

Service Discovery

The Service Discovery functionality within the RDA Registry allows Data Source Administrators to autogenerate RIF-CS service records for Open Geospatial Consortium (OGC) services referenced within their collection records.

The discovery process

The process works by identifying possible OGC service URLs within the 'address/electronic', 'relatedInfo/identifier' and 'relation/url' elements of published collection records within a data source. URLs which contain the following character sequences (case insensitive) are considered possible service URLs: "wms", "wfs", "ogc", "wcs", "wps", "wmts" and "ows".

For each URL discovered within a data source, the Service Discovery process will attempt a 'GetCapabilities' request. The request is sent without the 'version' parameter in order to request the latest version of metadata from the service. If a successful response is returned from the service, the Service Discovery process will use the returned metadata to construct a RIF-CS service record.

Once the RIF-CS service record has been created, relationships to all collection records which reference the service URL are added. This includes collections from other data sources. The business rules for the creation of the relationships from the service to the collections are as follows:

- Where the relationship from the collection to service is described using the 'relatedInfo' element, the Service Discovery process will attempt to reverse the relationship value in the 'relatedInfo/relation' element. If a reverse relationship cannot be generated the original value will be used. Where the 'relatedInfo/relation' element is not populated, the relationship will default to 'hasAssociationWith'
- Where the relationship from the collection to service is described using the 'address/electronic' element, the relationship will be set to 'makesAvailable'.

After the relationships have been added, the subjects from all the related collections are extracted and added to the service record.

The service record is then imported into the data source as a draft.

Running the Service Discovery process

1. Log into the RDA Registry and navigate to the Dashboard for the data source you wish to discovery services in.
2. Click the down arrow shown on the 'Import from Harvester' button.
3. Select the 'Run Service Discovery' option to schedule a Service Discovery task.
4. Upon starting the task, the system will create a new Activity Log entry:

'Background Task for <data source name> Service Discovery Started'

The progress of the task can be seen in the Harvester Status section of the page.

5. Upon completing the task, the system will create a new Activity Log entry:

'Background Task for <data source name> Service Discovery Completed'

Click the entry in the Activity Log to show the details.

- The 'Valid Records Received in Harvest' field will indicate how many service records have been created. If this field is not shown then no service records were created.
- The number of tested and invalid links will also be shown in the log entry. To find out which URLs failed, click the 'Show Task' link in the 'Harvester Status' section of the page. This will open the 'Task Content' popup. Scroll down until you find messages in the task which contain the text 'Not a Valid Service url:'.

Harvester Status

Status **IDLE**
 URI 2019-04-01 09:46:23 (6 hrs ago)
 Last Run Once Off
 Harvest Frequency

Harvest completed!

Show Task ← Import from Harvester

Activity Log All Logs

Background Task for AUTestingRecords2(546) ServiceDiscovery Completed 1 min ago

```

Background Task for AUTestingRecords2(546) ServiceDiscovery Completed
Batch ID: 78DE795A280AF12C3A90154DB30ED8FCBDBCB179
Time: 2019-04-01T15:32:20Z
TaskID: 521735
Default Import Status for Data Source: PUBLISHED
Target Status for Import: DRAFT
Valid Records Received in Harvest: 76
Record exist in other Datasource(s): 6
Records content unchanged: 70
Number of DRAFT records Before Import: 73
Number of DRAFT records After Import: 73
Number of service links tested: 154
Number of service links which didn't return a valid response: 78
  
```

6. Where service records have been created these will be shown on the Manage My Records page in a status of 'Draft'. These records should be reviewed and edited before being published. Note:

- The 'Group' and 'Originating Source' attributes of the service record will be set to a default value of 'ServiceDiscoveryProcess'. The 'Group' value should be changed prior to publishing so that the record appears in the correct group in Research Data Australia.
- The relationships to and from the service record will not be shown when previewing the 'Draft' record. This is a known issue with the Research Data Australia preview functionality. The relationships will display correctly once the service record has been published.

Re-running the Service Discovery process

The Service Discovery process can be run multiple times without losing content.

Where a 'Draft' service record already exists in your data source for a discovered service URL, the Service Discovery process will attempt to add additional related collections and subjects to the existing draft record. If only a published version of the record exists for a discovered service URL, the Service Discovery process will create a draft copy of the existing published record if performing any updates.

No other elements in existing records will be updated. This is done to ensure that any content manually changed/added is not overwritten. However, the drawback of this approach is that in order to completely refresh the information automatically retrieved from a service, you will need to delete any existing versions of the service record, before re-running the Service Discovery process.


Service URLs discovered in multiple data sources

The Service Discovery process uses the base URL (all parameters removed) for each discovered service to create a key for the generated RIF-CS record. This has the potential to cause key clashes where a service URL is present in collection records across more than one data source.

If another data source contains a discovered service record for one of your service URLs, you will not be able to generate and import the service record yourself. If you encounter this issue and would like to find out more information please contact service@ardc.edu.au.

ISO19115-3 Record Generation

Upon publishing a discovered service record, an ISO 19115-3 representation of the record will be generated. These records can be accessed by clicking on the 'iso19115-3' label displayed in the 'Registry Metadata' section of the RDA Registry view page for a discovered service record.

Registry Metadata	
Title	aasgeothermal_CAWellLogs
Status	Published
Data Source	AUTestingRecords2
Key	a81e1ee7-b31b-427b-1fe7-16d2e379f078
ID	1276199
URL "Slug"	aasgeothermalcawelllogs
Last edited by	Joel Benn
Date last changed	1 April 2019, 9:59 am
Date created	1 April 2019, 10:36 am
Feed type	Harvest
Quality Assessed	No
Other formats	iso19115-3 
<input type="button" value="Delete Record"/>	

The records are also exposed via the Research Data Australia OAI-PMH service and can be retrieved by requesting records using the metadataPrefix 'iso19115-3'. E.g. <https://researchdata.edu.au/registry/services/oai?verb=ListRecords&metadataPrefix=iso19115-3>

Manually creating service records

To assist users manually creating records for their services, the Service Discovery process has also been integrated into the manual entry form in the RDA Registry.

When manually adding a new OGC service record, the user will be presented with an optional 'OGC Service URL' field to enter the URL of their service (see image).

Upon clicking the 'Add New Service' button the Service Discovery process (as described above) will be run for the entered URL. If successful, the resulting RIF-CS record will be pre-populated into the manual entry form where the details can be edited and enriched before being saved or published. If unsuccessful, an empty manual entry form will be displayed.

Add New Registry Object
✕

Please provide the required information for the registry object

Key ?

ff73c078-6d59-aa6b-77fa-83231ea7eaf8
↻

Key must be unique and is case sensitive

Data source

AUTestingRecords
▾

Group ?

AUTestingRecords
▾

Type

OGC:WCS
▾

OGC Service URL (optional)

Where possible the form will be auto populated with information retrieved from the service

Add New service

GetCapabilities to RIF-CS mapping

The following mapping is used to crosswalk GetCapabilities responses into a RIF-CS service records.

GetCapabilities Element/Attribute	RIF-CS Element/Attribute
ServiceType	Service @type
Title OR Label OR Name	Name @type='primary'
Abstract	Description @type='brief'
Keyword	Subject @type='local'
OnlineResource OR Operation @name="GetCapabilities"	Location/address/electronic @type='url'
ContactInformation / ServiceContact	Location/address/physical
AddressType	@type= If AddressType is null/not present the type will default to 'postal'
ContactPosition OR PositionName	addressPart @type='fullName'
ContactOrganization	addressPart @type='organizationName'
Address	addressPart @type='addressLine'
City	addressPart @type='suburbOrPlaceOrLocality'
StateOrProvince OR AdministrativeArea	addressPart @type='stateOrTerritory'
Country	addressPart @type='country'

PostCode OR PostalCode	addressPart @type='postCode'
DeliveryPoint	addressPart @type='deliveryPointIdentifier'
ContactVoiceTelephone OR Voice	addressPart @type='addressPart @type='telephoneNumber'
ContactFacsimileTelephone OR Facsimile	addressPart @type='addressPart @type='faxNumber'
ContactElectronicMailAddress OR ElectronicMailAddress	Location/address/electronic @type='email'
AccessConstraints	Rights/AccessRights
Layer/EX_GeographicBoundingBox OR WGS84BoundingBox OR lonLatEnvelope	Coverage/spatial @type=' iso19139dcmiBox'