THE UNIVERSITY OF CHICAGO MEDICINE

Statistical Modeling of Clinical Data: the Clinical Research Data Warehouse (CRDW)

Julie Johnson, MPH, RN

Senior Healthcare Business Analyst

February 1, 2018

Disclaimer and Permissions

- This educational activity is being presented without the provision of commercial support and without bias or conflict of interest from the planners and presenters.

- The purpose of these materials is to provide insight into the CRDW and internal funding opportunities related to CRI services but is not meant to be exhaustive of all programmatic details.

- It is not permissible to share these materials outside of UCM or the BSD

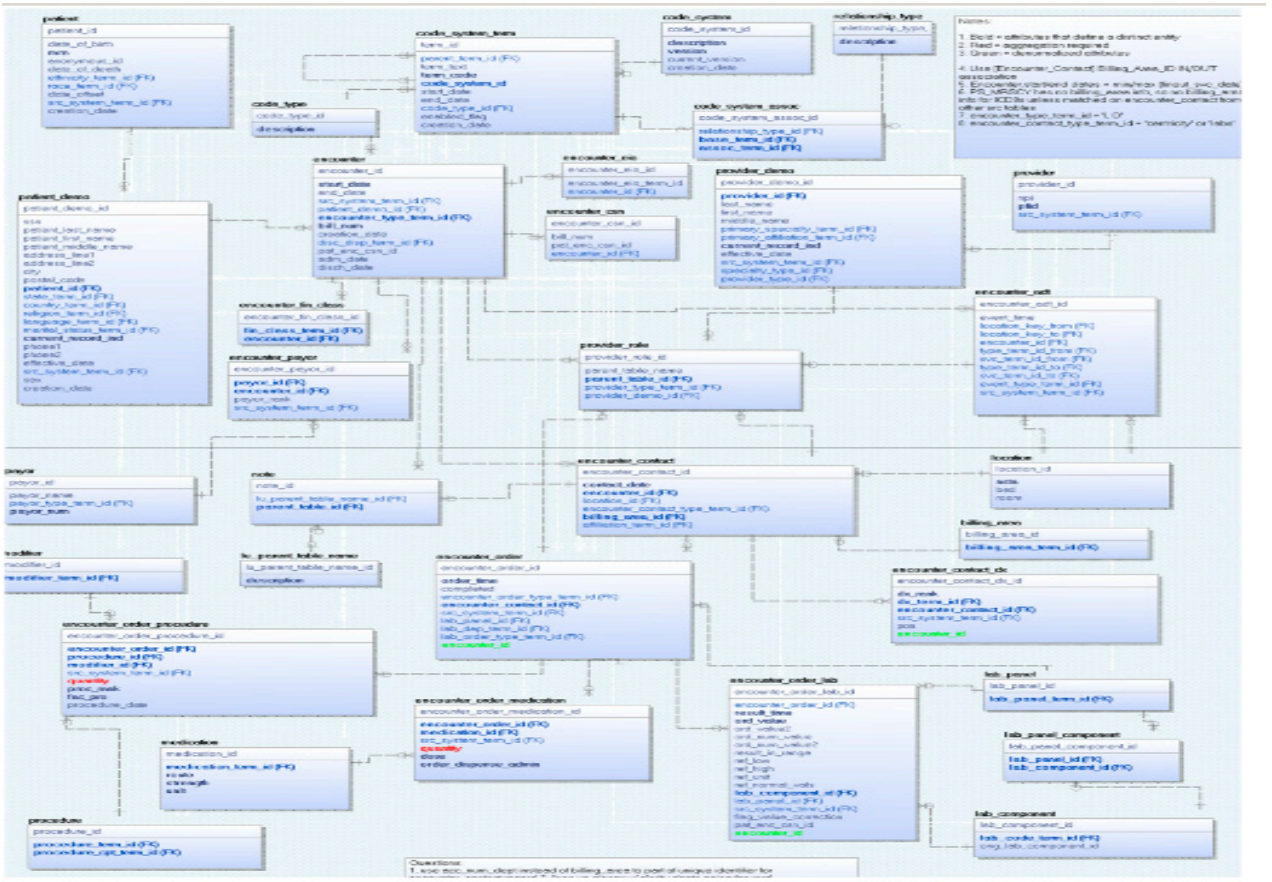
Agenda

- What is the CRDW?
- How do you build a cohort?
- Questions?

CRI: The Clinical Research Data Warehouse (CRDW)

- CRDW is a repository of University of Chicago medical data dating back to 2006.
- The CRDW team brings together data from disparate sources such as Epic, Centricity billing, Cancer Registry, LabVantage and REDCap to create cohesive datasets for research.

CRDW is a Physical Database

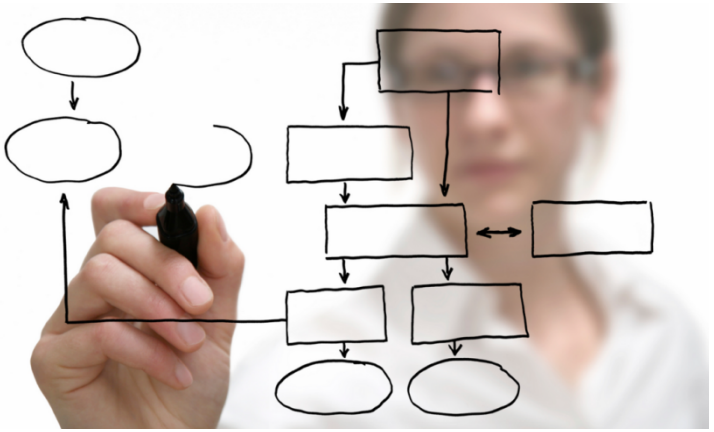


CRDW is a process

More than just hardware and software

The fastest hardware and the latest and greatest software is only a **small part of that process**

Without good processes in place for managing data, requesting data, governance of data, consistent standards, and quality control, the best hardware and software will do one thing: **serve up bad data faster**



What is IN the CRDW?

- The **most commonly-requested data elements** required to answer research questions are the following:
- Patient Demographic Info
- Encounter Info (Encounter type, Admit Date, Discharge Date, etc.)
- Diagnoses: ICD9/10
- Procedures: ICD9/10, CPT
- Flow Sheets: Vitals, Respiratory, Physical Assessment, etc.
- Medications: Outpatient and MAR
- Labs
- ADT: (Admission, Discharge, Transfer)

The Clinical Research Data Warehouse (CRDW)

The UChicago Clinical Research Data Warehouse

Patients	850,243
Diagnoses	23,959,858
Encounters	9,568,899
Labs	166,313,565
Medications	8,834,525
Procedures	47,731,079
Providers	70,941

What about notes?

- Notes are available in bulk and we can provide as part of a data request

- However, defining a cohort or doing additional filtering/logic using info from within the note is difficult, time-consuming, and expensive using Natural Language Processing

- Not impossible, but can be cost-prohibitive

Output from CRDW: 2 Flavors

- Static Data Set: 1x data extraction
  - Example: Finding a cohort of patients that match on a specific set of demographic, clinical, and billing variables.

- Dynamic Data (AKA Affiliated Data Mart): regularly updated data
  - Example: Bringing together a registry of patients maintained in REDCap w/ clinical data from Epic/Clarity, billing data from Centricity, and cancer data from the Cancer Registry

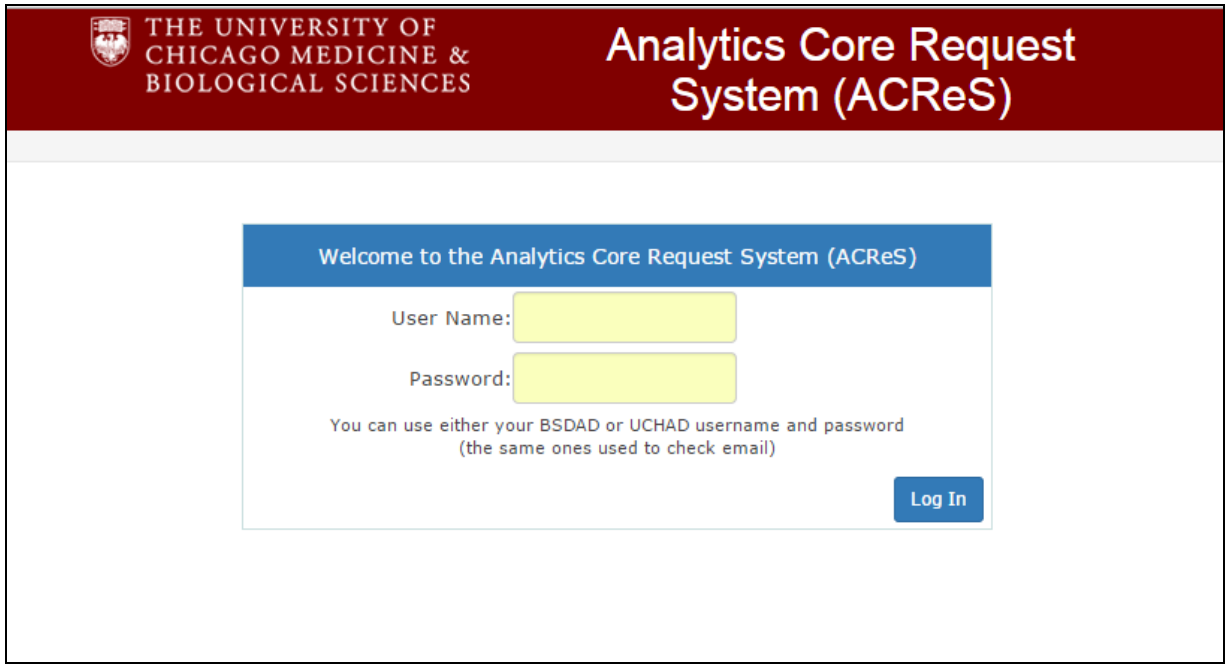
How do I request data?

How Do I Request Data? ACReS

- Analytics Core Request System
- https://cri-app02.bsd.uchicago.edu/ACReS

Step 1: Log into ACReS Website

- https://cri-app02.bsd.uchicago.edu/ACReS



CRDW Spec Document

PI: [redacted]
Contact: [redacted]
Title: [redacted]
Request #: 11242
IRB#: [redacted]
Data Type: PHI

Cohort Identification:
- Outpatients seen in Heme/Onc
- 4/1/2015-6/30/2015
- Designation of "palliative" intent in treatment plan

Index encounter is the first encounter that meets this criteria.

Data Set Output: Outpatient, Emergency, and Inpatient Encounters from Index through 6/30/2016

Cohort Info:
- MRN
- DOB
- DOD
- Race
- Gender
- Ethnicity
- Primary Language (as available)
- Religion (as available)

Encounter Info:
- MRN
- Bill_Num
- Enc_EID
- Admission_type (planned, emergent, urgent)
- Adm_DTTM
- Dis_DTTM
- Start_DTTM
- Attending_Provider
- Visit_location
- Admit_location
- Dis_location
- Fin_class
- Marital_status
- Primary_ICD9_Dx
- Primary_ICD9_Desc
- Primary_ICD10_Dx
- Primary_ICD10_Desc
- Flag field for ICU admission during encounter

Diagnosis Info:
- MRN
- Bill_Num
- ICD9_Dx
- ICD10_Dx
- Charlson Comorbidity Score

Procedure Info:
- MRN
- Bill_Num
- ICD9_Px
- ICD10_Px
- CPT

Medication Info:
Anti-neoplastic agents only
- MRN
- Bill_num
- Med_name
- Order_DTTM
- Start_DTTM
- End_DTTM
- Given_DTTM
- DC_DTTM

Radiation Therapy:
- MRN
- Bill_num
- DTTM (of occurrence)

Notes: Index Encounters
ECOG score via Natural Language Processing
- MRN
- Bill_Num
- ECOG score
  (Could also be PS: 0-4)

Cancer Registry Info:
- MRN
- Date of initial diagnosis
- Primary site
- Date of first contact
- Current treatment

Time Estimate (Standard; 4-6 weeks):

Task Description	**# Hrs**	
Initial meeting at Office Hours	1	Not billed
Generate, document, and review data request specifications	1	Not billed
Generate SQL queries to pull requested data	22	
Cancer Registry	6	
Analysis: NLP	10	
Create, QA, and deliver sample/final data set	2	
Billable Hours ($95/hr)	40	$ 3800.00

CRI Resources

- **Office Hours**
  - Tuesdays, Room N161, by appointment 10-4
  - Fridays, Room N161, by appointment 10-12

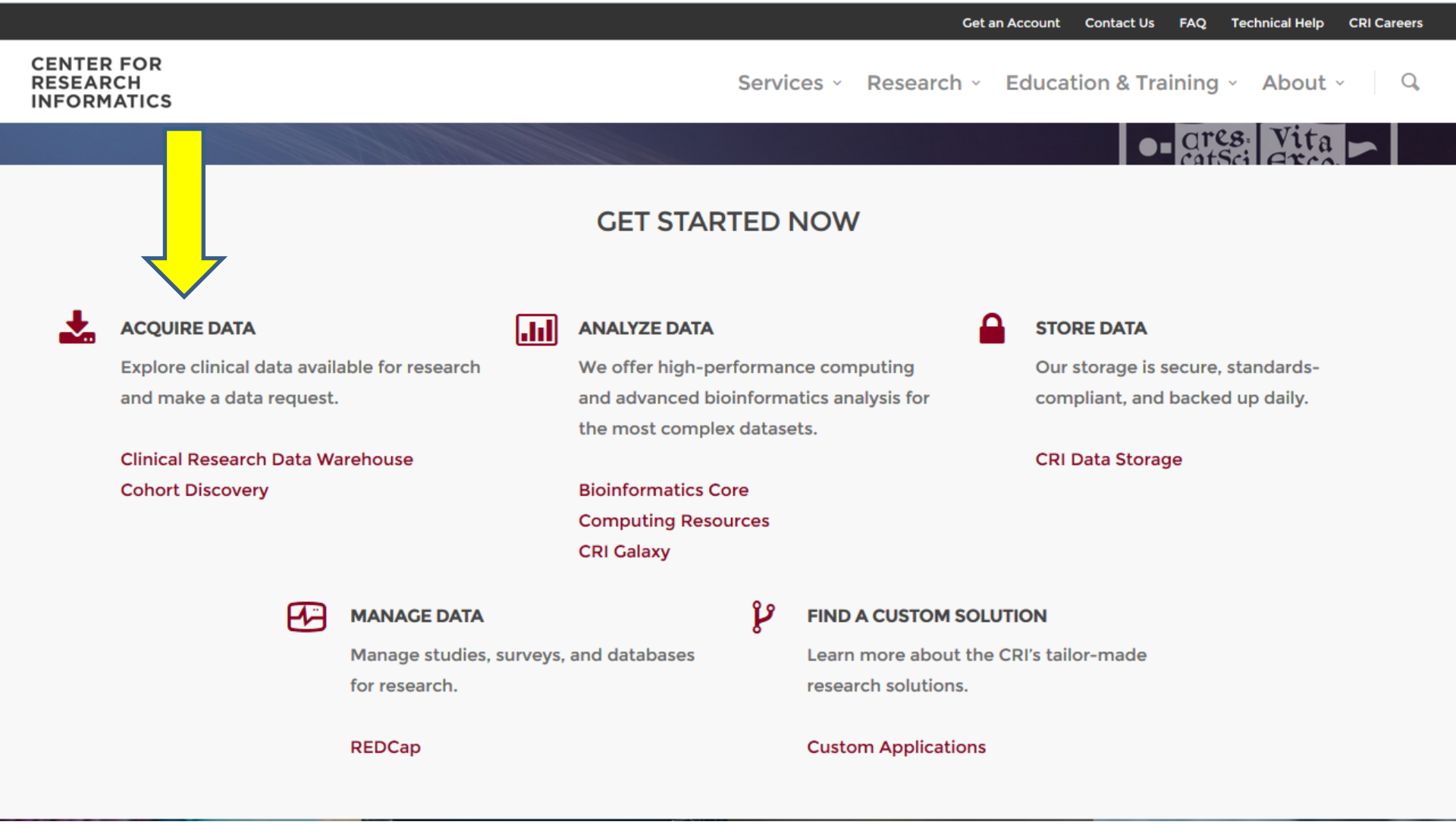
- **IRB guidance**
  - CRDW specific IRB language

- **Grant application guidance**

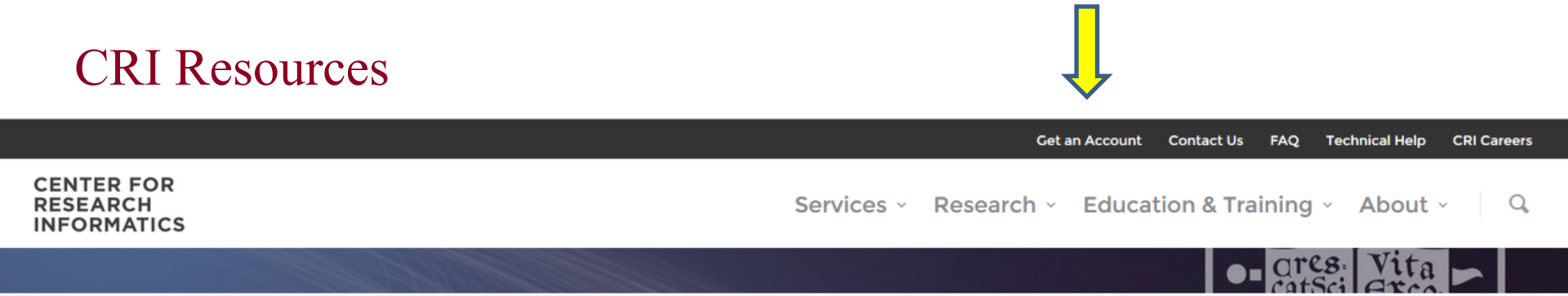
- **Semi-self service tools**
  - I2b2
  - SEE Cohorts

- **Data Storage Solutions**
  - REDCap
  - Bulkstorage

CRI Resources



CRI Resources



GET STARTED NOW

ACQUIRE DATA

Explore clinical data available for research and make a data request.

Clinical Research Data Warehouse
Cohort Discovery

ANALYZE DATA

We offer high-performance computing and advanced bioinformatics analysis for the most complex datasets.

Bioinformatics Core
Computing Resources
CRI Galaxy

STORE DATA

Our storage is secure, standards-compliant, and backed up daily.

CRI Data Storage

MANAGE DATA

Manage studies, surveys, and databases for research.

REDCap

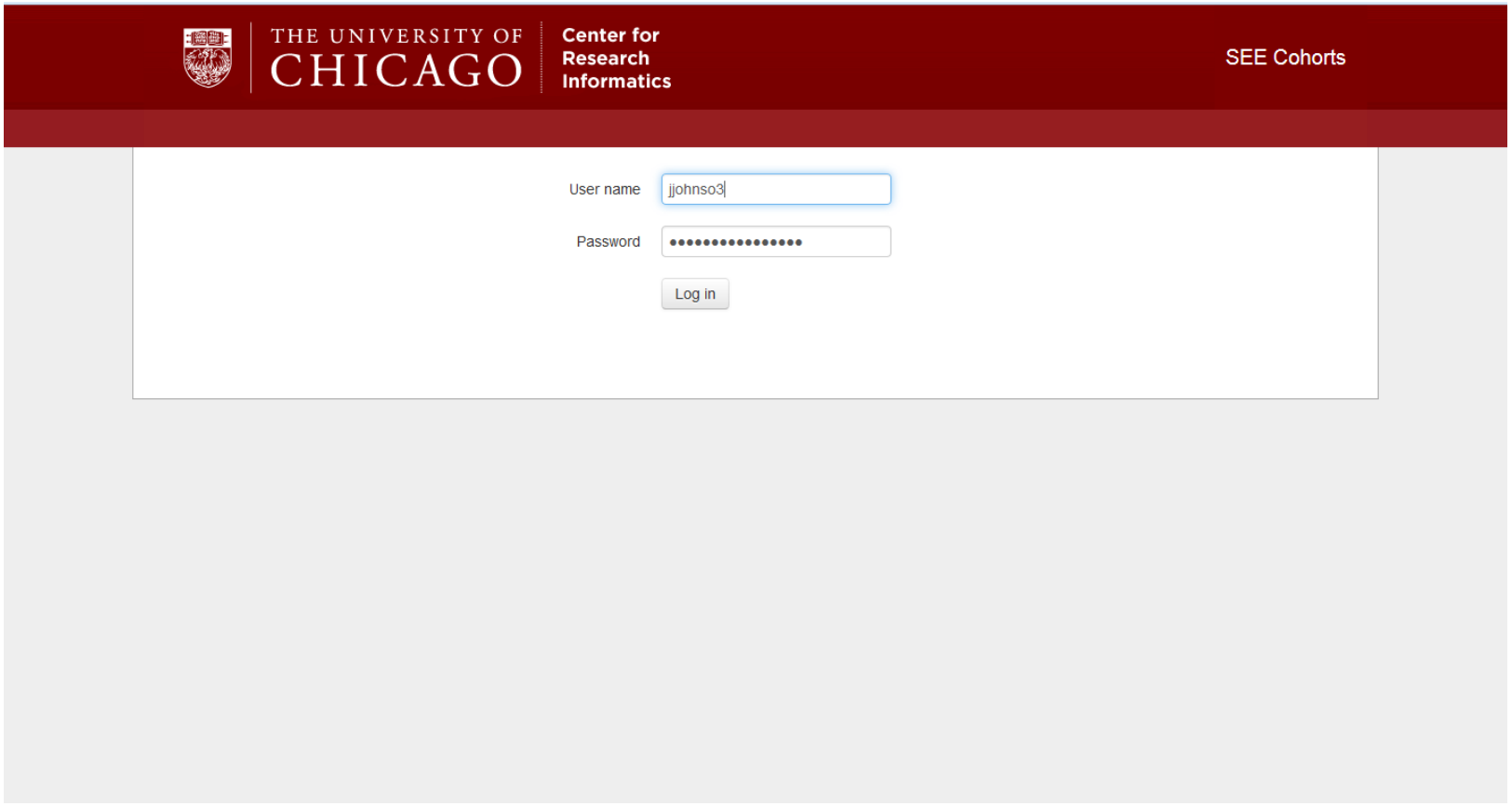
FIND A CUSTOM SOLUTION

Learn more about the CRI's tailor-made research solutions.

Custom Applications

SEECohorts

(https://seecohorts.cri.uchicago.edu/)



Contact Information

Julie Johnson
jjohnso3@bsd.uchicago.edu