Data Sharing Perceptions and Establishing Trust in Centralized Repositories

Case Studies: Implementing Networks

Rebecca Koskela

Implementing FAIR Data for People and Machines: Impacts and Implications

September 11, 2019

Outline

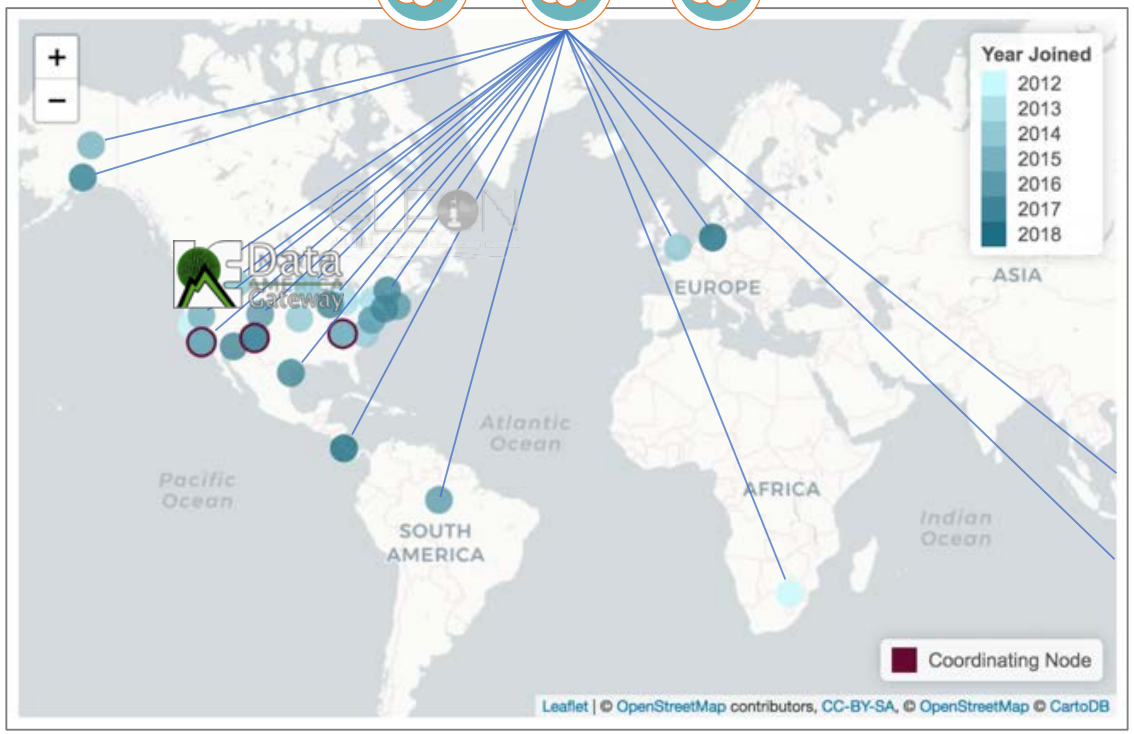
- DataONE Community Surveys
- Enabling FAIR Data Project
- CoreTrustSeal Certification of Repositories Cohort

DataONE: A Federation of Earth Science 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 PAHSTtlAromatic <- rowSums(PAH[,24:71], na.rm=T)  # sum the chemicals a
16 PAHSTPAH_adj <- rowSums(PAH[,c(24:26,28,30:45,47:71)], na.rm=T)  ; PAH[c
17 library(plyr)
18 PAH1 <- arrange(PAH, Sin)  ; head(PAH1)  # arrange by the Sample ID (Sin
19 TotalAromat <- PAH1[PAH1$Sin > 0, -c(3,8,15:22,24:71)]  # remove rows w
```

CN CN CN



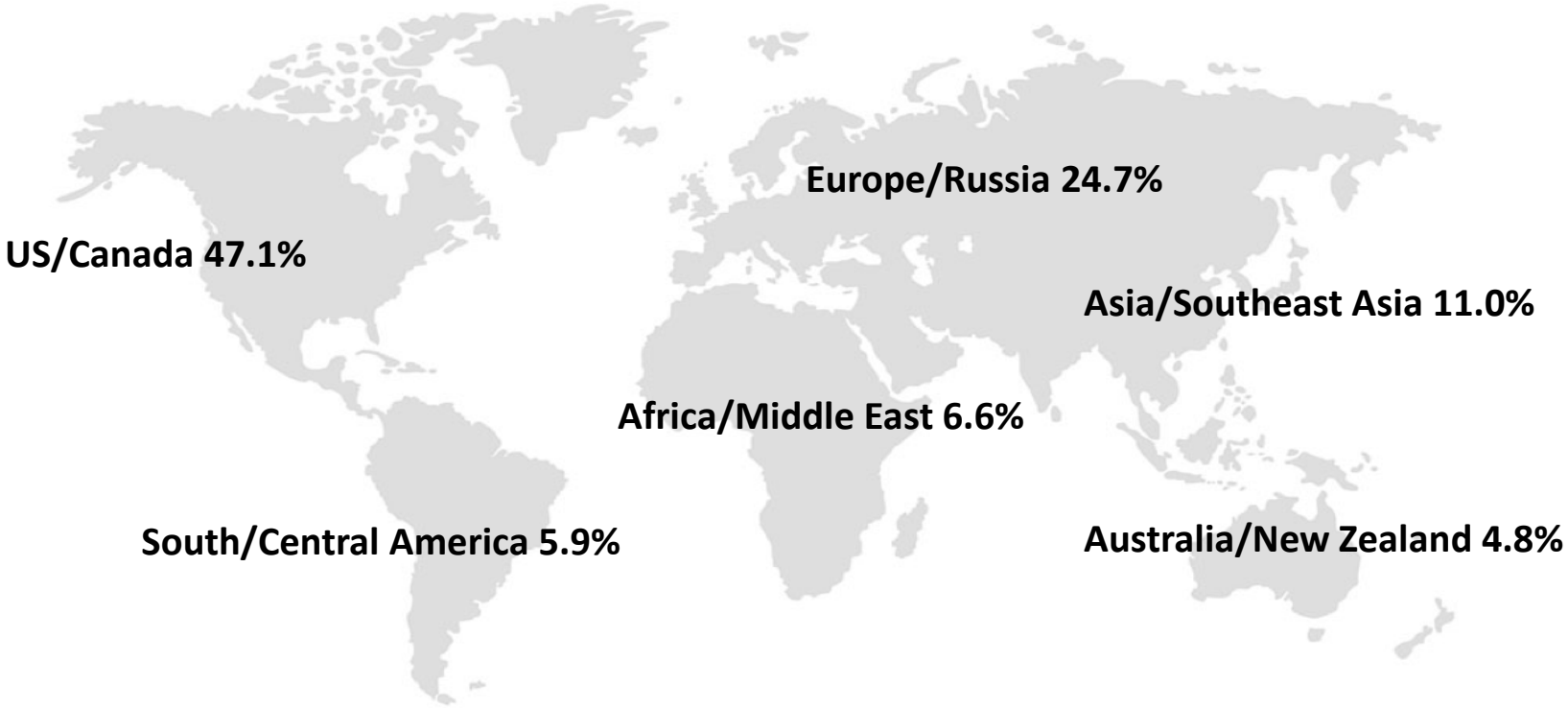
ORNL DAAC
PISCO
UC3 Merritt
MPC Minnesota Population Center
LTER
South African National Parks
EDAC Earth Data Analysis Center
IARC
TheCornellLab
SEAD
TERN
DRYAD
ONEShare
knb
AOOS Alaska Ocean Observing System
KU Biodiversity Institute The University of Kansas
esa
PPBio
ESS-DIVE Deep Insight for Earth Science Data
UIC University Library
NKN Northwest Knowledge Network
neon National Ecological Observatory Network
New Mexico EPSCoR
FEMC
tDAR
USGS science for a changing world
GRIIDC
BCO-DMO
NRDC

Data Sharing and Data Reuse Practices and Perceptions among Scientists

- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS One*, **6**(6), e21101. https://doi.org/10.1371/journal.pone.0021101
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*, **10**(8), e0134826. https://doi.org/10.1371/journal.pone.0134826
- Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., Sandusky, R., Langseth, M., & Lundeen, A. (2015). Research data services in academic libraries: Data intensive roles for the future. *Journal of eScience Librarianship.*, **4**(2). https://doi.org/10.7191/jeslib.2015.1085b
- Tenopir, C., Christian, L., Allard, S., &Borycz, J. (2018). Research data sharing: Practices and attitudes of geophysicists.Earth and Space Science, 5, 891–902.https://doi.org/

Third Scientist Survey

- **2184 responses from 87 countries**



Physical sciences (atmospheric/hydrology/physical/geology)	43.3%
Life sciences (bio/ecology/environmental/marine/ag)	26.3%
Other (other/social science)	19.8%
Computer science and engineering	10.7%

Scientists need help with long-term data storage

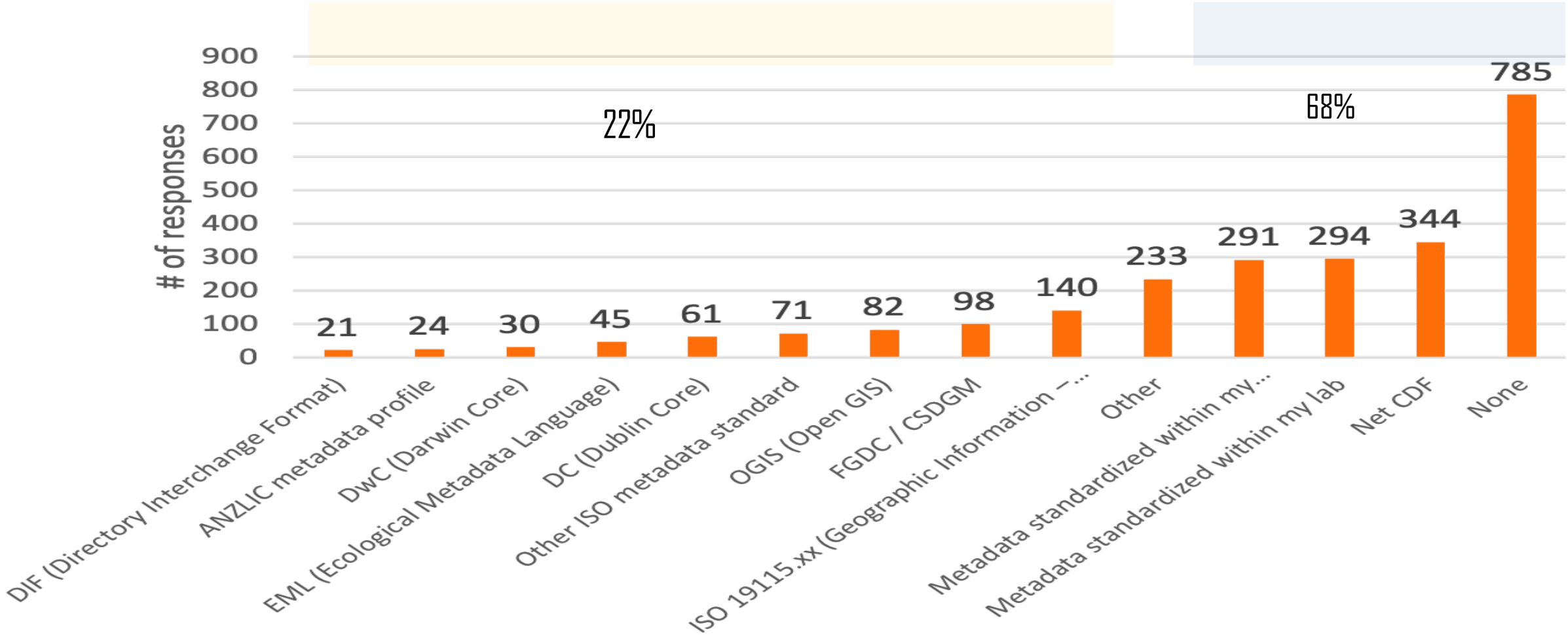
I am satisfied with the process of storing my data...

_ during the life of the project (short term). 74.5%

_ beyond the life of the project (long term). 52.4%

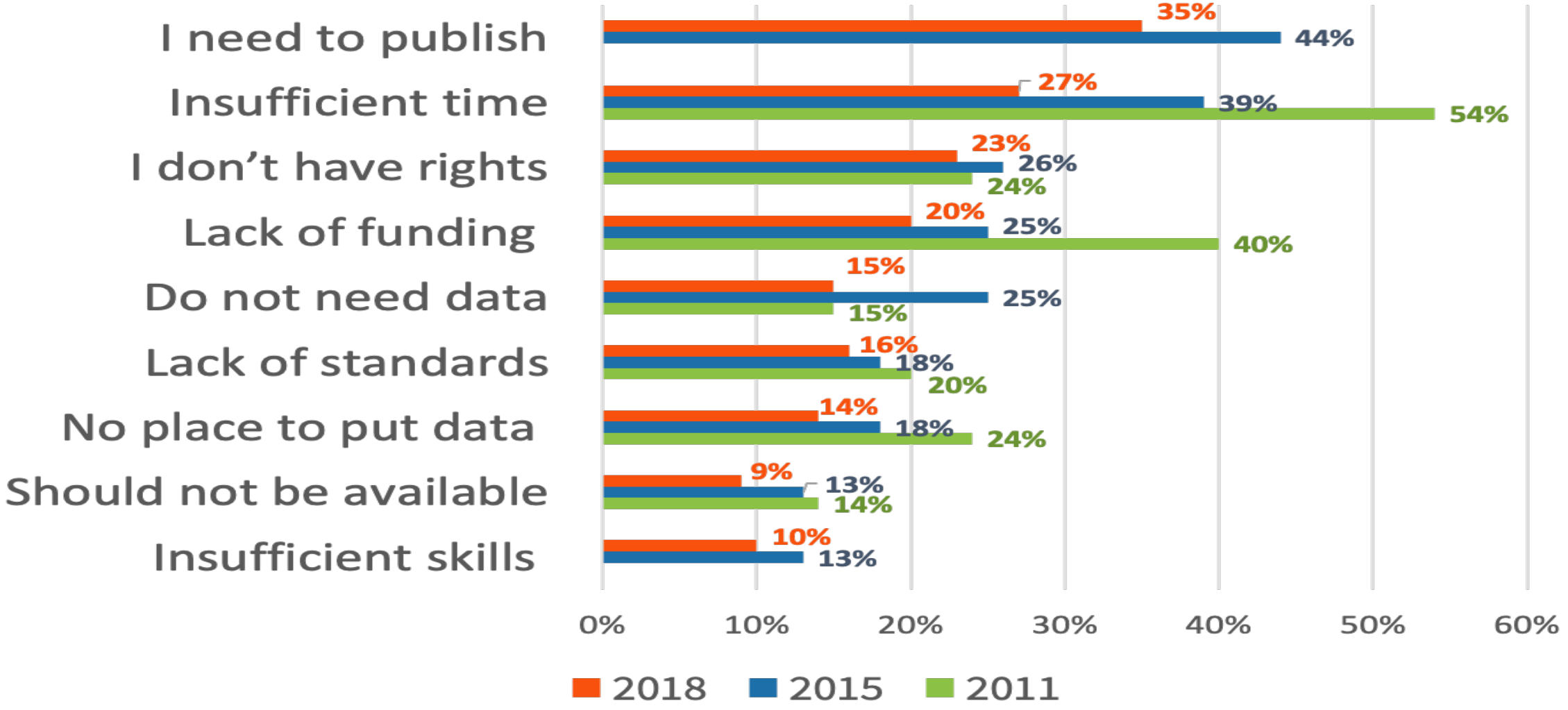
Metadata is still an issue

- Third Scientist Survey



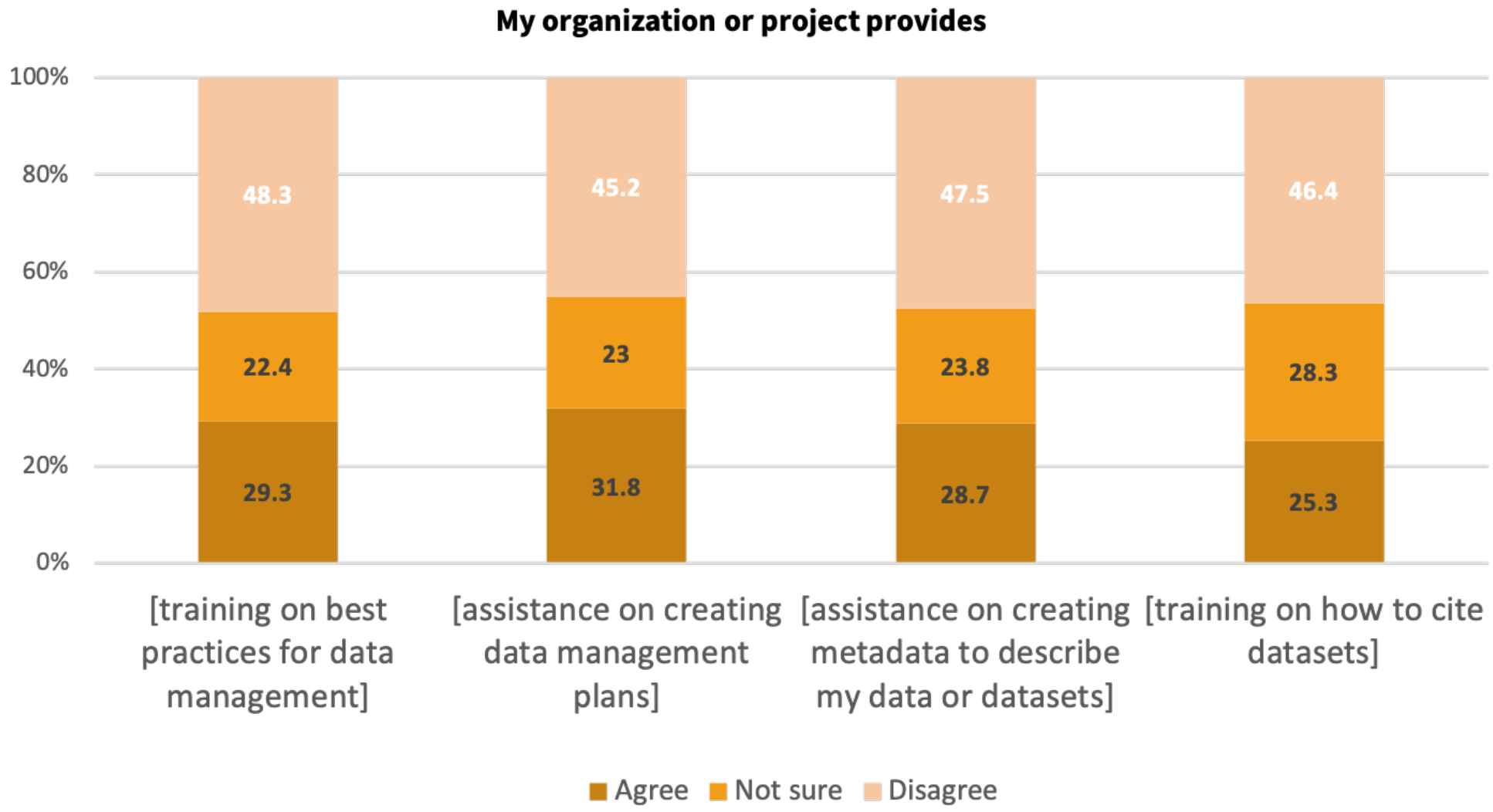
Barriers to data sharing remain

- Comparing Three Scientist Surveys



There isn't much assistance or training

- AGU Survey



Enabling FAIR Data across the Earth and Space Sciences Project

- Community-driven and community-led project
- Objectives:
  - FAIR-compliant data repositories will add value to research data, provide metadata and landing pages for discoverability, and support researchers with documentation guidance, citation support, and peer review.
  - FAIR-compliant Earth and space science publishers will align their policies to establish a similar experience for researchers. Data will be available through citations that resolve to repository landing pages. Data are not placed in the supplement.

Limitations

Technical Limita

Technical barriers to data

- A system does not op objectives and specifi
- Datasets are not com modifications
- Datasets do not conta
- Access to data & serv
- Datasets & services a reason)

Social Limitations: Trust

- Being unable to trust data from other sources is one of the cultural challenges preventing proper data management, preservation, and sharing (along with data ownership and the fear of being discredited or scooped)
- Trust is thus at the very heart of storing and sharing data:
  - Data funders want reassurances that their investment in the production of research data is not wasted, but will remain into the future; whilst data reuse data will give them a higher return on investment
  - Data depositors want to be sure their data are safe and remain accessible, usable, and meaningful over time
  - Data users want to know that data have been preserved properly and are of high quality

CORE TRUST SEAL

WORLD DATA SYSTEM

CoreTrustSeal Requirements

- https://www.coretrustseal.org/why-certification/requirements/

I. Licenses
II. Continuity of Access
III. Confidentiality/Ethics
IV. Organizational Infrastructure
V. Expert guidance
VI. Data integrity and authenticity
VII. Appraisal
IX. Documented storage procedures
X. Preservation plan
XI. Data quality
XII. Workflows
XIII. Data discovery and identification
XIV. Data reuse
XV. Technical infrastructure
XVI. Security

CoreTrustSeal Certification -- Cohort of Data Repositories

- A cohort of data repositories was assembled to apply for CoreTrustSeal certification as part of the American Geophysical Union Coalition Enabling FAIR Data Project.

- CoreTrustSeal certification demonstrates to a repository's users and funders that they have been evaluated by an independent authority and endorsed for their trustworthiness.

Why Certify

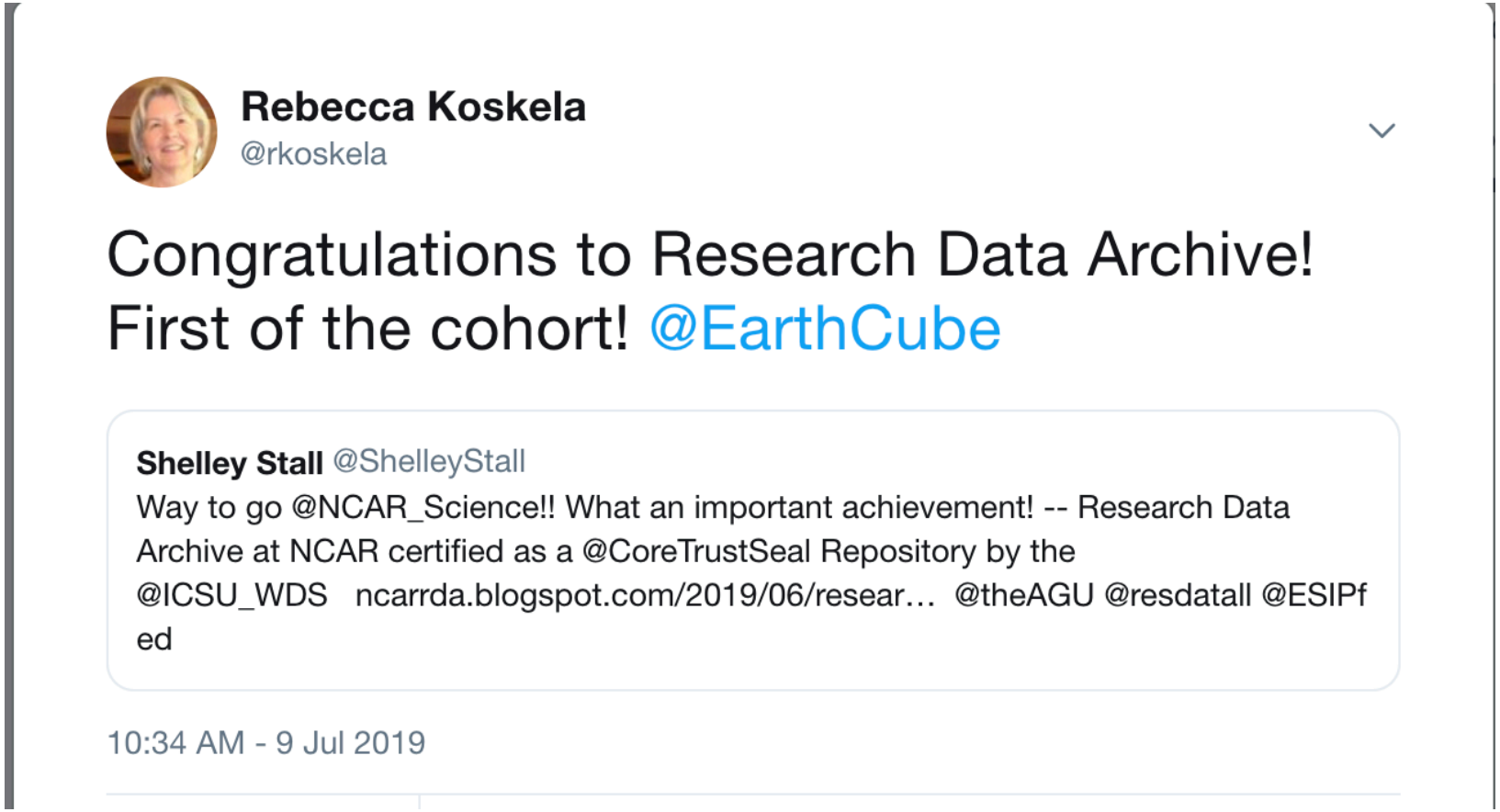
Why repository

- Builds stakeholder con
  users, publishers etc.)
- Bench marking of NGD
  against recognized sta
- Recognition as a truste
  designated community
- Differentiates NGDC fr
  other repositories

Why do repositories invest in certification efforts?

- Builds stakeholder confidence in the repository (funders, publishers, etc.)
- Raises awareness about digital preservation
- Improves communication within the repository
- Improves repository processes
- Ensures transparency
- Differentiates the repository from others

First Cohort Success



Rebecca Koskela
@rkoskela

Congratulations to Research Data Archive! First of the cohort! @EarthCube

> **Shelley Stall** @ShelleyStall
> Way to go @NCAR_Science!! What an important achievement! -- Research Data Archive at NCAR certified as a @CoreTrustSeal Repository by the @ICSU_WDS ncarrda.blogspot.com/2019/06/resear… @theAGU @resdatall @ESIPfed

10:34 AM - 9 Jul 2019

Summary

- Continue to survey the community for changing perceptions and practices
- Provide support and training for research data management
- Provide opportunities for repositories to gain certification, CoreTrustSeal is a good first step