



## Greenstone: Overview of features, capabilities, applications



Course material prepared by

Greenstone Digital Library Project  
University of Waikato, New Zealand

and National Centre for Science Information,  
Indian Institute of Science, Bangalore



## Agenda

- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ Example Greenstone collections
- ❖ Collection building
- ❖ Documentation and help

## What is a “digital library”?

*Collection of digital objects (text, video, audio) along  
with methods for access and retrieval, [user]  
and for selection, organization, and maintenance [lib]*

- ❖ Traditional user/librarian distinction is blurred
- ❖ Computers make information active
- ❖ Kitchens for knowledge preparation
- ❖ WWW ≠ DL!—organization, selectivity
- ❖ Nice Web site ≠ DL!—import new documents easily



## What are digital libraries for?

- ❖ Knowledge/content management
  - Manage and access internal information assets
- ❖ Scholarly communication, education, research
  - E-journals, e-prints, e-books, data sets, e-learning
- ❖ Access to cultural collections
  - Cultural, heritage, historical & special collections, museums, biodiversity
- ❖ E-governance
  - Improved access to government policies, plans, procedures, rules and regulations
- ❖ Archiving and preservation
- ❖ Many more ...

## DL Software: Key requirements

- ❖ Document types (book, journal article, lecture ...)
- ❖ Document formats (text, PDF, Word, PS, ...)
- ❖ Content acquisition (online and offline)
  - Metadata description, content tagging
  - Content uploading
- ❖ Indexing and retrieval
  - Structured/ full text indexing
  - Automatic metadata extraction
- ❖ Storage
  - Data compression
  - Efficient storage for metadata
  - Efficient location of metadata and documents
- ❖ Access and delivery
  - Structured search, browse, hierarchical browsing
  - CD-ROM distribution

## **DL Software: More requirements**

- ❖ Scaling up – for large collections
- ❖ Multilingual support
- ❖ Access management and security
- ❖ Usage monitoring and reporting
- ❖ Standards compliance
  - XML, Dublin Core, Unicode
- ❖ Interoperation
  - OAI, Z39.50 compliance, MARC, CDS/ISIS, ...

## **DL Software: Alternatives**

- ❖ What are your expectations?
- ❖ Develop local web-based application?
- ❖ Commercial DL solution?
- ❖ Adopt open source software?
  - Greenstone
  - Eprints
  - DSpace
  - (CDS/ISIS, Koha)



## Agenda

- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ Example Greenstone collections
- ❖ Collection building
- ❖ Documentation and help

Example

## Humanity Development Library

for sustainable development and basic human needs

❖ 160,000 pages	❖ CD-ROM
❖ 30,000 images	❖ US\$1
❖ 800 books	❖ Fully searchable
❖ 430 magazines	❖ Win3.1x upward
❖ 340 kg	❖ Stand-alone + intranet server
❖ US\$20,000	❖ Web browser user interface

***Global Help Project, Antwerp (+ UN agencies)***

## Accessing a Greenstone collection

Be adventurous—Click liberally!



**Book icon: read the corresponding book**



**Bookshelf: look at books on that subject**



**Open folder and view contents**



**View this document**



**Click on this to close the book**

## Reading the documents

Once you have found a document, you can



**Text: expand or contract**



**Table of contents: expand/contract**



**Highlight search terms, or not**



**Open this page in a new window**

## What is the Greenstone software?

- ❖ Software suite for building, maintaining, and distributing digital library collections
- ❖ Comprehensive, open-source
- ❖ Developed by New Zealand Digital Library Project at the University of Waikato
- ❖ Distribution and promotion partners:
  - UNESCO
  - Human Info NGO, Belgium
  - NCSI, Bangalore; UCT, Cape Town; Dakar, Senegal; Almaty, Kazakhstan; ...
  - You!

## Greenstone, Libraries and Open Access

*"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."*

—from [www.greenstone.org](http://www.greenstone.org)

## UNESCO: Distributing Greenstone DL software

### Sustainable development

*"Give a man a fish, feed him for a day  
Teach a man to fish, feed him for life"*

### Greenstone software on CD-ROM

- ❖ GNU licensed
- ❖ Fully documented ... in English/French/Spanish/Russian
- ❖ Language interfaces: Arabic Chinese Czech ... Thai Turkish
- ❖ Unix/Windows/Mac OS-X
- ❖ Trivial to install
- ❖ GUI interface for gathering, enriching, building ...
- ❖ Serve collections on Web or write them to CD-ROM
- ❖ Document formats: HTML, Word, PDF, PS, plain text, e-mail
- ❖ Metadata formats: XML, DC, OAI, MARC, ...

**download from <http://greenstone.org>**



## What we wanted

- ❖ "Collections" of digital material
- ❖ Individualized, depending on metadata etc
- ❖ Up to several Gb of text ...
- ❖ ... + associated images, movies, whatever
- ❖ Fully searchable
- ❖ Served on WWW, or published on CD-ROM
- ❖ Multi-platform (Unix + all Windows + Mac)
- ❖ Multi-format documents *and* metadata
- ❖ Multi-lingual: documents *and* interfaces
- ❖ Multimedia
- ❖ Metadata: standard *and* non-standard



## What is open-source software?

- ❖ “The **basic idea behind open source** is very simple: When programmers can read, redistribute, and modify the source code for a piece of software, the software evolves. People improve it, people adapt it, people fix bugs. And this can happen at a speed that, if one is used to the slow pace of conventional software development, seems astonishing.”

- from [www.opensource.org](http://www.opensource.org)

- ❖ Anyone can redistribute the software, even for a fee
- ❖ Source code must always be available

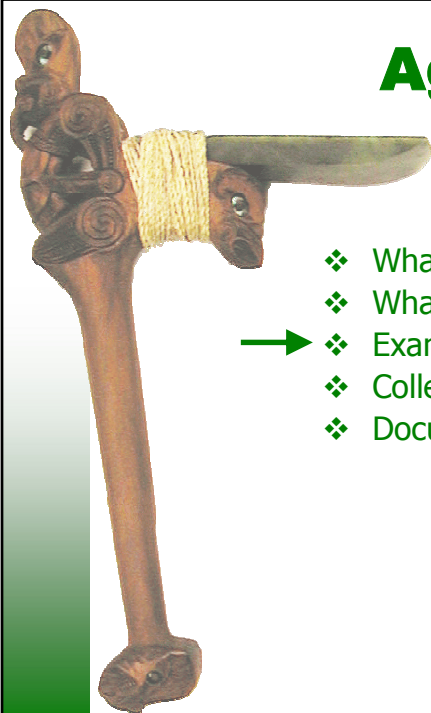
## The power of open source: Greenstone uses ...

- |               |  |
|---------------|--|
| ❖ Ghostscript | Interpreter for Adobe Postscript documents (Postscript plugin)                         |
| ❖ Kea         | Keyphrase extraction program (to generate metadata)                                    |
| ❖ pdftohtml   | Converter for PDF documents (PDF plugin)   |
| ❖ rtftohtml   | Converter for RTF documents (RTF plugin)   |
| ❖ TextCat     | Detects languages and document encodings   |
| ❖ wwWare      | Converter for Word documents (Word plugin)   |
| ❖ Xlhtml      | Converter for Excel/Powerpoint documents (plugins)                                     |
| ❖ XML::Parser | Parses XML documents, used to read and write Greenstone's internal XML document format |

## and ...

❖ MG	Creates compressed full-text indexes and performs searches
❖ GDBM	Database used for metadata etc
❖ wget	Downloading pages from the Web when creating collections
❖ YAZ	Client and server implementation of Z39.50
❖ Stemmer	English language stemmer
<hr/>	
❖ GCC	C/C++ compiler
❖ CVS	Version control system
❖ Perl	Used for plugins etc
❖ Apache	Web server used by many Greenstone installations

## Agenda



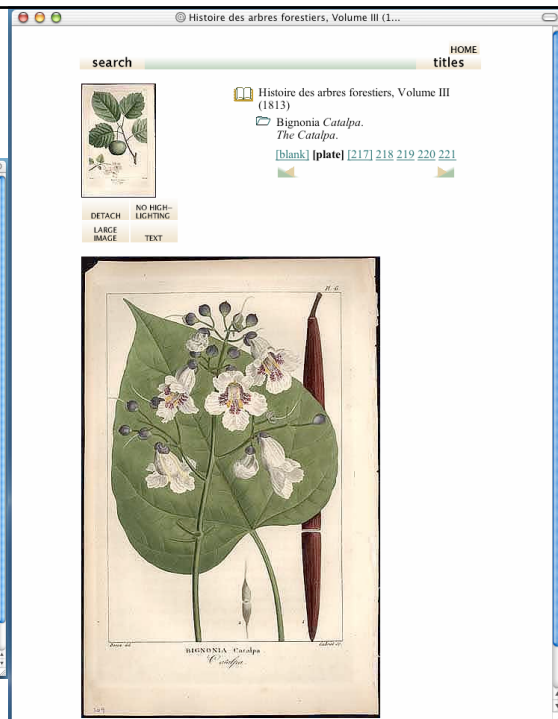
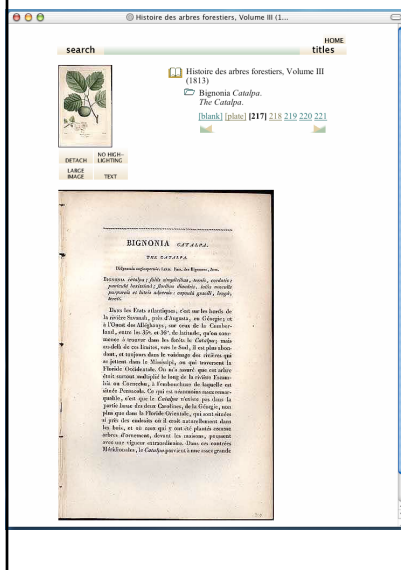
- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ Example Greenstone collections
- ❖ Collection building
- ❖ Documentation and help

## Example Greenstone collections

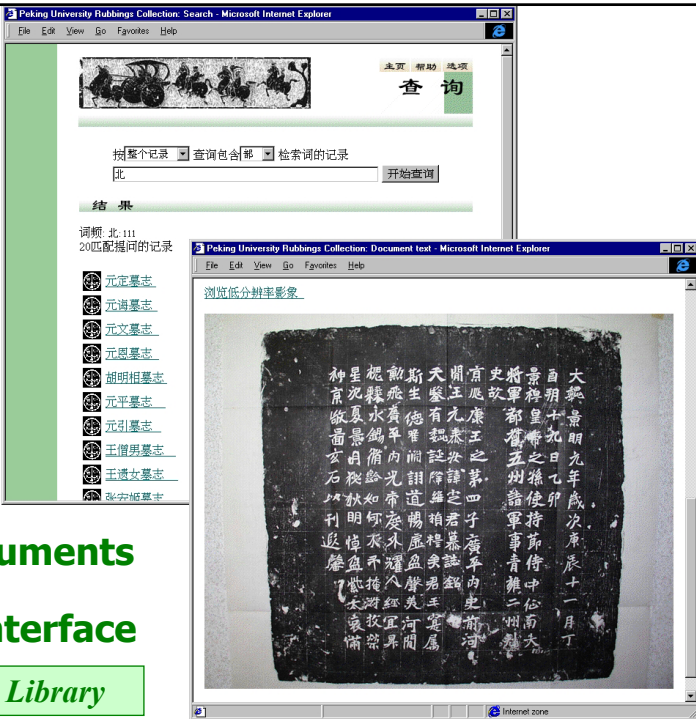
- ❖ Rapid growth in use
- ❖ International – China, Germany, India, UK, USA, Russia, etc.
- ❖ Increasing activity on Greenstone mailing list
- ❖ Promotion by UNESCO – “deployment of DL’s for sharing public domain information”
- ❖ Wide variety of DL collections have been developed in several languages
  - historical, educational, cultural, and research

### ❖ New York Botanical Garden

- Rare 19th century works on American trees
- Gorgeous full-color plates



[illegible]



**Chinese documents**  
(pictures of text)  
+ **Chinese interface**

*Peking University Library*



**Chinese**  
(Chinese & English  
interfaces)

**Classic Chinese literature**



# Arabic

(Arabic & English interfaces)

## Famous mosques

الموقع مساعدة الخيارات

العناوين أ-ي

العناوين أ-ي

جامع الناصر محمد جامع الناصر محمد بن قلاوون

جامع حسن باشا طاهر جامع حسن باشا طاهر انشا

جامع فابنباي جامع فابنباي انشاء الأشرف أبو النصر

جامع الناصر بالناصرين جامع ومدرسة السلطان لل

مشهد الإمام الجعفي مشهد الإمام الجعفي ثم انشاء

جامع اللنديني جامع اللنديني انشاء كريم الدين أحمد

الجامع اللنديني نسو هاج الجامع اللنديني نسو هاج يقع

الجامع للمري بغرض الجامع للمري بغرض يقع به

جامع سائر وسنجر الجوالي جامع سائر وسنجر ال

جامع فاني باي للرماح جامع فاني باي أمير أنور اب

السلطان حسن جامع السلطان حسن بنا في بركة السب

مشهد السيد يحيى الشديري مشهد السيد يحيى الشديري ثم

المدرسة السالحيية المدرسة السالحيية انشاء الملك

جامع الطاهر ببريس جامع الطاهر ببريس شيد الط

جامع خابر بك جامع خابر بك يقع بالقاهرة وقام...

عثمان كنذا جامع عثمان كنذا بشارع ع كسر

جامع الأشرف برسباي جامع الأشرف برسباي بنا

جامع الملكة صفية جامع الملكة صفية هالجامع انشا

جامع درغوق جامع فرج بن درغوق انشاء السلطان ال

جامع الأصغر جامع الأصغر بناه وزير الخليفة الفاطمي

HOME HELP PREFERENCES

titles a-z

search

### مدرسة وضريح ويبمارستان السلطان المنصور قلاوون

تم انشاء هذه المجموعة بين عامي 683-684هجرى / 1284-1285 ميلادي وتوجد بشارع المعز لدين الله الفاطمي بمدينة القاهرة

**وصف الجامع**

اهم جزء في هذه المجموعة القبة التي تطو الضريح وتتكون من قاعدة مثمثة مكونة من أربع دعام مربعة بها أربعة أعمدة رخامية مكسوة من الخارج بالرخام النقيع المطعم بالصفص وكلا من الدعائم والأعمدة تحمل عقودا معببة تطوها رقيقة مثمثة ويوجد بكل ضلع منها نافذة ويغلق الرقبة قبة مستديرة والواجهة الخارجية للمجموعة من الحجر الملون الأحمر والأبيض على شكل مربعات والمتمثلة تتكون من ثلاثة طوابق الأول والثاني يأخذان شكل مربع والعلوي مستدير وتم تجنيد المنقطة على يد الناصر محمد بن قلاوون عام 1303 ميلادي وتوجد المدرسة المنصورية امام باب القبة وتم تجنيد الإيوان الشرقي بهذه المدرسة بين عامي ميلادي ويتكون من ثلاثة أروقة أكبرها الأوسط والسقف محمول 1919-1916 على أعمدة رخامية تطوها عقود حيت هي والشبابيك المستديرة فوقها بالرخام الجصية اما البيمارستان لم يتبقى

# French

## UNESCO, Paris

SAHEL point DOC

ANTHOLOGIE du DÉVELOPPEMENT au SAHEL

Agriculture, pêche, élevage  
Environnement  
Développement social et humain  
Coopération, intégrations, marchés  
Culture, information, communication

Documents fournis par : AFATA, Banque Mondiale, Centre Soudanais, Centre du Sahel, CIRA, FAO, GRET, Institut du Sahel, Editions Mémoires et Larcos, PNUD

rechercher

sujets

titres a-z

organisations

comment

AIDE PRÉFÉRENCES

Rechercher les chapitres contenant (au moins un) mot(s)

Envoyer

### UNESCO - SAHEL point DOC - CD-ROM 1.0

Cliquez une couverture ... pour accéder au livre

Le Projet de la Collection PUBLICA de l'UNESCO

Ce CD-Rom SAHEL point DOC, première publication de la collection PUBLICA de l'UNESCO, met à la disposition du public, un choix de documents d'informations utiles et disponibles en langue française sur le développement du Sahel. Plus de 12.000 pages de documents y sont réparties en cinq thèmes que vous pouvez découvrir sous le bouton "sujets".

### Comment trouver les informations ?

AIDE Pour de l'aide à tout moment, activez le bouton "Aide".

Il y a cinq manières d'accéder aux informations contenues dans ce CD-ROM :

- rechercher des mots particuliers ou une phrase
- accéder aux publications par sujet
- accéder aux publications par titre dans l'ordre alphabétique
- accéder aux publications par organisation
- accéder aux publications par le bouton

## Spanish

*PAHO, WHO*



**Ciudades en Peligro - Ciudades más Seguras... antes de un Desastre (DIRDN): Parte II: ¿...**

File Edit View Go Communicator Help

PRINCIPAL AYUDA PREFERENCIAS

buscar títulos a-z temas organización

**Ciudades en peligro**  
Ciudades más seguras... antes de un desastre

TEXTO COMPLETO EXPANDIR ÍNDICE

SEPARAR NO RESALTAR

**Parte II: ¿Qué se está haciendo?**

- (introductory text...)
- Ecuador - Quito: Las nuevas leyes reflejan el cambio de actitud
- Papúa Nueva Guinea - Rabaul: Convivir con el peligro
- Sudáfrica - Una participación ciudadana ayuda a reducir la vulnerabilidad a las inundaciones
- Australia - Menos daños gracias a una educación ciudadana
- Paraguay - Un nuevo proyecto ribereño cambia la actitud local
- Dominica (Antigua) - Casas resistentes a huracanes a un precio reducido
- Holanda - La planificación de las emergencias es la base para una evacuación fluida durante inundaciones

Document: Done

## Russian

*Mari El Republic*  
<http://gov.mari.ru/gsd/>

Региональные  
документы

Марийский  
фольклор

Информация

Федеральные  
документы

Начало помощь настройки

поиск названия а-я указатель авторы а-я

Бирюков С. И. Песни, вои, частушки

Песни

- Рекрутская песня
- Величание
- Припев грузчиков

**Интернет - экономика**

весь текст раскрыть оглавление  
новое окно снять выделение

**Рекрутская песня**

Поется для лица, принятого и уезжающего в армию, накануне его отъезда, группой девушек на вечеринке (беседках).

Как по морю, морю синему,  
Посине морю волну чату.  
Что плывут, гребут четыре корабля,  
А на встречу им червонный корабель.  
На корабличке немножко людей,  
По моему счету восемь человек,  
А девятый на корме сидит,

Document: Done

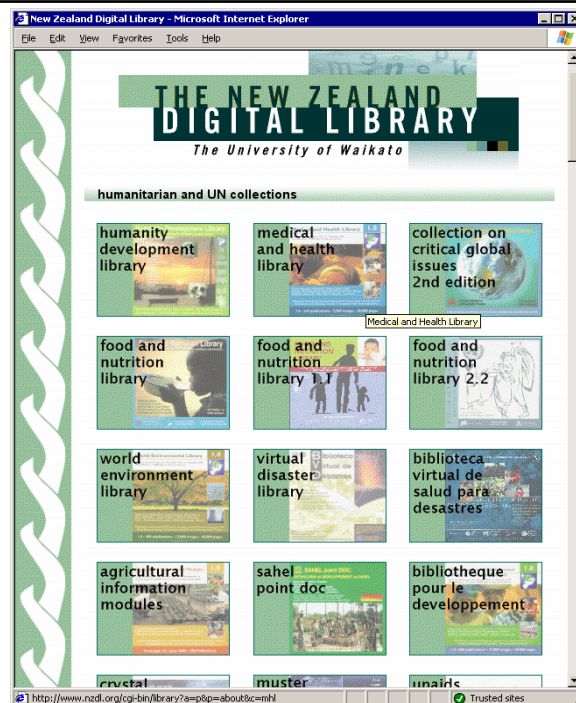
## Example Greenstone collections

- ❖ For brief description and links to DLs built using Greenstone:
  - “*Examples of practical digital libraries: Collections built internationally using Greenstone*”, By Ian Witten. D-Lib Magazine, March 2003.  
([www.dlib.org/dlib/march03/witten/03witten.html](http://www.dlib.org/dlib/march03/witten/03witten.html))
- ❖ [www.greenstone.org/examples](http://www.greenstone.org/examples)
- ❖ New Zealand DL Project ([www.nzdl.org](http://www.nzdl.org))

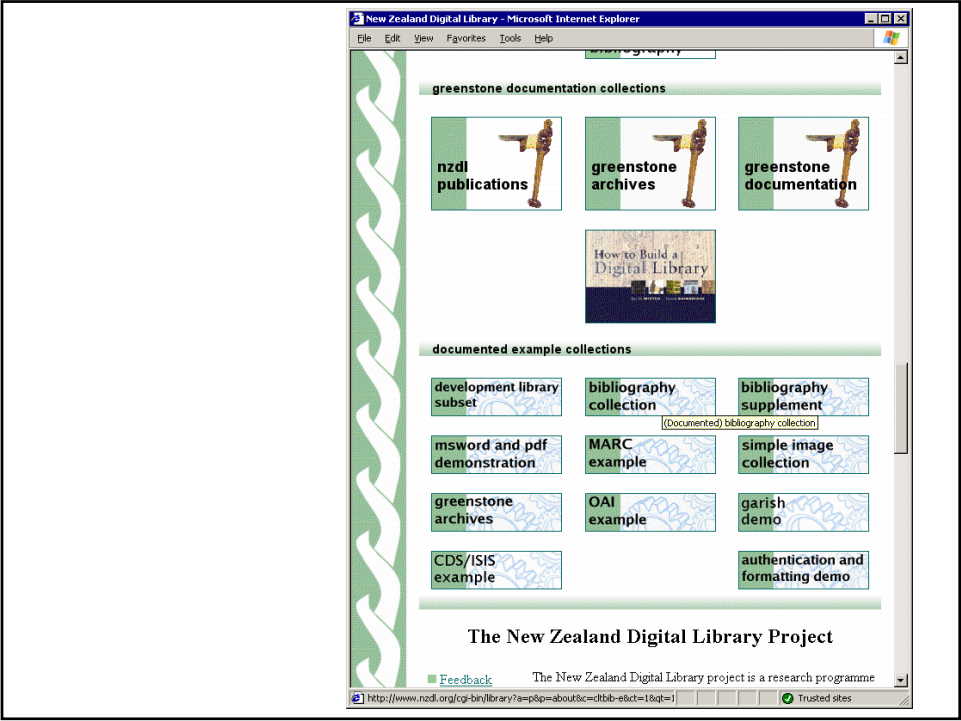
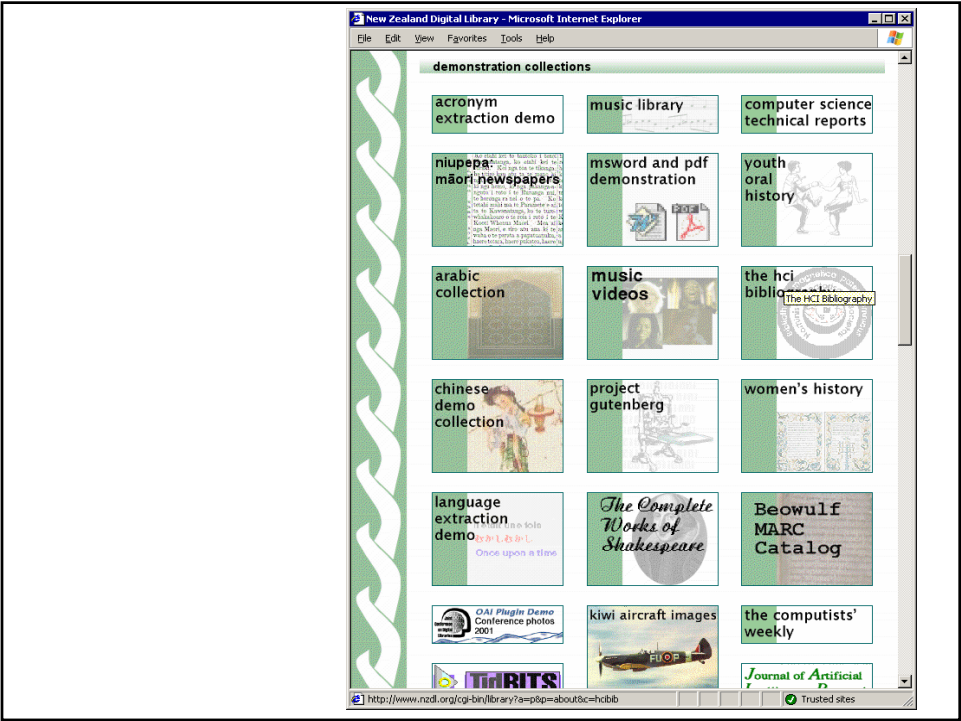
### Collections: on the Web

[nzdl.org](http://nzdl.org)

(demo, not  
service)









## Agenda

- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ Example Greenstone collections
- ➔ ❖ Collection building
- ❖ Documentation and help

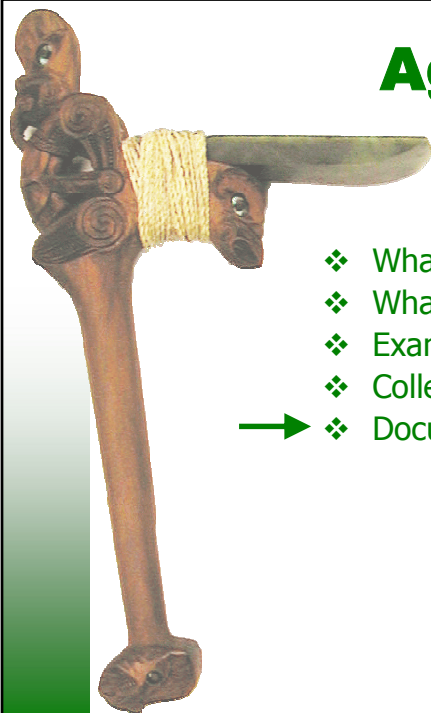
## Collection Building

- ❖ Input: a set of source documents, possibly in many different formats
- ❖ Greenstone “imports” these documents and converts them to its own internal (GA) format
  - Extracts as much metadata as possible
- ❖ Greenstone “builds” indexes and browsing structures using the GA files
- ❖ Start with a few documents, get the design right, then add the bulk of the documents

## Collection configuration

- ❖ Collection configuration file determines content conversion, extraction and building of indexes and browsing structures
  - indexes, classifiers, plugins
- ❖ Presentation of search/browse results and collection interface is determined by “format” strings and “macros”

## Agenda



- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ Example Greenstone collections
- ❖ Collection building
- ❖ Documentation and help

## Documentation and help

- ❖ Available at: [www.greenstone.org](http://www.greenstone.org)
  - Software
  - Demo collections
  - FAQ
  - Tutorial materials
- ❖ Documentation:
  - Installer's Guide, User's Guide, Developer's Guide, From Paper to Collection
- ❖ Mailing lists:
  - Greenstone Users List
  - Greenstone Developers List

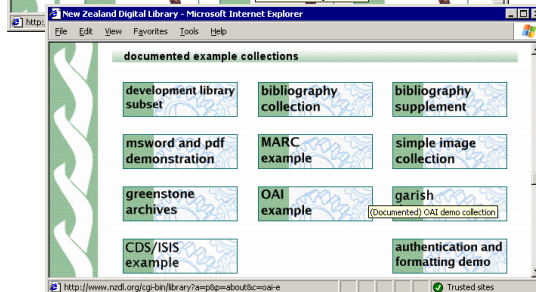
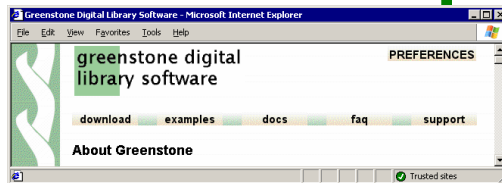
## Documentation and Help

### Manuals on the CD-ROM (docs)

- *Installer's Guide* (install.pdf, 36pp)
  - Versions of Greenstone, installation procedure, Greenstone collections, setting up the web server, configuring your site, personalizing your installation
- *User's Guide* (user.pdf, 90pp)
  - Overview of Greenstone, using Greenstone collections, the collector, administration, software features, glossary of terms
- *Developer's Guide* (develop.pdf, 113pp)
  - Understanding the collection building process, getting the most out of your collections, the Greenstone runtime systems, configuring your Greenstone site
- *From Paper To Collection* (paper.pdf, 30pp)
  - Scanners and scanning, OCR, 3 examples – from 1,000 to 100,000 pages, Creating an electronic collection

## Documentation and Help

- ❖ [greenstone.org](http://greenstone.org)
  - Download: software and tutorials
  - Example collections
  - Documentation
  - FAQ: general info section
  - support (+ join mailing list)
  - Configuration files for nzdl.org collections
- ❖ [nzdl.org](http://nzdl.org)
  - Documentation collections
  - Documented example collections



## Documentation and Help

### Mailing Lists

- *Greenstone Users List*  
For people installing and using standard Greenstone  
Join at: <https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-users>  
Mail to: [greenstone-users@list.scms.waikato.ac.nz](mailto:greenstone-users@list.scms.waikato.ac.nz)
- *Greenstone Developers List*  
For people customizing their version of Greenstone  
Join at: <https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-devel>  
Mail to: [greenstone-devel@list.scms.waikato.ac.nz](mailto:greenstone-devel@list.scms.waikato.ac.nz)

### Mailing List Archives

A Greenstone collection of mail from both mailing lists  
<http://www.nzdl.org/qsarchives>

**Questions?  
Comments?  
Discussion?  
Feedback form!**