



University of Tennessee, Knoxville
**TRACE: Tennessee Research and Creative
Exchange**

World Data System

Communication and Information

Summer 7-20-2023

Peers Speaking to Peers on Finding Sustainability

Mike Frame

Pedro Luiz Pizzigatti Correa

Rebecca Koskela

Giri Prakash

Follow this and additional works at: https://trace.tennessee.edu/utk_wds

Keynote speaker panel lunch:

Peers speaking to peers on how they “found” sustainability



USGS Science Mission Areas



Core Science Systems*

Delivers nationally focused Earth systems and information science that provides fundamental research and data that underpins all Mission Areas of the USGS



Ecosystems

Provides science to help America achieve sustainable management and conservation of biological resources in wild and urban spaces, and places in between



Energy & Minerals

Conducts research and assessments that focus on the location, quantity, and quality of mineral and energy resources, including the economic and environmental effects of resource extraction and use



Natural Hazards

Works with many partners to monitor, assess, and conduct targeted research on a wide range of natural hazards so that policymakers and the public have the understanding they need to enhance preparedness, response, and resilience

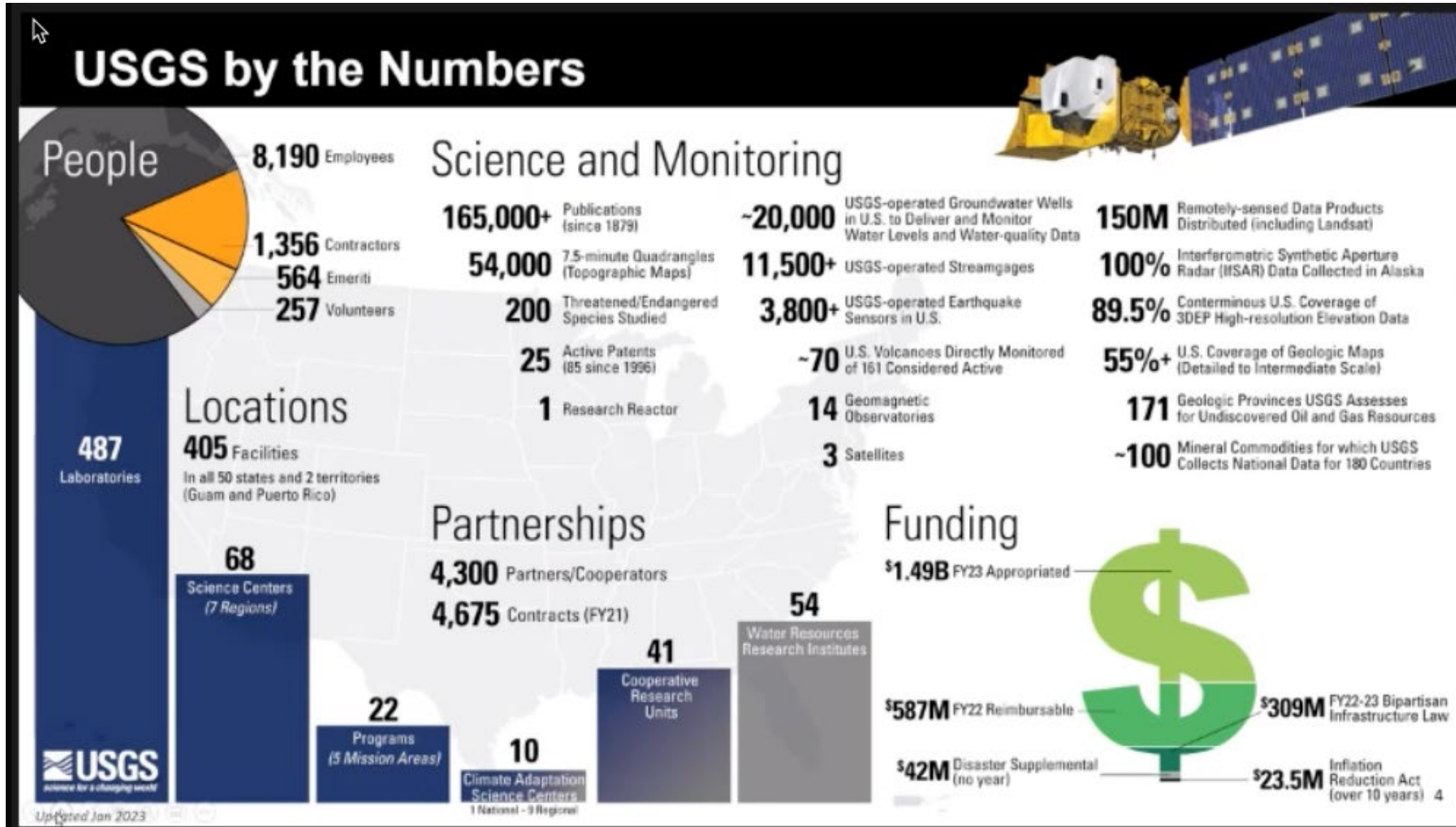


Water Resources

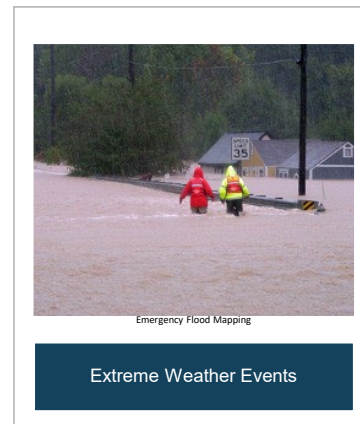
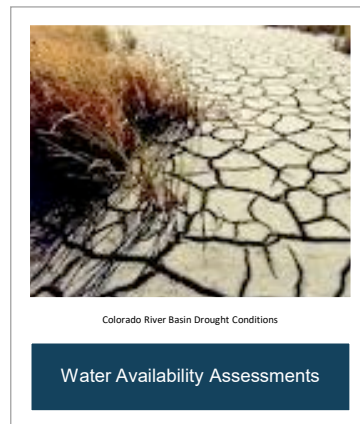
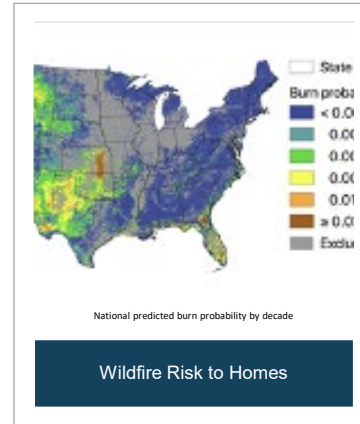
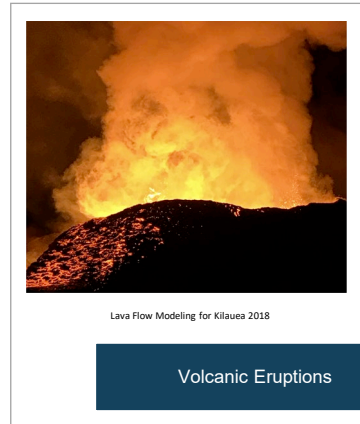
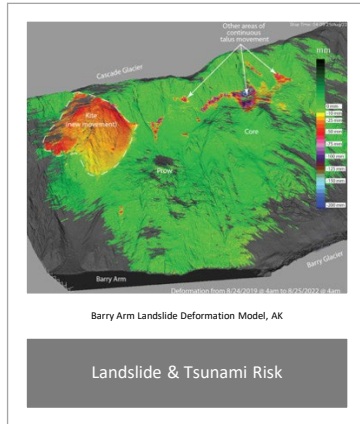
Works with partners to monitor, assess, conduct targeted research, and deliver information on a wide range of water resources and conditions including streamflow, groundwater, water quality, and water use and availability

Lead the Nation in 21st-century integrated research, assessments, and prediction of natural resources and processes to meet society's needs.

USGS – By the Numbers



USGS – Repositories & Sustainability



USGS – Repositories & Sustainability

Building Communities



Community for Data Integration
usgs.gov/cdi



USGS Powell Center
powellcenter.usgs.gov



Promoting Best Practices



USGS Data Management Website
usgs.gov/datamanagement

Developing Enterprise Tools

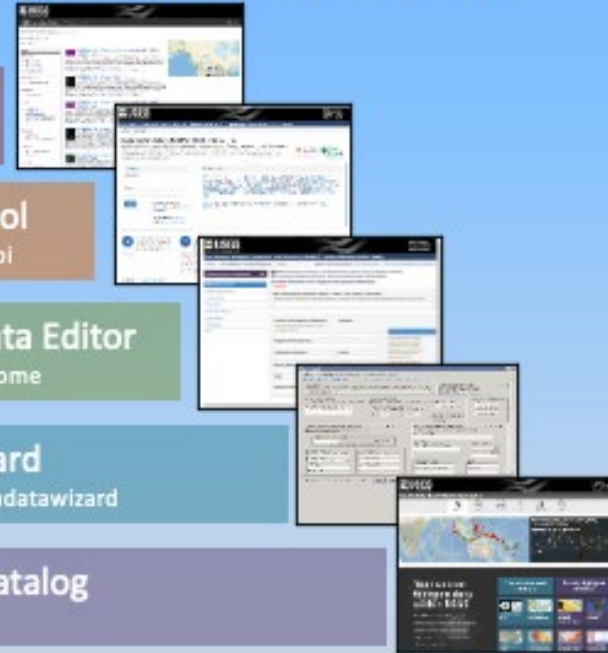
ScienceBase
sciencebase.gov

DOI Creation Tool
www1.usgs.gov/csas/doi

Online Metadata Editor
www1.usgs.gov/csas/ome

Metadata Wizard
sciencebase.gov/metadatawizard

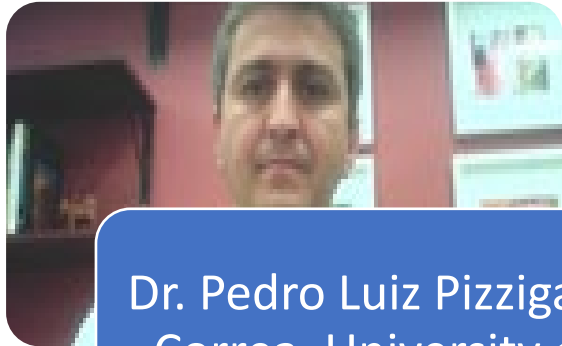
Science Data Catalog
data.usgs.gov



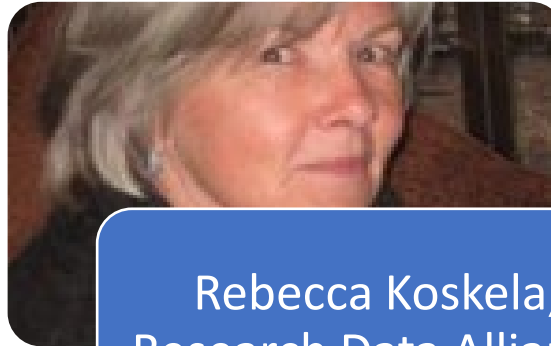
Acceptable Digital Repositories for USGS Scientific Publications and Data

Repository Name	Owner*	Funding Source*
USGS Sciencebase (USGS Trusted Digital Repository)	USGS	USGS
USGS Alaska Science Center Data Repository (USGS Trusted Digital Repository)	USGS	USGS
USGS Earth Resources Observation and Science (EROS) Center (USGS Trusted Digital Repository)	USGS	USGS
National Geologic Map Database (NGMDB) (USGS Trusted Digital Repository)	USGS	USGS
USGS Coastal and Marine Geoscience Data System (USGS Trusted Digital Repository)	USGS	USGS
USGS National Water Information System (NWIS)	USGS	USGS
Water Mission Area (WMA) National Spatial Data Infrastructure (NSDI) Node	USGS	USGS
LANDFIRE	U.S. Department of the Interior - Office of Wildland Fire and FS/USDA	U.S. Department of the Interior - Office of Wildland Fire and FS/USDA
Global Biodiversity Information Facility, United States (GBIF-US) Integrated Publishing Toolkit (IPT)	USGS	USGS, NSF
Biogeochemical Dynamics	NASA	NASA

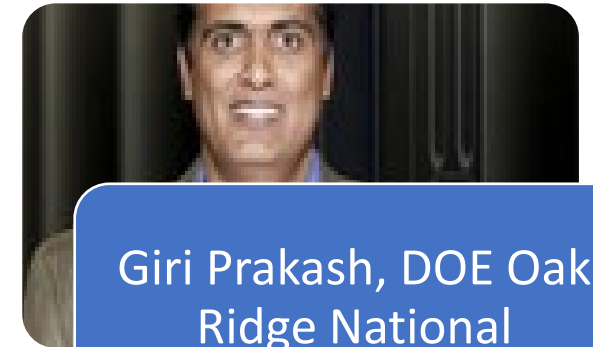
Panel Participants



Dr. Pedro Luiz Pizzigatti
Correa, University of
Sao Paulo, Brazil,
Professor




Rebecca Koskela,
Research Data Alliance
(RDA-US), Executive
Director



Giri Prakash, DOE Oak
Ridge National
Laboratory, Director of
ARM Data Center

Questions, Discussion, Feedback from ALL of YOU



WORLD DATA SYSTEM **REPOSITORY SUSTAINABILITY SUMMIT**

at the University of Tennessee in Knoxville, TN, USA

Keynote speaker panel lunch:
Peers speaking to peers on how they found sustainability
July 20th

Prof. Pedro Luiz Pizzigatti Corrêa - pedro.correa@usp.br

Digital Systems and Computer Engineering Department

Escola Politécnica da Universidade de São Paulo – EPUSP - Brazil

Big Data and Data Science Research Group of EPUSP wds.poli.usp.br



Study cases:

- Inter-American Biodiversity Information Network (IABIN);
- Data Portal of Brazilian Biodiversity – Ministry of Environment – Brazil.



- The Inter-American Biodiversity Information Network (IABIN) fosters technical collaboration and coordination among countries of the Americas in collection, sharing, and use of biodiversity information relevant to policy and decision-making on natural resources conservation and development (2003-2012).
- Funding
 - Global Environmental Facility – GEF;
 - World Bank.
- Support:
 - Organization of the American States.



THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP



OAS | More rights
for more people

Reference: <https://www.oas.org/en/sedi/dsd/iabin/>



Sub-Networks and Consortiums :

- Invasive Species: National Biological Information Infrastructure - NBII - (USGS);
- Species and Specimens: Instituto Nacional de Biodiversidad - INBio (Costa Rica);
- Protected Areas: Fundação O Boticário (Brazil);
- Pollinators: CoEvolution Institute (US) and Escola Politécnica of USP – EPUSP (Brazil);
- Ecosystems: Smithsonian Tropical Research Institute (Panama).

Tool:

- Catalog: NBII (USGS).

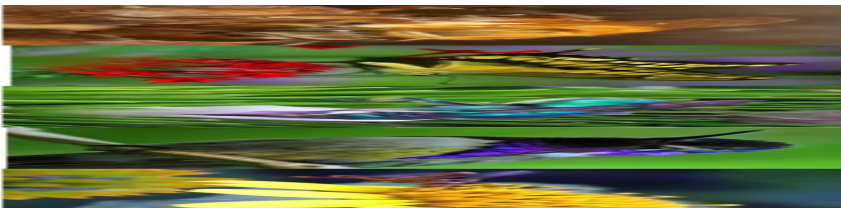


IABIN Activities:

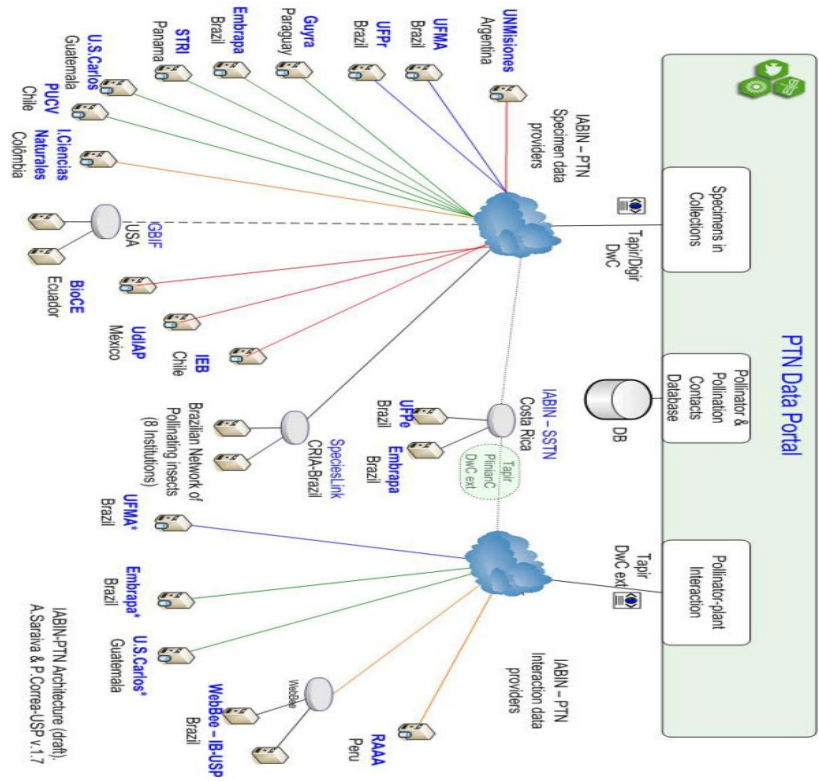
- Develop network IT infrastructure – Tools, Data standards;
- Grants for data content / tools creation;
- Training;
- Data providers meetings;
- Decision support tools.



Example of one sub-network:
Pollinators Thematic Network IABIN - PTN



Architecture
 Pollinators Information Network: IABIN - PTN -
 Architecture





iabin
Inter-American Biodiversity
Information Network

Sustainability: SWOT Analysis

- Strengths:
 - Community of experts (pollinators in special);
 - Infrastructure;
 - Unique Content (Assessment defined);
 - Partnership (national, regional, local);
 - Importance/urgency of issue (food security, political stability, conservation);



iabin
Inter-American Biodiversity
Information Network

Sustainability: SWOT Analysis

- Weaknesses:
 - Incompatibility of the Network's operating model with the institutions that inherited the legacy systems;
 - difficulty in maintaining a distributed network of subsystems without centralized coordination.



iabin
Inter-American Biodiversity
Information Network

Sustainability: SWOT Analysis

- Opportunities:
 - Training for development and generation of content for scientific data repository;
 - Collaboration peer to peer (Example: USP/USGS/ORNL/UT);
 - Development of new tools and repositories based on IABIN (Example: Brazilian Biodiversity Data Portal);




Sustainability: SWOT Analysis

- Threats

- Funding sustainability;
- Incomplete involvement of member States as an international initiative.



Data Portal of Brazilian Biodiversity – Ministry of Environment – Brazil - 2015

- Objective: Integrate biodiversity data collected by several federal research institutes, in addition to monitoring data from Brazilian Conservation Units.
- Funds: German Agency for International Cooperation –  Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
- Access: <https://portaldabiodiversidade.icmbio.gov.br/portal/>



Data Portal of Brazilian Biodiversity – Ministry of Environment

- Source code ported from the Atlas of Living Australia (ALA) by EPUSP in 2014;
- Supported by Ministry of Environment of Brazil .



Sustainability: SWOT Analysis

- Strengths
 - All data providers are institutions linked with Ministry of Environment – organizational facilities for data harvesting;
 - There is federal regular funds for operational support of the repository.



Sustainability: SWOT Analysis

- Weaknesses:
 - Difficult to get funds for maintenance;
 - Movement of staff responsible for data and system management.



Sustainability: SWOT Analysis


- Opportunities:
 - Receive all releases developed by international community of the ALA open source;
 - Facility to create national policies to increment the data reuse and reward for data publish.



Sustainability: SWOT Analysis

- Threats

- Difficulted of specific maintenance with demand only in Brazil;
- Missing skill in the staff related to software engineering and data management (focus only in the solution/data product and not on the staff development and research);
- High depend of the Atlas of Living Australia solution.



WORLD DATA SYSTEM **REPOSITORY SUSTAINABILITY SUMMIT**

at the University of Tennessee in Knoxville, TN, USA

Keynote speaker panel lunch:
Peers speaking to peers on how they found sustainability
July 20th

Prof. Pedro Luiz Pizzigatti Corrêa - pedro.correa@usp.br

Digital Systems and Computer Engineering Department

Escola Politécnica da Universidade de São Paulo – EPUSP - Brazil

Big Data and Data Science Research Group of EPUSP wds.poli.usp.br



Peers to Speaking to Peers on Searching for Sustainability

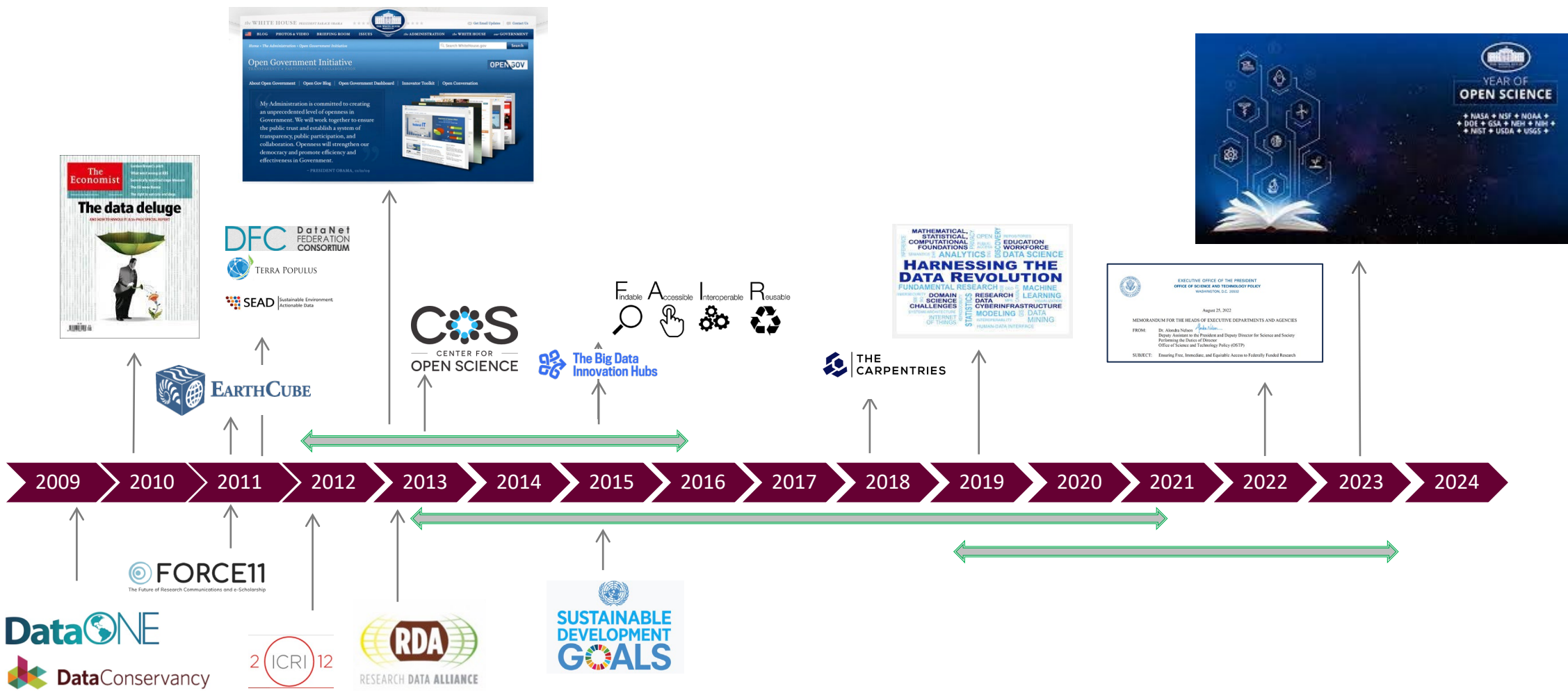
Rebecca Koskela

rebecca.koskela@rda-foundation.org

Executive Director, RDA US

20 July 2023

World Data System Sustainability Summit 2023



DataONE

DataNet Objectives

Program Solicitation 07-601

The new types of organizations envisioned in this solicitation will integrate library and archival sciences, cyberinfrastructure, computer and Information sciences, and domain science expertise to:

- provide reliable **digital preservation, access, integration, and analysis** capabilities for science and/or engineering data over a decades long timeline;
- continuously **anticipate and adapt to changes** in technologies and in user needs and expectations;
- engage at the frontiers of computer and information science and cyberinfrastructure with research and development to **drive the leading edge forward**; and
- serve as component elements of an **interoperable data preservation and access network**.

By demonstrating feasibility, identifying best practices, establishing viable models for long term technical and economic sustainability, and incorporating frontier research, these exemplar organizations can serve as the basis for rational investment in digital preservation and access by diverse sectors of society at the local, regional, national, and international levels, **paving the way for a robust and resilient national and global digital data framework**.

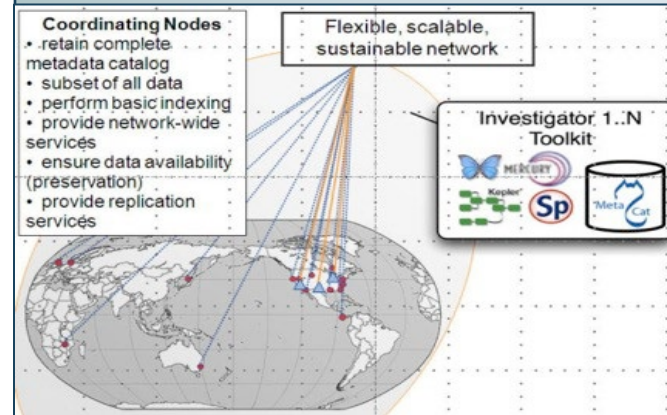
DataONE Vision and Approach

Enable new science and knowledge creation through easy access to data about life on earth and the environment that sustains it, plus access to key tools.

1. Build on existing cyberinfrastructure



2. Create new cyberinfrastructure



3. Support communities of practice



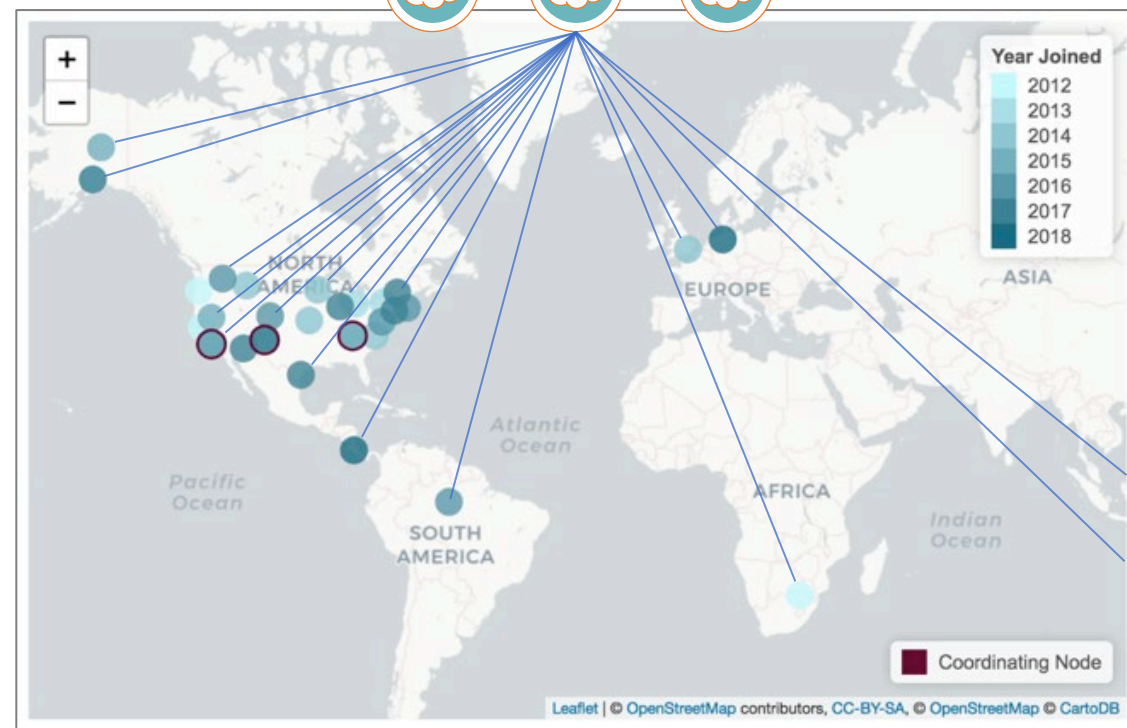
DataONE: A Federation of Repositories

Core Cyberinfrastructure



```

1- #####
2- ### GoA Hydrocarbon Data Cleaning #####
3- ### March 2015 ; Script by Rachael Blake ###
4- #####
5
6 # Set your working directory (this should be changed to download from th
7 setwd("C:/Users/rblake/Documents/NCEAS/GoA Portfolio Effects WG/Hydrocar
8
9- #####
10 # Calculate Total Aromatics
11 PAH <- read.csv("PAH.csv") # read in the PAH data file
12 head(PAH) ; str(PAH)
13
14 # Taking means of all chemical compound concentrations to get Total PAHs
15 PAH$TotalAromatic <- rowSums(PAH[,24:71], na.rm=T) # sum the chemicals a
16 PAH$TPAH_adj <- rowSums(PAH[,c(24:26,28,30:45,47:71)], na.rm=T) ; PAH[c
17 library(plyr)
18 PAH1 <- arrange(PAH, $In) ; head(PAH1) # arrange by the Sample ID ($In
19 TotalAromat <- PAH1$PAH1$In > 0. -c(3.8.15.22.24.71) # remove rows w
    
```



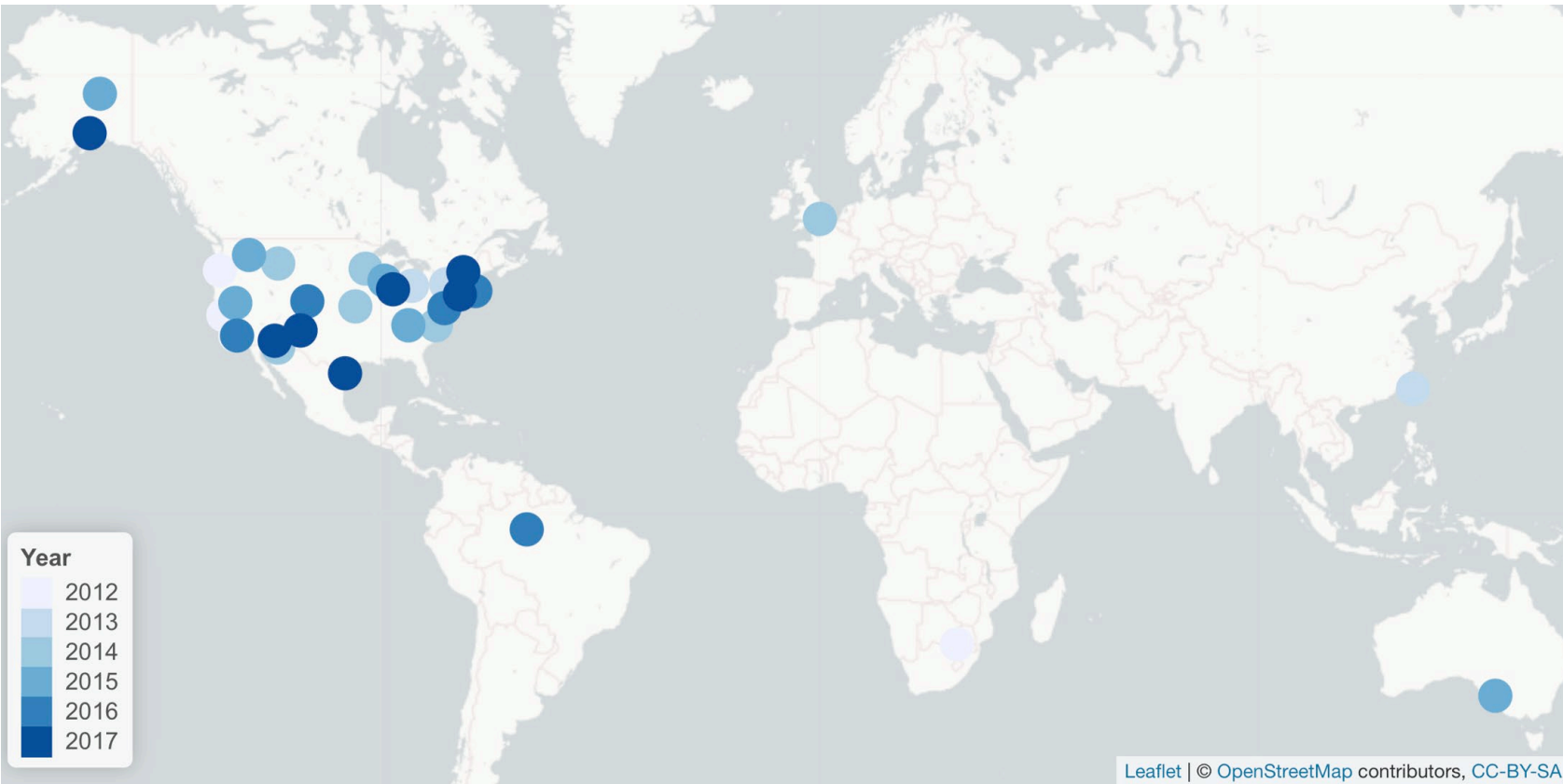
- GLEON
- LTRE Europe
- ORNL DAAC
- UC3Merritt
- ITER
- IARC
- SEAD
- DRYAD
- ONEShare
- AOOS
- KU BIODIVERSITY INSTITUTE
- Data Gateway
- PISCO
- MPC
- South African Botanical
- EDAC
- The Cornell Lab
- TERN
- kn
- esa

- PPBio
- UNIVERSITY LIBRARY
- NKN
- New Mexico EPSCoR
- tDAR
- EDD
- USGS
- BC-DM
- ESS-DIVE
- neon
- FEMC
- GULF-MEXICO PROGRAM INSTITUTE
- GRIDD
- NRDC

43 Member Nodes and Growing

Data Preservation and Services

Components for a flexible, scalable, sustainable network



Member Nodes

- diverse institutions
- serve local community
- provide resources for managing their data
- retain copies of data

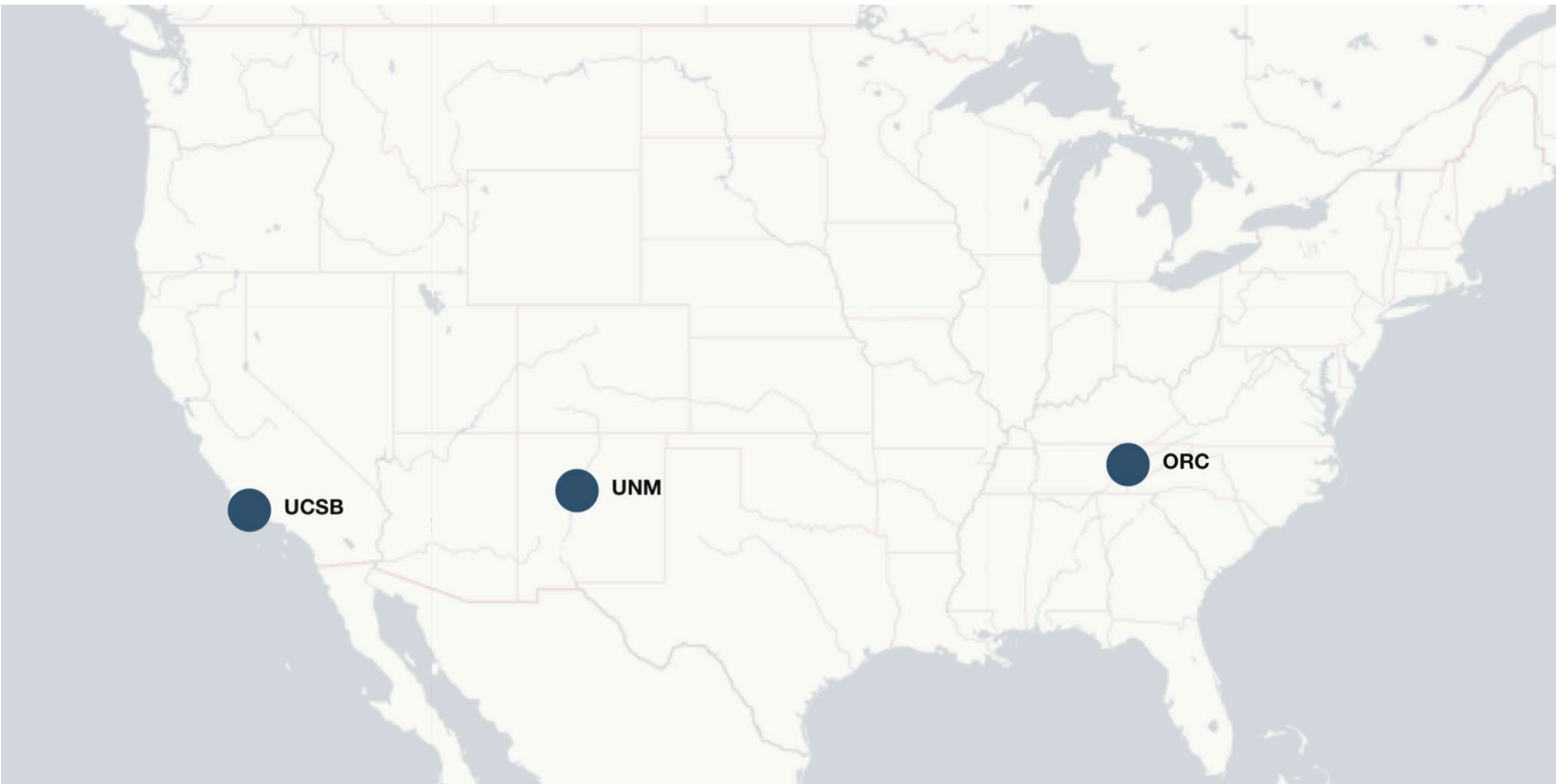


Leaflet | © OpenStreetMap contributors, CC-BY-SA, ©

Coordinating Nodes

Federation-wide Services

Components for a flexible, scalable, sustainable network



Coordinating Nodes ●

- retain complete metadata catalog
- indexing for search
- network-wide services
- ensure content availability (preservation)
- replication services



Network-wide services

Metadata search catalog

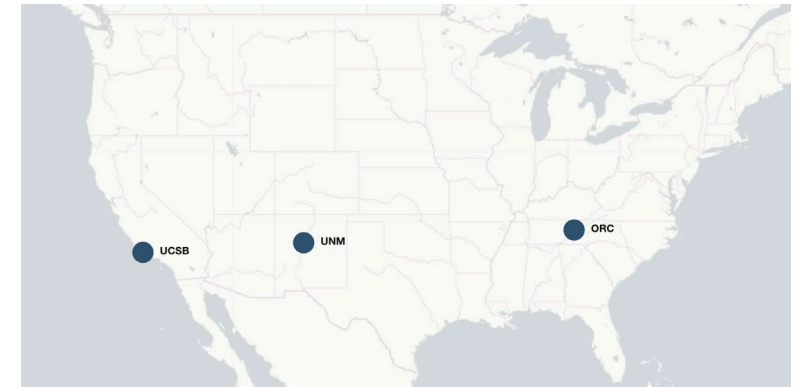
Citation and usage

Reproducibility and provenance

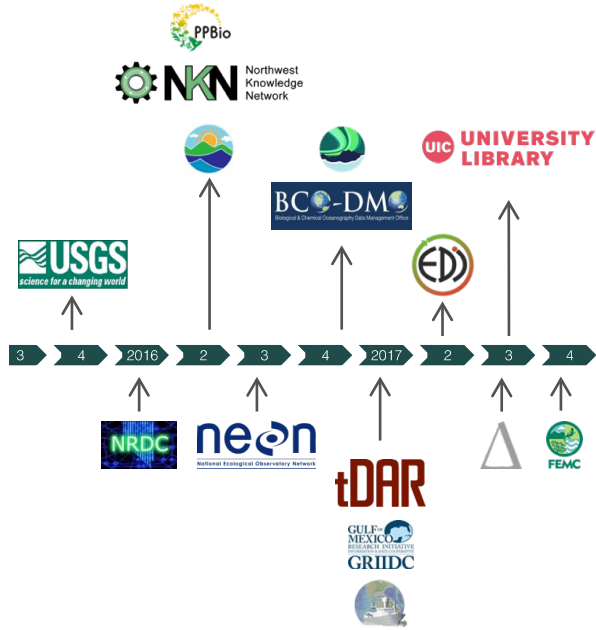
Replication and Audit

Provenance and Quality services

Coordinating Nodes



DataONE: Addressing Four Key Needs



Practical Tools for Harmonizing the Semantic Heterogeneity of Data

Data Table, Image, and Other Data Details

4 sources

Source Data

Alkane.csv

Citation

View >>

This data was generated by </> DataDownload.R.

This data prov_hasDerivations the data you are currently viewing, Total_Aromatic_Alkanes_PWS.csv and the data you are currently viewing, Total_Aromatic_Alkanes_PWS.csv.

This data was used by </> Total_PAH_and_Alkanes_GoA_Hydrocarbons_Clean.R.

An Empowered and Engaged Community

Discovery, Acquisition and Preservation of Relevant Data

Attribute Information

Variables

- site
- wet/dry
- post
- litter
- deli surv
- cats
- Formica
- total ants

Name: litter

Label: Leaf Litter Carbon Pool

Definition: Add tag to litter attribute. Help others find this dataset by adding semantic tags.

Storage Type: soil

Measurement Type: Soil Layer Top Depth

Definition: Depth from soil surface to top of soil layer

from the ECSO ontology (GUID ECSO_00000096)

Has related synonym: soil litter

soil litter

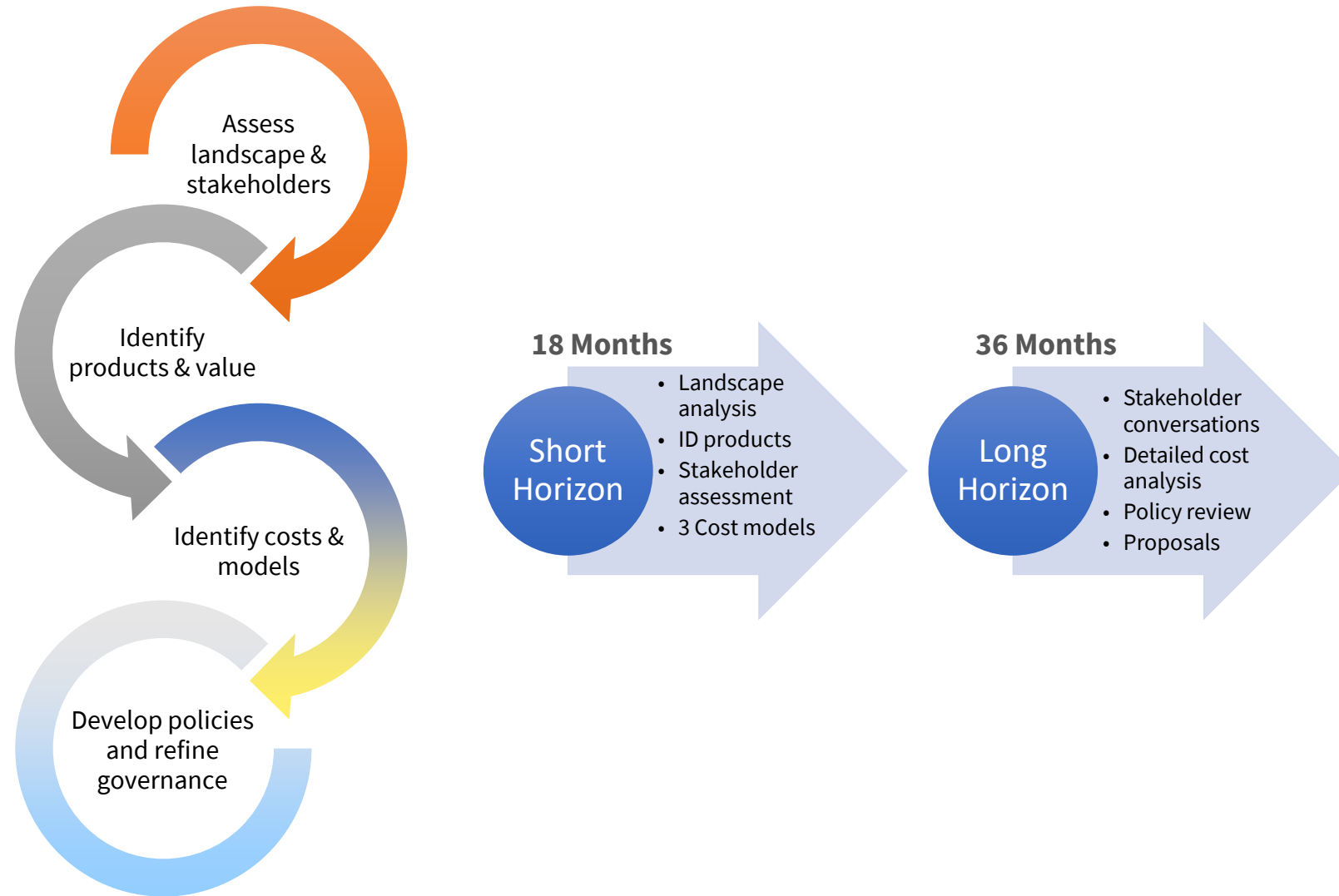
soil loss

soil order

Data and Practical Tools for Reproducible Science and Provenance



DataONE Phase 2: 18-36 month activities



DataONE

Making data more discoverable, accessible,

What data are you looking for?

ADVANCED SEARCH

NEW!

Introducing Products & Services from DataONE



DATAONE PLUS

Enhanced DataONE products and services in support of your research.

[LEARN MORE ABOUT DATAONE PLUS](#)



HOSTED REPOSITORIES

A dedicated repository solution for your group or institution's data, managed by DataONE.

[DISCOVER DATAONE HOSTED REPOSITORIES](#)

[SEE ALL FEATURES](#)

Current member repositories 44

Replication repositories 3

Contributing repositories 4

Upcoming member repositories 1

SUSTAINABILITY IS . . .

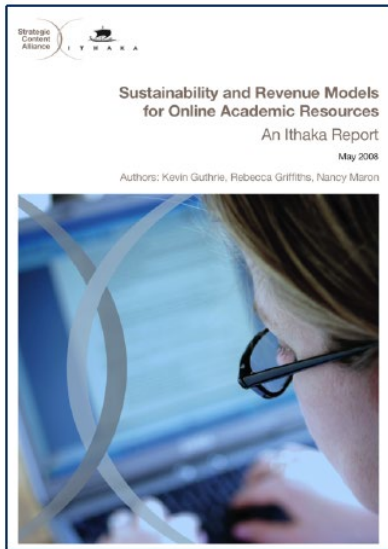
the ability to generate or gain access to the resources—financial or otherwise—needed to protect and increase the value of the content or service for those who use it.

A sustainability plan is an evidence-based strategy for identifying and securing these ongoing sources of support for your project or center over time.



Slide credit: Nancy Maron, ESA NSF PI Workshop on Creating and Implementing Sustainability Plans for Data Repositories, January 24, 2018

Sustainability is a Work in Progress



2008



2014

- Dec 2008: Blue Ribbon Task Force on Sustainable Digital Preservation and Access
- Mar 2010: Data Access and Interoperability Task Force (International)
- 2012 – 2015: COOPEUS (cooperation between the US and the EU in the field of environmental research infrastructures)
- Jun 2013: ICPSR Sustaining Domain Repositories for Digital Data
- Jun 2014: Ithaka Sustaining Digital Resources Course
- Nov 2014: ICPSR Integrating Domain Repositories into the National Data Infrastructure
- Jan 2017: Tempe Workshop "Return of Investment for Data Repositories"
- Dec 2017: OECD Business Models for Sustainable Research Data Repositories
- Jan 2018: ESA NSF PI Workshop on: Creating and Implementing Sustainability Plans for Data Repositories
- Nov 2020: Science Gateway Community Institute Introduction to Sustainability Workshop
- 2021 -2022: Sustainability models for integrated digital Earth Science beyond EarthCube

SUCCESSFUL PROJECTS HAD CERTAIN THINGS IN COMMON



Slide credit: Nancy Maron, ESA NSF PI Workshop on Creating and Implementing Sustainability Plans for Data Repositories, January 24, 2018

SUCCESSFUL PROJECTS HAD CERTAIN THINGS IN COMMON



- What is specifically, does your project help its audience do?
- Opportunistic (in a good way)
- Clearly defined value to clearly defined audience
- Relentless
- Time matters, you know if your project is a success? Number of users? Needs and/or changing environment?
- Number of institutional members delivering this
- What must you pay for and what might you be able to acquire through partnerships, volunteer support, or other in-kind contributions?
- What is meaningful to your work?
- How and how often will you measure this? Can it be recurring?
- How will you know if the mission and long-range plans for your host institution are not met?
- If your institution offers support, it is a major stakeholder; stay close in touch!

Research Data Alliance (RDA)

Research Data Alliance (RDA)

Global community-driven organization launched in March 2013 to accelerate data-driven innovation

Vision

Researchers and innovators openly share and **re-use** data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

RDA builds the **social and technical bridges** that enable open sharing and **re-use** of data.

64 FLAGSHIP OUTPUTS

including 8 ICT Technical Specifications

200+ ADOPTION CASES

across multiple disciplines, organisations & countries

91 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

41 Working Groups
65 Interest Groups
1 Community of Practice

13,671 INDIVIDUAL MEMBERS FROM 151 COUNTRIES

69% Academia & Research
14% Public Administration
11% Enterprise & Industry

65 ORGANISATIONAL MEMBERS
12 AFFILIATE MEMBERS

RDA Groups - Overview

RDA is a pragmatic and agile organization, and through the work of its Working and Interest Groups, directly and logically tackles numerous data infrastructure challenges.

Working Groups & Interest Groups: what's the difference?

Working Groups

Working Groups are short-term (18 months) and come together to develop and implement data infrastructure, which could be tools, policy, practices and products that are adopted and used by projects, organizations, and communities. Embedded within these groups are individuals who will use the infrastructure and help in making it broadly available to the public.

41

Interest Groups

Interest groups are open-ended in terms of longevity. They focus on solving a specific data sharing problem and identifying what kind of infrastructure needs to be built. Interest Groups can identify specific pieces of work and start up a Working Group to tackle those.

65

Research Data Management Life Cycle

[Ethical and Legal best practices for Drone Data in a global research context \(WELDD\) WG](#)

[Artificial Intelligence And Data Visitation \(AIDV\) WG](#)

Research Question

[Engaging Researchers with Data IG](#)

[Education and Training on Handling of Research Data IG](#)

[Earth, Space, and Environmental Science Complex Citations WG](#)

[RDA/WDS Publishing Data Workflows WG](#)

Publication

Research Data Management Lifecycle [example from USCS](#)

[Discipline-specific Guidance for Data Management Plans WG](#)

[DMP Common Standards WG](#)

[Exposing Data Management Plans WG](#)

[Active Data Management Plans IG](#)

Data Management Plan

[Data Citation WG](#)

Collection

[Data Discovery Paradigms IG](#)

Re-collection

Data Storage

Description

[Research Metadata Schemas WG](#)

[Metadata IG](#)

[FAIR for Virtual Research Environments WG](#)

Analysis

[FAIR for Research Software \(FAIR4RS\) WG](#)

Archive



[Data Repository Attributes WG](#)

[Preservation Tools, Techniques, and Policies IG](#)

[Archives and Records Professionals for Research Data IG](#)

[Domain Repositories IG](#)

1 Data Management

Develop best practices and tools for non-static, machine-readable data management plans which can evolve throughout the research data lifecycle, as well as be machine-readable by collaborators and stored with the data.

2

Data Collection

Capture scientific evidence that allows analysis to lead to the formulation of convincing and credible answers to questions of researchers.



Data Description

Documentation is an essential component of research data management and allows researchers to make sense of data in the future.



3

4

Identity, Store, and Preserve

Collections and streams of digital objects are growing at an incredibly rapid pace. We need to get clear on what these objects are, how we need to be documenting and storing them, and how they should all link and talk to one another before we get overtaken by a forest of data.



Disseminate, Link, and Find

An increasing number of publishers and journals are implementing policies that require or recommend published articles to be accompanied by the underlying research data.



5

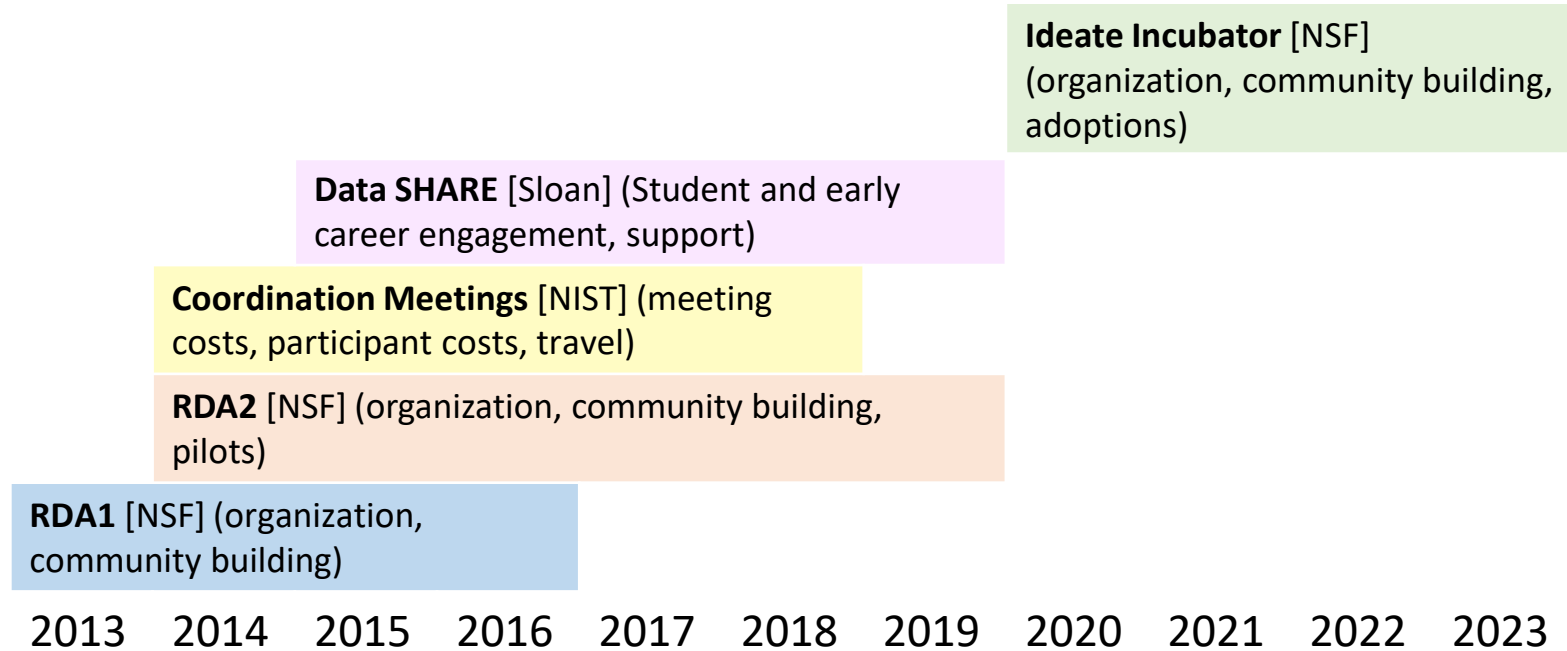
6

Policy, Legal Compliance, and Capacity

Currently all sites and scientific communities use their own set of policies, if any. A generic set of policies that can be revised and adapted by users to build up their own data collection does not exist.

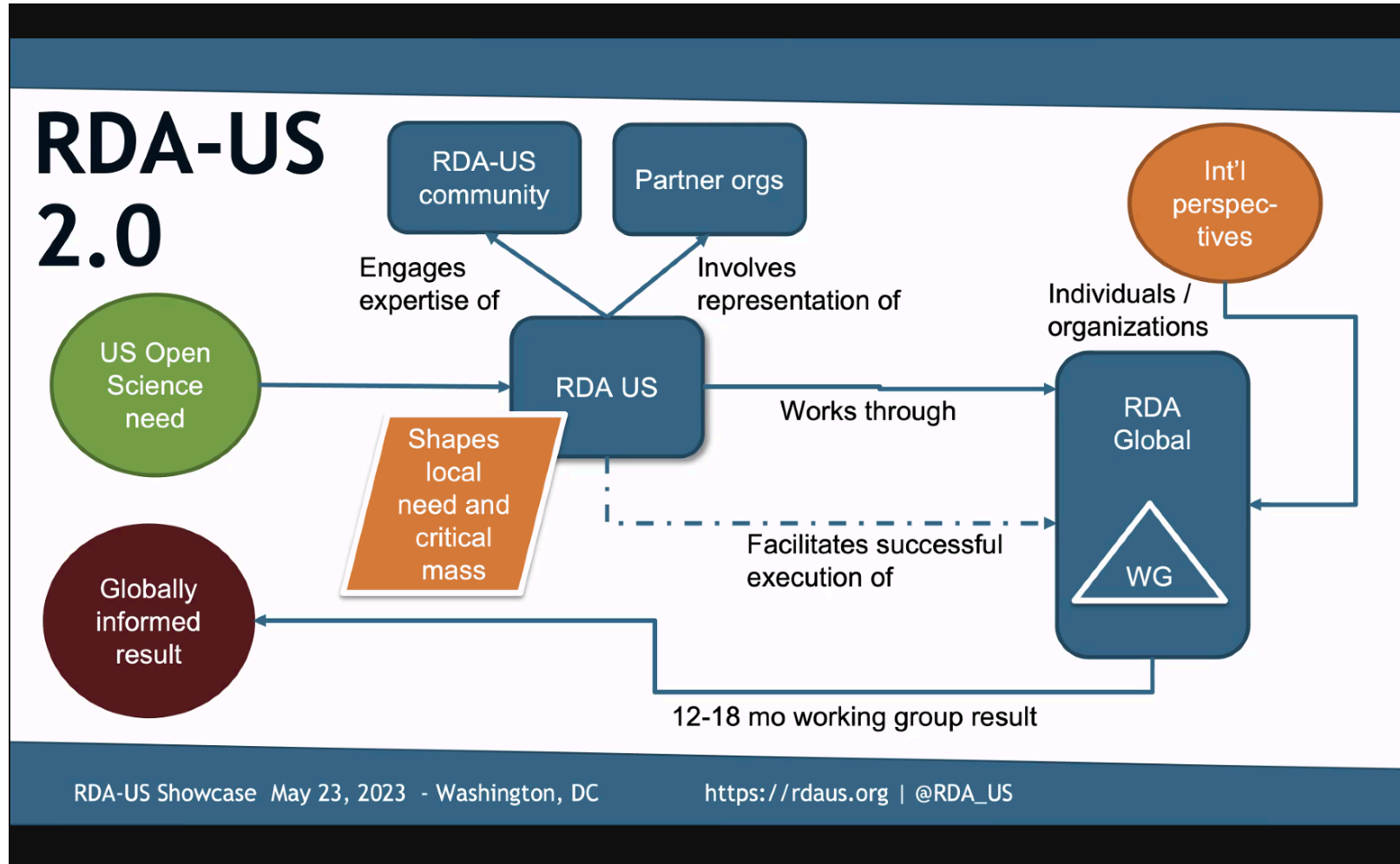


RDA/US Funding



But there's always a sustainability issue

RDA/US Going Forward



Contact

RDA Global

Email - enquiries@rd-alliance.org

Web - www.rd-alliance.org

Twitter - [@resdatall](https://twitter.com/resdatall)

LinkedIn - www.linkedin.com/in/ResearchDataAlliance

RDA US

Email - rdaus@rda-foundation.org

Twitter - [@RDA_US](https://twitter.com/RDA_US)



Thank you!

Rebecca Koskela
rebecca.koskela@rda-foundation.org
@rkoskela




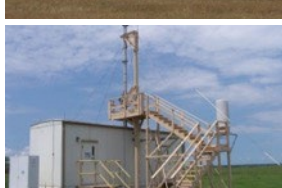

Sustainability Discussion – An example from the ARM Data Center

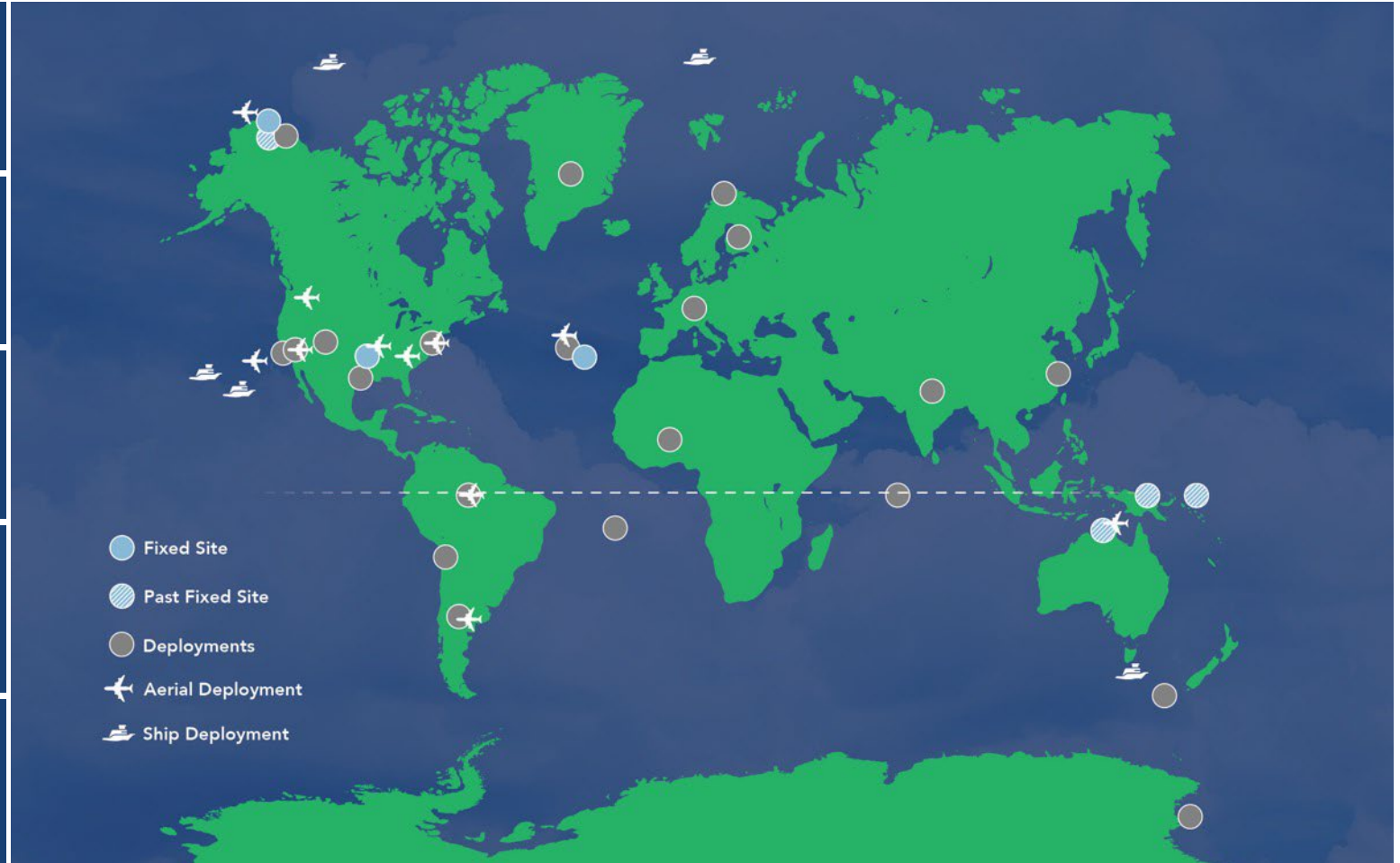
GIRI PRAKASH

ARM Data Center, Oak Ridge National Laboratory
palanisamyg@ornl.gov

World Data System Repository Sustainability Summit, July 20, 2023

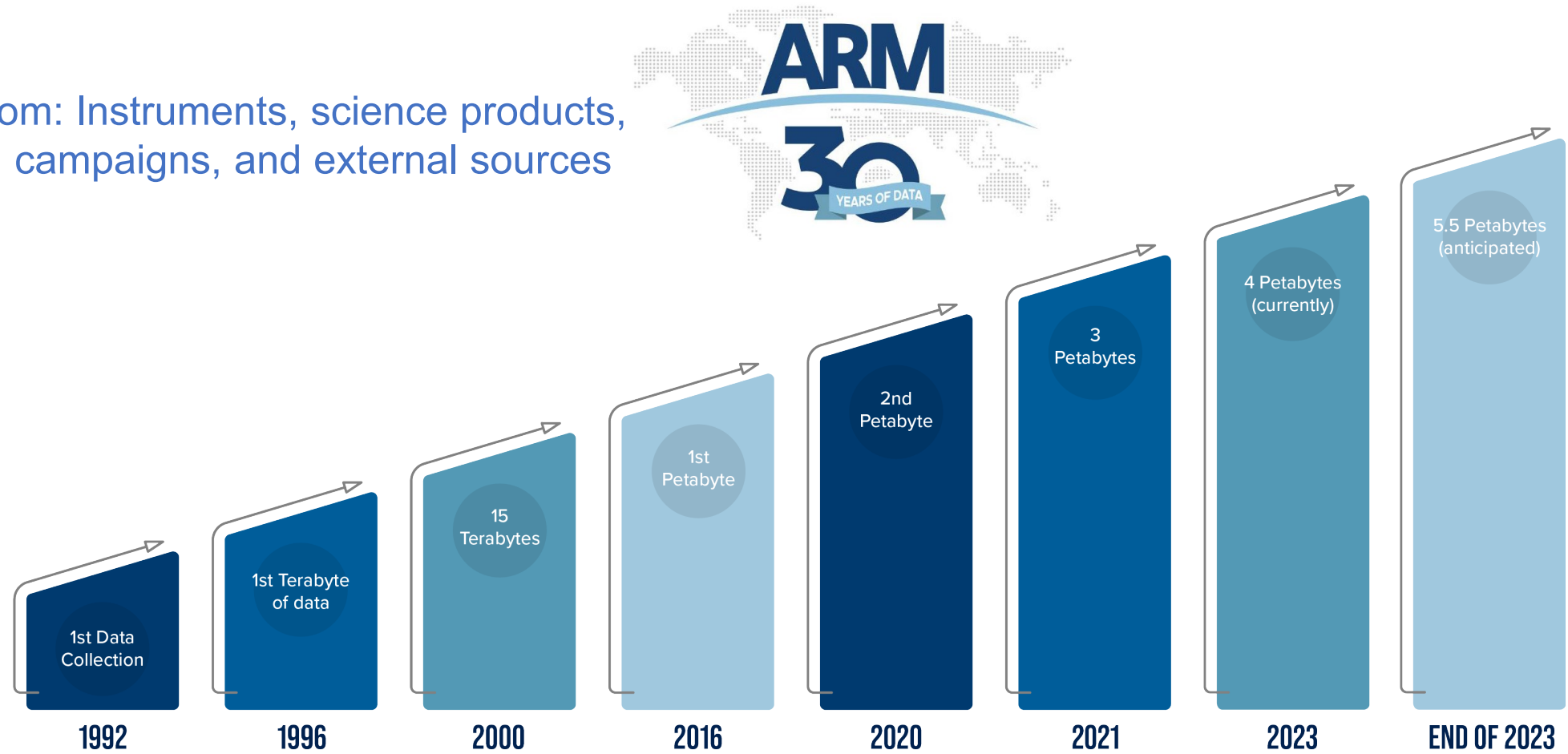
Atmospheric Radiation Measurement (ARM) Facility

	Background atmospheric state
	Surface energy balance
	Aerosol and hydrometeor profiles
	Near-surface aerosol properties
	Aerial measurements



Data and Users At a Glance

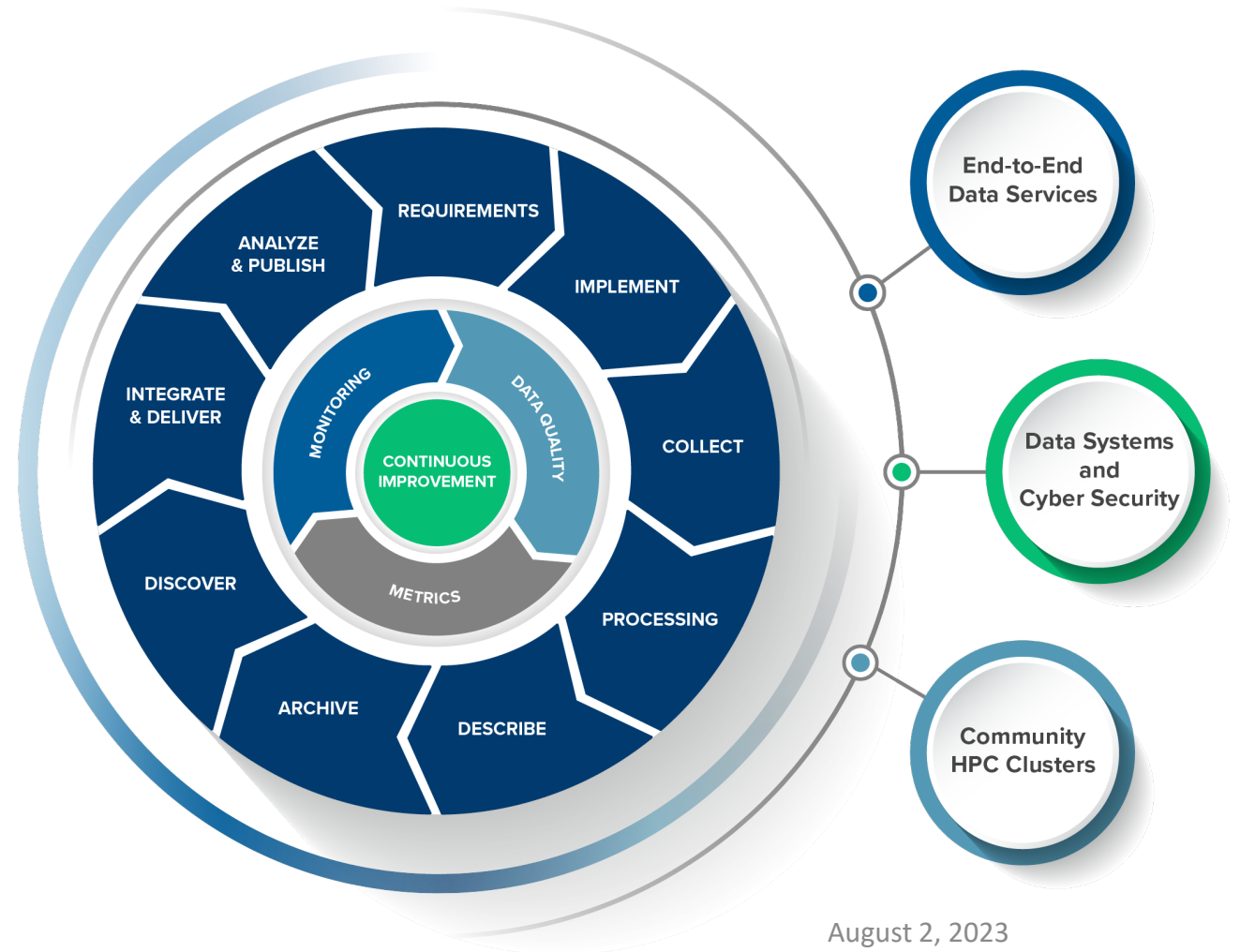
- ▶ Data From: Instruments, science products, models, campaigns, and external sources



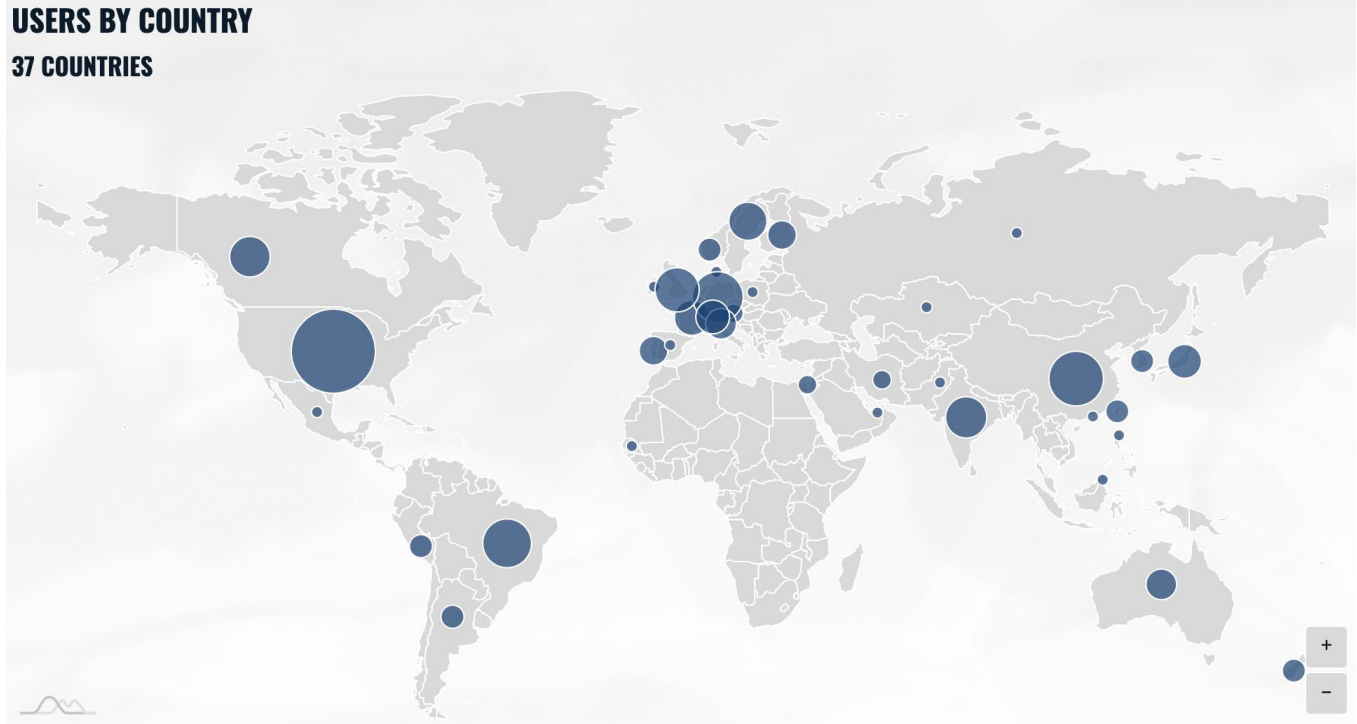
About The ARM Data Services

Provides a robust integrated data and computing ecosystem to advance understanding of atmospheric measurements

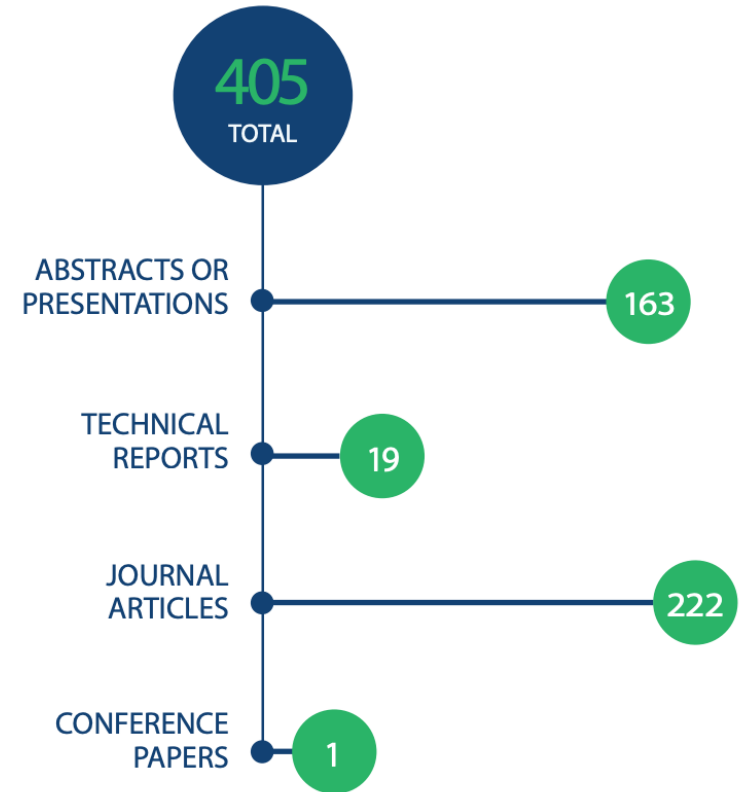
- ▶ Data flow operations and monitoring
- ▶ Advanced data collection systems
- ▶ High-performance computing (HPC)
- ▶ Comprehensive Data Processing
- ▶ Data Interoperability:
 - Advanced strategies for utilizing metadata
 - Data Discovery
 - Data workbench
 - FAIR, Standards, and Protocols
- ▶ User Management and Citations
- ▶ AI-based approach in data management



Data and Users: Key Metrics for Programmatic Sustainability



PUBLICATIONS USING ARM*

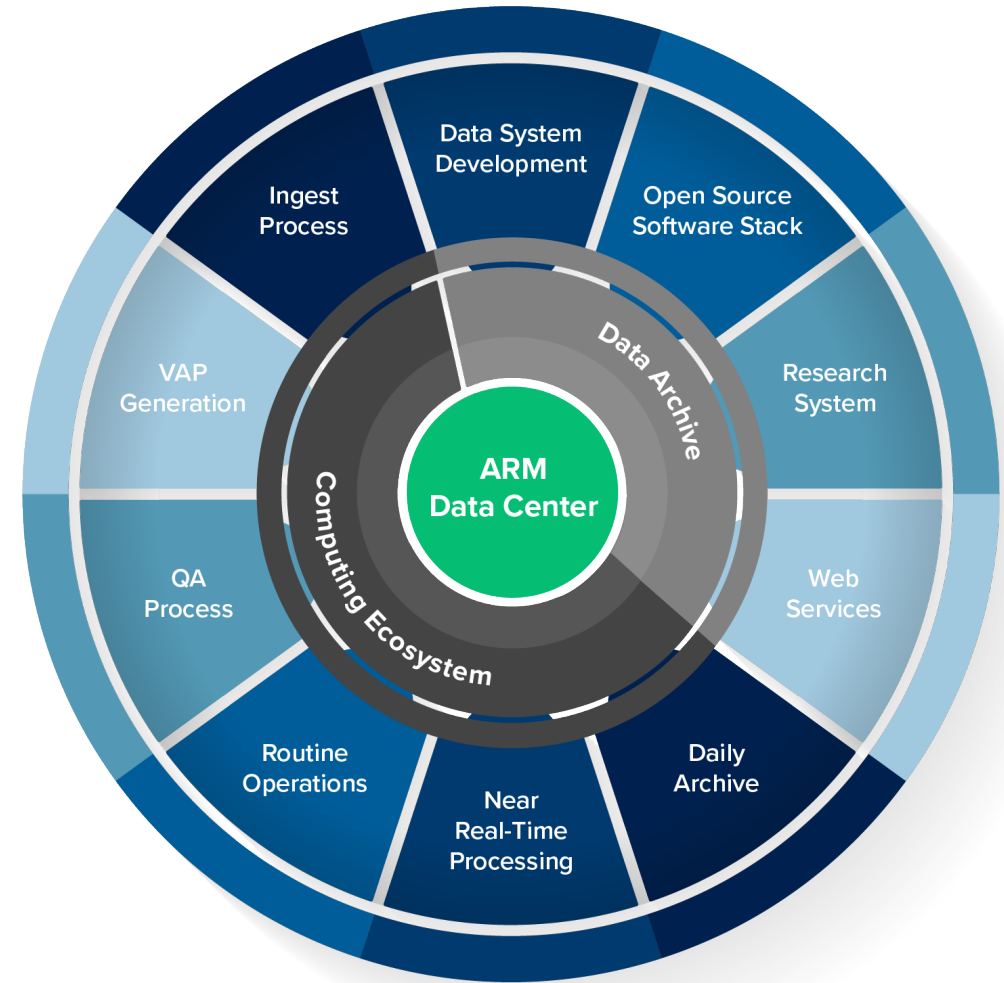


*Publication statistics were collected as of December 2022. Journal article numbers will continue to increase over time.

Streamlining Data Operations to Fuel Innovations

Objective: Optimizing funds allocation by reducing operational expenses and increasing investments in Research and Engineering activities.

- ▶ **Achieving Financial Efficiency:** A major operational streamlining of ARM Data Services during 2017-18 resulted in a significant reduction of over 30% in allocated funds for operations.
- ▶ **Driving Innovation:** The streamlining initiative facilitated funds for the development of new capabilities and successful proof-of-concepts.
- ▶ **Expanded Funding Potential:** These advancements opened doors to additional funding opportunities.

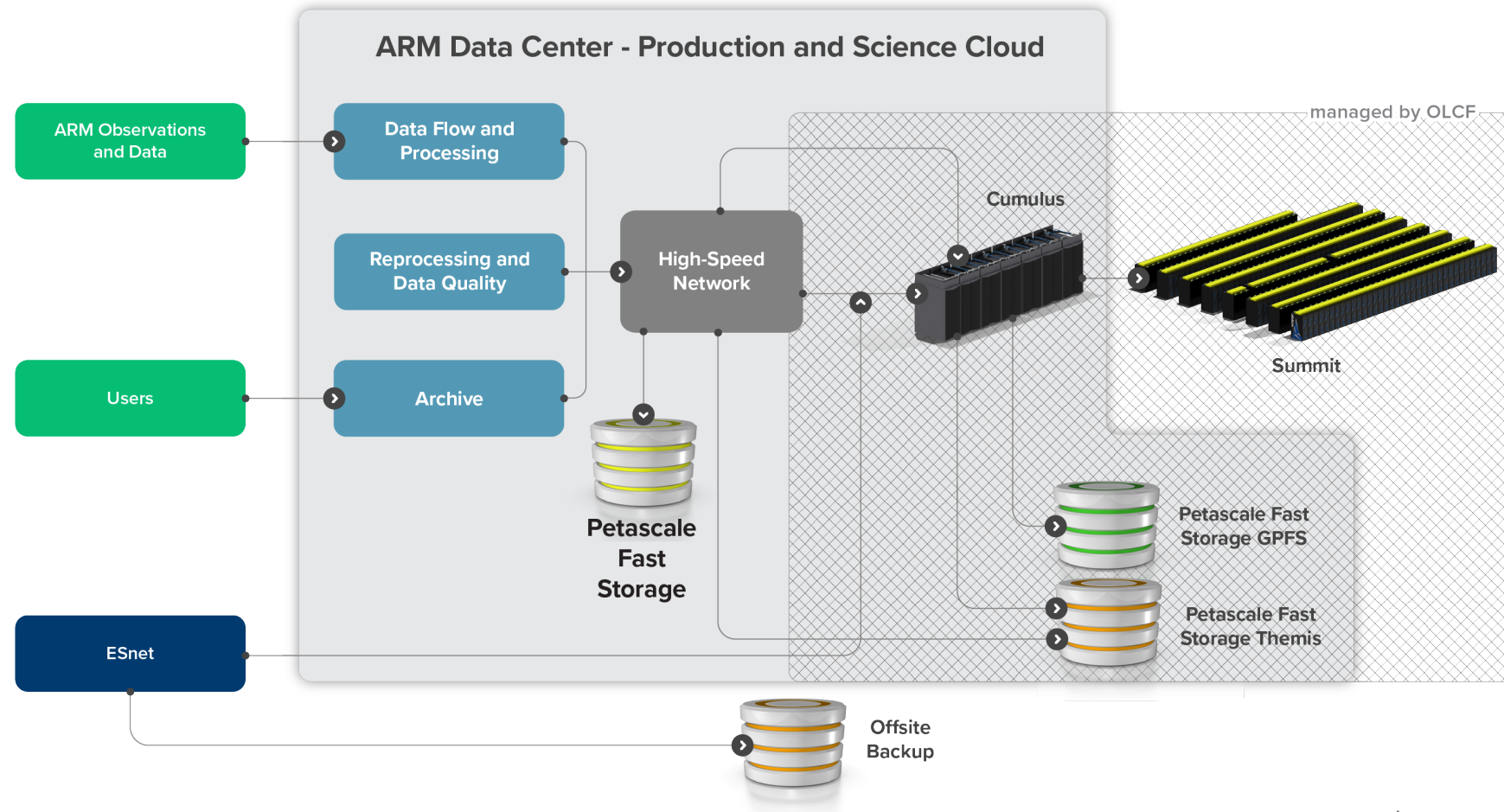


Collaboration and Reuse of Computing Capabilities

▶ ARM Data Center cyberinfrastructure utilizes synergy across DOE computing facilities

▶ Leveraging

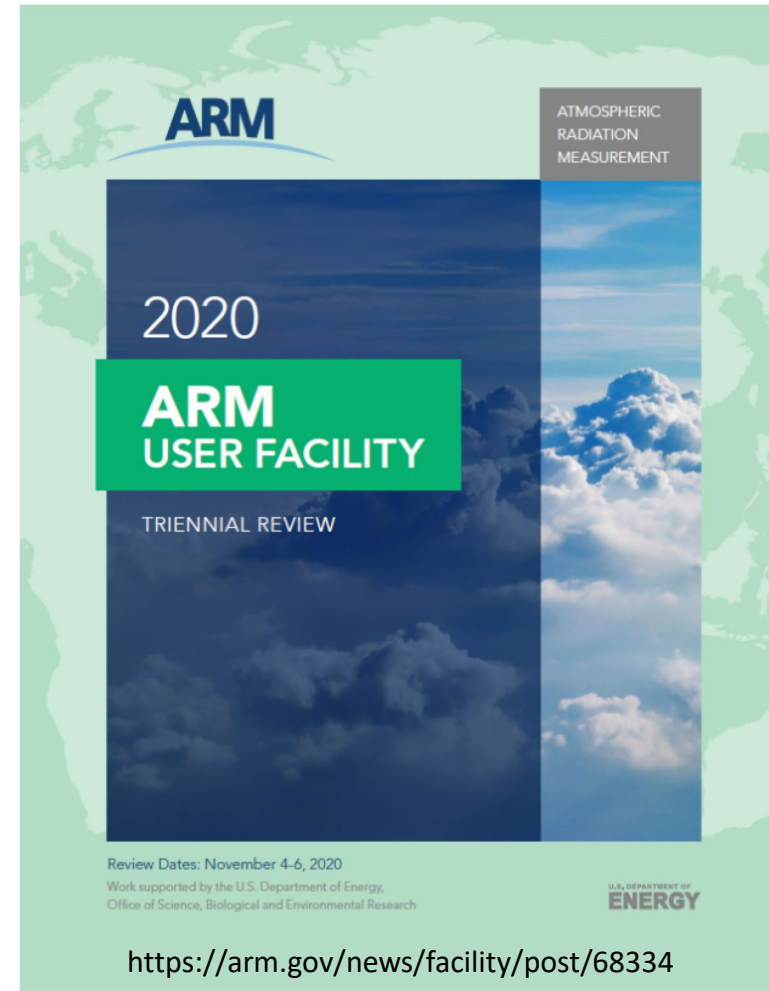
- Domain expertise
- Shared and established resources
- Cutting-edge solutions



Crucial Role of Programmatic Reviews in Ensuring Sustainability and Growth

Transforming Challenges into Opportunities

- ▶ ARM participates in Triennial Reviews
 - The last review resulted in 5 major recommendations to improve the usability and visibility of ARM Data
 - Resulted in new data, computing, and software development activities
- ▶ Here are some examples...

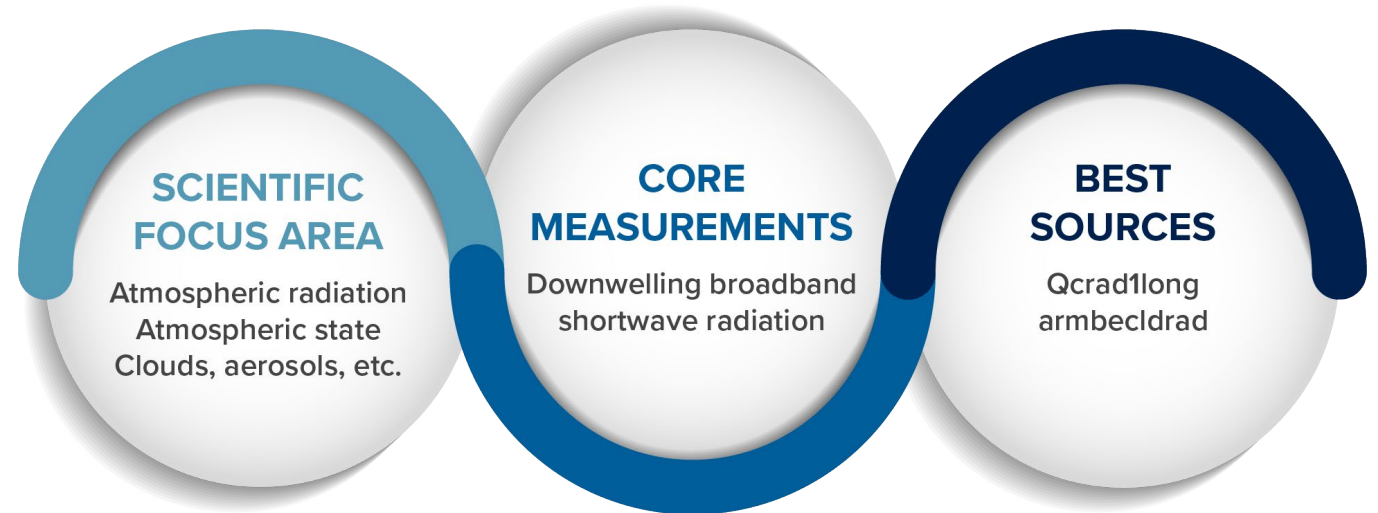


For ARM's Triennial Review, an external panel of experts evaluated ARM using a set of 26 questions posed by the U.S. Department of Energy. ARM's review team prepared a report intended to help the reviewers determine how to answer those questions; the report cover is shown above. This report is available on request.

Example 1: Selecting Quality Data Sources: Harnessing the Power of Rich Metadata

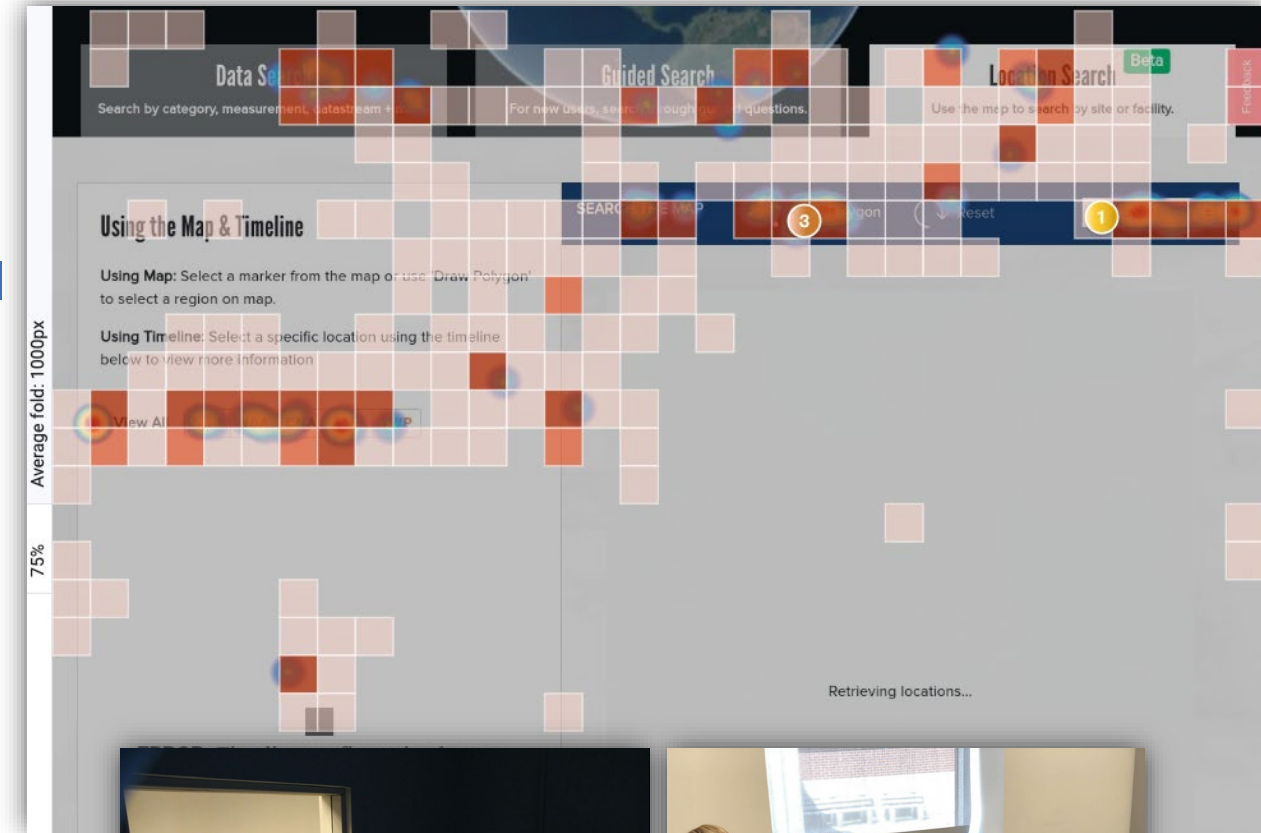
Over 11,000 Data products from 450+ instruments, science products, and model simulations

- ▶ Robust metadata workflow system effectively used for operations, discovery, and data interoperability
- ▶ Recommends best data sources for the core measurements (i.e., Data Epoch)
- ▶ Semi-automated process includes input from subject matter experts



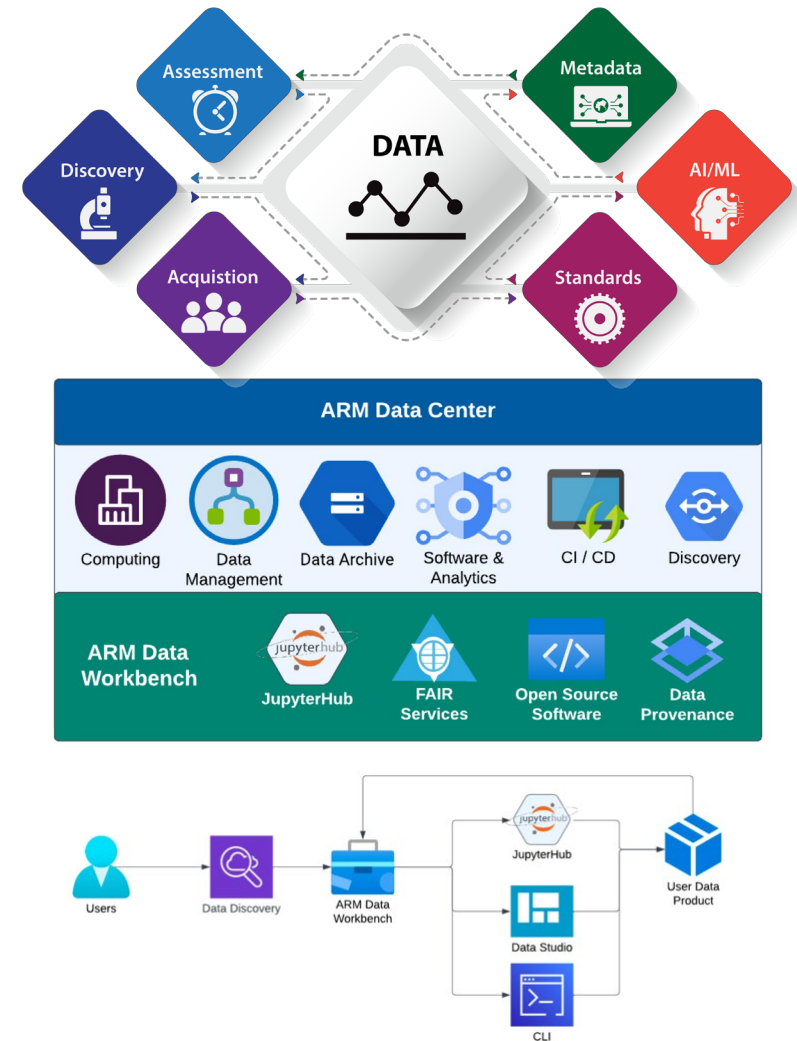
Example 2: Advanced Data Discovery: Leveraging Modern Architecture and Search Capabilities

- User-centric design and improvements using modern software architecture with Continuous Integration and Deployment (CI/CD)
- Intelligent search capabilities based on the actual data, guided search based on user experience
- Recommendations, data tagging based on epochs or golden periods



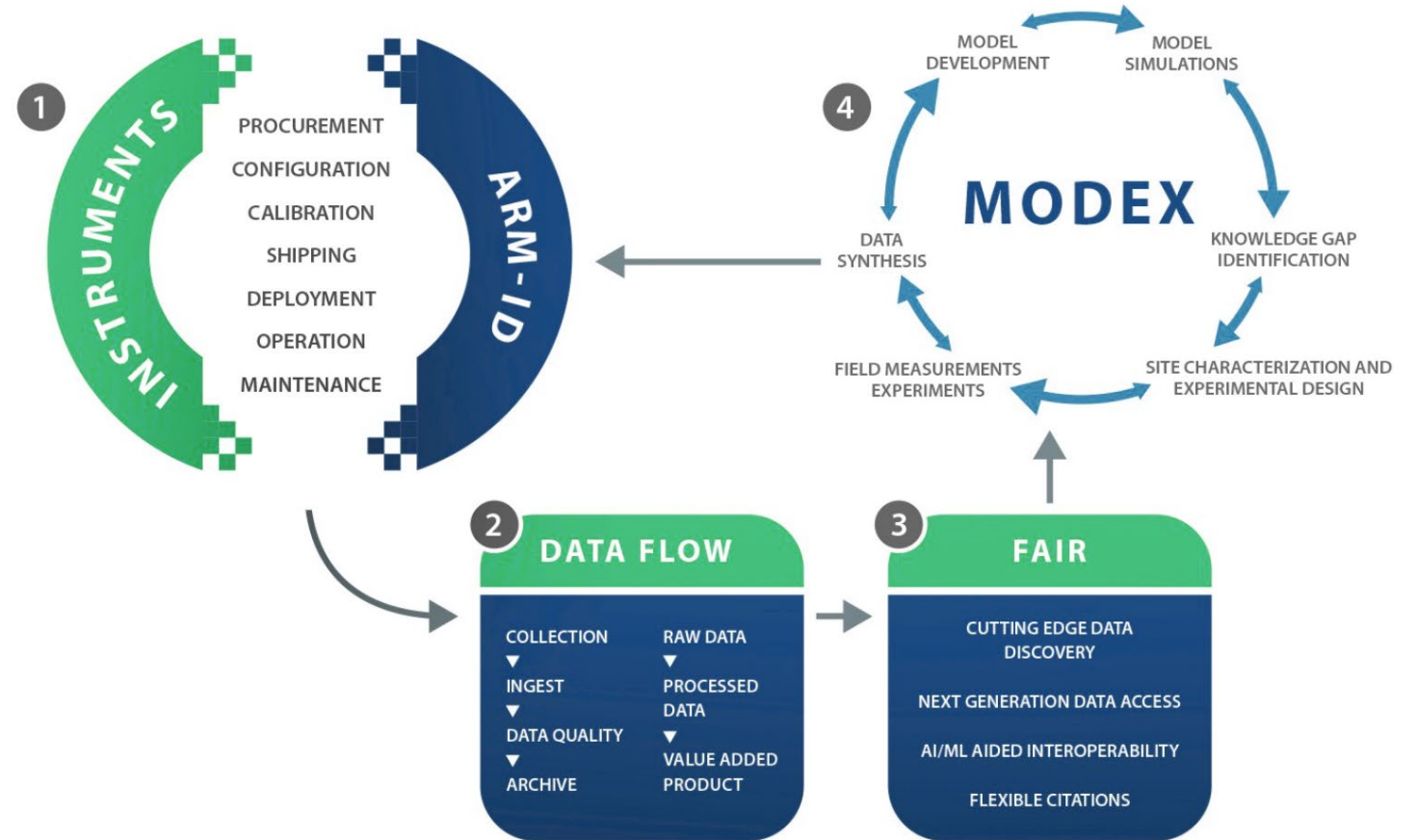
Example 3: Data Workbench: Enabling Data Interoperability

- ▶ Aims to achieve transformative knowledge discovery by providing modular computing, data, and software capabilities
- ▶ Facilitate easier interaction with ARM data and enable interoperability with other data sources
 - Provide a collaborative and dynamic computation environment for data analysis, scientific computing, and machine learning (e.g., JupyterHub)
 - Facilitate data access to external datasets (e.g., weather radar, satellite, model data, etc.)
- ▶ Enable FAIR-based access to ARM data and computing for initiatives such as AI4ESP

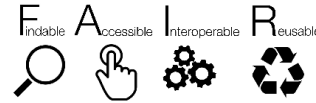


Example 4: Unlocking the Power of Data - The Role of AI in Enhancing Observational Data Centers

- ▶ Enabling interdisciplinary research through modernization of data pipelines from collection to distribution using AI-based approaches
- ▶ Near real-time data analysis and data collection configurations using edge computing
- ▶ Developing and extending community-based standards between data repositories and AI models
- ▶ Data tagging to identify benchmarking/training datasets




Promoting Sustainability through FAIRness Assessment and Community Engagement: Empowering Data Centers for the Future



- ▶ Review of data management capabilities and obtaining certifications
- ▶ Continuous collaboration with broader data networks
- ▶ Active contribution to national and international working groups

Questions?

- ▶ <https://www.arm.gov>
- ▶ "Ask Us"
- ▶ ARM Data Center: adc@arm.gov
- ▶ My contact: palanisamyg@ornl.gov

ATMOSPHERIC RADIATION MEASUREMENT USER FACILITY				
CONNECT WITH ARM	POLICIES	HELP	RESOURCES	WORKING WITH ARM
CREATE ACCOUNT	DATA POLICIES	ASK US	MEDIA	USE ARM FACILITIES
ORGANIZATION	CAMPAIGN GUIDELINES	ASK A UEC MEMBER	OUTREACH	ACKNOWLEDGE ARM
	LINKING POLICIES	DATA QUESTIONS	ACRONYMS	SUBMIT A PROPOSAL
Reviewed September 2021	PRIVACY & SECURITY NOTICE	FAQS	GLOSSARY	FIND EMPLOYMENT
	DIVERSITY, EQUITY, & INCLUSION	ACCOUNT MANAGEMENT		VIEW ARM PRIORITIES
	VULNERABILITY DISCLOSURE PROGRAM			