IJDC | *Research Paper*

Informed Consent Contexts in a Multidisciplinary Research Data Repository

Brian Jackson
Mount Royal University

Nurulamirah Azmi
Mount Royal University

Abstract

Secondary use of research data requires an understanding of the contexts in which it was collected. While depositors are often encouraged to describe methodological and structural contexts in the form of metadata and documentation, ethical contexts have received much less attention. As open data mandates and an ethos of FAIR (findable, accessible, interoperable, reuseable) data proliferate across disciplines, participant consent for unknown future secondary uses of data is increasingly sought, even for minimal risk research. Terms of broad consent generally establish limitations on data reuse, but those limitations may not be clear when data are accessed via an open repository. The absence of these contexts increases the risk that secondary uses of data will be inconsistent with the expectations of original research participants and may place unnecessary burden on research ethics boards.

This study examines the dataset records in a large, multidisciplinary data repository to determine the extent to which and how informed consent information is communicated to secondary users, and the degree to which conditions of access and use of data adhere to terms of informed consent. We identified all records published in Borealis: The Canadian Dataverse Repository between January 2022 and September 2024 containing individual-level human data. From those records, we analysed the frequency with which consent information was included and methods used to do so. We further compared terms of consent with the licensing, textual, and technological conditions placed on access and use of the data. Results indicate that informed consent contexts are infrequently provided alongside data and that access and use conditions align with terms of consent for a slim majority of the sample datasets. Based on these findings, we provide recommendations for the development of repository policy and guidelines that harmonise terms of consent and data use, the standardisation of language establishing access and use conditions, the adoption of metadata schema describing ethical contexts, and additional collaboration among data stewards and research ethics boards.

Submitted 20 January 2025 ~ *Accepted* 6 November 2025

Correspondence should be addressed to Brian Jackson. Email: bjackson@mtroyal.ca

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

The *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans - 2022* (TCPS2), the cardinal framework for human research ethics in Canada, is guided by three core principles—respect for persons, concern for welfare, and justice. Ensuring that evolving research practices remain faithful to these principles requires vigilance and debate among the three federal research funding bodies who oversee TCPS2, known as the Tri-Council, local research ethics boards (REBs), data stewards, and researchers, all of whose interests may at times conflict. The widespread adoption of funder and publisher open data mandates, and a burgeoning ethos of open scholarship, has raised important questions about how shared data should be managed while maintaining respect for participants, security of their welfare, and their trust (Corti & Bishop, 2020). Informed consent is frequently at the heart of these discussions.

Informed consent in research ostensibly ensures that subjects willingly provide information in full knowledge of the nature of the study, their rights during and after the transaction, and the risks of participation. Informed consent is mandated by both policy and law, for example by the TCPS2 in Canada (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada [CIHR, NSERC, & SSHRC], 2022), the Common Rule in the United States (Office for Human Research Protections, 2017), and the General Data Protection Regulation in the European Union (Regulation 2016/679, 2016), as well as by a range of additional legislation regulating the collection of human data. Despite widespread reliance on informed consent processes to uphold respect for persons and justice in research, questions have been raised about the efficacy and purpose of the practice. The detailed consent letters that often precede research data collection have been characterised as overwhelming and overly complex (Albala, Doyle, & Appelbaum, 2010), and there is evidence that details within those documents may be ignored or poorly understood (Cummings, Zagrodney, & Day, 2015; Pietrzykowski & Smilowska 2021; VandeVusse, Mueller, & Karcher, 2022). Some authors have questioned whether consent forms actually benefit participants or exist primarily to protect institutions and researchers (Smilan, 2022). As Barocas and Nissenbaum (2014) have argued, "we have crammed into the notice and consent protocol all our moral and political anxieties, believing that this is the way to achieve the level playing field, to promote the autonomy of data subjects…In our view, this became a futile effort at some point along the way" (p. 66). Others have considered whether consent should be required at all for non-sensitive data collection (Ferretti et al., 2022).

Criticisms of consent practices notwithstanding, the principle of informed consent remains a fundamental aspect of ethical research, demonstrating respect for the autonomy and welfare of participants and developing trust through transparency (Cumyn et al., 2021). Clear data management procedures being an important part of transparency, information provided during the consent process typically includes the nature of data collection, arrangements for access to and confidentiality of data, and the intended pathways of the data (CIHR, NSERC, & SSHRC, 2022). When those intended pathways include open repositories, there is an added complexity to the consent process. Not only does additional information about data sharing bloat already lengthy consent documents, but it adds uncertainty around privacy and potential uses of the data (Childs et al., 2014). In response, researchers and ethics oversight bodies have adopted varied practices of seeking consent for unknown future uses of data, most commonly broad consent.

Broad consent is distinct from specific consent, which seeks permission to use data only for a specified purpose, and blanket consent, which seeks open licence for all future uses of the data (Nielsen & Kongsholm, 2022). Broad consent, rather, sits between those two extremes by imposing constraints on secondary use to a specified field of study, type

of research, or similar limitation (Wendler, 2013). While broad consent satisfies some concerns in regard to ethical data sharing, the practice is not without criticism. Debates about broad consent have occurred over decades primarily within biomedical literature, where biobanking has been a common practice, and among qualitative researchers. Today, members of those communities remain the topic's majority discussants.

Objections to broad consent have been raised on the basis that consent for unknown future investigations with unanticipated risks cannot, by definition, be informed (Smilan, 2022). While unforeseen outcomes do occur in primary research, consent is generally established through a direct connection to an investigator and given, in part, due to trust in their institution (Broekstra et al., 2020). Consent documents further develop trust through transparency and provide impartial contacts to whom participants might seek remedy for adverse outcomes or a failure to honour the terms of the agreement (Williams & Pigeot, 2017). Where broad consent to share data is sought, the introduction of unknown third-party researchers and institutions alters the nature of that trust and disrupts the direct connection (Broom, Cheshire, & Emmison, 2009). For some sensitive qualitative research, even the prospect of data sharing might be injurious to the relationship between the researcher and the participant, and could negatively impact the original study (Mauthner, 2012).

Broad consent also raises questions about ownership and control of data. By consenting to the sharing of their data for unknown future uses, participants relinquish a greater degree of control than they would under a condition of specific consent (Parker, 2020). Even when data are completely anonymous or de-identified, and risks to privacy are low, participants may have moral objections to secondary uses of the data they provide that they are unable to voice (Warner et al., 2018). While alternative approaches such as dynamic consent allow participants to approve secondary uses on a case-by-case basis, those forms of consent also have limitations that may hinder reuse (Lay et al., 2024; McKeown et al., 2021). Where ethics oversight is required for secondary uses, decisions about data reuse are relegated to ethics boards whose purpose and motivations may be opaque to the participant (Kvale & Darch, 2022). And in some cases, participants may consent to share data that is linked to social life, spirituality, or customs of a community whose wider membership did not consent and who may have no reason to trust either the original investigator or subsequent agents of research who view the data (West–McGruer, 2020).

Proponents of broad consent have argued that these objections are paternalistic and may impede the progress of research. They contend that participants who provide broad consent do so freely, and that honouring those decisions demonstrates justice and respect for participant autonomy (Stodden, 2014). Some studies have found that a majority of participants actually prefer broad consent to specific consent due to the greater potential reach of their contributions (Garrison et al., 2016; Simon et al., 2011; Tomlinson et al., 2015). Going further, some argue that broad consent with the intent to share data should be required for some types of research. This view sees the adoption of ethical frameworks that position research data and biospecimens as a common good, that is, broadly accessible with appropriate ethical checks, as beneficial to the advancement of research and, thus, in the public interest (Mauthner, 2012; O'Doherty et al., 2021). For others, this suggestion strays too far from principles of individual autonomy outlined in numerous ethics codes including, for example, the World Medical Association's Declaration of Helsinki, which holds that the purposes of medical research "can never take precedence over the rights and interests of individual research participants" (World Medical Association, 2024, "General Principles"). The Canadian ethics framework similarly emphasises individual welfare, adopting a limited approach to broad consent.

A 2022 update to TCPS2 introduced guidance on broad consent for future unspecified use of data and biological materials to complement existing guidelines on secondary use of data and use of data repositories (CIHR, NSERC, & SSHRC, 2022). In the new guidelines, blanket consent is not permitted but broad consent is, with the stipulation that

participants must be allowed to opt-out of inclusion in datasets stored for future reuse. Researchers must also provide participants with enough information to make an informed decision, including the type of data that will be stored, a description of the repository and its governance, who might have access to the repository, and whether secondary uses will be subject to ethics oversight. Where researchers wish to deposit existing data in a repository but do not have participant consent to do so, they must attempt to re-contact participants to obtain consent or, if contact is impracticable, seek REB approval to deposit (Panel on Research Ethics, 2021). Going forward, then, human data deposited in repositories by Canadian researchers should have either broad consent from participants or approval from an REB if those datasets are intended for reuse.

How and if consent for deposit was obtained is a crucial part of the approval process for reuse and may impact both secondary researchers and REBs. The Tri-Council's Panel on Research Ethics (2024) makes an important distinction between anonymous and non-identifiable information, stating that "anonymous information is 'information [that] never had identifiers associated with it … and the risk of identification of individuals is low or very low'...TCPS defines information as non-identifiable 'if it does not identify an individual, for all practical purposes, when used alone or combined with other available information'" (p. 34). For researchers who wish to use secondary data, the guidelines state that "research that relies exclusively on secondary use of anonymous information is exempt from REB review …Research that relies exclusively on secondary use of non-identifiable information generally requires REB review" (Panel on Research Ethics, 2024, p. 34). The document further explains that "even information that is easily accessed by members of the public may be associated with expectations of privacy, particularly if the terms of consent are unclear" (p. 60). While this approach is laudable for its checks on reuse of human data, in practice it may not always be apparent if open datasets were collected anonymously or de-identified after collection, and the original terms of consent may be obscure (Poth, 2019). Thus, there is a responsibility placed on secondary researchers, data curators, and REBs to act in consideration of the welfare of participants with potentially very little context about their participation in research.

Access to the ethical contexts surrounding data collection and deposit would reduce the burden on researchers and REBs, but currently, the quality of documentation and metadata describing data held in open repositories varies greatly. Although some data repositories capture ethical elements, the practice is far from widespread (Hunt & Hofelich Mohr, n.d.). Some researchers have noted the absence of this information and have called for greater attention to consent contexts in data reuse. Both Mannheimer et al. (2019) and Bernier et al. (2023), for example, proposed that ethical provenance records should be maintained by data stewards alongside datasets, documenting participant consent and previous ethical reviews. Similarly, de Vries et al. (2014) observed the potential for misuse or misunderstanding of genomic data absent the context of its collection, stating "it may be necessary to accompany genomic data with relevant information about the normative context of research. This could include for instance information about the informed consent process" (p. 8). Following on these ideas, others have proposed the wider adoption by data repositories of metadata schema that incorporate information about the original consent process, allowing secondary users to make ethical judgements about the described data, simplifying the ethics review process, and assuaging some concerns about broad consent (Alter et al., 2020; Woolley, 2016).

It is imperative that the terms to which participants have agreed be upheld in both original and secondary research contexts in order to maintain trust and to fulfil obligations to demonstrate respect and concern for the welfare of participants. While there have been some calls to improve processes for communicating those terms in repository records, there has been little exploration about how it is currently done. This project aims to fill that gap by examining records in a large Canadian generalist data repository and asking:

- To what extent do the terms of participant consent appear in dataset records?
- How are consent contexts communicated in dataset records?
- To what degree do conditions of access and use described in dataset records reflect terms of participant consent?

Methods

We examined all records deposited in Borealis: The Canadian Dataverse Repository from January 2022 through September 2024. Borealis is a pan-Canadian data repository established and managed by Scholars Portal, a service provided by the Ontario Council of University Libraries. Nearly 70 universities and colleges across Canada subscribe to the service, many of which use Borealis as their primary institutional research data repository. Among these is the institution of the lead author, who is the primary administrator and curator of the university's repository. Subscribing institutions receive distinct, branded spaces within the repository, but all datasets are federated into centralised storage and search services. The repository houses more than 21,000 datasets across its member institutions, including a mix of open and restricted research datasets and government data.

The Borealis repository was selected primarily due to its scope. TCPS2 applies to all human research conducted at Canadian institutions that receive Tri-Agency funding, regardless of whether individual projects are funded. As most Canadian universities and colleges are both subject to TCPS2 and members of the Borealis repository, the choice of that repository provides a large sample of dataset records stemming from research governed by the same research ethics policy.

We used the following inclusion criteria to select dataset records for the study:

- The record includes individual-level data collected from human participants;
- The dataset includes primary data, collected from participants by the depositor or affiliated research team or organisation;
- The dataset description, metadata, and documentation are primarily in English (due to language constraints of this study's investigators);
- The dataset is held within the Dataverse of a college and university subject to TCPS2.

When necessary and possible, journal articles and other publications associated with datasets were consulted to confirm the above criteria but, where the nature of the data was uncertain, records were excluded from analysis. Both open and restricted access datasets were included.

For each qualifying record, we recorded descriptive metadata such as title and subject; the format and type of data; terms of access and use; whether terms of participant consent were present, either described or verbatim; where and how terms of consent were communicated; and the exact terms of consent as presented or described. Each team member collected data and subsequently reviewed the data collected by the other member in order to ensure accuracy.

We used a mixed methods approach to addressing our three research questions. In order to determine the extent to which terms of consent appear in dataset records, we used simple distribution tables and descriptive statistics. In order to determine how consent contexts are communicated in consent records, we first classified the location and methods of these communications into five categories:

- Described in a readme file
- Described in a dataset abstract/description
- Consent form attached as a standalone file
- Consent form included with survey or questionnaire
- Consent form appended to attached report or article

After classification and verification by each team member, we again used simple distribution tables to measure the frequency with which each type occurred.

In order to address the third research question, we examined both the verbatim terms of consent and the terms of use or licence agreements accompanying each dataset. To analyse the former, we used an evaluative qualitative text analysis approach based on Kuckartz (2014). After a close examination of the text copied from consent forms, we defined three evaluative categories: intention to deposit in a repository, details about the repository, and access to data. We then developed and described characteristic levels within each category as outlined in the following tables:

Table 1. Definitions of each characteristic level in evaluative category A: Intention to deposit in repository.

A. Intention to deposit in repository	Definition
A1. Confirmed intention to deposit	Indication of clear intention to deposit in a data repository or similar online platform, regardless of access type or whether specific repository is named
A2. Potential to deposit	Indication that data may be deposited in a repository or similar online platform, regardless of access type or whether specific repository is named
A3. No indication of intent to deposit	No indication that data will be deposited in a repository or similar online platform

Table 2. Definitions of each characteristic level in evaluative category B: Details about the repository.

B. Details about the repository	Definition
B1. Description of repository	Some details provided about the repository or similar platform, including name, affiliated institution, data storage location, and/or governance and oversight
B2. No information provided	No details specific to the intended repository

	provided
B3. No indication of intent to deposit	No indication that data will or might be deposited in a repository

Table 3. Definitions of each characteristic level in evaluative category C: Access to data.

C. Access to data	Definition
C1. Access for research	Some qualifying details provided signalling secondary access for research, including purpose of secondary use, procedures for accessing restricted data, and anticipated discipline of secondary use
C2. Public access	Indication that data will or might be available openly and publicly without additional qualifiers
C3. No details provided	Intent to or anticipation of deposit but no information about access or anticipated secondary users provided
C4. No indication of intent to deposit or share	No indication that data will or might be deposited in a repository or shared with other researchers

The characteristic levels within each category were designed to be comprehensive so that every dataset that included original terms of consent could be classified at a level within each category. Each team member coded the datasets separately. After comparison, the team members discussed areas of divergence in coding until a consensus was reached.

The Borealis repository allows depositors to both select a Creative Commons licence and textually describe terms of use for and access to uploaded datasets. Because relatively few datasets included textual terms of access and use, and the text for most was very brief, terms of access and use were simply classified into six non-exclusive categories:

- Licensing (including Creative Commons licensing and citation requirements)
- User may not request access
- User may request access
- User may request access for specified purposes
- User may request access for specified purposes with confirmation of ethics review
- Open data (data without technological access barriers) may only be used for specified purpose

After coding and classifying the consent language and the terms of access and use, we examined the records in order to describe how conditions of data access and use of relevant datasets reflect the terms of consent to which original participants agreed.

Results

Sample Characteristics

We identified 227 datasets meeting our criteria out of a total of 13,659 datasets published during the relevant time period (see Table 4). The most common subject classifications within our sample, among which some datasets include multiple classifications, were social sciences (48%); medicine, health, and life sciences (45%); agricultural sciences (11%); computer and information science (10%); and engineering (10%). Quantitative-only datasets accounted for a majority (152) of the sample. The remaining datasets contained qualitative data (23), a combination of qualitative and quantitative data (37), or were indeterminate due to access restrictions and sparse descriptions. Access to individual-level data files was fully restricted in 36 of the sample datasets, while only three datasets contained both open and restricted data files. Where restricted data types could be determined, qualitative data was disproportionately represented. Nearly 35% of the qualitative datasets, but only 5% of the quantitative datasets, were restricted. Most of the datasets contained tabular (201) or text (23) formats, with various other data types, including images, video, audio, and data exchange formats, accounting for a small minority of the sample.

Table 4. Descriptive breakdown of datasets by subject classifications, data type, and data format.

Characteristics of dataset sample (N=227)					
Subject classifications	n	Data type	n	Data format	n
Agricultural sciences	25	Quantitative	152	Audio	1
Arts and humanities	10	Qualitative	23	Data exchange format (e.g., json)	6
Astronomy and astrophysics	1	Mixed	37	Images	2
Business and management	7	Unknown	15	Tabular	201
Chemistry	1			Text	23
Computer and information science	23			Video	1
Earth and environmental sciences	6				

Engineering	23
Mathematical sciences	6
Medicine, health, and life sciences	102
Physics	1
Social sciences	110
Other	16

Terms of Consent

Of 227 datasets in our total sample, 45 contained information about participant terms of consent (Jackson & Azmi, 2025). Of those, four included consent forms that were attached and labelled but restricted, and so were excluded from further analysis. Within the remaining 41 records, consent documents were most commonly found appended to questionnaire files (30) or attached as standalone documents (8). Consent information was also found in a readme file, a dataset abstract, and appended to a report. When compared to the subject classifications of the total sample, the consent sample included higher representation from social sciences (67%) and lower representation from medicine, health, and life sciences (17%) but was otherwise similarly distributed.

The dataset sample included 39 datasets containing verbatim terms of consent. A quantitative breakdown of assigned codes can be found in Table 5. Nearly 40% of the consent documents provided no indication that research data would be shared in a repository or similar platform. Of the remaining consent letters, two-thirds stated affirmatively that data would be deposited for secondary use, while one-third signalled the potential for data deposit. Among the latter, some indicated that data would be deposited only if mandated by policy. One consent document, for example, included the statement, "data may be required to be open access (stored online for other researchers to view) as part of publication. If this is required, all identifying information will be removed to ensure your anonymity." Similar assurances of confidentiality or anonymity were observed in the majority of consent documents in which open data deposit was broached, but few provided details about specific types of identifying information that would be removed or the nature of the data that would be deposited. A small minority, however, did indicate that specific identifiers, such as names or geographical variables, or qualitative responses to survey questions would be removed, and one listed specific variables that would be shared.

Only 11 out of 24 consent documents indicating intent or potential to deposit data provided information about the repository. In nearly all cases, this information included only the name of the repository and affiliated institution. Three consent documents provided repository links for the information of participants, one to the home page of the intended repository and two to previously deposited datasets by the same researcher. The remaining 13 datasets indicated that data would be shared publicly or for future research but did not provide information about the intended repository. One consent document, for example, stated that "de-identified information will also be kept in a public research

database but you will not be identified in any way", but provided no description of the database.

The majority of consent documents that referred to data deposit indicated that data would be made openly or publicly available. Statements of that type were generally brief and without further description. For example, one document's sole reference to data sharing stated that "data will only be accessible to the research team until the time of publication when raw anonymous data will be made publicly available." Where consent letters signalled that data would be used for secondary research, which occurred in eight consent documents, how access would be managed was often ambiguous. In some cases, consent documents suggested that data would be both publicly accessible but used only by secondary researchers. For instance, one consent letter indicated that "the data will be stripped of short answer question responses, and uploaded to an open data website, Borealis, for future researchers to access." No consent documents placed limits on or anticipated the types of secondary research to which the data might be put, although one stipulated that future uses of data would require review by an ethics board. Three additional consent documents indicated that data would be deposited or archived in a repository without explanation of the nature of access to the platform or dataset.

Table 5. Quantitative breakdown of codes assigned to datasets containing verbatim terms of consent.

A. Intention to deposit in repository		B. Details about repository		C. Access to data	
Level	n	Level	n	Level	n
A1. Confirmed intention to deposit	16	B1. Description of repository	11	C1. Access for research	8
A2. Potential to deposit	8	B2. No information provided	13	C2. Public access	13
A3. No indication of intent to deposit	15	B3. No indication of intent to deposit	15	C3. No details provided	3
				C4. No indication of intent to deposit or share	15

Terms of Access and Use

Nearly all of the dataset records (218) contained terms of use that incorporated licensing information, including Creative Commons licences and/or citation requirements. Of the 36 records with additional textual terms of access and use, 18 indicated, without additional qualification or description, that users may request access to data, eight placed restrictions on types of reuse and required evidence of ethics review prior to data sharing, and seven indicated that access to data would be granted only for specified purposes, most often related research or verification. In two cases, terms of use stated that data would not be shared, and one open dataset contained textual restrictions on secondary uses of data to those related to research.

Within the subset of 39 dataset records with accessible informed consent documents, only four included textual terms of access and use that constrained secondary uses of data. Three of those did not include prior participant consent for deposit but did include terms limiting secondary uses to similar research studies, with one noting this limitation as a condition of ethics board approval for data deposit. Another record stated that data would be shared only upon request when, in fact, the data files were openly accessible. The fourth dataset containing textual terms of access and use included a consent letter that affirmed intent to publicly share data, but the associated data files were only partially available due to the sensitivity of some qualitative data.

Creative Commons licences were applied to a majority (38) of the consent-present subset. Public domain (CC0) and attribution only (CC-BY) licences, the least restrictive types of Creative Commons licences, were applied to 28 records. Additionally, non-commercial (CC-NC) clauses were found in nine records. Of the eight datasets that included terms of consent indicating that shared data would be reused for research purposes, all were openly accessible with only Creative Commons licences applied as terms of use. While two of these incorporated non-commercial use clauses, the remaining six were licensed under public domain or attribution only licences. Among the datasets in which terms of consent did not indicate an intent to deposit or share data, public domain licences were most common.

Table 6. Terms and access of use as outlined in terms of consent of included datasets.

Terms of access and use		Access to data described in terms of consent			
		Access for research	Public access	No details provided	No indication of intent to deposit or share
Access by request		0	1	0	2
Open data with research use only limitation		0	0	0	1
Creative Commons licensing	CC public domain licence	4	9	1	8
	CC attribution-only licence	2	1	1	8
	CC non-commercial licence	2	3	0	4
	All licences	8	13	3	14

Discussion

The context of collection is critical for data reuse. Without a firm understanding of the parameters under which data was collected, both methodological and ethical lapses can occur in secondary use. Data deposit guidelines typically emphasise the importance of methodological documentation, but ethics-related documentation and metadata are more often overlooked (Antes et al., 2018). The nature of deposit guidelines might explain, in part, why fewer than one fifth of the datasets in our sample included consent contexts anywhere in the dataset record or files. The absence of ethical contexts has potential ramifications for all of the agents of research involved, including original researchers, data stewards, secondary users of the data, and research ethics boards.

Recall that Canadian ethics policy requires ethics review for secondary research using de-identified datasets, asserting that participants may have expectations about privacy and the use of their data even if datasets are publicly accessible. Those expectations may be formed by the terms to which participants agreed when they consented to take part in the study, including any data access and reuse limitations outlined in consent documents. When reviewing proposed secondary uses of human data, local research ethics boards are directed to consider the original terms of consent, when available (Panel on Research Ethics, 2024). Absent the availability of consent contexts in dataset records, original researchers must be contacted to obtain participant terms of consent, or ethics boards must deliberate without them. As previous research has identified significant challenges in contacting original authors due to inoperable email addresses or lack of response (Vines et al., 2014), ethics boards are likely to experience the added burden of identifying risks on behalf of original participants without understanding the full context of their participation.

Where the terms of participant consent are referenced in dataset records, secondary researchers face the additional, albeit less significant, challenge of locating the precise terms to which participants have agreed. While most of the datasets in our sample that included terms of consent did so by attaching files, methods of doing so were inconsistent. Consent forms attached as standalone files were located relatively easily, but some consent documents were included as part of questionnaires or via external links embedded in attached documents. In many cases, file names and descriptions did not indicate that consent forms were available within those documents. A secondary user determined to uphold the original terms of consent would likely find these documents, but additional effort is needed to locate them, and a less determined user might not find them at all. A small number of datasets referred to informed consent processes in a readme file or dataset abstract/description, but no details were provided in these that would allow a secondary user to understand the expectations of participants. Some did clarify, though, that the deposit of the data had received prior ethics approval, which may reduce the degree of scrutiny required during ethics review for secondary use.

Of greater importance are the inconsistencies observed between the original terms of consent, dataset terms of access and use, and technological access barriers. While a slight majority of our sample datasets with consent terms that included potential secondary uses of data did honour those agreements in the establishment of access and use conditions, there were numerous dataset records where deviations were noted. Most common among these were cases in which terms of consent indicated that data would be shared for the purposes of secondary research but were in fact openly accessible with Creative Commons public domain or attribution licences associated with them, with no other use restrictions noted. Some datasets included non-commercial clauses, which are more restrictive but do not limit secondary uses to academic research studies. Only one fully open dataset included textual terms limiting use of data to secondary research purposes. Other inconsistencies may be attributable to misapplication of access barriers or Creative Commons licences, or failure to update records after embargo periods. For example, a

small number of datasets indicated that access to files may be granted upon request but were openly available for download without restriction.

Even when levels of access and use were consistent with terms of consent, few consent documents provided more than sparse details about the expected pathways of the data. A minority of consent letters indicated the name of the repository and the academic institution overseeing it, but none provided information about the stewardship and governance or data storage practices of the repository. In only a few cases did consent letters describe which data would be shared and how confidentiality would be maintained. These absences are not surprising as the TCPS2 detailed guidelines for establishing broad consent, which recommend the provision of information to participants about the nature of both the repository and the shared data, were not available at the time that much of the data in our sample was collected. Our sample reflects the uncertainty faced by researchers and REBs before the development of those recommendations, and we may see additional details about data repositories in consent letters going forward.

After the publication of the 2022 edition of TCPS2, the Interagency Panel on Research Ethics, which provides guidance on the application of TCPS2, further clarified the scope of the broad consent clause (Panel on Research Ethics, 2024). In its interpretation, the Panel clarifies that the broad consent clause "does not specifically apply in the context of archiving data at the completion of a research project, depositing data in a shared/open repository or when sharing datasets in the context of journal publications for purposes of reproducibility/transparency, verification or error detection" (p. 13). The same explanation goes on to note that when a participant does not consent to the storage of their data for reuse, "their data should not be deposited in a research data repository that serves such a purpose" (p. 13). In other words, data may be deposited in an open repository without broad consent if the intent is to do so for replication or verification purposes, but researchers should not deposit the data of participants who have specifically opted out of data sharing into open repositories established for the purpose of secondary data use.

This limitation poses a challenge for repositories with a broad mandate. Borealis, like other generalist and institutional repositories, serves several purposes, with its stated mission "to enhance open sharing, discovery, access, preservation, and reuse of research data" (Borealis, n.d.). Depositors rely on both technological barriers and non-standardised terms of use metadata fields to communicate data use limitations set by the terms of informed consent, their ethics approval, or their own preferences. In our sample, the majority of datasets were openly downloadable with only Creative Commons licences applied, signifying permission for reuse with few conditions. However, a significant minority of datasets, whether open, restricted, or of mixed access, were appended with textual terms of use. The detail and clarity of these terms varied widely, including declarations that data would not be shared, notices that data would be shared upon reasonable request (without definition of 'reasonable'), statements that data may be used for verification only, and comprehensive descriptions of data access procedures including prior ethics review. Like the mission of the repository, its use by depositors encompasses multiple purposes, including closed archiving, controlled data access, and open sharing for secondary use or replication, and those using the service have defined their intended purpose in various ways at the dataset, rather than repository, level.

For secondary users, though, this variation may be confusing or contradictory. Users may not be aware of the details of original terms of participant consent, types of permitted reuses may be unclear, the conditions upon which original researchers are willing to share restricted data often go undeclared, and data may not be described clearly enough to determine if ethics review is needed for secondary use. Even when these elements are clear, inconsistent use of the terms 'anonymous', 'anonymised', and 'confidential' in consent letters may obscure the expectations of participants and procedures for data reuse. TCPS2 distinguishes between these terms, defining anonymous data by its total absence of personal identifiers, anonymised data as data in which identifiers have been

removed, and confidentiality as "obligations to protect information from unauthorized access, use, disclosure, modification, loss or theft" (CIHR, NSERC, & SSHRC, 2022). In some consent letters, though, researchers promised anonymity even though personal identifiers were collected and removed prior to data sharing. While the common use of these terms may be for the benefit of participants, who might not distinguish between them in the same way that researchers do, it may lead to uncertainty about the nature of the data collection and the necessity of secondary ethics review.

In many of these situations, clarification requires mediation by the original researcher, repository stewards, and ethics boards, even for open datasets. These added efforts come with a cost of time and labour, some of which would be obviated by more detailed documentation and metadata describing the data, access and use conditions, and the ethical provenance of the data.

Data curators and stewards may be able to alleviate some of this potentially burdensome mediation and facilitate more seamless access to data in several ways, including:

- implementation of policies that require depositors to include and label original terms of consent and/or ethics approvals for the deposit of human data (Mannheimer et al., 2019)
- application of metadata schema describing terms of consent for secondary use and the reuse limitations imposed by those terms (Alter et al., 2020; Woolley, 2016)
- for repositories with self-deposit or minimal-mediation policies, the development of deposit guidelines outlining how depositors should describe conditions of access and use of data, including procedures for obtaining restricted data, the uses to which both restricted and open data may be put, and any requirements for citation and subsequent sharing
- development of data curation procedures that are attentive to harmonisation between terms of consent and terms of use and ensure that limitations on secondary use are described clearly and precisely.

Although these recommendations derive from a focus on Canadian ethical contexts, most would benefit repositories operating in environments with broadly similar ethics regulations, and, doubtlessly, some of those repositories have adopted some or all of them. In the United States, the revised Common Rule establishes broad consent rules that are similar to TCPS2 but distinct in their application. As examples, broad consent in the Common Rule only applies to identifiable data and biospecimens, unlike Canada, where broad consent must be sought for all data stored for future reuse, and the types of information that must be provided when seeking broad consent differ between the two countries (Office for Human Research Protections, 2017). In the European Union, the GDPR does not use the term 'broad consent' but adopts a similar approach in its definitions, stating, "It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research" (Regulation 2016/679, Recital 33, 2016). Like the United States, the GDPR only applies to personal data.

In international contexts, then, researchers may have fewer restrictions on sharing minimal-risk anonymous and de-identified data, but the broad principles of informed consent, many of which derive from the Nuremberg Code (Annas, 2018), are widely present and similar. These hold that participants should be made aware of the nature, duration, purpose, and risks associated with the research and that the procedures of research, including data management and sharing, should be faithful to the terms of consent. Stewards of human research data have some responsibility to establish

infrastructure and procedures that offer opportunities for those terms of consent to be upheld when data is stored and reused. Regardless of country or repository type, a more widespread focus on the coordination of terms of consent and data access controls for human data would help original researchers, ethics review boards, and secondary researchers navigate an often confusing element of an evolving research practice.

Conclusion

Like many evolving research practices, the growth of research data sharing has outpaced ethical regulations, which tend to be reactive (Jackson, 2018). The widespread establishment of data repositories and adoption of open data mandates predate guidance on broad consent in policy documents such as TCPS2 and the Common Rule, which were published in 2022 and 2017 respectively, and only came into effect in the years following. Although discussion of broad consent has taken place in the context of sensitive data in biomedical literature for decades, we now need to have similar discussions about how we share minimal- and moderate-risk data, where there are fewer regulatory and procedural hurdles to sharing but also where the rights of participants are equally valid. We should not be casual about the management and dissemination of participant data, or we risk losing the trust that has been built over decades.

Broad consent, as defined by the above policies, includes limits on how the data may be reused. While the precise types of restrictions that would constitute broad consent are only vaguely described in many ethics policies, researchers have a responsibility to consider how their research data will be shared and to communicate that as clearly and precisely as possible to participants. But data stewards and curators also have a part to play. The limitations on data reuse imposed by terms of consent should be apparent to secondary users. Repository managers can implement several solutions to improve consistency and clarity in reporting those limitations, including policies that require consent and ethics documentation to accompany all human datasets, consistent labelling of ethics documentation, the inclusion of consent-based metadata fields, structured language describing terms of use, and mandatory documentation outlining procedures for obtaining data when it is access-restricted. Data stewards can also work with researchers and ethics boards to ensure that the information provided to participants reflects the structures and scope of their repositories and accurately describes how data will be managed, stored, and shared after deposit (Kirilova & Karcher, 2017). Doing so would help to bring additional transparency to what can be, for both researchers and participants, an opaque process.

References

Albala, I., Doyle, M., & Appelbaum, P. S. (2010). The evolution of consent forms for research: A quarter century of changes. *IRB: Ethics & Human Research* 32(3), 7–11. Retrieved from https://www.jstor.org/stable/25703699

Alter, G., Gonzalez–Beltran, A., Ohno–Machado, L., & Rocca–Serra, P. (2020). The Data Tags Suite (DATS) model for discovering data access and use requirements. *GigaScience* 9(2), 1–10. doi:10.1093/gigascience/giz165

Annas, G. J. (2018). Beyond Nazi war crimes experiments: The voluntary consent requirement of the Nuremberg Code at 70. *American Journal of Public Health* 108(1), 42–46. doi:10.2105/AJPH.2017.304103

Antes, A. L., Walsh, H. A., Strait, M., Hudson–Vitale, C. R., & DuBois, J. M. (2018). Examining data repository guidelines for qualitative data sharing. *Journal of Empirical Research on Human Research Ethics* 13(1), 61–73. doi:10.1177/1556264617744121

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, big data, and the public good: Frameworks for engagement* (pp. 44–75).

Bernier, A., Raven–Adams, M., Zaccagnini, D., & Knoppers, B. M. (2023). Recording the ethical provenance of data and automating data stewardship. *Big Data and Society* 10(1), 1–11. doi:10.1177/20539517231163174

Borealis: The Canadian Dataverse Repository. (n.d.) *About.* Retrieved from https://borealisdata.ca/about/

Broekstra, R., Aris–Meijer, J., Maeckelberghe, E., Stolk, R., & Otten, S. (2020). Trust in centralized large-scale data repository: A qualitative analysis. *Journal of Empirical Research on Human Research Ethics* 15(4), 365–378. doi:10.1177/1556264619888365

Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology* 43(6), 1163–1180. doi:10.1177/0038038509345704

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada. (2022). *Tri-council policy statement: ethical conduct for research involving humans*. Retrieved from https://ethics.gc.ca/eng/documents/tcps2-2022-en.pdf

Childs, S., Mcleod, J., Lomas, E., & Cook, G. (2014). Opening research data: Issues and opportunities. *Records Management Journal* 24(2), 142–162. doi:10.1108/RMJ-01-2014-0005

Corti, L., & Bishop, L. (2020). Ethical issues in data sharing and archiving. In R. Iphofen (Ed.), *Handbook of research ethics and scientific integrity* (pp. 403–426). doi:10.1007/978-3-030-16759-2_17

Cummings, J. A., Zagrodney, J. M., & Day, T. E. (2015). Impact of open data policies on consent to participate in human subjects research: Discrepancies between participant action and reported concerns. *PLoS ONE* 10(5). doi:10.1371/journal.pone.0125208

Cumyn, A., Dault, R., Barton, A., Cloutier, A. M., & Ethier, J. F. (2021). Citizens, research ethics committee members and researchers' attitude toward information and consent for the secondary use of health data: Implications for research within learning health systems. *Journal of Empirical Research on Human Research Ethics* 16(3), 165–178. doi:10.1177/1556264621992214

de Vries, J., Williams, T. N., Bojang, K., Kwiatkowski, D. P., Fitzpatrick, R., & Parker, M. (2014). Knowing who to trust: Exploring the role of "ethical metadata" in mediating risk of harm in collaborative genomics research in Africa. *BMC Medical Ethics* 15(62), 1–10. doi:10.1186/1472-6939-15-62

Ferretti, A., Ienca, M., Velarde, M. R., Hurst, S., & Vayena, E. (2022). The challenges of big data for research ethics committees: A qualitative Swiss study. *Journal of Empirical Research on Human Research Ethics* 17(1–2), 129–143. doi:10.1177/15562646211053538

Garrison, N. A., Sathe, N. A., Antommaria, A. H. M., Holm, I. A., Sanderson, S. C., Smith, M. E., McPheeters, M. L., & Clayton, E. W. (2016). A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. *Genetics in Medicine* 18(7), 663–671. doi:10.1038/gim.2015.138

Hunt, S., & Hofelich Mohr, A. (n.d.). *Teaching an old repository new tricks: Developing a human participant data policy.* Data Curation Network. Retrieved from https:// https://datacuration.network/2024/02/08/teaching-an-old-repository-new-tricks-developing-a-human-participant-data-policy/

Jackson, B. (2018). The changing research data landscape and the experiences of ethics review board chairs: Implications for library practice and partnerships. *The Journal of Academic Librarianship* 44(5), 603–612. doi:10.1016/j.acalib.2018.07.001

Jackson, B., & Azmi, A. (2025). *Replication data for: Informed consent contexts in a multidisciplinary research data repository* [Data set]. Calgary, Alberta, Canada: MRU Data Repository. doi:10.5683/SP3/TEX9JY

Kirilova, D., & Karcher, S. (2017). Rethinking data sharing and human participant protection in social science research: Applications from the qualitative realm. *Data Science Journal* 16(43), 1–7. doi:10.5334/dsj-2017-043

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice & using software* (trans. by A. McWhertor). London: Sage Publications.

Kvale, L. H., & Darch, P. (2022). Privacy protection throughout the research data life cycle. *Information Research* 27(3). doi:10.47989/irpaper938

Lay, W., Gasparini, L., Siero, W., & Hughes, E. K. (2024). A rapid review of the benefits and challenges of dynamic consent. *Research Ethics* 21, 1–23. doi:10.1177/17470161241278064

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2019). Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges. *American Behavioral Scientist* 63(5), 643–664. doi:10.1177/0002764218784991

Mauthner, N. (2012). Are research data a "common" resource? *Feminists @ Law* 2(2), 1–22. Retrieved from https://journals.kent.ac.uk/index.php/feministsatlaw/article/view/60

McKeown, A., Mourby, M., Harrison, P., Walker, S., Sheehan, M., & Singh, I. (2021). Ethical issues in consent for the reuse of data in health data platforms. *Science and Engineering Ethics* 27(9), 1–21. doi:10.1007/s11948-021-00282-0

Nielsen, M. E. J., & Kongsholm, N. C. H. (2022). Blanket consent and trust in the biobanking context. *Journal of Bioethical Inquiry* 19, 613–623. doi:10.1007/s11673-022-10208-5

O'Doherty, K. C., Shabani, M., Dove, E. S., Bentzen, H. B., Borry, P., Burgess, M. M., & Burke, W. (2021). Toward better governance of human genomic data. *Nature Genetics* 53(1), 2–8. doi:10.1038/s41588-020-00742-6

Office for Human Research Protections. (2017). *Federal policy for the protection of human subjects ('common rule')*. U.S. Department of Health and Human Services. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html

Panel on Research Ethics. (2021). *Depositing existing data in public repositories: Guidance in applying TCPS 2*. Retrieved from https://ethics.gc.ca/eng/depositing_depots.html

Panel on Research Ethics. (2024). *TCPS 2022 interpretations.* Retrieved from https://ethics.gc.ca/eng/documents/TCPS2_interpretations-en.pdf

Parker, Lisa S. (2020). Genetic and genomics research. In A. S. Iltis, & D. MacKay (Eds.), *The Oxford handbook of research ethics* (pp. 338–367).

Pietrzykowski, T., & Smilowska, K. (2021). The reality of informed consent: Empirical studies on patient comprehension—systematic review. *Trials* 22(57). doi:10.1186/s13063-020-04969-w

Poth, C. N. (2019). Rigorous and ethical qualitative data reuse: Potential perils and promising practices. *International Journal of Qualitative Methods 18*, 1-4. doi:10.1177/1609406919868870

Regulation 2016/679. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).* Retrieved from http://data.europa.eu/eli/reg/2016/679/oj

Simon, C. M., L'Heureux, J., Murray, J. C., Winokur, P., Weiner, G., Newbury, E., Shinkunas, L., & Zimmerman, B. (2011). Active choice but not too active: Public perspectives on biobank consent models. *Genetics in Medicine* 13(9), 821–831. doi:10.1097/GIM.0b013e31821d2f88

Smilan, L. E. (2022). Broad consent—Are we asking enough? *Ethics & Human Research* 44(5), 22–31. doi:10.1002/eahr.500140

Stodden, V. (2014). Enabling reproducibility in big data research: Balancing confidentiality and scientific transparency. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, big data, and the public good: Frameworks for engagement* (pp. 112–132).

Tomlinson, T., de Vries, R., Ryan, K., Kim, H. M., Lehpamer, N., & Kim, S. Y. H. (2015). Moral concerns and the willingness to donate to a research biobank. *JAMA* 313(4), 417–419. doi:10.1001/jama.2014.16363

VandeVusse, A., Mueller, J., & Karcher, S. (2022). Qualitative data sharing: Participant understanding, motivation, and consent. *Qualitative Health Research* 32(1), 182–191. doi:10.1177/10497323211054058

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology* 24(1), 94–97. doi:10.1016/j.cub.2013.11.014

Warner, T. D., Weil, C. J., Andry, C., Degenholtz, H. B., Parker, L., Carithers, L. J., Feige, M., Wendler, D., & Pentz, R. D. (2018). Broad consent for research on biospecimens: The views of actual donors at four U.S. medical centers. *Journal of Empirical Research on Human Research Ethics 13*(2), 115–124. doi:10.1177/1556264617751204

Wendler, D. (2013). Broad versus blanket consent for research with human biological samples. *The Hastings Center Report* 43(5), 3–4. doi:10.1002/hast.200

West–McGruer, K. (2020). There's 'consent' and then there's consent: Mobilising Māori and Indigenous research ethics to problematise the western biomedical model. *Journal of Sociology* 56(2), 184–196. doi:10.1177/1440783319893523

Williams, G., & Pigeot, I. (2017). Consent and confidentiality in the light of recent demands for data sharing. *Biometrical Journal* 59(2), 240–250. doi:10.1002/bimj.201500044

Woolley, J. P. (2016). How data are transforming the landscape of biomedical ethics: The need for ELSI metadata on consent. In B. D. Mittelstadt, & L. Floridi (Eds.), *The ethics of biomedical big data* (pp. 171–197). doi:10.1007/978-3-319-33525-4_8

World Medical Association. (2024). *WMA Declaration of Helsinki – Ethical principles for medical research involving human participants*. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki/