A Lightweight Digital Library Framework

Lighton Phiri
Supervisor: A/Prof Hussein Suleman

October 3, 2011
Introduction
What is a digital library system?

• Digital Library System (DLS)
  - Information system
  - Collection of digital objects
  - Storage, management, access and preservation of digital objects

• Architectural Design
  - User interface
  - Services
  - Repository
Welcome to UCT CS Document Archive

Welcome to the UCT Computer Science Research Document Archive, which archives and makes accessible documents that are products and by-products of research in the department.

Search the Titles, Authors, Abstracts and Keywords:

Search

Browse

Browse the archive by Subject, Year, Lab or Type.

Latest Additions

View items added to the archive in the past week.

(http://pubs.cs.uct.ac.za/)
Cultural heritage collections

- Bleek and Lloyd Digital Collection

(http://lloydbleekcollection.cs.uct.ac.za/)
Large scale preservation

- Internet Archive

(http://www.archive.org/)
Problem Statement
Motivation

• Complexity of existing solutions
  – Preservation costs
    • Initial, maintenance and recovery costs
    • Funding
  – Resource limitation
    • Skills and education
  – Internet Bandwidth
    • Expensive in developing countries
Related Work
Frameworks

- DELOS DL reference model
  - Abstract framework
- Kahn/Wilensky Framework
  - Generic IS framework
  - DSpace and Fedora
- 5S Framework
  - Formal mathematical model
Popular FOSS tools for creating digital collections

- DSpace
- Fedora Commons
- Greenstone Digital Library Software
- E-Prints
- Duracloud
- Invenio
Research Questions
Previous observations

• 2007
  – Digital Libraries Without Databases
  – in-Browser Digital Library Services

• 2008
  – An African Perspective on Digital Preservation

• 2010
  – Hybrid Online-Offline Collections
Previous observations (cont.)

- simplyCT principles
  - Any metadata, any objects, any services
  - Minimalism
  - Preservation by copying
  - Repeatability
  - Superimposed information
  - Explicitly avoid the use of APIs
Hypotheses

● A formal simplistic abstract framework for Digital Library System design can be derived

● A Digital Library System architectural design based on a simple framework can be easily adopted and managed over time
Research questions

• Is it feasible to implement a Digital Library System using simple architectures?
  - How should simplicity for Digital Library System storage and service architectures be defined?
  - What are the comparative advantages and disadvantages between simple and complex architectures?
Methodology
Approach

- Pilot study
  - Digital library collection prototype
- Theory derivation
  - simplyCT principles
  - Case study of existing tools
- Design and Implementation
  - Extension of prototype DLS
  - Collection integration
Evaluation

- Efficiency
  - Controlled experiments
- Effectiveness
  - Real world case studies
- Usefulness
  - User studies
Work in Progress
Pilot study

• Bleek and Lloyd collection
  – Dublin Core Application profile
  – Fully qualified Dublin Core
  – Virtual and physical objects

• Metadata Object Relationships
  – <dcterms:hasPart>
  – <dcterms:isPartOf>

• Rudimentary digital collection
  – Search and browse
Anticipated Outcome
Anticipated outcome

- A formal definition of simplicity for digital library system design
- A proof of concept Digital Library System
- A set of principles behind the simplyCT framework
Thank you