Evaluation Section: Barriers and Concerns in Cohort Data Sharing

This evaluation section delves into the barriers and concerns associated with sharing cohort data, as revealed through inductive thematic analysis of interviews with researchers involved in managing and sharing such data. The analysis identifies three overarching themes: **Control over Data**, **Data Sharing Practices and Culture**, and **Incentives and Evaluation**. Each theme encompasses several subthemes, providing a nuanced understanding of the challenges inherent in promoting data sharing in research.

Theme 1: Control over Data

This theme focuses on the multifaceted aspects of control surrounding cohort data. It explores how researchers perceive ownership and responsibility for the data they generate, manage, and share, and the impact of these perceptions on their willingness to participate in data sharing initiatives. This theme is crucial to understanding the barriers to data sharing as it reveals the inherent tensions between individual interests and the broader goal of maximizing scientific progress through collaboration.

1.1: Perceived Ownership and Responsibility

Researchers often exhibit a strong sense of ownership and responsibility towards the cohort data they collect, manage, and share. This sense of ownership stems from years of dedicated work, securing funding, navigating ethical approvals, and overseeing data collection, quality control, and analysis. One interviewee, reflecting on the substantial effort involved in cohort development, stated: "We have worked now for ten years to set up this study and it was a lot of work, and now we have data." This perceived ownership manifests as a desire to maintain control over the data's use, ensuring its proper handling and accurate interpretation. As one interviewee explained: "The people at X set up the study and they had some ideas already in mind and this is what they want to do."

1.2: Fear of Misinterpretation and Misuse

The perceived ownership of cohort data is often intertwined with concerns about potential misinterpretation and misuse by external researchers. Researchers, deeply familiar with their data's intricacies, express anxiety about others lacking the necessary context and expertise to analyze it correctly. One interviewee highlighted this concern: "The problem is that the X study is a very complex study because it is so old and it has very different arms, and they were all research-driven... We feel at the moment that... in many cases, the active involvement of X scientists is needed." This fear of misinterpretation can lead researchers to prioritize collaborations where they retain a degree of control over the analysis process, ensuring the data is used appropriately and their scientific reputation is protected.

1.3: Data Protection and Legal Restrictions

Beyond personal perceptions of control, researchers are increasingly grappling with the complexities of data protection regulations, particularly the General Data Protection Regulation (GDPR) in Europe. While recognizing the importance of safeguarding personal data, interviewees expressed frustration with the ambiguity surrounding GDPR implementation and interpretation. One interviewee stated: "It's not really the GDPR itself but that nobody, this is my impression, but that nobody knows how to use them." This uncertainty can create significant barriers to data

sharing, particularly for international collaborations where legal frameworks and interpretations may differ.

1.4: Institutional Policies and Governance

Institutional policies and governance structures can further complicate data sharing efforts, particularly when multiple stakeholders are involved. Interviewees highlighted the challenges of navigating complex approval processes and obtaining permissions from various authorities, often citing a lack of clarity and coordination. One interviewee described this challenge: "So, between two authorities within the same town, the same house... No personal identification number, nothing... If it's a thousand, then of course I continue to the whole process et cetera, et cetera. But if it is one person, then perhaps I have to wait another 100 years." These institutional barriers can create significant delays and discourage researchers from engaging in data sharing initiatives.

Theme 1 Conclusion: Navigating Control in a Collaborative Landscape

The theme of control over data reveals the inherent tensions researchers face when balancing their desire to maintain control over their data with the need for collaborative research. Perceived ownership, fear of misinterpretation, data protection regulations, and institutional barriers all contribute to this tension, highlighting the need for clear guidelines, robust governance structures, and collaborative approaches that address both individual concerns and the broader goal of advancing scientific knowledge through data sharing.

Theme 2: Data Sharing Practices and Culture

This theme explores the established practices and prevailing culture surrounding data sharing within the research community. It examines how researchers approach data sharing, the influence of existing collaboration models, and the impact of these practices on the development and use of data sharing platforms. This theme is essential to understanding the barriers to data sharing as it reveals the deeply ingrained habits and norms that can hinder the adoption of more open and collaborative research practices.

2.1: Collaborative Research Models

Existing collaborative research models heavily influence data sharing practices. Researchers often prioritize collaborations with long-standing partners, fostering trust and shared understanding over time. One interviewee described this preference: "In the end, most of our data usage is from within and from collaborators who we work with ten, twenty, or sometimes thirty years." While valuable, this model can create a closed system where external researchers find it challenging to gain access to data, limiting the potential for novel research questions and diverse perspectives.

2.2: Authorship and Credit Attribution

Authorship and credit attribution remain paramount in the research world, often dictating researchers' willingness to share data. The prevailing system, with its emphasis on first and last authorship, can lead to extensive author lists in collaborative projects, sometimes exceeding 100 contributors. One interviewee, reflecting on this trend, stated: "I just finished to input 86 names for a paper... that I did in the morning." While recognizing the contributions of data generators, this hyper-authorship phenomenon raises concerns about accountability and dilutes the recognition for individual contributions, particularly those in the middle of lengthy author lists.

2.3: Transparency and Communication

Open and transparent communication is crucial for successful data sharing. Interviewees emphasized the importance of providing clear information about available data, access procedures, and data use policies. One interviewee stressed: "And another way to expose our data for the medical community is a very, very clear information webpage site, information on the web of how to access data and how to do it." Establishing clear communication channels and fostering a culture of transparency can help build trust among researchers and facilitate more effective data sharing.

2.4: Data Management and Standardization

Standardized data management practices are essential for efficient data sharing. This includes harmonizing data collection protocols, using controlled vocabularies, and providing detailed metadata. However, as one interviewee pointed out: "But often the data exists already and then you have to harmonize existing data and that is of course much more difficult." These efforts require significant resources and expertise, often lacking in individual research groups, highlighting the need for centralized support and infrastructure to facilitate data harmonization.

Theme 2 Conclusion: Shifting Practices and Cultivating a Collaborative Culture

The theme of data sharing practices and culture reveals the need for a shift away from traditional models that prioritize individual recognition and control towards more open and collaborative approaches. Fostering trust through transparent communication, establishing clear data management standards, and re-evaluating authorship models are crucial steps in cultivating a culture that embraces data sharing as a fundamental principle of scientific progress.

Theme 3: Incentives and Evaluation

This theme examines the current systems of incentives and evaluation in research and their impact on data sharing practices. It explores how researchers are rewarded for their work, the influence of funding mechanisms, and the potential for alternative metrics to encourage data sharing. This theme is crucial to understanding the barriers to data sharing as it reveals how the current reward structure can discourage researchers from engaging in open and collaborative research practices.

3.1: Impact of Traditional Metrics

Traditional research evaluation metrics, primarily focused on publication output and journal impact factors, create a strong incentive for researchers to prioritize individual publications over data sharing. One interviewee described this reality: "In our overall institution there is a shift towards first and last authorship... That they are more important and for a researcher, his career plans, I think they are not so highly valuable as first or last author papers." This focus on individual achievement can disincentivize researchers from contributing to collaborative projects where their individual recognition might be diminished.

3.2: Role of Funders in Promoting Data Sharing

Funding agencies play a crucial role in shaping research practices and promoting data sharing. By incorporating data sharing requirements into grant proposals and emphasizing the value of data reuse, funders can incentivize researchers to adopt more open and collaborative approaches. One interviewee highlighted the power of funding mandates: "I think it’s a bold step that the European

Commission has taken that they now require from the next framework programme, all data has to be fair, has to be in the European Open Science Cloud, has to be open." However, funding mechanisms also need to adapt to recognize and reward data generation and curation efforts, not just final publications.

3.3: Alternative Metrics for Data Sharing

Exploring alternative metrics for evaluating research contributions can encourage data sharing and recognize the often-invisible labor involved in data generation, quality control, and management. Data-level metrics, focusing on data quality, completeness, and reuse potential, can provide a more nuanced assessment of data generators' contributions. One interviewee advocated for this approach: "That is just what I was meaning, to have rules that qualify the data, more than the authors... to have rules that certify the quality of the data in some way, perhaps."

3.4: Data Citation and Recognition

Data citation, assigning unique identifiers (DOIs) to datasets, can enable researchers to receive credit for their data contributions, independent of publications. This approach allows for tracking data reuse and demonstrating the impact of data generation efforts. As one interviewee emphasized: "But for the platforms to combine the data or whatever, you need an ID, an identifier for the data set, you need a kind of identifier and if you use the public identifier then it goes hand in hand because then you can use (unintelligible; direct permission also) to that DOI system."

3.5: Shifting Institutional Culture and Practices

Academic institutions play a critical role in fostering a culture that values data sharing. This involves revising promotion and tenure criteria to recognize data contributions, providing support and resources for data management, and promoting collaboration across disciplines. One interviewee suggested: "Institution can do something, because can help to give some contribution for sharing data that is free, because of course you, to have good data you need a lot of work, you need person you need money to do this." Shifting institutional culture to value data sharing as a core research practice is crucial for long-term change.

Theme 3 Conclusion: Realigning Incentives and Reimagining Evaluation

The theme of incentives and evaluation highlights the need for a systemic shift in how research contributions are recognized and rewarded. Moving away from a singular focus on publication-based metrics towards a more holistic approach that values data sharing, data quality, and collaborative efforts is crucial for creating a research environment that encourages open science and maximizes the societal impact of research investments.

Overall Evaluation Conclusion: Towards a Collaborative Data Sharing Ecosystem

This evaluation reveals the complex interplay of control concerns, established research practices, and incentive structures in shaping data sharing behavior. Researchers exhibit a strong sense of ownership and responsibility for their data, coupled with concerns about misinterpretation, legal restrictions, and institutional barriers. The current academic system, with its emphasis on individual achievement and publication-based metrics, can hinder data sharing efforts.

To overcome these barriers, a multi-pronged approach is needed, involving clear data governance structures, robust data sharing platforms, and a cultural shift towards valuing collaboration and

data reuse. Funders and institutions play a crucial role in fostering this transition by incorporating data sharing requirements, recognizing diverse research contributions, and promoting a culture that values open science. Developing and implementing alternative crediting mechanisms, such as data-level metrics and data citation, can help incentivize data sharing and acknowledge the vital contributions of data generators and curators. By addressing the concerns and re-evaluating the reward systems, the research community can move towards a more collaborative data sharing ecosystem that maximizes the impact of cohort data for scientific progress and societal benefit.