Transcript for “Development of a policy on ingestion of human subjects datasets”

Slide 1: Title

Welcome to our presentation titled “Development of a policy on ingestion of human subjects datasets: An institutional assessment and large-scale repository scan.”

Slide 2: Authors

The authors, Shanda Hunt, Valerie Collins, and Alicia Hofelich Mohr, are curators for the Data Repository for the University of Minnesota. Shanda and Alicia handle the curation of all datasets that studied human subjects, and Valerie is the repository coordinator who reviews incoming data submissions for inclusion in DRUM.

Slide 3: Background

We’ll start with some background on our repository, DRUM, and the reasons why we’ve conducted this work, before moving on to discuss our evaluation of DRUM’s practices, and the results of our repository scan.

Slide 4: DRUM

The Data Repository for the University of Minnesota (DRUM) is a publicly available collection of digital research data generated by University of Minnesota researchers, students, and staff. Datasets published in DRUM are openly available, broadly disseminated, and downloadable without restriction.

Slide 5: Human Subjects Data

Within DRUM, there are 46 datasets with human subjects data. A human subject is a living individual about whom a researcher obtains, uses, studies, analyzes, or generates information. Because of the sensitive nature of human subjects data, and the potential issues with publishing such data in a publicly accessible repository, we assess all human subjects data before acceptance into DRUM. As part of this process, we ask researchers to share with us the consent form that participants signed as they joined the study.

Slide 6: Problematic Consent Language

What we've encountered in some of the consent forms that we've requested is problematic consent language. Institutional review boards have historically provided sample language that does not consider the possibility of sharing the data to broad audiences. Since researchers look to their IRBs for guidance on consent forms, we typically see this default language in the consent forms we receive.

However, IRBs are beginning to change their practices, as seen in this updated UMN template that was released last year. But, even as IRBs update their templates, these changes do not translate to researchers immediately, so we expect that we will continue to see consent forms that use language that has worked previously for researchers and which they've had no reason to question. So we're in the position, and will continue to be, of having to decide how to handle datasets that are submitted to us with problematic consent forms.

Slide 7: Problematic Consent Language

The next 2 slides will share examples from DRUM. The consent form language used by this researcher should look very familiar as it is the exact old template from our IRB. This is very common for us to see. There is often mention of records, reports, and publications - and if data is mentioned the language states it will be encrypted or shared in aggregate or shared only with the research team or shared only with researchers. In this instance, in 2018, we made the decision to not accept this dataset for ingest into

DRUM and advised that the researchers re-consent participants if they want to share in a public repository. The researcher decided not to, so we rejected the dataset. However, this is a point of inconsistency on our part, because, as we'll mention later, we have also interpreted identical language in other consent forms submitted to us in more favorable light, and accepted other datasets that have the same consent language.

Slide 8: Problematic Consent Language

Here is another example from DRUM where the researcher explicitly stated data would only be shared in aggregate, but they wanted to share individual-level data in DRUM. We advised that they return to the IRB, but the IRB had already determined that their work did not fall under the federal definition of human subjects research. In short, the researcher had not needed to consent their participants, but they had done so regardless. At that point we advised they had an ethical obligation to follow through with the agreement they made with participants.

Slide 9: Ethics

Why does all of this matter so much? Sharing human subjects data raises ethical considerations that don't exist with other types of research. Sharing human subjects data - when shared at the individual-level rather than in aggregate - increases the risk of re-identification, or exposing a participant's identity. Researchers are not trained in proper methods for de-identifying datasets.

Additionally, if the data are deposited into a public repository, there are no restrictions on how the data are used or who accesses it - this may be in direct conflict with what the participants agreed to in the consent form, or they may be completely unaware that this is how their data is being used. There is the possibility that if participants knew their data was being shared publicly, they would not have disclosed the same information about themselves.

However, when we've followed up with researchers or advised them to return to the IRB regarding sharing data under problematic or unclear consent language, we've encountered differing interpretations of the language and purpose of the consent form from researchers and IRB staff, which has emphasized for us over the past few years that we don't have any sort of human subjects data policy to fall back onto. So in order to develop a policy for DRUM, we wanted to analyze both our own processes over time as well as conduct a scan of other repositories' practices and policies

Slide 10: Evaluate DRUM practice

Our first task in considering what policies to incorporate about the ingestion of human subjects data was to evaluate our past practices - How has DRUM dealt with these submissions since our inception?

Slide 11: How has DRUM responded?

DRUM was launched in 2013, and, at the time this presentation was recorded, had 430 datasets, of which 46 contained human subjects data. Since 2013, we've had a review process in place for all datasets to ensure that no PII is included in the data. For human subjects datasets in particular, we request the consent form for our records. If we have any concerns regarding a dataset, we follow up with the researcher. So we have accepted almost all human subjects datasets that have been submitted to us, and we've rejected only three datasets. The new few slides will discuss some of the factors that go into our review process.

Slide 12: De-identification vs. Informed Consent

However, we've realized over time that while both our deposit agreement and our review process were designed primarily to catch deidentification issues, in practice, de-identification has not often been a concern. Rather, our concerns have more consistently centered around what the researcher told participants that they would do, with regards to sharing their data. This question of consent, then, is what we would like the policy we want to develop to address.

Slide 13: Evaluating Consent Forms

In earlier years, we reviewed that consent form only to ensure that the language did not explicitly *prevent* sharing data. We did not always request consent forms in the early years, so we have 28 consent forms to get a sense of what researchers are doing.

Seven consent forms that we've received do, actually, explicitly acknowledge that data might be shared. This language is often only a slight adjustment to existing boilerplate language.

We want to see more consent forms like the ones on the left, because unlike the language on the right, it leaves far less room for ambiguity for us, researchers, and research participants.

Slide 14: Which Disciplines are Submitting Human Subjects Data?

Our internal discussions and decisions have shifted over time into a combination of considering the consent form language, the nature of the data, and the research discipline altogether. We are an interdisciplinary repository. So when we discuss human subjects data in DRUM, it may be coming from disciplines outside of the health sciences, and be significantly less sensitive than the data that would be collected through medical research. So this has also been a guiding factor in our response.

Slide 15: Following up with Researchers

When we have questioned whether or not DRUM is an appropriate repository for a submitted dataset, we have historically referred the researcher back to the Institutional Review Board (IRB). Our experience, though, is that the IRB does not have the same mission that we do. They often give the "okay" as long as the dataset is de-identified and have told us that their involvement in the study is over after data analysis and manuscript publication. They do not want to review studies that they've already determined to be exempt or non-human subjects studies.

We've also used problematic datasets as individualized learning opportunities, but researchers tend to continue to submit datasets collected using restrictive consent forms, even after we discussed the issue with them at length. Due to our lack of official policy, DRUM has also had inconsistent acceptance practices - this only further confuses researchers who submit their data frequently. Education at an institutional level would be more effective, but implementing that can be challenging.

Slide 16: Repository Scan

The second arm of this study was to conduct a scan of existing repositories that share human subjects datasets.

Slide 17: Methodology

The initial criteria for identifying repositories was simply that they ingest human subjects data. Our goal was to analyze language related to participant consent on the website and in the deposit agreement. So we gathered a beginning list of possible repositories and narrowed it down through many scans of the websites.

Slide 18: Methodology

The initial list was created from the Data Curation Network, Big Ten Academic Alliance, PLOS ONE and NIH recommended repositories, and well-known general repositories. After duplicates and out-of-scope repositories were removed, we ended with 19 repositories for the analysis.

Slide 19: Identifying Relevant Information

We scanned entire repository websites, including external links to other websites and documents. If we could not locate a deposit agreement, we also emailed the repository to collect the deposit agreement or retrieve additional information about participant consent.

From the beginning of our analysis, we realized that the inconsistencies around how these guidelines were labeled was terribly confusing and would likely hinder the thoroughness of our analysis. Indeed, we combed the same websites three times and triple-checked our data collection table.

We identified - what we were calling deposit agreements - as required by the repository for data submission. Depositors had to agree to the terms laid out in these documents.

Deposit agreements were labeled:

- data sharing permission
- data submission request

- material transfer agreement
- institutional certification
- terms and conditions
- deposit agreement
- data submission agreement
- contribution form

Slide 20: The Issue of Labeling

Furthermore, this is how other web pages containing informed consent language were labeled. I'd like to recognize how ridiculous this slide is - you probably can't read it. There are 31 different labels under which we found information from the repositories regarding informed consent of participants. We were **looking** for this information - can you imagine being a researcher who is not looking for it?

We went into this study intending to analyze policies, but as you see here, only 5 website pages labeled themselves as policies. We wondered about the intentionality of this. Perhaps repositories are hesitant to declare policies around informed consent at this time, considering issues such as oversight.

Slide 21: Deposit Agreements

For our analysis, we were specifically looking for deposit agreements, i.e., a list of terms that the depositor had to agree to in order to submit data. Four repositories did not have them at all. Of the 15 that did, 12 of these agreements could be found on their websites, but 3 others required us directly inquiring with the repository whether or not they had one at all. In those circumstances, we were referred out to a number of sites: 2 repositories that actually housed the data (instead of the "landing repository") and a governance website, seemingly unrelated to the repository's funder or institution.

Slide 22: Deposit Agreements

Within the 15 deposit agreements, only 12 stated that data sharing should be consistent with informed consent. Of those 12, 9 go above and beyond to state that an ethics review board must also verify that data sharing is allowable or that limitations established in the consent form must be shared with the repository.

3 deposit agreements did not mention informed consent at all. This is especially problematic considering that the deposit agreement may be the only document that

submitters read.

Slide 23: Beyond the Deposit Agreement

We also scanned entire repository websites, including external links to other websites and linked documents. If we could not locate a deposit agreement, we also emailed the repository to collect the deposit agreement or retrieve additional information about participant consent. Through these methods (web scan, document scan, and email communication), we found language about the informed consent process for data sharing that was separate from the deposit agreement, and we categorized and counted those mentions.

Frequent Mentions:

- There were 21 instances where a repository stated that informed consent should address or give permission for data sharing or provide alternative mechanisms to share
- It was stated 15 times that depositors must submit the consent form or an ethics board verification that it's OK to share data
- There were 15 informed consent resources offered - for example, consent form templates or model language, an informed consent service, flow charts for data use restrictions, etc.

Moderate Mentions:

- There were 9 instances where repositories either stressed that informed consent should honor, safeguard, and/or protect participants or that the depositor should properly obtain consent to share
- There were also 9 mentions that addressed data use conditions, such as tiered consent, limitations, and restricted sharing options

Infrequent Mentions:

- There were only 2 instances where repositories shared how to handle removing a participant from the shared data when they withdraw consent to share

Not every mention was in a separate repository, so - for example - the 21 mentions about informed consent addressing data sharing came from 14 repositories. Some repositories had language about participant consent in a number of places on their websites, and some did not.

Slide 24: Comparison of Language

Another component we looked at was the consistency between informed consent

language found in the deposit agreement vs the language found elsewhere. It was great to see that in 10 cases, the language found on the website was, generally, consistent with the language found in the deposit agreement. However, in ⅓ of cases, the language found on the website or through other means did not reiterate or preface the informed consent language in the data agreement.

Slide 25: Repository Scan Takeaways

Most repositories that provided guidance around the consent process advised explicit data sharing language in the consent form - we were very happy to see this. However, there was enormous inconsistency in how the consent process guidance was labelled, where the information was located, and the level of detail provided. Also, the guidance was largely just that - guidance rather than a formalized policy, resulting in little or unclear enforcement. We also did not find any rejection mechanisms in place for circumstances where a depositor submits data collected with a consent form that says the data will not be shared. Finally, repositories that made it to our final round of review tended to be NIH repositories that specifically collected human subjects data within certain fields of medical research (although not all such repositories had guidance). Institutional or general repositories that collect data broadly tended to reflect DRUM's own practice of addressing only PII/confidential information in data submissions.

Slide 26: Limitations

It's important that we recognize our limitations - our initial recruitment list was not comprehensive - there could be repositories that ingest human subjects data and provide guidance on the informed consent process that we did not include in our analysis. As we've mentioned earlier in this presentation, the information we were seeking was not easy to locate - there was a lot of clicking through series of links; being led outside of the repository's website; digging through sub-sections, sub-sub-sections, and FAQs; and back-and-forth emails trying to get the correct information. These processes leave room for coding and analysis error.

Slide 27: Actionable Steps

As a result of the DRUM analysis and repository scan - as well as our past concerted efforts - we have put together 3 actionable steps you can take to ensure better practices

around sharing data that was collected via informed consent.

Slide 28: 1. Share Success Stories

First, share your success stories with researchers. It's important to share that very few researchers are resistant to our messaging about the importance of the consent form. They most often agree with us and work with us to find a solution that might work for their unique scenario. In this case, the researcher heard a presentation from us on data sharing mid-study. He and his research staff worked with us to re-write their consent form and send out to all participants. They reported it was a smooth process.

Slide 29: 2. Join Forces

Second, join forces within your institution. In this fuzzy organizational chart, what you cannot see is that right next to the Institutional Review Board is the Human Rights Protection Program. DRUM has forged a working relationship with this group. 2 of us serve on their Educational Advisory Group which tracks, delivers and advises on education to researchers about human subjects protections. We began requesting meetings with people in the Office of the Vice President for Research over 2 years ago; eventually we found an ear to bend and were invited to join this group. Not only are we privy to ongoing educational efforts at the institutional level, but this collaboration allows us to advocate for education around sharing data.

Slide 30: 2. Join Forces

For example, due to our membership on the Education Advisory Group, we were alerted to a "mini release" that would make updates to the Social Behavioral Templates. This slide shows just one recommendation we made, but they welcomed our input on the entire social behavioral consent form and protocol. Additionally, we shared a tool with them which we created to alert study participants to the reasons and processes around sharing their data.

Slide 31: 3. Enact 5 Best Practices

As more and more researchers share their human subjects data, we, as curators, gain valuable knowledge and insight into best practices. Our repositories should reflect those best practices to guide researchers. So the third actionable step is to enact these 5 best practices for sharing human subjects data.

Incorporate a deposit agreement that is easy to find and reflects other guidelines found on your website.

Require that depositors submit the participant consent forms with their data submission, and establish if/how you will evaluate those consent forms. There is a spectrum for handling consent forms - you might just keep the consent form as part of the data's underlying metadata record, or you could evaluate the sharing permissions laid out and address those with the researcher.

If you are unsure about how to reconcile the consent form with the type of data sharing that your repository does, defer to the IRB. Keep in mind, however, that IRBs in general are less knowledgeable about data sharing and can downplay the significance of informed consent.

Not all human subjects data submissions will have gone through an IRB process, but a participant agreement might still exist - you should have an established practice in place for how to handle these situations when they arise.

And finally, at every chance - establish early-intervention education to complement these policies. At the UMN we lecture in classrooms, at seminars and ethics events, to research groups and professional associations...pretty much anyone who will listen. Our hope is that the next wave of researchers is thinking about data sharing at the moment of informed consent.

It's important to note that DRUM has not yet formed an official policy around the sharing human subjects data in our repository.

Slide 32: Thank you

Thank you for listening. On this slide you will find a link to DRUM as well as our email addresses if you want to connect with us.