

Data standards; why standardise data and what are the benefits for oceanography?

JUSTIN BUCK (JUCK@BODC.AC.UK)

LOUISE DARROCH

THOMAS GARDNER

ALEXANDRA KOKKINAKI



Introduction

(why use data standards)



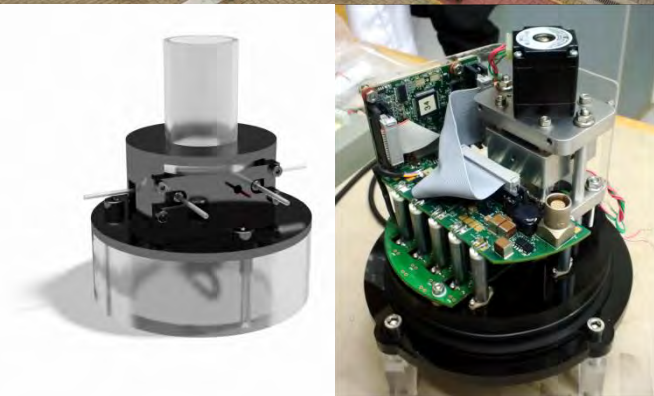
Evolving observation strategies

Autonomous ocean observation is massively increasing the number of sensors in the ocean.

Data practices need to evolve to ensure:

- Key metadata and technical data from novel sensors are never lost
- Efficient data processing
- Efficient data archival
- Seamless data delivery

Interoperability ... apply data standards from sensor through to delivery



Evolving data discovery and dissemination

Data Centres...



**British Oceanographic
Data Centre**

NATURAL ENVIRONMENT RESEARCH COUNCIL



**British
Geological Survey**

NATURAL ENVIRONMENT RESEARCH COUNCIL



Standard
data services

Aggregators...



Data standards

(what exists now)



Sensor Web Enablement (SWE)

<http://www.opengeospatial.org/>

OGC – Open Geospatial Consortium

SWE – Sensor Web Enablement initiative



OGC Defined prototyped and tested sensor web components:

- Sensor Model Language (SensorML)
- Observations & Measurements (O&M)
- Sensor Observation Service (SOS)
- Others (beyond the scope of this presentation)



Consistent naming conventions

Use example of
dissolved oxygen

What to call it?

- Oxygen
- Dissolved Oxygen
- O₂
- Disolved oxygen

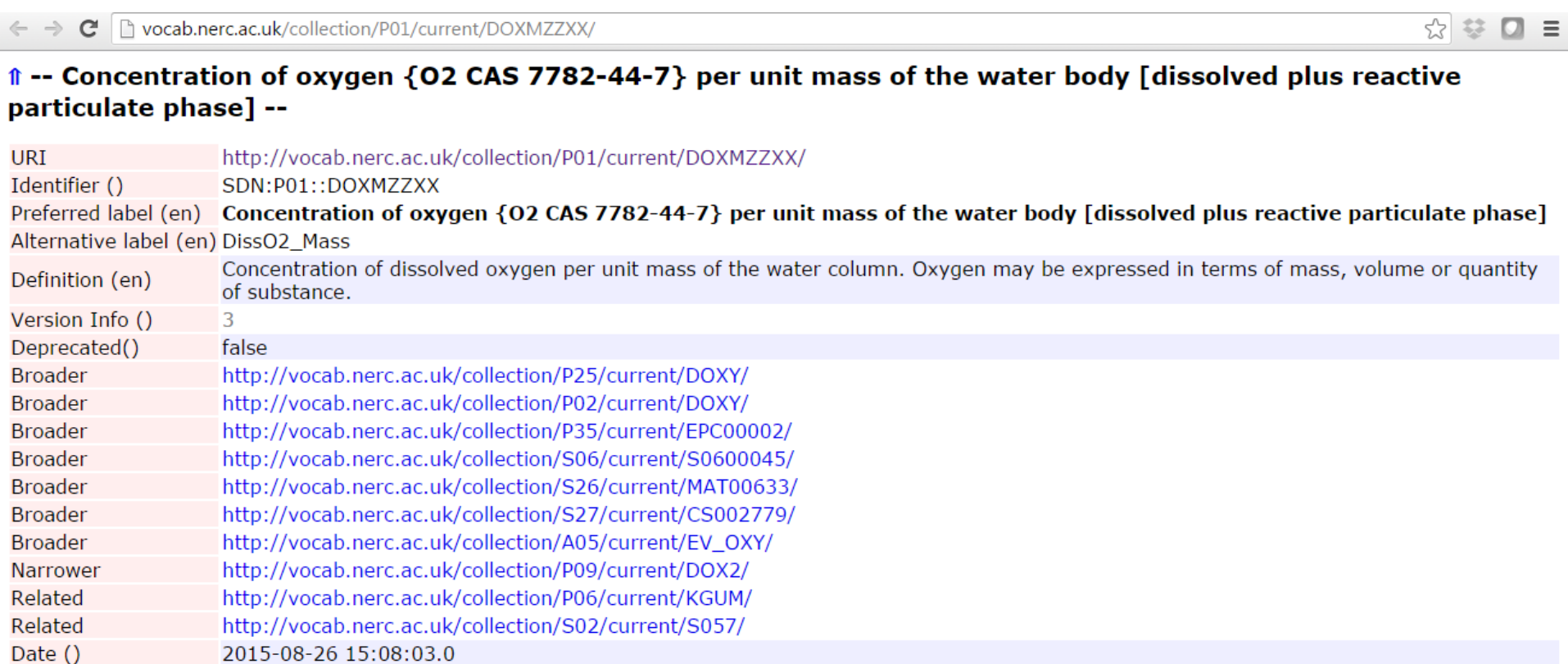
What are the units?

- micromole/kg
- micromoles per kg
- ml/l
- millilitres per litre

Consistent naming conventions

NERC vocabulary server resolves this issue:

https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/



The screenshot shows a web browser window with the address bar containing the URL `http://vocab.nerc.ac.uk/collection/P01/current/DOXMZZXX/`. The page content displays a detailed entry for a specific vocabulary term. The entry title is **Concentration of oxygen {O2 CAS 7782-44-7} per unit mass of the water body [dissolved plus reactive particulate phase]**. Below the title, a table lists various metadata fields and their values.

URI	http://vocab.nerc.ac.uk/collection/P01/current/DOXMZZXX/
Identifier ()	SDN:P01::DOXMZZXX
Preferred label (en)	Concentration of oxygen {O2 CAS 7782-44-7} per unit mass of the water body [dissolved plus reactive particulate phase]
Alternative label (en)	DissO2_Mass
Definition (en)	Concentration of dissolved oxygen per unit mass of the water column. Oxygen may be expressed in terms of mass, volume or quantity of substance.
Version Info ()	3
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/P25/current/DOXY/
Broader	http://vocab.nerc.ac.uk/collection/P02/current/DOXY/
Broader	http://vocab.nerc.ac.uk/collection/P35/current/EPC00002/
Broader	http://vocab.nerc.ac.uk/collection/S06/current/S0600045/
Broader	http://vocab.nerc.ac.uk/collection/S26/current/MAT00633/
Broader	http://vocab.nerc.ac.uk/collection/S27/current/CS002779/
Broader	http://vocab.nerc.ac.uk/collection/A05/current/EV_OXY/
Narrower	http://vocab.nerc.ac.uk/collection/P09/current/DOX2/
Related	http://vocab.nerc.ac.uk/collection/P06/current/KGUM/
Related	http://vocab.nerc.ac.uk/collection/S02/current/S057/
Date ()	2015-08-26 15:08:03.0

Linked data (<http://linkeddata.org/>)

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods.

More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of [data](#), [information](#), and [knowledge](#) on the Semantic Web using [URIs](#) and [RDF](#)."

```
<?xml version="1.0"?>
```

```
<rdf:RDF
```

```
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:si="http://www.w3schools.com/rdf/">
```

```
<rdf:Description rdf:about="http://www.w3schools.com">
```

```
  <si:title>W3Schools</si:title>
```

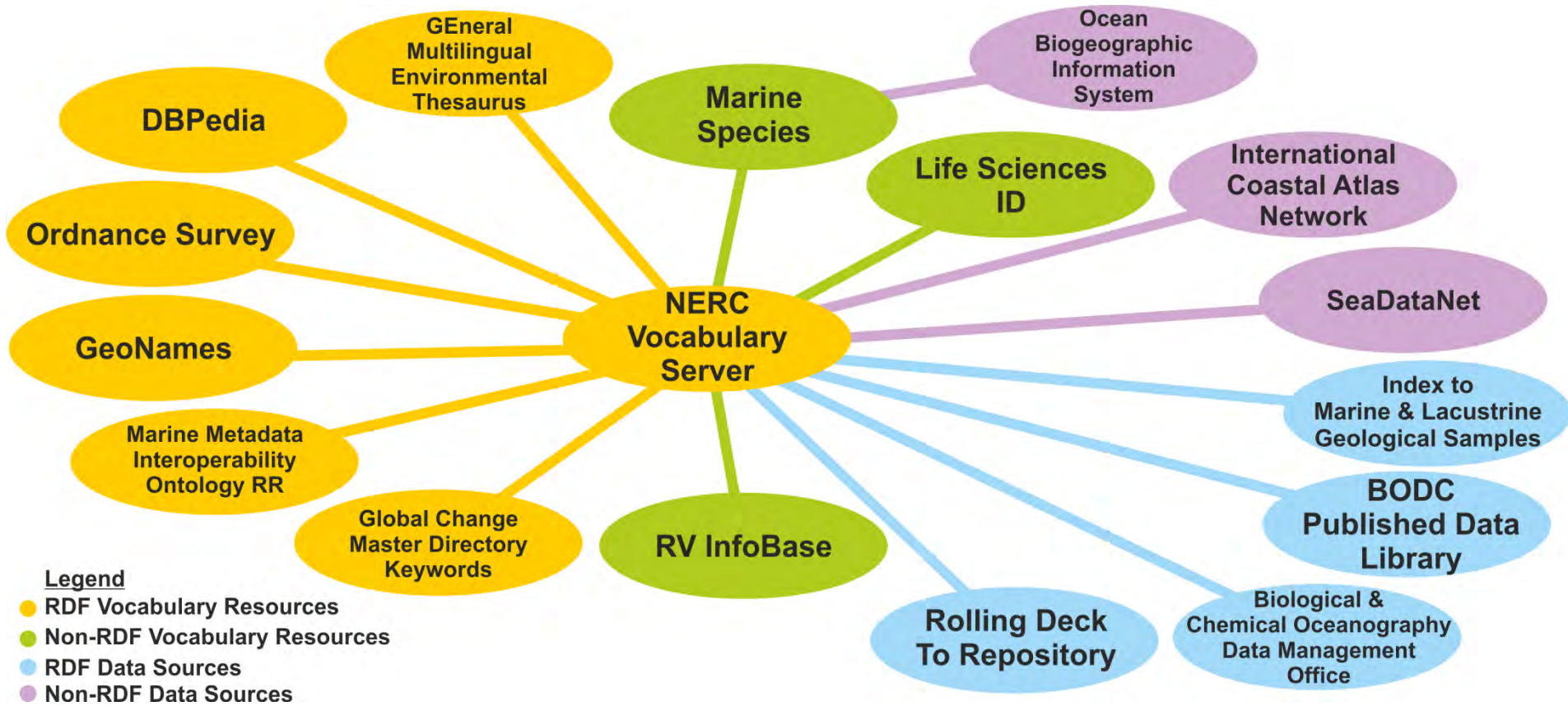
```
  <si:author>Jan Egil Refsnes</si:author>
```

```
</rdf:Description>
```

```
</rdf:RDF>
```



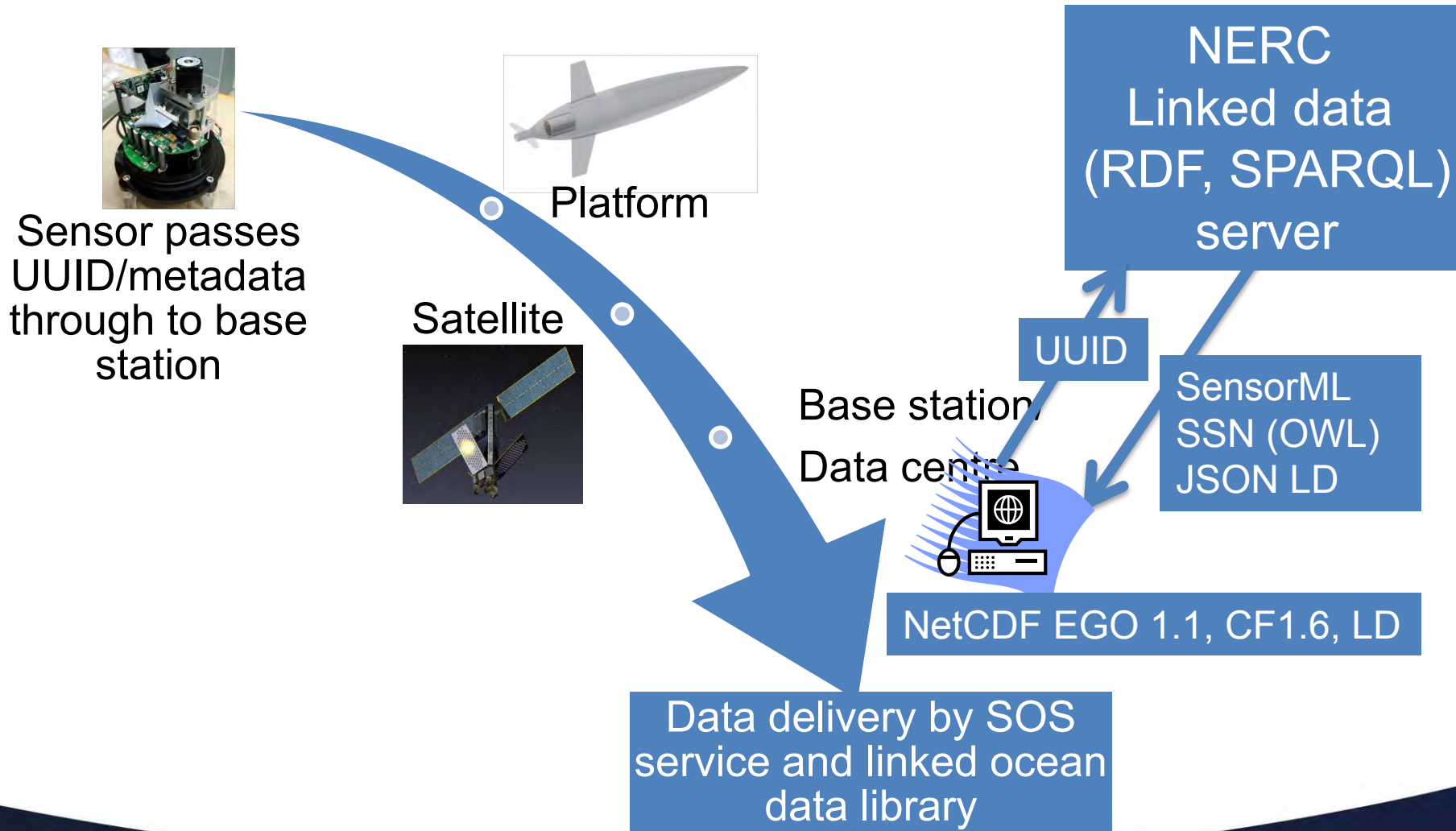
Linked Data



How standards will be applied & benefits (the on-going work)



Connecting data from sensor to delivery



Ontologies



ssn: What sensors measure, how they measure, and the qualities of such measurements

gr: Model, manufacturer, hasMakeAndModel, depth, height, weight

provo: What has occurred and how things were made what the entities are, what produced them and how

om-lite: An OWL representation of the Observation Schema O&M /OGC

SKOS: Supports the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists and taxonomies within the framework of the Semantic Web

time: temporal concepts, durations and datetime information

geo: geoSPARQL ontology for geospatial linked data

Marine SWE profile

(<http://meetingorganizer.copernicus.org/EGU2016/EGU2016-14690.pdf>)

```
<sm:classification>
```

```
<sm:ClassifierList>
```

```
<!-- Name of the manufacturer of the Sensor X -->
```

```
<sm:classifier name="InstrumentType">
```

```
<sm:Term definition="http://vocab.nerc.ac.uk/collection/W06/current/CLSS0002/">
```

```
<sm:label>Instrument Type</sm:label>
```

```
<sm:value>http://vocab.nerc.ac.uk/collection/L05/current/353/</sm:value>
```

```
</sm:Term>
```

```
</sm:identifier>
```

```
<sm:IdentifierList>
```

```
</sm:identification>
```

↑ -- Instrument type --

URI	http://vocab.nerc.ac.uk/collection/W06/current/CLSS0002/
Identifier ()	SDN:W06::CLSS0002

↑ -- active fluorometers --

URI	http://vocab.nerc.ac.uk/collection/L05/current/353/
Identifier ()	SDN:L05::353

ally computed data (numerical models) or

Preferred label (en)	active fluorometers
Alternative label ()	
Definition (en)	Fluorometers that measure photosynthetic parameters by taking measurements whilst manipulating the phytoplankton with controlled high-intensity illumination such as pump and probe and FRRF.
Version Info ()	2
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/L21/current/ICAT04/
Broader	http://vocab.nerc.ac.uk/collection/L19/current/SDNKG01/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0142/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0143/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0437/
Narrower	http://vocab.nerc.ac.uk/collection/W01/current/003/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0860/
Date ()	2013-02-26 17:11:07.0

Benefits for Oceanography

Standardised data delivered from platforms:

- Enabling the automation of routine data flows
- Remove ambiguity in data flows
- Enabling common tools and interfaces to be developed to access and process data

Exposure of data via common standards:

- Enables introduction of standardised APIs
- Enhance data discoverability and utility
- Opening up data beyond the marine domain for big data users
- Makes data machine comprehensible
- Enables machine to machine communication



Questions?



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Acknowledgments

This work is funded by the European projects SenseOCEAN and BRIDGES and supported by the National Environmental Research Council (NERC). SenseOCEAN is a Collaborative Project funded by the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement No. 61414. BRIDGES project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635359. NVS2.0 server is supported by NERC National Capability (NC) funding for NC-services, facilities and data (NC-SFD).

Examples

Sensors

TOOL0969, Aanderaa 4531 optode oxygen optode
: <http://linkedsystems.uk/system/prototype/TOOL0969/>

TOOLYSI, Chelsea Technologies Group 600R Multi-Parameter Water Quality Sonde: <http://linkedsystems.uk/system/prototype/TOOLYSI/>

Platform prototypes (with no metadata)

PL202 Unknown AUVs

<http://linkedsystems.uk/system/prototype/PL202/>

PL204 Unknown Argo Floats

<http://linkedsystems.uk/system/prototype/PL204/>

PL206 Unknown Seabed Observatory @

<http://linkedsystems.uk/system/prototype/PL206/>

PL208 Unknown Coastal Mooring @

<http://linkedsystems.uk/system/prototype/PL208/>

