Overviewing the Translating Research in Elder Care Measurement System (TMS) Data Platform

James Doiron, Manager, Health Research Data Repository, University of Alberta, Canada
Andrew DeCarlo, Metadata Technology North America
Shane McChesney, Nooro Online Research

NADDI 2015
Madison, WI
USA

April 09, 2015
Translating Research in Elder Care

- PI Dr. Carole Estabrooks (T1 CRC in KT); FoN U of A
- Assessing influence of organizational context on use of knowledge (esp research) and, in turn, influence of knowledge use on LTC residents’ outcomes
- Practical objective: contribute to improving the quality of care and quality of life/end of life for residents living in residential LTC settings and quality of work life for care providers
- Multi province (T1 = 3/T2 = 4) & nursing homes (T1 = 36/T2 = 91)
- 30+ investigators, as well as policy and decision makers, trainees, staff
TREC Measurement System Data Sources

1. TREC care provider surveys
   - HCAs (75%) & regulated staff
   - ~200 variables—likert, categorical, & open text
   - Alberta Context Tool (ACT) embedded; derived context scales
   - Also non-ACT derived scales (i.e., MBI, SF-8, etc)
   - Administered using CAPI (HCAs) and online (regulated staff)

2. Facility Profile Survey
   - ~180 variables—categorical & open text
   - Administered in paper or electronic format by research managers

3. Care Unit Profile Survey
   - ~100 variables—categorical & open text
   - Administered in paper or electronic format by research managers

4. RAI-MDS 2.0
   - LTC resident outcomes
   - ~500 variables
   - Collected continuously through projects (2007-2020)
   - Full Admission, Annual, and Quarterly assessments
   - Quality Indicators, Scales, RUGS
   - Administered by care providers; data acquired from data stewards
Hypotheses

TREC UNIQUE: 1. Context data & 2. Care unit level data
Reflecting on TREC 1.0 challenges.....

- Metadata was limited and not existing in a way that was easily able to be updated
- Quality assurance activities very time consuming and costly
- Internal reporting and analytic capacity was limited and not efficient
- Very limited capacity to support feedback and reporting
- Challenges in making data accessible to investigators, both for practical and privacy related reasons
### TREC 1.0 (2007-2012) vs TREC 2.0 (2012/14 – forward)

<table>
<thead>
<tr>
<th>TREC 1.0 (2007-2012)</th>
<th>TREC 2.0 (2012/14 – forward)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 provinces &amp; 36 NHs</td>
<td>4 provinces &amp; 91 NHs</td>
</tr>
<tr>
<td>2 waves of surveys data collection</td>
<td>3 waves of surveys data collection</td>
</tr>
<tr>
<td>4000 staff surveys (3K HCA)</td>
<td>18,000 staff surveys (13.5K HCA)</td>
</tr>
<tr>
<td>125,000 RAI records</td>
<td>400,000 RAI records</td>
</tr>
<tr>
<td>$4.7 million CAD (CIHR funded)</td>
<td>~$9 million CAD (provincially)</td>
</tr>
<tr>
<td>T1.0 = Two main projects (quantitative &amp; qualitative), and three interrelated pilot projects</td>
<td>T2.0 = Three main projects (TMS, Intervention, &amp; SNA)</td>
</tr>
</tbody>
</table>
How the TREC Measurement System Data Platform project came to be…

I just want my data to be immediately accessible, with high quality control, pristine documentation, flexible to support active research, and with unlimited potential to support reporting and feedback…is that too much to ask for?!

Hi Carole! I couldn’t help but overhear your data woes! Hey, you know what you need? You need a data integrator!

What in the world is a data integrator?!

I’m soooo glad that you asked…

Uh-oh, Chuck has that look in his eyes…..
I am Pascal from MTNA! I am the Data Integrator!!

Did I mention that I like beer and hockey?

Umm, yeah, ok. Hey, James! Please secure CFI funding and work with this Pascal guy to solve all my data woes.

Welcome aboard Pascal!!

Data Integrator?! CFI?! What the?!!

Hey, I wanna play too!!
CFI TMS Data Platform

- CFI = Canada Foundation for Innovation
- CFI TMS = $1 million CAD
- Funding = 40(CFI)/40 (matching)/20(In kind)
- 4 year active phase = 2012-2016
- TREC + HRDR + MTNA + Nooro = The Team
- We are a little over half in
- Plan to declare operational Oct ‘15
- IOF funding = 100K
CFI TMS Data Platform

- **Stage 1:** *Data collection and ingestion*, involves production of dataset ‘index files’ through the automation of the collection/ingestion processes and capture of DDI friendly metadata.

- **Stage 2:** *Harmonization and merging of analytical files*, involves using the metadata to produce analytical files. Automated data quality processing scripts transform and clean data to maximize workflow efficiency and data quality. These automated processes are managed by the workflow automation *Extract-Transfer-Load* tool. Longitudinal metadata editor captures information and define relationships across data sources and/or versions.

- **Stage 3:** *Aggregation and business intelligence operations for reports generation*, involves the timely delivery of research data outputs/reports to long-term care sector stakeholders. Regular reports can be automatically generated, while novel report requests will be able to be produced with minimal resources required and in a timely fashion. Business intelligence (BI) software will be used to deliver reports, and these will be able to be produced in multiple formats, including PDF, HTML, Word, and Excel.
Components / Infrastructure

KUSP Knowledge Utilization Studies Program

Portal

UofA

HRDR

VRE

Knowledge Users

Files

Metadata

Aggregates DWH

Files

Metadata

Micro DWH

Aggregates DWH

Researchers

TREC

RAI

HRDR = Health Research Data Repository
DHW = Data Warehouse

VRE = Virtual Research Environment
What is the HRDR?

- Comprehensive research data platform developed to support health based research
- The HRDR is not just one thing, but rather it is a collection of physical, technical, process, and educational related services

Two broad areas of focus are:
1) Risk reduction
2) Data management
What is the HRDR?

- Secure virtual research environment (VRE) – remote access 24/7
- Located in Faculty of Nursing, U of A
- Developed to support health related research projects throughout their lifecycle
- Analytic software provided (SPSS, SAS, Stata, R, Nvivo, Mplus, StatTransfer, MSOffice, etc)
- Currently 60+ projects and 150+ approved users
- Lots of different types of projects & data: Quantitative, qualitative, personal health information...
- A number of custom desktops built to support the CFI TMS project
Automate The Automatable

- RAI-MDS Data File Uploads
  - ~500 variables per row
  - ~91 participating facilities
  - ~K rows expected
  - several versions of instrument
  - several source information systems with different export formats

Minimum Data Set (MDS) 2.0©
Canadian Version

MDS 2.0 Form © interRAI Corporation 1997, 1999
Canadianized items Copyright © CIHI, 2002

FULL ASSESSMENT

SECTION AA and A: IDENTIFICATION INFORMATION

<table>
<thead>
<tr>
<th>AA1</th>
<th>UNIQUE REGISTRATION IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AA2</th>
<th>SEX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M. Male F. Female O. Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A3</th>
<th>ASSESSMENT REFERENCE DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year Month Day</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AA3a</th>
<th>BIRTH DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year Month Day</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AA4</th>
<th>ESTIMATED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Real-Time Feedback

- Replace SFTP Site With Python-Powered Secure Upload Site
  - real-time feedback to uploader
  - notification of upload to data team
  - automated scrubs and reports
Reject Privacy Violations At Source

Personally identifying information or data for non-participating units/facilities is rejected immediately and does not get saved to disk.
Data cleaning as an automated-yet-reproducible series of steps:
1. The raw file as uploaded
2. The "normalized" file (adapting the 7+ incoming formats into a single clean layout)
3. Unit Reconciliation
4. ...more stages can be added once they are reduced to practice
Data remains in text files for as long as possible
Each version of the file is retained as a history
Folders can be version controlled with Mercurial
Goals:
  - Auditability
  - Reproducibility
  - Who changed what when?
  - Why?
Free Limited Human Resources To Do Higher-Level Work

- Browse any row of any file at any stage
- See flagged issues / warnings
- Add functionality as new needs / issues / fixes are discovered

All from having rich human & machine-readable metadata.
Components / Infrastructure

KUSP
Knowledge Utilization Studies Program

Portal

UofA
Files
Metadata
Aggregates DWH

HRDR
Files
Metadata
Micro DWH
Aggregates DWH

VRE
Researchers

TREC

RAI

Knowledge Users

HRDR = Health Research Data Repository
DWH = Data Warehouse

VRE = Virtual Research Environment
MTNA's Involvement

- Data / Resource management
- Reporting / Dashboards
- Documentation
Where does this start?

Before we start working with the actual data we need to gather and organize the related metadata.

• Gather and consolidate documentation

• Ensure the metadata is machine actionable

• Do away with the old documents as an input

• Generate documents on the fly based on the metadata
What does this allow us to do?

Now that we have our metadata gathered together we can use it drive our data management.

Having our metadata defined allows us to move the data securely and safely.

• CSV has been cleaned by NOORO and comes into the environment

• The CSV file is associated with some of the predefined metadata that describes its record layout and all the variables in the CSV.

• We move the data to a SQL database where it is stored.
What does this allow us to do?

Provide specific data and documentation when it is needed.

• Subsets of data can be created based on data requests.

• Codebooks and syntax can be generated on the fly to accompany the subsets.

Create reports and applications.

• RShiny dashboards

• RMarkdown reports
TMS Data Platform: Immediate & going forward

Immediate:
- 2-1/2 years into a 4 year project
- Core ‘under the hood’ components are developed – time to take ‘er for a spin!
- TREC 2.0 W1 survey data collection finishing in next 3 weeks
- May-June process and clean survey data
- RAI data starting to come in
- Jul-Sep reporting and feedback preparation
- Oct-Nov –w1 dissemination and feedback

Going forward:
- Declare operational in October
- Secure IOF funds and hire for specialized position
- Training!
- TREC-Boards development and release
- Waves 2 & 3
- CFI 2.0??
Questions?

Contact:

James Doiron
Manager, Health Research Data Repository
University of Alberta
james.doiron@ualberta.ca

Andrew DeCarlo
Metadata Technology North America
andrew.decarlo@metadatatechnology.com

Shane McChesney
Nooro Online Research
shane@nooro.com