

Anobii2Fedora



Client Java per l'ingest automatico in Fedora

Corso di Linguaggi e Traduttori
Prof. Giovanni Semeraro

Fedelucio Narducci (#480101)

SOMMARIO

- **Fedora**
- **FOXML**
- **Client Java**
 - **Funzionamento**
 - **Arhitektura**
- **Prove di esecuzione**
- **Sviluppi futuri**
- **Demo**

2

DALLA PRECEDENTE PUNTATA...

○ **COSA ABBIAMO?**

Webservice per l'estrazione di dati da un account di Anobii (AnobiiRest)

○ **COSA CI SERVE?**

Memorizzare in maniera permanente i dati estratti

○ **DOVE MEMORIZZARLI?**

Fedora Digital Object Repository

3

COSA E' FEDORA?

- **Famosa distribuzione LINUX**

ci inquina i risultati di Google ☹



- **Flexible Extensible Digital Object and Repository Architecture**

Software per la gestione di contenuti, eseguito come webservice all'interno di un webservice Apache Tomcat



4

QUANDO USARE FEDORA?

Alcune caratteristiche:

- *Gestione avanzata dei contenuti e dei metadati*
- *Gestione di relazioni tra oggetti*
- *Contenuti disponibili in molteplici formati*
- *Contenuti generati automaticamente*
- *Formato aperto di rappresentazione dei dati*

5

COME FUNZIONA?

Tutto quello che è presente nel repository di FEDORA è un **DIGITAL OBJECT**

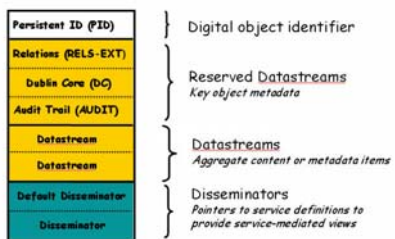
Ogni DIGITAL OBJECT ha tre componenti:

- **PID (Persistent ID)**
- **DATASTREAM**
- **DISSEMINATOR**

6

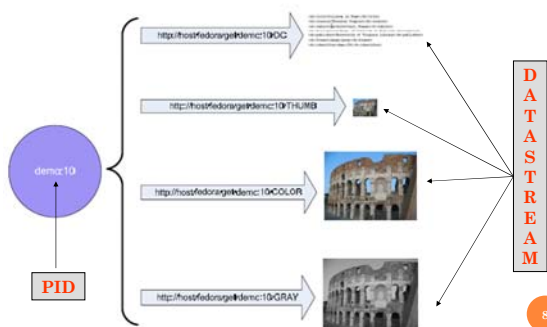
COSA E' UN Digital Object?

Fedora Digital Object Model Container View



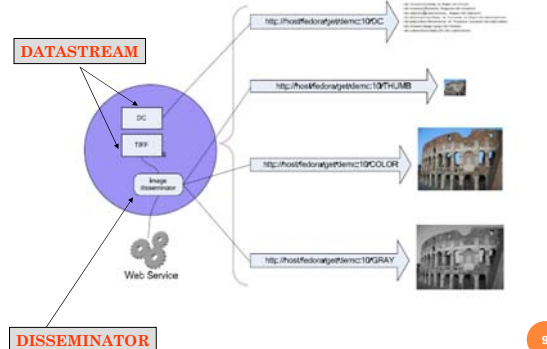
7

ESEMPIO 1/2



8

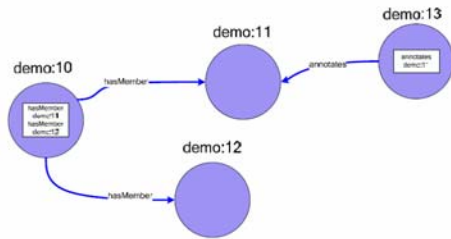
ESEMPIO 2/2



9

RELAZIONI TRA OGGETTI

- Possono essere stabilite relazioni tra Digital Object (definite nella Fedora Relationship Ontology)



RIASSUMENDO

- **PID**
identifica univocamente il Digital Object all'interno del repository
- **Datastream**
sorgente dati
 - *Managed Content (M)* (memorizzato nel repository - MIME type)
 - *Inline XML (X)* (memorizzato nel repository - XML)
 - *External Referenced Content* (esterno al repository - accesso mediato da Fedora)
 - *Redirect Referenced Content* (esterno al repository - reindirizzamento)
- **Disseminator**
servizio offerto dal digital object
 - *Behavior Definition Objects (bDef)*
 - *Behavior Mechanism Objects (bMech)*

} **DO SPECIALI** 11

GLI ELEMENTI DEL NOSTRO DOMINIO (1/2)

- **LIBRO**
digital object
- **UTENTE**
digital object
- **TAGS**
datastream (di utente)
- **COMMENTO**
datastream (di libro)

GLI ELEMENTI DEL NOSTRO DOMINIO (2/2)

LIBRO

PID: book:{ISBN}
DATASTREAM DUBLIN CORE Titolo Autore Editore ISBN Lingua
DATASTREAM REV_utente1_email.it Commento di utente1 sul libro (RDF - Review Vocabulary)
DATASTREAM REV_utente2_email.it Commento di utente2 sul libro (RDF - Review Vocabulary)

UTENTE

PID: user:{user1_email.it}
DATASTREAM DUBLIN CORE Email user1 (si potrebbero aggiungere dati anagrafici, etc., modellati con FOAF)
DATASTREAM TAGS Tag utilizzati da user1 su Anobbi e loro occorrenza (RDF - Vocabolario ad hoc)

13

RAPPRESENTAZIONE DEGLI OGGETTI IN FEDORA

- Un Digital Object all'interno di FEDORA, viene rappresentato in un **semplice linguaggio XML**

FOXML

(Fedora Object XML)

- E' **nessario conoscerlo? SI**
Un oggetto importato in FEDORA (o esportato), viene descritto in questo **linguaggio**

14

ESEMPIO DI DATASTREAM IN FOXML

Il **datastream Dublin Core** di un Digital Object, descritto in **FOXML**

```
<?xml-stylesheet href="http://www.dublincore.org/documents/2003/03/dc-foxml.xsl" type="text/xsl" />
<foxml:foxml xmlns:foxml="http://www.dublincore.org/2003/03/foxml" STATE="S" CONTROL_GROUP="N" VERSIONABLE="yes">
  <foxml:datastream ID="DC" MIME_TYPE="text/xml" LABEL="Dublin Dublin Core Record" SIZE="405" CREATED="2004-12-10T00:21:58.000Z">
    <foxml:comment>
      <foxml:dc:dc
        <dc:title>FOXML Reference Object</dc:title>
        <dc:creator>Daisy Foyette</dc:creator>
        <dc:subject>Fedora documents</dc:subject>
        <dc:description>
          FOXML, allowing how a digital object is encoded for persistent storage in a Fedora repository
        </dc:description>
        <dc:publisher>Cornell CES</dc:publisher>
        <dc:identifier>http://www.dublincore.org/2003/03/foxml/</dc:identifier>
      </foxml:dc:dc>
    </foxml:comment>
  </foxml:datastream>
</foxml:foxml>
```

15

FOXML NEL NOSTRO DOMINIO

E' necessario generare un file FOXML per ogni Digital Object da importare nel repository:

- o LIBRI
- o UTENTI

Come vengono generati i file FOXML?
Automaticamente, attraverso fogli di stile XSL creati *ad hoc*

16

XSL PER IL DO BOOK - *input*

```
<?xml version="1.0" encoding="UTF-8" ?>
<result>
  <libro>
    <titolo>La lunga marcia</titolo>
    <autore>Stephen King, Della Franna B. (Traduttore)</autore>
    <isbn>8804326743</isbn>
    <editore>Spring & Koffer</editore>
    <descrizione>
      Da confra con il Canada sino a Boston a piedi, senza soste. Una sfida mortale, con un regolamento implacabile, per cento volontari un passo falso, una caduta, un malore... e si viene abbattuti. Ma chi riesce a tagliare il traguardo ottiene il Premio. Tra i partecipanti, fra cui spicca il ricercatore Charney, si creano rapporti di sfida, di solidarietà, e di lucida follia, lungo il terribile percorso scandito dagli incantamenti della follia assopita amara della strada. Un modo on the road che solo King (Richard Bachman) poteva concepire...
    </descrizione>
    <pid>book:8804326743</pid>
    <lingua>Libri Italiani</lingua>
  </libro>
</result>

<?xml version="1.0" encoding="UTF-8" ?>
<result>
  <libro>
    <titolo>Terra</titolo>
    <autore>Zofias Erens</autore>
    <isbn>8807040034</isbn>
    <editore>Feltrinelli</editore>
    <descrizione>senza descrizione</descrizione>
    <pid>book:8807040034</pid>
    <lingua>Libri Italiani</lingua>
  </libro>
</result>
```

17

XSL PER IL DO BOOK - *xsl*

```
<?xml version="1.0" encoding="UTF-8" ?>
<result>
  <libro>
    <titolo>La lunga marcia</titolo>
    <autore>Stephen King, Della Franna B. (Traduttore)</autore>
    <isbn>8804326743</isbn>
    <editore>Spring & Koffer</editore>
    <descrizione>
      Da confra con il Canada sino a Boston a piedi, senza soste. Una sfida mortale, con un regolamento implacabile, per cento volontari un passo falso, una caduta, un malore... e si viene abbattuti. Ma chi riesce a tagliare il traguardo ottiene il Premio. Tra i partecipanti, fra cui spicca il ricercatore Charney, si creano rapporti di sfida, di solidarietà, e di lucida follia, lungo il terribile percorso scandito dagli incantamenti della follia assopita amara della strada. Un modo on the road che solo King (Richard Bachman) poteva concepire...
    </descrizione>
    <pid>book:8804326743</pid>
    <lingua>Libri Italiani</lingua>
  </libro>
</result>

<?xml version="1.0" encoding="UTF-8" ?>
<result>
  <libro>
    <titolo>Terra</titolo>
    <autore>Zofias Erens</autore>
    <isbn>8807040034</isbn>
    <editore>Feltrinelli</editore>
    <descrizione>senza descrizione</descrizione>
    <pid>book:8807040034</pid>
    <lingua>Libri Italiani</lingua>
  </libro>
</result>
```

18

XSL PER IL DO BOOK - output

```
<?xml version="1.0" encoding="UTF-8" ?>
<foml:digitalObject xsi:schemaLocation="info:fedora/fedora-system:def/foxml#
http://www.fedora.info/definitions/1.0/foxml1-0.xsd" id="book:8807040034"
xmlns:oa="http://www.oa.org/2001/XMLSchema-instance" xmlns:foml="info:fedora/fedora-
system:def/foxml#">
  <foml:objectProperties>
    <foml:property NAME="http://www.w3.org/1999/02/22-rdf-syntax-ns#type" VALUE="FedoraObject" />
    <foml:property NAME="info:fedora/fedora-system:def/model#state" VALUE="A" />
    <foml:property NAME="info:fedora/fedora-system:def/model#label" VALUE="Terra!" />
    <foml:property NAME="info:fedora/fedora-system:def/model#contentModel" VALUE="TEST" />
    <foml:property NAME="info:fedora/fedora-system:def/model#ownerId" VALUE="TerraAdmin" />
    <foml:extProperty NAME="http://www.openarchives.org/OAI/1.1/oai-identifier.xsd"
      VALUE="oai-cornell.edu:test:100" />
  </foml:objectProperties>
  <foml:datastream ID="DC" STATE="A" CONTROL_GROUP="X" VERSIONABLE="true">
    <foml:datastreamVersion ID="DC.0" MIMETYPE="text/xml" LABEL="Default Dublin Core Record"
      CREATED="2004-08-18T06:32:00" FORMAT_URI="http://www.openarchives.org/OAI/2.0/oai_dc/">
      <foml:xmlContent>
        <oa:oa_dc xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:oa_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
          <dc:title>Terra!</dc:title>
          <dc:creator>Stefano Benni</dc:creator>
          <dc:identifier>8807040034</dc:identifier>
          <dc:description>senza descrizione</dc:description>
          <dc:language>Libri Italiani</dc:language>
          <dc:publisher>F. Trinelli</dc:publisher>
        </oa_dc>
      </foml:xmlContent>
    </foml:datastreamVersion>
  </foml:datastream>
</foml:digitalObject>
```

19

LE API DI FEDORA

o API-M

- Sono le API utilizzate per le operazioni di management (creazione, modifica, cancellazione Digital Object e loro componenti)

o API-A

- Sono le API utilizzate per l'accesso in lettura al repository (es: visualizzare i disseminatori che un oggetto espone)

20

LE API DI FEDORA

o Come effettuare le operazioni sul repository?

- Interfaccia grafica
- Linea di comando
- SOAP (Fedora mette a disposizione un client SOAP Java)

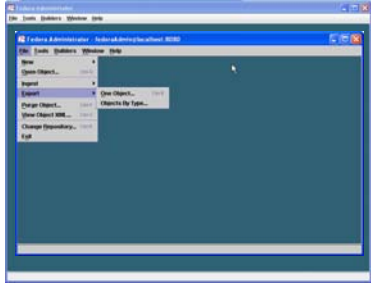
o Inoltre esistono

- API-M-Lite - API-A-Lite
 - o Accesso Rest
 - o Implementano pochissimi metodi

21

LE API DI FEDORA

◦ L'interfaccia grafica



22

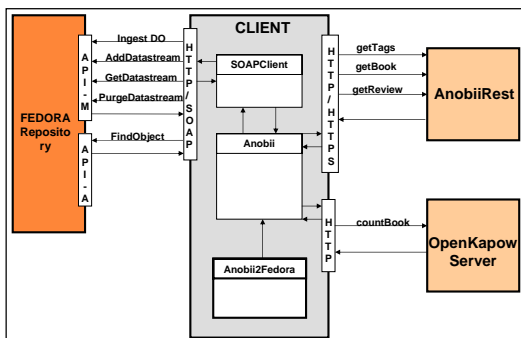
IL CLIENT

◦ Cosa fa?

- **connessione** al **repository** di Fedora
- **ingest** in Fedora delle **informazioni relative all'utente**
- **invocazione** del webservice per l'**estrazione dei tag**
- **ingest** in Fedora dei **tag** estratti al passo precedente
- **invocazione** del webservice per l'**estrazione delle informazioni relative ai libri** presenti nella libreria dell'utente
- **ingest** in Fedora delle informazioni estratte al passo precedente
- **invocazione** del webservice per l'**estrazione dei commenti** sui libri
- **ingest** in Fedora dei **commenti** estratti al passo precedente

23

IL CLIENT - ARCHITETTURA



24

IL CLIENT - CONFIGURAZIONE

Il client è totalmente **parametrizzato**

Prima dell'esecuzione, estrai i valori dei parametri da **tre file di configurazione**:

- **anobiiConf.xml**
- **anobiiRestConf.xml**
- **fedoraConf.xml**

25

IL CLIENT - CONFIGURAZIONE

anobiiConf.xml describe:

- il **path** del foglio di stile **"user.xsl"**
- il **path** del foglio di stile **"book.xsl"**
- il **path** del file temporaneo **"user.xml"** utilizzato per memorizzare il FOXML prima dell'ingest in Fedora
- il **path** del file temporaneo **"book.xml"** utilizzato per memorizzare il FOXML prima dell'ingest in Fedora
- l'**URL** del **robot di OpenKapow** a cui è delegato il compito di **contare i libri** presenti nella libreria dell'utente

26

IL CLIENT - CONFIGURAZIONE

anobiiConf.xml

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- Anobii2Fedora Configuration File
Last Modified: Ottobre 2007
Authors: Cataldo Musto, Lucio Narducci -->
<!-- Inserire di seguito i parametri di configurazione
-->
<!-- Inizio file di configurazione
-->
<configuration>
  <!-- Path del foglio di stile XSL -->
  <user_xsl>./conf/user.xsl</user_xsl>
  <!-- Path del file temporaneo FOXML -->
  <user_foxml>./temp/user.xml</user_foxml>
  <!-- Path del foglio di stile XSL -->
  <book_xsl>./conf/book.xsl</book_xsl>
  <!-- Path del file temporaneo FOXML -->
  <book_foxml>./temp/book.xml</book_foxml>
  <!-- URL del Robot per il conteggio -->
  <robot_url>http://service.openkapow.com/luna70/countbook2.xml</robot_url>
</configuration>
```

27

IL CLIENT - CONFIGURAZIONE

anobiiRestConf.xml describe:

- o il **protocollo** di comunicazione con **AnobiiRest** (http/https)
- o l'**host** per l'esecuzione dell'applicazione
- o la **porta** di comunicazione
- o la **cartella** di istallazione del servizio su **Tomcat**
- o gli **URL** per l'invocazione dei **servizi** estrazione tag, estrazione book, estrazione review, estrazione dei contenuti precedenti in formato RDF

28

IL CLIENT - CONFIGURAZIONE

anobiiRestConf.xml

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- AnobiiFedora Configuration File
      last Modified: Ottobre 2007
      authors: Cataldo Busto, Lucio Narducci -->
<!-- Inserire di seguito i parametri per la connessione al Servizio Anobii Rest
-->
<!-- Inizio file di configurazione
-->
<configuration>
  <!-- Protocollo di connessione -->
  <protocol>https</protocol>
  <!-- Host per l'esecuzione dell'applicazione -->
  <host>localhost</host>
  <!-- Porta di comunicazione -->
  <port>8143</port>
  <!-- Cartella base di installazione -->
  <folders>anobiiRest</folders>
  <!-- URL delle Restlet -->
  <tags_url>/tags</tags_url>
  <book_url>/books</book_url>
  <reviews_url>/reviews</reviews_url>
  <!-- URL per i contenuti in RDF -->
  <rdf_url>/rdf</rdf_url>
</configuration>
```

29

IL CLIENT - CONFIGURAZIONE

fedoraConf.xml describe:

- o il **protocollo** di comunicazione con **Fedora**
- o l'**host** per l'esecuzione dell'applicazione
- o la **porta** di comunicazione
- o la **cartella** di istallazione del servizio Fedora su **Tomcat**
- o l'**username** per l'accesso a **Fedora**
- o la **password** per l'accesso a **Fedora**
- o il **path** del file **truststore** in Fedora (viene utilizzato quello di default di Fedora)
- o la **password** del **truststore**

30

IL CLIENT - CONFIGURAZIONE

fedoraConf.xml

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- Ancho: Fedora Configuration File
Last Modified: Ottobre 2007
Autore: Cataldo Musto, Lucio Narducci -->
<!-- Inserire di seguito i parametri per la connessione al Repository Fedora
-->
<!-- Inizio file di configurazione
-->
<configuration>
  <!-- Protocollo di connessione -->
  <protocol>http</protocol>
  <!-- Host per l'esecuzione dell'applicazione -->
  <host>localhost</host>
  <!-- Porta di comunicazione -->
  <port>8001</port>
  <!-- Cartella base di installazione -->
  <folder>fedora</folder>
  <!-- Nome utente per l'accesso al repository -->
  <user>fedoraAdmin</user>
  <!-- Password -->
  <pass>admin</pass>
  <!-- Percorso dove è memorizzato il file truststore -->
  <truststore>C:/fedora2/serve/truststore</truststore>
  <!-- Password del Truststore -->
  <truststorepassword />
</configuration>
```

31

FUNZIONAMENTO

- **Ingest book**
se il libro è già presente, viene ignorato l'inserimento
- **Ingest user**
se l'utente è già presente, viene ignorato l'inserimento
- **Ingest review**
se è già presente un commento di quell'utente su quel libro, prima dell'inserimento del nuovo, viene cancellato il vecchio commento
- **Ingest tags**
se è già presente un elenco di tags usati da quell'utente, prima dell'inserimento del nuovo, viene cancellato il vecchio elenco

32

PROVE DI ESECUZIONE

FASE	# ITEM	OPERAZIONE ESEGUITA	STATO REPOSITORY	RISULTATO ATTESO	RISULTATO OTTENUTO
1	BOOKS: 93 USERS: 3 REVIEWS: 17 TAGS: 24	INGEST di utenti, libri, commenti, tags	VUOTO	Inserimento di DO (utenti e libri) e DS (commenti, tags) nel repository	Inserimento eseguito con successo

33

SVILUPPI FUTURI

- Istanziare relazioni tra oggetti (per nuove funzionalità)
- Gestire il versionamento dei datastream
- Importare dati da altre sorgenti (siti web, etc.)

37

FINE

38
