

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

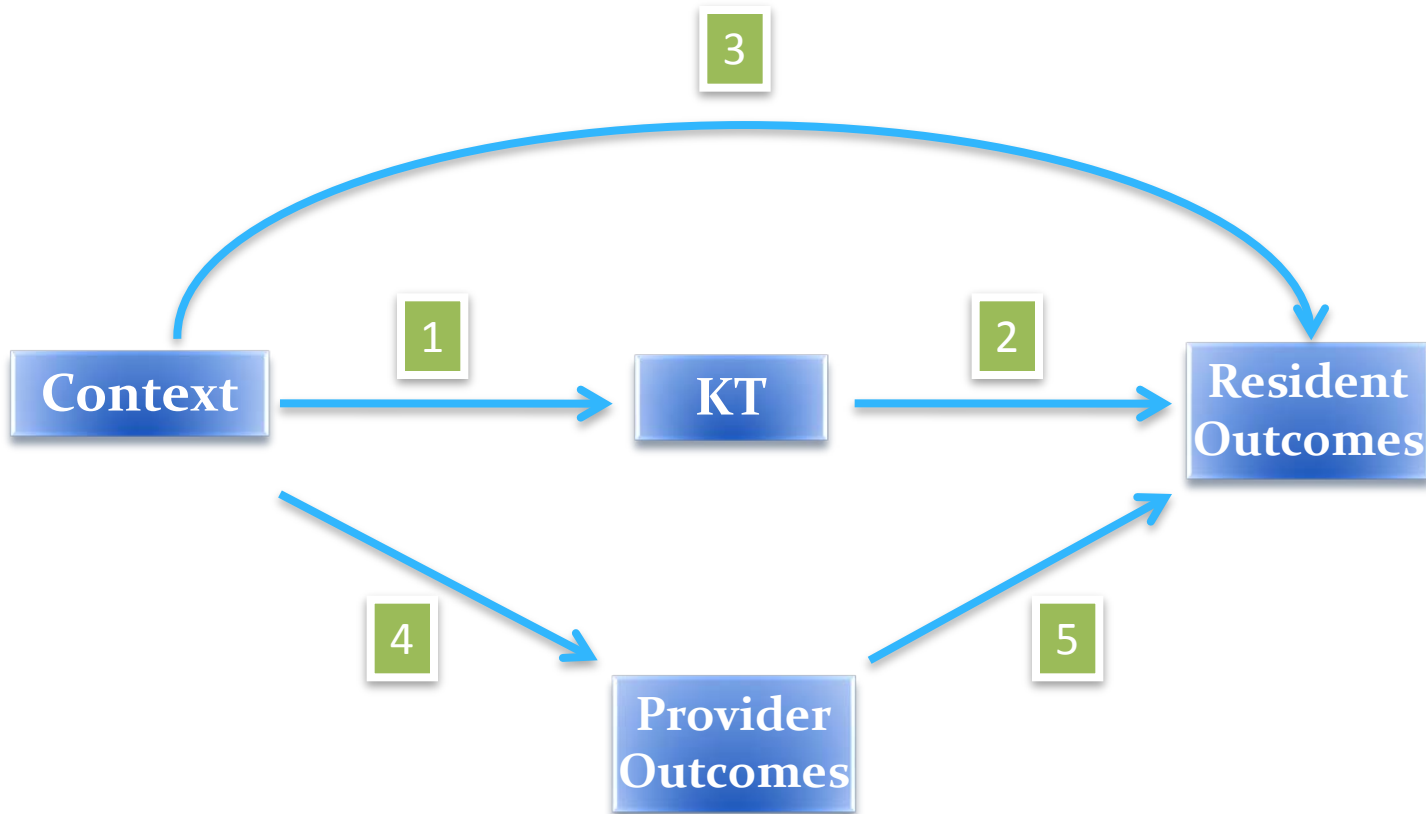
- ~100 variables—categorical & open text
- Administered in paper or electronic format by research managers

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



1.0



2.0

TREC 1.0 (2007-2012)	TREC 2.0 (2012/14 – forward)
3 provinces & 36 NHs	4 provinces & 91 NHs
2 waves of surveys data collection	3 waves of surveys data collection
4000 staff surveys (3K HCA)	18,000 staff surveys (13.5K HCA)
125,000 RAI records	400,000 RAI records
\$4.7 million CAD (CIHR funded)	~\$9 million CAD (provincially)
T1.0 = Two main projects (quantitative & qualitative), and three interrelated pilot projects	T2.0 = Three main projects (TMS, Intervention, & SNA)

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 capacity to support research.

Uh-oh, Chuck has that look in his eyes.....

Hi Carole! I couldn't help but to over hear your data woes! Hey, you know what you need?

I'm soooo glad that you asked...



I am Pascal from
M

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!!

...egrator?!
CFI?! What
the?!!

Hey, I wanna play
too!!



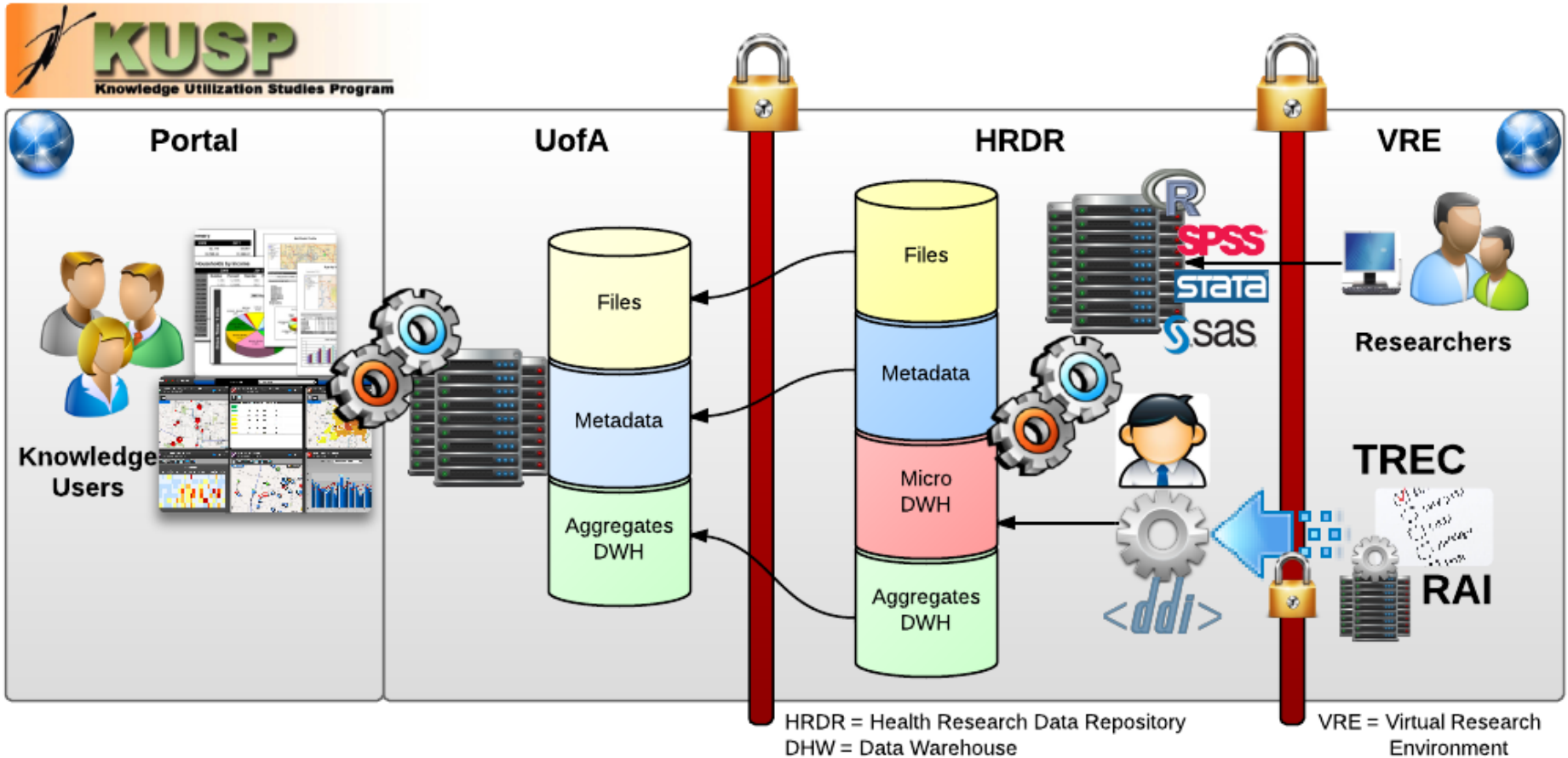
CFI TMS Data Platform

- 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



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

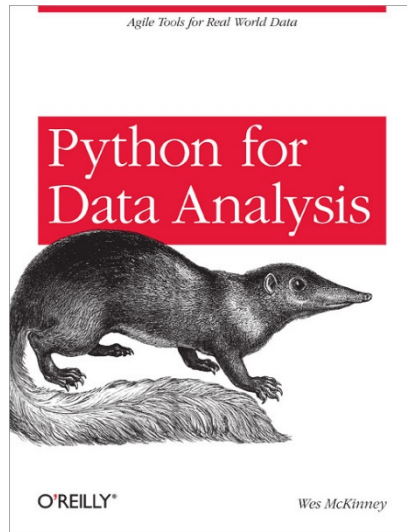
FULL ASSESSMENT

SECTION AA and A: IDENTIFICATION INFORMATION

AA1	UNIQUE REGISTRATION IDENTIFIER	<input type="text"/>		
	RESIDENT NAME	<input type="text"/>		
	ROOM NUMBER	<input type="text"/>	<input type="text"/>	
		a. Unit	b. Room #	
AA2	SEX	M. Male	F. Female	O. Other
A3	ASSESSMENT REFERENCE DATE	<input type="text"/>	<input type="text"/>	<input type="text"/>
		Year	Month	Day
AA3a	BIRTH DATE	<input type="text"/>	<input type="text"/>	<input type="text"/>
		Year	Month	Day
AA3b	ESTIMATED	<input type="text"/>		

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



Upload RAI File

Upload Notes

Enter Notes

Provider

File input

Browse... No file selected.

Choose an RAI file to upload by clicking on the above button or dragging the file onto it.

Submit

[Uploads](#) / Uploaded File

Data Check

Failed: see report for details

Filename

_scramble.csv

Uploaded

2014-02-27 12:47:17

Whitelist

[whitelist report](#)

Upload Notes

Notes

```
Unauthorized facility numbers exist: set([6057
```









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.

Inside The HRDR

RAI Management Tool Upload Uploads Users Units Instruments ▾

File Grid




				Stages		
		Provider	Date	0	1	2
	View	[REDACTED]	2014-04-09 00:49:55			
	View	[REDACTED]	2014-03-20 13:39:15			

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

Inside The HRDR

Stages



0	Raw		Explore	Download	
1	Columns renamed		Explore	Download	Generate
2	Unit Reconciliation		Explore	Download	Generate

Metadata Checks

Full Report

- 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?



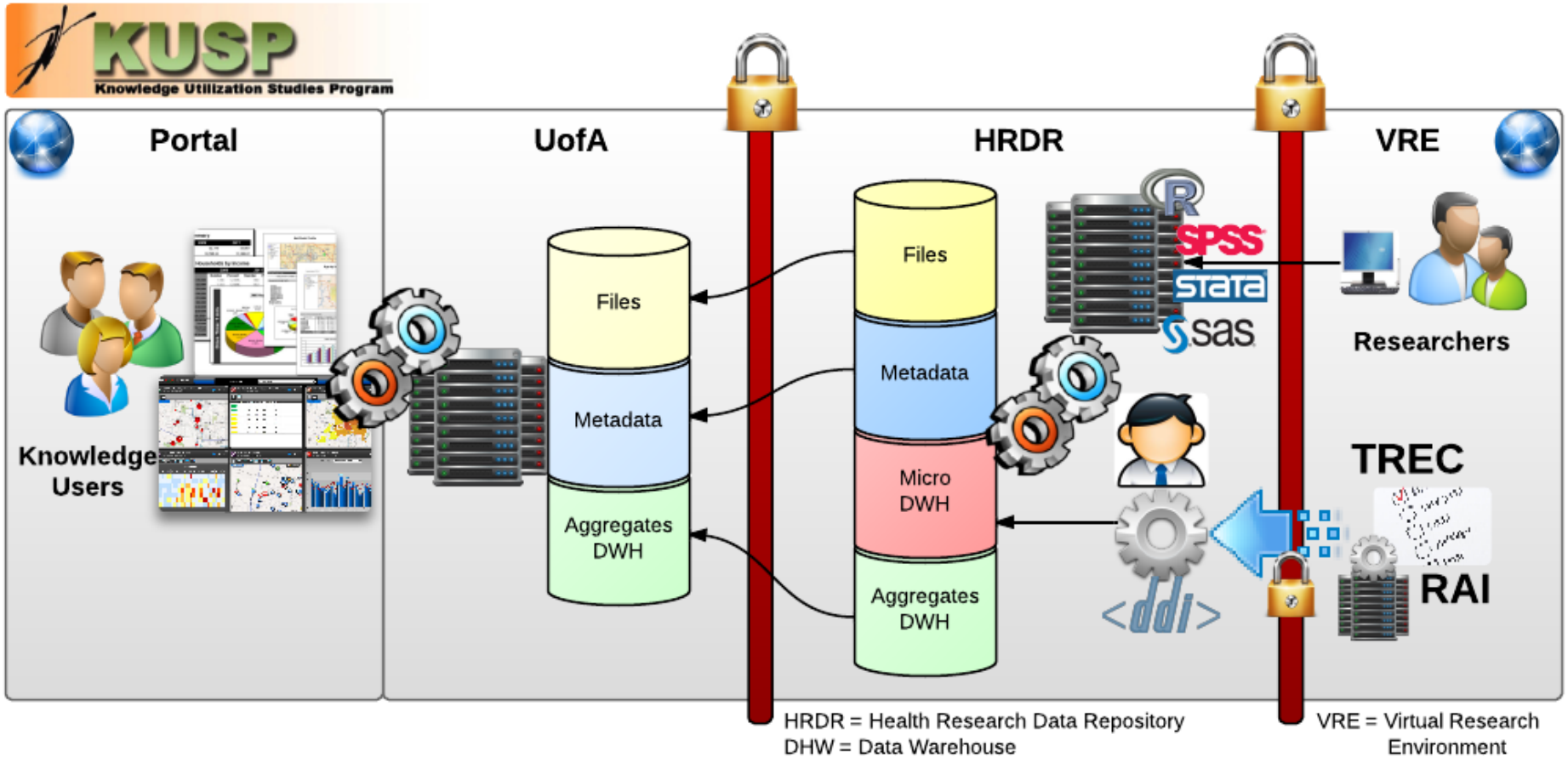
	Column	Name	Value	Acceptable Values
	AA1	Unique Registration Identifier	99999999999999999999	
	A2a	Unit	ZZZZZ9	
	A2b	Room #	ZZZZ999	
	AA2	Sex	Z	F,M,O
	AA3a	Birth Date (to be entered in YYYYMMDD format)	20100101	
	AA5a	Health Card Number	999999999	
	AA6	Facility number	86224	
	AA8	Reason for Assessment	05	00,01,02,03,04,05,06,07,08,09,10,17,18

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





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

