Steef de Rooi

From: Ties Jongsma
Sent: 22 February 2018 16:33
To: Steef de Rooi
Subject: FW: IBFD - Alfresco update PoC - TNS properties

From: Kagan Akbas [mailto:kagan.akbas@incentro.com]
Sent: 16 February 2018 14:11
To: Ties Jongsma
Cc: Floor Koole; David Wiering
Subject: Re: IBFD - Alfresco update PoC - TNS properties

Dear Ties & Floor,

I see, apologies for the lack of information about this. Alfresco indeed has a bulk importer tool that is able to handle large amount of files and import them into the repository.

As for your question about how to prepare the content and how the metadata is filled; if we wish to import metadata along with the files, then we need to have "shadow" Java property files in XML format. And these shadow properties must have the same name and extension as the file which it represents, but with the " `.metadata.properties.xml` " suffix.
For example, if there is a file called `IMG_1967.jpg`, the "shadow" metadata file is called `IMG_1967.jpg.metadata.properties.xml`.

The official documentation has a very nicely written explanation for how the metadata files should look like.

Here is an example:

```xml
<?xml version="1.0" encoding="UTF-8"?>

 <!DOCTYPE properties SYSTEM "http://java.sun.com/dtd/properties.dtd">

 <properties>

    <entry key="type">cm:content</entry>

    <entry key="aspects">cm:versionable,cm:dublincore</entry>

    <entry key="cm:title">A photo of a flower.</entry>

    <entry key="cm:description">A photo I took of a flower while walking around Bantry
Bay.</entry>

    <entry key="cm:created">1901-01-01T12:34:56.789+10:00</entry>

    <!-- cm:dublincore properties -->

    <entry key="cm:author">Peter Monks</entry>

    <entry key="cm:publisher">Peter Monks</entry>
```

```
    <entry key="cm:contributor">Peter Monks</entry>

    <entry key="cm:type">Photograph</entry>

    <entry key="cm:identifier">IMG_1967.jpg</entry>

    <entry key="cm:dcsource">Canon Powershot G2</entry>

    <entry key="cm:coverage">Worldwide</entry>

    <entry key="cm:rights">Copyright (c) Peter Monks 2002, All Rights Reserved</entry>

    <entry key="cm:subject">A photo of a flower.</entry>

  </properties>
```

I will be discussing with David on what the best practices are in terms of bulk importing, and what common obstacles have been encountered in the past. So that our importing process can go smoothly.

Please let me know if this gave any insights into your question.

Thanks.

Kind regards,
Kagan Akbas

2018-02-15 13:29 GMT+01:00 Ties Jongsma <T.Jongsma@ibfd.org>:

Beste David & Kagan,

Wij zouden nog informatie van jullie krijgen over hoe we een redelijke hoeveelheid bestaande content in Alfresco krijgen, via een bulk importer of iets dergelijks.

Hoe moeten we deze content klaarzetten, hoe krijgen we de meta data ingevuld, etc?

Graag jullie reactie.

Groet,

Ties

From: Kagan Akbas [mailto:kagan.akbas@incentro.com]
Sent: 05 February 2018 20:47
To: Floor Koole

Dear Floor, Ties,

I appreciate the detailed list of properties you've provided.

In terms of the *workflow* related properties, some of them are actually "native" properties built-in to the Alfresco system, such as "version number".
So the version number would be consistently incremented when a new file is uploaded, and this property wouldn't have to be added separately.

We will consider all of the properties together and see what the optimal way of using them will be.
And we will see which ones should be modified and/or added on top of existing properties.

Thanks!

Kind regards,
Kagan Akbas

2018-02-02 15:08 GMT+01:00 Floor Koole <F.Koole@ibfd.org>:

Dear David, Kagan,

As discussed last Wednesday, we would come up with a list of properties relevant for our TNS (Tax News Service) documents. TNS is the subcollection of our data that we want to use for this PoC.

Discussing these properties with Ties, we concluded that they can be divided into two types of properties.

The first type ('document properties') contains those document characteristics that we may want to use for sorting/viewing/reporting.

Regarding the second type (which I termed 'workflow-related'), we are not entirely certain whether use of such properties is the optimal way to implement the workflow process.

We presume that this depends on whether the workflow employs document properties in order to function, or whether another methodology is used. Please give your input on that.

Document properties

- Title

- Date of publication

- Country/Organization [single or multiple values]

- Author [currently saved as initials]

- Collection (human-friendly) [for TNS items always "Tax News Service"]

- Collection code [for TNS items always "tns"]

- Report type (human-friendly)

- Word file name (identifying a Word version of the same document)

- Html file name (identifying an html version of the same document)

- Xml file name (identifying an xml version of the same document)

- links to other documents

Workflow related

- Document part of current workflow – yes/no

- Document locked – yes/no

- Current owner – [person, if any]

- Version number [#.#]

- Status [draft/review/ready for publication]

Please let us know if you have any comments on or questions about the above.

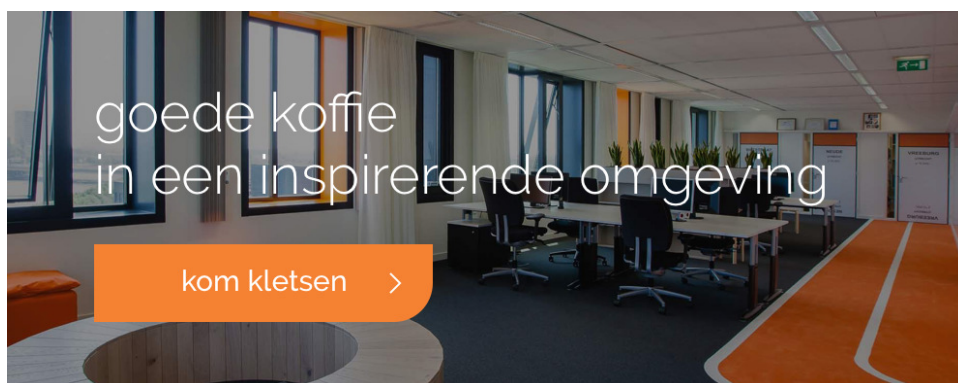Thanks, and kind regards,

Floor Koole

IBFD – IT Development

--

Kagan Akbas
Consultant Digital transformation

t: +31306665315
m: +3161319636
w: www.incentro.com

digital transformation | wtc papendorp
papendorpseweg 100 | 3528 BJ | utrecht

incentro

goede koffie
in een inspirerende omgeving

kom kletsen >

--

goede koffie
in een inspirerende omgeving

kom kletsen ›