Good data in, great results out

A look at centralized Pure updating and what can be done because of it

Joan Concilio, Jonathan Rogalski and Robyn Reed
Penn State College of Medicine
Oct. 31, 2018

pure@pennstatehealth.psu.edu

#prcn2018
The team

• Joan Concilio
  - Multimedia (Web Development) Specialist, Marketing and Communications, Penn State College of Medicine

• Jonathan Rogalski
  - Programmer Analyst - Lead, IT Enterprise Information Management, Penn State Health

• Robyn Reed
  - Biomedical Informatics and Emerging Technologies Librarian; Associate Librarian, Harrell Health Sciences Library
Part 1: Good data in
The previous system

• Profiles RNS
  - Most biographical information manually added (by investigators), who often disregarded structured data fields
  - No editing, quality control or central oversight
  - Often required redundant data entry; ran the risk of having different versions of information in different places
  - Reliant on individual programs to request addition or removal of investigators
  - Data not integrated into other projects, not comparable across departments
  - Treated solely as a Research IT product, not as a marketing/branding tool
• Not a bad system, but a very hands-off implementation
The issue: Well, that’s hard to extract from

Probably didn’t need all three of these
The opportunity

• Use the launch of a new system to begin to build a gold-standard data source for researcher information
  - Store information once (cleanly and consistently) through data quality control
  - Use stored information in multiple projects, both alone and paired with other data sets
  - Implement brand standards and editorial quality control to improve College’s and researchers’ image
• Help faculty understand how to use system to its greatest benefit and why doing so is important
The strategy: Centralized updating

- TL;DR – Hands-on, not hands-off
- Specific to the College of Medicine, not all of Penn State
- Requests come via webform to a shared mailbox
- Anyone can make a request – investigators or their staff designees
- Team of four people reviews and updates
  - Updaters include people from University Libraries (knowledgeable about research output), Research Development (knowledgeable about institution and collaboration) and Marketing and Communications/Web Solutions (knowledgeable about brand, editorial and compliance standards)
- Additions and departures handled centrally in conjunction with Dean’s Office
- No more photos of people in bathrooms wearing bear hats
The strategy: #LoveTheForm (research.med.psu.edu/pure)
The strategy: Detailed hierarchy

- Includes both visible (below) and invisible groupings
Part 2: Great results out
The big project

- Penn State Cancer Institute website relaunch (Q1 2018)
  - University-wide Cancer Institute needed a new online presence to be competitive for application as a National Cancer Institute (NCI)-designated cancer center
  - Website development and content would be managed through College of Medicine
  - Two audiences: One clinical/patient-focused, one research-focused
  - Research content provided opportunity for proof-of-concept of the uses of clean Pure data
The implementation

- So we have this nice, clean, standardized Pure information… now what?
  - Combine it with identity management data
  - Combine it with clinical provider data
  - Combine it with clinical trials data
  - … and present it in a searchable, user-friendly way

- It’s not dramatic without a time crunch!
  - Pure went publicly live on Jan. 31, 2018 (though of course we were testing internally before then)
  - Cancer Institute site was slated for go-live Feb. 27, 2018

- Made possible because of the contributions of several teams
The data flow

- Data pulled from Pure via Web Service API (daily)
- Data from other sources - identity management system, clinical trials systems, provider credentialing system - pulled into same intermediate repository (daily)
- Other data correlated with Pure data elements (daily)
- Centralized data from intermediate repository exposed to enterprise systems via custom Web Service Application
- Publishing platform performs web service calls to retrieve consolidated data and display to end users
The end result: Integrated search
The end result: Detailed researcher profiles

RESEARCHER PROFILE

Valerie Brown, MD, PhD
Associate Professor, Department of Pediatrics

Scientific Program: Experimental Therapeutics
Disease Teams:
- Immunotherapy
- Pediatric Cancer

Research Interests
- Hematopoietic Stem Cell Transplantation
- Therapeutics
- Precursor Cell Lymphoblastic Leukemia-Lymphoma
- Pediatrics

- Transplants
- Bone Marrow
- Leukemia
- Hematopoietic Stem Cells
- Graft vs Host Disease
- Stiilimus
- Proteins
- T-Lymphocytes

Clinical Trials
- PSCI-16-073: A Phase 3, Randomized, Adaptive Study Comparing the Efficacy and Safety of Delirulide vs Best Supportive Care in the Prevention of Hematologic Malignancy in Adult and Pediatric Patients with Hematologic Malignancy.

View Details
The next steps

• Rolling out this implementation on other College of Medicine websites
  - Two publishing platforms
  - Replacing multiple manually-generated types of content
    o Person lists (departmental faculty)
    o Publication lists by department or group
• Deepening the data pull from Pure into enterprise data warehouse
  - At both College and University levels
• Developing a shared code repository for both public and private reporting based on Pure data
  - Could be leveraged elsewhere in University
• Creation of data governance group spanning multiple teams
The things you can do

• Start with a small but significant department, area or college
• Get leadership buy-in early
• Build an interdisciplinary team
• Communicate early, communicate often (and expect flak)
• Know the (actual) needs of the researchers
• Set standards and workflows and document them well
• Find a big win
• Tout your big win
Here’s to clean data!

pure@pennstatehealth.psu.edu

pure.psu.edu