Clinical Informatics Research Division (CIRD)

**Faculty:** Alfredo Tirado-Ramos, PhD
Alex F. Bokov, PhD

**Home Department:** Epidemiology and Biostatistics

**Contact:** https://i2b2.uthscsa.edu/request
informatics@uthscsa.edu
CIRD's core service: i2b2
"Integrating Informatics from Bench to Bedside"
What you can do with i2b2

- Preparatory to research (prior to IRB approval)
  - See which data elements are available
  - Develop and improve inclusion criteria
  - Choose possible variables to use in your study
  - Prepare counts and aggregate tables for use in grant proposals, budgets, and protocols

- Research (after IRB protocol or determination)
  - All the above, plus...
  - Request visit-level data for the patients in your i2b2 query.
  - Request contact information for the patients in your i2b2 query (with an IRB-approved protocol).
The stages of an i2b2-powered project

- **Stage 1:** obtain an account (via the System Access Form, https://i2b2.uthscsa.edu/request) and perform preparatory queries.
- **Stage 2:** start filling out a Project Request Form in order to commence a research project
- **Stage 3:** schedule a consultation with CIRD
- **Stage 4:** finish preparing your protocol documents and submit your Project Request Form
- **Stage 5:** use the final version of your preparatory query to request de-identified data and/or patient identifiers (depending on the type of protocol you have)
Stage 1: Preparatory to Research

- Obtain an i2b2 account via https://i2b2.uthscsa.edu/request
  - Make sure to hold on to the password you choose and the URL for the project menu screen you will see at the end

- Watch these videos
  - https://youtu.be/1GroXZA_iTo
  - https://youtu.be/yt7djtOkk-0?t=31

- Explore i2b2-- try to create your eligible patient-set, and take notes on variables you might want to include in your analysis.
Stage 1: Preparatory to Research

As you explore i2b2, try to answer these questions:

- What are my main hypotheses or research goals?
- What are my inclusion/exclusion criteria?
- How many patients meet those criteria in i2b2?
- Do I need to recruit patients, or can I do the entire study using only the data?
- What variables will I need from i2b2?
- Who will analyze the data and how much effort did they say it will take?
- Is there anything I need that I am not certain is in i2b2?
- What funding opportunities are aligned with my project?

When you can answer the above questions or identify what else you need to know before you can answer them, you are ready to submit a Project Request Form.
Stage 2: Project Request Form

The Project Request Forms appear on a menu after you complete the System Access Form. Please hold on to this URL and to the password you chose. You will need them when you come back and fill out a Project Request Form.

Survey Queue

Congratulations, you have successfully completed the CIRD System Access Request form. You will shortly receive an email with the subject "+++Welcome to i2b2". This email will contain instructions for logging into i2b2, including your password.

Below are three identical CIRD Project Request Forms. You will need them when you are ready to request patient contact information for recruitment purposes or to pull deidentified visit-level data for retrospective analysis. You don't have to have everything ready to fill out your first Project Request Form— in fact, part of its purpose is to help you figure out what you need to do next. You can return to update the form at any time as your research plan evolves.

Once you have gotten as far as uploading a research plan, we invite you to schedule a [CIRD consultation](#).

<table>
<thead>
<tr>
<th>Status</th>
<th>Survey Title</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Completed</strong></td>
<td>CIRD System Access Agreement (Preparatory to Research) – 00SystemAccess</td>
</tr>
<tr>
<td>Begin survey</td>
<td>CIRD Project Request – Request01</td>
</tr>
<tr>
<td>Begin survey</td>
<td>CIRD Project Request – Request02</td>
</tr>
<tr>
<td>Begin survey</td>
<td>CIRD Project Request – Request03</td>
</tr>
</tbody>
</table>
Stage 2: Project Request Form

• If you only need **de-identified data**, you can fill out and submit a Declaration of Non Human Research (or Declaration of Non Research) right in the Project Request Form.

• If you need **identified data**, you need to submit a human subjects protocol directly to IRB and then forward their entire response, with **all** attachments, to informatics@uthscsa.edu, then indicate on the Project Request Form that you have done so.
Example language for describing how you will get data from i2b2
(I welcome suggestions/corrections from IRB members in the audience):

In the preparatory-to-research stage of this project, we constructed a feasibility query against the de-identified i2b2 data warehouse operated by the Clinical Informatics Research Division (CIRD). We will request that CIRD provide the identifying information listed in Form-J of this protocol for each of the patients matched by our i2b2 query, along with a study-assigned ID number with which we could link it to any non-identifying data we request for those patients. As authorized by IRB repository protocol HSC20150212HR, CIRD will provide all files to us via encrypted, password-protect secure file transfer, after which point we will assume responsibility for protecting the data. Our data protection procedures will be as follows: ...
Stage 3: Consultation

When you have completed the Project Request Form, or at least got as far as you could based on what you know, you can request a consultation:

https://i2b2.uthscsa.edu/consult
Stage 3: Consultation

These meetings have three goals:

- Reviewing the i2b2 query you created on your own and hopefully finalizing it.
- Helping you complete your Project Request Form if you are encountering obstacles.
- If it turns out your project needs capabilities that are not already available in i2b2, putting together an initial plan to prepare a grant proposal (or adjust an existing one).
Stage 3: Consultation, how to prepare

- Please bring your team and anybody else you want there. The more the merrier.
- Please bring at least one laptop that can connect to HSCWave.
- At least one participant should already have an i2b2 account, and in advance of the meeting should have already attempted to turn their inclusion/exclusion criteria into an i2b2 query.
- Please make note of any questions that occurred to you while using i2b2, and bring those questions with you.
Stage 4: Complete the submission of your protocol and Project Request Form
Stage 5: Request your data
Stage 5: Request your data

How the extracted data looks: one row per visit, one column per variable. Values that do not change are repeated at each visit (e.g. DOB). Values that are missing during some visits are blank.

<table>
<thead>
<tr>
<th>PATIENT</th>
<th>VS_DATE</th>
<th>DOB</th>
<th>SEX</th>
<th>LANG</th>
<th>RACE</th>
<th>VS_AGE</th>
<th>FRAC</th>
<th>SPRN</th>
<th>BMI</th>
<th>DIA_BP</th>
<th>SYS_BP</th>
<th>PULSE</th>
<th>RESP</th>
<th>TEMPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>14678</td>
<td>2011-04-17</td>
<td>1990-08-08</td>
<td>f</td>
<td>spanish</td>
<td>white</td>
<td>7557</td>
<td>N</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14678</td>
<td>2011-05-30</td>
<td>1990-08-08</td>
<td>f</td>
<td>spanish</td>
<td>white</td>
<td>7600</td>
<td>N</td>
<td>N</td>
<td>26.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14678</td>
<td>2011-07-11</td>
<td>1990-08-08</td>
<td>f</td>
<td>spanish</td>
<td>white</td>
<td>7642</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14678</td>
<td>2011-08-14</td>
<td>1990-08-08</td>
<td>f</td>
<td>spanish</td>
<td>white</td>
<td>7676</td>
<td>N</td>
<td>N</td>
<td>24.4</td>
<td>67</td>
<td>111</td>
<td>92</td>
<td>96.3</td>
<td></td>
</tr>
<tr>
<td>14678</td>
<td>2011-10-16</td>
<td>1990-08-08</td>
<td>f</td>
<td>spanish</td>
<td>white</td>
<td>7739</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10309</td>
<td>2012-09-29</td>
<td>1989-01-25</td>
<td>m</td>
<td>english</td>
<td>black</td>
<td>8648</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10309</td>
<td>2012-10-28</td>
<td>1989-01-25</td>
<td>m</td>
<td>english</td>
<td>black</td>
<td>8677</td>
<td>Y</td>
<td>N</td>
<td>26.2</td>
<td>53</td>
<td>123</td>
<td>66</td>
<td>16</td>
<td>97.1</td>
</tr>
<tr>
<td>10309</td>
<td>2012-12-10</td>
<td>1989-01-25</td>
<td>m</td>
<td>english</td>
<td>black</td>
<td>8720</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10309</td>
<td>2012-12-28</td>
<td>1989-01-25</td>
<td>m</td>
<td>english</td>
<td>black</td>
<td>8738</td>
<td>N</td>
<td>N</td>
<td>24.9</td>
<td>62</td>
<td>112</td>
<td>49</td>
<td>14</td>
<td>97.9</td>
</tr>
<tr>
<td>10309</td>
<td>2012-12-30</td>
<td>1989-01-25</td>
<td>m</td>
<td>english</td>
<td>black</td>
<td>8740</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10309</td>
<td>2013-08-15</td>
<td>1989-01-25</td>
<td>m</td>
<td>english</td>
<td>black</td>
<td>8968</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.1</td>
<td></td>
</tr>
</tbody>
</table>

Simulated data for illustrative purposes, from Bokov et al., Procedia Computer Science 2016, v80 pp1-9
What types of data are in i2b2?

- Epic EMR system: all UT Health including MARC and CTRC, ~400,000 patients
  - demographics, diagnoses, labs, vital signs, procedures, medications, allergies, provider specialty, visit type, insurance category
- Social Security Death Master File: dates of death
- NAACCR: cancer registry (copy of our site's contribution to the Texas Cancer Registry)
Things that will soon be in i2b2

- Median household incomes (December 2017)
- Sunrise EMR system: all UHS, ~1,200,000 patients (First Quarter, 2018 or whenever DEB approves purchase of faster server)
  - billed diagnoses, billed procedures, demographics, selected vital signs
  - additional data will be included over time
- Links to biosamples (First Quarter, 2018)
Things that are not in i2b2

- Information that is not recorded in the electronic medical record nor in other data-sources that can be linked to the medical record, e.g.: diet, exercise, individual education level, individual income, reliable linkage of family members to each other.
- Information that we don't yet have a way to fully de-identify, i.e. text notes.
- Information that is amenable to a text query, e.g. images (but we could link to an existing image repository).
- Visits outside UT Health or UHS (e.g. we do not have any VA data) and populations who are not seen at these sites.
- Upcoming appointments or any realtime data. We update quarterly at best.
Other Limitations

• We cannot run your queries for you, only help you refine or troubleshoot your queries. If you want to use the system, you or somebody on your team will need to learn to use i2b2.

• Use of existing i2b2 capabilities is free of charge. If you need custom variables to be created, new data sources linked, etc. we can prepare a time and cost estimate (and even help you write your grant).

• This is not a customer/vendor relationship. You are our academic collaborators. Our plate is almost full with funded commitments and in-preparation grants but the more clear, technically feasible, and self-contained your research plan, the faster this collaboration will bear fruit.
Pending Grant Proposals

- CTSA Renewal (some coverage for operational costs and for assisting researchers)
- CTSA Collaborative Innovation Award (using natural language processing to extract certain types of information from text notes in a study spanning three i2b2-enabled CTSA sites)
- Disparities in kidney cancer R21 (extraction and linkage of additional UHS data needed by project, expanded biorepository capabilities)
- Disparities in surgical complications/readmissions R21 (linkage of UHS data to ACS NSQIP surgical registry, calculation of risk scores, and updated geocoding to link census data)
Beyond our site

- act-network.org
- gpcnetwork.org
- trinetx.com
- Many institutions that are not yet partnered with us (especially CTSA and OAIC sites) have compatible i2b2 infrastructure that could greatly simplify multi site studies.
Summary

• i2b2 is a powerful research tool, linking rich data sources that are not available anywhere else in this combined, analyzable form.

• Like any technology it also has limitations-- in particular, the types of data that are available and the personnel-hours available for assisting researchers.

• The challenge is to design studies that can work within these limitations and make the maximum possible use of i2b2's strengths…

• ...so that with each successive study we can make i2b2 an even better tool for studies that are yet to come.
Thank you!

• For helping us understand the regulatory process:
  – Dr. Kim Summers
  – Ms. Meyad Baghezza
  – The IRB Team
• For helping us navigate institutional hurdles:
  – Dr. Jennifer Potter
  – Ms. Kathy James
• For being our research champions:
  – Dr. Ron Rodriguez
  – Dr. Robert Clark
  – Dr. Paula Shireman
  – Dr. Nicolas Musi
  – Dr. Sara Espinoza
• Dr. Amelie Ramirez and the Department of Epidemiology and Biostatistics
• My fellow members of the CIRD team, informatics@uthscsa.edu:
  – Dr. Alfredo Tirado-Ramos (division chief)
  – Ms. Laura Manuel (lead developer)
  – Ms. Olivia Suarez (IRB liaison and project manager)
  – Ms. Margie Guevara (admin)
  – Mr. Eric Moffet (developer)