United States Department of Agriculture, Economic Research Service

USDA's Economic Research Service

Data Product Quality Standards and Processes

FCSM Research and Policy Conference
October 25, 2022

Katherine Ralston

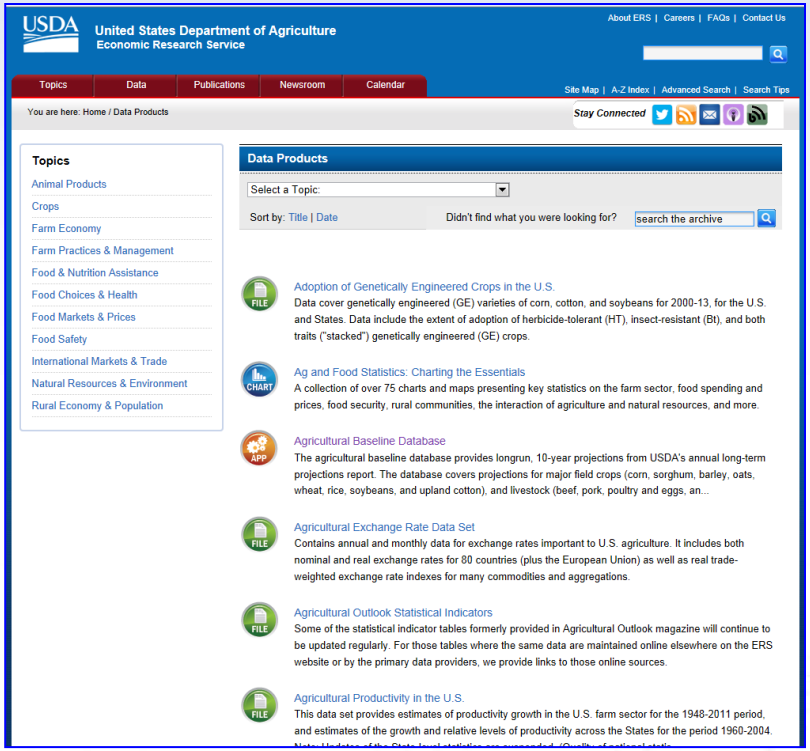
The findings and conclusions in this presentation are those of the author and should not be construed to represent any official USDA or U.S. Government determination or policy."

ERS Mission

Inform and enhance public and private decision making on economic and policy issues related to agriculture, food, the environment, and rural development.

Outline

- ERS's data product quality standards
- Data product review process
- Keys to success



www.ers.usda.gov/data-products.aspx

More Than 70 Data Products

www.ers.usda.gov/data-products.aspx

Types of ERS Data Products

Data Products include transformations of 'raw survey data' used as a research database and/or published to the ERS website, as well as development of data sets collected under ERS sponsorship or compiled from diverse sources where ERS adds value in the form of recompilation and subject-matter expertise.

Primary Data	Model Results	Summary Statistics	Repackaged Data
• ARMS • FoodAPS • Scanner data (best example of big data)	• Food Dollar • Commodity Supply and Use	• Foreign Agricultural Trade of the U.S.	• State Fact Sheets • Food Environment • Land use/ quality research data

Audiences and Data Users

- **Policymakers and their staffs**
- **Policy influencers**
- **Government**
- **Agribusiness professionals**
- Academics/researchers
- News media
- Informed lay people
- Developers and digital professionals

ERS Data Product Review Committee Charge (2012)

Develop a data policy for core, high-value statistics:

- Consistent with OMB guidance
- Reflect best practices at ERS, USDA, and other principal statistical agencies to ensure objectivity and quality

Develop a hierarchy of ERS data and information systems that recognize differences in quality, usability, purpose, transparency, and utility:

- Attributes can be mapped to measureable characteristics
- Recommended practices and policies will recognize the various types of data produced.
- Framework will form the basis for cost-benefit analysis to ensure efficient management of data resources.

Formulation of Data Product Quality Standards

Considered OMB Directives, Policies, and Guidance:

- Statistical Policy Directive Number 3
- Statistical Policy Directive Number 4
- Standards and Guidelines for Statistical Surveys
- Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies
- OMB Circular A-130
- Open Data Policy-Managing Information as an Asset

Emulated best practices from other Federal Statistical Agencies:

- BTS Statistical Standards Manual
- BEA Information Quality Guidelines
- NASS
- National Center for Science and Engineering Statistics/NSF

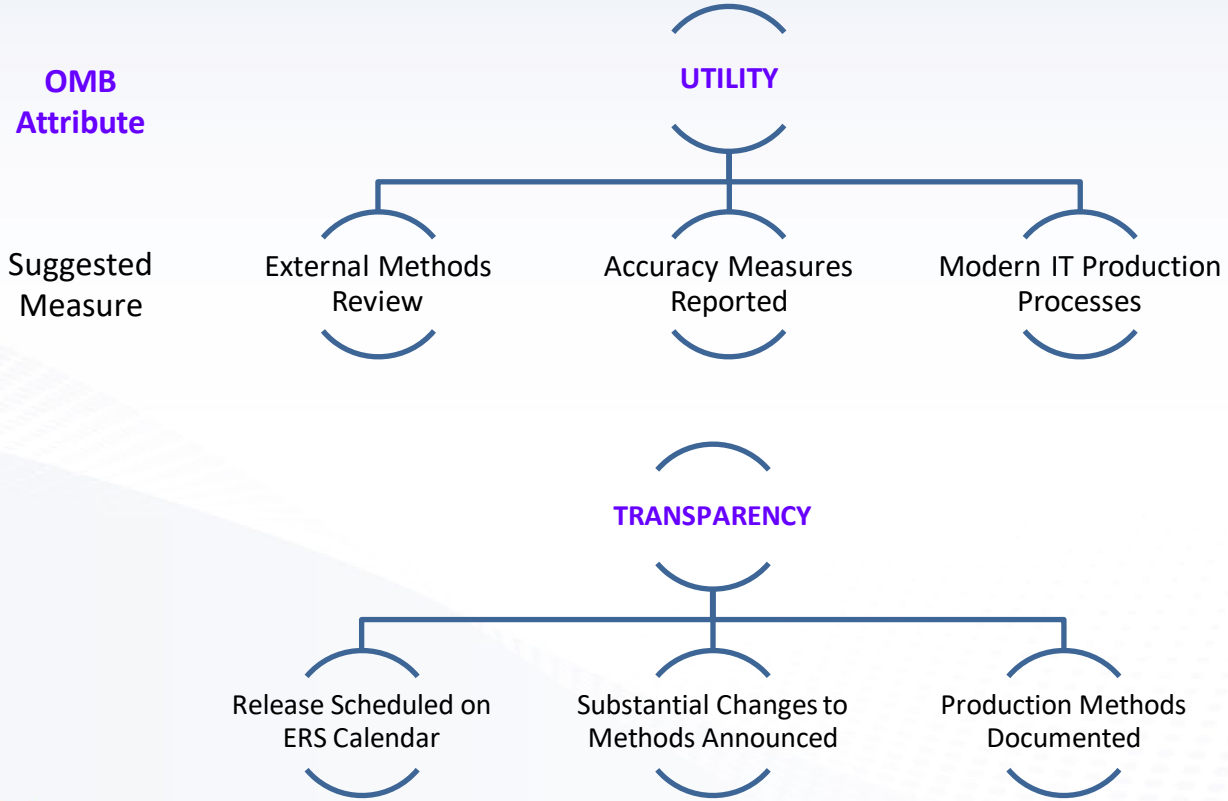
Hierarchy: Three Tiers of Data Products

Criteria developed to rank data products in terms of adherence with the OMB and USDA definitions of ***Influential*** Scientific, Financial, or Statistical Information and quality guidelines, as well as the importance to the agency's mission:

- **Premier data products:** influential and central to the agency's mission, as well as adhering to all components of quality guidelines as applicable.
- **Core data products:** central to the agency's mission but do not meet the criteria for influential. Includes foundational data that serve as inputs to premier data products.
- **Other data products:** serve key agency stakeholders and the public.

For Each OMB Attribute, Committee Developed Measurable Standards That Can Be Applied to Different Types and Tiers of ERS Data

Attributes: Purpose, Utility, Objectivity, Transparency, Integrity, and Accessibility



Data Product Quality Standards

- Operationalize OMB Directives
- Categories:
  - Purpose/MOUs
  - Utility
  - Objectivity
  - Transparency
  - Integrity
  - Accessibility
- 25 Standards (36 checklist items) https://www.ers.usda.gov/about-ers/policies-and-standards/data-product-quality/ers-data-product-quality-standards/

Purpose

1.1 MOU for data provided to others

1.2 MOU for data received from others

Utility

2.1 Preeminence and pertinence

2.2 Branding and data sources

2.3 Future releases

2.4 Stakeholders

2.5 Quality of communication

2.6 Feedback

2.7 Web Stats

Objectivity

3.1 Data quality review procedure

3.2 External review of methods

3.3 Accuracy measures reported

3.4 Research on Methods and Operations

3.5 IT investment

Transparency

4.1 Public notice of changes

4.2 Good documentation

4.3 Explanation of similar data elsewhere

4.4 Update and revision history available

4.5 Archival capability

Integrity

5.1 Disclosure protection

5.2 Physical security

5.3 Staff training

Accessibility

6.1 Machine readable Open format Section 508

6.2 Usability testing

6.3 Metadata

Initial Evaluation of Data Products Against Data Quality Attributes

				UTILITY						QUALITY			TRANSPARENCY			PURPOSE	INTEGRITY / OBJECTIVITY			USABILITY		
Data Product	Division	Owner	Tier	Contact	Update Frequency	Sole Source	ERS Web Usage*	Impact	Product Type	Relevancy Evaluation	External Methods Review	Modern IT Processes?	Accuracy Evaluation	Meta data	ERS Calendar	Methods Documentation	Mandate Used in USD.	Pre-Dissemination Review	Record Archival	Personnel Training	Accessible Output Formats	Open Data Compliant
Food Availability (Per Capita) Data System	FED		1	Jeanine Bentley	Annual (May)	Yes	24,428	High	nonparam calc	No	No	Yes	No	Yes	No	Yes	NO	Yes	Yes	No	XLS	
Food Dollar Series	FED		1	Patrick Canning	Annual (Feb)	Yes	7,304	High	modeling	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	XLS/PDF	
Food Price Outlook	FED		1	Richard Volpe	Monthly	Yes	45,527	High	modeling	No	No	Yes	Not recently	Yes	No	Yes	NO	Yes	No	Yes	No	XLS	
Food Security in the United States	FED		1	Alisha Coleman-Jensen	Annual (Sept)	Yes	7,575	High	primary stats	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	XLS/PDF	
Price Spreads from Farm to Consumer	FED		1	Hayden Stewart	Annual (Nov)	Yes	5,707	High	modeling	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	XLS/PDF	
Quarterly Food-at-Home Price Database	FED		1	Jessica Todd	Annual (Apr)	Yes	3,480	High	nonparam calc	No	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	No	XLS/PDF	
Food Environment Atlas	FED / ISD		1	Paula Dutko	Annual (Nov)	Yes	66,700	High	compilation of data	Yes	N/A	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	XLS/GIS	
Food Desert Locator	FED / ISD		1	Michele Ver Ploeg	Dec 2012 or Jan 2013	Yes	151,351	High	nonparam calc	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	XLS/GIS	
Agricultural Baseline Database	MTED		1	Paul Westcott	Annual (February)	No	43,979	High	modeling, judgment-based analyses		??	No	??	yes	yes	Yes	yes	yes	yes		html/xls/pdf	
Agricultural Trade Multipliers	MTED		1	Sarah Parsons	Annual	maybe	3,533	High	modeling		no	some		Yes		?		yes			XLS	
Commodity Costs and Returns	MTED		1	William McBride	4 times /yr (May, Oct;	Yes	38,684	High	primary data, summary stats	2006-07	Yes	SAS	??	yes	yes	Yes	yes	yes	yes		XLS	
Foreign Agricultural Trade of the United States (FATUS)	MTED		1	Stephen MacDonald	monthly	No	31,388	High	primary summary stats		no	intermediate	??	yes	yes	yes	no	??			XLS	
International Baseline Data	MTED		1	Ronald Trostle	annual (spring)	yes?	6,002	high	modeling, judgment-based analyses		??	no	??	yes	no	yes	yes	yes	yes		XLS	
U.S. Bioenergy Statistics	MTED		1	Thomas Capehart	Monthly	No	3,014	medium	compilation of data		no	no			no	yes	no	?			XLS	
World Ag Supply & Demand Estimates (WASDE)	MTED/WAOB		1	multiple staff	Monthly	yes	NA	high	nonparam calc; staff expertise	yes		no		yes	yes	yes	?	yes	yes		PDF/XML/XLS	
Agricultural Productivity in the U.S.	RRED		1	Eldon Ball	National=annual, States=3years	Yes	13,618	High	modeling & compilation of data	No	Yes	?	No	No	Yes	Yes	?	Yes	No	No	XLS/PDF	No
Farm Household Income and Characteristics	RRED		1	Mary Ahearn	3 times / year + annual	Yes	2,241	High	primary data, summary stats	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	XLS/PDF	No
Farm Income and Wealth Statistics	RRED		1	Mitch Morehart	3 times / year	Yes	25,367	High	modeling & compilation of data	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	XLS/PDF	No
ARMS Farm Financial and Crop Production Practices	RRED		1	Mitch Morehart	2 times / year	Yes	24,628	High	primary data; summary statistics	Yes	Yes	Yes	Yes	Partial	Yes	Partial	Yes	Yes	Partial	Yes	XLS/PDF	No
Population-Interaction Zones for Agriculture (PIZA)	ISD			Vince Breneman	Every 10 years	Yes	710	medium	summary stats, compilation of data	No	Yes	no	no	yes	no	yes	no	Yes	no	no	XLS/ArcGIS	No
Agricultural Exchange Rate Data Set	MTED		2	Mathew Shane	2 times a year?	maybe	4,616	??	nonparam calc??	baseline	??	no	??	??	no	yes	baseline	baseline	??	??	XLS	
Aquaculture Data	MTED		3	David Harvey	Monthly	no	10,686	??	summary stats, compilation of data	underway	no	yes	??	yes	yes	yes	??	yes	??	??	XLS/PDF	
Bilateral Fiber and Textile Trade	MTED		4	Stephen MacDonald	none	??	434	??	summary stats, compilation of data	when released	NA	NA	NA	??	no	yes	no	??	NA	NA	XLS	
China Agricultural and Economic Data	MTED		3	Fred Gale	Periodic	in English?	12,797	??	summary stats, compilation of data	??	NA	??	NA	yes	no	yes	no	no	no	no	XLS/PDF	
Commodity and Food Elasticities	MTED		4	James Hanson	none	no	10,184	low?	compilation of data	no	no	maybe	??	yes	no	yes	no	??	no	no	XLS/PDF	
Cotton and Wool Yearbook	MTED		2	Leslie Meyer	annual	yes	1,574	medium?	summary stats, compilation of data	underway**	TBD**	no	??	yes	yes	some	??	yes?	data in public domain	yes	no	XLS	
Dairy Data	MTED		2	Roger Hoskin	Monthly	no	8,712	medium?	summary stats, compilation of data	underway**	TBD**	some	??	yes	no	some	??	yes?	data in public domain	no	no	XLS	
Feed Grains Database	MTED		2	Thomas Capehart	Monthly	yes	166,278	high	summary stats, compilation of data	underway**	TBD**	yes	??	yes	yes	yes	??	yes	data in public domain	no	no	XLS/PDF	
Fruit and Tree Nut Data	MTED		2	Agnes Perez	Monthly	some	15,314	high	nonparam calc, sum stats, data com	underway**	TBD**	some	??	yes	yes	some	yes?	yes	??	some	no	XLS/PDF	
International Food Consumption Patterns	MTED		3	Andrew Muhammad	Periodic	no	5,534	medium?	summary stats, compilation of data	??	??	no	??	yes	no	yes	no	no	??	no	no	XLS	
International Macroeconomic Data Set	MTED		2	Mathew Shane	2 times a year?	no	18,186	medium?	summary stats, compilation of data	baseline	baseline	no	??	yes	no	yes	baseline	yes	yes, baseline	??	no	XLS	
Livestock & Meat Domestic Data	MTED		2	Kenneth Mathews	Monthly	no	21,736	medium?	summary stats, compilation of data	underway**	TBD**	no	??	yes	yes	some	??	yes?	data in public domain	??	no	XLS	
Livestock & Meat International Trade Data	MTED		2	Rachel Johnson	Monthly	yes	10,305	high	nonparam calc, sum stats, data com	underway**	TBD**	yes	??	yes	yes	yes	??	yes	data in public domain	yes	no	XLS/PDF	
Meat Price Spreads	MTED		2	William Hahn	Monthly	yes	31,940	high	nonparam calc, sum stats, data com	??	??	no	??	yes	yes	yes	no?	yes	??	??	no	XLS	
Milk Cost of Production Estimates	MTED		2	William McBride	Monthly	yes	8,036	high?	nonparam calc, sum stats, data com	??	??	yes	??	yes	yes	yes	yes?	yes?	??	??	no	XLS	
Oil Crops Yearbook	MTED		2	Mark Ash	annual	no	4,389	medium?	summary stats, compilation of data	underway**	TBD**	no	??	yes	yes	some	??	yes?	data in public domain	yes	no	XLS	
Photosanitary Regulation	MTED		3	Sarah Ferrier	annual??	no	838	medium?	summary stats, compilation of data	??	??	no	??	yes	no	yes	no	no	??	??	no	XLS	
Rice Chart Gallery	MTED		4	Nathan Childs	Monthly	no	1,638	low?	charts from monthly report	underway**	TBD**	no	??	yes	no	no	no	no	??	no	data in public domain	NA	NA	JPG	
Rice Yearbook	MTED		2	Nathan Childs	annual	no	3,116	medium?	summary stats, compilation of data	underway**	TBD**	no	??	yes	yes	some	??	yes?	data in public domain	yes	no	XLS	
Season-Average Price Forecasts	MTED		3	Linwood Hoffman	Monthly	yes	11,880	medium?	spreadsheet models	??	upon release	no	??	yes	yes	yes	no	no	??	no	yes	no	XLS	
State Export Data	MTED		2	Alberto Jurado	annual	no	5,856	medium?	nonparam calc, sum stats, data com	??	??	no	??	yes	yes	yes	no	no	??	??	no	XLS	
Sugar and Sweetener Yearbook Tables	MTED		2	Stephen Haley	Monthly	no	16,719	high?	summary stats, compilation of data	underway**	TBD**	no	??	yes	yes	some	??	yes?	data in public domain	no	no	XLS	
U.S. Food Imports	MTED		3	Alberto Jurado	annual??	no	3,607	medium?	summary stats, compilation of data	??	??	no	??	yes	no	yes	no	no	??	??	no	XLS	
Vegetables and Pulses Data	MTED		2	Suzanne Thornsbury	Monthly	yes	15,372	high	nonparam calc, sum stats, data com	underway**	TBD**	some	??	yes	yes	some	yes?	yes	??	yes	no	XLS	
Wheat Data	MTED		2	Gary Vocke	Monthly	no	25,671	medium?	summary stats, compilation of data	underway**	TBD**	yes	??	yes	yes	yes	no	yes?	data in public domain	no	no	XLS/PDF	
Farm Program Atlas	MTED/ISD		3	Anne Effland	Periodic	no	9,642	medium?	compilation of data; GIS display	yes	??	yes	??	yes	no	yes	no	yes	yes	??	no	JPG	
Agricultural Outlook Statistical Indicators	MULTIPLE			Maurice Landes	being discontinued	no	12,259		compilation of data	yes	no	no	no	NA	??	NA	NA	NA	NA	NA	NA		
Chart Gallery	MULTIPLE						29,564															
Charts of Note	MULTIPLE						NA															
Adoption of Genetically Engineered Crops in the U.S.	RRED		2	Jorge Fernandez-Cornejo	Annual	Yes	33,733	MEDIUM	primary data; summary statistics	No	No	?	No	No	Yes	Yes	No	?	Yes	No	No	XLS	No
Agricultural Research Funding in the Public and Private Sectors	RRED		2	Paul Heisey	Periodic	Yes	1,656	MEDIUM	compilation of data	No	No	?	No	No	No	Yes	No	?	Yes	No	No	XLS	No
Atlas of Rural and Small-Town America	RRED		2	John Cromartie	Periodic	Yes	21,708	MEDIUM	compilation of data	No	No	Yes	No	No	No	Yes	No	?	Yes	No	No	App/XLS	No
Commuting Zones and Labor Market Areas	RRED		2	Timothy Parker	2 Years	Yes	552	MEDIUM	compilation of data	No	No	Yes	No	No	No	Yes	No	?	Yes	No	No	XLS	No
County Typology Codes	RRED		2	Timothy Parker	6 years	Yes	5,087	MEDIUM	compilation of data	No	No	No	No	No	No	Yes	No	?	Yes	No	No	XLS	No
County-level Data Sets	RRED		3	Timothy Parker	Periodic	Partial	45,621	MEDIUM	compilation of data	No	No	Yes	No	No	No	Yes	No	?	Yes	No	No	XLS	No
Creative Class County Codes	RRED		3	Tim Wojan	Periodic	Yes	1,349	MEDIUM	compilation of data	No	No	No	No	No	No	Yes	No	?	Yes	No	No	XLS	No
Federal Funds	RRED		4	Richard Reeder	NONE	No	1,465	LOW	compilation of data	No	No	No	No	No	No	Yes	No	?	Yes	No	No	XLS	No
Fertilizer Imports/Exports	RRED		4	Wen-yuan Huang	NONE	No	7,035	LOW	compilation of data	No	No	No	No	No	No	Yes	No	?	Yes	No	No	XLS	No
Fertilizer Use and Price	RRED		4	Wen-yuan Huang	NONE	No	22,302	MEDIUM	compilation of data	No	No	No	No	No	No	Yes	No	?	Yes	No	No	XLS	No
Frontier and Remote Area Codes	RRED		3	John Cromartie	Periodic	Partial	1,876	MEDIUM	compilation of data	No	No	No	No	No	No	Yes	No	?	Yes	No	No	XLS	No
Major Land Uses	RRED		2	Cynthia Nickerson	5 Years	Partial	7,124	MEDIUM	compilation of data	Yes	No	No	No	No	No	Yes	No	Yes	Yes	No	No	XLS	No
Monthly Cash Receipts	RRED		3	Theodore Covey	Monthly	Yes	NA	MEDIUM	compilation of data	No	No	No	No	No	No	No	No	Yes	Yes	No	No	XLS	No
Natural Amenities Scale	RRED		3	David McGranahan	Periodic	Yes	1,462	MEDIUM	compilation of data	No	No	No	No	No	No	No	No	?	Yes	No	No	XLS	No
Normalized Prices	RRED		2	Jayson Beckman	Annual	Partial	761	MEDIUM	compilation of data	No	No	No	No	No	No	No	No	?	Yes	No	No	XLS	No
Organic Handlers: Procurement and Contracting	RRED		3	Catherine Greene	Periodic	Yes	613	MEDIUM	primary data; summary statistics	No	No	No	No	No	No	No	No	?	Yes	No	No	XLS	No
Organic Prices	RRED		3	Catherine Greene	Periodic	Partial	7,435	MEDIUM	compilation of data	No	No	No	No	No	No	No	No	?	Yes	No	No	XLS	No
Organic Production	RRED		3	Catherine Greene	Periodic	Yes	12,886	MEDIUM	primary data; summary statistics	No	No	No	No	No	No	No	No	?	Yes	No	No	XLS	No
Rural Definitions	RRED		3	John Cromartie	7 years	Yes	3,906	MEDIUM	compilation of data	No	No	No	No	No	No	No	No	Yes	Yes	No	No	XLS	No
Rural-Urban Commuting Area Codes	RRED		3	John Cromartie	9 years	Yes	5,182	MEDIUM	compilation of data	No	No	No	No	No	No	No	No	Yes	Yes	No	No	XLS	No
Rural-Urban Continuum Codes	RRED		3	Timothy Parker	10 years	Yes	11,338	MEDIUM	compilation of data	No	No	No	No	No	No	No	No	Yes	Yes	No	No	XLS	No
State Fact Sheets	RRED		2	Timothy Parker	Annual	Partial	145,537	HIGH	compilation of data	No	No	No	No	No	No	No	No	Yes	Yes	No	No	XLS	No
Urban Influence Codes	RRED		3	Timothy Parker	5 Years	Yes	3,512	MEDIUM	compilation of data	No	No	No	No	No	No	No	No	Yes	Yes	No	No	XLS	No
Commodity Consumption by Population Characteristics	FED		2	Biing-Hwan Lin	Every 2 yrs (at best)	Yes	5,255	MED-HI	nonparam calc	No	No	Yes	No	Yes	No	Yes	No	Yes	Yes	No	XLS	
Eating and Health Module (ATUS)	FED		4	Karen Hamrick	Every 5 yrs (at best)	Yes	3,183	MEDIUM	primary stats	No	No	Yes	No	Yes	No	Yes	No	Yes	Yes	No	XLS	

Applying the Standards In-Depth: Data Product Review Council

- Reviews Premier products every 5 years, others as time permits
- Develops guidance for meeting standards
- Helps track Action Plan progress for ERS management
- Improves review process to minimize time by researchers and reviewers

Data Product Review Process

- Selection of products, notification, kickoff meeting
- Data product managers submit response form describing how their product meets each ERS data quality standard
- Council provides comprehensive evaluation for adherence to the standards
- Council offers feedback and guidance to data product managers and their branch chiefs, and identifies areas for improvement if necessary
- Data product managers submit action plans for addressing findings, developed in consultation with branch chiefs, incorporated into annual workplan

Keys to Success

- Build from best practices and existing guidelines
- Build in flexibility to adapt to change
- Set feasible priorities
- Invest in staff training
- Pay attention to incentives facing staff
- Collaborate with Data Product Managers

Questions?

ERS Data Product Quality

Join us online: www.ers.usda.gov
Follow us on Twitter: @USDA_ERS