

EPrints Installation

- EPrints successfully runs on
 - Linux
 - Solaris
 - Mac OS-X
- Grant from Microsoft for Windows version
- Installation process is standardised
- One site installation can run many separate repositories
 - mixing and matching EPrints and other Web-based services on the same host is possible

eprints

EPrints - the Administrator's View

- SQL database
- Web server
- Scripts to configure repository activities
- Configuration files

eprints

EPrints Home Directory

8 items, Free space: 60.6 GB

eprints

EPrints Home Directory

Global configuration directory for the running EPrints server

8 items, Free space: 60.6 GB

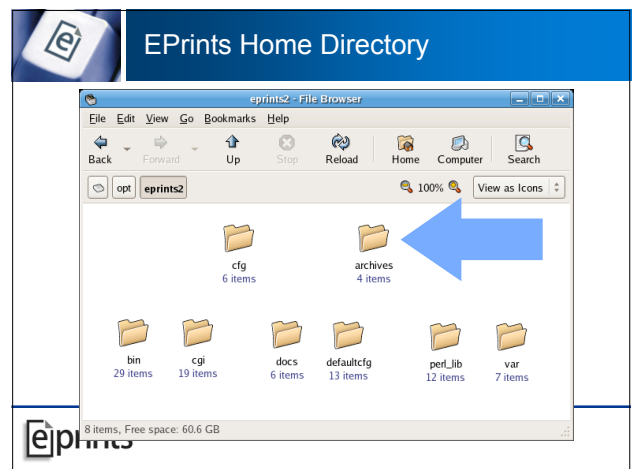
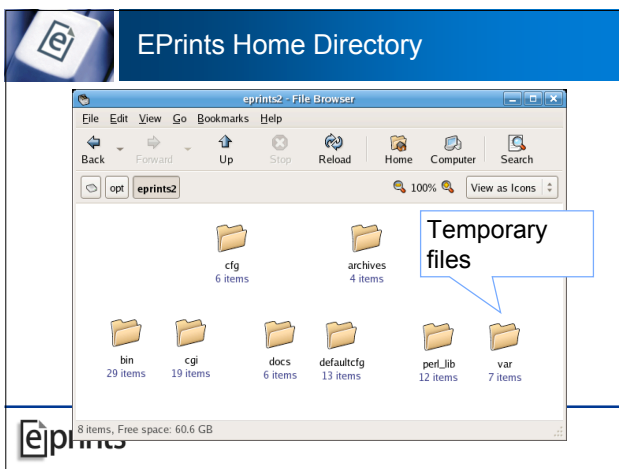
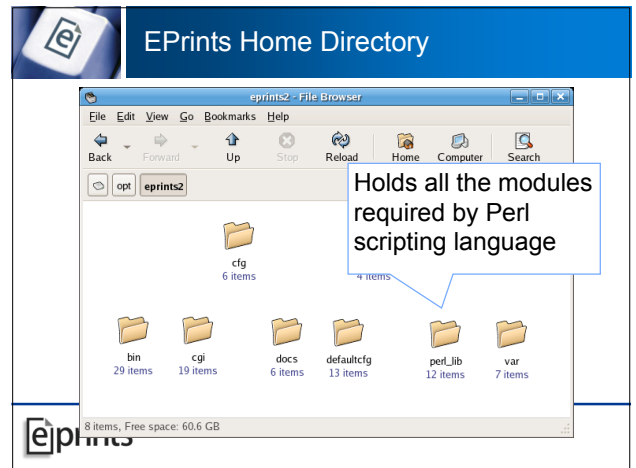
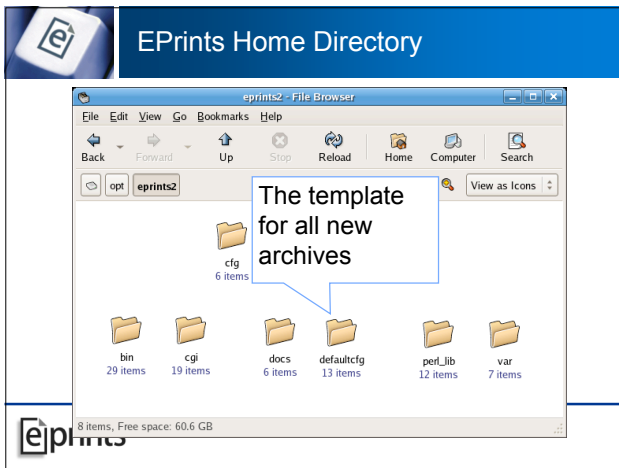
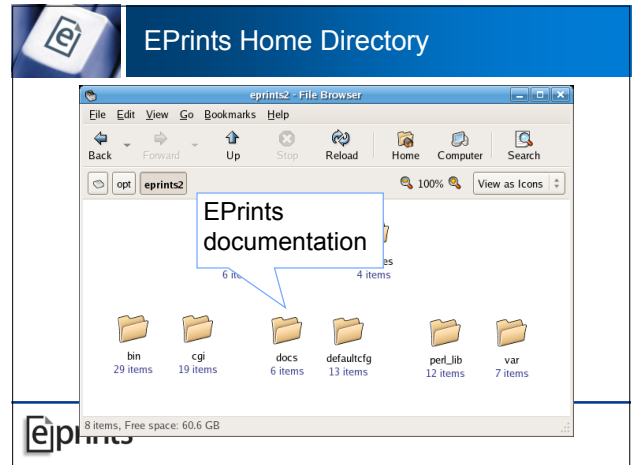
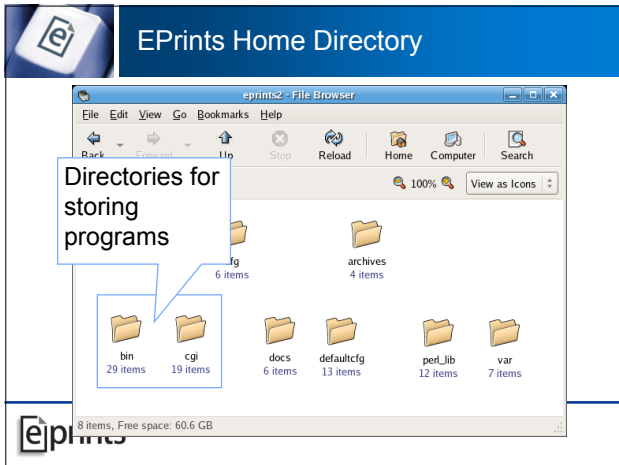
eprints

EPrints Home Directory

Holds the config subdirectories for each active archive (ie repository)

8 items, Free space: 60.6 GB

eprints



Contents of archives Directory

Contents of archives Directory

Contents of archives Directory

Contents of individual archive directory

Contents of individual archive directory

Contents of individual archive directory

Contents of individual archive directory

Temporary files

4 items, Free space: 60.6 GB

eprints

Contents of individual archive directory

4 items, Free space: 60.6 GB

eprints

Contents of individual cfg directory

18 items, Free space: 60.6 GB

eprints

Contents of individual config directory (2)

- ...there are LOTS of configuration files!
 - XML, DTD, Perl modules, apache...
- Rather than examine each individually, consider some common configuration tasks
 - Branding
 - Adding a new deposit type
 - Adding a new metadata field
- One step at a time
 - examining cause and effect
 - not how you would normally do things!

eprints

Task 1: Branding

The first thing most institutions do is brand their repository and fit it in with their existing look and feel

Southampton Repository

Home About Browse Search Register User Area Help

Welcome to Southampton Repository

Welcome to this generic eprint server, running the default configuration for [GNU EPrints repository software](#).

[More information](#) is available about the repository.

Search items:

Search full text:

eprints

Branding: Which Configuration Files?

template-en.xml
Site-wide HTML template

18 items, Free space: 60.6 GB

Branding: Which Configuration Files?

static/en/* .xpage
fixed content pages e.g.
homepage, about page,
help page
static/general/
images and stylesheets

Branding: Which Configuration Files?

entities-en.dtd
useful symbolic names
e.g. archivename,
adminemail

Branding: **template-en.xml**

- EPrints lets you define an HTML template (outline) which is used to build every Web page
- Customise the look and feel of the whole site
 - header
 - title, logo, navigation menu
 - footer
- pins** tell EPrints where on the page the title and page contents should be placed

Aside: What is XML?

- Quick answer: it's a bit like HTML
 - HTML is for making pages for people to read
 - XML is for making data for computers to use
 - The syntax is very similar, just stricter
 - All tags must have a matching closing tag
 - All attributes must be quoted
 - It doesn't know anything about Web pages
 - or anything else, come to that!

Branding: **static** directory

- You can also define fixed content for:
 - homepage, "about" page, help pages, "error" page...
- These configuration files are stored in the **static** directory
 - one subdirectory per language
 - e.g. english files go in **static/en**
- Images and stylesheets (and other language independent files) are stored in **static/general**

Aside: **static/en/index.xpage**

```
<?xml version="1.0" encoding="iso-8859-1" standalone="no" ?>
<!DOCTYPE page SYSTEM "entities-en.dtd" >
<ep:page
  xmlns="http://www.w3.org/1999/xhtml"
  xmlns:ep="http://eprints.org/ep2/page">
  <ep:title>Welcome to &archivename;</ep:title>
  <ep:body>

  <p>Welcome to this generic eprint server, running the default
  configuration for <a href="http://software.eprints.org/">GNU
  EPrints repository software</a>.</p>


  </ep:body>
</ep:page>
```

Southampton Repository

Welcome to Southampton Repository


Welcome to this generic eprint server, running the default configuration for [GNU EPrints repository software](http://software.eprints.org/).

[More information](#) is available about the repository.





Branding: Add a Logo

- Add the University of Southampton logo
 - first copy **logo.gif** into **static/general/images**
 - then add the logo to the header in **template-en.xml**




```
<div class="header noPrint">
  <table cellpadding="5" border="0" cellspacing="0" width="100%">
    <tr>
      <td>
        <!-- <div class="archivetitle">&archivename;</div> -->
        
      </td>
    </tr>
  </table>
</div>
```






Aside: entities-en.dtd


- Notice the entities in **template-en.xml**
 - &archivename;
 - &base_url;
- These are defined in **entities-en.dtd**
 - generated automatically by EPrints
 - definitions of character symbols e.g. copyright
 - contains useful symbolic names for various URLs and email addresses
 - lets you avoid hard-coding names and URLs






Aside: what is a DTD?

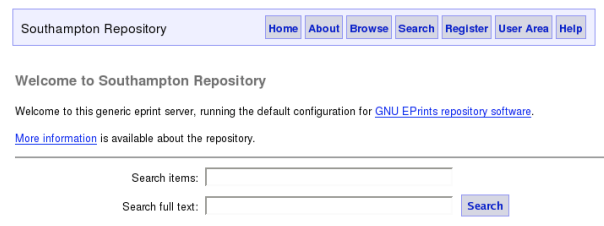

- A DTD is a definition file for XML
 - XML is just a naked standard for the syntax rules of a document or data file
 - A DTD provides it with the definitions needed for a particular vocabulary (e.g. HTML)
 - It also defines names for non-ASCII characters (e.g. copyright, euro, bullet)






Branding: Check the Homepage


- Check the homepage... no logo!







Branding: generate_static Command


- We need to run the **generate_static** command
- This takes the fixed content (**.xpage**) files in the **static** directory and wraps them in the template
- The resulting HTML pages are written to the repository's **html** directory
 - **myarchive/cfg/static/en/index.xpage**
 - becomes
 - **myarchive/html/en/index.html**
- Images and stylesheets copied across as well





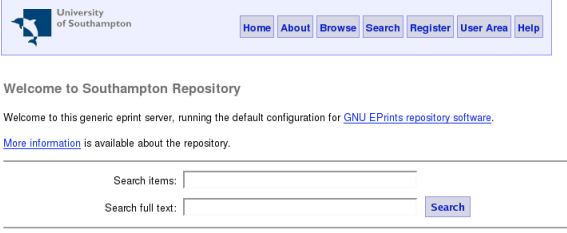
Branding: generate_static Command (2)

- Why?
 - these pages hardly ever change (hence "static")
 - for best server performance, serve static html pages
 - but, want to maintain "master" site template in single file
- **3.0** has hybrid approach:
 - **pin** dynamic bits of content onto a static page
 - e.g. login status (logged off, logged on as ...)



Branding: Check the Homepage

After running **generate_static**:



University of Southampton

Home About Browse Search Register User Area Help

Welcome to Southampton Repository

Welcome to this generic eprint server, running the default configuration for [GNU EPrints repository software](#).

[More information](#) is available about the repository.


Search items:

Search full text:

eprints

Branding: Check the View Pages

But the logo isn't showing up on the browse view pages!



Southampton Repository

Home About Browse Search Register User Area Help

Browse Items

Items may be browsed by the following:

- [Year](#)
- [Subject](#)

[Contact Information](#)

eprints

Branding: **generate_views** Command

- The browse view pages change much more frequently than the homepage *etc.*
 - but EPrints also serves these as static HTML pages for performance
 - often visited by crawlers e.g. Google
- To regenerate the view pages, we need to run **generate_views**
 - this is usually run nightly, or even hourly

eprints

Branding: Check the View Pages

After running **generate_views**:



University of Southampton

Home About Browse Search Register User Area Help

Browse Items

Items may be browsed by the following:

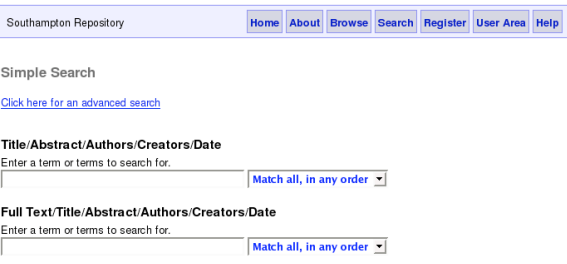
- [Year](#)
- [Subject](#)

[Contact Information](#)

eprints

Branding: Check the Search Page

But the logo isn't showing up on the search pages!



Southampton Repository

Home About Browse Search Register User Area Help

Simple Search

[Click here for an advanced search](#)

Title/Abstract/Authors/Creators/Date

Enter a term or terms to search for.

Full Text/Title/Abstract/Authors/Creators/Date

Enter a term or terms to search for.

eprints

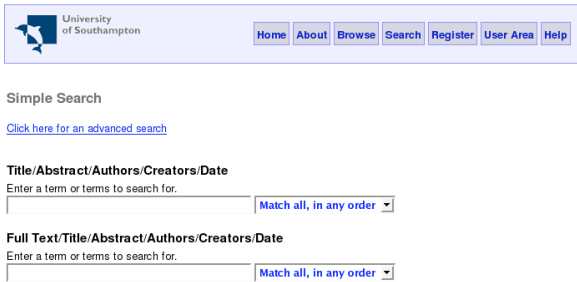
Branding: **force_config_reload** Command

- The search pages, and also user home page, deposit pages *etc.* are *dynamic*
 - created on-demand by EPrints
- For best performance, EPrints loads the template into memory at startup
 - dynamic pages are wrapped in this in-memory copy
 - so when we change the template, we need to get EPrints to refresh its copy
 - run **force_config_reload** or restart the Web server

eprints

Branding: Check the Search Page

After restarting the Web server:



Branding: Summary

- Copy logo image to **static/general/** directory
- Add logo to **template-en.xml**
- Regenerate static pages:
 - generate_static**
 - generate_views**
 - generate_abstracts**
- force_config_reload** to show logo on dynamic pages

Task 2: Adding a Deposit Type

- EPrints is pre-configured with several default deposit types
 - Article, Book, Book Section, Conference Item, Monograph, Other, Patent, Thesis
 - modeled on most common research outputs
- Each deposit type has a set of metadata associated with it
 - title, creators, editors, date of publication, abstract...
- And a set of document formats
 - PDF, PostScript, HTML, plain text

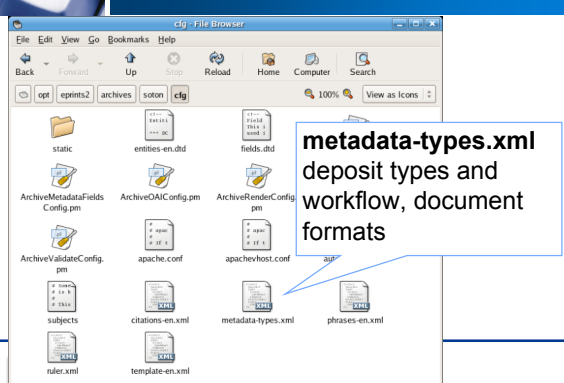
Task 2: Adding a Deposit Type

- Many institutions have other types of (research) output or collections/artefacts
 - data, teaching materials, multimedia
- e.g. the University of Southampton has:
 - a School of Art
 - a Textile Conservation Centre
 - a Music division in the School of Humanities
- What kinds of deposit might be needed?
 - metadata fields? document formats?

Some Suggestions

- New Deposit Types:
 - Composition, Performance, Show/Exhibition, Artefact
- Metadata:
 - composers, conductor, medium (oil, pencil, ink, watercolour, gouache, marble, clay, scrap metal...), producer, sound engineer, commissioning body, creation dates, venues/dates, genre (opera, jazz...)
- Document formats
 - image (JPG, TIFF...), audio (MP3, WAV, FLAC...), 3D model (?)

New Deposit Type: Which Config Files?



New Deposit Type: Which Config Files?

phrases-en.xml
display names for deposit types and document formats

New Deposit Type: Which Config Files?

citations-en.xml
citation styles for deposit types and document formats

New Deposit Type: Which Config Files?

ArchiveConfig.pm
specify list of required upload formats for each document type

New Deposit Type: metadata-types.xml

- The configuration file which describes the deposit types is **metadata-types.xml**
 - different types of **eprint** (deposit type), **user** (users, editors, administrators) and **document** (PDF, PS..)
 - defines the metadata fields that apply to each type
 - defines the order that the fields will appear in the deposit workflow
 - defines how the fields will be grouped into pages in the deposit workflow
- 3.0** adds conditionals to workflow
 - e.g. different workflows for different departments

eprints

Add a New Deposit Type

- Add a simple *Composition* deposit type
 - use existing metadata fields for now
- Restart Web server to re-read configuration files

```
<type name="composition">
  <page name="core" />
  <field name="title" required="yes" />
  <field name="creators" required="yes" />

  <page name="abstract" />
  <field name="abstract" />
  <field name="keywords" />
</type>
```

eprints

New Deposit Type: Check List of Types

- Begin a new deposit
 - the text for the Composition option looks strange!

Item Type *
Please select the most appropriate type for your deposit.


["eprint_type_name_composition" not defined]
["eprint_optdetails_type_composition" not defined]

Article
An article in a journal, magazine, newspaper. Not necessarily peer-reviewed as an online journal or news website.

eprints

New Deposit Type: **phrases-en.xml**

- EPrints needs to know how to display the type
- The **phrases-en.xml** configuration file is where *all* the phrases which appear in the EPrints Web interface are defined
- Each **ep:phrase** element has a **ref (id)**
 - often structured: **eprint_fieldname_abstract**
- Why?
 - phrases are not embedded in EPrints code
 - single file for editing phrases
 - referring to phrases by id enables multi-language support



New Deposit Type: Add Phrases

- Add phrases for the Composition deposit type

```
<ep:phrase ref="epprint_type_name_composition">Composition </ep:phrase>
<ep:phrase ref="epprint_opt_details_type_composition">A musical composition.</ep:phrase>
```

- Restart Web server (reloads all config files)

Item Type *
Please select the most appropriate type for your deposit.

☐ **Composition**
A musical composition.

☐ **Article**
An article in a journal, magazine, newspaper. Not necessarily peer-reviewed as an online journal or news website.


New Deposit Type: Check Citation

- As you work through the deposit process, EPrints displays the "citation" at the top of the screen
 - this shows you how the citation will appear on other pages
- For our new deposit type, we get an error

Core Bibliographic Information

Current item: **Error: Unknown Citation Style "epprint_composition"**

Please enter the information about this item. Fields marked with a * are fields that will be accepted.



New Deposit Type: **citations-en.xml**

- The citation style for each deposit type is defined in the **citations-en.xml** configuration file
- Very powerful and flexible but a bit hard to read
- Add citation style for Composition and restart

```
<ep:citation type="epprint_composition">
  <span class="citation">
    @creators@
    <ep:linkhere><em>@title;magicstop@</em></ep:linkhere>
    <ep:ifset name="keywords"> (@keywords@)</ep:ifset>
  </span>
</ep:citation>
```

New Deposit Type: Check Citation

Abstract and References

Current item: **Norris, David Owen Die! Sober Flirt.**

Please enter the information about this item. Fields marked with a * are fields that must be filled out before your deposit will be accepted.


[< Previous](#) [Save for Later](#) [Next >](#)

Abstract
A summary of the items content. If the item has a formal abstract then that is what should be entered here. No complicated text formatting is possible.

Keywords being entered...

Uncontrolled Keywords
Natural language terms to describe the content of the item

[< Previous](#) [Save for Later](#) [Next >](#)



New Deposit Type: Check Citation

Edit item: Documents Attached to Item

Current item: **Norris, David Owen Die! Sober Flirt. (obbligato, Mozart, Constanze, Sallert)**


Please enter the information about this item. Fields marked with a * are fields that must be filled out before your deposit will be accepted.

[< Previous](#) [Save for Later](#)

You must upload at least one of the following formats: HTML, PDF, Postscript, ASCII.

[Add a Format to this Item](#)

- Citation OK
- But default deposit formats not helpful!




New Deposit Type: Document Types

- Adding extra document types is a similar process to adding a new deposit type
 - add extra formats to **metadata-types.xml**
 - add phrases to **phrases-en.xml**
 - `document_type_name_mp3`
 - `document_type_name_wav`
 - also need citations
 - Can now deposit MP3/WAV
 - but also need to configure *required document formats* for Compositions

Please select the storage format


HTML
PDF
Postscript
ASCII
MP3
WAV
Other

You must upload at least one of the following formats: HTML, PDF, Postscript, ASCII.

 [Add a Format to this Item](#)

New Deposit Type: ArchiveConfig.pm

- Required document upload formats is just one of the many settings in the **ArchiveConfig.pm** configuration file
- Perl syntax, but easy to change simple things
 - skip submission buffer
 - web signup for depositing users
 - metadata input defaults
 - submission form customisation
 - definition of browse views, search forms and user privileges




New Deposit Type: Add Formats

- Add new document types to the list of required formats
 - restart Web server
 - also possible to define a list of required formats *for each deposit type*
 - more complicated

```
$c->{required_formats} = [
    "html",
    "pdf",
    "ps",
    "ascii",
    "mp3",
    "wav",
];
```

You must upload at least one of the following formats: HTML, PDF, Postscript, ASCII, MP3, WAV.

 [Add a Format to this Item](#)

New Deposit Type: Test Deposit

University of Southampton

[Home](#) [About](#) [Browse](#) [Search](#) [Register](#) [User Area](#) [Help](#)


Die! Sober Flirter

Norris, David Owen *Die! Sober Flirter*. (obligato, Mozart, Constanze, Salieri)

Full text available as:


MP3 - Requires a player such as [Windows Media Player](#)
1506 Kb

Item Type:	Composition
Uncontrolled Keywords:	obligato, Mozart, Constanze, Salieri
Subjects:	Subjects UNSPECIFIED
ID Code:	2
Deposited By:	User admin
Deposited On:	13 July 2006

 http://www.soton.ac.uk/music/news/2006_06_12.shtml


New Deposit Type: Summary

- ↔ Add type definition and any extra document formats to **metadata-types.xml**
- ↕ Add phrases for type and formats to **phrases-en.xml**
- ↔ Add citations for type and formats to **citations-en.xml**
- Specify which formats are required for each type in **ArchiveConfig.pm**
- ↕ **force_config_reload**



Task 3: Add a New Metadata Field

- Continuing our theme, add an extra field to the Composition type called **composition_genre**



New Metadata Field: Which Config Files?

metadata-types.xml
which fields apply to which types

New Metadata Field: Which Config Files?

ArchiveMetadataFieldsConfig.pm
defines type and properties of all fields

New Metadata Field: Which Config Files?

phrases-en.xml
display names and help text for fields,
display names for field options

Task 3: Add a New Metadata Field

▶ Add the new field to **metadata-types.xml**

```
<type name="composition">
  <page name="core" />
  <field name="title" required="yes" />
  <field name="creators" required="yes" />
  <field name="composition_genre" required="yes" />

  <page name="abstract" />
  <field name="abstract" />
  <field name="keywords" />
</type>
```

New Metadata Field: Check Workflow

▶ But when we restart the Web server...


```
----- EPrints System Error -----
soton: Could not find field "composition_genre" in dataset "deletion", although it is
part of type: "composition"
-----
EPrints System Error inducing stack dump
at /opt/eprints2/perl_lib/EPrints/Config.pm line 156
EPrints::Config::abort('soton: Could not find field "composition_genre" in d
ataset "d...') called at /opt/eprints2/perl_lib/EPrints/DataSet.pm line 408
EPrints::DataSet::new('EPrints::DataSet', 'EPrints::Archive-HASH(0xal71dec)',
'deletion', 'HASH(0xal7e974)', 'HASH(0xal7e998)') called at /opt/eprints2/perl_lib
/EPrints/Archive.pm line 634
EPrints::Archive::_load_datasets('EPrints::Archive-HASH(0xal71dec)') called
at /opt/eprints2/perl_lib/EPrints/Archive.pm line 200
EPrints::Archive::new_archive_by_id('EPrints::Archive', 'soton') called at /
opt/eprints2/perl_lib/EPrints/Session.pm line 157
EPrints::Session::new('EPrints::Session', 1, 'soton', 1) called at /opt/epr
ints2/bin/force_config_reload line 161
```

New Metadata Field:

ArchiveMetadataFieldsConfig.pm



▶ We've used a field in **metadata-types.xml** that EPrints doesn't know about

▶ All metadata fields must be defined in the **ArchiveMetadataFieldsConfig.pm** configuration file



New Metadata Field: ArchiveMetadataFieldsConfig.pm



- ArchiveMetadataFieldsConfig.pm** defines:
 - types and properties of all metadata fields
 - for eprints, users and documents
 - e.g. creators, title, abstract
 - default field values
 - automatic metadata fields
 - e.g. calculating the number of authors
 - Perl intensive

New Metadata Field: Add Definition

- Add a definition for the **composition_genre** field to **ArchiveMetadataFieldsConfig.pm**

```
{ name => "composition_genre",
  type => "set",
  options => [ "classical", "rock", "jazz", "blues",
               "folk", "world", "pop", "easy" ],
},
```

New Metadata Field: Check Workflow

- Web server restarts OK, but



Internal Server Error

The server encountered an internal error or misconfiguration and was unable to complete your request.

Please contact the server administrator, tmh@ecs.soton.ac.uk and inform them of the time the error occurred, and anything you might have done that may have caused the error.

More information about this error may be available in the server error log.



Apache/2.2.2 (Fedora) Server at flag.ecs.soton.ac.uk Port 80

New Metadata Field: Why it Failed



- EPrints uses the metadata configuration in **ArchiveMetadataFieldsConfig.pm** to:
 - construct its database tables
 - generate queries for selecting data from the database
- EPrints expects to find a **composition_genre** column in the database

```
DBD::mysql:db do failed: Unknown column 'composition_genre' in 'field list' at /opt/eprints2/perl_lib/EPrints/Database.pm line 1852.
at /opt/eprints2/perl_lib/EPrints/Database.pm line 1852
EPrints::Database::do("EPrints::Database::HASH(0x8c6a25c)", 'INSERT INTO inbo
x_ordervalue_en (eprintid,userid,dir,datest...) called at /opt/eprints2/perl_lib/
EPrints/Index.pm line 424
```

New Metadata Field: Update Database


- We need to either
 - rebuild the EPrints database tables for the new metadata configuration
 - will lose all data and uploaded files
 - use **erase_archive** and then **create_tables**
 - don't do this on a live repository!
 - useful development technique
 - add the field to the database by hand
 - won't lose any data
 - instructions for doing this on the EPrints wiki
 - http://wiki.eprints.org/w/Adding_a_Field_to_a_Live_Repository

New Metadata Field: Check Workflow

- Field now appears in deposit workflow
- Now just need to add some phrases!
 - field title and help text
 - name of each option

```
["eprint_fieldname_composition_genre" not defined] *
["eprint_fieldhelp_composition_genre" not defined]
["eprint_fielddopt_composition_genre_classical" not defined]
["eprint_fielddopt_composition_genre_rock" not defined]
["eprint_fielddopt_composition_genre_jazz" not defined]
["eprint_fielddopt_composition_genre_blues" not defined]
["eprint_fielddopt_composition_genre_folk" not defined]
["eprint_fielddopt_composition_genre_world" not defined]
["eprint_fielddopt_composition_genre_pop" not defined]
["eprint_fielddopt_composition_genre_easy" not defined]
```





New Metadata Field: Summary

- ↔ Define type and properties of new field in **ArchiveMetadataFieldsConfig.pm**
- ↕ Add field to deposit workflow in **metadata-types.xml**
- ↔ Add display name and help text, and display names for each field option, to **phrases-en.xml**
- **force_config_reload**
- ↕ Erase and rebuild database
 - or manually add new field

eprints



Other Config Files: **subjects**

- Plain text file that defines the subject tree for the classification system
- By default contains the top 2 levels of the US Library of Congress classification
- “subjects” is actually a misnomer
 - other hierarchical classifications can be defined
 - organisational structure is a common addition
 - our **composition_type** field could have taken its values from a hierarchy of musical genres

eprints



Other Config Files: **ArchiveOAIConfig.pm**

- Methods for handling Open Archive Initiative metadata harvesting protocol (OAI PMH)
 - main method **eprint_to_unqualified_dc** converts an EPrints data structure to an OAI structure
 - other informational definitions give policies etc
- This file should be extended for data archives, to allow non-DC information to be shared
- Perl intensive
- Hardly ever used (except for exotic data types)
 - we could expose **composition_genre** as dc:subject

eprints



Other Config Files: **ArchiveRenderConfig.pm**

- Methods for generating the abstract pages for each item
- Perl intensive
- eCrystals data repository heavily modified this configuration file
- We could embed an music player applet on each Composition page

eprints



Other Config Files: **ArchiveValidateConfig.pm**

- Methods for checking the metadata fields that a depositor is submitting
 - individual fields
 - a whole page (i.e. combination of fields)
 - a document (e.g. has the user submitted a format safe for preservation purposes?)
 - a complete eprint record
 - a user
- Perl intensive

eprints



Other Config Files: **ArchiveTextIndexing.pm**

- Methods for supporting free text indexing
 - definitions of lexical token separators
 - list of stop words
 - filter that translates a text into a bag of words
- Unlikely to be changed

eprints



Web Server Config Files

- ▶ **auto-apache.conf** is the main workhorse
- ▶ defines where the archive files are, how to handle script requests and errors etc.
- ▶ usually not changed
 - ▶ some tweaks may be necessary if you are hosting other Web-based services on the same server



Reflection: What do you need to do?

- ▶ Look back at the issues you raised for configuring EPrints
 - ▶ can you see where you would need to start working in the EPrints setup?
 - ▶ can you find some repositories which do things in the same way?
 - ▶ i.e. can you find someone to give you advice?

