

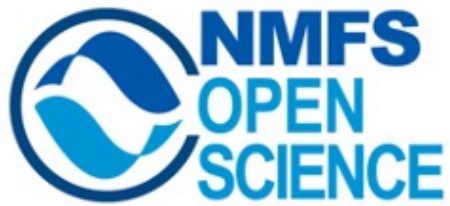
# NOAA Fisheries Open Science and the 2023 Year of Open Science

Eli Holmes, Ph.D

Northwest Fisheries Science Center

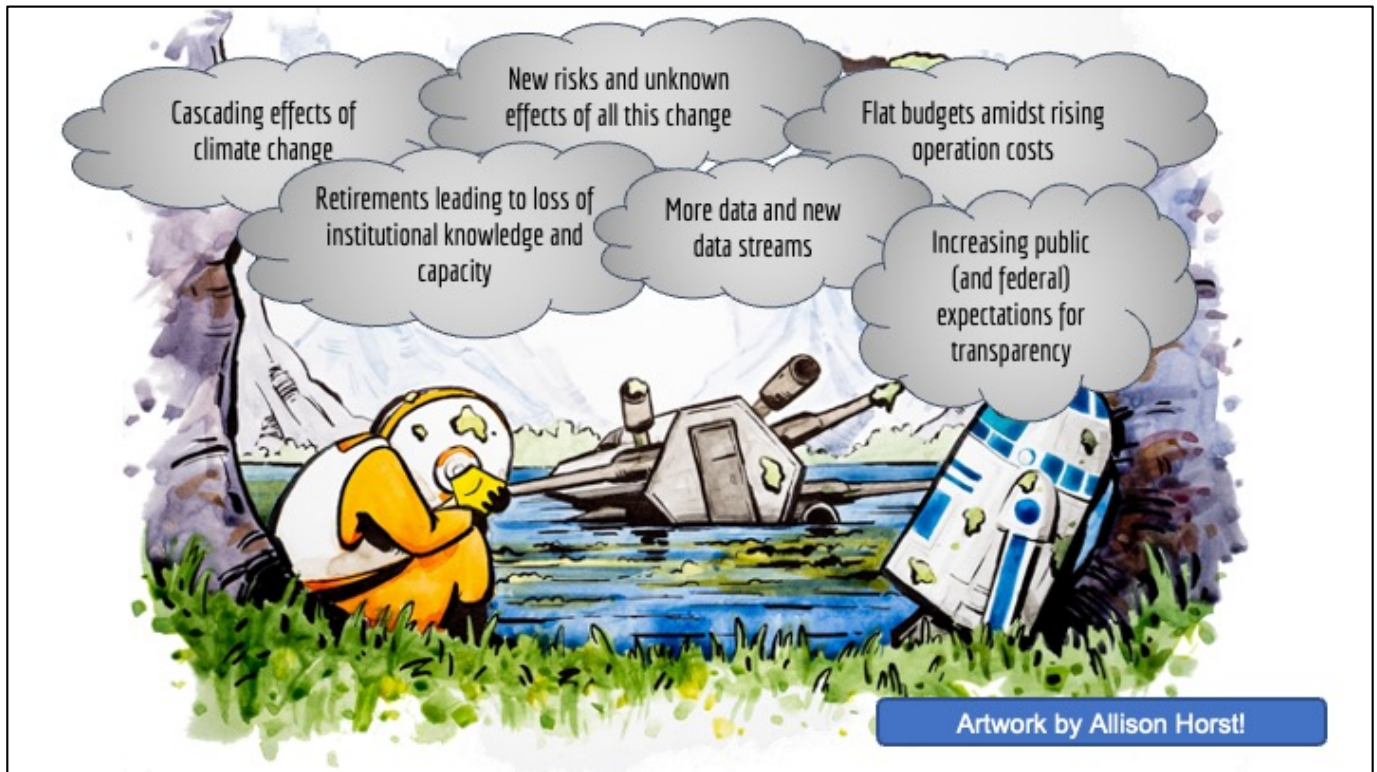
NMFS Openscapes, Co-PI

NMFS Open Science, Lead

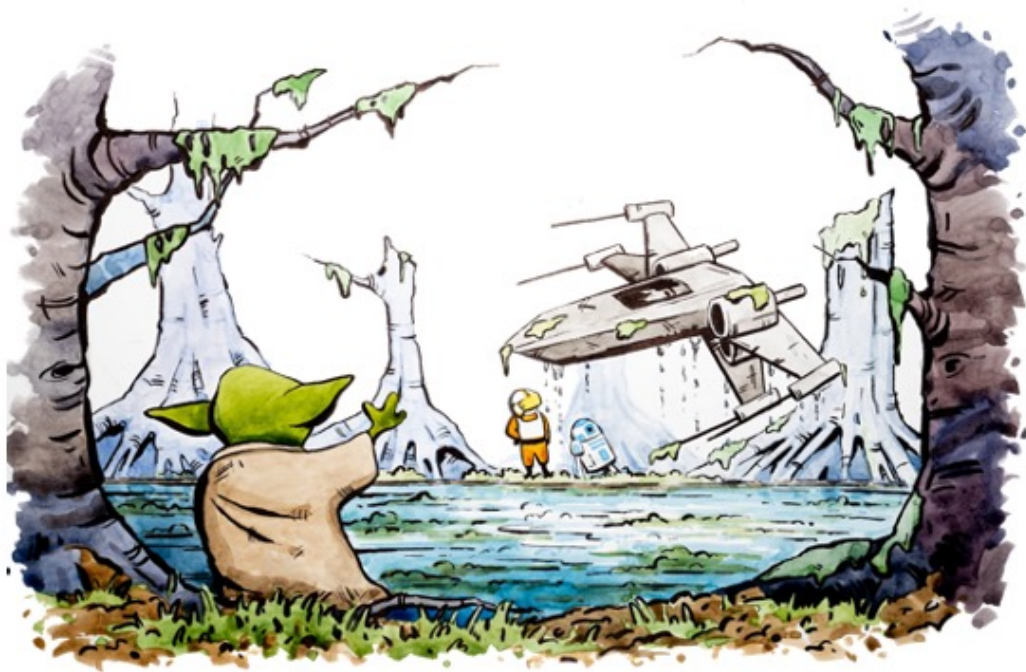


<https://nmfs-opensci.github.io/>





Art: <https://allisonhorst.com/>



Despite the title of my talk, my message is not that adopting Open Science will magically solve all these problems.



Rather the message is that Open Science is a way that teams and NOAA Fisheries can learn to collaborate together better and allow teams across the agency to work together to make real progress despite big challenges ahead.

**NOAA  
FISHERIES**



2020-2022  
Openscapes  
program

NMFS Open Science &  
Year of Open Science 2023



[nfms-openscapes.github.io](https://nfms-openscapes.github.io)  
[openscapes.org](https://openscapes.org)



# The White House announces 2023: A Year of Open Science

A multi-agency initiative across the US Federal Government to spark change and inspire open science engagement through events and activities that will advance adoption of open science.

- ✦ NASA
- ✦ National Oceanic and Atmospheric Administration
- ✦ National Science Foundation
- ✦ Department of Energy
- ✦ General Services Administration
- ✦ National Endowment for the Humanities
- ✦ National Institutes of Health
- ✦ National Institute of Standards and Technology
- ✦ US Department of Agriculture
- ✦ US Geological Survey



Gentemann, Chelle L., Shrestha, Sudhir, Ivey, Yvonne, & Hall, Cynthia. (2023, February 9). TOPS February 9 Community Forum. Zenodo.  
<https://doi.org/10.5281/zenodo.7626005>

<https://science.nasa.gov/open-science/transform-to-open-science>

<https://science.nasa.gov/open-science-overview/TOPS-community-panel>

<https://nasa.github.io/Transform-to-Open-Science/>

# What is Open Science?



White House [Office of Science and Technology Policy](#) (OSTP) official definition in 2023 Year of Open Science



- **Why NASA and federal agencies are declaring this the Year of Open Science. Here's how NASA is incentivizing open science, and how you can too.** [Chelle Gentemann](#) *Nature* **613**, 217 (2023) doi: <https://doi.org/10.1038/d41586-023-00019-y>
- <https://nsf.widencollective.com/portals/sd9nlyvd/YearofOpenScienceToolkit>
- <https://nasa.github.io/Transform-to-Open-Science-Book/index.html>

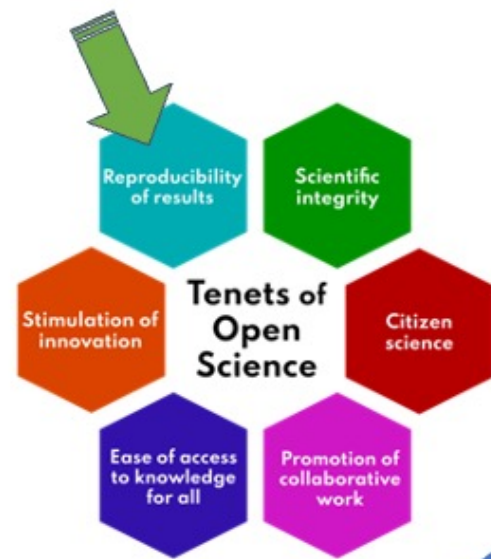
# The Reproducibility Crisis in Science

*Scientific fields have been rocked by the “reproducibility crisis” that has been building for the last 10 year or so, although really came to fore around 2015.*

**Journals begin requiring authors share the raw data and code**

*Recently scientific studies have shown that significant (over half) of studies cannot be replicated – even with the raw data and written methods.*

**Journals are moving toward requiring that authors share the “data to paper pipeline”**



Do a lit search of “Reproducibility Crisis” post-2020 to find reams of publications on this. These are a few that are particularly relevant to my talk.

Treves, A. (2022), “Best available science” and the reproducibility crisis. *Front Ecol Environ*, 20: 495-495. <https://doi.org/10.1002/fee.2568> Especially read the Supplemental information: <https://esajournals.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Ffee.2568&file=fee2568-sup-0001-Supinfo.pdf>

Discussing the controversy around the 2018 EPA Proposed rule. “Strengthening Transparency in Regulatory Science” Daniel J. Hicks (2023) Open science, the replication crisis, and environmental public health, *Accountability in Research*, 30:1, 34-62, DOI: [10.1080/08989621.2021.1962713](https://doi.org/10.1080/08989621.2021.1962713)

2021 EPA Transparency Final Rule. <https://www.federalregister.gov/documents/2021/01/06/2020-29179/strengthening-transparency-in-pivotal-science-underlying-significant-regulatory-actions-and>

Stagge JH, Rosenberg DE, Abdallah AM, Akbar H, Attallah NA, James R (February 2019). [“Assessing data availability and research reproducibility in hydrology and water resources”](#). *Scientific Data*. 6: 190030. [Bibcode:2019NatSD...690030S](#). [doi:10.1038/sdata.2019.30](#). “results might be



reproduced for only 0.6% to 6.8% of all 1,989 articles” largely due to missing data, methods or code.

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11), 2584-2589. Cannot replicate even with the raw data.  
<https://doi.org/10.1073/pnas.1708290115>

Data



- Analyses, plots, tables with no documentation (just the final product)
- Manual undocumented manipulations
- Many data file in different formats
- Scripts of various analyses
- Emails, emails, emails
- Lots of Google docs
- Files on individual folders
- Data of unknown provenance



Unreproducible product:

- Paper
- Decision
- Report

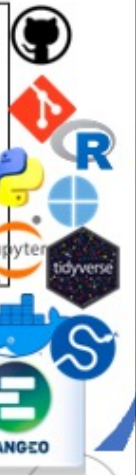


Decisions that impact protected species, human communities, fishing, land use

## How does one create a “reproducible scientific pipeline”?

- **Data:** Data management and documentation
- **Data wrangling:** Eliminating manual manipulation of data
- **Analysis:** Adopting a documented pipeline rather than a patchwork of poorly documented analyses

- **Version-control:** all changes and decision documented
- **Text and code integrated**
- Include a “repository” with a “make” file that reproduces the final product
- A “devcontainer” of the environment
- New skills, new tools, new ways of working



[Fisheries Information Management Modernization Workshop 2020, Tech Memo](#) September 17-19, 2019, NMFS Office of Science and Technology (OST)

2i2c



[Fisheries Information Management Modernization Workshop 2020, Tech Memo](#) September 17-19, 2019, Review and evaluate practical and tangible actions to modernize the data and information system of NMFS

The screenshot shows a video conference interface. On the left, a presentation slide titled "An Openscapes Future for Stock Assessment Reports at the AFSC's Marine Mammal Laboratory" is displayed. The slide features a landscape illustration, a circular diagram of the Openscapes framework, and the NOAA Fisheries logo. The text on the slide lists the presenters: Amelia Brower, Brian Fadely, Josh London, Tony Orr, Erin Richmond, Rod Towell, and Nancy Young. On the right, five video feeds of participants are visible: Joshua London, Amelia Brower, Tony Orr, Rod Towell, and a participant labeled "You".

**Slide Content:**

### An Openscapes Future for Stock Assessment Reports at the AFSC's Marine Mammal Laboratory

Amelia Brower, Brian Fadely, Josh London, Tony Orr, Erin Richmond, Rod Towell, and Nancy Young

**Diagram Labels:** Data, Science, Policy, Management, Assessment, Reporting, Communication, Engagement, Education, Outreach, Training, Capacity Building, Leadership, Governance, Institutionalization, Sustainability, Resilience, Adaptability, Innovation, Collaboration, Partnership, Network, Community, Ecosystem, Landscape, Seascape, Waterscape, Airspace, Cyberspace, Information Space, Knowledge Space, Opportunity Space, Potential Space, Future Space.

**Logos:** openscapes, NOAA FISHERIES

**Participants:** Joshua London - NOAA F..., Amelia Brower - NOAA F..., Tony Orr - NOAA Federal, Rod Towell - NOAA Federal, You

- View the video (NOAA internal)  
<https://drive.google.com/file/d/1dxe3qpuPLfaqbPkSkS36ca4nbq0w20EN/view?usp=sharing>

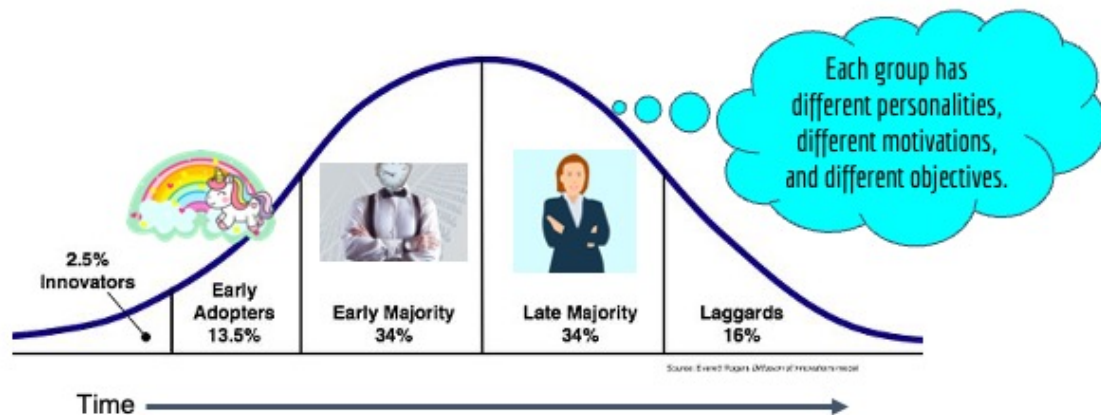


Fine for one team, but how to we  
spread new ways of working  
throughout an organization??



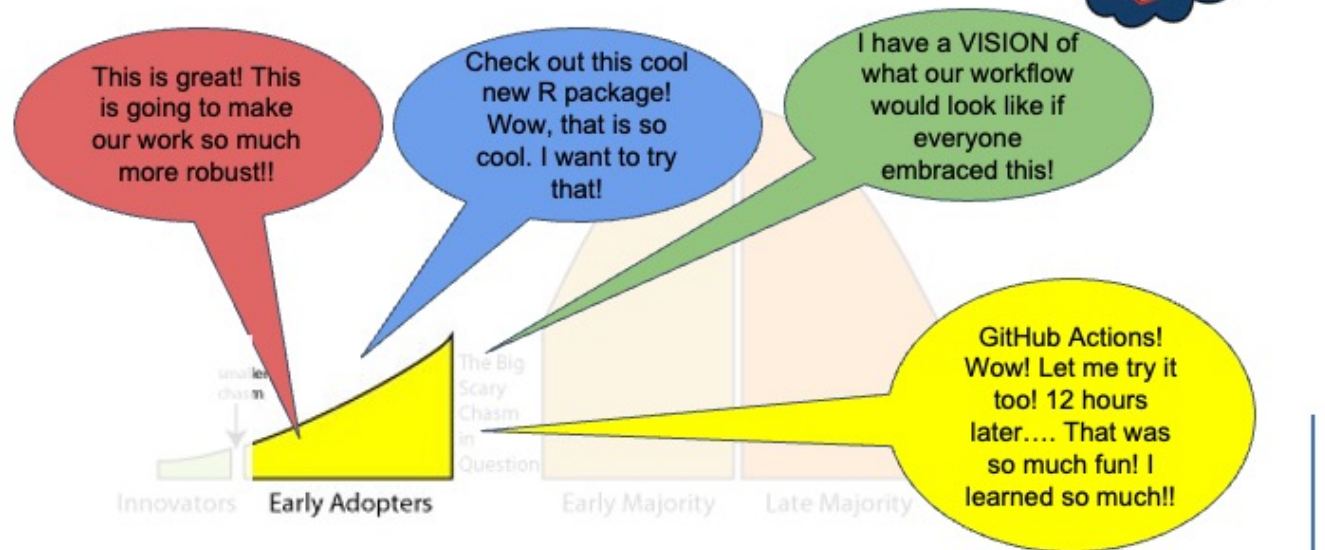
## EM Rogers (1962) “Diffusion of Innovation” theory

Predictable progression of stages as idea diffuses through a population

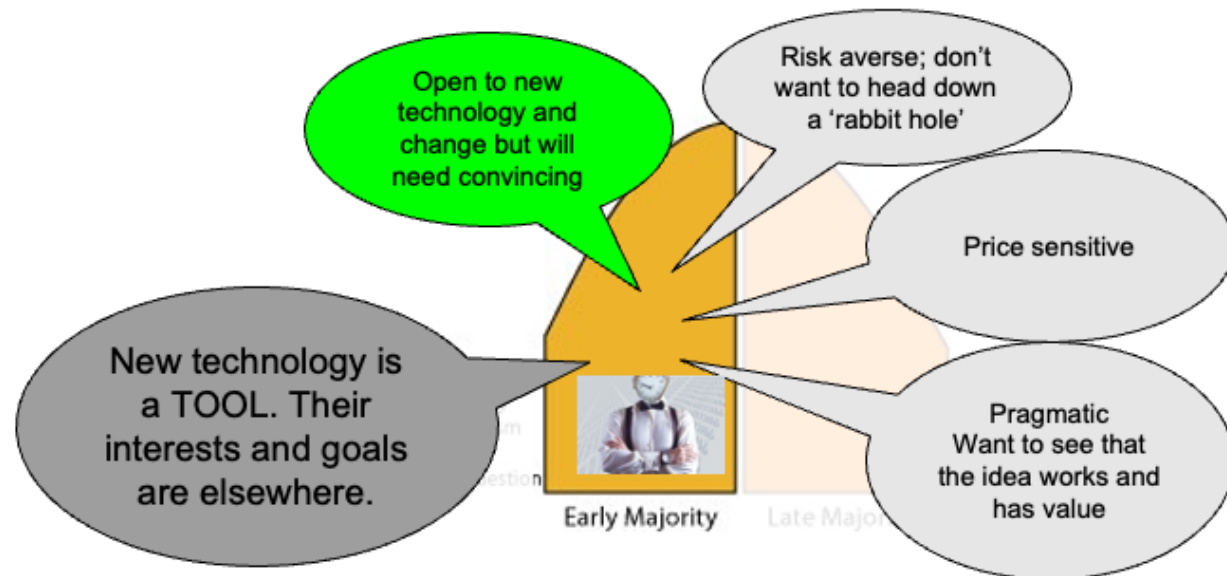


Rogers, E. M. (2003). *Diffusion of innovations*. New York, NY [u.a.]: Free Press. ISBN: 0-7432-2209-1, 978-0-7432-2209-9

## Open Science Early Adopters

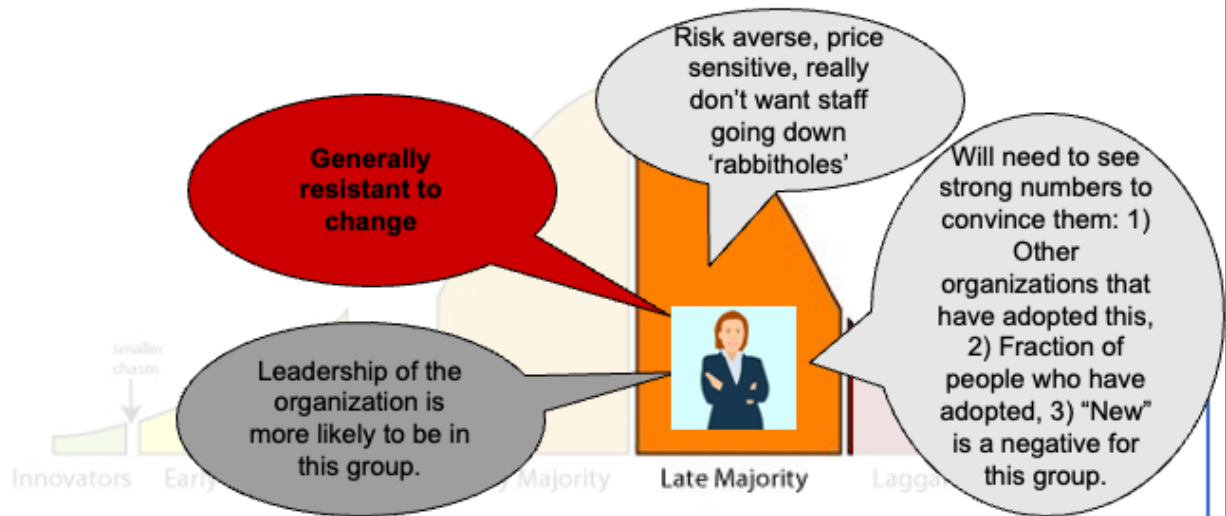


## Early Majority: Open to innovation but risk adverse



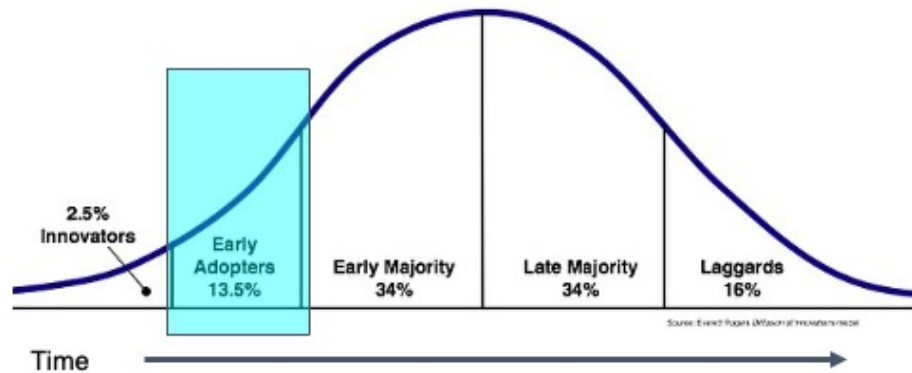


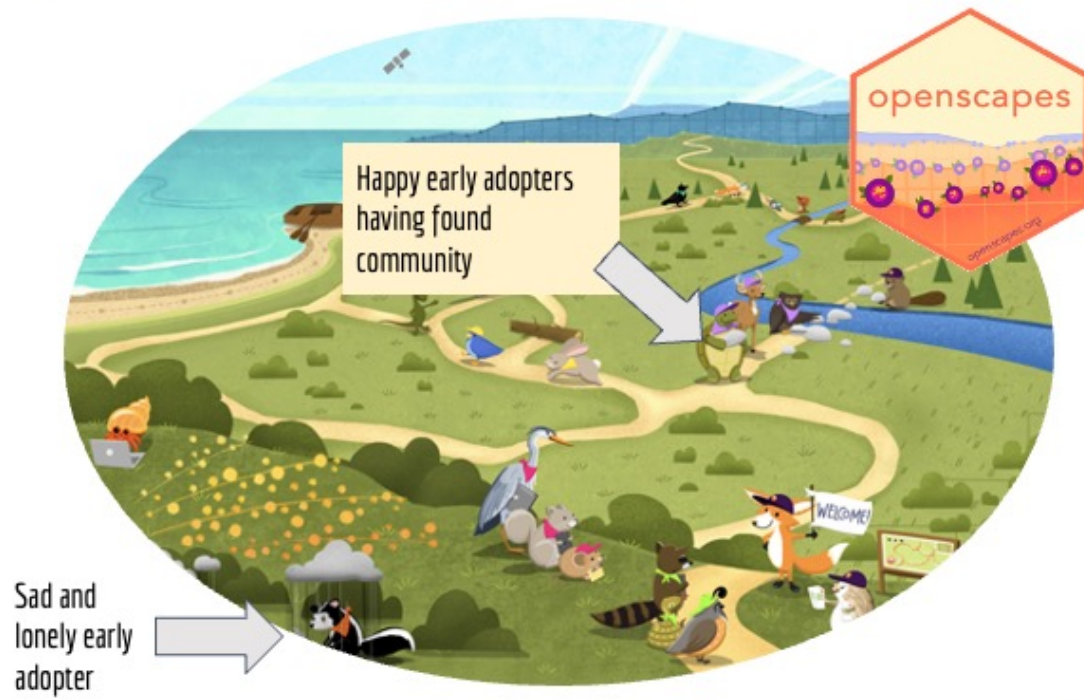
## Late Majority often includes organizational leadership



## The Early Adopters are critical to diffusion of innovation

1. Early Adopters develop the innovation into something of value
2. Their **energy and effort** is what drives the initial diffusion process, but that is a hard and slow process.





# NMFS Openscapes training in Open Science



At NMFS, a grassroots effort due to desire from staff for training in Open Science

9 NMFS Champions Cohorts (40 staff ea)

- 2020: Winter NEFSC
- 2021: Spring NWFSC
- 2021: Fall NWFSC, AFSC, SEFSC, NEFSC
- 2022: Winter AFSC
- 2022: Summer SEFSC/SERO
- 2022 Fall 4 cohorts 6 science ctrs, WCRO

<https://nmfs-openscapes.github.io/>





# What is Openscapes?

## Not your traditional training/workshop

- Cohort-based remote sessions for teams: introduce concepts and workflows; facilitate teams to talk about problems then go and solve them, with accountability and support.
- It's about getting stuff done. It's about identifying and making progress on barriers
- "A process to help you build better lanes of communication" -Laura Waters, SE Regional Office

## Sustainability built-in

- Strengthening a teaching & learning culture within teams & orgs. Not just for scientists; admin, IT staff, etc, welcomed. Equitable.

**\*No coding or software skills required\***

Openscapes works with many environmental  
orgs



<https://openscapes.org/>



Eli presents

UNESCO, NASA quotes. Eli + Julie just presented to UN. You can google this!

What is Openscapes? This is not your traditional training or workshop.

These are cohort-based remote sessions for teams, where Openscapes introduces concepts and workflows; facilitate teams to talk about problems then go and solve them, with accountability and support.

It's about getting stuff done. It's about identifying and making progress on barriers

Laura Waters at the SERO described it as "A process to help you build better lanes of communication"

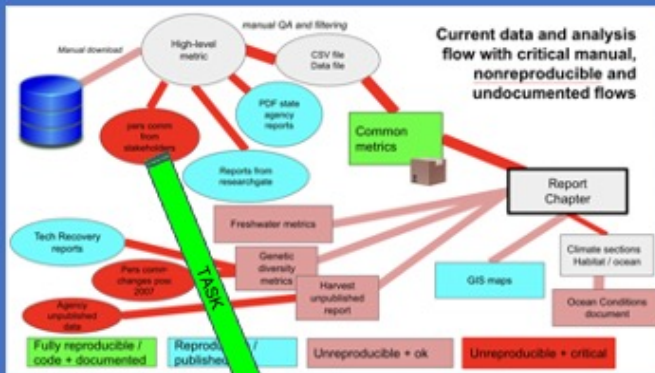
What's important too is that there is sustainability and scalability built-in. It's about strengthening a teaching & learning culture within teams & orgs. Not just for scientists: supervisors, admin, IT staff, etc, welcomed. Equitable.

Openscapes has led 10 Champions Cohorts so far, half of which have been with NMFS! At NMFS it's been a grassroots effort

During Openscapes sessions, scientists talk about issues that are hindering their work. Big issues: reproducibility and tracking, team awareness, on- and off-boarding, too much duplication of effort across the center, agency and year. Last need for team training to collaborate openly and effectively.

Next slides

# PNW Salmonid Viability Report (NWFSC) + Status Reviews (WCRO) Team



Set up single email address for pers comms re data requests

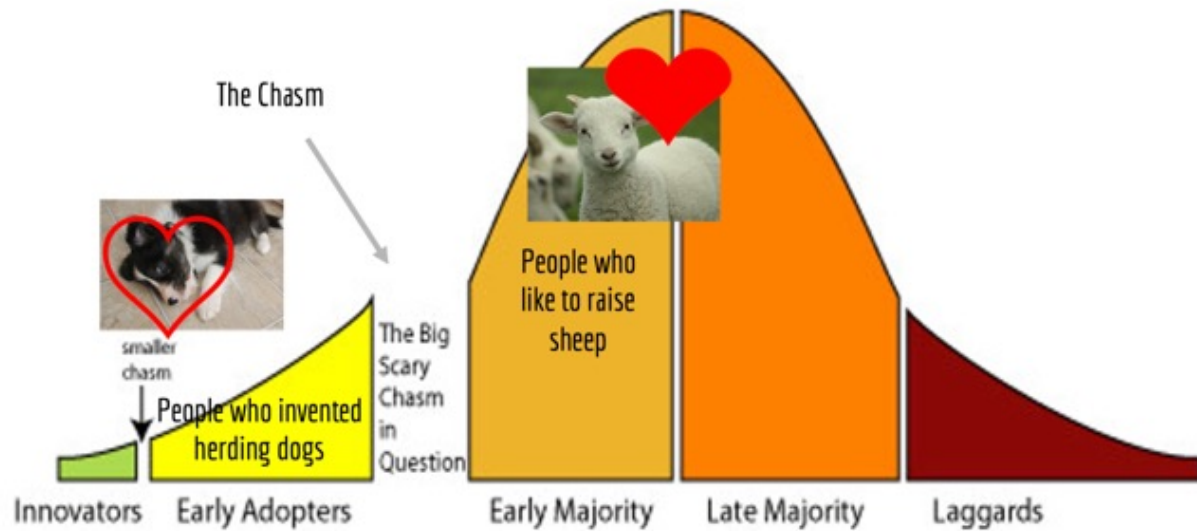




Moore, G. A. (2014). *Crossing the Chasm, 3rd Edition: Marketing and Selling Disruptive Products to Mainstream Customers*. Harper Business.



# The invention of the herding dog analogy



## How do you cross the Chasm?

Option 1. A charismatic communicator  
“salesperson” who is has deep connections with the  
“majority” but also understands the innovation



Hmm, that's kind of hard and not obvious how to do.

From: Moore, G. A. (2014, January 28). *Crossing the Chasm, 3rd Edition: Marketing and Selling Disruptive Products to Mainstream Customers*. Harper Business.

## How do you cross the Chasm?

Option 2. Judiciously choose a single market for the crossing. Put all your effort there.



From: Moore, G. A. (2014, January 28). *Crossing the Chasm, 3rd Edition: Marketing and Selling Disruptive Products to Mainstream Customers*. Harper Business.

## Choose a single market for the crossing

Create many use cases. Pick the one where you can reduce a major and clear pain point and there isn't a good alternative.

### 2022 -- Big Government Reports

- Big time savings
- Savings in staff time can be quantified
- Staff eager to automate soul-crushingly tedious work
- Solves a transparency and documentation problem

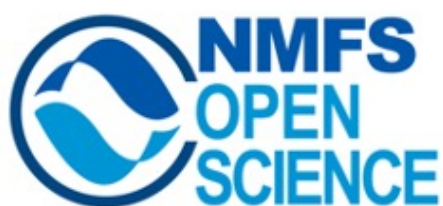
From: Moore, G. A. (2014, January 28). *Crossing the Chasm, 3rd Edition: Marketing and Selling Disruptive Products to Mainstream Customers*. Harper Business.



## 2023 Year of Open Science and Beyond

## NMFS Open Science

The overarching vision of NMFS Open Science is to support scientists, developers, and policy analysts within NOAA Fisheries (NMFS) in fulfilling NOAA's Open Science mandates: NOAA Data Strategy, DOC Open Source Code Policy, Federal Data Strategy, and the Federal Open Access Memo.



## NMFS Openscapes

is concerned Open Science training in workflow and technical skills needed at the individual and team level. We focus on helping all staff engaged in data-driven science and decision-making at NMFS. Support an active and engaged mentor group across NMFS.



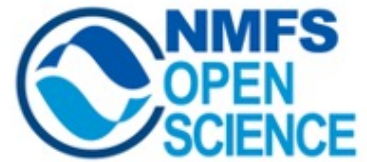


## NMFS Openscapes Plan 2023-2026

[illegible]

## NMFS Open Science is a strategic group

Triage the most pressing needs for scientists, developers, and policy analysts within all of NOAA Fisheries and take leadership roles to find solutions.



NMFS Openscapes	NOAA Fisheries Integrated Toolbox	NOAA Fisheries Integrated Modeling System	NMFS R User Group
The NOAA Fisheries logo (a blue circle with a white wave) and the Openscapes logo (an orange hexagon with a white wave and the word "openscapes" in white).	A hexagonal logo with a blue background, featuring a white wave and the text "NOAA Fisheries Integrated Toolbox" in white.	A hexagonal logo with a blue background, featuring a white wave and the text "FIMS" in white.	The NOAA Fisheries logo (a blue circle with a white wave) and a large blue "R" with the word "USERS" in white.

<https://nmfs-opensci.github.io/>

# Supporting the infrastructure for Open Science

Support for scientific software, package development, templates, utilities



Data science is highly dependent on soft infrastructure: development platforms, cloud virtual machines, and product delivery systems for data-science products. Support governance teams for these platforms.



GitHub Governance Team (live April 3, 2023. <https://sites.google.com/noaa.gov/nmfs-st-github-governance-team/home>)

