Exploring Open Source:
a Solution for
Records Management?

An InterPARES 3 General Study

Corinne Rogers, InterPARES Graduate Research Assistant,
Evelyn McLellan, Researcher and
Elizabeth Shaffer, Graduate Research Assistant
Open Access Week, UBC
October 19, 2010

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Outline

- InterPARES Project – background & context
- IP3 – General Study 08
- What, why & how of Open Source Software – Definitions, Landscape and Licensing
- Who is using OSS?
- OSS in Libraries and Archives – a success
- OSS for Records Management – is this also a success?

InterPARES Project

- International Research on Permanent Authentic Records in Electronic Systems (InterPARES)

- Developing knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form

- Providing the basis for standards, policies, strategies and plans of action to ensure longevity and trust in records' authenticity

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

InterPARES Project

- IP1: preservation of authentic records created and/or maintained in databases and document management systems

- IP2: reliability, accuracy, authenticity throughout records' lifecycle, emphasis on complex digital environments

- IP3: puts theory into practice

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

InterPARES 3 & General Study 08:

- GS08: identify and discuss open source software options for records management (EDRMS)

- Literature review – features, problems, concerns

- Map functionality of existing OSS RM to InterPARES Creator and Preserver Guidelines and RM standards (MoReq 2, ISAD(G) and ISO 15489)

OSS – *Assumptions & Concerns*

- Free, but no vendor backup or installation support
- Poor security because code is freely available
- Inconsistent support dependent on peer user groups
- Hidden costs: implementation, support, interoperability
- Issues with intellectual property rights
- Perpetuated by Microsoft's video criticizing Open Office (http://www.youtube.com/watch?v=kzdykNa2IBU)

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Proprietary – Assumptions & Concerns

- Source code is unavailable, hidden behind binaries (object code) – preserves developer control

- Development is secretive, slow, cumbersome

- High cost to use the software, costs to support and upgrade

- Promotes dependency on one provider – "the addiction model of software procurement"*

*http://www.redhat.com/about/whyopensource/

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Comparison: Proprietary v. OSS

Proprietary model

- Users do not have access to source code

- Restrictive licenses

- Costs associated with startup, support, leaving

- Software purchase implies vendor lock-in

Open source model

- Users have access to source code, can modify, reuse, redistribute

- Permissive licenses

- Different model of costing

- No vendor lock-in

"A technology revolution driven by market demand"*

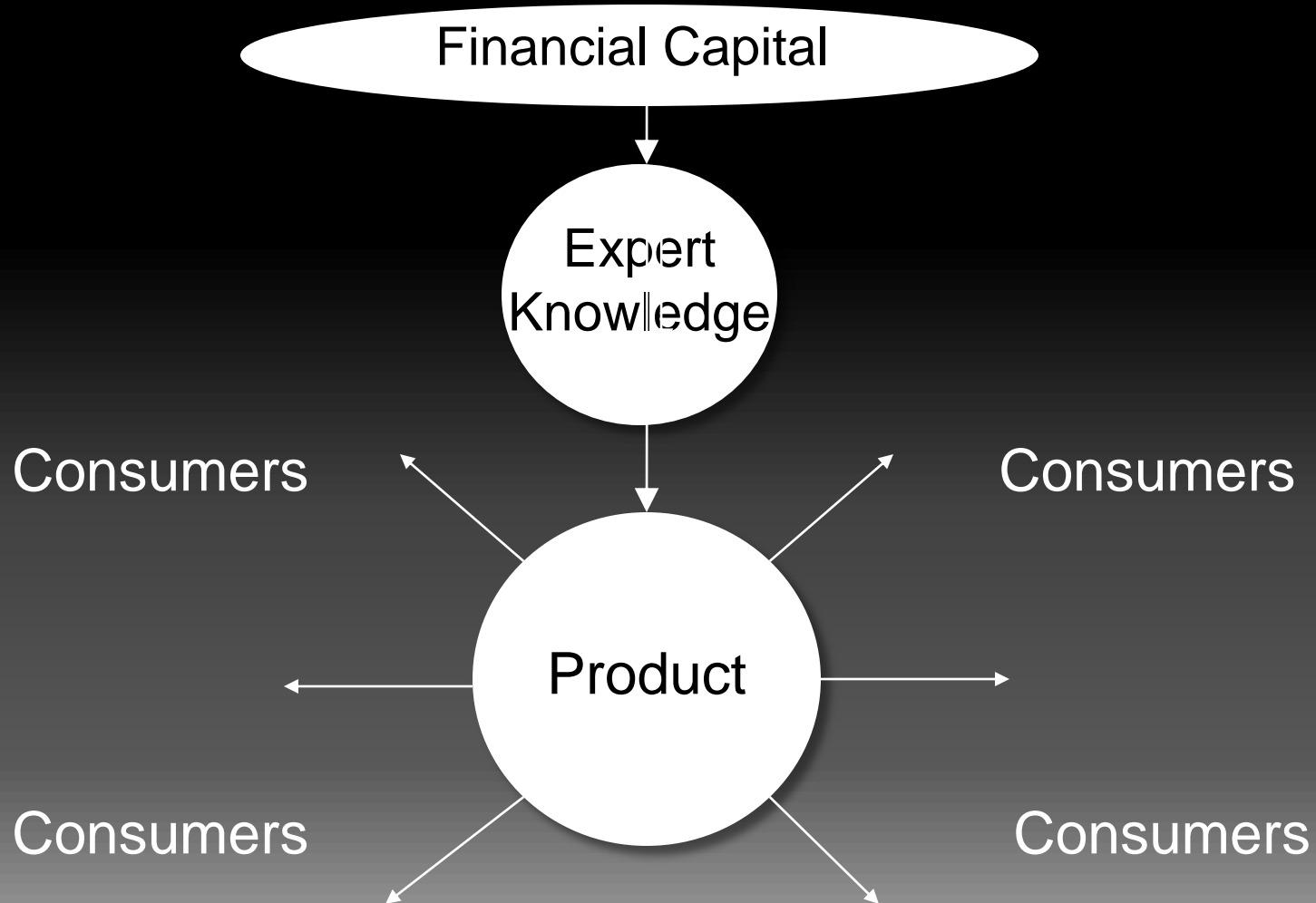**InterPARES 3 Project**
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Evolution…

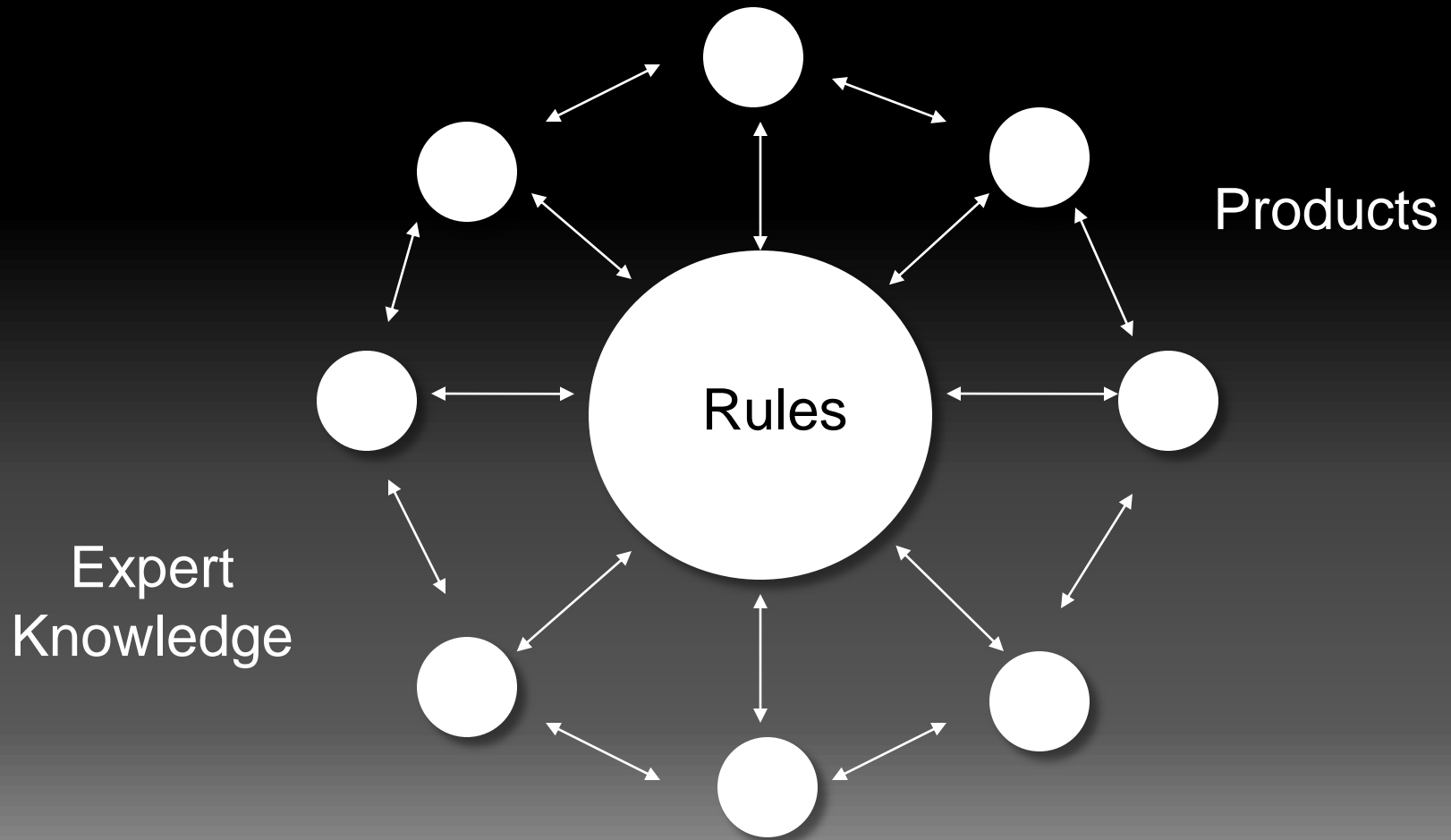- "Imagine if all past knowledge was kept hidden or its use was restricted to only those who are willing to pay for it. Education and research would suffer. Publishing books or sharing information of any sort would become difficult. Yet this is the mentality behind the proprietary software model. In the same way shared knowledge propels the whole of society forward, open technology development can drive innovation for an entire industry." http://www.redhat.com/about/whyopensource/

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Revolution…

"Just as the Copernican revolution was part of a broader social revolution that turned society away from hierarchy and received knowledge, and instead sparked a spirit of inquiry and knowledge sharing, open source is part of a communications revolution designed to maximize the free sharing of ideas expressed in code." (O'Reilly, 2008)

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

*From the **Industrial Model** …*

Financial Capital

Expert Knowledge

Consumers

Consumers

Product

Consumers

Consumers

…*to Connected Intelligence*

Rules

Products

Expert
Knowledge

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Transformation of Culture

How do we classify knowledge?

Centralization

Decentralization
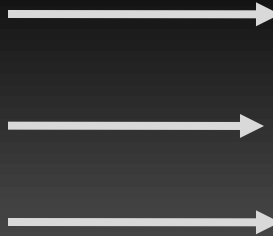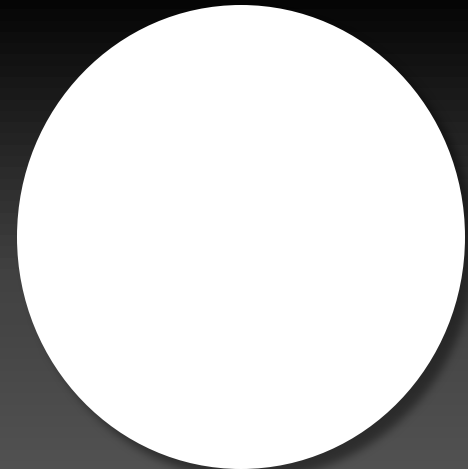
Assymetry   Commodification   Interdisciplinarity   Service

Open Source Software:
a model for the new paradigm?

Source code

- Distributed peer network

- Transparency of process

- Code can be used, modified and redistributed

- Code is licensed to make it available to the public

Licensing

Community

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Open Source Initiative

- Stewards of the open source definition
- Review and approval of licenses as OSD-compliant
- Community-building
- Education
- Public advocacy

Open Source Definition

1. Free Redistribution
2. Source Code
3. Derived Works
4. Integrity of Author's Source Code
5. No Discrimination Against Persons or Groups
6. No Discrimination Against Fields of Endeavor
7. Distribution of license
8. Licenses not specific to a product
9. Licenses do not restrict other software
10. Licenses are technology-neutral

Source code

Availability

Licensing

open source initiative

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Open Source Licensing

- All licenses must be in compliance with the OS definition

- Licenses are approved through a review process

- Purposes are consistency & transparency

Types of licenses

- Permissive - permit software to become proprietary (MIT, new BSD)
- Weakly protective (weak copyleft) - prevent the software component from becoming proprietary but permit it to be part of a larger proprietary system (LGPL, Mozilla Public License 1.1)
- Strongly protective (strong copyleft) - prevents the software from becoming proprietary (FLOSS - GPL)
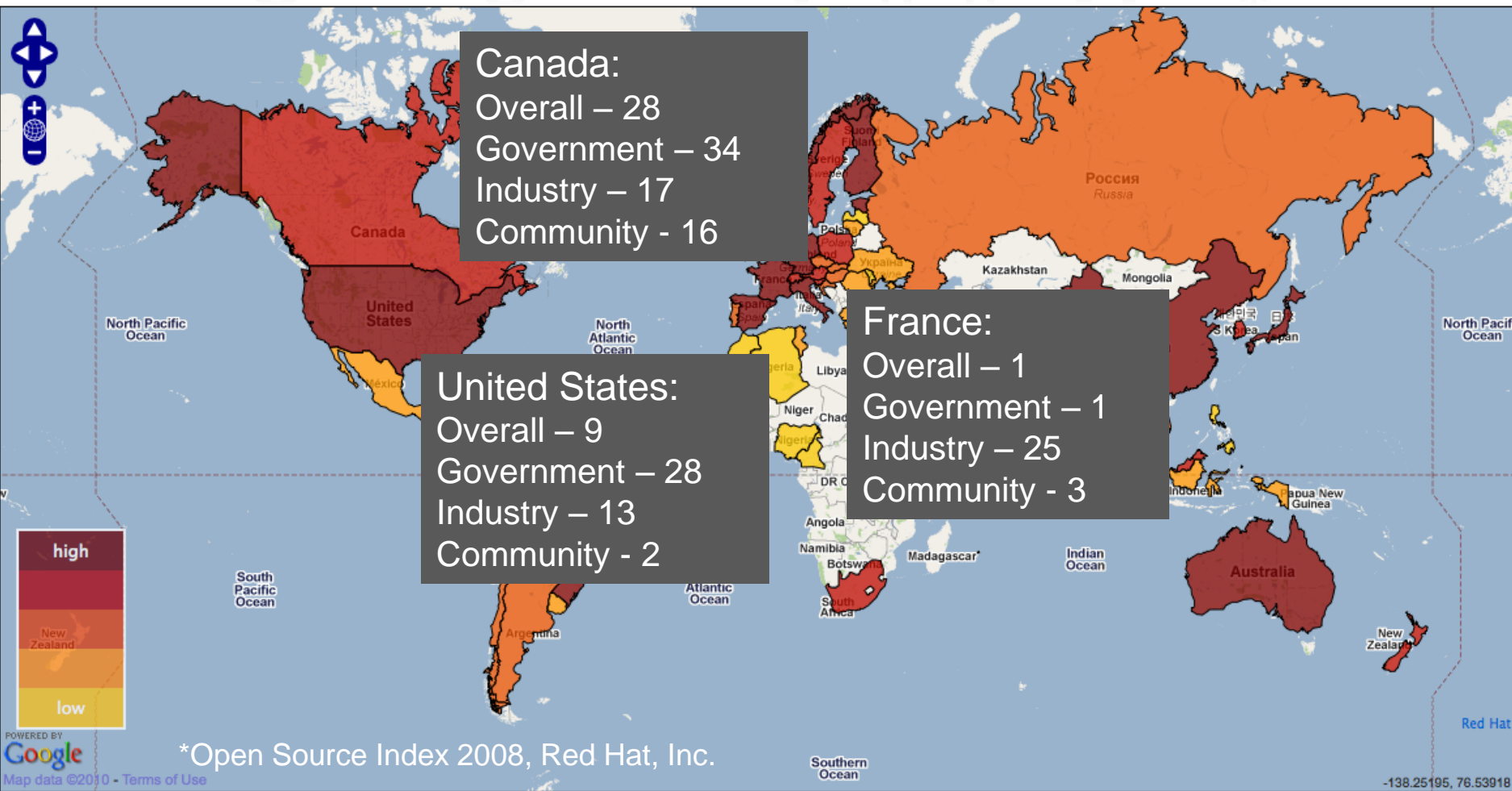
InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Types of licenses

The Free-Libre / Open Source Software (FLOSS) License Slide
by David A. Wheeler
September 27, 2007

There are a large number of Free-Libre / Open Source Software (FLOSS) licenses, but only a few are widely used. The widely-used licenses tend to be compatible, i.e., the software can be combined to produce a larger work. The following "license slide" figure makes it easy to see when common licenses can be combined:
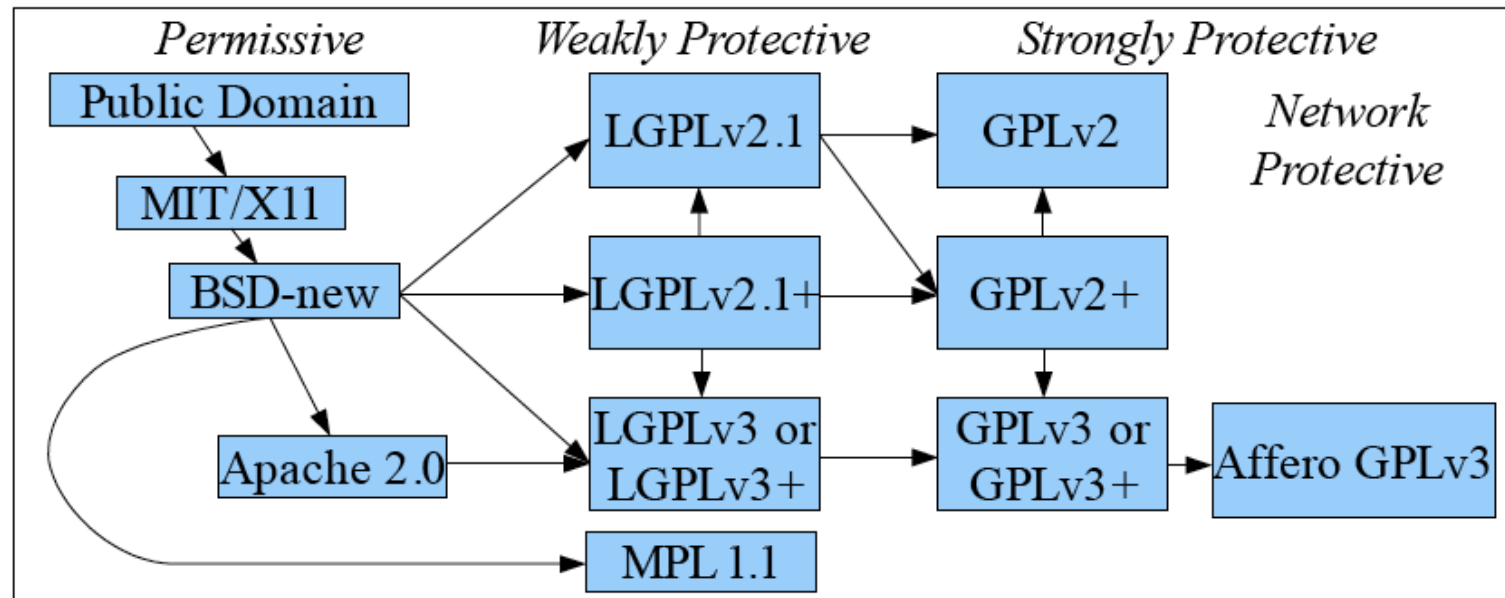
InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Licensing:
Copyright v. Copyleft

Copyright: Protects the individual creator from unrestricted distribution of his/her work

CopyLeft: protects the right to freely distribute a work without restrictions

Controlled access v. Free access

Open source licenses exist along a continuum:

Open - Proprietary   Open in Proprietary   Always Open

⟵——————————————————————————————⟶

Public domain   Weak copyleft   Strong copyleft

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Intellectual Property & Open Source

- Copyright
- Patent
- Trade Secrets

- "Viral" nature
- 3rd party infringement
- Validity of OS licenses

Open Source Activity Map*

Canada:
Overall – 28
Government – 34
Industry – 17
Community - 16

France:
Overall – 1
Government – 1
Industry – 25
Community - 3

United States:
Overall – 9
Government – 28
Industry – 13
Community - 2

*Open Source Index 2008, Red Hat, Inc.

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

OS in Libraries & Archives

What do Harvard, University of Florida, Stanford, Cornell, MIT, UC Berkeley and San Diego, the National Archives of the UK, Australia, the Netherlands and the Portuguese National Archives all have in common?

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Open digital repositories

http://www.dspace.org/why-use

DSPACE

ABOUT DSPACE
- About DSpace
- **Why Use DSpace?**
- Who's Using DSpace
- Use Case Examples
- Supporting Organization
- Service Providers

GETTING STARTED

USERS / DEVELOPERS

NEWS AND EVENTS

RESOURCES

Top Reasons to Use DSpace

"Largest community of users and developers worldwide"

"Free open source software"

"Completely customizable to fit your needs"

"Used by educational, government, private and commercial institutions"

"Can be installed out of the box

"Can manage and preserve all types of digital content"

View or download our informational brochure / spec sheet.

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Practical solutions in digital preservation

http://planets-suite.sourceforge.net/

PLANETS (Preservation and Long-term Access through Distributed NETworkS)

Planets is a four-year project co-funded by the European Commission to address core digital preservation challenges. The project has developed a suite of software tools and services to support preservation and long-term access to digital content.

Planets: http://planets-project.eu

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Practical solutions in digital preservation

Practical solutions in digital preservation

Open source records management?

- Is there a similar movement in the world of records management?
  - What products are available?
  - Do they adhere to existing RM standards?
  - How are they supported?
  - What is the uptake?
- If there are few products available, why?

Context of records management

- Active records not cultural assets to be shared
- Traditional business model
- Institution-based rather than collaborative
- Operating in relative isolation
- Security and privacy paramount
- EDRMS must integrate with other software
- EDRMS development is lucrative
- Institutional IT departments often want backing of well-established and familiar vendors

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Requirements for RM applications

- US Department of Defense DoD 5015.2-STD
- MoReq2
- National Archives of Australia, UK, NZ
- ICA Guidelines
- ISO 15489 & 23081

*i*Capture *i* Identify *i* Classify *i* Manage
Retain *i* Dispose *i* Search *i* Retrieve *i* Render

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Open source records management?

- Many content and document management systems, but not records management
- Only Alfresco offers a DoD-certified RM solution

"Alfresco reduces your ECM costs by up to 96% compared to proprietary systems like Documentum, Open Text and SharePoint. It's as simple to use as a shared drive or SharePoint and does not lock you in to a proprietary stack."   *www.alfresco.com*

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Alfresco

- Community version vs Enterprise version
- Provides its software under several different licenses depending on the user
- GPL - free to use
- OSI-approved - free to use w/FLOSS exception
- Flexible OEM commercial license
- Commercial license - subscription service

Alfresco Records Management

References

- O'Reilly, Tim (2008) "Open Sources 2.0/Beyond Open Source: Collaboration and Community/ The Open Source Paradigm Shift" http://commons.oreilly.com/wiki/index.php/Open_Sources_2.0/ Beyond_Open_Source:_Collaboration_and_Community/The_Open_Source_Paradigm_Shift.
- Thomas, John R. (2004) "Intellectual Property, Computer Software and the Open Source Movement." Congressional Research Service. March 11, 2004. Available at http://www.ipmall.info/hosted_resources/crs/RL32268_040311.pdf.
- Tiemann, Michael (2009) "How Open Source Software Can Save the ICT Industry One Trillion Dollars per Year," November 1, 2009.

Websites:
- http://apache.org
- http://archivematica.org/wiki/index.php?title=Main_Page
- http://en.wikipedia.org/wiki/Free_and_open_source_software
- http://lockss.stanford.edu/lockss/Home
- http://www.alfresco.com/
- http://www.dspace.org/
- http://www.fsf.org
- http://www.interpares.org
- http://www.openplanetsfoundation.org/
- http://www.opensource.org/
- http://www.planets-project.eu/
- http://www.redhat.com/

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

References

- The Open Source Definition (Annotated) - official definition of "open source software", with some explanations. http://www.opensource.org/docs/definition.php
- Free Software Definition - official definition of "Free software" (aka libre software; note the unusual capitalization). http://www.gnu.org/philosophy/free-sw.html
- Frequently Asked Questions (FAQ) about the GPL - Explains many issues relating to the GPL, and includes a detailed compatibility matrix for various versions of the GPL and LGPL (including some details about how they can be combined). http://www.gnu.org/licenses/gpl-faq.html
- Various Licenses and Comments About Them - Legal commentary by the Free Software Foundation (FSF) about many licenses. http://www.gnu.org/licenses/license-list.html
- "Commercial" is not the opposite of Free-Libre / Open Source Software (FLOSS)" - Explains why most FLOSS is commercial software. http://www.dwheeler.com/essays/commercial-floss.html
- Why Open Source Software / Free Software (OSS/FS, FLOSS, or FOSS)? Look at the Numbers! - Large collection of statistics on FLOSS programs. http://www.dwheeler.com/oss_fs_why.html
- Make Your Open Source Software GPL-Compatible. Or Else - Explains why FLOSS should be released under a GPL-compatible license, and includes many statistics showing that the GPL is the most popular FLOSS license. http://www.dwheeler.com/essays/gpl-compatible.html
- Maintaining Permissive-Licensed Files in a GPL-Licensed Project: Guidelines for Developers by the Software Freedom Law Center. http://www.softwarefreedom.org/resources/2007/gpl-non-gpl-collaboration.html

This list available at http://www.dwheeler.com/essays/floss-license-slide.html

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada

Last Thoughts

Questions?

Thank you

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems
TEAM Canada