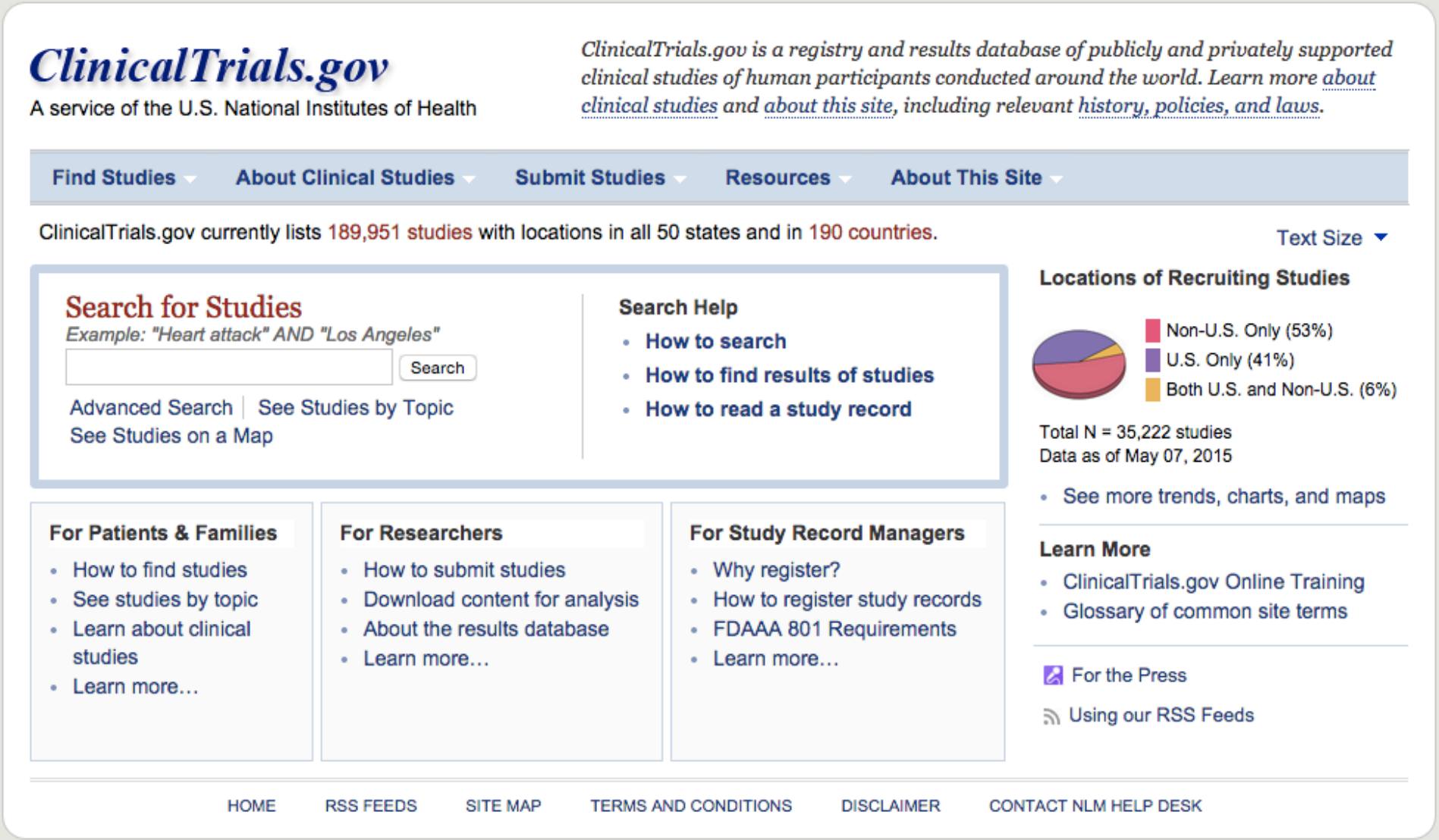
DiscoverCT

Jason Ost & Colin Gerber

What is a clinical trial registry?

- 1997: FDA Modernization Act
  - Creation of registry for all trials associated with investigational new drug (IND) applications
- 2000: ClinicalTrials.gov launched
- 2005: Int'l Cmte of Medical Journal Editors
  - Prospective registration of interventional trials as precondition to publication
- 2007: FDA Amendments Act
  - Mandated registration of all Phase II-IV trials of drugs, biologics, and devices regulated by the FDA



Almost 190,000 registered trials

112 million page visits per month

Majority of database is free text, unstructured

852 studies found for: bladder cancer

Modify this search | How to Use Search Results

List | By Topic | On a Map | Search Details

+ Show Display Options

Download | Subscribe to RSS

☐ Include only open studies ☐ Exclude studies with Unknown status

Rank	Status	Study
1	Recruiting	**Biomarker Identification for Bladder Cancer Patients** **Condition:** Bladder Cancer **Intervention:** Procedure: Sample Collection
2	Recruiting	**Fluorescence Cystoscopy and Optimized MMC in Recurrent Bladder Cancer (FinnBladder 9)** **Condition:** Bladder Cancer **Interventions:** Procedure: white light TUR-BT; Procedure: blue light TUR-BT; Drug: optimized MMC; Drug: single immediate chemotherapy instillation
3	Completed	**Erlotinib and Green Tea Extract (Polyphenon® E) in Preventing Cancer Recurrence in Former Smokers Who Have Undergone Surgery for Bladder Cancer** **Condition:** Bladder Cancer **Interventions:** Dietary Supplement: Polyphenon E; Drug: erlotinib hydrochloride; Other: Erlotinib placebo; Other: Polyphenon E

Fluorescence Cystoscopy and Optimized MMC in Recurrent Bladder Cancer (FinnBladder 9)

This study is currently recruiting participants. (see Contacts and Locations)

Verified January 2014 by Turku University Hospital

Sponsor:
Turku University Hospital

Collaborator:
Finnbladder

Information provided by (Responsible Party):
Peter Boström, Turku University Hospital

ClinicalTrials.gov Identifier:
NCT01675219

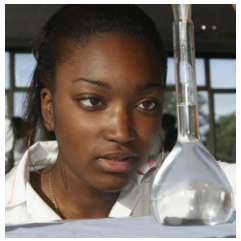
First received: August 25, 2012
Last updated: January 17, 2014
Last verified: January 2014
History of Changes

Purpose

Bladder cancer (BC), the second most common urological malignancy, is an important public health issue. One of the main challenges in the treatment of bladder cancer if the prevention of recurrences of non-invasive tumors, which is also associated with significant costs.

The current study will investigate optimal treatment of patients with bladder cancer with high risk of tumor recurrence but low risk of progression. The main interest is comparison of photodynamic (PDD) bladder tumor resection (TUR-BR)to traditional TUR-BT. Also the efficacy of adjuvant optimized mitomycin-C is compared to patients with no adjuvant treatment.

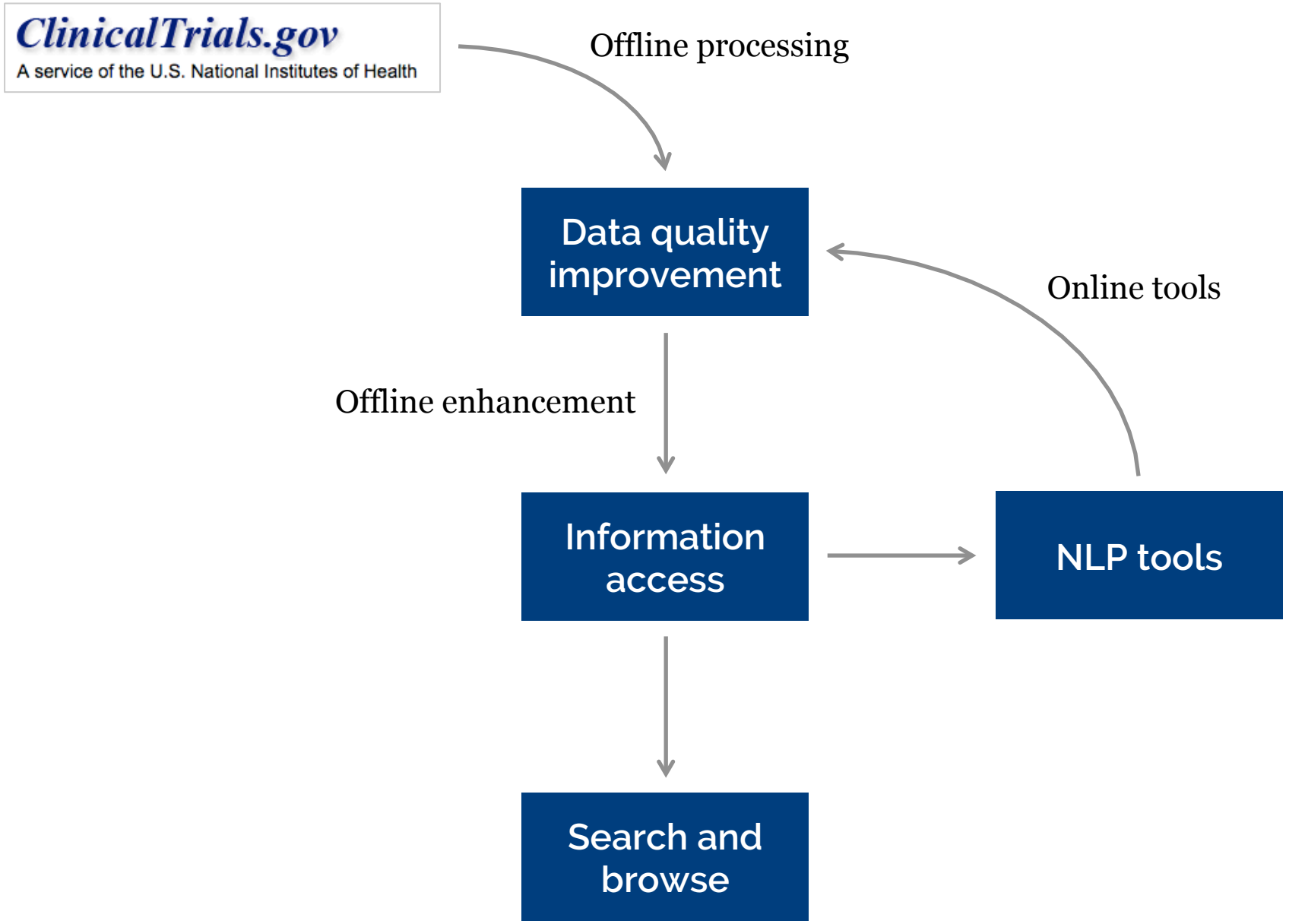
Our users



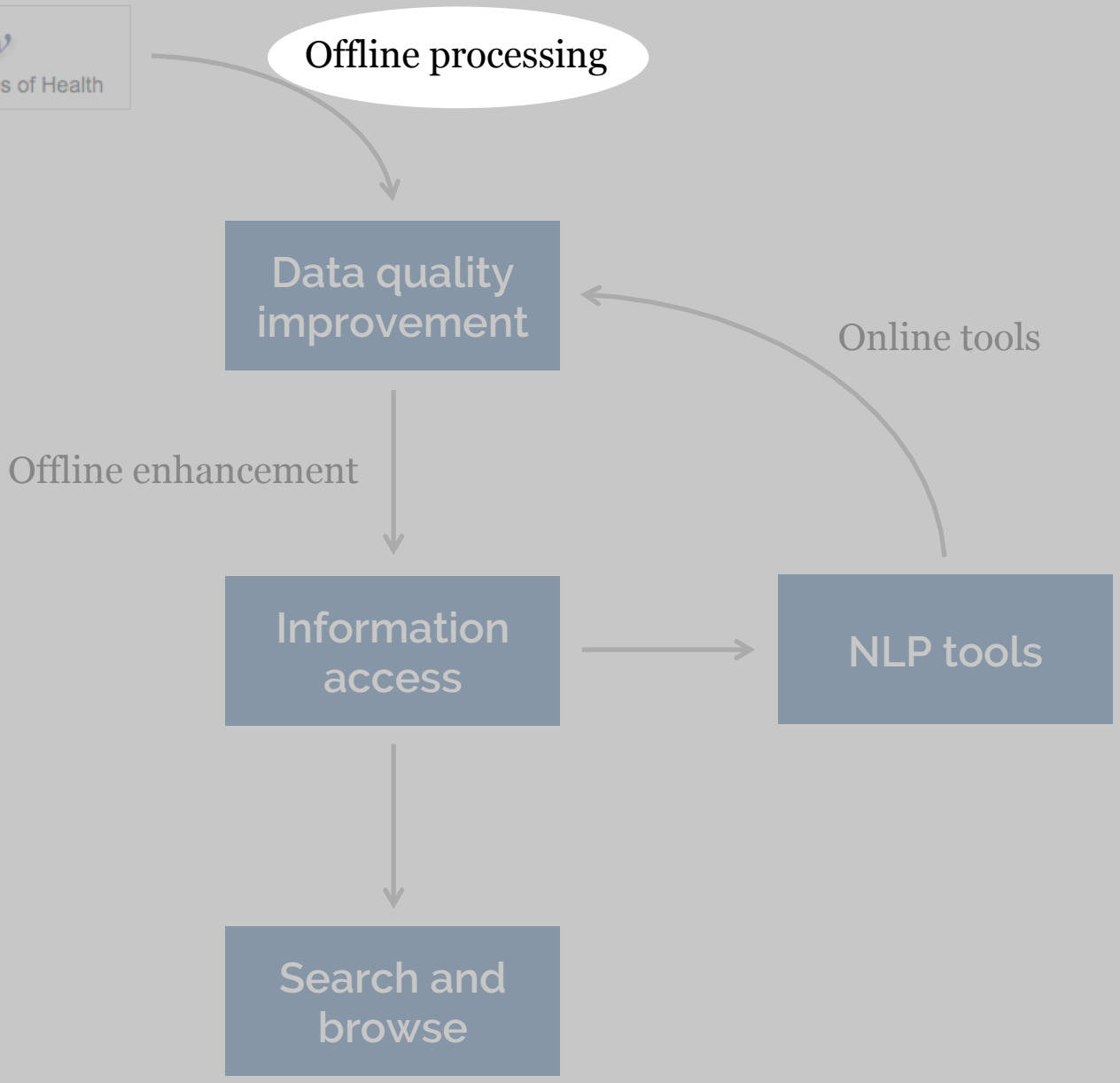
The Researcher needs an overview of the status of research in her field. She wants to know what others in her field have done and are doing so she can determine how she will contribute something new.

The Patient needs to find active trials in her area that offer potential treatments for her condition. She is not a medical expert and doesn't have the ability to sort through dozens of possible trials.

The Grant Program Officer and **The Policymaker** need high-quality information about research institutions and trial progress and outcomes in order to make decisions about funding.


ClinicalTrials.gov
A service of the U.S. National Institutes of Health
Offline processing
Data quality improvement
Online tools
Offline enhancement
Information access
NLP tools
Search and browse

ClinicalTrials.gov
A service of the U.S. National Institutes of Health

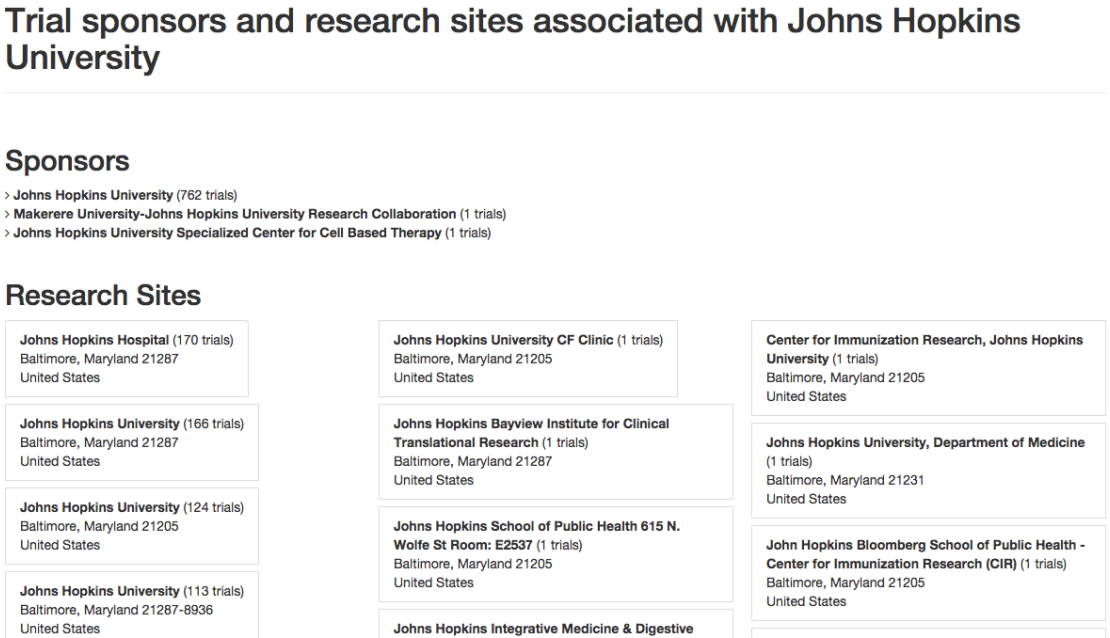


Offline processing

- Institution deduplication
- Medical Subject Headings (MeSH) assignment
- Trial and institution report cards

Institution deduplication

- Problem: many names and spellings for the same place
  - Johns Hopkins University
  - John's Hopkins University
  - John Hopkins
- Solution: semi-supervised machine learning approach to identify duplicate records based on text and location characteristics
- Results: identified 1,075 major institutions associated with 80% of trials (down from 540,000 unique entries in the registry)



Trial sponsors and research sites associated with Johns Hopkins University

Sponsors

› **Johns Hopkins University** (762 trials)
› **Makerere University-Johns Hopkins University Research Collaboration** (1 trials)
› **Johns Hopkins University Specialized Center for Cell Based Therapy** (1 trials)

Research Sites

Johns Hopkins Hospital (170 trials)
Baltimore, Maryland 21287
United States

Johns Hopkins University (166 trials)
Baltimore, Maryland 21287
United States

Johns Hopkins University (124 trials)
Baltimore, Maryland 21205
United States

Johns Hopkins University (113 trials)
Baltimore, Maryland 21287-8936
United States

Johns Hopkins University CF Clinic (1 trials)
Baltimore, Maryland 21205
United States

Johns Hopkins Bayview Institute for Clinical Translational Research (1 trials)
Baltimore, Maryland 21287
United States

Johns Hopkins School of Public Health 615 N. Wolfe St Room: E2537 (1 trials)
Baltimore, Maryland 21205
United States

Johns Hopkins Integrative Medicine & Digestive

Center for Immunization Research, Johns Hopkins University (1 trials)
Baltimore, Maryland 21205
United States

Johns Hopkins University, Department of Medicine (1 trials)
Baltimore, Maryland 21231
United States

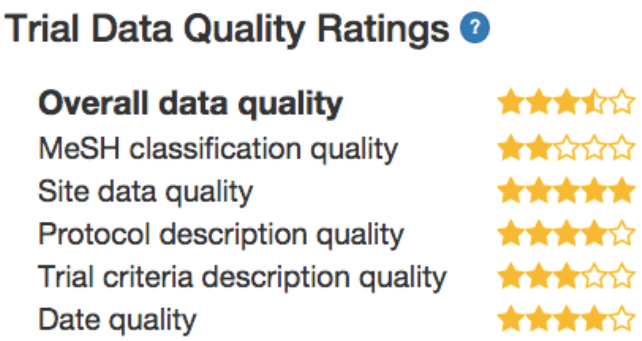
John Hopkins Bloomberg School of Public Health - Center for Immunization Research (CIR) (1 trials)
Baltimore, Maryland 21205
United States

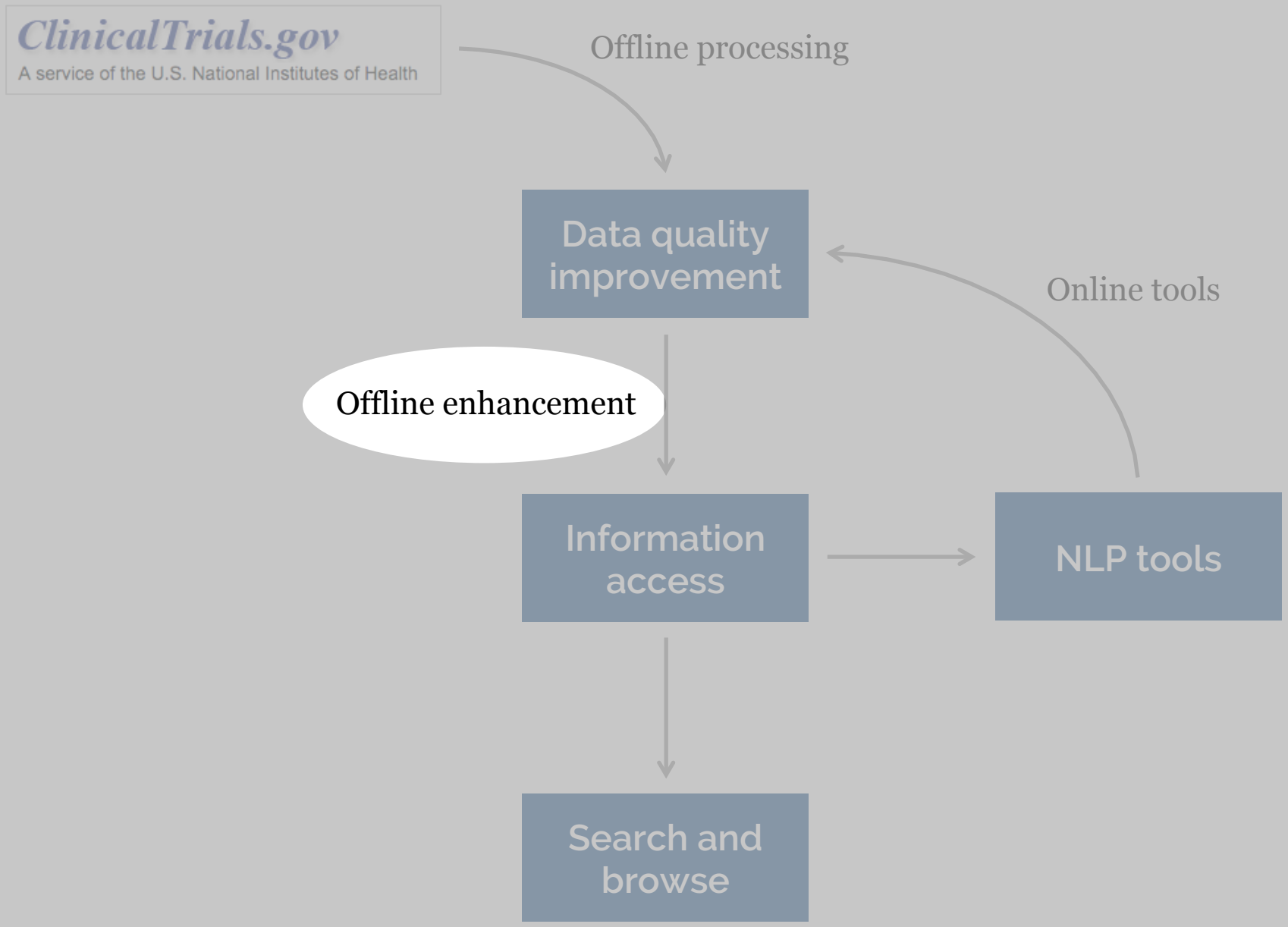
MeSH assignment

- Problem: 14% of trials have no MeSH terms, 31% have just one term
- Solution: develop KNN and maximum entropy models to predict MeSH terms based on similarity to well-described trials
- Results: assigned 159,000 terms to 106,000 trials

Report cards

- Problem: registry data can be poor quality due to lack of incentives for accuracy and completeness
- Solution: develop system to rate quality of trial data; improve transparency and encourage higher quality submissions
- Results: ratings for site location, date, protocol description, eligibility criteria, and MeSH classification quality



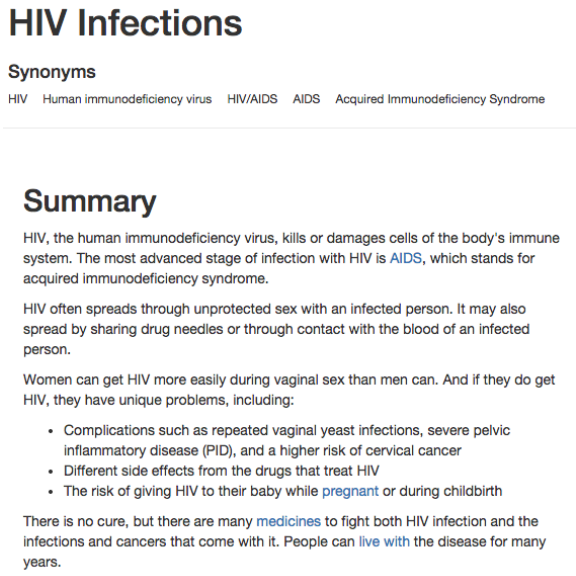

ClinicalTrials.gov
A service of the U.S. National Institutes of Health
Offline processing
Data quality improvement
Online tools
Offline enhancement
Information access
NLP tools
Search and browse

Offline enhancement

- Condition synonyms and descriptions
- Linking trials to PubMed
- Linking institutions to Freebase

Synonyms & Descriptions

- Problem: MeSH terms are scientific, not common; patients and other laypersons may benefit from less-advanced descriptions
- Solution: use Wikipedia and NLM's MedlinePlus health topics to get descriptions and related terms
- Results: added nearly 800 common synonyms and over 2,700 condition descriptions



HIV Infections

Synonyms

HIV Human immunodeficiency virus HIV/AIDS AIDS Acquired Immunodeficiency Syndrome

Summary

HIV, the human immunodeficiency virus, kills or damages cells of the body's immune system. The most advanced stage of infection with HIV is AIDS, which stands for acquired immunodeficiency syndrome.

HIV often spreads through unprotected sex with an infected person. It may also spread by sharing drug needles or through contact with the blood of an infected person.

Women can get HIV more easily during vaginal sex than men can. And if they do get HIV, they have unique problems, including:

- Complications such as repeated vaginal yeast infections, severe pelvic inflammatory disease (PID), and a higher risk of cervical cancer
- Different side effects from the drugs that treat HIV
- The risk of giving HIV to their baby while pregnant or during childbirth

There is no cure, but there are many medicines to fight both HIV infection and the infections and cancers that come with it. People can live with the disease for many years.

PubMed linkage

- Problem: publications are a key outcome for trials, yet only 18% of trials in the registry have linked publication references
- Solution: use trial investigator, date, and condition information to identify likely publications in PubMed
- Results: we identified another 28,000 publications associated with 6,000 trials

Publications associated with this trial

Linked by trial investigator

No publications linked by trial investigator

Other related publications

Likely related

- Hensing T, Chawla A, Batra R, Salgia R, et al. A personalized treatment for lung cancer: molecular pathways, targeted therapies, and genomic characterization. Adv. Exp. Med. Biol. 2014;799:85-117.

Probably related

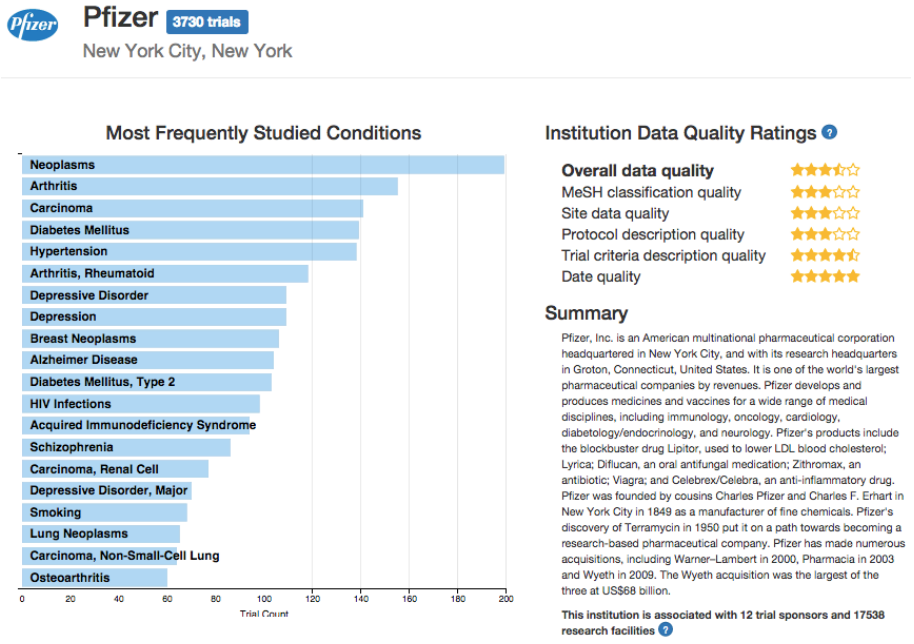
- Powell CA, Halmos B, Nana-Sinkam SP, et al. Update in lung cancer and mesothelioma 2012. Am. J. Respir. Crit. Care Med. 2013 Jul 15;188(2):157-66.

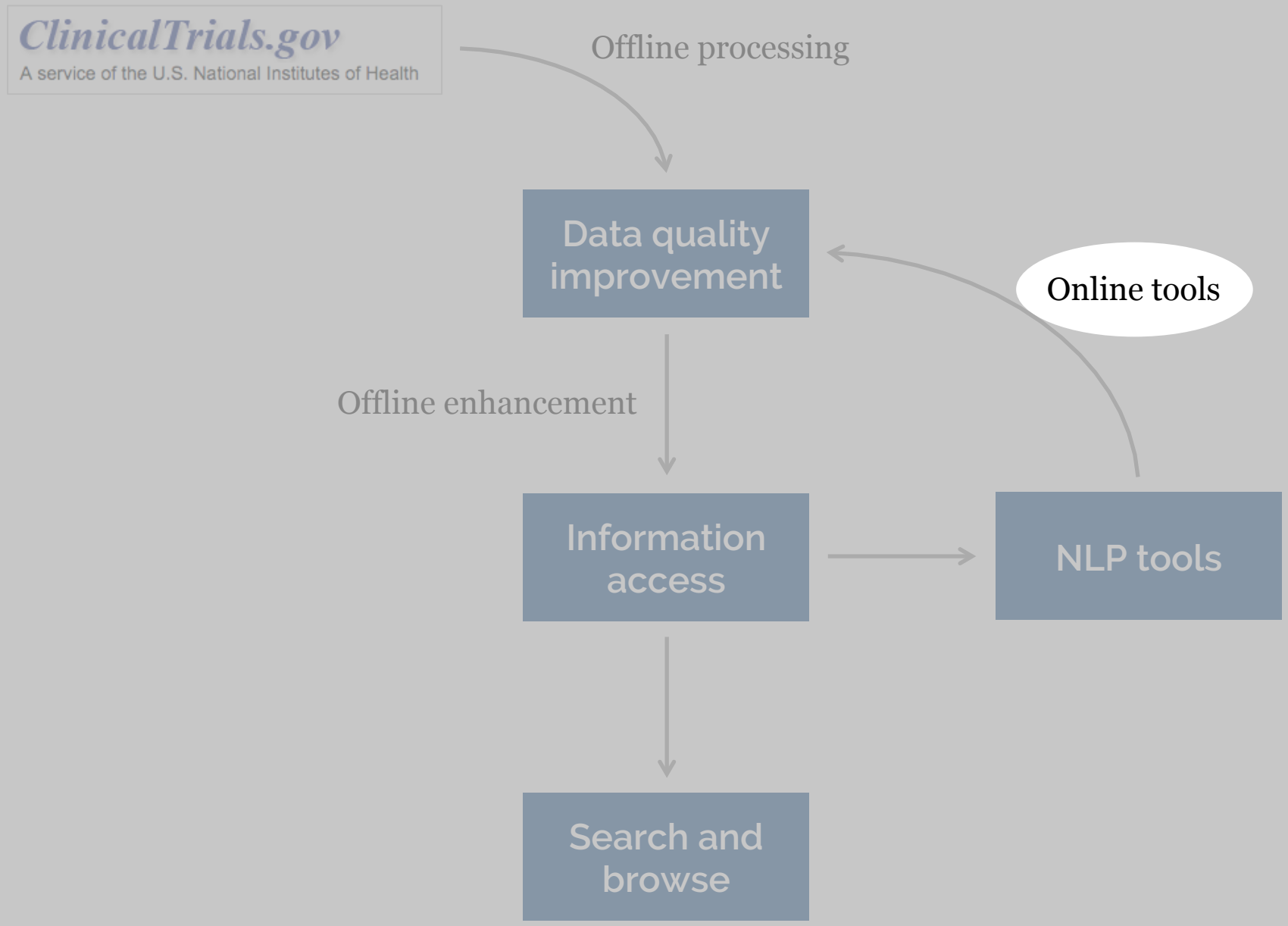
Possibly related

- Aisner J, Manola JB, Dakhil SR, Stella PJ, Sovak MA, Schiller JH, et al. Vandetanib plus chemotherapy for induction followed by vandetanib or placebo as maintenance for patients with advanced non-small-cell lung cancer: a randomized phase 2 PrECOG study (PrE0501). J Thorac Oncol 2013 Aug;8(8):1075-83.
- Brzezniak C, Carter CA, Giaccone G, et al. Dacomitinib, a new therapy for the treatment of non-small cell lung cancer. Expert Opin Pharmacother 2013 Feb;14(2):247-53.
- Butts C, Socinski MA, Mitchell PL, Thatcher N, Havel L, Krzakowski M, Nawrocki S, Ciuleanu TE, Bosqu Tecemotide (L-BLP25) versus placebo after chemoradiotherapy for stage III non-small-cell lung cancer (START): a randomised, double-blind, phase 3 trial. Lancet Oncol. 2014 Jan;15(1):59-68.
- Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, Shen L, Fan Y, Giri U, Tumula PK, et al. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin. Cancer Res. 2013 Jan 1;19(1):279-90.
- Champiat S, Ileana E, Giaccone G, Besse B, Mountzios G, Eggermont A, Soria JC, et al. Incorporating immune-checkpoint inhibitors into systemic therapy of NSCLC. J Thorac Oncol 2014 Feb;9(2):144-53.

Freebase linkage

- Problem: we had deduplicated references to institutions, but needed to ensure these represented real-world entities
- Solution: link each of our institutions to a Freebase entity representing a hospital, university, pharmaceutical firm, or other relevant institution
- Results: succeeded in linking the 1,075 institutions to Freebase entities, getting image, description, and location information where available



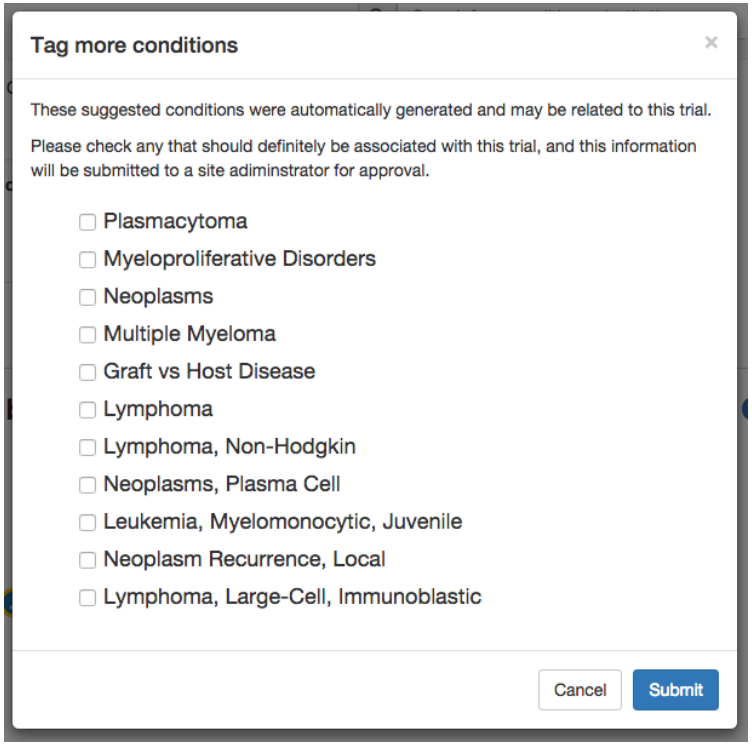

ClinicalTrials.gov
A service of the U.S. National Institutes of Health
Offline processing
Data quality improvement
Online tools
Offline enhancement
Information access
NLP tools
Search and browse

Online tools

- MeSH suggestion tagging
- MeSH recommendation engine
- Criteria concept discovery

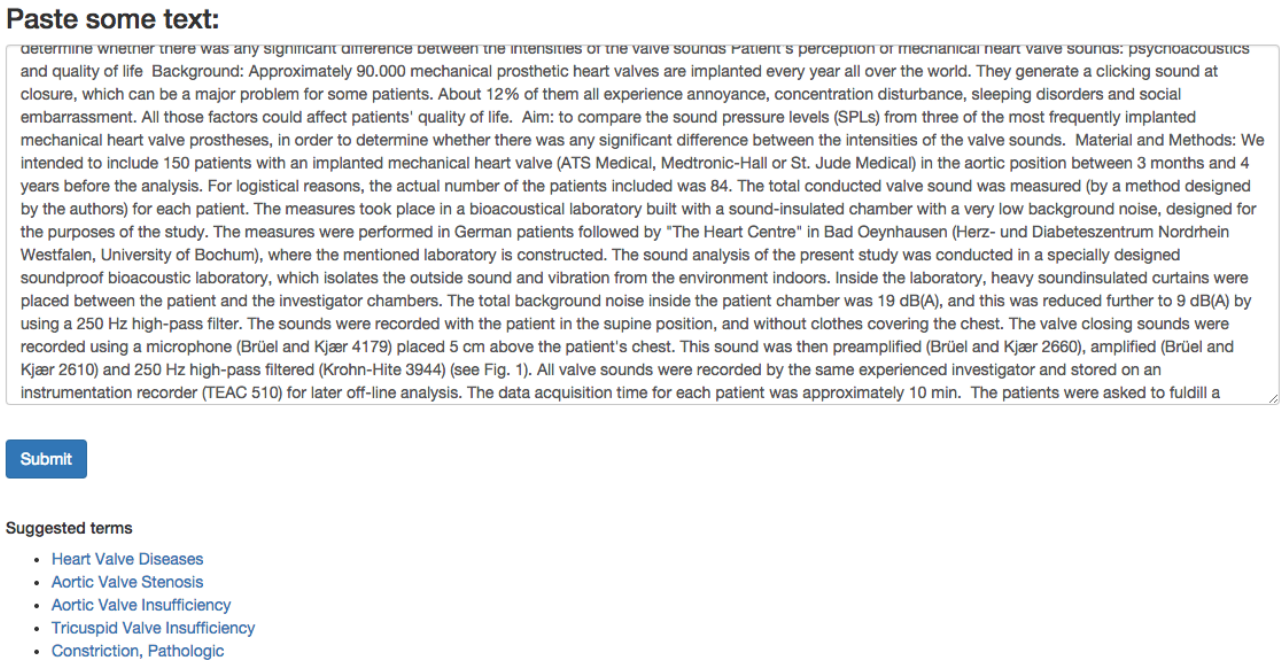
MeSH suggestion tagging

- Problem: the suggested terms from our offline MeSH assignment have some errors, so need human confirmation
- Solution: add interface to allow registered DiscoverCT.org users to confirm or reject suggestions
- Results: lightweight modal interface for accepting suggestions



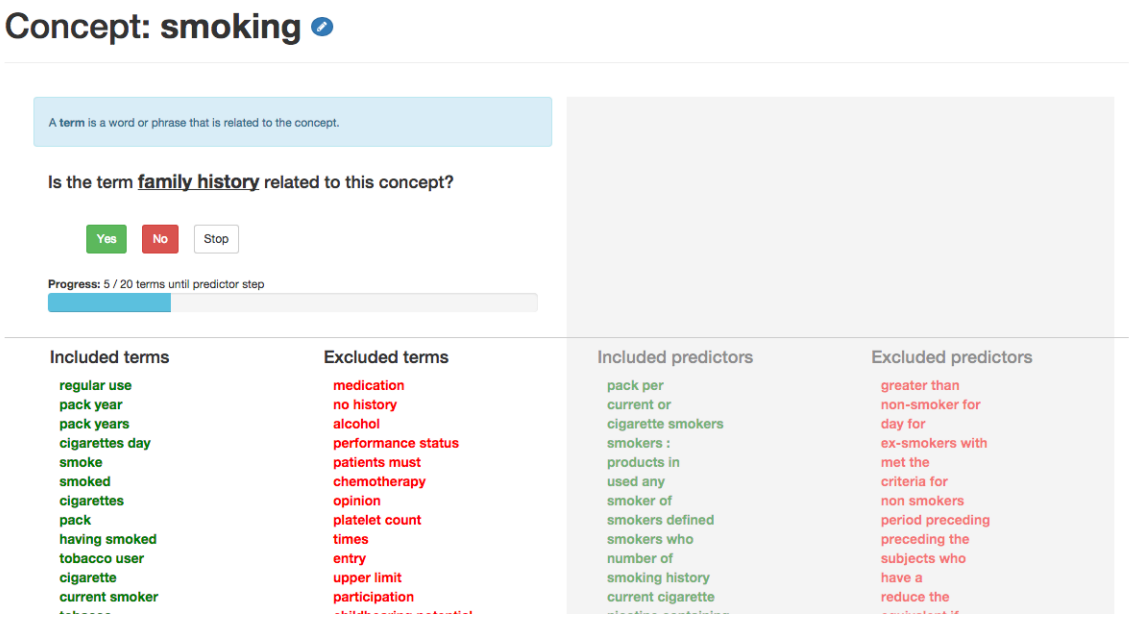
MeSH recommendation engine

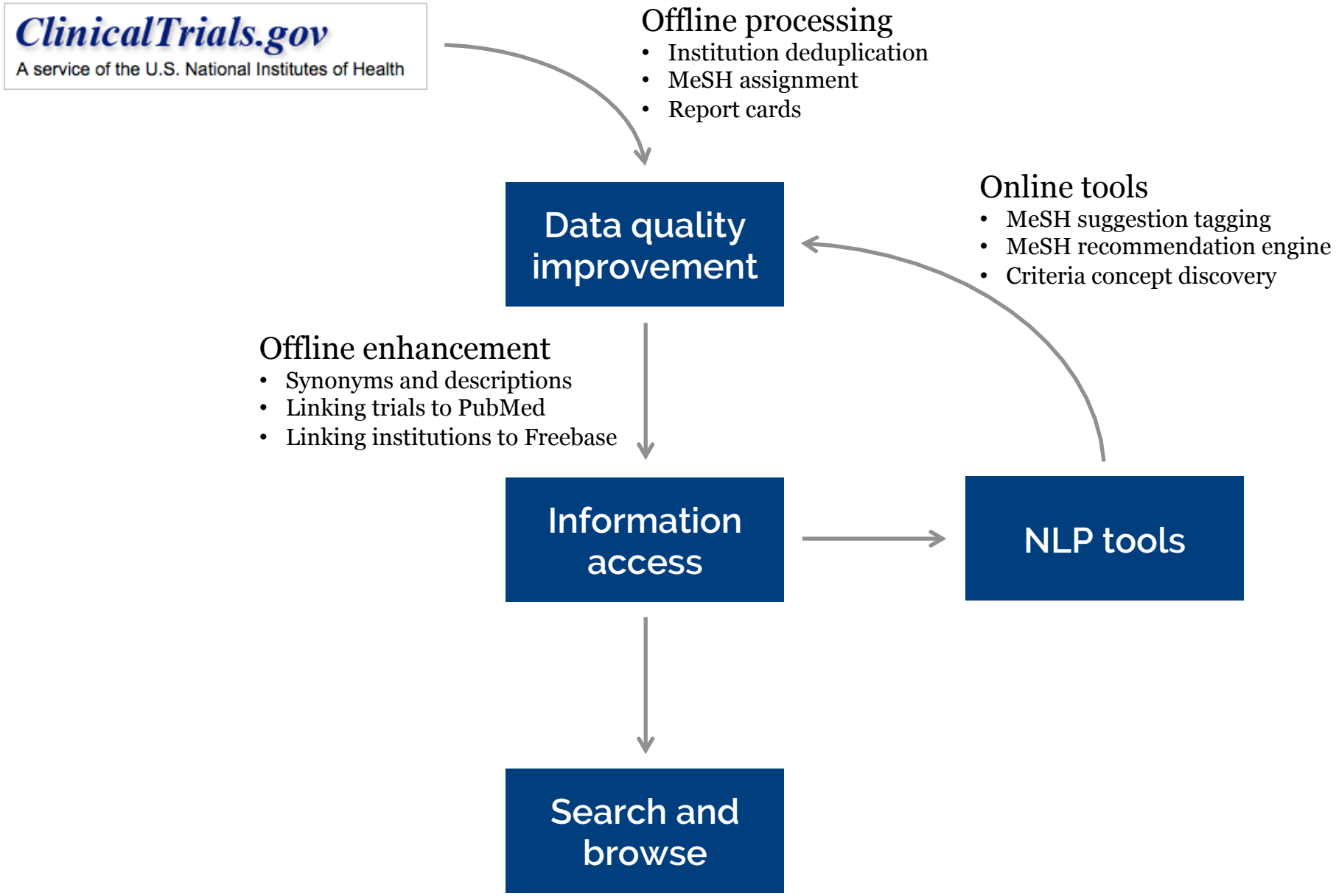
- Problem: trial investigators and researchers don't describe their work using MeSH terms because they are unfamiliar with the vocabulary
- Solution: develop lightweight version of offline KNN process to suggest relevant MeSH terms for any text a user enters
- Results: responsive interface that quickly generates relevant terms for text (up to 8KB)

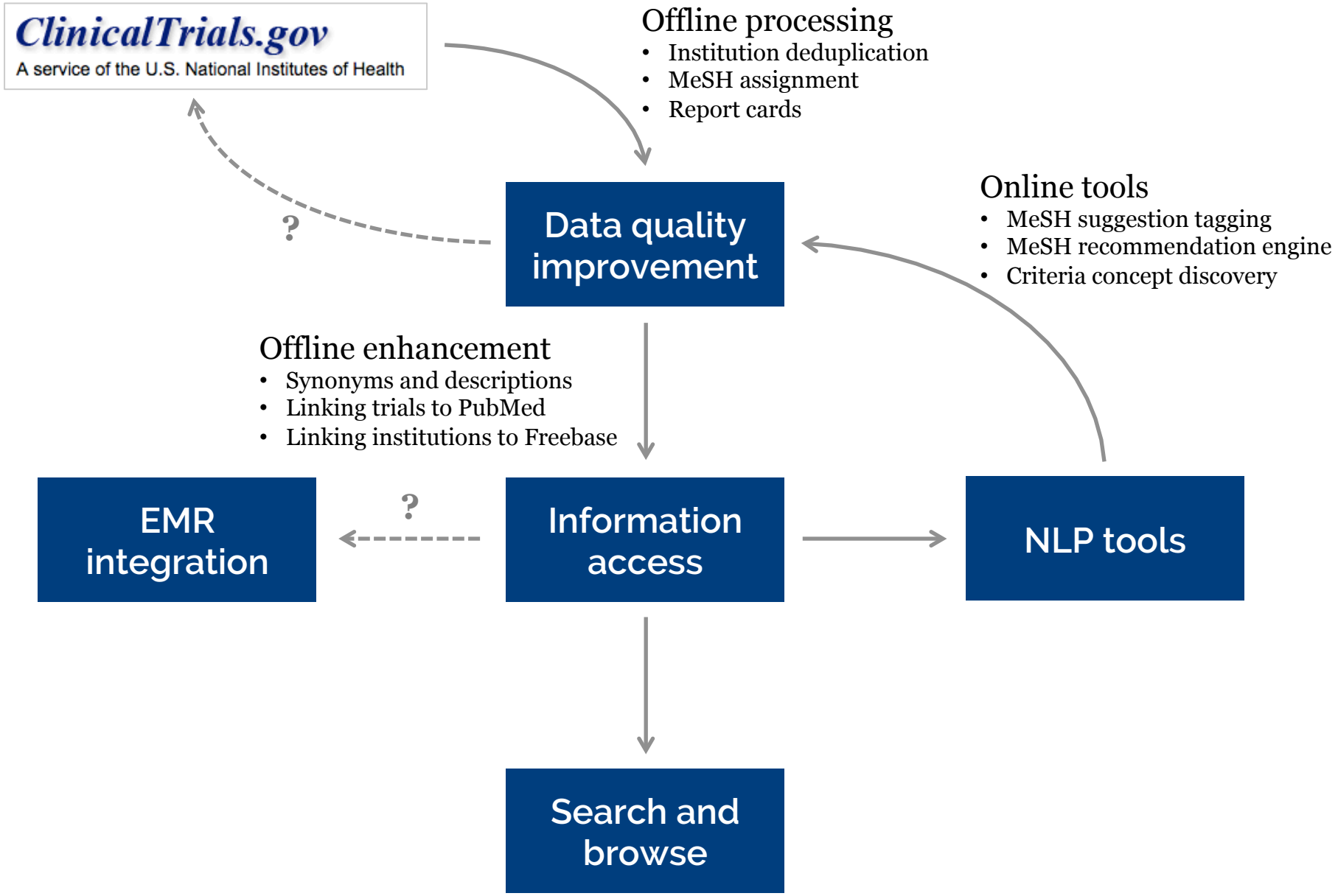


Criteria concept discovery

- Problem: eligibility criteria are free text and refer to concepts in different ways
- Solution: develop active learning interface so registered DiscoverCT.org users can teach an algorithm which terms should be included in a concept
- Results: user-friendly process for identifying a set of related terms that form a concept, which can then be used to filter trials in the search interface




ClinicalTrials.gov
A service of the U.S. National Institutes of Health
Offline processing
• Institution deduplication
• MeSH assignment
• Report cards
Data quality improvement
Online tools
• MeSH suggestion tagging
• MeSH recommendation engine
• Criteria concept discovery
Offline enhancement
• Synonyms and descriptions
• Linking trials to PubMed
• Linking institutions to Freebase
Information access
NLP tools
Search and browse


ClinicalTrials.gov
A service of the U.S. National Institutes of Health
Offline processing
• Institution deduplication
• MeSH assignment
• Report cards
?
Data quality improvement
Online tools
• MeSH suggestion tagging
• MeSH recommendation engine
• Criteria concept discovery
Offline enhancement
• Synonyms and descriptions
• Linking trials to PubMed
• Linking institutions to Freebase
EMR integration
?
Information access
NLP tools
Search and browse

Acknowledgements

- Professor Marti Hearst
- Anobel Odisho, MD
- Katey Basye, MIMS ’15
- Winston Chiong, Jack Colford, Jennifer Ahern, and all our test users

DiscoverCT.org