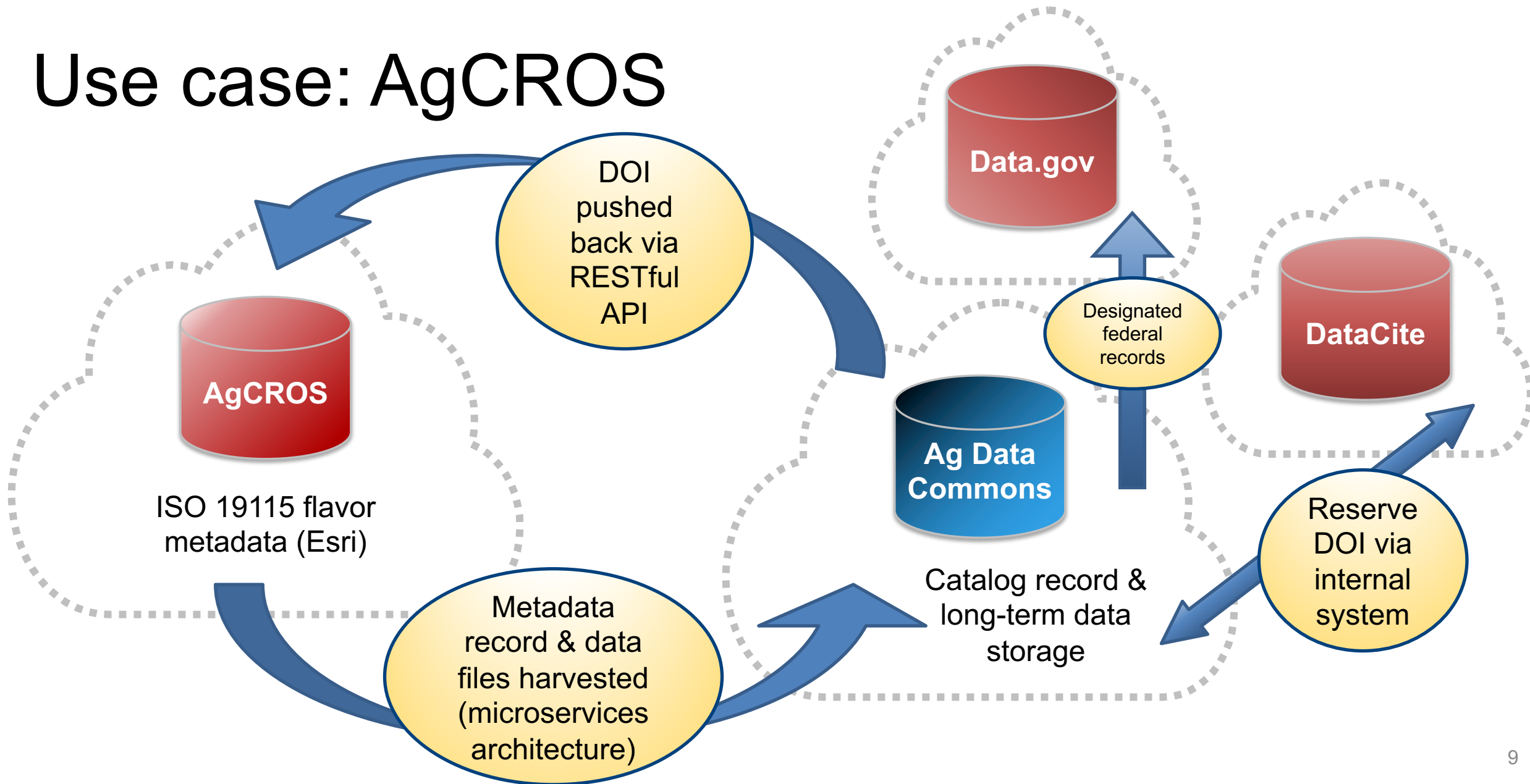


Ag Data Commons metadata fields may be output into a number of schemas to achieve a variety of functions

- ISO 19115
  - GeoData, AgCROS
- Project Open Data
  - Data.gov
- DataCite
  - DOIs



# Use case: AgCROS



# Essential metadata standardizes disparate ag resources

- Title
- Description
- Author and Author ID (ORCID)
- Publisher
- ISO Topic Category
- Temporal Coverage
- Geographic Bounding Box/Point
- Data Contact: Name and Email
- Tags / Keywords
- Bureau Code (ARS only)
- Program Code (ARS only)
- License
- Funding Source
- Resource Title
- Resource File Format
- Resource Link or Upload



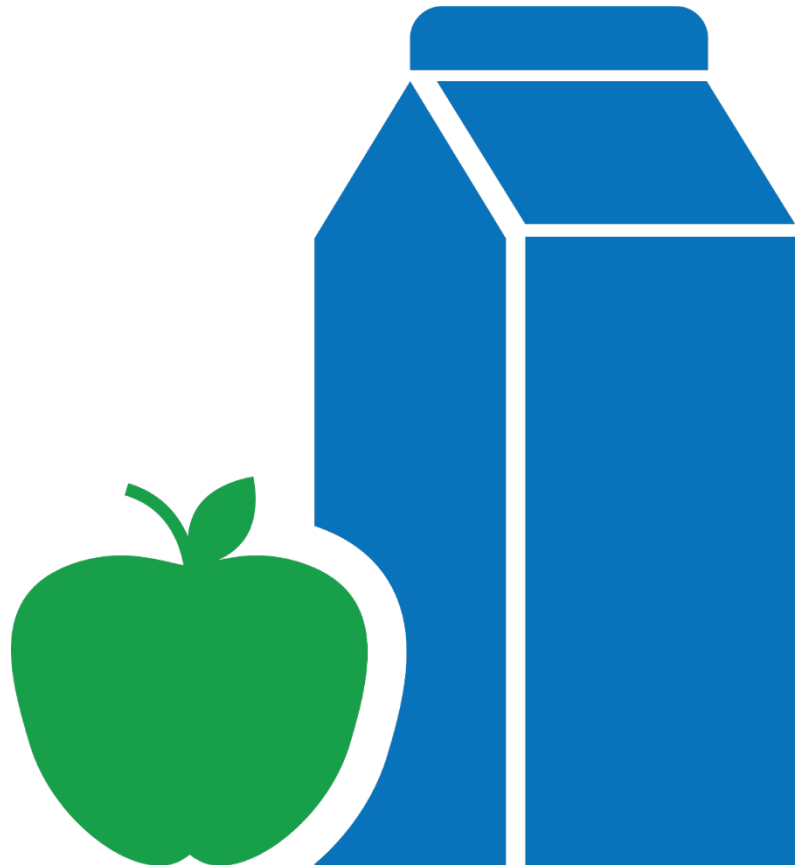
# NAL Thesaurus

- Produced by NAL, Inter-American Institute for Cooperation on Agriculture, and SIDALC orgs
- Updated annually
- Contains over 260,270 terms, including 150,842 descriptors in English and Spanish
- Available as linked open data

<https://agclass.nal.usda.gov/>



# Ag Data Commons links information



## Related content

- Group by project or affiliation
- Data hierarchies supported (parent/child relationships)
- Cite other datasets
- Link to project web content or methods documentation

## ORCID integration

- Authors can link to their profiles to prevent ambiguity

## Version guidance

# Features promote credit for USDA researchers

## Citations

- Automatic citation generation OR specify a citation for your own data
- Link to scholarly publications or data papers / PubAg
- Link to other related data content

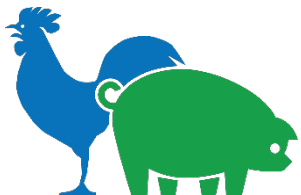
## DOI

- Assigned for locally held data files



# Challenges and opportunities

Increasing findability and accessibility

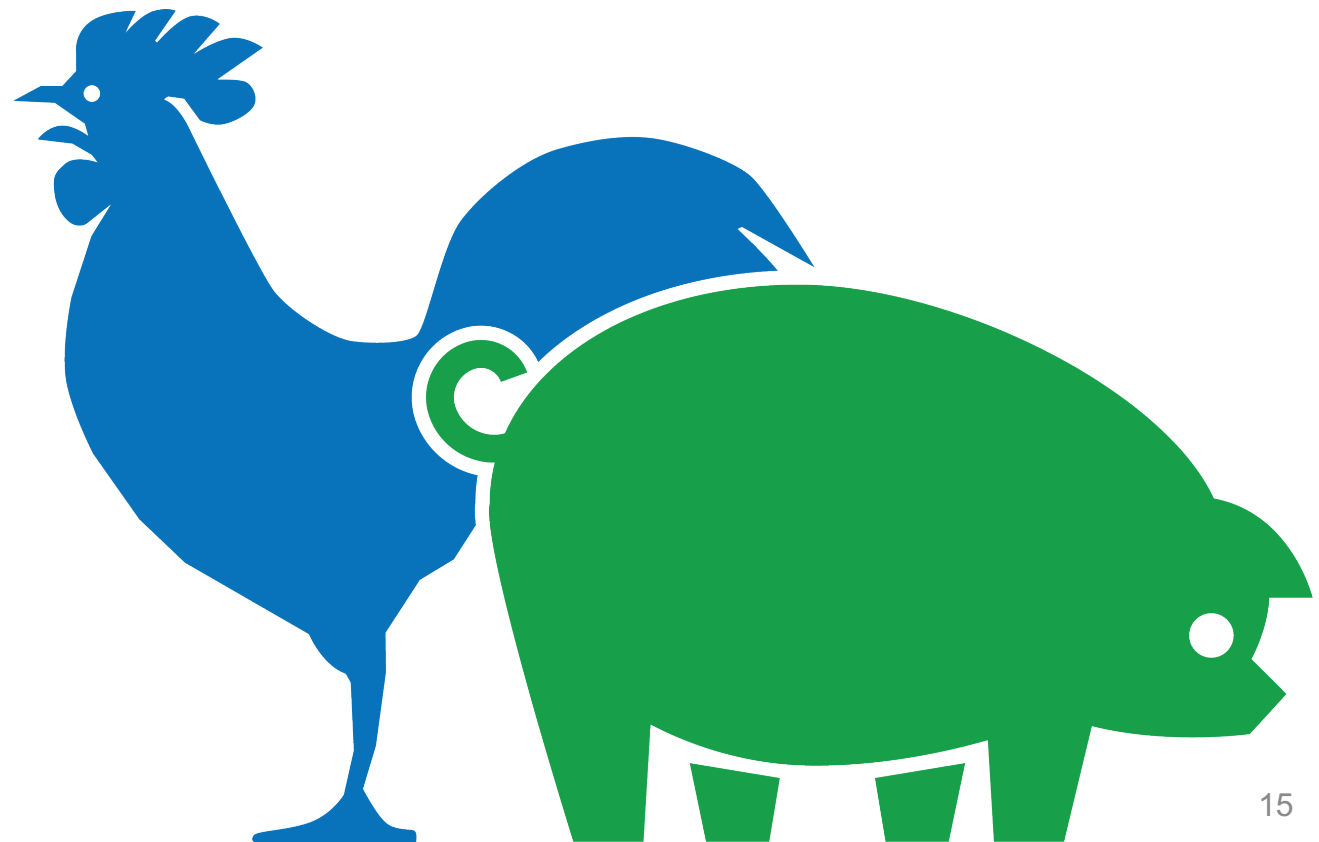


# Platform search & discovery features present challenges and opportunities

## Challenge: Search feature improvement plan

- Fielded search
- Boolean search
- Wildcard search

Opportunity: Platform upgrade expected in 2021



# Metadata presents challenges and opportunities



Diverse subject areas = diverse metadata needs

- We cannot be all things to all people

NALT application not automated

- Working with NAL's Indexing and Informatics Branch to improve this

Controlled vocabularies should link to the authoritative source (as opposed to flat terms)

- How do we represent each subject area's preferred vocabularies?



# Linking presents challenges and opportunities

## One-way vs two-way links

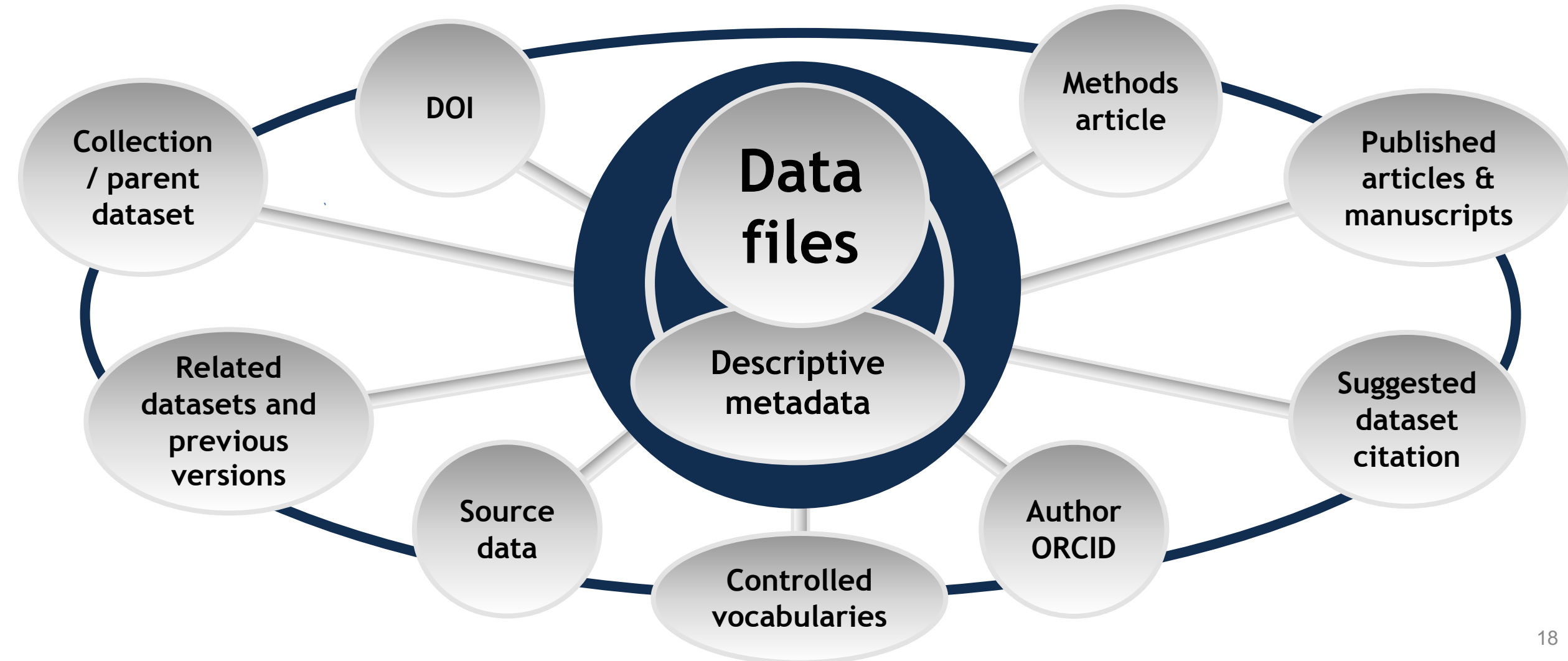
- How can our data become more discoverable outside of our platform?

## Capturing richness of linked information

- How can we add value without increased burden on data submitters?



# The Ag Data Commons links information



# Thank you!

Erin.Antognoli@usda.gov

Ag Data Commons: <https://data.nal.usda.gov>

National Agricultural Library: <https://www.nal.usda.gov>

NALT: <https://agclass.nal.usda.gov/>

