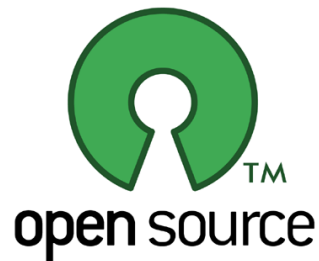


Open Source Software for Archives



ACA 2008

Fredericton, New Brunswick

Presentation Overview

- What is open-source software?
- Open-source business model Peter Van Garderen
- Open-source software for Archivists
- Who's using open-source software? Evelyn McLellan
- Open-source vs proprietary software
- Open-source implementation hurdles Glenn Dingwall
- Open-source software demos
 - ICA-AtoM Peter Van Garderen
 - Archivist's Toolkit Glenn Dingwall
 - Greenstone Evelyn McLellan

What is open-source
software?

Free Software!!







Copyleft



open source

1. Free Redistribution

The license shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources. The license shall not require a royalty or other fee for such sale.



2. Source Code

The program must include source code, and must allow distribution in source code as well as compiled form. Where some form of a product is not distributed with source code, there must be a well-publicized means of obtaining the source code for no more than a reasonable reproduction cost preferably, downloading via the Internet without charge. The source code must be the preferred form in which a programmer would modify the program. Deliberately obfuscated source code is not allowed. Intermediate forms such as the output of a preprocessor or translator are not allowed.

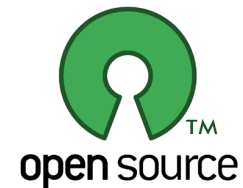
3. Derived Works

The license must allow modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software.



4. Integrity of The Author's Source Code

The license may restrict source-code from being distributed in modified form *only* if the license allows the distribution of "patch files" with the source code for the purpose of modifying the program at build time. The license must explicitly permit distribution of software built from modified source code. The license may require derived works to carry a different name or version number from the original software.



5. No Discrimination Against Persons or Groups

The license must not discriminate against any person or group of persons.



6. No Discrimination Against Fields of Endeavor

The license must not restrict anyone from making use of the program in a specific field of endeavor. For example, it may not restrict the program from being used in a business, or from being used for genetic research.



7. Distribution of License

The rights attached to the program must apply to all to whom the program is redistributed without the need for execution of an additional license by those parties.



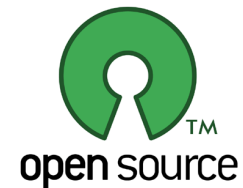
8. License Must Not Be Specific to a Product

The rights attached to the program must not depend on the program's being part of a particular software distribution. If the program is extracted from that distribution and used or distributed within the terms of the program's license, all parties to whom the program is redistributed should have the same rights as those that are granted in conjunction with the original software distribution.



9. License Must Not Restrict Other Software

The license must not place restrictions on other software that is distributed along with the licensed software. For example, the license must not insist that all other programs distributed on the same medium must be open-source software.



10. License Must Be Technology-Neutral

No provision of the license may be predicated on any individual technology or style of interface.



GNU GENERAL PUBLIC LICENSE

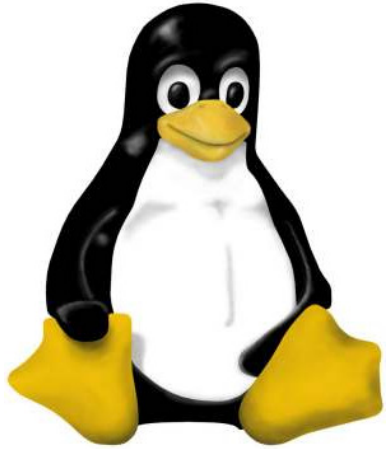
Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc.,
51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Lesser General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.



Free!



Open-Source Business Model

How Do We Pay For This?

- Individual contributions
- Grants and subsidies
- Membership fees
- Bounty system
- Advertising

How Do We Pay For This?

- Corporate sponsors
- Services
 - Development
 - Hosting
 - Support
 - Training

“Open source is a means to an end.
It’s a mechanism to grow the
broadest market, build the largest
ecosystem, reach the maximum set
of opportunities”

- Jonathan Schwartz (CEO Sun Microsystems). “Sun’s open-door policy” *eWeek* (April 21, 2008)



Peter Van Garderen
President/Senior Consultant



Jack Bates
Software Engineer



Richard Dancy
Systems Archivist



David Juhasz
Software Engineer



Evelyn McLellan
Systems Archivist



Austin Trask
Systems Engineer

ICA-AtoM Support

Artefactual Systems is the lead developer of the ICA-AtoM open-source archival description application. ICA-AtoM is open-source software and available as a free download at the ica-atom.org website.

If your institution needs assistance in using or customizing the ICA-AtoM software we provide the following support services:

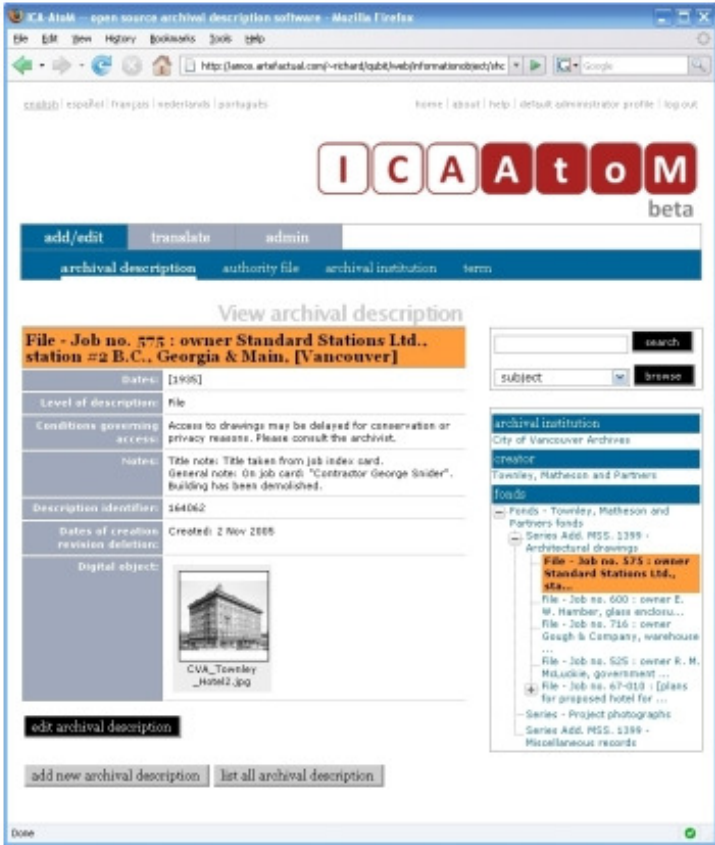
- **On-site installation or off-site hosting**
- **Training and technical support**
- **Customization and new feature development**

ICA-AtoM is fully web-based, multi-lingual archival description software that was commissioned by the International Council on Archives to make it easy for archival institutions worldwide to put their archival holdings online using the ICA's descriptive standards.

Through the flexibility of the underlying [Qubit Toolkit](#) it is easily configured to:

- use other national descriptive standards
- use descriptive standards from other professional communities (e.g. libraries, museums)
- act as a metadata management component to a digital archives repository
- serve as a multi-institutional union list portal

If your institution would like assistance in using or customizing the ICA-AtoM Software, please [contact us](#) for a complimentary cost estimate.



It's a race to Free

Burt Reynolds Roger Moore
Farrah Fawcett Dom Deluise

THE CANNONBALL RUN



DVD
VIDEO

“Basic economics tells us that in a competitive market, price falls to the marginal cost. There’s never been a more competitive market than the Internet, and every day the marginal cost of digital information comes closer to nothing.”

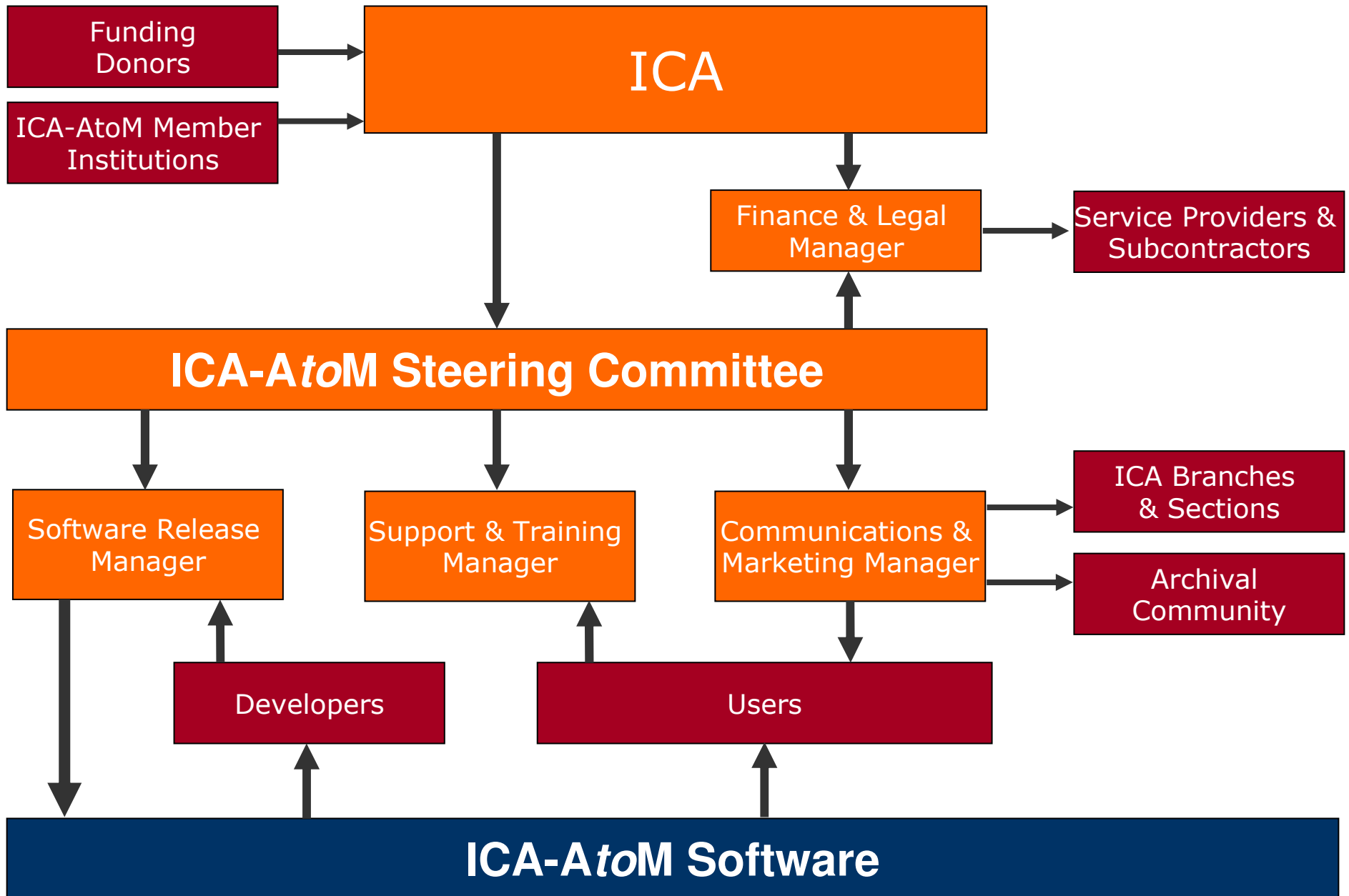
- Chris Anderson. “Why \$0.00 is the Future of Business” *Wired* (March 2008)

Open Source is Growing Up

- Free Software Foundation
- Apache Foundation
- Drupal Foundation
- Fedora Commons
- ICA-Atom Steering Committee

ICA-AtoM Governance Model

DRAFT FOR DISCUSSION. Version 3. May 20, 2008



What can archivists
use open-source
software for?

Traditional archives

- Accession management
- Arrangement and description
- Making descriptions available to the public

Digital archives

- Ingest, storage and dissemination
- Metadata extraction and management
- Implementing preservation strategies

Examples of digital preservation tools

- Repository software: Fedora, DSpace, DAITSS, Greenstone, LOCKSS
- Web crawling: Heritrix, Wget

Examples of digital preservation tools

- Object validation and metadata extraction: Jhove, DROID, New Zealand Metadata Extracting Tool
- Migration and Emulation: Xena and Dioscuri

Open-source software and “open” file formats

- “Open” file formats are non-proprietary, have freely available specifications and may be internationally accepted standards
- Certain open-source software may produce open file formats but “open-source software” and “open file formats” are two different things

Sustainable digital archives

Open-source software
+ “open” file formats
= sustainability

Who is using open-
source software?

- National Archives of UK, Australia, Netherlands, New Zealand, LAC
- Universities: Harvard, Yale, Cornell, Stanford, California, MIT, Florida State
- Consortium of Research Libraries of the British Isles

Open-source vs Proprietary Software

“Open source and proprietary software” Sept. 2007 UNESCO Report

| FOSS Proponents say | FOSS Opponents say |
|---|--|
| <u>Total cost of ownership</u> ◇ Open source has a much lower price (true) ◇ The total cost of open source is lower (maybe) | <u>Total cost of ownership</u> ◇ Some proprietary software are not compatible with open source (true) |
| <u>Features & Quality</u> ◇ Open source is more reliable (maybe) ◇ Open source is more secure (maybe) ◇ Open source is more powerful (maybe) ◇ Open source is more network friendly (true) ◇ Open source can be more customized ◇ Open formats and standards are better (true) ◇ Open source supports better curricula in technology (maybe) | <u>Features & Quality</u> ◇ Proprietary software has more features (true) ◇ Proprietary software is more user friendly (maybe) ◇ Open source is not mature enough for schools (false) ◇ There are no open source solutions for some school needs (true) ◇ Some curriculum software is incompatible with open source (true) |
| <u>Deployment & Maintenance</u> ◇ With open source you only pay for what you need (true) ◇ Open source makes license management easier (true) ◇ Open source means greater independence from companies (true) ◇ Open source lets teachers & students take software home (true) | <u>Deployment & Maintenance</u> ◇ Open source is harder to deploy (maybe) ◇ Proprietary software offers better service & support (maybe) |
| <u>Users & Migration</u> ◇ Some open source software are just as easy to learn and to use | <u>Users & Migration</u> ◇ Migration to FOSS is too expensive (maybe) ◇ Users are more familiar & comfortable with proprietary software (true) ◇ It's difficult to integrate open source & proprietary solutions (maybe) |
| <u>Free Markets & Choice</u> ◇ Software should be a commodity (maybe) ◇ Proprietary formats and standards lead to vendor lock in (true) ◇ Proprietary software leads to monopolies (maybe) | <u>Free Markets & Choice</u> ◇ Proprietary software may be needed to use some third-party programs (maybe) |

2007 UNESCO Report

Factors to Consider:

- Cost of Ownership
- Features and Functionality
- Deployment and Maintenance
- Users and Migration

Total Cost of Ownership

PROS

- Software purchase cost much lower (free!)

CONS

- Potentially higher associated costs
 - Deployment
 - Training
 - Maintenance

Evergreen ILS for BC Public Libraries

Saving about \$15 million dollars in vendor
license and support fees over the next few
years

Software Features & Quality

PROS

- Greater Compatibility and Interoperability
- Greater Customizability

CONS

- Narrower Range of Features
- Few Integrated Suites of Products

Deployment & Change Management

PROS

- Easier licensing
- No limitation on number of users

CONS

- Can be more difficult to install/deploy
- Somewhat less user friendly

Philosophical Issues

- Empowerment
- Transparency
- Universal Access

Other Considerations

- Do you have a choice?

Is the only option an open source (or proprietary) product? Examples:

- Web archiving – Heratrix
- Format Validation – DROID, JHove
- Format Normalization – XENA

Other Considerations

- Open Source Solution may not always be the best choice
- What is the best decision within your own institution's context:
 - Existing IT infrastructure
 - Available Resources
 - Potential Institutional Resistance

Open-Source Implementation Hurdles

Common Concerns

- Accountability
- Security

Support/Accountability

- One IT Director:
- Concerned about lack of support
 - There's no guarantee that an open-source community isn't just going to pick-up and abandon a product
 - We need to have somebody accountable
 - Who's going to provide support and training?

Support/Accountability

- No guarantee that software vendors won't abandon products
- Open-source projects are establishing foundations for long-term sustainability
- Can enter legal service agreements with third parties for support, training

Support/Accountability

- Choose OSS that is developed/used by major institutions (e.g. Harvard, MIT, National Archives of Australia, etc.)
- Choose OSS that has a large/active user community

Support/Accountability

→ Choose OSS that is backed by a formal governance/legal entity (e.g. Fedora Commons, Apache Foundation, ICA-AtOM Steering Committee)

Security

If the code is publicly-accessible, doesn't that make it less secure?

- Closed source code isn't more difficult to attack, but it is more difficult to fix
- OSS security holes are typically plugged faster because it's easier for someone in the user community to spot and fix leaks themselves

Security (continued)

- Secrecy \neq Security
- security is more a result of how software is deployed and managed
 - The same processes, protocols, architecture designs apply to OSS and proprietary software
 - Both can be deployed securely or insecurely

Security

- The U.S. DoD \$200 Billion integrated weapons system program
- “the army said they choose not to use Microsoft’s proprietary software because they didn’t want to be beholden to the company, opting instead to develop a Linux-based operating system based on publicly available code.”
 - “The Modern Face of Warfare” *Vancouver Sun* (January 26, 2008)

Security

- French Gendarmerie
 - Announced in January 2008 that it is migrating 70,000 police agent's desktops from Windows to Novell Suse Linux
 - Already 1145 PCs in French Parliament are running Ubuntu Linux

Linux Magazine (February 2008)