

# Breaking down content silo's with CMIS

Bas Peters

Open Source Conference 2011

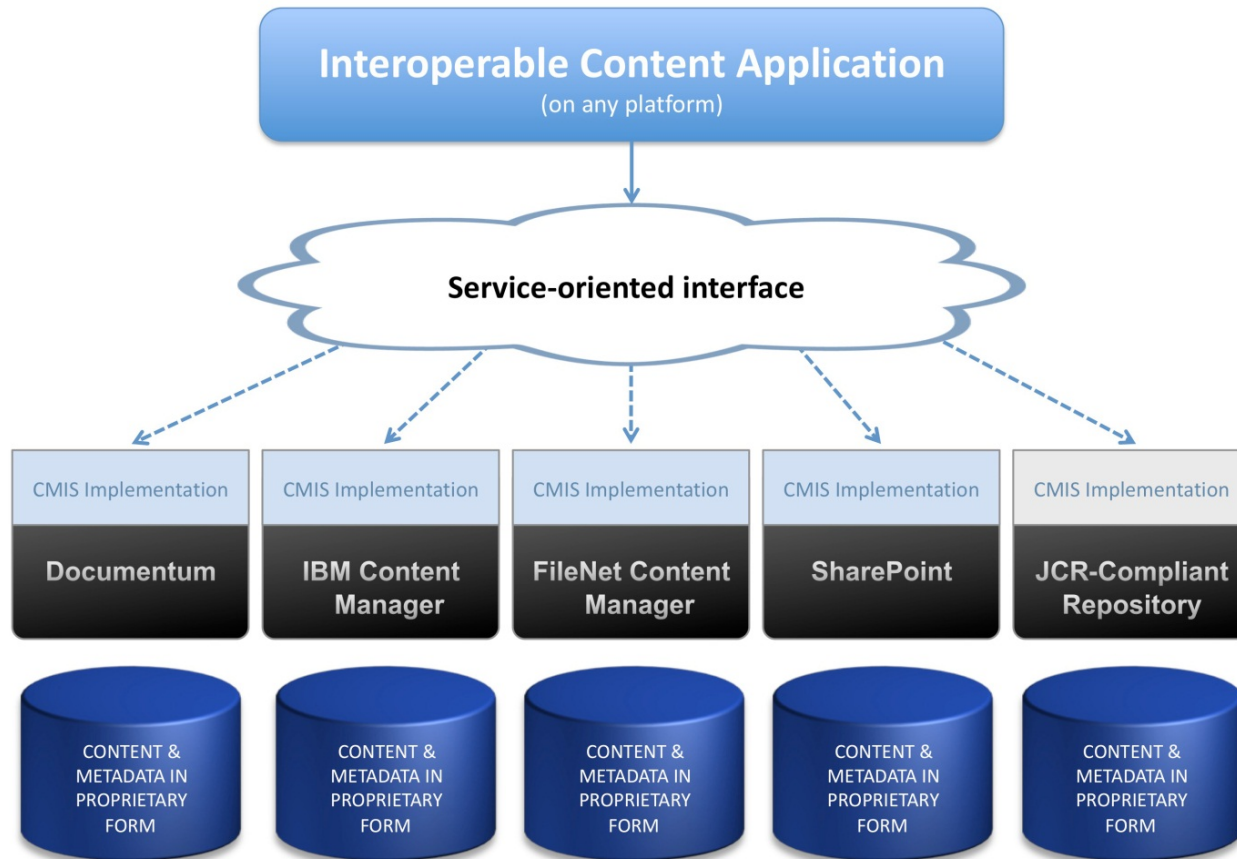


Breaking down content silo's with CMIS

# INTRODUCTION



# Improve ECM interoperability



# Some say that...

**“CMIS is as important to content management applications as the standardization of SQL was to database applications in the late 1980’s and early 1990’s”**



# Content as a service

## CMS Implementation

Document  
Management



Collaboration &  
Social Media



Web Content  
Management



Records  
Management



Email  
Archive

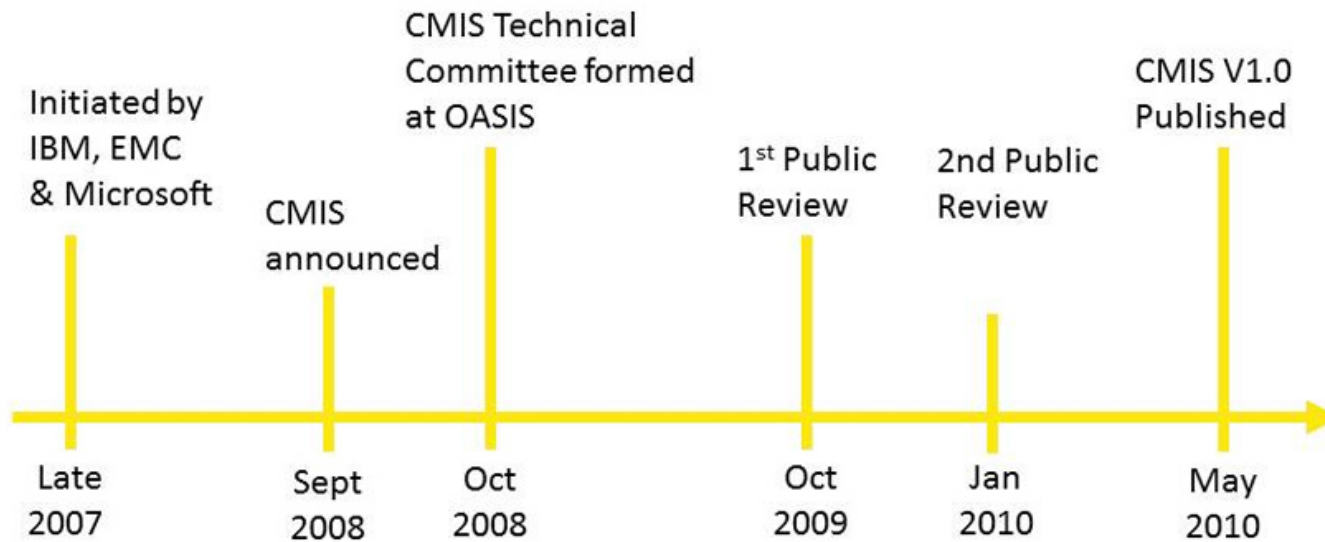


# OASIS

- **OASIS was founded under the name "SGML Open" in 1993**
- **Organization for the Advancement of Structured Information Standards**
- **Not-for-profit consortium that drives the development, convergence and adoption of open standards for the global information society**
- **Standards:**
  - Darwin Information Typing Architecture (DITA)
  - DocBook v5.0
  - ebXML Business Process
  - OpenDocument Format for Office Applications

# CMIS Timeline

Committee members included Microsoft, EMC, IBM, SAP, Alfresco, Oracle, Day and OpenText



<http://www.slideshare.net/alfresco/cm-is-spec-uncovered>

# Vendor Support



**Microsoft**<sup>®</sup>

**EMC**<sup>2</sup>  
where information lives

**ORACLE**

**OPEN TEXT**  
The Content Experts™



**nuxeo**  
Open Source ECM



Breaking down content silo's with CMIS

# USE CASES

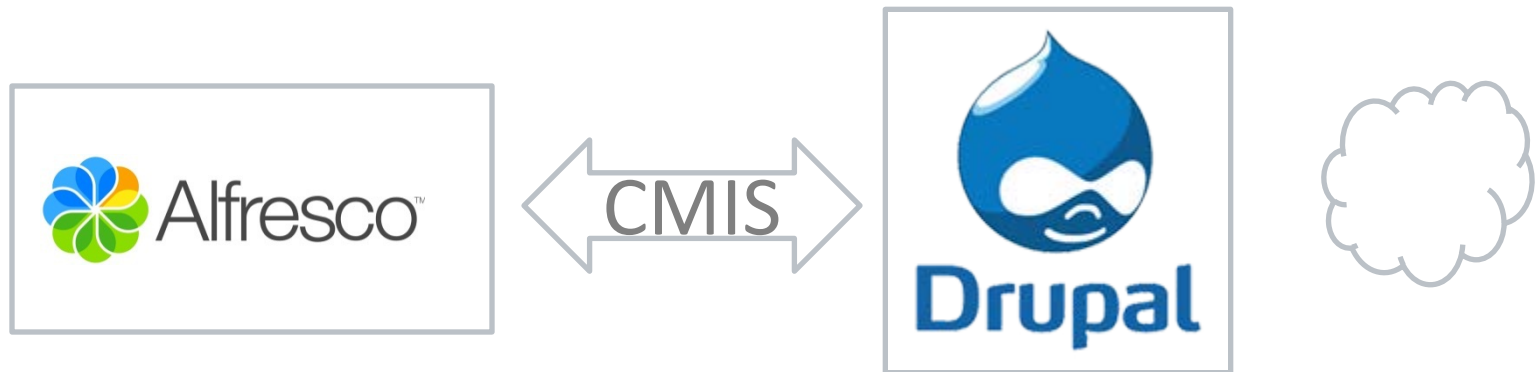


# Use Cases

- **Laurence Hart compiled a list of CMIS use cases:**
- Repository-to-Repository (R2R)
  - Application-to-Repository (A2R)
  - Federated Repositories

# Repository-to-Repository (R2R)

- Managing Records centrally that are stored in other repositories
- Publishing content from one repository to another



# Application-to-Repository (A2R)

- SharePoint or Alfresco Share as a front-end
- Enterprise Software Applications (BPM, CRM, Capture)
- Content Enabled Vertical Applications (CEVA)
- Productivity applications (Office integration)

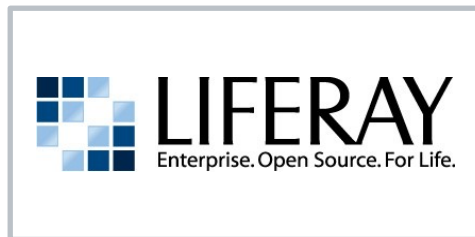


# Federated Repositories

## → Federated Search

- CMIS Client searches multiple repositories
- CMIS Client crawls multiple repositories

## → Federation instead of migration



Breaking down content silo's with CMIS

# THE STANDARD



# Parts

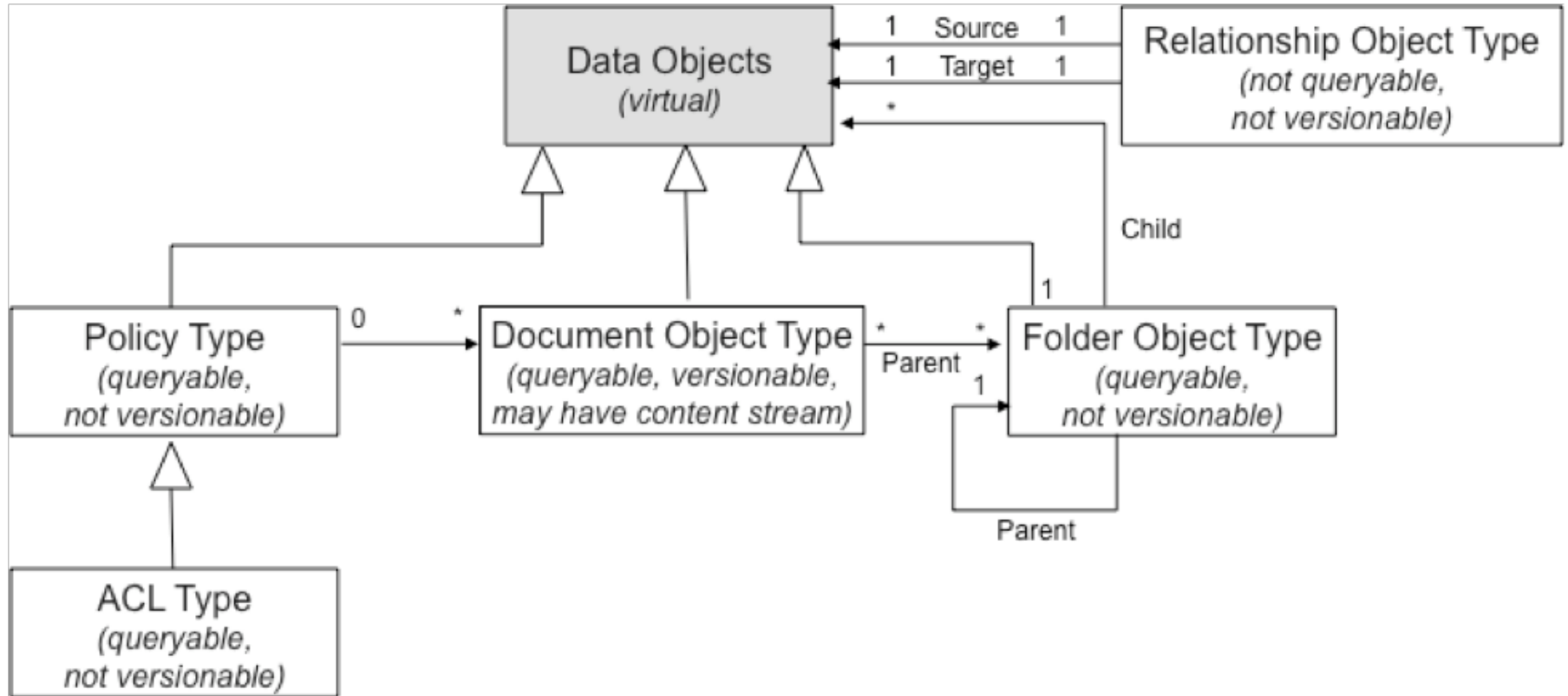
- **Domain model**
- **Services**
- **Query language**
- **Protocol bindings**
  - Web Services Binding
  - REST Atom Binding

Breaking down content silo's with CMIS

# DOMAIN MODEL



# Domain Model



# Base Object Type properties

Identifiers
objectId typeId baseTypeId name

Version
isLatestVersion isMajorVersion versionLabel checkinComment

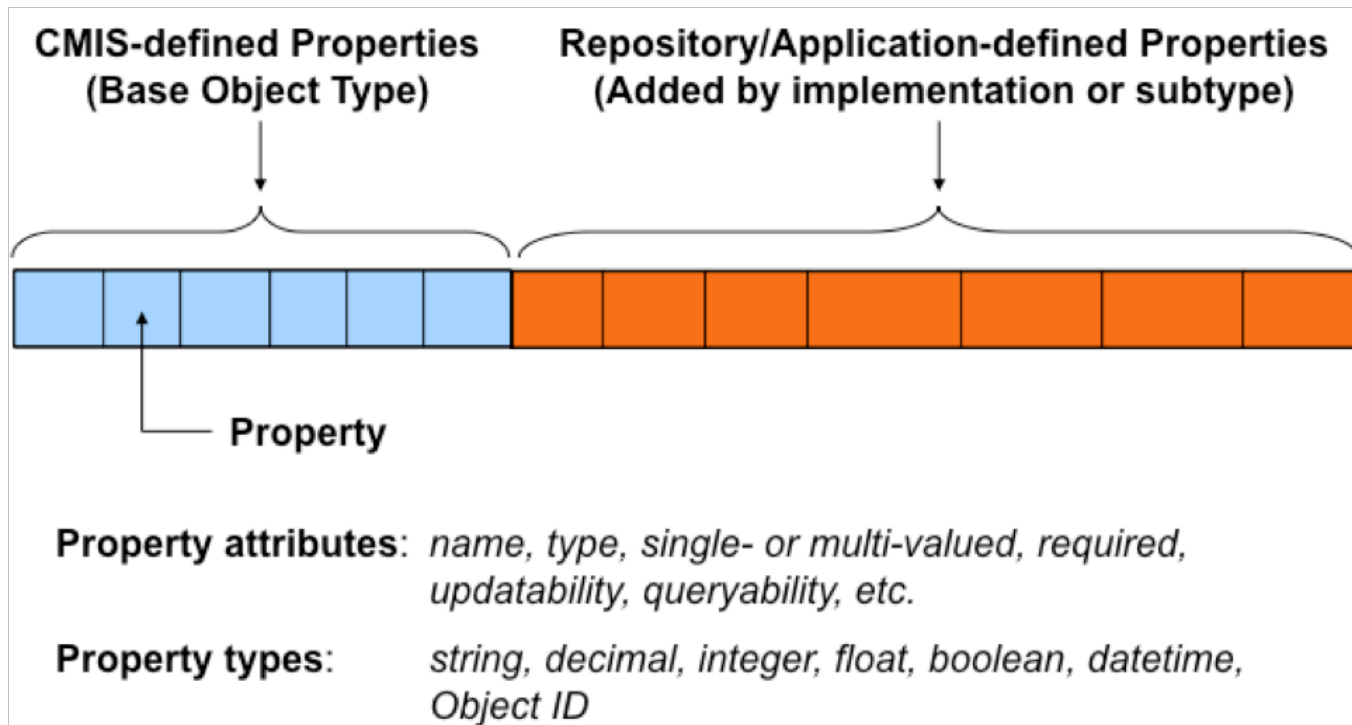
Content
contentStreamLength contentStreamMimeType

Hierarchy
parentId path allowedChildObjectTypes

Audit
createdBy creationDate lastModifiedBy lastModificationDate

# Custom Types

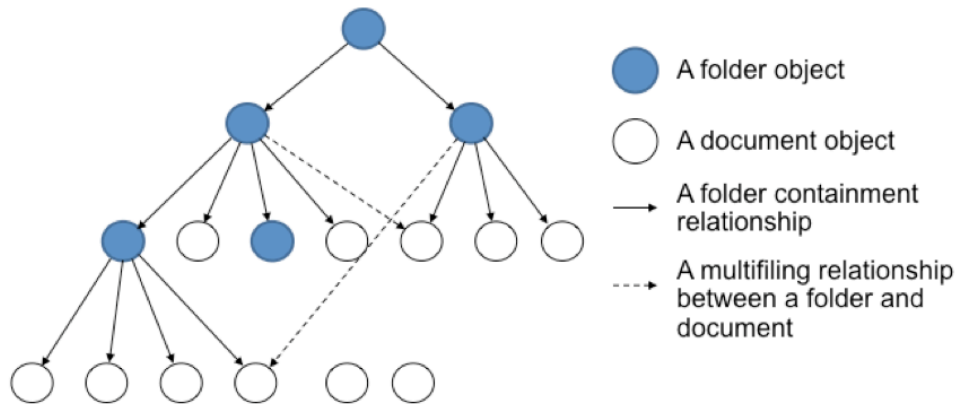
Custom types can extend the CMIS types



# Folder Hierarchy

→ Document objects live in a folder hierarchy

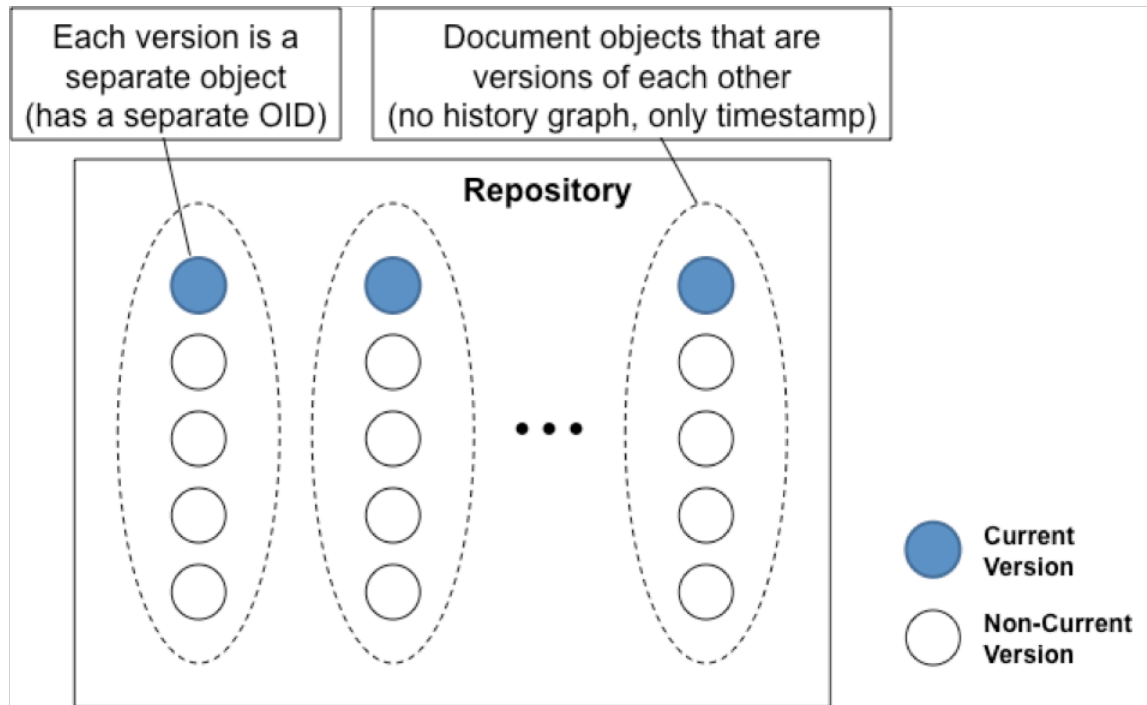
## Folders Structure



Copyright ©2011 Alfresco Software Inc.

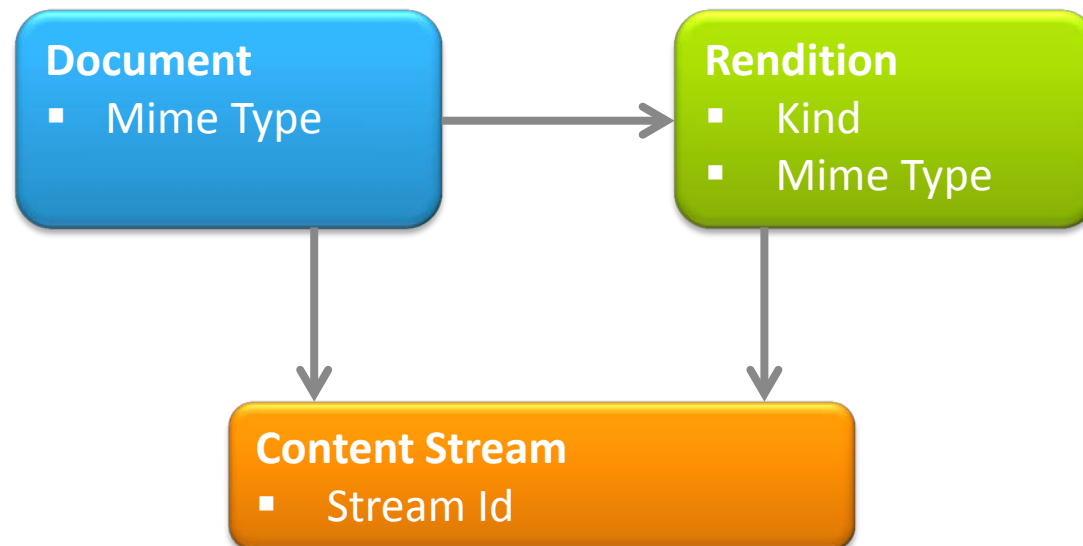
# Versions

→ **Document Objects are the only object type that can be versioned**



# Content Access

→ Document Objects provide access to content streams and enable retrieval of renditions



# Policy Object

- **A policy object represents an administrative policy that can be enforced by a repository**
  - ACL (Access Control Lists)
  - Retention policies
  - Aspects in Alfresco

Breaking down content silo's with CMIS

# CMIS SERVICES



# CMIS Services

## Repository Services

getRepositories  
getRepositoryInfo  
getTypeChildren  
getTypeDescendants  
getTypeDefinition

## Navigation Services

getFolderTree  
getDescendants  
getChildren  
getFolderParent  
getObjectParent  
getCheckedOutDocs

## Discovery Services

query  
getContentChanges

## Object Services

createDocument  
createFolder  
createRelationship  
createPolicy  
getAllowableActions  
getObject  
getProperties  
getFolderNypath  
getContentStream  
getRenditions  
updateProperties  
moveObject  
deleteObject  
deleteTree  
setContentStream  
deleteContentStream

## ACL Services

getACL  
applyACL

## Versioning Services

checkout  
cancelCheckout  
checkin  
getPropertiesOfLatestVersion  
getAllVersions

## Multi-filing Services

addObjectToFolder  
removeObjectFromFolder

## Policy Services

applyPolicy  
removePolicy  
getAppliedPolicies

## Relationship Services

getObjectRelationships

# Create Folder

http://localhost:8080/alfresco/s/cmisp/children

Content-Type: application/atom+xml

```
1  <?xml version="1.0" encoding="utf-8"?>
2  <entry xmlns="http://www.w3.org/2005/Atom"
3  xmlns:cmisra="http://docs.oasis-open.org/ns/cmisp/restatom/200908/"
4  xmlns:cmis="http://docs.oasis-open.org/ns/cmisp/core/200908/"
5  <title>XML Amsterdam 2011</title>
6  <cmisra:object>
7  <cmis:properties>
8  <cmis:propertyId
9  propertyDefinitionId="cmis:objectTypeId">
10 <cmis:value>cmis:folder</cmis:value>
11 </cmis:propertyId>
12 </cmis:properties>
13 </cmisra:object>
14 </entry>
```

Breaking down content silo's with CMIS

# QUERY LANGUAGE



# Query Language

## → Based on SQL 92, with ECM extensions and support for metadata and full text

- SELECT with a target list
- FROM with the object types being queried
- JOIN to perform a join between object types
- WHERE with the predicate
- IN and ANY to query multi-value properties
- CONTAINS to specify a full-text qualification
- IN\_FOLDER to search children within a folder
- IN\_TREE to search within a folder hierarchy
- ORDER BY to sort the results

# Query

http://localhost:8080/alfresco/s/cmis/queries

Content-Type: application/cmisquery+xml

```
1 <cmis:query xmlns:cmis="http://docs.oasisopen.org/ns/cmis/core/200908/">
2   <cmis:statement>
3     <![CDATA[SELECT * FROM cmis:document WHERE CONTAINS('sample') ]]>
4   </cmis:statement>
5   <cmis:skipCount>0</cmis:skipCount>
6   <cmis:maxItems>5</cmis:maxItems>
7 </cmis:query>
```

# Non-strict Queries

## → Retrieve aspects in Alfresco

```
SELECT d.*, t.*, a.* FROM cmis:document AS d
JOIN cm:titled AS t ON d.cmis:objectId = t.cmis:objectId
JOIN cm:author AS a ON d.cmis:objectId = a.cmis:objectId
WHERE IN_FOLDER(d, 40312a4b-7767-4586-a58b-18d050ffe0ca')
AND CONTAINS(d,'sample')
```

Breaking down content silo's with CMIS

# PROTOCOL BINDINGS



# Requirements

## → Neutral

- Any Programming Language
- Any Operating System
- Any Environment

## → Required

- Web Service Client Library
- HTTP Client Library

There are client libraries available in all general languages including Java, PHP, .NET, JavaScript and Python

# Web Services Binding

- **Service-oriented binding**
- **CMIS-specified XML schema**
- **Capable for system-to-system access**
- **Use MTOM to transfer content stream**
- **WS-Security 1.1, Username Token Profile 1.1**
- **Transactional Applications**

# REST Atom Binding

- Resource-based binding
- Atom Pub format with CMIS extensions
- Browse-able by standard Atom feed readers
- Suitable for ReSTful-style application-to-system access
- HTTP authentication
- Web-based Applications

Breaking down content silo's with CMIS

# FUTURE DIRECTIONS



# Future Directions

- **Browser binding**
- **Retention policies**
- **Annotations**



# Browser Binding

- Simplify uploading documents from a browser application (e.g. using multipart form)
- Simplify parsing (e.g. using JSON instead of XML)
- Simplify construction of required object metadata (e.g. constructing links)

# Retention Policies

- The rules for retention policies on invoices are maintained in the ERP system
- When a fax image is attached to the invoice data in the ERP system, the ERP system
  - Creates the appropriate CMIS Policy for the retention if it does not already exist
  - Applies the CMIS Policy for the retention to the CMIS Document with the fax image



# Annotations

- **Ability to create and retrieve annotations associated to documents:**
- To highlight a section of text
  - Pinpoint a specific item in an image
  - Provide a small note or comment during a particular stage within the process of the document's lifecycle

# Conclusion

- **CMIS is a well designed standard**
- **It supports relevant use cases**
- **All big vendors are involved, although open source seems to be leading**
- **It remains to be seen if it will become as big as SQL**

Breaking down content silo's with CMIS

**QUESTIONS?**

