Alfresco as a Metadata hub

Robin Bramley
Chief Scientific Officer, Ixxus
@rbramley
A Global Strategic-level Alfresco partner
Working with Alfresco since Alfresco v0.6
Over 140 Alfresco implementations
Alfresco & WCM Architecture Design Patterns
First worldwide implementation of the first Alfresco WCM product in 2007
Excellent Alfresco knowledge and highly trained and experienced staff
Alfresco Certified Engineers and Alfresco Certified Administrators
Authoring of leading Alfresco books
Multiple speaking engagements at the Alfresco Conferences (2011-2014), Alfresco Days and BeeCon 2016
Alfresco Million $ Club (May 2012) – Only 2 companies awarded
Alfresco Solution Partner of the Year 2013 and 2014
Alfresco Best Solution 2015 and Partner of the Year 2015 (Northern Europe)
Wholly owned subsidiary of Copyright Clearance Center
Metadata
meta-

1. A prefix appearing in loanwords from Greek, with the meanings “after,” “along with,” “beyond,” “among,” “behind,” and productive in English on the Greek model: metacarpus; metagenesis.

2. A prefix added to the name of a subject and designating another subject that analyzes the original one but at a more abstract, higher level: metaphilosophy; metalinguistics.

3. A prefix added to the name of something that consciously references or comments upon its own subject or features: a meta-painting of an artist painting a canvas.

@article {Dictionary.com2017,
  title = {Random House Dictionary},
  month = {Apr},
  day = {10},
  year = {2017},
  url = {http://www.dictionary.com/browse/meta},
}
“If music be the food of love, play on;”

Orsino, Shakespeare's Twelfth Night
Type Model

<<Alfresco>>

cm:content

rjb:BaseDocument

rjb:Article

rjb:Presentation

...
### Model admin

#### Model Manager

**Tools**
- Application
- Category Manager
- Module Browser
- Node Browser
- Search Manager
- Tag Manager
- Model Manager
- Sites Manager
- Repository
- Replication Jobs
- Users and Groups

#### RJB-Knowledge

<table>
<thead>
<tr>
<th>Name</th>
<th>Display Label</th>
<th>Parent</th>
</tr>
</thead>
<tbody>
<tr>
<td>rjb:BaseDocument</td>
<td>Document</td>
<td>cm:content</td>
</tr>
<tr>
<td>rjb:Presentation</td>
<td>Presentation</td>
<td>rjb:BaseDocument</td>
</tr>
<tr>
<td>rjb:Article</td>
<td>Article</td>
<td>rjb:BaseDocument</td>
</tr>
</tbody>
</table>

[Image of Model Manager interface]
Properties

• Properties allow us to capture metadata values
• Numbers, date / datetime, boolean, text & multi-lingual text
• *Mandatory or optional
• Multi-valued
• Constrained
Constraints

- Regex
- min/max length
- min/max value
- pick from a list
- or use a custom Java constraint class
Aspects

<<Alfresco>>
   cm:content

   rjb:BaseDocument
      rjb:Article
      Node
      <<Aspect>>
      Versionable

   acme:Document
      acme:Invoice
      Node
      <<Aspect>>
      Work In Progress

   ...
Associations

Company Home

Sites

Site A

Document Library

File 1

Document Library

Site B

File 2
Advanced search

Look for: Content

Keywords:

Name:

Title:

Description:

Mimetype: Unknown

Modified Date:
From: DD/MM/YYYY
To: DD/MM/YYYY

Modifier:
Configured Advanced Search
Categories

Presentations

Selected Items...

- **Bramley_2013_Scalability_with_notes.pptx** (Beating the benchmarks)
  - Modified 9 hours ago by Robin Bramley
  - 11 MB
  - No Description
  - alfresco, graph, scalability
  - EUROPE, Spain, English
  - Favorite: 0
  - Like: 0
  - Comment: 0
  - Share

- **Bramley_DevOps.pptx** (Applying DevOps to Alfresco)
  - Modified 10 hours ago by Robin Bramley
  - 4 MB
  - No Description
  - alfresco, puppet, vagrant, jenkins
  - EUROPE, Spain, English
  - Favorite: 0
  - Like: 0
  - Comment: 0
  - Share

- **Bramley_MemoryLane_LT.pptx** (Memory Lane ignite talk)
  - Modified 9 hours ago by Robin Bramley
  - 11 MB
  - No Description
  - alfresco
  - EUROPE, Spain, English
  - Favorite: 0
  - Like: 0
  - Comment: 0
  - Share

- **LuceneEurocon2011_Bramley_Solr_at_VMG.pptx** (Solr at VirginMoney Giving)
  - Modified 10 hours ago by Robin Bramley
  - 1 MB
  - Case study presented at Lucene EuroCon 2011 in Barcelona
  - solr, spring, java, case study
  - EUROPE, Spain, English
Friction

https://www.flickr.com/photos/gtweb/11328349694
All In One Edit

“Bring efficiency to content creation and manual enrichment tasks in an editorial workflow by being able to read / edit content and metadata in one single interface.”
“If music be the food of love, play on; Give me excess of it; that surfeiting, The appetite may sicken, and so die.”

Orsino, Shakespeare's Twelfth Night
Automation
Metadata Extraction

1. Apache Tika FTW!
2. XPath extraction from XML files
Inheritance from the parent

E.g. Glean context from the folder
Derived metadata

E.g. File naming conventions

FSA0057-D103-20170421-ARB-minutes.docx
Complex metadata gathering approaches

1. Integration with a Product Information Management system
2. Bulk metadata import (not necessarily automated)
3. Machine Learning
4. Semantic extraction from text
Trends & Standards
HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION:
THERE ARE 14 COMPETING STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES. YEAH!

SOON:

SITUATION:
THERE ARE 15 COMPETING STANDARDS.

https://imgs.xkcd.com/comics/standards.png
Resource Descriptor Framework

- RDF standard from the W3C
- Relies on *subject-predicate-object* triples
- John knows Jane
- DocumentA isA Contract
- Triples form a directed graph which can be navigated
Linked Online Data

Outbound serialization

Document Library

File 1

Category A

Category B

File 2

File 1

File 1

File 1

File 1

File 1

File 1

Type

Title

Category A

Category B

File 2
Complex needs
Where is the tipping point for Share?
Or the repo itself

Be aware of the limitations:

• Flexibility of the content model lagging behind current industry trends
• Internal model is proprietary
• Tricky to model some metadata standards
• Weak associative searches
• Many old file-based protocols don’t support metadata capture
Don’t forget Jeff Potts’ Anti-patterns


- Anti-Pattern #3: Highly relational solutions
- Anti-Pattern #4: JSON/XML object store
- Anti-Pattern #5: Storing lots of content-less objects
Summary
Conclusion

Alfresco Content Services with Share provides a **great foundation** for capturing and storing metadata

- Content modelling capabilities, properties exposed through forms, workflow

Platform is **easy to integrate** for upstream/downstream systems

- Open principles (Open Source, open standards, open architecture)
- Java/Spring underpinnings
- Node crawler (from the Transfer Service)
- Transformation capabilities