



Australian Government
Australian Research Council

SYSTEM TO EVALUATE THE
EXCELLENCE OF RESEARCH (SEER)
DRAFT TECHNICAL SPECIFICATIONS
PHYSICAL, CHEMICAL AND EARTH
SCIENCES (PCE) & HUMANITIES AND
CREATIVE ARTS (HCA) CLUSTERS

Consultation



Excellence in Research for Australia (ERA) Initiative

January 2009

RESEARCH in the national interest - enabling the future

© Commonwealth of Australia 2009

This work is subject to the laws of copyright. You may download, display, print and reproduce this material in unaltered form only (retaining this notice) for your personal, non-commercial use or use within your organisation. All other rights are reserved. Requests and inquiries concerning reproduction and rights should be addressed to: Commonwealth Copyright Administration, Copyright Law Branch, Attorney-General's Department, Robert Garran Offices, National Circuit, Barton ACT 2600, Fax: 02 6250 5989, or submitted via the copyright request form on the website <http://www.ag.gov.au/cca>. Requests and inquiries concerning reproduction and rights should be copied to: The General Manager, Research Excellence, Australian Research Council, GPO Box 2702, Canberra ACT 2601.

CONTENTS

1.	ERA-SEER OVERVIEW	6
1.1	Purpose of the Technical Specifications	6
1.2	Nature of ERA Submissions	6
1.3	Role of the SEER Platform	6
1.4	SEER User Summary	7
1.5	Scope of Specifications	7
	1.5.1 <i>In Scope</i>	7
	1.5.2 <i>Out of Scope</i>	7
1.6	Target Audience	7
1.7	Disclaimer	8
1.8	Version Control	8
1.9	Attachments	8
1.10	Related Documents	8
2.	SEER TECHNOLOGY PLATFORM.....	9
2.1	Summary	9
2.2	Process of ERA Submission.....	10
2.3	Anatomy of an ERA Submission	10
2.4	System Security	11
2.5	International and Special Characters.....	12
2.6	Service Availability	12
	2.6.1 <i>SEER</i>	12
	2.6.2 <i>SEER Helpdesk</i>	12
	2.6.3 <i>Repository Helpdesk</i>	12
3.	SUBMISSION DATA STRUCTURE AND FORMAT.....	13
3.1	Overview	13
3.2	ERA XML Schema Files.....	14
3.5	Usage	16
	3.5.1 <i>Introduction</i>	16
	3.5.2 <i>XML Namespaces and Versioning</i>	16
	3.5.3 <i>Encoding an Institution Submission</i>	16
	3.5.3 <i>Nominating Researchers</i>	16
	3.5.4 <i>Encoding Research Outputs</i>	16
	3.5.5 <i>Encoding Australian Competitive Grant Income</i>	17
	3.5.6 <i>Encoding Other Research Income</i>	17
	3.5.7 <i>Encoding Applied Measures</i>	18
	3.5.8 <i>Encoding Esteem Measures</i>	18

3.6	Codes	18
	3.6.1 <i>FoR Codes</i>	18
	3.6.2 <i>Institution Codes</i>	18
	3.6.3 <i>Ranked Journal Outlet Codes</i>	18
	3.6.4 <i>Language Codes</i>	18
	3.6.5 <i>Eligible Australian Competitive Grant Program Codes</i>	18
	3.6.6 <i>Research Theme Codes</i>	18
	3.6.7 <i>Patent Issuing Country Codes</i>	19
3.7	Uniquely Identifying Journal Articles.....	19
4	SYSTEM USER INTERFACE	20
4.1	Overview.....	20
4.2	Authentication and Authorisation	20
4.3	Supported Browsers	20
5.	INSTITUTIONALLY-SUPPORTED REPOSITORY REQUIREMENTS	21
5.1	Research Outputs	21
5.2	Specifications for Research Outputs.....	21
	5.2.1 <i>Research Output Digital Assets</i>	21
	5.2.2 <i>Research Output Metadata</i>	21
	5.2.3 <i>Multiple Institutionally-Supported Repositories</i>	21
	5.2.4 <i>Supported Formats</i>	21
	5.2.5 <i>Size Limits</i>	22
	5.2.6 <i>HTTP Ports</i>	22
	5.2.7 <i>Repository Service Availability and Helpdesk Arrangements</i>	22
	5.2.8 <i>Authentication</i>	22
	5.2.9 <i>Stability of Links</i>	23
6.	SYSTEM ENVIRONMENT	24
	APPENDIX A: ERA-SEER TECHNOLOGY PACK INVENTORY	25
	ERA Submission XML Schemas.....	25
	Code Tables	25
	Documentation	25
	Submission Examples.....	25
	APPENDIX B: GLOSSARY OF SUBMISSION DATA ITEMS	27
	Common Data Items.....	27
	<i>Research Outputs</i>	28
	<i>Research Income</i>	31
	<i>Esteem Measures</i>	32
	<i>Applied Measures</i>	33

PCE Cluster Data Items 33

HCA Cluster Data Items 34

Research Outputs 34

Esteem Measures..... 34

Applied Measures..... 35

APPENDIX C: GLOSSARY OF TERMS AND ACRONYMS 36

Technical Terms..... 36

1. ERA-SEER OVERVIEW

1.1 Purpose of the Technical Specifications

These *ERA-SEER Technical Specifications* are provided for eligible higher education providers (hereafter ‘institutions’) who are making submission(s) as part of the Excellence in Research for Australia (ERA) initiative. The policies and processes related to ERA submissions are addressed in the *ERA Submission Guidelines*. The *ERA Submission Guidelines* provide an overview of, and specify the process and information requirements for, institutional submissions to ERA. These *Technical Specifications* build upon the *ERA Submission Guidelines* to provide technical instruction for institutions electronically preparing and submitting ERA submissions.

The information technology platform that supports the implementation of ERA is called the System to Evaluate the Excellence of Research (SEER). These *Technical Specifications* relate specifically to institutional interaction with SEER, thereby serving as a technical complement to the *ERA Submission Guidelines*. This edition of the *ERA-SEER Technical Specifications* relates to the first two discipline clusters scheduled for submission via SEER, namely:

- *Cluster One*: the Physical, Chemical and Earth Sciences (PCE) cluster; and
- *Cluster Two*: the Humanities and Creative Arts (HCA) cluster.

This document addresses:

- submission data requirements;
- mechanisms of system interaction;
- institutional repository requirements; and
- relevant technologies and security mechanisms related to interactions with the SEER platform.

1.2 Nature of ERA Submissions

Institutions are required to collate and package their eligible information in a defined XML data format when participating in ERA. Institution submissions are then submitted to the SEER platform via the ‘bulk data upload’ feature of SEER’s web-based user interface.

Institution ‘research officers’ (see section 1.4) will also use the user interface to verify the validity of submitted data and to access reports related to their submission(s). SEER allows subsequent uploads of revised submissions to support corrections of submitted data by replacing previous uploads prior to submission finalisation.

There is no ‘system-to-system’ automated data exchange for the upload of institutional submissions – submission data transfers are undertaken via online user interaction. Note that subsequent interaction between SEER and institutional repositories will occur to facilitate the peer review of research outputs for relevant discipline clusters.

Institutions are expected to prepare their ERA submissions using the following resources:

- *ERA Submission Guidelines*;
- *ERA-SEER Technology Pack (inc. ERA-SEER Technical Specifications)*.

The *ERA-SEER Technology Pack* is a collection of electronic files that are intended to assist institutions preparing an ERA submission. The pack is available for download from www.arc.gov.au/era/seer.htm. Its contents are outlined in **Appendix A** of this document.

1.3 Role of the SEER Platform

The SEER technology platform aims to:

- enable institutions to successfully upload and verify electronic ERA submissions via a well-defined XML data format; and
- provide institutions with the ability to create and view reports on their submissions.

1.4 SEER User Summary

The following are identified users of SEER:

- *Research Officers (or their equivalent)*: Institutional employees responsible for maintaining their institution's Institution Profile and Institution Submission data;
- *Vice-Chancellors*: Represent the interests of an institution and are responsible for certifying that an Institution Submission is ready for inclusion in the ERA evaluation process;
- *Heads of Research (Pro/Deputy Vice-Chancellors, Research)*: Responsible for overseeing research activities and policy within their institution. They are also responsible for finalising an Institution Submission prior to its certification by the institution's Vice-Chancellor; and
- *Repository Administrators (or their equivalent)*: Institution employees with responsibility for the day-to-day operation and maintenance of institutional systems housing research outputs and other research items of interest to ERA.

References within this document to 'institutional users' refer to the abovementioned users. The following users are considered 'ARC users':

- *REC Chairs*: Research Evaluation Committee (REC) Chairs are REC members with the responsibility for assigning units of evaluation to expert reviewers and peer reviewers in the assignment phase, and with the responsibility of chairing the REC's during the committee phase. They are not geographically located within an ARC site and may be resident nationally or internationally;
- *REC Member/Expert Reviewers*: Responsible for reviewing research items submitted by institutions for the purpose of evaluation. REC Members/Expert Reviewers will not be geographically located within an ARC site and may be resident nationally or internationally; and
- *Peer Reviewers*: Responsible for evaluating a sample of research outputs associated with a particular Unit of Evaluation. Peer Reviewers are not geographically based within the ARC.

1.5 Scope of Specifications

1.5.1 In Scope

The following areas are covered in this document:

- data content and structures required for packaging an ERA PCE and/or HCA cluster submission;
- institutional system interfaces;
- mechanisms of data transfer (inc. access to institutional repositories);
- system security; and
- SEER platform availability and support procedures.

1.5.2 Out of Scope

The following are out of scope for this document:

- policy and business matters related to ERA submissions;
- guides to underlying standards and technologies;
- methods for aggregating/preparing submission information *within* institutional IT systems; and
- ERA processes or workflows that are not directly supported by SEER.

1.6 Target Audience

This is a detailed document that is intended for a technical audience, specifically:

- institution IT staff;
- institution research office staff; and

- third-party software developers (where applicable).

1.7 Disclaimer

The *ERA-SEER Technical Specifications* provide information for interoperating with SEER. They are not intended to provide IT systems implementation instructions for institutions. Institutions are responsible for their respective systems development and data gathering to meet the requirements of these specifications. The Commonwealth accepts no responsibility for any loss, damage, liability or expense incurred by any institution resulting from any use of these *ERA-SEER Technical Specifications* for any purpose.

The *ERA-SEER Technical Specifications* may be changed from time to time. The latest version of the *ERA-SEER Technical Specifications* can be found on the ERA website (www.arc.gov.au/era/seer.htm).

It is the responsibility of each institution to ensure that the latest version of the *ERA-SEER Technical Specifications* is used.

1.8 Version Control

Changes to this document will be managed through controlled versioning.

Versioning will be done individually on two key components, namely: this technical specifications document and the *ERA-SEER Technology Pack*. Institutions will be informed of all new technical specifications and/or *ERA-SEER Technology Pack* versions via the ARC website. It should be noted that minor revisions to the technology pack will not result in reissuing of this document.

1.9 Attachments

This document is distributed electronically within the *ERA-SEER Technology Pack* which includes:

- *ERA-SEER Technical Specifications* document (PDF);
- ERA cluster submission XML schema files (XSD);
- sample cluster submission XML documents (XML);
- schema help files (HTML); and
- SEER code (reference data) tables (TXT, CSV).

Release Notes are included with the *ERA-SEER Technology Pack*. These notes include instructions on how to open and use the packaged files.

1.10 Related Documents

Related policy documents issued by the ARC include:

- *ERA Submission Guidelines*;
- *ERA Indicator Principles*;
- *ERA Indicator Descriptors*; and
- *ERA Evaluation Guidelines*.

Related external references include:

- W3C specifications (e.g. XML), see <http://www.w3.org>;
- ABS (2008), Australian and New Zealand Standard Research Classification (ANZSRC), Cat. No. 1297.0 ABS, Canberra;
- ABS (2005), Australian Standard Classification of Languages (ASCL), Cat. No. 1267.0 ABS, Canberra; and
- Attorney-General's Department, Protective Security Manual (PSM 2005).

Other technical documentation may be sent to stakeholders to provide additional advice on integration with SEER if new issues or requirements arise. All documentation will be posted on the ERA website.

Key parties to the operation of the SEER platform include:

- *ERA Institutions*: Institutions eligible for participation in the formal ERA process as detailed in **Appendix A** of the *ERA Submission Guidelines*;
- *Publishers*: Third-party publishers that maintain institutional repositories of research outputs that may be referenced in an institution's submission;
- *Reviewers*: Those who have been assigned to undertake evaluation activities;
- *Bibliometric Providers*: Commercial entities who have been engaged by the ARC to provide citations reference data for one or more discipline clusters; and
- *Australian Research Council (ARC)*: Commonwealth statutory authority which is responsible for administering ERA.

Information related to 'identified users' of SEER are outlined in section 1.4 above.

2.2 Process of ERA Submission

The process of an ERA cluster submission is characterised by institution research officers preparing a submission as an electronic file, uploading the file to SEER for validation, and potentially repeating the process until such time as the institution is satisfied to finalise and then certify (via its Vice-Chancellor) the submission for evaluation. The SEER-supported processes that take place during ERA cluster submission are as follows:

- each institution packages its own research item metadata records into an XML document that is conformant with the relevant ERA cluster submission XML schema. This submission XML document is then uploaded to SEER by logging on to (and interacting with) the SEER web-based user interface;
- each submission XML document is scanned for syntactic and semantic validity against ERA specified guidelines. Institutional users view the results of validation for their submission on the SEER user interface. If errors are noted then the submission XML document may be corrected and re-uploaded at the convenience of the institutional user (replacing the previous submission);
- SEER may verify (on an ad hoc basis) the network accessibility of digital assets linked by submitted research outputs (as located in one or more institutional digital repositories);
- an institutional user may log on to the SEER user interface to view sample reports based on their submission;
- the institution 'Head of Research' or equivalent is required to log on to the SEER user interface to indicate 'finalisation' of a valid cluster submission, at which point the submission may no longer be modified;
- the institution Vice-Chancellor or equivalent is required to log on to the SEER user interface to indicate 'certification' of a finalised cluster submission, at which point the submission is ready for evaluation; and
- following the closing date for a cluster submission, SEER no longer allows input to related submissions and all certified submissions are prepared for assignment to REC members and/or expert/peer reviewers.

2.3 Anatomy of an ERA Submission

As noted, all institution submission data must be packaged within an XML document that conforms to the ERA submission XML schema provided for a given cluster. This packaging requires technical staff to collate relevant data from various information sources within the institution's research support systems and to

transform this data into an ERA submission XML document structure. Institutions are expected to implement this ETL (Extract, Transform, Load) process using their own systems development resources (utilising off-the-shelf ETL tools where applicable).

The ERA submission XML schema requires that a conformant submission XML document contains:

- a brief institution identifier;
- all *research item* metadata records being submitted for the ERA submission, consisting of:
 - *research outputs*;
 - *research income*;
 - *esteem measures*; and/or
 - *applied measures*.
- metadata about eligible researchers who are affiliated with the institution for the purpose of specifying eligible research outputs and esteem measures in the submission.

Note that SEER only serves to record snapshots of submissions as uploaded by institutional users. Each institution will remain the authoritative source for any data that is used for its submission. Although institution submission data may be re-submitted several times prior to certification, individual submission data items may not be edited interactively within the SEER system.

A full description of the ERA submission XML schemas for both PCE and HCA submissions is provided in section 3 of this document.

The actual submission file uploaded must be either:

- a complete ERA PCE or HCA submission XML document (e.g. 'pce-submission.xml'¹); or
- a ZIP² compressed archive of a complete ERA PCE or HCA submission XML document (e.g. 'hca-submission.xml.zip').

The maximum file size for an uploaded submission file must not exceed 25MB.

Note that an ERA submission may be accompanied by Background Statements. SEER provides a user interface screen through which institution research officers may enter the Background Statement for each eligible two-digit FoR level discipline within a cluster. Background Statements are limited to 10,000 characters for each discipline.

2.4 System Security

The SEER web-based user interface is only accessible to authenticated users. Each eligible system user will be granted an individual user account by the SEER Helpdesk (see section 2.6.2). A user account consists of a username and password combination which is required to authenticate to the SEER platform (via web form login page). Each user account will be associated with only one institution and is only able to access data related to that institution.

User login accounts will be created by the ARC and distributed upon request to ERA Liaison Officers prior to the opening of an ERA cluster round. User accounts will remain active unless instruction is received from an ERA Liaison Officer to disable an account or misuse is suspected. Users may change their own passwords via the SEER web-based user interface.

The mechanism of system-to-system authentication by the SEER platform with particular institutional repositories (for access to RODA files) is expected to be managed on a case-by-case basis in cooperation with individual institutions and/or publishers.

All network communications with the SEER platform are secured for confidentiality using the Secure Sockets Layer (SSL) protocol.

¹ Note that any desired filename may be used for a submission file.

² The standard ZIP file compression format, per MIME media type: application/zip.

2.5 International and Special Characters

International and special characters are supported via the UTF-8 character encoding standard for publication and outlet titles in research output metadata. UTF-8 is only required for internationalised research output metadata fields. There are no encoding requirements for RODA files.

Textual data that contains only 7-bit ASCII (i.e. Roman) characters will have the same encoding under both ASCII and UTF-8. This means that there are not likely to be any issues arising from systems which only support ASCII encoding. Similarly, there are not likely to be any issues converting from encodings which only support Roman characters for which a lossless conversion to ASCII exists. If systems which contain research output metadata use characters that do not exist in the ASCII character set then issues may arise when converting these characters.

Data stored in systems that support other Unicode encodings (e.g. UTF-16) should not present any issues because a simple conversion from characters in these encodings into UTF-8 will exist.

Research outputs with non-English language titles should list the title in the native script and then indicate the language of the title using the ABS classification code for the language. An English translation of relevant titles should be provided to ensure that a research output is appropriately assigned for evaluation. Where the native script of the title is not in roman characters a phonetic sounding of the title of the full title should be provided. The XML schema defines all the attributes required to provide this information.

2.6 Service Availability

2.6.1 SEER

The SEER system will be in operation 24 hours a day, 7 days a week during the PCE and HCA submissions (20 April 2009 until 8 May 2009). Note that a regular scheduled maintenance window exists from 7:30pm until 9:30pm AEST on Thursday evenings (during which time service may be disrupted). If the system is to be taken offline for any additional scheduled maintenance period(s) then advice will be given to participating institutions with 24 hours notice.

2.6.2 SEER Helpdesk

The ARC will provide Helpdesk support for institutions throughout and before the ERA cluster submission period during normal AEST business hours.

Queries regarding SEER should be directed to the SEER Helpdesk on (02) 6287 6755 or email seer@arc.gov.au.

2.6.3 Repository Helpdesk

Institutions are required to provide a Helpdesk contact to respond to any problems with institutionally-supported repository access. The Helpdesk contact is assumed to be the ERA Technical Officer unless the ARC is advised to the contrary. The ARC will liaise with the Helpdesk contact to resolve issues relating to access to research outputs during the following periods:

- PCE and HCA submissions period:
 - during a cluster submission period SEER may verify the accessibility of RODA files within institutionally-supported repositories via the individual links provided with submissions;
 - the repository Helpdesk should resolve any reported access problems within 48 hours;
- Post-submission period:
 - following closure of the submission period, institutions will be required to support requests for SEER access to their repositories to assist with evaluation processes.

3. SUBMISSION DATA STRUCTURE AND FORMAT

3.1 Overview

The format to be used for packaging and submitting an ERA PCE or HCA submission is as an (optionally compressed) XML document. An ERA submission XML document must conform to the relevant ERA cluster submission XML schema as supplied in tandem with these *Technical Specifications*. XML schemas are made available via '.xsd' files in the *ERA-SEER Technology Pack* as outlined in **Appendix A**.

The ERA **PCE submission data** structure is broadly composed of:

- *Institution Submission*: top-level grouping for an institution's submission data.
 - *Institution*: brief identifying details for the institution making the submission.
 - *Researchers*: metadata records about eligible researchers that the institution has nominated as affiliated at the census date.
 - *Research Output Items*: metadata records for all submitted research output items. For example, books, book chapters, journal articles and conference publications.
 - *Esteem Measures*: metadata records for all submitted esteem measures items. For example, editorial roles of A* or A ranked journals, elected fellowships of Learned Academies, and nationally competitive research fellowships.
 - *Research Income Items*: metadata records for all submitted research income items. For example, Australian Competitive Grants (Category 1), other public sector research income (Category 2), industry and other research income (Category 3), and Cooperative Research Centre research income (Category 4). Category 3 income is further broken down into Australian (Category 3i), international competitive peer-reviewed income (Category 3ii) and international other income (Category 3iii).
 - *Applied Measures*: metadata records for all submitted applied measures. For example, patents and research commercialisation income.

The ERA **HCA submission data** structure is broadly composed of:

- *Institution Submission*: top-level grouping for an institution's submission data.
 - *Institution*: brief identifying details for the institution making the submission.
 - *Researchers*: metadata records about eligible researchers that the institution has nominated as affiliated at the census date.
 - *Research Output Items*: metadata records for all submitted research output items. For example, books, book chapters, journal articles, conference publications, original (creative) works in the public domain, live performance works in the public domain, recorded (performance) public works; and curated or produced substantial public exhibitions, events or renderings.
 - *Esteem Measures*: metadata records for all submitted esteem measures. For example, editorial roles of A* or A ranked journals, contributions to prestigious works of reference, curatorial roles of prestigious events, elected fellowships of Learned Academies, nationally competitive research fellowships, and prizes or awards.
 - *Research Income Items*: metadata records for all submitted research income items. For example, Australian Competitive Grants (Category 1), other public sector research income (Category 2), industry and other research income (Category 3), and Cooperative Research Centre research income (Category 4). Category 3 income is further broken down into Australian (Category 3i), international competitive peer reviewed income (Category 3ii) and international other income (Category 3iii).
 - *Applied Measures*: metadata records for all submitted applied measures. For example, patents, registered designs, and research commercialisation income.

3.2 ERA XML Schema Files

The ERA cluster submission XML schema is provided in a number of files to subdivide the schema definition into modules.

3.2.1 Common XML Schema Files

The common XML schema files contain the definition of all the data elements that can be included in an ERA submission. The specific cluster schema files described in later sections are used to restrict the data elements available to those data elements which are eligible for that cluster.

A brief description of the each of the common SEER XML schema files follows:

- *common/seer.xsd* – defines the ‘seer’ namespace and includes all the other common XML schema files as part of that namespace;
- *common/institution.xsd* – defines the institution data element;
- *common/researcher.xsd* – defines the data elements used to specify eligible researchers;
- *common/researchOutput.xsd* – defines the data elements used to specify all eligible research output types;
- *common/researchOutlet.xsd* – defines the data elements used to specify all eligible research outlet types;
- *common/researchIncome.xsd* – defines the data elements used to specify eligible research income information;
- *common/esteemMeasure.xsd* – defines the data elements used to specify all eligible esteem measures;
- *common/appliedMeasure.xsd* – defines the data elements used to specify all eligible applied measures;
- *common/discipline.xsd* – defines the data elements used to specify Fields of Research (FoR) which are re-used in data elements specified in the researcher, researchOutlet, researchIncome and appliedMeasure schema files;
- *common/artefactTitle.xsd* – defines common data elements used to specify title metadata; and
- *common/type.xsd* – defines common types and enumerations that are used in the other schema files to restrict attribute values and enforce validation of data rules.

3.2.2 PCE Cluster XML Schema Files

The *pceInstitutionSubmission.xsd* XML schema file defines what constitutes a PCE cluster submission. This XML schema file specifies a new namespace for the PCE cluster and identifies that the PCE submission is composed of a common institution element; a common collection of researchers; a common collection research income data; a PCE-specific collection of research output items; a PCE-specific collection of applied measures; and a PCE-specific collection of esteem measures.

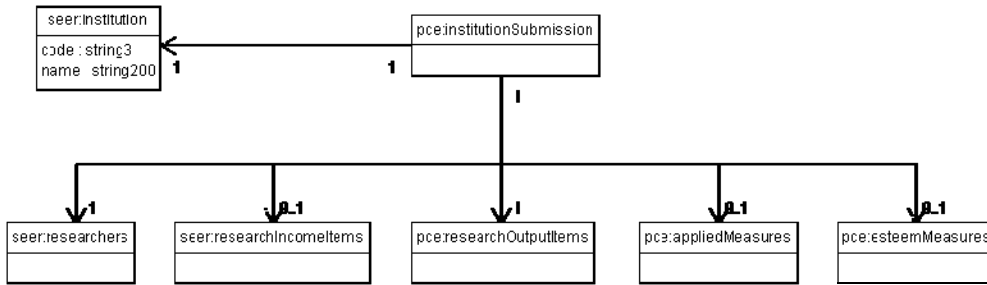


Figure 2: The composition of a PCE institution submission.

This schema file defines the PCE-specific collections by identifying which data elements defined in the common schema files are eligible for inclusion.

3.2.3 HCA Cluster XML Schema Files

The *hcaInstitutionSubmission.xsd* XML schema file defines what constitutes an HCA cluster submission. This XML schema file specifies a new namespace for the HCA cluster and identifies that the HCA submission is composed of a common institution element; a common collection of researchers; a common collection research income data; an HCA-specific collection of research output items; an HCA-specific collection of applied measures; and an HCA-specific collection of esteem measures.

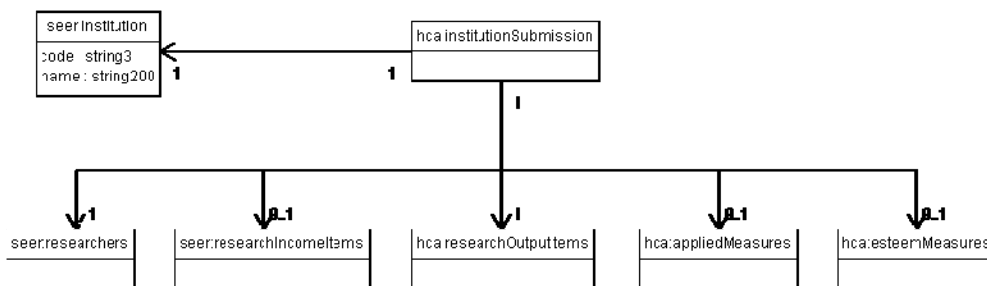


Figure 3: The composition of an HCA institution submission.

This schema file defines the HCA-specific collections by identifying which data elements defined in the common schema files are eligible for inclusion.

3.5 Usage

3.5.1 Introduction

From a policy perspective, the purpose and derivation of the content for an ERA cluster submission is discussed in the *ERA Submission Guidelines*. This section serves as a guide to the use of the ERA submission XML schema in terms of the technical process of developing conformant submission XML documents.

The ERA submission XML schemas are provided in the *ERA-SEER Technology Pack* and accompanied with examples of conformant XML submission documents and other related documentation as noted in **Appendix A**. Please find below a technical discussion of some aspects of the intended application of the ERA submission XML schemas. Note that XPath³ notation is used here to reference individual components of the schema.

Further documentation concerning the elements or attributes within the schema can be found in the `xml-doc/pce/index.html` and `xml-doc/hca/index.html` documentation and within the schema itself (via inline annotation elements in the schema).

Note that the scope of the ERA Cluster One and Cluster Two round is limited to research outputs from the PCE and HCA discipline clusters. Thus the scope of the data collected by the ERA submission XML schemas are correspondingly limited. Subsequent to the ERA Cluster One and Cluster Two round, revised XML schemas will be released to the sector to support collection of research items from disciplines beyond PCE and HCA. The revised schemas will be released to the sector in time to support future rounds of ERA.

3.5.2 XML Namespaces and Versioning

Versioning of any SEER XML schema is achieved via specification of in the XML namespace. The initial namespace for the SEER common XML schema is `http://seer.arc.gov.au/2009/seer/1`; the initial namespace for the PCE cluster XML schema is `http://seer.arc.gov.au/2009/pce/1`; and the initial namespace for the HCA cluster XML schema is `http://seer.arc.gov.au/2009/hca/1`.

3.5.3 Encoding an Institution Submission

The `/institutionSubmission` element taken from either the PCE or HCA namespace is the root element of a SEER XML data submission. The namespace of this element determines which cluster information is contained in the XML document.

The `/institutionSubmission/institution` element contains mandatory information describing the submitting institution. The allowed values for the `/institution/@code` attribute is defined in the provided SEER code table for institutions included in the *ERA-SEER Technology Pack*.

3.5.3 Nominating Researchers

Institutions are required to nominate researchers who meet the eligibility criteria set out in the *ERA Submission Guidelines* via the `/institutionSubmission/researchers` element. Included under this listing is the `researcher/@staffReference` attribute. This staff reference must be specified by an institution to differentiate each researcher within the submitted researchers list. The `researcher/@staffReference` attribute is utilised in the `/institutionSubmission/researchOutputItems/*/staffCreators/staffCreator/@staffReference` attributes and the `/institutionSubmission/esteemMeasures/*/@staffReference` attributes in order to relate an institution's submitted researchers to a particular research output or esteem measure.

3.5.4 Encoding Research Outputs

The `/institutionSubmission/researchOutputItems` element consists of a list of `researchOutputItemType` elements. Each research output item is associated with a research outlet and

³ <http://www.w3.org/TR/xpath>

consists of: a title; a list of creators; a list of staff creators; an optional list of electronic locations; an optional list of research themes; and an optional list of institutional units. Additional attributes and associated information is required depending on the specific type of research output.

3.5.5 Encoding Australian Competitive Grant Income

Institutions are required to provide information on each eligible grant for income from Australian Competitive Grants (Category 1). The `/institutionSubmission/researchIncome/australianCompetitiveGrants` element contains a list of `australianCompetitiveGrant` objects which specifies this information. Each `australianCompetitiveGrant` object contains up to three `incomeYear` elements which represents years within the research income reference period.

The `australianCompetitiveGrant/incomeYears/incomeYear/@acgrCode` attribute is a code representing an eligible Australian competitive grant program that can be found in the code table provided in the *ERA-SEER Technology Pack*. This attribute needs to be referenced on each year of a grant to cater for grant programs which have changed names over the life of a grant within the reference period.

The `australianCompetitiveGrant/incomeYears/incomeYear/@totalAmountRecieved` attribute is the amount of grant money that is received by an institution for the grant within a specified year.

For each year within the grant income is to be apportioned by four-digit Field of Research (FoR) code and a percentage allocation for that code.

3.5.6 Encoding Other Research Income

For each of the other research income categories (Category 2; Category 3i; Category 3ii; Category 3iii; and Category 4) the total income for the category is to be disaggregated for each year of the research income reference period and apportioned to a FoR code.

For Other Public Sector Research Income (Category 2) the `/institutionSubmission/researchIncomeItems/otherPublicSectorIncome` element contains a list of `incomeYear` elements. And the `otherPublicSectorIncome/otherIncomeYears/otherIncomeYear/incomeByDiscipline/valueByDiscipline` elements contain the breakdown of other public sector income by FoR within a specific year.

The `/institutionSubmission/researchIncomeItems/industryAndOtherResearchIncome` element is a container for the sub categories of Industry and Other Research Income (Category 3).

For Australian Industry and Other Research Income (Category 3i) the `/institutionSubmission/researchIncomeItems/industryAndOtherResearchIncome/australianIndustryAndOtherResearch` element contains a list of `incomeYear` elements. And the `australianIndustryAndOtherResearch/otherIncomeYears/otherIncomeYear/incomeByDiscipline/valueByDiscipline` elements contain the breakdown of income by FoR within a specific year.

For International Competitive, Peer Reviewed Research Income (Category 3ii) the `/institutionSubmission/researchIncomeItems/industryAndOtherResearchIncome/internationalCompetitivePeerReview` element contains a list of `incomeYear` elements. And the `internationalCompetitivePeerReview/otherIncomeYears/otherIncomeYear/incomeByDiscipline/valueByDiscipline` elements contain the breakdown of income by FoR within a specific year.

For Other International Research Income (Category 3iii) the `/institutionSubmission/researchIncomeItems/industryAndOtherResearchIncome/internationalOtherIncome` element contains a list of `incomeYear` elements. And the `internationalOtherIncome/otherIncomeYears/otherIncomeYear/incomeByDiscipline/valueByDiscipline` elements contain the breakdown of income by FoR within a specific year.

For Cooperative Research Centre Research Income (Category 4) the `/institutionSubmission/researchIncomeItems/crcResearchIncome` element contains a list of `incomeYear` elements. And the `crcResearchIncome/otherIncomeYears/otherIncomeYear/incomeByDiscipline/valueByDiscipline` elements contain the breakdown of Cooperative Research Centre Research income by FoR within a specific year.

3.5.7 *Encoding Applied Measures*

The `/institutionSubmission/appliedMeasures` element consists of a list of applied measures. Research commercialisation income (`/institutionSubmission/appliedMeasures/researchCommercialisation`) is an applied measure that is defined in the same way as the Category 2 – Category 4 research income.

3.5.8 *Encoding Esteem Measures*

The `/institutionSubmission/esteemMeasures` element consists of a list of esteem measures.

3.6 Codes

Code (enumerated) values that are required for an institution's submission have been included in the *ERA-SEER Technology Pack*. For the purposes of the ERA Cluster One and Cluster Two round the categories of code used include FoR codes and institution codes.

3.6.1 *FoR Codes*

FoRs are identified in ERA via *Group*-level ANZSRC FoR⁴ classifications. An ERA submission may specify a research discipline via one or more four-digit FoR codes⁵. The lists of PCE and HCA FoR codes that are valid for ERA Cluster One and Cluster Two round submissions are included in the *ERA-SEER Technology Pack*.

3.6.2 *Institution Codes*

SEER uses two-digit numeric codes to identify institutions. The code table of valid institution codes is provided in the *ERA-SEER Technology Pack*.

3.6.3 *Ranked Journal Outlet Codes*

The code table of valid ranked journal codes for PCE and HCA clusters will be provided in the ERA website prior to the submission period.

3.6.4 *Language Codes*

SEER uses the Australian Bureau of Statistics (ABS) Australian Standard Classification of Languages (ASCL)⁶ to track languages. The code table of valid ASCL languages is provided in the *ERA-SEER Technology Pack*.

3.6.5 *Eligible Australian Competitive Grant Program Codes*

SEER uses an alpha numeric code to identify eligible Australian competitive grant programs within a particular reference year. The code table of valid codes for eligible Australian competitive grant programs is provided in the *ERA-SEER Technology Pack*.

3.6.6 *Research Theme Codes*

SEER uses an alpha numeric code to identify research themes. The code table of valid codes for research themes is provided in the *ERA-SEER Technology Pack*.

4

<http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1297.0Main%20Features52008?opendocument&tabname=Summary&prodno=1297.0&issue=2008&num=&view=>

⁵ ANZSRC Group-level codes are four digits in length. If a Field-level code is supplied (i.e. six digits) then the last two digits will be truncated by SEER.

⁶ <http://www.abs.gov.au/AUSSTATS/abs@.nsf/MF/1267.0>

3.6.7 Patent Issuing Country Codes

SEER uses a one-digit numeric code to identify countries where a patent is registered. The code table of valid codes is provided in the *ERA-SEER Technology Pack*. This list is also enforced by an enumeration of valid values in the XML schema.

3.7 Uniquely Identifying Journal Articles

SEER requires that submitted journal articles from indexed publications are uniquely identified by specifying a bibliometric provider key for each article. Bibliometric providers are able to provide unique identification numbers for articles included in indexed publications.

4 SYSTEM USER INTERFACE

4.1 Overview

The SEER web-based user interface is the only interface by which institutions may upload an ERA submission. This user interface supports institutions in the process of uploading ERA submissions and obtaining high level feedback (reports) on them. The scope of the user interface that is made available to institutions includes but is not limited to:

- user authentication (logon);
- submission upload;
- management of Background Statements as part of a submission;
- submission content verification;
- submission finalisation and certification; and
- reporting.

4.2 Authentication and Authorisation

The SEER web-based user interface accepts only SSL secured communications (i.e. HTTPS). Each system user will be provided with an individual user account with which they may undertake system activities. User accounts for SEER are issued by the SEER Helpdesk.

User password security remains the responsibility of individual users. SEER user accounts are ongoing, although ERA may disable or revoke individual user accounts if concerns over system security or misuse arise. Note that SEER Pilot user accounts are no longer operable.

4.3 Supported Browsers

The SEER web-based user interface is being developed according to contemporary W3C standards; as such it is designed to be accessible from standards-compliant web browsers. In particular, the web-based user interface is verified to operate for users of the following web browsers:

- Internet Explorer v6 and v7; and,
- Mozilla Firefox v2 and v3.

Web browsers must be configured to enable JavaScript, cookies and pop-up windows when accessing SEER.

5. INSTITUTIONALLY-SUPPORTED REPOSITORY REQUIREMENTS

ERA recognises that the Australian Government has provided support for the establishment of digital repositories through the Australian Scheme for Higher Education Repositories (ASHER) program. ERA HCA submissions will require access from the SEER platform to institutionally-supported repositories for the purposes of undertaking peer review. During this phase all repository access will occur in cooperation with institutional technical contacts.

5.1 Research Outputs

For the purposes of this document, research outputs are defined as in the *ERA Submission Guidelines*.

As with all research items submitted during an evaluation round, research outputs are to be specified via metadata records that describe the characteristics of each output for the purposes of SEER processing. A research output metadata record may make reference to an electronic file associated with the output which is termed a RODA. The expectation of ERA is that where a research output is required for the purposes of peer review, institutions will store and reference their RODA files within their institutionally-supported digital repositories.

Research outputs that are not physically stored within repositories include those items that cannot be readily digitised or otherwise electronically stored. Information covering the process by which a non-digitised research output may be made available outside SEER is outlined in the *ERA Submission Guidelines*.

RODA files that may be accessed via SEER should conform to the requirements described in this document, guided by certain specifications and constraints, such as supported formats, file sizes and non-security related access issues.

5.2 Specifications for Research Outputs

5.2.1 Research Output Digital Assets

Research output submissions to SEER must include references to RODA files where a research output is being made available to ERA in digital form. Any nominated RODA files may be formally viewed by reviewers and other users authorised by the ARC.

In cases where this is not possible, other means may be used to ensure that reviewers, and other users authorised by the ARC have access to the additional information associated with research outputs.

URL-based automatic resolution of persistent identifiers or handles such as Digital Object Identifiers (DOIs) via sites such as dx.doi.org are suitable mechanisms for providing links. Multiple electronic locations can be associated with a research output; each of these locations must provide information indicating if the location identifies a metadata page or a RODA.

5.2.2 Research Output Metadata

The research output metadata required for an ERA HCA submission is dictated (in terms of data content/structure) by the ERA submission XML schema. This information supplements the contextual information provided in the *ERA Submission Guidelines*.

5.2.3 Multiple Institutionally-Supported Repositories

An institution may have more than one institutionally-supported repository for the purposes of an ERA submission round. If an institution submits research output metadata records that refer to RODA files across multiple repositories then the institution must be able to arrange for ERA access to each repository involved.

5.2.4 Supported Formats

SEER supports all research output file types that can be rendered across the internet provided that these files conform to specified size constraints and standard computer desktops may be configured with the required software to properly view the RODA file. As a minimum, SEER will assume that standard desktops may be configured to properly display the following file types:

- Adobe PDF (i.e, .pdf);

- GIF image (i.e. .gif);
- JPEG image (i.e. .jpg, .jpeg);
- Microsoft Word (i.e. .doc);
- PNG image (i.e. .png);
- Text (i.e. .txt); and,
- ZIP archive (i.e. .zip).

Research outputs with file types not identified in the list of supported formats may still be included as part of an ERA submission upload. For these other formats the Helpdesk contact (see section 5.2.7 below) may be contacted to discuss the nature of the application required to access such files.

5.2.5 *Size Limits*

ARC staff and ERA reviewers may access RODA files stored in an institution's digital repositories. These outputs will be relayed to them via SEER.

Research outputs stored in repositories for the purposes of ERA should be no greater than 15 MB in total per output. It is anticipated that the vast majority of research outputs will be under this limit and that many would be less than 2 MB in size. These file sizes and limits may be revised following performance and capacity testing.

It is expected that research outputs that exceed the maximum size will not be handled in an ERA cluster submission.

5.2.6 *HTTP Ports*

The ARC security rules only allow a small number of permissible TCP/IP ports as per the *Protective Security Manual (PSM 2005)*. Institutions must configure their ERA repositories to use only standard HTTP and HTTPS ports (port 80 and 443). These ports are usually permitted through firewalls.

5.2.7 *Repository Service Availability and Helpdesk Arrangements*

Each institution is required to set up a Helpdesk function to assist with repository access issues during the ERA process. The ARC will liaise with institutions to ensure unscheduled down-times are reduced to an absolute minimum at all times.

5.2.8 *Authentication*

Institutions are required to set up secure digital repositories to store research outputs that are not publicly accessible, whether because of copyright restrictions or because of commercial or cultural sensitivity. Where the institution or the researcher is the copyright owner, or where the copyright owner has given express permission for the research output to be stored in an 'open access' repository, the research output should be stored in an 'open access' repository or in an 'open access' part of an institution's ERA repository.

The SEER system requires read access to all objects within these repositories to retrieve RODA metadata pages and RODA files to relay back to ERA staff, REC members, expert and peer reviewers at the ARC. When accessing a 'closed' or otherwise 'non-public' institutionally-supported repository, SEER's preferred method of authentication is Basic Authentication (RFC 22617)⁷. Institutionally supported repositories that are not able to support the Basic Authentication (RFC 22617) method are permitted to utilise IP-based access restrictions. Based on feed back from institutions other authentication mechanisms will be considered where appropriate.

The ARC will require institutions to follow the following security rules when assigning SEER user account:

- Establish a special user account for SEER, granting read access to all objects within the institutionally supported repository/repositories;

⁷ RFC 2617 - HTTP Authentication: Basic and Digest Access Authentication (<http://www.faqs.org/rfcs/rfc2617.html>)

- User accounts employ the same username and password across all institutionally supported repositories that require SEER access;
- Strong passwords will be used – these must contain a combination of upper and lowercase characters, numbers and special characters (e.g. #, @, \$);
- Passwords must be between 7 and 15 characters long; and
- Passwords never expire.

All communications between SEER and institutionally supported repositories are required to employ SSL certificates to secure the communications.⁸

5.2.9 Stability of Links

Institutions should endeavour to ensure that submitted links within research output metadata records remain stable for the duration of an ERA cluster round. As the ERA evaluation process involves remotely located reviewers, it is a requirement that nominated links to institutionally-supported repositories be made available 24 hours per day, 7 days per week.

⁸ The ARC will manage and control access to research outputs and will only allow ARC staff, REC members and peer reviewers and other ARC authorised personnel access to research outputs that are stored in an institution's repository for the purposes of the ERA initiative.

6. SYSTEM ENVIRONMENT

There is only one environment being made available to support the ERA PCE and HCA submissions. This is termed the Live SEER Platform and may be accessed via web browser at the following URL:

- <https://seer.arc.gov.au>

Appendix A: ERA-SEER Technology Pack Inventory

ERA Submission XML Schemas

The ERA submission XML schemas uniquely describe the data format that is allowed for submissions in the ERA Cluster One and Cluster Two round. It is made available electronically in the *ERA-SEER Technology Pack* via the following files:

- seer.xsd;
- institution.xsd;
- researcher.xsd;
- researchItem.xsd;
- researchOutput.xsd;
- researchOutlet.xsd;
- researchIncome.xsd;
- otherMeasure.xsd;
- esteemMeasure.xsd;
- appliedMeasure.xsd;
- artefactTitle.xsd;
- types.xsd; and
- discipline.xsd
- pceInstitutionSubmission.xsd
- hcaInstitutionSubmission.xsd

Code Tables

Code tables are used to define enumerated fields (i.e. reference data) for use in submission XML documents. For the ERA PCE and HCA submissions these include:

- eligible institutions; eligible Australian Competitive Grants programs (2005 – 2007);
- research themes;
- PCE and HCA FoR discipline codes (ANZSRC Group-level);
- Australian Standard Classification of Languages(ASCL); and
- eligible patents issuing country list.

Documentation

This technical specifications document is included in the *ERA-SEER Technology Pack* in addition to schema help HTML files describing the technical elements of the ERA submission XML schema.

Submission Examples

Several sample submission files are included in the *ERA-SEER Technology Pack*. These are intended to demonstrate how the ERA submission XML schemas may be applied to create actual XML submission documents. Archetype examples included are:

- pce-basic-submission.xml;
- pce-outputs-only-submission.xml;
- pce-outputs-and-income-submission.xml;

- pce-full-submission.xml;
- hca-basic-submission.xml;
- hca-outputs-only-submission.xml;
- hca-outputs-and-income-submission.xml; and,
- hca-full-submission.xml.

Appendix B: Glossary of Submission Data Items

Common Data Items

Institution

Data Item	Description
code	The Institution Code for the Institution submitting the ERA Submission (provided in code tables in the ERA-SEER Technology Pack).
name	The official name of the institution submitting the ERA submission.

Affiliation

Data Item	Description
forGroupCode	The ANZSRC Field of Research group code
apportionment	The percentage of the research affiliation against each discipline that the researcher nominates for a Field of Research.

Researcher

Data Item	Description
staffReference	A reference given by an Institution to allow a Researcher to be nominated as a Creator of a Research Output. At least one Researcher must be nominated per Research Output submitted.
firstName	The first or given name of a researcher.
honorific	The preferred title of a person (e.g. “Dr”, “Professor”, etc.).
lastName	The last name, family name or surname of a person.
middleName	Any other given name(s) of a person that are not the person’s first name.
otherName	Any other names of the researcher under which the researcher may have published research outputs.
employmentLevel	Employment level of the Researcher. Based on Higher Education Staff Data Collection Element 408.
totalFTE	The total FTE of the researcher affiliated with the institution.
affiliationByDiscipline	A collection of researcher FoR and the percentage affiliation assigned to each FoR. Refer to ‘Affiliation’ table for details.
employmentStatus	The employment status of the researcher, based on Higher Education Staff Data Collection Element 412.
researcherFunction	The basis on which the researcher is affiliated to the institution. Either salaried by the submitting institution or non-salaried.

Research Outputs

Creator

Data Item	Description
orderInOutput	The order in which the Creator is listed as an author on a Research Output. (DC: Creator)
publishedName	The Name (as published) of the Creator listed on a Research Output (e.g. J. Block). (DC: Creator). Where the Name (as published) is not available, this should be left blank. The Researcher Names will be used instead for the research output

Institutional Unit

Data Item	Description
name	The full name of the Institutional Unit within the institution that is associated to the research.
code	A unique Code identifying the Institutional Unit within the institution for the purposes of allowing aggregation of data.

Research Theme

Data Item	Description
code	A research theme related to the research output.

Staff Creator

Data Item	Description
staffReference	A reference given by an Institution for use in nominating a Researcher as a Creator of a Research Output.

Discipline

Data Item	Description
forGroupCode	An ANZSRC Field of Research Group Code.

Artefact Title

Data Item	Description
nativeScript	The title of the Research Output.
language	The code for languages using ABS 1267.0 to identify the native Language of a Research Output Title (if other than English).
romanScript	Phonetic sounding of the full title in roman characters where the title of the output is in a foreign language, particularly if the title is in non-roman characters.
translated	The English translation of a Research Output Title, where the titleInNativeScript is in a language other than English.

Book Outlet

Data Item	Description
title	The title of book outlet related to the book chapter submitted.
edition	The edition of the book.
isbn	The International Standard Book Number of a Book in which a Book Chapter submitted as a Research Output appears.

publisher	The entity responsible for making the Research Output available. (DC: publisher)
-----------	---

Conference Publication Outlet

Data Item	Description
title	The formal name of a Conference Publication Outlet.
issue	The issue identifier of the Conference where the conference is submitted as a conference publication outlet.
volume	The volume identifier of the Conference where the conference is submitted as a conference publication outlet.
venue	The venue where the conference was held. (DC: Location)
disciplineOutput	The collection of ANZSRC Field of Research Group Codes associated with a Conference Publication.

Ranked Journal Outlet

Data Item	Description
eraJournalID	A unique identifier assigned by the ARC to Ranked Journal Outlets.

Unranked Journal Outlet

Data Item	Description
title	Titles of the journal outlet that is not already identified by ARC in the journal ranking list. See: Artefact Title table
disciplineOutput	The collection of ANZSRC Field of Research Group Codes associated with an Unranked Journal (Outlet). See: Discipline Output table
issn	The International Standard Serial Number of an Unranked Journal (Outlet).

Research Output

Data Item	Description
title	The title of the research output using complex format as specified in the Artefact Title.
electronicLocations	A collection of electronic locations related to each research output.
physicalLocation	The physical location of the Research Output if not in a repository. (DC: Location)
placeOfPublication	The location in which the Research Output was published. (DC: Coverage)
extent	The size or duration of the resource. (DC: Extent)
yearAvailable	The full (four-digit) calendar year in which the Research Output was first available if different to the published year.
yearPresented	The full (four-digit) calendar year in which the Research Output was presented (if different to the published year). (DC: Date).

yearPublished	The full (four-digit) calendar year in which the Research Output is published. (DC: Date).
isTranslated	An indicator that the Research Output submitted is a translation of an otherwise eligible Research Output.
referenceToOriginal	Bibliographic detail of the original research where the Research Output is the translation of the original output
isSensitive	An indicator that the Research Output requires special handling due to the sensitivity of its content.
typeOfSensitivity	The type of sensitivity that requires ARC consideration before being provided for peer review.
sensitiveNote	A description to address how the research output should be handled.
creators	A collection of one or more authors for the output. The information details required for each creator are specified in the Creator table. (DC: creator)
staffCreators	A collection of staff who contribute to the research output or other research items.
institutionalUnits	A collection of Institution Units associated to the research output.
researchThemes	A theme associated to the research
availableForPeerReview	An indicator whether the research output is identified as available for peer review.

Electronic Location

Data Item	Description
repositoryLink	The Uniform Resource Identifier (e.g. URL, Handle, DOI, etc.) that links to the Research Output.
linkType	The option to identify whether the link provided leads to a metadata page or directly to the content of the research output.

Journal Article

Data Item	Description
citationProvider	Name of the external citation provider.
uniqueIdentifier	The value of the metrics reference key specifically for the Provider identified in the 'citationProvider' field.
journalIssue	The issue number of the journal
journalVolume	The volume number of the journal
journalOutlet	Refer to Unranked Journal Outlet and Ranked Journal Outlet below.

Book Chapter

Data Item	Description
publisherType	The option for the type of publisher who publishes the book outlet associated with the book chapter.
bookOutlet	A link to the book outlet of the book in which the book chapter is published.
disciplineOutput	The collection of ANZSRC Field of Research Group Codes associated with a Book Chapter submitted as a Research Output.

Book

Data Item	Description
version	Distinguishes between the original and subsequent versions of a Book submitted as a Research Output.
edition	The edition of the book.
disciplineOutput	The collection of ANZSRC Field of Research Group Codes associated with a Book submitted as a Research Output.
isbn	The standard number for the Research Output.
publisher	(DC: Publisher)

Conference Publication

Data Item	Description
version	Distinguishes between the original and subsequent versions of a Conference Publication submitted as a Research Output.
coverage	Distinguishes between the complete Conference Proceedings and Extracts of a Conference Publication submitted as a Research Output.
conferencePublicationOutlet	A reference to the details submitted for the Conference Publication (Outlet).

Research Income

Australian Competitive Grant Income Year Type

Data Item	Description
referenceYear	The full (four-digit) calendar year.
acgrCode	The program code assigned by the ARC to an Australian Competitive Grant program awarded within each of the income submission reference year.
totalAmountReceived	The total amount received for each grant.
grantIncomeByDiscipline	A collection of ANZSRC Field of Research Group Codes together with the aggregated Research Income for the specified Disciplines.

Other Income Year Type

Data Item	Description
incomeByDiscipline	A collection of ANZSRC Field of Research Group Codes together with the aggregated Research Income for the specified Disciplines.
referenceYear	The full (four-digit) calendar year.
totalAmountReceived	The total amount the institution has received within a reference year.

Australian Competitive Grant

Data Item	Description
incomeYear	A collection of income for an income year as identified by each reference year.

CRC Research Income

Data Item	Description
otherIncomeYear	A collection of income per discipline for an income year as identified by each reference year.

Industry and other research Income

International Competitive Peer review

Data Item	Description
otherIncomeYear	A collection of income per discipline for an income year as identified by each reference year.

International Other Income

Data Item	Description
otherIncomeYear	A collection of income per discipline for an income year as identified by each reference year.

Australian Industry and Other Research Income

Data Item	Description
OtherIncomeYear	A collection of income per discipline for an income year as identified by each reference year.

Other Public Sector Income

Data Item	Description
otherIncomeYear	A collection of income per discipline for an income year as identified by each reference year.

Income Value Type

Data Item	Description
forGroupCode	A group code from FoR table.
apportionmentPerGroup	The apportionment by percentage of the total grant received by FoR.

Esteem Measures

Editorial Role

Data Item	Description
eraJournalID	A unique identifier assigned by the ARC to Ranked Journal Outlets and published as the ERA Ranked Journal List.
staffReference	An identifier assigned by the Institution to an eligible staff member for whom an Editorial Role is submitted.
referenceYear	The four-digit year relating submission data to the reference period for esteem measures.

Fellowship of Learned Academies

Data Item	Description
Name	The identifier of the Learned academy to be considered for esteem measure.
staffReference	A reference given by an institution identified the eligible staff member whose Fellowship is submitted as an esteem measure.
referenceYear	The four-digit year relating submission data to the reference period for esteem measures.

Nationally Competitive (Category 1) Research Fellowships

Data Item	Description
acgrCode	The identifier of the ACGR listed for the esteem measure. The code use for this esteem measure is the same as the organisation code used in the Australian Competitive Grant program (the esteem measure list will be provided)
staffReference	An identifier assigned by the Institution to an eligible staff member for whom an Australian competitive Fellowship is submitted.
referenceYear	The four-digit year relating submission data to the reference period for esteem measures.

Applied Measures

Patents

Data Item	Description
number	The registered number of the patent being submitted as part of a Cluster Submission.
name	The name of the applied measure item.
registeredCountry	The country code of the country in which the patent is recognised.
disciplineOutput	A collection of up to three FoR group Code
referenceYear	The four-digit year relating submission data to the reference period for applied measures.

Research Commercialisation Income

Data Item	Description
otherIncomeYear	A collection of income per income for an income year as identified by each reference year.

PCE Cluster Data Items

There are no additional data items specific to PCE submissions.

HCA Cluster Data Items

Research Outputs

Original (creative) Work in the Public Domain

Data Item	Description
outputType	The options to categorise the nature of the Original (creative) work. The option can be used for a range of original outputs that are not already included in the other research output type.
disciplineOutput	The FoR which describes the focus of the Research Output. (ANZSRC: Group).

Live Performance

Data Item	Description
discipline	The collection of ANZSRC Field of Research Group Codes associated with a Live Performance submitted as a Research Output.

Recorded Work

Data Item	Description
workType	The option to categorise the type of recorded work.
recordedFormat	The format that the output is recorded in e.g. VHS, NTSC, MP4, etc.
disciplineOutput	The collection of ANZSRC Field of Research Group Codes associated with a Recorded Work submitted as a Research Output.

Curated or Produced Substantial Exhibition, Event or Rendering

Data Item	Description
workType	Distinguishes between curation of physical exhibitions, web-based exhibitions, and festivals submitted as a Research Output.
disciplineOutput	The collection of ANZSRC Field of Research Group Codes associated with an Exhibition submitted as a Research Output.

Esteem Measures

Curatorial Role

Data Item	Description
eventName	The name of the curated event for which a Curatorial Role is claimed.
location	The geographic location of the event for which a Curatorial Role is claimed.
staffReference	An identifier assigned by the Institution to an eligible staff member for whom a Curatorial Role is submitted.
referenceYear	The four-digit year relating submission data to the reference period for esteem measures.
esteemJustification	A statement addressing the published criteria for inclusion provided in support of a claimed Curatorial Role.

Work of Reference

Data Item	Description
nameOfReferenceWork	The name of the prestigious work of reference for which a contribution is being claimed as a measure of esteem.
nameOfReferenceOutlet	The name of the outlet in which the output is included.
staffReference	An identifier assigned by the Institution to an eligible staff member for whom a contribution to a prestigious work of reference is submitted.
referenceYear	The full (four-digit) calendar year for which a contribution to a prestigious work of reference has been submitted.
esteemJustification	A statement addressing the published criteria for inclusion provided in support of a claimed contribution to a prestigious work of reference.

Recipient

Data Item	Description
staffReference	A reference given by an institution which identifies the eligible staff member who is the recipient.

Prestigious Awards and Prizes

Data Item	Description
awardName	The name of the award or prize being given.
organisationName	The name of the organisation giving the award.
isInternational	A flag to identify whether the Award or Prize is internationally recognised.
recipient	A link to the staff reference where the prize or award is applied to an individual rather than the institution.
referenceYear	The full (four-digit) calendar year for which a prize or award has been submitted.
esteemJustification	A statement addressing the published criteria for inclusion provided in support of a submitted award or prize.
disciplineOutput	The collection of up to three FoR group code if the prize or award is applied to the institution rather than an individual staff member.

Applied Measures

Registered Design

Data Item	Description
registeredId	The registered identifier of the registered design.
name	The name of the applied measure item.
disciplineOutput	A collection of up to three FoR group Codes
registryOrganisation	The organisation with which the design is registered.
referenceYear	The full (four-digit) calendar year for which a grant of a design has been registered.
justification	A statement addressing the criteria for the relevancy of the registered design as an applied measure.

Appendix C: Glossary of Terms and Acronyms

Technical Terms

Term	Description
Applied Measures Metadata	Data record describing a research item that embodies research-related patents, registered designs and research commercialisation income.
Authorisation	Determining a user's access to a resource. Authorisation almost always relies on the user having been authenticated.
Basic Authentication	A standard HTTP authentication protocol supported by most browsers where username and password is transmitted as Base-64 encoded text.
DOI	Digital Object Identifier: DOI is a standard for persistently identifying a piece of intellectual property on a digital network and associating it with related data, the metadata, in a structured extensible way. DOIs can be resolved through the DOI resolver at http://dx.doi.org .
Esteem Measures Metadata	Data record describing a research item that embodies indications that a researcher is held in particularly high regard by peers for their research work.
Handle System	Handle System enables a distributed computer system to store names, or handles, of digital resources and resolve those handles into the information necessary to locate, access, and otherwise make use of the resources. These associated values can be changed as needed to reflect the current state of the identified resource without changing the handle. This allows the name of the item to persist over changes of location and other current state information. An overview of Handle System may be found at http://www.ietf.org/rfc/rfc3650.txt
HTTP	HyperText Transfer Protocol: An application layer protocol that provides a standard for web browsers and web servers to communicate.
HTTPS	HyperText Transfer Protocol (Secure): HTTP exchanged over an SSL encrypted session.
IANA	The Internet Assigned Numbers Authority (IANA). Of relevance to the ERA SEER is the registry of MIME types that may be found at: http://www.iana.org/assignments/media-types/ This registry includes types from: RFC-2046 (http://www.ietf.org/rfc/rfc2046.txt), RFC-1521 (http://www.ietf.org/rfc/rfc1521.txt), and others.
MegaByte	1,048,576 Bytes

Term	Description
Metadata	A record that contains data known about an information resource. Thus any data that refers to a discrete informational asset may be referred to as the metadata of that asset, for example, a bibliographic record of a journal article is a form of <i>metadata</i> describing that article.
MIME	<p>Multipurpose Internet Mail Extensions:</p> <p>Originally designed as a specification for formatting non-ASCII messages so that they can be sent via e-mail. In addition, web browsers use MIME types to determine ‘helper’ applications or components that can be used to view the content.</p> <p>RFC-1341 (http://www.ietf.org/rfc/rfc1341.txt)</p>
REC	Research Evaluation Committee
Research Income Metadata	Data record describing a research item that embodies a funding event for research activity (e.g. metadata for an Australian Competitive Grant).
Research Item Metadata	Data record describing an entity that is eligible to be considered as evidence of research activity. Valid types for the ERA Cluster One and Cluster Two round include research outputs, research income, esteem measures and applied measures.
Research Output Metadata	Data record describing a research item that embodies a measurable/tangible output of research activity (e.g. a bibliographic record for journal article). The recognised types of research output are listed in the <i>ERA Submission Guidelines</i> .
Research Output Digital Asset (RODA)	A binary (electronic) file that represents an output of research activity such as a document, audio clip, video clip, data sample etc. Used here to refer to institutional research outputs that are available in digital form.
SSL	<p>Secure Sockets Layer:</p> <p>A protocol developed by Netscape for transmitting private documents via the internet. SSL works by using a private key to encrypt data that is transferred over the SSL connection. Both Netscape Navigator and Internet Explorer support SSL and many web sites use the protocol to obtain confidential user information, such as credit card numbers. By convention, URLs that require an SSL connection start with https: instead of http:.</p>
TCP/IP	<p>Transmission Control Protocol/Internet Protocol:</p> <p>A suite of protocols that computers use to exchange information over the internet.</p>

Term	Description
URI	<p>Uniform Resource Identifier (URI): A standard syntax for the textual naming (identification) of resources that exist on the internet. Two concrete implementations of the URI syntax include the Uniform Resource Locator (URL) and the Uniform Resource Name (URN). A URL is a URI that defines how a given resource may be retrieved over the internet (i.e. via a particular protocol such as HTTP).</p>
URL	<p>Uniform Resource Locator A global identifier for a network-retrievable document.</p>
UTF-8	<p>8-bit UCS/Unicode Transformation Format A variable-length character encoding for Unicode. It is able to represent any character in the Unicode standard. UTF-8 is the standard encoding method for email, web pages and XML documents.</p>
W3C	<p>World Wide Web Consortium: An international consortium of companies involved with developing standards for the internet and the web.</p>
XML	<p>eXtensible Markup Language: A text based specification used to promote interoperable exchange of data through standardised validation mechanisms and expression of data in a human-readable, self-describing manner.</p>
XML Namespace	<p>XML namespaces are used to ensure elements are named uniquely within an XML instance document. Namespaces remove any ambiguity when identically named elements from separate XML vocabularies are used within a single XML document. First published as a W3C Recommendation in 1999 and updated in August 2006 (http://www.w3.org/TR/REC-xml-names/).</p>
XSD	<p>XML Schema Definition XML schema, published as a W3C Recommendation in May 2001, can be used to express a scheme: a set of rules to which an XML document must conform in order to be considered 'valid' according to that schema. However, unlike most other schema languages, XML schema was also designed with the intent of validation resulting in a collection of information adhering to specific data-types, which can be useful in the development of XML document processing software.</p>