Long Lived: *slow, determined, indestructible*

Adding a Title to LOCKSS

Technical Work

Structure of Implementation

- Platform
  - PC into preservation appliance
  - Cheap to administer & run
- Daemon
  - Cooperate to detect & repair damage
  - Proxy cache gets content to readers
- Plug-ins
  - Adapts system to publisher

Daemon: What It Does

- Collect content:
  - Crawl publisher with help from plug-in
- Preserve content:
  - Compare content with other peers
  - Repair from other peers if damaged
- Distribute content:
  - Act as proxy cache for readers
  - Deliver publisher version if available

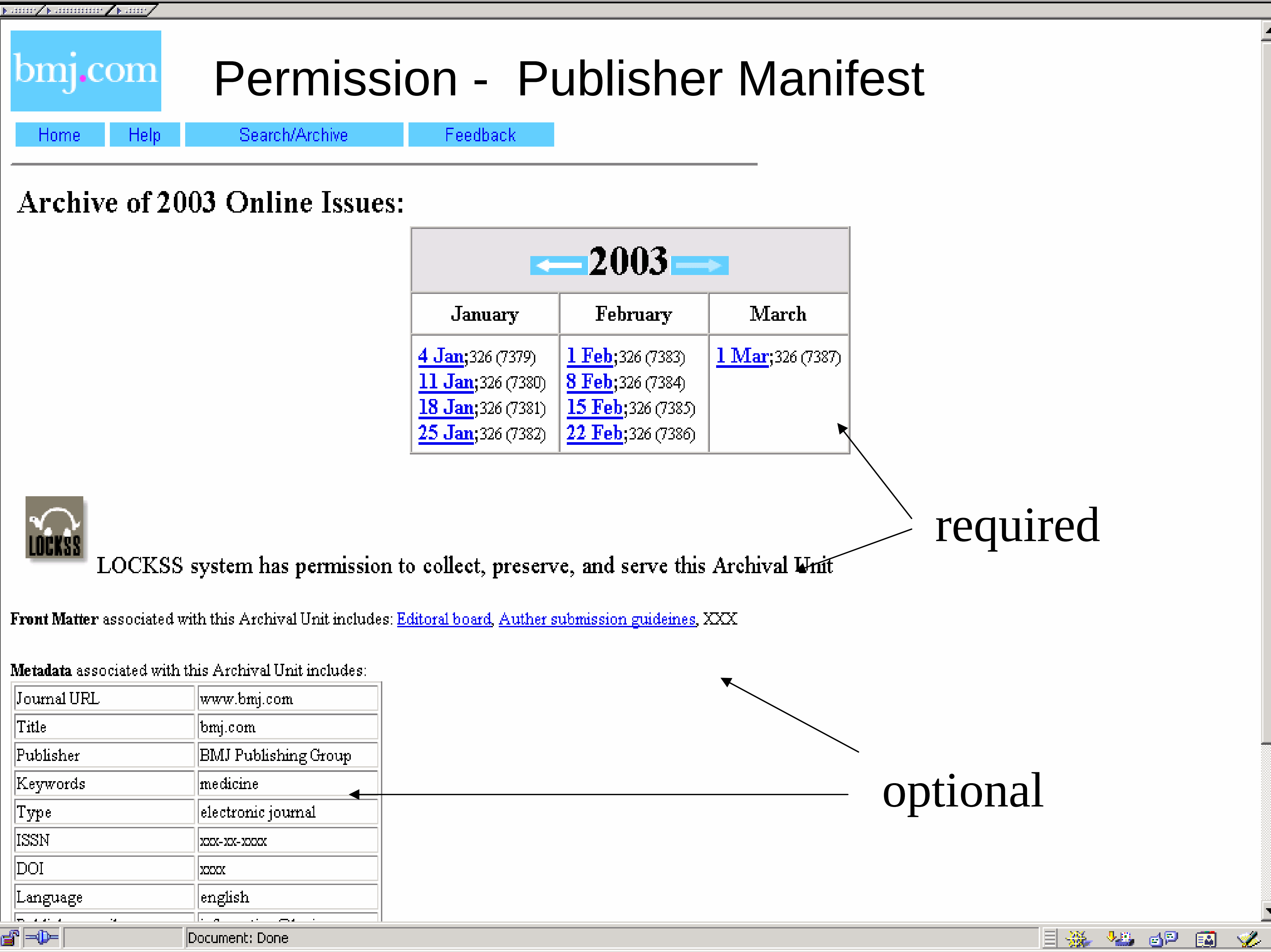
Add a Title to LOCKSS

- Publisher Manifest Page
- Write a plug-in

Publisher Permission LOCKSS Crawler

Slowly collect e-journals

- Publisher manifest
  - – List top level URLs/volume on a web page
  - – Include URLs for ‘front matter’, etc.
  - – Descriptive metadata
  - – Grant permission volume by volume



Permission - Publisher Manifest

Archive of 2003 Online Issues:

January	February	March
4 Jan;326 (7379) 11 Jan;326 (7380) 18 Jan;326 (7381) 25 Jan;326 (7382)	1 Feb;326 (7383) 8 Feb;326 (7384) 15 Feb;326 (7385) 22 Feb;326 (7386)	1 Mar;326 (7387)

LOCKSS system has permission to collect, preserve, and serve this Archival Unit

required

Front Matter associated with this Archival Unit includes: Editoral board, Auther submission guideines, XXX

Metadata associated with this Archival Unit includes:

Journal URL	www.bmj.com
Title	bmj.com
Publisher	BMJ Publishing Group
Keywords	medicine
Type	electronic journal
ISSN	xxx-xx-xxxx
DOI	xxxx
Language	english

optional

Plug-in Overview

- Adapts daemon to publisher:
  - Decide what/when to crawl
  - Handle publisher permission
  - Handle publisher authentication
  - Filter dynamic content for comparison
- From publisher or community:
  - Download as signed *.jar* file
  - Registry finds appropriate plug-in

Writing Plug-ins

- Identify publisher requirements:
  - –Crawl restrictions, boundaries
  - –Dynamic content
- Use plug-in tool
  - –Java based UI
  - –Generates XML file

Distributing Plug-in

- Plug-in repository
  - – Caches crawl like any other title
  - – Install new plug-ins after verifying signature
  - – Plug-ins are preserved
- Title Database
  - – Configured a plug-in for a specific volume/year for a title

What Format?

- LOCKSS is format agnostic
  - – Collect anything delivered over HTTP
- Formats become obsolete

Format Migration

- Format conversion API
  - Make converter plug-ins
- Conversion is done on access
- Original version preserved
- Current format served