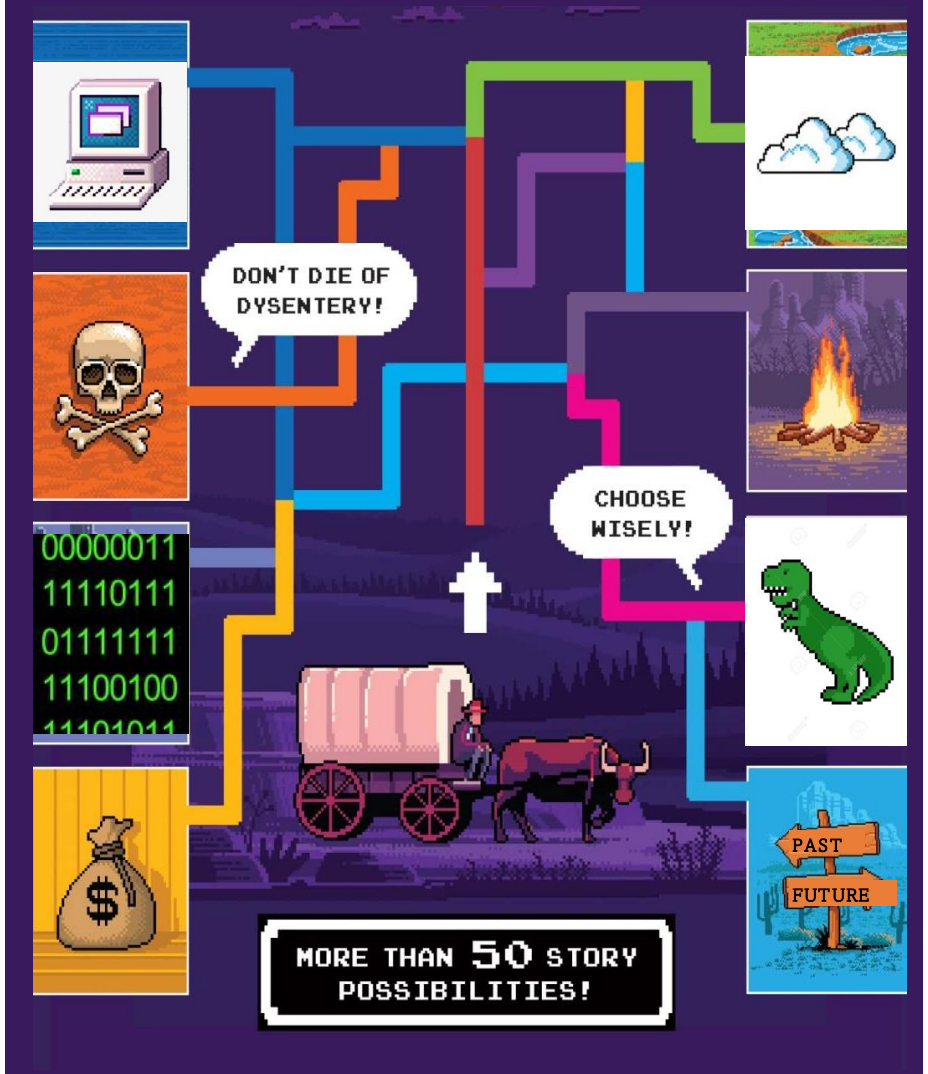


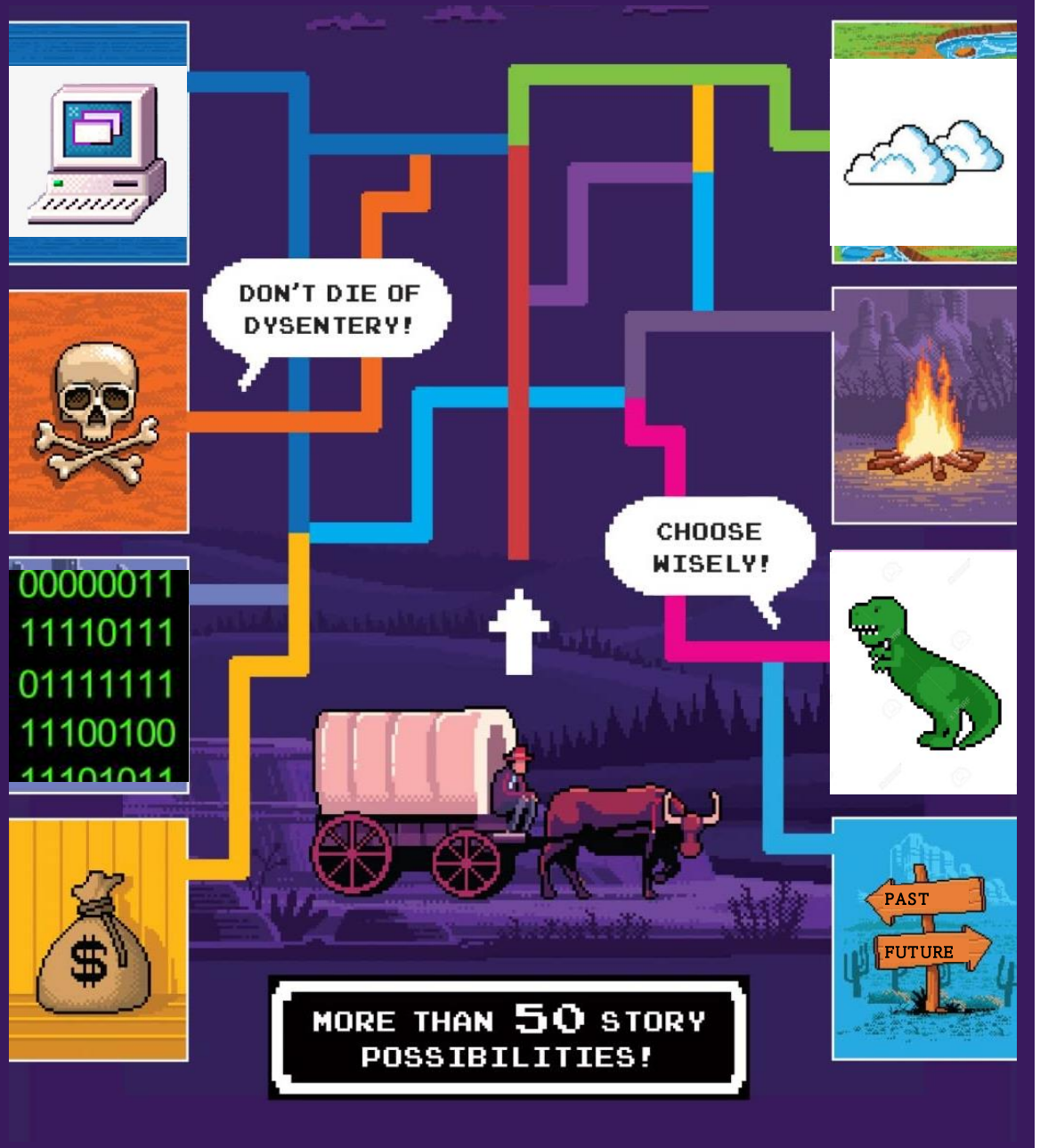
Choose your own (data science) adventure

FedCASIC, 4-5-22

Carla Medalia, Taylor Hanson, Luke Keller, Drew Zachary

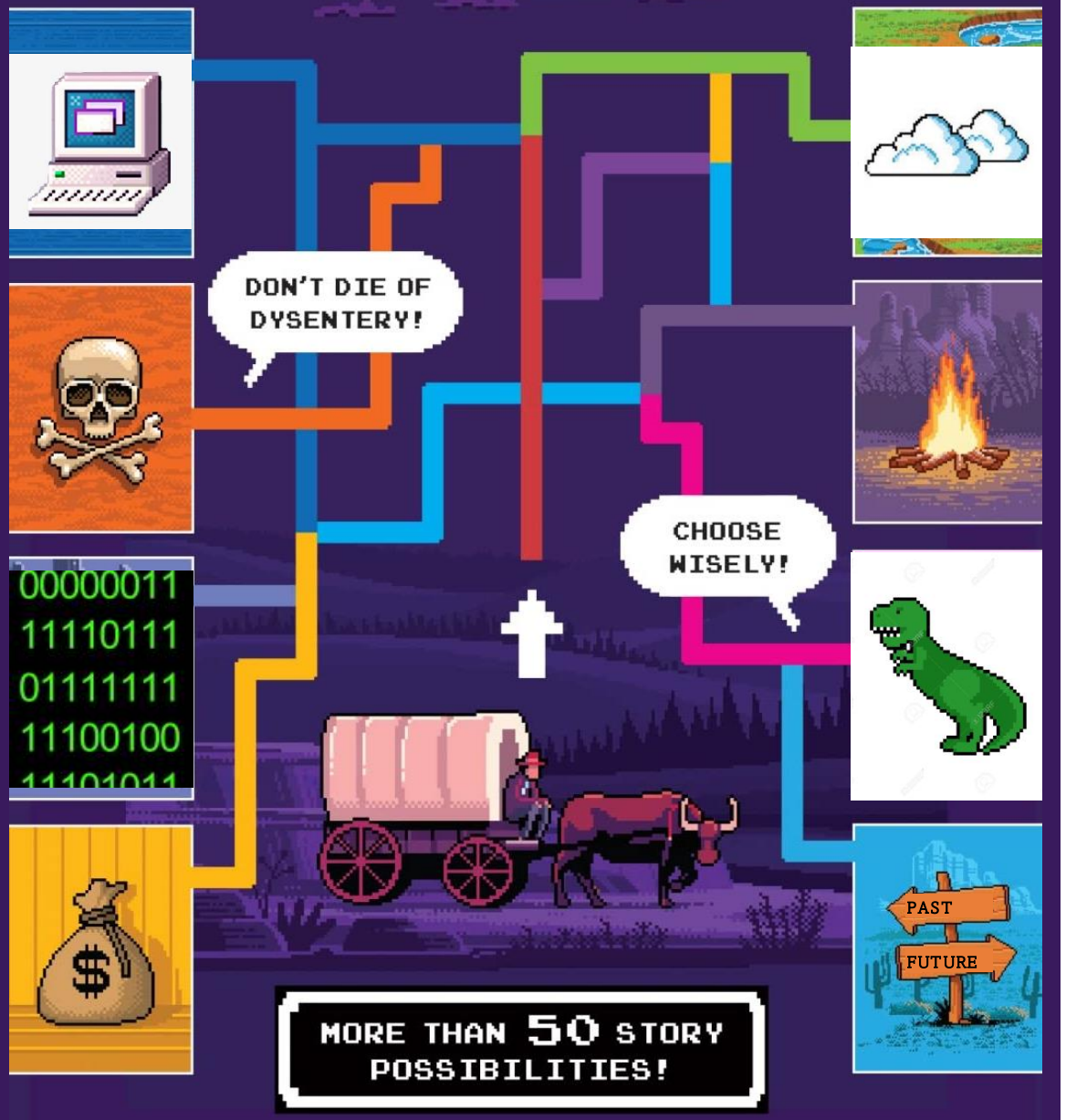
U.S. Census Bureau

Disclaimer: Any views expressed are those of the authors and not those of the U.S. Census Bureau.



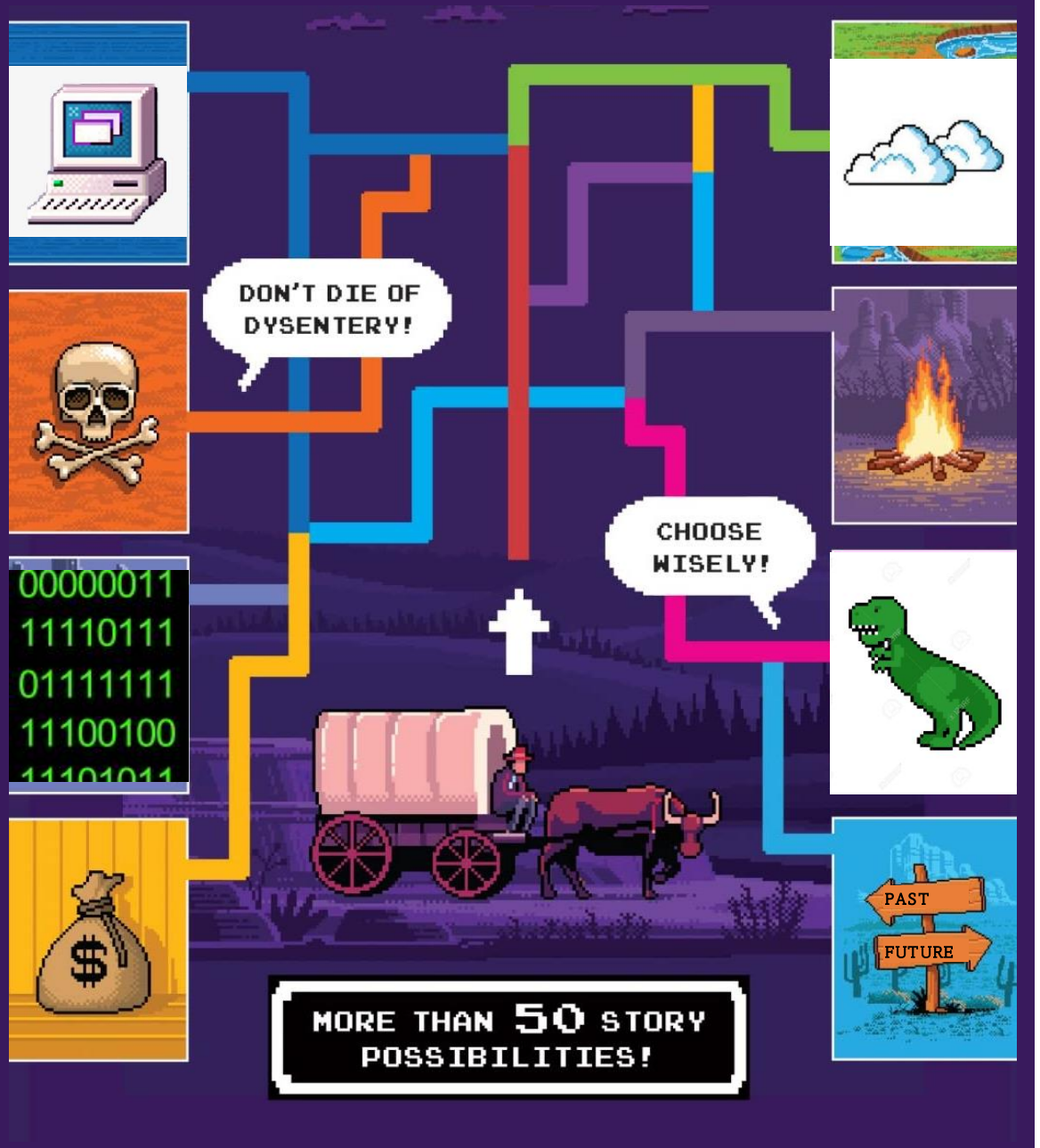
Today's goal

- How innovative teams form, work, and solve problems depends on their mission
- Here are 4 possibilities



Panelists and perspectives

- Carla Medalia: production environment
- Taylor Hanson: software engineering
- Luke Keller: emerging tech
- Drew Zachary: open innovation



Roadmap

- Presentations by panelists
- Panel discussion
- Open Q&A with audience

Innovating within a production environment

Carla Medalia
Assistant Division Chief, Business Development Staff
Economic Reimbursable Surveys Division
U.S. Census Bureau

Challenges that demand our attention

Statistical agencies faced with competing challenges

- Declining response rates
- Reduced (or static) budgets
- Demand for more data
- Demand for more timely/frequent data

Agencies must innovate to address these challenges

Production-focused teams need help to innovate

- Lack of time/resources
- Lack of technical experience
- Inefficient to reinvent the wheel
- Inefficient for everyone to be bureaucracy hackers

Agencies need innovative teams focused on data science

United States® Census Bureau

Mission: Business Development Staff

To leverage data science methodologies and novel data sources to improve the way the Census Bureau collects, produces, and disseminates data.

United States® Census Bureau

How we work

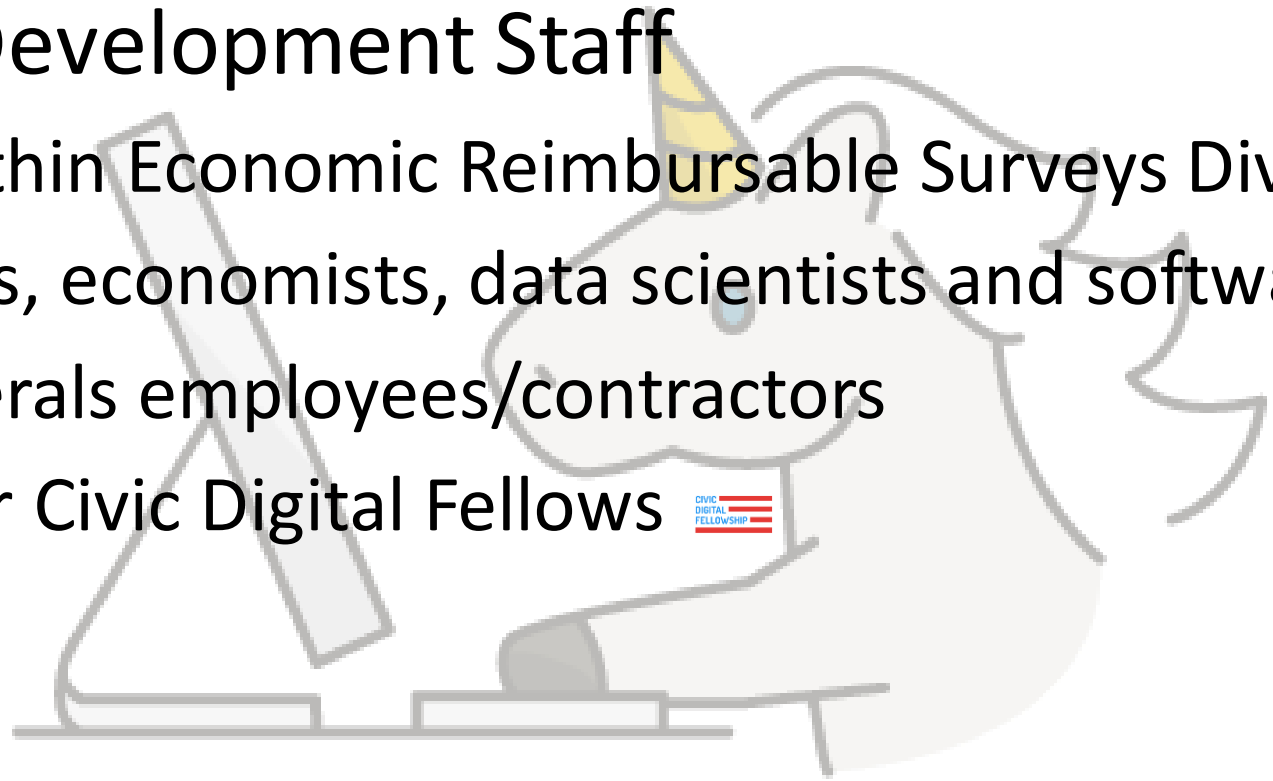
- We collaborate with production teams
  - Stakeholders are both inside and outside Census Bureau
- We leverage new tools, methodologies, and data sources
- We successfully navigate policy/legal/administrative hurdles
- We work in the open and share best practices

United States® Census Bureau

Who we are

Business Development Staff

- Housed within Economic Reimbursable Surveys Division (ERD)
- Statisticians, economists, data scientists and software engineers
- Mix of federals employees/contractors
- Two former Civic Digital Fellows



United States® Census Bureau

Tools we use

GitLab

tableau

django

Azure

aws

sas

UiPath

jupyter

Scrapy

pandas

STATA

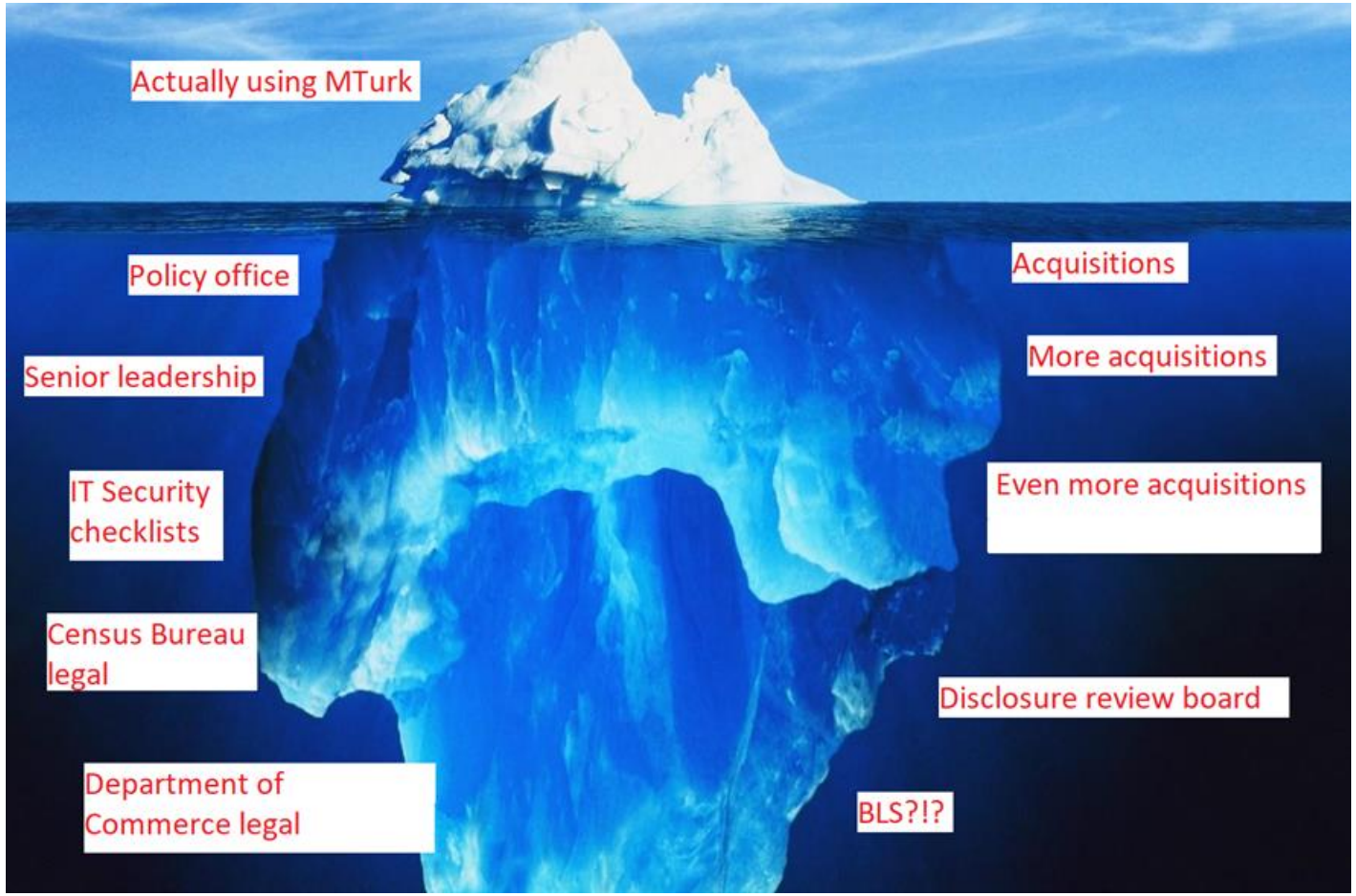
ANACONDA.

Methods we use

- Statistics/economics
  - Statistics/modeling
  - Data linkage
  - Using admin data
- Data science
  - Web scraping/crawling/APIs
  - Machine Learning
  - Natural Language Processing
  - Data visualization
- Software engineering
  - Dashboards
  - Data Pipelines
  - Graphical User Interfaces
  - Web applications
- User-centered design

United States® Census Bureau

But it’s more than that…



Source: Christian Moscardi

United States® Census Bureau

Thank you!

carla.medalia@census.gov

https://github.com/census-bds

United States® Census Bureau

SolAr

EAD Solution Architecture Team

Taylor Hanson

taylor.a.hanson@census.gov

SolAr's Missions

Application Development

"Build stuff"

- Research & Experimentation
- Proof of Concepts
- Rapid Development
- Full-service Project Development

Standards and Coaching

"Help others build stuff"

- Econ Architecture Review Board
- Consulting Services
- Project Kickstarts
- Application Templating
- Knowledge Sharing
- Office Hours

Enterprise and cross-divisional projects

"Build bigger and better stuff"

- Coordination between divisions
- Enabling workflows in other divisions
- Collaboration on large scale projects
- Representation of EAD in ECON/Census settings

How our team works

- SolAr has a flat structure – Chief SA -> Project Teams
- SolAr developers are senior level SMEs only
- We focus on broad, full stack talent and communication skills
- Each staff member is required to voice opinions, suggest POCs, and drive the vision
- Each staff member should eventually be able to ‘run the show’
- Staff members are added to a project team and deployed to a customer
- We strive for deploying for 3 months or less – no ownership, no O&M

SolAr Team Members

- SolAr currently has 16 permanent staff members
  - 11 Developers (Full stack, front end, back end, specialized)
  - 3 Project Managers
  - 1 Cloud Architect
  - 1 UX Engineer
- Most staff are in the ‘Solar Core Team’ and jump from project to project while some staff are in long term, dedicated project assignments

Innovation at xD

xD // DEPDIR

Luke Keller

xD
https://www.xd.gov

United States™ Census Bureau

WHAT IS XD?

xD is an emerging technologies group that's advancing the delivery of data-driven services through new and transformative technologies.

FY21-22 PORTFOLIO AREAS

xD is focusing on 2 **portfolios** in FY22.

Combating Bias in AL/ML

Building a Bias Toolkit to create a suite of tools for bias mitigation. Partnering with areas around the Bureau and government on case studies and deployments.

Privacy Enhancing Technologies (PETs)

What if we could query or train using data that we never see?
We're piloting a suite of emerging technologies to do just this.

PORTFOLIOS

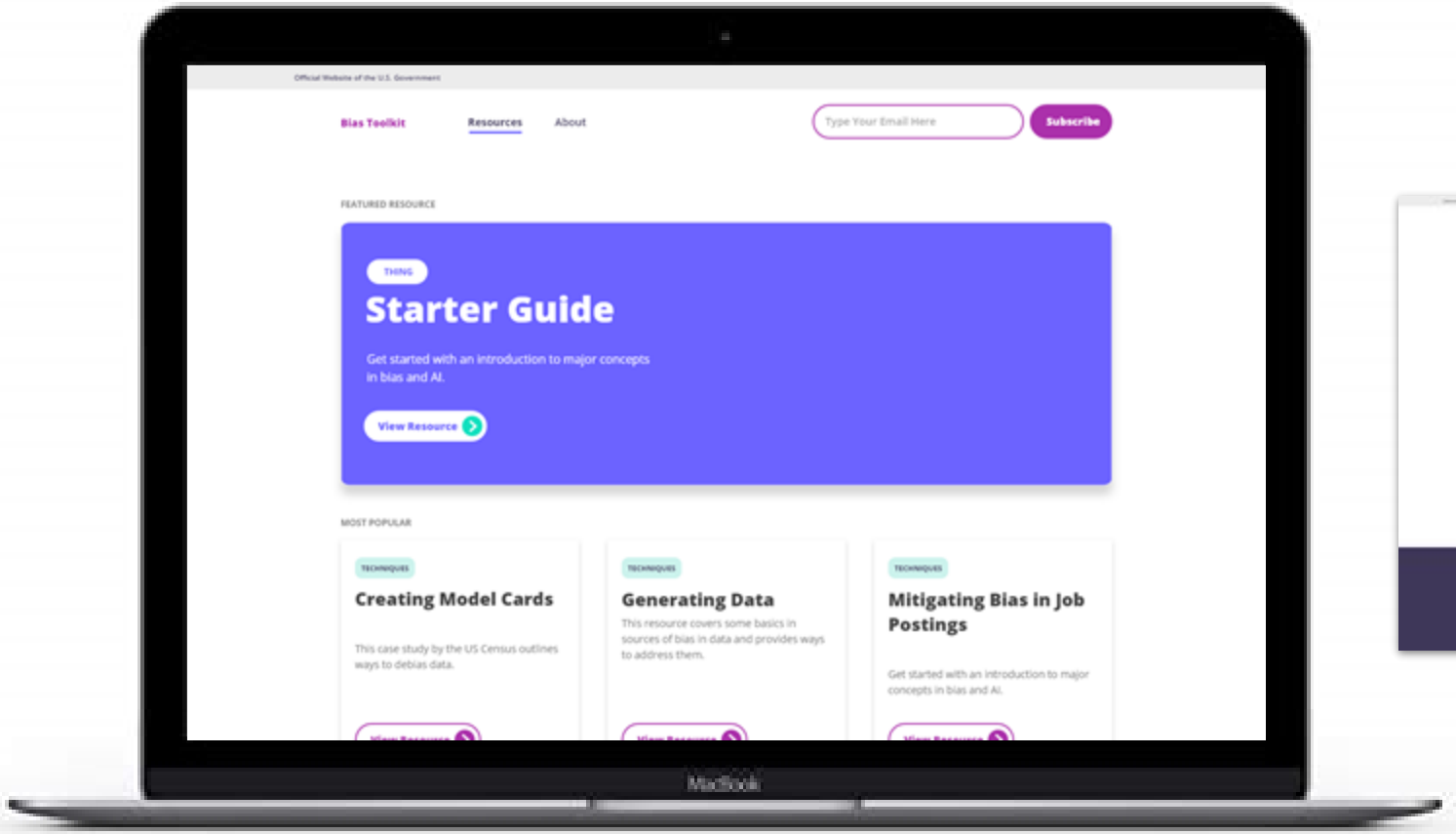
Combating Bias in AL/ML

OVERVIEW: WHAT IT IS

The Combating Bias Toolkit aims to curate a collection of tools that help mitigate and correct sources of bias in federal data and algorithms.

To date, these tools will have been built by DEPDIR but we're eager to include the work of other areas as well.

OVERVIEW: TOOLKIT PROTOTYPE



x 4

Resource Pages @ Launch

Planning for more post-launch as work continues

Model Card Generator

A tool and standard for documenting key information about machine learning models.

Addresses bias through model transparency.

Data Generator

A tool for generating synthetic data to compare outputs of machine learning models using different inputs to probe for bias.

Addresses the identification of bias in models.

Ableist Language Detector

A tool and website for identifying ableist language in job postings.

Addresses bias found in applicant pools for federal jobs.

PORTFOLIOS

Privacy Enhancing Technologies (PETs)

What if we could answer questions (query or train) using **data that we never 'see'**?

Pilot Project: United Nations PET Lab

OCIO and xD are deploying a Secure Multi-Party Computation Platform with the United Nations:

- The US is one of five countries in this pilot.
- The United Nations Global Platform is hosting a network node and each participating country is hosting domain nodes with public open trade data available for querying. (Currently deployed!)
- Each participating country will have the ability to query the platform and answer questions about aspects of trade imports and exports without ever seeing the underlying data of each domain node.

The Economist

Menu | Weekly edition | Search

Science & technology

Jan 29th 2022 edition

Data privacy

The UN is testing technology that processes data confidentially

How to analyse data without revealing their secrets

NEWS

UN Global Expo Highlights UN Privacy-Enhancing Technology Lab Participation

January 26, 2022

Team xD is thrilled to announce the U.S. Census Bureau's involvement in the United Nation's Privacy-Enhancing Technologies (PET) Lab pilot alongside Statistics Netherlands, ISTAT of Italy, and the U.K.'s Office for National Statistics.

As the U.S. representative for the pilot, the Census Bureau wanted to ensure it included a diverse set of subject-matter experts which include Team xD, the Office of Information Security (OIS) and other offices within the Office of the Chief Information Officer (OCIO). "This collaborative initiative introduces a new era in statistical research and data science by accelerating the use of data while maintaining the security, privacy and confidentiality," said Chief Information Security Officer Beau Houser.

This week at the U.N. Expo, we are demoing how researchers from participating countries can compute on trade data they cannot see using an open-source platform called PySyft, a peer-to-peer remote

xD OVERVIEW

Talent Acquisition

WHAT IS THE EMERGING TECH FELLOWSHIP?

The **Emerging Technology Fellowship** (ETF) is a unique talent acquisition program that recruits professionals with expertise in new and emerging data technology trends to drive data innovation at the Census Bureau.

Data Science • Artificial Intelligence (AI) • Secure Multiparty Computation • Privacy-Enhancing Technologies

WHAT IS THE EMERGING TECH ROTATION?

The **Emerging Technology Rotation** is a rotational program that enables Census Bureau employees to join xD for short or longer term opportunities to learn more about emerging technologies in a project-based way.

Data Science • Artificial Intelligence (AI) • Secure Multiparty Computation • Privacy-Enhancing Technologies

WAYS TO GET INVOLVED

We're working hard to improve emerging technology capabilities across the Bureau and government. Here are a few ways to get involved:

- Check out more at https://www.xD.gov
- Schedule a time to chat
- Pitch a project or working paper idea
- Help us recruit top AI talent into government
- Suggest people to connect with

We’d love to hear from you.

Get in touch

All Inquiries: inquiries@census.gov

Luke Keller: luke.keller@census.gov

CENSUS OPEN INNOVATION LABS

You picked the Open Innovation adventure!

Data science is
a team sport

Mission

COIL's mission is to source knowledge and solutions to key challenges for the Census Bureau and the public at large through human-centered design, open data, and technology.

Team Comp

We're hiring!!!

6 feds

Innov. Program Mgr

5-10 contractors

Designer

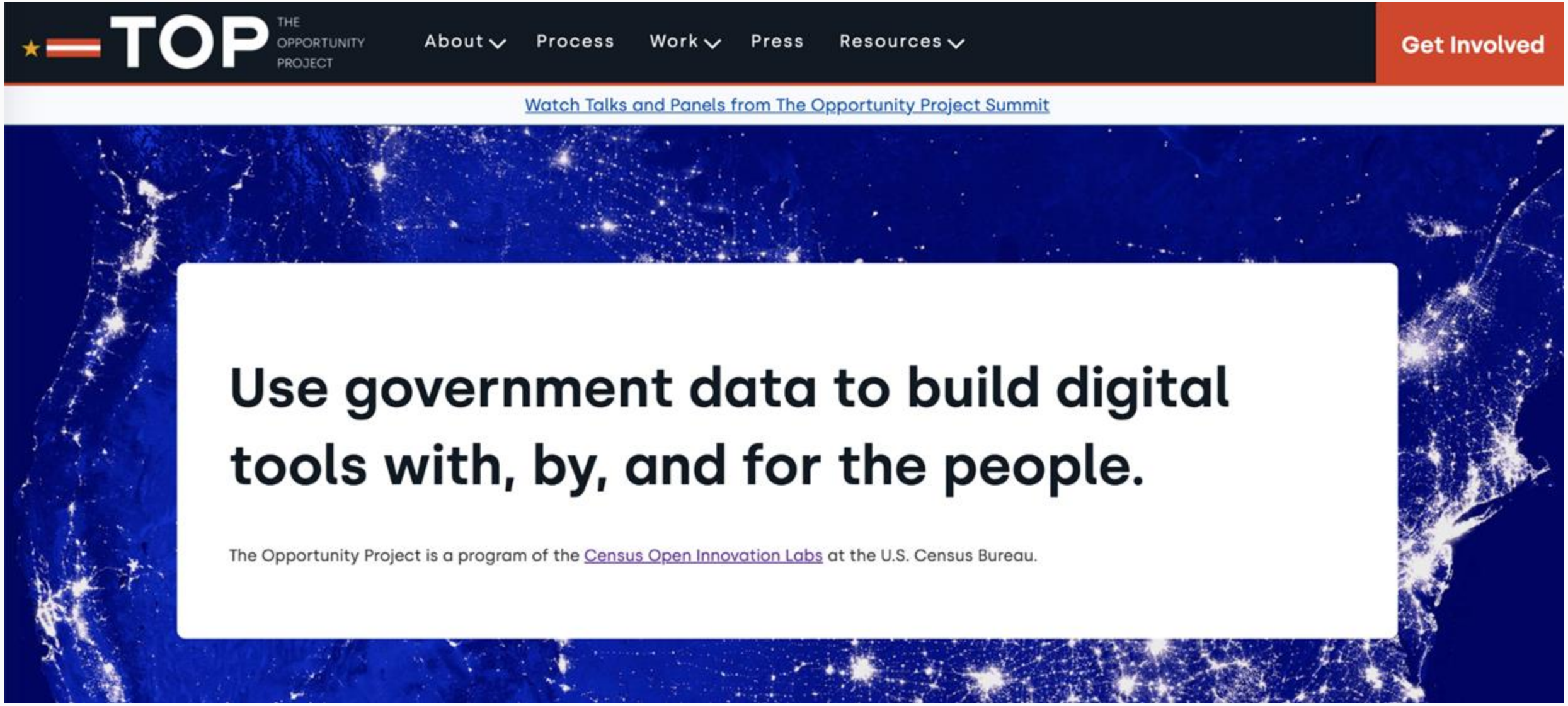
Web Developer

Capabilities

- Visual Design
- Human-centered design
- UX Research
- Data policy
- Program Management
- Technology development
- Strategic partner engagement
- Facilitation

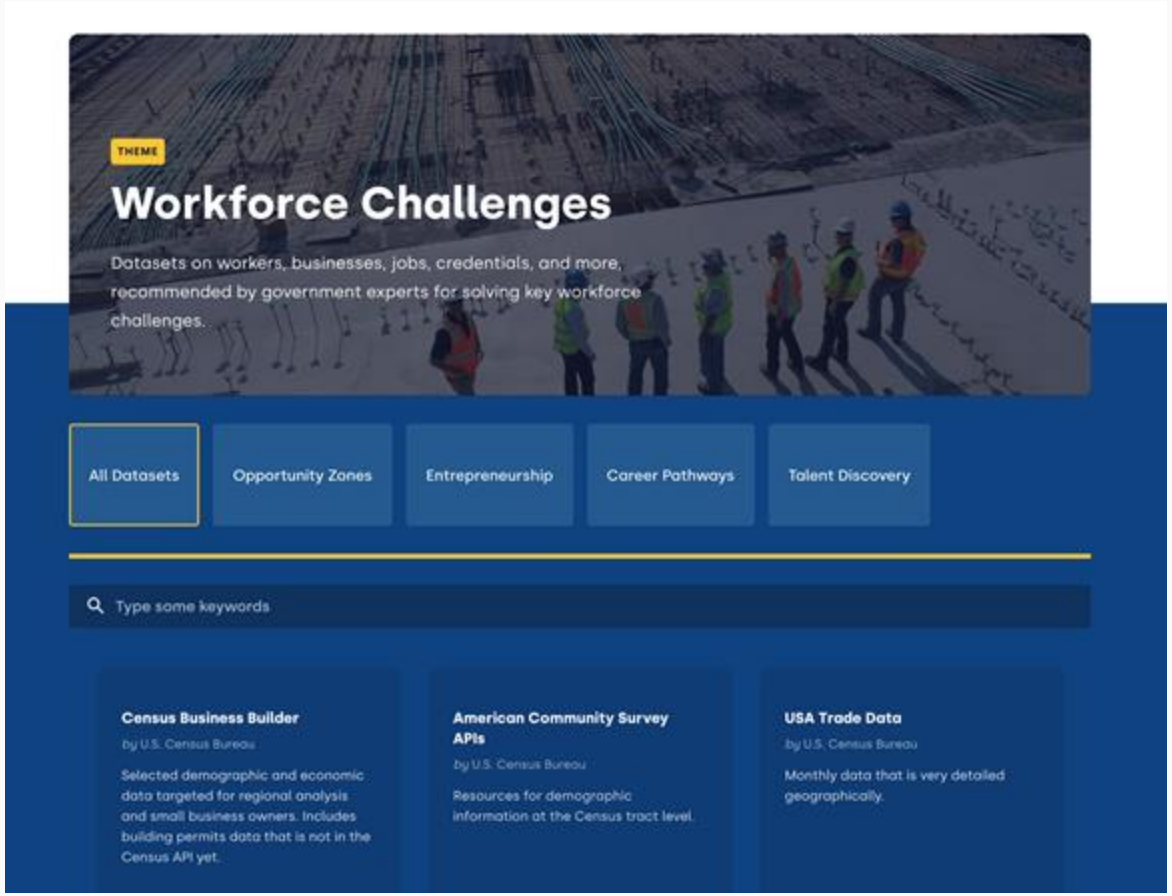
100s of internal and external partners

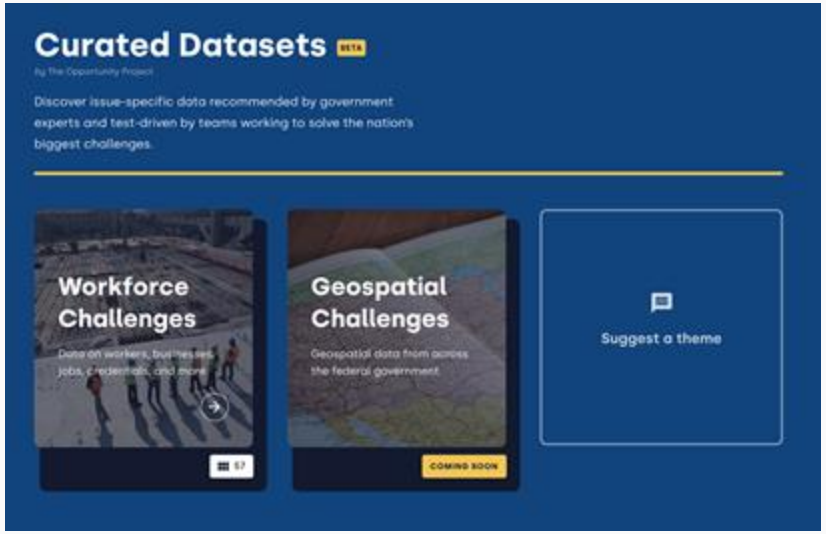
Some of our current work...


TOP
THE OPPORTUNITY PROJECT
About
Process
Work
Press
Resources
Get Involved
Watch Talks and Panels from The Opportunity Project Summit
Use government data to build digital tools with, by, and for the people.
The Opportunity Project is a program of the Census Open Innovation Labs at the U.S. Census Bureau.

TOP

Data Curation Hub ↗

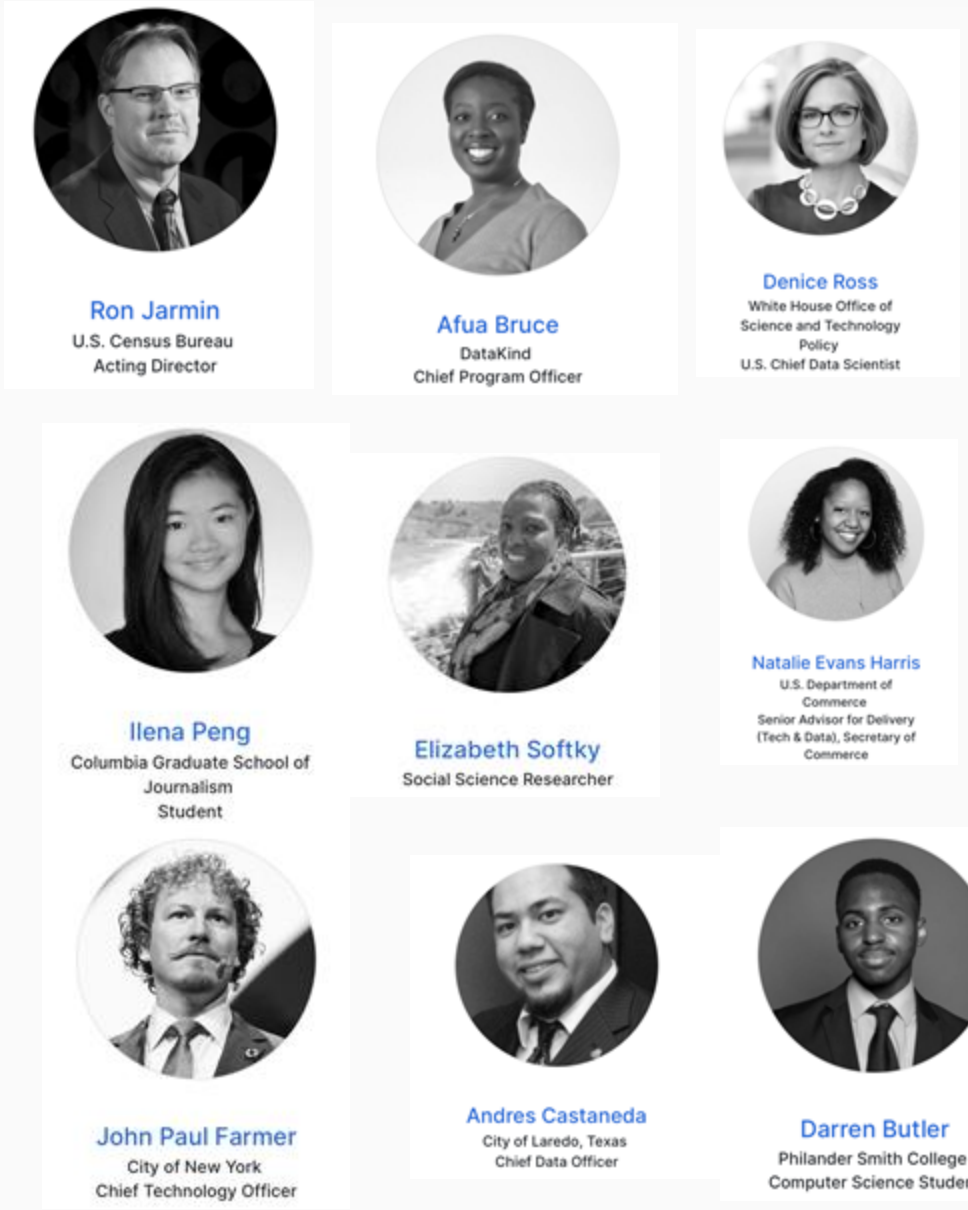




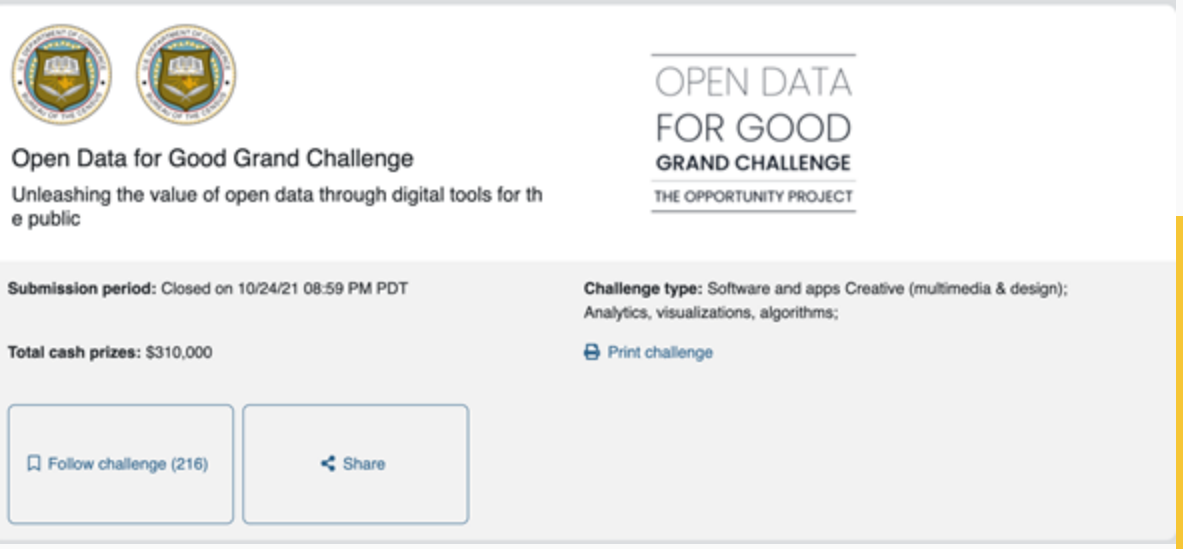
PRESENTED BY
CENSUS
OPEN
INNOVATION
LABS
THE OPPORTUNITY PROJECT
SUMMIT.
OPEN INNOVATION FOR ALL
DECEMBER 13 – 15, 2021
RSVP TO ATTEND THE VIRTUAL EVENT

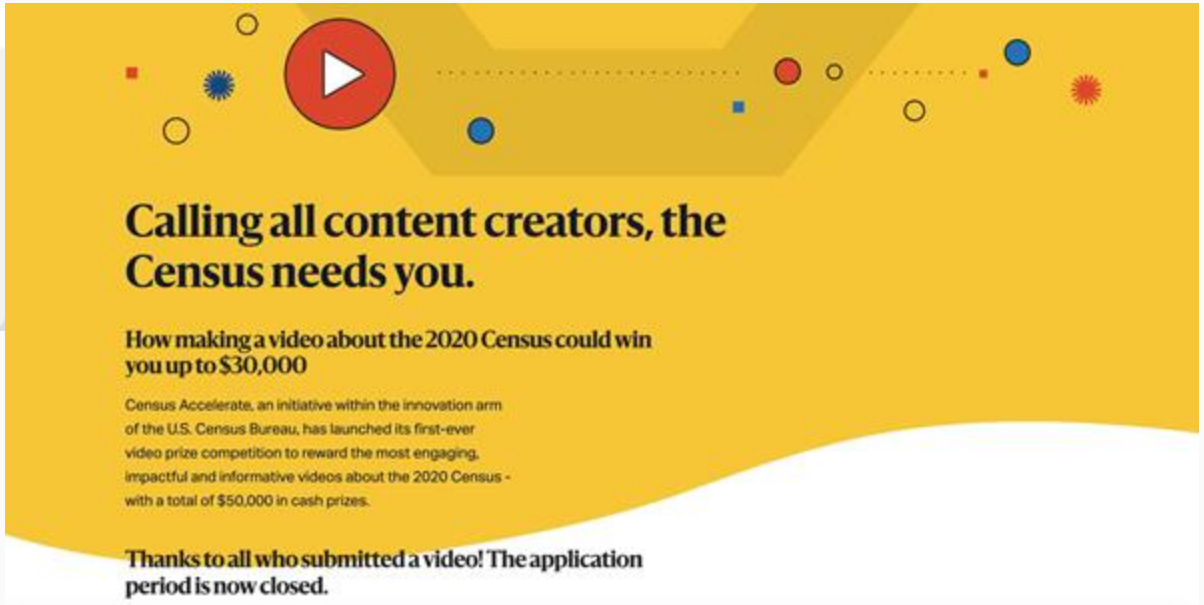
2021 Summit

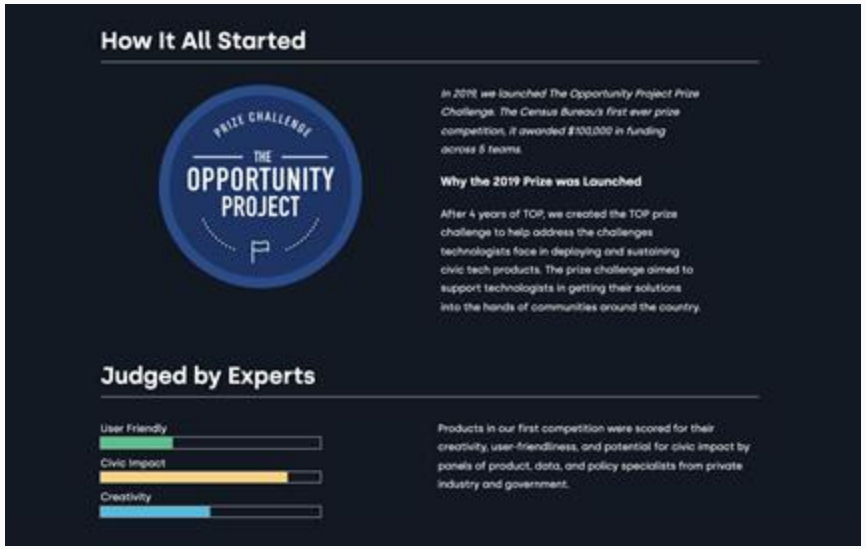
- 80+ speakers
- 2,200+ registered attendees
- 6,000+ Youtube views
- 17 hours of content
- 26 sessions on topics including
    - Meaningful partnerships with MSIs
    - Technology demos
    - Minority business access to capital
    - Thought diversity and polarization
    - Data equity
    - Tech talent hiring



Prize Challenges







Partners

REDFIN

mapbox

IBM

MIT

LinkedIn

mastercard
Center
for Inclusive
Growth

COLUMBIA
UNIVERSITY

CISCO

Microsoft
Health

esri

airbnb

Great!
SCHOOLS

civic
alliance

fitbit

lyft

sas

CORI
Center
on Rural
Innovation

alteryx

Georgetown
University

citi VENTURES

NEW Design Workshops: Data Equity and Inclusive Data Accessibility

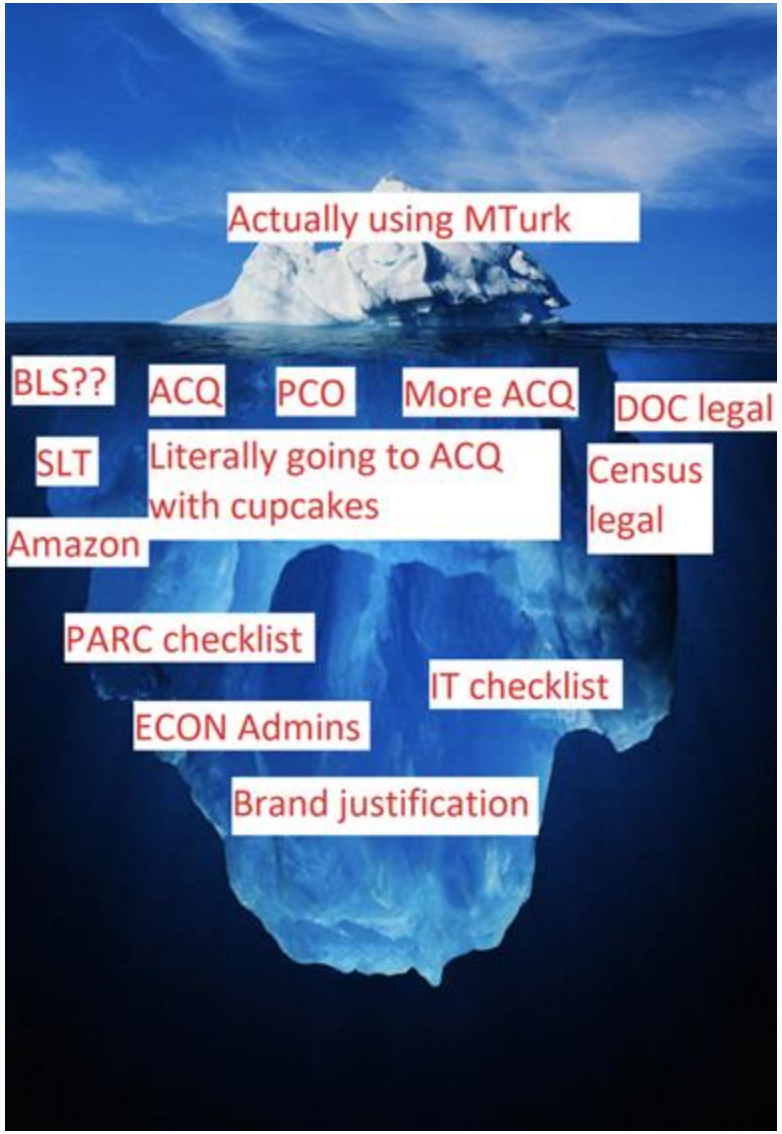
How do Census Bureau data consumers -- especially non-statistically trained -- envision making our data more accessible and useful to them?

- **Spring**: Pilot workshops
- **Summer**: Early findings reported
- **Fall**: Series launched, if successful

Collaboratory

Collaborative | Experimental

Create space that helps creative, forward-thinking individuals at Census meet on another, share Bureaucracy hacks, tips and tricks, and push each other forward.



By the Numbers

180
Products Created

4
Open source toolkits

750
Videos

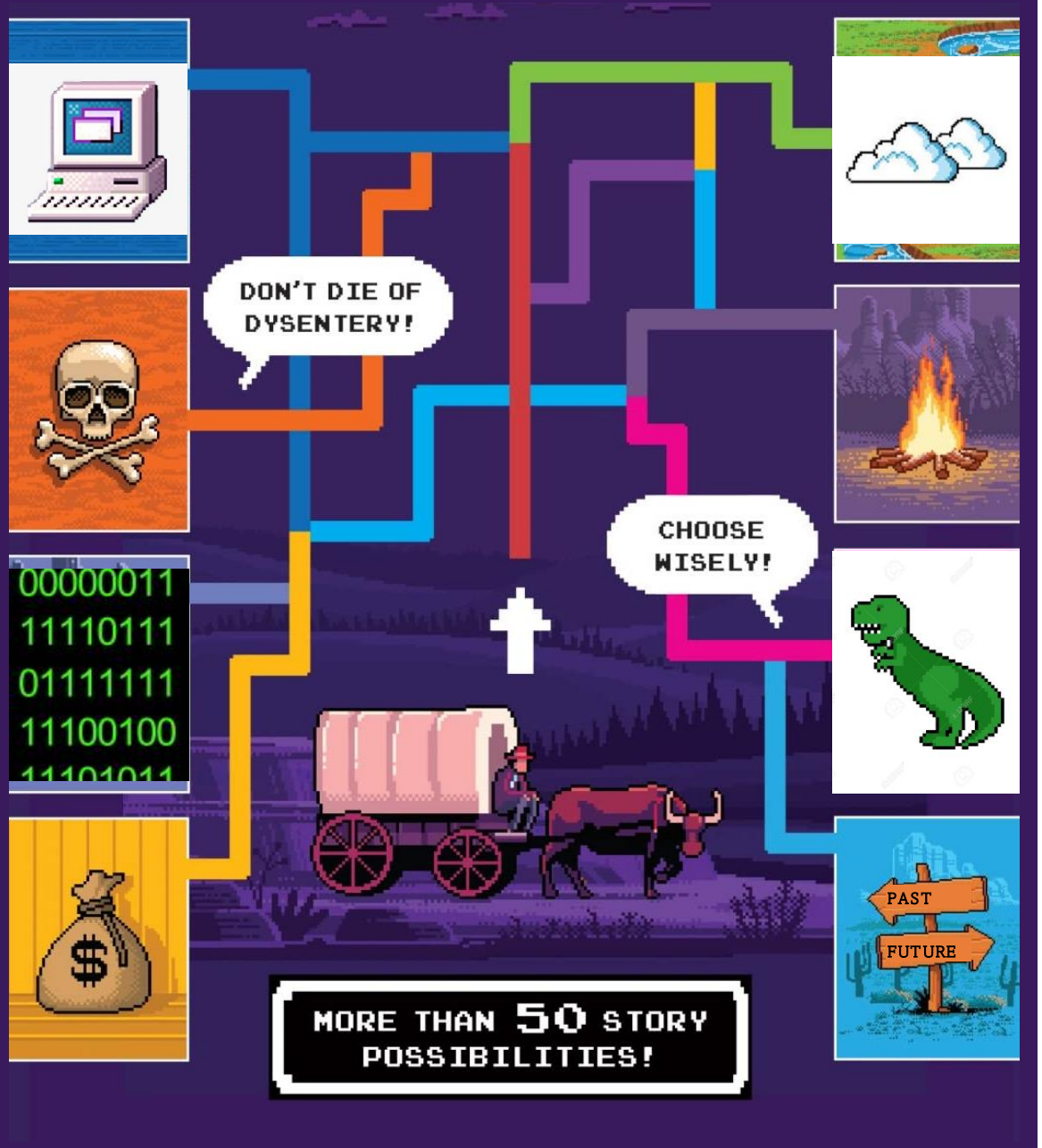
20,000+
Event livestream views

75+
managers trained in HCD

$410,000
In Prizes Awarded

Thank you!

drew.a.zachary@census.gov



Roadmap

- Presentations by panelists
- Panel discussion
- Open Q&A with audience