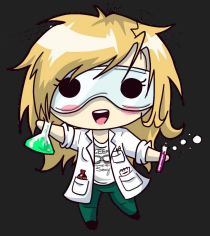
Applying Linked Data Principles to Static ISO Documents

Adam Shepherd, Technical Director
BCO-DMO

BCO-DMO
Biological & Chemical Oceanography Data Management Office

2019 ESIP Winter Meeting
January 16, 2019



BCO-DMO

Biological & Chemical Oceanography Data Management Office

ISO

NCEI

NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION

How do we enable static metadata records to evolve over time with new knowledge?

Example: Measuring "Photosynthetic Rate"

Dataset Name	Brief Description	Project	PI-Supplied Parameter Name
Host-symbiont reinfections	Symbiont removed and anemone reinfected with one of two strains of Symbiodinium	AnemoneOA	photo_output_symb
Imaging Pulse Amplitude Modulator Fluorometer Data	Imaging Pulse Amplitude Modulator Fluorometer Data	Resilient Acerv	mYield
Photosynthesic parameters - Rosario and Varadero	Photosynthesic parameters - Rosario and Varadero 2016 and 2017	Varadero Reef	alpha, Pmax_n, Pmax_g
Photosynthesis-Irradiance curve (P-E)	Photosynthesis-Irradiance curve	Varadero Reef	Slope_O2_prod, Rate_O2_prod
Photosynthetic and calcification rates	Photosynthetic and calcification rates of Pleurochrysis carterae	OA_Copes_Coccoliths	photo_rate, photoRate_LDcycle, photoRate_cellDensity_ugC, photoRate_cellDensity_pmolC
respiration and photosynthesis II	Biomass-normalized dark respiration and net photosynthesis rates in coral and algae as a function of pCO2 where LEDR (light-enhanced dark respiration) was measured	OA coral adaptation	photosynthesis
Thalassiosira pseudonana cyclostat	Thalassiosira pseudonana cyclostat experimental results	Stressors on Marine Phytoplankton	PI

- Definition added in May 2016
- Found in 7 Datasets
- 1st Dataset published May 2016
- Definition changed June 2016

https://www.bco-dmo.org/parameter/648663

Updating "Photosynthetic Rate"

Photosynthesis is a process by which plants and other organisms use light energy and chlorophyll to convert water, carbon dioxide, and minerals into oxygen and energy-rich organic compounds. Photosynthesis is one of the mechanisms responsible for Primary Production. The methods used to estimate rate of photosynthesis vary widely as do the units of measurement**, although estimates of photosynthesis are often reported as micromol CO2 m-2 s-1 or some variation of that**.

I took out this part: "although estimates of photosynthesis are often reported as micromol CO2 m-2 s-1 or some variation of that". CO2 is not a product of photosynthesis so when you measure photosynthesis you are not measuring CO2. You are measuring C or O2 usually. - mda, 6/13/2016

A compilation of dissolved noble gas and N2/Ar ratio measurements collected from 1999-2016 in locations spanning the globe

View/Open

- data_global-noble-gases.tsv (1.120Mb)
- ISO19115-2.xml (338.3Kb)
- Field_names.pdf (29.75Kb)
- Global_Hammeetal2019.mat (400.0Kb)
- Readme_Hammeetal2019.txt (20.55Kb)
- Dataset_description.pdf (197.0Kb)

Date
2018-08-24

Author
Hamme, Roberta C. ⊘
Jenkins, William J. ⊘

Citable URI
https://hdl.handle.net/1912/10538

Date Created
2018-08-13

Location
North Atlantic
Eastern Tropical Pacific - Transect from Peru to Tahiti
Oxygen minium zone; East Pacific Rise
Global oceans
westlimit: -159.9952; southlimit: -68.1081; eastlimit: 178.9985; northlimit: 78.9988

DOI
10.1575/1912/bco-dmo.744563

Inert gases dissolved in the ocean are powerful tracers of the impact of physical processes on gases, particularly air-sea gas exchange (by both diffusive and bubble-meditated processes), temperature change, atmospheric pressure variation, mixing between different water masses, and ice processes. We have compiled a global ocean database of dissolved neon, argon, and krypton measurements, supplemented by helium, xenon, and nitrogen/argon (N2/Ar) ratios in some locations. Samples were collected on board multiple research cruises spanning the period 1999 through 2016 and analyzed by mass spectrometry at four different shore-based laboratories (University of Victoria, Woods Hole Oceanographic Institution, University of Washington, and Scripps Institution of Oceanography). For a complete list of measurements, refer to the supplemental document 'Field_names.pdf', and a full dataset description is included in

A compilation of dissolved noble gas and N2/Ar ratio measurements collected from 1999-2016 in locations spanning the globe

View/Open

data_global-noble-gases.tsv (1.120Mb)
ISO19115-2.xml (338.3Kb)
Field_names.pdf (29.75Kb)
Global_Hammeetal2019.mat (400.0Kb)
Readme_Hammeetal2019.txt (20.55Kb)
Dataset_description.pdf (197.0Kb)

Date
2018-08-24

Author
Hamme, Roberta C. ⊘
Jenkins, William J. ⊘

Citable URI
https://hdl.handle.net/1912/10538

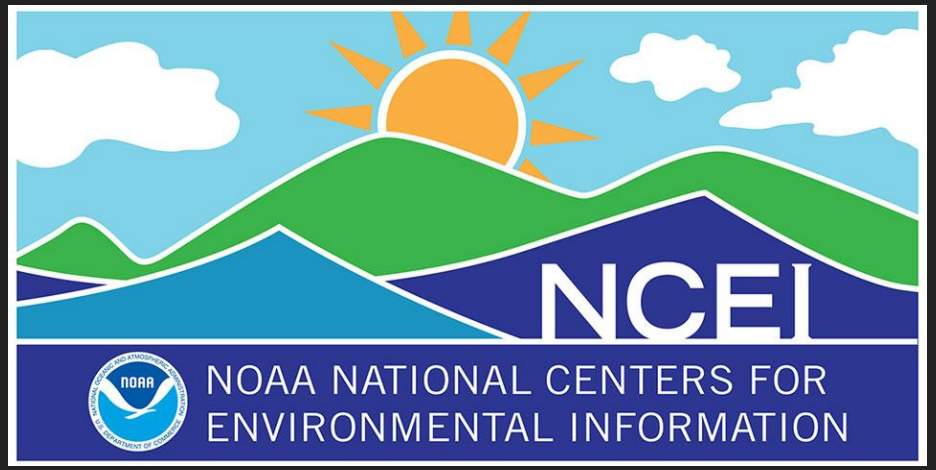
Date Created
2018-08-13

Location
North Atlantic
Eastern Tropical Pacific - Transect from Peru to Tahiti
Oxygen minium zone; East Pacific Rise
Global oceans
westlimit: -159.9952; southlimit: -68.1081; eastlimit: 178.9985; northlimit: 78.9988

DOI
10.1575/1912/bco-dmo.744563

Inert gases dissolved in the ocean are powerful tracers of the impact of physical processes on gases, particularly air-sea gas exchange (by both diffusive and bubble-meditated processes), temperature change, atmospheric pressure variation, mixing between different water masses, and ice processes. We have compiled a global ocean database of dissolved neon, argon, and krypton measurements, supplemented by helium, xenon, and nitrogen/argon (N2/Ar) ratios in some locations. Samples were collected on board multiple research cruises spanning the period 1999 through 2016 and analyzed by mass spectrometry at four different shore-based laboratories (University of Victoria, Woods Hole Oceanographic Institution, University of Washington, and Scripps Institution of Oceanography). For a complete list of measurements, refer to the supplemental document 'Field_names.pdf', and a full dataset description is included in

Static ISO record in the Archive w. old information



old definition of
"photosynthetic rate"

The gmx:Anchor tag

<gco:CharacterString>photosynthetic rate</gco:CharacterString>

becomes

<gmx:Anchor xlink:href="{URL goes here}">photosynthetic rate</gmx:Anchor>

Things instead of *Strings*

What is meant by a string

at a single point in time

may change in the future.

Photosynthetic Rate as a "Thing"

It has a label. "photosynthetic rate"

It has a definition. "Photosynthesis is a process..."

It can have synonyms. "photosynthesis"

It can be related to other things.

URI	http://vocab.nerc.ac.uk/collection/P21/current/MS8092/
Identifier ()	SDN:P21::MS8092
Preferred label (en)	**photosynthesis**
Alternative label ()	
Definition (en)	Photosynthesis is a process by which plants and other organisms use light energy and chlo compounds. Photosynthesis is one of the mechanisms responsible for Primary Production. measurement.
Version Info ()	3
Has Current Version	http://vocab.nerc.ac.uk/collection/P21/current/MS8092/3/
Has Version	http://vocab.nerc.ac.uk/collection/P21/current/MS8092/2/
Has Version	http://vocab.nerc.ac.uk/collection/P21/current/MS8092/1/
PAV Version ()	3
PAV Authored On ()	2016-06-20 12:31:50.0
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/P21/current/MS8687/
Narrower	http://vocab.nerc.ac.uk/collection/P21/current/MS3484/
Narrower	http://vocab.nerc.ac.uk/collection/P21/current/MS10803/
Narrower	http://vocab.nerc.ac.uk/collection/P21/current/MS4551/
Narrower	http://vocab.nerc.ac.uk/collection/P21/current/MS780/
Related	http://vocab.nerc.ac.uk/collection/P21/current/MS11730/
Date ()	2016-06-20 12:31:50.0

Photosynthetic Rate as a "Thing"

All important pieces of information

related to a dataset get an identifier.

http://lod.bco-dmo.org/id/parameter/648663

W3C Best Practices for Data on the Web

https://www.w3.org/TR/dwbp/#bp-summary

Best Practice 1: Provide metadata

Best Practice 2: Provide descriptive metadata

Best Practice 3: Provide structural metadata

Best Practice 4: Provide data license information

Best Practice 5: Provide data provenance information

Best Practice 6: Provide data quality information

Best Practice 7: Provide a version indicator

Best Practice 8: Provide version history

Best Practice 9: Use persistent URIs as identifiers of datasets

Best Practice 10: Use persistent URIs as identifiers within datasets

Best Practice 11: Assign URIs to dataset versions and series

Best Practice 12: Use machine-readable standardized data formats

Best Practice 13: Use locale-neutral data representations

Best Practice 14: Provide data in multiple formats

Best Practice 15: Reuse vocabularies, preferably standardized ones

Best Practice 16: Choose the right formalization level

Best Practice 17: Provide bulk download

Best Practice 18: Provide Subsets for Large Datasets

Best Practice 19: Use content negotiation for serving data available in multiple formats

Best Practice 20: Provide real-time access

Best Practice 21: Provide data up to date

Best Practice 22: Provide an explanation for data that is not available

Best Practice 23: Make data available through an API

Best Practice 24: Use Web Standards as the foundation of APIs

Best Practice 25: Provide complete documentation for your API

Best Practice 26: Avoid Breaking Changes to Your API

Best Practice 27: Preserve identifiers

Best Practice 28: Assess dataset coverage

Best Practice 29: Gather feedback from data consumers

Best Practice 30: Make feedback available

Best Practice 31: Enrich data by generating new data

Best Practice 32: Provide Complementary Presentations

Best Practice 33: Provide Feedback to the Original Publisher

Best Practice 34: Follow Licensing Terms

Best Practice 35: Cite the Original Publication

Best Practice #10:

Use persistent URIs as identifiers *within* datasets

https://www.w3.org/TR/dwbp/#identifiersWithinDatasets

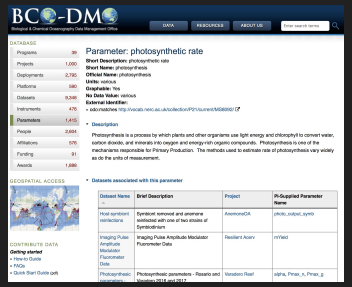
Use persistent URIs as identifiers within datasets

https://www.w3.org/TR/dwbp/#identifiersWithinDatasets

Use persistent URIs as identifiers of datasets	
Use persistent URIs as identifiers within datasets	
Assign URIs to dataset versions and series	

http://lod.bco-dmo.org/id/parameter/648663

In a browser https://www.bco-dmo.org/parameter/648663

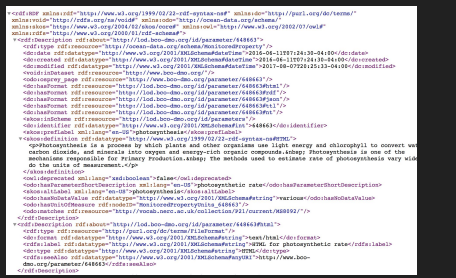


Supports Content Negotiation:

GET http://lod.bco-dmo.org/id/parameter/648663

Accept: text/html → HTML

Accept: application/xml → RDF/XML



Supports URL File Extensions:

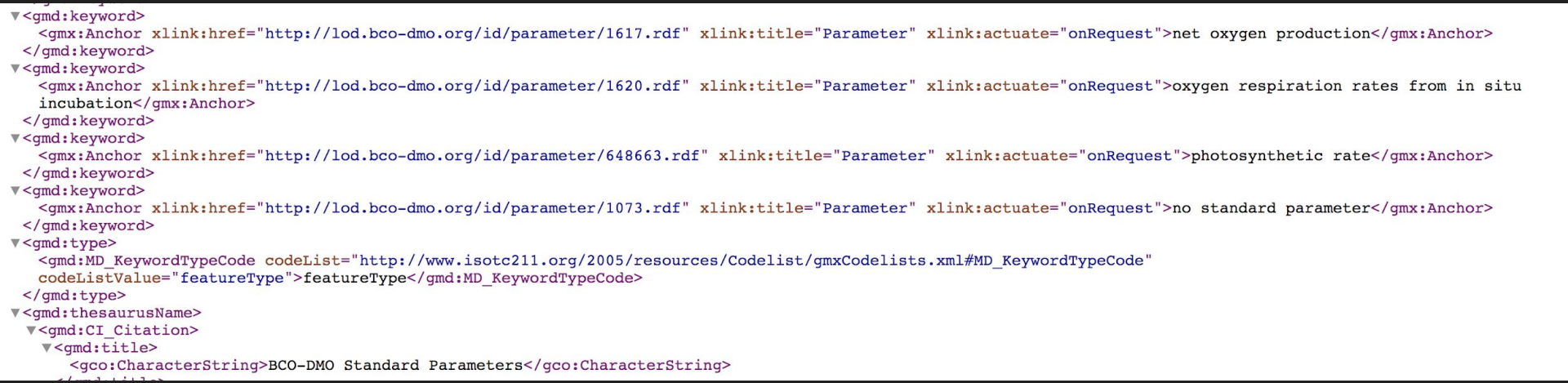
GET http://lod.bco-dmo.org/id/parameter/648663.xml → RDF/XML

GET http://lod.bco-dmo.org/id/parameter/648663.rdf → RDF/XML

The gmx:Anchor tag with URIs

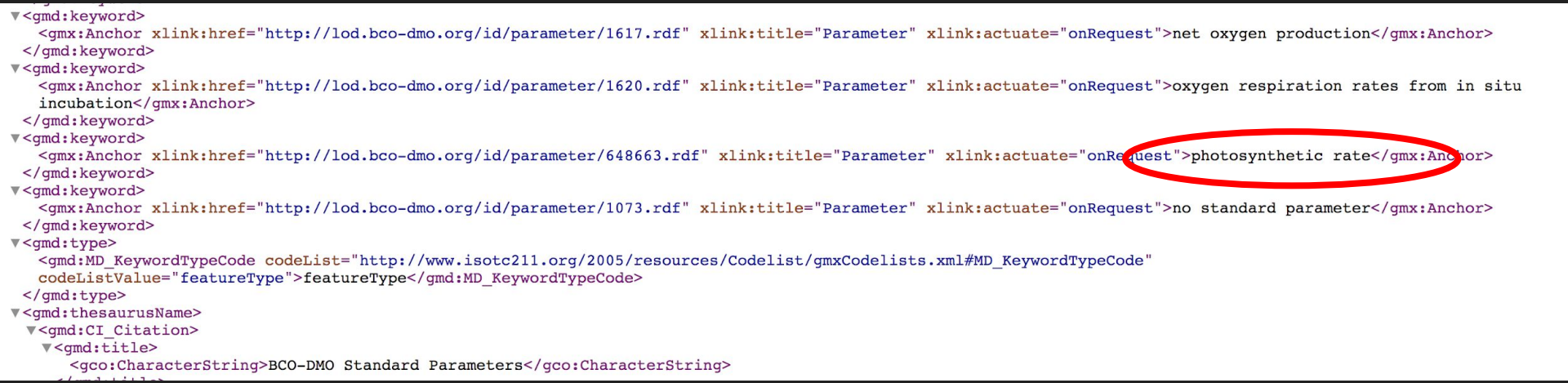
```
<gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/648663.rdf">...
```

NOAA ISO 19115-2 snapshot of Photosynthetic Rate



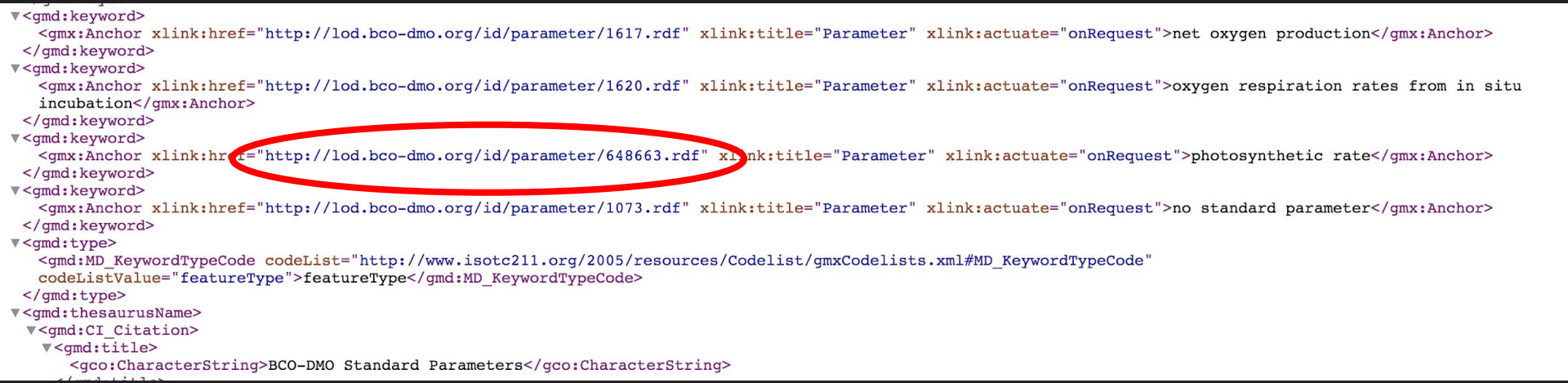
```
▼<gmd:keyword>
    <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1617.rdf" xlink:title="Parameter" xlink:actuate="onRequest">net oxygen production</gmx:Anchor>
  </gmd:keyword>
▼<gmd:keyword>
    <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1620.rdf" xlink:title="Parameter" xlink:actuate="onRequest">oxygen respiration rates from in situ
    incubation</gmx:Anchor>
  </gmd:keyword>
▼<gmd:keyword>
    <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/648663.rdf" xlink:title="Parameter" xlink:actuate="onRequest">photosynthetic rate</gmx:Anchor>
  </gmd:keyword>
▼<gmd:keyword>
    <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1073.rdf" xlink:title="Parameter" xlink:actuate="onRequest">no standard parameter</gmx:Anchor>
  </gmd:keyword>
▼<gmd:type>
    <gmd:MD_KeywordTypeCode codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#MD_KeywordTypeCode"
    codeListValue="featureType">featureType</gmd:MD_KeywordTypeCode>
  </gmd:type>
▼<gmd:thesaurusName>
  ▼<gmd:CI_Citation>
    ▼<gmd:title>
        <gco:CharacterString>BCO-DMO Standard Parameters</gco:CharacterString>
```

NOAA ISO 19115-2 snapshot of Photosynthetic Rate



```
▼<gmd:keyword>
    <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1617.rdf" xlink:title="Parameter" xlink:actuate="onRequest">net oxygen production</gmx:Anchor>
  </gmd:keyword>
▼<gmd:keyword>
    <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1620.rdf" xlink:title="Parameter" xlink:actuate="onRequest">oxygen respiration rates from in situ
    incubation</gmx:Anchor>
  </gmd:keyword>
▼<gmd:keyword>
    <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/648663.rdf" xlink:title="Parameter" xlink:actuate="onRequest">photosynthetic rate</gmx:Anchor>
  </gmd:keyword>
▼<gmd:keyword>
    <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1073.rdf" xlink:title="Parameter" xlink:actuate="onRequest">no standard parameter</gmx:Anchor>
  </gmd:keyword>
▼<gmd:type>
    <gmd:MD_KeywordTypeCode codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#MD_KeywordTypeCode"
    codeListValue="featureType">featureType</gmd:MD_KeywordTypeCode>
  </gmd:type>
▼<gmd:thesaurusName>
  ▼<gmd:CI_Citation>
    ▼<gmd:title>
        <gco:CharacterString>BCO-DMO Standard Parameters</gco:CharacterString>
```

NOAA ISO 19115-2 snapshot of Photosynthetic Rate



```
<gmd:keyword>
  <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1617.rdf" xlink:title="Parameter" xlink:actuate="onRequest">net oxygen production</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1620.rdf" xlink:title="Parameter" xlink:actuate="onRequest">oxygen respiration rates from in situ
  incubation</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/648663.rdf" xlink:title="Parameter" xlink:actuate="onRequest">photosynthetic rate</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://lod.bco-dmo.org/id/parameter/1073.rdf" xlink:title="Parameter" xlink:actuate="onRequest">no standard parameter</gmx:Anchor>
</gmd:keyword>
<gmd:type>
  <gmd:MD_KeywordTypeCode codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#MD_KeywordTypeCode"
  codeListValue="featureType">featureType</gmd:MD_KeywordTypeCode>
</gmd:type>
<gmd:thesaurusName>
  <gmd:CI_Citation>
    <gmd:title>
      <gco:CharacterString>BCO-DMO Standard Parameters</gco:CharacterString>
```

BCO-DMO Strategy for ISO 19115-2 XML

- Use <gmx:Anchor> instead of <gco:CharacterString>
  (Thing) (String)

- For humans, we use a file extension URL for readability (.rdf)

- For machines, gmx:Anchor requires xlink:href URL resolve to XML
  (Content-Type → application/rdf+xml)

http://lod.bco-dmo.org/id/parameter/648663.rdf

Questions?

Example ISO record at BCO-DMO:
https://www.bco-dmo.org/dataset/648416/iso/NOAA_ISO-19115-2.xml