Data Analysis Worksheet

# Data Analysis Overview

If you determined that you need to do a PIA, then the next phase is a data analysis of the system or process. A data analysis is similar to a data inventory, though an analysis is more comprehensive.

If there is a vendor or other third party involved in the process, you will need to bring in a vendor representative to ask questions surrounding their processing of patron data, including information security and who they share the data with on their end.

At the end of the data analysis, you should have a comprehensive overview as to what data is collected and processed, what data is shared with whom, and what regulations and policies need to be considered for the process.

# Data Analysis

## Data Inventory Templates

Most data inventory templates can serve as a foundation for a data analysis. Two library-oriented data inventory templates that can be adjusted to create a more comprehensive data analysis document:

Patron Data Lifecycle – Data Inventory - [Link to PLP Data Inventory Spreadsheet]

Library Data Risk Assessment Spreadsheet - <https://osf.io/vs4jh/>

## Data Flow Investigation

As mentioned above, the data analysis should be more thorough than a data inventory. The data analysis tracks not only the data lifecycle from collection to deletion, but also the components of the system or process that interacts with the data in any meaningful way. This includes technical and non-technical components, stakeholders and data subjects, and use of the data.

Data flow investigations take time, and will require a number of interviews with people who use the system or process, as well as the people whom the data is collected and reported for. In Michael L. Whitner’s article “Conducting a Privacy Audit”, Whitner presents 14 questions that can be adapted for a data flow investigation:

1. What are the purposes for which personal data is being collected? For example: customer administration, employee administration, advertising and marketing.
2. What individual’s personal data is being processed? For example: customers, employees, suppliers.
3. What types of personal data is being collected? For example: names, addresses, telephone numbers, occupational details, social security numbers or ID numbers, and financial information. Particular attention should be paid to any personal data that might be considered “sensitive.” In the EU, sensitive data includes that pertaining to racial or ethnic origin, political opinions, religious beliefs, trade union membership, health matters and sexual orientation.  *[Author’s note – these data points are also sensitive in the context of US libraries]*
4. How is personal data collected? For example: hard copy form, online, by telephone or from third parties.
5. Is the consent of the individual obtained? If so, by what means, and at what point in the collection process?
6. How relevant is the personal data for the purposes collected? Would anonymized personal data be equally relevant?
7. What steps are taken to ensure that the accuracy of personal data is maintained during the period of retention?
8. How long is personal data retained? Is this retention period really necessary — e.g., legally mandated? Or is personal data being held for longer than required to meet legal obligations or for reasonable business purposes?
9. Where and how is personal data stored?
10. What technical and organizational security measures are taken to protect personal data against unauthorized access, damage or erasure? For example: encryption, use of secure passwords, contingency plans and training.
11. Is the personal data disclosed to any third parties? If so, for what purposes? What additional security measures are taken to protect disclosed personal data from unauthorized access, damage or disclosure — e.g., written contracts with third parties that impose specific data security and privacy obligations?
12. Is personal data transferred outside the country in which it is collected? If so, what consents are obtained and what additional security measures are taken?
13. Are there procedures in place to allow individuals to access and control use of their personal data? For example: opportunities to correct inaccurate personal data, delete irrelevant personal data, and prevent their personal data from being used for marketing purposes.
14. How is personal data that is no longer required disposed of? Does the method of disposal/destruction ensure that the personal data cannot be accessed again?

Libraries who want to have a more prescriptive data flow investigation outline should refer to Section 5 “Map information flows” in the Office of the Australian Information Commissioner’s *Guide to undertaking privacy impact assessments* at <https://www.oaic.gov.au/assets/privacy/guidance-and-advice/guide-to-undertaking-privacy-impact-assessments.pdf>.

# Sources and Additional Resources

Briney, Kristin, Becky Yoose, John Mark Ockerbloom, and Shea Swauger. 2020. “A Practical Guide to Performing a Library User Data Risk Assessment in Library-Built Systems.” DLF Privacy and Ethics in Technology Working Group. <https://doi.org/10.17605/OSF.IO/V2C3M>.

Newfoundland and Labrador Centre for Health Information. 2013. *Full PIA Report Template.* Version 2.2.

Office of the Australian Information Commissioner. 2020. “Guide to Undertaking Privacy Impact Assessments.” <https://www.oaic.gov.au/privacy/guidance-and-advice/guide-to-undertaking-privacy-impact-assessments/>.

Whitener, Michael L. 2012. “Conducting a Privacy Audit.” *The Corporate Counselor.* 27(3). <https://iapp.org/media/pdf/knowledge_center/Conducting_a_Privacy_Audit_-_The_Corporate_Counselor_-_July_2012.pdf>.