

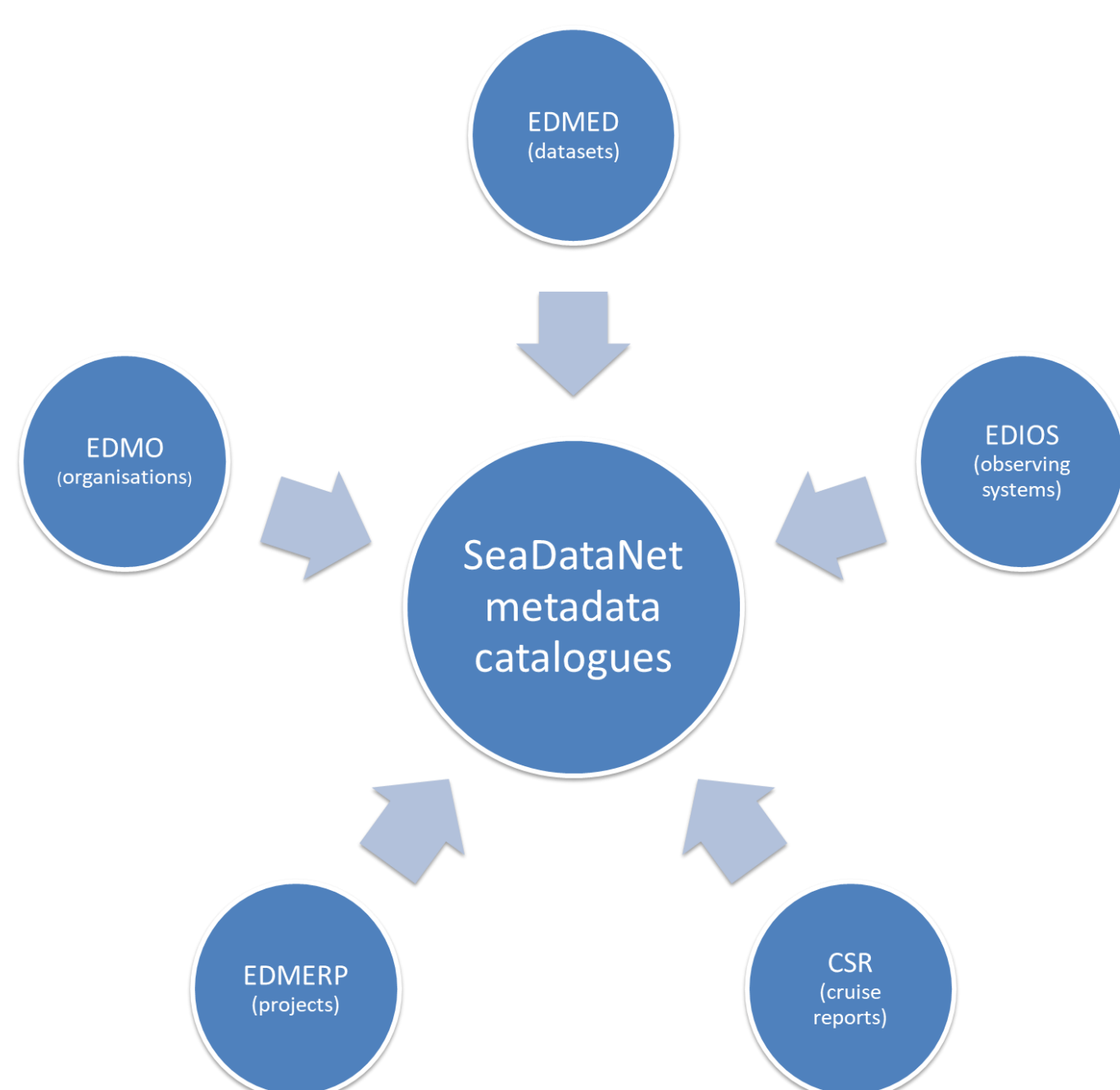


Why should we expose metadata catalogues via SPARQL endpoints?

Chris Wood & Alexandra Kokkinaki (British Oceanographic Data Centre, UK)

Adam Leadbetter (Marine Institute, Ireland)

Background & motive



- There are 5 SeaDataNet metadata catalogues, all of which hosted by separate institutions in different databases, with distinct web-interfaces for searching each catalogue
- However, although there are links between all 5 catalogues – it is currently not straightforward for end-users to determine these links
- Making these links available to the community via a standards-based metadata catalogue would be a valuable resource
- We have chosen to expose each of the catalogues via its own RDF-based triplestore, with appropriate ontological mappings

Technology



Ontologies

- Properties from these standard ontologies have been used
- Properties have been chosen to keep the data model as simple and intuitive as possible for end-users
- A vocabulary to help describe metadata from cruises and cruise reports has been created and published as there was no suitable current vocabulary

skos	org	vcard	rdfs	geo	schema
dct	prov	dbo	dcat	gl	rdf

Advantages for catalogue publishers

- Additional items of metadata are easily added and will be immediately exposed to users
- Search interfaces are more easily created & updated
- There is a large community to help with choosing correct vocabularies to describe the metadata, and there are mature RDF software tools to support publication of the databases over HTTP

Advantages for catalogue users

- The 5 SeaDataNet metadata catalogues now all use a common query interface, and can be queried directly over HTTP without using the existing search interfaces
- The catalogues are accessible to anyone with any knowledge of RDF databases
- Querying is simple as standard vocabularies have been used

Acknowledgements

- Funding was via EU H2020 grant 730960
- The authors would like to thank Richard Downer, Mark Hebden & Lesley Rickards (BODC), Rob Thomas (Marine Institute), and Dick Schaap & Peter Thijsse (MARIS, Netherlands) for support and helpful discussions

Federated queries

- One of the primary advantages of using triplestores to expose the catalogues is that multiple catalogues can be queried via a single SPARQL endpoint
- For example, it is not currently possible, without data replication, to determine the name of the contact organisation for a given EDMED dataset. However, with the imminent publication of the SeaDataCloud metadata catalogues as triplestores, the organisation could be easily found using a simple federated SPARQL query. In the example below, the result of the query will return “Swedish Meteorological and Hydrological Institute”

```

SELECT ?datasetName ?orgName WHERE {
  <http://www.bodc.ac.uk/resources/inventories/edmed/report/27/> dct:title ?datasetName ;
  dcat:contactPoint ?organisation .
  SERVICE <http://edmo.seadatanet.org/sparql/> { ?organisation skos:prefLabel ?orgName . }
}
  
```