An Institutional Strategy to Support Clinical Research with Centrally Managed Custom Data Repositories

Stephen B. Johnson, PhD[1], Thomas R. Campion, Jr., PhD[1], Nonie E. Pegoraro, MA[1], Leon Rozenblit, JD, PhD[2], Charles Tirrell[2], Curtis L. Cole, MD[1]
[1]Weill Cornell Medical College, NY, NY; [2]Prometheus Research, LLC, New Haven, CT

Weill Cornell Medical College

PROMETHEUS | RESEARCH™

Introduction

Healthcare organizations face significant technical and organizational challenges to support clinical research. These barriers have historically led to fragmentation of the research enterprise, which in the extreme force individual researchers to manage their data separately. In some cases, groups of researchers who share common interests have developed custom data repositories that meet their collective needs. [1] These approaches do not scale as an institutional strategy. When there are multiple local repositories, many data management activities are redundant, regulatory processes are convoluted, and opportunities for data sharing and collaboration in new areas are limited. One potential institutional strategy is to develop a centralized data repository, typically by making operational clinical data available to researchers for secondary use. [2] This approach often fails to meet the specialized needs of heterogeneous research groups for primary data collection, study management, local data integration and curation. Exclusive focus on central processing can reduce opportunities to improve research productivity. [1] We describe here a new strategy that focuses on the needs of local research groups, while attempting to achieve economies of scale at the institutional level with a sustainable funding model.

Methods

Architecture for Research Computing in Health (ARCH) partitions information technology for clinical research into external data sources, shared infrastructure and custom research data repositories. External sources include **electronic health records, research administration systems, electronic data capture and biobanks**. Shared infrastructure includes a loading zone where external data can be organized and integrated, and a working zone where data are transformed into structures needed locally. [2] Research data repositories provide highly customized, self-service facilities for extracting data for analysis and identifying cohorts as well as data quality assessment and exploration. Shared components are largely funded centrally while custom components are charged back. Both shared and custom components are implemented using an **open-source platform (RexDB®), which offers a model-driven architecture, meta-data management, configuration services and web-based query language (RexQL)**.

Results

The institution adopted the ARCH strategy with significant financial support from the Joint Clinical Trials Office and the Clinical and Translational Science Center to support hardware and staff for shared infrastructure, as well as a scientific advisory board to vet requests for creating new custom repositories. In addition, two research groups (anesthesia and digestive care) have adopted the strategy by investing in services to provide custom data transformations, user interfaces and reports. The demand for additional repositories from other research groups is high, with three planned for the coming year (neurogenetics, urology and myeloid biology) and new requests added every month.

Discussion

The ARCH strategy enables an institution to **centralize research infrastructure for secondary data use, regulatory compliance, data transformation, quality assessment, security, needs assessment and training.** Local research groups are fully empowered to collect and integrate specialized data, with customized workflow to maximize scientific collaboration and research productivity. A transparent model for cost sharing ensures financial sustainability.

References: 1 Hruby GW, McKiernan J, Bakken S, Weng C. A centralized research data repository enhances retrospective outcomes research capacity: a case report. J Am Med Inform Assoc. 2013 May 1;20(3):563-7. 2 Abend A, Housman D, Johnson B. Integrating Clinical Data into the i2b2 Repository. Summit on Translat Bioinforma. 2009 Mar 1;2009:1-5.

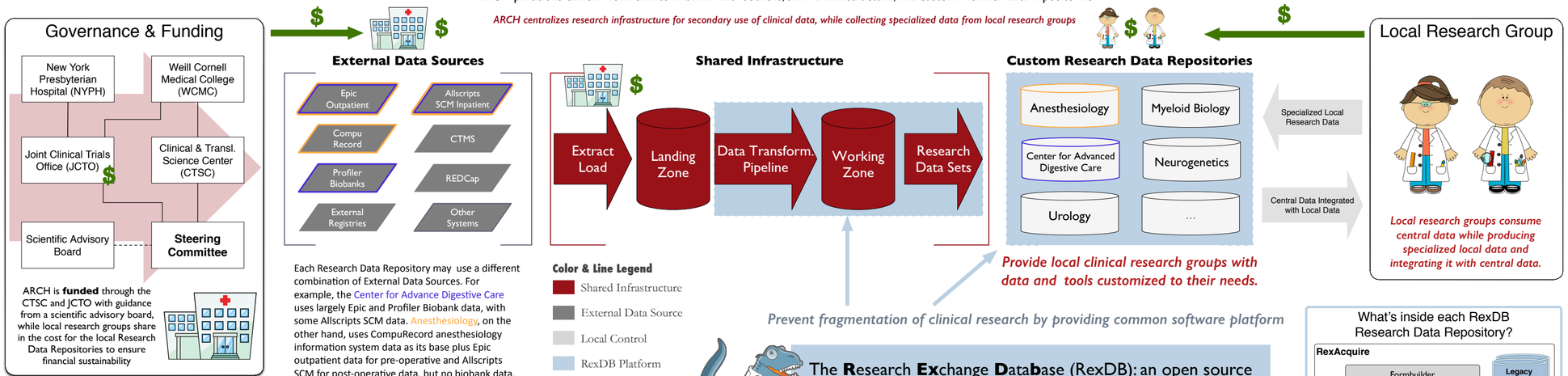
Funding: CTSC funding from NCATS through UL1 TR000457-06. RexDB research reported in this abstract was supported by the National Institutes of Health and the National Science Foundation (Award ID: 98771, grant numbers R43MH102900 and R43MH099826).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health and the National Science Foundation.

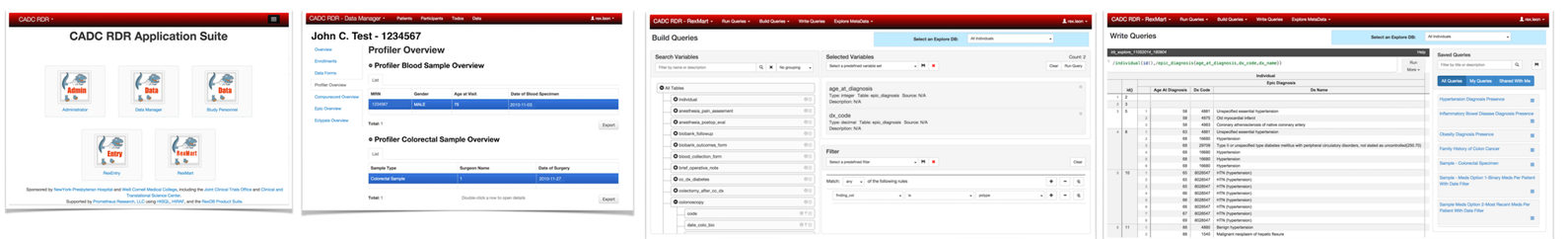
Overview: Architecture for Research Computing in Health (ARCH)

ARCH partitions clinical research into external data sources, shared infrastructure, and custom research data repositories

ARCH centralizes research infrastructure for secondary use of clinical data, while collecting specialized data from local research groups



Examples: End-User Applications Combining Data from Multiple Sources



What's inside each RexDB Research Data Repository?