CRITICAL PATH INSTITUTE | NORD® National Organization for Rare Disorders

eatris

European infrastructure
for translational medicine

Q&A: RDCA-DAP and EATRIS Present Rare Disease Drug Development Webinar | Jan. 21, 2021

Data Contribution to RDCA-DAP

What disease areas does RDCA-DAP cover – does it include rare cancers?
RDCA-DAP covers all rare diseases, including cancers. We do not have any cancer data in-house yet, but would be interested in sharing such data as it comes available.

What types of data will RDCA-DAP collect?
The RDCA-DAP team is interested in all types of rare disease data. While high quality clinical trial and natural history data are more rigorously collected and standardized, and may be of greater value for certain purposes, (e.g., historical control arms, aiding in interpretation of clinical trials), other types of data, such as data from patient registries, can be very informative in clinical trial design, understanding patient preference in terms of treatment and generally understanding the disease natural history and variance in a population. Used together these different data types provide a holistic view of rare diseases, so all data types are welcome.

Are biopharma companies sharing data with RDCA-DAP?
Yes, biopharma companies have expressed great interest in the platform, and we have several clinical trial datasets in house from pharmaceutical and biotech companies. We are also actively discussing and negotiating new datasets from both clinical trials and industry-sponsored natural history and registry studies with pharmaceutical companies. There is also a great deal of interest in the ability to access data and analytics through the platform.

How flexible is the process of bringing data into the platform – can you provide embargo periods before public access to data or limit use to certain groups?
Data sharing to RDCA-DAP is established through a Data Contribution Agreement for each individual dataset, in which terms are negotiated on a per case basis. We prefer a model where the custodian will not restrict access to their data through our platform. However, entities who share their data with us may decide to limit access in various ways thus retaining some control over use of the data. Specifically, data custodians may choose to place an embargo on use of the data so that it can only be shared after a certain date or event (e.g., drug approval, publication by the custodian), approve access to specific groups only (e.g., if the patient consent does not include sharing with specific types of researchers), approve access to the dataset themselves, etc. These terms are recorded at a dataset level.

Does RDCA-DAP accept only US data, or can data from other places (e.g., within Europe) be included?
Our goal is to be able to accept and use data from around the globe. We have successfully shared data from multiple jurisdictions and have many features in place to ensure that our database meets the needs of global privacy regulations. Our de-identification practices help us comply with HIPAA and GDPR laws by removing personal health information. Data contributors may also pseudonymize their patient-level data and destroy the key so that C-Path's copy of the data may be considered anonymized under GDPR. Further information on GDPR-compliance is discussed below in the section specific to interactions with entities in Europe.

Does RDCA-DAP support federated data sharing, or does it only accept physical copies of data? Can people access the data in other data federation platforms? Can data be kept in a secure environment?
RDCA-DAP does support federated data access both in and out. Physical data sharing to the platform is preferred as it allows us to map the data to standardized structures and ontologies, increasing the ability to search and integrate datasets. However, we recognize the challenges of sharing physical copies of data under certain circumstances and will set up systems to federate data from other sources where physical sharing cannot be achieved. This will increase the data availability within RDCA-DAP and reduce data silos, which is a goal of the project. Federated access to data will also be possible, after approval for use of the specific data requested through our standard data request system.

In addition to federated access, where platforms are unable to transform legacy structures to a standard and agreed-upon data model, C-Path will be developing application interfaces (API's) to share data in our data lake that may be standardized and annotated by our data management and data science teams.

RDCA-DAP's platform also allows for data to be used within a secure work environment within the platform. Our workspaces will include most common analytic tools and additional tools can be brought in or developed in the platform. Data that cannot be downloaded (due to permissions of the data custodian) will be able to be used and manipulated in a safe environment, and results will be able to be downloaded as permitted by the data custodian where data owner agreements allow for downloads.

Data In RDCA-DAP

What data sharing standards will be utilized? Do data standards need to be implemented prior to sharing?
Data may be accepted in most formats, with no need for incoming data to be standardized (although we encourage data custodians to adhere to standardized structures, ontologies and common data elements where possible). Once in RDCA-DAP, the data will be explored and tagged with standardized ontologies (e.g., OMIM, Monarch) to ensure searchability. Data will be mapped to data standards appropriate to the type of data for use in analysis, and data will be able to be exported in the Clinical Data Interchange Standards Consortium (CDISC) data structure to support regulatory submissions. Some data may remain unstructured, but searchable until it can be standardized.

Can you elaborate how you assure/validate data quality?
Different types of data will require different levels of data quality and data validation and will be used for different purposes within the platform. As data enter RDCA-DAP they go through several steps of curation and standardization. In this process, values that are outside of likely ranges are identified and we run other standardized checks on data quality and the data custodian is asked a series of questions to ensure that the data has been transmitted accurately. C-Path's DCC always stores raw data as a master copy and shares all code used in the curation and mapping processes to ensure full data provenance in pursuit of regulatory engagement. Moreover, all data access and manipulation will be audited by the platform.

Once a dataset is being used for an analysis (by our quantitative medicine scientists or outside users) additional standardization and curation may occur, and additional questions about the data provenance may arise. During data exploration for an analysis, a number of tests are typically completed to ensure similarity between related datasets and outliers are identified and explored. At that point, additional clarifications may be needed from the data custodian to ensure quality. This iterative process will result in continual data quality improvements. However, as we are not collecting the data ourselves, we cannot guarantee the accuracy of the data.

How will the platform account for the likelihood of duplicate records from the same patient in different studies? This is a likely occurrence in rare disease datasets.
RDCA-DAP recognizes the likelihood that individual patients may be represented in multiple datasets that are integrated into the database. In anonymized datasets, if the contributed data have not adopted global identifiers, it is not possible to recognize potential duplicate patients across datasets. However, the potential bias towards an artificially reduced analytics uncertainty can be adequately controlled by standard statistical and mathematical methods. Other quantitative drug development tools built by C-Path using databases with this same issue have accounted for this bias within the development of the models and tools, and this has not affected the utility of the tools developed. Moving forward, we strongly support the use of global identifiers (GUIDS) that can be used in all studies to allow us to link records from individual patients without identifying the person in question.

How does RDCA-DAP differ, collaborate or overlap with RARE-X and other data sharing platforms?
RDCA-DAP is set up to collaborate with all other data sharing platforms to make data available across the rare disease data ecosystem. We have several memoranda of understanding in place with other data platforms to explore how this can work to maximize the availability of data. We do not collect new prospective data, but aim to be able to share and integrate data across different data platforms and make it available through our analytics interfaces for use by the community. Depending on the needs of our partners, we may provide physical copies of their data or federated access to their data and can include their data sharing requirements and processes within our platform processes. We are open to sharing data back to be made available through other platforms, so long as the requirements of our data custodians can be met.

RDCA-DAP for European Collaborators

What steps does C-Path take to comply with GDPR?
Because C-Path hosts data securely in a cloud environment, our servers can be physically located in many regions of the world. Due to the rigor of the GDPR regulations, RDCA-DAP's data hub will be based in the Netherlands in 2021. Where necessary, and when the cloud service provider makes servers available, C-Path will store the data within another desired country or governmental jurisdiction, but will still have the security, infrastructure and tools that are available in other regions. This means that C-Path can access shared data from the US while hosting those data securely within in EU, for example, but may also aggregate those data with other data sources hosted elsewhere.

C-Path now has European branches as well as a US entity. Which part of C-Path should European entities interact with?
RDCA-DAP is being run out of our US organization, so all questions and collaborations can come to the US branch, through our email rdca-dap@c-path.org. For now, most scientific work is being done in the US, with specific data and regulatory activities occurring in the EU to meet with global regulations. This may change over time, but all C-Path branches will remain working together and can be reached through central contact points.

The UK has their own orphan drug legislation now. Have you started your cooperation already?
RDCA-DAP has not yet initiated a formal interaction with any regulatory agency except the US Food and Drug Administration at this time, although this is being explored. Other C-Path analytics/drug development tool projects have been developed and endorsed by global regulators, and the UK will likely be included in this in the future.

RDCA-DAP's Business Model

What is RDCA-DAP's financial model – how is the platform supported and will the data be monetized?
RDCA-DAP is funded through a collaborative grant to C-Path from the US Food and Drug Administration [Grant Number U18 FD005320] and developed through a collaboration between C-Path and the National Organization for Rare Disorders® (NORD). While the existing grant is for a defined term, the expectation is that there will be continued funding to support the platform over time. There is no cost for organizations sharing their data with RDCA-DAP, and we will not charge for access to data in the platform. It is possible that individual data contributors may charge for access to their datasets, but these transactions occur separate to RDCA-DAP.

-C-Path-