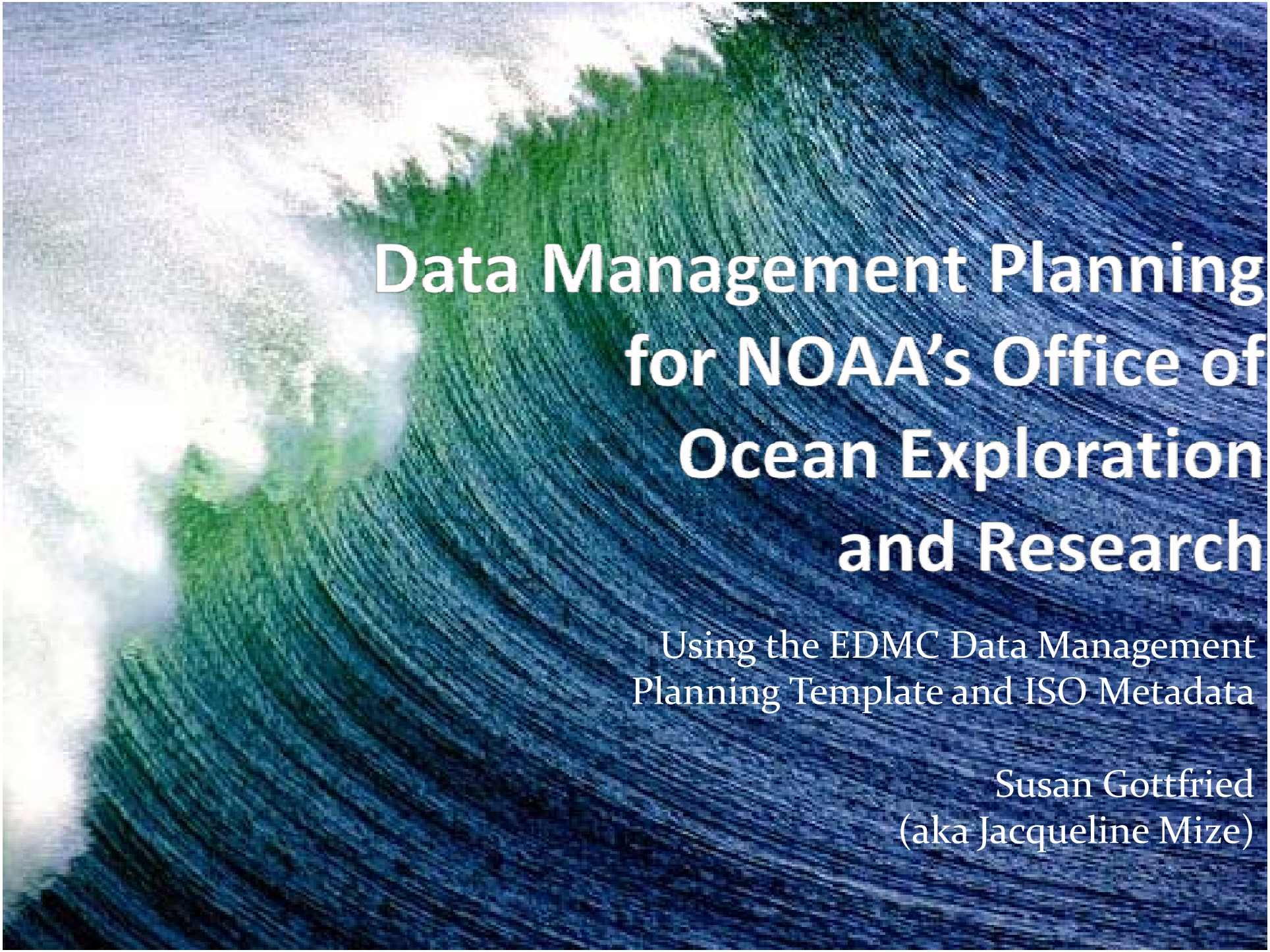
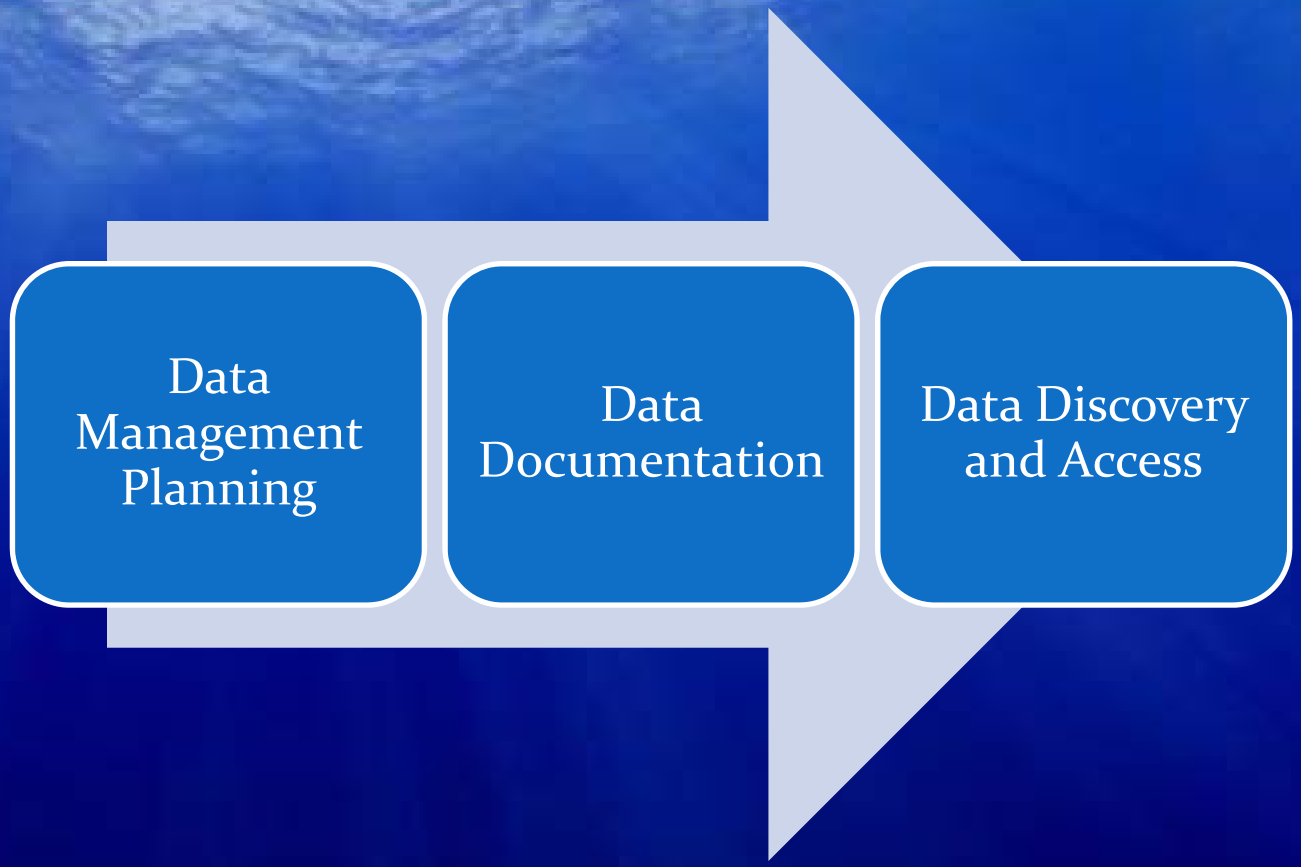


Data Management Planning for NOAA's Office of Ocean Exploration and Research

Using the EDMC Data Management Planning Template and ISO Metadata

Susan Gottfried
(aka Jacqueline Mize)

Data Management Workflow



OER Integrated Project Team for Data and Information Management

- The OER Integrated Project Team (IPT) for Data and Information Management has been evolving its systems and practices since 2003
- The team participation has evolved over time, but its core members are from OER and the NOAA Data Centers, including the NOAA Central Library.
- The team has only really managed data for missions considered "Signature" – or considered to have public interest for the website.

Data Management Planning

- Until 2008, the team has only met on the phone with Signature Expedition PIs to define plans and services
- In 2008, the team started writing customized data management plans and saw an improvement in compliance – from about 45% to about 65%.
- But the team still only is interfacing with a subset of funded PIs –everyone funded should be contacted and required to have a data management plan!

Data Management Planning 2.0

- In 2011, the team started writing ISO metadata, because of its relationships to other records – parents and children, its hierarchical structure, and its components.
- The most significant thing the team did beginning in 2011 is develop and deliver an ISO metadata record ***with the DMP to the PI*** and asked for feedback on all, but most importantly:
  - Abstract
  - Keywords
  - Data Use and Access Constraints (important for Marine Archaeology missions)

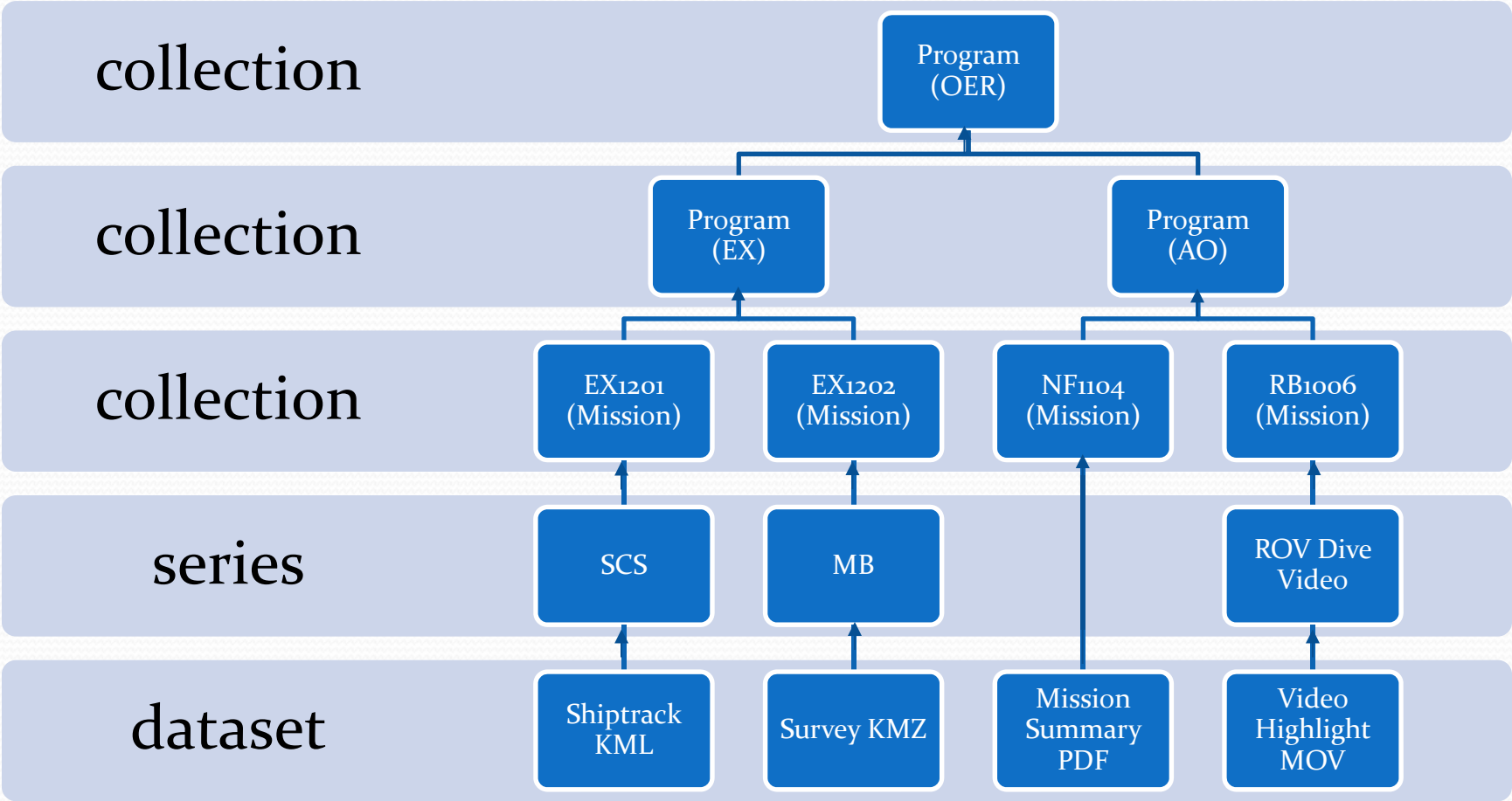
Components

- ISO allows the team to develop components (re-usable ISO segments) and so for the team has developed components for:
  - Organizations with contact information
  - Individuals with contact information and roles
  - Instruments with technical support and contact info
  - Platforms (ships, ROVs, Camera Platforms, etc) and their onboard instruments (calls to instrument components)
  - Keywords – standard keyword lists for example:
    - NOAA Programs (OER, OMAO, NODC, NGDC, NCL)
    - OER EX Missions
    - OER Marine Archaeology Missions

Templates

- Then the team built templates for metadata records for each Data Center that receives and hosts documented data and for different hierarchies of metadata:
  - Collection level (multiple data types collected)
  - Series level (specific data type collections)
  - Dataset level (data file or data product)
- Templates ensure that the metadata hierarchy - parent-child relationship - is ensured and maintained. Records always point up to their parent's Universally Unique ID (UUID).

ISO Metadata Hierarchy



The ISO Migration and the EDMC Template

- In 2012, the team started using the EDMC template – an easy transition and everyone is pleased.
- The team also recognized that the questions in the template could be directly related to an ISO record, and the template authors recognized that too, because the first few questions have the ISO field indicated.

EDMC DMP Template in Practice

Data Management Plan Template

[Extracted from Appendix A of the NOAA EDMC Data Management Planning Procedural Directive (2011): https://geo-ide.noaa.gov/wiki/index.php?title=Data_Management_Planning_PD]

This Template is provided as an example to assist DM Plan developers. It may be downloaded from the DMP Repository. The Repository also includes other plans as examples.

Note: Some of this information may be used to populate fields in a formal metadata record about the data collection. These are indicated for illustrative purposes only with [→ field name].

1. **General Description of Data to be Managed**

   1.1. Name of the Dataset or data collection project [→ gmd:title].

   1.2. Keywords that could be used to characterize the data, and vocabulary from which those keywords were obtained (e.g., GCMD, CF Conventions, etc.) [→ gmd:MD_Keywords]

   1.3. Summary description of the data to be generated [→ gmd:abstract].

   1.4. Anticipated temporal coverage of the data [→ gmd:EX_TemporalExtent].

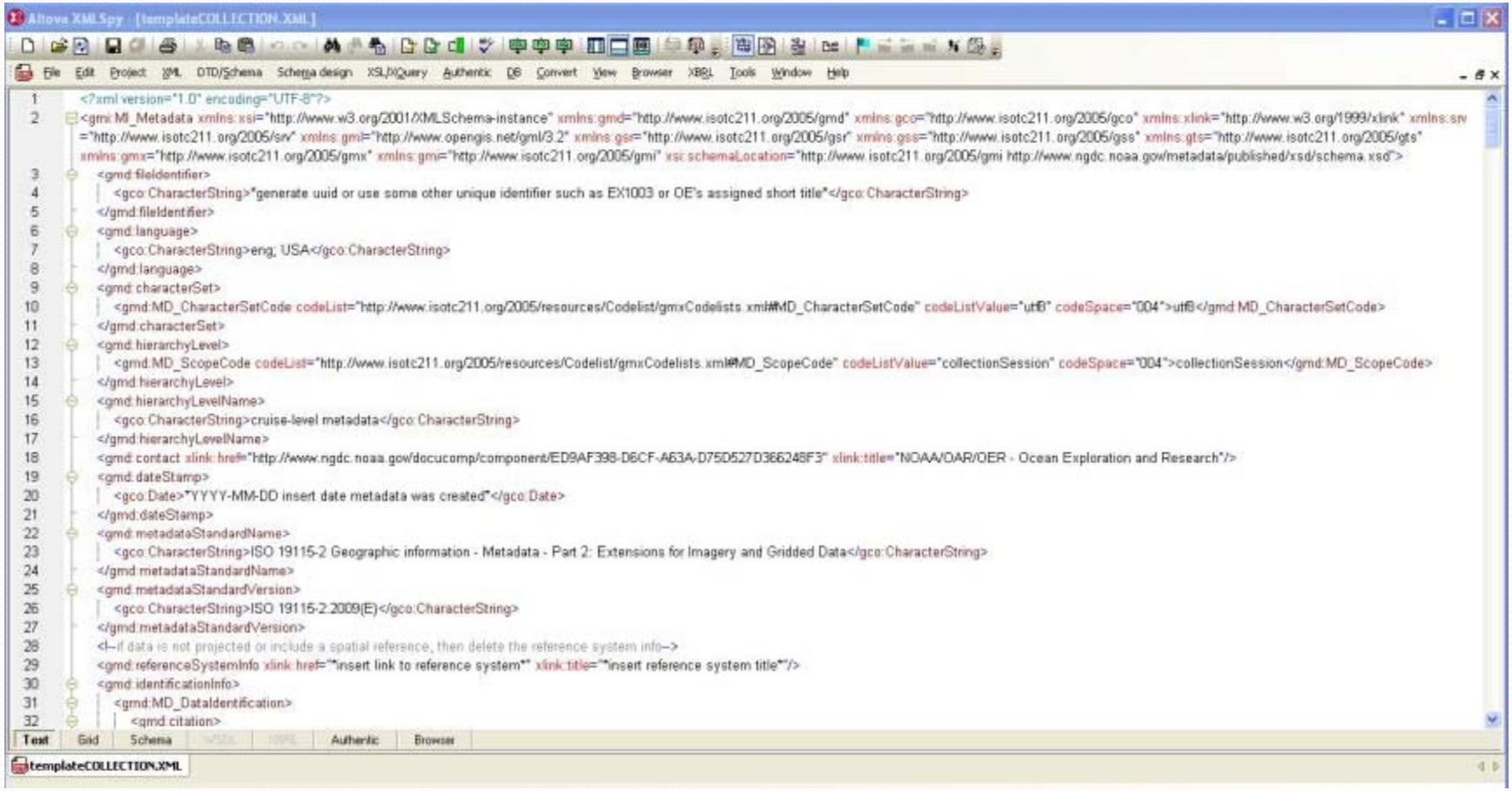
   1.5. Anticipated geographic coverage of the data [→ gmd:EX_Extent]

   1.6. What data types will you be creating or capturing? (e.g., digital numeric data, photographs, video, acoustic records, database tables, spreadsheets, paper records, physical samples, etc.)

Putting it into Practice

- So, the team combed through the questions and made notes about the other questions and where they belonged in the ISO record. What the team found was:
  - Some questions don't have a corresponding field.
  - Some questions for mandatory fields need to be asked.
  - Some blocks have multiple questions that should be separated so that they correspond to specific fields.
  - The questions should be organized according to where they fit within ISO.

Data Management Plan transformed into ISO 19115-2



Putting it into Practice

- To make it as convenient as possible, the team should build a user interface tool that asks these questions and then generates a data management planning report and a corresponding ISO record.
- Also, the team can provide templates already linked to its parents and, if the distribution centers have predictable or can provide the distribution link, that can be included.

An Example Metadata Entry Tool

1. General Desscription of Data to be Managed

1.1 Name of the Dataset or data collection project

insert title including who, what, when, where such as proposal title with dates and location

1.2 Keywords that could be used to characterize the data, and vocabulary from which those keywords were obtaind (e.g., GCMD, CF Conventions, etc.)

Keywords:	*gcmd keywords*		
Keywords:	*repeatable gcmd keyword*	Vocabulary used:	GCMD
Keywords:	*insert keyword*		
Keywords:	*insert keyword*	Vocabulary used:	Ocean Exploration and Research (OER) Discovery Ke
Keywords:	*ship name*	Vocabulary used:	None
Keywords:	*insert theme keyword or phrase*		
Keywords:	*insert theme keyword or phrase*		
Keywords:	*example:* coastal ecology		
Keywords:	*example:* geoarchaeology	Vocabulary used:	None
Keywords:	*insert place keyword*		
Keywords:	*insert place keyword*	Vocabulary used:	None

1.3 Summary description of the data to be generated

paraphrase the executive summary

1.4 Anticipated temporal coverage of the data

Start date/time (YYYY-MM-DDThh:mm:ss)	*YYYY-MM-DD start date*
End date/time (YYYY-MM-DDThh:mm:ss)	*YYYY-MM-DD end date*

1.5 Anticipated geographic coverage of the data

West Bounding Longitude	*template in decimal degrees*
East Bounding Longitude	*template in decimal degrees*
North Bounding Latitude	*template in decimal degrees*
South Bounding Latitude	*template in decimal degrees*

add gmd:EX_GeographicBoundingBox add gmd:EX_GeographicBoundingBox

1.6 What data types will you be creating or capturing?

digital video

Altova XMLSpy - [templateCOLLECTION.XML *]

File Edit Project XML DTD/Schema Schema design XSL/XQuery Authentic DB Convert View Browser XBRL Tools Window Help

1. General Desscription of Data to be Managed

1.1 Name of the Dataset or data collection project

insert title including who, what, when, where such as proposal title with dates and location

1.2 Keywords that could be used to characterize the data, and vocabulary from which those keywords were obtaind (e.g., GCMD, CF Conventions, etc.)

Keywords:	*gcmd keywords*		
Keywords:	*repeatable gcmd keyword*	Vocabulary used:	GCMD
Keywords:	*insert keyword*		
Keywords:	*insert keyword*	Vocabulary used:	Ocean Exploration and Research (OER) Discovery Ke
Keywords:	*ship name*	Vocabulary used:	None
Keywords:	*insert theme keyword or phrase*		
Keywords:	*insert theme keyword or phrase*		
Keywords:	*example:* coastal ecology		
Keywords:	*example:* geoarchaeology	Vocabulary used:	None
Keywords:	*insert place keyword*		
Keywords:	*insert place keyword*	Vocabulary used:	None

1.3 Summary description of the data to be generated

Text Grid Schema WSDL XBRL Authentic Browser

templateCOLLECTION.XML *

Altova XMLSpy - [templateCOLLECTION.XML *]

File Edit Project XML DTD/Schema Schema design XSL/XQuery Authentic DB Convert View Browser XBRL Tools Window Help

1. General Desscription of Data to be Managed

1.1 Name of the Dataset or data collection project

EDIT HERE

1.2 Keywords that could be used to characterize the data, and vocabulary from which those keywords were obtaind (e.g., GCMD, CF Conventions, etc.)

Keywords:	*gcmd keywords*		
Keywords:	*repeatable gcmd keyword*	Vocabulary used:	GCMD
Keywords:	*insert keyword*		
Keywords:	*insert keyword*	Vocabulary used:	Ocean Exploration and Research (OER) Discovery Ke
Keywords:	*ship name*	Vocabulary used:	None
Keywords:	*insert theme keyword or phrase*		
Keywords:	*insert theme keyword or phrase*		
Keywords:	*example:* coastal ecology		
Keywords:	*example:* geoarchaeology	Vocabulary used:	None
Keywords:	*insert place keyword*		
Keywords:	*insert place keyword*	Vocabulary used:	None

1.3 Summary description of the data to be generated

Text Grid Schema WSDL XBRL Authentic Browser

templateCOLLECTION.XML *

Altova XMLSpy - [templateCOLLECTION.XML *]

File Edit Project XML DTD/Schema Schema design XSL/XQuery Authentic DB Convert View Browser XBRL Tools Window Help

```
21      <gco:Date>*YYYY-MM-DD insert date metadata was created*</gco:Date>
22    </gmd:dateStamp>
23    <gmd:metadataStandardName>
24      <gco:CharacterString>ISO 19115-2 Geographic information - Metadata - Part 2: Extensions for Imagery and Gridded Data</gco:CharacterString>
25    </gmd:metadataStandardName>
26    <gmd:metadataStandardVersion>
27      <gco:CharacterString>ISO 19115-2:2009(E)</gco:CharacterString>
28    </gmd:metadataStandardVersion>
29    <!--if data is not projected or include a spatial reference, then delete the reference system info-->
30    <gmd:referenceSystemInfo xlink:href="*insert link to reference system*" xlink:title="*insert reference system title*"/>
31    <gmd:identificationInfo>
32      <gmd:MD_DataIdentification>
33        <gmd:citation>
34          <gmd:CI_Citation>
35            <gmd:title>
36              <gco:CharacterString>EDIT HERE</gco:CharacterString>
37            </gmd:title>
38            <gmd:alternateTitle>
39              <gco:CharacterString>*insert short title such as EX1003 or OE's cruise id*</gco:CharacterString>
40            </gmd:alternateTitle>
41            <gmd:date>
42              <gmd:CI_Date>
43                <gmd:date gco:nilReason="unknown"/>
44                <gmd:dateType>
45                  <gmd:CI_DateTypeCode codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#CI_DateTypeCode" codeListValue="publication" codeSpace="002">publication</gmd:CI_DateTypeCode>
46                </gmd:dateType>
47              </gmd:CI_Date>
48            </gmd:date>
49            <gmd:identifier>
50              <gmd:MD_Identifier>
51                <gmd:code>
52                  <gco:CharacterString>*same as alternative title*</gco:CharacterString>
53                </gmd:code>
54              </gmd:MD_Identifier>
```

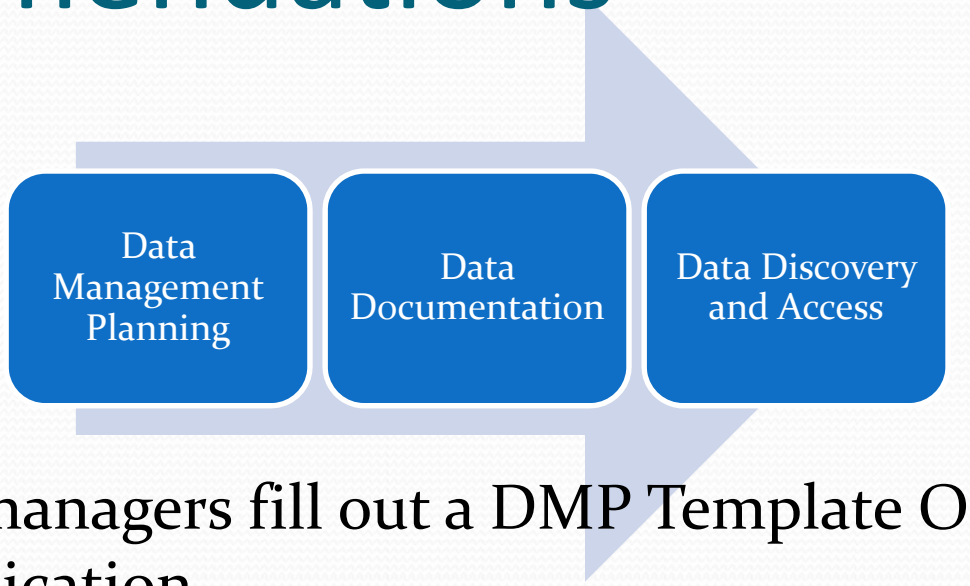
Text Grid Schema WSDL XBRL Authentic Browser

templateCOLLECTION.XML *

Recommendations

- Recognize that Data Management Plans can be generated for different tiers of metadata as described in the hierarchy slide.
- Slightly modify the EDMC DMP Template to be transformable into ISO metadata and to ask all mandatory field questions.
- Consider if templates by tier is an option or even necessary.
- Develop a template form (online?) to fill out a DMP for whichever tier corresponds to the need

Recommendations



- Have data managers fill out a DMP Template Online form or GUI application
- Export a final Data Management Plan
- Output an ISO record and publish in an available Web Accessible Folder (WAF)
- Export templates as needed for the next tier in the hierarchy that automatically reference the parent UUID.