Conversation with your data platform

Nirav Merchant
Dir. Data Science Institute
Co-PI CyVerse
University of Arizona

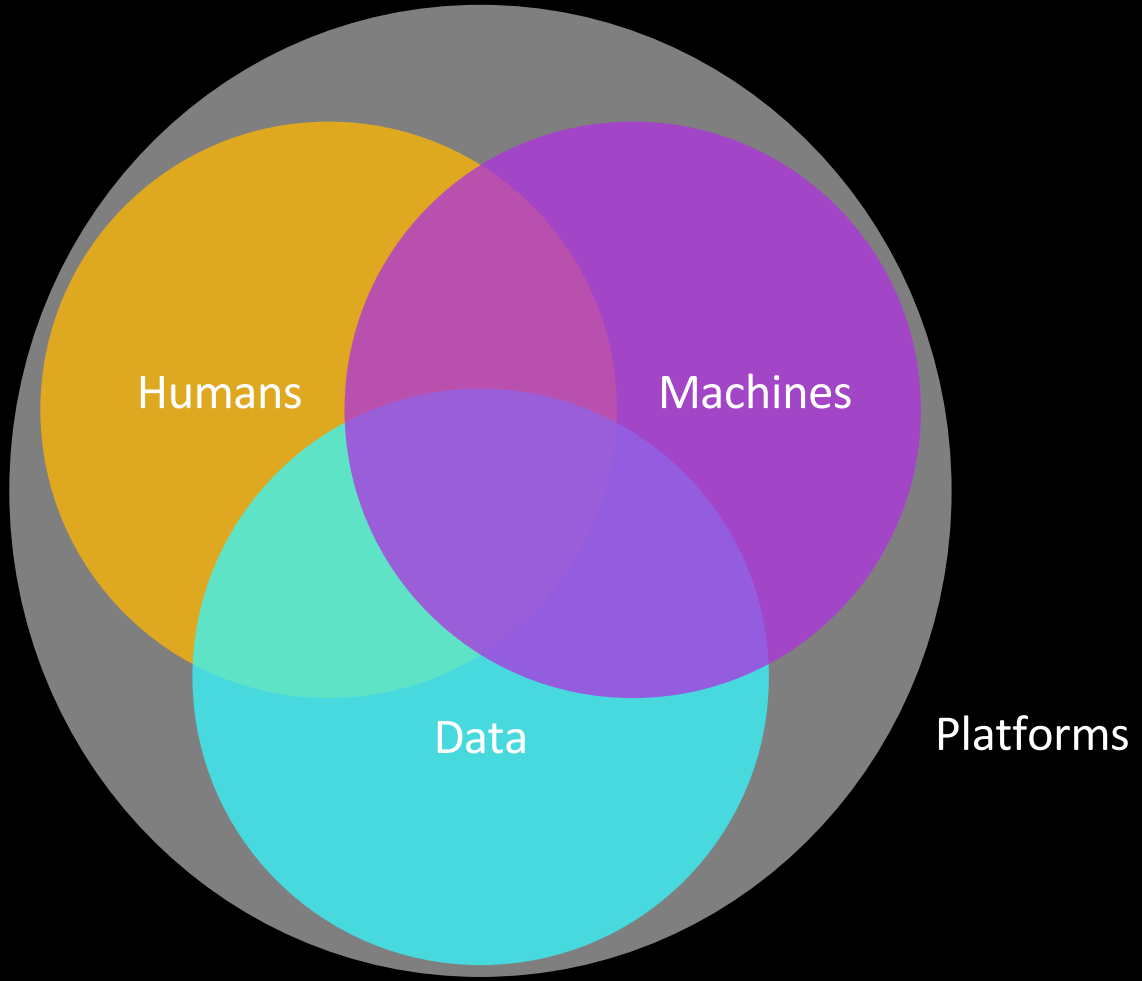
nirav@email.arizona.edu
www.cyverse.org
@cyverseorg

THE UNIVERSITY OF ARIZONA

iRODS UGM 2020

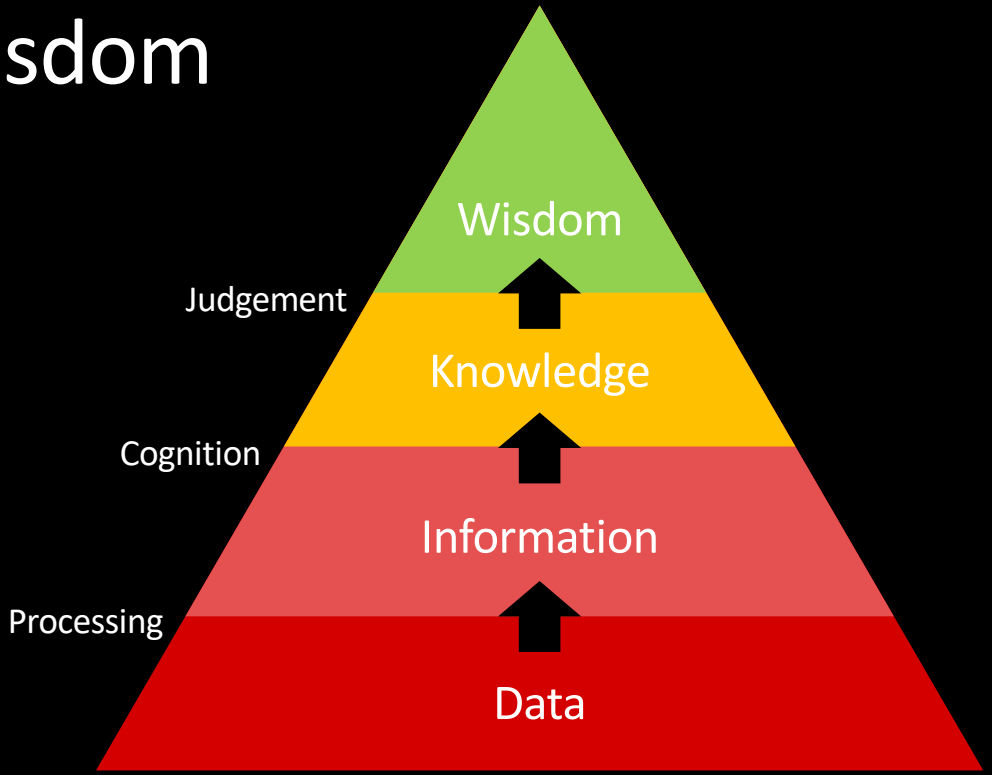
NSF BIO1743442

Data Platforms: Humans, Data and Machines


Humans
Machines
Data
Platforms

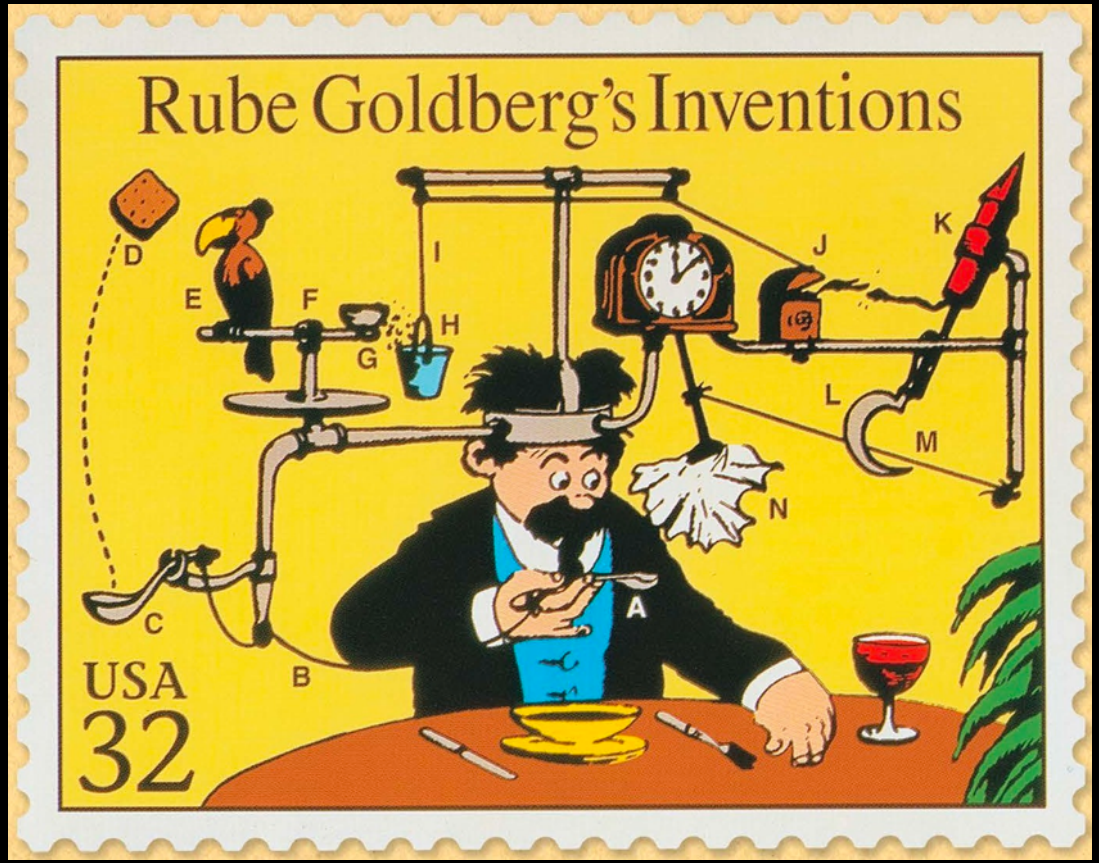

Expectations from your
Data Platform ?

A platform for transforming data to wisdom



DIKW pyramid: US Army Knowledge Managers

The reality: Your data platform



We have lived Rube Goldberg's life (building platforms)

Rube Goldberg works under an early animation camera.
Courtesy of the National Museum of American Jewish History


Need to go beyond
Data to Knowledge

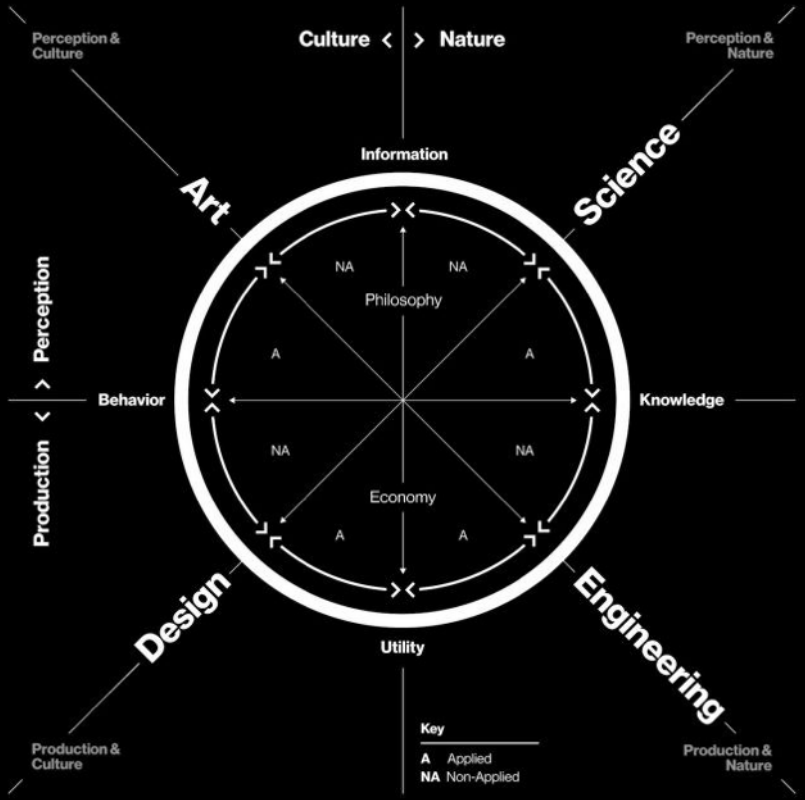
Krebs Cycle of Creativity (KCC)

Science
Converts information into knowledge

Engineering
Converts knowledge into utility

Design
Converts utility into cultural behavior and context

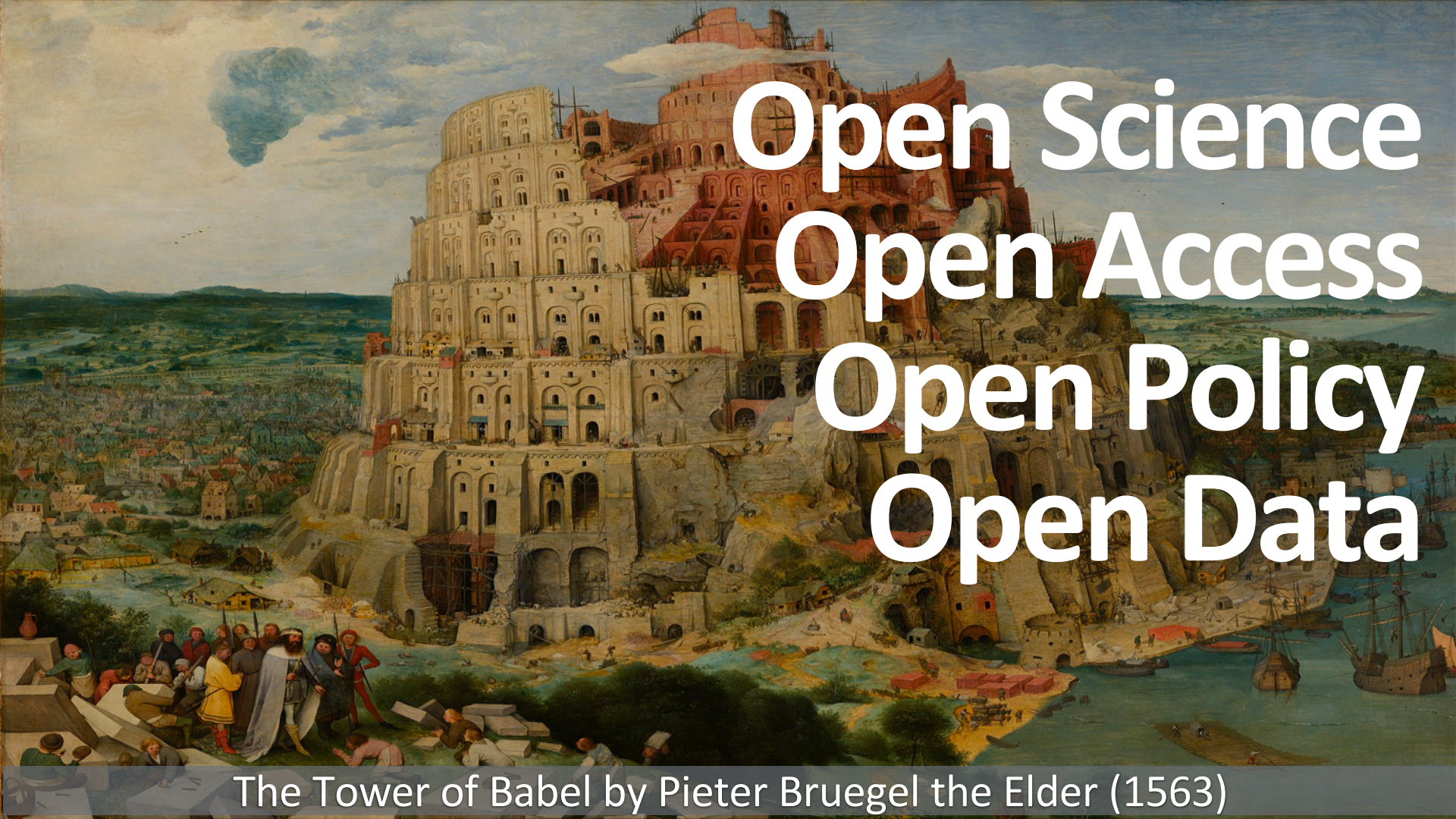
Art
Takes context and questions our perception of the world.



Neri Oxman: Bio-Architecture. Abstract. Netflix. 2019

KCC: From Rube to Atlas

- Data platforms that work for every use case (discipline) exist in only in marketing brochures or on TV (mythical)
- Supporting diverse communities is a common occurrence (requirement)
- While storage cost per TB is going down, managing it is getting expensive and harder in a distributed world



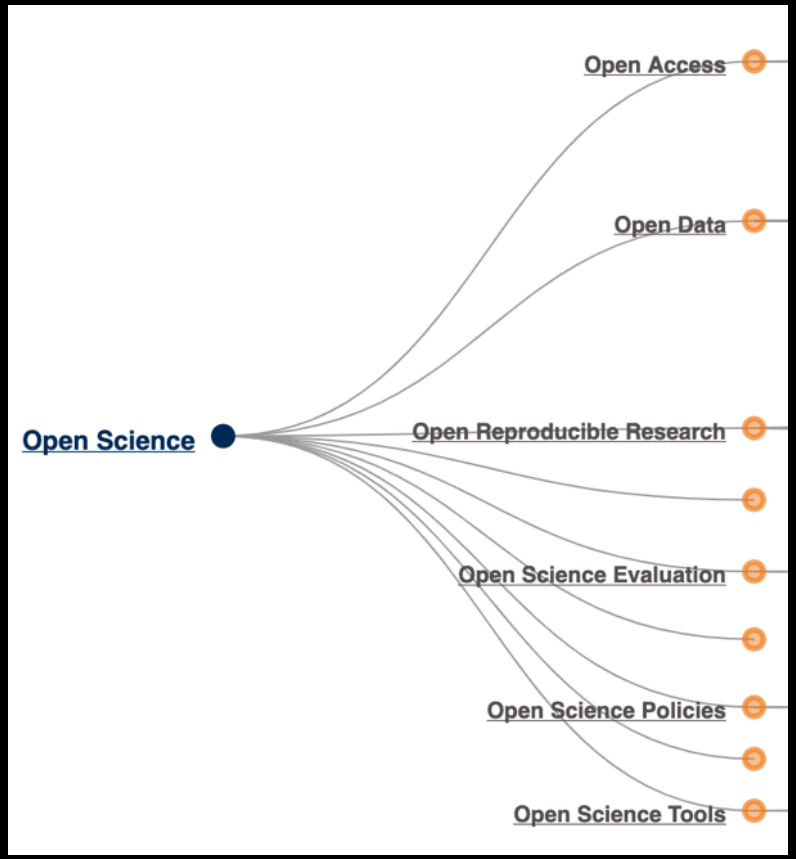
The Tower of Babel by Pieter Bruegel the Elder (1563)

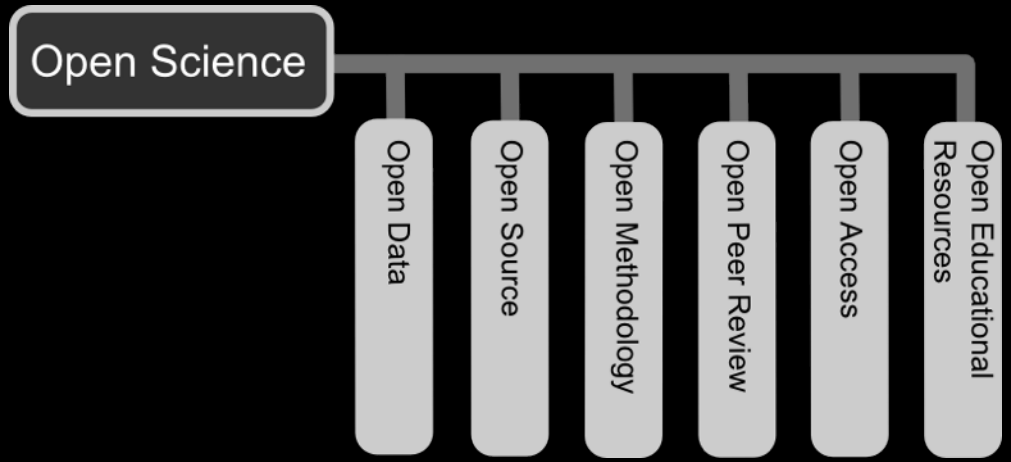
Why open science is critical to combatting COVID-19

Key messages

- In global emergencies like the coronavirus (COVID-19) pandemic, open science policies can remove obstacles to the free flow of research data and ideas, and thus accelerate the pace of research critical to combating the disease.
- While global sharing and collaboration of research data has reached unprecedented levels, challenges remain. Trust in at least some of the data is relatively low, and outstanding issues include the lack of specific standards, co-ordination and interoperability, as well as data quality and interpretation.
- To strengthen the contribution of open science to the COVID-19 response, policy makers need to ensure adequate data governance models, interoperable standards, sustainable data sharing agreements involving public sector, private sector and civil society, incentives for researchers, sustainable infrastructures, human and institutional capabilities and mechanisms for access to data across borders.

Open Science: Can your data platform do that ?





Managing Data Platforms: Rube to Atlas to ….

1

2,806 VOTES

↑↓

Zeus Chained Prometheus To

Zeus resented Pro
being willful and d
Prometheus to a r
wasn't enough. Ev
abdomen and mu

a Commons/Public Domain

His liver would then regenerated during the day
gain, every night. Eventually, he was freed
by Chiron, a centaur who gave his life for Prome
agle.

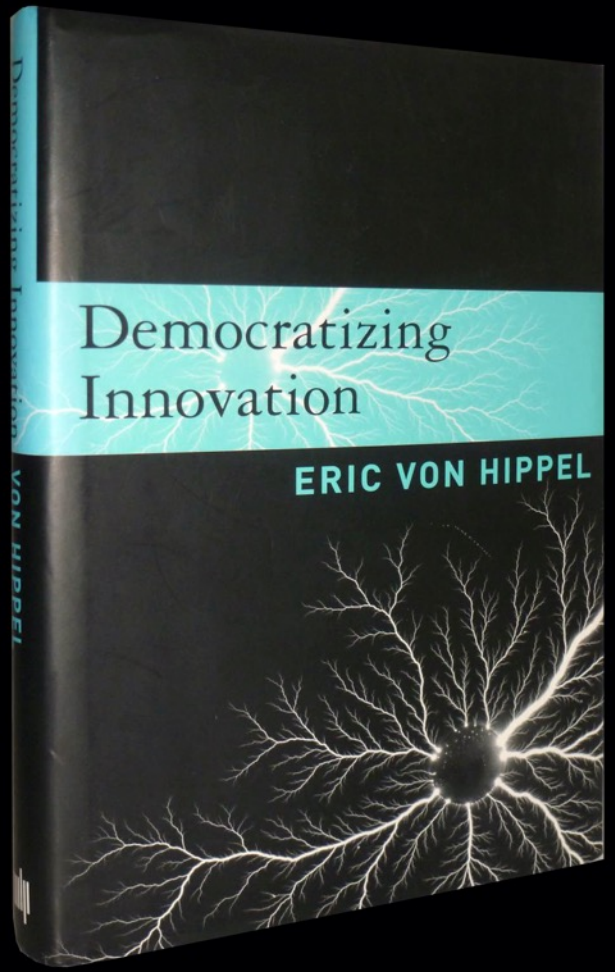
iRODS


Innovation and Creativity
Freedom of choice

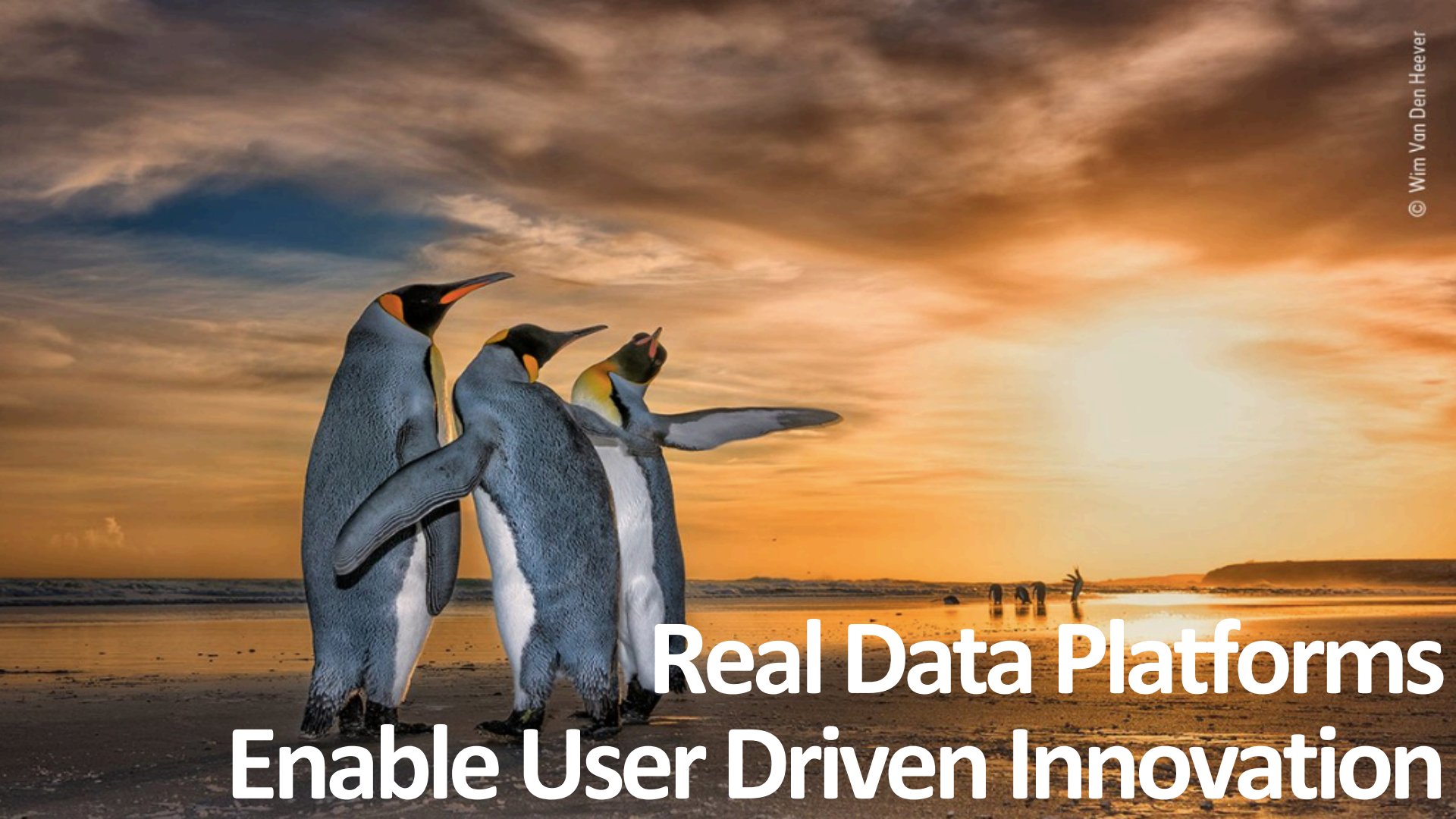
Democratizing Innovation

Innovating users often freely share their innovations with others, creating *user-innovation* communities and a rich intellectual commons.

Data Platforms are central to democratizing innovation



Von Hippel, Eric. Democratizing innovation. MIT press, 2005



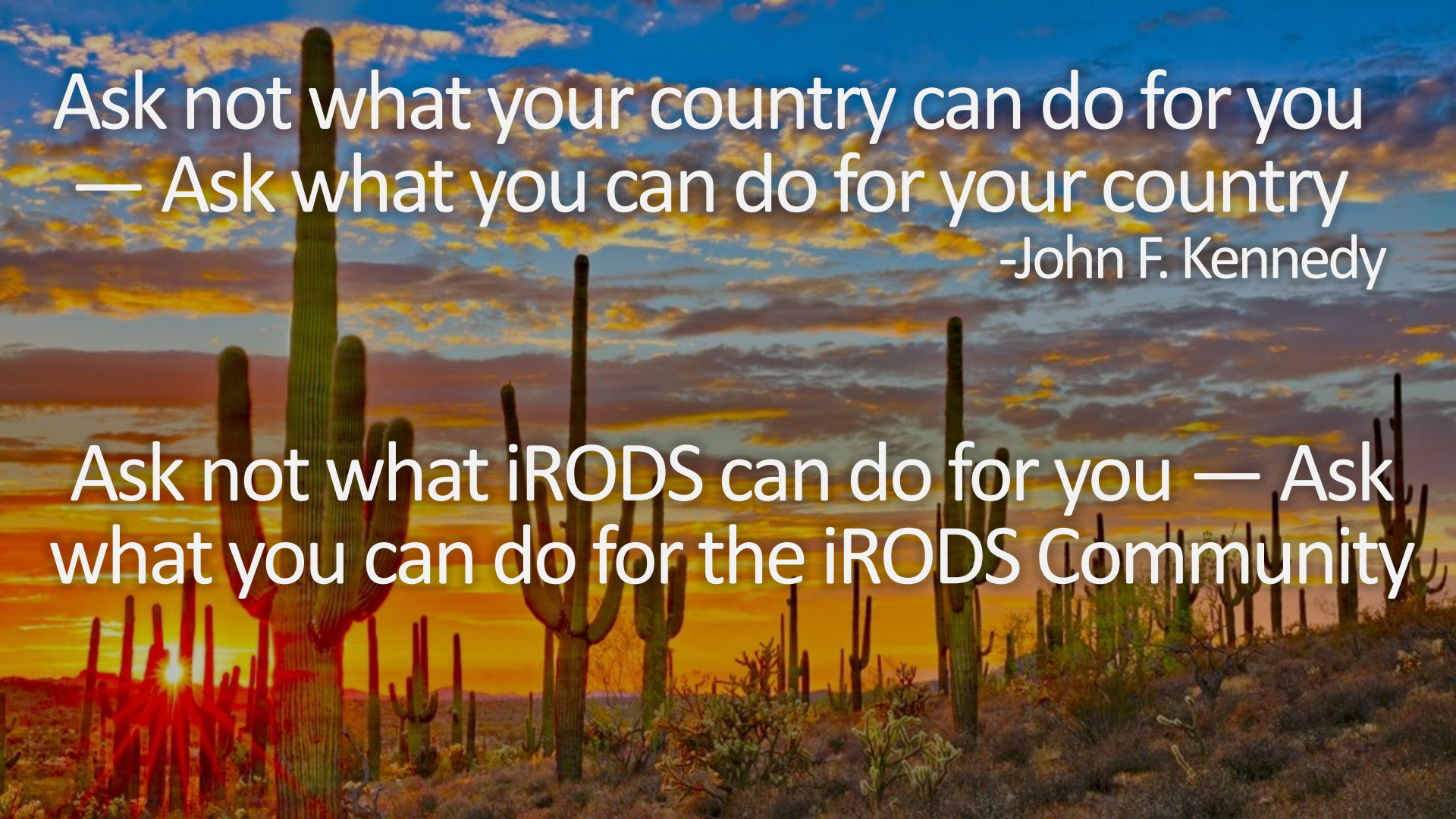
Real Data Platforms Enable User Driven Innovation

Data Platforms: Part of an Ecosystem

- No single provider of infrastructure, but a federation
- Distributed Data Grids, your data is everywhere
- Container Orchestration, your analysis come to your data
- Distributed Computing, your computation is everywhere
- Searching and indexing, your data is everywhere
- Integrating with all of the above is expected
- API based extensibility and automation, first class citizens

Data Platforms: New Generation of Apps

- Application stacks are becoming complex (Models **ML/AI**) that are event based, beyond HPC/batch workload
- Typically include web servers, opinionated frameworks (JavaScript etc.), databases and message buses
- Tools and platforms (R, Shiny, Jupyter etc.) being constantly extended by community, needing access to data


Ask not what your country can do for you — Ask what you can do for your country
-John F. Kennedy
Ask not what iRODS can do for you — Ask what you can do for the iRODS Community

iRODS: A Community Data Platform

- Given us a vendor neutral solution, we need to build an ecosystem of tools and solutions around it
- Allowed us to support large project with ease, we need to support long tail of science, making it easier to install client
- Allows cloud storage integration, we need to make it cloud native and first class citizen fluent in cloud access patterns
- Given training material and documentation, we need to create learning material and train our colleagues in its use (especially institutional data repositories, when budgets are dwindling)


If you want to go fast, go alone.
If you want to go far, go together.
African Proverb