THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES

Center for Research Informatics

CLINICAL RESEARCH DATA WAREHOUSE POLICIES AND WORKFLOWS

last updated April 30, 2015

1. TYPES OF DATA REQUESTS

Use of the CRDW is governed by an IRB-approved protocol describing the standards, governance, and oversight of the design, maintenance, and use of the CRDW. This protocol allows data from the CRDW to be provided for the following activities:

1. Requests that are preparatory for research

This is sometimes described as "cohort identification" and is the process by which an investigator assesses the number of subjects in the CRDW meeting given criteria to determine whether planned research is feasible.

2. Requests for de-identified data

An investigator may request de-identified data for a cohort, not including any protected health information. Since the data set is de-identified, it will not be possible to re-identify the data and therefore it will not be possible to receive data updates or supplement the data set with additional data elements at a later date.

3. Requests for coded data

This is similar to 2, but in this case the investigator may request the data for a cohort in a coded, rather than de-identified, manner. The data request cannot include any protected health information. The key to enable linkage of the coded data set with identifying information will remain with the CRDW Operations Staff. The investigator may not request and will not ever receive re-identification of the data set. Because the data are coded with an anonymous ID, updated data could be provided on an ongoing basis for this type of request.

4. Requests for a limited data set

This is similar to 3, but may include specific dates, date ranges, and zip codes as defined by HIPAA regulations. The data request could be a snapshot of what is available at the time of request or updated data could be provided on an ongoing basis. The investigator may not request and will not ever receive additional subject identifiers.

5. Subject identification for possible participation in a specific research project

For this purpose, individual patient contact information is provided to the investigator so that subjects can be approached for their interest in participating in a clinical or translational research project.

6. Data requests for an investigator-defined identified cohort

The investigator defines a group of identified human subjects for which a specific set of data is requested. The identified cohort could be individuals that have provided consent in the context of an approved IRB protocol, individuals identified through one of the above processes, or individuals identified through the investigator's clinical activities. The data request could be a snapshot of what is available at the time of request or updated data could be provided on an ongoing basis.

2. APPROVALS AND REQUIREMENTS FOR DATA REQUESTS

The below table is a summary of approvals and other requirements that must be met for each type of data request before CRDW data can be delivered.

	Data Request Type					
Requirements to receive CRDW data	**Preparatory for research**	**De-Identified data**	**Coded data**	**Limited data set**	**Contact information**	**Identified data**
Accept Data Use Agreement	✓	✓	✓	✓	✓	✓
Obtain IRB approval			*	✓	✓	✓
Obtain individual consent			†	†	†	✓
Obtain Data Use Committee approval		✓	✓	✓	✓	✓
Data received in Secure Data Zone			✓	✓	✓	✓
Access monitored by Data Use Committee	✓	✓	✓	✓	✓	✓

*Coded data requests require IRB approval if the cohort contains fewer than 50 children, contains fewer than 50 subjects total, or contains HIV or AIDS related data.

†The specific requirements for subject identification are the purview of the IRB Committee, who may place restrictions, such as notification of primary clinician, on the process. Individual written informed consent will be required to receive HIV or AIDS related data.

3. SUMMARY OF DATA ELEMENTS AVAILABLE

The below table is a summary of data elements available for each data request type.

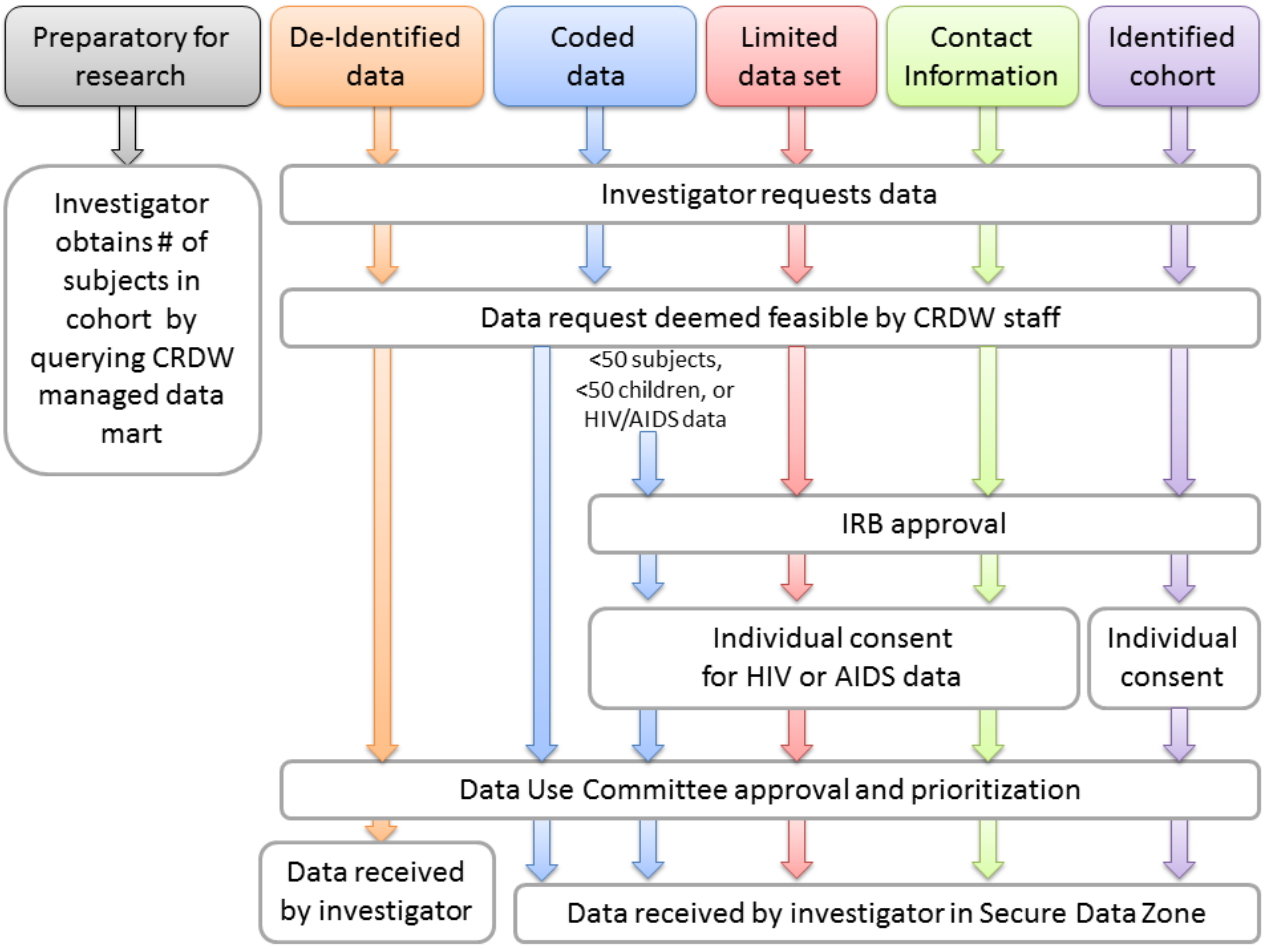
Available data elements	Data Request Type					
	Preparatory for research	De-identified	Coded data	Limited data set	Contact information	Identified data
• Cohort size	✓	✓	✓	✓	✓	✓
• Any data excluding identifiers, e.g., - clinical data - lab values - reports - age* - date-shifted dates - images - diagnosis - procedures		✓	✓	✓	✓	✓
• Anonymous ID			✓	✓	✓	
• Dates such as birth, death, admission, discharge, and service dates • Geographic information such as zip code, but excluding street address				✓	✓	✓
• Name • Street address • Phone and fax #s					✓	✓
• Social security numbers • Medical record numbers • Health plan beneficiary #s • Account numbers • Certificate/license numbers • Vehicle identifiers and serial #s, including license plate numbers • Device identifiers and serial #s • Web universal resource locators (URLs). • Internet protocol (IP) address #s • Biometric identifiers, including fingerprints and voiceprints • Full-face photographic images and any comparable images • Any other unique identifying #, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification.						✓
Example uses of data set	Feasibility of study	Chart review, retrospective only	Chart review, retrospective and prospective	Chart reviews (retro and pro) that require certain identifiers such as zip code	Recruitment for clinical study, clinical registries	Studies on an identified cohort, clinical trial

* ages over 89 will be aggregated into a single category of age 90 or older.

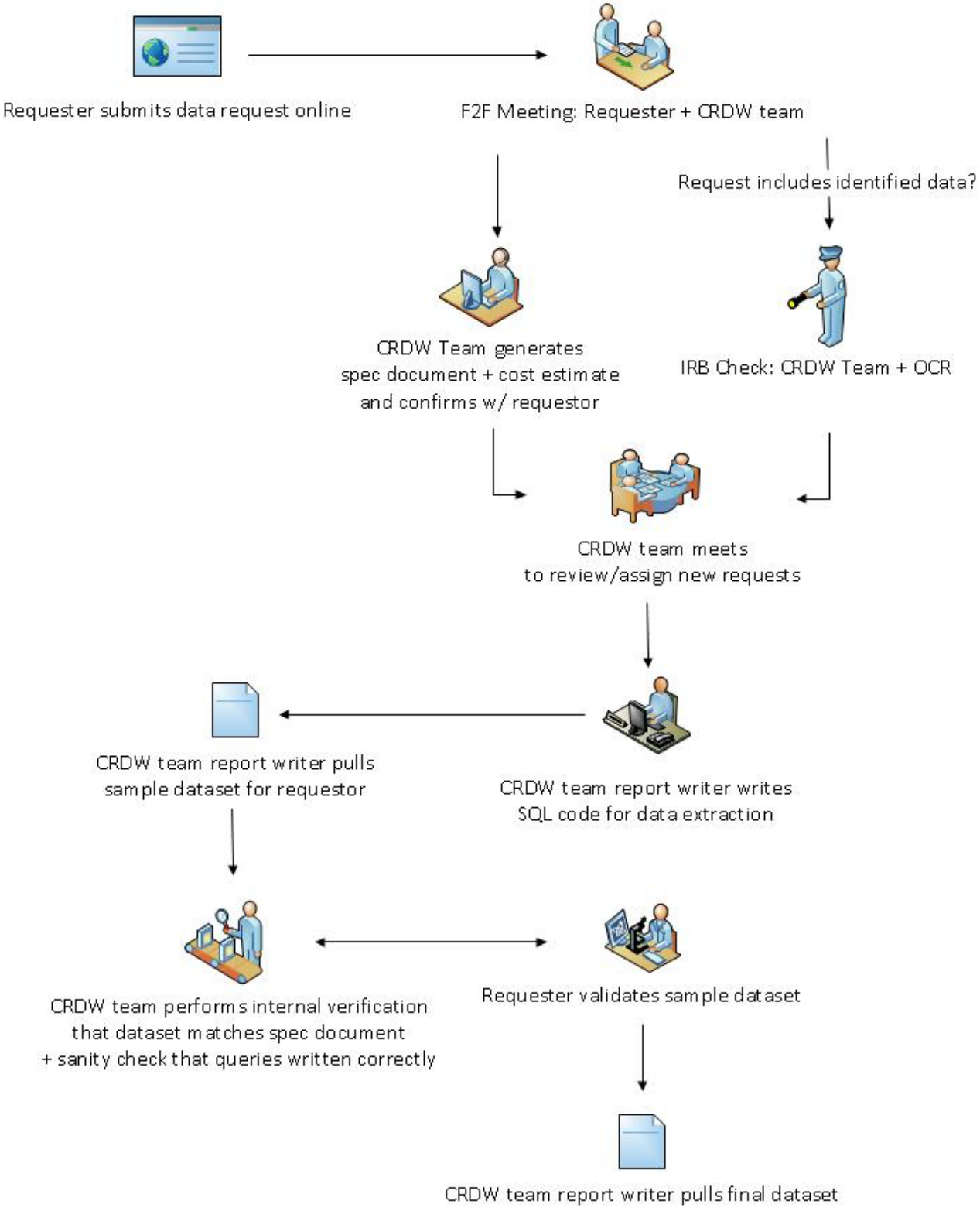
4. DATA REQUEST APPROVAL WORKFLOWS

Each CRDW data request passes through a review and approval process before it is filled.

The below shows the approval workflow for each data request type:



The below diagram details the approval workflow for data requests.



5. DATA REQUEST POLICIES

Policy on Request Processing

All standard data requests are processed in the order they are submitted. The current average turnaround time is 4-6 weeks, but varies based on availability of resources and complexity of request. Our chargeback model is outlined on our website at cri.uchicago.edu/crdw.

If your request requires IRB approval, you must first obtain that IRB approval before any work on your data request will begin. All requests for data will be reviewed against the corresponding approved IRB protocol, if applicable. If the data you are requesting does not match the methods approved in the IRB protocol, you will be asked to revise and resubmit your request.

Policy on Data Analysis

The CRDW staff will provide users with raw data to conduct their own analyses for research. At this time, we cannot provide analysis services such as aggregation (e.g., count of patient visits broken down by month), matching patterns of encounters (e.g., selecting for labs closest to the date of a patient's surgery), or custom formatting of data (e.g., transforming data into a different format to load into a third-party application).

Policy on Invalid or Incomplete Data Requests

All data points being requested must be clearly defined and the purpose expressly stated to ensure that the data request accurately reflects the data needed to perform the research defined in the request. A meeting with CRDW staff at office hours will help you to construct an appropriate data request. Below are examples of data requests that cannot be fulfilled.

Data points requested: "All relevant data points in the CRDW."

Reason this request cannot be fulfilled: The CRDW contains thousands of data points from Epic, Centricity, and the Cancer Registry, and returning all data points is not feasible. We do have standard dictionaries of labs, medications, diagnoses, and procedures, which can be browsed interactively using i2b2.

Data points requested: "Chemotherapy, staging, histology, etc."

Reason this request cannot be fulfilled: Each data point must be specifically named in the request. Researchers who are not domain experts in the area for which they are requesting data are encouraged to work with a collaborator with the necessary domain expertise to specify relevant data points.

Data point requested: "Cause of death."

Reason this request cannot be fulfilled: While this data point may exist in text notes, it is not a discrete field in any electronic system. Requested data points must correspond to a discrete field in an electronic system, or users may request de-identified text notes from which to do manual abstraction.