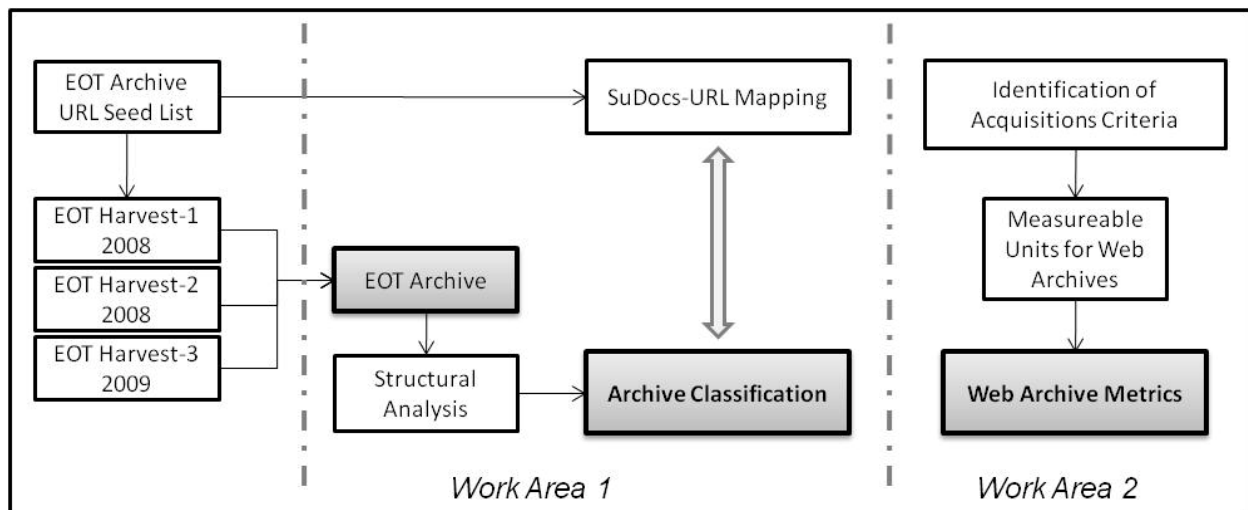# Project Overview

This two-year project will address two collection development needs relative to government information: Selection and Acquisition. A well-established classification scheme exists for federal government information: the Superintendent of Documents (SuDocs) Classification Numbering System. This project proposes to classify the materials in the 2008-2009 End-of-Term (EOT) Web in accord with the SuDocs system.

This proposed classification will enable librarians to select materials consistent with their collection development policies. Additionally, in response to the deficit of established metrics for materials in Web archives, this project proposes to identify metrics that fill the gap and support acquisition.

The project is comprised of two work areas: Archive Classification and Web Archive Metrics (Figure 1).

Figure 1. Project Work Areas

Work Area 1 - Archive Classification

Classification of the Archive will involve structural analysis and human analysis. Link analysis and visualization techniques will identify the organizational and relational structure of the EOT Archive. SMEs will map the Archive's seed URLs to the SuDocs classification system using a Web based application project staff will develop. The resulting classification map, the *SME Map*, will serve as the standard against which the effectiveness of the structural analysis method will be evaluated.

Research Question How effective is the organization of large-scale unstructured Web archives using a pre-defined classification system, the SuDocs classification numbering system, as evaluated by government information librarians?

Work Area 2 - Web Archive Metrics

Identification of metrics for Web archives will be informed by the project's SMEs who will participate in two focus groups to identify and refine the criteria libraries use for acquisition decisions. A tool will be developed to translate these criteria into measurable units appropriate to EOT Archive. An acquisitions exercise will test the effectiveness of this tool. The results will enable librarians to translate measurable units for selected materials in Web archives to units more familiar to libraries and their administrations.

Research Question What measurable units for the materials in Web archives best support management acquisition decisions in libraries?

SME Meeting - January 2010 - Boston, MA

SME Participation

Listed below are the major project work areas with specific SME activities and timeframes identified. The *Narrative* section of the proposal provides more information about the planned activities. As the project progresses specific details will be communicated.

Work Area 1: EOT Archive Classification	SME Activity	Proposed Timeframe
1.1 Data Storage Environment for Research and Testing	-	-
1.2 Structural Analysis	-	-
1.3 Mapping the EOT Seed URL List to the SuDocs Classification Numbering System	-	-
• QA of classification tool	Yes	Feb-Apr 2010
• SME training: Classification of seed-URLs	Yes	Apr 2010 - DLC - Buffalo
• SME Map Exercise: Classification of EOT Archive seed URLs to SuDocs author list	Yes	May 2010
1.4 Classification of Web Site Groupings	-	-
1.5 Evaluation of Work Area 1	-	-
• Focus group discussion and feedback	Yes	Oct 2010- DLC - DC
• SME review	Yes	Jan 2011
Work Area 2: Web Archive Metrics		
2.1. Identification of Acquisitions Criteria	-	-
• Focus group discussion: Identification of acquisitions criteria	Yes	Apr 2010 - DLC - Buffalo
2.2. Determination of Web Archive Measurement Units	-	-
2.3. Acquisitions Exercise	-	-
• Identify libraries' SuDocs selection lists	Yes	Feb 2011
2.4. Evaluation of Work Area		
• Focus group discussion: Evaluation of data in experimental collection reports	Yes	Oct 2011- DLC - DC

Contact Information - University of North Texas

Cathy Hartman	cathy.hartman@unt.edu
Kathleen Murray	kathleen.murray@unt.edu
Mark Phillips	mark.phillips@unt.edu