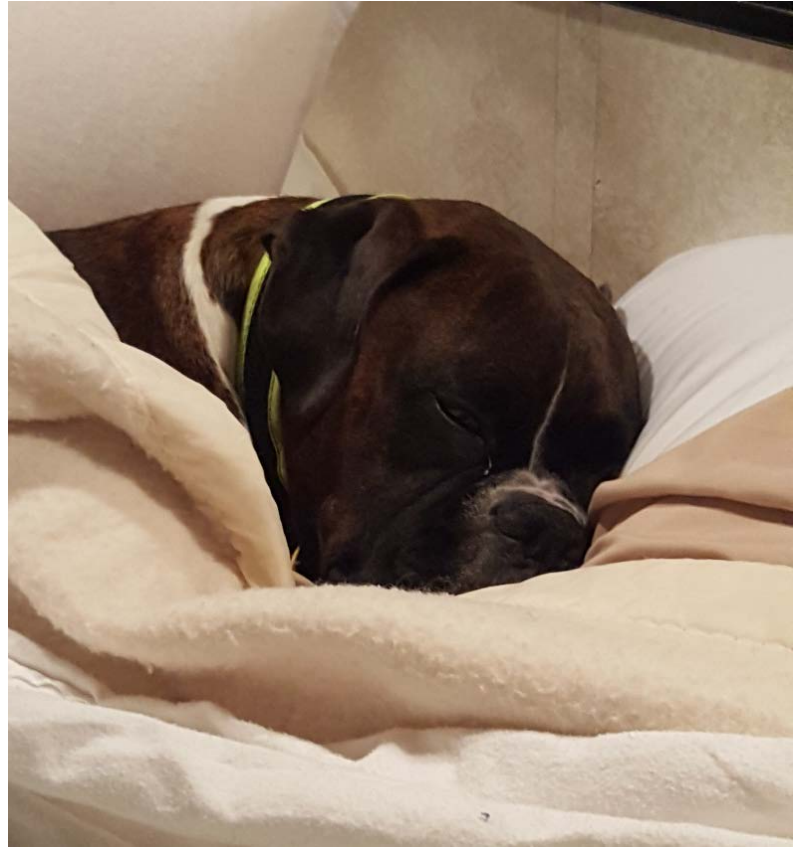




# Development of a Natural Language Processing (NLP) Web Service for Structuring and Standardizing Unstructured Clinical Information

Sandy Jones, Public Health Advisor  
Cancer Surveillance Branch  
Division of Cancer Prevention and Control  
Centers for Disease Control and Prevention (CDC)

**Good Morning.**



# Background and Purpose

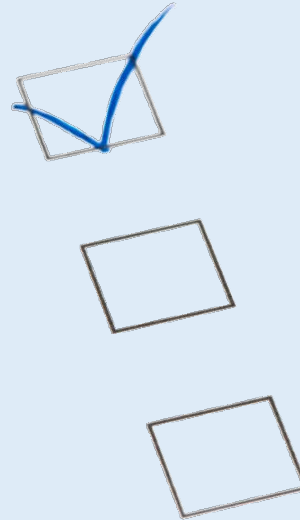
- Meaningful Use and development of Electronic Health Record (EHR) Systems ~ more standard clinical data
- U.S. Laboratories still produce narrative reports
- Food and Drug Administration (FDA) receives narrative data for vaccine, blood products, and drugs as part of their Spontaneous Reporting Systems
- NLP Web Service needed to convert narrative data into standard coded data





# PCORTF Collaborative Project – CDC and FDA

- **Two Year Project**
- **Project Goals:**
  - Develop a Natural Language Processing (NLP) Web Service for structuring and standardizing unstructured clinical information
  - Pilot for use in cancer registries and safety surveillance domains
  - Will be expandable to meet other domain requirements



# Stakeholders

**Centers for Disease Control and Prevention (CDC)**

**Food and Drug Administration**

**Hospitals, Cancer Centers, and Universities**

**Local, State, and Regional Public Health Agencies**

**Laboratories**

**Department of Energy**

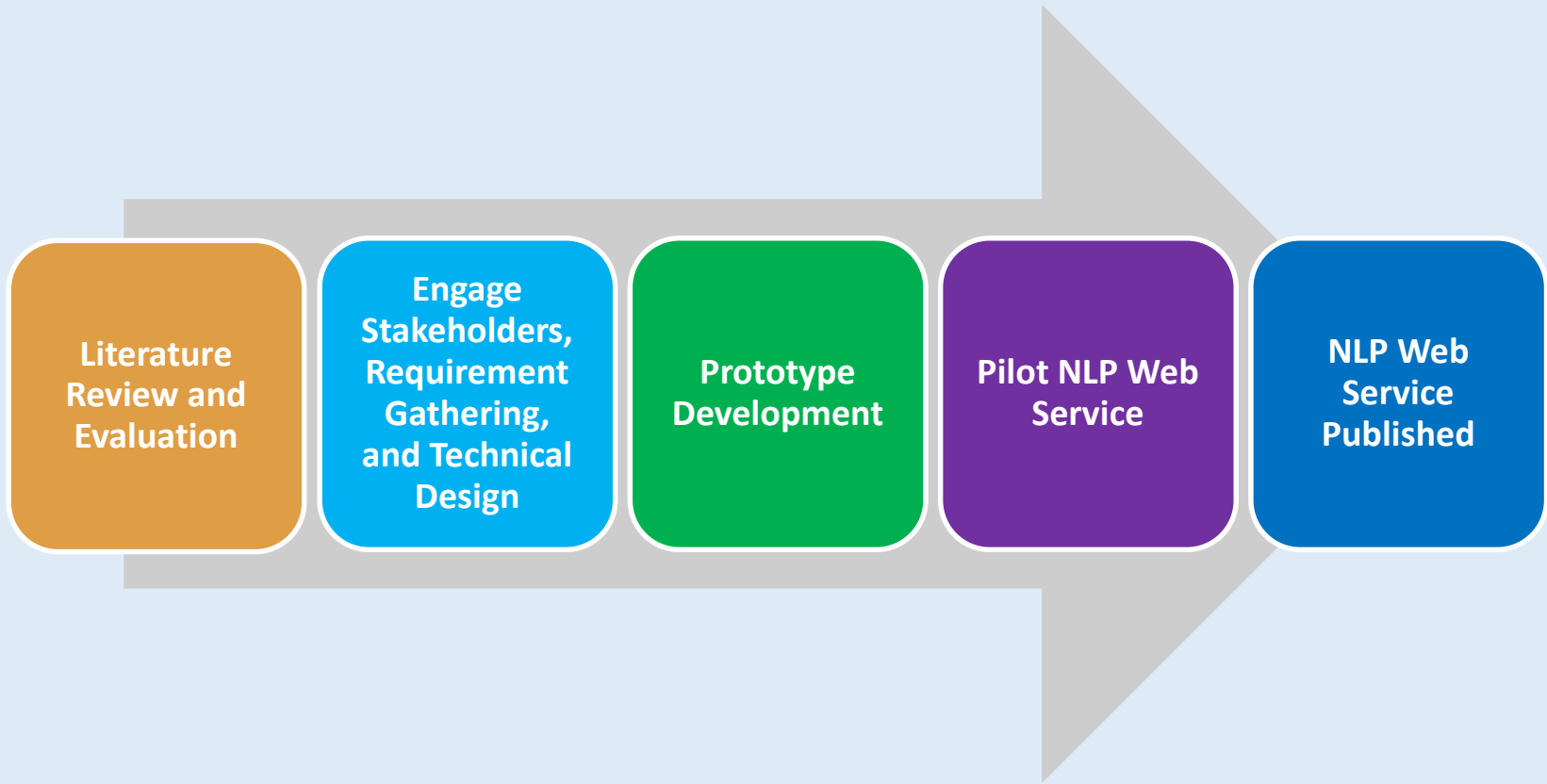
**National Cancer Institute**

**National Library of Medicine**

**PCOR Researchers**



# Overview of Project Activities



# Potential Challenges



Dependencies on open source software

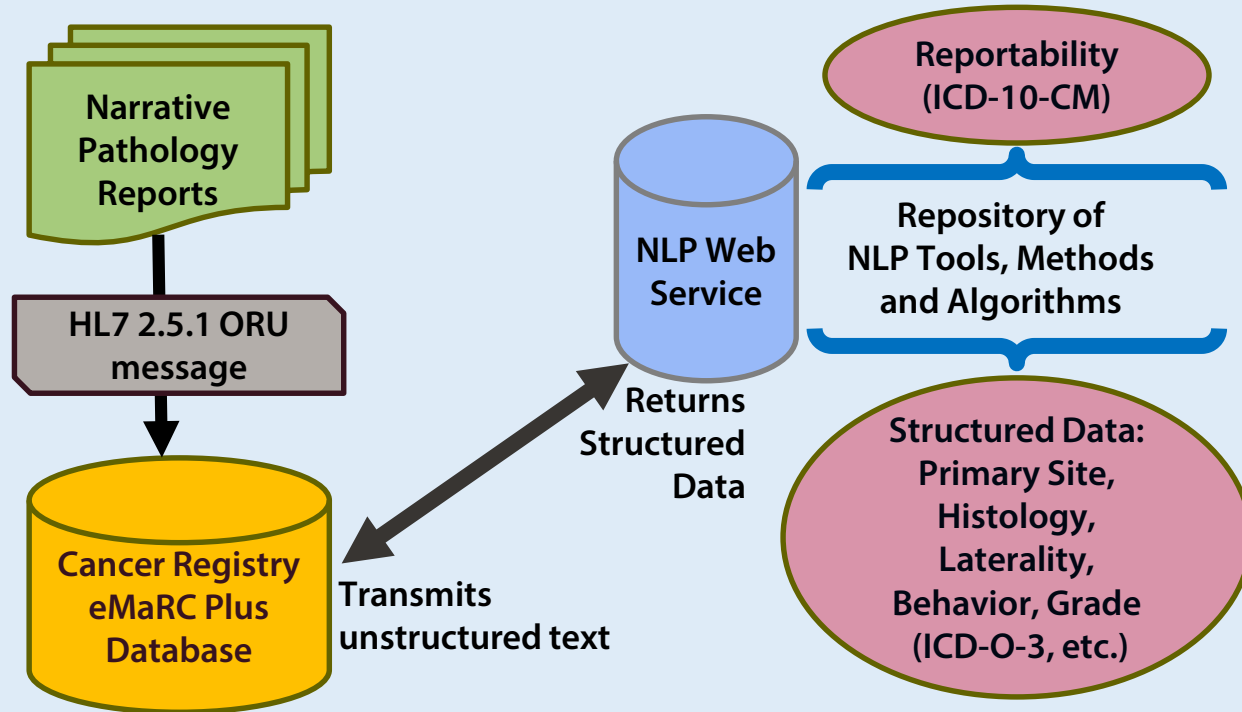
Availability of large test datasets

Limitations of open source NLP software and tools

Inclusion of stakeholders from different domains

Prioritization of requirements

# Conceptual NLP Design for Cancer Pathology





# Acknowledgement of Co-Authors

---



Taxiarchis Botsis



Wendy Blumenthal  
David Butterworth  
Joseph D. Rogers

**Questions?**



**For more information please contact Centers for Disease Control and Prevention**

1600 Clifton Road NE, Atlanta, GA 30333

Telephone: 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348

Visit: [www.cdc.gov](http://www.cdc.gov) | Contact CDC at: 1-800-CDC-INFO or [www.cdc.gov/info](http://www.cdc.gov/info)

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention and the Food and Drug Administration.

