

# UFDC and Greenstone

Mark Sullivan

July 2005



# UFDC / Greenstone

- ★ **What is UFDC?**
  - University of Florida Digital Collections
  - Complete suite of technologies
  
- ★ **What is Greenstone?**
  - One technology employed in UFDC



# Greenstone Architecture

- ★ **Two Main Components**
  - **Metadata Portion**
    - ★ Indexing Tools
    - ★ Searching Tools
    - ★ Metadata Retrieval
  - **Display Portion**
    - ★ Web Interface for users



# Greenstone Metadata Strengths

- ★ Everything is metadata, even file names, etc..
- ★ Very generic and extensible
- ★ Hierarchical structure
  - Document
  - Section
  - Page
- ★ Syntax for metadata retrieval



# Greenstone Metadata Weaknesses

- ★ Data Issues
  - Everything (except dc.date) defaults to string
  - Not a relational database... flat structure
  - Not a common database protocol [ G2 ]
- ★ Display layer mixed into metadata (html)
- ★ Some missing syntax (subjects – parent:siblings)
- ★ Requires specific format for loading



# Greenstone Display

- ★ **Problems with display**
  - Difficult to affect major layout changes
  - Support for table of contents weak
  - Generally rudimentary
  
- ★ **Greenstone Examples**
  - Site manager's for Table of Contents



# Architectural Decision

- ★ Greenstone

- Will be used for indexing and searching of metadata
- Will be used for retrieval of metadata

- ★ Presentation Layer

- Separate presentation layer
- Running on separate machine
  - ★ Windows, ASP.net, C#
  - ★ SQL Database



**WEB USER**

**PRESENTATION LAYER**

( C#, ASP.net, SQL )

**DATA LAYER**

( GREENSTONE 2.6 )





# WEB USER

## PRESENTATION LAYER

( C#, ASP.net, SQL )

Greenstone

JPEG 2000

Z39.50

# Architectural Decision

- ★ **One common interface**
  - Greenstone
  - JPEG 2000
- ★ **Collection hierarchy**
  - Collection Group Level
  - Collection Level
  - Sub-Collection Level
  - Item Level
- ★ **Separate interfaces**
  - Allows different interfaces to same collection



# Architectural Decision (continued)

- ★ Hierarchical table of contents
- ★ Support for different presentation preferences
  - Low Bandwidth
  - Multiple Languages (banner, home page, etc..)
- ★ Full Citation Page
- ★ Related Image Page
- ★ Static Browse Pages
- ★ Highlighted Item



# Work Required

## \* Interface

- Develop database for presentation data
- Develop structure for design changes
- Framework for Interface
  - \* Home Pages
  - \* Search Pages
  - \* Browse Pages
  - \* Results Pages
  - \* View Single Item

## \* Metadata

- Develop METS profile for UFDC items
- Tools to load new collections

## \* Backlog

- Read from various metadata formats
- (MXF, METS, SGML, INFO, MARC XML)

