# Challenge

- Goal
  - Accurate and computable information
- Challenges
  - We don't know ground truth
  - Information comes from others
  - Data providers have diverse incentives
  - Structured data is hard work for data providers
  - We don't have that much enforcement clout
  - Hard to get the right people to be attentive
  - The world changes

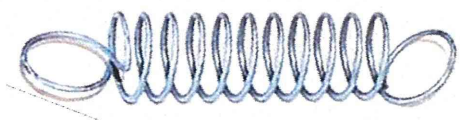
109

# Collecting Information

Reduce burden,
Quick, Easy

Expressive

Use existing
data



Clarity

Correctness

Standards

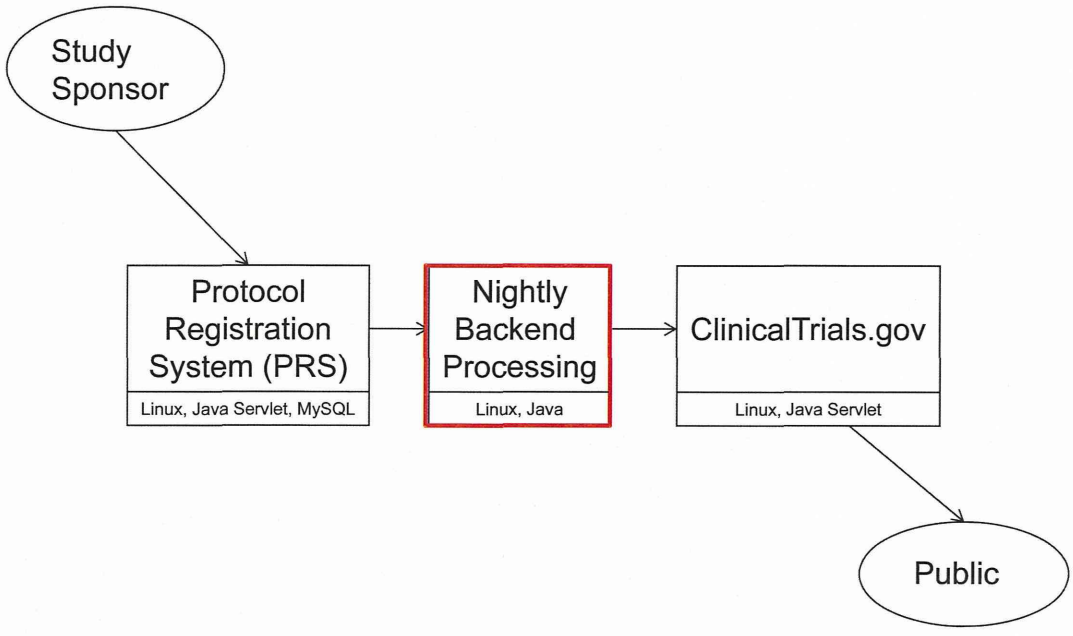
Controlled
Vocabularies

Free text blob,
Images

Partially Structured
Picklists
Controlled fields
Label: value

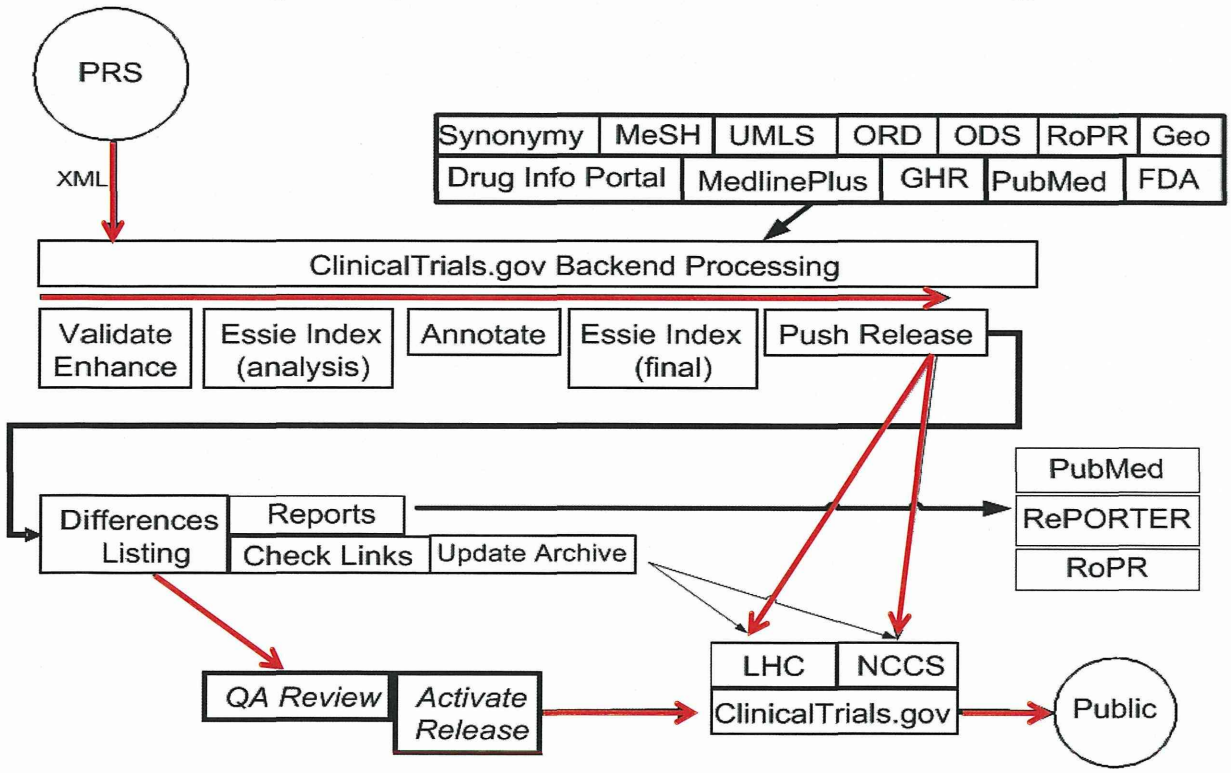
110

# Highest Level Overview




111

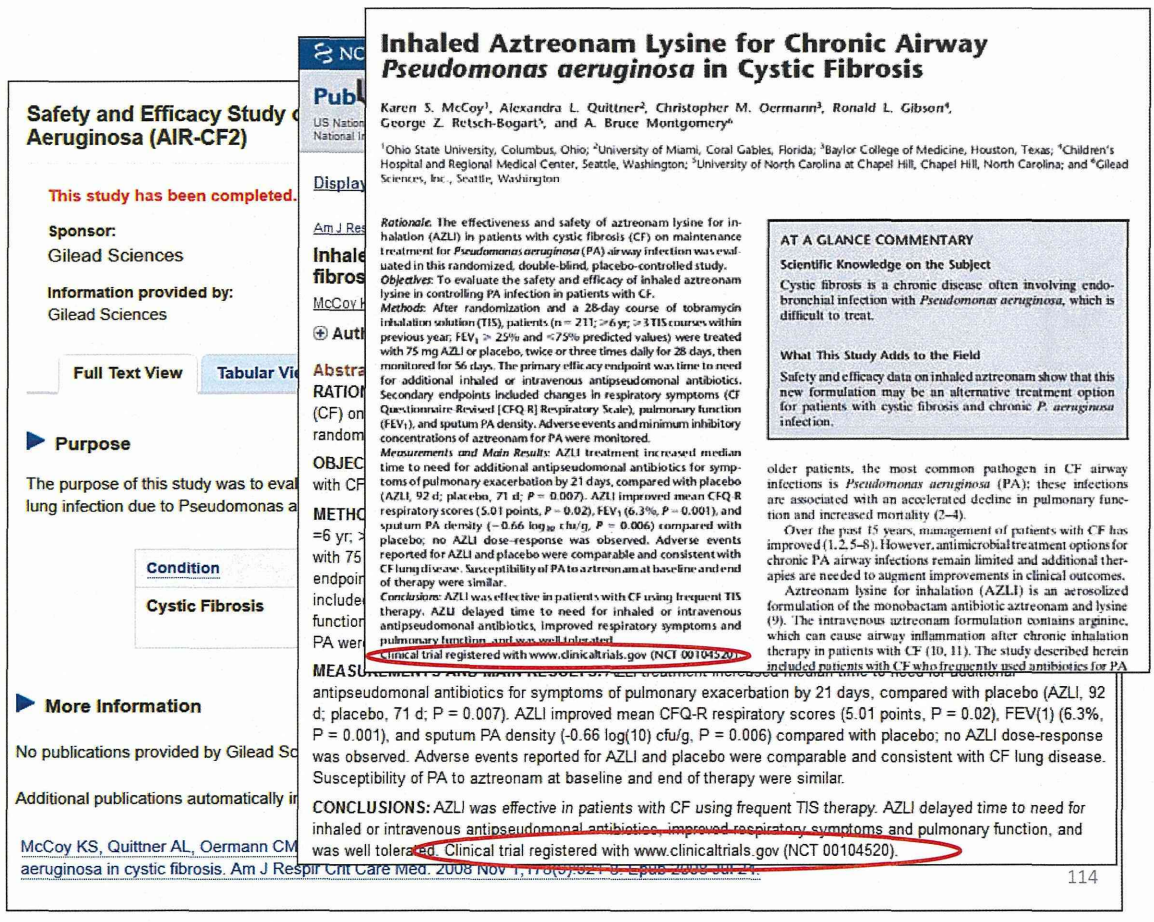
# Nightly Backend Processing



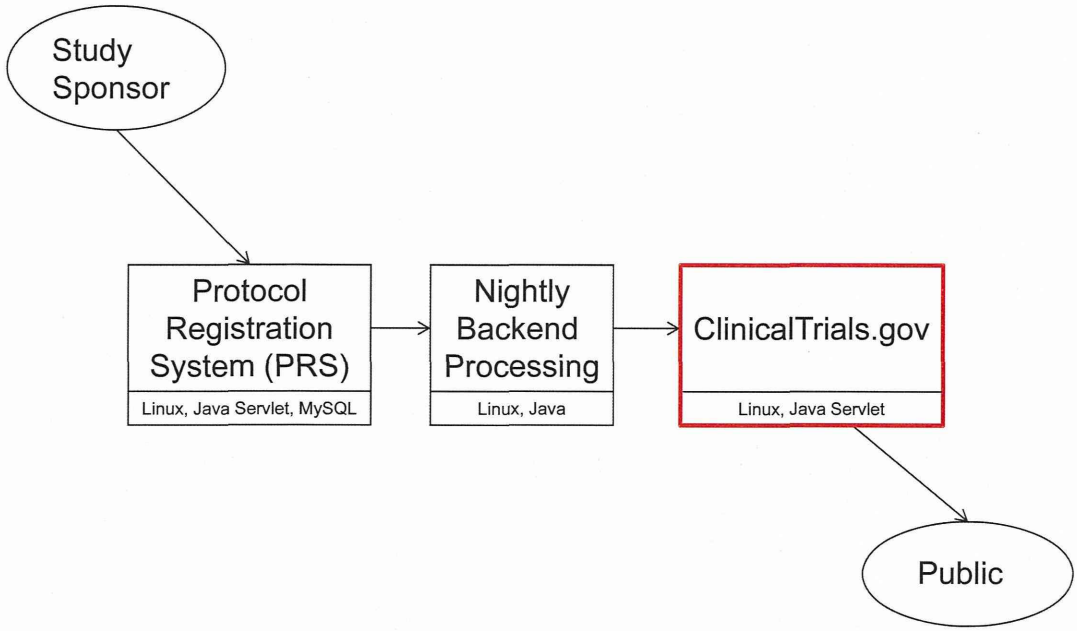
# Annotations for Display

- MeSH Terms and Ancestors
  - Search “crohn disease”, “Granulomatous colitis”, etc…
  - Add “Inflammatory Bowel Disease”, “Gastrointestinal Diseases”, …
- Add links to other resources
  - PubMed citations
  - MedlinePlus Health Topics
  - Genetics Home Reference
  - Drug Information Portal
- Attempt to normalize collaborator names
  - “Gates Foundation” → “Bill and Melinda Gates Foundation”

113



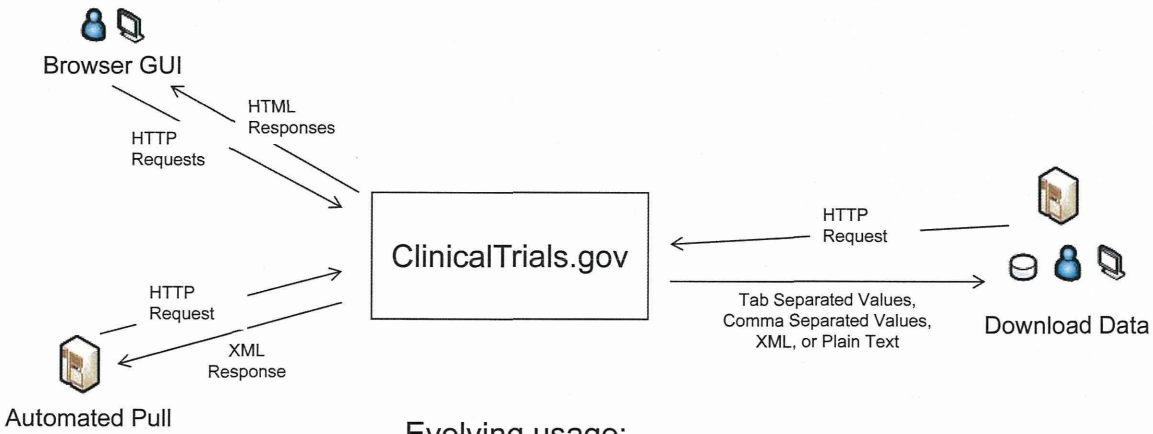
# Highest Level Overview




115

# ClinicalTrials.gov

## Public Site




Evolving usage:

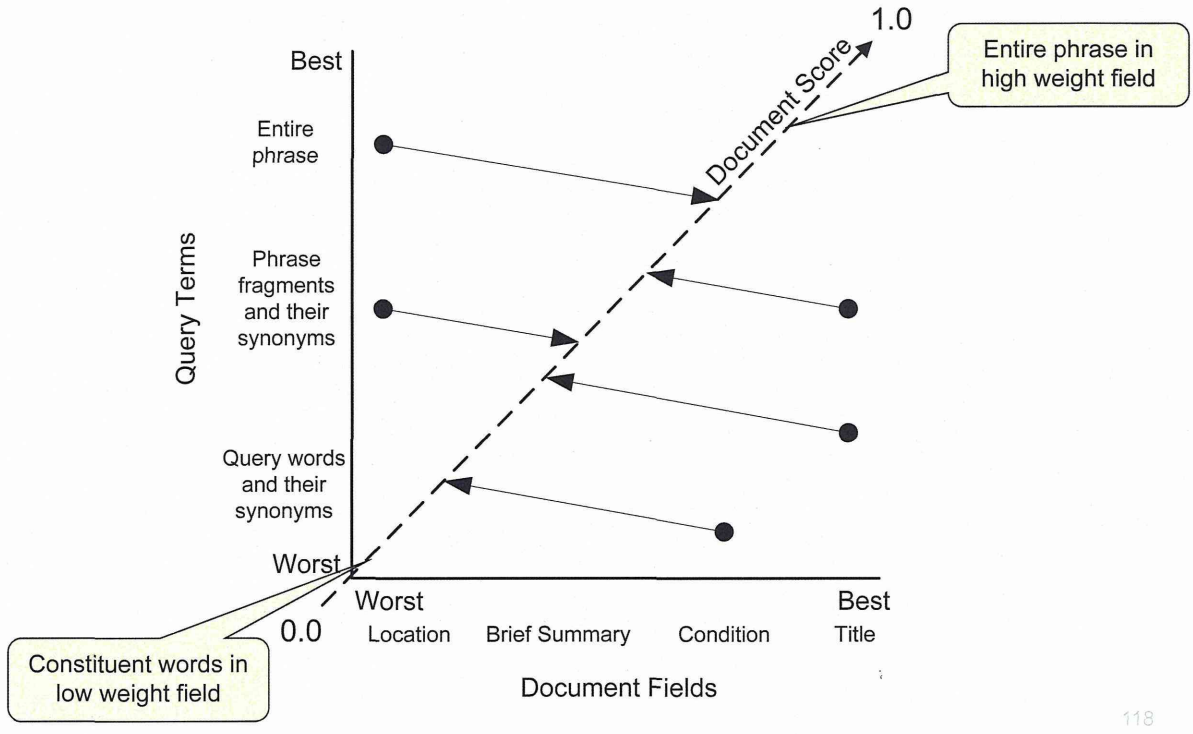
- Initially for "public" to find trials
- Increasingly used as research database serving the clinical research community.

116

# Search Overview: Essie

- Finding documents
  - Uses synonymy derived from UMLS
  - Lexical variants from UMLS
  - Fine grained tokenization **`JAK-2, d-ala(2)`**
- Ordering documents
  - Leverages knowledge of document structure
  - Query relaxation (prefer phrases over words)

# Document Ordering



# Example of Ranking

**266 studies found for**: aspirin for heart attack

| Rank | Status | Study |
|---|---|---|
| 1 | Completed | Warfarin After Anterior ST-Elevation Myocardial Infarction<br>**Condition:** Myocardial Infarction<br>**Interventions:** Drug: aspirin + clopidogrel; Drug: aspirin + clopidogrel + warfarin |
| 2 | Terminated<br>Has Results | Aspirin in Patients With Myocardial Infarction and Thrombocytopenia<br>**Conditions:** Thrombocytopenia; Myocardial Infarction<br>**Intervention:** Drug: Aspirin |
| ⋮ | | |
| **Rank** | **Status** | **Study** |
| 221 | Recruiting | Renal Acute MI Study<br>**Conditions:** Myocardial Infarction; Kidney Function<br>**Intervention:** |
| 222 | Active, not recruiting | Assessment of Revascularization Versus Conservative Treatment in Heart Transplant Patients for a Clinical Event Reduction<br>**Condition:** Heart-lung Transplant Rejection<br>**Interventions:** Device: Stent; Drug: optimal medical therapy |

119

# Archival Data: Tracking Changes in the Record

- Each record is expected to be corrected or updated throughout the trial's life cycle, and all changes are tracked on a public archive site that is accessible from each record (through a "History of Changes" link).
- Tabular View
  - Current Outcome Measures
  - Original (First Registered) Outcome Measures

120

Source: *N Engl J Med* 2011; 364:852-60.

# "World Chaos"

## Policies That Help Avoid World Chaos

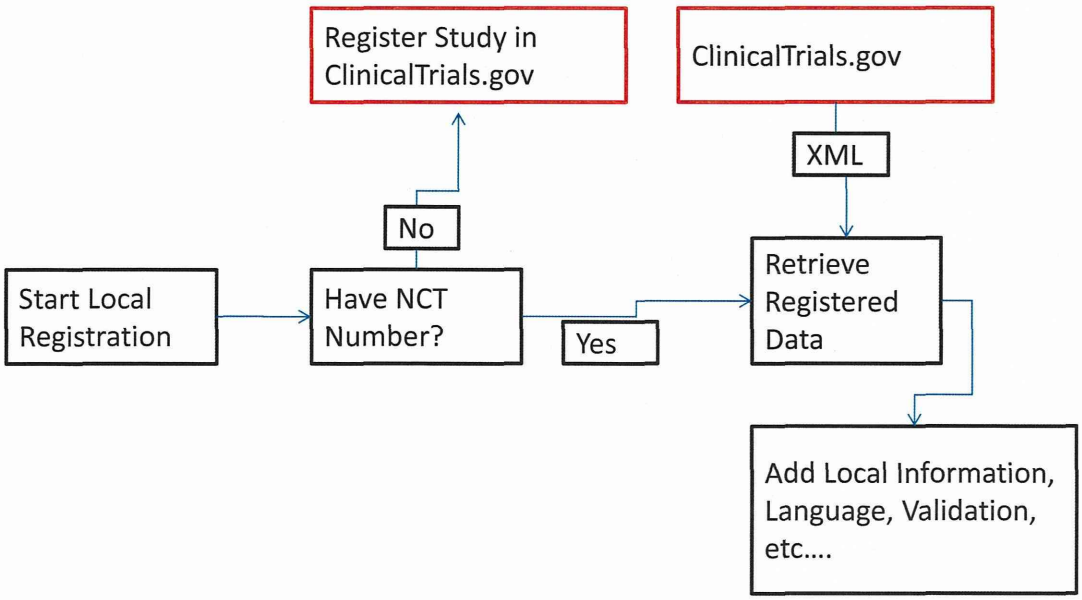
- **Ministry of Health, Israel**: "...controlled prospective clinical trials must be registered at the NIH website, http://www.clinicaltrials.gov."
- **Health Canada:** "...registration of clinical trials ... within 21 days of the trial's onset ... in ClinicalTrials.gov or ISRCTN [UK registry]"

# Collaborative Models: Data Collection

Collaboration is essential and complicated

- Register in ClinicalTrials.gov first *then* collect additional information in local system if necessary.
  - Record ClinicalTrials.gov identifier (NCT Number) in the country registry.
- What data items are desired that are not already collected in ClinicalTrials.gov?
  - ClinicalTrials.gov could add data elements which a country could require.

# Collaborative Data Collection



# Components of Global System

| | ClinicalTrials.gov |
|---|---|
| Incentives | U.S. FDAMA, ICMJE, Other laws, |
| Communication | Some **Need Help !** |
| Database | Done |
| Validation | Some **Need Help !** |
| Dissemination | Done **English Only!** |



# Options on Language Issues

- Parallel Registration
  - Register in English site, capture identifier, then register in language specific registry with cross links between the registries
    - Data provider does the translation
- Capture multiple languages in one database
- Translated Site
  - Using software and human translators, translate some or all of the English data to the target language
- Data Entry Assistance for non-English
  - Is this something the countries could help with ?

# Partner Models

- Content Validation Partner
  - Assist ClinicalTrials.gov with validation of organizations based in other countries
  - Assist ClinicalTrials.gov with validation activities for studies conducted outside of the United States
- Content Dissemination Partner
  - Take data from ClinicalTrials.gov,
  - translate portions of the data,
  - make information available in other language

# Content Validation