

# **Evaluation: Trust and Transparency in Data Sharing**

This evaluation delves into the theme of "Trust and Transparency in Data Sharing," a critical aspect underpinning the willingness of researchers and institutions to contribute to and utilize data sharing platforms. This theme explores the intricate concerns related to data security, privacy, ownership, and responsible use – all pivotal elements in fostering a culture where data sharing becomes the norm rather than the exception. Building this trust necessitates a multi-pronged approach, encompassing robust security measures, clearly defined data governance policies, transparent communication strategies, and robust accountability mechanisms. By addressing potential risks and demonstrating a commitment to ethical data stewardship, data sharing platforms can foster confidence among data contributors and unlock the vast potential held within shared datasets.

## **Subtheme 1: Data Security and Privacy Concerns**

At the forefront of researchers' minds when considering data sharing are the very real concerns surrounding data security and the potential for breaches of participant privacy. These concerns are not merely hypothetical; they are rooted in the recognition that sensitive cohort data, often containing personal health information, requires the utmost protection from unauthorized access and potential misuse. A common thread throughout the interviews is the emphasis on robust security protocols as a cornerstone of trustworthy data sharing. Researchers expressed a clear need for platforms to employ stringent security measures, encompassing data encryption, access controls, and secure storage solutions. The implementation of such measures serves to mitigate the risks of data breaches and unauthorized access, building a foundation of trust by demonstrating a commitment to safeguarding sensitive information. For instance, one researcher highlighted the importance of controlling data access, stating, "So at all times we keep the control of the data, so we make sure that the data is versioned, and we made an IT system where you, there is no possibility that you can download the data to your local PC, that helps." [4:4.22] This example underscores the need for platforms to prioritize data security measures that prevent unauthorized downloads and maintain data integrity. Another layer of concern revolves around the concept of data ownership and control. Researchers, particularly those heavily invested in data collection and analysis, expressed anxieties about relinquishing control over their datasets. The fear of potential misuse or unauthorized analysis by external parties emerged as a barrier to wider data sharing. Addressing this concern requires transparent data governance models that clearly outline data usage agreements, specify permitted analyses, and provide mechanisms for researchers to retain a degree of control over how their data is accessed and utilized.

## **Subtheme 2: Lack of Trust in Data Recipients and Platforms**

Beyond the technical aspects of data security, a recurring theme in the interviews was a palpable lack of trust in both the entities receiving shared data and the platforms themselves. This distrust often stemmed from concerns about data misuse, a lack of transparency in data governance, and uncertainty regarding the security measures employed by platforms. Researchers expressed a clear preference for data sharing within established consortia or collaborations built on years of mutual trust and understanding. This preference highlights the importance of pre-existing relationships and shared research goals in fostering a culture of data sharing. Within such collaborative frameworks, researchers reported a higher level of comfort in sharing their data, knowing that their collaborators understood the nuances of the data and were committed to using it responsibly. In contrast, sharing data on open platforms or with external researchers who were not part of their established networks elicited greater hesitation. This hesitancy often stemmed from a lack of familiarity with the recipients, uncertainty about their intentions, and concerns about potential misinterpretations of the data. The experience of one researcher illustrated this

point vividly: "That those who are sharing the data, are always sharing it for collaborative research, that they can trust that the data is treated as it should be treated and we have seen, I have seen over the years many situations where consortia have disappeared through it because it appears that they have been doing as well some other analyses on the data without the centers knowing." [6:6.4] This example highlights a critical challenge – the need for platforms to bridge the trust gap that exists when data sharing extends beyond established networks.

### **Subtheme 3: Ethical Considerations and Informed Consent**

The ethical dimensions of data sharing were woven throughout the interviews, emphasizing the need to balance the immense potential of shared data with the rights and expectations of research participants. Informed consent emerged as a cornerstone of ethical data sharing practices. Researchers highlighted the importance of obtaining consent that explicitly addressed data sharing, outlining the potential benefits, risks, and safeguards in place. This emphasis on informed consent underscores the need for clear and transparent communication with participants throughout the research process. The role of ethical review boards and data access committees was also frequently mentioned, signifying a shared understanding of the need for independent oversight to ensure that data access requests aligned with the original consent given by participants. For instance, a researcher stated, "If you want to do some work in X on CONSORTIUM B data, you write a research proposal and of course you must have an ethical vetting." [1:1.15] This statement reinforces the crucial role of ethical vetting in data sharing, ensuring that proposed analyses are aligned with the scope of informed consent and uphold the highest ethical standards.

### **Subtheme 4: Data Misinterpretation and Misuse Concerns**

A nuanced layer within the theme of trust relates to anxieties about the potential for data misinterpretation and misuse. Researchers acknowledged the inherent complexity of their datasets, often shaped by specific study designs, data collection methods, and contextual factors that might not be immediately apparent to external parties. This concern underscores the importance of comprehensive data documentation and metadata standards. Providing detailed information about the study context, data collection protocols, and variable definitions is essential to minimize the risk of misinterpretations and ensure that data is used appropriately. Furthermore, promoting data literacy emerged as a critical need. Equipping researchers with the skills to critically evaluate shared datasets, understand their limitations, and conduct robust analyses is essential to mitigating the risks of data misuse and maximizing the value of data sharing. One researcher aptly captured this concern: "Second reason is to know the local context and how data was collected and why you find this and that in the paper and it can go very, very wrong if you don't know this extra information" [1:1.16]

### **Subtheme 5: Community Engagement and Building Trust**

A consistent message throughout the interviews was the importance of community engagement in fostering a culture of trust in data sharing. This engagement encompasses not only the research community but also patient groups, policymakers, and the wider public. Open communication emerged as a cornerstone of community engagement. Researchers stressed the need for transparency in data governance policies, clearly communicating data access procedures, authorship guidelines, and the potential benefits of data sharing. A researcher emphasized this, stating, "And I think the main reason why this was a smooth process was that all the process was transparent, very clear and all the participants were put on the same level." [5:5.18] This transparency is essential to building trust and fostering a sense of shared ownership in the data sharing endeavor.

## **Conclusion: A Foundation for Progress**

In conclusion, the theme of "Trust and Transparency in Data Sharing" emerged as a dominant force shaping researchers' perceptions and willingness to engage with data sharing platforms. The analysis revealed a complex interplay of concerns surrounding data security and privacy, a lack of trust in recipients and platforms, ethical considerations, anxieties about data misinterpretation, and the need for proactive community engagement. Addressing these concerns is not merely a matter of implementing technical solutions; it demands a fundamental shift towards a culture of transparency, accountability, and shared responsibility in data stewardship. By prioritizing these principles, data sharing platforms can create an environment where researchers feel confident in contributing their valuable data, ultimately unlocking a future of accelerated scientific discovery and improved human health.

## **Evaluation: Overcoming Barriers to Data Sharing**

This theme delves into the multifaceted barriers hindering widespread cohort data sharing, encompassing legal and regulatory hurdles, technical challenges, resource limitations, and cultural norms. Understanding these obstacles is crucial for developing effective data sharing platforms and fostering a collaborative research environment. Our analysis highlights the urgent need for practical solutions, including harmonized regulations, standardized infrastructure, sustainable funding models, and a cultural shift towards open science.

### **Subtheme 1: Legal and Regulatory Barriers to Data Sharing**

Navigating the intricate and evolving landscape of legal and regulatory frameworks is a significant challenge for researchers engaged in data sharing initiatives. This subtheme specifically explores the complexities associated with data privacy, data protection, and ethical research conduct.

Our findings reveal a sense of ambiguity surrounding the interpretation and implementation of data protection regulations, particularly the General Data Protection Regulation (GDPR). Interviewees [2:2.2] and [3:3.1] expressed concerns over what they perceived as overly cautious legal interpretations of GDPR, especially in certain countries. They find that legal professionals often err on the side of caution, making data access and sharing difficult even within the same institution, hindering research progress. This cautious approach, while well-intentioned, can create roadblocks for researchers, particularly those involved in international collaborations. The lack of clear, consistent guidelines across borders adds another layer of complexity, as highlighted by interviewees [1:1.3] and [2:2.6] who experienced difficulties due to varying interpretations and implementations of GDPR across different countries and institutions. This inconsistency breeds confusion and uncertainty, potentially discouraging researchers from engaging in cross-border data sharing.

Adding to these complexities, collaborations with researchers outside the EU, particularly those from the US [3: 3.7, 3.8], often encounter difficulties due to differing data protection standards. The unfamiliarity with GDPR's stricter requirements among non-EU researchers poses a significant challenge in aligning international research practices with the EU's data privacy regulations.

These findings underscore the urgent need for harmonized regulations, streamlined procedures, and standardized data sharing agreements to facilitate compliant and efficient data sharing in the international research landscape.

## **Subtheme 2: Technical Challenges in Data Sharing**

This subtheme explores the technical hurdles inherent in data sharing, moving beyond legal frameworks to examine the practicalities of data exchange and analysis. Our analysis reveals that technical challenges, often intertwined with resource constraints, present significant obstacles to seamless data sharing.

A key concern identified is the need for secure and controlled data access. Interviewee [4: 4.8, 4.9, 4.10] highlighted the use of virtual environments and data versioning as mechanisms to maintain data integrity and prevent unauthorized downloads, demonstrating a proactive approach to mitigating data security risks. However, implementing such solutions requires specialized technical expertise and resources, which are not always readily available to research teams.

Furthermore, researchers [4:4.5, 4.6, 4.7] emphasized the importance of data standardization and interoperability for effective data sharing and analysis. While harmonization of existing cohort data is crucial, it presents greater challenges compared to prospective harmonization due to varying data collection methods, languages, and interpretations of concepts across different studies. Aligning questionnaire data on subjective experiences, for example, is considerably more complex than simply translating units of measurement.

Interestingly, the emergence of de facto international standards through the adoption of standardized procedures and variables in large-scale cohort studies, as observed by interviewee [6:6.10], offers a potential pathway for greater harmonization in the future. The credibility and comparability of research findings are enhanced when data sharing platforms adhere to and enforce widely accepted standards, as suggested by interviewee [4:4.11].

## **Subtheme 3: Time and Resource Constraints for Data Sharing**

This subtheme examines the significant resource demands associated with data sharing, highlighting the substantial time, financial investment, and specialized expertise required for successful implementation.

Our findings indicate that data sharing is a resource-intensive endeavor, placing a considerable burden on researchers already operating with limited resources. Interviewee [6:6.16, 6.17] emphasized that data sharing requires dedicated time and resources not just for the initial data collection but also for ongoing maintenance, preparation for sharing, and responding to data access requests. This ongoing commitment can be a significant challenge, particularly for researchers working on limited budgets and tight timelines.

The lack of dedicated funding and support specifically allocated for data sharing activities further exacerbates these constraints [6:6.16, 6.17]. While there is a growing push for data sharing, researchers often struggle to secure the necessary resources to meet these demands, highlighting the need for funding models that adequately support these efforts.

Our analysis also reveals a significant challenge in finding and retaining individuals with the appropriate data expertise, particularly within large institutions [7:7.11]. This shortage of qualified personnel often creates bottlenecks and delays in preparing and sharing data. The lack of dedicated IT support further complicates matters, often leading to researchers without formal IT backgrounds taking on the technical aspects of data sharing [4:4.2, 4.8], diverting their time and energy from their primary research activities.

## **Subtheme 4: Sustainability and Funding for Data Sharing Platforms**

This subtheme specifically addresses the long-term sustainability of data sharing platforms, exploring the financial and structural challenges they face.

Securing sustainable funding for data sharing platforms emerged as a critical concern [1:1.1, 6: 6.1, 7: 7.17]. Researchers highlighted the difficulties in obtaining long-term support for data curation, infrastructure maintenance, and staffing, often relying on short-term grants tied to specific projects. This reliance on project-based funding creates uncertainty and jeopardizes the long-term viability of these platforms.

The findings reveal a tension between the push for open data and the need for sustainable funding models that adequately compensate researchers for their data collection and curation efforts [5:5.3, 4: 4.4, 7:7.14]. While many researchers recognize the value of data sharing, the current academic reward system, primarily focused on first and last authorship, provides little incentive for sharing data openly.

## **Subtheme 5: Cultural Norms and Resistance to Data Sharing**

This subtheme explores the deeply ingrained cultural norms and attitudes within the research community that can hinder data sharing, highlighting the need for a cultural shift towards openness and collaboration.

Our analysis suggests that a sense of ownership over data and a perceived fear of losing control over valuable research outputs can create resistance to data sharing [2: 2.10, 4: 4.3, 7: 7.16]. Researchers, having invested significant time and effort in data collection, may be reluctant to share their data openly due to concerns about competition, lack of recognition for their contributions, and potential misuse of their data.

This reluctance is further compounded by the existing academic reward system, which often prioritizes individual achievement over collaborative efforts [2: 2.9, 5:5.7, 7:7.15]. Researchers may hesitate to share data openly if they perceive that doing so will not advance their careers or lead to tangible rewards.

Fostering a culture of trust and collaboration within the research community is crucial to overcoming these cultural barriers. This shift requires a collective effort from researchers, institutions, and funding bodies to recognize and reward data sharing practices and promote the benefits of open science.

## **Overall Conclusion**

Overcoming the barriers to data sharing is not merely a technical or logistical challenge but a multifaceted endeavor requiring a concerted effort to address legal, technical, resource-related, and cultural obstacles. Our analysis reveals a complex interplay between these factors, emphasizing the need for a holistic approach to developing effective solutions.

Creating a research environment where data sharing is the norm requires a multi-pronged approach that includes harmonizing data protection regulations, developing user-friendly and interoperable data sharing platforms, providing adequate funding and resources for data sharing activities, and fostering a culture that values and rewards data sharing practices.

# Maximizing the Value and Impact of Data Sharing

This theme explores the multifaceted objective of maximizing the utility, impact, and enduring value of shared cohort data, directly addressing the research question about barriers and concerns in data sharing and their implications for data sharing platforms. The analysis delves into the critical aspects of ensuring data quality, implementing standardization measures, promoting accessibility, and demonstrating the return on investment in data sharing initiatives. The theme underscores that realizing the full potential of data sharing necessitates robust data governance frameworks, transparent mechanisms for data citation, and a dedicated focus on maximizing the reusability and impact of shared datasets.

## Subtheme 1: Data Ownership and Control Issues

The complexities surrounding data ownership and control emerge as a paramount concern within the broader theme of maximizing the value of data sharing. This subtheme delves into the intricate aspects of intellectual property rights, the importance of clear data governance structures, and the continuous need for data originators to maintain a degree of control over their shared data. Participants highlighted the indispensable role of comprehensive data governance frameworks that clearly delineate data access protocols, define permissible data usage rights, and provide mechanisms for data owners to exert influence over the dissemination of their data.

One key aspect identified is the establishment of rigorous oversight mechanisms for data access requests. For instance, a participant described a multi-tiered governance structure involving internal research committees, steering committees, and external ethical boards to scrutinize project proposals, ensuring scientific rigor, strategic alignment, and adherence to ethical guidelines [2:2.1]. This example underscores the effort invested in establishing transparent and accountable data governance processes.

Beyond formal governance structures, the analysis revealed a nuanced understanding of data ownership, extending beyond legal definitions to encompass individual researchers' preferences and control over their data. A participant emphasized the "particularities of such an individual researcher," highlighting how personal relationships, financial considerations, or specific collaboration goals can influence data sharing decisions [4:4.23]. This insight emphasizes the need for platforms to accommodate diverse motivations and preferences regarding data control.

Furthermore, the analysis highlighted the evolving understanding of data ownership in collaborative research, particularly when dealing with sensitive health information. A participant noted, "we don't own any of those data, they are data that have been shared with us by those centers" [6:6.1], emphasizing the role of institutions as custodians rather than outright owners of shared data. This underscores the ethical imperative for platforms to facilitate responsible data governance that respects the original data owners and their intentions.

In conclusion, addressing data ownership and control issues is not merely about legal compliance but necessitates a comprehensive approach that balances the interests of data originators, researchers seeking access, and the broader scientific community. Data sharing platforms must prioritize transparency, clearly defined data governance frameworks, and mechanisms that empower data owners to make informed decisions about accessing and utilizing their valuable data assets.

## Subtheme 2: Lack of Awareness and Understanding of Data Sharing

This subtheme brings to light the persistent challenge of limited awareness and understanding of data sharing practices, benefits, ethical dimensions, and available platforms within research communities. This lack of awareness emerged as a significant barrier to maximizing the value of

data sharing, highlighting the crucial need for targeted interventions to promote data literacy and foster a culture of data sharing within research ecosystems.

One striking observation from the analysis is the lack of clarity surrounding data sharing regulations, even among researchers themselves. Despite the implementation of the General Data Protection Regulation (GDPR), a participant remarked, "nobody knows how to use them" [2:2.10], referring to the challenges in interpreting and applying the regulations in practice. This ambiguity creates hurdles in navigating data sharing processes effectively and underscores the need for clear, accessible guidance on regulatory compliance.

Beyond regulatory uncertainty, the analysis revealed a lack of understanding regarding the specific data available within existing cohorts. A participant recounted instances where researchers requested data that was not collected or readily accessible, highlighting the importance of transparent and comprehensive data catalogs [7:7.16]. Providing researchers with easily searchable and informative data catalogs will streamline data discovery and promote appropriate data requests, maximizing the utility of existing data resources.

In conclusion, addressing the lack of awareness and understanding regarding data sharing practices, regulations, and available platforms is paramount to fostering a data-driven research culture. Data sharing initiatives and platforms must prioritize educational initiatives, training programs, and outreach efforts to bridge knowledge gaps. By equipping researchers with data literacy skills, promoting a clear understanding of ethical and legal frameworks, and providing transparent access to data resources, the research community can move towards a future where data sharing becomes the norm rather than the exception.

### **Subtheme 3: Lack of Incentives and Recognition for Data Sharing**

This subtheme delves into the crucial issue of inadequate incentives and recognition for researchers who engage in data sharing, a factor that significantly hinders efforts to maximize the value of data sharing. The analysis underscores a critical disconnect between the increasing demand for data sharing and the lack of tangible rewards for researchers who contribute their valuable data resources. This imbalance creates a disincentive for researchers to engage in data sharing activities, potentially impeding the progress of research that relies on collaborative data access.

One prominent theme emerging from the analysis is the mismatch between the existing academic reward system and the contributions made through data sharing. The current system heavily emphasizes publications, with a strong focus on first and last authorship, neglecting to adequately acknowledge and reward the efforts involved in data generation and sharing. A participant noted, "you get rewarded by the publications you have, the authorship you have" [4:4.25], highlighting how the current system prioritizes individual achievements over collaborative efforts, even when data sharing contributes significantly to research advancements.

The lack of formal recognition for data contributions manifests in various ways. For instance, a participant highlighted the prevalence of excessively long author lists, where "sometimes you have more authors in a paper than words in the text itself" [4:4.25]. This observation underscores the tendency to simply add names to publications as a form of acknowledgment, diluting the significance of individual contributions and failing to provide meaningful credit for data sharing efforts.

Furthermore, the analysis revealed that the absence of clear guidelines and standards for attributing data contributions exacerbates the challenge. A participant questioned the role of journals in addressing this issue, asking, "as long as there are mechanisms for, for providing the merit as a co-author or non-author contributor" [6:6.21], highlighting the need for standardized mechanisms to acknowledge and reward data contributions transparently and fairly.

In conclusion, the lack of incentives and recognition for data sharing represents a systemic barrier to maximizing the value of shared data. A fundamental shift is required in how the research community values and rewards contributions beyond traditional publication metrics.

Implementing concrete measures to acknowledge data contributions, such as standardized data citation practices, alternative metrics for research evaluation, and institutional policies that recognize data sharing efforts, will be crucial in fostering a research culture that truly values and incentivizes data sharing as a valuable and integral part of scientific progress.

## **Subtheme 4: Data Governance and Management Complexity**

This subtheme tackles the intricate challenges associated with establishing, implementing, and navigating complex data governance and management frameworks in the context of data sharing initiatives. The analysis reveals that the administrative burden and intricate processes often involved in data access and sharing can act as significant impediments to realizing the full potential of data sharing, particularly as the scale and complexity of research collaborations continue to grow.

One prominent challenge highlighted by participants is the multi-layered review process involved in granting data access, particularly for external collaborations. A participant described a system with "a kind of a government structure" involving research committees, steering committees, data transfer agreements, and potentially external ethical board reviews [2:2.1]. While these measures are essential for ensuring responsible data use and protecting participant privacy, the complexity and potential for lengthy approval timelines can discourage researchers and hinder the agility of research collaborations.

Further emphasizing the tension between robust data governance and bureaucratic hurdles, a participant emphasized the need for "as less bureaucratic solution as possible" [4:4.24], recognizing the importance of streamlining processes while maintaining ethical and responsible data handling practices. This underscores the need for data sharing platforms to strike a balance between rigorous oversight and user-friendly, efficient procedures.

The analysis also revealed that the specific requirements for data access can vary significantly depending on the nature of the collaboration and the institutions involved. A participant noted the increased scrutiny applied to external data access requests compared to internal projects, particularly when data transfer outside the institution is involved, often necessitating formal data transfer agreements and involvement from legal departments [3:3.17]. This variability in procedures adds another layer of complexity to data sharing, particularly for researchers navigating collaborations across institutional and national boundaries.

In conclusion, simplifying and streamlining data governance and management processes is crucial to maximizing the value and impact of data sharing. Data sharing platforms should prioritize user-friendly tools, standardized data governance models, and clear, readily accessible guidelines to alleviate the administrative burden on researchers. By simplifying these processes while upholding ethical and legal standards, data sharing initiatives can encourage broader participation and facilitate more efficient, impactful research collaborations.

## **Overall Conclusion**

This in-depth analysis of the "Maximizing the Value and Impact of Data Sharing" theme reveals a complex interplay of factors that influence the success of data sharing initiatives. The subthemes discussed—data ownership and control issues, lack of awareness and understanding, insufficient incentives for data sharing, and complex data governance structures—all present significant barriers to realizing the full potential of shared data. Addressing these challenges is crucial for



fostering a research environment where data is readily shared, ethically managed, and effectively utilized to drive scientific progress.

The findings highlight the need for a multifaceted approach that addresses both technical and cultural aspects of data sharing. This includes developing clear, standardized data governance frameworks that balance data security and accessibility, promoting data literacy and awareness among researchers through targeted training and education programs, and establishing robust mechanisms for recognizing and rewarding data contributions. Additionally, data sharing platforms play a vital role in this ecosystem. They must prioritize user-friendly interfaces, robust data management tools, and transparent data access procedures to simplify the process for researchers and encourage broader participation in data sharing endeavors. By addressing these challenges head-on, the research community can unlock the full potential of data sharing, fostering a collaborative and transparent research culture that accelerates scientific discovery and maximizes the societal impact of research.