A GUIDE FOR SHARING HEALTH DATA

*For non-commercial research*

|  |
| --- |
| **CONTEXT: For non-commercial research in computational medicine**There is no one-size fits all guide to sharing health care data for computational medicine research across parties. This legal guide is meant to make the process more streamlined by giving researchers a tactical overview of what to expect and how to anticipate legal requirements.  |
| **CONTENT:** This document contains three tools that may be useful to streamline your review process:* **PART I: RESEARCH COVER SHEET (P.2):** A template 1-pager to include with your Data Use Agreement to orient reviewers to your project
* **PART II: DATA SHARING ISSUES GUIDE (P.3):** Areas and questions that we recommend you think through before the come up in the review process
* **PART III: RED FLAGS CHECKLIST (P.5)**: A list of aspects about your project that could delay review
* **Part IV (P.7):** Link tree to common research resources

You can further support your team with templates for crafting your own process map, an annotated DUA template, and a legal white paper on data sharing [available for free download at our website](chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/cmla).  |
| **INTENDED AUDIENCE:** We designed this Legal Guide for Data Sharing for Non-Lawyers to be used by: * Researchers and support staff
* Compliance teams
* Data security and governance teams

**HOW TO USE: Edit for your institution, share where you access information**This document should be customized for your organization and be shared wherever your team is most likely to access resources.  |
| **ABOUT: Computational Medicine Research Accelerator**The Computational Medicine Legal Accelerator (CMLA) is working to accelerate machine learning research aimed at improving clinical care through more streamlined legal processes for sharing health data between institutions. Our fundamental goal is to safely get more health data into the hands of machine learning researchers faster, in order to accelerate the development of computational medicine.  |

*PART I: RESEARCH COVER SHEET*

|  |
| --- |
| This research cover sheet is intended to give all parties (e.g. researchers, legal, support staff) who may be involved with the data sharing process an overview of the level of risk involved in your dataset. It is one structured format to centralize the details of data transfer and to ensure effective information flow. **Recommendation:** To best prepare your research for an efficient legal and security review, please complete this page and share it with those who would benefit from understanding more about your project.  |
| **What is the purpose of the project and how will your findings be used?**  |
| *[Insert your response here]* |
| Check all that apply to ensure a clear assessment of the risk profile of this data |
| Data |
|  | There is identifiable data in my dataset (consult Part III to understand what is considered “identifiable”)  |
|  | My data is from outside the US or a non-HIPAA covered institution |
|  | My data involves vulnerable populations as defined by federal regulations on human subjects research (i.e. pregnant folks/neonates/fetuses, prisoners, or children) |
| Access and Security |
|  | I will have collaborators from outside my institution working with this data |
|  | Data storage infrastructure will have to be built for this project |
| Process |
|  | I am interested in commercializing this research |
|  | I have consulted with my institution’s IRB or a central IRB regarding the need for IRB review |
| Project Details: This context is important to share with all parties to ensure an efficient approval and review process |
| **Data** |
| Sensitive Data Fields (if applicable) |
| For example: * *[Insert identifiable data fields here and/or attach data schema]*
* *[List any data fields relating to categories of persons protected under federal regulations on human subjects research*]
 |
| Source of Data |
| For example: * *[Country of origin, company, institution]*
 |
| **Security** |
| Storage |
| *For example:* * *[Where will the data be stored?]*
* *[Who will have access to the data? Please include affiliations and titles]*
 |
| Source of data |
| *For example:* * *[Country of origin, company, institution]*
 |

*STAKEHOLDERS LIST*

|  |
| --- |
| Taking the time to organize all stakeholders involved in the data sharing and acquisition process ahead of time will save you time in the long run. This contact list may not need to be updated for every project and should integrate any current intake forms/inboxes you already have at your institution. If you are unsure who to contact--please consult your internal resources. **Recommendation:** Share the Researcher Cover Sheet above (including the contact list) to all contacts involved to ensure alignment across teams. |
| Research Team |
| [Name and title] | [Email, phone number, contact] - Add row(s) below for more names as needed. |
| Data Security and Governance |
| [Name and title] | [Email, phone number, contact] - Add row(s) below for more names as needed. |
| Legal |
| [Name and title] | [Email, phone number, contact] - Add row(s) below for more names as needed. |
| Data Partners or Providers |
| [Name and title] | [Email, phone number, contact] - Add row(s) below for more names as needed. |
| Data Partners or Providers |
| [Name and title] | [Email, phone number, contact] - Add row(s) below for more names as needed. |
| Other Key Stakeholders |
| [Name and title] | [Email, phone number, contact] - Add row(s) below for more names as needed. |

*PART II: DATA SHARING ISSUES GUIDE*

This data sharing issues guide is intended to help you anticipate possible areas of friction and help your team proactively think through your options and answers. We recommend having a good understanding of the below issues in relation to your data sharing and data use agreement, so that you can quickly respond to questions during the review process. Your legal team can find a detailed white paper on these and other issues here.

|  |  |
| --- | --- |
| **Question** | **Guidance**  |
| **Data Security** |
| Where will the data be stored once it is transferred (the data environment)? | Knowing where data will be stored is critical to understanding whether the data recipient will be able to meet relevant requirements from the data provider. Waiting to determine where data will be stored can significantly delay a project, particularly if a new environment will need to be built or even approved by the data recipient. |
| Who owns or controls the data environment? | If this is a third party, like a cloud service provider, the data use agreement will need to allow for that third party access or use.The terms of any agreement with the third party should be reviewed to understand what rights they may give the third party and whether they conflict with the data use agreement. |
| How will the data be transferred? | The answer here will speak to necessary security requirements and also be helpful to think through so that it does not cause delay after the data use agreement and transfer is approved. |
| What are the security requirements set by the data provider for storing and transferring this data, and how will the data recipient meet them? | We recommend having this discussion simultaneous with the drafting of the agreement to ensure that the data use agreement reflects the appropriate agreement. |
| Who can access the data once it’s been transferred? | Please think through the entire environment, not just who is expected to access the data, but who has the ability to access it. |
| Are there any third parties involved in the research? | This should include third party services providers, research collaborators, and others. |
| **Publication** |
| What are the data recipient’s plans for publishing based on the transferred data? | Essentially, will any research results be (hopefully) published? |
| Does the data provider want to receive source attribution or publication review rights? Is that OK for the data recipient? | We recommend having this conversation early in the data transfer discussion. |
| **Intellectual Property** |
| Does the recipient plan to use the data solely for research purposes? | These answers help to determine what IP rights will need to be granted to the parties to begin research. |
| Does the data recipient intend to use the data for commercial purposes or any other use besides research or academic purposes? | “Commercial purposes” can be really broadly understood. We recommend flagging anything that could potentially be viewed as commercial use, or otherwise outside of the ordinary course of research. |
| Are you receiving federal funding? | You may be required to grant a non-exclusive, royalty-free license to the United States government under certain inventions arising from the use of such funding. Any research conducted through use of such funds may also be subject to the Common Rule and other federal regulations on human subjects research. |
| **Data** |
| Do you have a detailed and complete understanding of all the data fields in the set? Can this schema be easily shared with IRB and legal review teams?  | Knowing exactly what data you are receiving and sharing with all parties is one of the best ways you can ensure a smooth acquisition process. If your data contributor can’t provide this, then we strongly recommend creating a data dictionary.  |
| **People Process** |
| Who needs to review this data transfer? Common potential areas of review include legal, compliance, security? Does your legal team have capacity? | Anticipating your team’s capabilities and capacity will help keep expectations reasonable and aligned. |
| Will you need IRB approval or review? Have you already consulted with the IRB regarding the potential need for review? | Consulting with the IRB in parallel will be helpful. Data contributors sharing data that have previously been collected will not automatically trigger IRB approval under the Common Rule; however, your institution may require IRB review. In addition institutions receiving the data may need IRB approval from their own institution.  |

*PART III: RED FLAGS CHECKLIST*

This checklist highlights common causes of delay in legal review, these are not deal-breakers but knowing in advance if any of these issues apply to your data set will enable you to be proactive in addressing the concerns of your institution.

|  |  |  |
| --- | --- | --- |
| *LIST OF COMMON ROADBLOCKS*Note: Legal review can vary in duration depending on the complexity of the elements present below. |  | *EXAMPLES OF IDENTIFIABLE DATA*This list is not exhaustive and is based on the identifiers that must be removed in order for data to be considered de-identified under HIPAA’s “safe harbor” method of de-identification..  |
| ☐ | 1. **Data is from outside the US**

If any data are to be transferred from or shared with institutions outside of the US, additional laws, such as the European Union’s General Data Protection Regulation (“GDPR”), could apply. |  | * Names
* All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes (limited exceptions apply for ZIP codes)
* All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
* Telephone numbers
* Vehicle identifiers and serial numbers, including license plate numbers
* Fax numbers
* Device identifiers and serial numbers
* Email addresses
* Web Universal Resource Locators (URLs)
* Social security numbers
* Internet Protocol (IP) addresses
* Medical record numbers
* Biometric identifiers, including finger and voice prints
* Health plan beneficiary numbers
* Full-face photographs and any comparable images
* Account numbers
* Any other unique identifying number, characteristic, or code (except as permitted for re-identification purposes by the originally de-identifying entity in certain circumstances); and
* Certificate/license numbers
 |
| ☐ | **2. Your project does not cleanly meet the definition of research** HIPAA, like the Common Rule, defines “research” as “a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.” Activities like “quality improvement” or “quality assurance” often do not meet the definition of “research” under this rule. The line between “research” and “quality improvement/quality assurance” can sometimes be difficult to draw.  |  |
| ☐ | **4. The data deals with special categories of people under the Common Rule, including: Minors, individuals who are pregnant, prisoners, substance abuse information, mental health information, wards of the state, disabilities, people with HIV.**These populations are considered vulnerable and thus require additional protections and security. |  |
| ☐ | **5. The data is identifiable**Review the list of identifiable data. |  |

|  |
| --- |
| Personally Identifiable Information (PII) as any information that can be used to distinguish or trace an individual’s identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. The definition of PII is not anchored to any single category of information or technology. Rather, it requires a case-by-case assessment of the specific risk that an individual can be identified. |

*PART IV: OTHER RESOURCES*

This list contains resources that may be helpful to folks new to working with health data or the legal requirements for research.

**Recommendation:** Add your institution’s research resources (e.g. IRB, university research administration and legal counsel) to this table for easy reference.

|  |  |
| --- | --- |
| **Source** | **Description** |
| **Human Subjects and Common Rule** |
| [Office for Human Research Protections website](https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html) | US government site with text of the Common Rule regulation; the Belmont Report; and training on all things human subjects research.  |
| [National Institutes of Health (NIH) Grants and Funding website](https://grants.nih.gov/policy/humansubjects/research.htm) |  US government site with decision tools for determining human subjects research; exemptions; and how to manage biospecimens.  |
| [Federal Demonstration Project (FDP) Human Subjects Data Classification](https://thefdp.org/default/assets/File/Documents/human_subject_data_classification_tool.pdf) | A tool to help classify data fields as HIPAA Identifiers, limited data, de-identified data, or subject to FERPA |
| **Legal Landscape** |
| [Legal Whitepaper on Computational Medicine Research](https://drive.google.com/file/d/1LMa0PdhQHkfQxDJ6RdO3kSyZ2ieGMnFJ/view) | In-depth overview of the regulatory, contractual and operational requirements, and challenges in sharing health data between research institutions for computational medicine research.  |
| **Institutional Review Board (IRB)** |
| [ Your institution’s resource]  | [How this resource can be used]  |
| **Data Use Agreement (DUA)** |
| [Annotated DUA and template](https://drive.google.com/file/d/1HvVQGi9dyUutQ8AsPTC_gANWEZEDLYLF/view) | A CMLA resource that explains each component of the DUA in plain language with a modular template.  |
|  [FDP glossary](https://thefdp.org/default/assets/File/Documents/dtua_glossary.pdf) | Glossary of common legal terms involved in Data Use Agreements |
| [ Your institution’s resource]  | [How this resource can be used]  |
| **Process Flow** |
| [Process Maps for Data Sharing](https://docs.google.com/presentation/d/11Nx1mnHsU2KDAY57KiW8PDhOxAGNXDa3ai__582DrXw/edit?usp=sharing) | A CMLA resource and template that maps the steps in data transfer and acquisition at specific institutions.  |