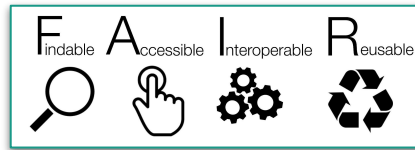# Sustainable Data Management Cluster

Investigates pathways for sustainable, increased collaboration and coordination among environmental data management systems that benefits both research networks and individual investigators
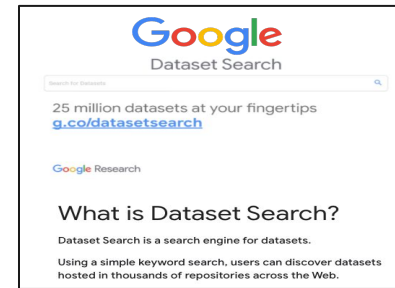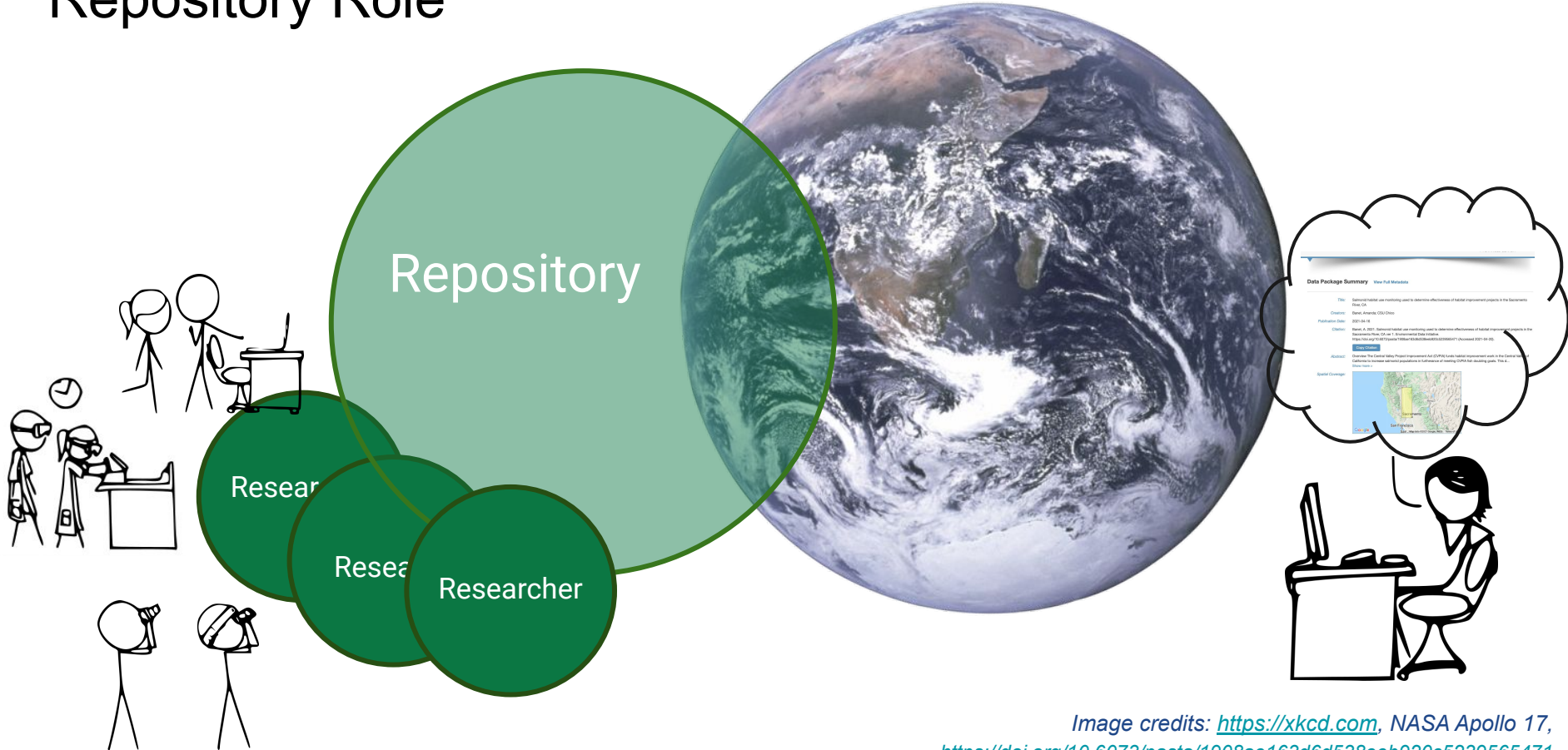
2020-2021 Project: Aligning TRUST, FAIR, CARE

Goals of Research Data Repositories

1. Publish and archive data from research projects

2. Make that data available for further use

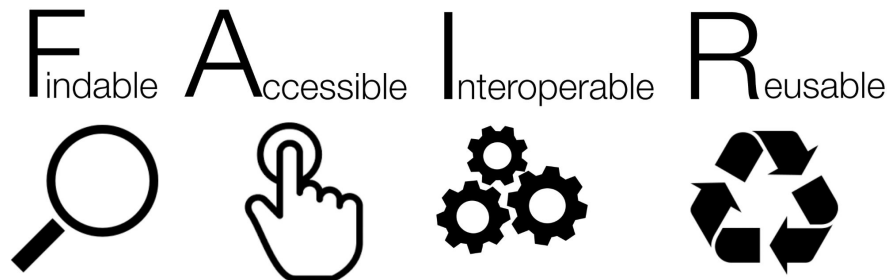3. Data are curated to the most complete level possible, and are easy to find

Repository Role

Repository

Researcher

Researcher

Researcher

- **Findable**
    - **unique and persistent IDs**
    - **rich metadata**
    - **metadata specify the data ID**
    - **Registered, indexed, easy to find**

- **Accessible**
    - **retrieved by ID, read and accessed via standardised protocols**
    - **open, free communications protocol**
    - **Protocol allows for authentication**
    - **metadata are accessible even if data are no longer available**

- **Interoperable**
    - **Use standardised, documented, and accessible semantic descriptions**
    - **vocabularies follow FAIR principles**
    - **Qualified references**

- **Reusable**
    - **Plurality of relevant attributes**
    - **there are clear conditions for data usage**
    - **detailed provenance information**
    - **Meet domain-relevant standards**

F indable A ccessible I nteroperable R eusable

https://datascience.codata.org/articles/10.5334/dsj-2020-041/
https://www.force11.org/group/fairgroup/fairprinciples

Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services.

CARE Principles for Indigenous Data Governance
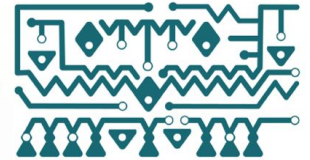
The increasing convergence of technology infrastructure and digital connectivity has raised the value of data across the globe

Indigenous data sovereignty reinforces the rights to engage in decision-making in accordance with Indigenous values and collective interests
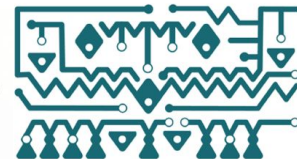
https://www.gida-global.org/care

CARE Principles

Collective Benefit

C1 For inclusive development and innovation

C2 For improved governance and citizen engagement

C3 For equitable outcomes

Authority to Control

A1 Recognizing rights and interests

A2 Data for governance

A3 Governance of data

Responsibility

R1 for positive relationships

R2 for expanding capability and capacity

R3 for indigenious languages and worldviews

Ethics

E1 For minimizing harm and maximizing benefit

E2 For justice

E3 For future use

People

CARE

Repository

TRUST

FAIR

Data

Image credits: https://moorea.berkeley.edu, https://sedac.ciesin.columbia.edu, ESIP

Process

~10 hrs

Conversations with the CARE group

Stephanie Carroll, Maui Hudson, Lydia Jennings, Ibrahim Garba, Andrew Martinez

From our cluster

Shelley Stall, Margaret O'Brien, Philip Tarrant, Bob Downs, Ruth Duerr, Erin Antognoli, Cyndy Parr, Nancy Ritchey, Paul Lemieux, Rebecca Koskela, Jonathan Sears, Lesley Wyborn, Kerstin Lehnert

CARE Principles
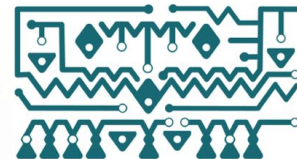
Collective Benefit

C1 For inclusive development and innovation
C2 For improved governance and citizen engagement
C3 For equitable outcomes

Authority to Control

A1 Recognizing rights and interests
A2 Data for governance
A3 Governance of data

Responsibility

R1 for positive relationships
R2 for expanding capability and capacity
R3 for indigenious languages and worldviews

Ethics

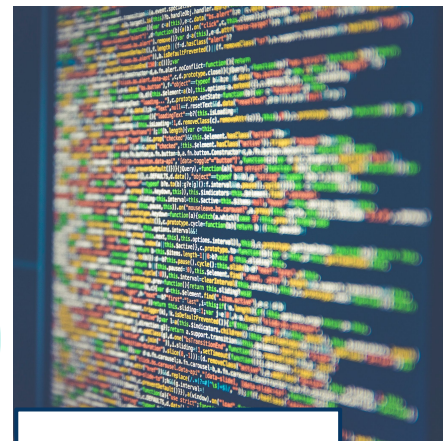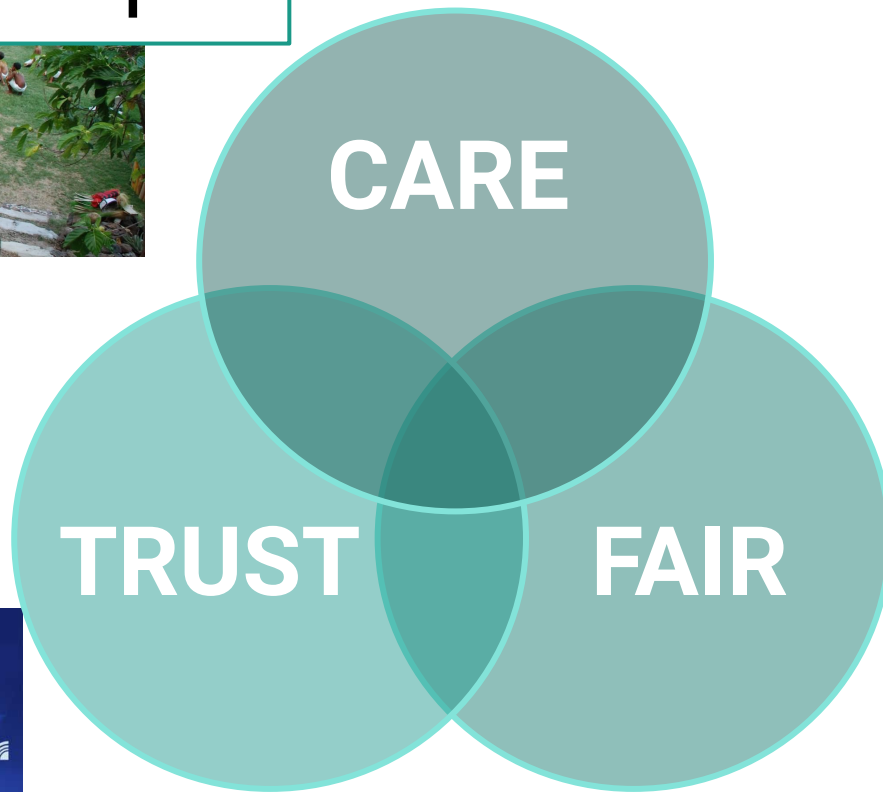**E1** For minimizing harm and maximizing benefit
E2 For justice
E3 For future use

https://www.gida-global.org/care

Repository Activity Categories (CARE-related)

Know your community and its data

Repository protocols

Communication and outreach

Technical aspects

Mixed

Know your community

1. Engage with indigenous communities (R1,R2,R3,E2)
2. Learn enough to determine if data access restrictions are necessary or for obfuscation of some elements (A1,A2,A3)
3. Understand indigenous legal rights (A2), consequences of publishing these data (E1,E3)
4. Be aware of changing roles/relationships over time (R1)

Repository protocols

1. Develop data policies and collection/curation protocols (A1)

2. Have contact protocols and follow up processes (C2)

3. Work with depositors to minimize restrictions (R1)

Communication and outreach

1. Consider your repository as part of a (broad) community (C2)
2. Identify that you hold data related to indigenous peoples (C1, C3)
3. Advertise governance protocols (A3)
4. Have transparent practices and defensible data management policies (A3)
5. Facilitate relationships between data provider and user communities (R1)
6. Share material with an indigenous workforce (R2)
7. Include indigenous representatives on repository advisory board if relevant (E1, E2)

Technical

1. Enable granular embargoes (C1)
2. Support tracking of data provenance (C1)
3. Link between data and papers, other outcomes (C2)
4. Create mechanism to control when/by whom data are accessed (A3)
5. Implement login systems that allow for transference of responsibilities (A3)
6. Receive (copies of) agreements between proprietors and package them permanently with data (A3)

Technical, cont.

7. Ensure that rights and restrictions are clearly communicated at deposition (R1)
8. Use a metadata format that handles multiple languages (R3)
9. Keep a list of data "actors" (e.g., by role) and be able to link data to them (R3)

Mixed activities

1. Work with IPs to develop keywords for their data discovery (C2, C3)
2. Ensure that depositing researchers have done their due diligence (C3)

3. Ensure metadata includes responsibilities of data users (R1)
4. Plan to evolve, in order to remain open and respect changing needs of IP (R1)
5. Describe appropriate use or limitations in metadata (R1)

Mixed activities, cont.

6. Allocate educational funds equitably, recognizing diff between "equal" and "equitable" (R2, E2)
7. Ensure tool design, selection of data formats, and metadata are appropriate for targeted communities and users (R3)
8. Implement text or badges that clarify types of use, e.g., Biocultural and Traditional Knowledge (E3)
9. Engage with IP review processes to assist in determining appropriateness of a repository for certain types of data, e.g, data are confidential, sensitive (A3, E1, E3)

Cluster Objective:

Checklist of Repository Activities

CARE Principles - ESIP 2021 Summer Mtg

File   Edit   View   Insert   Format   Data   Tools   Add-ons   Help   Last edit was seconds ago

Repository Mechanism or Activity	Collective Benefit			Authority to Control			Responsibility			Ethics		
	C1 For Inclusive development and innovation	C2 For improved governance and citizen	C3 For equitable outcomes	A1 Recognizing rights and interests	A2 Data for governance	A3 Governance of data	R1 For positive relationships	R2 For expanding capability and capacity	R3 For Indigenous languages and	E1 For minimizing harm and maximizing benefit	E2 For justice	E3 For future use
Engage with indigenous communities							x	x	x		x	
Learn enough to determine if data access restrictions are necessary or for obfuscation of some elements									x			
Understand indigenous legal rights, consequences of publishing these data					x							x
Be aware of changing roles/relationships over time							x					
Develop data policies and collection/curation protocols				x								
Have contact protocols and follow up processes	x											
Work with depositors to minimize restrictions							x					
Consider your repository as part of a (broad) community		x										
Identify that you hold data related to indigenous peoples	x		x									
Advertise governance protocols						x						
Have transparent practices and defensible data management policies						x						
Facilitate relationships between data provider and user communities							x					
Share material with an indigenous workforce									x			
Include indigenous representatives on repository advisory board if relevant										x	x	
Enable granular embargoes	x											
Support tracking data provenance	x											
Link between data and papers, other outcomes		x										
Create mechanism to control when/by whom data are accessed						x						
Implement login systems that allow for transference of responsibilities						x						
Receive (copies of) agreements between proprietors and package them permanently with data						x						
Ensure that rights and restrictions are clearly communicated at deposition							x					
Use a metadata format that handles multiple languages									x			
Keep a list of data "actors" (e.g., by role) and be able to link data to them									x			
Work with indigenous populations to develop keywords for their data discovery		x	x									

https://docs.google.com/spreadsheets/d/1jkDhSocA0w3znWvG1-epuBi2uCyv-N3LpH72Ogto4ns/edit#gid=353860576

Next Steps

1. Ensure that we have made all the necessary connections to CARE principles

2. TRUST - FAIR - CARE alignment

Provide or contribute to community catalogues

CARE

TRUST

FAIR

Transparent practices and defensible data management policies

Identify and respond to evolving community expectations

Implement relevant data metrics
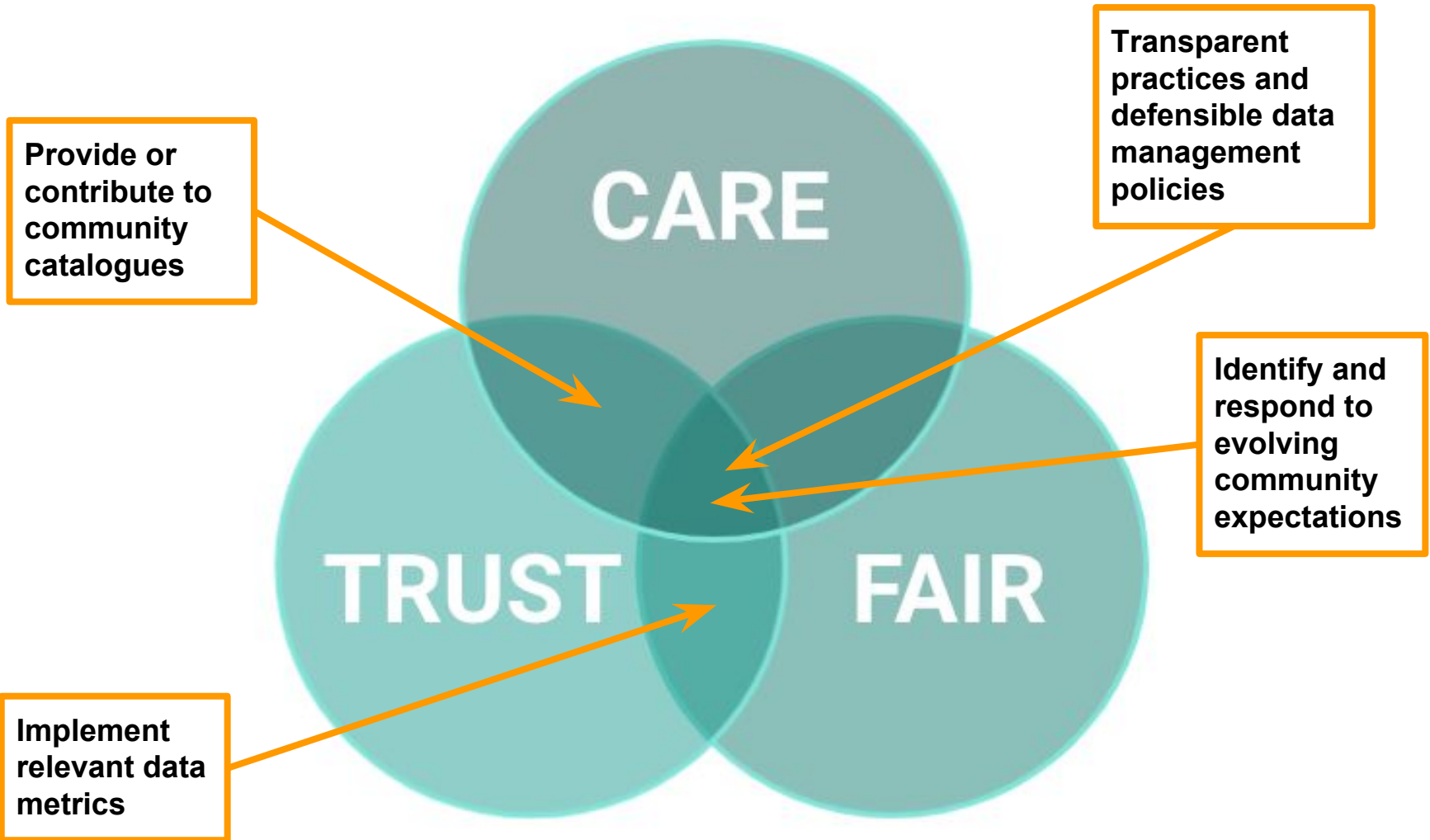
Discussion

1. What do the CARE principles mean to the operation of your organization?
2. How do you assess your repository's holdings, to know if CARE applies?
3. How is CARE different from FAIR and TRUST?
4. What might be intimidating about applying CARE principles?
5. What is exciting to anticipate, if we put them into practice?
6. Where are the gaps (vis-a-vis CARE) in your repo?
7. Which of these activities do you already do? What insights can you share?
8. Where is additional guidance or help needed?
9. What did the Cluster miss in their interpretation?

Discussion

What do the CARE principles mean to the operation of your organization?

How is CARE different from FAIR and TRUST?

Discussion

How do you assess your repository's holdings, to know if CARE applies?

Discussion

What might be intimidating about applying CARE principles?

What is exciting to anticipate, if we put them into practice?

Discussion

Where are the gaps (vis-a-vis CARE) in your repo?

Which of these activities do you already do? What insights can you share?

Discussion

Where is additional guidance or help needed?

What did the Cluster miss in their interpretation?

THANK YOU

Next Call: August 20, 4pm Eastern (3rd Friday)

https://wiki.esipfed.org/Sustainable_Data_Management

MAKING DATA MATTER
esipfed.org | #ESIPfed

Supported by: