

Excellence in Research for Australia (ERA)

SEER

(System to Evaluate the Excellence of Research)

Repository

Testing Strategy for 2010

1 Introduction

1.1 Purpose

The purpose of this document is to define the repository testing strategy and testing procedures for ERA 2010.

1.2 Audience

The Repository Testing Strategy is to be read and used by Institution Repository Managers, CAIRSS, and other internal executive stakeholders related to repository management. This document is to serve as the testing approach for the entire SEER 2010 Repository Testing life cycle only.

1.3 Testing Objective

The overall objective of the repository testing strategy is to provide institutions clear instructions, on the steps needed in order to manage and test their repositories.

2 Testing Strategy

The Repository Testing Strategy aims at facilitating a more comprehensive, systematic, uniform and independent approach to the depositing of and access to research outputs for the Excellence in Research for Australia (ERA) initiative for 2010.

2.1 Repository Testing Process

The key way the Repository Testing Strategy aims at achieving this is by providing institutions access to their repositories and the ability to manage repository details through a SEER user interface.

By providing each institution with the tools to manage their repository data for ERA 2010 with accuracy and authority, a higher degree of certainty ensues during the ERA Peer Review process.

There are three key processes of the Repository Testing and Management process. These are:

- Pre-Submission Repository Testing
- Submission Repository Testing
- Post-Certification Repository Management

2.1.1 Pre-Submission Repository Testing

The main goal of the Pre-Submission Repository Testing is to provide institutions with confidence in their management of repositories for ERA and with sufficient time to manage and test these repositories completely and accurately.

To achieve these goals, ERA has provided a platform called SEER Manage Repository Testing, in which repository managers will work. This platform effectively replicates what will be used by ERA peer reviewers to access an institution's research outputs. Working within such a platform enables institutions to test their repositories and fix any anomalies quickly and authoritatively in an environment that is identical to that used throughout the 2010 ERA Submission and Evaluation processes.

To achieve the desired outcome for this testing, institutions need to:

i. Register all repositories that will be used for ERA 2010

Within the SEER Manage Repository Testing application, institutions will be required to manually register each repository that will house research outputs submitted for ERA 2010.

ii. Populate these repositories with at least 10 research outputs that will subsequently be submitted for ERA 2010.

Once institutions register a repository, they can then populate them with sample research outputs that will be submitted for ERA 2010.

During this step, institutions should take care in ensuring that their research output links are aligned to the appropriate repository. For example, if the School of Environment and Built Design uses a different repository for its research than that used by the Faculty of Arts, then research outputs from both these faculties will be housed in different repositories. This should be reflected when institutions populate SEER Manage Repository Testing.

At this stage, it is also best to include research outputs with a broad range of file types (see section 3.0 for allowed data types). This way, the testing becomes wider in scope.

iii. Test all aspects of and record results for each research output in SEER Manage Repository Testing.

Once each repository is populated, institutions are able to test manually each individual research output in their registered repository and record the test outcome.

All aspects of a research output should be tested. For example, if a metadata page is being tested, institutions should make sure they test the actual research output link within the metadata page as well.

If a situation arises during this testing where the repository domain link needs to be updated, institutions must ensure they update manually the correct domain name within the research output links.

In order to advise and support institutions in managing their repositories, CAIRSS will be given temporary access to the institution's repository through the SEER user interface. This access will be solely for the purpose of testing and assisting institutions in setting up their repositories for the ERA submission process. An institution's repository authentication details will be hidden from CAIRSS users. If you do not wish CAIRSS to assist in the testing of your repositories, you may opt out by unchecking the relevant box within your SEER Manage Repository Testing interface.

2.1.2 Submission Repository Testing

During the ERA submission process, institutions will have access to their repositories and the research outputs within these through SEER Submit. When an institution uploads their data for submission, SEER Submit will automatically be populated with the research outputs submitted by that institution for ERA 2010. Institutions will be able to test the links to their research outputs within the Peer Review tab in SEER Submit. Institutions should ensure that all research outputs are accessible through SEER.

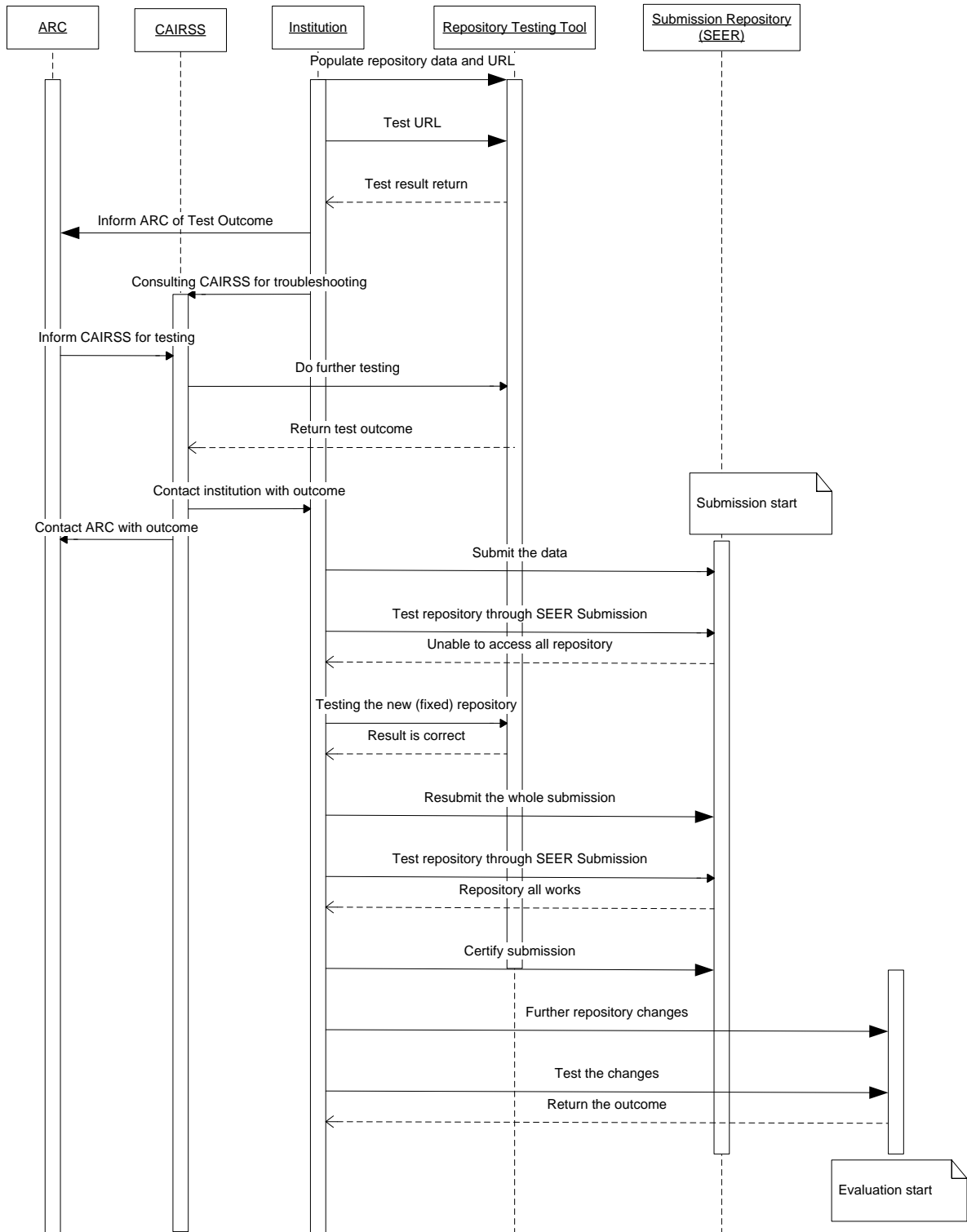
Should problems arise with the accessing of a research output, institutions can still use the SEER Manage Repository Testing platform to diagnose their repository- or research output-related issue.

Once an institution is satisfied that the data included in its ERA 2010 submission is correct, and the submission is certified, institutions will be able to manage and test research outputs within their repositories solely through the SEER Submit platform.

2.1.3 Post-Certification Repository Management

Once an institution has finalised and certified its data for ERA 2010, the SEER Submit platform cannot be used to add or remove research outputs; it can only manage existing data. This can include updating information pertaining to pre-existing data and testing that data. This means institutions can still update their repository domain details, authentication details and research output links without resubmitting their submission. Any changes made within the SEER Submit platform after certification will not require re-submission by the institutions.

2.2 Engagement Model



3 Expectations / Obligations

Where a research output is being made available for peer review and is also available within a repository, research output metadata must include references to Research Outputs in Digital Assets (RODA) files. In cases where direct repository access is not possible, other means may be used to ensure that ERA peer reviewers, and other users authorised by the ARC, have access to the additional information associated with research outputs. Digital Object Identifiers (DOIs) are supported.

For the purposes of an ERA submission, an institution may have more than one institutionally-supported repository. The institution must be able to arrange for ERA access to each repository involved.

Institutions are expected to submit the following RODA file types:

- Adobe PDF (i.e. .pdf);
- Images (i.e. .gif, .jpg, .jpeg, .bmp, .png, .tif, etc.);
- Microsoft Word (i.e. .doc, .docx);
- Microsoft PowerPoint (i.e. .ppt);
- Text (i.e. .txt); and
- Popular Video/Audio format (i.e. .mp3, .mp4, .avi)

Institutions are required to set up secure digital repositories to store research outputs where those outputs are not publicly accessible.

Only HTTP Basic and IP are acceptable repository authentication strategies.

Institutions must configure their ERA repositories to use only standard HTTP and HTTPS ports.

4 Technical Overview

4.1 RAM Detailed Description

The Repository Access Mechanism (RAM) in SEER uses the HTTP Client programming framework. When a reviewer selects a research output link through the SEER *Evaluation* application, the request is sent to the RAM module. The RAM module then independently crafts its own HTTP request on behalf of the user (proxy request). In this way, SEER is the originator and not the end user, thus ensuring anonymity is maintained. Once a response is received from the target server, the RAM again produces its own HTTP response and sends this back to the reviewer, again isolating the end user from the server.

This mechanism is depicted in *Figure 1* below.

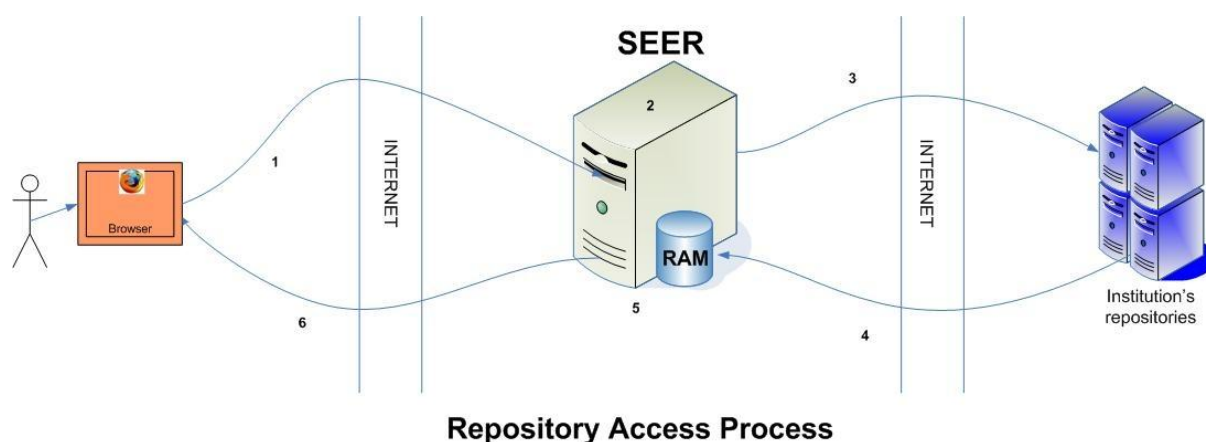


Figure 1 – The Conceptual Workings of the RAM

Ideally, Step 4 of *Figure 1* would return a simple web resource (such as a simple html page or pdf). During 2009, however, this was not often the case, leading to very complex HTML\JavaScript pages being returned from the repositories during Step 4. This meant that, in Step 5, SEER would often encounter problems translating documents back to the ERA peer reviewers.

4.2 Form vs Basic Authentication

If authentication into an institution's repository is required, SEER only allows two types:

1. BASIC authentication (<http://tinyurl.com/nxz4m>)
2. IP authentication.

Form-based authentication is not permitted for ERA.

When accessing an artifact that is secured through BASIC security, the user's browser will deliver a pop-up window asking for the username and password (credentials) to access the artifact. Without a successful username and password combination, the user will not be granted access.

Figure 2 displays a typical browser pop-up window asking for BASIC authentication details.



Figure 2 – BASIC pop-up window (Permitted by SEER)

Conversely, form-based authentication is backed by an application in the repository that will take the user input and determine whether or not to allow access. This type of authentication, which is not supported by SEER, is depicted in *Figure 3* below.

DREW

uLogin

Please enter your uLogin ID and password to continue. [Click here](#) for more information about uLogin accounts and passwords.

Please Log In

The iChain protected site you are attempting to access requires you to log in.

uLogin ID:

Password:

Login

uLogin is a service of the department of [Computing and Network Services](#) at [Drew University](#). Questions/Comments? Send mail to cns@drew.edu.

Drew University, 36 Madison Avenue, Madison NJ 07940, (973) 408-3000

POWERED BY eDirectory

Figure 3 – Form-based login (Not permitted by SEER)

The goal for all institutions is to ensure that only BASIC or IP authentication is used, or no authentication at all, to access material within a repository.

4.3 RAM and JavaScript

JavaScript is a dynamic language which can be used to provide interactive features to a web page. It is a *client* side technology. This means that it is executed in the client browser only; web servers do not execute JavaScript, they only create the scripts for client side execution. Drag and drop is a more elaborate example of JavaScript in action. JavaScript can be used to create almost any part of the web page.

The challenge for SEER is that it sometimes needs to 'guess' what the page will look like. For example, even a simple anchor tag may well be produced by a very complex JavaScript process.

The link below depicts a sample url which provides a link to a research output from an institution:

```
<a
href=https://repository.university.edu.au/ref=1234&sessionid=ABCDEFGHIJKL>Output</a
>
```

It is possible that the application generating this link on the institution side is represented instead by the following JavaScript:

```
var protocol = 'https';
var host = _host;
var ref=_refId;
var sessionId = _sessionId;
var displayText = 'Output';
var url = protocol + host + "refid = " + _refId + "&sessionId=" + _sessionId;

displayText.anchor("outputAnchorurl");
document.write(displayText);
```

As illustrated in the JavaScript described above, much work and logic would be necessary for SEER to re-write this URL correctly. SEER needs to determine the values of *host*, *ref* and *sessionId* – information which would likely be contained in another cryptic location elsewhere in the script. Once SEER has these details, and has pieced together all of the information required to create the link, then and only then can it serve this page back to the user.

This is only one example and does not represent the most complex scenarios encountered in 2009. Other examples can be cited which required much more development and effort on the part of SEER. In one or two cases, the structure was so complex that it became impossible for SEER to guarantee that errors would not be made or that anonymity would be maintained.

For the reasons presented above, SEER recommends that institutions provide both metadata, and RODA artifacts which are mostly JavaScript free. This will ensure consistent, anonymous access to research outputs.

5 Key Dates (Timelines)

Activity	Timeframe	Responsible	Platform Available	Troubleshooting Point Of Contact
Pre-Submission Repository Testing	12 March 2010 – 07 May 2010	Institutions	SEER Manage Repository Testing	CAIRSS
Submission Repository Testing	01 June 2010 – 29 July 2010	Institutions	SEER Submit and SEER Manage Repository Testing	ARC
Post-Certification Repository Management	30 July 2010 – Evaluation	Institutions	SEER Submit	ARC