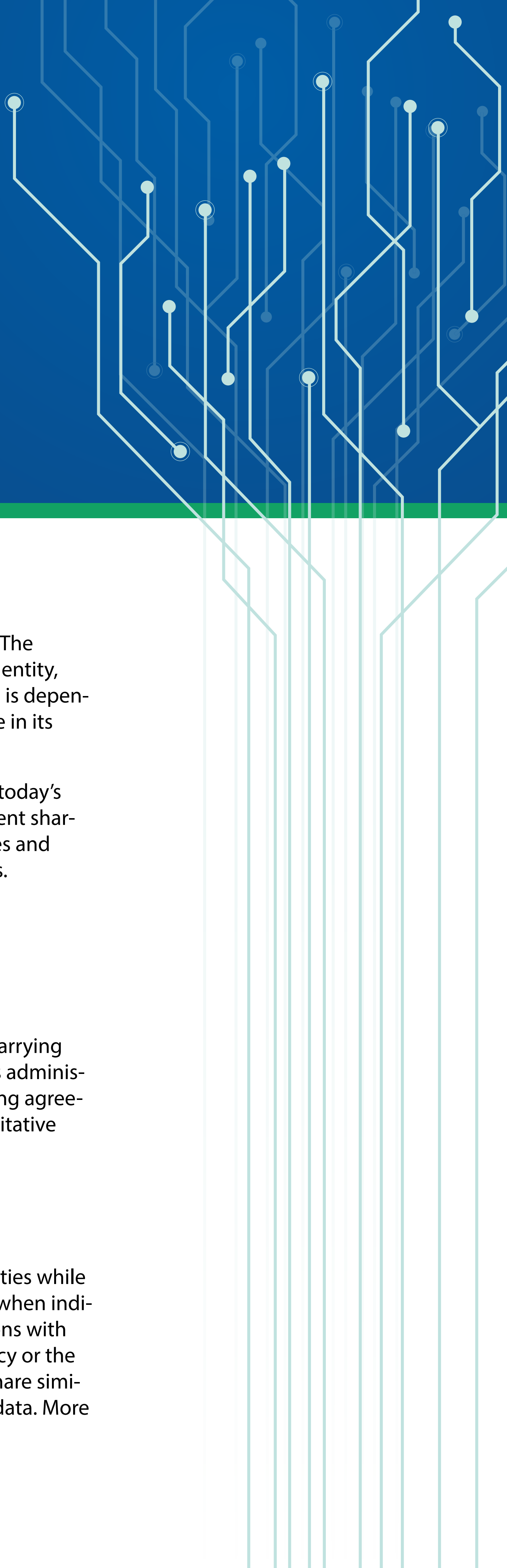
CSG The Council of State Governments

ADAPTAC



Understanding Data Sharing Agreements

Introduction

This paper is designed as an introduction to data sharing agreements. The intended audience is anyone who intends to collaborate with another entity, public or private, where the exchange of data is required. Data sharing is dependent on data quality, confidentiality, and the confidence agencies have in its accuracy.

Data sharing agreements play a crucial role within the connectivity of today's complex networks with collaborators depending on the fast and efficient sharing of information. Legally, they must adhere to all the relevant statutes and regulations regarding privacy, security, and intellectual property rights.

The types of data are generally categorized as:

1. Personal data
2. Demographic data

This paper will concentrate on the data collected for the purposes of carrying out the mission of an agency or organization, commonly referred to as administrative data. Additionally, third parties may be invited into a data sharing agreement for the purposes of evaluating the data from a quality and quantitative perspective.

What is Administrative Data?

Administrative data is data collected by public agencies or private entities while administering their policies and services. This could be data collected when individuals register for government programs, data that records transactions with customers, or data that is collected during the administration of a policy or the delivery of a service. While the two kinds of data, public and private, share similar best practices, this paper focuses on governmental administrative data. More specifically, it will focus on state government administrative data.

Personal data is a major concern; all Personally Identifiable Information (PII) should be collected only with the express consent of the individual and after demonstrating a basis for obtaining and processing the data. Demographic data, on the other hand, does not require permissions, but must be obtained from the individual and how they self-identify under the established sociodemographic categories being gathered. A lot of data quality is dependent on the collection of accurate demographic data to help decision makers with policy and resource allocation.

What is a Data Sharing Agreement?

A data sharing agreement is a legal contract that outlines the details of the process by which data should be shared between two parties. These documents protect all parties by defining the requirements of the agreement, the responsibilities of each party, data specifications, legal considerations, compensation, additional parties and protections.

Sections Commonly Found in Data Sharing Agreements

- Identification of Parties – Both the organization responsible for transferring the data and the organization receiving the data should be clearly identified, including details such as the physical address and key personnel.
- Purpose for data sharing – The receiving organization's purpose for the data should be explicitly stated. For instance, if data is to be used for research, the specific research question/s to be investigated should be stated.
    - Some data sharing agreements go a step further and place publishing restrictions on the results and findings. These data sharing agreements give the data's original owner the right to review and approve any results or findings before publication.
- Other Restrictions on Use – Any other restrictions on use should be detailed as well. This section might include things such as:
    - Dates for which the data is needed and how data will be handled when the agreed timeframe is reached.
    - The identities of those at the receiving organization given access to the data.
    - The receiving organization's freedom to share with third-party partner organizations.
    - The receiving organization's freedom to use the data for purposes outside the scope of the original agreement.
- Description of Data to be Shared – The exact data to be transferred should be described in detail. For example, the number of fields as well descriptions of those fields, the number of years covered and geographic scope (if applicable), and any other metadata that might be relevant.

- Security – There are several factors related to security that a data sharing agreement should cover:

    - How will the data be transferred? Will the transfer of physical material be required, or will it be a purely digital transfer?

    - How will the data be stored?

    - Is PII to be transferred? If so, what special precautions should be taken to secure this data?

    - Notification of breach procedure – If a breach does occur, what is the process for notifying the original owner and other parties?

    - In the event of a breach, who is responsible for the breach? And what steps are taken to resolve the breach. (see the Notification of Breach Procedure, below)

    - Who will be responsible for monitoring and enforcing security measures?

      The security section below will cover security considerations in more detail.

- Ownership retainment clause which outlines the rules of how long the data can be retained and used for other purposes by the recipient and when to destroy the data.

- Termination of Agreement – The conditions under which the transferring organization reserves the right to terminate or extend the agreement.

- Responsibility for Financial Costs – If there are any financial costs associated with the transfer of data, the party responsible for these costs should be explicitly stated. For example, if a software service will need to be purchased for the secure digital transfer of the data, the responsibility for this cost should be assigned in this section.

- Institutional Review Boards – Many academic and research institutions require any research related to human subjects to undergo preliminary review by an Institutional Review Board (IRB). IRBs are organizations that review the implications of research designs and possible findings to ensure any potential research meets ethical standards. If the transferring organization requires that any research using the transferred data under an IRB review, this should be highlighted here.

- Disclaimer on Accuracy, Completeness, Integrity, and Reliability of Data – This section should be both a guarantee by the transferring party that they have diligently ensured the accuracy and completeness of the data, and a protective clause that limits the transferring party’s liability for inaccurate or missing data.

Security Best Practices in Data-Sharing Agreements

Defining security practices in data-sharing agreements is fundamental where information is shared among separate entities. When the data to be shared includes sensitive or personally identifiable information (PII), it is especially important to communicate the standards and procedures to be utilized. This section outlines security concerns that should be considered and suggests best practices for including them in agreements.

TRANSFERRING DATA

3. **Secure Data Transfer Protocols:** Utilize secure data transfer protocols like HTTPS, SFTP, or encrypted email for transmitting sensitive data. These protocols provide protection if data is intercepted during transmission. The Transmission Method should be delineated, favoring methods that deliver data directly to a secure storage environment. For example, a web portal configured and maintained by a data recipient that allows a sender to directly upload data files is more secure than sending unencrypted Excel files by email.

4. **Data Integrity:** Ensure that the data sent is identical to the data received by using a checksum to compare the files. This protects the data from possible modification if it is intercepted in transit and ensures that no information is lost during transmission.

5. **Destruction of Transmitted Files:** Unless transmitted data is directly and immediately stored in a secure place like an encrypted database, files used to transfer information should be destroyed as soon as they are stored by the recipient. If it is necessary to convert or standardize datasets transmitted by senders to incorporate them in a secure storage system, the original files should be destroyed as soon as they are no longer needed (see **Data Retention Policy**).

STORING DATA

1. **Data Classification:** Categorize data based on sensitivity (e.g., public, internal, confidential, restricted) and grant individual access accordingly (see **Permissions**).

2. **Data Minimization:** Store only essential data required for the specific purpose outlined in the agreement. Limiting the data footprint reduces exposure and potential risks. Common practices in data minimization include limiting personal data sharing[1], reducing data collection on devices[2] and limiting data collection by third parties[3]. To effectively minimize your digital footprint, begin by deleting unused online accounts, as dormant profiles can be vulnerable to unauthorized access. Regularly clear your browsing history and cookies to reduce the data trail left behind during internet activities. Adjust privacy settings on social media platforms to limit information sharing to trusted contacts and be cautious about the personal details you post. Utilize encrypted messaging services to protect the content of your communications from potential interception. Additionally, consider disabling advertising identifiers on your devices

to prevent data brokers from compiling detailed profiles based on your online behavior. By implementing these strategies, you can significantly reduce the amount of personal information accessible online.

3. **Data Retention Policy:** Define a data retention policy within the agreement, specifying how long data should be stored. Regularly review and purge data that is no longer needed to reduce security risks. Miami University has compiled a resource on data retention policies. PII Protection

4. **Anonymization and Pseudonymization:** Whenever possible, anonymize or pseudonymize personally identifiable information (PII). This practice safeguards privacy while facilitating research and collaboration. Social security numbers (SSNs) should be converted to unique identifiers via hashing or an internal schema that obfuscates the original values. Whenever possible, raw SSNs should never be used as primary identifiers of individuals or records.

5. **Permissions:** Ensure that permission to access data is restricted to authorized individuals and reviewed regularly to align with the principles of data protection. To all data, apply the principle of least privilege, by which permission to access information is restricted to the minimum required by a person to perform their duties.

NOTIFICATION OF BREACH PROCEDURE

1. **Coordination:** Notification that a breach has occurred should be included in any incident response plan. Define the roles and responsibilities of each entity in the event of a breach. Establish clear timelines for each to notify the other once a breach is discovered.

2. **Reporting and Communication:** Identify third parties who could potentially be impacted by a breach as well as any regulatory entities to whom breaches must be disclosed. Determine who is responsible for making those notifications and the circumstances, if any, that might affect that determination. Detail the information that should be included in breach notifications, such as the nature of the breach, types of data compromised, potential risks to affected parties, and steps taken to address the breach.

In addition to the above considerations, a mechanism for ensuring that both parties adhere to the security practices defined in the agreement, as well as the consequences for failing to do so, should be identified. This could include regular audits by one or both parties, voluntary reporting, or oversight by a third party.

What States Have Already Done

In recent years, several states have adopted policies with the goal of increasing the efficiency of the data sharing process and facilitating research using administrative data. For example, some states have established designated data management bodies which are responsible for storing and sharing data as requested. Arkansas established the Arkansas Research Center (ARC), which houses their Statewide Longitudinal Data System (SLDI). The SLDI hosts data from several state agencies. According to their website:

> The Arkansas Research Center uses a Dual Database Architecture, which incorporates rigorous protocols for the protection of individual privacy and confidentiality. All data received from agencies is processed through a system that splits the file into two parts, personal and non-personal information. The system then generates a unique identifier for each record which is then encrypted. All data used for research and evaluation purposes are deidentified using this unique, dual database encryption approach. Having the deidentified data allows for rigorous research in a confidential and secure manner.

Pennsylvania established its Open Data Portal, which allows for public access to data on a wide variety of topics, ranging from voting statistics to employment records to drug use in schools. According to its website:

> The Commonwealth of Pennsylvania Open Data Portal was launched in August 2016 with an aim to make government data more accessible to the public to support government transparency and openness, spur social and economic benefits of government data, and empower citizens and businesses to innovate and create with government data. This data can be viewed, analyzed, visualized and exported all on one platform. This is the state's repository for publicly accessible open data owned by state agencies.

Similarly, Kentucky established the Kentucky Center for Statistics, which, according to its website, "…collects and links data to evaluate education and workforce efforts in the Commonwealth. This includes developing reports, responding to research requests, and providing statistical data about these efforts so policymakers, practitioners, and the general public can make better informed decisions."

Legal Considerations

Every state has laws regarding data privacy and the confidentiality of administrative data that will need to be accounted for when crafting data sharing agreements. The State Data Sharing initiative has provided an excellent database that catalogues laws regarding unemployment insurance and corporate tax data.

Third-Party Evaluators

The exponentially growing amount of data created and managed by agencies creates the potential for more nuanced ways of evaluating the impact of programs for participants. Many federal grant programs require that grantee organizations participate in evaluation activity by sharing administrative data with third-party evaluators. Third party evaluators also are often required to develop the agreements and protocols for sharing data related to an evaluation independently of the organization funding the evaluation. This decentralized process has made it challenging to develop centralized repositories for the management of data sharing agreements for administrative data. However, as data sharing agreements become more frequently needed, we can begin to identify trends and best practices from how these agreements are currently structured and negotiated.

Understanding the context of an evaluation and the role of administrative data is important for developing an effective data sharing agreement. Evaluations of program impact often utilize multiple data sources as part of the research design. Evaluators tend to solicit data that would allow them to create linkages across multiple data sources in their study. Since these data sources tend to be managed by different data systems, evaluators typically use PII to provide the link across data sources. Evaluators also can turn to probabilistic matching techniques in situations where PII is not available, but this technique is less reliable than utilizing PII. For example, the National Directory of New Hires is a federal repository of employment information and is often used by evaluators to assess employment outcomes for participants. Personally Identifiable Information is necessary to extract employment data from this directory.

An evaluator's data request should be tailored as narrowly as possible to the scope of the evaluation. This requires the evaluator to explain the research design in plain language that is accessible to diverse audiences. The process of tailoring the data request will often involve the following considerations:

- **Scope:** Evaluators will define the relevant population as specifically as possible. For example, an evaluator may be interested in participants of programs funded by Wagner-Peyser grants.
- **Instances:** Depending on the study design, an Evaluator may need to assess a given participant at multiple points in time. The data request should specify whether the data will be pulled once or multiple times, and the specific timeframe for each pull.
- **Duration:** The Evaluator will often define a particular time frame where participants are relevant for a study. This sometimes can be aligned to reporting period requirements or the specific study design.

An understanding of how data is organized within an organization also is critical for structuring effective data sharing agreements. For example, most U.S. Department of Labor (DOL), Employment and Training (ETA) grantees currently upload their Participant Individual Record Layout (PIRL)[1] data to the Workforce Integrated Performance System (WIPS). After grantees upload their PIRL file into WIPS, the system generates a Quarterly Performance Report (QPR) that provides aggregated characteristics of program participants, Workforce Innovation and Opportunity Act (WIOA) indicators of performance[2] outcomes of their program. States and territories that administer ETA grant programs may have multiple organizations that collect and generate data required for WIPS' reporting. In addition, many grantee organizations store other critical information that does not fall within the PIRL's parameters or the WIOA indicators of performance.

States vary in the level of integration across data systems serving a given organization. State agencies also can vary in their organizational structure and assign stewardship of administrative data between elements of an organization. A data sharing agreement should include all relevant parties in the organization providing data for the study. All parties to the agreement should have a clear

[1] ETA uses the Participant Individual Record Layout (PIRL) as data layout that instructs grantees about which data to collect on grant program participants.

[2] WIOA indicators of performance include: 1) Employment Rate – 2nd Quarter after Exit, 2) Employment Rate – 4th Quarter after Exit, 3) Median Earnings – 2nd Quarter after Exit, 4) Effectiveness in Serving Employers, 5) Credential Attainment, and 6) Measurable Skill Gains.

understanding of where the data is stored and who the relevant stewards of the data are. However, this can be challenging with the amount of variation in organizational structure across state and local agencies as well as private organizations.

The levels of documentation around data systems poses a related challenge for structuring effective data sharing agreements. While centralized systems, such as WIPS, include detailed documentation on how data is gathered and structured, organizations with less centralized systems can vary in the amount and quality of technical documentation. The quality of documentation can pose a particular challenge in the initial stages when determining the location of requested data and the relevant stakeholders. Evaluators either may not be aware of or have access to the documentation needed to finely tailor their data requests. This could lead to confusion around the most effective method for extracting the relevant data.

In addition to the technical challenges involved with managing a data sharing negotiation, evaluators must navigate a complex and expanding web of federal and state policies related to data privacy and confidentiality. For example, the Family Educational Rights and Privacy Act (FERPA) is a federal law that outlines, among other things, the rights of parents to have some control over the dissemination of PII and related information from educational records. Many, but not all, data requests which are within the scope of a federally funded evaluation are excluded from these dissemination rules, but state agencies may not be aware of this exclusion. Evaluators should make clear at the outset of a negotiation whether the party they are contracted with is eligible for such exceptions. Similarly, some states may apply different rules related to privacy and confidentiality depending on how they categorize the contractor. For example, if they consider the contractor a vendor for the purposes of managing the agreement, they may mistakenly apply rules that are not relevant for an evaluator and complicate the negotiation unnecessarily.

The management of Unemployment Insurance (UI) wage information is an example of the complexities involved with negotiating data sharing agreements within a given state agency. Employment and wage data stored by these programs is intentionally difficult to access for purposes beyond administering the state program. Because these programs and their data systems are managed by states, the structure of their data as well as the regulations around confidentiality and data security vary widely. There are several programs, such as the Longitudinal Employer-Household Dynamics program and the National Directory of New Hires, that attempt to aggregate UI wage information from multiple states, but these programs tend to provide aggregate wage information unless they have a particular agreement with the evaluators.

In addition, state policies vary in how they address breaches of data security and confidentiality. These policies vary specifically on the type of remediation required, the amount of insurance needed by evaluators to get access to data, and how they address issues of indemnification. This variation can create limitations on an evaluator’s ability to negotiate data sharing agreements with a given state agency. Outlining state-specific requirements at the initial phase of the negotiation is important for evaluators to successfully negotiate a data sharing agreement that adequately protects the state and participants in the study.

Conclusion

Understanding data sharing agreements is crucial in today’s data-driven world. Personal and organizational data flows across various platforms and geographic borders. These agreements define the legal, ethical, and practical frameworks governing how data is collected, shared, and used by different entities. By outlining responsibilities, risks, and compliance requirements, they help protect both the data subject’s privacy and the sharing parties from legal liabilities. For individuals, comprehending these agreements ensures that they are aware of how their data will be used and can make informed decisions about consent. Adhering to these agreements is essential for maintaining trust, mitigating risks, and complying with regulatory standards for many organizations.

Finally, data sharing agreements play a pivotal role in facilitating collaboration between businesses, research institutions, and governments. These agreements enable the seamless exchange of data, fostering innovation and driving advancements in fields such as healthcare, technology, and scientific research. However, with this increased collaboration comes the need for robust frameworks to safeguard sensitive data and ensure ethical use. As data privacy concerns continue to grow, developing transparent, fair, and enforceable data sharing agreements becomes increasingly vital to balancing the benefits of data sharing with the protection of individual rights and privacy.

Resources

1. Apple Support https://support.apple.com/en-us/102647

2. Electronic Frontier Foundation: 2017. How to Debug Your Content Blocker for Privacy. https://www.eff.org/deeplinks/2017/11/how-debug-your-content-blocker-privacy-protection

3. Wired 2023 The New Era of Social Media Looks as Bad for Privacy as the Last One. https://www.wired.com/story/x-alternatives-user-privacy-report/

Disclaimer

The Apprenticeship Data and Performance Technical Assistance Center (ADAPTAC) is a four year, four-million-dollar collaboration between the U.S. Department of Labor (DOL) Employment and Training Administration (ETA), The Council of State Governments, Mathematica Inc., and The Turnout, LLC. ADAPTAC began in 2021.

This workforce product was funded by a grant awarded by the DOL/ETA. The product was created by the recipient and does not necessarily reflect the official position of DOL/ETA. DOL/ETA makes no guarantees, warranties, or assurances of any kind, express or implied, with respect to such information, including any information on linked sites and including, but not limited to, accuracy of the information or its completeness, timeliness, usefulness, adequacy, continued availability, or ownership. This product is copyrighted by the institution that created it.