

機関リポジトリのシステム構築*

— 60分でできる DSpace セットアップ —

佐野 広明 (京都大学附属図書館)

2006年11月21日

1 この講義の目的

これから機関リポジトリのシステムを立ち上げようと考えている方、あるいはDSpaceのインストールを試みたいと考えている方などを対象に、どのようなプロセスでDSpaceのセットアップが行なわれていくかについて、実際にデモをしながら紹介する。

2 DSpace とは

DSpace^{*1}は、マサチューセッツ工科大学 (MIT) 図書館とヒューレット・パカードが開発した、デジタルリポジトリシステムである。Registry of Open Access Repositories (ROAR)^{*2}によると、DSpaceの採用機関は179機関と、GNU EPrintsの213機関に次いで多い^{*3}。日本国内に限ると、最も多くの機関で採用されているソフトウェアである。

最新版は、2006年7月にリリースされたDSpace 1.4である。京都大学の機関リポジトリでは、2006年9月より、このバージョンを採用している。次期マイナーバージョンアップのDSpace 1.4.1は、2006年のクリスマス前にはリリースされる予定である^{*4}。

3 京都大学の機関リポジトリサーバ

URL: <http://repository.kulib.kyoto-u.ac.jp/dspace/>

機種: HP Proliant DL385

CPU: AMD Opteron 275 × 2

メモリ: 10GB

HDD: 72GB SAS × 8 (RAID6 + オンラインスペア)

OS: SUSE Linux Enterprise Server 10(64bit 版)

リポジトリソフト: DSpace 1.4 + 日本語化 + オリジナルパッチ

その他: バックアップサーバ (SUSE Linux 10.1)、テストサーバ (CentOS 4.4 + VMware Server)

* 平成 18 年度 NAIST 電子図書館学講座 (2006 年 11 月 21 日 14:25 ~ 15:25)

^{*1} <http://www.dspace.org/>

^{*2} <http://archives.eprints.org>

^{*3} 2006 年 11 月 12 日現在

^{*4} ベータ版は、2006 年 11 月 8 日にリリースされた

4 DSpace セットアップデモ

OS のインストールから DSpace のセットアップまでを実演する^{*5}。

なお、コマンドの実行例および設定ファイルは、以下のような枠内に記述する。

実行コマンド

設定ファイルの内容

4.1 今回のデモで使用する環境

機種: VMware Server 上の仮想マシン

CPU: Pentium M 778 (1.60GHz)

メモリ: 512MB

HDD: 8GB

4.2 OS のインストール

SUSE Linux 10.1^{*6}(x86 版) をインストールする。

4.2.1 インストール DVD の作成

1. SUSE Linux 10.1 の ISO イメージファイル (SUSE-Linux-10.1-GM-DVD-i386.iso) を http://ja.opensuse.org/Released_Version からダウンロードする。
2. ダウンロードした ISO イメージファイルを、DVD 書き込みソフトで DVD-R に焼く (Windows 上でもよい)。

4.2.2 インストール実行

作成した DVD-R から起動する。このとき、必要ならば、BIOS の設定して CD-ROM/DVD-ROM から起動できるようにする。

1. "Installation" を選択
2. "English(US)" を選択
3. License Agreement → "Yes, I Agree to the License Agreement" にチェック
4. Installation Mode → "New Installation" にチェック
5. Clock and Time Zone → Asia, Tokyo を選択。Hardware Clock Set To は、Local Time を選択
6. Desktop Selection → 今回は、GUI は使わないので、Other → Select で、Alternate Desktops → Text Mode を選択。
7. Installation Settings で、"Expert" タブをクリック。Keyboard Layout に注意。先に選択した言語と連動しているので、実際のキーボードレイアウトと異なる場合には、変更すること。

^{*5} 実際に京大の公開用サーバで行なわれた作業とは若干異なります。また、各ソフトウェアのバージョンアップなどにより状況が変わる可能性があります。最新情報は、参考文献 [1] をご覧ください

^{*6} 詳細は、<http://ja.opensuse.org> を参照のこと。

8. パーティションの設定。今回は、ひとつのパーティションと swap パーティションを設定する。Partitioning で、”Create Custom Partition Setup”を選択 (パーティションは、それぞれの方針で決定すること)
9. Software は、”Minimum System”。最低限のものだけ入れることにする。
10. Confirm Installation → ”Install”
11. Accept をクリックすると、インストールが始まる。Details タブをクリックすると、インストール経過がわかる。
12. 再起動。ここからは、フィールド間の移動は Tab キーで行なう。
13. Hostname and Domain Name → ここでは、test.site とする。
14. Password for the System Administrator ”root”。システム管理者 (root) のパスワード設定
15. Network Configuration。Firewall は無効にする (これは後で、それぞれのセキュリティポリシーに従って設定する必要がある)。
Proxy の設定 → 必要ならば後で設定するので、この時点ではスキップする。
16. Test Internet Connection → この時点でネットワークに接続していない場合には、”No, Skip This Test”でスキップする。
17. User Authentication Method → ユーザ認証方法の選択。通常は、”Local”を選択。
18. New Local User → 一般ユーザーアカウントの作成。ここでは、dspace というユーザーを作成。dspace というグループを作成して、そこに所属させる。デフォルトでは、users というグループがあり、新規にユーザーを作成するとそこに所属してしまう。ユーザー dspace は、グループ wheel にも所属させる。
19. Release Notes → ”Next”
20. Hardware Configuration → DSpace とは関係ないので、そのまま”Next”
21. Installation Completed → ”Finish”
22. ログイン画面が表示される。ユーザー “dspace” でログインできるか、確認する。

4.3 DSpace に必要な各種ソフトウェアのインストール

4.3.1 パッケージからインストールするもの

前準備 YaST で以下の 2 つを “install source” として設定する。

Protocol: HTTP

Server Name: ftp.novell.co.jp

Directory on Server: pub/suse/suse/update/10.1

Protocol: HTTP

Server Name: ftp.novell.co.jp

Directory on Server: pub/opensuse/distribution/SL-10.1/inst-source

インストールするパッケージ 以下を YaST の “Software Management” でインストールする (ここに挙がっていないパッケージも依存関係でインストールされる)。

- java-1.5.0-sun
- java-1.5.0-sun_devel
- ant

- gcc (各種コンパイルに必要)
- readline-devel
- zlib
- zlib-devel (PostgreSQL の configure に必要)
- ant-apache-regexp (DSpace 1.4 から、build.xml に正規表現が入ったため必要)
- wget (ソフトウェアのダウンロードに必要)
- patch (パッチを当てするのに必要)
- xpdf (日本語全文検索に必要)

4.3.2 ソースからインストールするもの

ソースの置き場を確認

```
> su -
# cd /usr/local/src
```

PostgreSQL のインストール

```
# wget ftp://ftp.sra.co.jp/pub/cmd/postgres/8.1.4/postgresql-8.1.4.tar.gz
# tar xvzf postgresql-8.1.4.tar.gz
# chown -R dspace.dspace postgresql-8.1.4
# mkdir /usr/local/pgsql
# chown -R dspace.dspace /usr/local/pgsql
# exit
> cd /usr/local/src/postgresql-8.1.4/
> ./configure
> make install
```

環境変数の設定 ~/.bashrc に以下を追加する。

```
export JAVA_HOME=/usr/lib/jvm/java
export CLASS_PATH=$JAVA_HOME/lib/tools.jar
export POSTGRES_HOME=/usr/local/pgsql
export PGLIB=$POSTGRES_HOME/lib
export PGDATA=$POSTGRES_HOME/data
export MANPATH=$POSTGRES_HOME/man
export LD_LIBRARY_PATH=$PGLIB
export PATH=$POSTGRES_HOME/bin:$JAVA_HOME/bin:$PATH
```

~/.bashrc を読み込む

```
> source ~/.bashrc
```

データベースの初期化

```
> initdb
```

セキュリティ向上のため、/usr/local/pgsql/data/pg_hba.conf に、

```
host →dspace→dspace→127.0.0.1/32 →md5
```

を追加する (→ は、タブを示す)。

postmaster の起動

```
> pg_ctl start
```

データベースの作成

```
> createdb -U dspace -E UNICODE dspace
```

Tomcat のインストール

```
> su -
# cd /usr/local/src
# wget http://ftp.kddilabs.jp/infosystems/apache/tomcat/tomcat-5/v5.5.20/bin/
  apache-tomcat-5.5.20.tar.gz
# tar xvzf apache-tomcat-5.5.20.tar.gz -C /usr/local
# chown -R dspace.dspace /usr/local/apache-tomcat-5.5.20
# exit
```

UTF-8 対応のため、/usr/local/apache-tomcat-5.5.20/conf/server.xml の以下の部分に、「URIEncoding="UTF-8"」を追加する。

```
<!-- Define a non-SSL HTTP/1.1 Connector on port 8080 -->
<Connector port="8080" maxHttpHeaderSize="8192"
maxThreads="150" minSpareThreads="25" maxSpareThreads="75"
enableLookups="false" redirectPort="8443" acceptCount="100"
connectionTimeout="20000" disableUploadTimeout="true"
URIEncoding="UTF-8" />
```

Tomcat 起動の確認

```
> cd /usr/local/apache-tomcat-5.5.20/bin
> ./startup.sh
```

Web ブラウザで、http://[ホスト名 (または IP アドレス)]:8080/にアクセスし、正常動作を確認したら、

```
> ./shutdown.sh
```

で一旦 Tomcat を終わらせておく。

4.4 DSpace のインストール

4.4.1 ソースのダウンロード

```
> su -
# cd /usr/local/src
```

```
# wget http://jaist.dl.sourceforge.net/sourceforge/dspace/dspace-1.4-source.tar.gz
# tar xvzf dspace-1.4-source.tar.gz
# chown -R dspace.dspace /usr/local/src/dspace-1.4-source
# exit
```

4.4.2 PostgreSQL JDBC ドライバのダウンロード

```
> cd /usr/local/src/dspace-1.4-source/lib
> wget http://jdbc.postgresql.org/download/postgresql-8.1-407.jdbc3.jar
```

4.4.3 インストール先ディレクトリの準備

```
> su -
# mkdir /usr/local/dspace
# chown dspace.dspace /usr/local/dspace
# exit
```

4.4.4 設定ファイル dspace.cfg の修正

```
> cd /usr/local/src/dspace-1.4-source/config
> vi dspace.cfg
```

```
dspace.dir = /usr/local/dspace
dspace.url = http://repository.hoge.ac.jp:8080/dspace
dspace.hostname = repository.hoge.ac.jp
mail.server = sendmail.hoge.ac.jp
mail.from.address = hoge@hoge.ac.jp
feedback.recipient = hoge@hoge.ac.jp
mail.admin = hoge@hoge.ac.jp
```

dspace.dir DSpace のインストールディレクトリ (このデモでは、/usr/local/dspace)

dspace.url http://[ホスト名]:8080/dspace

dspace.hostname ホスト名

mail.server SMTP サーバ名 (DSpace をインストールしたサーバ以外のもの (学内の SMTP サーバなど) を指定するのが無難)

mail.from.address DSpace からのメールの送信者となるアドレス

feedback.recipient フィードバックフォームからメッセージを受け取るアドレス

mail.admin 管理者のメールアドレス。DSpace 1.4.1 では表示させないようにできる。

4.4.5 DSpace のインストール

```
> cd /usr/local/src/dspace-1.4-source
> ant fresh_install
> cp build/*.war /usr/local/apache-tomcat-5.5.20/webapps
```

4.4.6 初期管理者の設定

```
> /usr/local/dspace/bin/create-administrator
```

4.4.7 起動確認

```
> /usr/local/apache-tomcat-5.5.20/bin/startup.sh
```

としてから、Web ブラウザで、`http://[ホスト名 (または IP アドレス)]:8080/dspace/` にアクセスしてみる。正常に動作しているのを確認したら、

```
> /usr/local/apache-tomcat-5.5.20/bin/shutdown.sh
```

として、Tomcat を終了させる。

4.5 DSpace の日本語化

DSpace の日本語化については、基本的に参考文献 [2] に基づいている。

4.5.1 ユーザーインターフェースの日本語化

メッセージファイルの日本語化

```
> cd /usr/local/src/dspace/config/language-packs
> wget http://dspace.cvs.sourceforge.net/*checkout*/dspace/language-packs/ja/
  Messages_ja.properties?revision=1.2
> mv Messages_ja.properties?revision=1.2 Messages_ja.properties
```

この `Messages_ja.properties` は、

```
> native2ascii -reverse Messages_ja.properties Messages_ja.properties.utf
```

とすると、UTF-8 のファイルに変換できるので、エディタなどで修正してから、

```
> native2ascii -encoding UTF-8 Messages_ja.properties.utf Messages_ja.properties
```

とすれば、また変換できる。

4.5.2 日本語メタデータの検索

Lucene: Java で記述された全文検索エンジンのライブラリ

Sen: Lucene で利用できる形態素解析エンジン

Sen のインストール

```
> su -
# cd /usr/local
# wget --no-check-certificate https://sen.dev.java.net/files/documents
  /1373/31864/sen-1.2.2.1.zip
# unzip sen-1.2.2.1.zip
# chown -R dspace.dspace sen-1.2.2.1
# exit
> cd /usr/local/sen-1.2.2.1/conf
> vi sen.xml
> vi sen-processor.xml
```

(wget のオプション “-no-check-certificate” は、ダウンロードできない場合に付けてみてください。)
sen.xml と sen-processor.xml は、“< charset>euc-jp</ charset>” 部分を、

```
< charset>utf-8</ charset>
```

に修正する。

変換辞書の作成

```
> cd /usr/local/sen-1.2.2.1/dic
> vi dictionary.properties
> wget http://chasen.aist-nara.ac.jp/stable/ipadic/ipadic-2.6.0.tar.gz
> ant
```

dictionary.properties は、“sen.charset=EUC-JP” 部分を、

```
sen.charset=utf-8
```

に修正する。

Lucene-ja のインストール

```
> su -
# cd /usr/local
# wget --no-check-certificate https://sen.dev.java.net/files/documents
  /1373/11260/lucene-ja-1.4.3sen1.2-2.zip
# unzip lucene-ja-1.4.3sen1.2-2.zip
# chown -R dspace.dspace lucene-ja
# exit
```

(wget のオプション “-no-check-certificate” は、ダウンロードできない場合に付けてみてください。)

検索に使用する品詞を決める

```
> cd /usr/local/lucene-ja/lib
> jar xvf lucene-ja.jar
> cd org/apache/lucene/analysis/ja/
> vi analyzer-sen.xml
```


ここで、<accept> 要素の下の <pos> 要素で検索に使用しない要素をコメント化 (<!--> で囲む) する。

```
> cd /usr/local/lucene-ja/lib
> mv lucene-ja.jar lucene-ja.jar.org
> jar cvf lucene-ja.jar org
```

ライブラリを DSpace ソースにコピーする

```
> cd /usr/local/lucene-ja/lib
> cp lucene-ja.jar sen.jar commons-logging.jar /usr/local/src/dspace-1.4-source/
lib
```

DSAnalyzer の設定

/usr/local/dspace/config/dspace.cfg に以下を追加する。

```
# Japanese analyzer
search.analyzer = org.apache.lucene.analysis.ja.JapaneseAnalyzer
```

/usr/local/dspace/bin/dsrun の最後の行を以下のように変更する。

```
java -Xmx256m -classpath $FULLPATH -Dsen.home=/usr/local/sen-1.2.2.1 "$@"
```

~/.bashrc に環境変数 JAVA_OPTS を追加

```
export JAVA_OPTS="-Xmx512M -Xms64M -Dfile.encoding=UTF-8 -Dsen.home=/usr/
local/sen-1.2.2.1"
```

```
> source ~/.bashrc
```

DSpace の更新

```
> cd /usr/local/src/dspace-1.4-source
> ant clean
> ant -Dconfig=/usr/local/dspace/config/dspace.cfg update
> cp build/*.war /usr/local/apache-tomcat-5.5.20/webapps
> rm -r /usr/local/apache-tomcat-5.5.20/webapps/dspace
> rm -r /usr/local/apache-tomcat-5.5.20/webapps/dspace-oai
```

4.5.3 全文検索の日本語対応

PDF や MS Word ファイルなどに含まれる日本語の全文検索を実現する。

mediafilter へのパッチの適用

```
> cd /usr/local/src/dspace-1.4-source/src/org/dspace/app/mediafilter
> wget http://www12.ocn.ne.jp/~zuki/Japanization/dspace14/mediafilter.patch
> patch < mediafilter.patch
```

BitstreamStorageManager.java へのパッチの適用

```
> cd /usr/local/src/dspace-1.4-source/src/org/dspace/storage/bitstore
> wget http://www12.ocn.ne.jp/~zuki/Japanization/dspace14/
  BitstreamStorageManager.patch
> patch < BitstreamStorageManager.patch
```

pdftotext(xpdfに含まれる)のインストール先をDSpaceに知らせる
/usr/local/dspace/config/dspace.cfgの最後に以下を追加し、DSpaceを更新する。

```
pdftotext.path = /usr/bin/pdftotext
```

DSpaceのコンテンツ索引を再作成する

```
> /usr/local/dspace/bin/filter-media -f
```

4.6 DSpaceのカスタマイズ例

4.6.1 ニュース編集

DSpaceの管理者で、http://[ホスト名]:8080/dspace/dspace-admin/)にログインすると管理者メニューが利用できる。サイドメニューに「ニュース編集」があるので、それを選択することによって、「トップ・ニュース」「サイドバー・ニュース」を修正することができる。HTMLタグが使用でき、日本語も入力可能である。

4.6.2 JuNii2への対応

DSpaceのOAIプロバイダ機能として、新NIIメタデータ記述要素(JuNii2)でのメタデータ出力を追加するには、/usr/local/src/dspace-1.4-source/src/org/dspace/app/oai/JUNIICrosswalk.javaを作成する。このプログラムによって、どういった項目を出力させるかを指定できる(このプログラムの参考例を、参考文献[1]に置く予定)。

/usr/local/dspace/config/oaicat.propertiesに、以下を追加する。

```
Crosswalks.junii=org.dspace.app.oai.JUNIICrosswalk
```

また、NIIの参加組織IDなどを設定ファイルに保存する場合には、/usr/local/dspace/config/dspace.cfgに以下を追加する。

```
#junii oai configuration
junii.userid = A*****
junii.fano = FA*****
junii.institution = *****
```

“A*****”は利用者番号、“FA*****”は参加組織コード。junii.institutionは、日本語(漢字)の場合、native2asciiコマンドでエンコードする。

4.6.3 データの一括投入

既にメタデータがタブ区切りテキストなどの形である場合、一括してインポートすることができる。京大では、貴重資料画像のコレクションがその例。

必要なデータは、以下のように構成する。

```
kichosho/
  1/
    contents
    dublin_core.xml
    コンテンツファイル (PDF など)
  2/
    contents
    dublin_core.xml
    コンテンツファイル (PDF など)
  ...
```

“contents” は、1 行に 1 ファイル名を記したテキストファイル。メタデータのみをインポートをしたい場合には、このファイルを空にすれば良い。

dublin_core.xml として、

```
<dublin_core>
<dcvalue element='title' qualifier='none'>國女歌舞妓繪詞</dcvalue>
<dcvalue element='title' qualifier='transcription'>クニジョカブキエコトバ
  </dcvalue>
<dcvalue element='creator' qualifier='none'>京都大学附属図書館</dcvalue>
<dcvalue element='creator' qualifier='transcription'>キョウト ダイガクフゾクト
  ショカン</dcvalue>
..... 中略.....
<dcvalue element='language' qualifier='iso'>jpn</dcvalue>
<dcvalue element='relation' qualifier='ispartof'>http://ddb.libnet.kulib.
  kyoto-u.ac.jp/minds.html</dcvalue>
</dublin_core>
```

のような XML ファイルを作成する。

必要データが整ったら、以下のコマンドをユーザー “dspace” で実行する。

```
> /usr/local/dspace/bin/dsrun org.dspace.app.itemimport.ItemImport --add --
  eperson=hoge@hoge.ac.jp --collection=123456789/2 --source=/home/dspace/
  kichosho --mapfile=/home/dspace/mapfile
```

“-mapfile” は、インポート実行時に作成されるアイテム ID のリスト。

4.6.4 ログの変更

例として、「電子図書館大学 学術成果コレクション」というロゴ画像を作成し、変更してみる。

/usr/local/src/dspace-1.4-source/jsp 以下のファイルを、/usr/local/src/dspace-1.4-source/jsp/local 以下に同じ構造で作成する。

```
> cd /usr/local/src/dspace-1.4-source/jsp
> mkdir local
> cd local
> mkdir layout
```

```
> mkdir image
> cd layout
> cp /usr/local/src/dspace-1.4-source/jsp/layout/header-default.jsp .
```

その上で、`/usr/local/src/dspace-1.4-source/jsp/local/layout/header-default.jsp` の以下の部分を修正する。

```
<a href="<%= request.getContextPath() %>/">" width="198" height="79" border="0"/></a
></td>
<td class="tagLine" width="99%"> <!-- Make as wide as possible. cellpadding
  repeated for broken NS 4.x --%>
<a class="tagLineText" target="_blank" href="http://www.dspace.org/"><fmt:
  message key="jsp.layout.header-default.about"/></a>
```

上記の、`dspace-blue.gif` をロゴの画像ファイル名に変更。width, height をそれに合わせる。ロゴの画像ファイルは、`/usr/local/src/dspace-1.4-source/jsp/local/image/` 以下に置く。

4.6.5 その他のカスタマイズ

- Apache との連携
- リンクを別ウィンドウで開くようにする
- 簡略表示項目の変更
- その他 ...

参考文献

- [1] DSpace メモ [URL:<http://repository.kulib.kyoto-u.ac.jp/dspace-memo/>]
- [2] 機関リポジトリソフトウェアの日本語対応 [URL: <http://www12.ocn.ne.jp/~zuki/Japanization/index.html>]
- [3] 学術機関リポジトリ構築ソフトウェア実装実験プロジェクト報告書 [URL:<http://www.nii.ac.jp/metadata/irp/NII-IRPreport.pdf>]
- [4] Dspace インストール日記 [URL:http://d.hatena.ne.jp/ocha_repo/]
- [5] DSpace How-To Guide [URL: <http://hdl.handle.net/1920/1045>]
- [6] Making DSpace Your Own [URL: <http://hdl.handle.net/1920/1046>]
- [7] DSpace Installation and Systems Administration Guide [URL: <http://www.ub.uib.no/bora/tekndok/DSpace%20Installation%20v4.0.pdf>]