Open Data Foundation

Managing social, behavioral and economic data and metadata
Guidelines for Tools Development and Operating Environments

Author/Maintainer
Pascal Heus, Open Data Foundation

Contributors
Jack Gager(Metadata Technology), Jannick Jensen (Danish Data Archive), Herve Lhours (UKDA)

Table of Contents

Overview

This documents contains technical information and recommendations for the development, deployment and operation of open source products and services for the management of social, behavioral and economic data and metadata. The objective is to provide harmonized guidelines for implementers and to foster the adoption of common technologies in order to maximize interoperability and integration of the open tools.

This document is open to suggestions/comments/contributions and will be revised on a regular basis.

Product Development

Environment Overview

Java is the recommended language for component developments with Adobe Flex 3 / AIR is the preferred platform for light desktop based end-user applications. Other stand-alone applications should be implemented on top of the Eclipse Rich Client Platform. Occasionally, a classic Swing or AWT user interface can be used for prototyping or for small applications that are not expected to integrate with other products. In case high-performance computing is needed, C/C++ can be used but the code must be portable and binaries made available for common platforms (Windows/Mac/Linux). It such case, it is also recommended to expose the C/C++ functionalities in Java using JNI[1].

For web site development, XHTML and Javascript / AJAX should be used in combination with pHp. We find the Yahoo! User Interface Library a good option for general scripting needs and also suggest web sites to take advantage of Google Analytics to track usage / traffic. Adobe Flex is recommended for rich client web based application. Apache / Tomcat servers is the preferred combination for the delivery of web sites and services with Orbeon forms as an optional product for implementation of XForms based products. Back-end database systems should be based on MySql, eXist (for light XML databases) and Oracle Berkeley DB-XML (for production systems). Free text indexing and search functionalities, when not directly available, should be built upon Apache Lucene (and related sub-projects).

A typical product implementation consists in a low level Java packages implementing the core functionalities that are made available to users through integration in stand-alone of web based applications. When operating over the network, the package is exposed to applications using standard web services technologies such as SOAP, REST, etc.

---
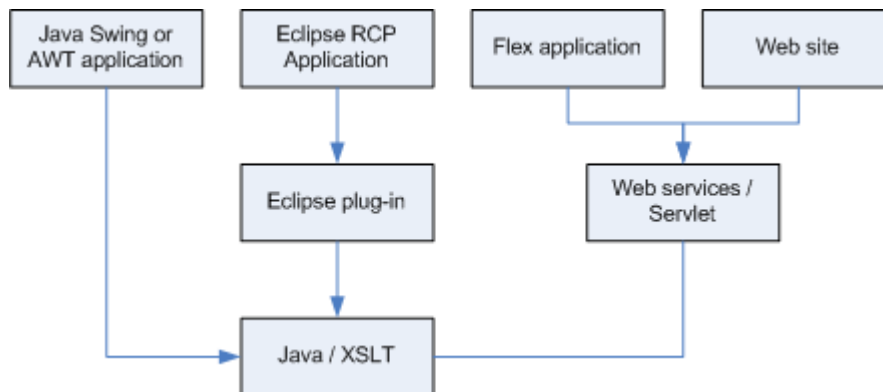
1   See http://en.wikipedia.org/wiki/Java_Native_Interface

Development Guidelines

The following table summarizes the characteristics of a typical environment for the development and operation of open source solutions.

Application Programming Language	Preferred: Java 5.0+ & Adobe Flex 3 / AIR Optional: C, C++
Web Programming Languages	PHp 5+ and Java Server Pages (JSP/J2EE)
Integrated Development Environment	Eclipse 3.2+
Naming Conventions / Techniques	Java: Java code should follow Sun's "Code Conventions for the Java Programming Language"[2] All source code must be commented using Javadoc[3] syntax following Sun's recommendations[4] All Java Java classes must be accompanied by a JUnit 4.0[5] unit testing use cases C/C++ All source code should be commented using Doxygen [need naming conventions for C/C++] pHp No official coding conventions have been identified for pHp. A widely used documentation tool is phpDocumentor[6]
Source code sharing and maintenance	Both CVS and Subversion are widely used by developers. Our preference goes to using Subversion. We do not recommend using MS Visual Source Safe given the limited functionalities and proprietary aspects of the system. Several public open source repositories are available. The Open Data Foundation has a G-Forge based repository open for public use[7]. Alternate options include SourceForge and ObjectWeb
Target platforms	All products should be platform independent. Core components should be available as Java packages or

---

2   See http://java.sun.com/docs/codeconv/
3   See http://java.sun.com/j2se/javadoc/
4   See http://java.sun.com/j2se/javadoc/writingdoccomments
5   See http://www.junit.org/
6   See http://www.phpdoc.org/
7   See http://forge.opendatafoundation.org

	Adobe Flex Components. UI based applications should operate in the the Eclipse Rich Client Platform (RCP) or Adobe AIR environments. Web based solutions should operate in a Java (JSP) and/or pHp environment and work with at least MS-Internet Explorer 6 and Mozilla Firefox.2 Rich web UI should be based on recommended products like Adobe Flex and the Adobe Integrate Runtime (AIR). In general, viewers and light applications can be web based but applications that requires serious processing or manage large quantities of information (1Mb+) should execute locally (stand alone).
Internationalization (i18n) / Localization (l10n)	All components, in particular end-user application interfaces, should be developed with glottalization in mind and support as much as possible translation into multiple languages.
Licensing	Source code should be released under a permissive open license that facilitate adoption by all users, including commercial companies. Default licensing model is LGPL.
Product Documentation	All documentation should be produced in open formats such as the Open Document Format (ODF) or XHTML

Project Sharing

Open Data Foundation Forge

http://forge.opendatafoundation.org

The Open Data Foundation provides a free public repository for hosting open source projects relevant to its user communities. The facility is based on G-Forge Advanced Server and provide typical functionalities such as source code management (CVS/Subversion), package release system, forums, wiki, documentation, bug tracking, project management.

ODaF Preferred Solution
While large pubic repositories may seemingly provide higher visibility, hosting your project at the ODaF Forge will ensure it reaches its intended community of users.

SourceForge

Http://www.sourceforge.net

SourceForge.net is the world's largest Open Source software development web site, hosting more than 100,000 projects and over 1,000,000 registered users with a centralized resource for managing projects, issues, communications, and code. SourceForge.net has the largest repository of Open Source code and applications available on the Internet, and hosts more Open Source development products than any other site or network worldwide.

ObjectWeb

http://www.objectweb.org/

The OW2 Consortium is an open source community committed to making available to everyone the best and most reliable middleware technology. OW2 mission is to develop open source code middleware and to foster a vibrant community and business ecosystem.

Licensing

This section provides information on recommended licensing models for the development of tools.

Commonly used licenses

Eclipse Public License (EPL)

http://www.eclipse.org/legal/epl-v10.html

The Eclipse Public License (EPL) is an open source software license used by the Eclipse Foundation for its software. It replaces the Common Public License (CPL) and removes certain terms relating to patent litigation.

The Eclipse Public License is designed to be a business friendly free software license, and features weaker copyleft provisions than contemporary licenses such as the GNU General Public License (GPL). The receiver of EPL-licensed programs can use, modify, copy and distribute the work and modified versions, in some cases being obligated to release their own changes.[1]

The EPL is approved by the Open Source Initiative (OSI)[2] and the Free Software Foundation (FSF) [3].

See http://en.wikipedia.org/wiki/Eclipse_Public_License

GNU Lesser Public License (LGPL)

http://www.gnu.org/copyleft/lesser.html

The GNU Lesser General Public License (formerly the GNU Library General Public License) or LGPL is a free software license published by the Free Software Foundation. It was designed as a compromise between the strong-copyleft GNU General Public License and permissive licenses such as the BSD licenses and the MIT License. The GNU Lesser General Public License was written in 1991 (and updated in 1999) by Richard Stallman, with legal advice from Eben Moglen.

The LGPL places copyleft restrictions on the program itself but does not apply these restrictions to other software that merely links with the program. There are, however, certain other restrictions on this software.

The LGPL is primarily intended for software libraries, although it is also used by some stand-alone applications, most notably Mozilla and OpenOffice.org.

See http://en.wikipedia.org/wiki/LGPL

ODaF Preferred Solution
This license is highly permissive and widely used amongst the open source development community.

BSD Licenses

http://www.opensource.org/licenses/bsd-license.php

BSD licenses represent a family of permissive free software licences. The original was used for the Berkeley Software Distribution, a Unix-like operating system for which the license is named. The original owners of BSD were the Regents of the University of California because BSD was first written at the University of California, Berkeley. The first version of the license was revised, and the resulting licenses are more properly called modified BSD licenses. Permissive licenses, sometimes with important differences pertaining to license compatibility, are referred to as "BSD-style licenses". Several BSD-like licenses, including the New BSD license, have been vetted by the Open Source Initiative as meeting their definition of open source.

The licenses have few restrictions compared to other free software licenses such as the GNU General Public License or even the default restrictions provided by copyright, putting it relatively closer to the public domain. The BSD licenses have been referred to as copycenter, as a comparison to standard copyright and copyleft free software: "Take it down to the copy center and make as many copies as you want."

See http://en.wikipedia.org/wiki/BSD_licenses

Other licenses

– http://en.wikipedia.org/wiki/List_of_FSF_approved_software_licenses

– http://www.opensource.org/

Examples

LGPL example

```
/*
 * <DESCRIPTION>
 *
 * Author(s): Pascal Heus (pheus@opendatafoundation.org)
 *            Jack Gager (jgager@metadatatechnology.com)
```

```
*
* This product is made possible through financial and
* technical support of the UK Data Archive Data Exchange Tools
* project (http://www.data-archive.ac.uk/dext/) and the
* Open Data Foundation (http://www.opendatafoundation.org)
*
* Copyright 2007 University of Essex (http://www.esds.ac.uk)
*
* This program is free software; you can redistribute it and/or modify it
* under the terms of the GNU Lesser General Public License as published by
* the Free Software Foundation; either version 3 of the License, or
* (at your option) any later version.
*
* This program is distributed in the hope that it will be useful,
* but WITHOUT ANY WARRANTY; without even the implied warranty of
* MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
* See the GNU Lesser General Public License for more details.
*
* You should have received a copy of the GNU General Public License
* along with this program.  If not, see <http://www.gnu.org/licenses/>.
*
*/
```

Products Guide

This sections provides general information on recommended products that can be used for the development and implementation of tools

Overview

Task/Activity	Recommended tools and techniques
Deploy a web server	Use Apache HTTP server in combination with pHp.
Develop a simple web site	Use XHTML and pHp. For development, use the Eclipse PDT or your favorite editor.
Develop interactive user interface for a web site	Use the Yahoo! User Interface (YUI) Javascript library
Develop a web site with a rich client interface	Use Adobe Flex or alternatively other open source solutions such as OpenLaszlo or the Eclipse RCP
Implement web services	Deploy Apache Tomcat and use Java Server Pages (JSP) to wrap the java packages. Expose their functionalities through SOAP/REST based interfaces. For development, use the Eclipse Java EE tools.
Develop a web front end using the XForms specification	Deploy and use Orbeon Forms
Maintain/distribute an open source product	Publish on the Open Data Foundation G-Forge using the LGPL or similar license
Convert an XML Schema into Java classes	Use Apache XML Beans or alternatively Eclipse EMF.
Develop Java Packages	*See product development section of this his document*
Index unstructured metadata or documents for free text search	Use Apache Lucene and related tools

Web server software

Apache HTTP Server

http://httpd.apache.org/

The Apache HTTP Server Project is an effort to develop and maintain an open-source HTTP server for modern operating systems including UNIX and Windows NT. The goal of this project is to provide a secure, efficient and extensible server that provides HTTP services in sync with the current HTTP standards.

Apache has been the most popular web server on the Internet since April 1996. The November 2005 Netcraft Web Server Survey found that more than 70% of the web sites on the Internet are using Apache, thus making it more widely used than all other web servers combined.

The Apache HTTP Server is a project of the Apache Software Foundation.

Apache Tomcat

http://tomcat.apache.org/

Apache Tomcat is the servlet container that is used in the official Reference Implementation for the Java Servlet and JavaServer Pages technologies. The Java Servlet and JavaServer Pages specifications are developed by Sun under the Java Community Process.

Apache Tomcat is developed in an open and participatory environment and released under the Apache Software License. Apache Tomcat is intended to be a collaboration of the best-of-breed developers from around the world. We invite you to participate in this open development project.

Apache Tomcat powers numerous large-scale, mission-critical web applications across a diverse range of industries and organizations

Apache Hibernate

http://www.hibernate.org/

Hibernate is a powerful, high performance object/relational persistence and query service. Hibernate lets you develop persistent classes following object-oriented idiom - including association, inheritance, polymorphism, composition, and collections. Hibernate allows you to express queries in its own portable SQL extension (HQL), as well as in native SQL, or with an object-oriented Criteria and Example API.

Unlike many other persistence solutions, Hibernate does not hide the power of SQL from you and guarantees that your investment in relational technology and knowledge is as valid as always. The LGPL open source license allows the use of Hibernate and NHibernate in open source and commercial projects.

Orbeon Forms

http://www.orbeon.com

 Orbeon Forms is an open source forms solution that handles the complexity of forms typical of the enterprise or government. It is delivered to standard web browsers (including Internet Explorer, Firefox, Safari and Opera) thanks to XForms and Ajax technology, with no need for client-side software or plugins. Orbeon Forms allows you to build fully interactive forms with features that include

as-you-type validation, optional and repeated sections, always up-to-date error summaries, PDF output, full internationalization, and controls like auto-completion, tabs, dialogs, trees and menus.

Orbeon Forms is built around a user-friendly Ajax-based XForms engine, which brings standard W3C XForms to mainstream browsers, and the XPL engine, a mature, high-performance XML pipeline engine for processing XML data.

pHp5

http://www.php.net/

PHP is a widely-used general-purpose scripting language that is especially suited for Web development and can be embedded into HTML.

Rich client Platforms

Adobe Flex 3 / AIR

http://www.adobe.com/devnet/flex/

http://labs.adobe.com/technologies/flex/

Adobe® Flex™ is a cross-platform development framework for creating rich Internet applications (RIAs). Flex enables you to create expressive, high-performance applications that run identically on all major browsers and operating systems.

http://labs.adobe.com/technologies/air/

Adobe® AIR™ lets developers use their existing web development skills in HTML, AJAX, Flash and Flex to build and deploy rich Internet applications to the desktop.

ODaF Preferred Solution
We recommend Flex for end user applications.

Flex combines the philosophy of open source with a industry standard technology well supported by a commercial vendor. Flex provides a rapid development environment that is well suited to manage XML objects. The newest version 3 combined with the AIR offers a platform independent operating environment and facilitates the reuse of components for both desktop and web based applications. While the Flex SDK is open source, it is recommended to acquire the Flex Builder, a commercial product sold by Adobe.

Eclipse Rich Client Platform

http://wiki.eclipse.org/index.php/Rich_Client_Platform

While the Eclipse platform is designed to serve as an open tools platform, it is architect ed so that its components could be used to build just about any client application. The minimal set of plug-ins needed to build a rich client application is collectively known as the Rich Client Platform.

Applications other than IDEs can be built using a subset of the platform. These rich applications are still based on a dynamic plug-in model, and the UI is built using the same toolkits and extension points. The layout and function of the workbench is under fine-grained control of the plug-in developer in this case.

When we say that the Rich Client Platform is the minimal set of plug-ins needed to build a platform application with a UI, we mean that your application need only require two plug-ins, org.eclipse.ui and org.eclipse.core.runtime, and their prerequisites.

However, rich client applications are free to use any API deemed necessary for their feature set, and can require any plug-ins above the bare minimum

ODaF Preferred Solution
We recommend the Eclipse RCP for the development of heavy duty administrative / management stand alone applications.

Eclipse provides an ideal platform independent environment for the integration of shared components (plug-ins), comes equipped with a built-in HTML based help system and supports automatic updates and multi-lingual interfaces.

OpenLazlo

http://www.openlaszlo.org/

OpenLaszlo is an open source platform for creating zero-install web applications with the user interface capabilities of desktop client software.

OpenLaszlo programs are written in XML and JavaScript and transparently compiled to Flash and, with OpenLaszlo 4, DHTML. The OpenLaszlo APIs provide animation, layout, data binding, server communication, and declarative UI. An OpenLaszlo application can be as short as a single source file, or factored into multiple files that define reusable classes and libraries.

OpenLaszlo is "write once, run everywhere." An OpenLaszlo application developed on one machine will run on all leading Web browsers on all leading desktop operating systems.

Development Tools

Eclipse

http://www.eclipse.org/

Eclipse is an open source community whose projects are focused on building an open development platform comprised of extensible frameworks, tools and runtimes for building, deploying and managing software across the lifecycle. A large and vibrant ecosystem of major technology vendors, innovative start-ups, universities, research institutions and individuals extend, complement and support the Eclipse platform.

Eclipse Rich Client Platform (RCP)

http://www.eclipse.org/home/categories/rcp.php

Eclipse RCP is a platform for building and deploying rich client applications. It includes Equinox, a component framework based on the OSGi standard, the ability to deploy native GUI applications to a variety of desktop operating systems, such as Windows, Linux and Mac OSX and an integrated update mechanism for deploying desktop applications from a central server.

Eclipse Modeling Framework (EMF)

http://www.eclipse.org/modeling/emf/

The EMF project is a modeling framework and code generation facility for building tools and other applications based on a structured data model. From a model specification described in XMI, EMF provides tools and runtime support to produce a set of Java classes for the model, along with a set of adapter classes that enable viewing and command-based editing of the model, and a basic editor.

EMF builds include XML Schema Infoset Model (XSD), now a component of the Model Development Tools (MDT) project, and an EMF-based implementation of Service Data Objects (SDO). Other subprojects are packaged separately. XSD provides a model and API for manipulating components of an XML Schema, with access to the underlying DOM representation of the schema document.

Eclipse Graphical Editing Framework (GEF)

http://www.eclipse.org/gef/

The Graphical Editing Framework (GEF) allows developers to take an existing application model and quickly create a rich graphical editor.

Eclipse pHp Development Tools (PDT)

http://www.eclipse.org/pdt

The PDT project provides a PHP Development Tools framework for the Eclipse platform. This project encompasses all development components necessary to develop PHP and facilitate extensibility. It leverages the existing Web Tools Project in providing developers with PHP capabilities.

Eclipse Plugins

EclipseNSIS

http://nsis.sourceforge.net/EclipseNSIS_-_NSIS_plugin_for_Eclipse

EclipseNSIS is a plugin for the Eclipse platform which enables editing, compiling and testing NSIS scripts.

Eclox

http://home.gna.org/eclox/

Eclox is a simple doxygen frontend plug-in for eclipse. It aims to provide a slim and sleek integration of the code documentation process into eclipse.

Eclox is a free software distributed under the term of the GNU General Public Licence (GPL).

Subclipse

http://subclipse.tigris.org/

Subclipse is an Eclipse Team Provider plug-in providing support for Subversion within the Eclipse IDE. The software is released under the Eclipse Public License (EPL) 1.0 open source license.

Subversion

http://subversion.tigris.org/

The goal of the Subversion project is to build a version control system that is a compelling replacement for CVS in the open source community. The software is released under an Apache/BSD-style open source license.

UMLet

http://www.umlet.com/

UMLet is an open-source UML tool with a simple user interface: draw UML diagrams fast, export diagrams to eps, pdf, jpg, svg, and clipboard, share diagrams using Eclipse, and create new, custom UML elements.

UMLet runs stand-alone or as Eclipse plug-in on Windows, OS X and Linux.

Libraries

Apache XML Beans

http://xmlbeans.apache.org/

XMLBeans is a tool that allows you to access the full power of XML in a Java friendly way. The idea is that you can take advantage of the richness and features of XML and XML Schema and have these features mapped as naturally as possible to the equivalent Java language and typing constructs. XMLBeans uses XML Schema to compile Java interfaces and classes that you can then use to access and modify XML instance data. Using XMLBeans is similar to using any other Java interface/class, you will see things like getFoo or setFoo just as you would expect when working with Java. While a major use of XMLBeans is to access your XML instance data with strongly typed Java classes there are also API's that allow you access to the full XML infoset (XMLBeans keeps XML Infoset fidelity) as

well as to allow you to reflect into the XML schema itself through an XML Schema Object model.

Given the ease of use of XML Beans, we strongly recommend producing XML Beans packages for XML specifications you are using. ODaF is currently working with the DDI, SDMX and other standards and will make such libraries available in its public repository at http://forge.opendatafoundation.org

Apache Digester

http://commons.apache.org/digester/

Many projects read XML configuration files to provide initialization of various Java objects within the system. There are several ways of doing this, and the Digester component was designed to provide a common implementation that can be used in many different projects.

Basically, the Digester package lets you configure an XML -> Java object mapping module, which triggers certain actions called rules whenever a particular pattern of nested XML elements is recognized. A rich set of predefined rules is available for your use, or you can also create your own. Advanced features of Digester include:

– Ability to plug in your own pattern matching engine, if the standard one is not sufficient for your requirements.
– Optional namespace-aware processing, so that you can define rules that are relevant only to a particular XML namespace.
– - Encapsulation of Rules into RuleSets that can be easily and conveniently reused in more than one application that requires the same type of processing.

Digester is particularly useful when combined with Apache Lucene to index XML document for free text search. See http://www.ibm.com/developerworks/java/library/j-lucene/ for further information.

Bug Tracking

Web based bug tracking tools are very useful tools for development teams, in particular in virtual environments where individuals are often working from different locations. Most open source repositories such as the ODaF Forge, SourceForge and other provide bug tracking facilities and these are useful for interaction with end users. It is also often useful to operate stand alone interanal such as the ones presented below. Most of these tools provide integration with the Eclipse development environment.

BugZilla

http://www.bugzilla.org/

Bugzilla is a "Defect Tracking System" or "Bug-Tracking System". Defect Tracking Systems allow individual or groups of developers to keep track of outstanding bugs in their product effectively. Most commercial defect-tracking software vendors charge enormous licensing fees. Despite being "free",

Bugzilla has many features its expensive counterparts lack. Consequently, Bugzilla has quickly become a favorite of hundreds of organizations across the globe.

BugZilla requirements includes Perl, a database servers such as MySql or PostgreSQL and a web server. It is typically installed in a Linux or Solaris environments.

Mantis

http://mantisbt.org/

Mantis is a free popular web-based bugtracking system (feature list). It is written in the PHP scripting language and works with MySQL, MS SQL, and PostgreSQL databases and a webserver. Mantis has been installed on Windows, Linux, Mac OS, OS/2, and others. Almost any web browser should be able to function as a client. It is released under the terms of the GNU General Public License (GPL).

Testing and Packaging

JUnit

http://www.junit.org/

JUnit is a regression testing framework written by Erich Gamma and Kent Beck. It is used by the developer who implements unit tests in Java. JUnit is Open Source Software, released under the Common Public License Version 1.0 and hosted on SourceForge.

NSIS

http://nsis.sourceforge.net/

NSIS (Nullsoft Scriptable Install System) is a professional open source system to create Windows installers. It is designed to be as small and flexible as possible and is therefore very suitable for internet distribution.

Being a user's first experience with your product, a stable and reliable installer is an important component of succesful software. With NSIS you can create such installers that are capable of doing everything that is needed to setup your software.

NSIS is script-based and allows you to create the logic to handle even the most complex installation tasks. Many plug-ins and scripts are already available: you can create web installers, communicate with Windows and other software components, install or update shared components and more.

XML Products

Saxon

http://saxon.sourceforge.net/

Saxon is an open-source XSLT and XQuery processor created by Michael Kay.

Saxon comes in two packages: Saxon-B implements the "basic" conformance level for XSLT 2.0 and XQuery, while Saxon-SA is a schema-aware XSLT and XQuery processor. Both packages are available on Java and .NET. The open-source provides provides all the features of the XLST 2.0 languages except schema-aware processing. It reflects the syntax of the final XSLT 2.0, XQuery 1.0, and XPath 2.0 Recommendations of 23 January 2007.

Documentation

Doxygen

http://www.stack.nl/~dimitri/doxygen/

Doxygen is a documentation system for C++, C, Java, Objective-C, Python, IDL (Corba and Microsoft flavors) and to some extent PHP, C#, and D.

It can help you in three ways:

1. It can generate an on-line documentation browser (in HTML) and/or an off-line reference manual (in $\mbox{\LaTeX}$) from a set of documented source files. There is also support for generating output in RTF (MS-Word), PostScript, hyperlinked PDF, compressed HTML, and Unix man pages. The documentation is extracted directly from the sources, which makes it much easier to keep the documentation consistent with the source code.

2. You can configure doxygen to extract the code structure from undocumented source files. This is very useful to quickly find your way in large source distributions. You can also visualize the relations between the various elements by means of include dependency graphs, inheritance diagrams, and collaboration diagrams, which are all generated automatically.

3. You can even `abuse' doxygen for creating normal documentation (as I did for this manual).

Doxygen is developed under Linux and Mac OS X, but is set-up to be highly portable. As a result, it runs on most other Unix flavors as well. Furthermore, executables for Windows are available.

Javadoc

http://java.sun.com/j2se/javadoc/

Javadoc is a tool for generating API documentation in HTML format from doc comments in source code

OpenOffice

OpenOffice.org is a multiplatform and multilingual office suite and an open-source project. Compatible with all other major office suites, the product is free to download, use, and distribute.

OpenOffice documents are compatible with the OASIS Open Document Format for Office Application[8].

Wiki?

To be determined. Probably MediaWiki

Database

MySql

http://mysql.com/

The MySQL® database has become the world's most popular open source database because of its consistent fast performance, high reliability and ease of use. It's used in more than 11 million installations ranging from large corporations to specialized embedded applications on every continent in the world. (Yes, even Antarctica!)

Not only is MySQL the world's most popular open source database, it's also become the database of choice for a new generation of applications built on the LAMP stack (Linux, Apache, MySQL, PHP / Perl / Python.) MySQL runs on more than 20 platforms including Linux, Windows, OS/X, HP-UX, AIX, Netware, giving you the kind of flexibility that puts you in control.

Whether you're new to database technology or an experienced developer or DBA, MySQL offers a comprehensive range of certified software, support, training and consulting to make you successful.

Exist

http://exist.sourceforge.net/

 eXist is an Open Source native XML database featuring efficient, index-based XQuery processing, automatic indexing, extensions for full-text search, XUpdate support, XQuery update extensions and tight integration with existing XML development tools. The database implements the current XQuery 1.0 working drafts, with exception of the schema import and schema validation features defined as optional in the XQuery specification.

eXist provides a powerful environment for the development of web applications based on XQuery and related standards. Entire web applications can be written in XQuery, using XSLT, XHTML, CSS and maybe Javascript (for AJAX functionality). XQuery server pages can be executed from the filesystem or stored in the database.

 While we do recognize that this is a very popular product that is very easy to setup and use out of the box, we do not recommend it for production system managing large databases or XML documents as performance and reliability have a tendency to degrade at the metadata collections size increase. This is otherwise an excellent choice for entry

---

8   See http://en.wikipedia.org/wiki/OpenDocument and http://www.oasis-open.org/committees/office/

level XML database or for educational purposes.

Oracle Berkeley DB Product Family

http://www.oracle.com/database/berkeley-db

Oracle Berkeley DB[9] is a family of open source embeddable databases that allows developers to incorporate within their applications a fast, scalable, transactional database engine with industrial grade reliability and availability. As a result, customers and end-users will experience an application that simply works, reliably manages data, can scale under extreme load, but requires no ongoing database administration. As a developer, you can focus on your application and be confident that Oracle Berkeley DB will manage your persistence needs.

Oracle Berkeley DB

http://www.oracle.com/database/berkeley-db/db/

The World's Most Popular High Performance, Embeddable Database

Oracle Berkeley DB is the industry-leading open source, embeddable database engine that provides developers with fast, reliable, local persistence with zero administration. Oracle Berkeley DB is a library that links directly into your application. Your application makes simple function calls, rather than sending messages to a remote server, eliminating the performance penalty of client-server architectures. Oracle Berkeley DB eliminates the overhead of SQL query processing, enabling applications with predictable access patterns to run faster.

Oracle Berkeley DB Java Edition

http://www.oracle.com/database/berkeley-db/je

High Performance, Pure Java, Embeddable Database Engine

Oracle Berkeley DB Java Edition is a open source, embeddable, transactional storage engine written entirely in Java. Like Oracle Berkeley DB, Oracle Berkeley DB Java Edition executes in the address space of the application, without the overhead of client/server communication. It stores data in the application's native format, so no runtime data translation is required. It provides an easy-to-use, programmatic interface, allowing developers to store and retrieve information quickly, simply and reliably.

Oracle Berkeley DB Java Edition was designed from the ground up in Java. It takes full advantage of the Java environment to simplify development and deployment. The architecture of Oracle Berkeley DB Java Edition supports very high performance and concurrency for both read-intensive and write-intensive workloads.

---

9   This used to be known Sleepycat. They were acquired by Oracle in 2006 but remain open source

Oracle Berkeley DB XML

http://www.oracle.com/database/berkeley-db/xml

High Performance, Embeddable XML Database Engine

Oracle Berkeley DB XML is an open source, embeddable XML database with XQuery-based access to documents stored in containers and indexed based on their content. Oracle Berkeley DB XML is built on top of Oracle Berkeley DB and inherits its rich features and attributes. Like Oracle Berkeley DB, it runs in process with the application with no need for human administration. Oracle Berkeley DB XML adds a document parser, XML indexer and XQuery engine on top of Oracle Berkeley DB to enable the fastest, most efficient retrieval of data.

ODaF Preferred Solution

Berkeley DB XML is a reliable solution that delivers high level of performance and scales-up to large databases. This product however only provides core functionalities and does not come with a does with any administrative or user interfaces. It is therefore intended for software developers. Less technical users should consider eXist as an alternate solution.

Note that ODaF is developing Java classes that facilitate the use of Berkeley DBXML for the management of metadata catalogs. Contact us for more information.

Text / Document Indexing and Searching

Relational and XML database systems are very good at storing and indexing key elements of information but often perform very poorly when it comes to free text search. The following set of products are recommended as solution or complementary tools for implementing free text search on metadata or documents.

Apache Lucene

http://lucene.apache.org

Apache Lucene is a high-performance, full-featured text search engine library written entirely in Java. It is a technology suitable for nearly any application that requires full-text search, especially cross-platform.

Lucene provides a very easy to use and implement Java based text / document indexing engine. It is widely used and well supported by the Apache Foundation. The product has also been ported to several environments sucha s C++, .NET or Perl. Combine with complementary projects such as Apache Solr or Digester, it provides a complete toolkit for quickly searching through unstructured information.

Apache Solr

http://lucene.apache.org/solr/

Solr is an open source enterprise search server based on the Lucene Java search library, with XML/HTTP and JSON APIs, hit highlighting, faceted search, caching, replication, and a web administration interface. It runs in a Java servlet container such as Tomcat.

Luke Index Toolbox for Lucene

http://www.getopt.org/luke/

Luke is a handy development and diagnostic tool, which accesses already existing Lucene indexes and allows you to display and modify their contents in several ways:

- browse by document number, or by term
- view documents / copy to clipboard
- retrieve a ranked list of most frequent terms
- execute a search, and browse the results
- analyze search results
- selectively delete documents from the index
- reconstruct the original document fields, edit them and re-insert to the index
- optimize indexes
- and much more...

Security

While not directly part of the ODaF development framework, security is an important component for most of today's product implementation. We in general recommend following standards and best practices outlined under the OASIS security committees[10]. The following products are aligned or implement these standards.

Sun's XACML Implementation

http://sunxacml.sourceforge.net/

This is an open source implementation of the OASIS XACML standard, written in the JavaTM programming language

This project provides complete support for all the mandatory features of XACML as well as a number of optional features. Specifically, there is full support for parsing both policy and request/response documents, determining applicability of policies, and evaluating requests against policies. All of the standard attribute types, functions, and combining algorithms are supported, and there are APIs for adding new functionality as needed. There are also APIs for writing new retrieval mechanisms used for finding things like policies and attributes.

---

10 See http://www.oasis-open.org/committees/tc_cat.php?cat=security

Shibboleth

http://shibboleth.internet2.edu/

Shibboleth is standards-based, open source middleware software which provides Web Single SignOn (SSO) across or within organizational boundaries. It allows sites to make informed authorization decisions for individual access of protected online resources in a privacy-preserving manner.

Java / .NET Bridge

While not commonly used for the development of open source solutions, Microsoft .NET is a solid technology for the implementation of both web based and desktop application. One simple way to use .NET with Java based products is to establish a dialog between applications through a generic XML layer for example using web services. This is not however always possible and it is sometimes necessary to directly integrate Java packages in .NET. The products listed below are available to do so, though most of them are commercial packages.

The following article also provides useful information on how to integrate Java and .NET:

– "Integrating Java* and Microsoft .NET*", Jon Jagger, Content Master Ltd., Aug 2007
   http://softwarecommunity.intel.com/articles/eng/3105.htm

– "Portable GUIs Improve Application Flexibility", John Sharp, Content Master Ltd, Aug 2007,
   http://softwarecommunity.intel.com/articles/eng/3039.htm

– "Calling Java Classes Directly from .NET", Wayne Citrin, Jan 2004,
   http://www.devx.com/interop/Article/19945/1763

Note that the products listed below have not been evaluated and this topic needs further input. Suggestions / contributions are most welcomed.

Visual J# Binary Converter Tool (JBIMP.EXE)

http://msdn2.microsoft.com/en-us/library/y9teabc2(VS.80).aspx

The Microsoft Visual J# Binary Converter Tool, or jbimp.exe, is a tool for converting compiled Java-language .class files (bytecode) into .NET Framework assemblies. jbimp.exe is a command line tool, and can be accessed via the Visual Studio .NET Command Prompt.

IKVM.NET (free)

http://www.ikvm.net

IKVM.NET is an implementation of Java for Mono and the Microsoft .NET Framework. It includes the following components:

– A Java Virtual Machine implemented in .NET

- A .NET implementation of the Java class libraries
- Tools that enable Java and .NET interoperability

jnbridge (commercial)

http://www.jnbridge.com/

JNBridge interoperability tools enable you to access Java from .NET, or access .NET from Java.