Introduction to the Research Informatics Center
Overview

- SoM data science infrastructure
- Obtaining data for research
- Data coordination services
- Examples of RIC collaborations
SoM Data Science Infrastructure

<table>
<thead>
<tr>
<th>PHS</th>
<th>RIT</th>
<th>RIC</th>
<th>QSU</th>
<th>DBDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center for Population Health Sciences</td>
<td>Research Information Technology</td>
<td>Research Informatics Center</td>
<td>Quantitative Sciences Unit</td>
<td>Department of Biomedical Data Science</td>
</tr>
<tr>
<td>Epidemiologic Data</td>
<td>Technology Infrastructure</td>
<td>Informatics Services</td>
<td>Biostatistics Services</td>
<td>Analytic Tools and Methods</td>
</tr>
<tr>
<td>- Population-level data</td>
<td>- Clinical data for research and infrastructure</td>
<td>- Data consultation</td>
<td>- Statistical design and analysis for medical studies</td>
<td>- Data Studio</td>
</tr>
<tr>
<td>- World data sources</td>
<td>- STARR tools and services</td>
<td>- Data coordination</td>
<td></td>
<td>- Innovative study design</td>
</tr>
</tbody>
</table>
When RIC Can Help

- Define the research question
- Explore preliminary data and patient counts
- Obtain compliance
- Identify funding sources
- Identify data and statistical analysis requirements
- Data collection and aggregation
- Statistical analysis
- Publication and scientific impact
- Research Informatics Center
Why work with RIC?

If you **plan to use clinical data for research** at Stanford Medicine, and need guidance accessing the data, you should work with RIC.

**RIC is the source of clinical data for research** use at Stanford Medicine.

Epic data is **intended only for clinical use** (including care as part of a clinical trial).

Use of **clinical data for research** triggers HIPAA and RIC handles the necessary HIPAA compliance.
## Clinical Data Sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clarity</strong></td>
<td>Clarity database: Electronic Health Record (EHR) reporting data set from SHC and LPCH <strong>Epic</strong> systems</td>
</tr>
<tr>
<td><strong>Epic</strong></td>
<td>Bedside monitoring data (LPCH only)</td>
</tr>
<tr>
<td><strong>Images</strong></td>
<td>Images from PACS and labeled image data sets</td>
</tr>
<tr>
<td><strong>Ancillary</strong></td>
<td>Ancillary systems data sets (Pathology, Genetics, Dermatology, etc.)</td>
</tr>
<tr>
<td><strong>Curation Application for Disease Database (CADD)</strong></td>
<td>Curation Application for Disease Database (CADD)</td>
</tr>
<tr>
<td><strong>SCIRIDB</strong></td>
<td>Stanford Cancer Institute Research Database (SCIRIDB)</td>
</tr>
<tr>
<td><strong>OMOP</strong></td>
<td>Project-specific data sets, registries, and CDMs (Common Data Models such as OMOP, derived from Epic)</td>
</tr>
</tbody>
</table>
Ancillary Systems Data (Available Now + Pending*)

- OnCore (Clinical trial data)
- ARIA (Radiation Oncology)
- PowerPath (Pathology)
- STAMP and FoundationOne (Molecular mutation data)
- TraceMasterVue (EKG) and XCelera (Cardiac imaging)
- OnBase (Scanned documents in the Epic Media tab, such as dermatology photos)
- SafeTrace, UNOS, and Nihon Koden NeuroWorkbench
New Initiative: STARR-OMOP

- STARR-OMOP-deid database offers pre-IRB access to the de-identified OMOP dataset
- Identified STARR-OMOP data available as custom extract
- Good for researchers familiar with common data science practices
- Standardized Data
  - Ideal for multi-site studies
  - Useful for collaboration with sites using different EHR systems (e.g., Epic and Cerner)
- Improved Data Quality
  - Standard ontologies
  - NLP-enabled notes
- Scalable Science

Learn more here: https://med.stanford.edu/starr-omop.html
## Data Sources Best Fit and Access

<table>
<thead>
<tr>
<th></th>
<th>STARR-STRIDE</th>
<th>STARR-OMOP</th>
<th>Clarity (Epic data)</th>
<th>Ancillary Clinical Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete data</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Varies</td>
</tr>
<tr>
<td>Self-service</td>
<td>Yes</td>
<td>Yes (de-id data only)</td>
<td>Paid custom service available</td>
<td>Paid custom service available</td>
</tr>
<tr>
<td>Training provided</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Best fit for</td>
<td>Case studies, prospective studies, patient chart review</td>
<td>Population Health research, network studies, AI models</td>
<td>QA/QI</td>
<td>Custom, large research projects</td>
</tr>
<tr>
<td>Main advantage</td>
<td>Fully integrated with Stanford IRB and Privacy Office processes</td>
<td>Allows running models and comparing results between institutions</td>
<td>Complete data</td>
<td>Offers specialized access to critical clinical elements for analysis not available in Epic</td>
</tr>
<tr>
<td>Point of Access</td>
<td>STARR Self-Service Cohort/ Chart Review Tools</td>
<td>Email Request to Priya Desai ( <a href="mailto:prd@stanford.edu">prd@stanford.edu</a> )</td>
<td>RIC Consult Request</td>
<td>RIC Consult Request</td>
</tr>
</tbody>
</table>

More info here: [https://med.stanford.edu/starr-omop/summary.html#clarity_stride_omop](https://med.stanford.edu/starr-omop/summary.html#clarity_stride_omop)
Self-Service: STARR Cohort Discovery Tool

Build a cohort based on:

- **Demographics**
  Search by Age, Gender, Race, Ethnicity, or Language

- **Clinical Events**
  Search by ICD-9/10 Codes, CPT Codes, Lab Codes, Drug, Clinical Documents, Clinic/Provider Encounters, Admission, or Vital Status

- **Temporal Constraints**

User guide: [https://med.stanford.edu/starr-tools/](https://med.stanford.edu/starr-tools/)
Self-Service: STARR Chart Review Tool

- Up to 1,500 patients per cohort
- View and download patient data in .csv format
  - Encounters
  - Diagnoses
  - Procedures
  - Lab Results
  - Medications
  - Clinical Notes
  - Flowsheets
  - Pathology Reports
  - Radiology Reports

User guide: [https://med.stanford.edu/starr-tools/](https://med.stanford.edu/starr-tools/)
Custom Data Extracts and Data Coordination

- Clinical data custom extracts services
  - RIC provides paid custom extracts of data not available through the free STARR self-service tools or the STARR-OMOP database
  - Weekly office hours are provided for initial project development
  - PTA salary support funding required (up to $90/ hour)

- Data coordination services
  - Available for $126 per hour in FY21
  - Stanford Cancer Institute faculty PIs receive 8 hours of free data coordination services per year
Request a Consultation with RIC

https://med.stanford.edu/ric.html
# Data Coordination Services

<table>
<thead>
<tr>
<th>Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Extraction</td>
</tr>
<tr>
<td>Data Integration</td>
</tr>
<tr>
<td>Data Curation</td>
</tr>
<tr>
<td>Data Analysis</td>
</tr>
</tbody>
</table>
CADD: Patient data curation
CADD Features

- Enables investigators to formulate research questions over time, based on elements they have curated
- Supports analysis capabilities for outcomes, response to treatments, survival
- Can be used as a training dataset for machine learning
- Allows integration of legacy and other data sources
SCIRDB

- Stanford Cancer Registry
- Bone Marrow Transplant / CAR-T Database
- Solid Tumor and Lymphoma Databases
- Specialized Oncology Clinical Datasets (Pathology, Mutations, Imaging, Radiation Therapy)
SCIRDB: Biostatistical analysis console

Kaplan-Meier Overall Survival by SEX

Days Post-Dx for SEX

Outcomes
The team worked with clinicians to define “non-small cell lung cancer” patient cohort (consulting the Stanford Tumor Registry, diagnosis codes and pathology keyword search techniques).

Identified a subset of patients that had some KRAS mutation results.

A dataset was provided to the researchers, allowing them to dig deeper into the KRAS mutation sub-types, other pathological mutations, and determine clinical outcomes for each patient.
A disease-specific clinical research data warehouse combining multiple data sources was created to facilitate secondary data use and enhance observational research.
Project Example – Imaging Data

Obtaining Imaging data for research:
https://med.stanford.edu/ric/consultation-service/imaging-data.html
Imaging data and analysis tools

- Commercial tool for cancer lesion measurement and tracking
- Standard RECIST criteria
- > 5,000 lesion measurements (image annotations)
- Integration with ePAD for automated response assessment and waterfall plots
What should our group do next?

- We recommend you form a **Data Science Team** within your Disease Group
- We can work with you to build a **disease specific data infrastructure** that suits your needs. Consider the following strategies:
  - Identify **variables of interest** (data dictionary) – What do you plan to analyze?
  - **Cohort identification** – Inclusion and exclusion criteria
  - Identify **datasets of interest** (such as demographics, diagnosis, treatment, medication, lab results, clinical notes, encounters, or procedures)
  - Define **milestones/deliverables/timeline**
- We can then setup a consultation and provide you a cost estimate – Financial model (FTE and Service Center model)
Thank you!

Visit our website to request a consultation:

https://med.stanford.edu/ric/