The Role of Transparency in Promoting Scientific Integrity

Open Census Initiative

FCSM Briefing

10/25/2023

United States® Census Bureau

Agenda

- Introduction
  - Vision
- Proposed Solution
  - Open Census Initiative
  - Initiative Implementation Challenges
  - US Government Models
- Solution In Action
  - Open Data Initiative
  - AWS Open Data Registry
- Open Census Next Steps
- Open Census and the FCSM Network

United States® Census Bureau

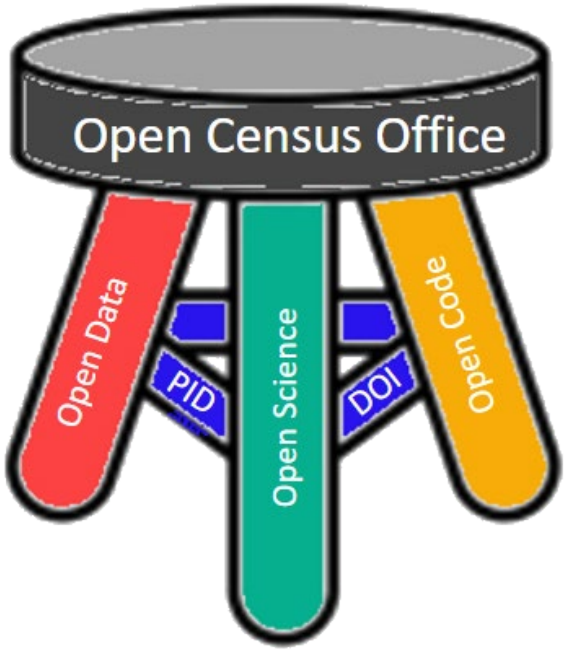
Introduction | Vision

- Become a leader in US Government Open Platforms for Open Science, focused on public accessibility.
  - Open Data, Open Code, Open Science
- Provide an intuitive and secure ecosystem for Census Bureau researchers, software engineers, and statisticians to develop, publish, and disseminate their work, driving the highest standards of research integrity and transparency.
  - Increase peer review, collaboration, citations, and reuse of federally funded research
- Develop USCB standards and best practices for cataloging published products.
  - Metadata, Persistent Identifiers (PIDs), and Digital Object Identifiers (DOIs)
- Seamlessly publish USCB supplemental data products in cloud native formats.
  - Encourage and enable further research using USCB data

United States® Census Bureau

Proposed Solution | Open Census Initiative

- Become the central repository for statistical research dissemination
  - Establish transparency, accessibility, collaboration, and integrity for all publicly released federally funded Census research data products, publications, and software.
- Three distinct but connected Initiatives
  - **US Census Open Science Initiative**
    - Open-Source Research Platform to access and search publicly supported research
  - **US Census Open Data Initiative**
    - Commercial Cloud Data Repositories to access Census data
  - **US Census Open Code Initiative**
    - Code Collaboration Repository to access Census software products and projects



- Connected via Standardized Metadata
  - An upload to any of the platforms or repositories will be accompanied by a USCB Standardized set of
    - Persistent Identifiers (PIDs)
      - Digital Object Identifiers (DOIs)

United States® Census Bureau

Challenges | Open Census Implementation

- Ensure Information Accessibility to the American Public
- Securely Manage Intellectual Property
- Maintain Information Quality, Accuracy, and Integrity
- Provide Reproducibility and Citation Capabilities to USCB Researchers
- Assure Seamless Integration with Existing USCB Data Sources
- Develop Effective and Efficient Path to Onboard Data

United States® Census Bureau

Models |US Government Landscape

- Government Partners with an Open Science Solution
  - Department of Energy (DOE) – Office of Scientific and Technical Information
  - National Institutes of Health (NIH) – National Library of Medicine
  - Department of Agriculture (USDA) – National Agricultural Library
  - National Aeronautics and Space Administration(NASA) – Transform to Open Science Home | NASA TOPS
- Common Characteristics
  - Centralized landing page
  - Smooth and coherent UI
  - Standardized metadata cataloging
  - Allows for reuse with prebuilt citations
  - Provides everyday resources for the American public

United States® Census Bureau

Solution in Action | Open Data Initiative

- Goal: Publish USCB data in a cloud native format to easily store and share.
- Amazon Web Services (AWS) Open Data Registry
  - Centralized repository of publicly available datasets from a variety of organizations, including government agencies, research institutions, and businesses.
  - Datasets are available for anyone to use, without any restrictions, and are accessible from AWS resources, making them easy to use for analysis and machine learning.
  - Offers a convenient way to find and use large datasets and a valuable resource for anyone who needs or wants access to large amounts of data for research or educational purposes.
  - Metadata standards are established and are in process for evolving how the USCB approaches Metadata across all publicly released products.

United States® Census Bureau

Solution in Action | AWS Open Data Registry

- Product Releases on AWS Open Data Registry
  - 9/15/2023
    - 2010 Redistricting Noisy Measurement File (NMF) State-Partitioned
    - 2010 Demographic and Housing Characteristics (DHC) NMF County-Partitioned
  - 10/1/2023 - 2020 Redistricting NMF State-Partitioned
  - 10/23/2023 - 2010 Demographic and Housing Characteristics (DHC) NMF County-Partitioned

United States® Census Bureau

Solution in Action | Census GitHub

- Product Releases on Census GitHub

  DAS 2020 DHC Production Code

  - source code for the version of the Top-Down-Algorithm (TDA) used in the "Decennial Disclosure Avoidance System" for the 2020 DHC (Demographic & Housing Characteristics) release

  DAS 2020 Redistricting Production Code

  - source code for the Disclosure Avoidance System (DAS) used to protect against the disclosure of individual information based on published statistical summaries.

- Coming Soon...

  DAS 2020 Detailed DHC-A Production Code

  DAS 2020 Detailed DHC-B Production Code

  DAS 2020 S-DHC Production Code

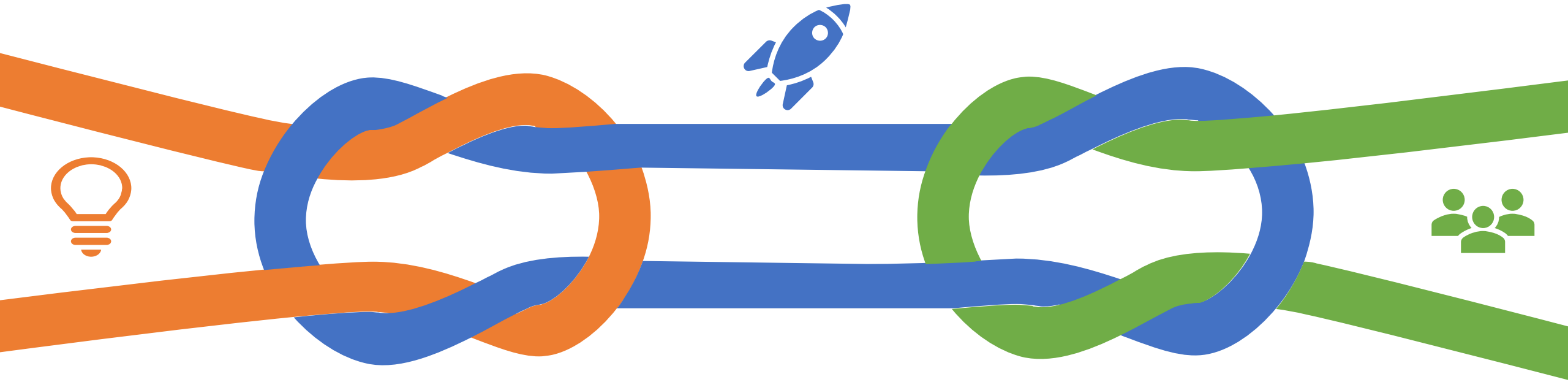
United States® Census Bureau

Open Census Initiative Next Steps

- Near Term
  - Inform and involve statistical partners
  - Continue refining Open Data approach, building off the success of the AWS Open Data Registry
  - Gather requirements for Open Science and Open Code from statistical partners
  - Develop Concept of Operations for the vision and operation of Open Census
- Long Term
  - Establish Requirements for Open Science and Open Code platforms
  - Develop partnerships with entities working in Open Science and Open Data
  - Publish Research on Open Census associated platforms

United States® Census Bureau

Open Census and the FCSM Network



Open Census Initiative

USCB Open Census Initiative developed and funded. Planning and Requirements gathering begun with the goal of supporting an Open Data and Open Science in FY24.

Public/Private Partnerships

Leverage existing and develop new partnerships with public and private entities in the Open Science, Data, and Code industry space to gather requirements, implement best practices, and achieve operational readiness.

FCSM Network Support

Once operational, participation, support, and feedback from industry partners such as the FCSM Network will increase the effectiveness of the Open Census solutions

United States® Census Bureau