



Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY

WELCOME!

A decorative horizontal line with a small circle on the left and a small square on the right.

2023 VIRTUAL RESEARCH SYMPOSIUM

**Showcasing Progress: Data and Evidence
for Better Government**

A decorative graphic on the left side of the slide consisting of several concentric, wavy lines in shades of pink and orange, resembling a fingerprint or a stylized wave pattern.



JESSICA YABSLEY

Senior Director of Communications,
Data Foundation



Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY



@data_foundation / #DataLive



MOLLY IRWIN, PH.D.

Member, Data Foundation Board of Directors



Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY



@data_foundation / #DataLive



MARY TSCHIRHART, PH.D.

Director, Trachtenberg School of Public Policy and
Public Administration, The George Washington
University



Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY



Morning Agenda

9:00 AM

Opening Remarks

9:20 AM / Making the Case for Data Quality

Group

10:20 AM

Break

**10:25 AM / The Time is Now: People Need
Access to Data**

Group

**11:25 AM / Building Data Capacity on
International Worker's Rights Programs**

Individual

11:40 AM

Lunch



Making the Case for Data Quality

Lisa Mirel

Statistical Advisor, National Center for Science and Engineering Statistics, National Science Foundation

Erika Liliedahl

Senior Evidence Analyst, Office of Management and Budget

Darius Singpurwalla

Mathematical Statistician, National Center for Science and Engineering Statistics, National Science Foundation



[@data_foundation](#) / [#DataLive](#)



Trachtenberg School
of Public Policy
& Public Administration
THE GEORGE WASHINGTON UNIVERSITY

Question Poll

Describe your role with data
(select all that apply)

- Data producer of blended data
- Data producer of survey data
- Use data to inform research in academia or the private sector
- Use data to inform government funded research





Disclaimer: The views expressed in this presentation are those of the author and do not necessarily reflect the views of the National Center for Science and Engineering Statistics or the National Science Foundation

Making the Case for Data Quality

Lisa B. Mirel

May 24, 2023

Data Foundation Symposium

National Center for Science and Engineering Statistics

Social, Behavioral and Economic Sciences

National Science Foundation

Question Poll

Have you ever heard of the Data Quality Framework from the Federal Committee on Statistical Methodology (FCSM)?



Background



Understanding data quality is essential for data-driven decision-making

Data users who understand the “fitness-for-use” of data products are more likely to use them appropriately

Higher-impact uses of data require higher-quality data



All data have strengths and weaknesses



Data quality for surveys is relatively well-established, but data quality for integrated data and other nonstatistical data are less developed

FCSM Framework for Data Quality

Builds on experience of the Federal Statistical System



Organizes the elements of data quality around the structure of the Information Quality Act

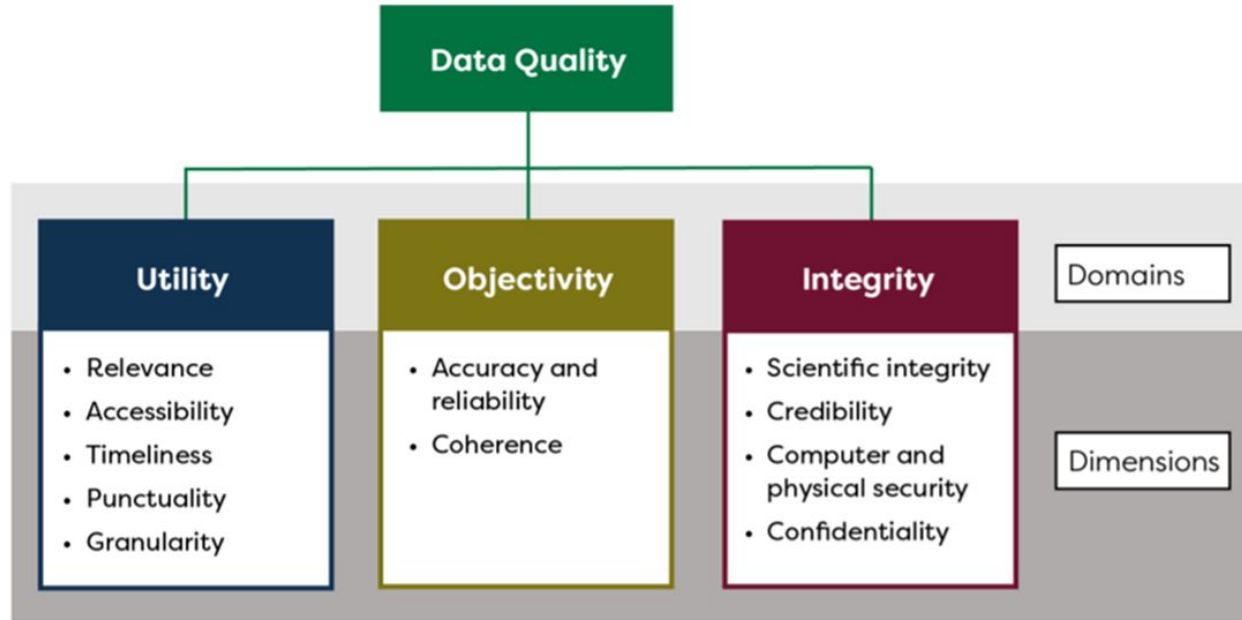


Explains for a broad audience the importance of understanding data quality to determine fitness for purpose, identifying and mitigating key data quality threats, and evaluating trade-offs



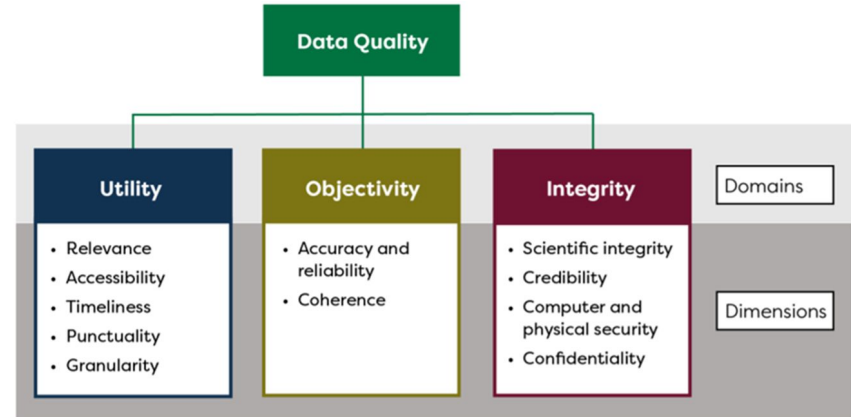
Provides strategies for documenting and reporting data quality in a systematic and consistent way

FCSM Framework for Data Quality

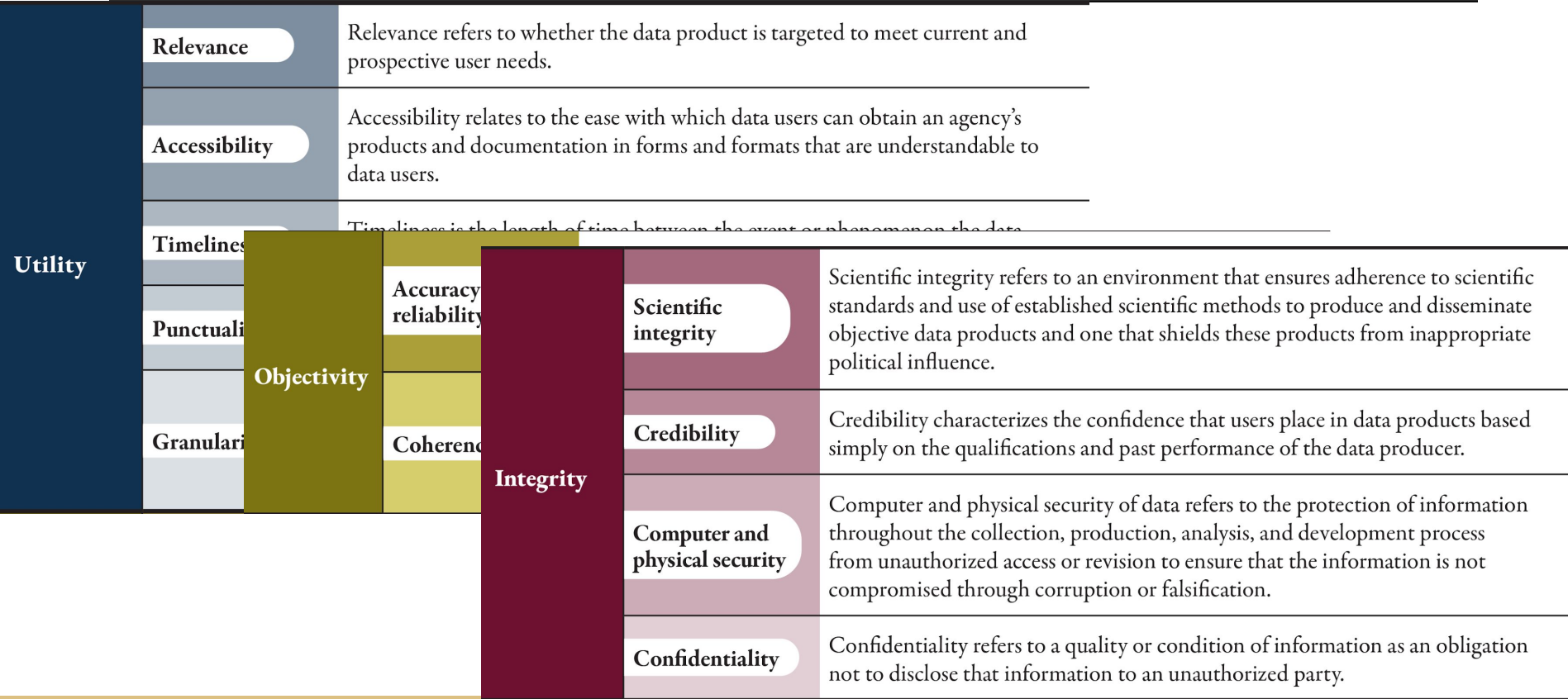


Domains of Data Quality

- **Utility:** The extent to which information is well-targeted to valuable needs; reflects the usefulness of the information to the intended users
- **Objectivity:** Whether information is accurate, reliable, and unbiased and is presented in an accurate, clear and interpretable, and unbiased manner
- **Integrity:** The maintenance of rigorous scientific standards and the protection of information from manipulation or influence as well as unauthorized access or revision

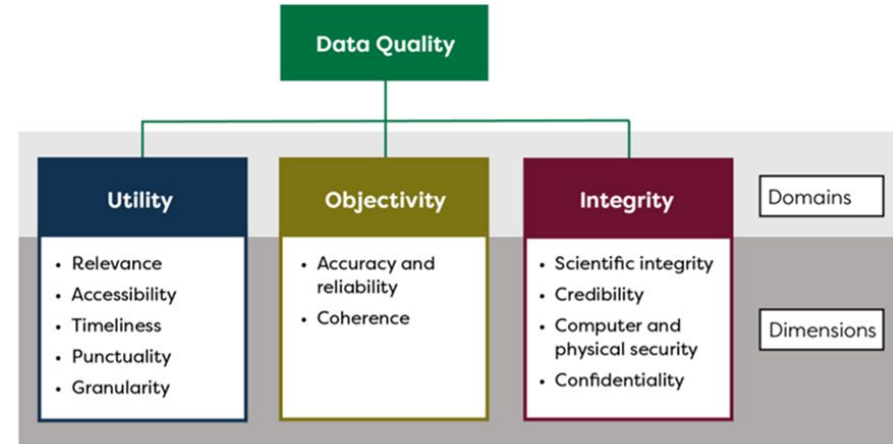


Dimensions of Data Quality



Question Poll

Think of a data resource that you use often, which dimension(s) of the framework do you wish you had more information on or was better represented in the data resource documentation?



Implementing the Framework

- Looks overwhelming but many data quality threats can be mitigated or dismissed after brief consideration for a data program
- There are few universal rules for weighing importance of one data quality concern over another: tradeoffs are expected
- Documenting a data source's strengths and weaknesses as it relates to the intended use is a good habit that helps your successors and supports transparency

Key Takeaways



All data have problems, but do the problems matter for the decision at hand?



It is important to consider all possible data quality problems, deal with problems that can reasonably be addressed, and document how each problem is dealt with (so successors have a record)

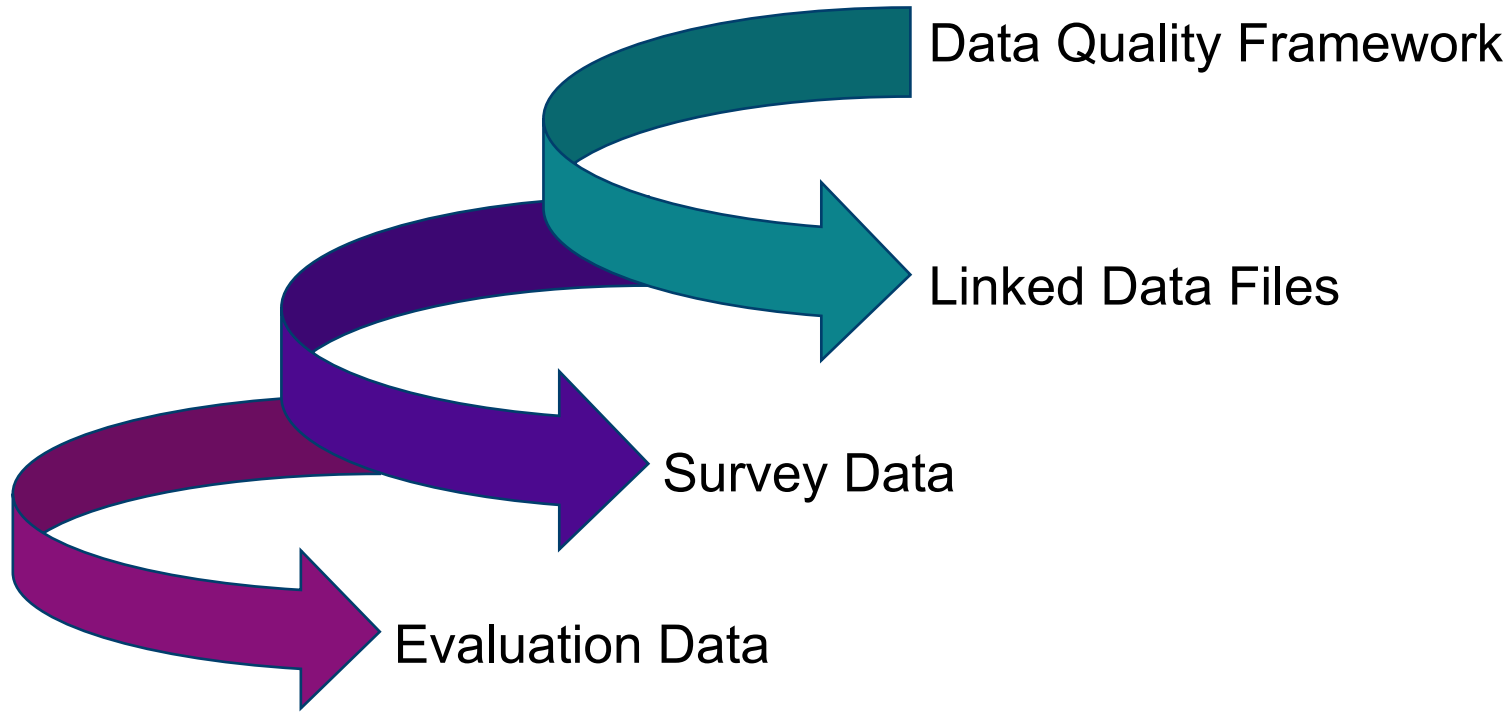


Include data quality in guides for power users, and summarize the problems for an elevator speech to tell occasional users how far they can take the data without misleading decisions that have important consequences

Conclusion

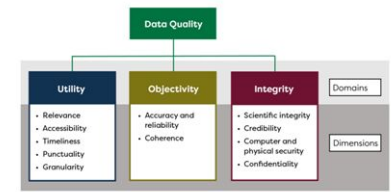
- By using the structure and terminology of the Framework, we will have a common basis for sharing information about data quality across agencies and with the public
- A common language will support transparency about data and analyses and will provide a common basis for considering improvements in data and analysis

Three Case Studies



Case Study: Linked Data Files

Relevance



Linked Mortality Files Citation List as of 2/16/2023¹

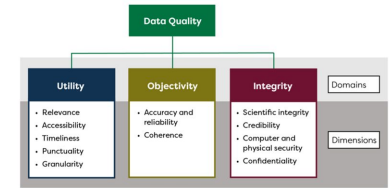
- [The National Health and Nutrition Examination Survey \(NHANES\)](#)
- [The National Health and Nutrition Examination Survey II \(NHANES II\)](#)
- [The National Health and Nutrition Examination Survey III \(NHANES III\)](#)
- [The National Health Interview Survey \(NHIS\)](#)
- [The Second Longitudinal Study of Aging \(LSOA II\)](#)

The National Health and Nutrition Examination Survey (NHANES)

2023

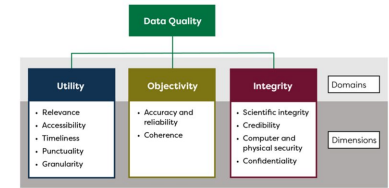
1. Zhu, X., et al., *Associations of Serum Carotenoids With Risk of All-Cause and Cardiovascular Mortality in Hypertensive Adults*. *J Am Heart Assoc*, 2023. **12**(4): p. e027568.
2. Zhang, Y., et al., *Associations of different isomeric forms of serum lycopene with cardiovascular-disease and all-cause mortality*. *Int J Vitam Nutr Res*, 2023.
3. Zhang, N., et al., *Associations of Dietary Zinc-Vitamin B6 Ratio with All-Cause Mortality and Cardiovascular Disease Mortality Based on National Health and Nutrition Examination Survey 1999-2016*. *Nutrients*, 2023. **15**(2).
4. Zhang, J., et al., *Prognostic nutritional index as a risk factor for diabetic kidney disease and mortality in patients with type 2 diabetes mellitus*. *Acta Diabetol*, 2023. **60**(2): p. 235-245.
5. Zhang, D., et al., *Low muscle mass is associated with a higher risk of all-cause and cardiovascular disease-specific mortality in cancer survivors*. *Nutrition*, 2023. **107**: p. 111934.

Accessibility



File details	Restricted use	Partially synthetic public use
Dates	Exact date of death, birth date, and interview date	Date of death represented by quarter/year or person months of follow up, depending on survey
Cause of death	Detailed underlying and multiple code of death information	Most common underlying cause of deaths and two indicators for multiple cause of death, diabetes and hypertension
Participants	Both adults and children	Only adults
Perturbation	No perturbation	Perturbed information for cause of death or follow up time for select decedents; vital status is not perturbed

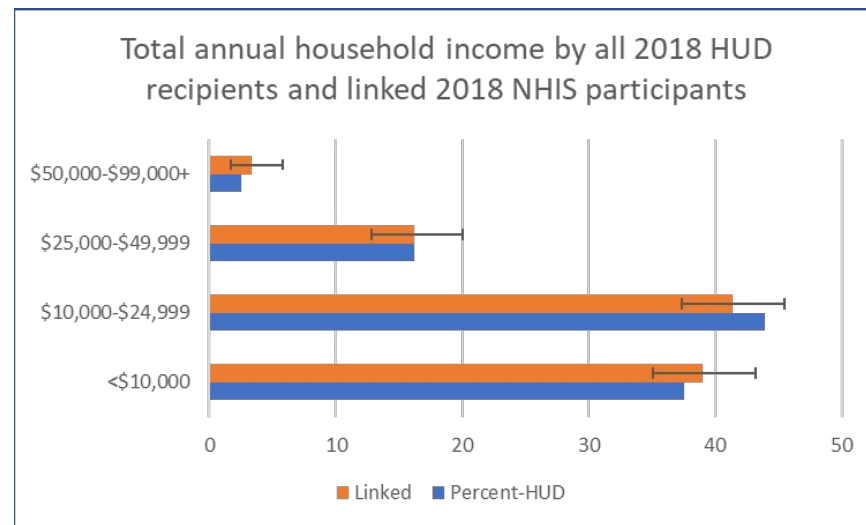
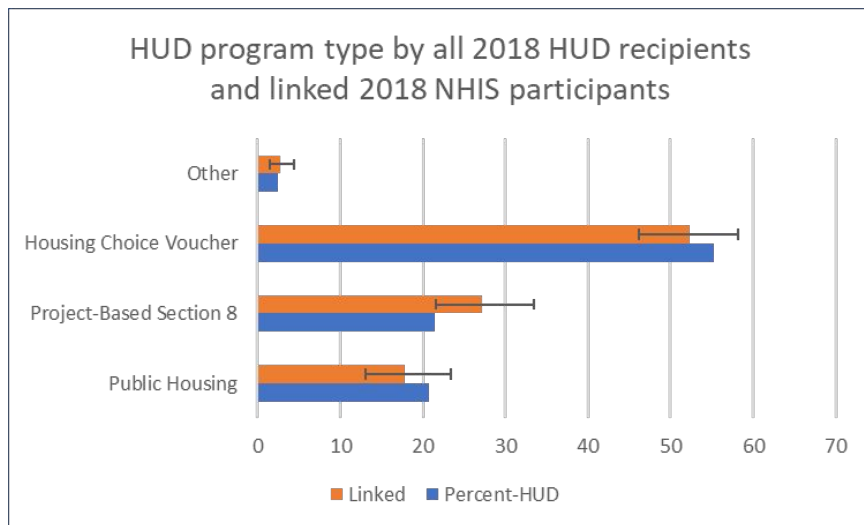
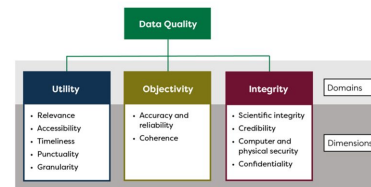
Timeliness



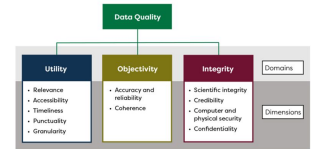
NCHS Survey	Linked National Death Index data	Linked Medicare data	Linked Medicaid data	Linked End Stage Renal Disease (ESRD) data	Linked Department of Housing and Urban Development (HUD) data	Linked Department of Veteran Affairs (VA) data
National Health Interview Survey (NHIS)						
NHIS 1986-1993	through 2019	***	***	***	***	***
NHIS 1994-1998	through 2019	1991-2018	2014-2019	through 2018	***	***
NHIS 1999-2004	through 2019	1999-2018	2014-2019	through 2018	through 2019	***
NHIS 2005-2014	through 2019	1999-2018	2014-2019	through 2018	through 2019	through 2020
NHIS 2015-2016	through 2019	2014-2018	2014-2019	through 2018	through 2019	through 2020
NHIS 2017-2018	through 2019	2014-2018	2014-2019	through 2018	through 2019	through 2020
National Health and Nutrition Examination Survey (NHANES)						
NHANES 1999-2004	through 2019	1999-2018	2014-2019	through 2018	through 2019	***
NHANES 2005-2012	through 2019	1999-2018	2014-2019	through 2018	through 2019	through 2020
NHANES 2013-2016	through 2019	2014-2018	2014-2019	through 2018	through 2019	through 2020
NHANES 2017-2018	through 2019	2014-2018	2014-2019	through 2018	through 2019	through 2020
NHANES I Epidemiologic Follow up Survey (NHEFS)	through 2019	1991-2013	***	through 2018	***	***
NHANES II	through 2015	***	***	***	***	***
NHANES III	through 2019	1991-2018	2014-2019	through 2018	***	***
National Hospital Care Survey (NHCS)						
NHCS 2014	2014-2015	2014-2015	***	***	2013-2015	***
NHCS 2016	2016-2017	2016-2017	2015-2017	***	2015-2017	***

Note: 1999-2014 linked NCHS-CMS Medicaid data (in MAX format) are available from a previous linkage for some of the surveys listed
 *** Survey not included in linkage. Linked data not available.

Accuracy and Reliability: Benchmarking

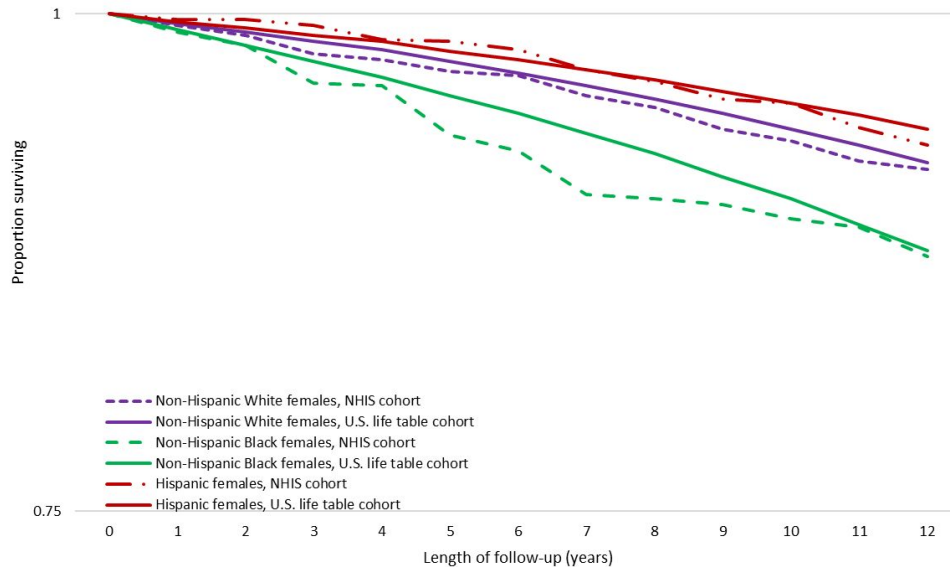


NHIS: National Health Interview Survey
 HUD: Department of Housing and Urban Development
 Source: 2018 NHIS linked to 2018 HUD administrative data



Integrity/Coherence: Example from the NHIS Linked Mortality Files

Survival curves for females, aged 50-59 years, by race/ethnicity and sex:
2006 NHIS LMF and U.S. life table cohorts



NHIS: National Health Interview Survey; LMF: Linked Mortality Files

Compared life expectancy models for national and linked data populations

Alignment of estimates support robust analyses using the linked data

Keralis JM, et al. *A comparison of the mortality experience of U.S. adults estimated with the 2006–2018 National Health Interview Survey Linked Mortality Files and the annual U.S. life tables*. National Health Statistics Reports; no 186. Hyattsville, MD: National Center for Health Statistics. 2023. DOI: <https://dx.doi.org/10.15620/cdc.126565>.

Scientific Integrity: Concordance

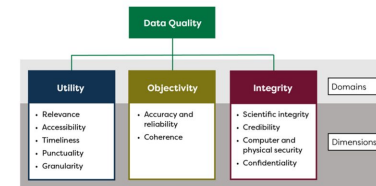


Table 1. Comparison of National Health Interview Survey report to T-MSIS record of Medicaid/CHIP coverage, 2016-2018*

NHIS report of Medicaid/CHIP coverage	T-MSIS record of full-scope or comprehensive Medicaid/CHIP coverage in same month of NHIS interview		
	Yes	No	Total
Yes	14,177	2,187	16,364
No	4,216	74,209	78,425
Total	18,393	76,396	94,789

Overall agreement: 93.2%

Sensitivity: 77.1%

Specificity: 97.1%

Positive predictive value: 86.6%

Negative predictive value: 94.6%

NOTES: NHIS is National Health Interview Survey. T-MSIS is Transformed Medicaid Statistical Information System. CHIP is Children's Health Insurance Program.

SOURCE: Linked National Health Interview Survey and CMS Medicaid data

*Results do not include 17-year-olds



Contact: lbmirel@nsf.gov





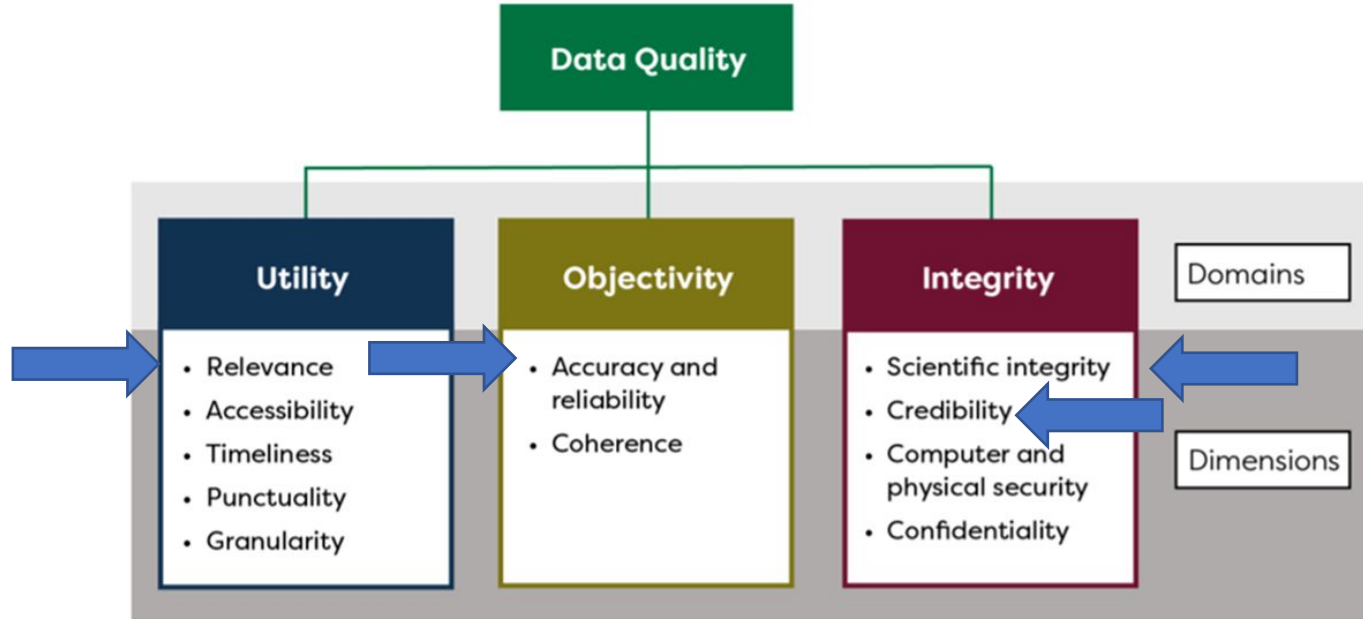
Disclaimer: The views expressed in this presentation are those of the author and do not necessarily reflect the views of the National Center for Science and Engineering Statistics or the National Science Foundation

Case Study: Survey Data

- Darius Singpurwalla
 - May 24, 2023
 - Data Foundation Symposium
-

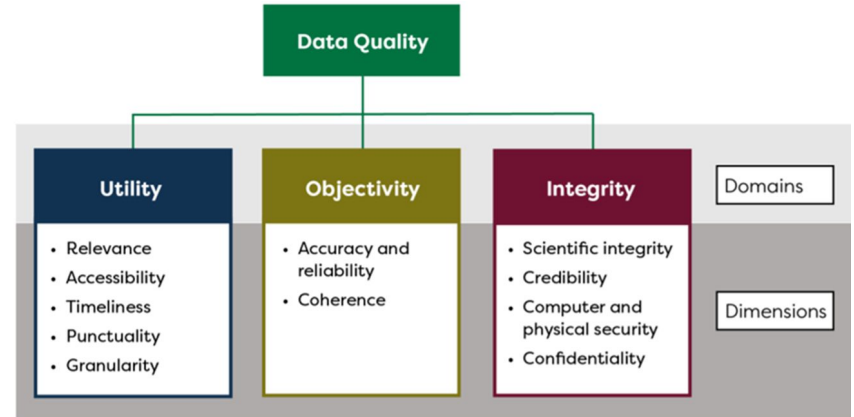
National Center for Science and Engineering Statistics
Social, Behavioral and Economic Sciences
National Science Foundation

FCSM Framework for Data Quality



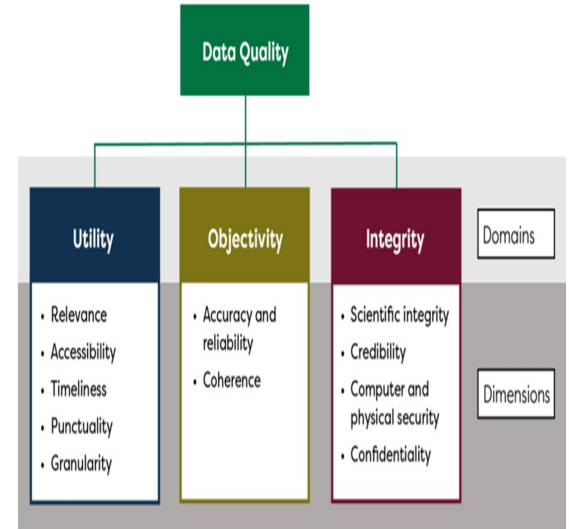
Domains/Dimensions Salient to a New Collection

- **Utility / Relevance**
 - Does data product meet current and prospective user needs?
- **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 under similar conditions.
- **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
- **Integrity / Credibility**
 - Credibility characterizes the confidence that users place in data products based simply on the qualifications and past performance of the data producers.



Project Milestones

Survey Steps	Suggested FCSM Data Quality Dimension
Survey Basics	Credibility (Integrity)
Define Research Objectives	Relevance
Identify Target Population	Accuracy
Survey Method	Scientific Integrity
Identify and Define Constructs for Measurement	Relevance
Literature Review	Relevance
Previous SEI chapters	
Secondary Analysis of Data	
Operationalize Research Questions	
Revise and Finalize Survey	Accuracy and Reliability
Structure the Survey	
Cognitive Test/Pre-test Questionnaire	Accuracy
Field Survey	
Analyze Results from Survey	



Scientific Knowledge Questions

Question Option 1

- The 'Big Bang Theory' is a theory that describes how the universe was created.
 - True
 - False
 - I don't know

Question Option 2

- What does the Big Bang Theory tell us?
 - How the universe began
 - The state of the Los Angeles higher education during the 2010's.
 - The entire universe was once condensed in the form of a primeval atom, or a dense mass.

Comparing Quality of Different Sources

*Data quality assessments are vital to understanding the fitness for **purpose** of a certain resource.*

Because the framework is designed to cover all forms of data, some components may not be relevant **(or of less importance)** to a specific data collection, estimation method, or analysis. Specific quality issues and **tradeoffs** among quality challenges vary by data source and application.

- Utility

- Relevance (++)
- Accessibility (+)
- Timeliness (-)

- Objectivity

- Accuracy and Reliability (++)

- Integrity

- Scientific integrity (+)
- Credibility (+)
- Confidentiality (-)

- Utility

- Relevance (+)
- Accessibility (+)
- Timeliness (++)

- Objectivity

- Accuracy and Reliability (-)

- Integrity

- Scientific integrity (+)
- Credibility (++)
- Confidentiality (++)

Data Quality Framework: Evaluation Case Study

**Erika Liliedahl, Evidence Team
Office of Management and Budget**



Program Evaluation

- An assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency
- May address:
 - questions related to the implementation of a program, policy, or organization;
 - the effectiveness of specific strategies related to or used by a program, policy, or organization; and/or
 - factors that relate to variability in the effectiveness of a program, policy, or organization or strategies of these.
- Can examine questions related to understanding the contextual factors surrounding a program or how to effectively target specific populations or groups for a particular intervention

Sources: Evidence Act--5 U.S.C. § 311(3), OMB M-19-23, OMB M-20-12, OMB Circular A-11

Why a Data Quality Case Study on Program Evaluation?

- Federal agencies often face limited resources for program evaluation
- Evaluators have gotten entrepreneurial in identifying data sets that will support answering research questions of interest
- Agencies often rely on administrative data, that is already collected by the government for a different purpose, to aggregate participant outcomes and estimate program impacts.
- *Case Study*: How would the FCSM data quality framework apply to a program evaluation that uses a pre-existing data source?

Selecting a Program Evaluation Case

- In response to the Evidence Act, agencies now have Annual Evaluation Plans, available on [evaluation.gov](https://www.evaluation.gov)
- Annual Evaluation Plans include research questions, proposed evaluation methods and approaches, proposed data, and more
- Agencies then proceed with proposed activities, if funding and other resources are available
- Many evaluations to choose from to consider for a case study!

Program Evaluation Case Study

- Quasi-experimental impact analysis using administrative data to evaluate the Department of Labor (DOL) Transition Assistance Program (TAP)
- Quarterly wage data contained within the National Directory for New Hires (NDNH) database was used to measure group employment outcomes and aggregate program impacts
- The study sample included transitioning service members from the Army branch who completed a Department of Labor Employment Workshop between October 1, 2014 and June 30, 2019 prior to transitioning from their military service to civilian life.
- Select data within the NDNH was linked with Army data to develop the “treatment” and “control” groups (through matched comparison)
- *Case Study*: How would the FCSM data quality framework apply to the use of NDNH for the TAP program evaluation?

Data Considerations for Evaluation

Typical Evaluation Considerations	FCSM Data Quality Framework	
	Dimensions	Domains
A description of the data and their origin	<ul style="list-style-type: none"> Relevance 	Utility
The universe from which the data are collected and any known coverage issues (both population and amounts)	<ul style="list-style-type: none"> Relevance Granularity 	Utility
Information included (for example, the availability and quality of information)	<ul style="list-style-type: none"> Accuracy and reliability Credibility 	Objectivity Integrity
Options for data linkage	<ul style="list-style-type: none"> Granularity 	Utility
Procedures and practices surrounding access	<ul style="list-style-type: none"> Accessibility Computer and Physical Security Confidentiality 	Utility Integrity
Timeliness of the data—from collection to availability and from application to receipt	<ul style="list-style-type: none"> Timeliness 	Utility
Key strengths and limitations for analysis	<ul style="list-style-type: none"> Any dimension identified as a key strength and limitation 	Utility, Objectivity, and Integrity

Accessibility—Utility Domain

- Access to data contained within the National Directory of New Hires (NDNH) database is explicitly limited by statute
 - Title IV-D of the Social Security Act specifies that researchers may be given access only to de-identified NDNH information to conduct research found by the Secretary of Health and Human Services (HHS) to be likely to contribute to achieving the purposes of part A or part D of the Social Security Act (42 U.S.C. §653(j)(5)), that is, to contribute to achieving the mission of Temporary Assistance for Needy Families and Child Support Enforcement programs
- Users must:
 - Sign an agreement or memorandum of understanding that describes the purpose, legal authority, justification, expected results of the match, description of the records, retention and disposition of information, reimbursement, and user's performance reporting requirements
 - Have a security addendum in place that details the security requirements and safeguards that users must have in place before receiving NDNH information
 - Reimburse OCSE for the costs of obtaining, verifying, maintaining, and comparing the information
 - Provide identifiers for the study sample and then HHS provides de-identified results based on the specifications by group requested

Accuracy & Reliability—Objectivity Domain

- Quarterly wage and unemployment insurance data contained in NDNH is derived from the state UI agencies so susceptible to the same coverage and underreporting issues (nothing different or special to state's version)
- Excludes self-employed individuals, most independent contractors, railroad employees, some part-time employees of nonprofit institutions, employees of religious orders, and some students employed by their schools
- Does not include employment data for people without valid Social Security numbers. However, in the case of this study, SSNs were available.
- An advantage over NDNH and other earnings data sources is that it includes those employed by federal or military agencies
- Does not have information on job characteristics (such as industry, occupation, and hours worked) which limits analysis and understanding
- Is not useful as a historical data source or to examine a long time horizon, as it contains only up to two years of data at the time of application for access.
 - For program evaluation this may present critical issues when establishing outcomes at different time periods, including the base period

Scientific Integrity—Integrity Domain

- Although the NDNH provides a rich dataset, it does not allow the development of any type of dataset (restricted or public use) to allow for additional research or for re-analysis by others not included in an approved and active agreement
- The restricted access and de-identified files provided to researchers limit the potential uses and analysis during the evaluation, such as
 - Limiting the sophistication of model specifications for a program evaluation
 - Prohibiting the ability to correct problems with data linkages after the de-identified file with NDNH data has been returned
- As the files returned are de-identified, it is impossible for researchers to later incorporate additional years of data or link to new sources of data

Lessons from Evaluation Case Study



- What is documented about data sets is important and instrumental in determining whether a data set is a good fit for a particular program evaluation, for example:
 - Will the data allow for analysis of a sufficient sample to detect program impacts?
 - Will the data facilitate subgroup analysis on important groups?
- The FCSM data quality framework provides a consistent framework to discuss the difficult yet realistic data quality balance for program evaluation, such as:
 - Balancing timeliness vs data accuracy
 - Determining how to measure outcomes in meaningful ways
 - Abilities, authorities, and capabilities needed to obtain data
- Important contextual information on the data is critical to understanding the data and implications of use
 - What data cleaning or validation (e.g., missing data imputations, elimination of obvious data reporting errors or inconsistencies, etc.) would be necessary in order to ready data for analysis on the outcomes of interest?

Lessons from Evaluation Case Study



- It is crucial that owners of data are transparent as possible about data quality so that data users, including program evaluators, can make informed choices
- It is imperative that researchers are transparent about the fitness for use of existing data for program evaluation, to strengthen credibility of study design and findings, and to account for other factors including data privacy.
- The FCSM data quality framework can serve as a valuable tool to determine the utility of administrative data in determining program effectiveness and provides considerations for the data's use that are paramount to rigor and ethics as principles of program evaluation.
- FCSM data quality framework also gives data collectors, data creators, and data users a common language to discuss data quality

Polls: We want your input!



- Which case study resonated the most with you and the work you are doing?
 - Blended data
 - Survey design
 - Program evaluation
- Do you see yourself utilizing the Data Quality Framework in your work?
- What other lessons do you think will surface from using the Data Quality Framework?

Discussion



- What are particular challenges you are encountering when linking data and assessing data quality?
- How can data quality assessments be prioritized to ensure high-quality linkages?
- How can these examples best be applied at your organizations?
- Other questions or comments?

Contact Information



- lbmirel@nsf.gov
- dsingpur@nsf.gov
- eliliedahl@omb.eop.gov



Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY

WE'RE TAKING A BREAK

—

EVENT WILL RESUME SHORTLY



@data_foundation / #DataLive



The Time is Now: People Need Access to Data

Mary Ann Miller Bates

Executive Director, California
Cradle to Career Data System

Paige Kowalski

Executive Vice President,
Data Quality Campaign
(moderator)

Michael Vente,

Chief Performance Officer,
Senior Director of Research
and Data Governance,
Colorado Department of
Higher Education



@data_foundation / #DataLive



Trachtenberg School
of Public Policy
& Public Administration
THE GEORGE WASHINGTON UNIVERSITY



Building Data Capacity on International Worker's Rights Programs

Jennifer Werth

International Relations Officer,
Bureau of International Labor
Affairs (ILAB), Department of
Labor

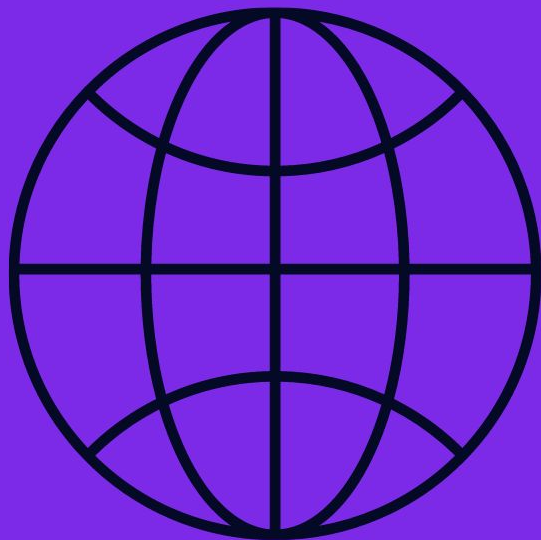


Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY



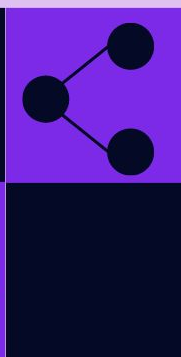
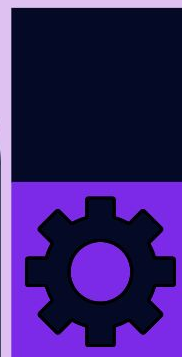
@data_foundation / #DataLive



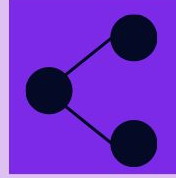
Jennifer Werth

USDOL Bureau of International Labor Affairs, Office of Trade and Labor Affairs

Building Data Capacity in International Worker's Rights Programs



For 20 years



The Office of Trade and Labor Affairs had a robust technical assistance and cooperation (TAC) portfolio and a robust policy team.

But no monitoring and evaluation team to support their work.



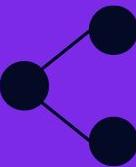
May 2021: M&E is established to support the whole office



tasked with systematically examining the office's data to determine "what works and why?"

“How many projects have we funded and where were they?”

A seemingly simple question the new M&E team was tasked with answering






Where was the data?

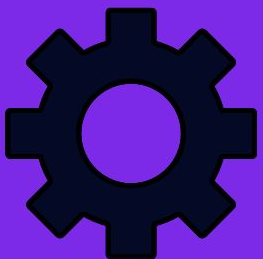
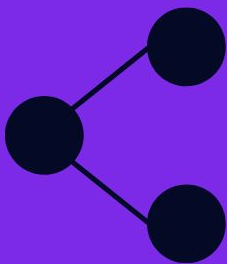
20 years worth
of information
spread across
11 or more
spreadsheets of
various quality +
project files

no
comprehensive
listing of all the
projects funded
by the office, no
common metrics
across projects

build the office's
grant history to
explain its
outcomes from
scratch and
explain our story



Participatory Process



Consolidation

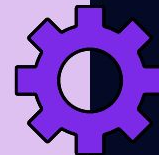
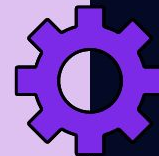
- Get information into one central data set, develop the fields to determine how we organize the data. Going from active projects to all projects.

Listening Sessions

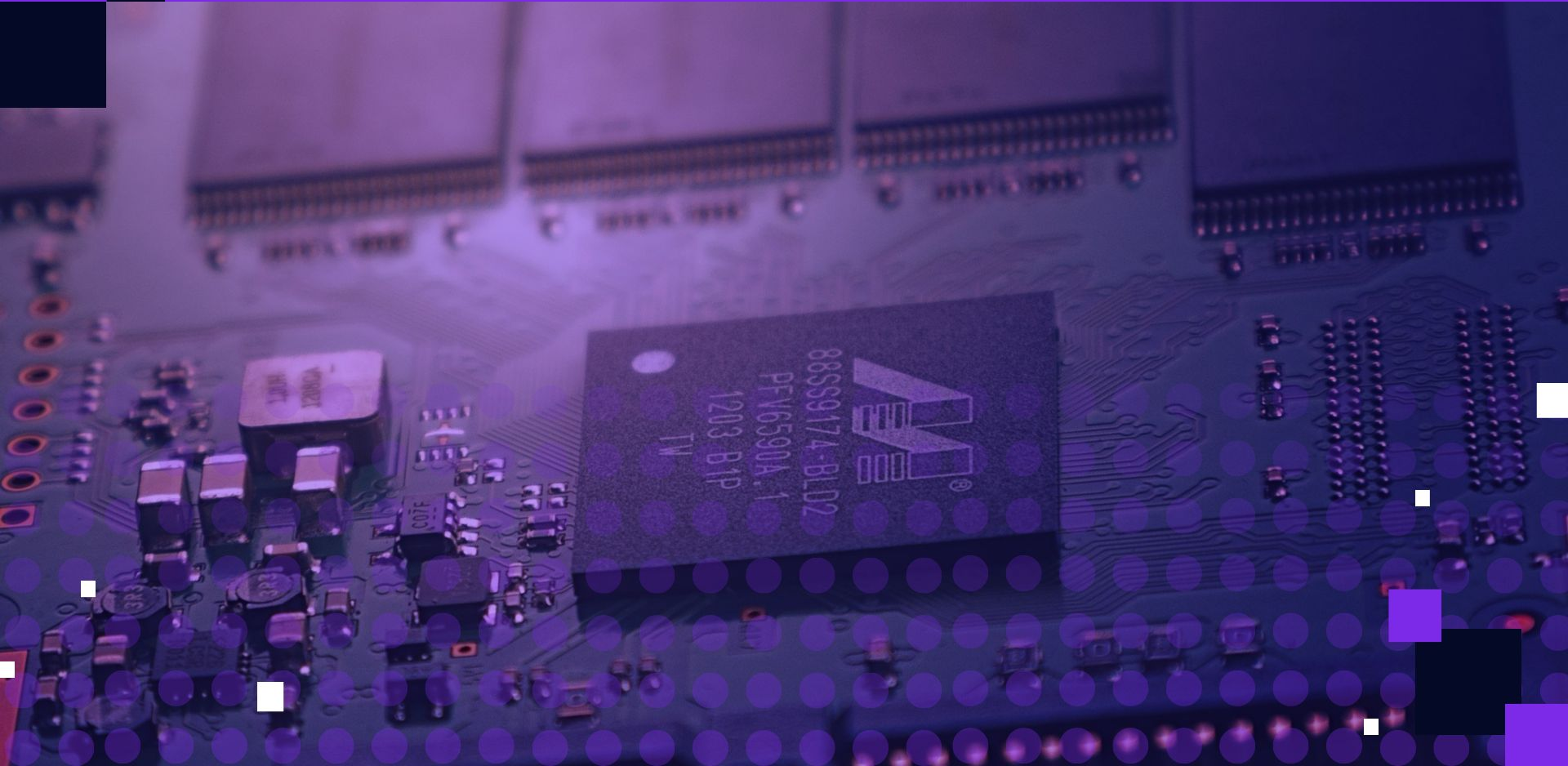
- Discuss what kind of information teams would like to see (TAC, policy team, stakeholders, management). How can what we create be responsive to their needs?

Visualization

- Used co-creation, design thinking, and ideation to create a dashboard that was useful to different groups. Piloted different aspects of the tool.



Result: Interactive Dashboard



Dataset & Dashboard Impact



On Leadership

Can better discuss trends and lessons learned, Greater understanding of how technical assistance fits into the work.



On Technical Assistance Team

Ability to respond to data calls, improving productivity. Sense of ownership and better storytelling.

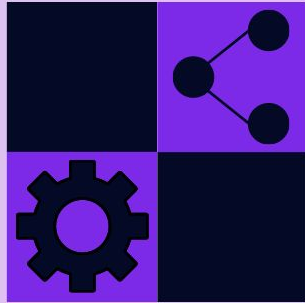


On ILAB

Shifting towards evidence based culture in our office, impacting our sister office. Raising awareness for the importance of data



Data as a strategic asset



Theory of Sustained Change

Prioritizing complexity aware methods.

Standardize & Aggregate

Comparability of data across projects.

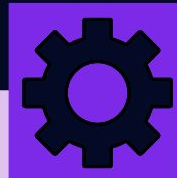
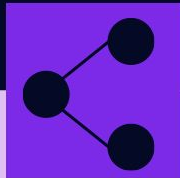
Metrics for Enforcement

Beginning to measure trade policy work.

Evaluation Techniques

More rigorous evaluations addressing new questions, including equity questions and thematic questions

Thank you!



Contact Me For Questions or
More Information:
werth.jennifer.k@dol.gov

[Visit ILAB's MEL Website](#)



Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY

LUNCH BREAK

   @data_foundation / #DataLive



Afternoon Agenda

12:00 PM

Welcome Back Remarks

12:05 PM / Opportunities to Strengthen State and Local Data and Evaluation Capacity in Human Services

Group

12:50 PM

Break

12:55 PM / Evidence in Action: Using Data to Support Educator Decision-Making

Group

1:55 PM / Better Metrics to Guide Public Health Policy: Lessons Learned From COVID-19 for Public Health Data Systems Improvement

Individual

2:10 PM

Break

2:15 PM / Using the Criminal Justice Administrative Records System to Improve Administration of the Criminal Justice System

Group

3:15 PM / Analysis of Alternative Poverty Measures Applied to the Case of Montana

Individual

3:30 PM / From Evidence Generation to Better Decisions: MCC's Evidence Lifecycle

Group

4:30 PM

Closing Remarks



Trachtenberg School
of Public Policy
& Public Administration
THE GEORGE WASHINGTON UNIVERSITY



@data_foundation /
#DataLive



Opportunities to Strengthen State and Local Data and Evaluation Capacity in Human Services

Kathy Stack

Senior Policy Fellow, Tobin Center
for Economic Policy

Jonathan Womer

Senior Advisor, The Policy Lab at
Brown University



Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY



@data_foundation / #DataLive

Resourcing State and Local Integrated Data and Evaluation

Kathy Stack kstack9117@gmail.com

Jonathan Womer Jonathan_womer@brown.edu

2023 Research Symposium - Data and Evidence for Better Government
May 24, 2023



BROWN
The Policy Lab

Understanding LA Systems That Affect Families

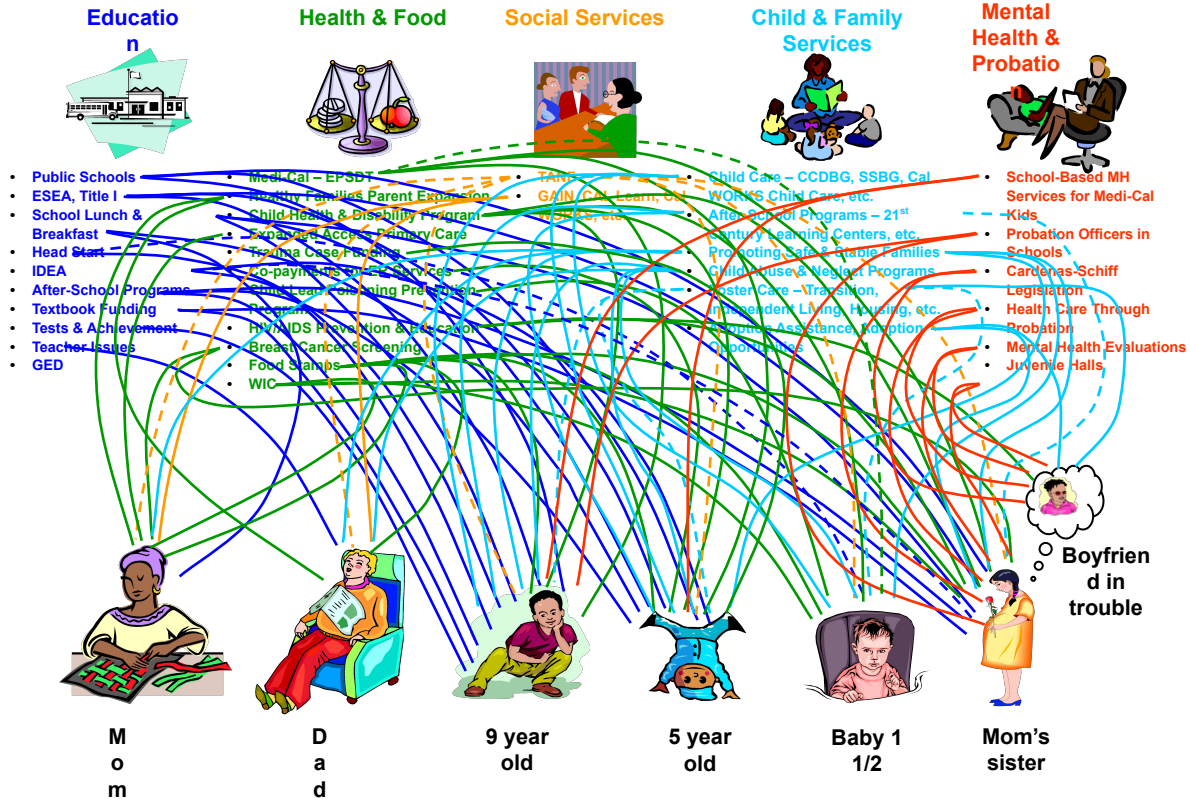
A Look at How 40+ Programs Might Touch One Los Angeles Family

Margaret Dunkle

The George Washington University
& The LA County Children's Planning Council

2002

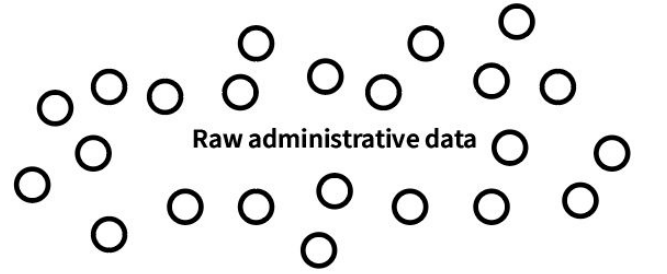
The case for
integrated
data



Data to Information in an integrated data system

Not very expensive

Less aggregated



More sensitive



De-identified administrative data

Data or program dashboard

Public transparency portal

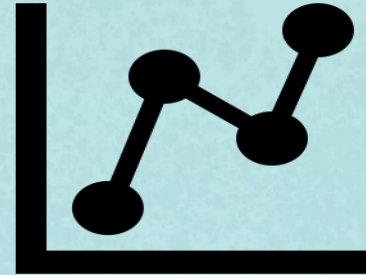
More aggregated

Less sensitive

State Integrated Data System (rough) Statistics

50 state and DC scan take-aways from survey of websites and reports. 84 identified.

- Median State - SLDS
- California has 6
- At least 19 have Medicaid data
- Location
 - 33 Education Agencies
 - 16 Universities
 - 16 Central Service Agencies
 - 11 HHS Agencies
 - 7 Other Government Agencies



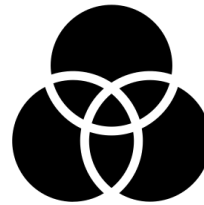
Support Structures

Federal Stimulus and Evidence Based Policymaking

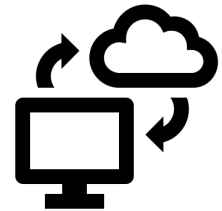
State and Local Fiscal Recovery Fund (SLFRF); Infrastructure Investment and Jobs Act (IIJA) eligible use



Program Evaluation and Evidence: develop learning agendas, select evidence-based interventions and conduct program evaluations and use evidence clearinghouses.



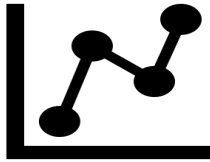
Data and Analysis: gather assess and use data for effective policymaking and taking of program performance.



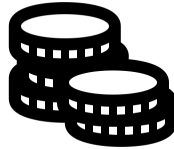
Capacity-Building: hiring staff, academics and consultants with expertise in evaluation, data, technology and data management systems.

Foundations for Evidence-Based Policymaking Act of 2018

OMB [M-21-27](#), June 30, 2021, Evidence-Based Policymaking: Learning Agendas and Annual Evaluation Plans. (Also in [M-22-12](#) page 7.)



Evaluation is Broad: Per section 101(a) of the Evidence Act, “[t]he term ‘evaluation’ means an assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency.”



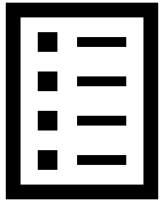
Allowable Cost: “. . . evaluation costs are allowable costs (either as direct or indirect) of Federal awards . . .” Page 16 of M-21-27.



Evaluation Policies: Federal agency evaluation policies must apply to “recipients of Federal awards that are performing work on behalf of the agency.” Page 12 of M-21-27

Standing Federal Funding Streams

Influences the application of data and evaluation



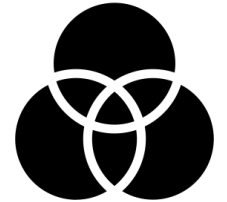
Project Funding: Most common approach. Not great at funding fixed costs so often supplemented with state appropriations.



Medicaid and HHS: Can get 90% federal match. Emphasis on day-to-day improvement of case management.



P-20+ / Workforce SLDS: Long standing research on the effects of education on employment. Emphasis on researchers and formal evaluations.



Statewide Cost Allocation: Treats data as a utility service like telecommunication. Emphasis on public view of the data.

Building a State or Local integrated data system

It isn't just
funding ...



Leverage existing systems: Tax, Medicaid management, audit and fraud, and state longitudinal data systems are good places to start.



Data Governance: Programs own the data and privacy rules must be navigated. *Data flows at the speed of trust.*



Existing Communities: Universities, NGOs, GIS, revenue and caseload forecasting, labor market information

State and Local Examples

Medicaid and HHS

Emphasis on
improving case
management and
understanding
clients to improve
policy



- **Washington State's Integrated Client Databases:** Includes data from 30+ state agency data systems. About 50 analytic projects are ongoing at any one time. Also supports some core agency functions including care management, fraud detection and caseload forecasting. Uses significant state appropriations with HHS funds.
- **Allegheny County Data Warehouse:** Research projects Often involve a standing university partnership group composed of local universities (Carlow University, Carnegie Mellon University, Duquesne University, University of Pittsburgh, and Robert Morris University).
- **Both:** Staff and service providers receive alerts when clients experience life events such as births, legal charges or school truancy. These features, combined with predictive analytics, enable staff to make better decisions for clients.

P-20+ / Workforce SLDS

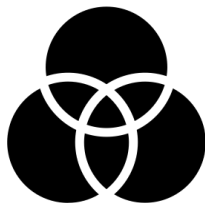
Historically
focused on the
effects of
education and
training on
employment



- **Kentucky Center for Statistics (KYSTATS):** KYSTATS develops a biennial research agenda, with input from each partner agencies. On an annual basis, it publishes about 45-50 reports per year, fulfill around 300 data requests.
- **Georgia Information Tunnel:** The Georgia “tunnel” links data from a single SLDS directly to district-level student information systems, allowing district administrators, principals, teachers, and parents to access state education data through their district’s existing program. (Data Quality Campaign Example, December 2022)
- **California College Guidance Initiative (CCGI):** CCGI works with individual K–12 districts and California’s higher education systems to allow students, families, and educators, allowing them to make informed decisions about college and career pathways. (Data Quality Campaign Example, December 2022)

Statewide Integrated Data Systems

Treats data as a utility service and emphasizes public view of the data.



- **State Coverage:** Nine states (Delaware, Indiana, Massachusetts, Michigan, New Jersey, North Carolina, Ohio, South Carolina, and Virginia) appear to manage integrated data systems that provide data and analytics for state agencies across major social service, health, education, criminal justice and workforce programs.
- **Indiana Management Performance Hub (MPH):** MPH focuses on improving government operations and decision-making, economic development and public transparency. Inside Indiana OMB and started with executive order. Number of university partnerships and a researcher residency program.
- **InnovateOhio Platform:** This IDS works to boost public transparency, improve economic development, facilitate research, provide operational program decision support. Inside of Ohio Department of Administration and started with executive order. State data kept in Ohio's secured data-sharing platform for analysis and research. Funded in part through cost allocation.

Federal Opportunities to do More

Federal Opportunities for Encouragement

Possible Administrative Actions



- **Set expectation that grantees strengthen data and evaluation capacity** with existing funds and create financial incentives to do so.
- **Clarify permissible ways to combine funds** for IDS, including central service cost allocation.
- **Issue regulations on Presumption of Use** per the Advisory Committee on Data for Evidence Building (ACDEB)
- **Improve Technical Assistance** per the Advisory Committee on Data for Evidence Building (ACDEB)
- **Create an Intergovernmental Data Working Group:** Seek shared solutions to common barriers including intergovernmental data-sharing, standard data-use agreements, and shared procurement vehicles.
- **Modernize grants guidance to reflect cloud-based technology and services**

Federal Opportunities for Encouragement

Potential Congressional Actions



- **NEED Act Expansion of SLDS Grant Program**
- **New Programs: *Block Grant to States*** (ACDEB Recommendation) and ***LINC Act Grants*** on Health and Social Services Data
- **Repurposing Unused Grant Funds**
- **Fund National Secure Data Service Projects with State and Local Governments**

Thoughts?

thepolicylab@brown.edu



THANK YOU TO OUR PARTNER!

Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY



@data_foundation / #DataLive



Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY

THANK YOU FOR TUNING IN TO OUR VIRTUAL RESEARCH SYMPOSIUM!

Learn more at datafoundation.org, or follow
us on social media @data_foundation



@data_foundation / #DataLive