



Manual  
for  
Setting up

CSIR Open Access Institutional Repository

Drafted by

CSIR Open Access Committee

## Content

1.0 Scope.....	4
2.0 Eprints 3.x Installation on CentOS 5/6 OR RHEL 4/5 .....	5
2.1 CentOS (Community ENTERprise Operating System) .....	5
2.2 Now Install Eprints .....	7
2.3 Configuring and Creating the Archive.....	7
2.3.1 Creating an Archive.....	7
2.3.2 Running epadmin .....	7
2.4 Regular Maintenance .....	9
2.4.1 Generate Views.....	9
2.4.2 Generate Statistics .....	9
2.4.3 Generate Abstracts.....	9
3.0 Content of the Institutional Repository (IR) .....	10
3.1 Content determined for the annual performance of the individual author .....	10
3.2 Optional .....	10
3.3 Responsibility .....	10
3.3.1 Coordinators .....	10
3.3.2 Authors .....	11
4.0 Customization .....	12
4.1 Home page.....	13
4.1.1 Masthead.....	13
4.1.2 Welcome Message .....	14
4.1.3 Number of Records .....	14
4.1.4 Recent Submissions.....	14
4.1.5 Join Search Harvesters .....	14
4.1.6 Submission Policy and Copyright Issues .....	14
4.1.7 Links to the Institute's and Library Websites.....	14
4.1.8 Footer .....	14
5.0 Subjects (Defining Classes) .....	15
5.1 Guidelines .....	15
5.2 Illustrative Example .....	15
5.3 Menu Level Operation.....	15
5.4 Defining in the File .....	16
6.0 Divisions.....	17
7.0 Types of Documents.....	18
7.1 Adding Data Elements for Data Entry .....	18
8.0 Adding Fields .....	19
8.1 Guide/Supervisor .....	19
8.2 CSIR Author ID .....	19
8.3 Field for Citation Link .....	20
8.4 Unique OAI-Identifier to IR.....	20
9.0 Standards for Metadata Description .....	21
9.1 General.....	21
9.2 Title.....	21

9.3 Abstract .....	21
9.4 Creators/Authors/Editors .....	21
9.5 Corporate Creators .....	22
9.6 Publication Details/ Citation .....	22
9.6.1 Journal Title .....	22
9.6.2 Title of Book.....	22
9.6.3 Publisher.....	22
9.6.4 Number of Pages.....	22
9.6.5 Series Name .....	23
9.6.6 Official URL .....	23
9.6.7 Page Range.....	23
9.6.8 Identification Number .....	23
9.6.9 Related URLs .....	23
9.7 Project .....	23
9.8 Contact Email Address.....	23
9.9 Keywords.....	23
9.10 Additional Information/Description .....	24
9.11 Subject Category.....	24
10.0 Copyright Issues .....	25
10.1 The Problem.....	25
10.2 The Solutions .....	25
10.2.1 If the copyrights are transferred.....	25
10.2.2 If the copyrights are not transferred.....	25
10.2.3 Solution for items with 'white' publishers.....	25
11.0 Work-flow .....	26
12.0 Backup .....	27
13.0 Statistics.....	29
Annexure I.....	30
Annexure II: DSpace and EPrints Fields for Data Entry – Comparison .....	31
Annexure III.....	33

## 1.0 Scope

This manual provides guideline for the

1. Installation of the
  - a) OS - CentOS 5/6 OR RHEL 4/5, and
  - b) Eprints 3.x
2. Configuration of the eprints for the repositories in CSIR laboratories
3. Guidelines to the coordinators and authors on the
  - a) Content of the repository
  - b) The metadata description standards
  - c) Maintenance

## 2.0 Eprints 3.x Installation on CentOS 5/6 OR RHEL 4/5

Following guidelines are based on Eprints v3.2.8 installation/configuration on CentOS 6.  
(For any technical assistance, please contact [alok@urdip.res.in](mailto:alok@urdip.res.in))

### 2.1 CentOS (Community ENTERprise Operating System)

CentOS is a free operating system based on Red Hat Enterprise Linux (RHEL). Red Hat Enterprise Linux is available only through a paid subscription service that provides access to software updates and varying levels of technical support. CentOS developers use Red Hat's source code to create a final product very similar to Red Hat Enterprise Linux. CentOS is available free of charge. Technical support is primarily provided by the community via official mailing lists, web forums, and chat rooms.

Pre-requisite for Eprints 3.x installation: Install CentOS 6 with Perl, MySQL and Apache Web Server. Though many required packages are already installed with CentOS.

To check whether any package (say xxx) is installed; use this command as root user  
# rpm -qa | grep xxx

```
# yum install httpd
# yum install mysql-server
# yum install gcc
# yum install wget
# yum install lynx
# yum install mod_perl
# yum install perl-DBI
# yum install perl-DBD-MySQL
# yum install gzip,
# yum install unzip,
# yum install perl-XML-Parser
```

antiword and xpdf is required for Fulltext indexing. xpdf is an open-source PDF viewer for the X Window System.

antiword is not available through "yum" on CentOS, so do the following:

```
#cd /root
# wget http://pkgs.repoforge.org/antiword/antiword-0.37-1.2.el5.rf.i386.rpm
#rpm -Uvh rpm -ivh antiword-0.37-1.2.el5.rf.i386.rpm
```

It is suggested on centos.org to use poppler and poppler utils instead of xpdf. These packages were installed using

```
# yum install poppler poppler-utils
```

texlive The LaTeX front end for the TeX text formatting system.

```
# yum install texlive-latex
On previous version of CentOS package name is tetex-latex
```

```
# yum install ImageMagick
# yum install glib2-devel
# yum install libxml2-devel
```

-----

Installed following perl modules using CPAN:

-----

```
perl -MCPAN -e shell
cpan> install Data::ShowTable
cpan> install MIME::Base64
cpan> install Unicode::String
cpan> install Term::ReadKey -----> if problem, the do this ***
cpan> install Readonly
cpan> install MIME::Lite
cpan> install XML::LibXML
cpan> install CGI --->if this creates problem and suggests to install by force then do like this
cpan> force install CGI
```

-----

```
*** # mkdir /usr/local/build
# cd /usr/local/build
# wget http://search.cpan.org/CPAN/authors/id/K/KJ/KJALB/TermReadKey-2.14.tar.gz
# tar xvf TermReadKey-2.14.tar
# cd TermReadKey-2.14
# perl Makefile.PL
# yum install gcc.i386 (compiler did not come standard with CentOS 5)
# make
# make install
```

-----

```
# cd /root
# wget http://gdome2.cs.unibo.it/rpm/gdome2-0.8.1-1.i386.rpm
# wget http://gdome2.cs.unibo.it/rpm/gdome2-devel-0.8.1-1.i386.rpm
# rpm -Uvh gdome2-0.8.1-1.i386.rpm gdome2-devel-0.8.1-1.i386.rpm
```

There is a bug in gdome-config. To fix it do this:

To see if the bug is a problem run:

```
# gdome-config --libs
```

If you get something like:

```
/usr/bin/gdome-config: line 86: --libs: command not found
```

```
/usr/bin/gdome-config: line 87: --cflags: command not found
```

Then you need to fix the bug. As root edit /usr/bin/gdome-config

```
vi /usr/bin/gdome-config
```

Around line 88 find these two lines:

```
the_libs="$the_libs -L${exec_prefix}/lib -lgdome `--libs` `xml2-config --libs`"
```

```
the_flags="$the_flags -I${prefix}/include -I${prefix}/include/libgdome `--cflags` `xml2-config --cflags`"
```

And change them to this:

```
the_libs="$the_libs -L${exec_prefix}/lib -lgdome /usr/bin/glib-config --libs` `xml2-config --libs`"
```

```
the_flags="$the_flags -I${prefix}/include -I${prefix}/include/libgdome /usr/bin/glib-config --cflags` `xml2-config --cflags`"
```

## 2.2 Now Install Eprints

```
-----  
# cd /urs/local/build  
# wget http://files.eprints.org/662/1/eprints%2D3.2.8.tar.gz  
# tar xvfz eprints-3.2.8.tar.gz  
# cd eprints-3.2.8  
# ./configure --with-smtp-server=YourSMTPServer --prefix=/var/lib/eprints3 --with-  
user=apache --with-group=apache  
# ./install.pl  
# ln -s /var/lib/eprints3/bin/epindexer /etc/init.d/epindexer  
# chkconfig --add epindexer  
# chkconfig epindexer on  
-----
```

Open httpd.conf and add this line to the very end of the file

```
-----  
# vi /etc/httpd/conf/httpd.conf  
Include /var/lib/eprints3/cfg/apache.conf  
-----
```

Do following to start Mysql and Apache to start at the OS boot

```
-----  
# /sbin/chkconfig mysqld on  
# /sbin/chkconfig httpd on  
-----
```

## 2.3 Configuring and Creating the Archive

### 2.3.1 Creating an Archive

EPrints 3 can run multiple archives under one install. Multiple archives will require giving additional DNS aliases to the machine running EPrints. EPrints can then create all the parts of the apache configuration file needed to run the virtual hosts.

### 2.3.2 Running epadmin

Make sure MySQL is actually running.

Change to your eprints user (probably eprints).

```
# su apache
```

Change directory to the eprints directory (**/var/lib/eprints3**) and run

```
bin/epadmin create
```

You will get the following prompts (note that when you see something in [square brackets], it's the default value and can be selected by simply hitting <enter>):

- Archive ID - the system name for your archive. It's probably a good idea to think of something short and memorable. Once entered, an archive/<archive\_id> directory will be created, and the standard configuration files will be copied in.
- Configure vital settings - Hit enter to say 'yes'. This will lead to more prompting about core settings:

- Hostname - What someone will type into a web browser to get to your archive. Make sure that your systems team has a DNS alias pointing to your server for this.
  - Webserver Port - Which port do you want to serve the archive on? The default is 80, so unless you can think of a good reason not to, just hit enter to accept the default.
  - Alias - You can enter any number of aliases that will take users to this archive. Enter a '#' when you don't want to enter any more. You could have your archive served on eprints.myorganisation.org and eprints.myorg.org. As with the Hostname, your systems team need to be informed about these aliases too.
    - Redirect your chosen alias to Hostname - Yes, usually that is your intention.
  - Please enter the path part of the repository's base URL - Make it simple for you and your visitors ... and confirm the suggestion; multiple archives could be managed by different domains or ports!
  - If you will use https for your user pages (including login) enter the https hostname - No doubt, for secure usage you need another name.
  - Administrator Email - Enter the email address of the repository administrator. This will allow your repository users to send email to the right person.
  - Archive Name - The full name of your archive. By default, this will be used on many of the pages, and in the title bar of the browser.
  - Write these core settings - If you don't say 'yes', then you entered all that data for nothing (relaunch the process by eadmin config\_core archive\_id).
- Configure database - EPrints makes extensive use of a MySQL database. Enter 'yes' to launch eadmin config\_db archive\_id.
    - Database Name - The internal name of your database. It makes sense to use the Archive ID for this, but you don't have to. You don't need to create this database, eadmin will do it for you.
    - MySQL Host - The address of the server that the database is running on. If the database is on the same machine as the EPrints installation, confirm 'localhost'.
    - MySQL Port - You probably don't need to enter a value. If you have problems connecting to the database, talk to your systems team.
    - MySQL Socket - As with MySQL Port, it's unlikely that you need to enter anything.
    - Database User - The username with which to log into the MySQL Database. You don't need to create this user, eadmin will do it for you. If you enter a MySQL username that already exists, it will be overwritten by epstats!
    - Database Password - Just confirm the suggested password for the Database User, eprints will keep track of it!
    - Write these database settings - You should write them, or you'll lose them.
- Create database <Database Name> - Say yes to launch eadmin create\_db archive\_id, and eadmin can create the database and populate it with all the right tables. If you've already created a database and a user for this archive, say 'no'.
    - Database Superuser Username - 'root' isn't a bad suggestion, but should match your specifications during the MySQL installation.
    - Database Superuser Password - To create the database and the user, eadmin needs the MySQL Root Password (specified during MySQL installation). This is not saved anywhere. It is only used to log into MySQL, create the database and create the user with useful access rights. The password is forgotten afterwards, because it's not needed anymore!



- Create database tables - say 'yes' to have epadmin create all the database tables, i.e. initiate the same as with epadmin create\_tables archive\_id.
- Create an initial user - It's a good idea to create a user account for yourself at this point, but you can start the same process by epadmin create\_user archive\_id anytime ...
  - Enter a username - The username you will be used to log into EPrints on the home page of the new archive.
  - Select a user type (user|editor|admin) - There are three user roles in EPrints. You probably want to be an administrator, so enter 'admin'.
  - Enter Password - A password for this user. Remember to choose a password that will be hard for someone else to guess!
  - Email - Enter your email address so that users can get in contact with you.
- Do you want to build the static web pages - There are a number of pages in EPrints which change very rarely. These are the static pages. The Home page and the About page are examples of static pages. Stylesheets are also static. These pages need to be built, so say 'yes' (otherwise you have to start that process manually later by generate\_static archive\_id).
- Do you want to import the LOC subjects - If you will be using the Library Of Congress subject hierarchy, say 'yes'. Otherwise you will need to create your own subject hierarchy and to import them using import\_subjects.
- Do you want to update the apache config files? (you still need to add the 'Include' line in the very central http.conf, e.g.) - Your archive has a number of files which it uses to configure the web server. These should be updated, so say yes or start generate\_apacheconf archive\_id later.
- Before exiting, epadmin will display information about configuring the webserver.

If you want to add some more users, use the command

```
epadmin create_user <repository id>.
```

## **2.4 Regular Maintenance**

EPrints front end web pages and abstracts *\*are not\** automatically updated when you make changes to the repository. To apply your changes and update the web pages:

### **2.4.1 Generate Views**

```
eprints@host$ bin/generate_views yourarchivename
```

### **2.4.2 Generate Statistics**

```
eprints@host$ bin/generate_static yourarchivename
```

### **2.4.3 Generate Abstracts**

```
eprints@host$ bin/generate_abstracts yourarchivename
```

Finally a restart of your apache server is recommended, because a lot of settings will only be read initially!

**Now in Internet browser you may type hostname used at the time of creating archive to see the working of IR@Eprints.**

## 3.0 Content of the Institutional Repository (IR)

### 3.1 Content determined for the annual performance of the individual author

Every laboratory/institute of CSIR must establish an IR. This would at least contain metadata for the following:

(a) Published research literature

- Papers in the journals
- Conference proceedings
- Chapters in books
- Books and other published monographs

with the post-reviewed final manuscripts (including text, tables and figures) in pdf format (whose copyrights have been transferred to the publishers e.g article in a commercial publisher's journal) or publisher's pdf files (whose copyrights continue with the author – e.g. Current Science) and

(b) Unpublished literature

- Thesis and dissertations of the students who have worked in the CSIR laboratory for their degree or have a research guide/ supervisor from CSIR laboratory – with full-text (including figures and tables) in the form of pdf files for each of the chapters and preliminary pages and other content of the work.

### 3.2 Optional

The institute/laboratory may like to deposit other types of literature (e.g. Annual reports, technical reports, presentations, etc).

### 3.3 Responsibility

#### 3.3.1 Coordinators

The 'Coordinator' is the individual recognized for establishment and maintenance of the IR of a laboratory by the Director of the laboratory.

The coordinator is responsible for

- Establishment, smooth functioning of the IR within the laboratory
- Customization
- User profile maintenance
- Linking IR with subject, national, international and institutional harvesters
- Ensure availability of IR on Internet all the time (24x7)
- User training for metadata entry and file upload
- Checking and updating of record metadata so submitted by the authors and copyright issues that are connected with the record
- Making the record submitted by the author available on Internet
- Delivery of the documents (on receipt of requests) that are not open on Internet
- Periodically check functionality of the services
- Maintain statistics of downloads, and gaps, etc.
- Backup
- Entry of legacy data
- Liaison with other coordinators on issues of concern
- Updating the software
- Outreach

The coordinator may undertake these tasks him-/her-self or form a team within the laboratory with specific responsibilities and ensure that the timeline is maintained.

### **3.3.2 Authors**

The 'authors' are the individuals working as permanent staff of the CSIR irrespective of the Scientific or Technical groups.

Among the list of authors for a given work, the first author from the laboratory would be responsible for entry of metadata and uploading the full-text.

If the work belongs to authors from more than one laboratory (of CSIR), first authors among the list of authors of each laboratory would be responsible in making the content available in respective laboratory's IR. This is an essential duplication (each laboratory can claim for the credits of the work).

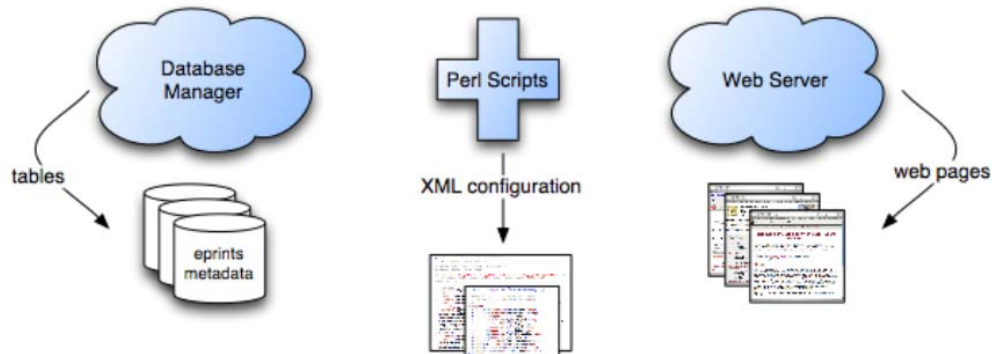
Within 15 days of publication of a literature the authors would upload the metadata along with full-text (post-referred item) on the IR. It is the responsibility of research supervisor/guide from CSIR laboratory to upload the thesis/dissertation of the student they guided within the 15 days of award of a provisional degree for the student's work.

## 4.0 Customization

The coordinator will carry out (may be in cooperation with the IT staff of the institute) following customization before populating the repository.

This chapter touches only essential aspects of customization for the IRs of CSIR. In order to do the customization, an understanding of the basic architecture is useful.

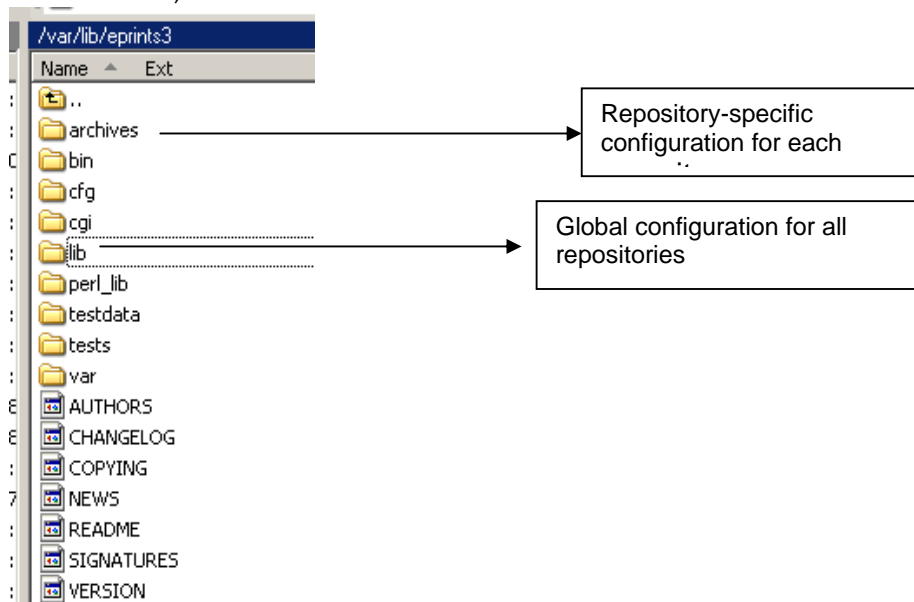
1) Given below is the Architecture of the Eprints Installation



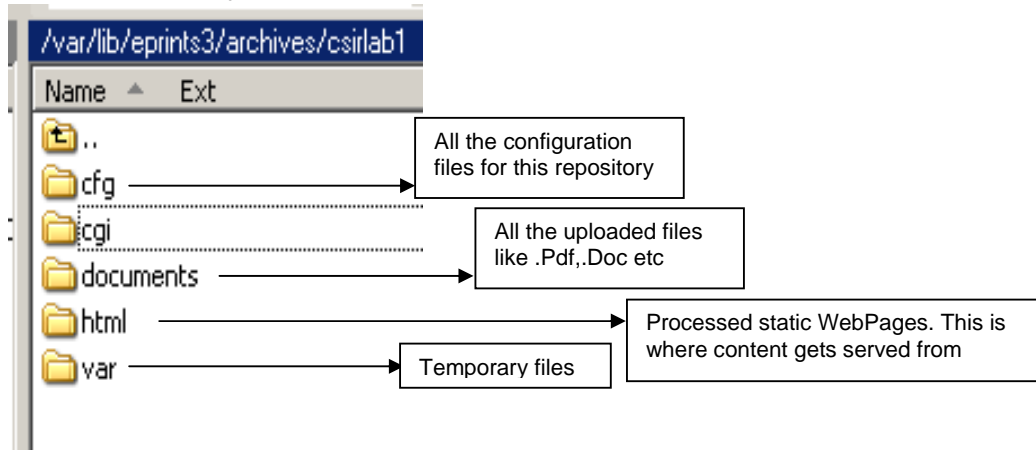
The important components of the Architecture are as follows:

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

2) After installation the folder structure are as follows(Considering Eprints is installed in /var/lib):



After installation and configuration of IR, for example say “csirlab1”, a folder with “csirlab1” will be created under “eprints3/archives/” folder. The structure of this folder will be like this:



## 4.1 Home page

The branding of the repository starts with giving a welcome look and useful information on the home page. Positioning of the information on the home page and presentation style may have flexibility according to one's choice.

To change the layout of the pages, there is a default template defined in the “eprints3/archives/cfg/lang/en/templates/default.xml” file. If one is familiar with html tags and having basic skills in designing, then this layout can be changed. Here one can change the look and feel of the Top/Bottom and how the middle (content portion) is placed. Required images or stylesheets to static content directory can be copied. Pins can be used to tell EPrints where to insert content in the layout. Once you are done, Run:

```
$ eprints3/bin/epadmin reload csirlab1
```

and if you have used any additional images/stylesheets, move them to the live area of the site by running:

```
$ eprints3/bin/generate_static csirlab1
```

The proposed changes are as follows:

### 4.1.1 Masthead

The header – the top area – is usually occupied by the masthead. This is usually an image file. It is recommended to have similar masthead that of the institute's website to maintain consistency. The masthead may also contain the name of the repository.

Suppose you have a picture to be used here is “mylogo.gif”. Copy this image file in to “eprints3/archives/cfg/static/images” folder. Open file “branding.pl” from “eprints3/archives/cfg/cfg.d/” folder and locate the following line and replace the “sitelogo.gif” with “mylogo.gif”

```
$c->{site_logo} = "/images/sitelogo.gif";
```

Once you are done, Run:

```
$ eprints3/bin/epadmin reload csirlab1  
$ eprints3/bin/generate_static csirlab1
```

### 4.1.2 Welcome Message

The main page area may have a welcome text that tells about the institute and the repository in short say in 4-5 lines. If you desire to have more information, may be the text continues with a link 'more' to a new page (some sample text messages in the Annex I).

### 4.1.3 Number of Records

The dynamic information on the total number of records in a repository on the home page demonstrates its strength. This incrementing number also gives a frequent visitor a confidence that there is something for him/her to look into.

A plugin has been developed for this (to be got from Dr. Francis Jaykanth).

### 4.1.4 Recent Submissions

It is customary to provide information on the home page itself about the recent publications added to the repository. This data is dynamically captured from the database and brought to the main page. This is to help the visitor reach to the content directly rather than going through search/browse mechanism every time. A plugin has been developed for this (to be got from Dr. Francis Jaykanth).

### 4.1.5 Join Search Harvesters

Web harvesting help end-user in Web-based search and retrieval, commonly referred to as search engines, by indexing the content that is pertinent to the audience for which the harvest is intended. Content is thus virtually integrated and made searchable as a separate Web application. Such harvesters exist at institutional, national and international levels as well as for a given focused discipline. Links from home page of IR to the following search harvesters are thought appropriate. Also one should ensure arrangements that these harvesters harvest the data from your IR (and check periodically that this is being done):

- CSIR Harvesters at
  - o URDIP (<http://www.csircentral.net>) and
  - o NAL (<http://csirpubs.nal.res.in/>)
- Indian National Archive maintained by CASSIR (<http://casin.ncsi.iisc.ernet.in/oai/>),
- Globally popular harvesters –
  - o OAIster (<http://oaiSTER.umd.umich.edu/o/oaister/>)
  - o Google (<http://www.google.com/>) and
  - o Google Scholar (<http://scholar.google.com/>)

Do find harvester(s) in the area of your IRs domain. Such an arrangement helps in increasing number of hits and downloads from your IR.

### 4.1.6 Submission Policy and Copyright Issues

One can also define fixed content for: home page, about page, help pages, error page and information page. These configuration files are stored in the "eprints3/archives/cfg/lang/en/static" directory. Different Policies like Metadata Policy, Data Policy, Content Policy, Submission Policy and Preservation Policy are defined in the Policies.xpage. To define the policies, web tool is available at "http://www.openoer.org/tools/en/policies.php". Generate the policies and copy in the policies.xpage file.

### 4.1.7 Links to the Institute's and Library Websites

It is customary to promote the institute's website to which the IR belongs and other research information groups (for example KRC's website) that the institute provides.

### 4.1.8 Footer

The footer – as the name suggests – is at the bottom of the page and should at least contain information on (i) the agency responsible for development and maintenance of the website, (ii) contact information, and (iii) the software used.

## 5.0 Subjects (Defining Classes)

EPrints has a facility to pre-define the subject areas under which the literature is to be deposited. This is a useful grouping of the subjects that help visitor in browsing the items present in the repository by subject of his/her choice. This is similar to the classification schemes for library books wherein main classes and sub-classes or further division downwards exist in the form of 'parent-child' relations / family tree.

The laboratories of CSIR have freedom to define these.

### 5.1 Guidelines

A broad definition of the subjects - maximum two levels - is ideal. There are advantages in restricting this to such a broad level:

- Justice to place complex type of publications in one group only (for exclusiveness).
- The incrementing number of records in each subject category impresses visitor by its size.
- Small schema aesthetically looks beautiful and permits the user to make quick decision.
- The purpose of 'browsing' (this schema permits) on repository is to get information that is specifically needed and also know what exists in periphery.
- It may be borne in mind that the subject categories so created are mutually exclusive of each other in an array.

### 5.2 Illustrative Example

#### *Chemistry*

- *Physical chemistry*
- *Analytical chemistry*
  - *Qualitative analysis*
  - *Quantitative analysis*
- *Organic chemistry*
  - *Biochemistry*
  - *Polymer chemistry*
- *Inorganic chemistry*

Once you make the tree of the subjects as illustrated above, there are two ways to define this in the IR.

### 5.3 Menu Level Operation

Log-in as administrator and click on "Admin" link -> select "Config. Tool" tab -> click on "Edit Subject" button, A Edit subject (Top level) screen will appear wherein under Subject ID: ROOT two (Library of Congress Subject Areas and University Structure) are defined. In the text box Add child to subject node: ROOT, type subject string (this is internal code for the subject that you are likely to define). For above illustrative example, type 'Chem' as a Subject ID String and click 'Add Child' button. Now you see a option to type 'Subject Name' for a subject string 'Chem'. Type 'Chemistry' here, opt for 'no' radio button (because you do not want any one to add their record under this main subject (class) and click 'Save Changes' button. A new subject in the root by name 'Chemistry' would get created. This now permits you to add a child for the parent 'Chem' by repeating the process of adding subject string initially and subject name. Repeat this process until you complete your definition. In the similar fashion, you can delete the subjects. The 'child' requires to be deleted first before the 'parent'.

## 5.4 Defining in the File

Create a text file with any name say "subjects.txt" in the given format

Key : Name of Subject : Root/Child : Can add Item(1) Or not(0)

Now import the subjects/divisions:

```
$ eprints3/bin/import_subjects csirlab1 subjects.txt
```

Example: subjects.txt

```
Section: Subjects
#####
#
# SUBJECTS
#
#####

Chem:Chemistry:ROOT:0
Phys:Physical Chemistry:Chem:1
Anal:Analytical Chemistry:Chem:0
Qual:Qualitative Analysis:Anal:1
Quan:Quantitative Analysis:Anal:1
Org:Organic Chemistry:Chem:0
Bioc:Biochemistry:Org:1
Poly:Polymer Chemistry:Org:1
Inor:Inorganic Chemistry:Chem:1
```

Whenever you change anything, always restart the web server so that all the changes are reflected.



## 6.0 Divisions

EPrints has a facility to predefine and browse by Divisions (these are defined in the same file that of subjects. However, CSIR being a dynamic system, the new divisions get formed from existing staff or get amalgamated. Such dynamics is likely to affect the IR's functionality (no new input is likely from the divisions so amalgamated, the inputs of the staff moved to the new divisions can not be grouped easily under new division, etc.). Besides, the interdisciplinary nature of research today, requires inputs from various disciplines and such inputs from various divisions for a published literature are common. Therefore division of literature in IR does not warrant.

Three steps are required to be taken:

1. Remove the divisions so defined if any (by default eprints have predefinition of divisions under heading 'University structure'). Delete these by Log-in as administrator and click on "Admin" link -> select "Config. Tool" tab -> click on "Edit Subject" button. Select 'University Structure' to reach to the bottom of the child and use delete buttons to delete complete structure. Alternately, edit 'subjects' file (where subject categories have been defined) and remove the lines pertaining to Division.

2. Removing the browse button below masthead:

Here is how the menu items of eprints look like.



We want to remove 'Browse by Division' button.

Go to the file:

...../eprints3/archives/<repository name>/cfg/lang/en/templates/default.xml

In that file following text would be seen. Please delete the red portion and save the file.

```
Section: Division

<td align="left"><ul class="ep_tm_menu"><li><a
href="{<config{http_url}&#92;&#92;}"><epc:phrase
ref="template/navigation:home" /></a></li><li><a
href="{<config{http_url}&#92;/information.html"><epc:phrase
ref="template/navigation:about" /></a></li><li><a
href="{<config{http_url}&#92;/view/year/"><epc:phrase
ref="bin/generate_views:indextitleprefix" /><epc:phrase
ref="viewname_eprint_year" /></a></li><li><a
href="{<config{http_url}&#92;/view/subjects/"><epc:phrase
ref="bin/generate_views:indextitleprefix" /><epc:phrase
ref="viewname_eprint_subjects" /></a></li><li><a
href="{<config{http_url}&#92;/view/divisions/"><epc:phrase
ref="bin/generate_views:indextitleprefix" /><epc:phrase
ref="viewname_eprint_divisions" /></a></li><li><a
href="{<config{http_url}&#92;/view/creators/"><epc:phrase
ref="bin/generate_views:indextitleprefix" /><epc:phrase
ref="viewname_eprint_creators" /></a></li></ul>
</td>
```

Restart apache..

The divisions link will disappear.

3. Removing the data entry field from worksheets

Comment out following line in "...eprints3/archives/cfg/workflows/eprint/default.xml".

```
<!-- <component><field ref="divisions"/></component> -->
```

## 7.0 Types of Documents

Librarians experience visitors' choice for a literature by type of the documents. It is useful in having the browse facility by the type of documents.

Open a file "*.....eprints3/archives/cfg/cfg.d/views.pl*" and add the following at the end before the end tag "】".

```
Section: Types of documents:
{
  id => "types",
  menus => [
    {
      fields => [ "type" ],
      reverse_order => 1,
      allow_null => 0,
      new_column_at => [10,10],
    }
  ],
  order => "creators_name/title",
  variations => [
    "creators_name;first_letter",
    "type",
    "date;res=year",
    "DEFAULT" ],
}
```

### 7.1 Adding Data Elements for Data Entry

Also edit the "*eprints3/archives/cfg/lang/en/templates/default.xml*". Locate the line where all the links are defined for "home", "About", "Browse by....." and add the new link for "browse by Document Type":

```
<li><a href="{ $config{http_url} }/view/types/"> <epc:phrase
ref="bin/generate_views:indextitleprefix" /> <epc:phrase ref="viewname_eprint_types" /> </a>
</li>
```

Once you are done, Run:

```
$ eprints3/bin/epadmin reload csirlab1
$ eprints3/bin/ generate_view csirlab1
```

## 8.0 Adding Fields

### 8.1 Guide/Supervisor

EPrints permit normal flow of a data entry by opening up templates that contain needed fields for a given document type. However, for the thesis type of document the field Guide/Supervisor's name is missing. This is an important piece of information and therefore we need to add this field to the definitions. To do so, following is required to be done.

To add an additional field say "thesis\_guide" in the document type "Thesis". Then edit "*eprints3/archives/cfg/cfg.d/eprint-fields.pl*" and add following at the end:

```
{
    'name' => 'thesis_guide',
    'type' => 'name',
    'family_first' => 1,
},
```

Edit the "*eprints3/archives/cfg/workflow/eprint/default.xml*" and locate the line

```
<epc:if test="type = 'thesis'">
```

below this line add the following:

```
<field ref="thesis_guide" required="yes" />
```

Open database for this archive. Alter the table ***eprint*** and add following columns:

- thesis\_guide\_family                    varchar(255)
- thesis\_guide\_given                    varchar(255)
- thesis\_guide\_lineage                varchar(255)
- thesis\_guide\_honourific            varchar(255)

Edit the "*eprints3/archives/cfg/lang/en/phrase/zz\_webcfg.xml*" and add following line:

```
<epp:phrase id="eprint_fieldname_thesis_guide">Guide  
Name</epp:phrase>
```

Once you are done, Run:

```
$ eprints3/bin/epadmin reload csirlab1            and restart the Apache.
```

### 8.2 CSIR Author ID

The repositories would eventually be linked to the R&D portal of CSIR. This link for a particular author can be established by the unique employee ID that is being thought of in ERP Solution. It is therefore essential to add this as a subfield in every author name (currently a text box for email address exists that could be replaced). This is important because the rendering of author names differs in scientific documents than what otherwise would exist on CSIR's R&D Portal.

To add a new sub-field “cisrid” for each creators to store the CSIR ID of the creator. Then edit “*eprints3/archives/cfg/cfg.d/eprint-fields.pl*” and add following under creators, below sub-name => ‘id’:

```

    {
        'sub_name' => 'id',
        'type' => 'text',
        'input_cols' => 20,
        'allow_null' => 1,
    },
    {
        'sub_name' => 'csirid',
        'type' => 'text',
        'input_cols' => 20,
        'allow_null' => 1,
    }

```

Add this {

Create a new table “**eprint\_creators\_csirid**” with following fields:

- eprintid            int(11)            Primary Key    Not Null
- pos                int(11)            Primary Key    Not Null
- creators\_csirid    varchar(20)

(NOTE: Don’t change the Table/Fields name)

Edit the “*eprints3/archives/cfg/lang/en/phrase/zz\_webcfg.xml*” and add following line:

```

    <epp:phrase id="eprint_fieldname_creators_csirid">CSIR
    ID</epp:phrase>

```

Once you are done, Run:

**\$ eprints3/bin/epadmin reload csirlab1**            and restart the Apache.

### 8.3 Field for Citation Link

As a value addition, the record on IR can be linked to Google Scholar to fetch the number of citations a particular document received. This needs to be defined as a field with default content (event: on click search in Google Scholar (GS) to search text in the title field and retrieve GS details in separate tab).

.....Under Construction.....

### 8.4 Unique OAI-Identifier to IR

(IMPORTANT for Harvesting) Give unique OAI-Identifier to IR. Open “*eprints3/archives/cfg/cfg.d/oai.pl*” and replace “**generic.eprints.org**” with hostname of the IR like “*csirlab1.csircentral.net*” at the following line:

**\$oai->{v2}->{archive\_id} = "generic.eprints.org";**

## 9.0 Standards for Metadata Description

Entry of clean unambiguous metadata is the key to successful information retrieval. The librarians have traditionally been doing this job under the concept 'cataloguing'. A clean data entry helps in viewing a data that is self explanatory for making decisions as well as building valuable indexes for comprehensive as well as exclusive retrieval of desired information. It is therefore essential that certain standards/ practices are to be followed. Following tips for each of the fields would help in making the data entry flawless. Note that such tips are also available (or could be incorporated/ modified as needed) in the data entry work sheets of the software that you would use.

### 9.1 General

Every punctuation mark ( , ; . ? ! ) follows space and then the next word. The punctuation mark must be immediately next to the word without space.

### 9.2 Title

The title should not end with a full stop, but may end with other punctuation mark if original title has it. There is no way to type italic text, please enter it normally. A subtitle should proceed with a colon [:]. Use upper case for the first letter of the first word of title & subtitle and words with proper nouns.

**Example: Dynamics of chromophoric dissolved organic matter in Mandovi and Zuari estuaries: A study through in situ and satellite data**

### 9.3 Abstract

Copy-paste here a formal abstract or summary of the document if exists after editing for removing any carriage returns and complicated text. No complicated text formatting is possible. Use 'sup( )' superscripted and 'sub( )' for subscripted words/symbols. Convert mathematical symbols to normal text.

**Example:** Mandovi estuary is a tropical estuary strongly influenced by the southwest monsoon. In order to understand, sources and fate of particulate organic nitrogen, suspended particulate matter (SPM) collected from various locations, was analyzed for particulate organic carbon (POC) and particulate organic nitrogen (PON), **delta sup(13)C sub(POC)**, total hydrolysable amino acid enantiomers (L- and D- amino acids) concentration and composition. **Delta sup(13)C sub(POC)** values were depleted (-32 to -25 ppt) during the monsoon and enriched (-29.6 to -21 ppt) in the pre-monsoon season implying that OM was derived from terrestrial and marine sources during the former and latter season, respectively. .... Based on D-amino acid yields, bacterial OM accounted for 16-34% (23.0 + or -6.7%) of POC and 29-75% (47.9 + or -18.7%) of PON in monsoon, and 30-78% (50.0 + or -15%) of POC and 34-79% (51.2 + or -13.3%) of the PON in premonsoon in the estuary. ....

### 9.4 Creators/Authors/Editors

It is always (wrongly) considered that the name element has first name, middle name and last name (surname or family name) and also written in *this* order. Internationally, in scientific world, the name that is written last is picked as family name and other names are turned to initials. Indian names are complex by nature. Some variations identified are:

- Names with spaces
- Names with no last/family names
- Names indicating social grouping

- Female names followed by Father/ Husband's name (and no family name)
- Names with suffixes

Refer Annexure III for some tips on how should one come over these variations (and use them while rendering in this repository).

## **9.5 Corporate Creators**

At times, the individuals are not named but the responsibility of creation goes to the organizational/corporate entities. In such a situation, write the name of the organization followed by the place.

**Example: National Institute of Oceanography, Goa**

## **9.6 Publication Details/ Citation**

This field describes the source where the document under reference is published. Various data elements form part of this field. Standards for some important elements are described below.

### **9.6.1 Journal Title**

The title of a journal or magazine. Do not use abbreviations. Enter in full. First letter of every word should be in upper case.

**Example: Journal of Experimental Marine Biology and Ecology**

### **9.6.2 Title of Book**

Title usually found on the title page (first page of book).

**Example: Tsunamis: Detection, Monitoring, and Early-Warning Technologies**

### **9.6.3 Publisher**

A person, firm or corporate body responsible for placing an item on the market. Do not use abbreviations, enter in full.

**Example: Oxford University Press**

### **9.6.4 Number of Pages**

Just the total number of pages of the document in case if it is a monograph/book/thesis type. Do not prefix or suffix 'p.' or 'pp.'. Sum total and write total pages in case if the pagination for introductory pages, main text, annexure, etc., is numbered differently. In case if the pagination is not continuous (at times pagination per chapter is numbered), write 'vp' for various pages.

**Example: 239**

### **9.6.5 Series Name**

The name of the series to which the document belongs and series number. Use upper case for the first character of every word in series except for articles (a, an, the), prepositions (of, for), etc. In EPrints, Series name forms a part of Publication Details.

**Example: Elsevier Oceanography Series, 25**

### **9.6.6 Official URL**

The canonical URL for this item. Some journals require you to provide a link to their version.

**Example: [www.ias.ac.in/currsci/25jul2011/222.pdf](http://www.ias.ac.in/currsci/25jul2011/222.pdf)**

### **9.6.7 Page Range**

Numerals only. The sequence of pages of the item. Do not enter 'pp. '

**Example: 221-234 Write numbers in full (not 221-34)**

### **9.6.8 Identification Number**

The unique identification number or a DOI. Such identification numbers are normally available to the monographs whereas DOI have become common for the journal articles on the Internet.

### **9.6.9 Related URLs**

URLs that are related to this item. This is to be used only when there is any other record related to the record under description and its URL. For example, a second edition of a book is an update of first edition of the same book. Links could be provided to these two records interchangeably.

### **9.7 Project**

This field could be used to link the repository record to the CSIR R&D Portal to retrieve data from IRs to the Portal and link it up to the respective projects. Type the project number of which this document is an outcome of the research project.

### **9.8 Contact Email Address**

If the full-text is not available to the public, then requests to view the full-text are sent to this email address. The email address is not made available to the public. Traditionally the library takes care of honouring the reprint requests to the authors of the institute.

### **9.9 Keywords**

These could either be natural language terms that describe the item. However, use of standard vocabulary (thesaurus) is strongly recommended. Standard vocabularies help in better retrieval of items than natural language terms.

### **9.10 Additional Information/Description**

This field is an open text for any additional information on the item that the standard fields do not describe.

### **9.11 Subject Category**

In EPrints, this is the last part of the description. These are predefined categories as indicated in the configuration section. The user has to select one of the categories to place the item in the repository.



## 10.0 Copyright Issues

### 10.1 The Problem

Commercial publishers, when accept the paper for publications, normally ask authors to transfer copyrights to the publisher. This practice limits the institutions to archive the published papers (normally .pdf formatted files) on the repositories.

### 10.2 The Solutions

#### 10.2.1 If the copyrights are transferred....

Most of the publishers today, permit archiving post-print (i.e. final draft post-refereeing) manuscript on the IRs. There are several publishers now who have changed their policies and even permit the authors to place the publisher's .pdf formatted files too. "Publisher copyright policies & self-archiving" (<http://www.sherpa.ac.uk/romeo/>) website is a part of Sherpa/RoMeo Project that compiles policies of most prominent publishers and updates from time to time. One can search either by journal or by publisher on this website to ascertain publisher's policies. The website has grouped publishers and journals in to following categories:

- **Green/Blue** publisher: permitting to archive post-print (i.e. final draft post-refereeing) or the PDF file of the publisher
- **Yellow** publisher: permitting to archive pre-print (i.e. pre-referring) and
- **White** publisher: A publisher that does not formally support archive.

This website has also developed a page for the publishers allow authors to deposit the publisher version or PDF of their article in an Institutional Repository, without fee or an embargo or with an embargo (restriction to place the paper after certain period of publication). (<http://www.sherpa.ac.uk/romeo/PDFandIR.php?la=en>)

#### 10.2.2 If the copyrights are not transferred....

It is worth noting here that most of the Indian publishers do not ask authors to transfer the copyright to them for publishing a paper. In case of all such papers, the publisher's pdf files could be placed on the IR.

In short, experience shows that an institute can upload 99% of the post-prints (final draft post referring document) publications on the IR without any problems.

#### 10.2.3 Solution for items with 'white' publishers

As indicated above, there are still some publishers who do not permit authors to archive their papers on IRs. However, there is no restriction of storing metadata about the paper and an abstract. A repository in such rare situations will have the metadata of such papers on an archive and 'restricted access' to the pdf file. Repository software permits such items to have a restricted access and then facilitates visitors to 'request for a reprint' of this paper. The process is extremely simple. The visitor clicks on a button, fills a form with his/her email address and submits the same. The contact person would receive such a request who can again by a click of button post the e-print of the paper to the requester as email attachment thus making the document available in offline mode to the requester.

## 11.0 Work-flow

Having mandated the establishment of repository and that information about every published paper and thesis must be archived, there needs to be a mechanism to check that the repository has all the publications of the institute/ laboratory.

There might be several mechanisms to have these checks and balances. However, one of the CSIR laboratories has been most successful in capturing all publications of the laboratory by linking this submission to the Annual Performance Appraisal System. CSIR as a part of ERP is building R&D Portal as one of the modules under which this link is to be established.

## 12.0 Backup

Backup is an important task to protect data against accidental loss, database corruption, hardware failures, and even natural disasters. Once the repository is populated and running, one of the important aspects of maintenance task is the routine work of taking backup. The best practice is to regularly backup repository archive/ database and keep multiple sets of the same. An ideal situation would be to have a back up every day replacing earlier files of that day (The back up taken on earlier Monday is replaced by backup of this Monday and so on) so that at a given time seven generations of back ups are available. It is essential that the back up should be taken on a system that is not housed in the IT centre.

Irrespective of the IR software used, we need to backup two things i.e. (i) the directory where all the configuration files and archive files exist and (ii) the database.

EPrints does not have an inbuilt feature to generate backup regularly.

The backup command can be run manually or can be automated by executing a database backup query from a PHP/shell script and setting a Cron/scheduled tasks (win) ( a time-based job scheduler), option.

Backup in Eprints:

- 1) The directory /opt/eprints3/ (or whatever it is called) including the symbolically linked directories.
- 2) MySQL database that your archive uses.

Following script would help in taking daily (7 generations) backup in Eprints.

```
-----
#!/bin/sh

# Backup script for eprints archive.
TODAY="date +%a ";

# Archived files will be in this directory
OUTPUT_DIRECTORY_PATH="/home/";

SERVER_IP=192.168.9.122; #put your IP
FILENAME_HEADER="IRBackup_on_"$TODAY;
OUTFILE_DIRECTORY_BACKUP=$OUTPUT_DIRECTORY_PATH/$FILENAME_HEADER"
_directory.tar.gz";
OUTFILE_DATABASE_BACKUP=$OUTPUT_DIRECTORY_PATH/$FILENAME_HEADER"
database.txt";

clear
echo "*****",
echo "The script takes backup of entire IR archive at "$SERVER_IP;
echo "*****",

echo "OK, starting now..."

sleep 3

# EPrints directory backup
( cd /opt ; tar -czvf $OUTFILE_DIRECTORY_BACKUP eprints3/* )

# Database backup
echo "Enter database password for IR archive";
```

```
/usr/bin/mysqldump --opt -u root -p nal >> $OUTFILE_DATABASE_BACKUP #change the  
database name
```

```
ls -lh $OUTFILE_DIRECTORY_BACKUP
```

```
ls -lh $OUTFILE_DATABASE_BACKUP
```

```
echo "The backup process has been completed..."
```

```
echo "Exiting..."
```

---

## 13.0 Statistics

Statistics of IR not only help to observe the usage of the repository as a whole (for the manager) and per item download (for the author), but also gives clue on how changes to the repository affect its usage.

.....Under Construction.....

### Welcome messages in a repository

**Eprints@CMFRI** is the Open Access Institutional Repository of [Central Marine Fisheries Research Institute](#). Research outputs of CMFRI - journal papers, conference papers, reports, theses, patents etc. - are uploaded/self-archived by CMFRI scientists who do research on fisheries and related areas. Interested users can freely download and use documents as most of them are directly accessible and full-text downloadable. 'Request Copy' forms can be used for documents to which direct full-text download is restricted due to publisher embargo. See [repository policies](#) here. Thanks for using the repository! (<http://eprints.cmfri.org.in/>)

NAL's Institutional Repository is the digital archive of the research output of our scientists. This knowledge base covers journal articles, conference papers, technical reports, presentation/lectures, preprints, Thesis, images etc. One can browse the documents by author, division, subject, date and document type. Both simple and advanced search facilities have been provided. ICAST hopes to make this archive exhaustive very soon. NAL scientists are welcome to submit their publications on their own by registering. (<http://nal-ir.nal.res.in/>)

Welcome to Digital Repository Service of National Institute of Oceanography (CSIR), India. DRS@nio tries to collect, preserve and disseminate different institutional publications (journal articles, conference proceeding articles, Technical reports, thesis, dissertations, etc). You can search, browse and access publications of NIO from this collection.

## Annexure II: DSpace and EPrints Fields for Data Entry – Comparison

One template for all	Fields in DSpace	Fields in EPrints	Journal article	Chapter in a book	Monograph	Book	Thesis
	Document type	Document Type	*+	*+	*+	*+	*+
*+	Title	Title	*+	*+	*+	*+	*+
*	Other (Alternative) title						
*	Abstract	Abstract	*	*	*	*	
		Thesis type (Masters/PhD/Eng/Other)					*+
		Monograph type (TR, Proj Rep, Documentation, Manual, Working paper, Discussion paper, Other)			*+		
*	Authors	Creators (Authors)	*	*	*+	*	*+
		Corporate Creators	*	*	*	*	*
		Editors		*		*	
*	Citation	Publication details					
		Refereed version	*+	*+		*+	
		Status (published/in press/submitted/unpublished)	*+	*+	*+	*+	*+
		Institution			*		*+
		Department			*		*+
		Journal title	*+				
		Title of the book		*+			
*	ISSN	ISSN	*				
*	ISBN	ISBN		*		*	
		Place of publication		*	*	*	
*	Publisher	Publisher	*	*+	*+	*+	
		Official URL	*	*	*	*	*
		Volume	*			*	
		Number	*	*		*	
		Page range	*	*			

*+	Date of issue	Number of pages	*	*	*	*
		Date	*	*	*	*+
		Date type (publication/submission/completion)	*		*	*+
*	Identifiers	Identification number	*	*		
*	Series/Report No	Series Name		*	*	
		Related URLs	*		*	*
*	Sponsors	Funding Agencies	*	*	*	*
		Project	*	*	*	*
		Contact email address	*	*	*	*
*	Subject keywords	Keywords	*	*	*	*
*	Description	Additional info	*	*	*	*
		Comments and suggestions	*	*	*	*
		Subject category	*+	*+	*+	*+
*	Language					

**Legend in this table**

Field exists (\*)  
 mandatory field (+)

Independent fields in DSpace



## Annexure III

Some problems with the Indian names:

Names with spaces

Name	Gets converted to	Recommended practice (remove spaces between compound names and write that in family name column)
<i>North Indian Names</i>		
R. Nagin Chand	Chand, R.N.	R. NaginChand
Naresh Kumar	Kumar, N.	NareshKumar
R. Sen Gupta	Gupta, R.S.	R. SenGupta
<i>South Indian Names</i>		
M. Dileep Kumar	Kumar, M.D.	M. DileepKumar
P.A. Loka Bharathi	Bharathi, P.A.L.	P.A. LokaBharathi

Names with no last/family names

Name	Gets converted to	Recommended practice (write first name in the family name column and other names/initials in other names)
Shankar Doraiswamy (first name and father's name)	Doraiswamy, S.	D. Shankar
Pankajakshan Thadathil (first name and place name)	Thadathil, P.	T. Pankajakshan

Names indicating social grouping

Name	Gets converted to	Recommended practice (write social grouping name as family name)
C. Satyanarayana Murty	Murty, C.S.	This is accepted
L.V. Gangadhara Rao	Rao, L.V.G.	This is accepted
N. Purna Chandra Reddy	Reddy, N.P.C.	This is accepted
Shanta Nair	Nair, S.	This is accepted

Name followed by Father/ Husband's name (and no family name)

Name	Gets converted to	Recommended practice (write first name in the family name and husband/father's names in other names)
Chandralata Raghukumar	Raghukumar, C.	R. Chandralata
Soniya Sukumaran	Sukumaran, S.	S. Soniya
Geeta Subramanian	Subramanian, G.	S. Geeta

Names with suffixes

Name	Gets converted to	Recommended practice (remove spaces between name and suffixes and write in family name with other names in other names column)
C.B. Lalitambika Devi	Devi, C.B.L.	LalitambikaDevi, C.B.
Rishemjit Kaur	Kaur, R.	RishemjitKaur
Jaswant Singh	Singh, J.	JaswantSingh

Write your names correctly every where and preserve YOUR identity!