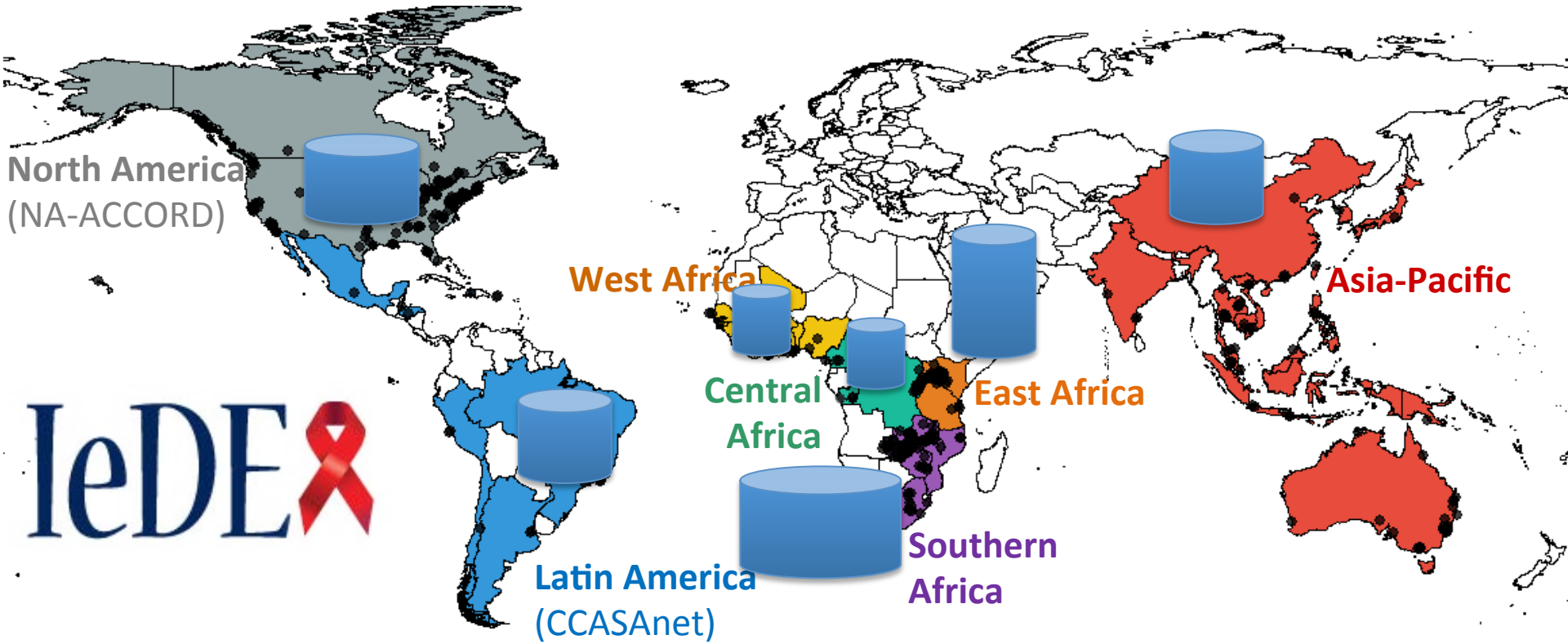
Data Curation in IeDEA

Stephany Duda, PhD
Department of Biomedical Informatics, Vanderbilt University
June 22, 2016

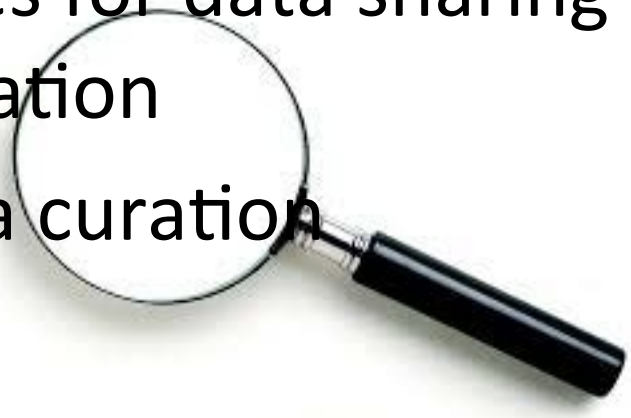
Start of IeDEA



Siloes of Observational HIV Data

IeDEA Targets Data

1. Convene a working group to review and discuss these topics
   → IeDEA Data Harmonization Working Group
2. Establish shared data standards
3. Identify collectively managed data
4. Improve workflow processes for data sharing and multi-regional collaboration
5. Create shared tools for data curation



Collaborative Data Curation

- Index of IeDEA Sites (tblCENTER) and basic site characteristics
- Shared survey-based data collection
  - Site Assessment
  - Surveys on site capacity and patterns of practice for TB, NCDs, adolescent transition and disclosure

IeDEA & REDCap
Research Electronic Data Capture

REDCap™
Research Electronic Data Capture

http://**projectredcap**.org/

Introduction | Software | Consortium Partners | Become a Partner | Video Resources | Citing REDCap | Library | REDCapCon

REDCap is a mature, secure web application for building and managing online surveys and databases. While REDCap can be used to collect virtually any type of data, it is specifically geared to support data capture for research studies. The REDCap Consortium is composed of **1,912 active institutional partners** in **102 countries** who utilize and support REDCap in various ways.

The REDCap application allows users to build and manage online surveys and databases quickly and securely, and is currently in production use or development build-status for more than **272,000 projects** with over **370,000 users** spanning numerous research focus areas across the consortium. To find out if your institution is already running REDCap, you will find contact information on the Consortium Partners page. Learn more about REDCap by watching a brief summary video (5 min) or overview (14 min).

Map of REDCap Consortium Partners

View fullscreen map

Recent publications using REDCap:

Influence of the WIC Program on Loss to Follow-up for Newborn Hearing Screening *Pediatrics. 2016 Jun 15. pii: e20154301.*

Beyond Traditional Newspaper Advertisement: Leveraging Facebook-Targeted Advertisement to Recruit Long-Term Smokers for Research *J Med Internet Res. 2016 Jun 15;18(6):e117. doi: 10.2196/jmir.5502.*

A Prospective Study of the Lymphedema and Fibrosis Continuum in Patients with Head and Neck Cancer *Lymphat Res Biol. 2016 Jun 15.*

A retrospective analysis of medical record use in e-consultations *J Telemed Telecare. 2016 Jun 14. pii: 1357633X16654607.*

Implementation of Developmental Screening Guidelines for Children with Congenital Heart Disease *J Pediatr. 2016 Jun 11. pii: S0022-3476(16)30237-2. doi: 10.1016/j.jpeds.2016.05.029.*

Survey on the Current State of Endotracheal Intubation Among the Critically Ill HEMAIR Investigators *J Intensive Care Med. 2016 Jun 13. pii: 0885066616654452.*

View all 2,476 articles

IeDEA has contributed to the global reach of REDCap.

is a registered trademark of DHHS.

Index of IeDEA Sites (tblCENTER)

IeDEA 2014 tblCENTER

Record Status Dashboard (all records)

Displayed below is a table listing all existing records/responses and their status for every data collection instrument (and if longitudinal, for every event). You may click any of the colored buttons in the table to open a new tab/window in your browser to view that record on that particular data collection instrument. Please note that if your form-level user privileges are restricted for certain data collection instruments, you will only be able to view those instruments, and if you belong to a Data Access Group, you will only be able to view records that belong to your group.

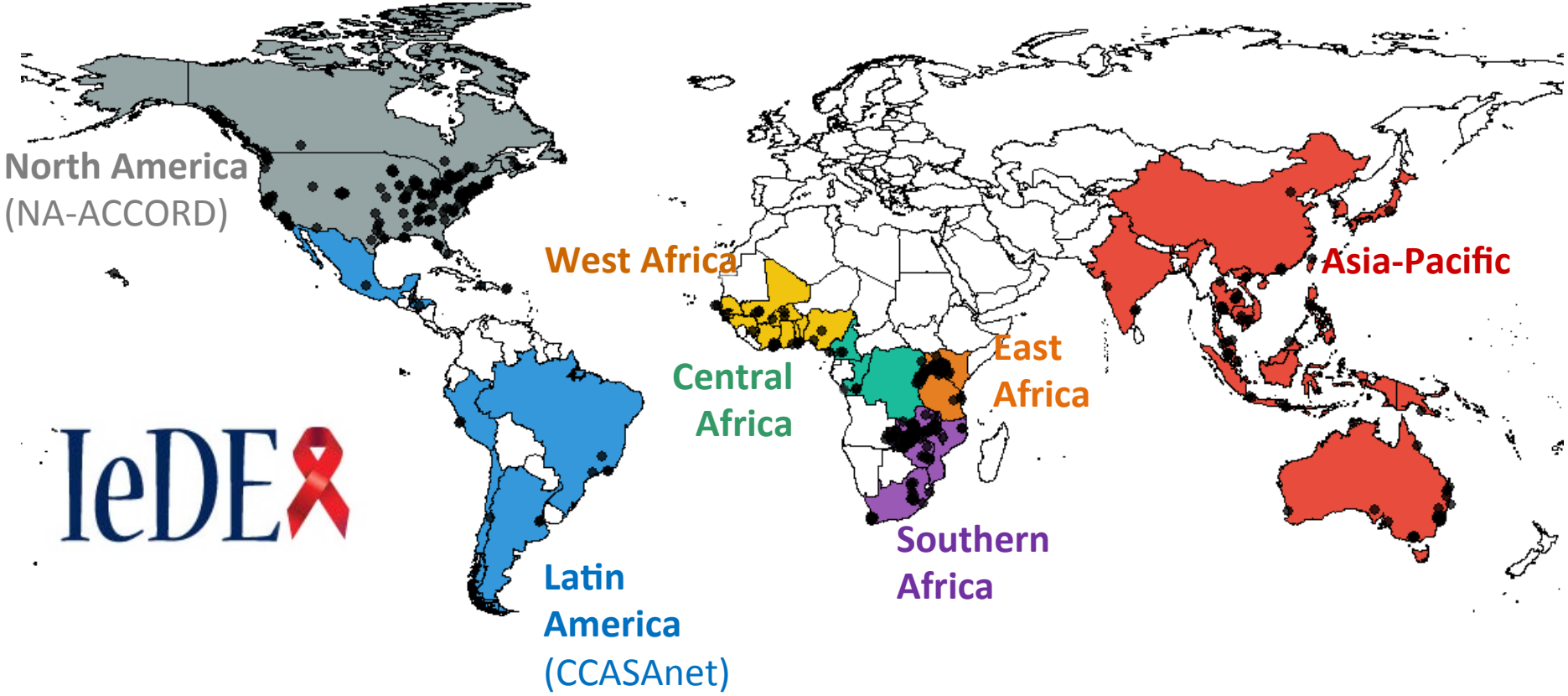
Legend for status icons:
Incomplete | Incomplete (no data saved) ?
Unverified
Complete

Displaying record "1" through "100" of **537** records

Displaying: Instrument status only | Lock status only | All status types

Record ID	tblCENTER
1 (Dr Ellis GMP)	
2 (Holdsworth House Medical Practice)	
3 (Northside Clinic)	
4 (Dept Clin Immunol, Royal Perth Hospital)	
5 (Prahran Market Clinic)	
6 (Holden Street Clinic)	
7 (SHAIDS (Lismore Sexual Health Service))	
8 (East Sydney Doctors)	
9 (Sexual Health Program, Cairns Base Hospital)	
10 (RPA Sexual Health)	

Map Generated from tblCENTER



Personal Open source Business

Scripts are available:
https://github.com/IeDEA

IeDEA

International Epidemiologic Databases to Evaluate AIDS

http://IeDEA.org/

Repositories People 1

Filters Find a repository...

iedea-des-base

1 0

iedea-des-base

Updated on Sep 3, 2014

qa-checks-r

R 0 1

See below for instructions and the `Downloads` for documentation of IeDEA-DES QA Checks

Updated on Apr 10, 2014

Personal Open source Business

Scripts are available:
https://github.com/IeDEA

IeDEA

International Epidemiologic Databases to Evaluate AIDS

http://IeDEA.org/

Repositories People 1

Filters Find a repository...

Shared bank of **data quality checking scripts** to assess the internal consistency of IeDEA-formatted data.

iedea-des-base

iedea-des-base

Updated on Sep 3, 2014

qa-checks-r

See below for instructions and the `Downloads` for documentation of IeDEA-DES QA Checks

Updated on Apr 10, 2014

Shared Vision for Additional Tools

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

INDIANA UNIVERSITY

amfAR
MAKING AIDS HISTORY

Instituts thématiques
Inserm
Institut national de la santé et de la recherche médicale

EINSTEIN
Albert Einstein College of Medicine

u^b
UNIVERSITÄT BERN

Kirby Institute

UNSW AUSTRALIA

NIH National Institute of Allergy and Infectious Diseases

CUNY SCHOOL OF PUBLIC HEALTH

JOHNS HOPKINS UNIVERSITY

ISPED Bordeaux school of public health
Institut de Santé Publique d'Epidémiologie et de Développement

VANDERBILT UNIVERSITY

New Group Initiative: Harmonist Toolkit

Build a standards-enabled toolkit for secure data exchange, data quality checking, and scientific project and portfolio management.

REDCap
Research Electronic Data Capture

Goals of Harmonist

1. Simplify structured data requests and secure data exchange
2. Improve data quality and data conformity with standards (by making our data quality tools easy to use)
3. Automate (some) components of workflow for IeDEA data and projects
4. Provide tools for scientific portfolio management
5. Track project and network milestones and metrics

Data Modules for Harmonist

R Server
(for data processing)

Data Quality Reports

Data Profile Reports

REDCap Projects

REDCap Plugins

Dashboards
- Region Leaders
- Data Managers
- Researchers
- Network Overview

Region-Specific Status Pages/To-Do Lists
Email Notifications
Metrics and Reports

Person Database

Concept Sheets

Data Requests

Data Submissions

Common Data Model (Computable Representation)

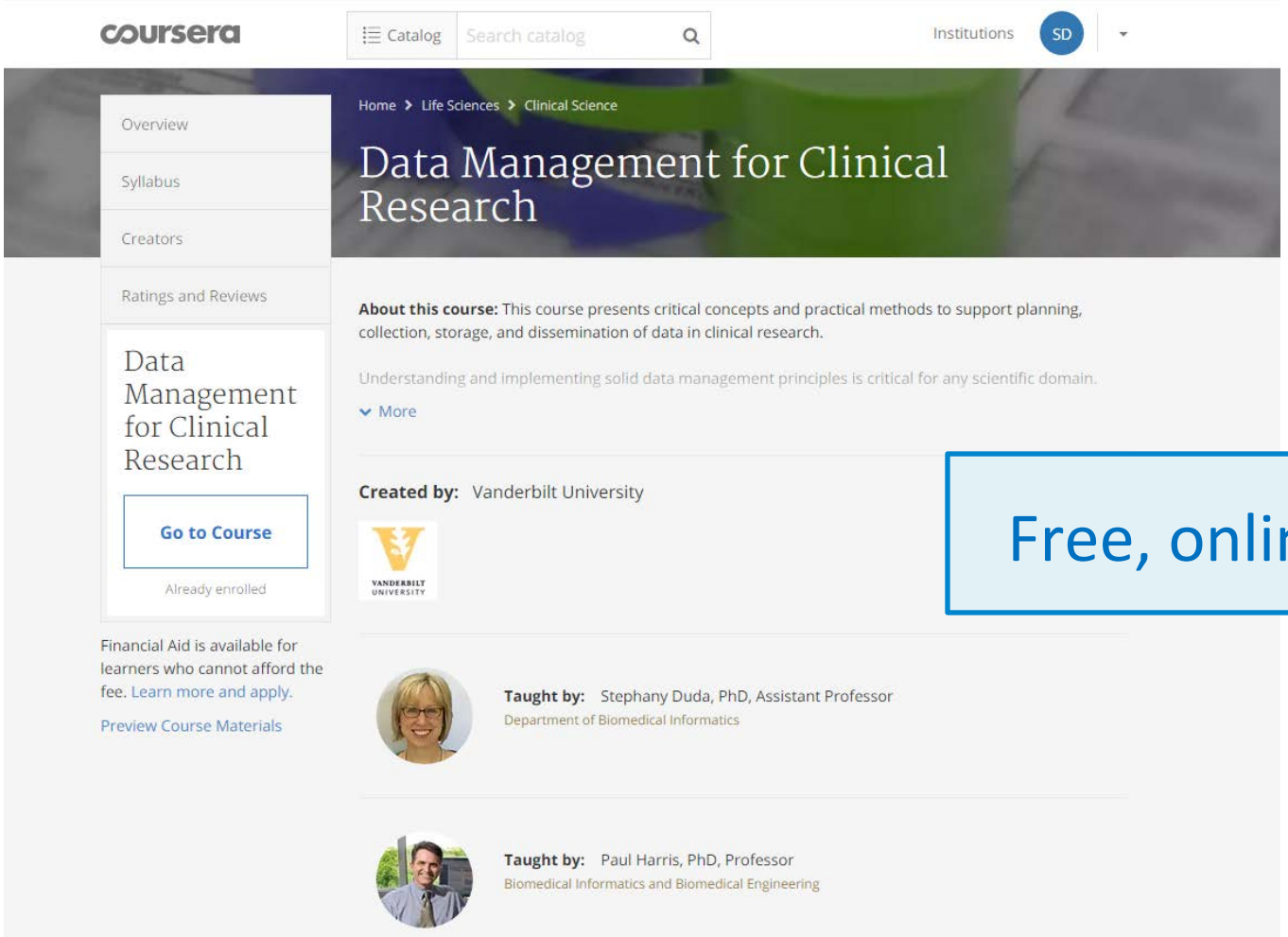
Extra Variables

Publications

Sharing is Caring!

- IeDEA actively shares resources, tools, and lessons learned
  - Data standards and code available online
  - Methods and tools presented at HIV and informatics conferences
  - Collaborations with partner networks (e.g., EuroCoord)
- We provide educational opportunities for others

Sharing our Lessons Learned: Coursera



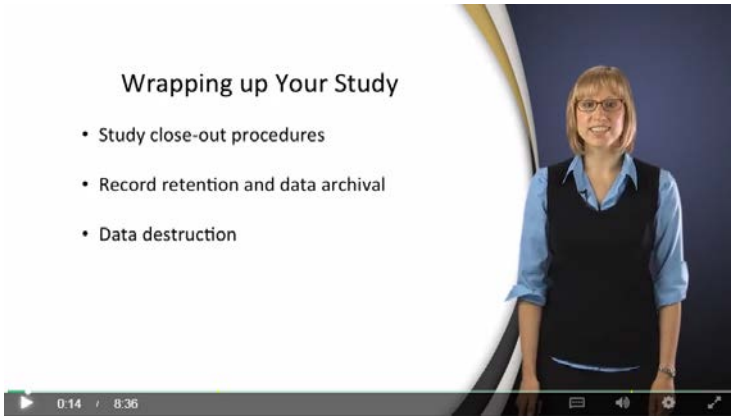
Free, online MOOC

https://www.coursera.org/learn/clinical-data-management/

Coursera Course Topics

Course Topics

- Principles of clinical research data management
- Data sources
- Electronic data capture
- E-Forms implementation
- Regulatory concerns (DSMB, HIPAA, FISMA, 21 CFR Part 11, IRB, etc.)
- Data quality
- Study workflow

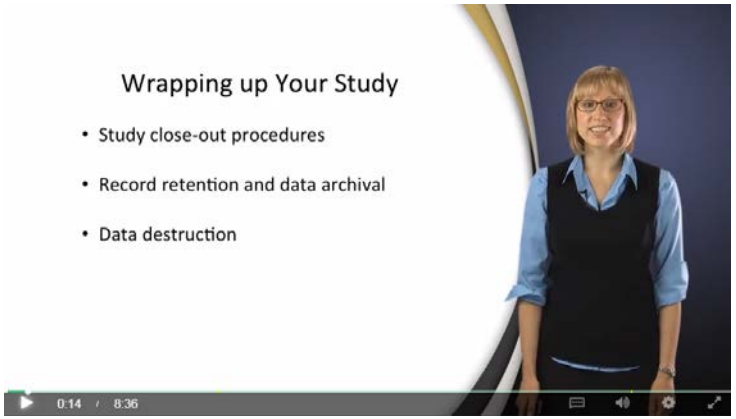


Many IeDEA Concepts

Coursera Course Topics

Course Topics

- Principles of clinical research data management
- Data sources
- Electronic data capture
- E-Forms implementation
- Regulatory concerns (DSMB, HIPAA, FISMA, 21 CFR Part 11, IRB, etc.)
- Data quality
- Study workflow



Student Feedback:
Add more international topics?

Coursera and IeDEA

5-week course in 2014

+ additional international content with IeDEA collaborators

6-week course in 2016

Since May 2016 Launch:

Total Visitors	Active Learners
4,736	1,273

IeDEA Targets Data

1. Convene a working group to review and discuss these topics
   → IeDEA Data Harmonization WG
2. Establish shared data standards
3. Collectively manage key data
4. Improve workflow processes for data sharing and multi-regional collaboration
5. Create and share tools for data curation

Continue!