



Classification of the End-of-Term Archive:  
Extending Collection Development  
Practices to Web Archives

IIPC General Assembly – May 3-7, 2010 – Singapore

# EOT Archive Project

---

- ▶ Who
  - ▶ Library of Congress, the GPO, the Internet Archive (IA), the University of North Texas (UNT) Libraries, and the California Digital Library (CDL)
- ▶ What
  - ▶ Entirety of the federal government's public Web presence
- ▶ When
  - ▶ Before & after the 2009 change in administrations
- ▶ How
  - ▶ Nomination Tool: Websites
  - ▶ Website Harvests: IA, UNT, & CDL
  - ▶ Harvest Consolidation: Library of Congress

# Project Objectives

---

- ▶ EOT Archive Classification
  - ▶ Objective: Classify materials in accord with the Superintendent of Documents (SuDocs) Classification Numbering System
  - ▶ Outcome: Enable librarians to utilize existing selection practices to identify materials in the EOT Archive
- ▶ Web Archive Metrics
  - ▶ Objective: Identify a set of metrics for materials in Web archives
  - ▶ Outcome: Enable characterization of materials in Web archives in units of measurement more familiar to libraries and their administrations

# Research Questions

---

1. 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?
2. What measurable units for the materials in Web archives best support management acquisition decisions in libraries?