

# Lessons learned:

Supporting major longitudinal research projects  
with metadata, &  
using metadata for rules-based data ingestion  
in a secure data repository

Shane McChesney  
[shane@nooro.com](mailto:shane@nooro.com)



NADDI 2016 Edmonton  
April 7<sup>th</sup> 2016

# Show, don't tell

Data Sharing and Management Snafu in 3 Short Acts

<https://www.youtube.com/watch?v=N2zK3sAtr-4>

- Karen Hanson, Alisa Surkis and Karen Yacobucci of NYU  
Health Sciences Library

Illustrates the structured data documentation problem in a  
brilliant and compelling way.

# Core Findings

Researchers, stakeholders:  
get programmers into the room!

Free course in dealing with programmers:  
Bloomberg Special Double Issue  
*What Is Code?*

<http://www.bloomberg.com/graphics/2015-paul-ford-what-is-code/>

# Core Findings

Programmers:

**keep** researchers & stakeholders  
in the room...

...or at least on the call.

Get reality checks daily / weekly... okay, biweekly.  
(But at least twice as often as anyone thinks is necessary.)

**Always be working on  
The Most Important Thing To Do Next**

# A Few Quotes

Helpful perspectives for metadata implementors, interspersed between my lessons learned.

Perfect  
is the  
enemy  
of good.

# Timelines

Anything over 2 weeks is too long.

Software projects should deliver running code on a continual basis.

# Timelines

Academic / government projects  
seem to take 2-5x as long as business /  
market research projects.

Sometimes they have to... sometimes?



# Teams Change

Plan for that:  
document, document, document.

Remember:  
“It's not about you.”

# Names are not codes, codes are not UUIDs

Truely unique identifiers should not be  
overthought, let the computer do it:

Name: Alberta

Code: AB

UUID: 170ce6e4-9f18-4432-9b6d-74b94ca8f883

Put UUIDs on everything, you'll be glad you did.

Corrolary: autoincrementing integers suck.



# Pain Points

“Doctor, it hurts when I do this...”

- Ralph Furley, *Three's Company*

# “Then don't do that.”

If creating metadata is a manual task after field has closed, you are doing it wrong.

It should be a condition of funding for primary researchers to provide metadata with their data.

The Cure:

Start with metadata

Run Data Collection Instruments  
Directly From Metadata

# Metadata-Driven Survey Players

- 2 types of online surveys
  - Self-completed web surveys
  - In-person interviews using iPads



surveymedia

```
{
  "editor_class": "basic_options",
  "name": "VAR014",
  "options": [
    {
      "text": {
        "en": "Yes"
      },
      "value": "1",
      "weight": 50
    },
    {
      "text": {
        "en": "No"
      },
      "value": "2",
      "weight": 50
    }
  ],
  "ordering": "default",
  "orientation": "horizontal_grid",
  "required": true,
  "text": {
    "en": "Do you provide at least one third of
your long-term care services (i.e., <i>at least 6
days a month</i>) in <b>{{facilityname}}</b>?"
  },
  "type": "option"
},
```

# Metadata-Driven Survey Players



Do you provide at least one third of your long-term care services (i.e., *at least 6 days a month*) in **Valinor Retirement Residence (Test Site)**?

Yes

No

8%

← Previous

Contents

Next →

(multilingual, offline capable, mobile-ready)

# On Metadata Complexity



“It has been said that democracy is the worst form of government... except for all the others that have been tried.”

- Sir Winston Churchill

[https://en.wikiquote.org/wiki/Winston\\_Churchill](https://en.wikiquote.org/wiki/Winston_Churchill)



# Metadata

management can be complex,  
but it beats the alternative.

The good news:  
The tools are always improving.

# VALHUE

## Value and Limitations in Hospital Utilization and Expenditures

"...one of the largest systematic collections of Patient-Reported Outcome Measures (PROMs) in Canada, and aims to evaluate the changes in patients' health status pre- and post- elective surgery."



## Patient Reported Outcomes

[Home](#) [PROMs Blog](#) [What are PROMs?](#) ▼ [Who is working on PROMs?](#) ▼ [Our Research](#) [Resources](#) [About Us](#) ▼» [Home](#) » [VALHUE](#)

# VALHUE

## Value and Limitations in Hospital Utilization and Expenditure (VALHUE) Project Summary

### Significance and Policy Context

There is little understanding of whether shorter wait times are associated with better clinical outcomes, and whether policy-linked initiatives to increase access to surgery are associated with consistent changes in health improvement. The literature regarding the effect of wait times on health is significantly under-developed. VALHUE aims to address this gap in the literature.

This study is intended to inform clinicians, Vancouver Coastal Health (VCH) decision makers, and policymakers about the nature of relationships between wait times and trajectory of change in health status, pain and depression. We expect that the results will be immediately relevant in BC and other provinces implementing changes to prioritizing health expenditures. This study is the first in Canada to examine the trajectory of patients' health while on the wait list, and also to differentiate the impact of surgical care by different types of conditions.

Funded by the Canadian Institutes for Health Research (CIHR), VALHUE is a significant PROMS project within VCH that began in January 2012. Current funding for this project ends September 2014.

### Objective and Specific Aims

This joint project with the University of British Columbia aims to quantify the association between time spent on the surgical waitlist and health status. This is to be evaluated by measuring changes in health status for those patients undergoing selected surgical procedures.

The VALHUE project uses validated Patient Reported Outcome instruments to measure change in general health status, acute and chronic pain, and depression while waiting for care and after surgical treatment. This study will fill an important gap in the current understanding on the effects that changes in wait times and surgical volumes have on patients' health.

VALHUE has two specific aims:

**Aim 1:** Surveying patients twice prior to surgery, we will test for changes in patients' health while on the surgical wait list.

**Aim 2:** Surveying patients pre- and post-surgery, we will test for changes in patients' health as a result of wait times.

### Tags

[population health](#) [CAHSPR](#) [arthroplasty](#) [hip replacement](#) [ISOQOL](#) [UBC](#) [surgery](#) [Canada](#) [generic instruments](#) [UK](#) [health status](#) [cholecystectomy](#) [instrument selection](#) [VALHUE](#) [CHSPR](#) [Vancouver](#) [BC](#) [generic instruments](#) [pediatrics](#) [PROMs](#) [cataracts](#) [elective surgical procedures](#) [disectomy](#) [chronic care](#) [patient experience](#) [resio](#) [HRQoL](#) [knowledge](#) [mHealth](#) [self-management](#) [condition-specific](#) [instruments](#) [pain](#) [symptom severity](#) [VCH](#) [hysterectomy](#) [Electronic PROs](#) [PREMs](#) [burden](#) [condition-specific instruments](#) [Australia](#) [conference](#) [prostatectomy](#) [patient satisfaction](#) [uveal melanoma](#) [overactive bladder](#)



## Evidence and Perspectives on Funding Healthcare in Canada

[Home](#) [Key Issues](#) ▼ [About Us](#) ▼ [Our Research](#) ▼ [Newsletter](#) ▼ [Contact us](#)[» Home](#) [» About us](#) [\\*Old\\*](#) [» Jason Sutherland](#)

# Jason Sutherland

**PhD, Statistics, Simon Fraser University, 2003****MSc, Statistics, Simon Fraser University, 1997****BA, Mathematics, University of British Columbia, 1992**

Jason Sutherland joined UBC's School of Population and Public Health and the UBC Centre for Health Services and Policy Research in 2009. Dr. Sutherland is a highly regarded and recognized leader in researching the link between financial incentives and hospital behaviours and is a Michael Smith Foundation for Health Research Scholar.

Dr. Sutherland is currently leading a CIHR-funded study of the effects of changing the incentives of hospital funding. Dr. Sutherland is also leading another CIHR-funded study investigating the links between funding policy and patient reported health outcomes (PROMS). He is currently an associate editor of Health Policy, an international journal for health policy research.

In collaboration with the Ontario Ministry of Health and Long-Term Care, Jason has led efforts to develop and implement a case mix measurement tool for inpatient rehabilitation, known as called Rehabilitation Patient Groups. Affecting a broad array of rehabilitation patients in Ontario, this health system measurement tool is applied to hospitals that treat patients with stroke, joint replacement and brain injury. This case mix measurement tool is now applied by the Canadian Institute for Health Information to hospital based rehabilitation.

Before joining UBC, Jason was an Assistant Professor at Dartmouth College within The Dartmouth Institute for Health Policy and Clinical Practice (2007-2009). From 2004-2007, Jason was an Assistant Professor in the Department of Medicine, Division of Biostatistics at Indiana University. At this time, he was also an Assistant Research Scientist at the Center of Excellence on Implementing Evidence Based Practice, Roudebush Veteran's Affairs Medical Center in Indianapolis, Indiana.

The VALHUE questionnaire is a combination of almost 30 validated health status instruments.

The appropriate path for each patient through the overall questionnaire varies based on the patient's particular diagnosis codes and the surgical procedures (ORMIS PX codes) that the patient is scheduled for.

# TREC

Translating Research In Elder Care





## FACULTY OF NURSING

# Translating Research in Elder Care

### ► Home

### Projects

### Team

### Publications

### Resources

### News

### Album

### Contact



## TREC AGM June 4-5, 2015

Royal Mayfair, Edmonton

More info...



Translating Research in Elder Care (TREC) is a research program focused on developing solutions for improving the quality of care provided to nursing home residents, enriching the work life of their caregivers, and enhancing system efficiency.



*"A life that has been lived  
should not be valued less than  
a life that's going to be lived."*  
-Carole Estabrooks, Scientific  
Director

## FACULTY OF NURSING

# Translating Research in Elder Care



### Home

### ▼ Projects

#### ► TREC Measurement System

#### SCOPE

#### Advice Seeking Networks in LTC

#### Trajectories

### Team

### Publications

### Resources

### News

### Album

### Contact

[Home](#) > [Projects](#) > TREC Measurement System

## TREC MEASUREMENT SYSTEM

Favourable work environments lead to increased use of best practices, higher quality of care and outcomes for older people, improved staff health and well-being in long term care facilities.

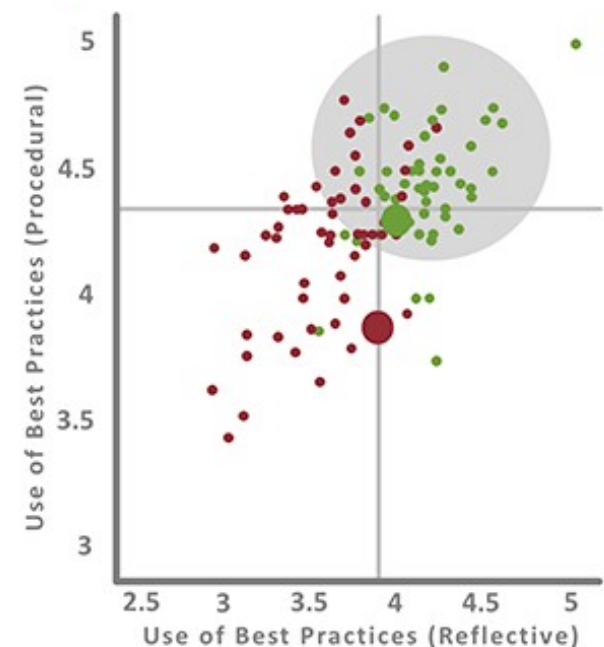
### Why is it important?

We know from TREC 1.0 that context influences the use of best practices and provider outcomes (e.g., job satisfaction, burnout) which in turn can affect resident outcomes. It is possible to influence context and we want to learn more about which factors lead to positive changes in nursing home care.

### What does the TREC Measurement System study involve?

The TREC Measurement System (TMS) study is a longitudinal study that explores the strength of linkages between work environments, best practice use and resident outcomes using staff survey data and RAI-MDS 2.0 data. The study will take place in 90 nursing homes across British Columbia, Alberta and Manitoba. Recruitment of sites will ensure representation across these locations, and that facilities of different sizes and owner-operator models (public, voluntary and private) are included.

Data will be collected from facilities in three waves spaced approximately 15 months apart. It will include surveys on facilities and units, and from regulated and unregulated staff. Participating facilities will play an important role in building the first longitudinal database of which we are aware of in Canada or elsewhere. We will work closely with Facility Managers, Administrator and Directors of Care to ensure that these data are collected with minimal impact on resident care.





## FACULTY OF NURSING

## Knowledge Utilization Studies Program



## About

## People

## Research

## Publications

## ▼ Resources

## KUSP FYI

## ▶ Health Research Data Repository

## Top Ten KT Readings

## KU Resource Guide

## EQUATOR

## Opportunities

## Partners

## Knowledge Utilization Colloquia

## Alberta Context Tool

[Home](#) > [Resources](#) > Health Research Data Repository

## Health Research Data Repository

Based within the Faculty of Nursing at the University of Alberta, the Health Research Data Repository (HRDR) is a secure and confidential virtual research environment (VRE) created to support health related research projects, to support collaboration across research disciplines, to house health research data and meta-data throughout their life-cycle, and to promote the secondary use and re-purposing of health research data. The HRDR also has a mandate to promote and offer educational opportunities regarding research data management best practices. The mission of the HRDR is to stimulate both qualitative and quantitative health related research while building a collaborative culture of respect relating to data management and confidentiality.

KUSP currently houses research data for a number of its projects including TREC and SCOPE. We are committed to have our future data accessible in the HRDR's secure environment available to other researchers.

## Contact Us

**James Doiron**  
**Manager, HRDR**

5-174 Edmonton Clinic Health  
Academy (ECHA)  
University of Alberta  
Edmonton AB T6G 1C9

**Phone:** 780-248-1653

**Email:** [james.doiron@ualberta.ca](mailto:james.doiron@ualberta.ca)



# Automating Data Ingestion

Start With Metadata

# Planning



“No battle plan ever survived contact with the enemy.”

- Field Marshall Helmuth Karl Bernhard Graf von Moltke  
[https://en.m.wikiquote.org/wiki/Helmuth\\_von\\_Moltke\\_the\\_Elder](https://en.m.wikiquote.org/wiki/Helmuth_von_Moltke_the_Elder)

# Time & Budget

...are usually that enemy.

# Planning



“Everybody has a plan... until they get punched in the face.”

- Mike Tyson

<http://goodmenproject.com/business-ethics-2/everybody-has-a-plan-until-they-get-punched-in-the-face/>

# Scope Creep & Complexity

...can knock out the best laid plans.

# TREC Measurement System Data Sources

- I: In-Person Interviews (CAPI)
  - Health Care Aides
  - Long-Term Care facility directors & other management for overall facility and care unit profiles
- trained interviewers in a carefully managed and scheduled set of on-site interviews
- II: Online Surveys
  - Nurses
  - Unit Managers
  - Allied Health Professionals
  - Practice Specialists
- controlled in-person distribution of unique links and information packages

- **III: Non-Survey Data:  
RAI-MDS 2.0 Patient Assessments**



## Speaking the same language for high quality care

- | [Getting Started](#)
- | [Training](#)
- | [InterRAI Worldwide](#)
- | [Bibliography](#)
- | [Newsletter](#)
- | [OECD Report](#)

## Events

### **The 2016 WORLD interRAI CONFERENCE**

April 11-14, 2016

Toronto, Ontario Canada

[www.worldinterrai.org](http://www.worldinterrai.org)



**interRAI is an international collaborative to improve the quality of life of vulnerable persons through a seamless comprehensive assessment system.**

Our consortium strives to promote evidence-informed clinical practice and policy decision making through the collection and interpretation of high-quality data about the characteristics and outcomes of persons served across a variety of health and social services settings. [Read More >>](#)





## Long-Term Care Facilities (LTCF)

The interRAI Long-Term Care Facilities Assessment System (interRAI LTCF) enables comprehensive, standardized evaluation of the needs, strengths, and preferences of persons receiving short-term post-acute care in skilled nursing facilities as well as persons living in chronic care and nursing home institutional settings.

The interRAI LTCF (or previous versions) is currently used in Canada, Europe (Belgium, England, Finland, France, Germany, Iceland, Italy, Netherlands, Norway, Spain, Sweden, Switzerland), Asia (Hong Kong, Korea, and Japan), and Pacific Rim (Australia, New Zealand).

# Admission, Annual, & Quarterly Patient Assessments

- RAI-MDS 2.0 records
  - Patient Assessments From ~90 Long Term Care homes
  - ~500 variables per assessment
  - 3 flavours (admission, annual, quarterly)
  - collected continuously from 2007-2020
  - administered by care providers
  - stored in several software systems
  - submitted to government quarterly
  - ~350K rows expected to be recieved in ? files at ?? times, in ??? formats from ??? distinct source systems
- by default, no care unit identifier in exports (open ended field)
- No existing metadata found...

## 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"/>	<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
	ESTIMATED	<input type="text"/>		

# 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"/>											

```
280 ▼ {
281     "name": "AA8",
282     "width": 2,
283     "text": "Primary reason for assessment",
284     "label": "Reason for assessment",
285     "start": 36,
286     "valid": "01-Admission assessment",
287 ▼   "values": {
288       "1": "Admission assessment (before d
289       "3": "Significant change in status a
290       "2": "Full annual assessment",
291       "4": "Significant correction of prio
292     },
293     "type": "99"
294   },
295 ▼ {
296     "name": "A9a",
297     "width": 1,
298     "text": "Responsibility/Legal Guardian
299     "label": "Legal guardian",
300     "start": 50,
301     "valid": "0, 1",
302 ▼   "values": {
303     "1": "Yes",
304     "0": "No",
305     "9": "Unknown"
306   },
307     "type": "9"
308   },
309 ▼ {
310     "name": "A9b",
311     "width": 1,
312     "text": "Responsibility/Legal Guardian -
313     "label": "Durable power of attorney/fina
314     "start": 51,
315     "valid": "0, 1",
316 ▼   "values": {
317     "1": "Yes",
318     "0": "No",
```

# Build Applications

(Automate the Automatable,  
Control the Uncontrolled Processes)

# Now We Can Identify & Reject Privacy Violations At The Point Of Ingestion...

RAI Management Tool

Upload

Uploads

Users

Instruments

Uploads / Uploaded File

Data Check

Failed: see report for details

Filename

\_scramble.csv

Whitelist




[whitelist report](#)

Uploaded

Unauthorized facility numbers exist: 6057

# Automate Stages of Data Scrubbing

## Stages

0	Raw		<a href="#">Explore</a>	<a href="#">Download</a>	
1	Columns renamed		<a href="#">Explore</a>	<a href="#">Download</a>	<a href="#">Generate</a>
2	Unit Reconciliation		<a href="#">Explore</a>	<a href="#">Download</a>	<a href="#">Generate</a>

[Metadata Checks](#)

[Full Report](#)

...and add more stages as they are reduced to practice.

“Our needs are  
unique...”

...just like everyone else's.

# Facility & Unit Reconciliation

## RAI AA6 Facility Aliases

As raw RAI records are imported from many sources, you will encounter various CIHI codes and other identifiers in the incoming data that are meant to refer to this facility. Enter those known "aliases" for this facility below, one per line, so that they are permanently preserved in project metadata. The automated RAI ingestion process will use them to link new incoming data to this facility.

TESHT  
TESST  
TEST1

Please enter a commit message to describe the change you have made.

Added new facility aliases found in source data today.

Cancel

OK

## Notes

Test facility.

SAVE

## Unit Definitions



The myth of the  
single codebook

# Research Stakeholders at all stages of the GSBPM

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			



# Benefits Of Using Metadata for Researchers / Investigators

- **Reproducibility**
  - during & after the project
- **Budget**
  - reduce manual labour
  - reduce rework
- **Productivity**
  - free resources for higher value uses
- **Discoverability**
  - online visibility = citations
- **Quality Control**
- **Documentation**

# Manage Change

## In Dimensional Data Over Time

# Tying Together Surveys & Patient Assessments in the HRDR

- **Who** carried out the research?
  - TREC interviewers
  - self completed
  - LTC staff doing patient assessments
- **When** did it happen?
  - survey data collected at multiple time points over 3 waves over 6 years.
  - RAI assessments done quarterly and on admission/re-admission
- **Where (what location)** does that data apply to?
  - Country
  - Province
  - Health Region
  - Facility
  - Unit
- All of these can change over time in surprising ways.

# Custom Data Management Tools To Manage These Dimensions

[Home](#) / [Facilities](#)

## Facilities

valinor

SEARCH

CLEAR SEARCH

ADD FACILITY

Results for **valinor**

Facility	Code	Units
<a href="#">Valinor Retirement Residence (Test)</a>	TST001	5

# Commit Messages on All Changes... plus versioning (document history) and undo

## RAI AA6 Facility Aliases

As raw RAI records are imported from many sources, you will encounter various CIHI codes and other identifiers in the incoming data that are meant to refer to this facility. Enter those known "aliases" for this facility below, one per line, so that they are permanently preserved in project metadata. The automated RAI ingestion process will use them to link new incoming data to this facility.

TESHT  
TESST  
TEST1

Please enter a commit message to describe the change you have made.

Added new facility aliases found in source data today.

Cancel

OK

## Notes

Test facility.

SAVE

## Unit Definitions



# Support

...the Field Team  
During Data Collection

Nightly email receipts to Interviewers, reports to Regional Coordinators, summary reports to the Data Managers

From Nooro TREC Support <support@trec-survey.ca>★  
Subject **TREC Regional Receipt Summary - XX**  
To shane@nooro.com★  
2015-03-15 02:30 AM  
Other Actions ▼



## **TREC Regional Receipt Summary**

**Instrument:** HCA CAPI Survey (Live)

**Region:** XX

**Regional Coordinator:** Shane McChesney

**Activity From:** 2015-03-14 12:00 AM

**Activity To:** 2015-03-15 12:00 AM

## **Uploads For Facility ABN\_008**

### **Completed Interviews**

<b>Interviewer</b>	<b>Interview ID</b>	<b>Mode</b>	<b>Most Recent Upload</b>
Nooro	b08435b0	TEST	2015-03-14 09:20 AM

# Managed, Versioned Recovery From User Error / Real-Life Variances



logout nooro

[Home](#) / Update Response Location

## Step 1 Search for Interview by UUID

Enter a response UUID or a portion of a response UUID to find survey responses that match.

ffee26b8-538c-3389-85fa-fced4e4169cf

GO!

# Step 2 Choose a Response

From the search results below, select the survey response you wish to update and click its "Continue" button.

Response ffee26b8-538c-3389-85fa-fced4e4169cf

CONTINUE

**Interviewer Name** Christine [REDACTED]

**Facility UUID** 2292dbed-8372-4285-83cf-bb3746fe2bdb

**Facility Name** [REDACTED]

**Unit UUID** c2bc6401-72f8-42a6-be40-4662430d2ff9

**Unit Name** Second Floor

**Interviewer Comments** Respondent was a little unclear about some of the questions even after I clarified using the prompts. No appropriate unit category available on screen so I selected [REDACTED]  
She worked on [REDACTED] (4th floor).

**System Comments**

**Created** 2014-09-22 14:54:34.071000-04:00

**Max Page Reached** 34

**Interview Location** Manitoba

**Interview Context** [REDACTED]

Interview location and context are from the mobile interviewer settings view.

### Step 3 Select New Location and Update Response

Selected Response: ffee26b8-538c-3389-85fa-fced4e4169cf

Select the target location and sublocation from the combined list below. This is the information that the response will be updated to. Enter all or part of the location details to filter the list.

Filter...

- » **Fourth Floor**
- » Second Floor
- » Third Floor

This system will automatically add a note logging the details of this change, including the old and new location details. Add a note describing your rationale for this change. For example:

Updated interview location as per 2015-04-03 email from RobJ.

Updated interview location based on interviewer comment and request from RC.

Once you have selected the desired location and entered your notes above, please click 'Apply' to save your changes, or click 'Cancel' to return to the beginning without making any changes.

## APPLY

CANCEL

# Commit messages! ->

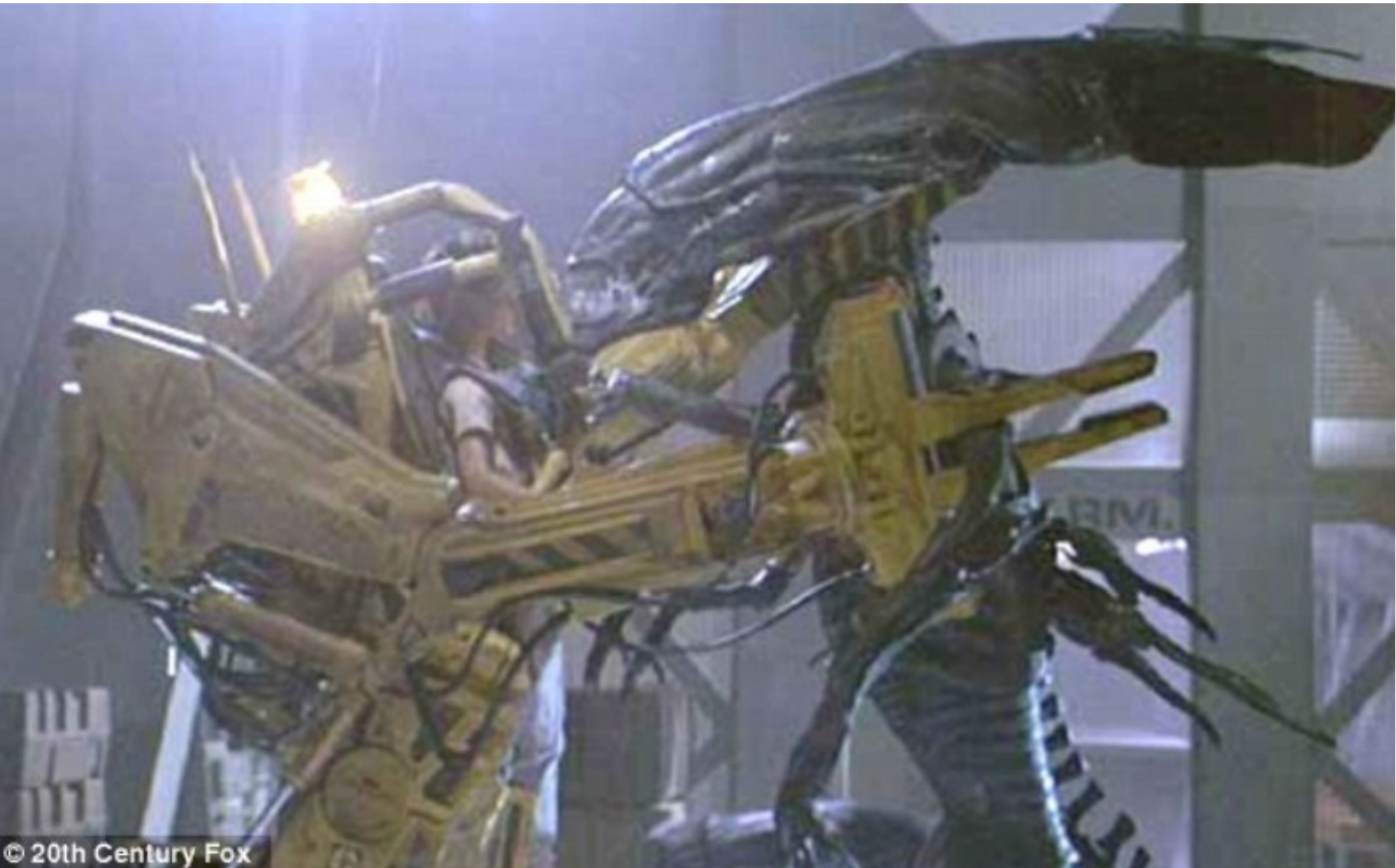
# Support

...the

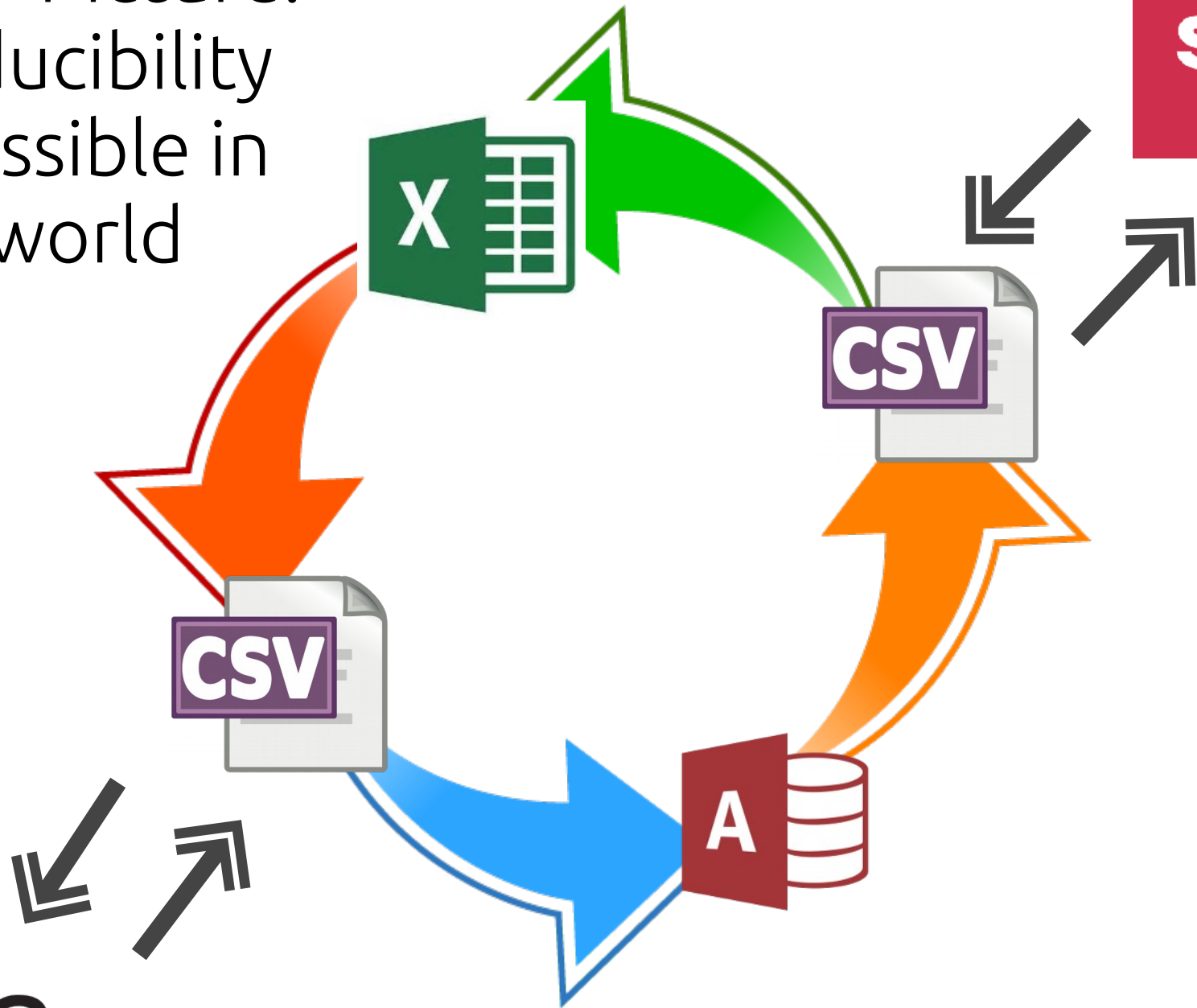
## Data Management Team



# Better Equip the Data Wranglers



“Before” Picture:  
Reproducibility  
is impossible in  
this world

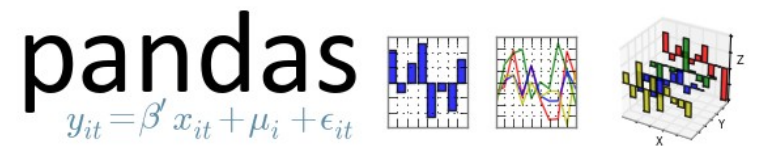
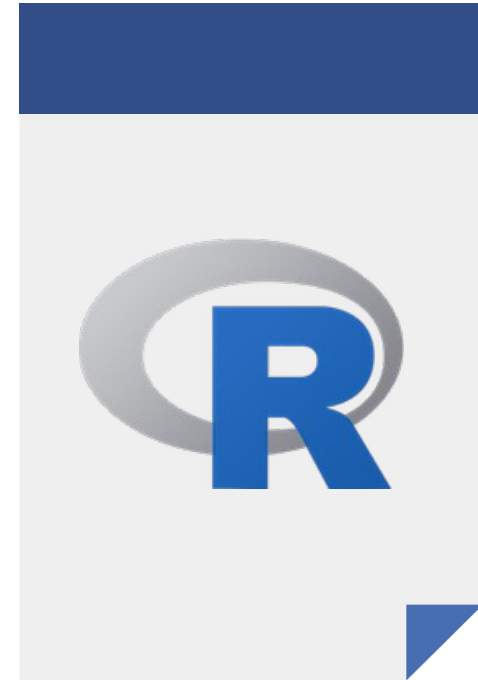




# Older “Windowed” Workflow / Tools



# Newer “Scriptable” Workflow / Tools



# Technical Debt

“Every line of code you write is a liability.”

- Trevor Morgan

# Technical Debt

“Yeah, but we'll make it up in volume!”

- Me? You?

# Programmers Helping Analysts: A Quick Scripting Case Study

By doing a few simple steps in Python and collaborating with analysts on the following R code, they were able to take it and replace 1000's of lines of manually coded SAS script with a metadata-driven loop over the questions to validate all fields for missing values.

```
# Prereqs: R, rjson
install.packages("rjson")
```

```
# Load the rjson library.
library(rjson)
```

```
# Read a copy of our survey definition
survey <- fromJSON(file='20150205_hca_survey.json')
```

```
# Pull a flat purpose-built JSON list produced from metadata
questions <- fromJSON(file='20150205_hca_questions_export.json')
```

```
# Open the latest data file
hcadata <- read.csv("20150205_0301_hca_all.csv", header=T)
```

```
dim(hcadata)
```

```
hcadata_frame <- data.frame(hcadata)
```

# Automate It

If you build on a foundation of Open Source scripting tools like Python and R, you can create workflows that can be run automatically:  
    nightly, for instance, or  
    whenever a new file comes in,  
freeing up your knowledge workers for actual, you know, *knowledge work*.

# Interoperability

<https://ddialliance.org>



- With proper metadata, processes built by one organization in one programming language can flow seamlessly into solutions built by another group in completely different language:
  - Metadata Technology is also assisting TREC with post-cleaning reporting tools.
  - Their tools in turn enable export to the broader ecosystem of DDI and traditional statistical software
- Metadata means old instruments can be brought back to life:
  - Nooro has been able to turn DDI-formatted metadata into functioning online surveys in under a day (including a 1995 US Election telephone survey)
  - NADDI 2015/6 & IASSIST 2015 conference evaluations handled in DDI by Colectica and fielded by Nooro

# Thank you & safe travels!

Shane McChesney  
[shane@nooro.com](mailto:shane@nooro.com)



NADDI 2016 Edmonton  
April 7<sup>th</sup> 2016