V2.0 | 7/17/2025 | Authors/Owners: T. Burdick, R. Mosher, K. Anton

Dartmouth Health

Research Informatics and Data Handbook

Acknowledgement

The authors wish to thank the following individuals for their contribution to this Handbook:

Brock Christensen (Geisel School of Medicine)
Carol Felone (Analytics Institute)
Cristina Ferrazzano Yaussy (ORO Research Quality and Safety)
Jennifer Driscoll (ORO Clinical Trials Office)
Jennifer Snide (Analytics Institute)
Jessica Chevalier (Research Operations Manager)
John Mecchella (Clinical Informatics)
Kristen Anton (Research Informatics)
Michael Hill (IRB)
Monica Polk (ORO Clinical Trials Office)
Pamela Burghardt (Analytics Institute)
Rebecca Butcher (CPDE)
Rebecca Malila (Privacy Institute)
Rebekah Thomas (Research Operations Manager)
Richard Eikstadt (Analytics Institute)
Robyn Mosher (Research Informatics)
Scot Wingo (Analytics Institute)
Shannon Sewards (IRB)
Timothy Burdick (Research Informatics)
Todd Morrell (Clinical Informatics)
Tonya LaClair (IRB)

Table of Contents

Scope

This document describes informatics procedures and activities applicable to all research conducted at Dartmouth Health, including Dartmouth-Hitchcock Medical Center and affiliated Clinics.

What is the purpose of this handbook?

The Dartmouth Health Research Informatics and Data Handbook is a guide for requesting and working with data and informatics systems from Dartmouth Health (DH) for research.

The use of clinical data for research is regulated by institutional, state, and federal guidelines and statutes, and this Handbook provides information that researchers need to conduct research responsibly and in accordance with DH policies and legal requirements.

Please read all sections of this handbook carefully.

Principal investigators and study team members are responsible for conducting research with the appropriate approvals, contracts, procedures, and data protections.

The DH Office of Research Operations (ORO) and SYNERGY Research Informatics team is available for consultation to answer specific questions and facilitate the acquisition and use of DH research data or informatics systems. Researchers may request assistance by submitting a SYNERGY Informatics Consult Request form, found in the 'Data and Informatics' section of the SYNERGY website.

Please note that hyperlinks to DH policies included in this document are accessible only to DH employees.

What are the key considerations for acquisition and use of DH data for research?

There are important considerations that will determine what is required from the research team as research data is acquired and managed. These topics are covered in this handbook. For example:

- What training is required for investigators and staff who plan to work with DH data for research?
- How is research data classified, and how does that classification impact how research data is managed?
- With whom can researchers share data, and under what conditions?
- Does use of research data require researchers to obtain any special contracts?
- Does the research study require review and approval by the DH Research Informatics and Data Governance Group (RIDGe)?

- Is an Accounting of Disclosures needed?
- Can researchers get the research data they need by collaborating with the DH Analytics Institute?
- What computerized systems are available, or required, for research use?

What does a typical data acquisition workflow look like?

Study teams will likely interact with several Office of Research Operations (ORO) groups in the process of activating a research study at DH. These groups include the DH Human Research Protection Program (HRPP)/Institutional Review Board (IRB); the DH Clinical Trials Office (CTO); and the DH Research Informatics and Data Governance Group (RIDGe).

Study teams will also interact with several electronic systems that will shepherd the required research documents through the approval process to activation; systems may include eIRB, eREG, and OnCore.

The diagram below is a high-level overview of a typical research study activation workflow, designed to provide a general idea of the process. The landscape of clinical research is vast, and the exact steps to follow, and the documents and approvals needed, will depend upon the characteristics of the research protocol and the research data that is required.

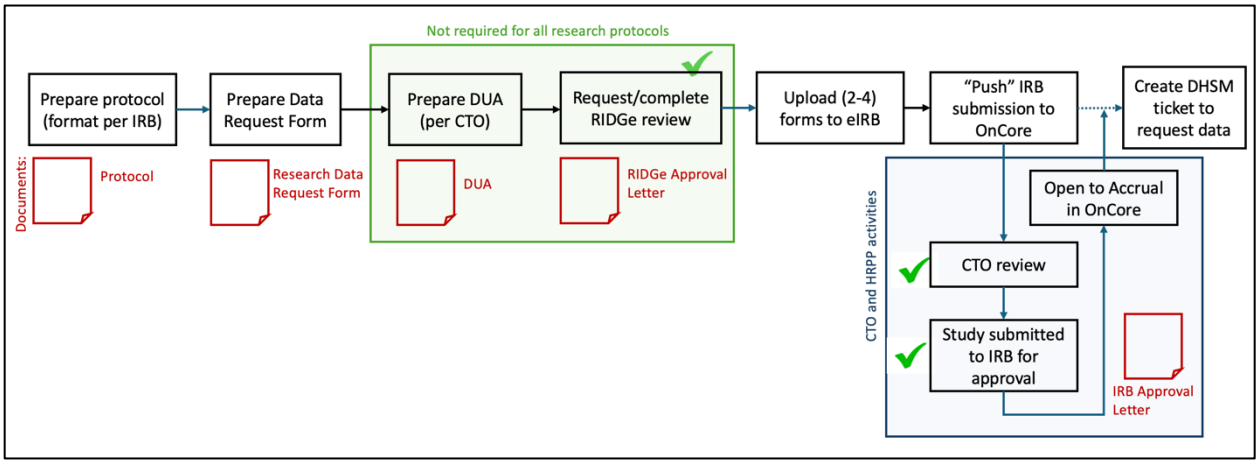


Figure 1. High-level overview of the process for opening a new investigator-initiated study at DH, including documents to be generated (in red), and possible steps and corresponding documents (in shaded green box). The blue shaded box denotes activities that may be performed by ORO teams, and green check marks indicate areas where reviews may be required. The characteristics of a study will determine what documents are needed, what contracts may be required, and what steps are needed to active the study.

What training does a study team need to conduct research with DH data?

All members of the study team who are DH workforce, both investigators and staff, must complete the Collaborative Institutional Training Initiative (CITI) biomedical research online training program. The CITI site can be accessed at http://www.citiprogram.org/. Training is valid for a three-year period, after which time the training must be refreshed. Note that researchers whose paymaster is Dartmouth College follow Dartmouth College requirements for research training.

If research includes human subjects, the ORO requires specialized training in accordance with DH Policy ID# 23146, entitled "Research Training Requirements for Principal Investigators and Personnel Engaged in Human Subjects Research."

Study team members may need to complete additional training based on other federal, state, or institutional policies as well as specialized training for interacting with secure software platforms. For example, cybersecurity training is required for working with Protected Health Information (PHI) in Health Insurance Portability and Accountability Act (HIPAA)-compliant DH computing platforms (details provided in this Handbook in the section for software platforms).

Opportunities for additional research training are detailed in the 'Education and Training' section of the SYNERGY ORO website.

Who is considered a member of the DH workforce for the purposes of research?

It is important to understand who, in the context of DH research, is considered a member of the DH workforce, with the rights and responsibilities that membership conveys. Many data-focused DH policies reference DH workforce and the DH Covered Entity. HIPAA regulations define a 'covered entity,' in part, as a healthcare provider involved in the transmission of protected health information (PHI). DH is a Covered Entity.

Under HIPAA regulations (specifically 45 CFR part 160.103), the term 'workforce' is defined as employees, volunteers, trainees, and other persons whose conduct, in performance of work for a covered entity or business associate, is under the direct control of the covered entity or business associate, whether or not they are paid by the covered entity or business associate.

DH Workforce (staff who are members of the DH Covered Entity) specifically includes system members of:

- Dartmouth Hitchcock Clinic
- Mary Hitchcock Memorial Hospital
- Alice Peck Day Hospital
- Mt. Ascutney Hospital and Health Center

- New London Hospital
- Cheshire Medical Center/DH-Keene
- Hanover Psychiatry
- Visiting Nurse and Hospice of VT/NH
- All Northern Locations
- Community Group Practices in Southern NH (Manchester ASC, Nashua ASC)
- Southwest Vermont Medical Center
- Valley Regional Hospital
- Hampstead Hospital

The DH research team often includes an 'affiliate' of DH. A DH affiliate is someone with a formal relationship to the DH Covered Entity, typically contractual, who is not a member of the DH Workforce. Affiliates of DH, e.g., Dartmouth College employees, Dartmouth students (undergraduate, graduate), Geisel students who are not MD students, students from other academic institutions, and community members are not considered members of the Dartmouth workforce. Additionally, investigators and staff who have a dual affiliation (DH and Dartmouth College) but who are working under their Dartmouth College affiliation for research are not considered members of the DH workforce. This handbook provides information for accessing, using, disclosing, and sharing DH data with individuals or organizations that are not part of the DH Covered Entity.

As stated, generally DH Workforce does not include undergraduate or graduate students from Dartmouth College. The one exception is Geisel MD students who are considered DH workforce when:

1. The Geisel MD student is working under the direct guidance of a mentor who is both DH workforce and Geisel faculty, *and*
2. The Geisel MD student and the faculty mentor are documented study team members in the IRB protocol.

Principal investigators and study team members are responsible for understanding who is included in the study team, and what contracts and procedures are required to ensure the privacy and security of DH data. The Clinical Trials Office team (cto@hitchcock.org) will assist with this assessment and development of appropriate contracts.

DH Research Informatics and Data Governance Group (RIDGe)

What is the DH Research Informatics and Data Governance Group (RIDGe)?

Formed in 2023, the DH Research Informatics and Data Governance group (RIDGe) brings together expertise from across the DH research and informatics community to expedite the activation of DH investigator-initiated research projects while ensuring DH data are appropriately handled. RIDGe participants include the leadership of ORO Research Informatics, HPPR/IRB, the Analytics Institute, the Clinical Trials Office, Clinical Informatics, Research Quality and Safety, and the Privacy Office who come together weekly to review research protocols that require special approval for data or informatics platforms (see below). DH researchers do not need to bring studies to the Privacy Office for review (RIDGe review includes Privacy Office concerns).

RIDGe has replaced the circulation of protocols among offices for approval with a collective discussion of data-related issues, within one meeting. A research study undergoing RIDGe review can expect to receive communication from RIDGe within 1 – 3 weeks of submission. Communication with RIDGe is often iterative, and the efficiency of the RIDGe review will depend on the principal investigator's timeliness in responding to RIDGe communications.

What projects require RIDGe review?

RIDGe will review a research study if the protocol is proposing:

- Manual chart review for data abstraction
- Accessing PHI (e.g., from eDH and other electronic health record sources) for preparatory to research activities, where preparatory to research activity is defined as the use or disclosure of PHI (without removal of the PHI from the DH Covered Entity) for the purpose of determining the scope, feasibility, and/or potential participants for a future research study
- Use of clinically sensitive data, business confidential data, or social security numbers (see Handbook section on data with special consideration for more information)
- Use of third-party software, websites, or information technology not approved for use by DH Information Systems (IS) Security
- Storage of research data on systems or platforms not approved for use by DH IS Security
- A protocol modification that proposes any of the above activities

On occasion, other teams reviewing a protocol, e.g., IRB, Analytics Institute, or CTO, may request that RIDGe review a study.

Study teams are encouraged to request a consult if unsure whether the research study requires RIDGe review. The study team can request assistance via the SYNERGY Informatics Consult Request form found in the 'Data and Informatics' section of the SYNERGY website.

What projects do not require RIDGe review?

RIDGe reviews are required only for research protocols meeting the criteria as described in the prior section. There are data-focused activities and studies that do not require RIDGe review. These include:

- Projects that are not considered research
    - Studies that do not meet the regulatory definition of 'Research,' including Quality Improvement (QI) activities
    - Studies that do not meet the regulatory definition of "Research with Human Subjects," as determined by the IRB
- Projects with close oversight by the Clinical Trials Office (CTO)
    - Sponsor Initiated Trials, including Cooperative Group trials
    - Expanded Access (also called Compassionate Use) trials

CAUTION: While RIDGe review is not required for studies that do not meet the definition of research, these studies still require CTO review, which is triggered through appropriate workflows in OnCore.

Appendix A-3 includes an instrument to assess the differences between research that requires IRB oversight and quality improvement activities; QI is addressed in more detail in the 'Quality Improvement (QI) Activities' section of this Handbook.

Study teams are encouraged to request a consult if unsure whether the research study requires RIDGe review.

Is a Research Data Request Form (DRF) needed?

The DRF is required for RIDGe review of studies requesting to access patient data (e.g., manual chart review) and is used by RIDGe, the IRB, and the Analytics Institute.

The DRF is not required for studies that do not involve accessing data from the medical record, for instance studies requesting use of 3rd party software for research.

The Research Data Request Form (DRF) contains metadata about the research study, e.g., Principal Investigator (PI), cohort, data acquisition activity, and
provide the specific data elements to be included in the data set. Documenting and defining each data element to be collected can be time consuming, however, the up-front work will

ensure that the study will generate a meaningful data set for analysis. The DRF is designed to provide RIDGe with the information needed to assess and approve the study, so it is imperative that all appropriate sections of the DRF be completed with sufficient detail. Researchers can consult with the department's Research Operations Manager, Vice Chair for Research, or other faculty mentor for assistance completing the DRF. DRF templates are available on the SYNERGY website, as described in STEP 3 of the 'How to request RIDGe review of a research project' section, below.

How to request RIDGe review of a research project

Follow the steps below to prepare the research study for RIDGe review and initiate the review.

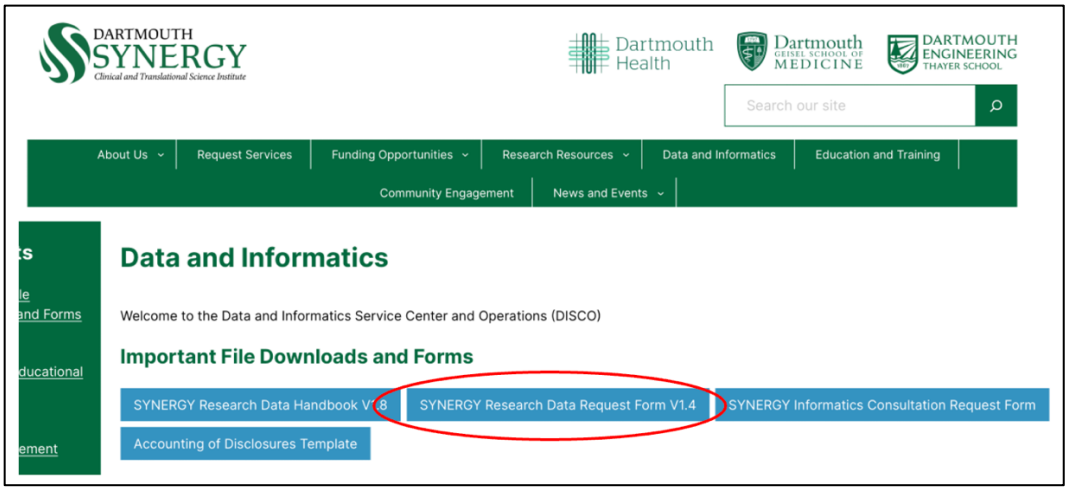
STEP 1: Prepare the research study Protocol in the format required for submission for IRB review (HRP-503, found in eIRB in the Library of Templates).

STEP 2: Download the most current version of the Data Request Form.

Navigate to the SYNERGY website.

Navigate to the 'Data and Informatics' page of the SYNERGY website by clicking on the 'Data and Informatics' tab (shown in red circle, below).

Download the current version of the SYNERGY Research Data Request Form by selecting the appropriate link (shown in red circle, below) and selecting the file to download. The Data Request Form is in Microsoft Word format.

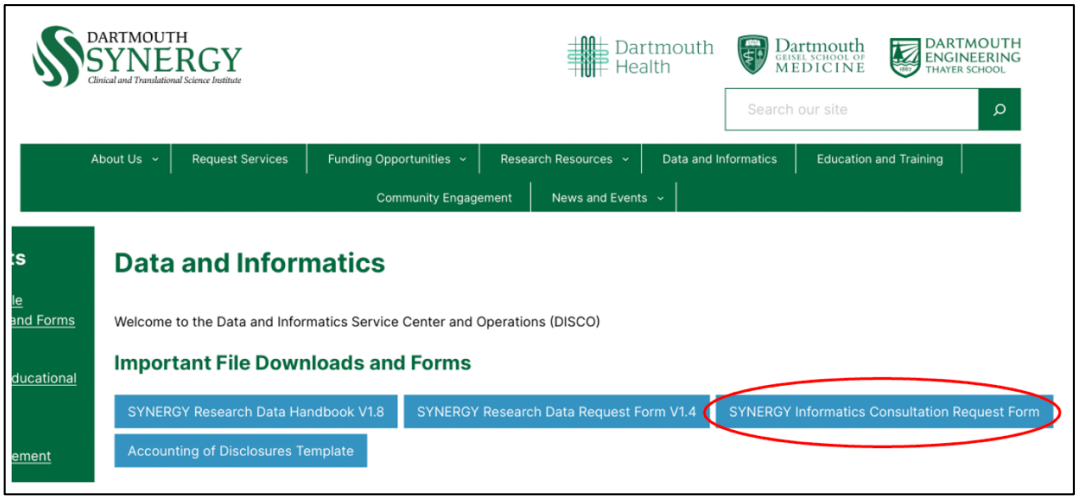


STEP 3: Complete the Data Request Form for the research project. A complete and accurate Data Request Form will minimize iteration between the study team and the ORO Research Informatics team in preparation for the RIDGe review.

> *TIP: Check that the common information in the Protocol and Data Request Form, for example, descriptions of cohort and data collection methodology, are identical. Substantive discrepancies may cause delays in RIDGe review.*

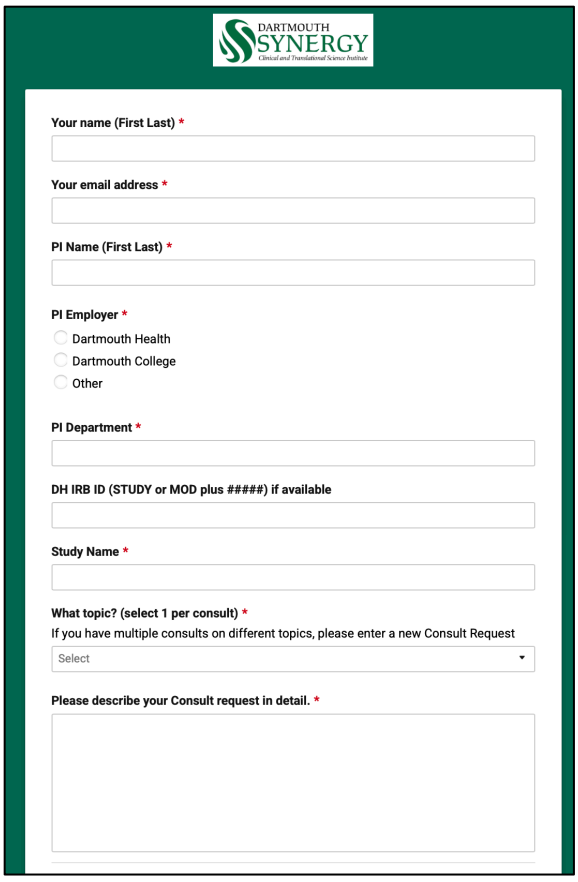
STEP 4: Request a SYNERGY Informatics consultation for RIDGe review.

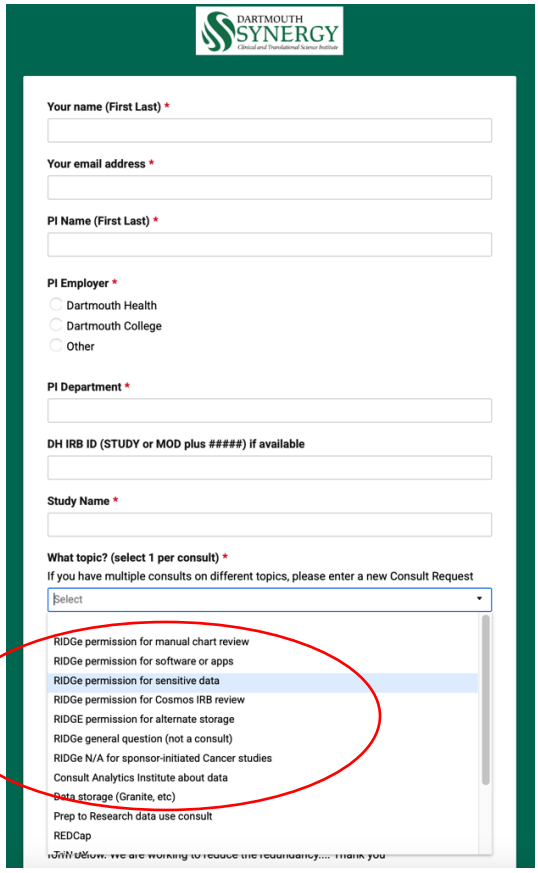
Navigate to the 'SYNERGY Informatics Consultation Request Form' by selecting the 'SYNERGY Informatics Consultation Request Form' tab (shown in red circle, below).



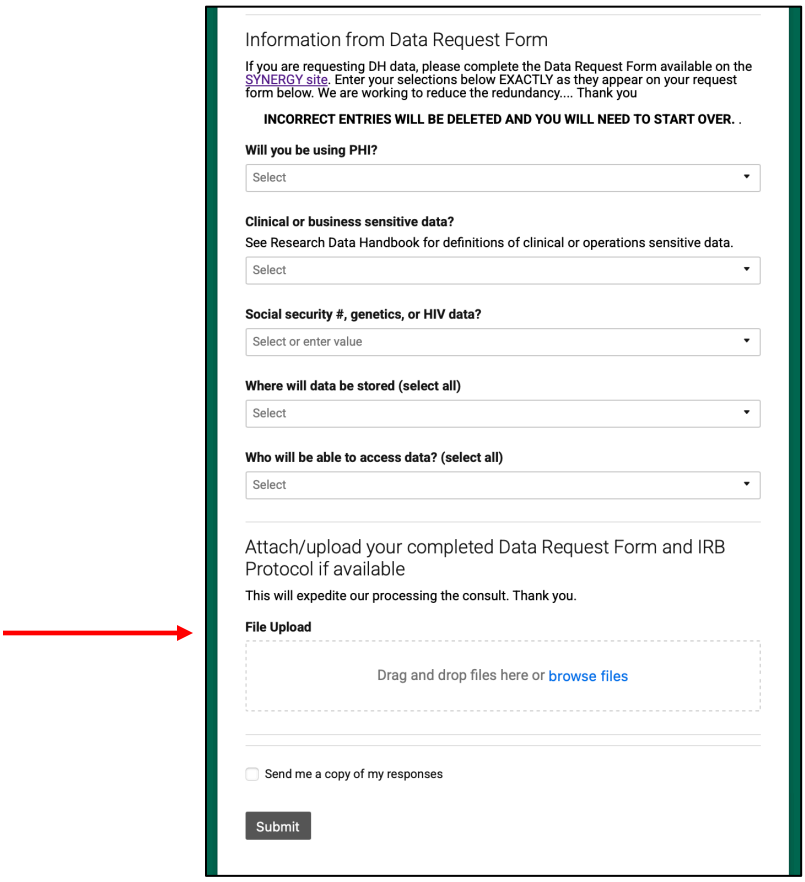
This link will bring up a fillable form where researchers will enter information that provides the Research Informatics team with information about the research team, study, and the specific consult request. The top portion of the form, shown below left, requires input of personal and basic study information.

In the 'What Topic' field, select the most appropriate 'RIDGe Permission' option (field and options shown, below right).





Populate select information from the completed Data Request Form into the bottom portion of the form, shown below.



When all information has been entered, attach the Protocol and Data Request Form in the 'File Upload' section of the form (see red arrow, above). If preferred, provide the data element specification in spreadsheet format instead of within the Data Request Form and upload that document here as well.

When the completed the form and required documents have been uploaded, select 'Submit.'

Upon submission, the information will automatically move into the RIDGe queue. If there are any questions, someone from Research Informatics will reach out to the study team for clarification.

How will the outcome of the RIDGe review be communicated?

Following RIDGe review, the researcher will receive a determination letter by email that may approve the project or request revision.

Upon RIDGe approval, the study team should upload the RIDGe determination letter, with other required documents, into eIRB.

The Research Informatics team will contact the study team to schedule a meeting if RIDGe identifies issues with the proposed study. The teams will collaborate to develop revisions that enable the study to proceed.

Classification of Data

What is data classification?

Data classification is the process of organizing data into categories based on the level of sensitivity, value and criticality to the institution. Data at DH is classified into one of three sensitivity levels: Confidential Data (Tier 1), Internal Use Only Data (Tier 2), and Public Data (Tier 3). The DH 'Data Classification Policy' (Policy ID# 10084) establishes a tiered framework for classifying and controlling DH Enterprise Data.

Confidential Data (Tier 1): Data for which the unauthorized disclosure, alteration, or destruction would cause a grave level of risk. Confidential Data is always restricted and must be carefully controlled. These data are to be accessed on a "need to know" basis. Confidential Data include (but are not limited to):

- Data protected by state and/or federal regulations
    - Protected Health Information (PHI)
    - Personally Identifiable information (PII)
    - Social Security Numbers (SSN)
    - Germline genomic data
    - HIV testing status or results
- Data protected by confidentiality agreements
    - Non-disclosure agreements and certain other contracts
- Internal DH information
    - Business sensitive information

Internal Use Only Data (Tier 2): Data for which the unauthorized disclosure, alteration, or destruction would cause a low to moderate level of risk. Internal Use Only Data is not for release to the public and is considered sensitive. These data include (but are not limited to):

- Internal company newsletters
- Project plans and documentation
- Operations policies

Public Data (Tier 3): Data for which the unauthorized disclosure, alteration, or destruction would cause little to no risk.

All research personnel are responsible for understanding the classification of data with which they are working, and what protections are required for that data.

What is HIPAA Protected Health Information (PHI)?

HIPAA Protected Health Information (PHI) is any information in the medical record or designated record set that can be used to identify an individual and that was created, used, or

disclosed in the course of providing a healthcare service such as diagnosis or treatment. It is the data that is collected and held by a covered entity.

In accordance with DH policy, **PHI is Confidential (Tier 1) Data**, protected with the highest level of security. It is extremely important to understand PHI since mistakes can lead to unlawful violation of patient privacy and significant federal fines.

PHI contains one or more of the following 18 identifiers:

1. Names (including initials)
2. Geographic subdivisions smaller than a state
3. Dates (except year)
4. Telephone numbers
5. Fax numbers
6. E-mail addresses
7. Social Security Numbers (full or partial)
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers/serial numbers
13. Device identifiers/serial numbers
14. Web URLs
15. IP address numbers
16. Biometric identifiers (fingerprints or voiceprints)
17. Full face photographs or images
18. Any other unique number, code, or characteristic that can be linked to an individual, including the Epic CSN (Contact Serial Number) and HAR (Hospital Account)

If any **1** of these **18** identifiers exists within a research data set, the data set contains PHI and is considered a fully identifiable data set.

Research personnel are responsible for following the safeguards required for the protection of PHI, including those relating to Data Use Agreements, Accounting of Disclosures, and proper, secure data storage that are described in this Handbook.

What is Personally Identifiable Information (PII)?

Personally Identifiable Information (PII) is a broad term that encompasses any information that can be used to identify an individual, such as name, address, and social security number. PII is considered PHI under HIPAA regulations when the data relates to health information and the data contains elements that can be used to identify a specific individual. Not all PII is PHI.

Patients participating in some research protocols, e.g., patient surveys, interviews, and focus groups, voluntarily share PII or feedback about health encounters. Study teams providing

incentives like gift cards to study participants are required to collect PII (including social security number) on W-9 forms. These data are not considered HIPAA-protected PHI, however the PII is considered Confidential Data (Tier 1) and should be appropriately protected.

What data are handled with special consideration?

In addition to PHI, Clinically Sensitive and Business Confidential data are handled with special consideration, as mandated by DH policy and/or federal and state laws. These data are

Confidential (Tier 1) Data, protected with the highest level of security, may only be used under certain circumstances, if at all. Research protocols requesting the data types below must be reviewed by RIDGe before the protocol can be submitted for IRB approval. Clear justification for inclusion of 'special consideration' data must be included with the request for RIDGe review.

Social Security Number (SSN): By federal statute, Social Security Numbers may not be disclosed or used for research. Note that SSNs can be used for federal (IRS) reporting of study payments, as long as they are not being used for the research itself.

Substance Use Disorder (SUD) Clinic: By law, use of data pertaining to care delivered in a designated SUD treatment clinic (e.g., dates of treatment, location of clinic, diagnoses, medications, urine drug testing and other laboratory test results, clinic notes) may not be included in a research data set. For instance, DH Rivermill Addiction Treatment Program is covered under this statute. Data describing treatment for SUD that is delivered outside of an SUD clinic, e.g., treatment for SUD in a primary care clinic, Emergency Department, or during a non-psychiatric medical hospital admission, is exempt from this prohibition and may be included in research.

HIV and Germline Genetic Testing: NH State law strictly limits access to and use of HIV and Genetic Testing data for research purposes. As described in "Understanding NH Title X – Public Health Chapter 141- H – Genetic Testing," DH interprets the legal mandate to differentiate between germline genetics, and somatic genetics (including tumor or virus/pathogen genetics). Patient germline genetics are considered clinically sensitive, where somatic genetics are not considered clinically sensitive.

When PHI is included with HIV and/or germline genetic testing data in the research data set, researchers must obtain written patient consent to access and use HIV or germline genetics data. Written authorization is likewise required for the generation of a data set that includes patient identifiers with information that indicates that an HIV test or germline genetics is performed, even if the testing result is not included in the data set.

When the research data set is de-identified or anonymized (see Data Classification section for further information), researchers may include HIV or germline genetics data in the research data set without written patient authorization.

CAUTION: The DH Patient Consent form (clinical consent) includes a section for the patient to allow the use of pathology specimens for subsequent research. This broad consent is not sufficient for use of germline genetics data. Patients must provide explicit consent for the use of their germline genetics data for a specific study.

Business Confidential Data: DH policy prohibits the disclosure or use of Business Confidential Data for research purposes, where 'Business Confidential Data' is defined as information stored, processed, or managed by an organization that is sensitive and only accessible to authorized users. Although DH financial, administrative, and contractual data are typically not covered under Human Subjects protection, DH may be legally obligated to ensure data confidentiality, and the disclosure of the data would cause serious risk. Business Confidential Data is considered **Confidential Data (Tier 1)**.

Examples of Business Confidential Data include:

- Claim payment and reimbursement data
- Medical product costs
- Compliance data (e.g., CMS quality data, The Joint Commission results)
- Financial reports and budget information
- Intellectual Property
- Merger/acquisition-sensitive information

Business Confidential Data may, in some situations, be used internally for QI/QA initiatives, however it may never be publicly shared. For research purposes, study teams may consider using proxy values. For example, projects examining cost reduction with a change of practice may consider using utilization rates such as level-of-service or other CPT billing codes as proxy for cost. A researcher may use Medicare charge/reimbursement data as a proxy for institutional cost/reimbursement data.

TIP: In the context of DH research and for completing data request forms, PHI and sensitive data are defined as exclusive data types. PHI is a specific category of information defined by HIPAA that contains elements that can be used to identify an individual. Sensitive data is a broader term that encompasses other types of information that require protection due to privacy concerns and potential harm if disclosed.

How are data sets classified?

For DH research, data sets are classified in one of four levels depending upon the degree to which the data in the data set are identifiable or have the potential for re-identification of the individual study participants.

Fully Identifiable Data: Data are considered identifiable when any of the 18 PHI data types is included in the data set. A data set containing any PHI is considered Confidential Data (Tier 1).

Limited Data: A Limited Data Set (LDS) may contain only particular elements of PHI, specifically dates (e.g., procedure dates, date of birth, date of death); address in the form of city, state, and/or zip code; age in years, months, days or hours. A Limited Data Set contains PHI and is considered Confidential (Tier 1). Sharing a Limited Data Set requires a data use agreement between DH and the recipient.

There are clear advantages to working with a limited data set. A research study should collect a limited (rather than fully identified) data set, when possible, to:

- Comply with the Minimum Necessary Standard of the HIPAA Privacy Rule
- Protect patient confidentiality
- Avoid the need for Accounting of Disclosures
- Simplify the execution of a Data Use Agreement (DUA)
- Ease data storage security requirements and cost
- Reduce the risk and harms of a data breach

De-identified Data: Data are considered de-identified when enough personally identifiable information is removed or obscured so that the data do not identify an individual and there is no reasonable basis to believe that the information can be used to re-identify an individual. De-identified data may retain a re-identification code (a "key") maintained separate from the data set. Although de-identified data are considered Public Data (Tier 3), DH researchers are required to execute a data use agreement before sharing the data outside of the DH Covered Entity. The Master Data Use and Transfer Agreement (MDUTA), a DUA enacted between DH and Dartmouth College to facilitate the sharing of research data between the institutions, may cover data sharing between DH and Dartmouth College for a particular project. The CTO (cto@hitchcock.org) will confirm that the MDUTA is applicable and sufficient.

Anonymized Data: Data are considered anonymized when the data is made untraceable to an individual by removing any information that could be used to re-identify the person. Anonymized data is often stripped of demographic information to definitively prevent the possibility of reidentification. For anonymized data, no re-identification code (a "key") exists. Anonymized data is considered Public Data (Tier 3). At this time, RIDGe approval is required for sharing anonymized data outside of the DH Covered Entity without a DUA.

The Analytics Institute is available to assist with the de-identification or anonymization of research data sets. There is no cost to the researcher. A researcher may request a consult with the Analytics Institute through the 'Data and Informatics' section of the SYNERGY website. Wait time for completion of a data set de-identification will depend upon the queue at the time of submission of the request.

Data Set	DH Data Classification	DUA required
Fully Identifiable	Tier 1	Yes
Limited Data	Tier 1	Yes
De-identified	Tier 3	Yes
Anonymized	Tier 3	No*

Table 1. Summary of data set classification and data use agreement requirement.

* The release of anonymized data without a DUA is under discussion among CTO, ORO leadership, and the Office of the General Counsel. At this time, the release of anonymized data without a DUA in place will require RIDGe review.

Research personnel are responsible for understanding the classification of the research data set and how to appropriately protect the research data.

Data Acquisition

This section describes methods for building research data sets.

Consented Patient Research

Consented Patient Research refers to clinical studies where participants have given their informed consent to take part. These research studies must be approved by the DH IRB. When participant authorization is given, researchers may collect the data (including PHI) specified in the research protocol.

For complete information about Consented Patient Research, contact the DH IRB (IRB@hitchcock.org) and refer to the HRP-103, Investigator Manual.

The conditions for storing and managing research data described in this Handbook apply to data collected with participant authorization.

Obtaining or Creating a Limited Data Set

Best practices in research data management include minimizing identifiable information through de-identification techniques. The approach is often legally required and offers many advantages:

- Patient privacy protection

- Regulatory compliance (e.g., HIPAA)
- Ethical standards (Minimum Necessary Standard)
- Practical advantages, i.e., smaller, cleaner data sets are easier to manage, analyze, share, and less expensive to store securely

Creating a Limited Data Set requires the removal of all PHI elements from the data set, except, as needed, dates (e.g., procedure dates, date of birth, date of death); address in the form of city, state, and/or zip code; age in years, months, days or hours. Removal of PHI can be accomplished using one of the two de-identification methods and approaches offered by the US Department of Health and Human Services (HHS). In summary:

- The "Expert Determination" method provides that a covered entity may determine that health information is not individually identifiable if a person with appropriate knowledge of and expertise with generally accepted statistics and scientific principles and methods for rendering information not individually identifiable determines the data to be de-identified
- Following the "Safe Harbor" method, a data set is determined to be de-identified when all of the 18 personal identifiers of the individual or relatives, employers, or household members of the individual are removed

Two critical steps in the de-identification process are assigning an anonymized study identifier to each participant (replacing the identifying data elements) and obfuscating dates.

Patient Identification: To allow for re-identification, if required, the patient's Medical Record Number (MRN) may be replaced with a unique anonymized, coded study identifier. A crosswalk ("key") between MRN and coded identifier must be stored in a repository that is separate from the data itself (format and physical location). If the crosswalk is destroyed, and the data set does not contain any other PHI, the data set is considered anonymized.

Date Obfuscation or "Shifting": The Analytics Institute Honest Data Brokers may take additional steps to obfuscate the dates included in a Limited Data Set. Date obfuscation is done either by shifting dates or by expressing dates as time from an index or incident date. By consistently shifting dates in terms of time (e.g., number of days) and direction (forward or backward), the dates retain accurate relative time. Likewise, expressing all dates as time from index (e.g., number of days from index date or event) allows all dates to retain relative time between events. Both obfuscation methods preserve the researcher's ability to create timelines within a project. The Analytics Institute Honest Data Brokers have tools for facilitating obfuscation of dates by random time and shift, and Honest Data Brokers retain the information for clarifying the dates if required (dates may be restored to the data set with IRB approval).

The Analytics Institute Honest Data Brokers can provide a Limited Data Sets to researchers. They are experienced with the methods used for de-identification and re-identification of research data and have tools to perform these transformations. There is no cost to the researcher. A researcher may request a consult with the Analytics Institute through the 'Data

and Informatics' section of the SYNERGY website. The wait time for completion of a data set de-identification will depend upon the queue at the time of submission of the request.

Conditions for Disclosure: A researcher does not need to track Accounting of Disclosures for patients included in Limited, De-identified, on Anonymized Data Sets (see 'Accounting of Disclosures' section of this handbook).

A researcher does need a Data Use Agreement (DUA) in place to share a Limited Data Set outside of the DH Covered Entity. Dartmouth College is not part of the DH Covered Entity, so Limited Data Sets require a DUA in place before being shared with Dartmouth College or any investigators affiliated with other organizations. Note that the CTO (cto@hitchcock.org) can confirm that the MDUTA is applicable and sufficient for sharing data with Dartmouth College investigators.

Submission of data to public repositories to comply with federal open science mandates is discussed in the 'Data Sharing for Open Science Requirements' section of this handbook.

Manual Chart Review

Manual Chart Review is a data collection method where researchers systematically extract relevant clinical information from patient medical records (Epic, eDH) to answer targeted research questions.

The HIPAA Minimum Necessary Standard requires that DH limit the use, disclosure, and request of PHI to the minimum amount necessary to achieve the research goal. To meet this standard and provide more efficient access to clinical data for research, the Analytics Institute Honest Data Brokers are available to generate a data set (preferably de-identified or limited) for the research investigator. Manual chart review is required when research data exist only in unstructured formats (e.g., clinical notes, medical images, diagrams) that are inaccessible through automated extraction methods.

Manual Chart Review is allowed when *both* of the following criteria are met:

1. The study team can identify the patient cohort without "fishing" in the medical record
    - The study team has or is prospectively building a list of patients that meet the inclusion/exclusion criteria, e.g., from a patient registry, or
    - The Analytics Institute can provide the list of patient MRNs who meet the inclusion/exclusion criteria
2. Some data elements for the research data set cannot be accessed by the HDBs and must be accessed by Manual Chart Review

An overview of data that the Analytics Institute Honest Data Brokers can provide to researchers is listed in the 'Readily Reportable Data for Research' section of this Handbook.

Who is permitted to conduct Manual Chart Review?

The eDH chart reviewer must be a member of the DH Covered Entity workforce with authorization to access the medical record. The abstractor of the medical record must also be a member of the research study team, and it must be clearly stated in the study protocol that there will be a dedicated abstractor.

Research investigators who do not have eDH access must include a DH researcher in the study team who is able to perform the chart review. DH does not allow Dartmouth College researchers to access DH medical records.

How do researchers obtain permission to conduct Manual Chart Review?

For approval to perform Manual Chart Review, the study team must submit a request for RIDGe review of the proposed activities:

- Navigate to the 'Data and Informatics' page of the SYNERGY website
- Navigate to the 'SYNERGY Informatics Consultation Request Form' by selecting the 'SYNERGY Informatics Consultation Request Form' tab.
- In the Request Form, complete the required fields, selecting 'RIDGe permission for manual chart review' in the field 'What Topic' field
- Upload the Data Request Form and Protocol
- Submit the request

When completing the data request form, clearly specify whether each data element will be obtained through automated systems or manual chart review.

Note that manual chart review may only begin after the study is 'Open to Accrual' in OnCore or is approved by RIDGe as 'preparatory to research' activity.

'Break the Glass' Encounters

A 'Break the Glass' encounter refers to a situation in which a healthcare provider needs emergency access to a patient's electronic health record when they wouldn't normally have permission to view it. This mechanism balances the competing priorities of protecting patient privacy while ensuring that critical care isn't delayed by access restriction during emergencies.

As a rule, researchers are not allowed to access 'Break the Glass' encounters for research purposes (exceptions must be approved by RIDGe). Data generated during the encounter, such as lab results or medication orders, are typically accessible in the medical record from the Analytics Institute Honest Data Brokers or by manual chart review.

The screenshot below is what a user would see when accessing a 'Break the Glass' encounter. The user needs a clinical/business need to access the associated information.

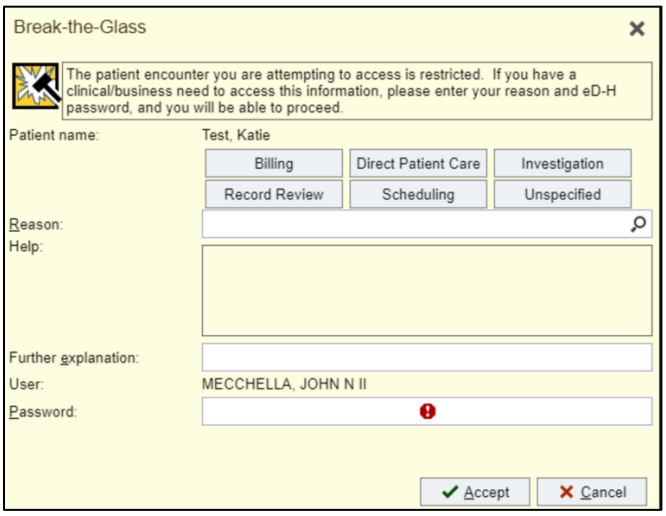


Figure 2. 'Break the Glass' encounter interface in eDH. Only clinical or business needs, documented for the record, permit access to the associated restricted information.

Data Acquired using Deep6 AI

PHI collected from the medical record using Deep6 AI Software is handled with the same protections as any PHI used for research. The use of Deep6 must be included in the study protocol. Participant authorization via informed consent is required if the researcher plans to access any sensitive data using Deep6.

RIDGe approval is required for use of Deep6 to view or use PHI.

Readily Reportable Data for Research

The Analytics Institute includes a team of Honest Data Brokers who can efficiently provide researchers with data from eDH for research purposes. Below is an overview of data elements available via the Analytics Institute; a full list of data points is too voluminous for this document, and data points captured in eDH change regularly. The Analytics Institute is available for discussion of the data needed for a particular research study.

Structured Data

---

1) Patient demographics
    a) MRN
    b) Name
    c) Sex (assigned a birth and gender identity)
    d) DOB
    e) Race
    f) Ethnicity

g) Preferred language
h) Address
i) Hospital Service Location
j) RUCA category (rural/urban location scale)
k) Home phone
l) Cell phone
m) Current primary insurance

2) Visit/Encounter data[1]
   a) Encounter date
   b) Age at encounter
   c) Encounter department
   d) Encounter provider and specialty
   e) Visit type
   f) Telehealth category and provider attestation responses

3) Hospital visit data (Inpatient, Outpatient and Emergency)
   a) Admission date
   b) Discharge date
   c) Length of stay
   d) Admitting department
   e) Discharge department
   f) Patient class

4) Flowsheets
   a) Systolic and diastolic blood pressure, heart rate, height, weight, BMI
   b) Pain score

5) Social history
   a) Smoking use (if completed: start date, quit date, # packs per day, # years)
   b) Smokeless tobacco use
   c) Drug use (status and substance)
   d) Birth control use and partner gender identity

6) Conditions
   a) Diagnosis dates
   b) Encounter/Visit diagnosis (ICD 9/10 code and name)
   c) Diagnosis billed in Professional or Hospital Billing
   d) Primary admit or discharge diagnosis for hospital visit
   e) Problem list and Medical History diagnosis (ICD 9/10, name, and date added to chart)

[1] See 'What data are handled with special consideration' in this handbook regarding data from substance use treatment encounters.

f) DRG codes (inpatient only)

7) Procedures
   a) Patient age at procedure
   b) Procedure dates and time stamps
   c) Procedure provider
   d) Procedure location (if done at DHMC)
   e) CPT Procedure code/description (ICD 9/10 codes for inpatients only)
   f) Operating room case/log data
   g) Anesthesia data
   h) Implant data

8) Episodes of Care
   a) Pregnancy Episodes data
   b) Transitional Care Management Episodes data
   c) Others

9) eDH Registries
   a) Current membership (y/n)
   b) Earliest activation date
   c) Deactivation date
   d) Appointment/Order Question responses
   e) Question ID/name
   f) Patient response

10) Lab results[2]
   a) Component ID/Name
   b) LOINC (Logical Observation Identifiers Names and Codes) – lab test identifier code
   c) Order date
   d) Result date
   e) Result if numeric/short text
   f) Ordering provider/authorizing provider

11) Imaging data
   a) Image code/name
   b) Order date
   c) Interpretation date
   d) Ordering provider
   e) Order-specific question responses
   f) Interpreter name

[2] See "What data are handled with special consideration" in this handbook regarding use of HIV and genetics testing

g) DXA bone mineral density, T-score, and Z-score

12) Medications
    a) Prescribed medication name (brand, generic), dose, instructions, # dispensed, # refills
    b) Medication administration
    c) Medication on reviewed list
    d) Start, stop, refill, and dose adjustment dates
    e) Pharmaceutical class and subclass; Therapeutic class
    f) RxNorm code
    g) Beacon/therapy plans

13) Smart data elements
    a) Smartdata ID/name
    b) Smartdata value (from note, order template, etc.)
    c) Questionnaires
    d) Questions asked
    e) Questions answered
    f) Answers

14) Referrals
    a) Referral order name
    b) Referred to: (internal/external, provider name and address)
    c) Referral order questions
    d) Referring provider name
    e) Referral date and time

Coded Data

Clinical research often includes diagnoses, procedures, and treatments. Much of this information is provided in the form of standardized codes. This coded data is typically available to the Analytics Institute and can be used to build the data set. There are three main types of codes, described below.

ICD codes: Developed by the World Health Organization (WHO), ICD (International Classification of Disease) codes are a standardized system of alphanumeric codes used globally to classify and identify diseases, symptoms, and other findings. Researchers typically collect ICD codes for diagnosis (one or more code assigned to each patient visit) and procedures. ICD codes are also assigned for billing purposes. Note that the transition from ICD-9 to ICD-10 occurred on October 1, 2015, and the code sets are different. ICD codes are available at http://www.icd10data.com/.

CPT codes: Developed by the American Medical Association (AMA), CPT (Current Procedural Terminology) codes are a standardized system of alphanumeric codes used in the United

States to describe medical, surgical, and diagnostic services and to facilitate billing. CPT codes are available at www.aapc.com/codes/cpt-codes-range/.

DRG codes: DRG (Diagnosis-Related Group) codes are a patient classification system that standardizes prospective payment to hospitals for inpatient stays. Each admission is coded with a single DRG code that covers the entirety of the hospital admission. Much of the work of a healthcare coding department is ensuring the right DRG code is assigned to a hospitalization.

A research project may, for instance, use ICD-10-PCS codes for hospital inpatient procedures, a DRG code for inpatient hospitalization, and CPT codes for physician time.

Unstructured Data

The Analytics Institute cannot obtain data that is embedded in text and images. Capturing these data will require Manual Chart Review:

- Unstructured data from notes, e.g., progress notes, admission history and physical, surgical notes
- Data embedded in Radiology reports and images, e.g., tumor size, margin dimensions
- Procedure details, e.g., surgical approach, instruments used
- Clinical outcomes, e.g., medical side effects, procedure complications
- Pathology reports

Data Delivery from Analytics Institute: The Analytics Institute delivers data to the researcher in one of two standard ways:

- Via ShareFile as an email link, where only the recipient (with Hitchcock.org email address) can download what is sent via ShareFile, and the recipient is responsible for handling the data from that point forward
- Via the Granite research platform

Other delivery options may be set up as required for special cases. Please consult the Analytics Institute Honest Data Broker assigned to the project.

For a consult with the Analytics Institute to discuss data for a research study:

- Navigate to the 'Data and Informatics' page of the SYNERGY website
- Navigate to the 'SYNERGY Informatics Consultation Request Form' by selecting the 'SYNERGY Informatics Consultation Request Form' tab.
- In the Request Form, complete the required fields, selecting 'Consult Analytics Institute about data' in the field 'What Topic' field
- Submit the request

Data Available from Across the Dartmouth Health Network

As new organizations join DH, they begin contributing data to the Epic system. The table below includes the DH Clinic and when that clinic data was integrated into the Epic Data Warehouse for Research.

Organization	**Date data begin in Epic (eDH)**
Lebanon, Manchester, Nashua clinical data	04/2011
Lebanon, Manchester, Nashua billing data	10/2015
Cheshire Medical Center and clinics	11/2017
Alice Peck Day Hospital and clinics	05/2019
New London Hospital and clinics	05/2022
Mt. Ascutney Hospital	Projected Fall 2025
Southwest Vermont Medical Center	TBD
Hempstead Hospital	TBD
Valley Regional Hospital	TBD

Table 2. Summary of dates when data from DH clinics became/ will become available for research.

Images for Use in Research

DH research may include radiology images, e.g., plain x-ray, CT, MRI, and ultrasound. Researchers may also use non-radiological images, e.g., pathology slides and wound photographs. Images should be fully de-identified whenever possible; however researchers may use identified images with proper approvals for use of PHI.

Image De-identification: The DH Imaging De-ID team performs the de-identification of all imaging (radiology and non-radiology), using institutionally approved software, and in this process, the image is linked to the coded study participant identifier. As a rule, researchers may not use any other methods for de-identifying radiologic images; any exception must be approved by RIDGe.

Requesting Image Deidentification: When the Analytics Institute generates the research cohort, they will submit the de-identification request, with the coded study participant identifier when relevant, to the DH Imaging De-ID team directly. If the researcher already has the cohort, the study team may submit the de-identification request using Imaging De-ID Request Form.

For questions about imaging de-identification, storage, and viewing, contact the DH Imaging De-ID team at image-deid-team@hitchcock.org.

Qualitative Data

Qualitative data refers to non-numerical information that captures qualities, characteristics, or properties that cannot be easily measured or counted. Qualitative research focuses on describing, understanding, and interpreting phenomena rather that quantifying them. Qualitative

methods include (but are not limited to) patient surveys, focus groups, and semi-structured interviews. Community engagement studios may also be used to collect qualitative data in a research study or may be used in an advisory capacity.

Research participants voluntarily and directly share their data when participating in qualitative research (i.e., interviews). Qualitative research data is not considered PHI; however, the data should be treated as confidential data and securely managed.

Preferred Software for Qualitative research: Dedoose. DH IS Security has approved the use of the cloud-based software application Dedoose for IRB-approved, consented research studies where the patient will be providing HIPAA authorization. It is recommended that the researcher include a statement in the consent, e.g., 'For this study, a tool called Dedoose will be used to assist with analyzing a transcript of your interview. Your data will be maintained in a confidential way.' Dedoose offers a HIPAA-compliant workflow that may be used if required (involves added cost). If Dedoose will not work for the study, RIDGe approval for using 3rd Party Software is required to use the alternative chosen software application.

NVivo, a commonly used alternative software package for analyzing qualitative data, has not been approved by DH IS Security for use on DH studies. Researchers with a compelling need to use NVivo may request approval from RIDGe and will need to notify patients in the consent that NVivo has not been reviewed by DH.

Data for Preparatory to Research Activities

'Preparatory to Research' refers to the use or disclosure of PHI, without removal of the PHI from the DH Covered Entity, for the purpose of determining the scope, feasibility, and/or potential participants for a future research study. Preparatory to Research activities may include:

- Developing a research question or hypothesis
- Preparing a research protocol
- Writing a grant application
- Assessing the feasibility of conducting a research study

For information about contacting patients as part of Preparatory to Research activity, contact the IRB (and refer to the HRP-103, Investigator Manual) for complete information.

How can de-identified data be accessed for Preparatory to Research Activities?

The Analytics Institute: The Analytics Institute Honest Data Brokers can supply aggregate counts, demographic distributions, and other de-identified data. These aggregate statistics summarize grouped data without revealing individual patient information. IRB and RIDGe approvals are not necessary when working with de-identified aggregate data and summary statistics.

It is important for researchers to prepare a detailed description of the cohort and data they are requesting the Analytics Institute to provide. Appendix 1 of the Data Request Form, which includes inclusion and exclusion criteria, provides an effective template. Submit a request to the Analytics Institute by creating a DHSM ticket, entering 'prep 2 research' in the 'IRB #' field:

> DHSM > New Ticket > Data Reporting > Report or Data Extract > Custom Request > select radio button top right 'for research' | Add IRB study ID or 'prep 2 research' | complete form and attach Data Request Form (DRF) or document describing the cohort > Save (to submit the ticket) top left

Wait time for completion of a data set de-identification will depend upon the queue at the time of submission of the request.

Cohort Discovery Tools: DH supports the use of tools for the exploration of de-identified clinical data. These tools allow researchers to independently investigate the data:

- **TriNetX**, a global health research network and real-world data platform that connects healthcare organizations, pharmaceutical companies, and researchers to facilitate clinical research and drug development, with particular focus on clinical trials accrual and opportunities
- **Epic Cosmos**, Epic Systems' de-identified database that aggregates electronic health record data from Epic's network of healthcare organizations, has three access options:
    - SlicerDicer
    - Data Scientist
    - Data Architect

SlicerDicer serves as Epic's user-friendly interface for population-level data visualization and is generally adequate for exploratory analyses. Access to Epic Cosmos requires certification as either a Data Architect or Data Scientist. These certified roles utilize Epic Cosmos through the Data Science Virtual Machine platform, where users can execute queries against SQL-generated data subsets. Data Architects directly query the Cosmos SQL database to construct data marts, and Data Scientists perform subsequent analyses on these subsets using SQL, R, or Python. Certification requires Epic-hosted training at $400/day (one day for the Data Scientists, 2 days for Data Architects). SYNERGY maintains certified Data Architects and Data Scientists available for project collaboration as needed.

To use Epic Cosmos, a researcher must attest to have read the Epic Cosmos Rules of the Road, which will help the user understand publication requirements and project budgeting. This attestation may be completed when a user sets up a Cosmos account, however appropriate resourcing for Epic Cosmos use in the context of a funded research project is essential.

TriNetX and SlicerDicer (for Epic Cosmos exploration) are available at no cost to the DH community, including complimentary training for proficiency with these platforms. Access and training requests can be submitted through the 'Data and Informatics section' of the SYNERGY website.

Can PHI be accessed for Preparatory to Research activities?

DH guidance and the HIPAA Privacy Rule allow researchers to use protected health information (PHI) for preparatory to research activities without IRB approval and without obtaining written authorization (patient consent or waiver) if the researcher is able to demonstrate that:

- RIDGe has approved the use of PHI in the 'preparatory to research' data request
- The use is sought solely to review PHI as necessary to prepare a research protocol or for similar purposes preparatory to research
- The study team will comply with the Minimum Necessary Standard of HIPAA
- The study team will not access sensitive or restricted data
- The researcher will not remove any PHI from the DH Covered Entity (data is accessed only by DH workforce and stored only on DH drives)

To proceed with Preparatory to Research activities, a researcher must request RIDGe review of the activity (includes a completed Data Request Form).

Upon receipt of the RIDGe approval letter, the researcher creates a DHSM ticket to request data, entering 'prep 2 research' in the 'IRB #' field:

> DHSM > New Ticket > Data Reporting > Report or Data Extract > Custom Request > select radio button top right 'for research' | Add IRB study ID or 'prep 2 research' | complete form and attach Data Request Form (DRF) and RIDGe approval letter > Save (to submit the ticket) top left.

Quality Improvement Activities

Quality Improvement (QI) employs systematic, data-driven approaches to enhancing processes, systems, and outcomes. While QI may constitute regulated research, they are classified as human subjects research only when they meet the HHS 45 CFR 46 definition: systematic investigation involving human subjects through direct interaction or identifiable private information collection, designed to develop or contribute to generalizable knowledge (e.g., publication).

Appendix A-3 contains a tool used to determine whether a project is considered QI or Clinical Research with Human Subjects.

When should a QI project be submitted for IRB review?

Researchers should submit QI projects to HRPP for IRB review when:

- The project is intended for peer reviewed publication
- There Is uncertainty about whether the project constitutes QI or Human Subjects Research

If the QI project is submitted to the HRPP for IRB review, it is helpful to use QI terminology when describing the activities:

Research Terminology	Quality improvement (QI) terminology
Study	QI project
Study Coordinator	QI Project Manager
Participants or Subjects	Patients
Hypothesis	Goal
Intervention	Test of change, Plan-Do-Study-Act (PDSA) cycle, or Implementation of best practice
Control	Standard care, Current care
Endpoint	Outcome, metric, measure
Knowledge gap	Performance gap

Table 3. Overview of terminology typically used in research studies vs QI. Appropriate terminology helps to differentiate QI activities from research for institutional reviews.

The IRB will determine whether the activity is Not Research (NR) or Human Subjects Research (HSR). The IRB will provide the determination in writing.
The Clinical Trials Office (CTO) reviews all QI activity.

> *TIP: While RIDGe review is not required for studies that do not meet the definition of research, these studies still require CTO review, which is triggered through appropriate workflows in OnCore.*

For a project that is QI, best practice dictates that a researcher

- Create a QI Project Charter (templates available from the DH Value Institute)
- Obtain the sponsorship from the appropriate Department Chair, Vice Chair, or other leader

For help assessing the QI project:

- Navigate to the 'Data and Informatics' page of the SYNERGY website
- Navigate to the 'SYNERGY Informatics Consultation Request Form' by selecting the 'SYNERGY Informatics Consultation Request Form' tab.
- In the Request Form, complete the required fields, selecting 'Other' in the field 'What Topic' field and adding 'Assistance with QI activities' in the 'Please describe the Consult request in detail' field
- Submit the request

Systems and Platforms for Management of Research Data

Researchers are required to store research data in both a DH IS Security-approved software application and in a DH IS Security-approved location.

What DH systems are available for collecting and managing research data?

For reasons of safety and data integrity, it is preferable for researchers to collect and manage their research data in a database. Two database management systems are available to DH researchers, DH REDCap and EDC. Excel spreadsheets are widely used in research, and although spreadsheets are accessible and can be useful, they have important limitations, especially as data grow in size and complexity. For example, databases provide granular access controls, audit trails and encryption, and Excel files are easily copied, emailed and shared. Databases are safely accessed concurrently by multiple users, with native version control, enabling data consistency. Concurrent use of Excel may result in errors in and/or loss of data. While spreadsheets are allowed, use of a spreadsheet to manage research data is not preferred.

DH REDCap: REDCap ("**R**esearch **E**lectronic **D**ata **Cap**ture") is a web-based, metadata-driven system designed to support clinical and translational research conducted in an academic environment. The DH instance of REDCap is a tool available to DH employees for purposes of quality improvement, clinical operations and research. Accounts are available to Dartmouth College researchers when working with DH protected health information (PHI). Data within REDCap is stored securely; however, it is the responsibility of the Project Owner to manage the access controls within REDCap to ensure data is managed appropriately.

The Dartmouth community supports multiple different instances of REDCap, one managed by DH ("DH REDCap") and others managed by Dartmouth College (collectively known as "DC REDCap"). **Only the DH REDCap instance is approved for use with DH data without additional contracts in place.**

DH REDCap is:

- Approved by DH IS Security for capture and management of DH PHI
- Hosted on the DH technical infrastructure, ensuring that data is regularly and automatically backed up
- Hosted behind the DH firewall, increasing data security

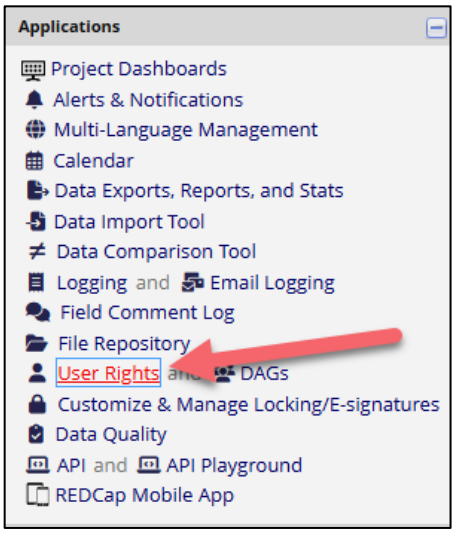
REDCap includes full audit trail logging, recording all operations on the data, including viewing and exporting. The audit log records operation, date and time, and the user performing the operation, permitting review of the audit trail as necessary.

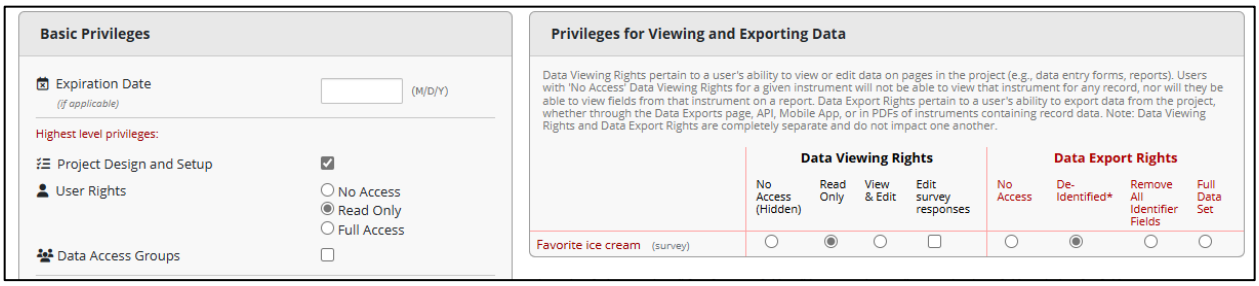
DH REDCap is not 21 CFR part 11 compliant, so it should not be used for data management for FDA-regulated studies when there is an IND or IDE.

For the duration of a DH REDCap project, the Project Owner is responsible for the following access control measures:

1) If the project contains PHI, identifiable/sensitive data fields, those fields are required to be marked as containing identified information.
2) Assignment and active management of the roles and authorizations via the user rights matrix for project members use of specific forms and functions
3) Users are assigned minimum necessary rights and access to PHI based on job requirements, IRB authorization, or data sharing agreements.
    - To decrease the risk of a PHI breach, the Project Owner is required to restrict data export privileges as above. If data export is required for the project, consider the export of de-identified data only.
4) If PHI will be exported, it is the responsibility of the Project Owner to ensure that data will only be exported to secure locations (e.g., DH managed machines, DH Secured Servers, or approved Dartmouth College Research Computing Environment [i.e., Granite]) and not to personal machines.
    - Exporting confidential data to inappropriate locations, such as a personal computer, may constitute a data breach.
5) Ensure collection of the minimum necessary data required for the project
6) Move the project to “Inactive” or “Archive” status once the project is complete and promptly remove or disable user access for persons and entities that no longer need access to DH REDCap.
7) If the project is a Research project:
    - IRB approval is required before storing data in DH REDCap
    - The IRB approved protocol should contain a list of the study team members who will have access to the project

Managing User Rights





DH REDCap Administrators reserve the following rights:

1) Ability to audit any project at any time for review of confidential data and review audit history of exporting, and user access settings.
2) Development of reports for review by the ORO, IRB, IS Security and/or D-HH Privacy Office on the activity and authorized users of specific projects.

The Office of Research Operations (ORO) and SYNERGY provide resources to help with the use of DH REDCap. To begin organizing a project in DH REDCap, request a SYNERGY Informatics consult:

- Navigate to the 'Data and Informatics' page of the SYNERGY website
- Navigate to the 'SYNERGY Informatics Consultation Request Form' by selecting the 'SYNERGY Informatics Consultation Request Form' tab.
- In the Request Form, complete the required fields, selecting 'REDCap' in the field 'What Topic' field
- Submit the request

One of the Research Informatics team will reach out to the researcher to help with REDCap access, training, and configuration. Additionally, Research Informatics offers REDCap Office hours; the office hour schedule is available at the REDCap login page (redcap.hitchcock.org).

CAUTION: Dartmouth College REDCap and Qualtrics are not DH IS Security-approved software applications; therefore, PHI cannot be stored in these databases without prospective patient consent.

EDC: EDC ("**E**lectronic **D**ata **C**apture") is a data capture and management tool that facilitates Investigator Initiated Clinical Trials (IIT). EDC is part of the Advarra suite of tools that includes OnCore. The EDC intranet page (available to DH employees) includes full information. EDC is available to DH research studies; at present external users cannot access EDC.

EDC:

- Is approved by DH IS Security for capture and management of DH PHI
- Is 21 CFR part 11 compliant (required for FDA-regulated studies)

- Is integrated with eDH to support direct population of specific lab values from eDH to EDC
- Is integrated with OnCore to reduce duplicate data entry
- Facilitates electronic PI sign-off for eCRFs
- Is hosted on the DH technical infrastructure, ensuring that data is regularly and automatically backed up
- Supports single-site (DH) studies only

All DH Investigators who personally hold an Investigational New Drug Application (IND) or Investigational Device Exemption (IDE) (i.e. regulatory sponsor) **must** use the EDC to capture and store the data collected for their investigational protocols. Note that Expanded Access Protocols (EAPs) may be an exception to this requirement.

Currently, the use of EDC is limited to interventional clinical trials. Use of EDC is optional for IITs that involve a drug/device therapeutic intervention/diagnostic that is not FDA regulated.

ORO has dedicated staff who will work with researchers to design, build, and test the study database. For help implementing a project in EDC, submit a DHSM ticket:

DHSM > Research Resources > EDC > Calendar build/revision request

> *TIP: Data and process engineering are critical to the success of a research study. Documenting and defining each data element to be collected, and organizing the process (e.g., calendar of events) for the capture and use of each of those data elements, is time consuming. However, the up-front work will ensure that the study will generate a meaningful data set for analysis. This process is important regardless of what software is used for managing research data.*

Where can my research data be safely stored?

For storing DH PHI, a researcher must use DH-approved infrastructure:

- DH H: drive
- DH shared drive (e.g., DH I: drive)
- Local hard drive of a DH managed computer (DH C: drive)
- Granite
- Within DH REDCap (not Dartmouth College REDCap)
- Within DH EDC

DH Investigators are prohibited from storing Confidential (Tier 1) data sets (data sets containing **any** PHI or sensitive data) on a DartFS server and Dartmouth College computers.

Storing research data on the approved infrastructure, listed above, will protect the data and facilitate the project's RIDGe approval.

Data Storage Location	Institutional Network	Approved for Storage of PHI	Approved with Patient Consent	DUA Required	Automatically backed up	Additional Costs to Project
DH H: Drive	DH	Yes	Yes	No	Yes	No
DH I: Drive	DH	Yes	Yes	No	Yes	No
DH C: Drive (DH computer)	DH	Yes	Yes	No	No	No
DH REDCap	DH	Yes	Yes	No	Yes	No
EDC	DH	Yes	Yes	No	Yes	No
Granite	Dartmouth College	Yes	Yes	Yes	Yes	Yes
DC REDCap	Dartmouth College	No	Yes	Yes	Yes	No
Dartmouth Google Drive	Dartmouth College	No	Yes	Yes	Yes	No
DartFS/ Computing	Dartmouth College	No	Yes	Yes	Yes	No

Table 4. Summary of DH data storage locations, with conditions and required approvals.

Granite

Granite is Dartmouth's secure research platform built on tiCrypt software, enabling encrypted storage, access, and analysis of sensitive data. Created to address Dartmouth College researchers' need for secure DH patient identifiable data access, Granite serves as the preferred environment for DH PHI data sets. The platform is jointly supported by DH and Dartmouth College and hosted on Dartmouth College servers.

The security of research data on Granite is the joint responsibility of the PI and the study team.

There is a cost associated with the use of Granite; information for cost planning can be found on the Granite website: dartgo.org/granite-info

The Principal Investigator (PI) of the research study is responsible for overseeing the management of study data on Granite, including approving the movement of data safely from Granite to other platforms.

Can researchers outside of DH and Dartmouth College access Granite?

Investigators from outside institutions may access Granite if they are collaborating with a Dartmouth College or DH study team. The collaborator will need either a Dartmouth College NetID or DH employee ID to be granted access privileges.

Removing De-identified Data from Granite

Aggregate counts and other de-identified data may be removed from Granite without special handling.

The study PI is responsible for de-identifying non-aggregated data, using one of the two de-identification methods and approaches offered by the US Department of Health and Human Services (HHS). In summary:

- The "Expert Determination" method provides that a covered entity may determine that health information is not individually identifiable if a person with appropriate knowledge of and expertise with generally accepted statistics and scientific principles and methods for rendering information not individually identifiable determines the data to be de-identified
- Following the "Safe Harbor" method, a data set is determined to be de-identified when all of the 18 personal identifiers of the individual or relatives, employers, or household members of the individual are removed

To remove de-identified data from Granite, the research study PI should send an email to Granite User Support (granite.support@dartmouth.edu) in the following format:

> Email subject: Granite de-identified data removal users
>
> Email content:
>
> 1) I am the Principal Investigator on the Granite project: *(Granite virtual machine name and Granite project name)*.
> 2) The study IRB # is: *(IRB number)*
> 3) I am giving the following people, who are on the research IRB protocol, permission to remove data from Granite: *(1 or 2 names of study members)*.
> 4) These designated data managers will only remove de-identified data (no PHI) such as aggregate data, counts, tables, figures, or cleaned data rows.
> 5) I understand, as Principal Investigator (PI) that I am responsible for data management in accordance with DH guidelines, research best practices, HIPAA, and the approved IRB protocol. As PI, I am responsible for verifying that the data to be pulled from Granite is fully de-identified (no PHI) following the HHS guidance of "Expert Determination" and/or "Safe Harbor" methods. Once removed from Granite, the data should be stored/shared only in accordance with the approved IRB protocol.

Removing Identifiable Data from Granite

A study PI may move identifiable data (PHI) from Granite to a DH IS Security-approved storage space (as described above, DH H: drive, DH I: drive, DH C: drive). The study team member transferring the data must be a member of the DH workforce.

Identifiable data being transferred from Granite to DH must be encrypted in transmission, with a minimum of 128-bit encryption. Instructions for encrypting data in transport are available in the Granite User Guide: 'Transferring OUT sensitive data.'

To remove identifiable data from Granite, the research study PI should send an email to Granite User Support (granite.support@dartmouth.edu) in the following format:

> Email subject: Granite PHI data removal users
>
> Email content:
> 1) I am the Principal Investigator on the Granite project: *(Granite virtual machine name and Granite project name)*.
> 2) The study IRB # is: *(IRB number)*
> 3) I am giving the following people, who are on the research IRB protocol and are DH workforce, permission to remove PHI data from Granite: *(1 or 2 names of study members)*.
> 4) The data will be saved to a DH managed computer on the I: drive, H: drive, or C: drive.
> 5) I understand, as Principal Investigator (PI) that I am responsible for data management in accordance with DH guidelines, research best practices, HIPAA, and the approved IRB protocol. Once pulled from Granite, the data should be stored/shared only in accordance with the approved IRB protocol.

Methods for safe transfer of DH research data

The safe methods for the transfer of data files is governed by the DH 'File Transfer Policy' (Policy ID# 4640) that outlines the requirements for transmission of files containing Organizational Data to ensure its security and privacy. Organizational Data is defined as all material, in any form, related to the operation of DH, including, but not limited to health information, genetic information, financial information, employee information, proprietary products and product development, marketing and general business strategies, and any information marked 'confidential.'

The ORO recommends two safe options for transferring DH research data from Granite (Dartmouth College network) into the DH network to a secure, approved DH location:

- ShareFile
- SFTP

ShareFile is the DH platform for securely uploading, storing, and sharing files. It is appropriate for PHI, PII, and confidential business information. A user may share files and folders with users within and outside of the DH network. ShareFile information is available on the DH intranet: ShareFile.

Secure File Transfer Protocol (SFTP) is a network protocol that provides file transfer over a secure channel (such as Secure Shell Protocol, or SSH). File transfer services are managed by

the DH IS Application Team, however the ORO Research Informatics team can assist a PI with SFTP set-up:

- Navigate to the 'Data and Informatics' page of the SYNERGY website
- Navigate to the 'SYNERGY Informatics Consultation Request Form' by selecting the 'SYNERGY Informatics Consultation Request Form' tab.
- In the Request Form, complete the required fields, selecting 'Other' in the field 'What Topic' field and include 'SFTP set-up' in the 'Please describe the Consult request in detail' field.
- Submit the request

For information on encrypting files and/or folders in transit, see the Granite User Guide: 'Transferring OUT sensitive data.'

Software for Research

The mission of DH Information Systems (IS) Security includes ensuring the confidentiality, integrity, and availability of data and data systems within DH. DH IS Security must balance innovation with the paramount concerns of patient safety, data security, and operation reliability in an environment where failures can have life-or-death consequences.

Healthcare is particularly targeted by cyberattacks due to the value of patient data and the critical nature of services. DH is no exception. Ransomware attacks on healthcare have increased drastically. DH scrutinizes new applications intensely for several critical reasons, including (but not limited to):

- Patient safety concerns: software bugs or malfunctions could directly impact clinical decision-making or patient care
- Data security requirements: HIPAA, state, and other federal regulations mandate high security, and consequences of data breach are severe
- System interoperability challenges: new software must integrate and not interrupt existing workflows

In the context of Dartmouth Health, 'approved software' refers to computer programs and applications that have been reviewed and approved by DH IS Security team. Approved software can be installed on a DH-managed computer and can be used to manage or process DH PHI. In contrast, 3rd Party Software is defined as a computer program that is not a direct part of the DH business system and/or has not been reviewed and approved by DH IS Security for use on DH-managed computers with DH PHI.

Examples of typical 3rd Party Software include:

- Large Language Models (LLMs) for Generative AI, for instance OpenAI's GPT series, Google's Gemini models, Anthropic's Claude models, and Meta's Llama series

- Plug-in modules for DH IS Security-approved applications, for instance statistical packages for Excel like QI Macro
- Smartphone applications (e.g., Signal messaging app)
- Web- or cloud-based applications or software (e.g., Google Sheets, Medidata Rave, NVivo)
- Custom software created by researchers at Dartmouth or other institutions
- Statistical software, e.g., SAS, R, ArcGIS

Introducing new software for handling DH data into the system infrastructure and/or network (e.g., software with Application Program Interfaces (APIs) that interact with eDH or other DH systems, or software that manages DH Confidential Data) may introduce security vulnerabilities that increase risk of data security and operational reliability. Importantly, resource limitations and mission-critical use of those resources limit the time DH IS Security can devote to vetting new 3rd Party Software for use with DH data. The high level of scrutiny of new 3rd Party Software, described in New Technology Acquisitions (accessible to DH employees via the DH intranet), which includes threat analysis by an external consulting firm, can be frustrating to researchers, however it reflects the responsibility and constraints of a healthcare environment.

DH maintains an extensive ecosystem of approved software for clinical and research use. Software packages that are approved for use on a DH-managed computer with DH PHI include (but are not limited to):

- Microsoft Suite (including Access, Project, Visio)
- Statistical software: SPSS, SAS, R, Stata
- Python
- Matlab
- tiCrypt (Granite)
- REDCap (DH installation)
- EDC (Advarra's Electronic Data Capture)
- Dedoose (qualitative and mixed methods research software package)

Using software applications that are already approved by DH is highly recommended.

Using 3rd Party Software for research

In practice, the best way to facilitate research that may involve 3rd Party software that has not been approved by DH IS Security is to obtain patient consent for the use of the 3rd Party Software. DH IRB can offer language in consent templates,

In rare cases, DH will review and approve 3rd Party systems for researchers. The process is intensive:

- DH will need to execute a Business Associate Agreement (BAA) with the vendor of the 3rd Party Software. This step involves contracting and legal departments.

- DH IS Security will need to conduct a cybersecurity review, including hiring an outside consulting firm to conduct background checks on the vendor.
- Due to large number of requests for new software, the time and high cost of the three-tier cybersecurity evaluation and legal review, and competing high-priority institutional requests, this process requires at least 6 months. In some cases, DH is not able to review or approve requests.
- If you are interested in using an unapproved 3rd Party system, please contact the ORO Research Informatics team to discuss.

For industry-sponsored studies, the sponsor manages all software systems used for the study. While many industry-sponsored studies utilize 3rd party software, the CTO provides direct oversight, and RIDGe review is not required for these studies.

For qualitative and other research that does not involve DH PHI, a researcher may be approved to run the software outside of the DH IS infrastructure. Consult with ORO Research Informatics to confirm.

ORO Research Informatics team can assist with identifying options for a research study:

- Navigate to the 'Data and Informatics' page of the SYNERGY website
- Navigate to the 'SYNERGY Informatics Consultation Request Form' by selecting the 'SYNERGY Informatics Consultation Request Form' tab.
- In the Request Form, complete the required fields, selecting 'RIDGe permission for software or apps'
- Submit the request

Data Use – Data Disclosure – Data Sharing

A researcher is responsible for understanding what is needed when handling DH data for research. Special considerations are required for handling Confidential (Tier 1) data.

Proper use of research data containing PHI requires:

- All data users are members of the DH Workforce
- Data is hosted within the DH Firewall, i.e., on a DH server (H: drive, I: drive), a DH managed computer (C: drive), or within DH REDCap or EDC

Dartmouth College systems are not inside the DH firewall. Granite and the Dartmouth College instance of REDCap (DC REDCap) are not inside the DH firewall.

Disclosure of research data containing PHI

As defined by the DH 'Release of Protected Health Information' Policy (Policy #12977), Data Disclosure refers to any release, transfer, or sharing in any manner (including oral) of PHI outside of DH.

PHI is disclosed when it is:

- Shared with individuals who are not DH workforce, i.e., outside of the DH Covered Entity, without appropriate consent and/or contracts
- Sent to or stored on computers (including email servers) outside of the DH network firewall

DH Workforce does not include students from Dartmouth College or any other institution. The one exception is Geisel MD students who are considered DH workforce when:

1. **The Geisel MD student is working under the direct guidance of a mentor who is both DH workforce and Geisel faculty, *and***
2. **The Geisel MD student and the faculty mentor are documented study team members in the IRB protocol**

When these conditions are met, sharing data with the Geisel MD student is not considered a disclosure of research data.

Again, students are not considered members of the DH Workforce and may not access and use DH PHI for research. Examples include:

- A Geisel MD student working on a research study without a DH PI listed on the DH IRB protocol
- A Geisel student who is not in the MD program (e.g., PhD, MSc, MPH)
- A Dartmouth student, undergraduate or graduate (other than the Geisel MD program)
- A student from any other academic institution

CAUTION: For research involving PHI, a student using a 'hitchcock.org' email address is not automatically a member of the DH Workforce.

When can PHI for research be disclosed without obtaining written authorization?

The HRP-103, Investigator Manual is available to guide researchers through regulations, policies and procedures related to the conduct of Human Research. A thorough discussion of considerations around patient consent and waiver of HIPAA authorization are available in the Investigator Manual, and researchers are encouraged to refer to it. This section of the Handbook provides select information pertinent to informatics-focused concerns.

The HIPAA Privacy Rule permits covered entities to use or disclose PHI for research purposes when an individual authorizes the use or disclosure of information about him or herself in writing. The individual may also consent to the use of 3rd Party Software to manage his or her

data when information about the 3rd Party Software and its intended use are explicitly included in the written authorization.

There are four situations where a health care covered entity may use or disclose protected health information for research purposes without obtaining written authorizations from individuals:

1. IRB waivers
2. Reviews preparatory to research
3. Research on a decedent's information
4. Limited, De-identified, and Anonymized Data Sets

While all DH research involving protected health information requires DH IRB review, only the first situation involves specific IRB action and findings. In other situations, administrative units within the covered entity enforce the governing regulations.

> *TIP: A researcher can use TriNetX without IRB approval or waiver of authorization. TriNetX is a web-based platform for exploring de-identified data shared by dozens of organizations. TriNetX returns queries with de-identified, aggregate counts, not patient-level data.*

Use of Data from DH Patients who have Died (Decedents)

Refer to the HRP-103, Investigator Manual for complete information.

In summary, when accessing and using data from DH patients who have died (decedents):

- Written authorization from the decedent's family is not required
- A HIPAA Waiver of Authorization is not required
- An Accounting of Disclosure must be provided if the patient died less than 50 years from the date of data query and PHI will be disclosed to individuals or organizations outside of the DH Covered Entity

The IRB protocol should indicate the reason for using decedent data including why the PHI is necessary for the research.

Unauthorized Disclosure of PHI – Data Breach

An Unauthorized Disclosure of PHI occurs when data is shared outside of the DH Covered Entity without a HIPAA Authorization, the IRB approval of a waiver of authorization, or executed Data Use Agreement (DUA) in place.
Example scenarios of Unauthorized Disclosure:

- DH IRB approves research data containing PHI to be stored in DH REDCap, and the study team stores the research data in Dartmouth College REDCap
- A university student (not authorized Geisel MD) is added to the study team and accesses research data containing PHI without a DUA in place

- A data file containing PHI is downloaded from Granite (PI did not de-identify the data prior to download), and the file is shared with collaborators at an outside institution without a DUA in place

A Data Breach is an Unauthorized Disclosure of PHI for which the DH Privacy office is not able, after risk assessment, to demonstrate a low probability that the confidentiality of the PHI was compromised. A Data Breach is a serious incident that has reporting requirements mandated by the state and federal governments.

If a study team suspects that an Unauthorized Disclosure or Data Breach has occurred, the research study Principal Investigator must immediately:

- Complete a Privacy Incident Report (online)
- Notify DH Privacy Office (privacyoffice@hitchcock.org)

Accounting of Disclosures

Under HIPAA, an Accounting of Disclosures (AoD) refers to a written record of when a Covered Entity has shared a patient's PHI with someone other than the patient themselves. The accounting must be provided to the patient upon request. The AoD includes details like the date of disclosure, the recipient, and the reason for disclosure.

An Accounting of Disclosures is required for:

- Disclosures made without patient authorization and without an IRB waiver of authorization
- Disclosures to researchers who do not have proper authorization or waiver documentation
- Disclosure of 'preparatory to research' data to any individual or entity outside of the DH covered entity
- Disclosure of decedent information if the patient died less than 50 years from the date of data query and PHI has been disclosed to individuals or organizations outside of the DH Covered Entity

An Accounting of Disclosures is not required for

- Disclosures made with valid patient authorization for research
- Disclosures made under an IRB-approved waiver of authorization that meets HIPAA requirements
- Disclosures that are part of a Limited Data Set with a Data Use Agreement
- Disclosures that are part of a De-identified or Anonymized data set
- Disclosures made to the individuals themselves
- Disclosures for treatment, payment, or health care operations

The Analytics Institute Honest Data Brokers track all disclosures of PHI for research. If the Analytics Institute provides the researcher with a complete data set, the Honest Data Brokers will complete the AoD.

For help determining if the research requires AoD, request a SYNERGY Informatics consult:

- Navigate to the 'Data and Informatics' page of the SYNERGY website
- Navigate to the 'SYNERGY Informatics Consultation Request Form' by selecting the 'SYNERGY Informatics Consultation Request Form' tab.
- In the Request Form, complete the required fields, selecting 'Other' in the field 'What Topic' field
- In the 'Please describe your Consult request in detail,' enter 'AoD consult'
- Submit the request

How is Accounting of Disclosures documented?

If the Analytics Institute is not providing the complete data set (for instance the research team obtains PHI by running reports, performing manual chart review, using clinic registries, etc.), and the research meets the requirements for needing an Accounting of Disclosure listed in the prior section, the Principal Investigator is responsible for completing the Accounting of Disclosures.

To document Accounting of Disclosures:

- Populate the Accounting of Disclosures Excel Template with information describing the research study, PHI disclosed, and all Medical Record Numbers (MRN).
- Save the Excel File with the following naming convention: Name of PI_AoD_Year-Month-Date
    - For example: Burdick_AoD_2023-06-19
    - If the MRN listing need to be updated, coordinate with the Analytics Institute staff to update the listing and ensure the Accounting of Disclosures is complete
- Use ShareFile to securely send the AoD document to Analytics Institute staff at: AI_ResearchData_Information@hitchcock.org

  TIP: There is a ShareFile plug-in for DH Microsoft Outlook that makes transferring files via ShareFile easy. A researcher can request the plug-in by submitting a DHSM Help Desk ticket.

Data Use Agreement

A Data Use Agreement (DUA) is a legal data sharing contract. One organization agrees to share (disclose) data, and the recipient organization(s) agree to terms for managing and safeguarding the data. The terms often include description of the data elements to be shared, methods for data transfer, requirements for secure data storage and data destruction,

restrictions on forward sharing of the data, and responsibilities in the event of a data breach. A DUA must be executed before any data is shared.

When is a DUA needed?

A DUA is required for sharing DH data outside of the DH Covered Entity. In some cases, anonymized data may be shared without a DUA following RIDGe approval. The Clinical Trials Office (CTO) DUA information page on the DH intranet provides clear definitions, guidance, and PI responsibilities for the DUA process.

What is the Master Data Use and Transfer Agreement (MDUTA)?

The MDUTA is a contract executed to facilitate research data sharing between DH and Dartmouth College. It establishes data sharing expectations and eliminates the need for individual DUA negotiations for each project. The Clinical Trials Office (CTO@hitchcock.org) confirms MDUTA sufficiency and applicability for each research project.

How can a researcher obtain a DUA?

The Clinical Trials Office (CTO) Contracting Team facilitates the review and approval of DUAs. Begin the process by reaching out to the CTO via CTO Request Intake Form or emailing CTO (cto@hitchcock.org).

Data Sharing for Open Science Requirements

Federal open science mandates are policy requirements issued by U.S. government agencies that require federally funded research to make data, publications, and other research outputs publicly accessible. Key mandates include the NIH Data Management and Sharing Policy, Office of Science and Technology Policy Memorandum on Public Access, NSF Public Access Initiative, and the FDA Amendments Act requiring clinical trial data transparency. These mandates represent a significant shift toward open science, driven by goals of maximizing the return on public investment in research, accelerating scientific discovery, and improving research reproducibility. They require researchers to plan for data sharing from the study design phase and have implications for consent process, data de-identification, and repository selection. The mandates vary in their specific requirements.

In general, data sets submitted to NIH repositories and those required to be publicly accessible under open science guidelines must be de-identified or anonymized, however the requirements vary by data type and repository. They key principle for all repositories is that publicly accessible research data should not pose unreasonable risks of de-identification while still maintaining scientific utility for secondary research purposes.

To comply with publication requirements, research data may also be shared using Dartmouth College Library Research Data Services, which offers mechanisms for making data sets public, including Dataverse. Dataverse is Dartmouth's institutional repository for depositing and sharing research data.

The Research Informatics team is available for consultation for researchers concerned about how to safely share data in public repositories.

Summary of considerations for working with Dartmouth College colleagues, software, and infrastructure

Because Dartmouth College is not included in the DH Covered Entity, there are special considerations for DH researchers working with Dartmouth College colleagues and resources.

DH Workforce: Under HIPAA regulations (specifically 45 CFR part 160.103), DH Workforce is defined as employees, volunteers, trainees, and other persons whose conduct, in performance of work for the DH Covered Entity, is under the direct control of the DH Covered Entity, whether or not they are paid by the covered entity or business associate.

Dartmouth College employees, Dartmouth students (undergraduate, graduate), Geisel students who are not MD students are not considered members of the Dartmouth workforce. Additionally, investigators and staff who have a dual affiliation (DH and Dartmouth College) but are working under their Dartmouth College affiliation for research are not considered members of the Dartmouth Health workforce.

Data Classification: Data that is classified as 'Confidential – Tier 1' by the DH Data Classification policy (e.g., PHI, business confidential data) are not accessible by Dartmouth College investigators and study team members. Disclosure of Confidential – Tier 1 data to Dartmouth College employees requires an Accounting of Disclosures. DH PHI may be shared with Dartmouth College researchers, with the appropriate Data Use Agreement (e.g., the MDUTA) and with IRB and CTO approval.

Data Set Classification: Tier 1, 2, and 3 data sets all require a DUA for sharing with any researchers who are not members of the DH workforce (the MDTUA may be sufficient for data sharing with Dartmouth College researchers; contact the DH CTO, CTO@hitchcock.org, for confirmation). Only anonymized data sets may be shared without a Data Use Agreement in place.

Platforms and Systems: Generally, Dartmouth College data management applications, e.g., Dartmouth College REDCap, and storage infrastructure, e.g., DartFS and computing clusters,

are not approved for storage of DH PHI. One exception is Granite, which is co-owned and supported by DH and Dartmouth College and serves as the preferred secure environment for computing with DH PHI. Granite is hosted on the Dartmouth College infrastructure, so a DUA is required (MDUTA may apply).
If study participants consent to have their PHI managed on Dartmouth College machines, or if there are appropriate contracts and approvals in place (DUA, IRB and RIDGe approval), DH data may be managed on the Dartmouth College infrastructure as designated by the contracts and approvals.

Data Use Agreement: The MDUTA is a contract executed to facilitate research data sharing between DH and Dartmouth College. It establishes data sharing expectations and eliminates the need for individual DUA negotiations for each project. The Clinical Trials Office (CTO@hitchcock.org) confirms MDUTA sufficiency and applicability for each research project.

Appendices

Appendix A-1 Glossary

Term	Definition
3rd Party Software	A computer program that is not a direct part of the DH business system and/or has not been reviewed and approved by DH IS Security for use on DH-managed computers with DH PHI.
Accounting of Disclosures	Under HIPAA, an Accounting of Disclosures (AoD) refers to a written record of when a Covered Entity has disclosed a patient's PHI with someone outside of the Covered Entity.
Anonymized Data Set	A data set for which the data has been made untraceable to an individual by the removal of any information that could be used to re-identify the person. The data may not contain any of the 18 PHI elements. Any coded 'key' that could be used to re-link the data to the patient has been destroyed.
Business Confidential Data	Data that are not limited by HIPAA and are not 'clinically sensitive data' but are restricted by contracts, BAAs or DH institutional priorities. This includes certain safety metrics, clinical outcomes, patient satisfaction surveys, and financial data.
Consented Patient Research	Clinical or medical studies where participants have given their informed consent to take part. Consents may be written or verbal depending on IRB requirements.
Data Breach	An unauthorized disclosure of PHI, in which data is shared outside of the DH Covered Entity without IRB approval, executed Data Use Agreement (DUA), or another contract (e.g., BAA) in place.
Data Classification	Data classification is the process of organizing data into categories based on the level of sensitivity, value and criticality to the institution.
Data Disclosure	Data Disclosure refers to any release, transfer, or sharing in any manner (including oral) of PHI outside of the DH

	Covered Entity. This includes a person outside of the DH workforce seeing or accessing the data, and data being stored on a computer that is not in the DH information system.
Data Request Form	A document that captures metadata about a study, e.g., Principal Investigator (PI), cohort, data acquisition activity, and a listing of the specific data elements to be collected for the research.
De-identified Data Set	A data set for which enough personally identifiable information has been removed or obscured so that the data do not identify an individual and there is no reasonable basis to believe that the information can be used to identify an individual. Generally, this involves removing all of the 18 PHI elements.
DH Workforce	DH Workforce includes employees, volunteers, trainees, and other persons whose conduct, in performance of work for a covered entity, is under the direct control of the DH Covered Entity, whether or not they are paid by the covered entity. Geisel MD students mentored by members of the DH workforce may be considered DH workforce. No other students are considered DH workforce.
Expanded Access Protocol	Sometimes called “compassionate use,” expanded access is a potential pathway for a patient with a serious or immediately life-threatening disease or condition to gain access to an investigational medical product for treatment outside of clinical trials when no comparable or satisfactory alternative therapy options are available.
Granite	Granite, Dartmouth’s installation of tiCrypt software, is a secure research environment that allows researchers to access and analyze sensitive data on an encrypted platform. Granite is the preferred platform for working with DH PHI.
Identifiable Data Set	A data set containing one or more of the 18 PHI data types.
Investigator Initiated Trials (IIT)	Clinical research studies developed and led by individual researchers or academic institutions based on their own scientific interests.

Limited Data Set	A data set that may contain only particular elements of PHI, specifically dates (e.g., procedure dates, date of birth, date of death); address in the form of city, state, and/or zip code (not street level); age in years, months, days or hours.
Minimum Necessary	The HIPAA Minimum Necessary Standard requires that the use, disclosure, and request of PHI be limited to the minimum amount necessary to achieve the research goal.
Protected Health Information (PHI)	Protected Health Information (PHI) is any information in the medical record or designated record set that can be used to identify an individual and that was created, used, or disclosed in the course of providing a healthcare service such as diagnosis or treatment. There are 18 PHI data types.
Preparatory to Research	The use of data for the purpose of determining the scope, feasibility, and/or potential participants for a future research study. With 'preparatory to research,' PHI may not be removed from or shared outside of the DH Covered Entity.
Quality Improvement Activities	Quality Improvement (QI) research is a systematic, data-driven approach to enhancing processes, systems, and outcomes within a healthcare organization. The DH IRB determines whether or not QI is regulated research.
RIDGe	The DH Research Informatics and Data Governance group (RIDGe) brings together expertise from across the DH research and informatics community to expedite the activation of DH investigator-initiated research projects while ensuring DH data are appropriately handled.
Sensitive Data	Clinically sensitive data refers to information that requires special protection, beyond HIPAA, due to its potential to cause harm, embarrassment, or discrimination if improperly handled or disclosed. Examples include germline genetics, HIV status, and data associated with a participant's treatment in a substance use disorder clinic. PHI does not carry the "clinically sensitive" designation.
SFTP	Secure File Transfer Protocol (SFTP) is a network protocol that provides file transfer over a secure channel (such as Secure Shell Protocol, or SSH).

Sponsor Initiated Trials (SIT)	Clinical research studies where an external organization (typically a pharmaceutical company, medical device manufacturer, or other commercial entity) takes on the role of the “sponsor” who initiates, manages and/or funds the trial.

Appendix A-2 Acronyms and Initialisms

AI	Analytics Institute
AoD	Accounting of Disclosures
CTO	Clinical Trials Office
DRF	Data Request Form
DUA	Data Use Agreement
EAP	Expanded Access Protocol
EDC	Electronic Data Capture
HHS	Health and Human Services
HSR	Human Subjects Research
IDE	Investigational Device Exemption
IND	Investigational New Drug
IIT	Investigator-initiated Trial
IRB	Institutional Review Board
LDS	Limited Data Set
MDUTA	Master Data Use and Transfer Agreement
MRN	Medical Record Number
NHSR	Not Human Subjects Research
PHI	Protected Health Information
PI	Principal Investigator
PII	Personally Identifiable Information
RIDGe	Research Informatics and Data Governance group
SFTP	Secure File Transfer Protocol

SIT	Sponsor-initiated Trial
SPA	Sponsored Project Administration

Appendix A-3 Instrument to Assess the Difference between QI and Clinical Research with Human Subjects

from Ogrinc G, Nelson WA, Adams SM, O'Hara AE. An instrument to differentiate between clinical research and quality improvement. IRB. 2013 Sep-Oct;35(5):1-8. PMID: 24350502.)

Instrument to Assess the Differences between QI and Clinical Research with Human Subjects

This table is intended to compare and contrast the general characteristics of quality improvement (QI) and clinical research activities and is for use by Institutional Review Boards (IRBs), administrative reviewers, investigators, and improvers. This table is intended to guide discussion among these individuals and is not intended to supplant the judgment of IRBs or local QI ethics review committees. Please start by considering these overarching questions:

1. Will the activities of this project occur within the standard of care? If NO, then proceed to IRB review.
2. Is there risk? If YES, use chart below to determine whether this project requires QI review or IRB review.
3. Is this project primarily intended to generate generalizable knowledge? If YES, proceed to IRB review.
4. Does this project involve vulnerable populations? If YES, use chart below to determine whether this project requires QI review or IRB review.

For each item, choose the column to which the project most closely relates—QI or research. You may only choose one answer. Leave the item blank if neither choice applies.

Attribute	*Quality Improvement*	*Clinical Research with Human Subjects*
Intent and Background	☐ Describes the nature and severity of a specific local performance gap	☐ Identifies a specific deficit in scientific knowledge from the literature
	☐ Focus is to improve a specific aspect of health or health care delivery that is currently NOT consistently and appropriately being implemented at this site	☐ Proposes to address or identify specific hypotheses in order to develop new knowledge or advance existing knowledge
Methods	☐ Mechanisms of the intervention are expected to change over time (i.e., an iterative activity) in response to ongoing feedback	☐ Specific protocol defines the intervention, interaction, and use of collected data and tissues, plus project may rely on the randomization of individuals to enhance confidence in differences
	☐ Plan for intervention and analysis includes an assessment of the system (i.e., process flow diagram, fishbone, etc)	☐ May use qualitative or quantitative methods to make observations, make comparisons between groups, or generate hypotheses
	☐ Statistical methods evaluate system level processes and outcomes over time with statistical process control or other methods	☐ Statistical methods primarily compare differences between groups or correlate observed differences with a known health condition
Intended Benefit	☐ Intervention would be considered within the usual clinician-patient therapeutic relationship	☐ Intervention, interaction, or use of identifiable private information occurs outside of the usual clinician-patient therapeutic relationship
	☐ Direct benefit to participants is indicated (e.g., for the decrease in risk by receiving a vaccination or by creating a safer institutional system)	☐ Direct benefit to each individual participant or for the institution is not typically the intent or is not certain
	☐ Potential local institutional benefit is specified (e.g., increased efficiency or decreased cost)	☐ Potential societal benefit in developing new or advancing existing generalizable knowledge