

RUBRIC Toolkit: Establishing a Pilot Repository

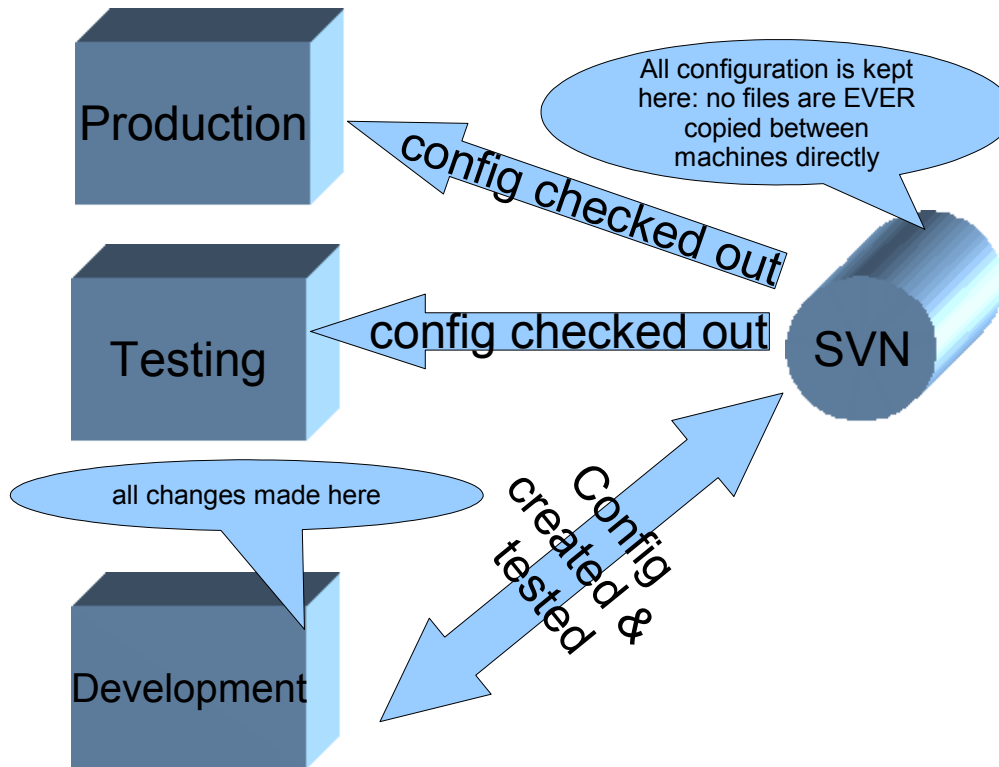
Planning a Pilot

Developing a pilot Institutional Repository (IR) helps to plan and fine tune issues of policy, workflow, marketing and resourcing prior to a public launch.

When planning the pilot phase, consider:

- **The length of time allocated to the pilot phase**
Ensure a deadline for migrating the pilot over to a live instance is set at the beginning of the process. Deliverables include:
 - customisations (branding, look and feel)
 - import of data
 - data analysis (such as quality control checking)
 - workflow development and training
 - policy development
- **Responsibility for the pilot**
Consider technical, administrative, marketing and budgetary roles. Will separate people be involved? How will this group communicate? Will there be a hierarchy? What are the implications for their current roles? What level of authority and decision making are granted?
- **Software selection**
[System Options](#)¹ provides some assistance on the process of evaluating and selecting appropriate software; but you will need to establish your own institution's process for software evaluation and procurement.
- **Interoperability**
Identify systems the IR will have a relationship with, e.g. library catalogue, research database, or an authentication system such as LDAP
 - define the nature of the relationships with other identified systems
 - outline the requirements of each system relationship (e.g. information/reporting, harvesting, exporting, importing)
 - determine which way the data will flow between identified systems
 - determine the frequency of data updates or transfers between the identified systems
 - what processes need to be established for data and workflow purposes?
 - who will be responsible for mapping the processes and how long will it take?

Sample software selection model



Multi-tiered Development & Version Control

Using Multiple Tiers

RUBRIC advises the use of a three-tiered approach for evaluating, configuring, and testing repository software rather than maintaining a single copy of repository software on a server and patching it repeatedly, doing upgrades and configuration changes on the live system.

1. Production tier	Pilots are for primarily for demonstrating the look and feel of the repository, as well as being used as a marketing tool
2. Testing tier	Testing is where code that has been developed in the development tier is placed to ensure it functions as desired. This tier is kept as similar to the production tier as possible
3. Development tier	Development is where technical staff learn about the software being

RUBRIC is supported by the Systemic Infrastructure Initiative as part of the Commonwealth Government's Backing Australia's Ability - An Innovative Action Plan for the Future (<http://backingaus.innovation.gov.au>)

evaluated, about piloting and eventually managing in production. Every change to configuration is done first in development and the changes committed into a [Subversion repository](#)² before any code is moved to the next tier

[Subversion revision control software](#)³ provides well managed version control rather than relying on copying configuration files from machine to machine. It is not essential to use the Subversion software, but some form of similar version control is strongly recommended.

[Virtualised infrastructure](#)⁴ provides a way to easily work with tiers 1 and 2. Virtualisation is not essential for these tiers, but it is strongly recommended for the development tier as it encourages developers to treat instances of repository software as disposable and be practiced at installing and configuring software.

RUBRIC suggests installing a repository with Tier 3 and proceeding as follows:

- 1. Technical staff install the repository software on virtual machines**
Development machines are used to try out new versions of software, develop all configurations, such as colour schemes, templates and index configuration.
 - all configuration is kept under version control via Subversion (SVN)
 - all changes are made on a development machine and checked-in to Subversion
- 2. Technical Staff edits the configuration of the live repository**
Technical staff check out the same configuration as is running in production, change it then commit it back to SVN for testing.
- 3. Technical staff tests new configuration**
Technical staff checks out the changes they made on their development machines into testing and has independent confirmation that their changes are correct. Depending on the setup in use, this configuration may then need to be copied to a release branch. See the Subversion book by [Collins-Sussman et al](#)⁵. (2004) for guidance.
- 4. Technical staff deploys to production**
Technical staff checks out the latest copy of the production-ready configuration. If there has been a mistake, then it is easy for the development staff to roll-back to an earlier, working version of the configuration.

Configuration

Institutional branding can be added to the pilot IR with minor cosmetic adjustments:

- logo
- institutional name
- contact information
- help files and an FAQ

RUBRIC is supported by the Systemic Infrastructure Initiative as part of the Commonwealth Government's Backing Australia's Ability - An Innovative Action Plan for the Future (<http://backingaus.innovation.gov.au>)

It is also advisable to review the available database fields in the software and whether these are sufficiently comprehensive for your purposes. Information management experts in your organisation may want to review usability, consistency and interoperability standards with other systems.

Configuration Check List for IR Managers:

- **look and feel:** is there an institutional 'look' that the repository needs to adhere to?
- **impacting context:** such as a purpose statement
- **browsing:** are users able to browse by all the fields necessary? If not, how do you create new browse fields?
- **searching:** is there a simple search available? Is it obvious?
- **advanced searching:** is the advanced search easy to use? Is it accurate?
- **help:** is there a clear way for users to find online help?
- **information:** is there information for users who wish to submit to the repository? Who do they contact?

This configuration table will assist with planning the configuration work and prioritising tasks.

Ease of Configuration:

- configuration on a repository should be an up-front and infrequent task
- configurations will probably involve staff with a high level of technical skill
- configuration is easier to manage if stored in files with version control software utilised
- it is not recommended that configuration work, such as customising input forms be done on a live server
- when selecting software, a graphical user interface to make configuration easier is of no benefit if the system overall is harder to manage
- develop, test and deploy configuration using a multi-tiered approach

[Technical Reports](#)⁶ on the RUBRIC website provides further information on configuration.

Customizing Screens

When branding a pilot IR, consider the following:

- simplicity keeps users happier than fancy functionality
- keep wording as unambiguous as possible
- different browsers (ie Internet Explorer, Mozilla Firefox) may display the same page differently. If you heavily customise the skin of your web interface, test that the changes work in as many different browsers as possible

- suggestions for a front page to the repository include:
 - introduction to the repository: which group it represents, its purpose, etc
 - “Simple Search”
 - links to Browsing, Advanced Search, Help, FAQs and Contacts

It is advisable to keep good records of your customisations for two reasons:

- to easily recreate the customisations as part of your disaster recovery process
- to easily review and reassess requirements for customisations when you upgrade or migrate to alternative software over time

Document information about customisation to ensure you can easily recreate the look and feel in new software versions and in the event of disaster recovery.

[Making DSpace Your Own](#)⁷ (Donahue and Salo 2006) is another example to assist with software customisation.

Metadata Requirements

Data about an item being deposited to an IR must be captured on entry. Known as “metadata”, this is structured information provided about or alongside a resource to inform users what that resource is about and how they can use it. It captures descriptive information about the item (for example, the author, title, date, publisher) as well as value-added information (such as a subject classification, administrative data, rights data).

Every IR manager needs to know:

- why metadata is important
- who manages the metadata, including quality control and policy
- what metadata schemas are available
- how to choose a metadata schema for resource discovery purposes
- when to apply the metadata
- how to link to resource discovery services

The Metadata section provides a detailed overview of these aspects of metadata and its role in an IR.

Pilot to Production

To transition from the pilot system to a production system, refer to

- Pilot to Production and

RUBRIC is supported by the Systemic Infrastructure Initiative as part of the Commonwealth Government's Backing Australia's Ability - An Innovative Action Plan for the Future (<http://backingaus.innovation.gov.au>)

- Managing a Repository

These sections give you the tools to plan the process of going live and consider the associated management issues of a live instance.

References and Further Reading

Refer to the Further Reading section at the end of the Toolkit for bibliographic details of works referenced in this section.

“RUBRIC Toolkit: Establishing a Pilot Repository” produced May 2007



Copyright⁸ 2007 RUBRIC⁹

- 1 ../docs/System_Options.htm
- 2 [http://en.wikipedia.org/wiki/Subversion_\(software\)](http://en.wikipedia.org/wiki/Subversion_(software))
- 3 [http://en.wikipedia.org/wiki/Subversion_\(software\)](http://en.wikipedia.org/wiki/Subversion_(software))
- 4 http://en.wikipedia.org/wiki/Virtual_Infrastructure#VMware_Virtual_Infrastructure
- 5 <http://www.oreilly.com/catalog/0596004486/>
- 6 <http://www.rubric.edu.au/techreports/index.htm>
- 7 <http://hdl.handle.net/2144/240>
- 8 <http://creativecommons.org/licenses/by-sa/2.5/au/>
- 9 <http://www.rubric.edu.au/>