



# Open Source Document Management – Warum und warum jetzt?

---

Sebastian Wohlrapp

Zürich, 5. September 2006

OpenTuesday@**DIGICOMP**

# Sebastian Wohlrapp

---

- ◆ Cambridge Technology Partners Deutschland
- ◆ mediaworx berlin AG
- ◆ Novell / Cambridge Technology Partners Schweiz
- ◆ Optaros
  
- ◆ CMS and Collaboration Practice Lead



# Agenda

---

- ◆ About Optaros
- ◆ Open Source
- ◆ Introduction to Document Management
- ◆ Open Source Value Add for Document Management
- ◆ Looking into leading Open Source Document Management Solutions (market view)
- ◆ How to evaluate and implement Open Source Document Management Solutions
- ◆ Alfresco Demo

# Agenda

---

- ◆ **About Optaros**
- ◆ Open Source
- ◆ Introduction to Document Management
- ◆ Open Source Value Add for Document Management
- ◆ Looking into leading Open Source Document Management Solutions (market view)
- ◆ How to evaluate and implement Open Source Document Management Solutions
- ◆ Alfresco Demo

## Our mission

---

# Realize the Benefits of Open Source

Optaros is a **global consulting and systems integration** firm that provides enterprises with **best-fit solutions** to IT business challenges, maximizing the benefits of **Open Standards, Open Source** software and **Global Sourcing**. Whether your organization has had little exposure to open source software or has open source-based solutions currently in place, Optaros has the business and IT problem solving experience to advise on the areas where open source can be effective and how to proactively manage the use and benefits of open source software.

## Optaros - who we are

---

<b>Founded:</b>	2004
<b>Capitalization:</b>	VC funded (Charles River Ventures, General Catalyst, ...) – first round more than CHF 10'000'000.-
<b>CEO:</b>	Bob Gett
<b>Leadership team:</b>	Mavis Chin, Stephen Wally, Brian Otis, Seth Gottlieb, Miguel Picornell, Dave Gynn, Frederic Weill, Philippe Grosjean, Heinz Rudin, Bruno von Rotz
<b>Community:</b>	<b>Global</b> company, run community based across the world. One of the largest global consulting and integration firm focusing on Open Source
<b>Offices:</b>	Boston, Geneva, Zurich, New York, San Francisco, Atlanta, Chicago*, Frankfurt*

\* 2006 (planned)

# Optaros Solution Areas and Clients

Content Management & Document Management

Portal & Next Generation Internet / Web 2.0

CRM & Self Service

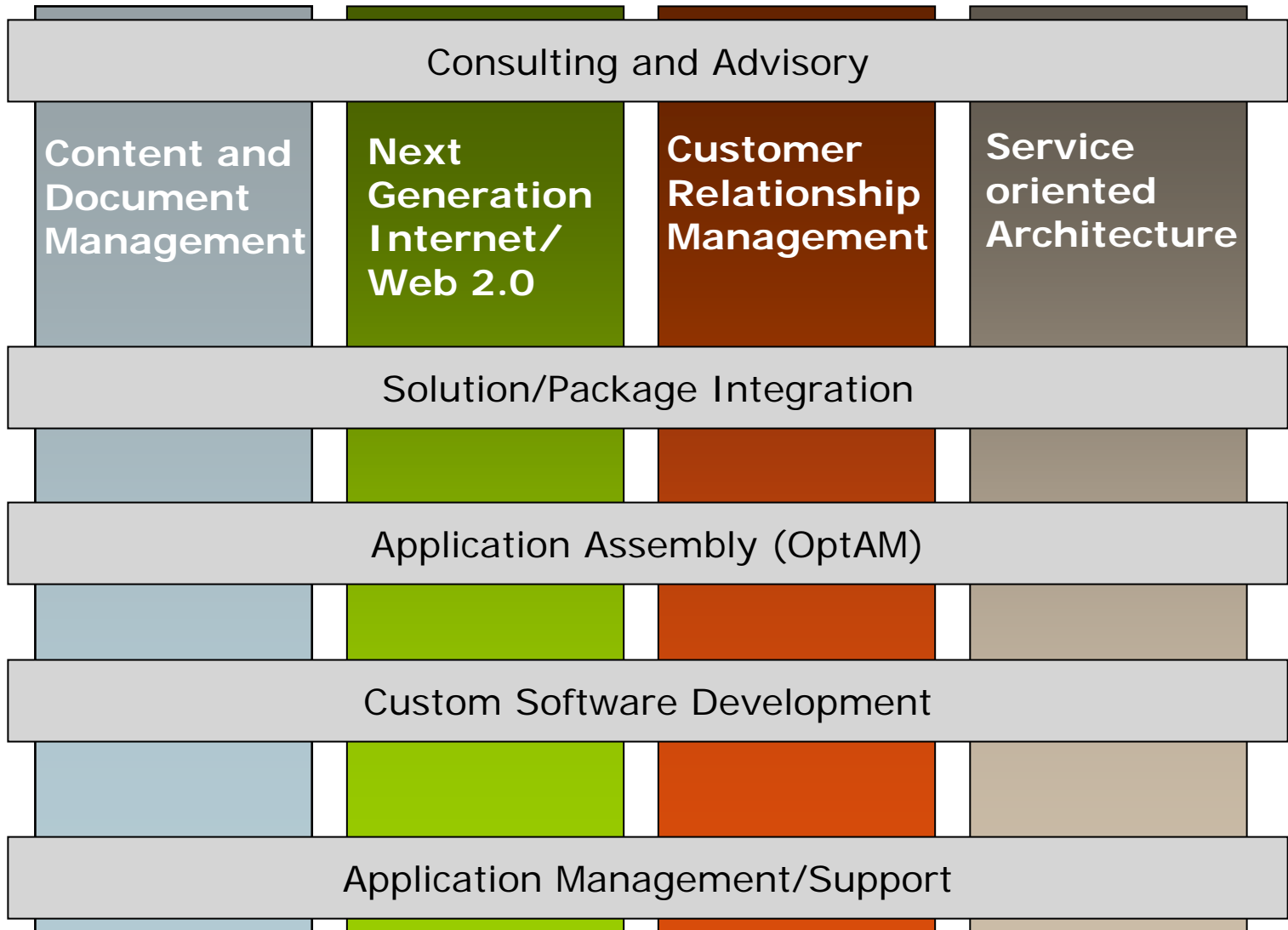
SOA & Workflow Management

Identity Management & Infrastructure Optimization

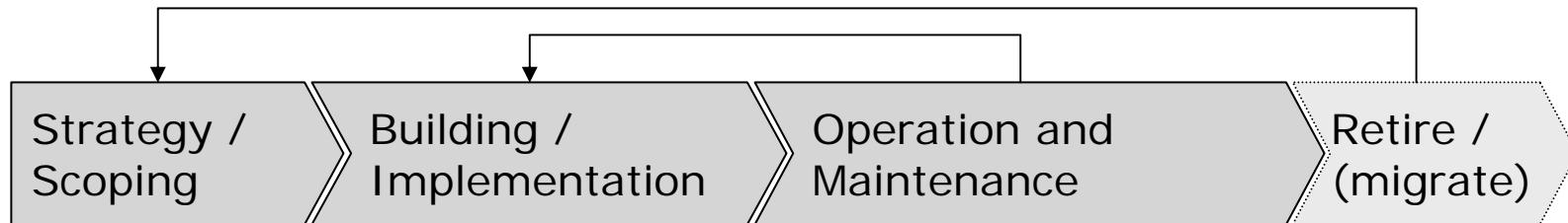
Java EE solution implementation

Consulting & Advisory

## Our focus – four practice areas



# Optaros services along the „application lifecycle“



Strategy, advisory, scoping, evaluation, Project&Program Management

- Application Assembly
- Solution Integration
- Custom development

Application Management and Support,  
Testing Services

# Mastering technology to implement better business applications (examples)

Applications



Solution components



Application Server ++



Database



Operating System



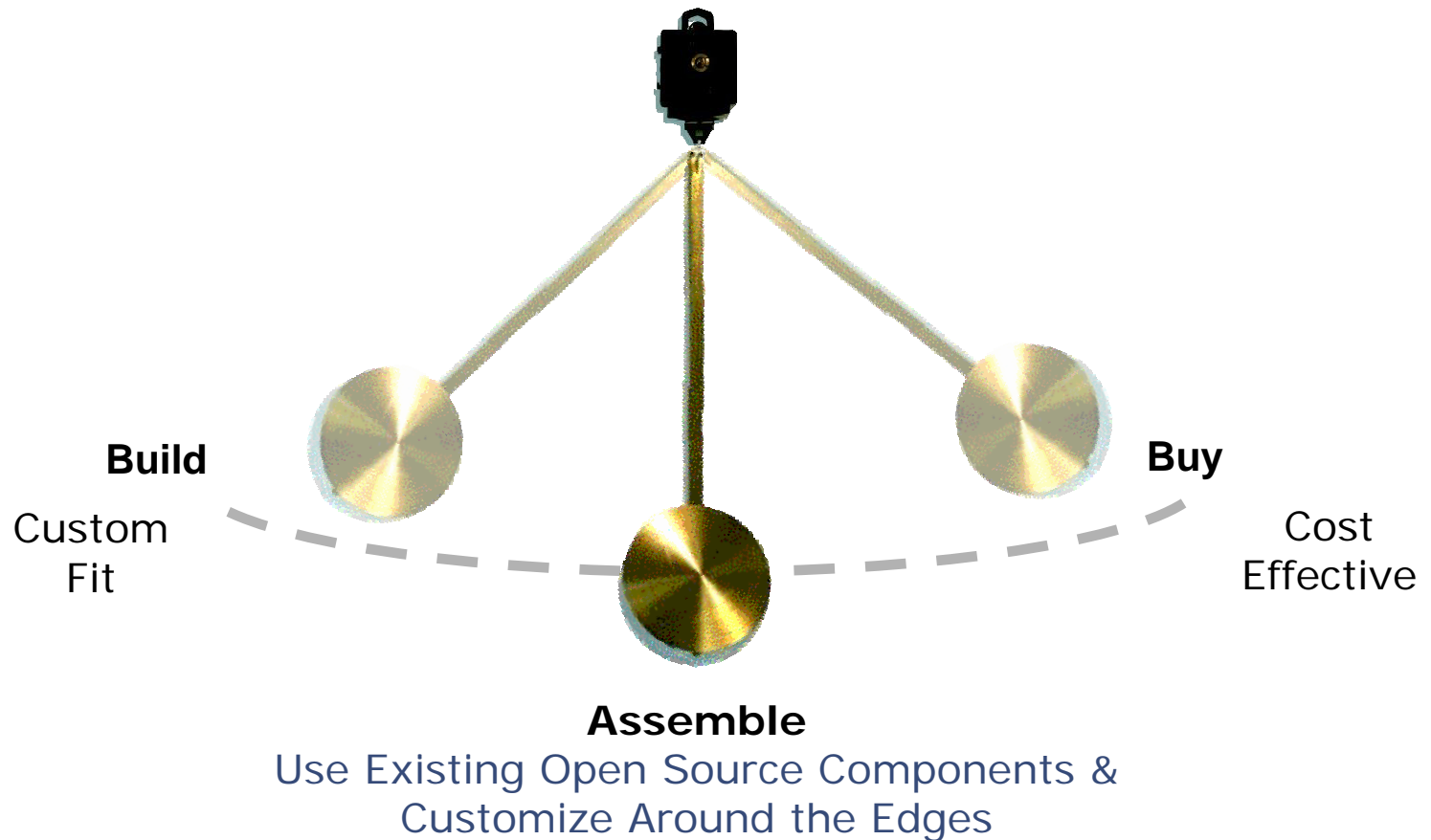
Hardware



Open Source ← → Commercial

If building is too expensive, and buying doesn't deliver the solution you need – go for OptAM

## From "Buy vs. Build" to Application Assembly



# Optaros in Switzerland

The map shows the following logos and locations:

- Zurich:** ABB, SWISSEX Westlaw, sunrise, Stadt Zürich Organisation und Informatik.
- Geneva:** ILO, a coat of arms, a cross and scales icon.
- Other locations:** swisscom, RMG (the risk management group), serono (biotech & beyond).

- Established since 2005
- Currently 28 employees in Zurich and Geneva
- Adding 2-4 employees per month
- Objective: Presence with 45-50 consultants in the next 10-12 months

# Agenda

---

- ◆ About Optaros
- ◆ **Open Source**
- ◆ Introduction to Document Management
- ◆ Open Source Value Add for Document Management
- ◆ Looking into leading Open Source Document Management Solutions (market view)
- ◆ How to evaluate and implement Open Source Document Management Solutions
- ◆ Alfresco Demo

# The Open Source Definition

- 1) **Free Redistribution**
  - No restrictions for using software as a component
  - No royalties or fees
- 2) **Source Code**
  - THE basis for changes and inspections
- 3) **Derived Work**
  - Modifications and derived work is possible
  - Same license as the original work
- 4) **Integrity of the Author's Source Code**
- 5) **No Discrimination Against Persons or Groups**
- 6) **No Discrimination Against Fields of Endeavor**
  - No restrictions on what the technology is used for is permitted
- 7) **Distribution of License**
- 8) **License must not be specific to the Product**
  - The initial rights attached to the program are inherited
- 9) **License must not restrict other Software**
- 10) **License must be technology-neutral**

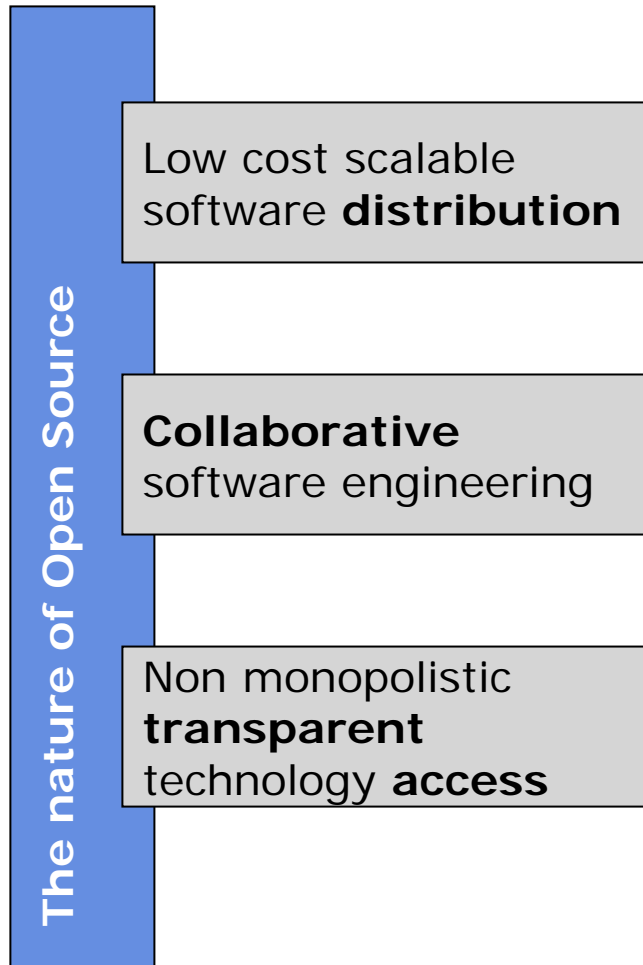


# What is Open Source really?

- ◆ More than **180'000 projects**
- ◆ Wide **range** of information technology platforms, solutions and components
- ◆ Production ready tools but also obsolete, useless projects and pure "think ware"



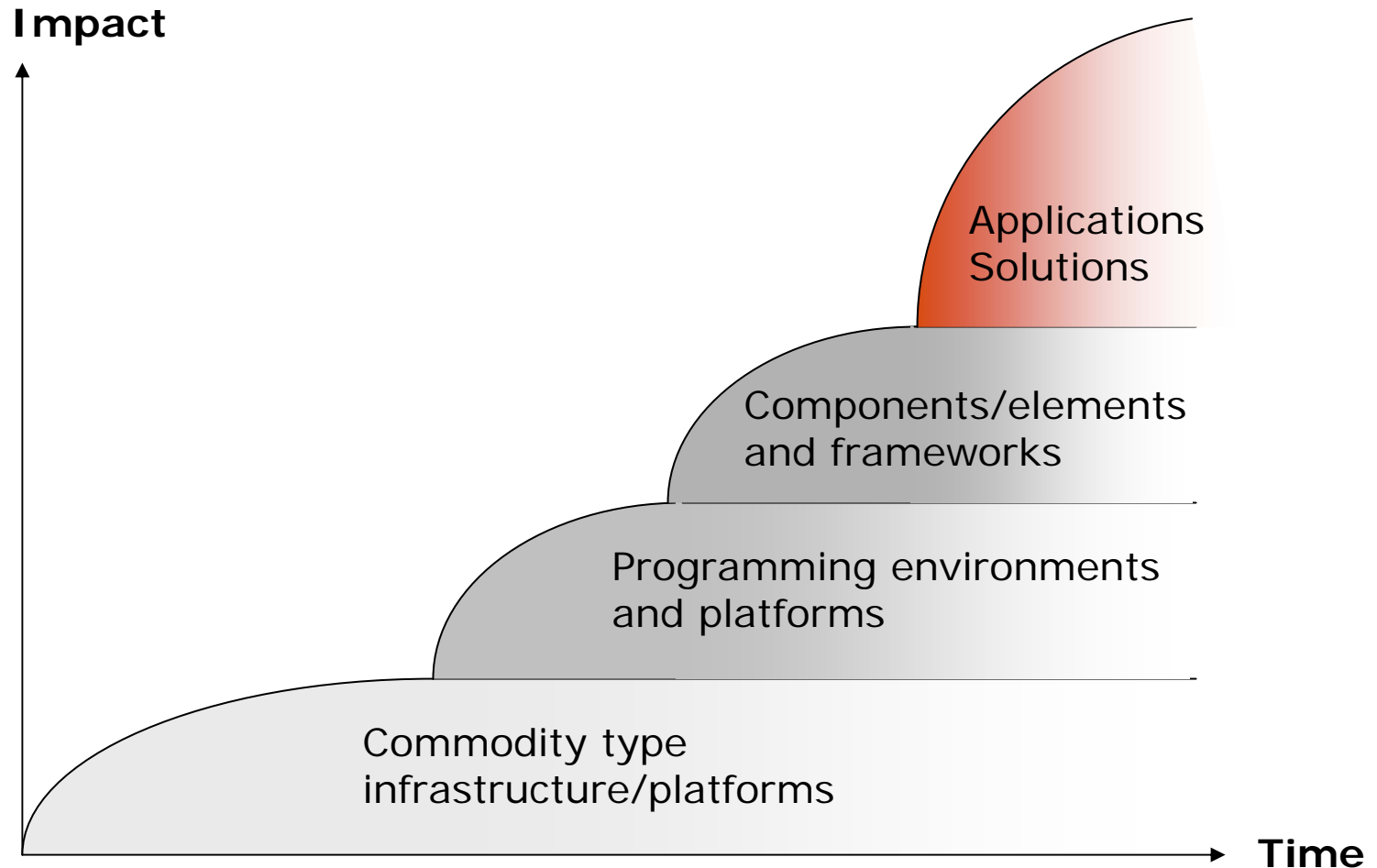
# The (r)evolutionary elements of Open Source



# Some key conditions favor the development and the success of Open Source

- ◆ The “**closer**” the “problem” to the typical issues of a **developer** is (or the more users and developers are the same) , the better are the chances for Open Source projects
- ◆ The bigger the **number of potential users** the more traction Open Source projects get
- ◆ **Unsatisfying economical frame conditions** in a specific solution area (e.g. expensive commercial software, lack of language versions, etc.) support the development of Open Source alternatives
- ◆ The existence of a product, that many/most people **hate** due to various reasons (e.g. monopolistic behavior of the vendor) supports the fast development and the rapid adoption of Open Source alternatives
- ◆ **Clearly visible gaps** in the existing Open Source solution portfolio establish the ground for commercial Open Source solutions

# Applications are seen as the next big wave in Open Source



# Agenda

---

- ◆ About Optaros
- ◆ Open Source
- ◆ **Introduction to Document Management**
- ◆ Open Source Value Add for Document Management
- ◆ Looking into leading Open Source Document Management Solutions (market view)
- ◆ How to evaluate and implement Open Source Document Management Solutions
- ◆ Alfresco Demo

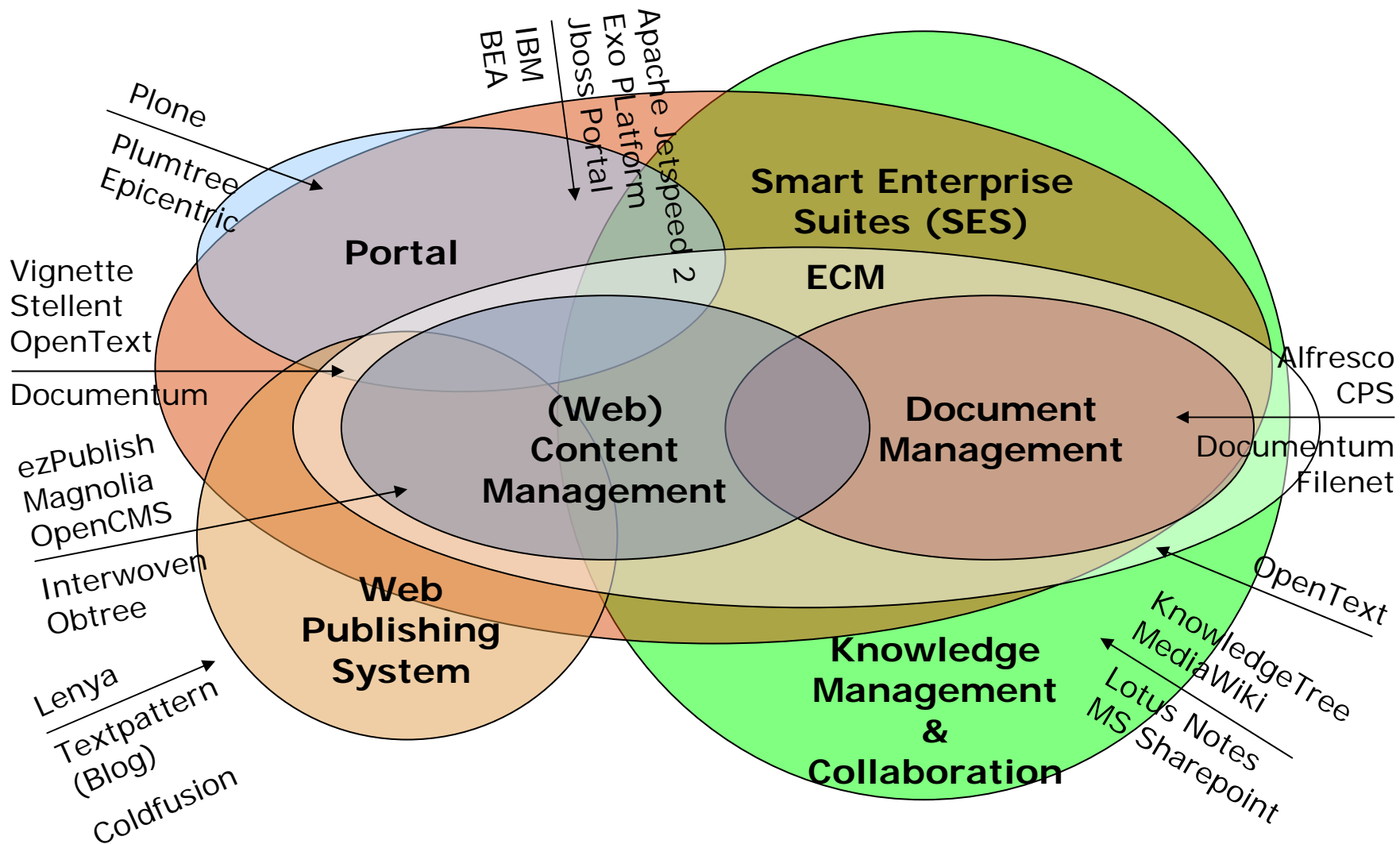
## Enterprise Content Management – a definition

---

“Enterprise Content Management is the technologies used to **Capture, Manage, Store, Preserve, and Deliver content and documents** related to organizational processes”

(aiim org, 2005)

# Enterprise Content Management ...and its relatives



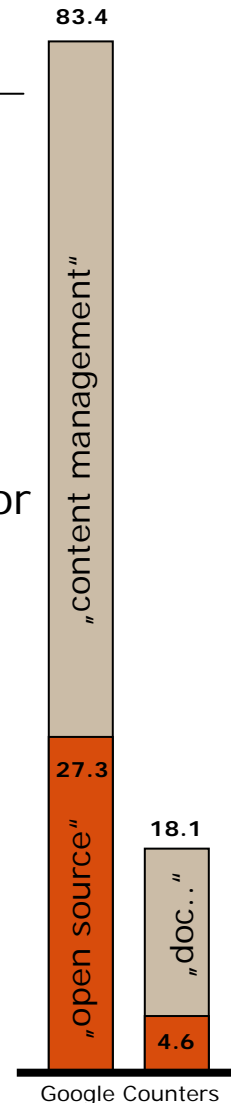
# Content vs. Document Management

## Content Management

- Deals with **content components**
- Are concerned with **creating** content as well as **managing** it
- Create **publications** from components under control or even beyond own control
- Focus is on **delivery** and look and feel
- Handles usually content in HTML or plain text with some meta data
- Often contains mix of "what" and "how"

## Document Management

- Deals with **documents** (files)
- **Manage** files that other applications created
- **Access** the file in a controlled way
- Focus is on **storage** and **management**
- Handles all kinds of file types, but mainly PDF, MS Office and partially OpenOffice
- Strict separation of meta data and file content



Google Counters

# The Document Management landscape

ACOM Solutions - EZeDocs  
 Active Innovations - infoRouter  
 activePDF - activePDF Server  
 Ademero - W...  
 Altimate Sys...  
 Amplefile - D...  
 Anuva - Anu...  
 Asset Mana...  
 Astoria Soft...  
 Auersoft - V...  
 Automated V...  
 Bf.Collabora...  
 Blue Angel T...  
 Bottomline T...  
 Bridgeline S...  
 Captaris - A...  
 CCH - ProS...  
 CEO Image...  
 Cerenade - ...  
 CHL Softwa...  
 Cincom Sys...  
 Citadon - C...  
 ClearVision...  
 ColumbiaSo...  
 Core Imagin...  
 Currit Softw...  
 Cypress - C...  
 Data Reduc...  
 DATAssimila...  
 DBSentry - S...  
 Dieselpoint...  
 DMAC - Uni...  
 Document C...  
 DocumenTr...  
 DocuSource...  
 Docutron Sy...  
 DocuXplore...  
 Doxess - D...  
 DynaPortal...  
 ebrary - Isaac  
 eFinitive Systems - KwikTag  
 EFS Technology - LaserNet  
 Ektron - Ektron DMS400  
 eMASON - IWorkS Image Workflow System  
 EMC Documentum

Esker - Esker DeliveryWare  
 Exact Software - Macola  
 Exegenix - Exegenix Conversion Solutions

Omnivore - Omnivore  
 Open Text - Livelink Enterprise Suite  
 Optical Image Technology - DocFinity

- Hundreds of commercial Document Management Systems
- Existing since the late 80ies
- Archiving was a major focus initially
- Records Management, Digital Asset Management, Web Content Management, etc. was added later
- Few large players (EMC Documentum, Filenet, InfoText, IBM, etc.)
- Many many midrange players, often country or language specific
- Continuous consolidation since the late nineties










Only few (8-15)  
 Open Source  
 Document  
 Management  
 Systems

...ager  
 ...n  
 ...CMS  
 ...cedure Manager  
 ...ent4000  
 ...EB  
 ...Management Tool  
 ...SD Folders  
 ...agement  
 ...ent Presentment  
 ...Synergy eForms  
 ...n  
 ...ment Management  
 ...AGE  
 ...Images  
 ...File  
 WebAsyst - Document Depot  
 WindFire Technology - Xtorm  
 Workshare - Workshare Professional 4  
 WorldView - Document Management Services  
 Xpert Technologies - XPERT Knowledge  
 Xpress Data Systems - EDMXpress

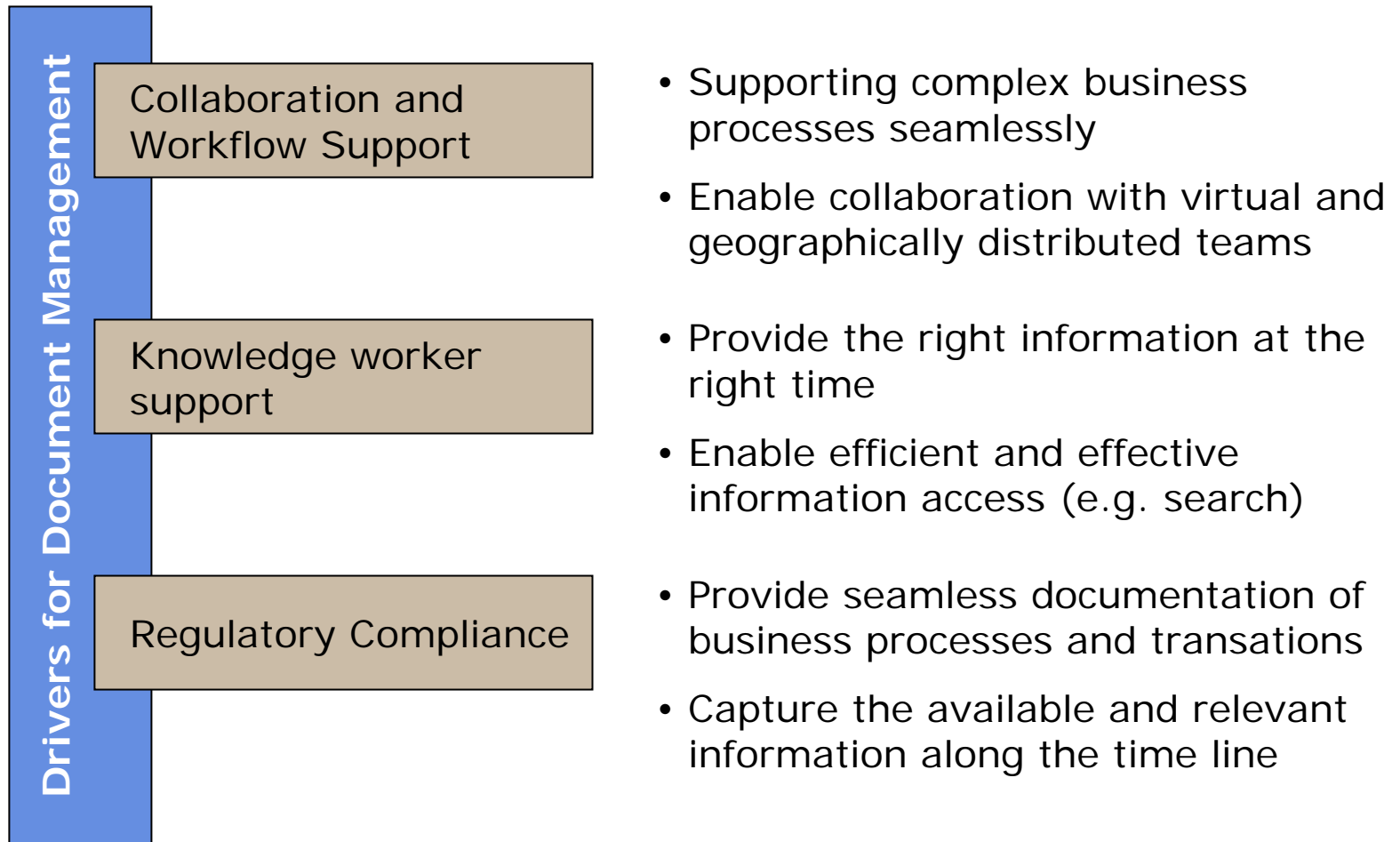
# So what are the chances for success of Open Source Content/Document Management?

OSS CMS

OSS DMS

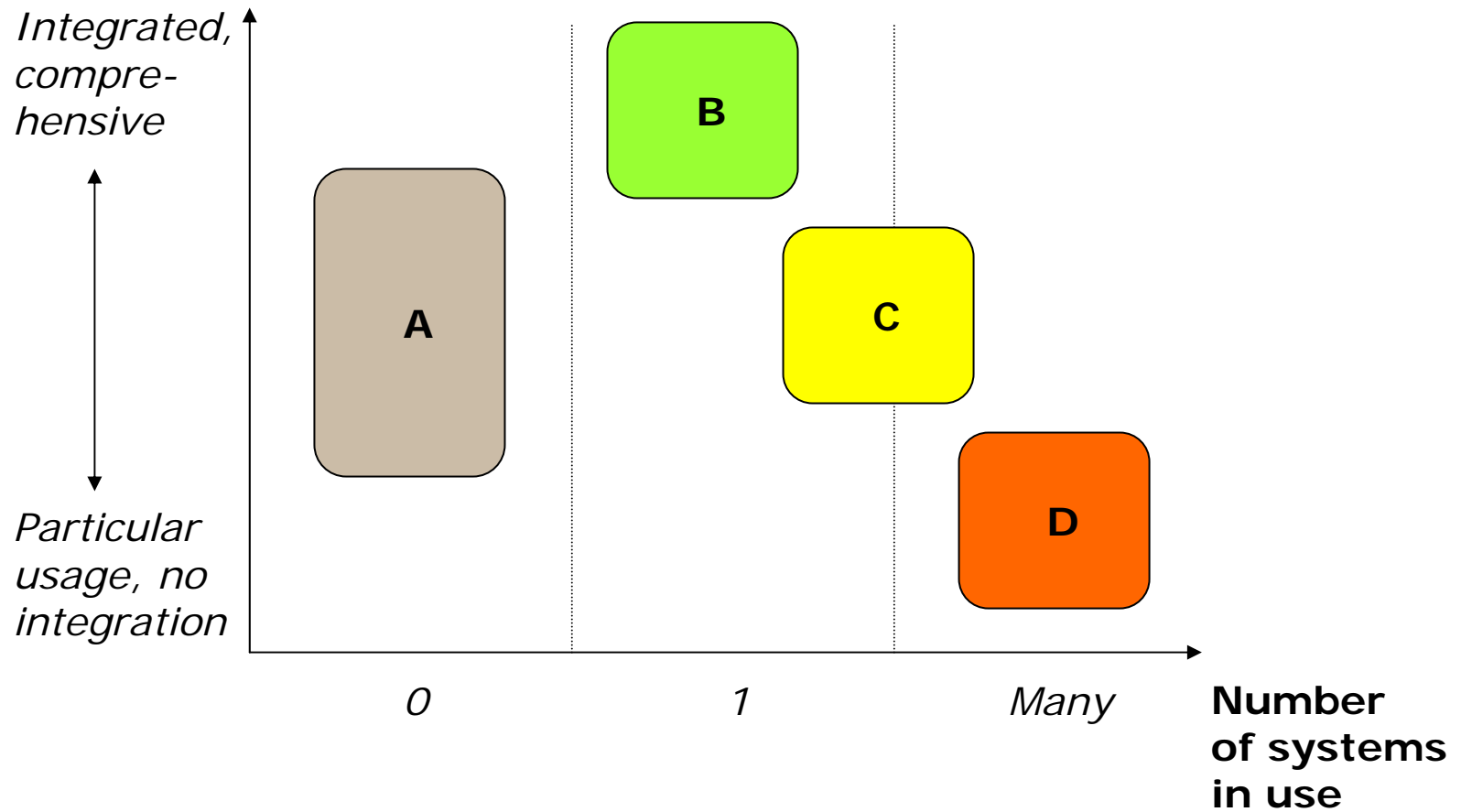
- ◆ The “**closer**” the “problem” to the typical issues of a **developer** is (or the more users and developers are the same) , the better  the chances for Open Source projects
- ◆ The bigger the **number of potential users** the more traction Open Source projects get  
- ◆ **Unsatisfying economical frame conditions** in a specific solution area (e.g. expensive commercial software, lack of language versions, etc.) support the development of Open Source alternatives  
- ◆ The existence of a product, that many/most people **hate** due to various reasons (e.g. monopolistic behavior of the vendor)  supports  the fast development and the rapid adoption of Open Source alternatives
- ◆ **Clearly visible gaps** in the existing Open Source solution portfolio establish the ground for commercial Open Source solutions 

# Today's drivers for Document Management



# State of Document Management: Typical scenarios we see in larger companies

## Usage and integration



## Short history of Document Management

---

- ◆ Started with **storage of images of paper documents**
- ◆ 80ies optic filing created **the discipline of "imaging"**
- ◆ 90ies turned imaging into **workflow** by routing and distributing uncoded information
- ◆ **Groupware Tools** help to compensate for OSes lack of file handling capabilities
- ◆ Not before Win95: **Long (descriptive) file names**
- ◆ **Meta data** become more important
- ◆ **Office integration**
- ◆ Integration into **enterprise applications**
- ◆ SOA, Web clients

# Document Management – standard features

## Create and Store

- Integration with standard authoring tools like MS Office
- Import doc through upload, WebDAV, CIFS
- Meta data

## Organize

- Version control
- Lifecycle Management
- Workflow
- Document references and smart folders

## Collaborate

- Role- and Access based security
- Notification

## Search & Access

- Attribute based and full-text
- Browse Folder/Tree Navigation
- Bookmarks and Links

## Transform

- Output channel or device specific formats like PDF or HTML

# Agenda

---

- ◆ About Optaros
- ◆ Open Source
- ◆ Introduction to Document Management
- ◆ **Open Source Value Add for Document Management**
- ◆ Looking into leading Open Source Document Management Solutions (market view)
- ◆ How to evaluate and implement Open Source Document Management Solutions
- ◆ Alfresco Demo

# Open Source and Document Management – the value add

## Benefits of using Open Source for DMS

Reduce total cost of ownership (TCO) of DMS

Avoid vendor dependencies

Investment protection due to Open Standards compliance

Leverage proven Open Source infrastructure layer

Increase flexibility for implementation/development

- No or low license costs
- Low support and maintenance costs
- Full access to source code
- Power of the communities
- Adherence to open standards
- Interoperability a given
- Strong support for Open Standards
- Reuse of other proven Open Source components reduces risk and cost
- Openness and component orientation allows for more flexible implementations

# An important influencer - Industry Standards and Open Standards

<b>Standard</b>	<b>Description</b>	<b>Standard type</b>
<b>MS Office</b>	Proprietary (native) format in which MS Office documents are stored	Industry
<b>PDF</b>	Proprietary file format for representing documents in a device independent/ resolution independent format	Industry
<b>CIFS</b>	Proprietary approach to access files (CIFS=Common Internet File System)	Industry
<b>XML</b>	Extensible Markup Language is a general-purpose markup language for creating special-purpose markup languages	Open
<b>Docbook</b>	XML vocabulary to create documents in a presentation-neutral form capturing the logical structure of the content	Open
<b>ODF</b>	Open Document Format for saving and exchanging editable office documents	Open
<b>WebDav</b>	Web-based Distributed Authoring and Versioning; also HTTP extension	Open
<b>JSR-170</b>	Java Specification Request to define a standard API to access content repositories independently of its implementation	Open

# Agenda

---

- ◆ About Optaros
- ◆ Open Source
- ◆ Introduction to Document Management
- ◆ Open Source Value Add for Document Management
- ◆ **Looking into leading Open Source Document Management Solutions (market view)**
- ◆ How to evaluate and implement Open Source Document Management Solutions
- ◆ Alfresco Demo

# Open Source Document Management – some of the players

	Version	Architecture	License
<b>Alfresco</b>	1.3	Java	Commercial/MPL
<b>Contineo</b>	2.1 beta	Java	GPL
<b>CPS</b>	3.4	Zope	GPL
<b>KnowledgeTree</b>	3.0	PHP	GPL
<b>Owl</b>	0.8	PHP	GPL
<b>Plone</b>	2.1.1	Zope	GPL
<b>Open sTeam</b>	1.6.0	Pike/MUD	GPL



OWL





# Alfresco

## Architecture

- Distributed Java, Spring and Hibernate
- JSR 168 and JSR 170 support
- Webservices, LDAP and MS Active Directory
- Aspect oriented

## Creating

- CIFS/WebDAV for intuitive file management
- Meta data automation and management

## Collaboration

- Forums for Discussions
- Preconfigured Workspaces, Dashboards
- Transformers

## Search & Access

- Access control on file and folder level
- Lucene indexing (OpenOffice.org) and search engine with taxonomy support

## Community

- Young project
- Under control of Alfresco software



# Contineo

## Architecture

- Java, Struts, Torque
- Apache Velocity Templating Engine
- Open to multiple DB

## Creating

- Files or archives upload through web client
- Key word extraction for meta data

## Collaboration

- Keyword browsing
- Similar documents
- Download tickets
- Check in/out, Versioning

## Search & Access

- Group based access model
- Lucene
- Framebased UI

## Community

- Very small
- Good support on base technologies

# Collaborative Portal Server (CPS)



## Architecture

- Zope3
- ZODB separating storage and placement
- Event-driven
- CPS frame work

## Creating

- Web interface
- Support of Zope's ExternalEditor
- Meta data and taxonomy

## Collaboration

- Workspaces (ACL)
- Robust versioning system
- Workflow support in stacks
- Subscription

## Search & Access

- Full-text index for common office formats
- Readable URLs

## Community

- Small but helpful, Nuxedo

# KnowledgeTree



## Architecture

- PHP
- MySQL, Filesystem
- WebDAV not free

## Creating

- Files or archives upload through web client
- Document types determine meta data

## Collaboration

- Check in/out, versioning
- Archiving

## Search & Access

- Full-text index for common office formats
- Save searches
- Folder and document type based navigation
- RBAC
- Fast and DCM oriented UI

## Community

- Documentation not up to date
- Just new user forum

# Owl

OWL

## Architecture

- PHP
- MySQL, PostgreSQL or Oracle for meta data
- File system for documents

## Creating

- Files or archives upload through web client
- Meta data history
- Folder hierarchy

## Collaboration

- Commenting
- Monitoring folders, Dashboard on usage
- Version by adding suffixes

## Search & Access

- Email Link or Attachment, Download in bulk
- Search gives ranking, finds related docs
- Full-text search only for Word and PDF

## Community

- Totally community driven
- Little documentation but active forum



# Plone

## Architecture

- Zope CMF
- ZODB
- XML-RPC

## Creating

- Web client, FTP, WebDAV
- ExternalEditor, Plone Desktop (commercial)
- Strong meta data support and validation

## Collaboration

- Versioning not supported
- Folders and smart folders, Links
- Robust workflow from CMS background
- Notes on documents

## Search & Access

- Folder based, readable URLs
- Full-text index common office formats
- SmartSearch (AJAX)

## Community

- Extremely large and prolific inkl books

# Open sTeam

## Architecture

- MUD, Pike
- MySQL
- Java and C++ APIs

## Creating

- Files upload through web client
- Meta data support
- Rooms and Knowledge Areas

## Collaboration

- Virtual rooms as central meeting points
- Notification, Rucksack
- Versioning and check in/out
- Chat and message boards

## Search & Access

- Full-text search
- Configurable links to docs
- Open protocols















## Community

- German universities, 20 diploma theses

# Open Source Document Management – how do they compare?

	Strengths	Weaknesses
<b>Alfresco</b>	<ul style="list-style-type: none"> <li>▪ Strong versioning</li> <li>▪ Standards support</li> <li>▪ Content rules</li> </ul>	<ul style="list-style-type: none"> <li>▪ Documentation</li> </ul>
<b>Contineo</b>	<ul style="list-style-type: none"> <li>▪ Simplicity</li> <li>▪ Familiar Java technologies</li> </ul>	<ul style="list-style-type: none"> <li>▪ Documentation</li> <li>▪ Small community</li> </ul>
<b>CPS</b>	<ul style="list-style-type: none"> <li>▪ Strong versioning</li> <li>▪ Strong workflow</li> </ul>	<ul style="list-style-type: none"> <li>▪ Lack of WebDAV support</li> <li>▪ Closed ZODB repository</li> </ul>
<b>KnowledgeTree</b>	<ul style="list-style-type: none"> <li>▪ Simplicity</li> </ul>	<ul style="list-style-type: none"> <li>▪ Lack of standards support</li> </ul>
<b>Owl</b>	<ul style="list-style-type: none"> <li>▪ Simplicity</li> </ul>	<ul style="list-style-type: none"> <li>▪ Unattractive UI</li> <li>▪ Weak search</li> <li>▪ Lack of standards support</li> </ul>
<b>Plone</b>	<ul style="list-style-type: none"> <li>▪ Nice UI</li> <li>▪ Good search</li> </ul>	<ul style="list-style-type: none"> <li>▪ Poor versioning</li> <li>▪ Closed ZODB</li> </ul>
<b>Open sTeam</b>	<ul style="list-style-type: none"> <li>▪ Strong collaboration support</li> </ul>	<ul style="list-style-type: none"> <li>▪ Closed DB, lack of standards</li> </ul>

# Open Source Document Management – how do they compare?

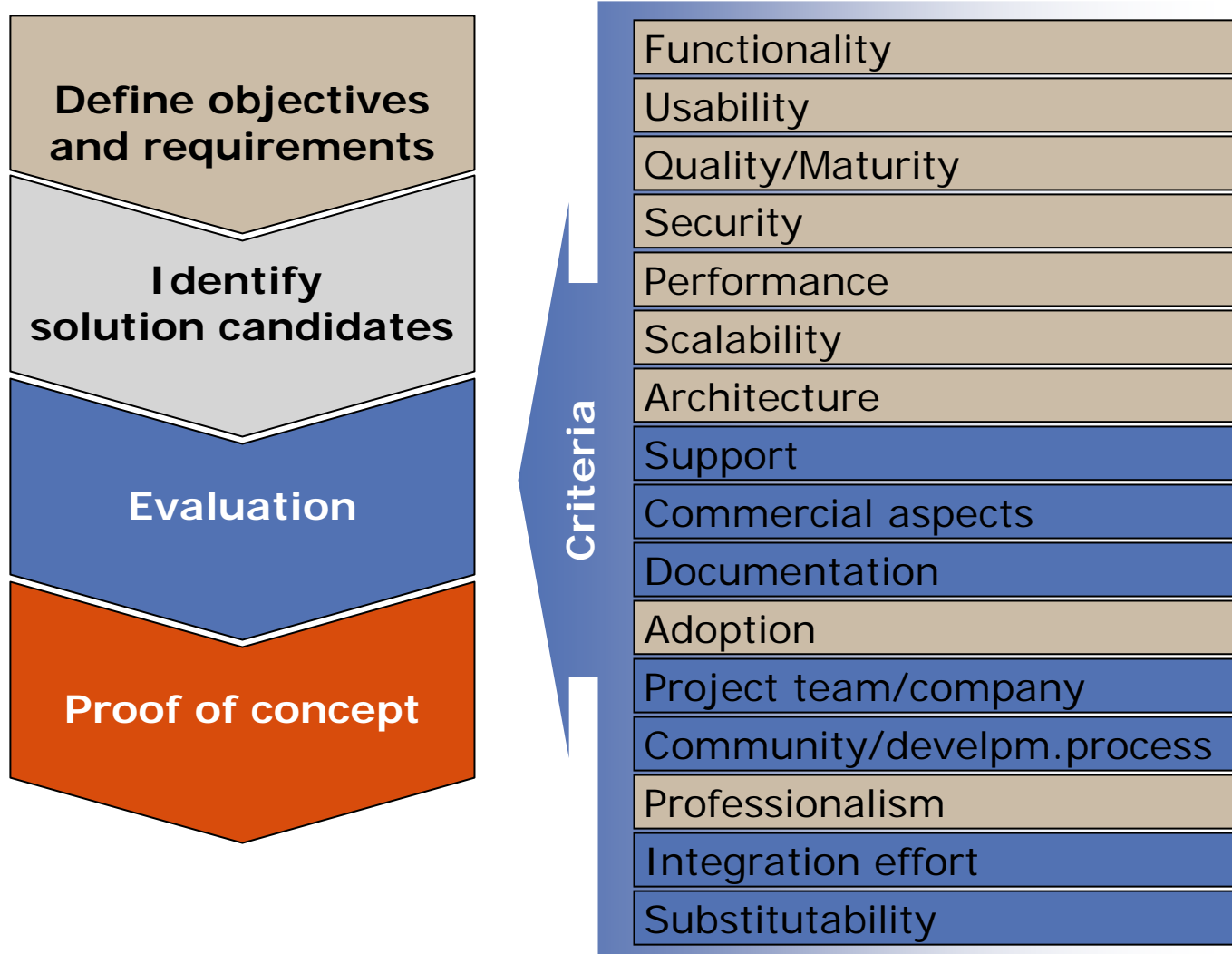
	Standards Support	Community	Resources
<b>Alfresco</b>	<ul style="list-style-type: none"> <li>▪ CIFS</li> <li>▪ WebDAV</li> <li>▪ FTP</li> <li>▪ JCR</li> <li>▪ LDAP</li> </ul>		
<b>Contineo</b>	<ul style="list-style-type: none"> <li>▪ RDBMS</li> </ul>		
<b>CPS</b>	<ul style="list-style-type: none"> <li>▪ LDAP</li> </ul>		
<b>KnowledgeTree</b>	<ul style="list-style-type: none"> <li>▪ LDAP</li> </ul>		
<b>Owl</b>	<ul style="list-style-type: none"> <li>▪ RDBMS</li> </ul>		
<b>Plone</b>	<ul style="list-style-type: none"> <li>▪ WebDAV</li> <li>▪ FTP</li> <li>▪ Section 508</li> <li>▪ W3C AA</li> <li>▪ LDAP</li> </ul>		
<b>Open sTeam</b>	<ul style="list-style-type: none"> <li>▪ Java and C++ API</li> </ul>		

# Agenda

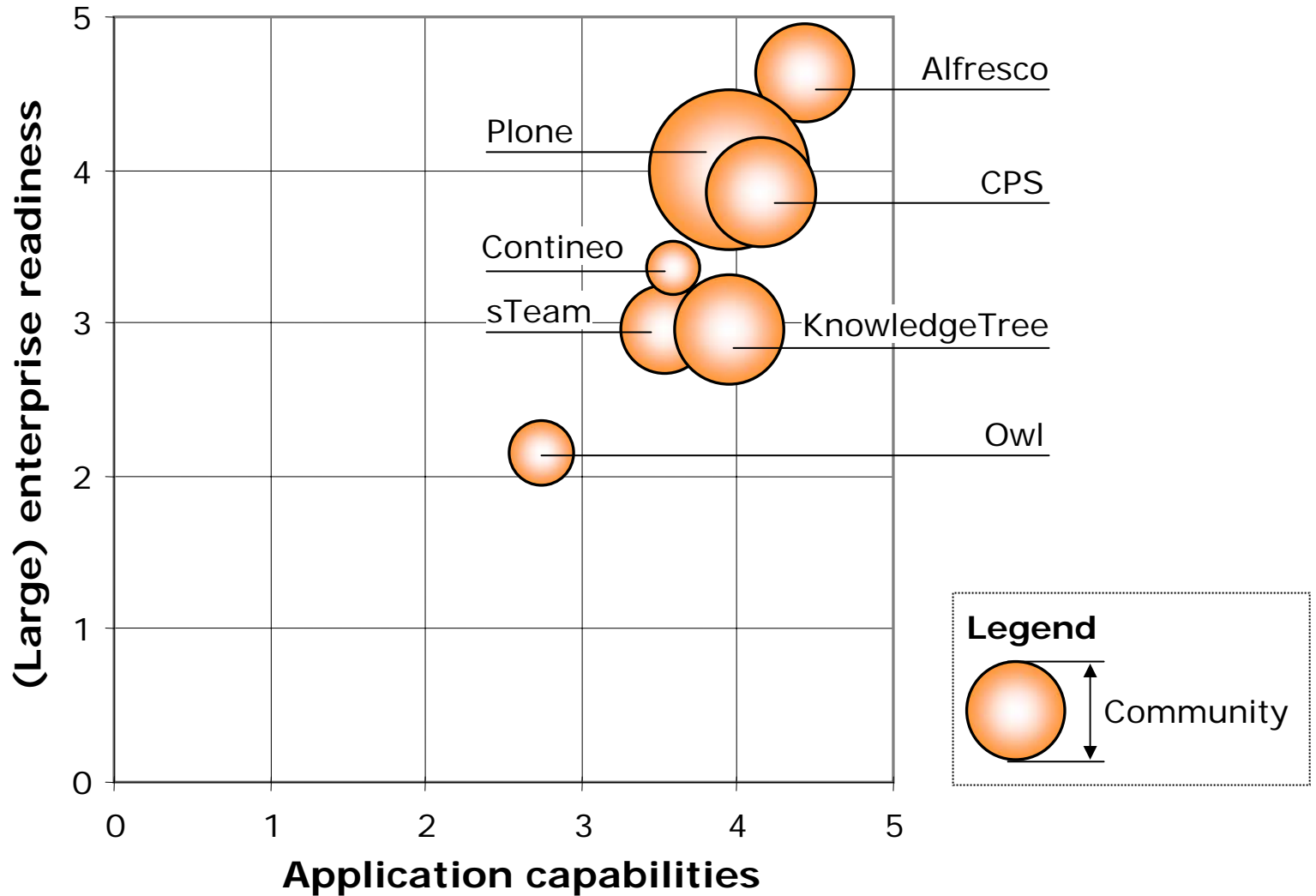
---

- ◆ About Optaros
- ◆ Open Source
- ◆ Introduction to Document Management
- ◆ Open Source Value Add for Document Management
- ◆ Looking into leading Open Source Document Management Solutions (market view)
- ◆ **How to evaluate and implement Open Source Document Management Solutions**
- ◆ Alfresco Demo

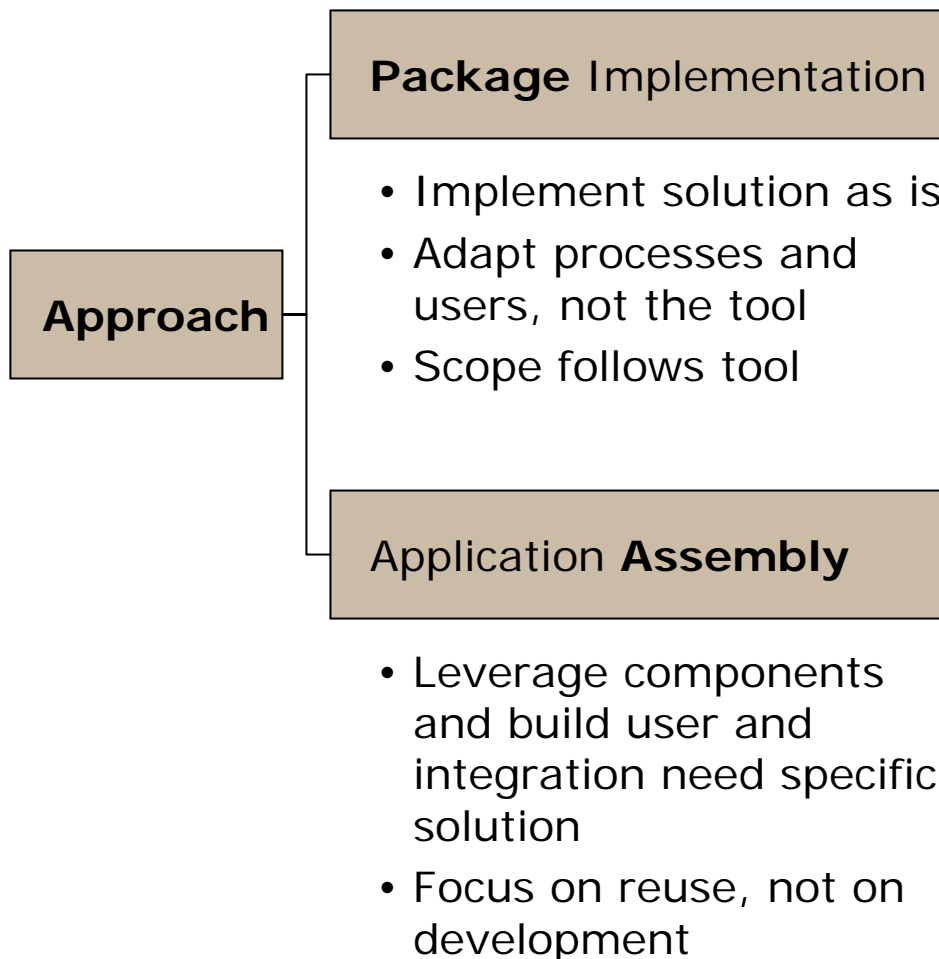
# Evaluating and selecting Open Source Document Management solutions



# Open Source Document Management – what's the right one for you?



# Implementing Open Source Document Management – two approaches



## When to apply it?

- Requirements are „standard“
- Selected solution is very close to the requirements
- Low budget
- Complex integration and/or adaptation needs
- Enduser productivity and application specific process are more important than project cost

# Implementation experience matters – what Optaros learned in recent projects

---

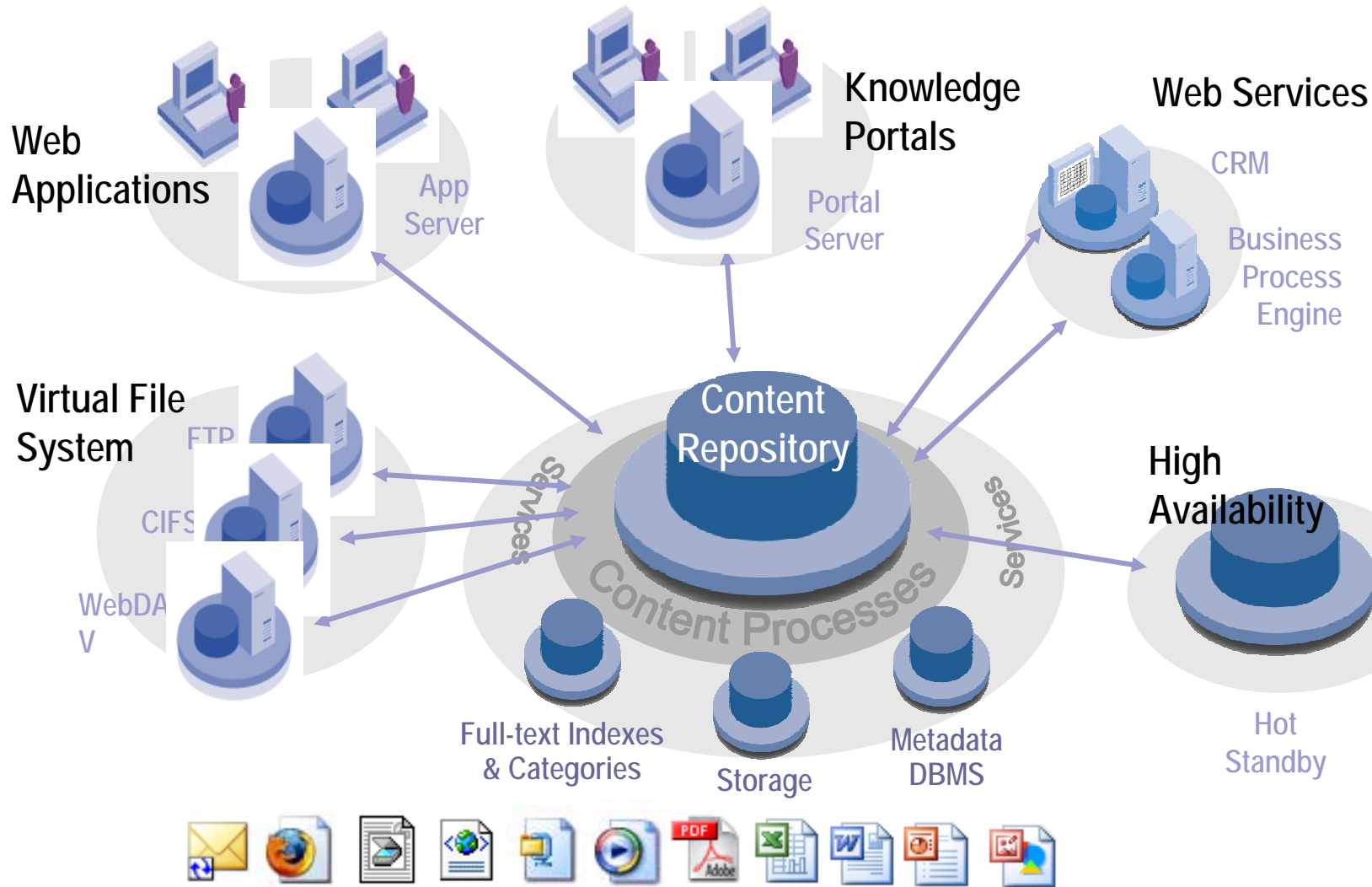
- ◆ Always start with a **proof of concept** when picking a new technology
- ◆ **Understand how the technology works** before making scoping and adaptation decisions
- ◆ Leverage wealth of **open source component repository** when looking for additional tools outside the document management solution scope
- ◆ Interact with the **community** and the **Open Source project team** before making key decisions with architecture and customization impact
- ◆ Apply an **iterative methodology** to be able to show the end user what he will get early on

# Agenda

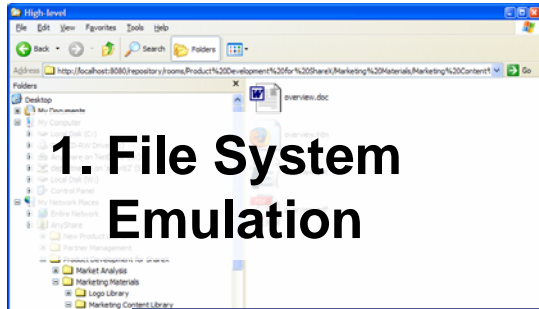
---

- ◆ About Optaros
- ◆ Open Source
- ◆ Introduction to Document Management
- ◆ Open Source Value Add for Document Management
- ◆ Looking into leading Open Source Document Management Solutions (market view)
- ◆ How to evaluate and implement Open Source Document Management Solutions
- ◆ **Alfresco Demo**

# Alfresco - Overview

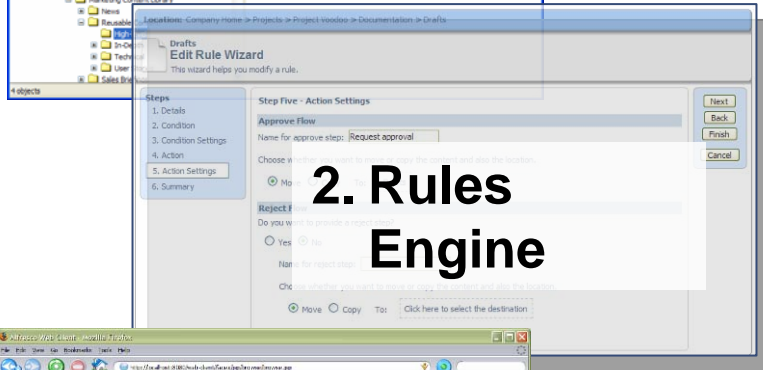


# Users Today – what they understand ...



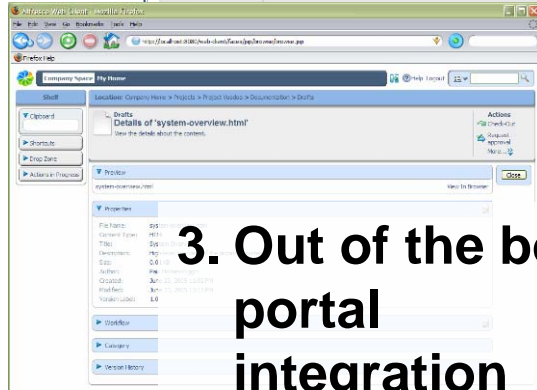
**1. File System Emulation**

◆ Shared file drives



**2. Rules Engine**

◆ Email – Rules



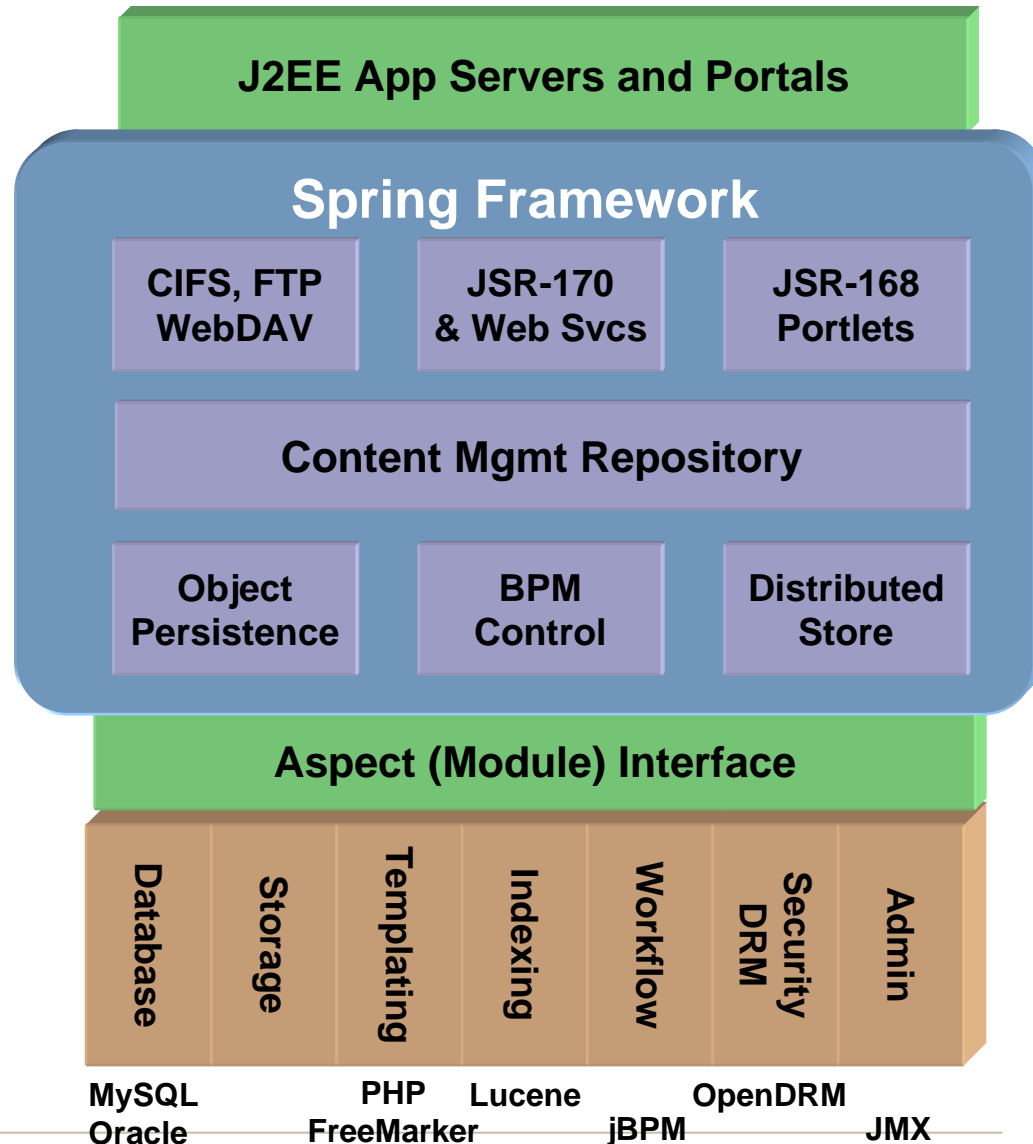
**3. Out of the box portal integration**

◆ Portals

- Google-like search
- Yahoo-like browsing

# Alfresco – first new ECM architecture since 6 years

- ◆ **Infrastructure Components**
  - Spring, Hibernate, Lucene, MyFaces, Aspect-Oriented Programming
- ◆ **Latest Standards**
  - JSF, JSR168, JSR170, JSR 283, iECM, Web Services, BPEL
- ◆ **Modular, light-weight architecture**
- ◆ **Enterprise-scale**
  - 5X faster than commercial
- ◆ **Distributed architecture**
- ◆ **High Availability**

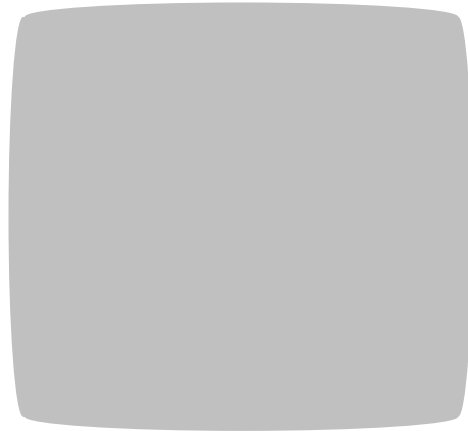
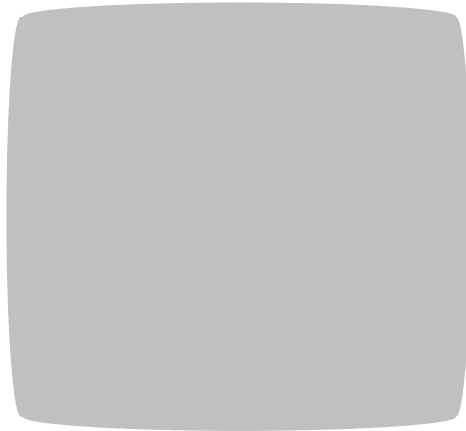


## Alfresco - what is "Better"?

- ◆ Developer Productivity
  - Aspect-Oriented Programming Wizards
  - Templating
  - Web Services
  - Java
  - 1 Model Windows, Linux, Mac Server and Client
- ◆ End User Productivity
  - Ease-of-Use Native Application
  - Intelligent Virtual File System
- ◆ Administrator Productivity
  - Simple Install
  - Zero Footprint
  - Single Sign-On
- ◆ CIO
  - Dramatically Lower Cost
  - No Proprietary Tie-In – JSR-168, JSR-170
  - Complements Enterprise Architecture
  - High Availability
- ◆ Compliance Officer
  - Secure, Auditable Document Lifecycle Management

# Alfresco - Demo





## Contact

---

- ◆ **Sebastian Wohrapp**  
[swohrapp@optaros.com](mailto:swohrapp@optaros.com)

Optaros (Schweiz) AG  
Weinbergstrasse 147  
8006 Zürich  
+41 (44) 3 62 11 11

- ◆ **München**  
Theresienstrasse 6-8  
D-80333 München  
+49 (89) 28 89 03 79
- ◆ **Genf**  
Geneva Business Center - 1  
12, Avenue des Morgines  
CH-1213 Petit-Lancy/Genève  
+41 (0) 22 731 84 20
- ◆ **Boston**  
60 Canal Street, 4th Floor  
Boston, MA 02114  
+1 (617) 227-1855

<http://www.optaros.com/de>  
<http://optaros-update.blogspot.com>