Data Frames and FAIRness

Michael Ratcliffe
Geography Division
Census Bureau

GeoGov Summit Roundtable

September 6, 2023

My remarks will center around two projects:

- The Building Trust and FAIRness into the Process for Finding and Using Government Data ("FAIRness Project")
  - Advance efforts to improve and enhance data users' (internal and external to the federal government) ability to find and assess the utility of federal data.
- Census Bureau's Frames Program
  - To create Enterprise-wide frames linkable in nature, agile in structure, accessible for production or research on a need-to-know basis, and that adhere to best practices in terms of technology usage, data management, and methodology.

FAIRness Project

- Joint project of the Chief Data Officers Council and the Federal Committee on Statistical Methodology
- Co-Leads:
  - Thomas (Tod) Dabolt, Chief Data Officer, Department of Interior
  - Michael Ratcliffe, Senior Advisor for Frames, Census Bureau
- Project Goal:
  - Advance efforts to improve and enhance data users' (internal and external to the federal government) ability to find and assess the utility of federal data.
    - Aligned with the Evidence Act requirement for a comprehensive federal data catalog that supports implementing the new data access authorities and responsibilities under Title III of the Evidence Act.
    - Builds from FAIR (Findable, Accessible, Interoperable, and Reusable/Reproducible) principles.

FAIRness Project Deliverables

- Three deliverables:
  - **Updated DCAT-US Schema to a US profile of DCAT version 3**, aligning the schema with international standards and geospatial metadata in the GeoPlatform, and improving adoption by software vendors and open-source technologies.
  - **A proposed governance model** for updating and maintaining the federal metadata schema profiles in accordance with data inventory and data cataloging standards.
  - **A strategic two-year sequencing plan** to improve and integrate government-wide metadata cataloging to facilitate data users being able to easily find, access, assess for fitness-for-use, and use federal data.

  - For more information, visit the project site on GitHub (***https://github.com/DOI-DO/dcat-us***)

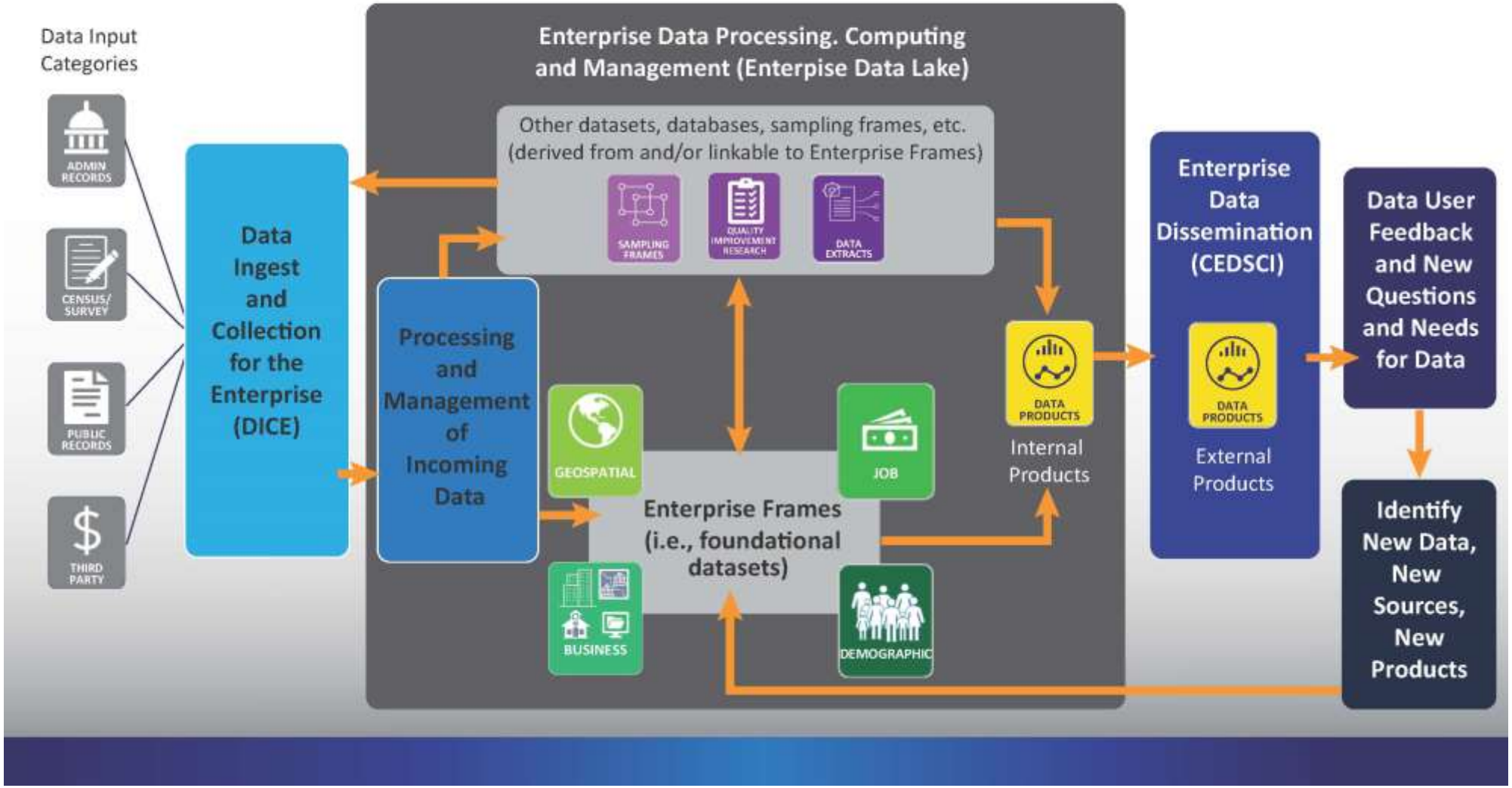
Frames Program
Vision

- *To create Enterprise-wide frames linkable in nature, agile in structure, accessible for production or research on a need-to-know basis, and that adhere to best practices in terms of technology usage, data management, and methodology.*
- Linkable in Nature
  - Each frame will include the necessary unique identifiers and keys for linkage to each other. For example:
    - Location information will be available on each person, job, and business record.
    - A person record may have linkage information for a job.
    - A job record may have linkage information to a business.

United States® Census Bureau

U.S. Census Bureau GEOGRAPHY

Creating a Business Ecosystem to Modernize the Census Bureau's Statistical Foundation



Frames Program
Benefits and Use Cases

Enumerate	Enumerate non-responding housing units.
Replace	Replace questions on surveys and reduce need to ask for information. For example, use of parcel data and other geospatial data to augment or replace ACS questions about acreage and year built.
Update	Use information in one frame to update information in another. For example, information in the Business Register about businesses providing residential services and facilities can be used to update group quarters in the Master Address File.
Facilitate	Facilitate analyses of people, households, businesses, or locations across time and over space. Ability to link to the Job Frames to pull information about a person's job and/or place of work.
Enrich	Enrich reporting on demographic and economic changes in a particular geographic area or areas resulting from demographic and/or economic processes.

United States® Census Bureau

U.S. Census Bureau GEOGRAPHY

Linked Data Use Case: Did a government business incentive program reduce poverty in selected neighborhoods?

- Questions to be answered through analysis:
  - Which businesses relocated to the neighborhood or opened new establishments in response to incentives?
  - Did the business(es) hire workers from within the neighborhood? When were they hired?
  - Where do workers reside? How many reside within the neighborhood? How many commute from other locations?
  - Did existing businesses hire additional workers? How have businesses' revenues changed?
  - How many new businesses were established to meet needs of workers?
  - Where do individual residents of the community spend their money? Are dollars staying in the community?
  - If the poverty rate declined and median income increased, was that due to gains on the part of longer-term households or due to higher income households moving in (i.e., gentrification)?

Geospatial Data:
- Spatial areas in question
- Addresses
- Parcel and structure data

Demographic Data:
- Characteristics of residents and workers
- Residential locations of workers over time

Business Data:
- Date of establishment
- Length of time at location
- Payroll
- Revenue

Jobs Data:
- Workers linked to establishments
- Employment history
- Commuting patterns

Assessors' Data & Real Estate Data:
- Housing value (assessed and market)
- Tenure
- Monthly rent

Building Permit Survey & Survey of Construction:
- New housing
- Commercial development

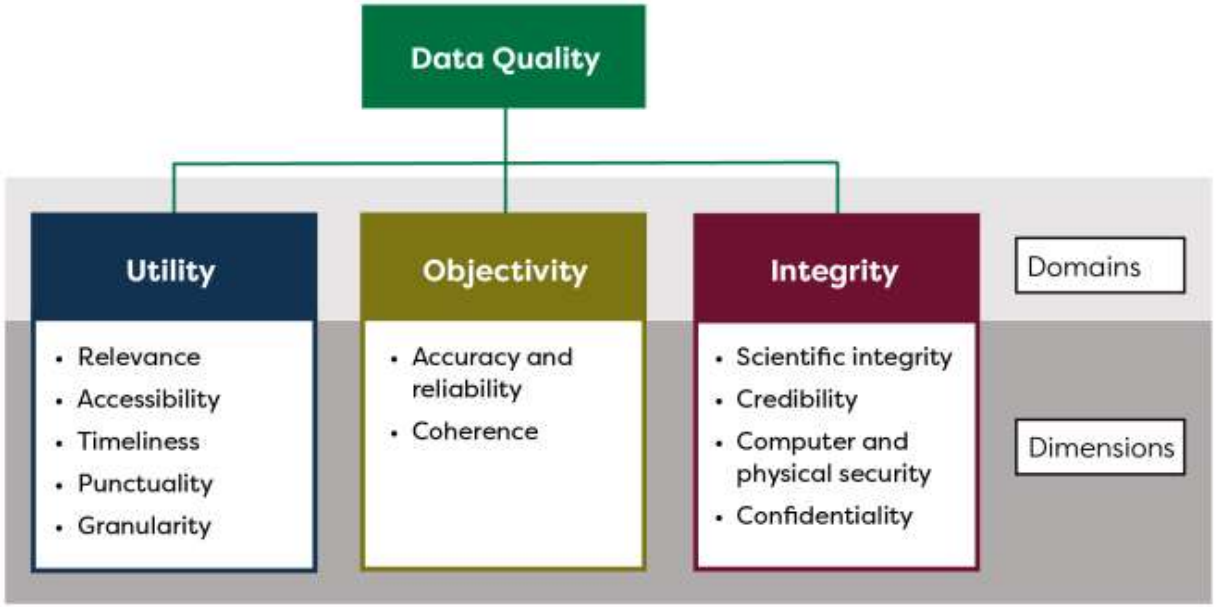
Point of Purchase Data:
- Where do residents spend money?
- Individual consumers

United States® Census Bureau

U.S. Census Bureau GEOGRAPHY

8

The domains and dimensions in the FCSM's Framework for Data Quality intersect with FAIR principles. Incorporating these into metadata records will improve search and discovery as well as assessment of fitness for use.



Federal Committee on Statistical Methodology. 2020. A Framework for Data Quality. FCSM 20-04. Federal Committee on Statistical Methodology. September 2020. FCSM-20-04 A Framework for Data Quality

Domain	Dimension	Definition
Utility	Relevance	Relevance refers to whether the data product is targeted to meet current and prospective user needs.
	Accessibility	Accessibility relates to the ease with which data users can obtain an agency's products and documentation in forms and formats that are understandable to data users.
	Timeliness	Timeliness is the length of time between the event or phenomenon the data describe and their availability.
	Punctuality	Punctuality is measured as the time lag between the actual release of the data and the planned target date for data release.
	Granularity	Granularity refers to the amount of disaggregation available for key data elements. Granularity can be expressed in units of time, level of geographic detail available, or the amount of detail available on any of a number of characteristics (*e.g.* (demographic, socio-economic).
Objectivity	Accuracy and reliability	Accuracy measures the closeness of an estimate from a data product to its true value. Reliability, a related concept, characterizes the consistency of results when the same phenomenon is measured or estimated more than once under similar conditions.
	Coherence	Coherence is defined as the ability of the data product to maintain common definitions, classification, and methodological processes, to align with external statistical standards, and to maintain consistency and comparability with other relevant data.
Integrity	Scientific integrity	Scientific integrity refers to an environment that ensures adherence to scientific standards and use of established scientific methods to produce and disseminate objective data products and one that shields these products from inappropriate political influence.
	Credibility	Credibility characterizes the confidence that users place in data products based simply on the qualifications and past performance of the data producer.
	Computer and physical security	Computer and physical security of data refers to the protection of information throughout the collection, production, analysis, and development process from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.
	Confidentiality	Confidentiality refers to a quality or condition of information as an obligation not to disclose that information to an unauthorized party.

United States® Census Bureau

Thank You!

Michael Ratcliffe

Senior Advisor for Frames

Geography Division

Michael.r.Ratcliffe@census.gov