Fedora 4.7 Triplestore Integration Notes
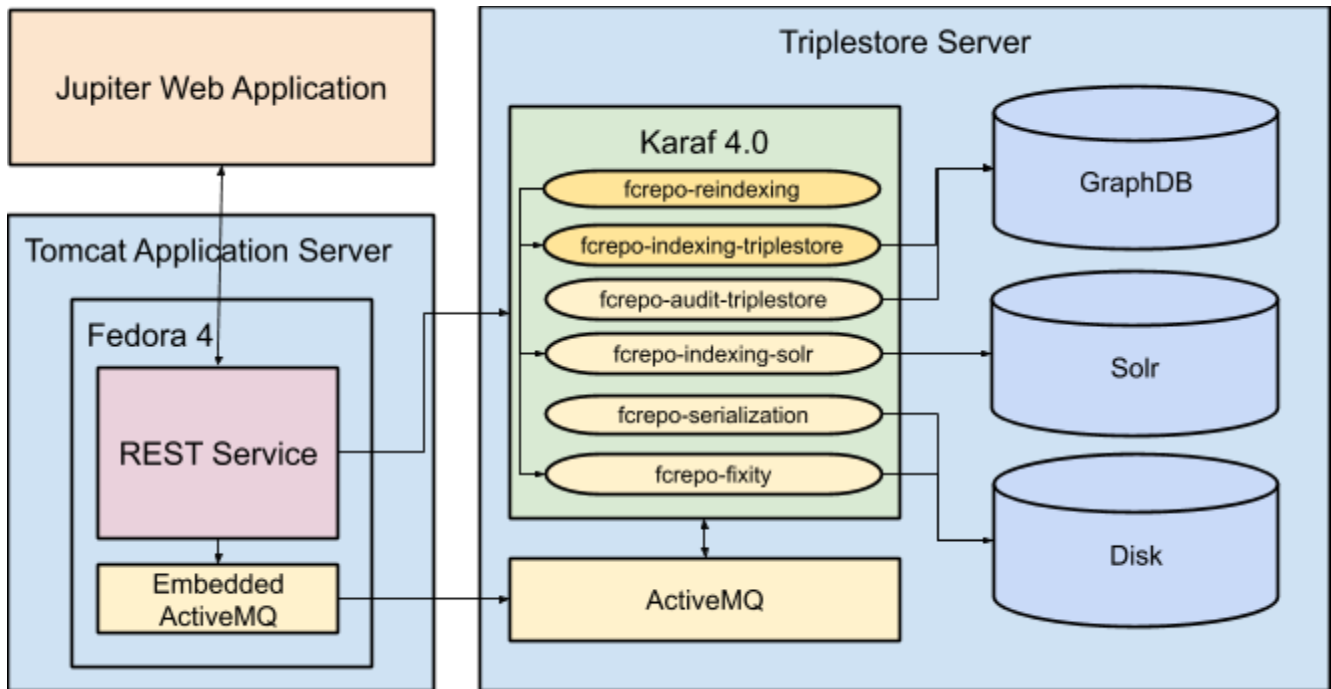
- https://wiki.duraspace.org/display/FEDORA474/Setup+Camel+Message+Integrations

Triplestroes

- Apache Jena Fuseki: http://sheff.library.ualberta.ca:3030/
- Apache Marmotta: http://sheff.library.ualberta.ca:8080/marmotta
- RDF4J: http://sheff.library.ualberta.ca:8080/rdf4j-workbench
- Blazegraph: http://sheff.library.ualberta.ca:9999/
- GraphDB: http://graphdb.ontotext.com/

Start / Stop Triplestores

- Fuseki
  - `$ sudo /opt/fuseki/fuseki [start|stop|status|restart]`
- Blazegraph-2.1.4
  - `$ sudo /opt/blazegraph/bin/blazegraph.sh [start|stop|status|restart]`
- Marmotta / RDF4J (both running on tomcat server)
  - `$ sudo service tomcat [start|stop]`

Start / Stop Karaf

- `$ sudo /opt/karaf/bin/[start|stop|status]`

Karaf Web Console

- http://sheff.library.ualberta.ca:8181/hawtio

Fedora 4 (era-test)

- http://gillingham.library.ualberta.ca:8080/fedora/rest/

Triplestore Works

- Install Apache Karaf and fcrepo-camel-toolbox: fcrepo-indexing-tripletore, fcrepo-reindexing, fcrepo-audit and fcrepo-fixity and howtio
- Install Apache Jena Fuseki triplestore
- Install Tomcat7 and deploy Apache Marmotta, RDF4J triplestores
- Install Blazegraph triplestore
- Deploy jolokia webapp agent on Tomcat7 and add jolokia agent to Fuseki configuration
- Configure fcrepo-reindexing and fcrepo-indexing-triplestore
- Index Fedora4 data into Fuseki
- *Test fcrepo-indexing-triplestore to make sure that automatically update is working properly*
- *Configure fcrepo-audit and fcrepo-audit-triplestore and test*
- *Configure fcrepo-fixity and test*
- *Index Fedora4 data into Marmotta*
- Index Fedora4 data into RDF4J
- Index Fedora4 data into Blazegraph
- *Setup triplestore production server*
- *Index Fedora4 production server to selected triplestore*

Karaf Installation

- Server: **sheff.library.ualberta.ca**
- IP: 129.128.222.21
- Hawt.io Karaf Console: http://sheff.library.ualberta.ca:8181/hawtio
  - username/password, karaf/karaf
- Install apache **karaf 4.0.10** and start http://karaf.apache.org/manual/latest/#_quick_start
- Configure remote debugger (Eclipse)
  - /apache-karaf-4.0.10/bin/setenv
    - adding export KARAF_DEBUG=true # Enable debug mode
  - Or /bin/start debug
  - Configure ssh tunnel if firewall is not opened
    - ssh -f pcharoen@sheff -L 5005:sheff:5005 -N

- - Create Remote Java Application debugger configuration point to localhost:5005

  - Start / Stop Karaf
    - Start: `/bin/start | /bin/start debug`
    - Stop: `/bin/stop`
    - Console: `/bin/client`
  - Client command line to start feature
    - `$ ./client feature:start fcrepo-serialization`
    - `$ java -Dkaraf.instances=/opt/karaf/instances -Dkaraf.home=/opt/karaf`
      `-Dkaraf.base=/opt/karaf -Dkaraf.etc=/opt/karaf/etc`
      `-Djava.io.tmpdir=/opt/karaf/data/tmp`
      `-Djava.util.logging.config.file=/opt/karaf/etc/java.util.logging.properties`
      `-classpath`
      `/opt/karaf/system/org/apache/karaf/org.apache.karaf.client/4.0.10/org.apache.`
      `karaf.client-4.0.10.jar:/opt/karaf/system/org/apache/sshd/sshd-core/0.14.0/ss`
      `hd-core-0.14.0.jar:/opt/karaf/system/jline/jline/2.14.1/jline-2.14.1.jar:/opt`
      `/karaf/system/org/slf4j/slf4j-api/1.7.12/slf4j-api-1.7.12.jar`
      `org.apache.karaf.client.Main feature:start fcrepo-serialization`

Install Fedora Camel Toolbox dependencies

See v4.7.4 document, and Github project and sub project documents
- https://wiki.duraspace.org/display/FEDORA474/Setup+Camel+Message+Integrations
- https://github.com/fcrepo4-exts/fcrepo-camel-toolbox/tree/fcrepo-camel-toolbox-4.7.2

(working with v4.0.10, but **not working** with v4.1.1 and v4.1.2)
(See below for Karaf Provisioning)

Install Fedora Camel Toolbox

```
$> feature:repo-add camel 2.18.0
$> feature:repo-add activemq 5.14.1
$> feature:install camel
$> feature:install activemq-camel

# display available camel features
$> feature:list | grep camel

# install camel features, as needed
$> feature:install camel-http4

# install fcrepo-camel-toolbox (as of v4.7.2)
$> feature:repo-add mvn:org.fcrepo.camel/toolbox-features/4.7.2/xml/features
$> feature:install fcrepo-service-activemq
$> feature:install fcrepo-indexing-triplestore
$> feature:install fcrepo-audit-triplestore
$> feature:install fcrepo-fixity
$> feature:install fcrepo-indexing-sole
```

- Install hawt.io monitoring web console, hawtio is available at http://localhost:8181/hawtio/

  ```
  $ karaf> feature:repo-add hawtio
  $ karaf> feature:install hawtio
  ```

- Configure components either edit the configuration file or use hawtio web interface
  - /etc/org.fcrepo.camel.audit.cfg

    ```
    # The baseUri to use for event URIs in the triplestore. A `UUID` will be appended
    # to this value, forming, for instance: `http://example.com/event/{UUID}`
    event.baseUri=http://era.library.ualberta.ca/event

    # The base URL of the triplestore being used.
    triplestore.baseUrl = localhost:3030/audit/update
    ```
  - /etc/org.fcrepo.camel.indexing.triplestore.cfg

    ```
    # The baseUrl for the fedora repository.
    fcrepo.baseUrl = localhost:8080/fedora/rest/

    # The base URL of the triplestore being used.
    triplestore.baseUrl = localhost:3030/index/update
    ```
  - /etc/org.fcrepo.camel.reindexing.cfg

    ```
    # The baseUrl for the fedora repository.
    fcrepo.baseUrl = localhost:8080/fedora/rest/
    ```

- Fedora 4.7.4 on Gillingham
  - Start | Stop

    ```
    $ sudo service tomcat7 (start|stop)
    ```
  - Remove remote access filter in ./tomcat7/conf/server.xml

```
<!--
<Valve className="org.apache.catalina.valves.RemoteAddrValve"
allow="127\.\d+\.\d+\.\d+|::1|142.244.34.\d+|129.128.217.88|129.128.217.108|129.128.46.143
|129.128.222.21"/>
→
```

- Apache Jena Fuseki
  - Start | stop

    ```
    $ ./fuseki (start|stop)
    ```
  - Allow remote access in ./fuseki/run/shiro.ini

    ```
    $ #/$/** = localhostFilter
    $ /$/** = anon
    ```
  - Install Jolokia and add the agent in /etc/default/fuseki for Hawt.io connection

    ```
    JAVA_OPTIONS="Xmx1200m -javaagent:/opt/jolokia/agents/jolokia-jvm.jar"
    ```
- Reindex repository data to external triplestore running on sheff
  https://wiki.duraspace.org/display/FEDORA451/Integration+Services

  ```
  $ curl -XPOST localhost:9080/reindexing/prod -H "Content-Type: application/json" -d
  '["activemq:queue:triplestore.reindex"]'
  ```

- SPARQL Examples

  ```
  # count objects
  ```

```
SELECT (count(*) as ?count)
WHERE {
      ?s ?p ?o .
}


# find model:hasModel and count
SELECT ?o (COUNT(*) AS ?count)
WHERE {
            ?s <info:fedora/fedora-system:def/model#hasModel> ?o .
      }
GROUP BY ?o
```

Open Firewall Commands on Sheff for Remote Debugger

- `$ sudo firewall-cmd --add-port=5005/tcp --zone=public --permanent`
- `$ sudo service firewalld restart`

Sheff Firewall Rules

- 8080: Apache Tomcat for RDF4J and Apache Marmotta triplestores
- 8181: Hawt.io Karaf dashboard and monitoring
- 3030: Apache Fuseki triplestore
- 9999: BlazeGraph triplestore
- 5005: Remote debugger
- The server needs to access Gillingham on port 8080 (Fedora 4) to retrieve data and create index on triplestores.
- Piyapong: 129.1128.46.143
- Metadata Team: Sharon, John, Mariana, Zach and Robert

Karaf Web Console

- `feature:install webconsole`
- `http://localhost:8181/system/console`

Karaf Local Repository

- Karaf scans local repository for feature to install before downloading from remote repository.
- The default local repository is ${user.home}/.m2/repository/ (/root/.m2/repository/).
- Set local repository in /etc/org.ops4j.pax.url.mvn.cfg:
    - `org.ops4j.pax.url.mvn.localRepository=/home/pcharoen/.m2/repository`
- To update local repository, run # `mvn clean install` to install feature on the local repository.

Karaf Provisioning

(http://karaf.apache.org/manual/latest/#_provisioning)
- Boot features (http://karaf.apache.org/manual/latest/#_boot_features)
    - /etc/org.apache.karaf.features.cfg
      `featuresRepositories = \`

```
        mvn:org.apache.karaf.features/standard/4.0.10/xml/features, \
        mvn:org.apache.karaf.features/spring/4.0.10/xml/features, \
        mvn:org.apache.karaf.features/framework/4.0.10/xml/features, \
        mvn:org.apache.karaf.features/enterprise/4.0.10/xml/features,\
        mvn:io.hawt/hawtio-karaf/2.0.0/xml/features, \
        mvn:org.apache.camel.karaf/apache-camel/2.18.0/xml/features, \
        mvn:org.apache.activemq/activemq-karaf/5.14.1/xml/features, \
        mvn:org.fcrepo.camel/toolbox-features/4.7.2/xml/features

#
# Comma separated list of features to install at startup
#
featuresBoot = \
    instance, \
    package, \
    log, \
    ssh, \
    aries-blueprint, \
    framework, \
    system, \
    feature, \
    shell, \
    management, \
    service, \
    jaas, \
    shell-compat, \
    deployer, \
    diagnostic, \
    wrap, \
    bundle, \
    config, \
    kar, \
    webconsole, \
    hawtio, \
    camel, \
    activemq-camel, \
    camel-http4, \
    camel-quartz2, \
    fcrepo-service-activemq, \
    fcrepo-indexing-triplestore, \
    fcrepo-fixity, \
    fcrepo-audit-triplestore, \
    fcrepo-service-ldcache-file, \
    fcrepo-ldpath, \
    fcrepo-indexing-solr, \
    fcrepo-serialization, \
    fcrepo-reindexing
```

Karaf Remote Shell

- `$ /karaf/bin/client -h triplestore.library.ualberta.ca -u karaf`

Hawt.io Monitoring Tools

- URL: http://sheff.library.ualberta.ca:8181/hawtio/welcome
- Fuseki (jolokia javaagent)
  - Host: localhost, port: 8778, path: jolokia
- Fedora4 (deploy jolokia.war on Fedora4 tomcat)
  - Host: gillingham.library.ualberta.ca, port: 8080, path: jolokia
- Tomcat7 (deploy jolokia.war on tomcat7)
  - Host: localhost, port: 8080, path: jolokia
- Blazegraph (jolokia javaagent)
  - Host: localhost, port: 8779, path: jolokia
- Solr (Blazegraph external search index) (jolokia javaagent)
  - Host: localhost, port: 8780, path: jolokia

Container Configurations

- Install Chrome LocalStorage Manager extension.

Import data from sheff.library.ualberta.ca-8181-2017-09-22_14-13-20.txt by open the file and copy the content and paste in local storage data (JSON) input box then click OK.

Camel Component Modification

- Checkout source code from Git repository.
- Make changes.
- Compile and install using JDK8, set JAVA_HOME if needed.
- `# mvn -DskipTests clean install`

Add Log Messages

- Add in ReindexingRouter onException to print out object path and JMS headers.

Add Filters

- Add in TriplestoreRouter (`direct:index.triplestore`) to filter out, content, thumbnail, fedora3foxml, era1stats, batch and lease objects.

Maven Repository Jar File Installation to Maven Local Repository

Install jar file from Maven project to specific Maven Repository

fcrepo-reindexing

```
$ mvn org.apache.maven.plugins:maven-install-plugin:2.5.2:install-file
-Dfile=target/fcrepo-reindexing-4.7.2.jar -DgroupId=org.fcrepo.camel
-DartifactId=fcrepo-reindexing -Dversion=4.7.2 -Dpackageing=jar
-DlocalRepositoryPath=/opt/karaf/mvn/repository/
Or
$ cd fcrepo-reindexing
$ ./install_jar.sh
```

fcrepo-audit-triplestore-blueprint

```
$ mvn clean install
$ mvn org.apache.maven.plugins:maven-install-plugin:2.5.2:install-file
-Dfile=target/fcrepo-audit-triplestore-blueprint-4.7.2.jar -DgroupId=org.fcrepo.camel
-DartifactId=fcrepo-audit-triplestore-blueprint -Dversion=4.7.2 -Dpackageing=jar
-DlocalRepositoryPath=/opt/karaf/mvn/repository/
```

toolbox-features

To fix Local Maven Repository Unknown Protocol wrap
- Add `<feature prerequisite="true">wrap</feature>` to
  /fcrepo-camel-toolbox/toolbox-features/src/main/resources/features.xml
- Run mvn clean install -DskipTests on sub project /fcrepo-camel-toolbox/toolbox-features/
- Deploy features to the local repository by copying
  .m2/repository/org/fcrepo/camel/toolbox-features/4.7.2/ to
  /mvn/repositroy/org/fcrepo/camel/toolbox-features/4.7.2/

  ```
  $ cp .m2/repository/org/fcrepo/camel/toolbox-features/4.7.2/
  /mvn/repositroy/org/fcrepo/camel/toolbox-features/4.7.2/
  ```

Reindexing

Replacing localhost with IP address if necessary.

Reindex All

- Reindex repository data to external triplestore running on sheff
  https://wiki.duraspace.org/display/FEDORA451/Integration+Services
  ```
  $ curl -XPOST localhost:9080/reindexing/prod -H "Content-Type: application/json" -d
  '["broker:queue:triplestore.reindex", "broker:queue:solr.reindex", "broker:queue:fixity",
  "broker:queue:serialization"]'
  ```

Reindex Triplestore

- ```
  $ curl -XPOST localhost:9080/reindexing/prod -H "Content-Type: application/json" -d
  '["broker:queue:triplestore.reindex"]'
  ```

Reindex Solr

- ```
  $ curl -XPOST localhost:9080/reindexing/prod -H "Content-Type: application/json" -d
  '["broker:queue:solr.reindex"]'
  ```

Reindex Fixity

- ```
  $ curl -XPOST localhost:9080/reindexing/prod -H "Content-Type: application/json" -d
  '["broker:queue:fixity"]'
  ```

Reindex Serialization

- ```
  $ curl -XPOST localhost:9080/reindexing/prod -H "Content-Type: application/json" -d
  '["broker:queue:serialization"]'
  ```

Indexing

Fuseki

Gillingham

- *Reindexing request using curl command does not respond, the request hanging (sync)*
- *Reindexing stop after indexing a number of objects*
- *Reindexing again for a number of objects then stop*
- *The repository has ~40,000 items, ~205,600 objects*
- *JMS messages on Fedora in 2 queues*
    - *reindexing*
    - *triplestore.reindex*
- *The messages from reindexing has been moved to triplestore.reindex automatically for a number of messages then stopped. The messages in triplestore.reindex started to send back to Camel component and begin to update the index on the triplestore of Sheff*
- *Moving the messages from reindexing manually from reindexing queue to triplestore.reindex queue makes the indexing process started again*
- *Moving the messages manually using howtio (connecting to jolokia.war on Fedora server) ActiveMQ user interface (JMX) and moveMatchingMessagesTo(java.lang.String,java.lang.String) operation with empty Selector and Destination move all messages from reindexing queue to triplestore.reindex queue*
- ~~Solve the problems above by setting fcrepo-reindexing configuration (/etc/org.fcrepo.camel.reindexing.cfg): **reindexing.stream = activemq:queue:triplestore.reindex**~~
- SPARQL query first attempt, number of items on the user interface is 39,733 (Solr index)

o	count
1 "GenericFile"	"40644"^^xsd:integer

	o	count
2	"Hydra::AccessControls::Lease"	"14"^^xsd:integer
3	"Hydra::AccessControls::Embargo"	"1583"^^xsd:integer
4	"Collection"	"444"^^xsd:integer
5	"Batch"	"47853"^^xsd:integer
6	"Hydra::AccessControls::Permission"	"127192"^^xsd:integer

7,837,266 triples
- Reindex all data in Fedora 4 repository, 215,874 objects took ~9 hours
- SPARQL query results by model:hasModel

	o	count
1	"GenericFile"	"40646"^^xsd:integer
2	"Hydra::AccessControls::Lease"	"14"^^xsd:integer
3	"Hydra::AccessControls::Embargo"	"1583"^^xsd:integer
4	"Collection"	"444"^^xsd:integer
5	"Batch"	"47855"^^xsd:integer
6	"Hydra::AccessControls::Permission"	"127195"^^xsd:integer

Plano

- Reindexing request response with message, Indexing started at /dev
- The repository has 2,956 items, 15,195 objects
- Reindexing finished, 465,010 triples
- SPARQL Query to group by model:hasModel
  ```
  SELECT ?o (COUNT(*) AS ?count)
  WHERE {
        ?s <info:fedora/fedora-system:def/model#hasModel> ?o .
  }
  GROUP BY ?o
  ```
  Results:

o	count

1	"GenericFile"	"2956"^^xsd:integer
2	"Collection"	"387"^^xsd:integer
3	"Batch"	"3030"^^xsd:integer
4	"Hydra::AccessControls::Permission"	"8821"^^xsd:integer

Number of objects in triplestore is 15,194 objects close to 15,195 objects in repository. There might be an object that does not have model:hasModel property or binary object without metadata.

Solr Indexing

- Create Fedora ldpath custom transformation, indexing-solr-transformation.txt
- Use Fedora REST API to register the custom transformation
  (https://wiki.duraspace.org/display/FEDORA42/RESTful+HTTP+API+-+Transform)
  ```
  curl -u hydranorth:_u9_Ap-F -X PUT -H "Content-Type: application/rdf+ldpath"
  --data-binary "@indexing-solr-transformation.txt"
  "http://gillingham.library.ualberta.ca:8080/fedora/rest/fedora:system/fedora:transf
  orm/fedora:ldpath/indexing-solr/fedora:Container"
  ```
- Configure fcrepo-indexing-solr Camel component to use the transformation
  (/etc/org.fcrepo.camel.indexing.solr.cfg)
  ```
  fcrepo.defaultTransform = indexing-solr
  ```
- Delete all documents: `curl http://localhost:8983/solr/${core}/update?commit=true -H "Content-Type: text/xml" --data-binary '<delete><query>*:*</query></delete>'`

Fixity

- Request on object path
  - `$ curl -XPOST localhost:9080/reindexing/prod -H"Content-Type: application/json" -d '["broker:queue:fixity"]'`
- Output in /tmp/fixityErrors.log
- Configure output in /etc/org.fcrepo.camel.fixity.cfg
  - `fixity.failure=file:/opt/karaf/data/log/?fileName=fixityErrors.log&fileExist=Append`

Audit

- Modify the fcrepo-audit-triplestore component to use message id as a subject then deploy the component to Maven local repository. The karaf feature installation will scan the Maven local repository before downloading from the remote repository.

Marmotta

- Use simple security.profile (default) to allow localhost access for indexing

- Change marmotta.home in web.xml pointing to data directory (Ex: /var/data/marmotta)
- fcrepo-indexing-triplestore configuration
  - Triplestore base url: localhost:8080/marmotta/sparql/update

Blazegraph

Configurations

https://drive.google.com/drive/folders/1n67g3kOpmYQaD4BZxn78_8ux89-dRO_F?usp=sharing

Setup

- Add javaagent in /bin/blazegraph.sh

    ```
    ...
    cmd=java \
    -javaagent:/opt/jolokia/agents/jolokia-jvm.jar=port=8779,host=localhost \
    ${JAVA_OPTS} \
    ...
    ```
- Start Blazegraph
  - `$ sudo /opt/blazegraph/bin/blazegraph.sh [stat|stop|status|restart]`
- Create namespace, fcrepo
- fcrepo-indexing-triplestore configuration
  - Triplestore URL: `localhost:8080/blazegraph/namespace/fedora/sparql`
- Install external full-text index
  - https://wiki.blazegraph.com/wiki/index.php/SOLR_External_Fulltext_Search
  - Start solr with javaagent
    - `/opt/solr/bin/solr start -force -a "-javaagent:/opt/jolokia/agents/jolokia-jvm.jar=port=8780,host=localhost"`
  - Solr indexing, need to modify *label2JSON.sh* to query data using REST API and transform results to json and insert to Solr.

Create namespace (using API)

- Export properties from existing namespace
  - `curl --header 'Accept: application/xml' http://localhost:9999/blazegraph/namespace/gillingham/properties > gillingham.xml`
- Edit and change namespace in the properties.
- Create namespace
  - `curl -v -X POST --data-binary @gillingham.xml --header 'Content-Type:application/xml' http://localhost:9999/blazegraph/namespace`
- Delete namespace
  - `curl -v -X DELETE http://localhost:9999/blazegraph/namespace/${namespace}`

Export data instructions

- Stop Blazegraph server before export.

- Properties File (`data.properties`) pointing to blazegraph.jnl file.
  - `com.bigdata.journal.AbstractJournal.file=/opt/blazegraph/data/blazegraph.jnl`
- Export Utility (blank {namespace} to export all namespaces)
  - `java -cp blazegraph.jar com.bigdata.rdf.sail.ExportKB data.properties [{namespace}]`
  - See `com.bigdata.rdf.sail.ExportKB` Javadocs for options
    - Export one or more KBs from a Journal. The only required argument is the name of the properties file for the Journal. By default all KB instances found on the journal will be exported into the current working directory. Each KB will be written into a subdirectory based on the namespace of the KB.
    - Parameters:
      - args [options] propertyFile namespace* where options is any of:
    - -outdir
      - The output directory (default is the current working directory)
    - -format
      - The RDFFormat which will be used to export the data. If not specified then an appropriate format will be selected based on the KB configuration. The default for triples or SIDs is RDFFormat.RDFXML. The default for quads is RDFFormat.TRIX.
      - [RDF/XML, N-Triples, Turtle, N3, TriX, TriG, BinaryRDF, N-Quads, JSON-LD, RDF/JSON, RDFa]
      - See in class org.openrdf.rio.RDFFormat
    - -includeInferred
      - Normally only the told triples/quads will be exported. This option may be given to export the axioms and inferences as well as the told triples/quads.
    - -n
      - Do nothing, but show the KBs which would be exported.
    - -help
      - Display the usage message and exit.
    - where propertyFile is the properties file for the Journal.
    - where namespace is zero or more namespaces of KBs to export from the Journal. If no namespace is given, then all KBs on the Journal will be exported.

Export data command

- `$ java -cp blazegraph.jar -format 'Turtle' com.bigdata.rdf.sail.ExportKB blazegraph.properties fedora`

Delete all triples

- `# curl --get -X DELETE -H 'Accept: application/xml' 'http://localhost:9999/blazegraph/namespace/${namespace}/sparql'`

GraphDB

Github

- https://github.com/ualbertalib/di_internal/tree/triplestore

Start GraphDB

- `# ./graphdb -Dgraphdb.connector.port=8080 -Dgraphdb.workbench.importDirectory=/data/graphdb/import -d`

Stop GraphDB

- `# pkill -9 -f graphdb`

GraphDB Security

Security

- Setup -> Users and Access
  - Security is ON
    - admin/root (default password)
  - Free Access is ON
    - Free Access configuration
      - Repository read/write access
        - audit: read and write
        - fedora: read, write

Create user

- `$ curl 'http://localhost:7200/rest/security/user/${username}' -H 'Origin: http://localhost:7200' -H 'Accept-Encoding: gzip, deflate, br' -H 'Accept-Language: en-US,en;q=0.9' -H 'Content-Type: application/json;charset=UTF-8' -H 'Accept: application/json, text/plain, */*' -H 'Cache-Control: no-cache' -H 'Referer: http://localhost:7201/user/create' -H 'X-GraphDB-Repository: repository name' -H 'X-GraphDB-Password: ${password}' -H 'DNT: 1' --data-binary '{"appSettings": {"DEFAULT_SAMEAS":true,"DEFAULT_INFERENCE":true,"EXECUTE_COUNT":true,"IGNORE_SHARED _QUERIES":false},"grantedAuthorities": ["ROLE_USER", "WRITE_REPO_audit", "READ_REPO_audit", "WRITE_REPO_fedora", "READ_REPO_fedora"]}' --compressed -u ${admin_username}:${admin_password}`

Delete user

- `$ curl 'http://localhost:7200/rest/security/user/${username}' -X DELETE -u ${admin_username}:${admin_password}`

Camel Component Configuration

org.fcrepo.camel.indexing.triplestore.cfg

- ```
  triplestore.baseUrl =
  http4://localhost:7200/repositories/${repositoryId}/statements
  ```

Query data programmatically

- ```
  $ curl -G -H "Accept:application/x-trig"
    -d query=CONSTRUCT+%7B%3Fs+%3Fp+%3Fo%7D+WHERE+%7B%3Fs+%3Fp+%3Fo%7D+LIMIT+10
    http://localhost:7200/repositories/yourrepository
  ```
- ```
  $ curl -X POST --data-binary @file.sparql -H "Accept: application/rdf+xml"
    -H "Content-type: application/x-www-form-urlencoded"
    http://localhost:7200/repositories/worker-node
  ```
  where, **file.sparql** contains an encoded query:
  ```
  query=CONSTRUCT+%7B%3Fs+%3Fp+%3Fo%7D+WHERE+%7B%3Fs+%3Fp+%3Fo%7D+LIMIT+10
  ```

Delete all data in a repository

- ```
  $ curl -X DELETE --header 'Accept: application/json'
  'http://localhost:7200/repositories/${repository_id}/statements'
  ```

Export data

Exporting data in TriX format is including graph object.

- ```
  $ curl -X GET -H "Accept:application/x-trig"
  "http://localhost:7200/repositories/fedora/statements?infer=false" | gzip >
  fedora.trig.gz
  ```

Import data

- Put export data package (Ex: fedora.trig.gz) in graphdb-import directory checking from import on the workbench.
- Select repository to import
- Use Import tool, import server files user interface
- Select the export data package to import and click on import and import without changing settings

Data Migration

- Package repository data directory, Ex: /graphdb/data/repositories/fedora
  - ```
    $ cd /graphdb/data/repositories
    ```
  - ```
    $ tar -zcf fedora.tar.gz
    ```
- Remove repository data directory on the destination repository
  - ```
    $ cd /graphdb/data/repositories
    ```
  - ```
    $ rm -rf fedora
    ```
- Extract the source repository data package on the destination repository
  - ```
    $ tar -xf fedora.tar.gz
    ```

Create Repository using REST API

- Get Repository info
  ```
  $ curl -X GET --header 'Accept: application/json'
  'http://localhost:7200/rest/repositories/fedora'
  ```
- Repository properties sample, **fedora.json**
  ```
  {
    "id": "fedora",
    "location": "",
    "params": {},
    "sesameType": "graphdb:FreeSailRepository",
    "title": "Fedora 4 Triplestore Repository",
    "type": "free"
  }
  ```
- Create fedora Respository
  ```
  $ curl -X PUT -H 'Content-Type: application/json' -H 'Accept: text/plain' -d
  @fedora.json 'http://localhost:7200/rest/repositories'
  ```

Modify Saved queries

- Need to stop GraphDB before editing the settings.js file
- The queries are saved in /Users/Library/Application
  Support/GraphDB/work/workbench/settings.js (json): queries objects.

```
...
 "queries" : {
    "SPARQL Select template" : {
      "name" : "SPARQL Select template",
      "body" : "SELECT ?s ?p ?o\nWHERE {\n\t?s ?p ?o .\n} LIMIT 100"
    },
    "Find Objects by LastModified" : {
      "name" : "Find Objects by LastModified",
      "body" : "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>\nPREFIX fedora:
<http://fedora.info/definitions/v4/repository#>\nselect ?s ?p ?date \nwhere {\n  ?s
fedora:lastModified ?date .\n  FILTER (?date > \"2018-04-13T00:00:00.000Z\"^^xsd:dateTime
&& ?date < \"2018-04-13T23:59:59.999Z\"^^xsd:dateTime)\n}\n"
    },
    "List All FixityServices" : {
      "name" : "List All FixityServices",
      "body" : "SELECT ?s ?p ?o \nWHERE {\n  ?s
\t<http://fedora.info/definitions/v4/repository#hasFixityService> ?o .\n}\nLIMIT 50\n"
    },
    "Find Objects by ContentModel" : {
      "name" : "Find Objects by ContentModel",
      "body" : "PREFIX dc: <http://purl.org/dc/elements/1.1/>\nPREFIX fedora:
<http://fedora.info/definitions/v4/repository#>\nPREFIX model:
<info:fedora/fedora-system:def/model#>\nselect ?s ?p ?o\nwhere {\n    ?s model:hasModel
\"IRItem\" .\n    ?s ?p ?o .\n}\norder by ?s\nlimit 100"
```

```
    },
    "List All Predicates" : {
      "name" : "List All Predicates",
      "body" : "SELECT ?p (COUNT(*) AS ?count)\nWHERE {\n    ?s ?p ?o .\n}\nGROUP BY
?p\nORDER by ?p\n"
    },
    "Add statements" : {
      "name" : "Add statements",
      "body" : "PREFIX dc: <http://purl.org/dc/elements/1.1/>\nINSERT DATA\n      {\n
GRAPH <http://example> {\n          <http://example/book1> dc:title \"A new book\" ;\n
dc:creator \"A.N.Other\" .\n          }\n      }"
    },
    "Delete All Triples" : {
      "name" : "Delete All Triples",
      "body" : "DELETE WHERE {\n    ?s ?p ?o .\n}"
    },
    "List All Versions" : {
      "name" : "List All Versions",
      "body" : "SELECT ?s ?p ?o \nWHERE {\n  ?s
\t<http://fedora.info/definitions/v4/repository#hasVersions> ?o .\n}\nLIMIT 50\n"
    },
    "Find Triple using RegEx" : {
      "name" : "Find Triple using RegEx",
      "body" : "select ?s ?p ?o\nwhere { \n  ?s ?p ?o .\n  FILTER regex(str(?o),
\"alberta\", \"i\")\n}\n"
    },
    "Remove statements" : {
      "name" : "Remove statements",
      "body" : "PREFIX dc: <http://purl.org/dc/elements/1.1/>\nDELETE DATA\n{\nGRAPH
<http://example> {\n    <http://example/book1> dc:title \"A new book\" ;\n
dc:creator \"A.N.Other\" .\n    }\n}"
    },
    "List Explicit Context" : {
      "name" : "List Explicit Context",
      "body" : "select ?s ?p ?o\nfrom <http://www.ontotext.com/explicit>\nwhere {\n    ?s
?p ?o .\n} order by ?s \nlimit 100"
    },
    "Find by ContentModel and LastModified" : {
      "name" : "Find by ContentModel and LastModified",
      "body" : "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>\nPREFIX fedora:
<http://fedora.info/definitions/v4/repository#>\nPREFIX model:
<info:fedora/fedora-system:def/model#>\nselect ?s ?p ?date \nwhere {\n  ?s
fedora:lastModified ?date .\n  ?s model:hasModel \"IRItem\"\n  FILTER (?date >
\"2018-04-01T00:00:00.000Z\"^^xsd:dateTime && ?date <
\"2018-04-06T23:59:59.999Z\"^^xsd:dateTime)\n}"
    },
    "Clear graph" : {
      "name" : "Clear graph",
```

```
      "body" : "CLEAR GRAPH <http://example>"
    },
    "List All MineTypes" : {
      "name" : "List All MineTypes",
      "body" : "SELECT ?s ?p ?o \nWHERE {\n  ?s
\t<http://fedora.info/definitions/v4/repository#mimeType> ?o .\n}\nLIMIT 50\n"
    },
    "Count Triple by ContentModels" : {
      "name" : "Count Triple by ContentModels",
      "body" : "SELECT ?o (COUNT(*) AS ?count)\nWHERE {\n  ?s
<info:fedora/fedora-system:def/model#hasModel> ?o .\n}\nGROUP BY ?o\n"
    },
    "Delete Object by ContentModel" : {
      "name" : "Delete Object by ContentModel",
      "body" : "DELETE\nWHERE { \n  ?s <info:fedora/fedora-system:def/model#hasModel>
\"ActiveFedora::DirectContainer\"; \n  ?p ?o .\n}\n"
    }
  },
...
```

Modify Namespaces settings

- Need to stop GraphDB before editing the owlim.properties file
- The settings is in `/Users/pcharoen/Library/Application Support/`**`GraphDB/data/repositories/fedora/storage/owlim.properties`**. The namespaces setting is for each repository. It is not global settings.
- Copy Fedora namespaces and paste to owlim.properties

Namespaces

```
# GraphDB namespaces
Namespace wgs : http://www.w3.org/2003/01/geo/wgs84_pos#
Namespace owl : http://www.w3.org/2002/07/owl#
Namespace gn : http://www.geonames.org/ontology#
Namespace xsd : http://www.w3.org/2001/XMLSchema#
Namespace fn : http://www.w3.org/2005/xpath-functions#
Namespace rdfs : http://www.w3.org/2000/01/rdf-schema#
Namespace rdf : http://www.w3.org/1999/02/22-rdf-syntax-ns#
Namespace sesame : http://www.openrdf.org/schema/sesame#

# Fedora namespaces
Namespace fedora : http://fedora.info/definitions/v4/repository#
Namespace fedoramodel : info:fedora/fedora-system:def/model#
Namespace fedoraconfig : http://fedora.info/definitions/v4/config#
Namespace fedorawebac : http://fedora.info/definitions/v4/webac#
Namespace ldp : http://www.w3.org/ns/ldp#
Namespace acl : http://www.w3.org/ns/auth/acl#
Namespace mycombe : http://mycombe.library.ualberta.ca:8080/fedora/rest/
Namespace event : http://era.library.ualberta.ca/event/
```

```
Namespace bibo : http://purl.org/ontology/bibo/
Namespace cc : http://creativecommons.org/ns#
Namespace dc : http://purl.org/dc/elements/1.1/
Namespace dcterms : http://purl.org/dc/terms/
Namespace ebu : http://www.ebu.ch/metadata/ontologies/ebucore/ebucore#
Namespace etd_ms : http://www.ndltd.org/standards/metadata/etdms/1.0/
Namespace lang : http://id.loc.gov/vocabulary/iso639-2/
Namespace mrel : http://id.loc.gov/vocabulary/relators/
Namespace lcn : http://id.loc.gov/authorities/names/
Namespace obo : http://purl.obolibrary.org/obo/
Namespace ore : http://www.openarchives.org/ore/terms/
Namespace pcdm : http://pcdm.org/models#
Namespace prism : http://prismstandard.org/namespaces/basic/3.0/
Namespace schema : http://schema.org/
Namespace scholar : http://scholarsphere.psu.edu/ns#
Namespace skos : http://www.w3.org/2004/02/skos/core#
Namespace status : http://www.w3.org/2003/06/sw-vocab-status/ns#
Namespace swrc : http://ontoware.org/swrc/ontology#
Namespace ual : http://terms.library.ualberta.ca/
Namespace ualdate : http://terms.library.ualberta.ca/date/
Namespace ualid : http://terms.library.ualberta.ca/id/
Namespace ualids : http://terms.library.ualberta.ca/identifiers/
Namespace ualrole : http://terms.library.ualberta.ca/role/
Namespace ualthesis : http://terms.library.ualberta.ca/thesis/
Namespace works : http://pcdm.org/works#
Namespace vivo : http://vivoweb.org/ontology/core#
Namespace pcdmuse : http://pcdm.org/use#
Namespace hydramodels : http://projecthydra.org/works/models#
```

SPARQL Examples

- https://www.w3.org/2009/Talks/0615-qbe/

Select content model

```
SELECT ?s ?p ?o
WHERE {
  ?s  <info:fedora/fedora-system:def/model#hasModel> ?o .
  FILTER (?o = "IRItem")
}
LIMIT 50
```

Count triples

```
SELECT (count(*) as ?n)
WHERE {
  ?s ?p ?o .
}
```

Count by content models

```
SELECT ?o (COUNT(*) AS ?count)
WHERE {
  ?s <info:fedora/fedora-system:def/model#hasModel> ?o .
}
GROUP BY ?o
ORDER by ?o
```

List all predicates

```
SELECT ?p (COUNT(*) AS ?count)
WHERE {
   ?s ?p ?o .
}
GROUP BY ?p
ORDER by ?p
```

List fixity services

```
SELECT ?s ?p ?o
WHERE {
  ?s  <http://fedora.info/definitions/v4/repository#hasFixityService> ?o .
}
LIMIT 50
```

List versions

```
SELECT ?s ?p ?o
WHERE {
  ?s  <http://fedora.info/definitions/v4/repository#hasVersions> ?o .
}
LIMIT 50
```

List mimetypes

```
SELECT ?s ?p ?o
WHERE {
  ?s  <http://fedora.info/definitions/v4/repository#mimeType> ?o .
}
LIMIT 50
```

Find object permission

```
SELECT ?s ?p ?o
WHERE {
 ?s <http://www.w3.org/ns/auth/acl#accessTo>
<http://gillingham.library.ualberta.ca:8080/fedora/rest/prod/xw/42/n9/50/xw42n950r> .
}
```

Find using regex

```
select ?s ?p ?o
where {
  ?s ?p ?o .
  FILTER regex(str(?o), "alberta", "i")
}
```

Find objects

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX fedora: <http://fedora.info/definitions/v4/repository#>
SELECT ?s ?p ?o
FROM <http://www.ontotext.com/explicit>
WHERE {
    ?s dc:type "FedoraObject";
    ?p ?o .
}
ORDER BY ?s
LIMIT 100
```

model:hasModel "IRItem"

```
SELECT ?s ?p ?o
WHERE {
        ?s <info:fedora/fedora-system:def/model#hasModel> "IRItem";
        ?p ?o .
}
ORDER BY ?s
```

Find objects by lastModified

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select ?s ?p ?date
where {
  ?s <http://fedora.info/definitions/v4/repository#lastModified> ?date .
  FILTER (?date > "2018-04-13T00:00:00.000Z"^^xsd:dateTime && ?date <
"2018-04-13T23:59:59.999Z"^^xsd:dateTime)
}
```

Or

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select * {
  ?s <http://fedora.info/definitions/v4/repository#lastModified> ?date
  FILTER (?date > "2018-04-13T00:00:00.000Z"^^xsd:dateTime && ?date <
"2018-04-13T23:59:59.999Z"^^xsd:dateTime)
}
```

Find IRItem object by lastModified desc order by lastModified

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select ?s ?p ?date
where {
  ?s <http://fedora.info/definitions/v4/repository#lastModified> ?date .
  ?s <info:fedora/fedora-system:def/model#hasModel> "IRItem"
  FILTER (?date > "2018-04-01T00:00:00.000Z"^^xsd:dateTime && ?date <
"2018-04-06T23:59:59.999Z"^^xsd:dateTime)
}
order by desc(?date)
```

With all properties

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select ?s ?p ?o
where {
  ?s <http://fedora.info/definitions/v4/repository#lastModified> ?date .
  ?s <info:fedora/fedora-system:def/model#hasModel> "IRItem"; ?p ?o .
  FILTER (?date > "2018-04-01T00:00:00.000Z"^^xsd:dateTime && ?date <
"2018-04-04T23:59:59.999Z"^^xsd:dateTime)
}
```

Find Last Modified Objects in Audit

```
PREFIX premis: <http://www.loc.gov/premis/rdf/v1#>
select ?d ?o
where {
  ?s premis:hasEventDateTime ?d .
  ?s premis:hasEventRelatedObject ?o .
}
order by desc(?d) ?o
limit 100
```

Delete all triples command line

```
# curl --get -X DELETE -H 'Accept: application/xml'
'http://localhost:9999/blazegraph/namespace/${namespace}/sparql'
```

Delete all triples

```
DELETE WHERE { ?s ?p ?o }
```

Delete an object

```
DELETE
WHERE {
```

```
  <http://gillingham.library.ualberta.ca:8080/fedora/rest/prod/9p/29/0b/44/9p290b448> ?p
?o
}
```

Delete objects by contentModel

```
DELETE
WHERE {
  ?s <info:fedora/fedora-system:def/model#hasModel> "ActiveFedora::DirectContainer"; ?p ?o
.
}
```

Solr

Installation

- http://lucene.apache.org/solr/guide/7_3/installing-solr.html#installing-solr

Start / Stop

- `# /bin/solr start -force # login as root`
- `# /bin/solr stop`

Create / Delete Core from command line

- `# /bin/solr create -c fedora -force`
- `# /bin/solr delete -c fedora`

SSH Tunnel for User Interface

- `# ssh -f pcharoen@triplestore-test -L 8983:triplestore-test:8983 -N`

JDBC

- Apache Jena JDBC for Fuseki
- Github: https://github.com/ualbertalib/jena/tree/master/jena-jdbc
- Driver: https://github.com/ualbertalib/jena/releases

Make JDBC Driver Jar File with Dependencies

- Clone source code from Github
- `$ cd ./jena/jena-jdbc/jena-jdbc-bundle`
- `$ mvn clean package`

Fedora 4 (era-test) Connection URLs

- Fuseki connection for read / write
  - `jdbc:jena:remote:query=http://localhost:3030/data/query&update=http://localhost:3030/data/update`
- Blazegraph connection for read / write
```

○ jdbc:jena:remote:query=http://localhost:9999/blazegraph/namespace/fcrepo/spar ql&update=http://localhost:9999/blazegraph/namespace/fcrepo/sqarql

# Reindexing Triplestore

## 2017.12.07

- From: Thursday, December 7, 2017 at 3:43:00 pm
  To: Friday, December 8, 2017 at 8:15:00 am
  Result: 16 hours, 32 minutes and 0 seconds
- No of Triples

| n |
|---|
| 14536281 |

- No of Triples by ContentModel

| o | count |
|---|---|
| Collection | 541 |
| GenericFile | 44341 |
| Hydra::AccessControls::Permission | 142202 |
| Hydra::AccessControls::Embargo | 7505 |

- Failed
  - FcrepoTriplestoreIndexer: 4
  - FcrepoIndexer: 2
- Error Objects
  - /prod/b1/54/4b/p1/b1544bp15w
    - SAXParseException: An invalid XML character (Unicode: 0x1) was found in the element content of the document.
    - The error data is a value of dcterms:rights (base64).
  - /prod/c8/s4/5q/87/c8s45q876s
    - SAXParseException: An invalid XML character (Unicode: 0x1) was found in the element content of the document.
    - The error data is a value of dcterms:description (base64).

## 2017.02.02

- Jupiter test data from Gillingham2 14,528 triples

| o | count |
|---|---|
| IRItem | 20 |
| IRCollection | 235 |
| IRFileSet | 16 |
| ActiveFedora::IndirectContainer | 282 |
| ActiveFedora::DirectContainer | 16 |
| ActiveFedora::Aggregation::Proxy | 282 |

# Fedora 4.7

## Configuration Customization

### Add Namespaces

- https://wiki.duraspace.org/display/FEDORA4x/Best+Practices+-+RDF+Namespaces

### ActiveMQ Bridge for Camel Components

- https://wiki.duraspace.org/display/FEDORA474/Setup+Camel+Message+Integrations
- Scroll down to Supporting Queues section
- Modify **Fedora 4.7** embedded ActiveMQ configuration, /fedora/WEB-INF/classes/config/activemq.xml to forward fedora topic to fedora queue. Then, add fedora_bridge networkConnector to distribute topic and queue to external ActiveMQ Broker.

```
        <!-- The topic forwardings are for fcrepo-indexing-triplestore and
fcrepo-audit-triplestore camel components  -->
        <destinationInterceptors>
          <virtualDestinationInterceptor>
            <virtualDestinations>
              <compositeTopic name="fedora" forwardOnly="false">
                <forwardTo>
                  <queue physicalName="fcrepo.indexing.triplestore"/>
                  <queue physicalName="fcrepo.audit.triplestore"/>
                </forwardTo>
              </compositeTopic>
            </virtualDestinations>
          </virtualDestinationInterceptor>
        </destinationInterceptors>

        <!-- Distributed ActiveMQ Broker -->
        <networkConnectors>
          <networkConnector name="fedora_bridge" dynamicOnly="true"
uri="static:(tcp://${fcrepo.triplestore.activemq.broker})">
            <dynamicallyIncludedDestinations>
              <queue physicalName="fcrepo.indexing.triplestore"/>
              <queue physicalName="fcrepo.audit.triplestore"/>
              <topic physicalName="fedora"/>
            </dynamicallyIncludedDestinations>
          </networkConnector>
        </networkConnectors>
```

Java Opts

- Gillingham2: /etc/tomcat/conf.d/jvm_opts.conf

```
## Fedora 4 Configurations
FCREPO_HOME=/home/pcharoen/fedora_data
JAVA_OPTS="${JAVA_OPTS} -Dfcrepo.home=${FCREPO_HOME}"
JAVA_OPTS="${JAVA_OPTS} -Dfcrepo.log.directory=/var/log/tomcat7"
JAVA_OPTS="${JAVA_OPTS} -Dfcrepo.log.jcr=DEBUG"
JAVA_OPTS="${JAVA_OPTS} -Dfcrepo.log.oai=DEBUG"
JAVA_OPTS="${JAVA_OPTS} -Dfcrepo.log.maxHistory=10"
JAVA_OPTS="${JAVA_OPTS} -Dfcrepo.log.totalSizeCap=3G"


JAVA_OPTS="${JAVA_OPTS}
-Dfcrepo.modeshape.configuration=classpath:/config/file-simple/repository.json"
JAVA_OPTS="${JAVA_OPTS}
-Dfcrepo.modeshape.index.directory=${FCREPO_HOME}/fcrepo.index.directory"


# Parallel processing of streams can boost the retrieval speeds of RDF on a multiprocessor
machine.
JAVA_OPTS="${JAVA_OPTS} -Dfcrepo.streaming.parallel=true"


# Allow import/export tools to update triples
JAVA_OPTS="${JAVA_OPTS} -Dfcrepo.properties.management=relaxed"

## Saxon tranformer factory - XSLT 2.0
JAVA_OPTS="${JAVA_OPTS}
-Djava.xml.transform.TransformerFactory=net.sf.saxon.TransformerFactoryImpl"


## Triplestore ActiveMQ
JAVA_OPTS="${JAVA_OPTS}
-Dfcrepo.triplestore.activemq.broker=triplestore.library.ualberta.ca:61616"
```

Fedora 4 Import and Export Tools

- Document https://wiki.duraspace.org/display/FEDORA4x/Import+and+Export+Tools
- Data migration will need to set Java property,
  https://wiki.duraspace.org/pages/viewpage.action?pageId=87469300 to allow
  import-export-tools to update server managed triples.
  (-Dfcrepo.properties.management=relaxed)
- See this thread,
  https://groups.google.com/forum/#!searchin/fedora-tech/import$20problem%7Csort:date/fedor
  a-tech/mX0nuJrexfw/cde4PVgMAgAJ
- Download import/export tool from https://github.com/fcrepo4-labs/fcrepo-import-export
- Help command to see all import-export-tools options
  - $ java -jar fcrepo-import-export-0.3.0-SNAPSHOT.jar -h

Export Data

Export data from

With binaries

```
$ java -jar fcrepo-import-export-0.3.0-SNAPSHOT.jar --mode export --resource
http://gillingham2.library.ualberta.ca:8080/fedora/rest/prod -u fedoraAdmin:_gGv4_afB_
--dir ./data/ --binaries
```

Without binaries

```
$ java -jar fcrepo-import-export-0.3.0-SNAPSHOT.jar --mode export --resource
http://gillingham2.library.ualberta.ca:8080/fedora/rest/prod -u fedoraAdmin:_gGv4_afB_
--dir ./data_no_binaries/
```

Export Data with a Bagit Support

Default Profile

bagit-config.xml
```
bag-info.txt:
  Source-Organization: York University Libraries
  Organization-Address: 4700 Keele Street Toronto, Ontario M3J 1P3 Canada
  Contact-Name: Nick Ruest
  Contact-Phone: +14167362100
  Contact-Email: ruestn@yorku.ca
  External-Description: Sample bag exported from fcrepo
  External-Identifier: SAMPLE_001
  Bag-Group-Identifier: SAMPLE
  Internal-Sender-Identifier: SAMPLE_001
  Internal-Sender-Description: Sample bag exported from fcrepo
```

```
$ java -jar fcrepo-import-export-0.3.0-SNAPSHOT.jar --mode export --resource
http://gillingham2.library.ualberta.ca:8080/fedora/rest/prod -u fedoraAdmin:_gGv4_afB_
--dir data_bagit/ --binaries --bag-profile default --bag-config bagit-config.yml
```

Aptrust Profile

bagit-config-aptrust.yml
```
bag-info.txt:
  Source-Organization: York University Libraries
  Organization-Address: 4700 Keele Street Toronto, Ontario M3J 1P3 Canada
  Contact-Name: Nick Ruest
  Contact-Phone: +14167362100
  Contact-Email: ruestn@yorku.ca
  External-Description: Sample bag exported from fcrepo
  External-Identifier: SAMPLE_001
  Bag-Group-Identifier: SAMPLE
```

```
  Internal-Sender-Identifier: SAMPLE_001
  Internal-Sender-Description: Sample bag exported from fcrepo
aptrust-info.txt:
  Access: Restricted
  Title: Sample fcrepo bag
```

```
$ java -jar fcrepo-import-export-0.3.0-SNAPSHOT.jar --mode export --resource
http://gillingham2.library.ualberta.ca:8080/fedora/rest/prod -u fedoraAdmin:_gGv4_afB_
--dir data_bagit_aptrust/ --binaries --bag-profile aptrust --bag-config
bagit-config-aptrust.yml
```

Change data format to application/rdf+xml

See Format Options in the document for supported formats

```
$ java -jar fcrepo-import-export-0.3.0-SNAPSHOT.jar --mode export --resource
http://gillingham2.library.ualberta.ca:8080/fedora/rest/prod -u fedoraAdmin:_gGv4_afB_
--dir data_bagit_aptrust/ --binaries --bag-profile aptrust --bag-config
bagit-config-aptrust.yml -x .rdf -l application/rdf+xml
```

Import Data

Import data to http://localhost:8080
- Remove web application security block in web.xml to allow import tools writing to Fedora without authentication
- --map parameter maps export host URL to import host URL
- Set JAVA_OPTS=-Dfcrepo.properties.management=relaxed

(See
https://wiki.duraspace.org/display/FEDORA474/How+to+allow+user-updates+to+certain+server+managed+triples)

With binaries

```
$ java -jar fcrepo-import-export-0.3.0-SNAPSHOT.jar --mode import --resource
http://localhost:8080/fedora/rest/ --dir ./data/ --binaries --map
http://gillingham2.library.ualberta.ca:8080/fedora/rest/prod/,http://localhost:8080/fedora
/rest/prod/ -u fedoraAdmin:_gGv4_afB_
```

Without binaries

```
$ java -jar fcrepo-import-export-0.3.0-SNAPSHOT.jar --mode import --resource
http://localhost:8080/fedora/rest/ --dir ./data_no_binaries/ --map
http://gillingham2.library.ualberta.ca:8080/fedora/rest/prod/,http://localhost:8080/fedora
/rest/prod/ -u fedoraAdmin:_gGv4_afB_
```

Fixing Namespace ns00x

- Export from Fedora using import-export-tools
- Stop Tomcat server

- Add namespaces in spring configuration,
  https://wiki.duraspace.org/display/FEDORA475/Best+Practices+-+RDF+Namespaces
- Remove all data in fedora.home data directory
- Start Tomcat server
- Import data to Fedora using import-export-tools and exported data

Network monitoring

```
$ iftop -P
```

With OAI provider installed

- The OAI provider updates provider information to Fedora. The repository will generate JMS message message and send out to subscribers. This will result errors on camel routes, indexing-triplestore and indexing-solr.

ActiveMQ Server

Fedora JMS topic forwarding. See ActiveMQ Bridge for Camel Components

ActiveMQ

- `# cd /opt/activemq`
- `# bin/activemq start`
- `# bin/activemq stop`

Jolokia

JMX-HTTP bridge for Hawtio monitor system.
- Download jolokia java agent from
  http://search.maven.org/remotecontent?filepath=org/jolokia/jolokia-jvm/1.4.0/jolokia-jvm-1.4.0-agent.jar
- Use the script below to start jolokia java agent and attach to the ActiveMQ server.

  ```sh
  #!/bin/sh
  # jolokia
  # start/stop jolokia for activemq
  # ./jolokia [start/stop]
  export ACTIVEMQ_PID=`ps -ef | grep activemq | grep -v grep | awk '{print $2}'`
  java -jar /usr/share/activemq/bin/jolokia-jvm-1.3.7-agent.jar $1 $ACTIVEMQ_PID
  ```

- Start / Stop jolokia java agent, `./jolokia [start/stop]`

Start up sequence

Macbook test enveronment

- ActiveMQ
    - `# activemq start`
- Feodra 4, Tomcat server
    - `# tomcat7 start`
- GraphDB server
    - `Start GraphDB applicaton`
- Karaf server
    - `# /opt/karaf/bin/start`