



Data Service Discoverability

A Metadata Perspective

Alex Ball

University of Bath

Data Service Discoverability Workshop, Santorini
21–22 April 2016

research data sharing without barriers
rd-alliance.org



Data services

Definitions

- ▶ A **service** is a capability provided by someone else.
- ▶ A **research data service** is one that involves research data in some way...


We can imagine varieties across different dimensions:

- ▶ Is the service aimed at **people** (👤) or **tools** (⚙️)?
- ▶ Does the service take **metadata** (📄) or **data** (🗄️) as input?
- ▶ Does the service give **metadata** (📄) or **data** (🗄️) as output?

A single provider may offer more than one service.


Data services – examples

 → Data description registry → 

 Browsing/searching


 Harvesting

  → Data repository (ingest) → 

 Manual deposit

 SWORD deposit

 → Data repository (access) →  

 Browsing/searching

 Harvesting

 → Online data analysis library → 

 Workflow scripting

Discovery pathways

General search for service by capability

Search for catalogue relating to capability

List of services

Filter list



Pick one based on description



Follow URL



Interact with Web form



Pick one based on properties

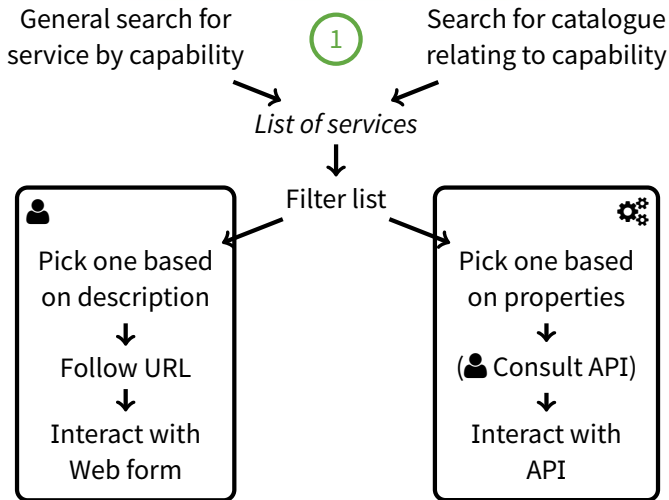


(User icon) Consult API

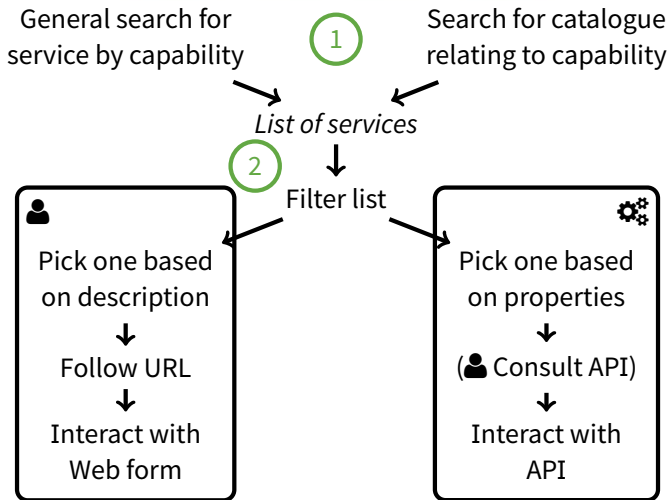


Interact with API

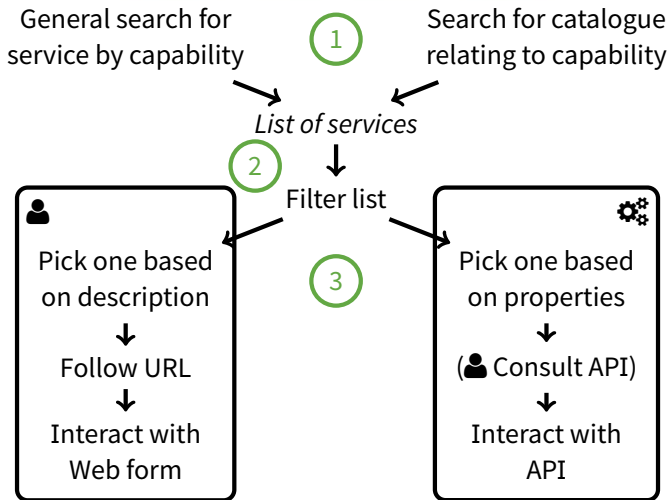
Discovery pathways



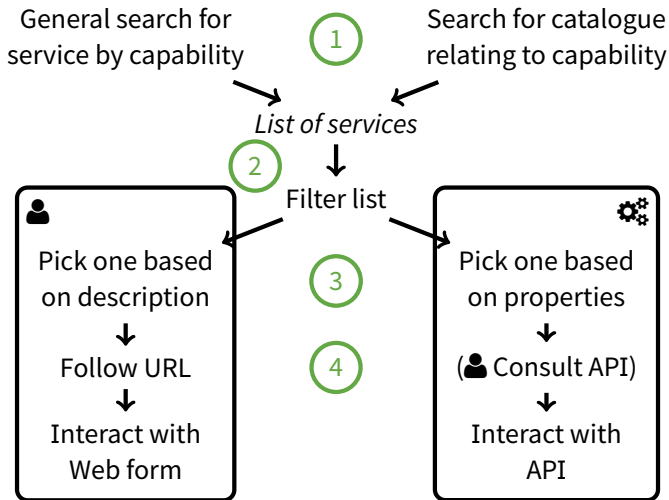
Discovery pathways



Discovery pathways



Discovery pathways



Discovering metadata standards

DCC Disciplinary Metadata Catalogue

1. Listing of metadata standards
2. Designed to have subject/discipline as main filter
3. Description of
 - ▶ what sort of data the metadata describes
 - ▶ who maintains it, and how often
4. Useful information
 - ▶ documentation
 - ▶ vocabularies
 - ▶ crosswalks
 - ▶ tools
 - ▶ services with expertise

Discovering metadata standards

DCC Disciplinary Metadata Catalogue

1. Listing of metadata standards
2. Designed to have subject/discipline as main filter
3. Description of
 - ▶ what sort of data the metadata describes
 - ▶ who maintains it, and how often
4. Useful information
 - ▶ documentation
 - ▶ vocabularies
 - ▶ crosswalks
 - ▶ tools
 - ▶ services with expertise

Discovering metadata standards

DCC Disciplinary Metadata Catalogue

1. Listing of metadata standards
2. Designed to have subject/discipline as main filter
3. Description of
 - ▶ what sort of data the metadata describes
 - ▶ who maintains it, and how often
4. Useful information
 - ▶ documentation
 - ▶ vocabularies
 - ▶ crosswalks
 - ▶ tools
 - ▶ services with expertise

Discovering metadata standards

DCC Disciplinary Metadata Catalogue

1. Listing of metadata standards
2. Designed to have subject/discipline as main filter
3. Description of
 - ▶ what sort of data the metadata describes
 - ▶ who maintains it, and how often
4. Useful information
 - ▶ documentation
 - ▶ vocabularies
 - ▶ crosswalks
 - ▶ tools
 - ▶ services with expertise

Organisational profile documents

Autodiscovery works either by

```
http://www.bath.ac.uk/
```

```
<link rel="openorg" href="http://www.bath.ac.uk/organisation-profile.opd" />
```

or a redirect from <http://www.bath.ac.uk/.well-known/openorg>

Conventions for specifying datasets:

<http://opd.data.ac.uk/docs/datasets>

- ▶ Each institution lists its data resources
- ▶ Correct 'output' identified by subject and format/API (conformsTo)

Organisational profile documents

```
http://www.bath.ac.uk/organisation-profile.opd
```

```
@prefix oo: <http://xmlns.com/foaf/0.1/> .  
@prefix dcterms: <http://purl.org/dc/terms/> .  
# And so on...
```

```
"profile.ttl" rdf:type oo:OrganizationProfileDocument ;  
foaf:primaryTopic <http://www.bath.ac.uk/#org> .
```

```
<http://www.bath.ac.uk/#org> rdf:type org:FormalOrganization ;  
rdfs:label "University of Bath (The)" ;  
foaf:homepage <http://www.bath.ac.uk/> ;  
# And so on...
```

```
<http://www.bath.ac.uk/equipment/equipment-list.xls> rdf:type dcat:Download ;  
oo:organization <http://www.bath.ac.uk/#org> ;  
dcterms:subject <http://purl.org/openorg/theme/equipment> ;  
dcterms:conformsTo <http://equipment.data.ac.uk/uniqueip> ;  
dcterms:license <http://creativecommons.org/publicdomain/zero/1.0/> ;  
oo:contact <mailto:research-equipment-sharing@bath.ac.uk> ;  
oo:corrections <mailto:research-equipment-sharing@bath.ac.uk> .
```

Outstanding questions

Is the **OWL-S framework** appropriate for describing data services?

- ▶ Promising for fully automated interactions with services
- ▶ Would rely on service providers writing the profiles
- ▶ Could we get by with something simpler?

How should we **describe inputs/outputs**? Do we need to?

- ▶ Implicit in the nature of the service
- ▶ Implicit in support for standard protocols
- ▶ Controlled list of 'known' protocols, otherwise link to formal definitions (WSDL, OWL-S, ...)?

Outstanding questions

Are they **stateless** or **state-based** services?

- ▶ Client state: depends on the complexity of the transaction
- ▶ Server state: reproducibility versus machine learning?

What about **scientific workflows**?

- ▶ Not sure about using them to describe services/processes
- ▶ But data services might be used as part of a scientific workflow
- ▶ Talk to vendors about incorporating data service registries into products?

Do we need **discipline-specific classification** of data services?

- ▶ Need category-specific classifications to which disciplines contribute

Outstanding questions

Is **scalability** a functional characteristic?

- ▶ Would be useful to know about
 - ▶ maximum input size
 - ▶ order-of-magnitude turnaround time if not 'instant'

Is **citation** instrumental in making services discoverable?

- ▶ Citation more appropriate for workflows (myExperiment, SageCite)
- ▶ Certain expectations for cited material also helpful for services: persistent URLs and global IDs, version control...



Metadata
Standards
Catalog
WORKING GROUP

Thank you for your attention

Alex Ball: <http://alexball.me.uk/>

Metadata Standards Catalog Working Group:
<https://rd-alliance.org/groups/metadata-standards-catalog-working-group.html>

research data sharing without barriers
rd-alliance.org

