Data Confidentiality: Real World Considerations

Jordan M. Robbins

July, 2023

Sharing Data: Key Considerations

- As researchers, practitioners, and leaders, we rely on data to understand the world in which we work, make informed decisioned, and guide projects and initiatives.

- While improved access to data remains a key goal in our field, real world considerations will often impact how we share data, who we can share it with, and what data we share.

- Key considerations:
  - How do we limit risk?
  - Are there policies, laws, and regulations limiting access?
  - Is there a legitimate need to know?
  - What is the scope of the data being requested?
  - What is the quality of the data?

How Data is Shared

- Risk associated with data sharing can be moderated by limiting access to raw data or identified data.

- Raw vs. Transformed:
  - Raw data can be transformed to limit disclosure of sensitive information.
  - Includes techniques such as aggregating individual responses and removing sensitive information.
- Identified vs. De-Identified:
  - De-identification or anonymization is used to maintain confidentiality and prevent data from being linked back to a given source (e.g. person, group, or organization).
  - When disclosure poses a potential risk to an individual de-identification can be used.

	Identified Data	De-Identified Data
Raw Data	Higher Risk	Moderate Risk
Transformed Data	Moderate Risk	Lower Risk

Policies, Laws, and Regulations

- In many cases policies, laws, and regulations may limit what data can be accessed and by whom.
- Policy and Security Protocols:
  - Policies and security protocols can dictate who can access data and what can be shared.

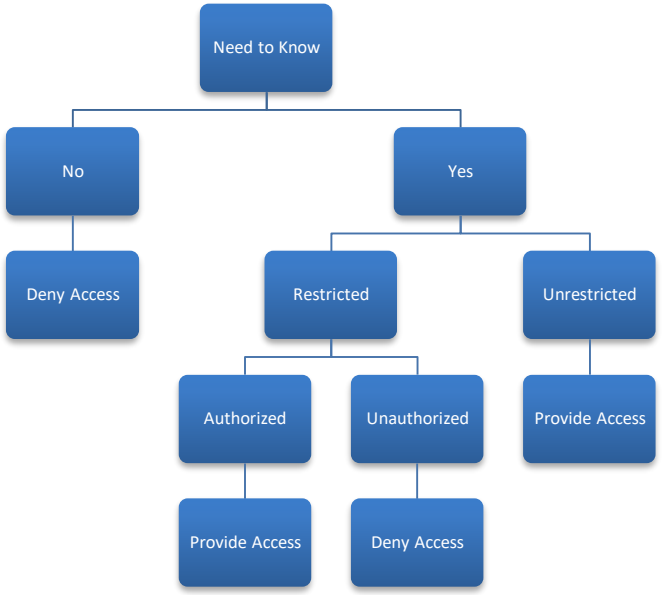
- Legal Considerations:
  - Federal and state laws may govern what data can be disclosed.
  - Does sharing data increase exposure to liability?

- Consent:
  - Is consent needed to release information to others?
  - Informed consent during data collection may specify how data is being used and by whom.

- These constraints do not necessarily prevent sharing from occurring, but it is important that we understand these limitations so that we can work with them when possible.

Need to Know

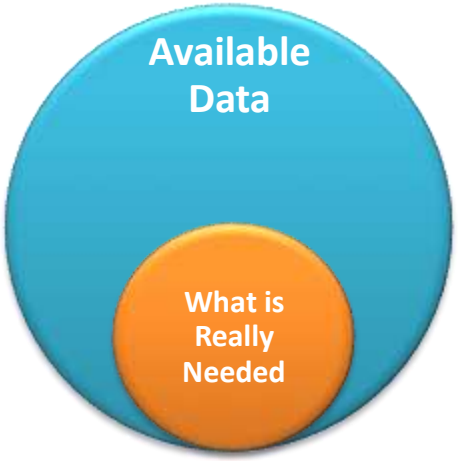
- Is there a legitimate reason that the data should be shared?
  - Job specific duties within an organization
  - Address the mission or goals of the organization
  - Fulfill the legal obligations or requirements an organization has (e.g. disclosure to governing bodies, the public, etc.)
  - Benefits the public or furthers understanding of issues (when allowed)



- While having a legitimate reason to access the data may help reduce risk and limit exposure to liability, we must also take into account the nature of the request and data quality

Scope of Request & Data Quality

- The scope of the request should guide what data is shared.
  - Recommend limiting the data to what is needed.
  - Requires knowledge of what is in a dataset, and where the data is located.
  - Requires and understanding of the questions being asked and how the data element defined.
  - Requests can involve specific elements from multiple data sources.



- Data quality issues can pose unique challenges.
  - Varies over time and between components as data governance policies evolve and the reasons for the data collection change.
  - Can produce misleading analyses and interpretations.
  - Important that an organization understands limitations in the data and communicates this to others.

Key Takeaways

- Understand your organizational environment and responsibilities including laws, policies, and related considerations.
- When possible, share what is needed, but limit oversharing of data.
- Understand your data and its limitations.
- Communicate limitations and unknowns.
- Take steps to reduce risk where possible.

Questions?