



# Australian Research Data Commons (ARDC) Party Infrastructure Description

Final Draft

27 August 2010

Please send any feedback on this consultation draft to the Project Team:

**Basil Dewhurst**

Project Manager  
ARDC Party Infrastructure Project  
National Library of Australia  
[bdewhurst@nla.gov.au](mailto:bdewhurst@nla.gov.au)

**Natasha Simons**

Business Analyst  
ARDC Party Infrastructure Project  
National Library of Australia  
[nsimons@nla.gov.au](mailto:nsimons@nla.gov.au)

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

## Table of Contents

1	Purpose .....	1
2	Project Background .....	1
3	ARDC Party Infrastructure Project.....	2
3.1	Project objective .....	2
3.2	Project stages.....	2
4	ARDC Party Infrastructure.....	3
4.1	Persistent Identifiers/Party Registry Keys .....	3
4.2	Party Infrastructure Overview .....	4
4.3	Party Data Contribution .....	5
4.3.1	OAI-PMH Repository .....	6
4.3.2	Static OAI-PMH Repository.....	6
4.4	Party Data Ingest.....	7
4.4.1	Static OAI-PMH Repository Gateway (Ceres) .....	7
4.4.2	NLA Harvester .....	7
4.4.3	Identity Matching Service.....	8
4.4.4	Party Administration Tool .....	8
4.4.5	People Australia Database.....	9
4.5	Party Data Dissemination .....	9
4.5.1	OAI-PMH Repository .....	9
4.5.2	Search/Retrieve via URL (SRU) Interface.....	9
4.5.3	Trove People and Organisations view.....	10
4.6	Party Data Consumption .....	10
4.6.1	OAI Harvester.....	10
4.6.2	SRU .....	11
4.7	Party Infrastructure Usage .....	11
4.7.1	Interoperability .....	11
4.7.2	Record Format .....	12
4.8	Roles and Responsibilities .....	14
4.8.1	Scenario 1 .....	15
4.8.2	Scenario 2 .....	16
4.8.3	Scenario 3 .....	17
4.8.4	Scenario 4 .....	19
4.8.5	Summary of record sources and identifier/key usage .....	20

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

## 1 Purpose

The purpose of this document is to describe the Australian Research Data Commons (ARDC) Party Infrastructure being developed by the National Library of Australia (NLA) with funding from the Australian National Data Service (ANDS).

The ARDC Party Infrastructure Project will adapt the NLA's Trove People and Organisations infrastructure, hereafter described as the Party Infrastructure, to enable authority control<sup>1</sup> for research data. The infrastructure will encourage better management and curation of data generated by Australian research and supports the long-term access, use and re-use of this data.

This document provides an overview of the project and describes the Party Infrastructure and its various components. It details contribution requirements and the means by which party identifiers will be assigned, managed and disseminated in the Australian Research Data Commons.

The National Library is using this document to consult with the research sector and seeks feedback on it and the infrastructure described.

## 2 Project Background

The Australian Research Data Commons (ARDC) Party Infrastructure Project is a National Library of Australia (NLA) project funded by the Australian National Data Service (ANDS). The project aims to enable improved discovery of research outputs and data through persistently identifying Australian researchers and research organisations. This will enable a more coherent view of research data collections in the Australian Research Data Commons (ARDC) and their relationships with people and organisations in the innovation system.

Work on this project will focus on adapting the Party Infrastructure<sup>2</sup> to enable authority control for research outputs and for data. As such it will encourage better management and curation of data generated by Australian research and supports the long-term access, use and re-use of this data.

---

<sup>1</sup> Name authority control in libraries is a process that standardises name headings and groups different forms of name into a single record to improve discovery of resources.

<sup>2</sup> National Library of Australia 2010, *People Australia overview*, <http://wiki.nla.gov.au/x/DwH8>.

## 3 ARDC Party Infrastructure Project

### 3.1 Project objective

The objective of the Party Infrastructure project is to develop a sustainable infrastructure based on People Australia for publicly identifying parties in the research sector that:

- Enables the management of party records by information officers in research organisations by extending the Party Administration Tool;
- Improves the automatic matching algorithms currently in use;
- Provides standards based machine interfaces to data about researchers and research organisations;
- Extends the data standards used in the service to cater for the record standards used in the research sector; and,
- Provides access to public, persistent identifiers for researchers and research organisations to support the Australian Research Data Commons.

### 3.2 Project stages

The Party Infrastructure Project is planned to run for approximately 18 months beginning January 2010 and is divided into two stages described below. A Project Advisory Group formally involves key stakeholders from the research sector who provide input and an Implementers Group will also be established.

#### *Stage 1: Consultation/specification (January 2010 - June 2010)*

The goals of Stage 1 are:

- Conduct community consultation on research sector requirements for the identification and description of researchers and research groups;
- Establish Project Advisory group for consultation purposes;
- Consult the Australian Access Federation (AAF), Australian Research Council (ARC), and the National Health and Medical Research Council (NHMRC) on identifier equivalence services;
- Complete extensions to the People Australia machine-to-machine interfaces to enable provision of party data in RIF-CS format to the ANDS collection registry;
- Enable the preliminary bidirectional exchange of collection metadata between NLA services and the ANDS collections registry;
- Develop specifications for Stage 2 development and create initial documentation for the Party Administration Tool; and
- Create end of stage report and options for delivering Stage 2 which include recommendations from the Advisory group.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

*Stage 2: Development and Implementation (July 2010 - June 2011)*

The goals of Stage 2 are:

- Build Party Infrastructure (as defined in this document and agreed at the end of Stage 1);
- Deliver Party Administration Tool and identity matching service that enables repository and data managers to manage the party data they contribute and to obtain assigned identifiers;
- Create technical documentation and administrator's guide;
- Bring a small number of initial implementers on board;
- Create mid-stage 2 report; and
- Create end project report.

## 4 ARDC Party Infrastructure

ANDS has identified four kinds of related objects and entities that need to be described in the ARDC, namely: collections, parties, activities and services.<sup>3</sup>

To participate in the ARDC, metadata about these objects and entities, known as registry objects, needs to be provided by research organisations to the ANDS Collections Registry and Research Data Australia.

### 4.1 *Persistent Identifiers/Party Registry Keys*

Each registry object requires a unique key prior to submission to the ANDS services.<sup>4</sup> Keys are used to uniquely identify a registry object in the ARDC and in the expression relationships between registry objects. ANDS has a strong preference for the use of persistent identifiers for keys.<sup>5</sup>

In order to publicly and persistently identify parties in the ARDC and provide keys for party registry objects, the ARDC Party Infrastructure project will deliver a sustainable infrastructure that enables the identification of parties involved in the research lifecycle.

The National Library is committed to the use of persistent identifiers for the permanent identification of internet resources and has a history of managing and supporting persistent identification.<sup>6</sup> The People Australia service assigns each party

---

<sup>3</sup> Australian National Data Service (ANDS) 2009, *Australian National Data Service Content Providers Guide*, Australian National Data Service (ANDS), Melbourne, Vic, <http://ands.org.au/guides/content-providers-guide-20may2010.pdf>, page 9.

<sup>4</sup> Ibid., page 56.

<sup>5</sup> Ibid., page 13.

<sup>6</sup> National Library of Australia 2001, *Persistent Identifier scheme for digital collections at the National Library of Australia*, National Library of Australia, Canberra, <http://www.nla.gov.au/initiatives/nlapi.html>

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

with a public, persistent and citable identifier. Identifiers for parties take the form <collectionId>-<field1> where <collectionId> is the string nla.party and <field1> is an integer. These identifiers are resolvable at <http://nla.gov.au> making them globally unique. *Table 1* below provides two People Australia program examples.

Party	Persistent Identifier	Resolvable URL
CSIRO	nla.party-458251	<a href="http://nla.gov.au/nla.party-458251">http://nla.gov.au/nla.party-458251</a>
Huxley, Thomas Henry (1825-1895)	nla.party-869246	<a href="http://nla.gov.au/nla.party-869246">http://nla.gov.au/nla.party-869246</a>

**Table 1: Example Persistent Identifiers**

People Australia persistent identifiers for parties are made freely available via Trove and via machine-to-machine interfaces including the People Australia Search and Retrieval via URL (SRU) and Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) services.

## 4.2 Party Infrastructure Overview

The ARDC Party Infrastructure will provide of a number of services that enable the ingest, manipulation, persistent identification and dissemination of party records and persistent identifiers/party registry keys.

Four workflow scenarios for using the infrastructure are detailed below in Section 4.7.2 however the preferred workflow for obtaining party identifiers and providing party records to ANDS via the ARDC Party Infrastructure is as follows:

- 1) The ARDC Contributor provides party records to the ARDC Party Infrastructure.
- 2) The ARDC Party Infrastructure assigns persistent identifiers to these parties and provides these identifiers and associated records to contributors.
- 3) The ARDC Party Infrastructure provides party records to ANDS.
- 4) The ARDC Contributor implements the persistent identifiers as party keys in their collections/services/activities metadata feed to ANDS.

To enable the workflow above the ARDC Party Infrastructure is comprised of a number of machine-to-machine interfaces for ingesting and disseminating party records and their associated identifiers as well as human interfaces for managing party records. An overall diagram representing the party infrastructure is provided below in *Figure 1*. Descriptions of each component of the infrastructure follow.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

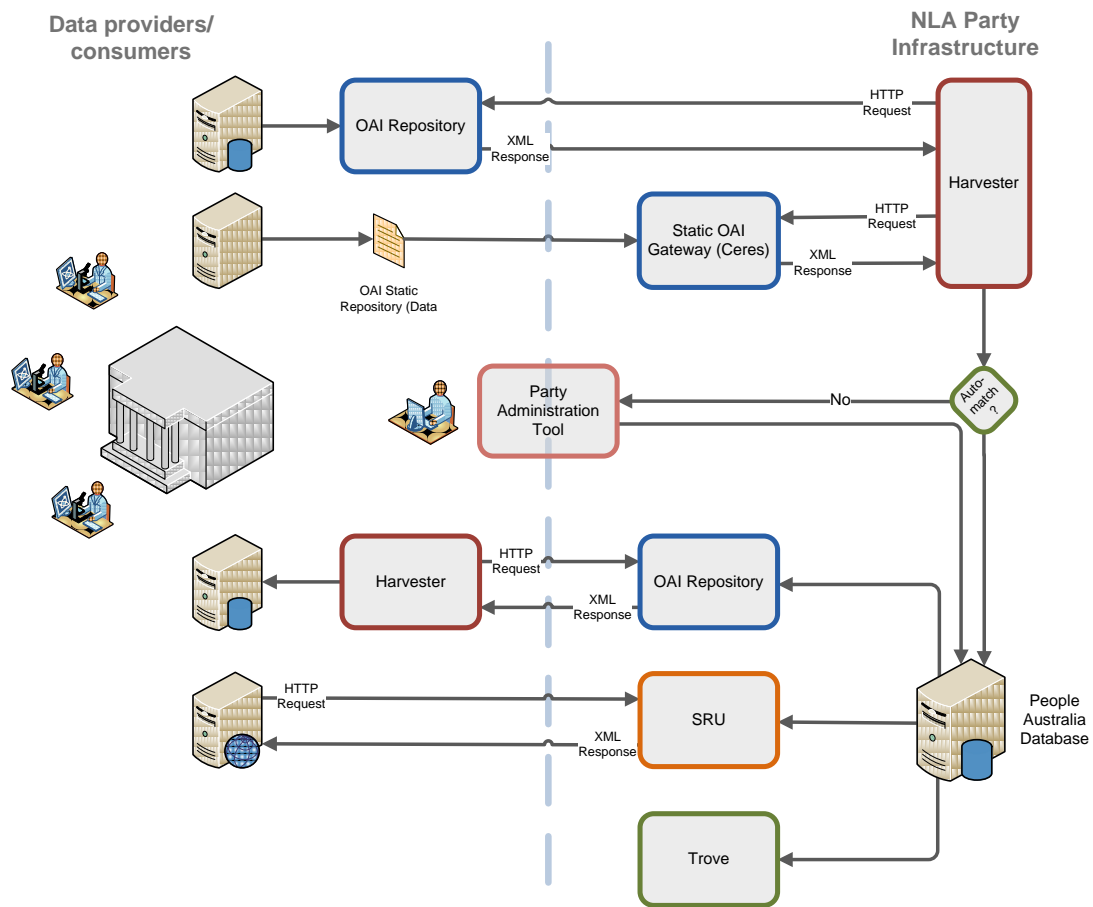


Figure 1: NLA Party Infrastructure Detail

## 4.3 Party Data Contribution

Data providers will contribute records using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)<sup>7</sup>. There are two types of OAI-PMH repositories that the Party Infrastructure supports, these are described below.

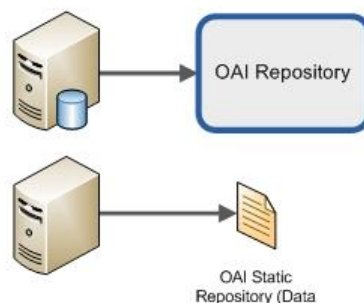


Figure 2: Contribution components

<sup>7</sup> Open Archives Initiative 2010, *Open Archives Initiative Homepage*, <http://www.openarchives.org>

## 4.3.1 OAI-PMH Repository

An OAI-PMH Repository is a network accessible server that can process the six OAI-PMH requests and provide metadata records for harvesting by OAI-PMH compliant harvesters as defined in the OAI-PMH version 2.0 specification<sup>8</sup>.

OAI-PMH Repositories provide metadata records in XML format and enable incremental harvesting of these. OAI-PMH Repositories can be configured to support a range of metadata formats. In order to be compliant with the OAI-PMH repositories must support unqualified Dublin Core<sup>9</sup>. Unqualified Dublin Core is however not suitable for describing parties as required by the party infrastructure and so, for repositories contributing party records, support for richer metadata formats such as EAC-CPF or RIF-CS would be required. See section 4.7.2.

OAI-PMH Repositories can optionally group records into sets for their provision to the party infrastructure and the ANDS Collections Registry. It is recommended that contributors to the party infrastructure provide a separate set of party records.

OAI-PMH repositories are harvested using the NLA Harvester described below in Section 4.4.2.

## 4.3.2 Static OAI-PMH Repository

A Static OAI-PMH Repository is essentially an XML file, conforming to the 'Specification for an OAI Static Repository'.<sup>10</sup> Static Repositories contain metadata records and repository information and are made available at a URL. Static Repositories provide a simple, low-barrier approach for exposing metadata records through the OAI-PMH. They are useful when collections are small and comprised of relatively static content; data can be made available through a network-accessible Web server; and, a technically simpler implementation strategy (when compared to implementing a full OAI-PMH Repository) is required.

Static Repositories do not support sets but do support multiple metadata formats and, at base, are required to implement unqualified Dublin Core. As described in Section 4.3.1 support for either EAC-CPF or RIF-CS would be required if using a Static Repository to provide data to the party infrastructure.

---

<sup>8</sup> Open Archives Initiative 2002, *The Open Archives Initiative Protocol for Metadata Harvesting*, <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>

<sup>9</sup> Dublin Core Metadata Initiative 2008, *Dublin Core Metadata Element Set, Version 1.1*, <http://www.dublincore.org/documents/dces/>. The schema for unqualified Dublin Core used by the OAI-PMH is available at [http://www.openarchives.org/OAI/2.0/oai\\_dc.xsd](http://www.openarchives.org/OAI/2.0/oai_dc.xsd)

<sup>10</sup> Open Archives Initiative 2002, *Specification for an OAI Static Repository and an OAI Static Repository Gateway*, <http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>



# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

A Static Repository becomes accessible via OAI-PMH through a Static OAI-PMH Repository Gateway as described in Section 4.4.1.

## 4.4 Party Data Ingest

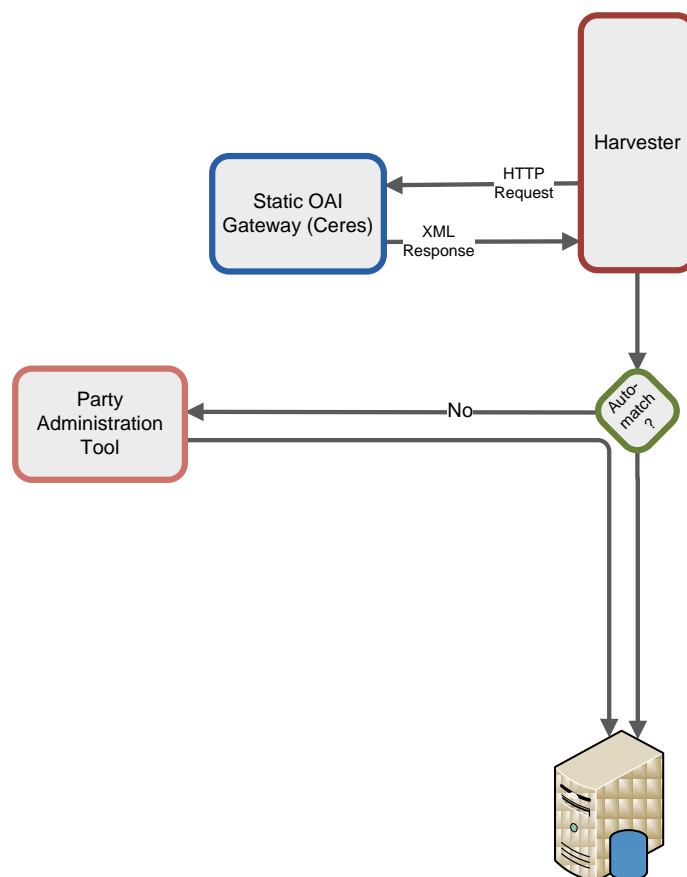


Figure 3: Ingest components

### 4.4.1 Static OAI-PMH Repository Gateway (Ceres)

In order to support contributors providing data using OAI Static Repositories the National Library has implemented an OAI-PMH Static Repository Gateway called Ceres.

Ceres indexes Static OAI-PMH Repositories and creates an OAI-PMH interface to these which are then harvestable by the NLA Harvester described in Section 4.4.2 below.

### 4.4.2 NLA Harvester

Party records for ingest into the party infrastructure are harvested from a contributor's OAI-PMH Repository or Ceres by the NLA Harvester.

The harvester is able to apply a number of processing steps to transform the structure and content of ingested XML files.

In the case of party records, these records are ingested in or transformed into EAC-CPF in the harvester and are then passed to the Identity Matching Service, described in Section 4.4.3, for further processing.

### 4.4.3 Identity Matching Service

Records ingested into People Australia by the harvester are passed to the Identity Matching Service; this is depicted as the auto-match process in *Figure 3*. This service attempts to automatically match incoming party records to existing parties in the People Australia database.

Currently the matching rules used in the Identity Matching Service are tailored for content provided by Libraries and related organisations and will need to be revised in Stage 2 of the project to maximise the chance that ingested records are able to be correctly matched to existing identities.

Successful matching relies on enough contextual information about a party being provided. It is also dependent on the richness of records in the People Australia data corpus to which these provided records are being matched. So in order to maximise the chance of a match, it is critical that institutions provide enough public information about a party to support matching. If sparse records that cannot be automatically matched are provided then contributors will need to devote the required resourcing to hand matching using the Party Administration Tool described in Section 4.4.4.

If a record is automatically matched to an existing party during the auto-match process it is assigned a persistent identifier. If a record is unable to be matched it is passed to the Party Administration Tool described in Section 4.4.4 for matching by hand.

### 4.4.4 Party Administration Tool

The Party Administration Tool allows for the management of records contributed to the People Australia program. Currently it allows authorised National Library staff to hand match ingested records with parties that already exist in the People Australia Database or to create new party records if they are new. The tool also allows NLA party records to be merged if records are duplicates, or split if it is found that one records relates to two of more parties.

The Party Administration Tool is to be extended as part of this project to allow data managers in the research sector to manage their party records. It will also be extended to enable reporting, provide for the management of a data manager's profile and to provide access to documentation about the tool.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

## 4.4.5 People Australia Database

The People Australia database stores all ingested records and maintains the relationships between Persistent Identifiers and these records. The OAI-PMH, Search/Retrieve via URL (SRU) and Trove People and Organisations interfaces described below provide access to this data.

## 4.5 Party Data Dissemination

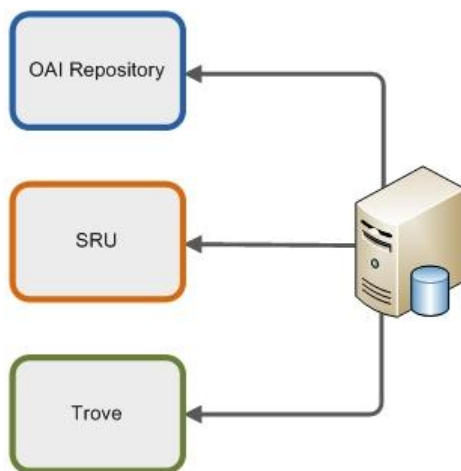


Figure 4: Dissemination components

### 4.5.1 OAI-PMH Repository

The People Australia OAI-PMH Repository located at <http://www.nla.gov.au/apps/peopleaustralia-oai/OAIHandler> provides records in the EAC, RIF-CS and Dublin Core metadata formats for harvesting by other services and in the near future it will support records in EAC-CPF format.

The People Australia OAI-PMH Repository supports sets. Currently these sets provide access to metadata records grouped by data contributor.

### 4.5.2 Search/Retrieve via URL (SRU) Interface

SRU is a standard XML-based search protocol for Internet search queries.<sup>11</sup> The People Australia SRU interface provides records in EAC, RIF-CS, Dublin Core and ATOM formats and is available at <http://www.nla.gov.au/apps/srw/search/peopleaustralia>. In the near future the interface will support records in EAC-CPF format also.

<sup>11</sup> The Library of Congress 2010, *Search/Retrieve via URL*, <http://www.loc.gov/standards/sru>.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

SRU uses the Contextual Query Language (CQL) to structure queries. The People Australia SRU interface supports CQL version 1.1, Level 1 queries and provides for precise searching across 45 indexes.

## 4.5.3 Trove People and Organisations view

Trove (<http://trove.nla.gov.au>) is the National Library's new integrated discovery service offering users access to approximately 90 million records from Australian and international collections.

The Trove People and Organisations view (<http://trove.nla.gov.au/people>) allows users to search over 880,000 party records and find persistent identifiers, biographical or other contextual information about parties (where available) as well as related resources by and about them.

## 4.6 Party Data Consumption

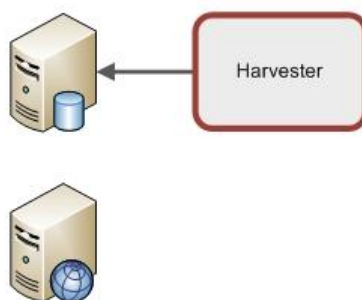


Figure 5: Data consumption components

### 4.6.1 OAI Harvester

Contributors who set up their own OAI-PMH harvester are currently able to harvest their records in the EAC, RIF-CS and Dublin Core metadata formats from the Party Infrastructure by harvesting metadata records from their sets. *Table 2* below provides examples.

Contributor	Set name	OAI Request for records in EAC format
Encyclopaedia of Australian Science	AU-VU:EOAS	<a href="http://www.nla.gov.au/apps/peopleaustralia-oai/OAIHandler?verb=ListRecords&amp;set=AU-VU:EOAS&amp;metadataPrefix=eac">http://www.nla.gov.au/apps/peopleaustralia-oai/OAIHandler?verb=ListRecords&amp;set=AU-VU:EOAS&amp;metadataPrefix=eac</a>
Australian Mines Atlas	AU-AMG	<a href="http://www.nla.gov.au/apps/peopleaustralia-oai/OAIHandler?verb=ListRecords&amp;set=AU-AMG&amp;metadataPrefix=eac">http://www.nla.gov.au/apps/peopleaustralia-oai/OAIHandler?verb=ListRecords&amp;set=AU-AMG&amp;metadataPrefix=eac</a>
Libraries Australia	AuCNLKIN	<a href="http://www.nla.gov.au/apps/peopleaustralia-oai/OAIHandler?verb=ListRecords&amp;set=AuCNLKIN&amp;metadataPrefix=eac">http://www.nla.gov.au/apps/peopleaustralia-oai/OAIHandler?verb=ListRecords&amp;set=AuCNLKIN&amp;metadataPrefix=eac</a>

Table 2: Example OAI Requests

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

## 4.6.2 SRU

Contributors using the SRU interface are currently able to query for all or some of their records in a range of record formats (EAC, RIF-CS, Dublin Core and ATOM).

As well as enabling the provision of records and their associated identifiers, SRU is ideal for use as part of web applications. Example queries follow in *Table 3* below and detailed SRU Documentation will be produced.

Query Description	SRU Query
Keyword search for records matching 'Mawson' in EAC format	<a href="http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=cli.anywhere+%3D+" mawson"&amp;version='1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fjefferson.village.virginia.edu%2Ffeac"'>http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=cli.anywhere+%3D+"Mawson"&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fjefferson.village.virginia.edu%2Ffeac</a>
Search for records matching urname 'Mawson' and firstname 'Paquita' or surname 'Delprat' and firstname 'Guillame' in EAC format	<a href="http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=%28pa.firstname+%3D+%22Paquita%22+and+pa.surname+%3D+%22Mawson%22%29+or+%28pa.firstname+%3D+%22Guillaume%22+and+pa.surname+%3D+%22Delprat%22%29&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fjefferson.village.virginia.edu%2Ffeac">http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=%28pa.firstname+%3D+%22Paquita%22+and+pa.surname+%3D+%22Mawson%22%29+or+%28pa.firstname+%3D+%22Guillaume%22+and+pa.surname+%3D+%22Delprat%22%29&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fjefferson.village.virginia.edu%2Ffeac</a>
Search for records matching all people born between 1945 and 1960 in ATOM format	<a href="http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=pa.existencedate+%3E+%221945%22+and+pa.existencedate+%3C+%221960%22&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fwww.w3.org%2F2005%2FAtom">http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=pa.existencedate+%3E+%221945%22+and+pa.existencedate+%3C+%221960%22&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fwww.w3.org%2F2005%2FAtom</a>
Search for all Encyclopaedia of Australian Science records in EAC format	<a au-vu:eoas"&amp;version='1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fjefferson.village.virginia.edu%2Ffeac"' href="http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=bath.possessingInstitution+%3D+">http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=bath.possessingInstitution+%3D+"au-vu:eoas"&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fjefferson.village.virginia.edu%2Ffeac</a>
Search for record with local identifier 'P000553' in Dublin Core format	<a href="http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=rec.identifier+%3D+%22P000553%22&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=info%3Asrw%2Fschema%2F1%2Fdc-v1.1">http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=rec.identifier+%3D+%22P000553%22&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=info%3Asrw%2Fschema%2F1%2Fdc-v1.1</a>
Search for record with party identifier 'nla.party-607285' in EAC format	<a href="http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=rec.identifier+%3D+%22nla.party-607285%22&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fjefferson.village.virginia.edu%2Ffeac">http://www.nla.gov.au/apps/srw/search/peopleaustralia?query=rec.identifier+%3D+%22nla.party-607285%22&amp;version=1.1&amp;operation=searchRetrieve&amp;recordSchema=http%3A%2F%2Fjefferson.village.virginia.edu%2Ffeac</a>

Table 3: Example SRU Requests

## 4.7 Party Infrastructure Usage

In order to contribute to the ARDC Party Infrastructure there are two sets of requirements that need to be met these are outlined below.

### 4.7.1 Interoperability

In order for researchers or research organisations to be assigned a persistent identifier, records about them need to be provided to the ARDC Party Infrastructure. To streamline this process of contributing records, the ARDC

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

Party Infrastructure will use the OAI-PMH for ingest. This is the protocol recommended by ANDS for providing metadata about collections, services and activities to the ANDS Collections Registry and Research Data Australia.

## 4.7.2 Record Format

The People Australia program uses the Encoded Archival Context for Corporate bodies, Persons and Families (EAC-CPF) standard to exchange records about parties.<sup>12</sup>

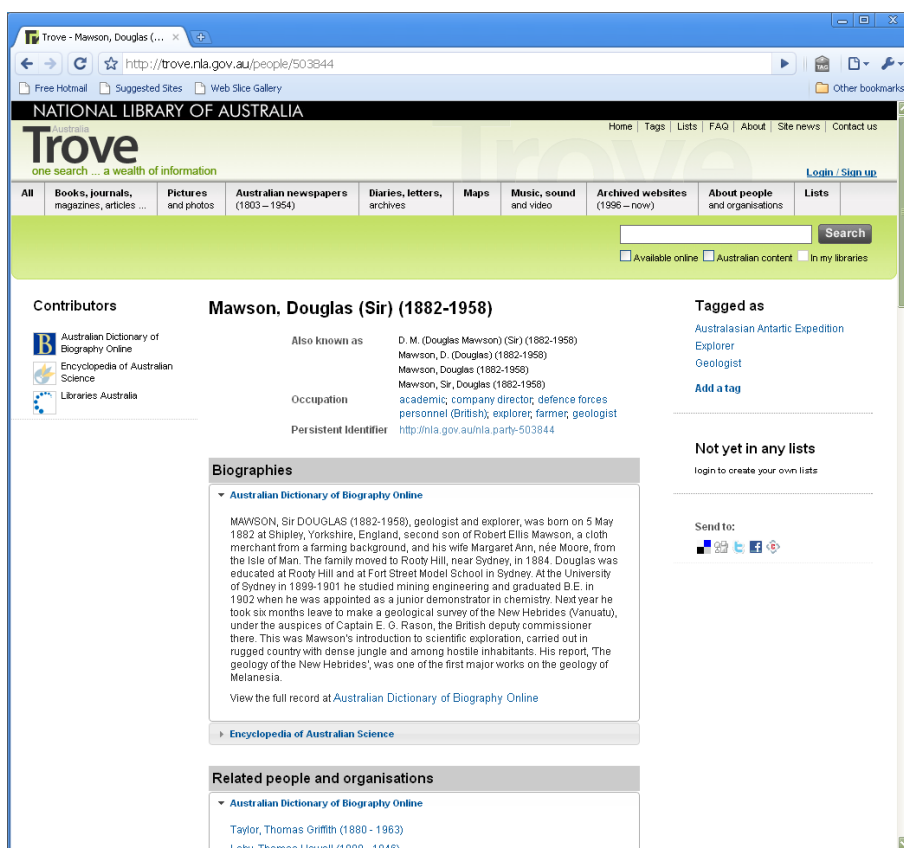
EAC-CPF is a rich standard designed to support the detailed description of parties which groups name variants for a party and, where available, includes biographical and contextual information, references to related resource records, references to related parties and details about a party's function, field of activity or occupation.

In the People Australia database the EAC container records relate EAC party records from a number of contributors. *Figure 6* below depicts how EAC records (in yellow) are grouped into an overarching party record (in blue). It also shows how Local Identifiers are associated with NLA party identifiers. *Figure 7* depicts the Trove display of the Douglas Mawson EAC record.

Name:	Mawson, Douglas (Sir) (1882-1958)
Party ID:	nla.party-503844
Name:	Mawson, Douglas, Sir, 1882-1958 Mawson, D. (Douglas), 1882-1958 D. M (Douglas Mawson), Sir, 1882-1958
Source:	Libraries Australia Name Authority File
Local ID:	anbd.aut-an35335937
Name:	Mawson, Douglas (1882 - 1958)
Source:	Encyclopaedia of Australian Science
Local ID:	P000631
Name:	Mawson, Sir Douglas (1882 - 1958)
Source:	Australian Dictionary of Biography Online
Local ID:	A100444

**Figure 6: EAC Record for Douglas Mawson**

<sup>12</sup> Encoded Archival Context (EAC) Working Group 2010, *Encoded Archival Context for Corporate bodies, Persons and Families (EAC-CPF) Homepage*, <http://eac.staatsbibliothek-berlin.de/>.



**Figure 7: Douglas Mawson Record in Trove**

The use of EAC-CPF as the format for contribution will mean that ingested records are more likely to match. It should be noted that the richer the content of a party record, the more likely it is to be automatically matched with a party record about the same person or organisation.

While records are stored in EAC format the NLA can ingest party records in other formats which are then mapped to and stored in EAC format. Contributors who choose to use other formats and provide more sparse records would need to devote resources to hand-matching of records.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

## 4.8 Roles and Responsibilities

Table 4 below summarises all possible roles and responsibilities of ARDC Contributors, the ARDC Party Infrastructure and the Australian National Data Service.

Party	Role	Responsibilities
ARDC Contributor	Data provider	<ul style="list-style-type: none"><li>• Creates and manages records about data collections, research activities, services and parties.</li><li>• Provides records about parties in RIF-CS or EAC-CPF format via an OAI Repository.</li><li>• Obtains and stores records/persistent identifiers for parties via Trove or the ARDC Party Infrastructure OAI or SRU interfaces.</li><li>• Uses persistent identifiers to link collection, service, activity and party records with related parties.</li><li>• Provide collection, activity and service records to ANDS Collection Registry/Research Data Australia.</li></ul>
ARDC Party Infrastructure	Party aggregator	<ul style="list-style-type: none"><li>• Harvests and ingests party records from ARDC Contributors.</li><li>• Harvests and ingests party records from ANDS Collection Registry/Research Data Australia.</li><li>• Automatically matches ingested party records and assigns identifiers if possible.</li><li>• Provide Party Administration Tool for use by the ARDC Contributor for the purposes of matching and identifier creation for unmatched records.</li><li>• Assigns and maintains globally unique, public and persistent identifiers.</li><li>• Provides party records to the ANDS Collections Registry/Research Data Australia</li><li>• Provides human and machine-to-machine access to party records and associated identifiers.</li></ul>
Australian National Data Service	ARDC aggregator	<ul style="list-style-type: none"><li>• Harvests and ingests records about data collections, research activities, services and parties into the ANDS Collection Registry/Research Data Australia.</li><li>• Provides human and machine-to-machine access to records about data collections, research activities, services and parties via ANDS Collection Registry/ Research Data Australia.</li></ul>

**Table 4: Party Infrastructure Roles and Responsibilities**

The currently defined interactions with the ARDC Party Infrastructure are described in the four scenarios below along with the specific roles and responsibilities that relate to each. Other local approaches may suggest other scenarios for use in future but at base each scenario provides for local party records to be assigned global, public and persistent identifiers. At this stage Scenarios 1 and 2 are the preferred scenarios for interaction with the party infrastructure. Scenarios 3 and 4 are provided for ARDC Contributors who will transition to either Scenario 1 or 2 as they develop and implement systems capable of full interaction with the infrastructure in future.

It is important to note that in an operational mode the Party Infrastructure will be able to be embedded into local workflows easily. Party records will be able to be supplied by an ARDC Contributor at any time and identifiers created and provided as soon as records are automatically or hand matched. Supplied party



# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

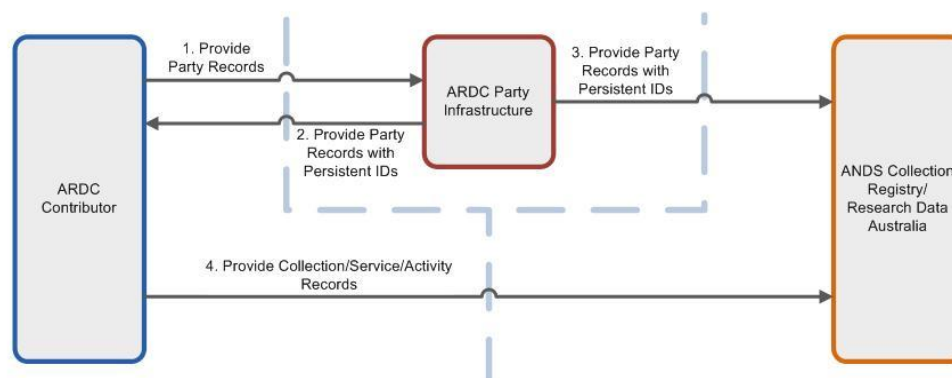
records will be able to be updated by the ARDC Contributor as more information about the party becomes available over time.

## 4.8.1 Scenario 1

The ARDC Contributor is using a metadata store system designed to work with the Party Infrastructure. They want to:

- Obtain Party Identifiers for use in their internal systems and the ARDC;
- Have the ARDC Party Infrastructure contribute party records and identifiers to ANDS on their behalf; and,
- Provide collection, activity and service records to the ANDS Collection Registry/Research Data Australia that incorporate public, persistent Party Identifiers.

Figure 8 below provides a visual representation of this scenario with the arrows indicating record flow and numbers indicating the sequence.



**Figure 8: Contribution Scenario 1**

The following list provides a detailed overview of roles and responsibilities for this scenario:

- ARDC Contributor creates and manages records about data collections, research activities, services and parties.
- ARDC Contributor provides access to records about parties in RIF-CS or EAC-CPF format via an OAI Repository.
- ARDC Party Infrastructure harvests and ingests party records from ARDC Contributor.
- ARDC Party Infrastructure automatically matches ingested party records if possible.
- ARDC Party Infrastructure provides Party Administration Tool for use by the ARDC Contributor for the purposes of record matching and creation for unmatched records.
- ARDC Party Infrastructure assigns and maintains globally unique, public and persistent identifiers to records.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

- ARDC Contributor obtains persistent identifiers for parties via Trove or the ARDC Party Infrastructure OAI or SRU interfaces and stores them.
- ARDC Contributor uses persistent identifiers for parties as registry keys to link collection, service, activity and party records with related party records.
- ARDC Party Infrastructure provides party records to the ANDS Collections Registry/Research Data Australia with public persistent identifiers as registry keys and identifiers.
- ARDC Contributor provides collection, activity and service records to ANDS Collection Registry/Research Data Australia using the persistent identifiers for parties as registry keys.
- Australian National Data Service harvests and ingests records about data collections, research activities, services and parties into the ANDS Collection Registry/Research Data Australia.

## 4.8.2 Scenario 2

The ARDC Contributor has systems designed to work with the Party Infrastructure. They want to:

- Obtain Party Identifiers for use in their internal systems and the ARDC;
- Contribute their own party records and public, persistent identifiers to ANDS; and,
- Provide collection, activity and service records to the ANDS Collection Registry/Research Data Australia that incorporate public, persistent Party Identifiers.

Figure 9 below provides a visual representation of this scenario with the arrows indicating record flow and numbers indicating the sequence.

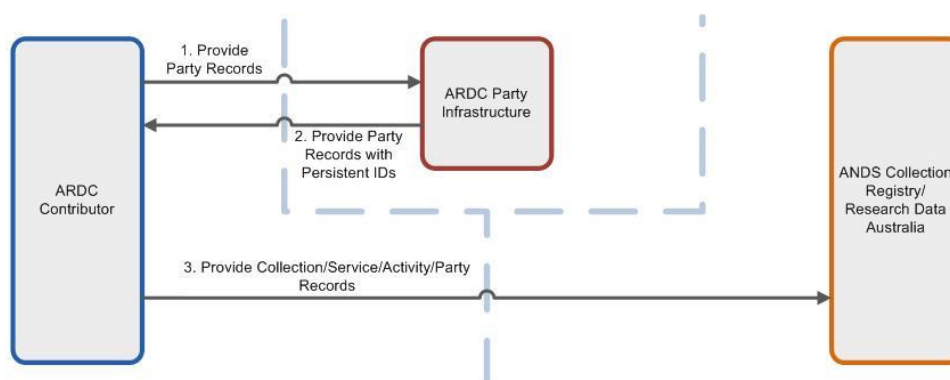


Figure 9: Contribution Scenario 2

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

The following list provides a detailed overview of roles and responsibilities for this scenario:

- ARDC Contributor creates and manages records about data collections, research activities, services and parties.
- ARDC Contributor provides access to records about parties in RIF-CS or EAC-CPF format via an OAI Repository.
- ARDC Party Infrastructure harvests and ingests party records from ARDC Contributor.
- ARDC Party Infrastructure automatically matches ingested party records if possible.
- ARDC Party Infrastructure provides Party Administration Tool for use by the ARDC Contributor for the purposes of record matching and creation for unmatched records.
- ARDC Party Infrastructure assigns and maintains globally unique, public and persistent identifiers to records.
- ARDC Contributor obtains persistent identifiers for parties via Trove or the ARDC Party Infrastructure OAI or SRU interfaces and stores them.
- ARDC Contributor uses local identifiers for parties as registry keys to link collection, service, activity and party records with related parties.
- ARDC Contributor provides party records to the ANDS Collections Registry/Research Data Australia with local identifiers for parties as registry key and the public persistent identifiers for parties as identifiers.
- ARDC Contributor provides collection, activity, service records to ANDS Collection Registry/Research Data Australia using the local identifiers as registry keys.

## 4.8.3 Scenario 3

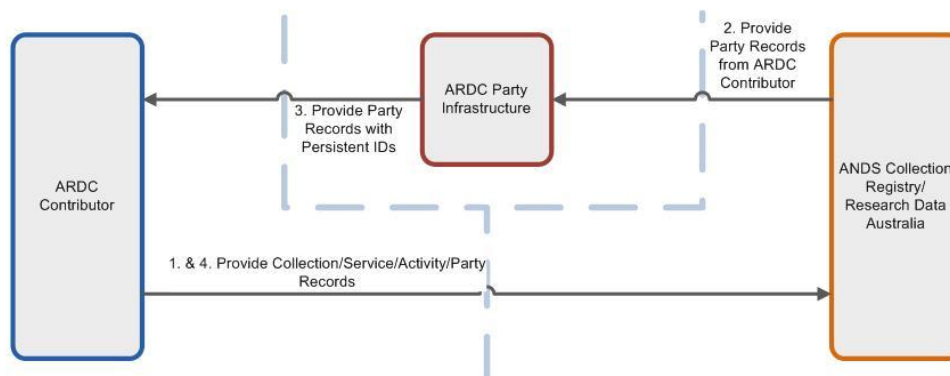
The ARDC Contributor has systems designed to provide collection, activity, service and party records to the ANDS Collection Registry/Research Data Australia but are not yet able to implement an OAI Repository feed of party information directly to the ARDC Party Infrastructure. They want to:

- Obtain Party Identifiers for use in their internal systems and the ARDC;
- Contribute their own party records and identifiers to ANDS; and,
- Provide collection, activity and service records to the ANDS Collection Registry/Research Data Australia that incorporate public, persistent Party Identifiers.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

Figure 10 below provides a visual representation of this scenario with the arrows indicating record flow and numbers indicating the sequence.



**Figure 10: Contribution Scenario 3**

The following list provides a detailed overview of roles and responsibilities for this scenario:

- ARDC Contributor creates and manages records about data collections, research activities, services and parties.
- ARDC Contributor provides access to records about parties in RIF-CS or an OAI Repository or by using the 'Direct' method.
- Australian National Data Service harvests and ingests records about data collections, research activities, services and parties into the ANDS Collection Registry/Research Data Australia.
- Australian National Data Service provides machine-to-machine access (OAI) to records about parties via ANDS Collection Registry/ Research Data Australia.
- ARDC Party Infrastructure harvests and ingests party records from ANDS Collection Registry/Research Data Australia.
- ARDC Party Infrastructure automatically matches ingested party records if possible.
- ARDC Party Infrastructure provides Party Administration Tool for use by the ARDC Contributor for the purposes of record matching and creation for unmatched records.
- ARDC Party Infrastructure assigns and maintains globally unique, public and persistent identifiers to records.
- ARDC Contributor obtains persistent identifiers for parties via Trove or the ARDC Party Infrastructure OAI or SRU interfaces and stores them.
- ARDC Contributor uses local identifiers for parties as registry keys to link collection, service, activity and party records with related parties.
- ARDC Contributor provides party records to the ANDS Collections Registry/Research Data Australia with local identifiers for parties as registry key and the persistent identifiers for parties as identifiers.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

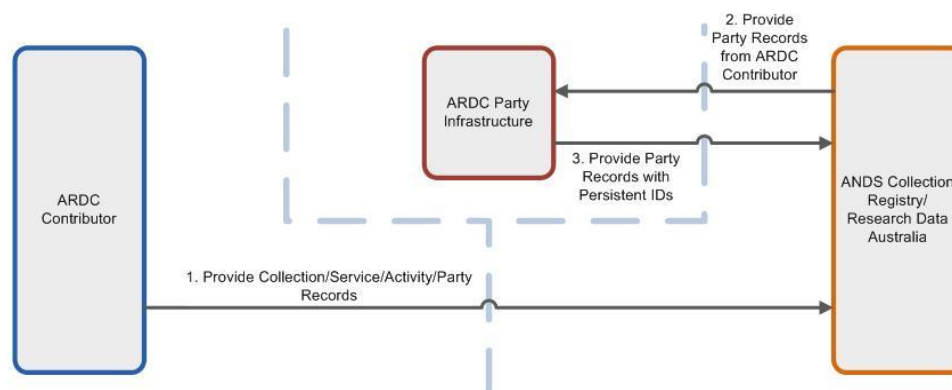
- ANDS harvests and ingests records about data collections, research activities, services and parties into the ANDS Collection Registry/Research Data Australia.

## 4.8.4 Scenario 4

ARDC Contributor's systems are unable to store ARDC Party Identifiers. They want to:

- Have Party Identifiers assigned for later use in their internal systems and in the ARDC;
- Contribute their own party records and identifiers to ANDS; and,
- Provide collection, activity, service and party records to the ANDS Collection Registry/Research Data Australia.

Figure 11 below provides a visual representation of this scenario with the arrows indicating record flow and numbers indicating the sequence.



**Figure 11: Contribution Scenario 4**

The following list provides a detailed overview of roles and responsibilities for this scenario:

- ARDC Contributor creates and manages records about data collections, research activities, services and parties.
- ARDC Contributor provides access to records about parties in RIF-CS or EAC-CPF format via an OAI Repository or by using the 'Direct' method.
- Australian National Data Service harvests and ingests records about data collections, research activities, services and parties into the ANDS Collection Registry/Research Data Australia.
- Australian National Data Service provides machine-to-machine access (OAI) to records about parties via ANDS Collection Registry/ Research Data Australia.
- ARDC Party Infrastructure harvests and ingests party records from ANDS Collection Registry/Research Data Australia.

# Australian Research Data Commons (ARDC) Party Infrastructure

Final Draft

- ARDC Party Infrastructure automatically matches ingested party records if possible.
- ARDC Party Infrastructure provides Party Administration Tool for use by the ARDC Contributor for the purposes of record matching and creation for unmatched records.
- ARDC Party Infrastructure assigns and maintains globally unique, public and persistent identifiers to records.
- ARDC Party Infrastructure provides party records to the ANDS Collections Registry/Research Data Australia with persistent identifiers for parties as registry keys and identifiers.
- Australian National Data Service harvests and ingests records about data collections, research activities, services and parties into the ANDS Collection Registry/Research Data Australia.

## 4.8.5 Summary of record sources and identifier/key usage

For each of the scenarios presented above, *Table 5* below summarises the source of the ANDS Collection Registry/Research Data Australia Party Record and the Party registry keys and identifiers to be used.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
<b>ANDS Collection Registry/Research Data Australia Party Record Source:</b>	ARDC Party Infrastructure	ARDC Contributor	ARDC Contributor	ARDC Contributor then ARDC Party Infrastructure
<b>RIF-CS Party Registry Key:</b>	ARDC Party Infrastructure Persistent Party Identifier	ARDC Contributor Local Key	ARDC Contributor Local Key	ARDC Contributor Local Key then NLA Persistent Party Identifier (and any others)
<b>RIF-CS Party Identifier:</b>	ARDC Party Infrastructure Persistent Party Identifier (and any others)	ARDC Party Infrastructure Persistent Party Identifier (and any others)	ARDC Contributor Local Identifier initially and then NLA Persistent Party Identifier (and any others)	ARDC Contributor Local Identifier
<b>RIF-CS Related Registry Keys for Parties:</b>	ARDC Party Infrastructure Persistent Party Identifier	ARDC Contributor Local Key	ARDC Contributor Local Key or NLA Persistent Party Identifier	ARDC Contributor Local Key

**Table 5: Record sources and identifier/key usage**