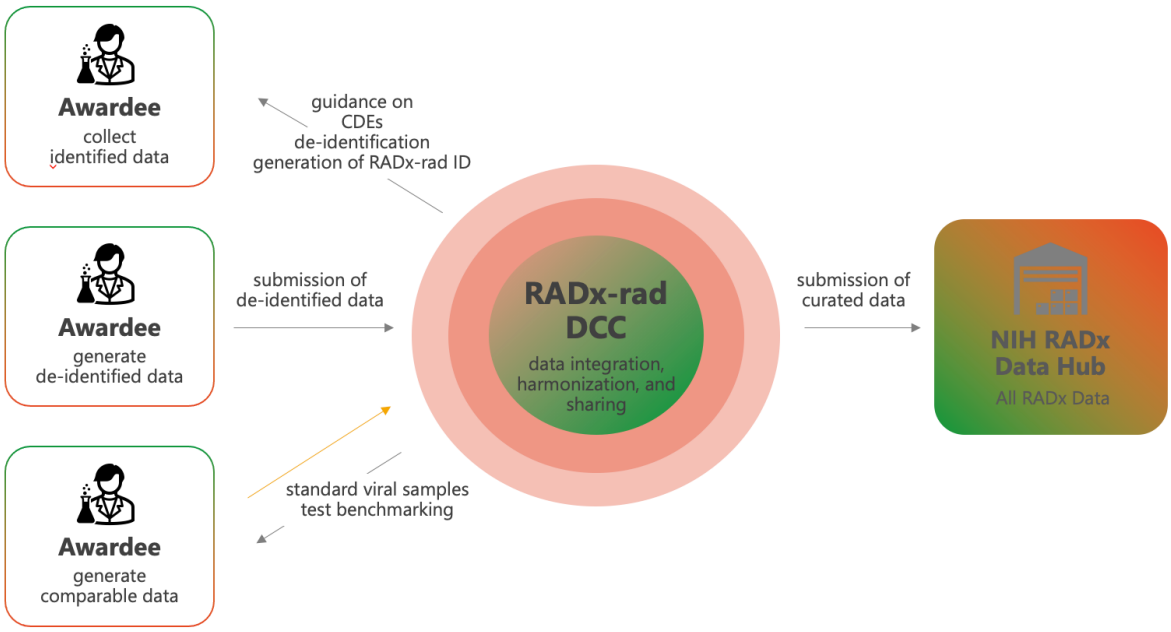
Data flow for RADx-rad awardees

1. **Key Definitions.** The *RADx Data Hub* is the ultimate destination of RADx-rad data. It is managed by the NIH. The *RADx-rad DCC* will ensure that data submitted to the hub will comply with standards agreed upon by *the RADx program* at NIH. The DCC will help awardees transform their data into Common Data Elements (CDEs), generate participant RADx-IDs, and help awardees remove identifiers so that data can be shared in a way that helps protect the privacy of participants. Identified data will not be stored at the RADx-rad DCC.



2. **Common Data Elements**. NIH has defined a minimum set of *CDEs* that should be prospectively collected for human participants by every RADx-rad awardee. We are committed to working with the NIH and RADx-rad awardees to discuss how to accommodate all needs. A separate document on CDE selection and/or development discusses the processes that the DCC will follow to determine what other CDEs might be relevant to RADx-rad as a whole, and to each of the 13 funded programs. We will also work with individual awardees to assist in CDEs and other items that are specific for their project.
3. **Types of Data**. The standardization at the point of data collection, via CDEs, will help ensure that data are comparable across awards. RADx-rad is a collection of programs in which there are new diagnostic methods (most based on virus detection, but some based on symptoms), as well as predictive models that use existing diagnostic methods and phenotype data for prognosis. As a result, data may come from very specialized instruments, some of which are completely novel and some others are adaptations of existing instruments, as well as from electronic health records (EHRs), customized

surveys, environmental sources (e.g., wastewater), etc. Although some standardization will happen at the point of collection, we anticipate that, for pre-existing data such as EHRs, some mapping will be necessary, and we will assist awardees in this process.

4. **IRB**. Understanding that it is important that awardees include all data collection items in their protocols and explain their provenance and chain of custody in consent forms, we have developed office-hours and individual consultation for awardees.
5. **Identified Data**. The current version of the minimum set of required CDEs includes identifiers (e.g., address). Data de-identification will happen at the awardee site.
6. **"De-identified" Data**. The DCC developed a checklist of items that should not be submitted to the RADx-rad DCC helps awardees remove identifiers from their data. This is a list typically used to consider data "de-identified," although we recognize that this does not guarantee protection against certain types of privacy attacks. For this reason, all data received from awardees will be held in a protected environment that is compliant with HIPAA regulations. In the future, a program that checks for simple potential identifiers will be sent to awardees so they can double check their files. In the rare case in which the RADx-rad inadvertently receives data that contain identifiers (e.g., a participant's name is included by mistake in another field), we will notify the awardee so that they can proceed with their institution's processes to notify participants if necessary.
7. **RADx-rad Study Subject ID**. The DCC will provide awardees with a tool to generate unique RADx-rad participant IDs and only include this type of ID in the shared data.
8. **Curation**. Although data may be submitted in the right format, impossible or improbable data values may be detected at the DCC. We will work with the awardees to ensure that these errors are fixed. This will be an iterative process that will require active participation from awardees and will result in a curated data set.
9. **Submission to the NIH RADx Data Hub**. The DCC will serve as an intermediary host for the awardee's de-identified data, and it will submit the files to the RADx Data Hub according to timelines that are mutually agreed upon (e.g., after embargo). Should problems with the data be detected at the NIH RADx Data Hub, the DCC will be charged with interacting with the awardee to resolve them.