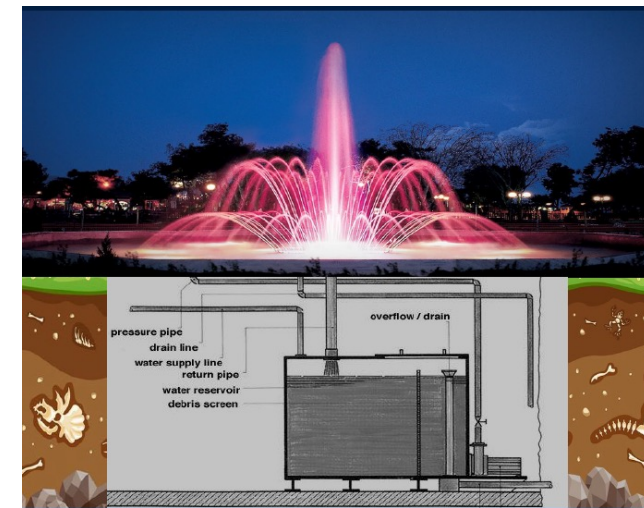
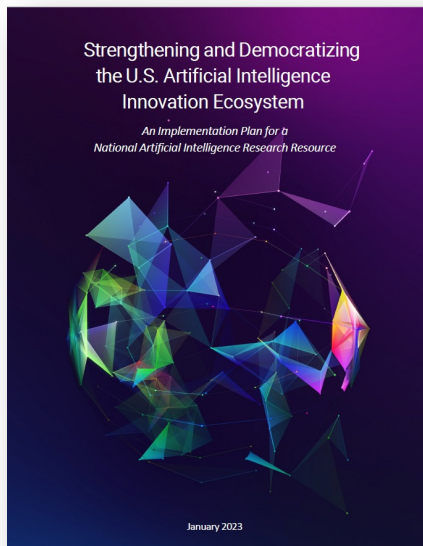


# FAIR data for AI and AI for FAIR data

AKA Democratization of Data  
Julia Lane, NYU



Terry McConnell, UNESCO

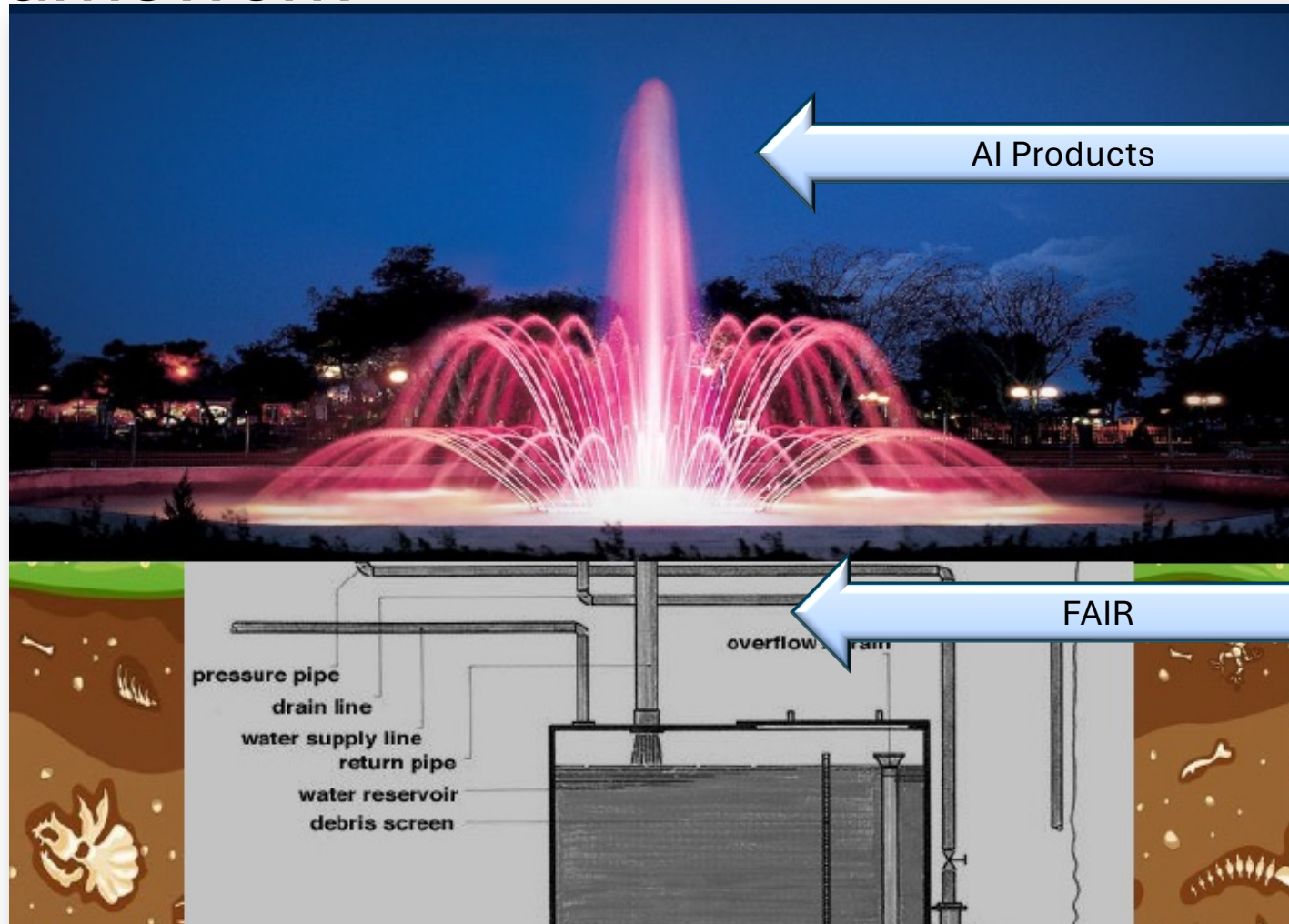
# Outline

- Organizing framework
- US experience
- Theory of change
- Incentivize and democratize
- A possible agenda

# Outline

- Organizing framework
- US experience
- Theory of change
- Democratization
- A possible agenda

# Framework



AI Products

FAIR



Democratization

# Outline

- Organizing framework
- **US experience**
- Theory of change
- Incentivize and democratize
- A possible agenda

# Motivating Example: NAIRR



JANUARY 24, 2023

## National Artificial Intelligence Research Resource Task Force Releases Final Report



OSTP NEWS & UPDATES PRESS RELEASES

Today, the National Artificial Intelligence Research Resource (NAIRR) Task Force released its [final report](#), a roadmap for standing up a national research infrastructure that would broaden access to the resources essential to artificial intelligence (AI) research and development.

While AI research and development (R&D) in the United States is advancing rapidly, opportunities to pursue cutting-edge AI research and new AI applications are often inaccessible to researchers beyond those at well-resourced companies, organizations, and academic institutions. A NAIRR would change that by providing AI researchers and students with significantly expanded access to computational resources, high-quality data, educational



OCTOBER 30, 2023

## Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence



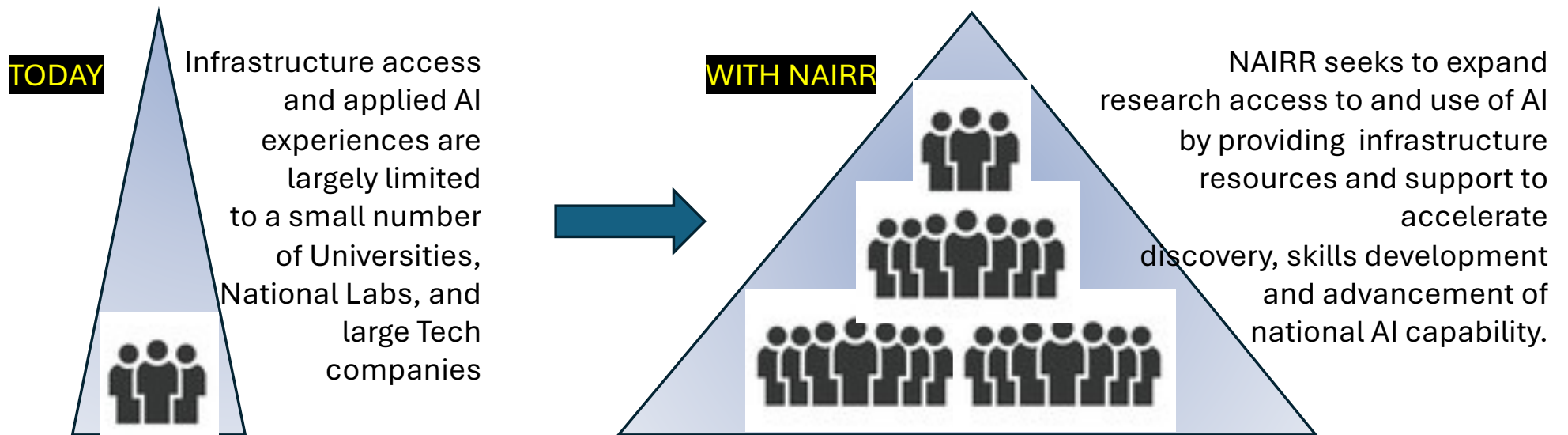
BRIEFING ROOM PRESIDENTIAL ACTIONS

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. Purpose. Artificial intelligence (AI) holds extraordinary potential for both promise and peril. Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure. At the same time, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security. Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor

# Rationale

The NAIRR was created to expand the base for AI basic and applied research, to build a larger population of skilled people for US jobs, and to expand depth of knowledge about the capabilities of advanced technology.



Source: Suzette Kent

# Outline

- Organizing framework
- US experience
- **Theory of change**
- Incentivize and democratize
- A possible agenda



**Baker wasn't alone in winning the Nobel Prize for Chemistry.** The Royal Swedish Academy of Sciences awarded it to Demis Hassabis, the cofounder and CEO of Google DeepMind, and John M. Jumper, a director at the same company, too. Google DeepMind was awarded for its research on AlphaFold, a tool which can predict how proteins are structured, while Baker was recognized for his work using AI to design new proteins. [Read more about it here.](#)

Meanwhile, the physics prize went to Geoffrey Hinton, a computer scientist whose pioneering work on deep learning in the 1980s and '90s underpins all of the most powerful AI models in the world today, and fellow computer scientist John Hopfield, who invented a type of pattern-matching neural network that can store and

[. Read more about it here.](#)

## ARTIFICIAL INTELLIGENCE

**A** emulating, and the other  
**N** aker for tools to help  
**O** w proteins.

AI's us

By Me

Octob

**But there is one problem. AI needs masses of high-quality data to be useful for science, and databases containing that sort of data are rare, says Baker.**

The prize is a recognition for the whole community of people working as protein designers. It will help move protein design from the “lunatic fringe of stuff that no one ever thought would be useful for anything to being at the center stage,” he says.

# Key findings from NAIRTF

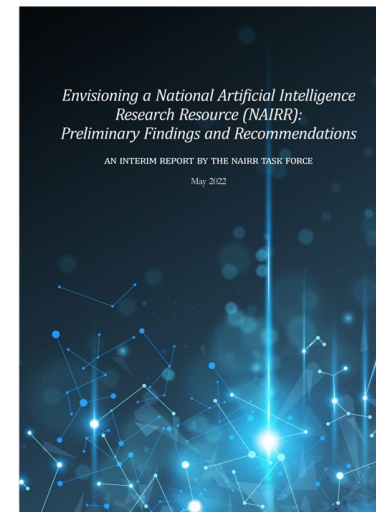
**Finding 4-1: Rigorous AI R&D is often not possible without high-quality, trusted, dense, and transparent data resources.**

**Finding 4-2: There are substantial data quality challenges within and across most research domains.**

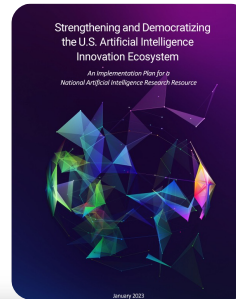
**Finding 4-3: Data curation is a substantial challenge for researchers in all domains.**

**Finding 4-4: There are substantial costs to combining and linking heterogeneous data.**

**Finding 4-5: There are opportunities to learn from, complement, or build on ongoing activities to leverage data for the public good.**



# Recommendation: Democratize data



## Role of the Operating Entity in Incentivizing and Curating Contributed Datasets and Other Resources

Since the quality of many AI models depends on high-quality training and test data, the Operating Entity should establish a data service that facilitates access to and additional use of existing curated datasets of value and interest to the NAIRR user community. Curation of AI data, models, tools, and workflows should be done by the user community in an AI data commons, facilitated by the NAIRR search and discovery platform. Such a community system, governed by terms of use as well as a review system, would facilitate data sharing and curation by members of the community. In the context of a commons model, researchers who contribute to the commons are incentivized through data curation and code sharing, and whose contributions are recognized and valued by relevant communities, could be incentivized through high-profile NAIRR recognition and preferential access to NAIRR resources.



# Theory of Change

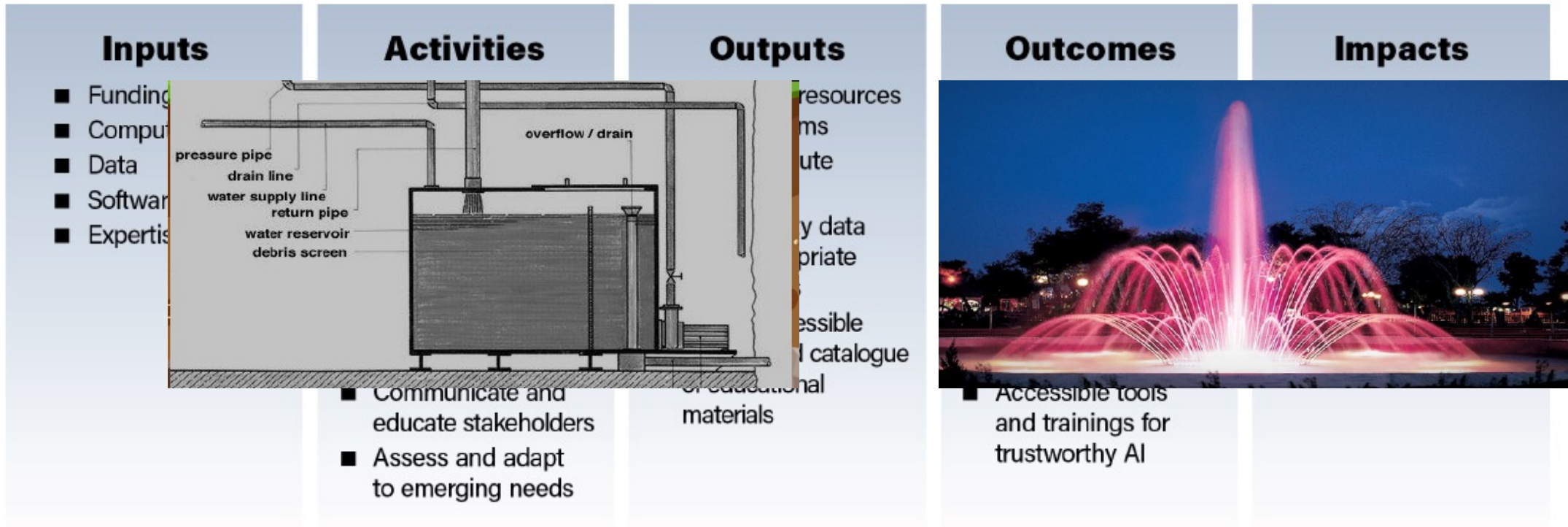


Figure 6. Example Elements of a Theory of Change for the NAIRR

# Outline

- Organizing framework
- US experience
- Theory of change
- **Incentivize and democratize**
- A possible agenda

Incentives

hdsr.mitpress.mit.edu/pub/6blzwgpg/release/4?readingCollection=a6db7dff

HOME ISSUES ▾ SECTIONS ▾ COLUMNS ▾ COLLECTIONS ▾ PODCAST SUBMIT ▾ ABOUT ▾ MASTHEAD ▾

Special Issue 4: Democratizing Data ▾ Published on Apr 02, 2024 DOI 10.1162/99608f92.03719804 SHOW DETAILS ▾

# An Invisible Hand for Creating Public Value From Data

by *Julia Lane, Alfred Spector, and Michael Stebbins*

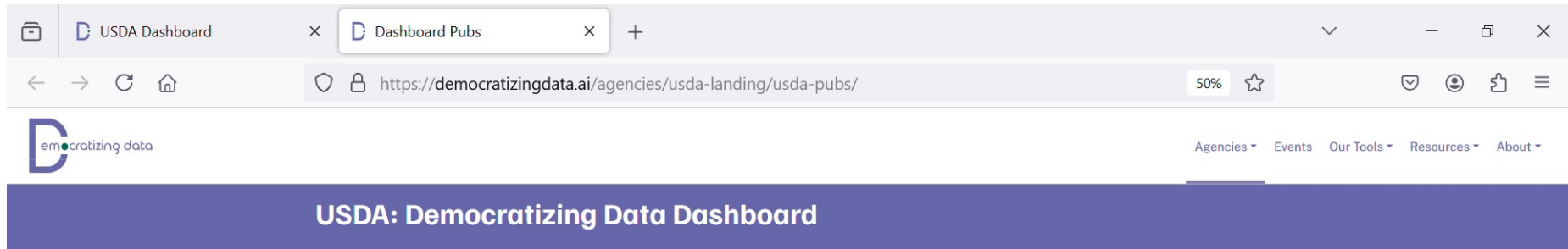
last released 5 months ago


CITE [#] SOCIAL DOWNLOAD CONTENTS

## ABSTRACT

In this article we describe how data usage statistics could be designed to reward the contributions of both researchers and government agencies, functionally creating a carrot that rewards engagement. We provide a concrete use case—the Democratizing Data pilot project—and discuss how the results could be used to develop a community-driven incentive structure, or an ‘invisible hand,’ to generate better evidence for policymakers and lower costs for the taxpayer. We also note the potential to extend the approach and platform to other data-driven initiatives, notably investments in artificial intelligence.

# Incentivize and democratize





**About the dashboard**

This dashboard provides an initial view of data usage in the research communities to answer following questions about USDA data assets: (1) who are the data experts on particular topics, (2) how have the data assets been used, and (3) how have the data been combined and with what results.

**Sort filters by**

Publications  
Citations

**Select Classification and Research Areas**

Research Subfields (Science Matrix)

**Word Cloud**

Select a subject from the word cloud. The bigger a word or phrase appears, the more publications on that subject.

**Select a Data asset | Table View**

List of data assets ranked by use in Publications.

NASS Census of Agriculture	4,931
RUCC	1,510
Household Food Security Survey Module	994

**Stats at a glance**

PUBLICATIONS: 8,525 | JOURNALS: 2,497 | INSTITUTIONS: 6,540 | AUTHORS: 29,639

Select a view: Publications & Journals | Regions, Institutions & Authors | Usage Over Time

**CLEAR FILTERS**

**List of publications in which the data asset is cited**

Data asset: All, Subject: All, Year: All

8,525 Publications

Publication	Citations
'Slow to change': Farmers' perceptions of place-based barriers to sustainable agriculture	23
'Broad but not deep': regional food hubs and rural development in the United States	4
'Farmers Don't Retire': Re-Evaluating How We Engage with and Understand the 'Older' Farmer's Perspective	4
'From a circle of introductions': adult learning and empowerment of women agricultural landowners	9
'If the change is going to happen it's not by us': Exploring the role of NGOs in the politicization of Ugandan agriculture	10

DOWNLOAD THE FULL LIST (Coming soon)

2,497 Journals

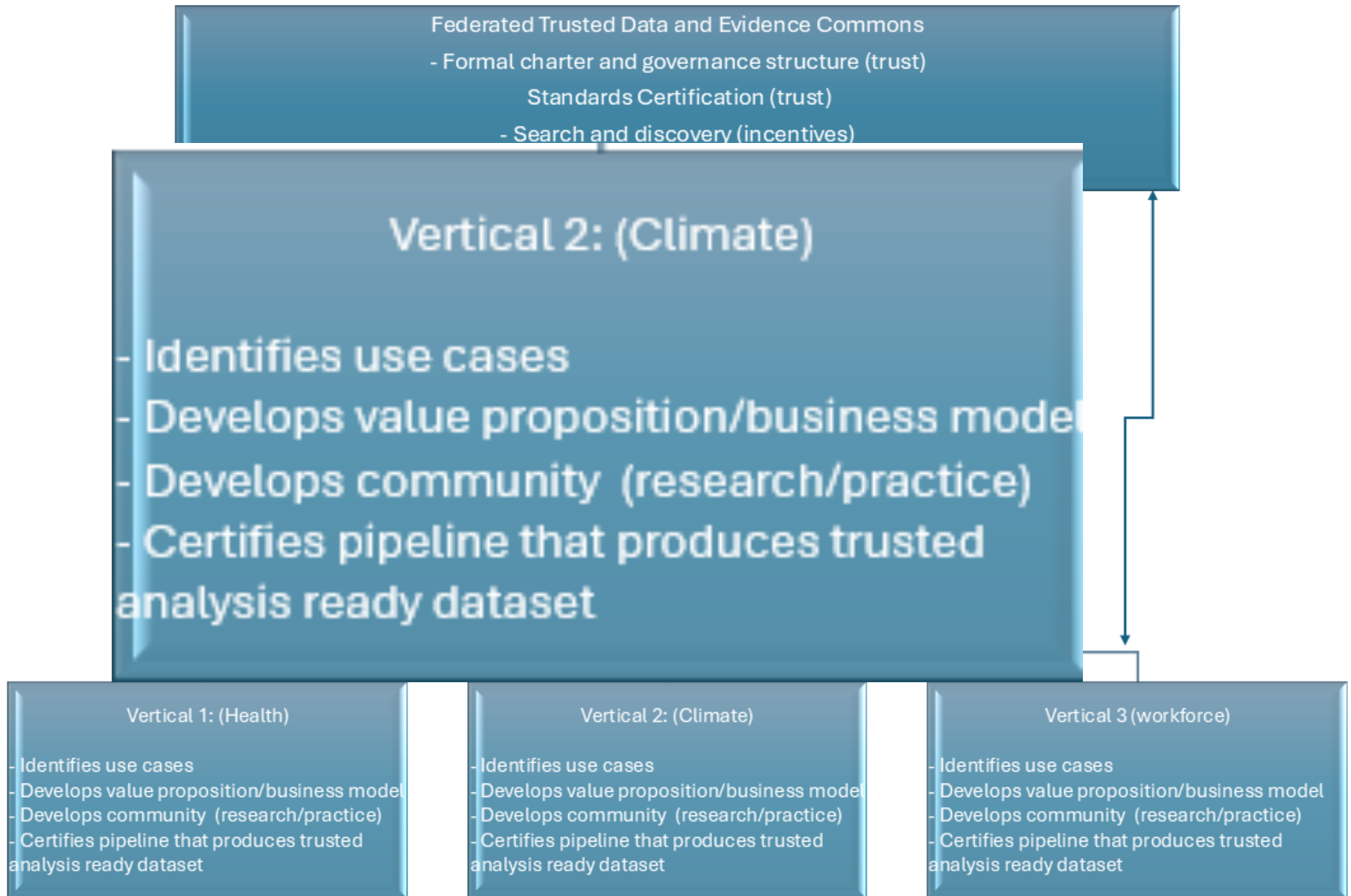
Journal	Publications	Citations
Sustainability (Switzerland)	66	1,551
International Journal of Environmental Research and Public Health	49	1,177
Nutrients	43	1,812
Journal of Hunger and Environmental Nutrition	38	687
Science of the Total Environment	29	1,401

DOWNLOAD THE FULL LIST (Coming soon)

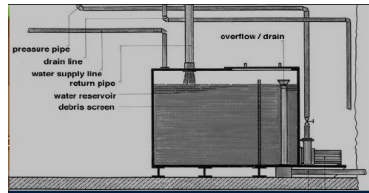
# Outline

- Organizing framework
- US experience
- Theory of change
- Incentivize and democratize
- A possible agenda





# Internationalize

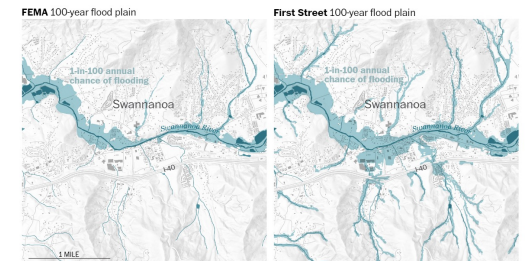



## Democratizing Data: A Search And Discovery Platform For Public Data Assets

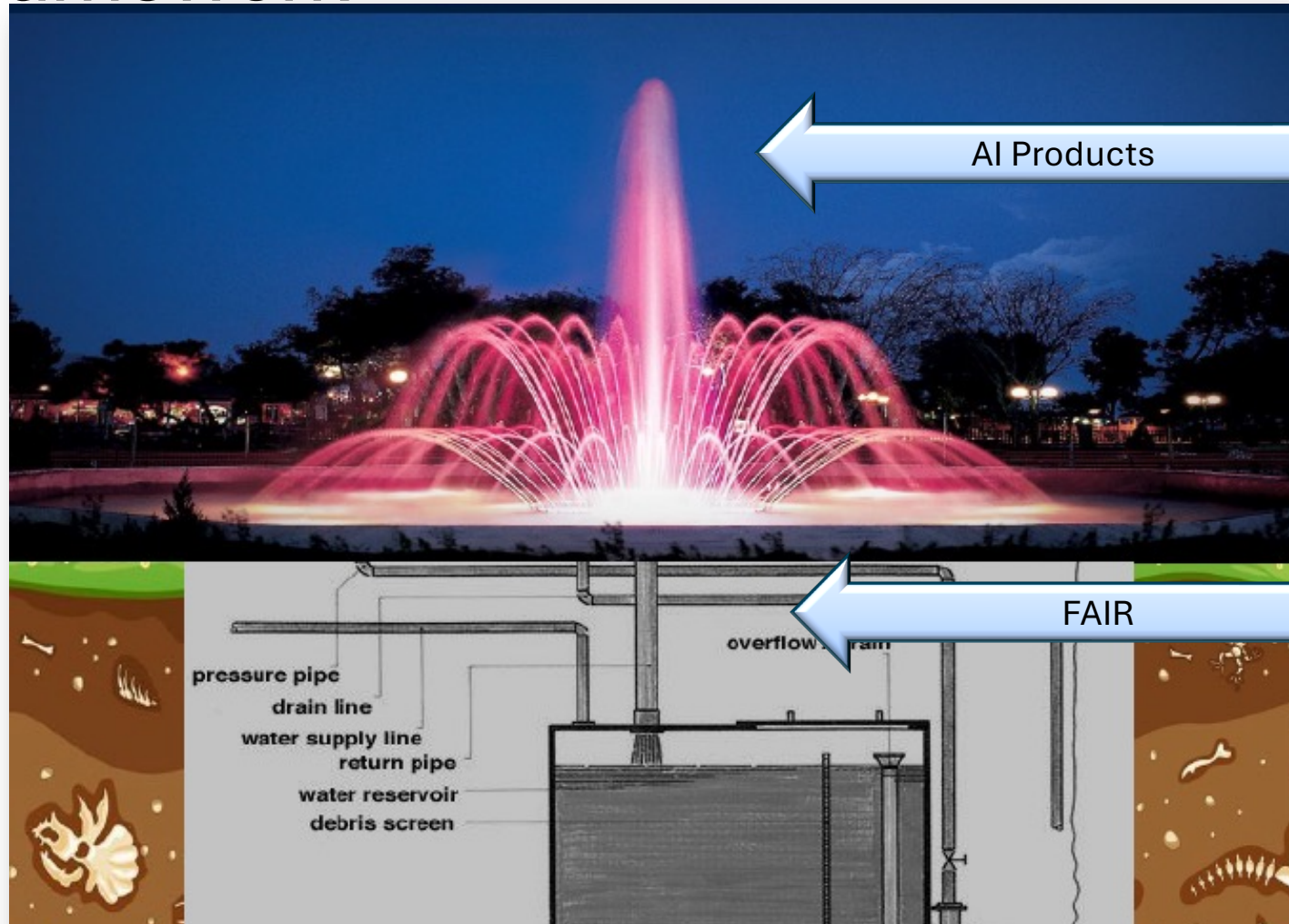
### Step (1) to Digitalise B-Plans

- Source that contains centralized information of:
- Building plan PDFs
  - Geographical locations of the building plans

The Washington Post  
under First Street's model, according to The Post's analysis.

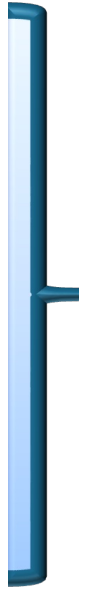


# Framework



AI Products

FAIR



Democratization

# Comments and questions

- Julia Lane
- [Julia.lane@nyu.edu](mailto:Julia.lane@nyu.edu)
- <https://julialane.org/>
- <https://www.linkedin.com/in/julia-lane/>