

# Defining and Assessing Relevance of Data and Information Products

---

PANEL PRESENTATION TO FEDERAL COMMITTEE OF STATISTICAL  
METHODOLOGY

OCTOBER 23, 2024



# Agenda

---

1. Welcome and Introductions
  - Sharon Boivin (NCSES)
  - Grace Kena (BJS)
  - Jacob Malcom (DoI)
  - Rolf Schmitt (BTS)
  - Lexi Shankster (NCES)
2. Current Efforts to Define and Assess Relevance
3. Proposed Framework
4. Panel Questions and Discussion

# Definitions of Relevance

---

the quality or state of being closely connected or appropriate.

Example: "this film has contemporary relevance" (Oxford Dictionary)

*Relevance* is a comparative property: the more positive cognitive effects and the less processing effort, the more relevant the utterance. (Sperber and Wilson, 1995)

The meaning of "relevance" in U.S. law is reflected in Rule 401 of the [Federal Rules of Evidence](#). That rule defines relevance as "having any tendency to make the existence of any fact that is of consequence to the determinations of the action more probable or less probable than it would be without the evidence."

*Relevance* means measuring processes, activities, and things that matter to policy makers, and public and private sector data users (Principles and Practices for a Federal Statistical Agency, 2013)

# Making Relevance

---

My data is more relevant when:

- the evidence bang for my audience's cognitive buck is increased

- It is easy to access, process and use

- My data is important to choices under consideration

# Proposal: Two Types of Relevance to Consider

---

**Contextual**: Beginning with use case or action/decision, identify relevant data needed

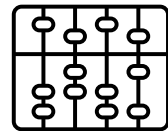
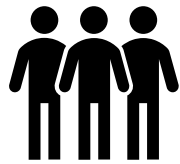
- Identify decision-makers and action-takers
- Identify decision set, action set  
Identify data needed to inform decision/actions

**Absolute**: Beginning with data, make data more relevant to unspecified audiences, actions/decisions

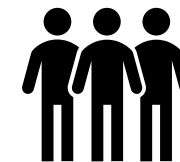
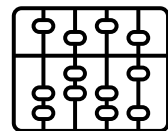
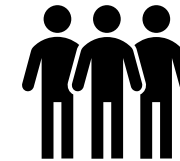
- Improve reusability of data
- Improve accessibility of data
- Lower the cognitive effort needed to interpret data (plain language, documentation, visualizations,...others?)

# Need a specific audience for contextual relevance

---



Data Collection  
and Product  
Design



FAIR principles



# Who is the specific audience?

---

Proposal: Relevance of data collections/products shall be assessed with respect to the following primary stakeholders:

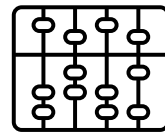
1. Federal decision-makers (program administrators, legislators, President)
2. **Data providers for the given data and/or data product** (for NCES data, for example, this could include students, educational staff, educational institutions, districts and states depending on the product)
3. **Specific audiences or use cases** identified in legislation authorizing data collection

# What about everyone else?

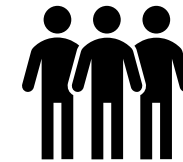
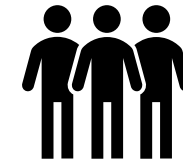
---

Relevance of data collections and products shall be assessed with respect to the “general public” through application of the FAIR principles

*Findable*  
*Accessible*  
*Interoperable*  
*Reusable*



FAIR principles





# Proposal: Apply M 23-22 Principles to Relevance

---

OMB Memorandum M23 -22 identifies best practices for improving user experience such as:

- testing digital products for accessibility
- ensuring content is authoritative and easy to understand
- using plain language
- writing and testing content for the *intended audience*
- starting with users' wants and needs, designing for users

# Implications

---

This framework implies we need the following:

Specific use cases for data collections and data products in different decision-making environments

Robust channels for feedback from data providers and users, program administrators, and legislators

- User testing with data providers on contextual relevance
  - What decisions will/could data providers make with this data?
  - What decisions will/could be made that affect the data providers using this data?

User testing with data users on absolute relevance

- Is the documentation clear to data users outside of the primary use case(s)?

Also implies relevance cannot be defined in absolute terms

- Cannot say “we scored a 98% on the relevance dimension”
- We need to communicate with data users to understand a dataset’s relevance in their context

# Unresolved questions

---

How do we assess our portfolio for relevance rather than individual products?

- What are we **not** creating/disseminating that is of importance to decision-makers?

How do we support the information ecosystem in a way that provides legislators important insights?

- When do we design for researchers versus legislators, for example?

How do we forecast relevant information and use cases?

- Example: the onset of COVID made HVAC systems in schools of interest to policymakers