

# Open Source “Digital Library” Systems: Streetprint & Greenstone



## An introduction and comparison

Prepared by Martha Chantiny,  
Head, Desktop Network Services

114<sup>th</sup> Annual ASEE Conference Workshop 2  
University of Hawaii at Manoa, Hamilton Library  
June 24, 2007

# Systems to make digital collections available, searchable & harvestable

- NOT a complex Content Management System
- Intended for the one-person project/organization
- Does not require technological or programming wizard level skills
- Mostly platform agnostic

# History

[http://wiki.greenstone.org/wiki/gsdoc/others/Greenstone\\_history.htm](http://wiki.greenstone.org/wiki/gsdoc/others/Greenstone_history.htm)

## **Greenstone** – created 1997-98

- ◆ Greenstone software grew out of the New Zealand Digital Library Project at the University of Waikato
- ◆ Developed and distributed in cooperation with UNESCO
- ◆ Licensed under the Gnu Public License (GPL)
- ◆ First CD-ROM the *Humanity Development Library 1.3* issued in April 1998

# History

<http://www.crcstudio.arts.ualberta.ca/streetprintorg/streetprintMLA.pdf>

## **Streetprint** – created 2003

- ◆ Grew out of Canada Research Chair Humanities Computing Studio project to create “computerized and integrated multi-media humanities research centre to unite humanities research and digital technology “
- ◆ Funded by Canada Research Chairs Infrastructure Fund \$54,860 in June 2001
- ◆ Licensed under the Gnu Public License (GPL)
- ◆ First collection was digitized British street literature *Streetprint: Revolution and Romanticism*

# Pros & Cons

## FEATURES

### Greenstone

- ◆ CD runtime output
- ◆ Any metadata scheme
- ◆ Cross-collection searching
- ◆ Large user base, active listserv
- ◆ Multi-lingual
- ◆ Highly scaleable

### Streetprint

- ◆ CSS & templates provided
- ◆ PHP (easy to change functionality)
- ◆ Comment feature
- ◆ Easy to use

## Pros & Cons

# DRAWBACKS

### Greenstone

- ◆ Complex mix of specialized code – more difficult to understand - mix of GUI and coding
- ◆ Requires more SysAdmin level knowledge & skills

### Streetprint

- ◆ No batch load
- ◆ Code not bug free or finished
- ◆ Only Dublin Core
- ◆ No cross-collection search
- ◆ Small user base, inactive lists
- ◆ How scaleable?

# Installation - Streetprint

1. A server (Linux or Mac with OSX or a Windows server)
2. Web server software (e.g. Apache)
3. MySQL version 3.23 or higher
4. PHP version 4.1 or higher
5. Know how to download & expand tar.gz or zip files
6. Knowledge of (or access to someone who can) change permissions to allow user running apache to write to the Streetprint directory, create databases, & assign rights in MySQL, configure web server software for PHP
7. Desktop computer with any web browser

# Installation - Greenstone

1. A server (Linux or Mac with OSX or a Windows server)
2. Web server software (e.g. Apache)
3. PERL/CGI
4. GCC Gnu C++ compiler and GDBM Gnu database manager compiler (if the precompiled binaries don't work for you)
5. Knowledge of how to download and correctly expand tar.gz or zip files
6. Ability to (or access to someone who can) change web server config & cgi-bin config files; add users
7. Desktop computer with Java Runtime Environment software, installed GLI java applet and web browser





---

# Setting up a collection –Streetprint Database creation

---

**Geek Alert!** This chapter of the Streetprint Engine manual is by far the geekiest of the lot. If you need any help with this section, we recommend you consult your nearest internet wizard or technical support staff member.

# Setting up a collection –Streetprint Name the DB in MySQL

**Streetprint Engine 2 Installer**

There are just a few settings we need...

**MySQL server name:**   
This is nearly always localhost, so if you don't know use localhost.

**MySQL username:**   
Fill in the username you need to connect to your MySQL database, if you don't know, try the username of your ftp account as they are often the same.

**MySQL password:**   
Fill in the password you need to connect to your MySQL database, if you don't know, try the password of your ftp account as they are often the same.

**MySQL Database name:**   
Fill in the name of the database that Streetprint will store it's data in, make sure it is empty or at least doesn't contain important (it is very likely that it will be destroyed).  
If this database does not exist, we will try to create it.

# Setting up a collection –Streetprint

A STREETPRINT SITE

**STREETPRINT**  
engine  
SITE ADMINISTRATION

[Home](#) > [Login](#)

Welcome to A Streetprint Site.

**login**

Username:

Password:

Remember Me:

A STREETPRINT SITE

**STREETPRINT**  
engine  
SITE ADMINISTRATION

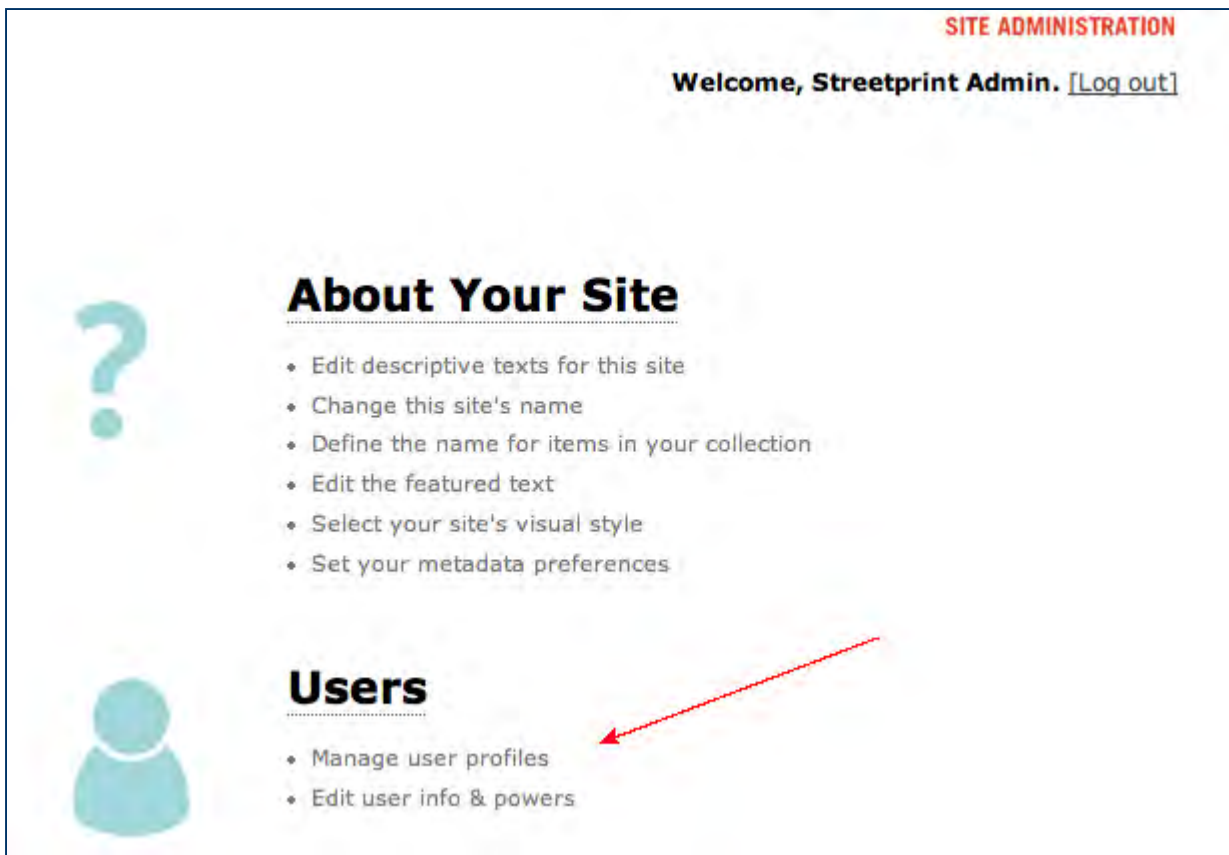
Welcome, Streetprint Admin. [[Log out](#)]

[Home](#)

[SITE CONFIGURATION](#)

[MY COLLECTION](#)

# Setting up a collection –Streetprint User profiles



**SITE ADMINISTRATION**

Welcome, Streetprint Admin. [\[Log out\]](#)

## About Your Site

- Edit descriptive texts for this site
- Change this site's name
- Define the name for items in your collection
- Edit the featured text
- Select your site's visual style
- Set your metadata preferences

## Users

- Manage user profiles
- Edit user info & powers

# Setting up a collection –Streetprint User profiles

[Home](#) > [Site Configuration](#) > [Users](#) > [Add User](#)

**user details**

**add user**


**Username**

**Password:**

**Confirm Password:**

**Real Name**

**Email**

**Role**  

**Bio**


**Show On Contact Page:**

[\[back\]](#)

# Setting up a collection - Greenstone

## greenstone digital library software

HELP

- 
- The Librarian Interface** Helps you create new collections, modify or add to existing ones, or delete collections
  - Administration Page** Allows you to add new users, summarizes the collections in the system, gives technical information on the Greenstone installation
  - About Greenstone** Tells you about the Greenstone software and the New Zealand Digital Library Project where it originated

Greenstone is a suite of software for building and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet or on CD-ROM. Greenstone is produced by the **New Zealand Digital Library Project** at the **University of Waikato**, and developed and distributed in cooperation with UNESCO and the **Hunan Info** NGO. It is open-source software, available from <http://greenstone.org> under the terms of the GNU General Public License.

The aim of the software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. Digital libraries are radically reforming how information is disseminated and acquired in UNESCO's partner communities and institutions in the fields of education, science and culture around the world, and particularly in developing countries. We hope that this software will encourage the effective deployment of digital libraries to share information and place it in the public domain.

# Setting up a collection - Greenstone User Accounts



[admin home](#)  
[Greenstone home](#)

## Configuration files

[gsdlsite.cfg](#) → Info only  
[main.cfg](#) → Editable

## Logs

[usage log](#)  
[error log](#)

## User management

[list users](#)  
[add a new user](#) ← Info only  
[change password](#)

## Technical information

[general](#)  
[arguments](#)  
[actions](#)  
[browsers](#)  
[protocols](#) } Info only

## Administration

Greenstone version number: 2.72

Maintenance and administration services available include:

- ♦ view on-line logs
- ♦ create, maintain and update collections
- ♦ access technical information such as CGI arguments

These services are accessed using the side navigation bar on the lefthand side of the page.

## Collection Status

Collections will only appear as "running" if their build.cfg files exist, are readable, contain a valid collection's index directory (i.e. NOT the building directory).

# Setting up a collection - Greenstone User Accounts

## Add a new user

username  Usernames must be between 2 and 30 characters long. They can contain letters, numbers, and '\_'.

password  Passwords must be between 3 and 8 characters long. They can contain letters, numbers, and '\_'.

account status

groups  Groups is a comma separated list, do not put spaces after the commas.

comment



[admin home](#)  
[Greenstone home](#)

### Configuration files

[gds/site.cfg](#)  
[main.cfg](#)

### Logs

[usage log](#)  
[error log](#)

### User management

[list users](#)  
[add a new user](#)  
[change password](#)

## List of current users

user	account status	groups
DNSstudent	enabled	<u>all-collections-editor,remote-superuser,remote-collection-builder</u>
administrator		administrator,all-collections-editor,colbuilder,remote-superuser
remote-superuser		remote-superuser,all-collections-editor
remote-superuser		remote-superuser,all-collections-editor
all-collections-editor		all-collections-editor,remote-superuser,remote-collection-builder
all-collections-editor		all-collections-editor,remote-superuser,remote-collection-builder

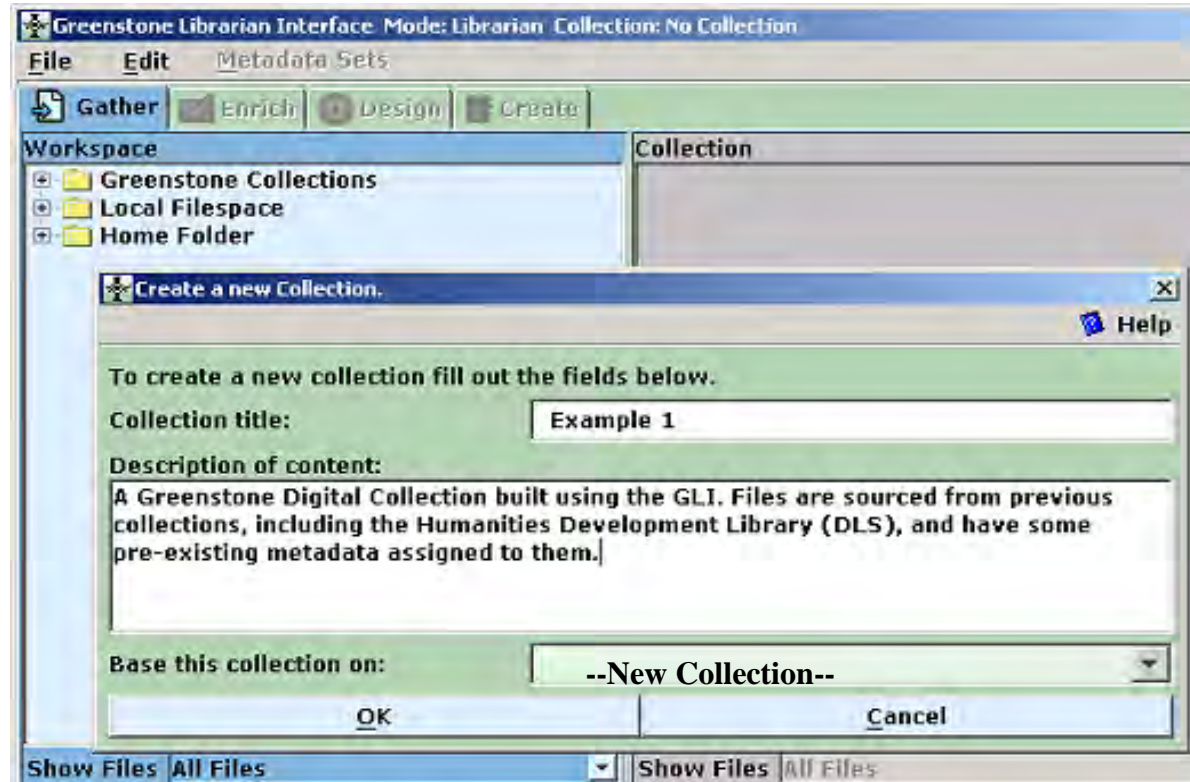
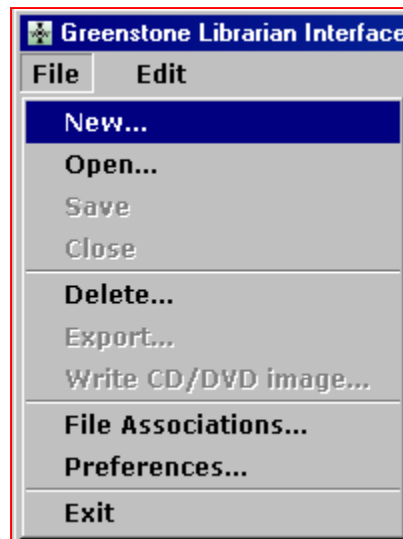


# Setting up a collection –Greenstone Database creation

The installation begins by asking which directory to install in. By default, all Greenstone software is installed in the directory */usr/local/gSDL* if it is the root user who is doing the installation. Next you choose to install binaries or source code. You are then asked some questions about your webserver setup. A valid cgi executable directory (normally called “cgi-bin” on Unix systems) is required and you need to enter the web address of the cgi directory. A directory called “collect” will be created during the install process.

# Setting up a collection –Greenstone Database creation

“Greenstone Librarian Interface” or GLI applet should be installed at this point



# Setting up a collection – Greenstone Description, Contact, About

Greenstone Librarian Interface v2.72 Mode: Expert Collection: word and pdf (wordandp)

File Edit Help

Download Gather Enrich Design Create Format

General Search Format Features Translate Text Cross-Collection Search Collection Specific Macros

**General** Help for this screen

Creator's email

Maintainer's email

Collection title: word and pdf

Collection folder: wordandp

URL to 'about page' image: Browse...

URL to 'home page' image: Browse...

This collection should be publicly accessible

Collection description:

A description of the purpose and content of the collection

Preview Collection

# Setting up a collection – Streetprint Description, Contact, About

EDIT

## about your site

### front page

Write your own text for the front page of your Streetprint site.

HTML accepted. Paragraph and line break tags will be added for you automatically.



**Page Title**

Streetprint

**Site Name**

A Streetprint Site

**Welcome blurb:**

Welcome to A Streetprint Site. (Put your welcome text here!)

# Setting up a collection – Streetprint Description, Contact, About

## about page

Write your own text and specify options for the "about" page of your Streetprint site.

HTML accepted. Paragraph and line break tags will be added for you automatically.



### About the Project:

A detailed description of your Streetprint site and/or your collection's history.

Put paragraphs describing your streetprint site here.

### About Digitization Procedures:

An optional field. Describe the methodology and equipment used for scanning the artifacts in your Streetprint site.

Put text describing your archiving practices and methodologies here.

Submit About Page Changes

## fineprint

Write a line or two of "fine print". This text will appear at the bottom of every page, along with the standard Streetprint Engine fineprint.



Fine Print

Submit Fineprint Changes

[back]

# Adding Data - Streetprint

[Home](#) > [Site Configuration](#) > [Type Management](#)

## types

- **document types**
- **categories**
- **image types**
- **media types**
- image resolutions

[\[back\]](#)



**Important note:** By default, most PHP servers limit uploaded file sizes to 2MB or smaller. If you plan to add media files which are larger than 2MB, you may need to contact your server administrator and have them change this setting first.

# Adding Data - Streetprint Document type

**document types**  
Default [\[edit\]](#)

---

**add document type**

**Name**

**Description:**


**Default Image Type:**

---

**delete document type**

Default

[\[back\]](#)



This can be used to distinguish items in the collection, e.g. different report series, or different sizes

# Adding Data - Streetprint Defining types

## types

### image types

Back [\[edit\]](#)

Front [\[edit\]](#)

Inside [\[edit\]](#)

### add image type

Name

Description:

Add Image Type

### delete image type

Back

Front

Inside

Delete Image Type

[\[back\]](#)



## types

### media types

MP3 Audio File [\[edit\]](#)

Quicktime Movie [\[edit\]](#)

### add media type

Name

Description:

Extension

Add Media Type



# Adding Data - Streetprint

## Defining types

Category =  
Subject Heading

By default this  
creates a pull-  
down list in the  
Add/Edit mode

The screenshot shows a web interface titled "types". It has two main sections: "add category" and "delete categories".

The "add category" section includes a form with two input fields: "Name" and "Description:". Below the form is a button labeled "Add Category".

The "delete categories" section includes a button labeled "Delete Category".

At the bottom left of the interface is a link labeled "[back]".

On the right side of the interface, there are two overlapping mobile device screens displaying a list of items, likely representing the pull-down list mentioned in the text.

# Adding Data - Streetprint Metadata

## about your site

### editing metadata settings

**What is metadata?** Simply put, metadata is "data about data." While your Streetprint site is unique, it can be useful to provide ways for other organizations and databases to be able to "read" your information in a common format.

The Streetprint Engine currently provides basic metadata using the [Dublin Core Metadata Initiative's metadata terms](#). If the checkbox below is checked, DCMI-compliant metadata tags will be added to the source code of your Streetprint pages. While this change will not be visible to most browsers, any software which is "metadata aware" can use this additional information.

Checking the box below also creates a [metadata repository](#) for your Streetprint site which conforms to guidelines from the [Open Archives Initiative](#). You may optionally register this repository in the OAI's directory by using the form on [this page](#). Programmers who would like to fine-tune their site's metadata settings can do so by editing the "SPDublinCore.obj" file in Streetprint's objects directory.

Team Streetprint welcomes feedback on our current beta implementation of metadata. [Visit our forums](#) and let us know what you think!

**Enable DCMI Metadata**

[\[back\]](#)



# Adding Data – Greenstone Plug-ins (\*.pm)



# Adding Data – Greenstone

## more about plug-ins

*From the documentation:*

Two types, metadata and document plugins to process files, some

- group several files into one document,
- some split one file into several documents, and
- some have a one to one mapping.

### **ImagePlug**

- uses the imagemagic utilities to
- create derivatives(thumbnail images) and,
- extract image metadata (width height format)

ImagePlug can easily be extended to extract more extensive image metadata if required

# Greenstone customization Plug-ins

Please configure the arguments for ImagePlug.

ImagePlug	
<input type="checkbox"/> process_exp	{?i}(\.jpe?g \.gif \.png \.bmp \.x \.tiff?)\$
<input type="checkbox"/> noscaleup	
<input type="checkbox"/> thumbnailsize	100
<input type="checkbox"/> thumbnailtype	gif thumbnailsize: Make thumbnails of size nxn.
<input type="checkbox"/> screenviewsize	1
<input type="checkbox"/> screenviewtype	jpg
<input type="checkbox"/> converttotype	
<input type="checkbox"/> minimumsize	100

BasPlug	
<input type="checkbox"/> block_exp	
<input type="checkbox"/> smart_block	
<input type="checkbox"/> associate_ext	
<input type="checkbox"/> associate_tail_re	
<input type="checkbox"/> input_encoding	auto - Use text categorization algorithm to automatically identify the encoding of each source d...
<input type="checkbox"/> default_encoding	utf8 - Either utf8 or unicode -- automatically detected.
<input type="checkbox"/> extract_language	
<input type="checkbox"/> default_language	en
<input type="checkbox"/> extract_acronyms	
<input type="checkbox"/> markup_acronyms	

# Adding Data – Greenstone configuring plug-ins

I am building a collection that is mainly JPEG files, and 1 PDF file of a pamphlet. Is it possible to type a description page in MS Word for each JPEG in the collection, and associate the word document with the JPEG so that when the JPEG image is retrieved, the user also has the option of clicking an icon and reading MS Word document description of the image.

Do you want the text in the Word document to be indexed? What you could try is naming your documents like photo1.jpg, photo1.doc i.e. same filename for word doc and image doc. Then have WordPlug and ImagePlug in the collection. On one of the plugins use the -associate-ext option - if you add it to WordPlug, set the value of the option to jpg. If you use the option on ImagePlug, set the value to doc. This will mean that one version is processed as a document, the other is associated with it. Which way round you do this affects the processing of the files. (whether you get the word document indexed or not, whether you get the image processed into thumbnails etc) You'll also need to play around with format statements.



---

# Diverging Paths

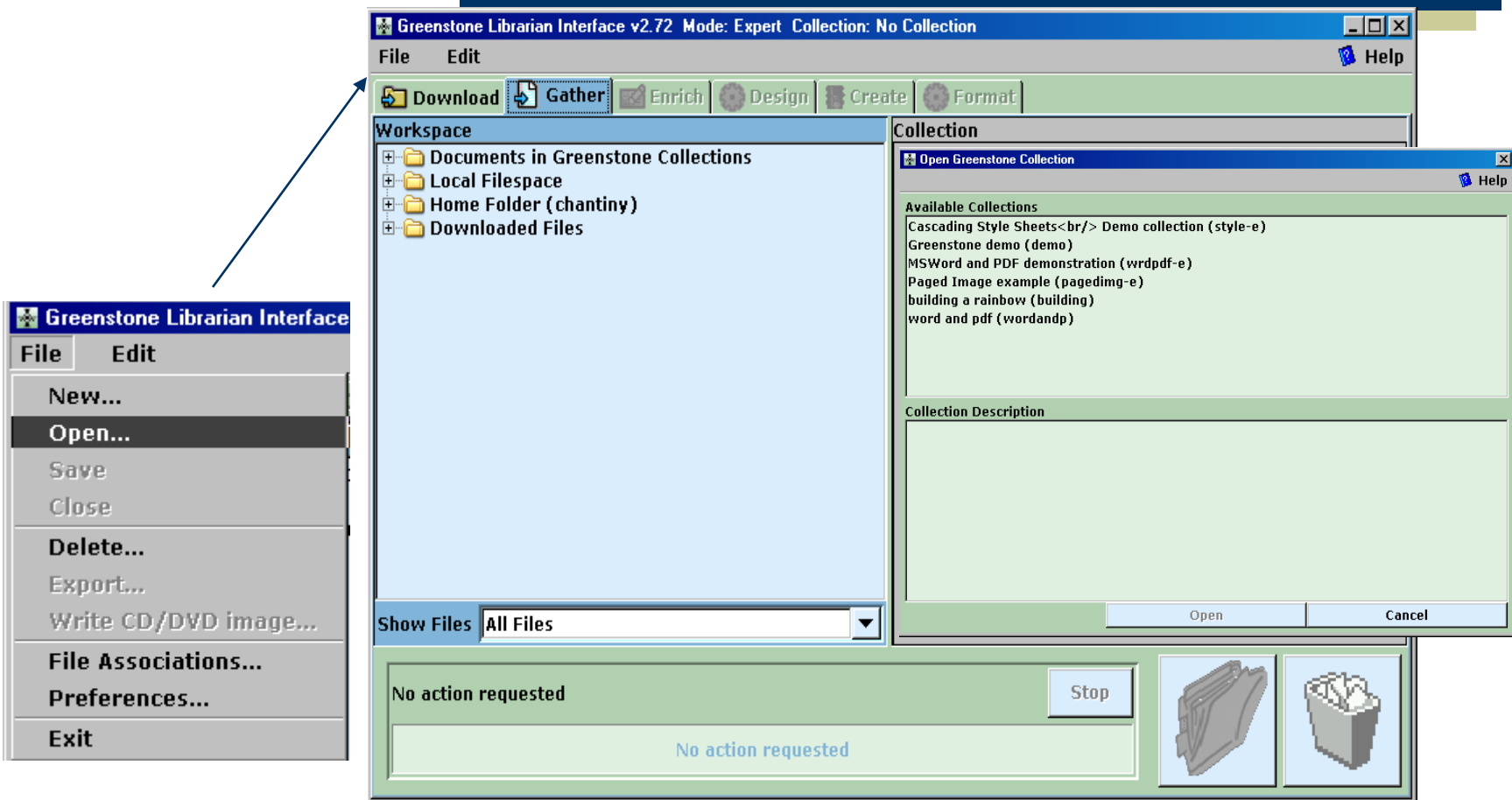
---

At this point using the 2 systems involves very different workflows

Greenstone = quasi-batch

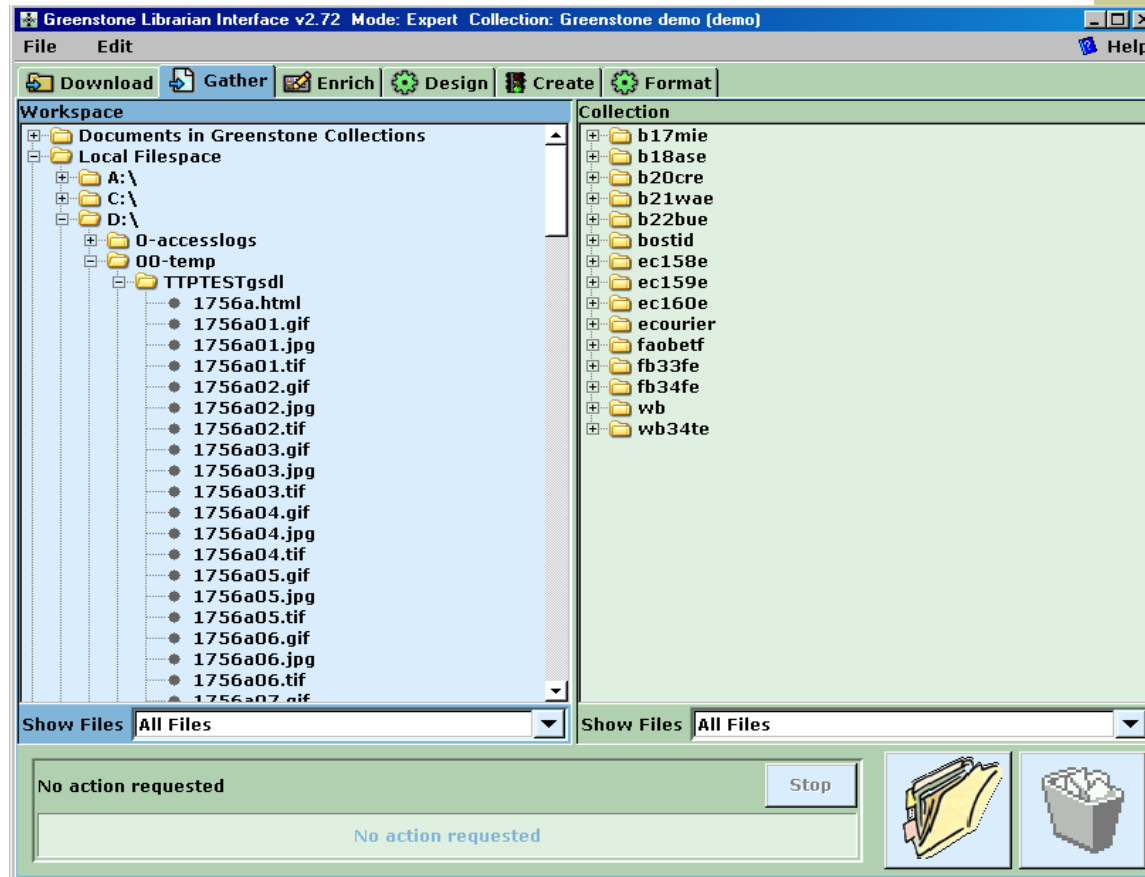
Streetprint = record by record

# Input - Greenstone





# Input – Greenstone drag & drop all file types



# Input – Streetprint default form for metadata entry

A STREETPRINT SITE



Welcome, Streetprint Admin. [[Log out](#)]

[Home](#) > [My Collection](#)



## Add to Collection

- Add new texts to your Streetprint site.



## View/Edit Collection

- View and edit the information for existing texts.
- Add/edit images for an existing text.
- Publish/unpublish texts on your site.
- Delete texts from your site.

A STREETPRINT SITE



Welcome, Streetprint Admin. [[Log out](#)]

[Home](#) > [My Collection](#) > [View/Edit Text](#) > [Add to Collection](#)

**ADD**

### details

#### Add a text to your collection

Optional fields are in *italics*:

Title:

Author:

Reference Number:

Publisher:

Document Type:

Category:

Date: Year:  Month:  Day:  Circs:

Date Details:

Dimensions:

Page/notes:

City:

Illustrations:

Location:

Notes:

[[back](#)]

# Input – Streetprint error checking & publish

**ADD** details

**Add a photo to your collection**

Your item 'test record' was not successfully added. Please see errors below.

*Optional fields are in italics.*

**Title** test record

**Creator**  
Creator cannot be blank.

**Reference Number**

**Document Type:** Collection, Misc.: - Oahu

**Category:**  
Lanai (Hawaii)  
Maui (Hawaii)  
Military bases  
Molokai (Hawaii)  
Mountains  
Oahu (Hawaii)

**Date** Year:  Month:  Day:  Circa:

**Place** Hamilton Library late at night

SITE ADMINISTRATION

Welcome, Streetprint Admin. [\[Log out\]](#)

[Home](#) > [My Collection](#) > [View/Edit Collection](#) > Editing test record

**EDIT** details

**Your photo 'test record' was successfully edited.**

**test record**

This photo is **NOT** currently published. [\[Publish\]](#)

**Title:** test record

**Creator:** martha

**Reference Number:**

**Document Type:** Collection, Misc.: - Oahu

**Category:** Oahu (Hawaii)

**Date:** No date

**Place:** Hamilton Library late at night

# Input – Streetprint adding images

Home > My Collection > View/Edit Collection > Editing test record > Adding Images

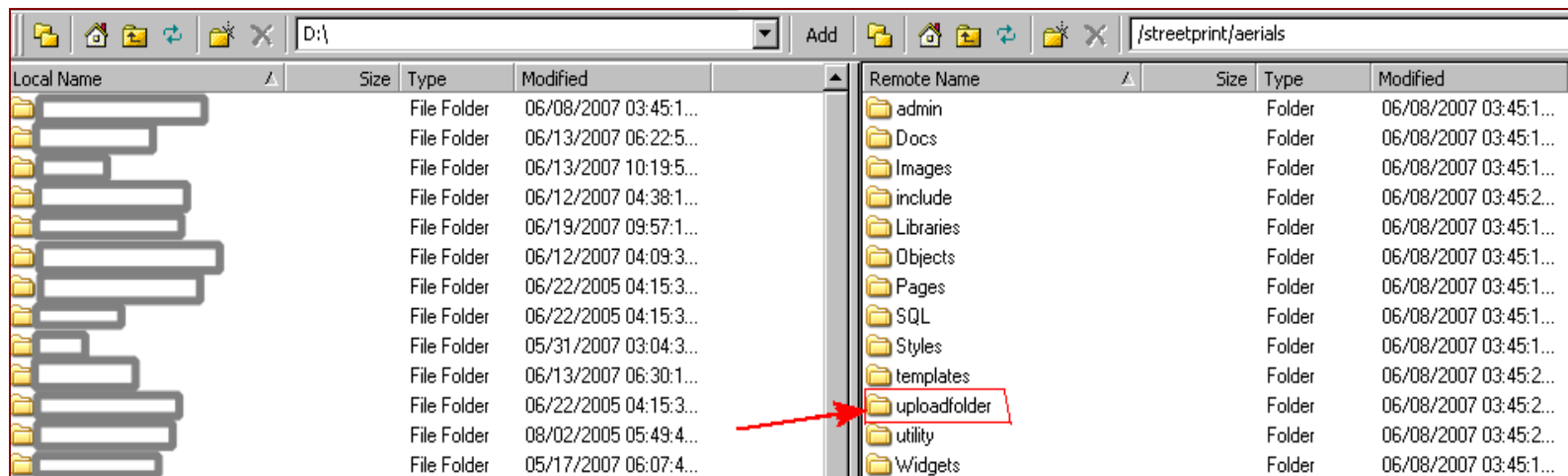
## image details

### adding images - step 1 of 2

**First, move your web-ready images for "test record" into Streetprint's upload folder.** (This folder is inside your Streetprint root directory and named 'uploadfolder'.)

The Streetprint Engine will then list all the images in this directory for easy uploading.

[Proceed to Step 2](#)



# Input – Streetprint adding images

**adding images - step 2 of 2**  
[How does this work? Click for help.](#)

---

**Adding images for "test record"**

**A) Specify image details:**

IMAGE TYPE:  FIRST PAGE # ON IMAGE:  # OF PAGES:  CAPTION:

**B) Select files for this image from the upload folder:**

NAME	DIMENSIONS (pixels)	RESOLUTION
<input checked="" type="checkbox"/> workyourspace.jpg	200 x 200	<input type="text" value="Thumbnail"/>
<input checked="" type="checkbox"/> gregg.jpg	800 x 522	<input type="text" value="Low"/>

**C) Upload the selected files for this image;**

A minimum of two files per record need to be pre-processed and uploaded

# Input – Streetprint adding full text

[Home](#) > [My Collection](#) > [View/Edit Collection](#) > [Editing test record](#) > [Add/Edit Full Text](#)

## details

**add your full text**

Full Text:

[back]

Cut and paste  
(or type in!?)  
only plain ASCII

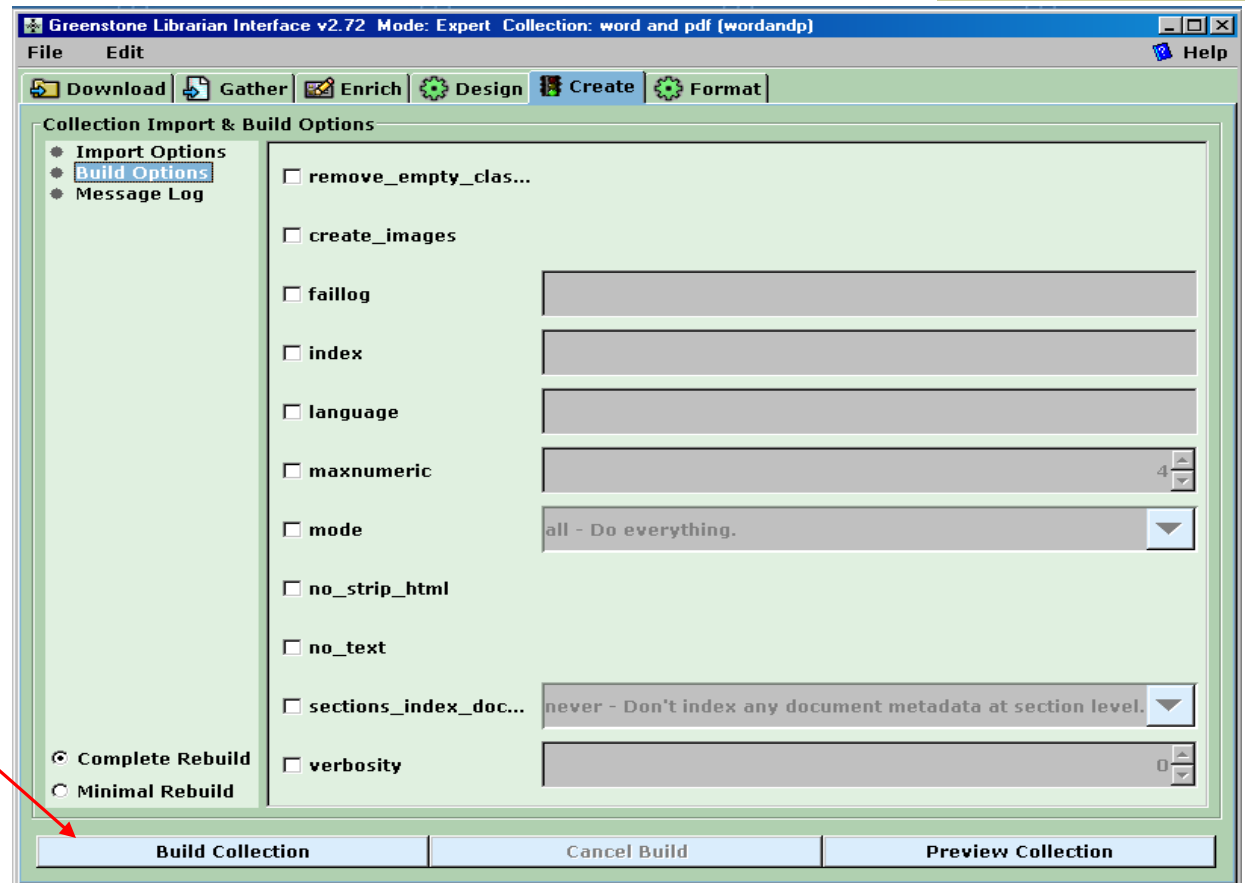
# Input – Streetprint adding media

The screenshot shows two distinct input sections. The top section, titled 'media' in orange, contains the text 'No media files.' followed by a green 'Add Media' link. Below this are three form fields: 'Select a File' with an empty text box and a 'Browse...' button; 'Select Media Type' with a dropdown menu currently set to 'MP3 Audio File'; and 'Media File Description' with a large empty text area. At the bottom of this section is an 'Add Media File' button. The bottom section, titled 'first lines' in orange, contains the text 'No first lines.' followed by a green 'Add First Line' link. Below this is a 'First line' label and a large empty text box.

Unlike images –  
media files must  
be uploaded from  
the PC

# Adding Data – Greenstone build

After adding  
files ...  
build collection





# Adding Data – Greenstone

## adding & editing metadata

The screenshot shows the Greenstone Librarian Interface v2.72. The window title is "Greenstone Librarian Interface v2.72 Mode: Expert Collection: word and pdf (wordandp)". The menu bar includes "File", "Edit", and "Help". The toolbar contains "Download", "Gather", "Enrich", "Design", "Create", and "Format".

The main interface is divided into two panes. The left pane, titled "Collection", lists files: br0001.doc, br0001.pdf, br0002.doc, br0002.pdf, br0003.doc, br0003.pdf, br0005.doc, br0005.pdf, br0006.pdf, br0007.doc, br0007.pdf, br0008.doc, br0008.pdf, br0009.pdf, br0010.doc, and br0010.pdf. A "Show Files" dropdown is set to "All Files".

The right pane displays a table of metadata elements and their values for the selected file (br0001.doc). The table has two columns: "Element" and "Value".

Element	Value
dc.Title	
dc.Creator	
dc.Subject and Key...	A name given to the resource.
dc.Description	
dc.Publisher	
dc.Contributor	
dc.Date	
dc.Resource Type	
dc.Format	
dc.Resource Identifier	
dc.Source	
dc.Language	
dc.Relation	
dc.Coverage	
dc.Rights Management	
ex./srclink	&lt;/a&gt;
ex.Encoding	utf8
ex.FileFormat	Word
ex.FileSize	2221
ex.Identifier	HASH01f20dc6a87908ff5751ca6d
ex.Language	en
ex.Plugin	WordPlug
ex.Source	br0001.doc
ex.Title	53D CONGRESS, } HOUSE OF BEPBESENTATIVES. { E...
ex.URL	http://C:/Program Files/Greenstone2.72/collect/wo...
ex.assocfilepath	HASH01f2.dir
ex.srcicon	_icondoc_
ex.srclink	&lt;a href=&quot;_httpprefix_/collect/[collection]/in...

At the bottom of the right pane, it says "No Metadata Element Selected". A "Manage Metadata Sets..." button is located at the bottom left of the interface.

If you haven't carefully manipulated and coded your files, the extracted metadata isn't always very useful



# Break?