Collaborating to Address a Digital Backlog of Web Archives

● ● ●

Jesse Johnston, Grace Thomas, Rick Fitzgerald
Library of Congress

@jesseajohnston
Slides: https://goo.gl/ACpTbH

LIBRARY OF
CONGRESS

Case Study: Applied Research

- *How can we reduce the amount of underdescribed Web materials?*
  - *While: building on organizational strengths, maintaining standard descriptive practice, and achieving higher throughput to describe large-scale digital collections?*
- *Where can we apply inspirations or take lessons from archival description practices?*

@jesseajohnston
Slides: https://goo.gl/ACpTbH

LC Web Archives

- Program since 2000
- Thematic & Event
- Permission to crawl required
- Much available: loc.gov/websites/collections

New Processes & Products

Collaborative Inputs

Recommenders - nominate collections

Cataloger - developed MODS framework

Web Archivist - develop python scripts

Web Services - push to loc.gov site

LIBRARY OF
CONGRESS

Minimal Description

Pulled information from crawl database into minimal MODS framework

More details here

Example (cute overload! ;)): loc.gov, MODS

```
            <identifier>lcwaN0010888</identifier>
▼<titleInfo>
    <title>Cute Overload! ;)</title>
  </titleInfo>
▼<name authority="naf" type="corporate">
  ▼<namePart>
       <!--  TODO: Insert name authority here.  -->
    </namePart>
  </name>
▼<language>
    <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
  </language>
▼<physicalDescription>
    <form authority="marcform">electronic</form>
    <internetMediaType>text/html</internetMediaType>
    <digitalOrigin>born digital</digitalOrigin>
  </physicalDescription>
  <targetAudience>general</targetAudience>
  <typeOfResource>text</typeOfResource>
  <genre authority="marcgt">web site</genre>
▼<originInfo>
  ▼<place>
       <placeTerm type="text">United States</placeTerm>
    </place>
  </originInfo>
▼<abstract>
    Cute Overload was a widely read blog created by Megan Frost featuring
```

Relational Series

ARCHIVED WEB SITE
Beliefnet : Inspiration, Spirituality, Faith

Beliefnet.com offers daily inspiration with news articles on faith, religion, politics, health, family entertainment, sustainable living, and more. Web Site. electronic | Electronic (Form).

Contributor: Beliefnet (Firm)
Date: 2009-01-13

ARCHIVED WEB SITE
Beliefnet: The source for Spirituality, Religion and Morality
beliefnet

Beliefnet: The source for Spirituality, Religion and Morality, a Web Site produced by BeliefNet, a religious organization, is part of the Library of Congress September 11 Web Archive and preserves the web expressions of individuals, groups, the press and institutions in the United States and from around the world in the aftermath of the attacks in the United States on September 11, 2001. Web...

Contributor: Beliefnet
Date: 2001-09-13

ARCHIVED WEB SITE
Beliefnet : spirituality, prayer, God, angels, meditation...

Beliefnet is the largest spirituality and faith website. Whether your are exploring your own religion or other spiritual traditions, you will find inspiration, prayer tools, spritual teachers and an welcoming community. You will also find devotional tools such as bible study guides and online meditations. Web Site. electronic | Electronic (Form).

Contributor: Beliefnet (Firm)
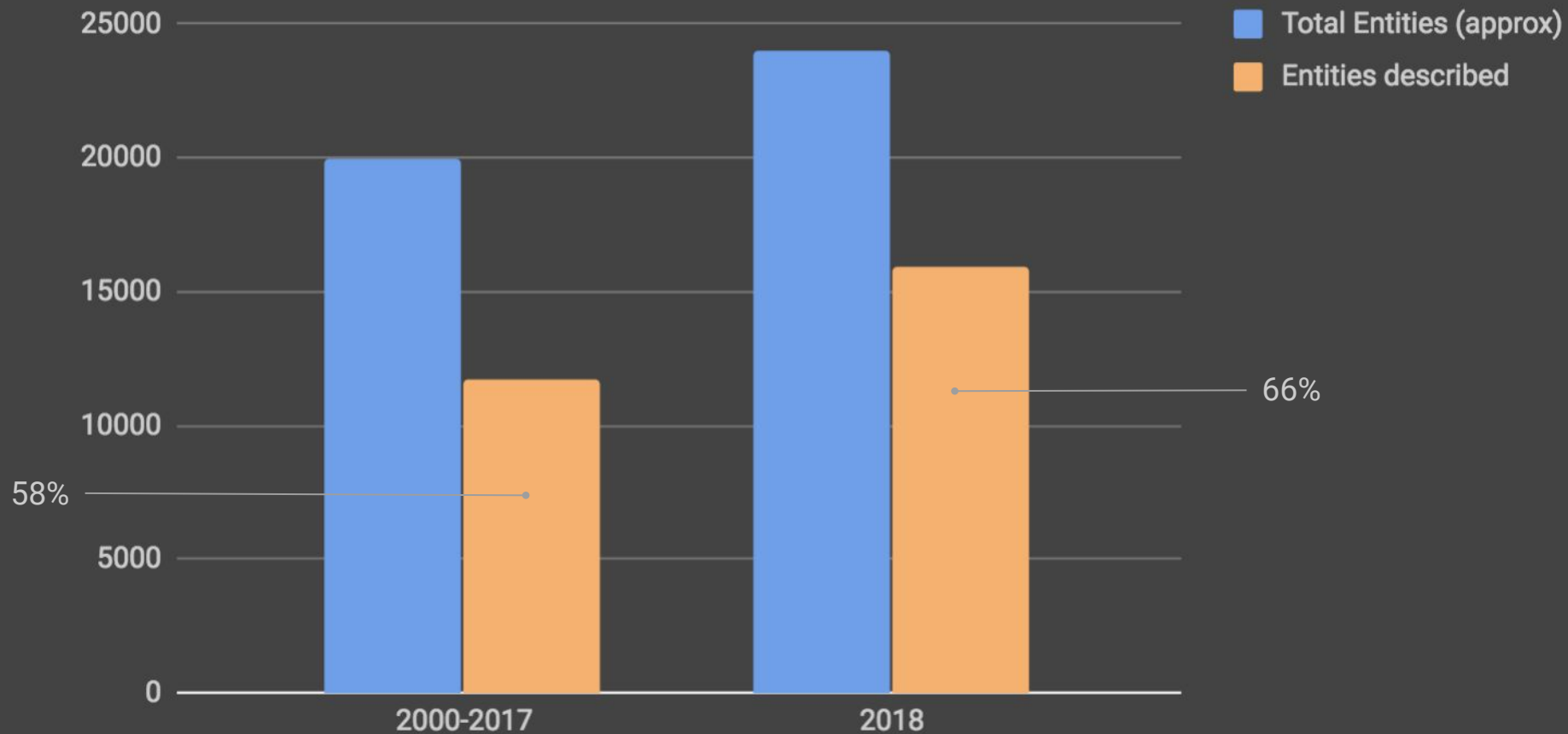**Date:** 2006-03-20

Part of...
Crisis in Darfur 2006 Web Archive (241)
Iraq War 2003 Web Archive (342)
Collections and Services Directorate (508)
African and Middle Eastern Division (1,594)
Public Policy Topics Web Archive (2,358)
Research and Reference Services Division (2,376)
Humanities and Social Sciences Division (9,146)
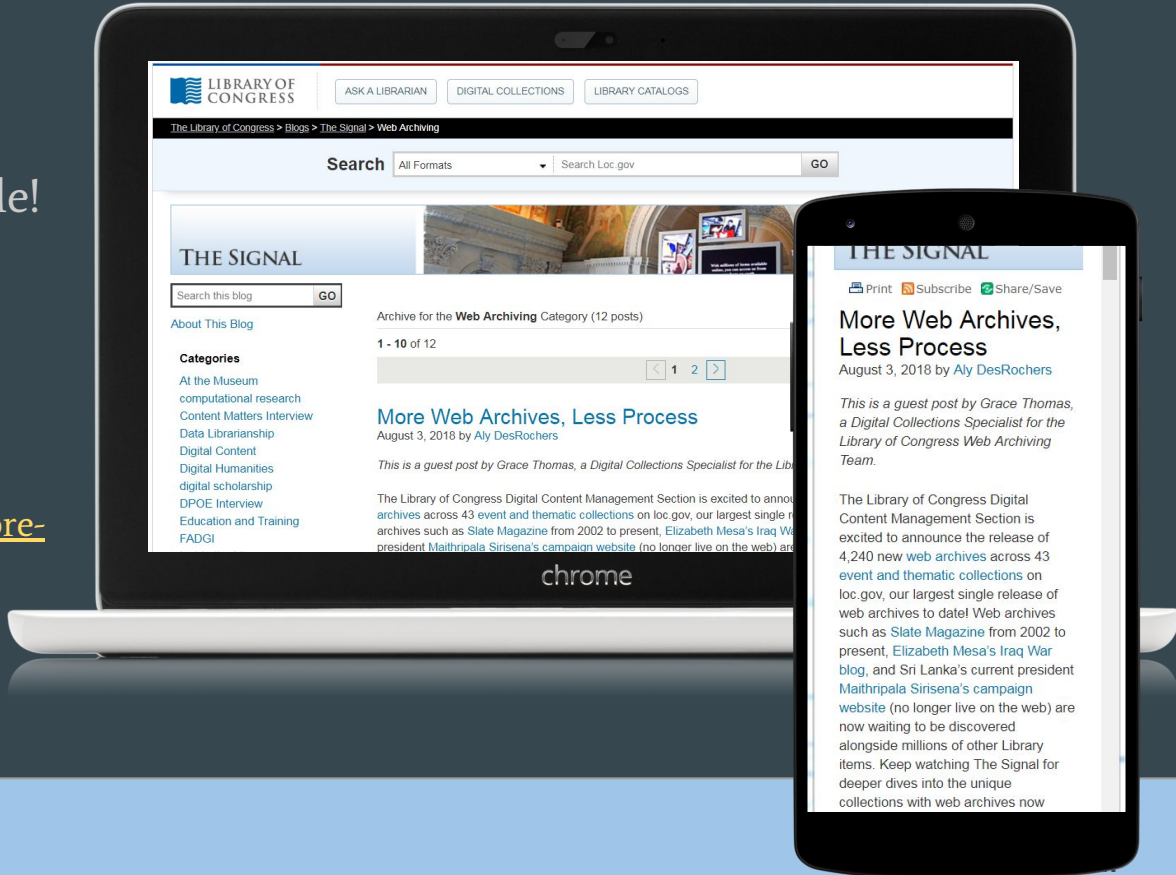
LIBRARY OF CONGRESS

Extent of Web Archives Described

Implications

- More web archives available!
- Alternative to hierarchical description

Longer version available on The Signal:

http://blogs.loc.gov/thesignal/2018/08/more-web-archives-less-process/

LIBRARY OF CONGRESS

ASK A LIBRARIAN   DIGITAL COLLECTIONS   LIBRARY CATALOGS

The Library of Congress > Blogs > The Signal > Web Archiving

Search All Formats   Search Loc.gov   GO

THE SIGNAL

Search this blog   GO

About This Blog

Categories
At the Museum
computational research
Content Matters Interview
Data Librarianship
Digital Content
Digital Humanities
digital scholarship
DPOE Interview
Education and Training
FADGI

Archive for the **Web Archiving** Category (12 posts)

1 - 10 of 12

< 1 2 >

More Web Archives, Less Process
August 3, 2018 by Aly DesRochers

This is a guest post by Grace Thomas, a Digital Collections Specialist for the Lib...

The Library of Congress Digital Content Management Section is excited to anno...
archives across 43 event and thematic collections on loc.gov, our largest single r...
archives such as Slate Magazine from 2002 to present, Elizabeth Mesa's Iraq Wa...
president Maithripala Sirisena's campaign website (no longer live on the web) are

THE SIGNAL

Print Subscribe Share/Save

More Web Archives, Less Process
August 3, 2018 by Aly DesRochers

This is a guest post by Grace Thomas, a Digital Collections Specialist for the Library of Congress Web Archiving Team.

The Library of Congress Digital Content Management Section is excited to announce the release of 4,240 new web archives across 43 event and thematic collections on loc.gov, our largest single release of web archives to date! Web archives such as Slate Magazine from 2002 to present, Elizabeth Mesa's Iraq War blog, and Sri Lanka's current president Maithripala Sirisena's campaign website (no longer live on the web) are now waiting to be discovered alongside millions of other Library items. Keep watching The Signal for deeper dives into the unique collections with web archives now

chrome

Contact

Jesse Johnston

Grace Thomas

Rick Fitzgerald

webcapture@loc.gov

http://www.loc.gov/websites/collections