Data Quality Literacy 

**A Knowledge Brief**

# *Data Quality Literacy Series 07*

# **Understanding Commercial Data**

*Understanding commercial data, the benefits of using commercial data, and its limitations.*

**Commercial Data**, also called private-sector data or third-party data, broadly refers to data created and provided by commercial entities rather than government agencies.

## Commercial Data Examples

| **Structured Commercial Data** | | |
| --- | --- | --- |
| **Structured Survey Data** | **Structured** **Administrative Records** | **Other Structured Data** |
| Media Market Data (e.g., Nielsen) | Market Research Probability Survey or Opt-in Panel data (e.g., Ipsos, Gfk, Kantar, Mintel) | Customer Satisfaction Surveys | Banking and Stock Records (e.g., Bloomberg, S&P, Moody’s, Compustat, CRSP) | Commercial Transactions (e.g., Refinitiv) | Point of Sales Data (e.g., IRI/Circana) | Credit Card Records (e.g., Experian) | Housing Data (e.g., Zillow) | Private Employment Data (e.g., ADP) | State and Local Tax Data (e.g., CoreLogic) |Climate and Self-reported Environmental Data (e.g., CDP) | E-commerce Transactions | Mobile Phone Location Sensors | GPS Sensors | Utility Company Sensors | Weather or Pollution Sensors |
| **Semi-Structured**  **Commercial Data** | XML or JSON Files; Data from Computer/Online Systems (e.g., web logs); Emails; Articles from Full-text databases | |
| **Unstructured**  **Commercial Data** | Social Media Data (e.g., Facebook, Twitter, LinkedIn); Internet Searches (e.g., Google), Videos (e.g., YouTube), Traffic Webcams, Satellite Images | |

## The Benefits of Using Commercial Data

* It can have content or a level of granularity that federal statistics do not provide.
* It can be provided more timely and frequently than federal statistics.
* Commercial data vendors can have business relationships with private firms, so they can acquire and synthesize proprietary data.
* It can be more cost-effective than collecting data on your own and can reduce the response burden.
* It can complement and enhance the analysis of federal statistics.

## Limitations of Commercial Data

* Commercial survey data generally have lower response rates than government surveys. Many firms have chosen opt-in Internet panels over probability surveys, which may cause concerns about the representativeness of the sample.
* Administrative data collected for transactional purposes tend to be less stable in data definition and data-generating processes.
* Commercial data are often vulnerable to changes or discontinuation without notice and subject to manipulation for private interest.
* Vendors often provide the latest data at a point in time in a dashboard interface rather than datasets and data file format may not be compatible with statistical software.

To learn more, refer to *Data Quality Literacy Series 08: Evaluating Commercial Data Quality.*

**References**

Harris-Kojetin, B. A., & Groves, R. M. (Eds.). (2017). *Innovations in Federal Statistics: Combining Data Sources while Protecting Privacy*. <https://nap.nationalacademies.org/catalog/24652/innovations-in-federal-statistics-combining-data-sources-while-protecting-privacy>

Harris-Kojetin, B. A., & Groves, R. M. (Eds.). (2018). *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*. <https://nap.nationalacademies.org/catalog/24893/federal-statistics-multiple-data-sources-and-privacy-protection-next-steps>

Liu, G., Bordelon, B., Nagar, R., Sarti, J., Nguyen, U., & Boettcher, J. (2024). *Data Quality Literacy: A Guidebook*. Institute of Museum and Library Services (IMLS) Grant Project. <https://doi.org/10.31219/osf.io/ruawm>

Muth, M., Sweitzer, M., Brown, D., Capogrossi, K., Karns, S., Levin, D., Okrent, A., Siegel, P., and Zhen, C. (2016). ​​Understanding IRI Household-Based and Store-Based Scanner Data. Economic Research Service. *USDA Technical Bulletin 1942*. <https://www.ers.usda.gov/publications/pub-details/?pubid=47636>

| The Knowledge Brief is compiled by Grace Liu, inspired by the National Forum presentation from Bill Sermons, Todd Hines, and Patrick W. McLaughlin, reviewed by the IMLS Data Quality Literacy project team, and designed by Niko Galioto. This project was made possible in part by the Institute of Museum and Library Services [[RE-252357-OLS-22](https://www.imls.gov/grants/awarded/re-252357-ols-22)].  Visit the project website to learn more!  [https://www.dataqualityliteracy.org](https://www.dataqualityliteracy.org/) |
| --- |