

VIEWPOINT

Developing Best Practices for Regulatory Data Collections

16-01 | May 10, 2016

Comprehensive, high quality, interoperable data are increasingly important for financial stability analysis, market monitoring, and policymaking. They are also critical for risk management at financial firms. Members of the Financial Stability Oversight Council (FSOC) collect a significant amount of data for these purposes and their needs are increasing. Risk managers use the same or similar data to assess enterprise risk; clearly, their needs also are increasing. The OFR prepared this report to initiate a discussion about identifying best practices in financial regulatory data collection, with a view to aligning the interests of government officials and the industry.

Financial and economic analysis, monitoring vulnerabilities in the financial system, oversight of financial markets, and policy decisions are only as good as the data supporting them. Most fundamentally, the data must be fit for their intended purpose, in other words, suited to help answer questions raised.

Employing best practices in data collection is likely to improve data quality. Experience shows that precise definitions of financial contracts and parties in financial transactions help ensure quality. Adhering to best practices in the collection process is crucial to producing consistent, accurate, and suitable data. The process requires meticulous planning, data transmission controls to avoid introducing errors, and validation techniques to ensure the data are fit for their purpose.

Best practices in collecting data can also facilitate data sharing among regulators. For example, the same data needed for market oversight are likely also suitable for financial stability monitoring. By identifying such overlaps early in collection design, authorities who are engaged in each endeavor can collaborate in sharing at the start. Sharing existing data can also serve a valid governmental purpose — such as forming a more comprehensive understanding of the entire financial system — even if that purpose differs from the original purpose of the data collection. In turn, appropriate sharing can limit or reduce the regulatory reporting burden

This OFR Viewpoint represents the views of the Office of Financial Research. It is not an OFR policy statement and is not binding. OFR Viewpoints do not necessarily represent official positions or policy of the U.S. Treasury Department. OFR publications may be quoted without additional permission.

on industry, by reducing or even eliminating instances of collecting the same data more than once.

This paper describes a set of best practices for permanent regulatory collections of structured data. It supports ongoing OFR programs devoted to data quality, data scope, and data accessibility (see OFR, 2015a). This set of best practices is based on current knowledge and the literature that regulators and market participants have published about their efforts to collect structured data en masse using appropriate data standards (for example, see Burnside, 2000; Federal Reserve Bank of New York, 2012; Food and Drug Administration, 2015; and Glasser, 2013).

These best practices will evolve as the OFR continues to learn from experience conducting data collections from financial market participants, in collaboration with financial regulators. For example, the OFR recently completed a voluntary pilot data collection on the bilateral repurchase agreement (repo) market with the Federal Reserve System, with input from the Securities and Exchange Commission (SEC). This pilot was undertaken as a basis for informing the development of a permanent collection in the future. The pilot helped the OFR explore whether the proposed collection would meet its intended objectives. It also helped in understanding the types of data needed and the capabilities the OFR and the market participants would need for an ongoing collection. This type of pilot may become a best practice, particularly when data involve new market products or activities, new data reporters, or new technologies.

This paper is organized into sections discussing preliminary determinations about best practices (including transmission protocols and data quality), common pitfalls, regulatory considerations, and conclusions.

Best Practices Begin with Preparation

A **data reporter** is a company that reports data to a government regulator.

D ata collection is, by nature, a detailed undertaking. Each situation for regulatory data collection is different. Before beginning, regulators need to (1) define the business purpose, (2) design a template, (3) ensure key terms are clear and precise, and (4) prepare collection specifications so data reporters understand what is required.

1 Define the business purpose. A well-defined business purpose for collecting the data is essential. Just as an architect needs to know whether a design is for a home or business, a regulator designing a collection needs to know what knowledge gap will be filled, understanding deepened, or problem solved. The following set of considerations is a good place to start with any data collection, even when the answers to all of these questions are not readily known:

- What questions will be answered with these data? What information is needed to answer the questions?
- Do other regulators have similar questions that could be answered through this data collection, given the jurisdictional authority of the collector? Could related questions arise in the near future? In some cases, making changes to an existing data collection is more expensive for the data reporter than providing additional information at the outset.
- Are there relevant legal restrictions, for example, on the kinds of entities that may be asked to participate in the data collection? What kinds of data fields may be allowed or not allowed from a legal point of view? Does the regulator have a legal ability to mandate the collection, or can each data reporter make a best effort?
- Are these data available elsewhere? Can the data be derived from other collected data? Is industry already undertaking a similar collection? If so, can the existing collection be leveraged (standards, terms, structure, for example)?
- Is information needed from the entire population or would a sample suffice? If using a sample, should it be a sample of reporting entities or a sample of data elements, or both?
- What aspects of the population need to be represented? What sampling methodology best ensures that the sample is a true representation of the population?
- What level of granularity is needed? Are summary statistics sufficient or are detailed components required?
- What periodicity is needed? Is the same periodicity needed for all data terms? How long and how often will the data be collected? Is this a one-time survey or a survey that will continue for years?
- Are data that meet the regulator's business purpose already available in firms' business, risk, or accounting management systems?
- What is the approximate volume of the data to be collected? Do the data require additional infrastructure to be collected, loaded, and stored?
- What happens with these data after they are collected and analyzed? If the data will be used in publications, additional information may be valuable to provide context or assist in making representative or anonymized datasets.
- What type of IT systems and support are needed to store and service the data? What type of business support is needed to first build and then maintain the data?
- What level of security is needed for the data? Can the data be publicly disclosed or is it confidential supervisory information? Do the data contain personally identifiable information as defined under the Gramm-Leach-Bliley Act? Each of these considerations is critical for determining how the data are stored, maintained, and accessed.

A Good Place to Start

What questions need to be answered?

Are there any legal restrictions?

Are similar data or collections available anywhere else?

What is the frequency and term of the collection?

What is the size and breadth of the collection?

How will the data be used?

Can IT give the support needed?

Who will have access to the data?

How will the data be accessed?

What level of security is needed?

Are there incentives and benefits to the public or industry?

Data Collection Preparation Process



- What data governance is needed? How will the data be stored and how will access be maintained?
- What types of tools are needed to access the data so that they can be utilized effectively?
- What are the likely benefits for the private companies who are being required to report the data? Collecting additional information, or making aggregates or subsets of the information publicly available, could make the data collection more valuable to the reporters. Adding value for the reporters often improves data quality and response rates.
- Do data reporters have incentives to make the data ambiguous or inaccurate? For example, such incentives led to the misreporting of rates for the London Interbank Offered Rate. What strategies could prevent intentional misreporting, such as requiring the reporting of transactions that are used to calculate a rate rather than reporting a rate?

A regulator should understand the business process involved in producing the data as a prerequisite to working with data reporters. The point in the reporter's business process where the data are collected is sometimes an important consideration. Understanding the business process not only helps the people defining the requirement but it can help to ensure that the requirements are fully understood by the reporter.

2 Design a template. Ultimately, the data collection template forms the standard for developing the collection, so the template requires vetting by subject matter experts. Regulators, data reporters, and others can help to identify commonly used terms and policies. Data reporters can be tapped to create fictional data for discussion purposes in advance of a collection, to differentiate theoretical ideas from realworld activity. This approach allows information technology specialists to ensure fields are properly formatted and that the data stream can be automated.

3 Develop clear and precise definitions. Clear and precise definitions may be the most important part of any data collection. Technical jargon and complex concepts can contribute to semantic misunderstandings. The following steps can help ensure that the desired data and the reported data are the same:

- Reference existing industry data standards whenever possible.
- Provide a detailed glossary or section describing data elements as a point of reference for all definitions.
- Define all information to be collected or refer to a common dictionary, even when the meanings seem obvious.
- Review with data reporters data elements to be collected and their definitions.

- Give a visual representation of how the data elements relate to one another.
- Assess legal or policy issues, such as confidentiality, that may limit reporting by the reporter, or the use and sharing of the data by the regulator.
- Issue guidance to participants on how to submit data points that are blank, null, missing, unknown, or "not applicable."
- Issue guidance on the point in time (Eastern Standard Time, close of business) or time period (calendar year, fiscal year) that the data should reflect.

The importance of clear and precise definitions cannot be overstated. For example, simply requesting the value of a firm's inventory would be insufficient because of multiple alternative valuation methods, such as last in, first out and first in, first out inventory accounting methods. Similarly, collecting the value of a firm's investment portfolio requires specifying the valuation method (mark-to-market or historical cost accounting). If the valuation method isn't specified, or at least identified, then inconsistent reporting is more likely and erroneous analysis can result. A straightforward approach may be to rely on an outside, existing dictionary, such as generally accepted accounting principles.

A visual representation can be a powerful tool in helping to ensure that definitions are understood and expectations are matched. For example, loans and leases can be defined as the amount of loans and leases that the

Visual Representation Example

reporting entity has the intent and ability to hold for the foreseeable future or until maturity or payoff. But without more context, it might be hard to know if this definition of loans and leases takes allowances for losses into account. The "visual representation example," which was taken from the Report of Condition and Income or Call Report submitted by U.S. banks each quarter, answers that question, as well as other possible questions, by putting the data in the context of an accounting statement. Providing the visual representation answers many potential questions.

ast business day of the quarter.	the a	amour	nt outs	standir	ng as	of the
Schedule RC—Balance Sheet						
Dollar Amounts in Thousa	nds	RCFD	Tril	Bil	Mil	Thou
Assets						
1. Cash and balances due from depository institutions (from Schedule RC-A):		0081				
a. Noninterest-bearing balances and currency and coin ¹				<u> </u>	<u> </u>	+
b. Interest-bearing balances ²		0071				
2. Securities:		1754				
a. Held-to-maturity securities (from Schedule RC-B, column A)		1754 1773			<u> </u>	+
b. Available-for-sale securities (from Schedule RC-B, column D)		1//3				
3. Federal funds sold and securities purchased under agreements to resell:		0007				
		B987			<u> </u>	+
b. Securities purchased under agreements to resell ³			_			
4. Loans and lease financing receivables (from Schedule RC-C):	-	RCFD				
a. Loans and leases held for sale		5369				
b. Loans and leases, net of unearned income B528						
c. LESS: Allowance for loan and lease losses 3123						
d. Loans and leases, net of unearned income and allowance (item 4.b minus 4.c)		B529			<u> </u>	\vdash
5. Trading assets (from Schedule RC-D)		3545			<u> </u>	\vdash
6. Premises and fixed assets (including capitalized leases)		2145			<u> </u>	\vdash
7. Other real estate owned (from Schedule RC-M)						
8. Investments in unconsolidated subsidiaries and associated companies						\vdash
9. Direct and indirect investments in real estate ventures		3656				

4 Create collection specifications. An important step in the data collection process is to create a clear, simple set of instructions, known as collection specifications. Collection specifications include data names, definitions, types, acceptable enumerations, and referential integrity restrictions that are essential for data.

Common Collection Specifications

1

2

Create a document about your collection's purpose, process, schedule, and operations.

Describe data names, definitions, formats, rules, value ranges, transmission methods, etc.

Provide a mechanism for data reporters to ask questions and obtain help. Some of the more established regulatory collections, such as bank Call Reports (see Federal Financial Institutions Examination Council, 2015), include the following:

- 1. A narrative document describing the broad purpose of the data collection, an overview of the process, reporting deadlines, and the operational framework.
- 2. A detailed data specification that includes precise descriptions of:
 - names for data;
 - data definitions;
 - data formats and types to identify if an item is a number, a character string, or date;
 - any rules for the data such as an address must always include a city;
 - allowable values for data that are enumerated;
 - categorized lists of permitted values for data terms;
 - data transmission requirements such as extensible markup language schema or comma-separated value layout, or a standard such as eXtensible Business Reporting Language (XBRL); and
 - required data validation and quality level.
- 3. A mechanism for data reporters to ask questions and obtain help, such as a website, e-mail address, or phone number. Common questions from data reporters should be published as frequently asked questions and used to improve the data collection processes and data definitions.

Transmission Processes

Collect structured data. Although regulators may find a report form or visualization helpful as a frame of reference, the data should be collected as individual elements that are components of a structure, not as a digitized version of a report in paper form. Breaking the collection into component parts provides greater flexibility in using the data. Data collected at a detailed level and stored in a structured manner can serve more than one purpose, enabling regulators to use the same data in different ways. This type of approach has been the conceptual driver for many industries' pursuit and use of data as an asset. Data that are collected and cultivated as an asset can be used to answer multiple questions, both known and not yet considered. This approach is the bedrock of data-driven management — using digital data streams and analytics for more efficient decisionmaking.

Follow transmission standards. Standards are essential to facilitate the digital transmission of data. In many industries, data standards are already in place and regulators can use them to address challenges in data definition and consistency. Businesses often use data standards in electronic transactions or messages with other companies. Messages transmit business requests or responses to requests, and may list terms

and conditions for conducting transactions. Some standard-setting bodies such as the Mortgage Industry Standards Maintenance Organization, Financial Information Exchange, Financial Products Markup Language, and the International Organization for Standardization have created data standards for messages that are used for reporting required by regulators.

Transmitting data system-to-system allows a regulator to rely on fixed technology controls to protect data from exposure, alteration, and fraud. System-to-system transmission also means regulators can efficiently collect large amounts of data from multiple reporters simultaneously with fewer, if any, needs for increases in staff. Automated data collection facilitates accessing and analyzing data in real time.

Existing standards can help facilitate the creation of a data collection. For example, the Generic Statistical Business Process Model can help with statistical collections, XBRL provides rules for reporting accounting information, Statistical Data and Metadata Exchange highlights essential elements for economic data, and the Dublin Core Metadata Initiative helps to inform thinking on metadata collections. Referencing an existing standard allows regulators to leverage the collective knowledge of industry experts. The SEC requires firms to submit some forms in XBRL, including 10-K and 10-Q forms, and relies on the Financial Accounting Standards Board to update the taxonomy annually (see Securities and Exchange Commission, 2009).

There are benefits to creating a standard by consensus with feedback from the industry. Such a standard reflects the systems, data, and business processes already used by the industry, improving a regulator's likelihood of success with a new data collection.

Data Quality

Data quality refers to accuracy, timeliness, and completeness. Data collection efforts are more likely to result in quality data if they use pilot data collections, test periods, and automated validation controls to identify any problems. Ongoing feedback from data reporters is also critical.

Pilot data collections and test periods. A pilot data collection offers an opportunity for regulators to work closely with data reporters and work out the details of a collection before deciding on a permanent collection. For example, the OFR and the Federal Reserve System, with input from the SEC, launched a pilot data collection in 2014 on the bilateral repo market. Nine bank holding companies participated on a voluntary basis. The pilot helped the participants identify specific challenges of collecting this type of market data. This information will be useful in defining the parameters and data collection template for a permanent data collection. The experience suggests that pilots may evolve into a best practice, particularly when data about new market products or activities, new reporters, or new technologies are involved. Pilot data collections

- + Test periods
- + Automated validation controls
- + Continuous feedback

Quality data

A test period for data submission is a standard industry practice that regulators and data reporters can use to test processes. During a test period, reporters provide test data to see if systems work and to make sure data are formatted properly. A test period confirms that properly formatted data can be used and improperly formatted data or data security issues are handled appropriately. This kind of round-trip testing gives regulators and data reporters confidence that the process is ready for real data to begin to flow.

Controls. At the point of data collection, automated controls are needed to reduce errors and confirm that the data received are valid. Upfront validation, in the form of submittal failures and messaging, is more efficient for data reporters than answering validity questions after submission. Problems are also easier and less costly for regulators to correct than when poor quality data enter and permeate a system. Creating a validation framework at the planning stage of the data collection process is easier for data reporters than adding one after collection has begun.

An automated submittal validation control screens incoming data and blocks incorrectly formatted data. It should use industry data standards whenever possible and create proprietary standards only to fill the gaps. The system generates a message to the data reporter when a submission has failed. Other validation controls use error messages that escalate in severity from warnings to rejection of the data submission. Many organizations use both approaches to prioritize data cleanup when introducing new data, or when requirements change over time.

Regulators also need procedures to restate historical data on occasion. When restatements are necessary, regulators need to develop and maintain documents that reflect the data's lineage and history to ensure that restated data are properly identified for end users.

Ongoing feedback. Work does not end once the data collection process is tested and finalized. Regulators should gather ongoing feedback and incrementally refine controls, specifications, and data to improve the process over time. Other changes will be needed in the collection process when unforeseen circumstances arise, business needs change, or new technology is introduced. Regulators face additional costs and challenges in making changes after a collection has begun.

Common Pitfalls in Regulatory Data Collection

• Failure to use an industry data standard when one exists. Ignoring an existing industry standard that could meet supervisory or regulatory needs can result in data requirements that differ from the way the industry conducts business. Definitions of similar data terms may be different, or the tools and technologies required may not integrate with industry methodologies. Both situations can mean higher costs and lost productivity. Failure to adopt industry standards when possible — such as a universal identifier, a widely accepted list of terms, or standard codes — may produce data that are difficult to merge with other datasets.

- **Missing or incomplete data requirements.** Without detailed requirements, data reporters will have extensive questions about data definitions, performance requirements, compliance issues, and timelines. Operational problems are likely if data submissions fail, or data are submitted incorrectly. Data collected may result in flawed or unusable analysis. For example, the OFR's 2015 Financial Stability Report noted that the lack of agreement on reporting requirements has hindered the global effort to improve transparency in derivatives markets (see Office of Financial Research, 2015b, p. 77).
- Inaccurate instructions and lack of resources to support data reporters. Without clear, documented instructions and a mechanism for data reporters to get additional information, confusion may emerge when the collection process begins. Regulators may be overwhelmed by multiple data reporters asking the same questions. Frustrated data reporters may eventually question the credibility of the data collection initiative and the regulatory agency or organization sponsoring it.
- Focus on collecting reports, not data. When a regulator creates a data collection program that is report-focused, the data tend to be narrowly defined to answer a single question. A report-focused approach may encourage data reporters to infer a "right answer" to each question in the report, which affects the quality of the data. Although such a report could serve an important "point-in-time" compliance purpose, data collected for a specific report may have limited, narrow use and may be difficult to use with other datasets. A report-focused approach may result in an eventual collection of separate reports with overlapping, duplicative, or divergent data requirements.
- Inadequate preparation. Failing to prepare adequately including necessary upfront analysis and other industry outreach can compromise the data collection initiative from the start. Conscientious preparation is important to establish and carry out best practices with regulatory data collections. Preparation also includes the use of pilot data collections and testing. Without these steps, data reporters may lack confidence that the data they send will meet the project requirements. Operational and integration issues are likely to crop up as the data collection begins. These problems will hamper use of the data and test the patience of the data reporter and the regulator.
- **Ignoring data quality.** An approach that relies on collecting general data then determining what they mean, or conducting data quality exercises too late in the project, will extend the implementation window. As noted, a pilot collection of data can help in analyzing and properly defining data requirements and validation rules.

Ways to Avoid Common Pitfalls

Use an industry standard when one exists.

Provide detailed requirements.

Give clear instructions and methods for help.

Focus on collecting data, not reports.

Conduct early outreach to industry.

Prepare analysis upfront.

Perform a pilot collection.

Data Regulatory Considerations

A lthough many best practices and common pitfalls apply equally to data collection by private industry and government, regulatory reporting has some unique issues. Rules (including the Paperwork Reduction Act of 1995) require or encourage U.S. regulators to estimate how much a new required data submission will cost industry participants. To comply, regulators must estimate the amount of effort and the costs for industry's ongoing production of the data. The cost of not having careful reporting regimes and common standards can be significant and should be considered.

A regulator striving for internal efficiency may be ready for the foundational changes necessary to begin collecting raw data suitable for reuse for other business needs, as allowed by federal statute.

With this concept in mind, regulators should consider ways to maximize the value of existing data collection processes. Using a central facility, like the Federal Financial Institutions Examination Council structure, for data collection and aggregation can be an efficient way to collect data once for multiple purposes among several regulatory agencies (see Federal Financial Institutions Examination Council, 2006). The Office of Management and Budget also administers an interagency process that facilitates coordination in promulgating regulations. In addition to this framework, a regulator could avoid duplication by collaborating with an agency that has a similar data collection, or by enhancing an existing data collection. Within confidentiality requirements and data usage restrictions, collaboration among U.S. agencies or with international regulators can make collection more efficient and potentially reduce the burden on data reporters. Finally, regulators should consider the alternatives - for example, purchasing data for needs that are ad hoc or temporary, or conducting surveys rather than undertaking collections.

Improving collaboration among regulators would help in reducing impediments to data sharing, such as legal requirements, contractual provisions, or old technology and processes. One potential improvement would be to create a shared data utility with mutual agreements to align data protection, usage requirements, and policies of the organizations involved.

Conclusions

U sing best practices of industry and government can help regulators collect financial data that will answer immediate questions and potentially those in the future. Regulators need to think beyond project management, data storage management, and information technology practices and focus on the requirements for data collection.

Collection specifications that are comprehensive and well defined form the bedrock of the data collection effort. Regulators need to approach each step in the collection process with attention to detail, while striving for simplicity to build a solid foundation. Also, regulators should remember that analytical needs change frequently and therefore they should build flexibility into the collection system.

More collaboration among regulatory data collectors should be explored to maximize the value of current requirements for data reporting. Regulators should assess obstacles to collaboration and data sharing, such as data policies, statutory requirements, and technologies, to find ways to harvest more value from current data collections.

Collaboration coupled with a methodically designed data collection process will give regulators a dynamic source of information for operations, policies, recommendations, and actions. Although data collection practices are the essential first step, other parts of regulators' data fulfillment processes must also be examined to find greater efficiencies. Research is needed in data quality, value, semantics, transmission protocols, data processing, security and confidentiality, data publishing, and regulatory changes. Improvements in these areas will help regulators leverage data for financial stability monitoring.

The OFR is working with FSOC member agencies to improve the scope, quality, and accessibility of financial data, especially related to new and emerging sources of potential vulnerability (see OFR, 2015b, pp. 113-6). The OFR may further refine the best practices described in this paper as it continues to learn from experiences conducting data collections from financial market participants, in collaboration with the financial regulators.

References

Burnside, Rob. "Towards Best Practice for Design of Electronic Data Capture Instruments." Methodology Advisory Committee paper. Australian National Statistical Service, November 2000. www. nss.gov.au/nss/home.nsf/75427d-7291fa0145ca2571340022a2ad/2b7df9253ec05610ca2571ab-002470f1?OpenDocument (accessed May 2, 2016).

Federal Financial Institutions Examination Council. *Call Report Instruction Book Update.* June 2015. www.fdic.gov/ regulations/resources/call/crinst/2015-06/615Inserts_063015.pdf (accessed May 2, 2016).

Federal Financial Institutions Examination Council. "Improved Business Process Through XBRL: A Use Case for Business Reporting." January 2006. xbrl.us/wp-content/ uploads/2007/12/20060202FFIEC WhitePaper.pdf (accessed May 2, 2016).

Federal Reserve Bank of New York. "Regulatory Reporting Best Practices." Presentation at the Regulatory Reporting Best Practices Seminar, April 26, 2012. www.newyorkfed.org/medialibrary/media/newsevents/events/ banking/2012/2012BestPracticesPresentation.pdf (accessed May 2, 2016).

Food and Drug Administration's Center for Drug Evaluation and Research. "Data Standards Strategy 2015-2017." July 2015. www.fda.gov/downloads/Drugs/ DevelopmentApprovalProcess/ FormsSubmissionRequirements/ ElectronicSubmissions/UCM455270. pdf (accessed May 2, 2016). Glasser, Dessa T. "Leveraging Data for Financial Stability Monitoring." *Journal of Banking Regulation* 14 (July/November 2013): 195-208.

Office of Financial Research. 2015 Annual Report to Congress. Washington: OFR, 2015a. www.financialresearch.gov/annual-reports/files/ office-of-financial-research-annual-report-2015.pdf (accessed May 2, 2016).

Office of Financial Research. 2015 Financial Stability Report. Washington: OFR, 2015b. www. financialresearch.gov/financial-stability-reports/files/ OFR_2015-Financial-Stability-Report_12-15-2015.pdf (accessed May 2, 2016).

Securities and Exchange Commission. "Toward Greater Transparency: 21st Century Disclosure Initiative." January 2009. www.sec.gov/spotlight/disclosureinitiative/report.pdf (accessed May 2, 2016).

U.S. Congress. Dodd-Frank Wall Street Reform and Consumer Protection Act. Public Law 111-203, Washington: July 21, 2010. www.gpo. gov/fdsys/pkg/PLAW-111publ203/pdf/ PLAW-111publ203.pdf (accessed May 2, 2016).

U.S. Congress. *Paperwork Reduction Act of 1995.* Public Law 104-113, Washington: May 22, 1995. www.gpo. gov/fdsys/pkg/PLAW-104publ13/ html/PLAW-104publ13.htm (accessed May 2, 2016).