

Crystallography Service Sample Database Administrator's Guide

J.P.Hagon
Computer Systems Support
School of Chemistry

March 31, 2012

Introduction

This guide is intended for staff who administer the Newcastle University Crystallography Service Sample Database. It includes both a general description of the web interface and associated administration procedures along with a more technical description of the software interface and the database itself so that administrators can recover from situations such as a forgotten administrator password.

Contents

1	General Description of the System	3
1.1	Introduction	3
1.2	Users	3
1.3	Groups	3
1.4	Samples	3
1.5	Public Pages	5
2	Web Management Guide	5
2.1	Introduction	5
2.2	Adding Static Pages	6
2.3	Uploading General Files to the Server	7
2.4	Creating Users and Groups	8
2.5	Sample Management	8
3	The Database	10
3.1	Introduction	10
3.2	The Samples Table	13
3.3	The Users Table	14
3.4	The Stores, Hazards and Sensitivities Tables	15
3.5	The Groups Table	15
3.6	Other Tables	15
3.6.1	The Pages Table	16
3.6.2	The Assets Table	16

3.6.3	The Popups Table	16
3.6.4	The Flags Table	16

1 General Description of the System

1.1 Introduction

The system consists of a *front-end* which is used by users and administrators to submit sample requests and upload analysis data. This front-end is implemented using the ruby programming language and version 3 of Ruby on Rails. The front-end is hosted on an *Apache* server running on an *Ubuntu* linux system. Technical details will be described elsewhere.

The *back-end* consists of a set of ruby programming libraries and a SQL database — in this case SQLite3. More technical aspects of the back-end will be described elsewhere.

1.2 Users

The system has a relatively simple user setup with just one basic user type. However, there are three levels of authority that a user can have:

Standard Most users of the system will have a standard account which allows them to submit sample requests and view their own sample data.

Group Leader These users have the additional privilege of being able to see all of the sample data for their own group in addition to their own samples. A group of users may have more than one designated group leader.

Administrator An administrator, in addition to standard user privileges, can do many administration tasks. These include adding/deleting users, changing user privileges, submitting/updating/deleting samples, editing public web pages on the server and uploading files to the server.

Users can either self-register or be added by an administrator. An administrator can also disable a user account without actually deleting it. Only the most basic information about a user is stored in the database, namely first name, last name and email address. the email address serves as a login id. The user can set his own password. If the user forgets his password, the system can email him a secure link to the server via which the password can be reset.

1.3 Groups

All users must be associated with a group. Typically this will be a research group associated with a particular person. When a user self-registers, he must select an appropriate group. If such a group does not exist, an administrator must set one up for him. Usually one or more users will be designated *group leaders* and will have access to information about all the group's samples.

1.4 Samples

The primary purpose of the system is to track and keep a record of samples submitted to the crystallography service. A typical workflow is shown in Figure 1. Emails are sent automatically by the system when a sample status is updated by an administrator.

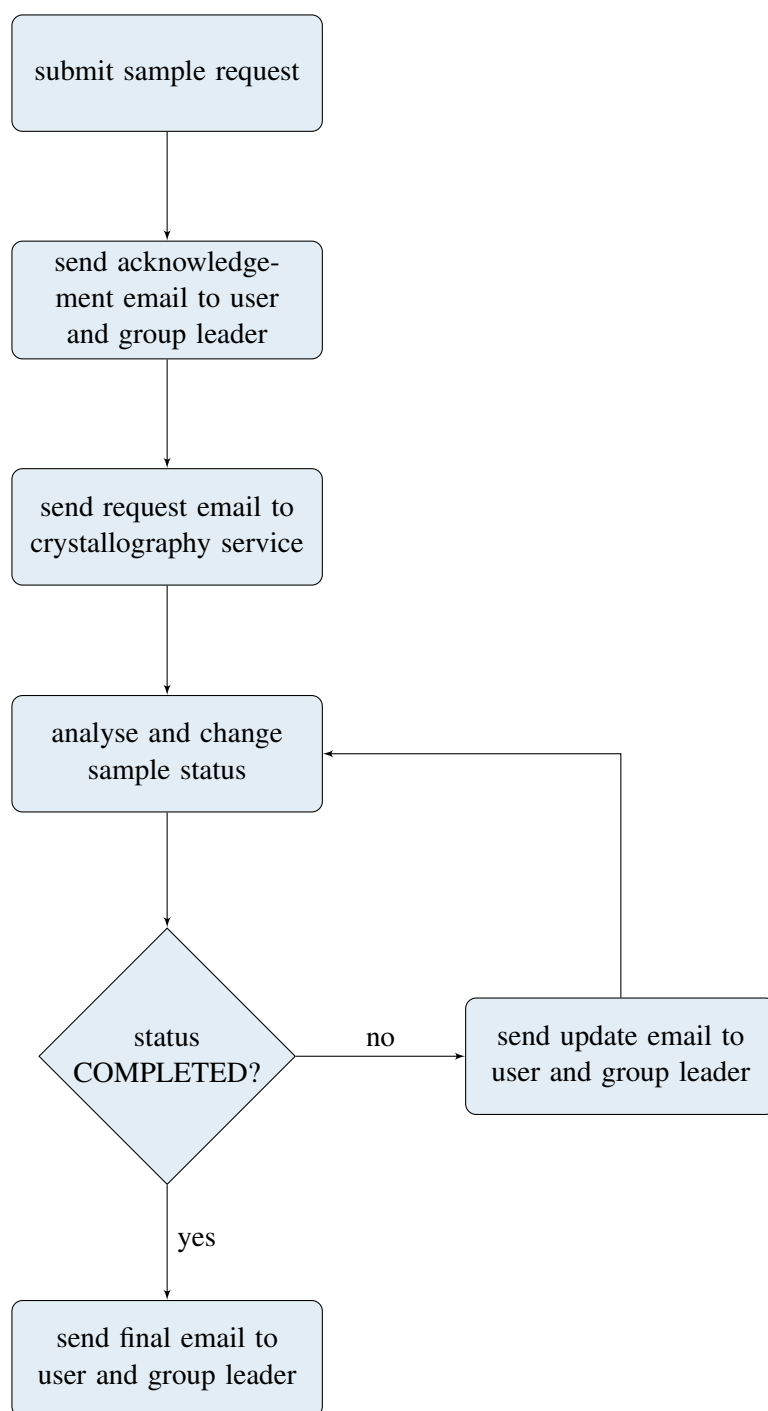


Figure 1: Typical workflow for sample processing cycle. Emails are sent automatically by the system whenever the status of a sample is updated.

1.5 Public Pages

Most information on the server can be viewed only by registered users. However, there are some pages which are more generally accessible. Such pages include the home page, general information pages and the sample queue. Public pages can be created and edited by an administrator using tools provided by the server software. rather than write pure HTML, an administrator can use a text-based markup language called Textile which can produce sophisticated web pages with all the usual constructs such as headings, paragraphs, floating elements, tables and images.

2 Web Management Guide

2.1 Introduction

In this section we describe the web management interface to the sample tracking database. When a manager is logged-in, the home page of the system looks similar to that shown in Figure 2

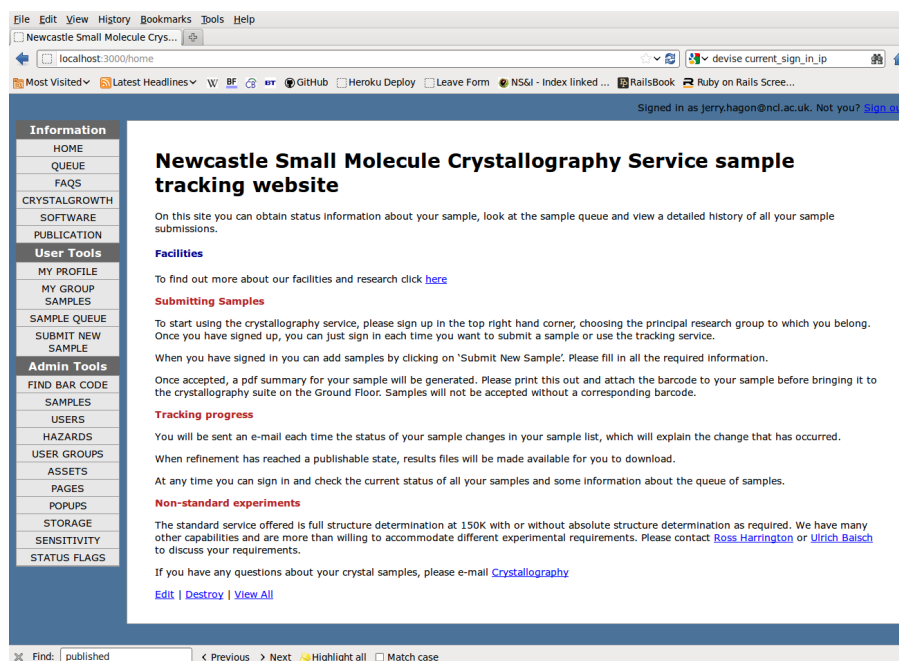


Figure 2: Administrator's view of home page.

There are three parts to this browser view:

- (i) a main display showing the contents of a page of information;
- (ii) a menu on the left side of the browser window;
- (iii) login information and a `sign_out` link above the main display on the right.

The left side menu consists of three sections:

Information These links point to *static* pages which can be created by an administrator. the administrator can also add extra links to the information section. We describe how to do this in §2.2.

User Tools These tools allow a user to view his sample list, submit a new sample and view his profile information. Additionally, if a user is also a group leader, he will have access to the *My Group Samples* link for listing all samples in the user's group.

Admin Tools This is the main set of web-based tools for administrators. We will describe each of these in the next section.

2.2 Adding Static Pages

Clicking the *PAGES* link in the *Admin Tools* sub-menu produces the pages index shown in Figure 4. This shows a list of pages. For each of these pages is a set of buttons allowing the administrator to show, edit or delete the page as shown in Figure 3. These buttons are used throughout the database editing pages on the web server. At the bottom of the list is a link to create a new page.



Figure 3: The *show*, *edit* and *delete* buttons.

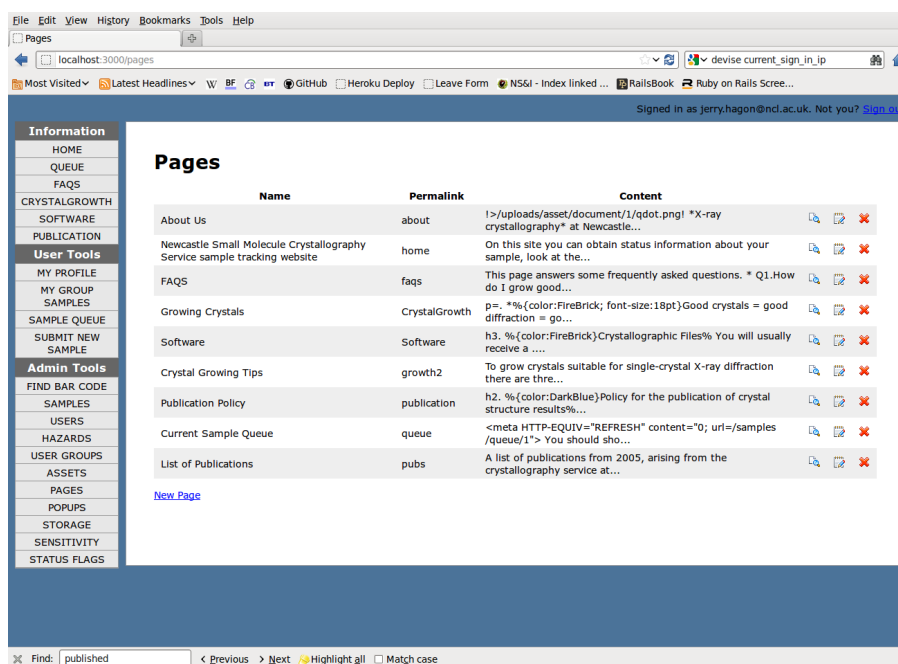


Figure 4: The pages index view.

Figure 5 shows the page editor — a very simple form for entering/changing text. This example shows the home page data. the content is entered in a markup language called Textile*. You can also specify the name of the page (this will be used to set the HTML title attribute and a permalink†. At the bottom are links to the Textile Reference Manual, the page view and the pages index. the page can be referred to via the URL:

```
<server name>/permalink
```

thus making static page addressing very simple. An alternative URL which can be generally used for any page is:

```
<server name>/pages/<id>
```

but the permalink-based URL is what you'd almost always use in practice. If, for some reason, you want to use the id-based URL, but don't know what the id is, then just click the 'show' icon in the pages index for the page you're interested in and look at the URL in the web browser window.

Note that there is also a checkbox labelled *Menu*. If this is checked then the page is added to the left hand *Information* menu of static pages. The *Priority* parameter is an integer that determines the order of the page in the menu. If two pages have the same priority their order is determined alphabetically. A good practice is to initially assign priorities in units of, say, 10. Subsequently, if a new page is created, there are then plenty of 'spare' priority numbers which can be used to add menu links in between existing links — otherwise existing priority numbers may need to be tweaked.

Edit Page

Name
Newcastle Small Molecule Crystallography

Permalink
home

Menu
☒

Priority
1

Content
On this site you can obtain status information about your sample, look at the sample queue and view a detailed history of all your sample submissions.
h4. Facilities
To find out more about our facilities and research click
(class=here)here (information about NCL Xray research):http://crystal.ncl.ac.uk/about
Submitting Samples
To start using the crystallography service, please sign up in the top right hand corner, choosing the principal research group to which you belong. Once you have signed up, you can just sign in each time you want to

Update Page

Show | View All | Textile Reference Manual | Textile Quick Reference

Figure 5: The page edit view.

2.3 Uploading General Files to the Server

You can upload arbitrary files to the server. These files are referred to as 'assets' and can be uploaded via the *ASSETS* link in the *Admin Tools* menu. Clicking on this link will take you to the assets index page which looks very similar to the pages index described in the previous section. each index entry tells you the pathname of the file on the server, together with a description of what the file contains. Often these files will be images or documents (e.g. PDF files) that you want to link to on one of the static web pages created as described in the previous section.

At the bottom of the asset index list is a link to create a new asset. Clicking this takes you to a simple menu where you can browse for a file to upload to the server. Clicking the *Create Asset* button will then upload the file to the server. It will then be listed in the asset index with an entry under the *Document* column pointing to its location in the file system. This location has the general form:

```
/uploads/asset/document/<id>/<filename>
```

*You are allowed to mix HTML and Textile together.

†A tag which is used as a basis for a concise URL.

Here, <filename> is the actual name of the uploaded file as it was when it was uploaded. <id> is the database id the document has in the assets table described in detail in §3. The actual URL of the document is then:

<servername>/uploads/asset/document/<id>/<filename>

2.4 Creating Users and Groups

A user must belong to a group, so it is advisable to create a group for a user before the user is created. Creating a user group is straightforward via the *USER GROUPS* link in the *Admin Tools* menu. As usual, this will take you to an index of existing groups, with a link to create a new group at the bottom. Creating a new group merely requires that you enter two fields:

- (i) a 3-letter group abbreviation (it *must* be three letters);
- (ii) a longer group description.

Users can be created in two ways; they can self-register by clicking on the link top-right on the home page or they can be created by an administrator. In either case the form used to create and register a new user is the same.

2.5 Sample Management

In this section we give a complete description of the sample management workflow.

- (i) First, a user fills out a sample submission form online. This form is shown in Figure 6. This figure shows the submission form from the point of view of both a non-administrative user and an administrator.

Figure 6: User's view (left) and administrator's view (right) of the sample submission form.

The bits that an ordinary user doesn't see are those parts of the sample fields that are subsequently filled in by an administrator in the course of sample processing.

There is a lot of validation built into the form making it very unlikely that a form will be submitted with incomplete information. Table ?? gives a complete list of validation checks. If a validation check fails on submission of the form, the user will be presented

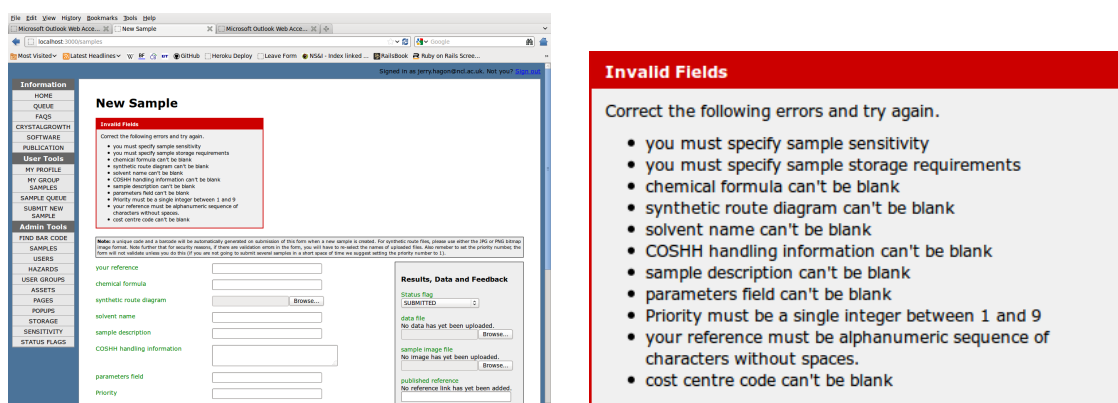


Figure 7: Illustrating what happens when an invalid form is submitted. The user sees an *Invalid Fields* box (shown enlarged on the right). In this case, an administrative user has failed to fill in any of the fields — hopefully a very rare occurrence!

with the *Invalid Fields* box which will tell him which fields have not been entered correctly and what is required. An extreme example of this is shown in Figure 7 which shows what happens when none of the fields in the form are filled-in.

- (ii) On successful submission of the sample form, the user (and group leaders if the user is not a group leader himself[‡]) will receive a confirmation email containing a link to the *Sample Receipt* — a PDF file containing the user input information as well as a unique code identifying the user, user group and sample and a unique barcode. The email has the following general form with the tags in angle brackets filled-in automatically:

Dear User

New Sample Submission Code: <SAMPLE CODE> (your ref <SAMPLE USERREF>)
Submitted By: <USER FULL NAME>

your sample analysis request has been received. Please download a receipt using the link below. Please quote the sample code in any correspondence.

There is a tear-off slip at the bottom of the receipt which you should attach to your sample. You will be informed via email of any changes in the status of your sample.

<LINK TO SAMPLE RECEIPT>

Copies of this email are sent to both sample submitters and their research group leaders (where different).

Newcastle Crystallography Service

A tear-off slip at the bottom of the sample receipt containing the barcode, sample code

[‡]If there is more than one group leader then the other group leaders will also receive an email.

and provided user reference as well as essential COSHH information can be attached to the actual sample itself. Figure 8 shows a typical sample receipt.

The receipt is generated on-the-fly from the supplied sample information. To perform the generation of the PDF, the well-known Prawn ruby library is used. Once the sample has been submitted, a user can regenerate the sample receipt at any time via a PDF link button in the sample list on his profile page shown in Figure 9.

- (iii) Next, the user brings his sample (with attached slip) for analysis and at this point he should also see it in the sample queue.
- (iv) Initially the sample will have a status of SUBMITTED. At various points in the analysis, this status will be changed by crystallography staff. Whenever the status is changed, an email is sent to both the user and group leaders informing them of the change. The text of the email looks similar to this:

Dear User

This is to inform you that the status of sample
<SAMPLE CODE> (your ref <SAMPLE USERREF>) submitted by
<USER FULL NAME>
has changed as follows:

New Status: <NEW STATUS FLAG> <NEW STATUS FLAG DESCRIPTION>

Old Status: <OLD STATUS FLAG> <OLD STATUS FLAG DESCRIPTION>

<LINK TO FULL SAMPLE INFORMATION>

Copies of this email are sent to both sample submitters and their
research group leaders (where different).

Newcastle Crystallography Service

- (v) When crystallography staff have uploaded all results files (docx, res and sample image) to the server and added any additional text feedback for the sample, they will flag the sample as COMPLETED and the process will end. The sample data will continue to be available to users (and administrators) indefinitely after that.

Results files (docx and res) will usually be made available in a single zip file, with the sample image in a separate file. The files are uploaded by administrators using the sample edit form shown in Figure 10. Note that only administrators have access to the sample edit form. Users will see the information in a slightly different way, not as a form but as a standard 'show' page. This view is shown in Figure 10. Note that if administrators do not fill in the feedback section, then a default message 'No feedback given.' is seen on the relevant part of the sample show page. This is also illustrated in Figure 11.

3 The Database

3.1 Introduction

The core of the system is the database which holds information about users, samples etc. In this section we describe the whole database structure (or *schema* in database parlance). The

Newcastle Crystallography Service

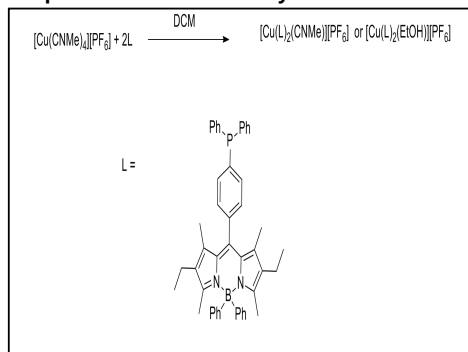
Bedson Building ♦ Newcastle University ♦ NE1 7RU

Sample Code: LJH-LD-12-0002

Your Ref: LD261

Please check the details on this receipt. Changes can be made only by Crystallography staff. Please use the tear-off slip at the bottom of the page and attach it to your sample. You will be automatically informed via e-mail of any changes to your sample status.

Proposed Structure and Synthetic Route



Sample Details and Requirements

Powder Diffraction Required? No
Chiral Structure? No
Your Priority Number: 2

User Details

Submission Date: 2012-03-02 15:29:54 UTC
Submitted By: Laura Davies
Research Group: Lee Higham Research Group
Contact E-Mail: l.h.davies@ncl.ac.uk
Cost Centre Code: n/a
Assigned Bar Code: FQYSY1J7L15

Supplied COSHH Information

Name of Solvent: Ethanol/Pentane
Description of Sample: Copper Bodipy phosphine complex
Handling Procedures: Flammable, Irritant
User Comments: Structure should be similar to LD260. As the crystals were grown in ethanol, the bound MeCN ligand may exchange with the ethanol.

Hazards:

Highly Flammable (F)
 Harmful (Xn)
 Irritant (Xi)

Storage:

bench dark

Sensitivity:

light



LJH-LD-12-0002 (LD261)



F Q Y S Y 1 J 7 L 1 5

Supplied COSHH Information

Name of Solvent: Ethanol/Pentane
Description of Sample: Copper Bodipy phosphine complex
Handling Procedures: Flammable, Irritant
User Comments: Structure should be similar to LD260. As the crystals were grown in ethanol, the bound MeCN ligand may exchange with the ethanol.

Hazards:

Highly Flammable (F)
 Harmful (Xn)
 Irritant (Xi)

Storage:

bench dark

Sensitivity:

light

School of Chemistry Crystallography Service, Newcastle University

Figure 8: Example of a sample receipt (the actual size is A4). Note the tear-off slip at the bottom. The receipt is generated on-the-fly from the supplied sample information. The green border is not part of the PDF rendering, it is used here merely to show the A4 page border.

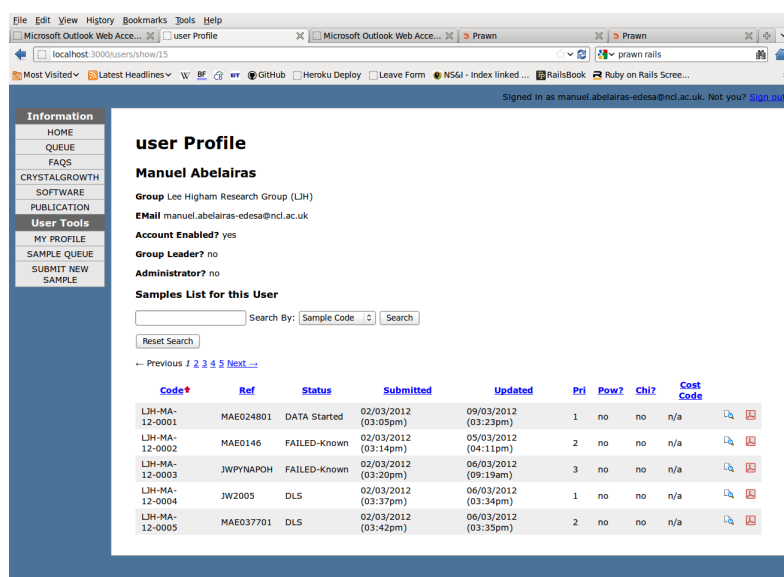


Figure 9: User profile page showing sample list with PDF icon (shown enlarged bottom right).

The screenshot shows a web browser window displaying the 'Edit Sample LJH-LD-12-0002' form. The form is divided into several sections: 'your reference' (LD261), 'chemical formula' (C96H95B2CuF6N5P3 OR C96H98B2CuF4), 'synthetic route diagram' (Browse...), 'solvent name' (Ethanol/Pentane), 'sample description' (Copper Bodipy phosphine complex), 'COSHH handling information' (Flammable, Irritant), 'parameters field' (n/a), 'Priority' (2), 'Powd' (checkbox), 'Chiral' (checkbox), 'cost centre code' (n/a), and 'Comments' (Structure should be similar to LD260. As the crystals were grown in ethanol, the bound MeCN ligand may exchange with...). There are also sections for 'Chemical hazard information', 'Sample sensitivity information', 'Sample storage information', and a 'Results, Data and Feedback' section with fields for 'Status flag' (COMPLETED), 'data file', 'sample image file', 'published reference', 'Colour' (red), 'Size' (0.15 0.10 0.04), 'Shape' (lath), and 'Feedback'.

Figure 10: A fully complete sample edit form.

easiest way to get an overall view of the database schema is to study Figure 12. This shows all the tables, fields and relationships in a single diagram. We now give a brief description of each table.

Note that all tables except join tables have an autoincremented integer field called `id` which serves as the unique primary key for each record in the table. The `id` field will not be listed explicitly in the description of each table. All non-join tables also have two other fields, `created_at` and `updated_at` in a datetime format. Again, we will not explicitly list these fields in the description of the tables which follows.

The figure shows a web application interface for a sample database. The left panel displays the user's view of a completed sample, 'Sample LJH-MA-12-0018'. It includes a sidebar with navigation links (HOME, QUEUE, FAQs, CRYSTALGROWTH, SOFTWARE, PUBLICATION, User Tools, MY PROFILE, SAMPLE QUEUE, SUBMIT NEW SAMPLE). The main content area shows sample details: userref: MCQUINAPOTF, cif: C39H25F3N2O4S, a chemical structure diagram, and a 3D molecular model. Below the model, fields for coshh_name, coshh_desc, coshh_info, params, priority, powd, and chiral are listed. The right panel is an enlarged view of the 'Analysis Results' section, showing status: COMPLETED, sample data: available, sample image: available, published reference: not yet available, and other crystal data (colour: colourless, size: 0.40 0.30 0.30, shape: block, feedback: No feedback given).

Figure 11: The user's view of a completed sample (left). On the right is an enlarged view of the results section. Note also that a thumbnail image of the sample is displayed. This thumbnail, when clicked, will show the full-size image which can then be downloaded if desired.

3.2 The Samples Table

The samples and users tables are the key parts of the database as is evident from Figure 12. They are related to each other via a *one-to-many* relationship, i.e. *one* user can have *many* samples. The samples table consists of the following fields:

code a string, automatically generated by the system having the general form AAA-AA-YY-1111 where the AAA and AA represent 3-letter codes for group and submitter respectively; the YY represents the year and the 1111 represents a number which is incremented for that group but reset to zero at the start of each calendar year.

cif a string representing the chemical formula of the sample in cif format.

synth a string representing the file name of an image file specifying the details of the synthesis.

coshh_name a string representing the name of the solvent (if any).

coshh_info another string describing any procedures in case of contact with the sample.

coshh_desc a text field providing a brief description of the sample (e.g. organic amide).

params a string representing unit cell parameters or CSD/Newcastle code for possible by-products or previously obtained, unpublished results.

priority an integer between 1 and 9 to give an indication of priority.

powd a boolean parameter indicating if the sample requires powder diffraction (y/n).

chiral another boolean indicating whether the molecule is chiral (y/n).

costcode a string providing a cost centre code for charging if relevant.

barcode a string field for an automatically generated Code39 standard barcode.

user_id this integer holds the id field of the user who requested the sample analysis.

flag_id an integer holding the id field of the status flag of the sample.

userref a string for a user-defined reference. This is required to be an alphanumeric sequence of characters *without spaces*.

zipdata a string holding the name of a zipfile containing the results of the analysis.

sampleimage a string holding the name of an image file of the sample molecule after it has been identified by the analysis.

reference a text field for a published reference (typically in the form of a DOI).

comments a text field for any general comments the user wishes to make about the sample.

colour a string holding information about the colour of a sample after analysis.

size a string holding information about the size of a sample after analysis.

shape a string holding information about the shape of a sample after analysis.

feedback a text field containing any additional comments on the sample by crystallography staff.

3.3 The Users Table

The users table, in addition to maintaining a record of users and their samples, also serves as a key part of the authentication and authorization system which will be described later. the users table is related to the samples table via a *one-to-many* relationship, i.e. *one* user has *many* samples.

email a string holding the email address of the user. This serves also as the user login id.

encrypted_password a string holding the user's password in an encrypted form.

reset_password_token a string containing a special token used if the user has forgotten his password and needs to reset it.

reset_password_sent_at a datetime field recording the time a token enabling a user to reset his password was sent.

remember_created_at a datetime field specifying the time at which a user requested that his login id be remembered by the browser so he need not type in his credentials.

sign_in_count an integer holding the number of times a user has logged-in.

current_sign_in_at a datetime field holding the sign-in time for the current session.

last_sign_in_at a datetime field holding the last sign-in time for the user.

current_sign_in_ip a datetime field holding the user's ip address for the current session.

last_sign_in_ip a datetime field holding the previous login ip address for the user.

group_id an integer representing the id field of the group to which the user belongs.

admin a boolean field indicating whether the user is an administrator (y/n).

firstname a string holding the user's first name.

lastname a string holding the user's last name.

leader a boolean field indicating whether the user is a group leader (y/n).

enabled a boolean field indicating if the account is enabled (y/n).

3.4 The Stores, Hazards and Sensitivities Tables

These tables are each very similar and have the same basic structure. They are used to specify storage, hazard and sensitivity properties for a sample. They all have a *many-to-many* relationship with the samples table. This is because a sample can have, for example, *many* storage requirements, but also a single storage requirement can be associated with *many* samples. All these tables have essentially the same fields:

name a string defining a short name for the property.

description a text field describing the property at greater length.

For historical reasons, the hazards table uses the names hazard_abbr and hazard_desc for the name and description fields. Also the hazard_desc field is a text field rather than a string.

Associated with these tables are three further *join tables* which facilitate the many-to-many relationship between a sample and its properties. These join tables are called samples_stores, samples_hazards and samples_sensitivities. They all contain two fields corresponding to the sample id field and the associated property id field. For example, samples_stores contains the fields sample_id and store_id. Both these fields are integers of course.

3.5 The Groups Table

This table represents groups of users, normally research groups but also perhaps external companies etc. It is a simple table, but important in the way the whole system works. It contains the following fields:

group_abbr a 3-letter string as an abbreviation for the group. Amongst other things this is used to form part of the sample code string mentioned earlier.

group_desc a string giving a more complete description of the group.

3.6 Other Tables

There are several other tables which are less important than the ones discussed so far in the sense that they are strictly not necessary for a working sample tracking system. However, they do assist in making the system much easier to manage and also help making the system much friendlier for users. These tables are the assets, pages and popups tables.

3.6.1 The Pages Table

The purpose of this table is to provide a means by which administrators can add ‘static’ content to the sample tracking web site. Each static page has its content stored in this table. The fields are:

name a string storing a name for the page. This is typically used to provide a title for the page in a web browser window.

permalink another string used to provide a short, quick URL for the page.

content a text field which contains the page content. This is expected to be written in Textile markup language (although a mixture of pure HTML and Textile can be used).

menu a boolean specifying whether this page should appear on the *Information Menu*.

priority an integer specifying a priority for ordering the page on the *Information Menu*.

3.6.2 The Assets Table

The assets table keeps a record of general files which have been uploaded to the server. These files are typically graphical images, pdf documents etc. and will usually be referenced in one of the static pages created by administrators which are stored in the pages table. An ‘asset’ is simply one of these uploaded documents and the assets table keeps a record of it. The fields are:

document the full path name of an uploaded document. This path name is ultimately assigned using the carrierwave file uploading plugin to ruby on rails.

description a text field giving a brief description of the document.

3.6.3 The Popups Table

This table stores descriptive information about the primary fields in the samples table. It has two fields:

name this string should have the same name as one of the sample fields for which a detailed description is required.

description a text field giving a detailed description of the associated sample field in the corresponding name field.

The popups table, as its name implies, provides descriptive text in popup boxes whenever a user hovers the mouse over the appropriate field in the sample submission form.

3.6.4 The Flags Table

This table stores a set of status flags together with a more verbose description of what the flag means.

name a string containing the name of the status flag, e.g. SUBMITTED, COMPLETED etc. There can be any number of flags but the aforementioned flags must be present because when the sample is originally submitted it is, by default, given the status SUBMITTED. Also, when analysis is finished, the sample queue list will omit all samples which have had the COMPLETED flag set or have a status flag which begins with the string FAILED.

description a text field giving a detailed description of the associated flag name.

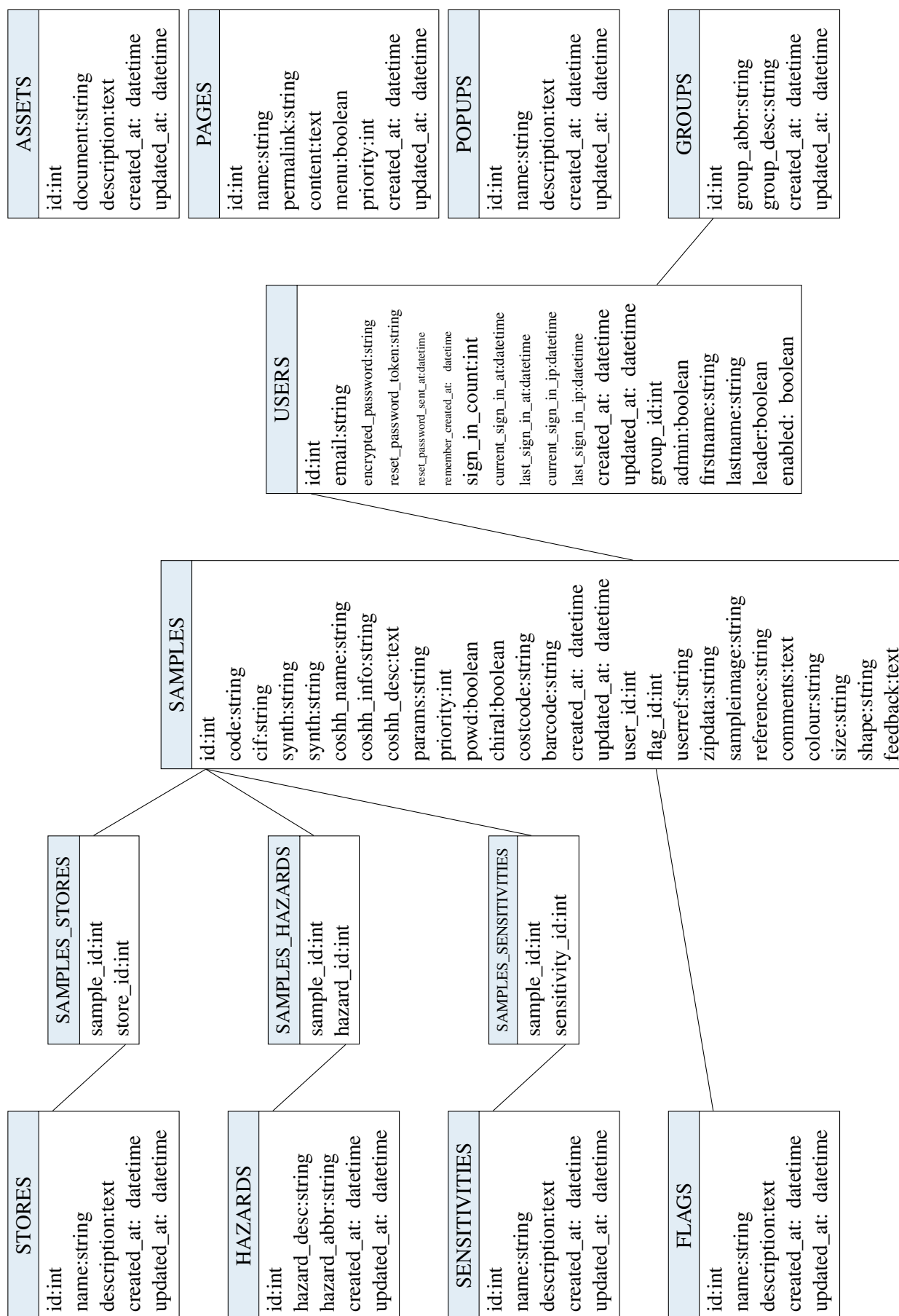


Figure 12: The overall database schema. Relationships between tables are indicated with lines joining the relevant fields. Note that the tables SAMPLES_STORES, SAMPLES_HAZARDS and SAMPLES_SENSITIVITIES are *join tables* which serve only to facilitate a many-to-many relationship between the tables they link.