

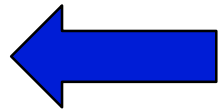


- Vittore Casarosa
  - casarosa@isti.cnr.it
  - Office: 050 621 3115
  - Mobile: 348 397 2168
  - Skype: vittore1201
- “Ricevimento” at the end of the lessons or by appointment
- Final assessment
  - 70% oral examination
  - 30% project (development of a small digital library))
- Reference material:
  - Ian Witten, David Bainbridge, David Nichols, How to build a Digital Library, Morgan Kaufmann, 2010, ISBN 978-0-12-374857-7 (Second edition)
  - Material provided by the teacher
- **<http://cloudone.isti.cnr.it/casarosa/BDG/>**

# Modules



- Computer Fundamentals and Networking
- A conceptual model for Digital Libraries
- Bibliographic records and metadata
- Information Retrieval and Search Engines
- Knowledge representation
- Digital Libraries and the Web
- Hands-on laboratory: the Greenstone system

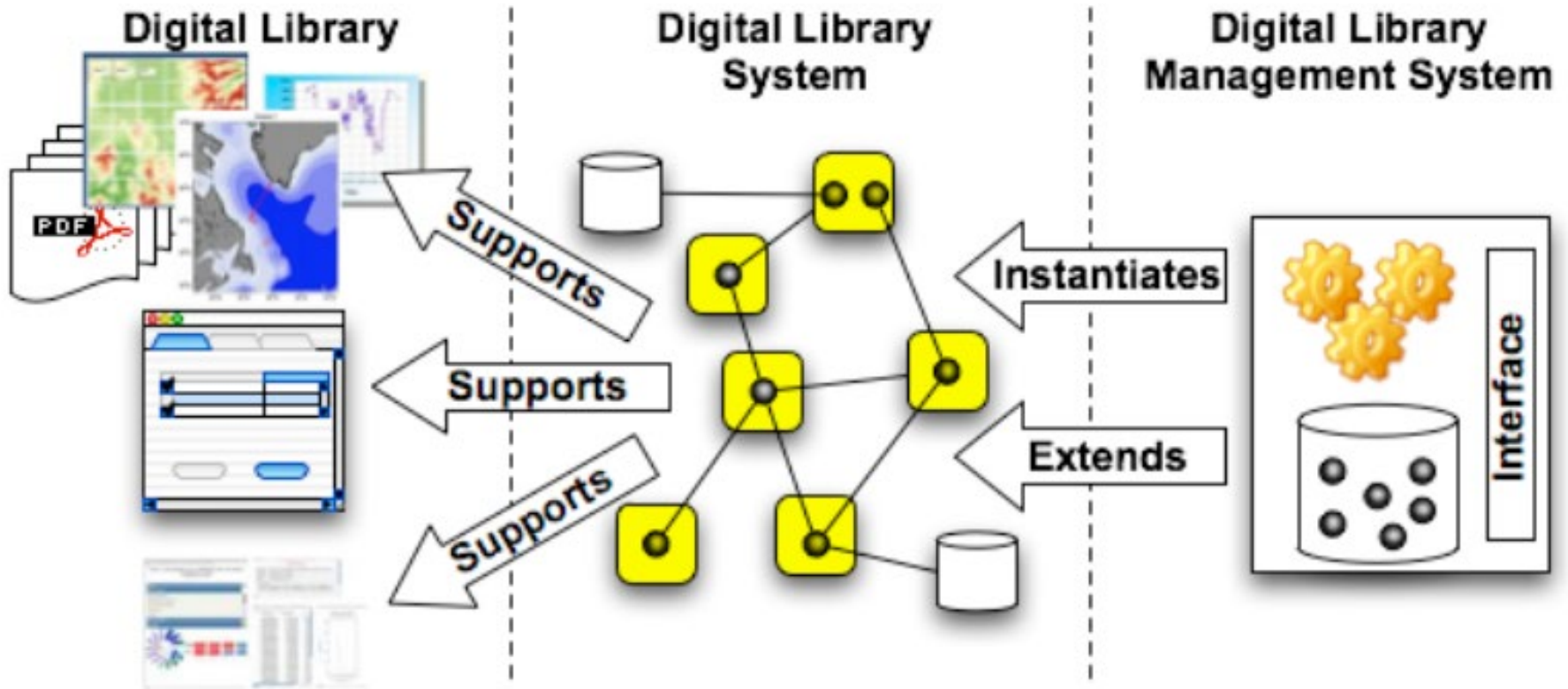
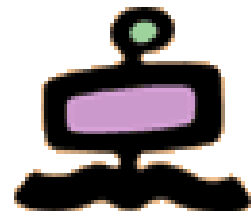


# Greenstone



- Hands-on laboratory: the Greenstone system
  - What is Greenstone
  - Greenstone installation
  - The Greenstone Librarian Interface
  - Development of a small digital library

# A Three-Entity Framework



# Different types of DLMS



- **Extensible Digital Library System**

- A complete Digital Library System that is fully operational with respect to basic/ foundational functionality required. It is based on an open software architecture, so that further software components can be incorporated on top of the ones already there with ease (Greenstone, DelosDLMS)

- **Digital Library System Warehouse**

- A collection of software components that encapsulate the core suite of DL functionality and a set of tools that can be used to combine these components in a variety of ways (in Lego-like fashion) to create Digital Library Systems offering a tailored integration of functionalities. New software components can easily be incorporated into the Warehouse for subsequent combination with those already there (BRICKS, DILIGENT)

- **Digital Library System Generator**

- A highly parameterized software system that encapsulates templates covering a broad range of functionalities, including a defined core suite of DL functionality as well as any advanced functionality that has been deemed appropriate to meet the needs of the specific application domain. Through an initialization session, the appropriate parameters are set and configured; at the end of that session, an application is automatically generated, and this constitutes the Digital Library System ready for installation and deployment (MARIAN)



# Introducing the Greenstone Digital Library Software

**Ian H. Witten**

Computer Science Department  
Waikato University  
New Zealand

<http://greenstone.org>  
<http://nzdl.org>

# Greenstone



## Access

- ❖ Accessible via any Web browser
- ❖ Server runs on anything (all Windows + Unix + Mac)
- ❖ Collections can be published on CD-ROM/DVD
- ❖ Trivial to install
- ❖ GUI interface for building and publishing collections

## Searching/ browsing

- ❖ Collection-specific
- ❖ Full-text and fielded search
- ❖ Flexible browsing facilities
- ❖ Metadata-based (Dublin Core recommended)
- ❖ Creates all access structures automatically

## Extensible

- ❖ Plugins — new document, metadata formats
- ❖ Classifiers — new metadata browsers

## Multi-\*

- ❖ Multilingual: Documents *and* interfaces
- ❖ Multimedia: image, video, audio collections exist
- ❖ Multiformat: Documents *and* metadata



- Metadata
- ❖ Can use any metadata set, Dublin Core supplied
  - ❖ Plugins for

<b>XML</b>	<b>Refer</b>
<b>MARC</b>	<b>OAI</b>
<b>CDS/ISIS</b>	<b>METS</b>
<b>ProCite</b>	<b>DSpace</b>
<b>BibTex</b>	

- ❖ Web
- ❖ Can publish Greenstone collections on CD-ROM

- Serving
- ❖ Can publish Greenstone collections on OAI
  - ❖ Export collections to METS
  - ❖ Export collections to DSpace (*ready for DSpace's batch import program*)

- Documents
- ❖ Plugins for

<b>PDF</b>	<b>ZIP</b>	<b>Images (GIF, JPEG, TIFF ...)</b>
<b>PostScript</b>	<b>Excel</b>	<b>MP3</b>
<b>Word, RTF</b>	<b>PPT</b>	<b>Ogg Vorbis</b>
<b>HTML</b>	<b>Email</b>	<b>MediaWiki</b>
<b>Plain text</b>	<b>Source code</b>	<b>UnknownPlug</b>
<b>Latex</b>	<b>RealMedia</b>	<b>(e.g. for audio, MPEG, Midi)</b>



# The Greenstone system



- Greenstone is “free software” that can be downloaded from the Web
  - <http://www.greenstone.org/download>
- The Greenstone download installs two “programs”
  - The Greenstone server (the Apache web server) supporting access to and navigation through “collections” (digital libraries) via any browser
  - The Greenstone Librarian Interface (GLI) to build “collections” (digital libraries)

# The Greenstone system



- There are two versions
  - Greenstone 2, latest version 2.87, maintained but no longer developed
  - Greenstone 3 is actively maintained and developed; the latest version is 3.11
- Operating system (for both versions):
  - Windows (any version)
  - Linux (any version)
  - Mac OS X

# The Greenstone Librarian Interface (GLI)



- ❖ Building collections
- ❖ Interactive Java program
- ❖ Runs on anything
- ❖ Build a collection on the computer you are on
- ❖ Includes metadata editor

*Caveat: cannot deal with such huge metadata collections as Greenstone can*

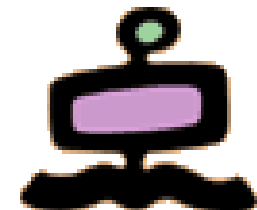


**(Tutorial exercise: small collection of HTML files)**

**Invoke GLI: build a small collection of HTML files**

- ❖ **Gather**
- ❖ **Create**
- ❖ **Look at extracted metadata**
- ❖ **Set up shortcut in the Librarian interface**

# Create a new collection



The screenshot shows the Greenstone Librarian Interface (GLI) window. The title bar reads "Greenstone Librarian Interface Mode: Librarian Collection: No Collection". The menu bar includes "File" and "Edit". The toolbar contains icons for "Download", "Gather", "Enrich", "Design", and "Create". The "Gather" button is highlighted. The main workspace is divided into "Workspace" and "Collection" panes. The "Workspace" pane shows a tree view with folders: "Documents in Greenstone Collections", "Local Filespace", "Home Folder (lib)", "Dow", and "sam". A dialog box is open in the foreground, titled "Create". It contains the following text and fields:

To create a new collection fill out the fields below.

Collection title:

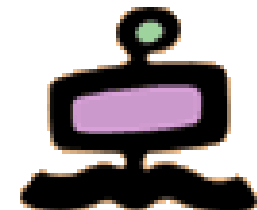
Description of content:

Base this collection on:

Buttons: OK, Cancel

At the bottom of the GLI window, there is a status bar with "Show Files" and "All Files" buttons. Below this, a message box displays "No action requested" with a "Stop" button. To the right of the message box are icons for a folder and a trash bin.

# Gather: Gather the files together



Greenstone Librarian Interface Mode: Librarian Collection: About hobbits (about01)

File Edit Help

Download Gather Enrich Design Create

**Workspace**

- Local Filespace
- Home Folder (ihw)
- Downloaded Files
- sample\_files
  - beatles
  - difficult\_documents
  - dspace
  - hobbits
    - bilbo\_files
    - Frodo\_files
    - merry\_files
    - pippin\_files
    - sam\_files
    - bilbo.html
    - Frodo.html
    - merry.html
    - pippin.html
    - sam.html
  - images

**Collection**

- bilbo.html
- Frodo.html
- merry.html
- pippin.html
- sam.html
- bilbo\_files
- Frodo\_files
- merry\_files
- pippin\_files
- sam\_files

Show Files All Files Show Files All Files

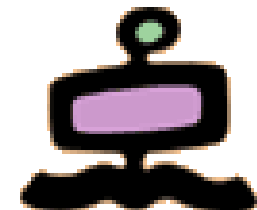
Copying ...sample\_files\hobbits\bilbo\_files\bilbo2.gif

92%

Stop



# Create: Build the collection



Greenstone Librarian Interface Mode: Librarian Collection: About hobbits (about01)

File Edit Help

Download Gather Enrich Design Create

maxdocs 1

Complete Rebuild 50%

Incremental Rebuild

Build Collection Cancel Build Preview Collection

```
***** Import Started *****
The file bilbo.html is being processed by HTMLPlug.
The file Frodo.html is being processed by HTMLPlug.
The file merry.html is being processed by HTMLPlug.
The file pippin.html is being processed by HTMLPlug.
The file sam.html is being processed by HTMLPlug.
***** Import Finished *****
5 documents were considered for processing:
  5 documents were processed and included in the collection.
```

# Preview: admire the result



HOME HELP PREFERENCES

## About hobbits

about

search titles a-z filenames

Search for  that contain  of the words

Begin Search

### About this collection

A little collection about little people

### How to find information in the About hobbits collection

There are 3 ways to find information in this collection:

- search for particular words
- access publications by title
- access publications by filename

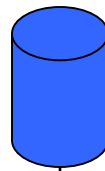
You can *search for particular words* that appear in the text from the "search" page. This is the first page that comes up when you begin, and can be reached from other pages by pressing the *search* button.

You can *access publications by title* by pressing the *titles a-z* button. This brings up a list of books in alphabetic order.

You can *access publications by filename* by pressing the *filenames* button. This brings up a list of entries, sorted by original filename.



Set up environment variables

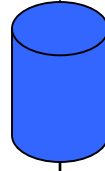


Details about the collection



Makecol

Put source docs into a subdirectory

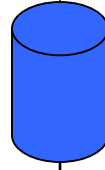


collect.cfg (plugins)



Import

Docs in Greenstone Archive format

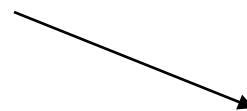
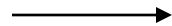


collect.cfg

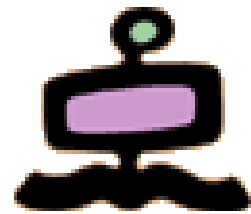


Build

Greenstone collection



# Building a collection



Create a directory for the collection (with subdirectories), put collect.cfg file in "etc" subdirectory

Convert to archive format  
Extract metadata

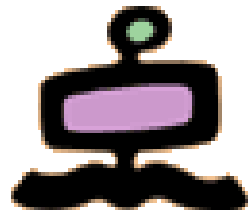
create indexing & browsing structures, compress ...

Search

Results



# Greenstone useful links



- The starting place (it includes the downloads)
  - <http://www.greenstone.org/>
- Download
  - <http://www.greenstone.org/download>
- Documentation
  - <http://wiki.greenstone.org/doku.php>
- Extensive set of implemented (many well documented) Digital Libraries
  - <http://www.nzdl.org/cgi-bin/library.cgi>  
(go down the page and look at  
“Documented Example collections”)

# Greenstone exercises



- Tutorial exercises

GS3 and GS2 (check the tabs)

- [https://wiki.greenstone.org/doku.php?id=en:tutorials#tab\\_\\_greenstone3](https://wiki.greenstone.org/doku.php?id=en:tutorials#tab__greenstone3)

- Sample files for GS3

- [http://wiki.greenstone.org/gsdoc/tutorial/gs3-current/sample\\_files/sample\\_files.zip](http://wiki.greenstone.org/gsdoc/tutorial/gs3-current/sample_files/sample_files.zip)

# Steps to build a collection



- Put together (from your PC or downloading from the Web) all the documents that will be in your collection
- **Design how your collection will be “navigated”**
- Based on the “Design”, organize the documents in a hierarchy of folders and subfolders
- Based on the “Design”, decide which metadata will be needed
- **Open the GII and define your new collection** (name and description)
- Import a **SMALL representative subset** of your documents into Greenstone (drag and drop from “Gather”)
- “Enrich” your collection with metadata
- Based on the “Design” define the indexes and the classifiers to search and navigate your collection
- “Create” the collection (Greenstone will build indexes etc.)
- Admire the result 😊
- **Personalize the “Format”**