



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

ClinicalTrials.gov currently lists **189,951 studies** with locations in all 50 states and in **190 countries**.

Text Size ▾

Search for Studies

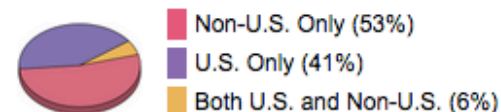
Example: "Heart attack" AND "Los Angeles"

[Advanced Search](#) | [See Studies by Topic](#)
[See Studies on a Map](#)

Search Help

- [How to search](#)
- [How to find results of studies](#)
- [How to read a study record](#)

Locations of Recruiting Studies



Total N = 35,222 studies
Data as of May 07, 2015

- [See more trends, charts, and maps](#)

Learn More

- [ClinicalTrials.gov Online Training](#)
- [Glossary of common site terms](#)

[For the Press](#)

[Using our RSS Feeds](#)

For Patients & Families

- [How to find studies](#)
- [See studies by topic](#)
- [Learn about clinical studies](#)
- [Learn more...](#)

For Researchers

- [How to submit studies](#)
- [Download content for analysis](#)
- [About the results database](#)
- [Learn more...](#)

For Study Record Managers

- [Why register?](#)
- [How to register study records](#)
- [FDAAA 801 Requirements](#)
- [Learn more...](#)

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

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](#)

[Full Text View](#)

[Tabular View](#)

[No Study Results Posted](#)

[Disclaimer](#)

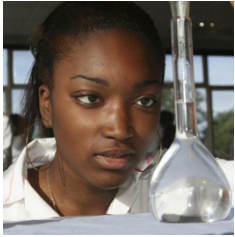
[? How to Read a Study Record](#)

► 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 is 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.



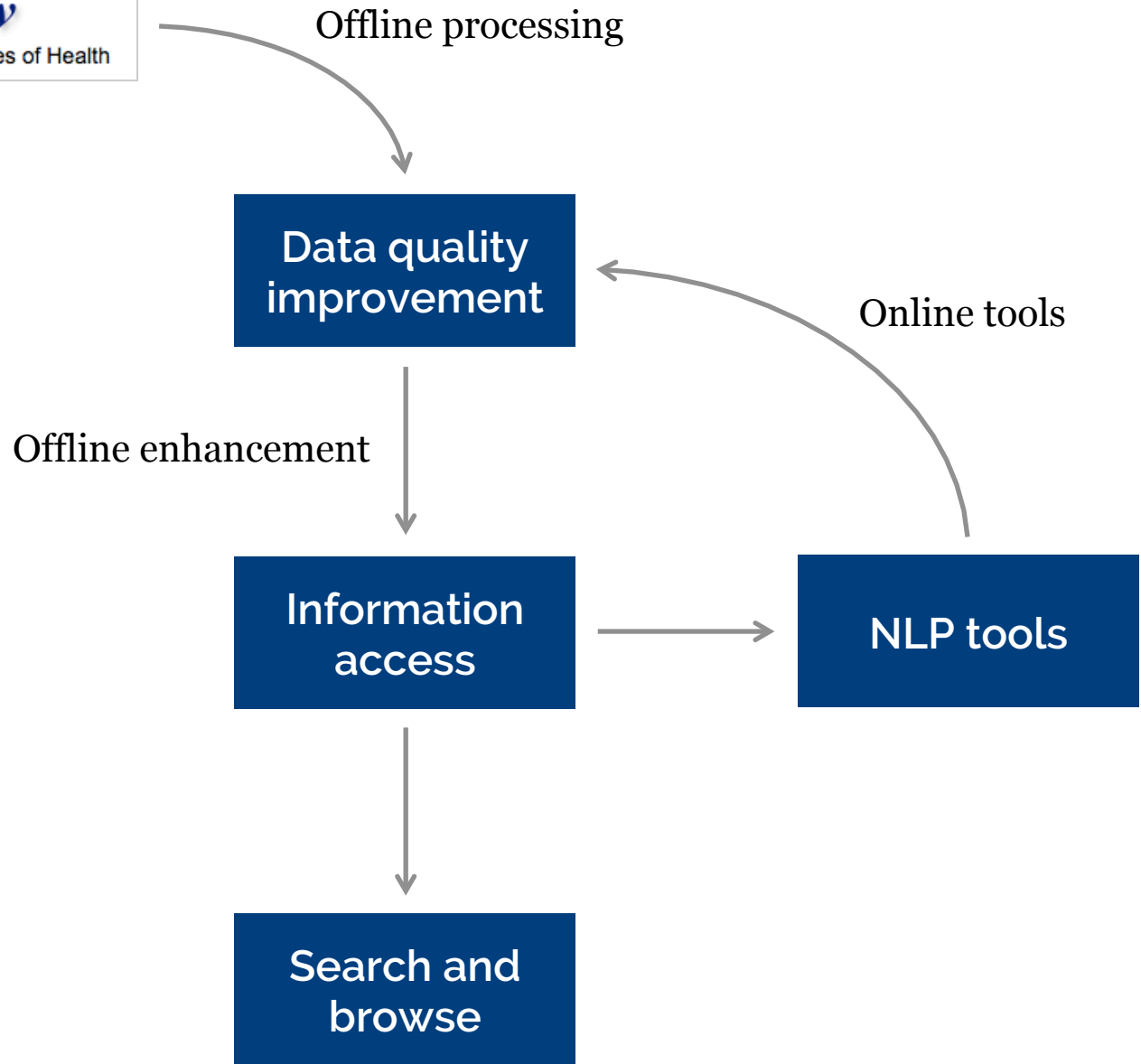
The Patient needs to find active trials in her area that offer potential treatments for her condition. She is not a medical expert and doesn't have the ability to sort through dozens of possible trials.



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



ClinicalTrials.gov

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

Offline processing

Offline enhancement

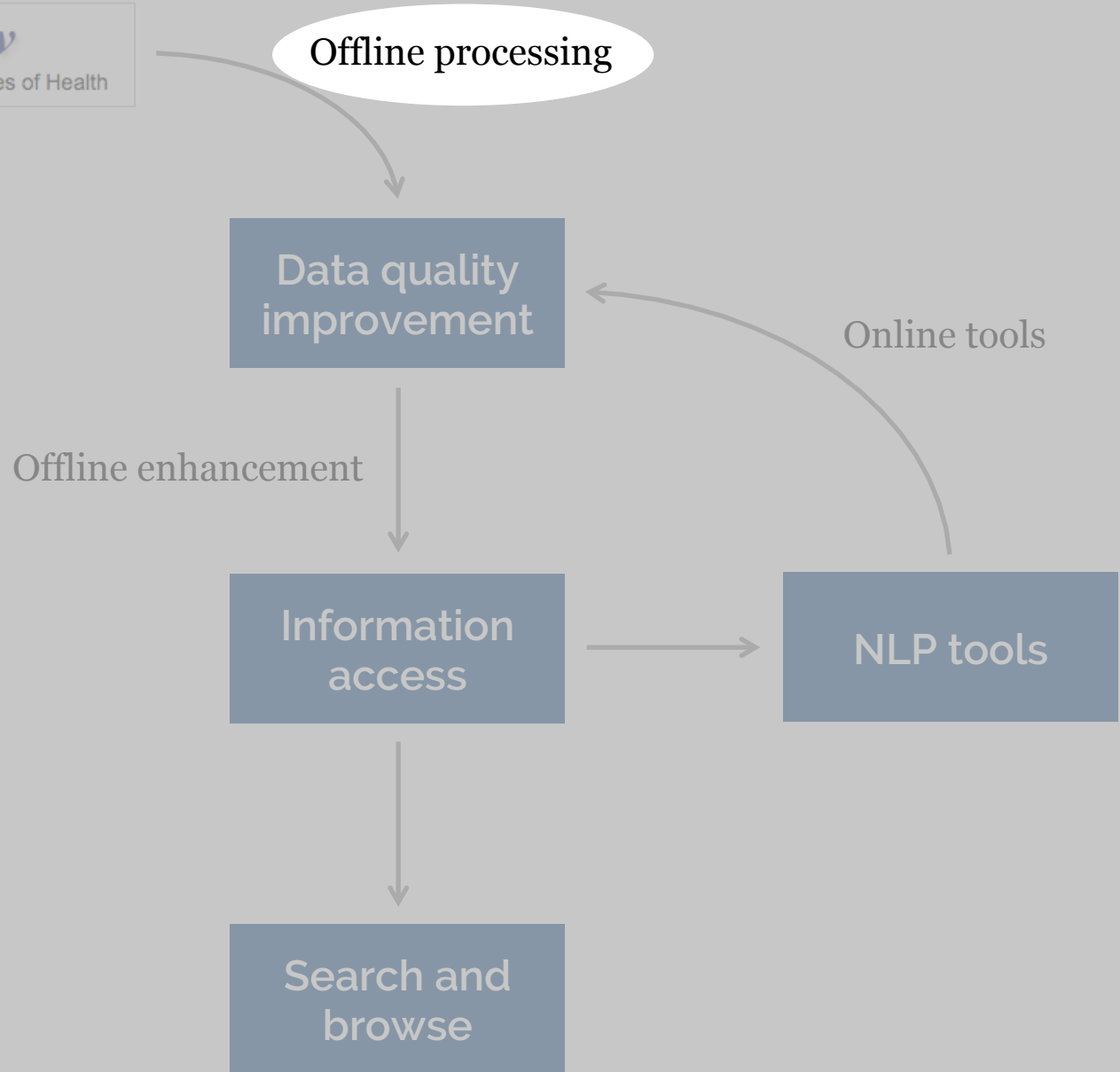
Online tools

Data quality improvement

Information access

NLP tools

Search and browse



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

Conditions addressed by this trial

Officially assigned:

- Hematologic Neoplasms
- Myelodysplastic Syndromes
- Preleukemia
- Leukemia

Suggested by DiscoverCT:

- Plasmacytoma
- Myeloproliferative Disorders
- Neoplasms
- Multiple Myeloma
- Graft vs Host Disease
- Lymphoma
- Lymphoma, Non-Hodgkin
- Neoplasms, Plasma Cell
- Leukemia, Myelomonocytic, Juvenile
- Neoplasm Recurrence, Local
- Lymphoma, Large-Cell, Immunoblastic

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

Trial Data Quality Ratings

Overall data quality	★★★★☆
MeSH classification quality	★★★☆☆
Site data quality	★★★★★
Protocol description quality	★★★★☆
Trial criteria description quality	★★★★☆☆
Date quality	★★★★☆

Institution Data Quality Ratings

Overall data quality	★★★★☆
MeSH classification quality	★★★★☆☆
Site data quality	★★★★☆
Protocol description quality	★★★★☆☆
Trial criteria description quality	★★★★☆☆
Date quality	★★★★★

ClinicalTrials.gov

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

Offline processing

Data quality improvement

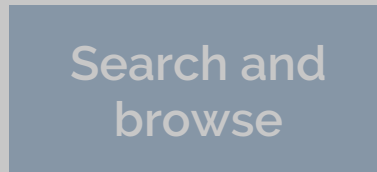
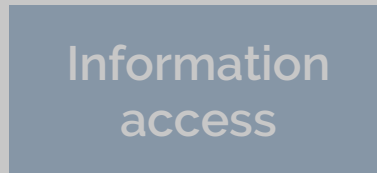
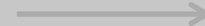
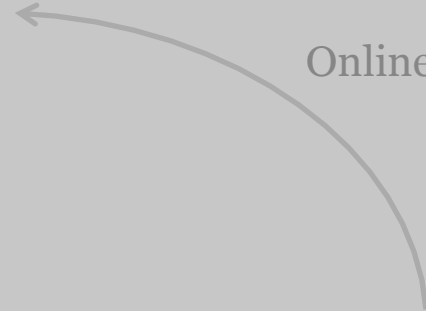
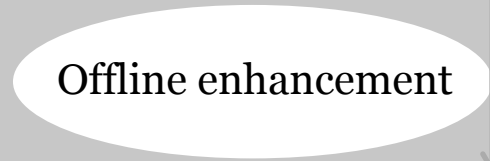
Online tools

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

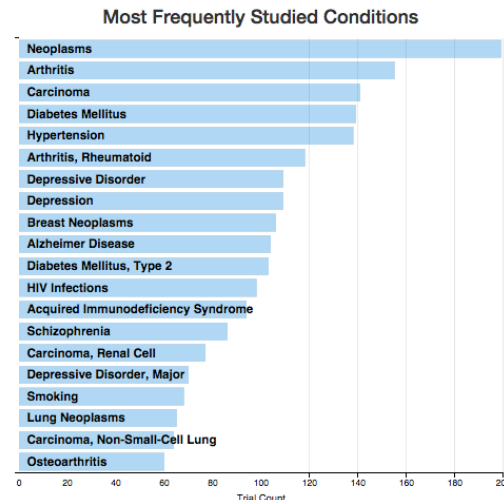
- Powell CA, Halmos B, Nana-Sinkam SP, et al. [Update in lung cancer and mesothelioma 2012](#). *Am. J. Respir. Crit. Care Med.* 2013 Jul 15;188(2):157-66.

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



Institution Data Quality Ratings

Overall data quality	★★★★☆
MeSH classification quality	★★★★☆
Site data quality	★★★★☆
Protocol description quality	★★★★☆
Trial criteria description quality	★★★★☆
Date quality	★★★★★

Summary

Pfizer, Inc. is an American multinational pharmaceutical corporation headquartered in New York City, and with its research headquarters in Groton, Connecticut, United States. It is one of the world's largest pharmaceutical companies by revenues. Pfizer develops and produces medicines and vaccines for a wide range of medical disciplines, including immunology, oncology, cardiology, diabetology/endocrinology, and neurology. Pfizer's products include the blockbuster drug Lipitor, used to lower LDL blood cholesterol; Lyrica; Diflucan, an oral antifungal medication; Zithromax, an antibiotic; Viagra; and Celebrex/Celebra, an anti-inflammatory drug. Pfizer was founded by cousins Charles Pfizer and Charles F. Erhart in New York City in 1849 as a manufacturer of fine chemicals. Pfizer's discovery of Terramycin in 1950 put it on a path towards becoming a research-based pharmaceutical company. Pfizer has made numerous acquisitions, including Warner-Lambert in 2000, Pharmacia in 2003 and Wyeth in 2009. The Wyeth acquisition was the largest of the three at US\$68 billion.

This institution is associated with 12 trial sponsors and 17538 research facilities

ClinicalTrials.gov

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

Offline processing

Data quality improvement

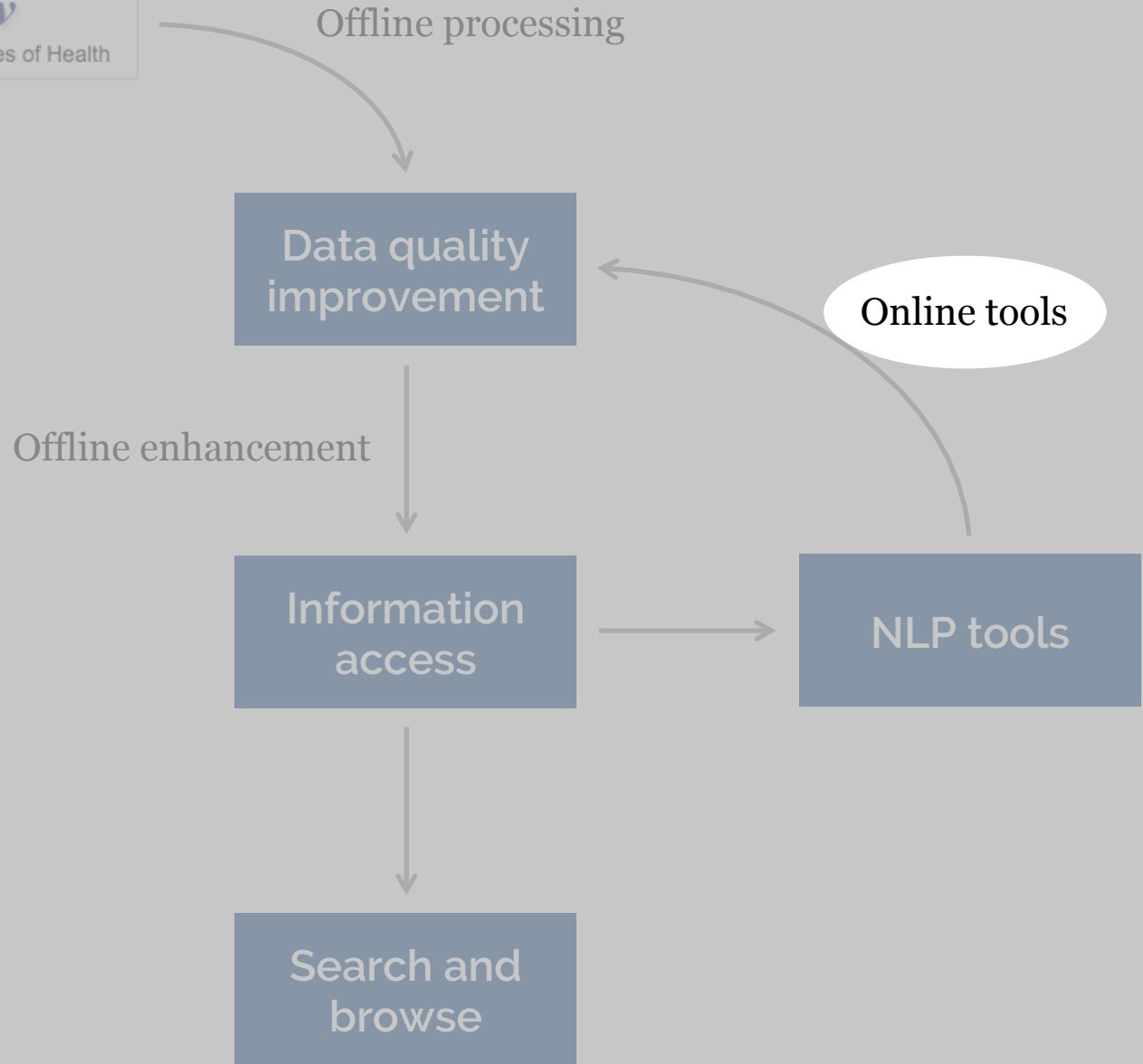
Online tools

Offline enhancement

Information access

NLP tools

Search and browse

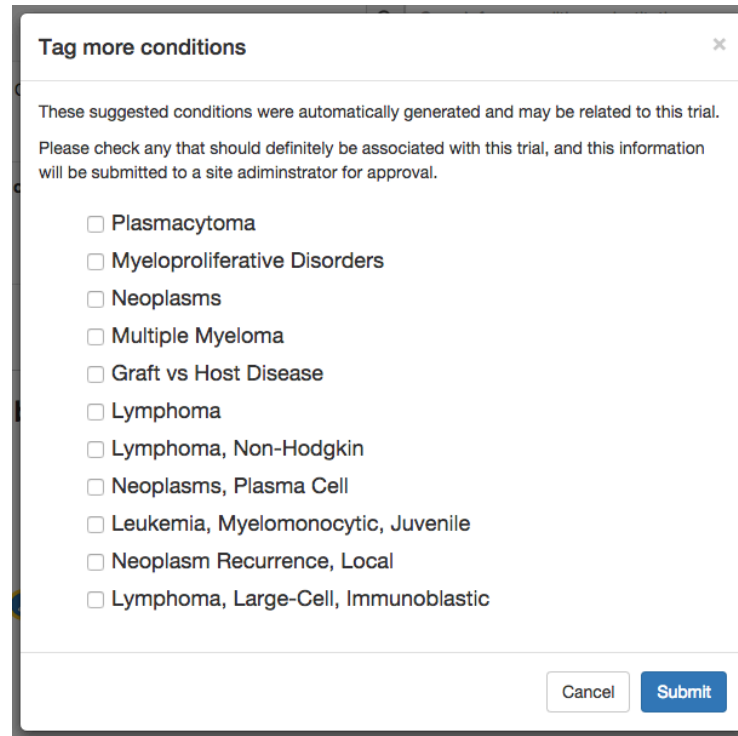


Online tools

- 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



The image shows a modal window titled "Tag more conditions" with a close button (x) in the top right corner. The modal contains the following text and list:

These suggested conditions were automatically generated and may be related to this trial.

Please check any that should definitely be associated with this trial, and this information will be submitted to a site administrator for approval.

- Plasmacytoma
- Myeloproliferative Disorders
- Neoplasms
- Multiple Myeloma
- Graft vs Host Disease
- Lymphoma
- Lymphoma, Non-Hodgkin
- Neoplasms, Plasma Cell
- Leukemia, Myelomonocytic, Juvenile
- Neoplasm Recurrence, Local
- Lymphoma, Large-Cell, Immunoblastic

At the bottom right of the modal, there are two buttons: "Cancel" and "Submit".

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)

Paste some text:

determine whether there was any significant difference between the intensities of the valve sounds. Patient's perception of mechanical heart valve sounds: psychoacoustics and quality of life. Background: Approximately 90.000 mechanical prosthetic heart valves are implanted every year all over the world. They generate a clicking sound at closure, which can be a major problem for some patients. About 12% of them all experience annoyance, concentration disturbance, sleeping disorders and social embarrassment. All those factors could affect patients' quality of life. Aim: to compare the sound pressure levels (SPLs) from three of the most frequently implanted mechanical heart valve prostheses, in order to determine whether there was any significant difference between the intensities of the valve sounds. Material and Methods: We intended to include 150 patients with an implanted mechanical heart valve (ATS Medical, Medtronic-Hall or St. Jude Medical) in the aortic position between 3 months and 4 years before the analysis. For logistical reasons, the actual number of the patients included was 84. The total conducted valve sound was measured (by a method designed by the authors) for each patient. The measures took place in a bioacoustical laboratory built with a sound-insulated chamber with a very low background noise, designed for the purposes of the study. The measures were performed in German patients followed by "The Heart Centre" in Bad Oeynhausen (Herz- und Diabeteszentrum Nordrhein Westfalen, University of Bochum), where the mentioned laboratory is constructed. The sound analysis of the present study was conducted in a specially designed soundproof bioacoustic laboratory, which isolates the outside sound and vibration from the environment indoors. Inside the laboratory, heavy soundinsulated curtains were placed between the patient and the investigator chambers. The total background noise inside the patient chamber was 19 dB(A), and this was reduced further to 9 dB(A) by using a 250 Hz high-pass filter. The sounds were recorded with the patient in the supine position, and without clothes covering the chest. The valve closing sounds were recorded using a microphone (Brüel and Kjær 4179) placed 5 cm above the patient's chest. This sound was then preamplified (Brüel and Kjær 2660), amplified (Brüel and Kjær 2610) and 250 Hz high-pass filtered (Krohn-Hite 3944) (see Fig. 1). All valve sounds were recorded by the same experienced investigator and stored on an instrumentation recorder (TEAC 510) for later off-line analysis. The data acquisition time for each patient was approximately 10 min. The patients were asked to fulfill a

Submit

Suggested terms

- Heart Valve Diseases
- Aortic Valve Stenosis
- Aortic Valve Insufficiency
- Tricuspid Valve Insufficiency
- Constriction, Pathologic

Criteria concept discovery

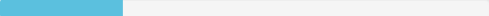
- Problem: eligibility criteria are free text and refer to concepts in different ways
- Solution: develop active learning interface so registered DiscoverCT.org users can teach an algorithm which terms should be included in a concept
- Results: user-friendly process for identifying a set of related terms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

Concept: **smoking** 

A term is a word or phrase that is related to the concept.

Is the term **family history** related to this concept?

Progress: 5 / 20 terms until predictor step



Included terms	Excluded terms	Included predictors	Excluded predictors
regular use	medication	pack per	greater than
pack year	no history	current or	non-smoker for
pack years	alcohol	cigarette smokers	day for
cigarettes day	performance status	smokers :	ex-smokers with
smoke	patients must	products in	met the
smoked	chemotherapy	used any	criteria for
cigarettes	opinion	smoker of	non smokers
pack	platelet count	smokers defined	period preceding
having smoked	times	smokers who	preceding the
tobacco user	entry	number of	subjects who
cigarette	upper limit	smoking history	have a
current smoker	participation	current cigarette	reduce the

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

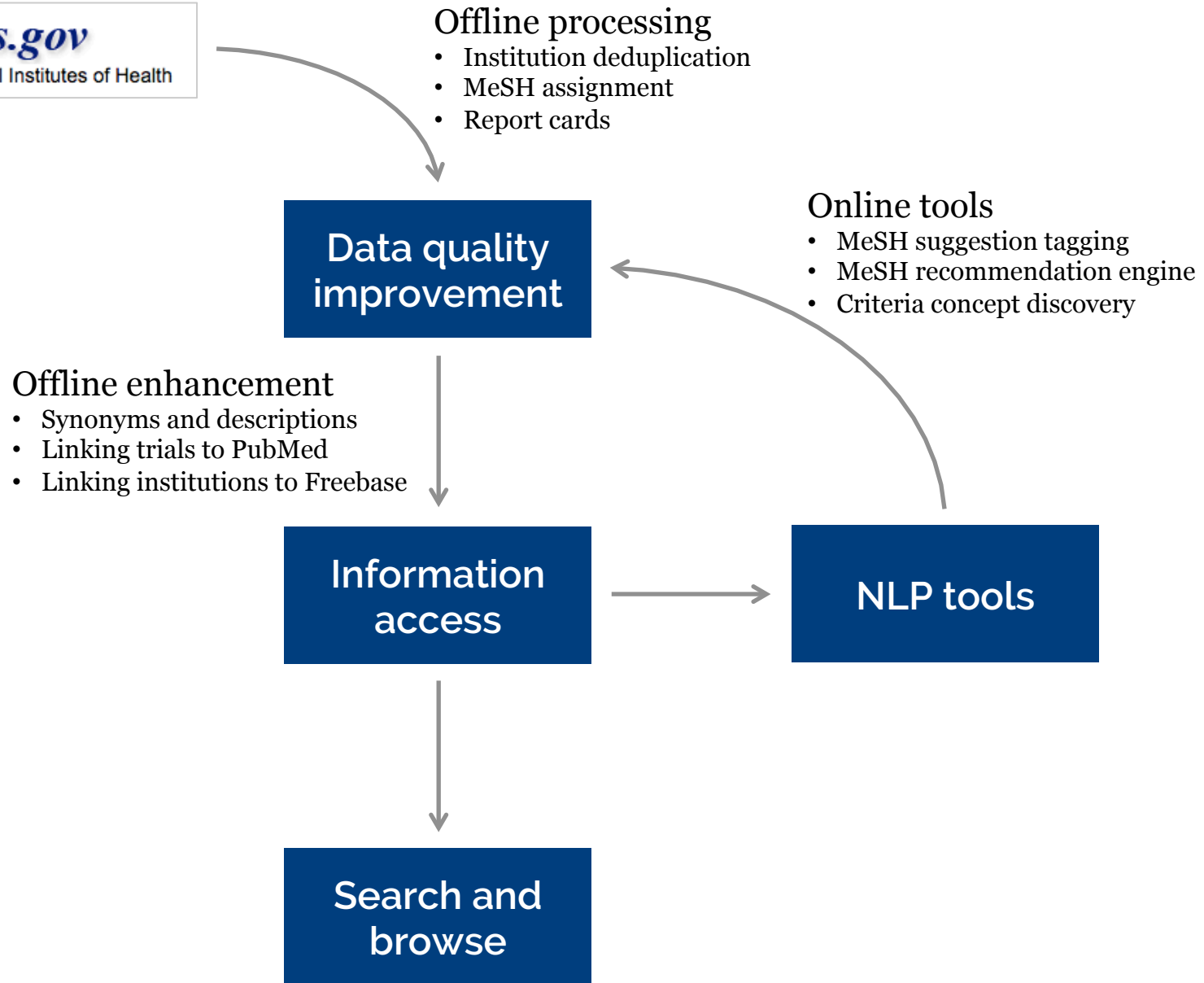
Offline enhancement

- 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

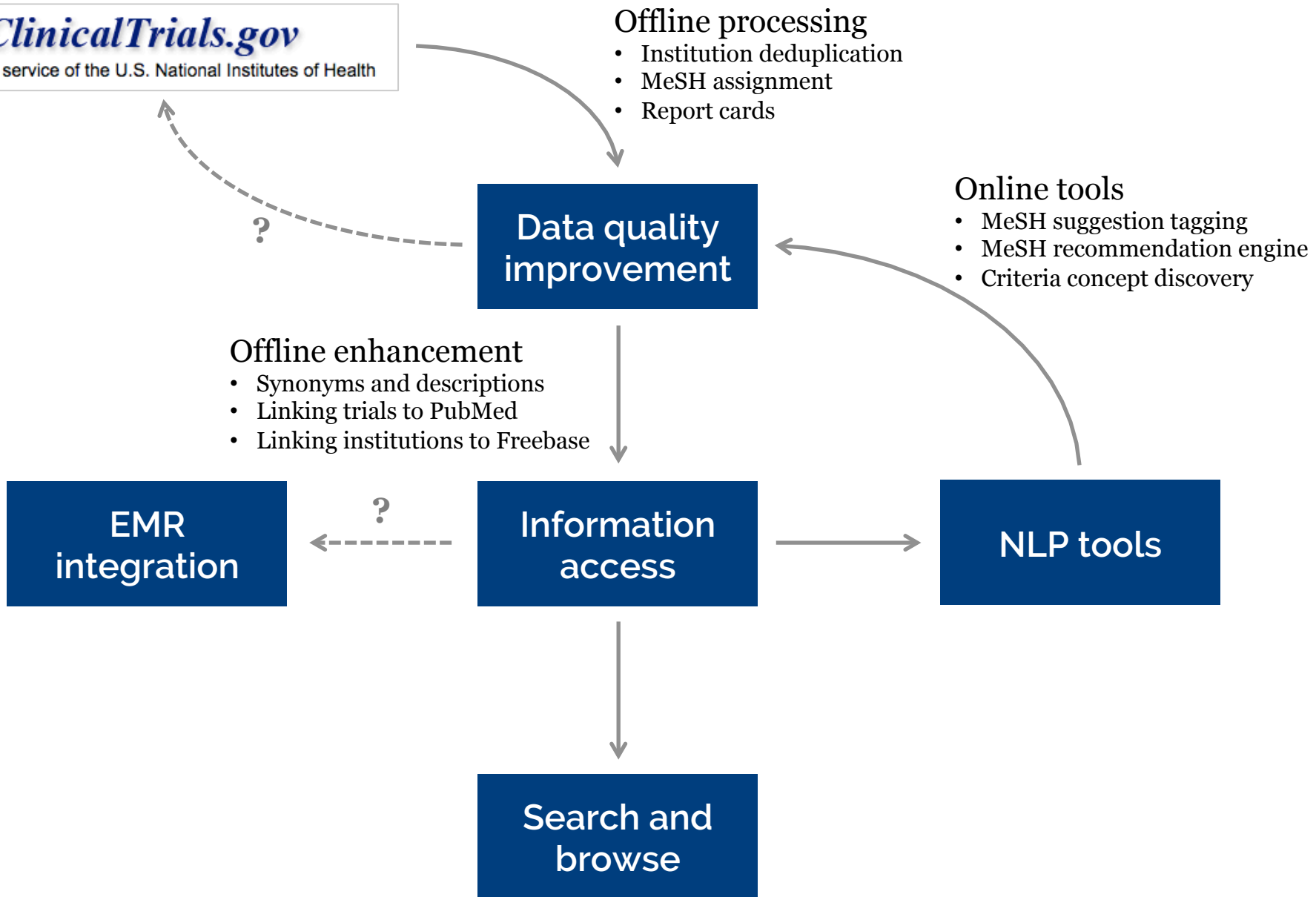
- Synonyms and descriptions
- Linking trials to PubMed
- Linking institutions to Freebase

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



DiscoverCT

[Create account](#) [Login](#)

Search for clinical trials:

Or check out our [other tools](#)

Major institutions that host or sponsor clinical trials:

