



Big Earth Metadata Initiative: Catalog Services for the Web Analytics Scenario

Ted Habermann, John Kozimor, Sean Gordon

Earth Science Group

The HDF Group

thabermann@hdfgroup.org

John Farley

Raytheon

Concept – Abstract property or aspect of an object.

Location – World Wide Web Resources have unique locations that are expressed as Universal resource locators (URLs).

Documenting a World Wide Web Resource requires a location expressed as a URL:

http://en.wikipedia.org/wiki/Uniform_resource_locator

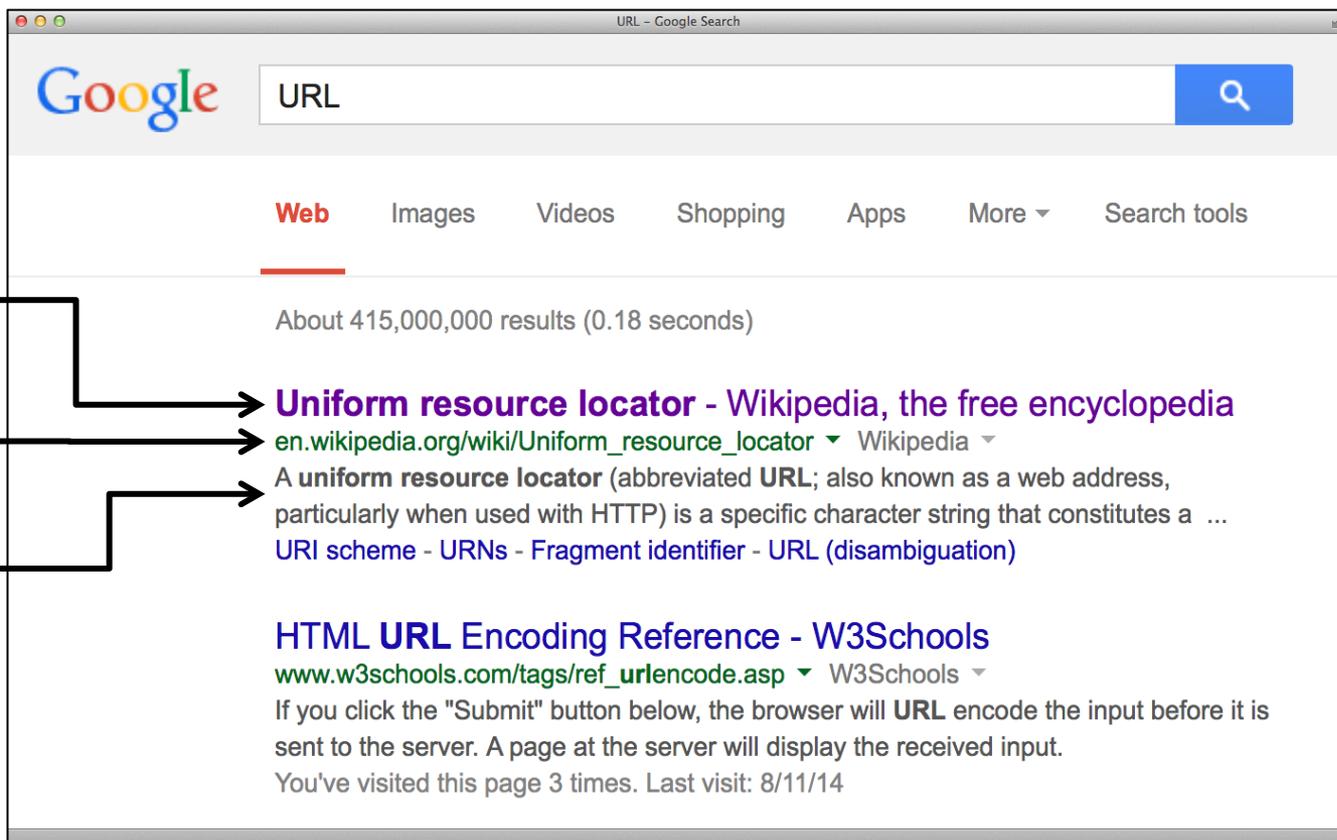
Spiral – A collection of concepts required to support a particular documentation need or use case.

Name

Location

Description

Type



The screenshot shows a Google search for "URL". The search bar contains "URL" and the search button is visible. Below the search bar, the "Web" tab is selected. The search results show "About 415,000,000 results (0.18 seconds)". The first result is "Uniform resource locator - Wikipedia, the free encyclopedia" with the URL "en.wikipedia.org/wiki/Uniform_resource_locator". The second result is "HTML URL Encoding Reference - W3Schools" with the URL "www.w3schools.com/tags/ref_urlencode.asp".

Annotations on the left side of the screenshot:

- Name**: Points to the search term "URL" in the search bar.
- Location**: Points to the URL "en.wikipedia.org/wiki/Uniform_resource_locator" in the first search result.
- Description**: Points to the text "A uniform resource locator (abbreviated URL; also known as a web address, particularly when used with HTTP) is a specific character string that constitutes a ... URI scheme - URNs - Fragment identifier - URL (disambiguation)" in the first search result.
- Type**: Points to the text "HTML URL Encoding Reference - W3Schools" in the second search result.

Recommendation – A set of concepts that a group/organization believes are required for achieving a document

The Directory Interchange Format (DIF) is metadata sets. A DIF holds a collection of fields, which DIF, while others expand upon and clarify the controlled keywords (sometimes known as database, where validation for mandatory fields is normalized searches for users. The DIF is contained here can be used to guide the author through the terms in orange indicate **required fields**. Terms displayed in green.

Field:

- Entry ID:
- Entry Title:
- Parameters (Science Keywords):
- ISO Topic Category:
- Data Center:
- Summary:
- Metadata Name:
- Metadata Version:
- Data Set Citation:
- Personnel:
- Related URL:
- Instrument (Sensor Name):
- Platform (Source Name):
- Temporal Coverage:
- Paleo-Temporal Coverage:

DIF

Attribute Convention for Data Discovery 1-1 – Federation of Earth Science Information Partners

Version and Status

This version is designated as Version 1.1.

This page always has the current version of the Attribute Convention for Data Discovery (ACDD). As it is updated, the version number at the top of the page will change.

Development

Any development version of the ACDD definitions is maintained at [Attribute_Convention_for_Data_Discovery_1ACDD_Working](#).

The netCDF Group at Unidata has recommended netCDF attributes for data discovery. Many of these attributes correspond to general metadata conventions: THREDDS and add the ones from ISO 19115-2. Note that the attribute names use the Unidata definitions. See Table by an M, O, or C in parentheses. An 'M' indicates that the element is mandatory. An 'O' indicates that the element is optional. A 'C' indicates that the element is optional. A 'C' indicates that the element is optional. A 'C' indicates that the element is optional.

Metadata Link

The netCDF metadata model is focused on providing "use metadata" for the data included in the file (or granule). Other metadata datasets make users aware of that additional metadata we recommend adding a global attribute named "Metadata_Link" to the netCDF file. The netCDF file is not included in the current version of the NetCDF Attribute Convention for Dataset Discovery.

Highly Recommended

Attribute	Description	THREDDS	ISO 19115-2
file@	A short description of the dataset.	dataset@name	gms:MI_Metadata@id=identification@MD_DataIdentification
summary@	A paragraph describing the dataset.	metadata@documentation@type="summary"	gms:MI_Metadata@id=identification@MD_DataIdentification
keywords@	A comma separated list of any words and phrases.	metadata@keyword	gms:MI_Metadata@id=identification@MD_DataIdentification

Recommended

Attribute	Description	THREDDS	ISO 19115-2
id@	The combination of the "hanging authority" and the "id" should be a globally unique identifier for the dataset.	dataset@id	gms:MI_Metadata@id=identification@MD_DataIdentification
keywords_vocabulary@	If you are following a guideline for the words/phrases in your "keywords" attribute, put the	metadata@keyword@vocabulary	gms:MI_Metadata@id=identification@MD_DataIdentification

ISO 19115-2
Annex G (normative)
Discovery metadata for geographic resources

G.1 Metadata for resource discovery

The metadata listed below is intended to be used to manage product and/or service metadata catalogues; the metadata to be exposed to facilitate discovery. Not all metadata elements are mandatory, obligations follow those defined in this international standard.

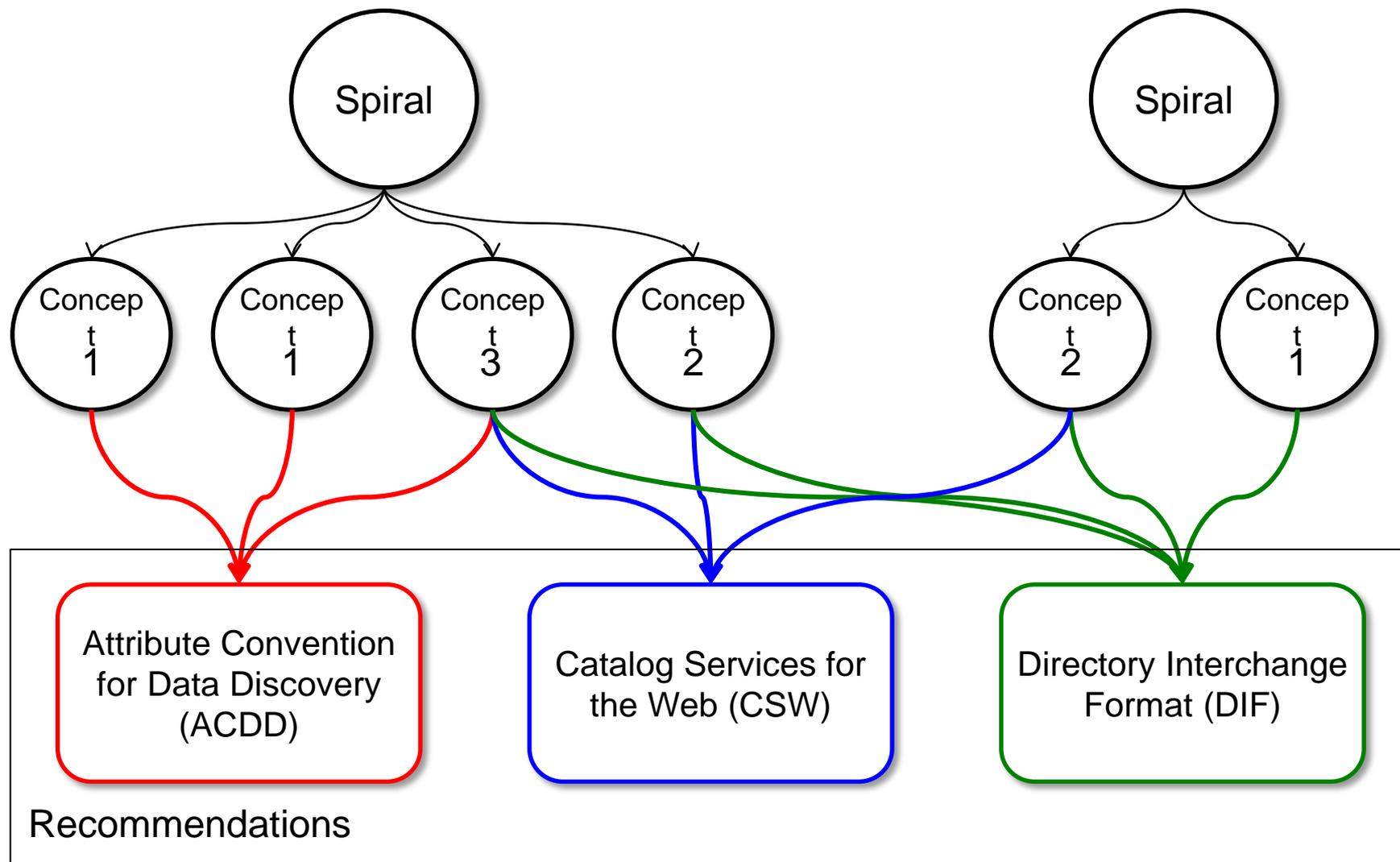
G.2 Metadata for the discovery of geographic datasets and series

Table G.1 — Metadata for the discovery of geographic datasets and series

Metadata Element	Obligation/max occurrence	Comment
Metadata reference information:	OR	Unique identifier for the metadata
MD_Metadata.metadataIdentifier		
Resource title: (MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.title)	MT	Title by which the resource is known
Resource reference date: (MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.date)	OR	A date which is used to help identify the resource
Resource identifier: (MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.identifier > MD_Identifier)	ON	Unique identifier for the resource
Resource point of contact: (MD_Metadata.identificationInfo > MD_DataIdentification.pointOfContact > CI_Responsibility)	ON	Name of the person, position, or organization responsible for the resource
Geographic location: (MD_Metadata.identificationInfo > MD_DataIdentification.extent > EX_Extent.geographicElement > EX_GeographicExtent > EX_GeographicBoundingBox > EX_GeographicDescription)	C – not equal to dataset (not mandatory)N	Geographic description or coordinates (with/without/plus) which describes the location of the resource
Resource language: (MD_Metadata.identificationInfo > MD_DataIdentification.language)	C – mandatory if a language is used in the resource (if a language is used)	The language used in the resource (if a language is used)
Resource topic category: (MD_Metadata.identificationInfo > MD_DataIdentification.topicCategory > TopicCategoryCode)	C – not equal to dataset or services (not mandatory)N	A selection of the 20 elements in the MD_TopicCategory enumeration which describe the topic of the resource.
Spatial resolution: (MD_Metadata.identificationInfo > MD_Resolution.spatialResolution > MD_Resolution.equivalentScale or MD_Resolution.distance)	ON	The nominal scale and/or spatial resolution of the resource



How Things Fit Together



Subject: CSW Support

To: Data / Metadata System Manager

From: Above

I recently learned that it is critical that we support something called CSW for our metadata catalog. I think it is related to the OGC (not sure what that is).

We need to get this done ASAP! Presentation tomorrow!

Thanks,
The Management

The screenshot shows the ESIP Federation website. A callout box labeled 'CONCEPTS' points to the 'Concept' column in a table. The table has two columns: 'Concept' and 'Description'. The 'Description' column contains the text: 'A word or phrase that describes a... aspect of a resource. Can be one... several types. Note: The general identification keywords usually have a type of... and are referred to as "theme key... Other types and vocabularies are... for other information. Service Ent... Resource Format (SERF) require... Science and a Service GCMD Ke... This concept is called "Subject" in... CSW Specification.'

Below the table, a summary box lists the following recommendations:

- Initially:
 - Data Discovery – 8 recommendations
 - Data Accessibility/Usability (Services) – 3 recommendations
 - Data Understanding – 6 recommendations

http://wiki.esipfed.org/index.php/Category:Documentation_Recommendations

HDF Recommendation Comparisons

Recommendation comparison tool helps identify concepts that are included in multiple recommendations: **high priority concepts.**

Recommendation Tool

52.10.182.225:443/recTool/?recommendations=CSWCoreQueryable,CSWCoreReturnables,CSWAdditionalQueryables

Select Recommendations:

ACDD	CSW	ECHO	DIF	FGDC
<input type="checkbox"/> Highly Recommended	<input checked="" type="checkbox"/> Core Queryables	<input type="checkbox"/> Discovery Mandatory	<input type="checkbox"/> Required	<input type="checkbox"/> Mandatory
<input type="checkbox"/> Recommended	<input checked="" type="checkbox"/> Core Returnables	<input type="checkbox"/> Discovery Recommended	<input type="checkbox"/> Recommended	<input type="checkbox"/> Recommended If Applicable
<input type="checkbox"/> Suggested	<input checked="" type="checkbox"/> Additional Queryable Properties			

Submit

Search:

Showing 1 to 25 of 25 entries

Item	Score	Catalog Services for the Web (CSW) Core Queryables	Catalog Services for the Web (CSW) Core Returnable Properties	Catalog Service for the Web (CSW) Additional Queryable Properties
Abstract	2	X	X	
Bounding Box	2	X	X	
Metadata Identifier	2	X	X	
Metadata Language	2		X	X
Modified Date	2	X	X	
Resource	2	X	X	
Resource	2	X	X	
Resource	2	X	X	
Association	1	X		
Author	1		X	
Contributor	1		X	
Coordinate Reference System (CRS)	1	X		
Keyword	1	X		
Keyword	1			X
Organization	1			X
Parent	1			X
Publication	1			X
Publisher	1			X
Publication Date	1		X	
Related Information	1		X	
Resource	1			X
Resource	1			X
Revision	1			X
Rights	1		X	
Security Constraints	1			X
Source Citation	1		X	

CONCEPTS

Translation Version: 0.1 (Feb. 9, 2015)



What does it mean for me?

The impact of a recommendation on an organization depends on the metadata dialects that are used in the organization.

Recommendation Tool

52.10.182.225:443/drTool/?recommendations=CSWCoreQueryable,CSWCoreReturnables,CSWAdditionalQueryable&dialects=ISO,DIF,ECHO

CF RADIAL FORMATS Recommendation Tool National Collaborative Resea... Recommendation Tool Recommendation Tool 500 Internal Server Error

Select Recommendations:

ACDD	CSW	DIF	FGDC
<input type="checkbox"/> Highly Recommended	<input checked="" type="checkbox"/> Core Queryables	<input type="checkbox"/> Required	<input type="checkbox"/> Mandatory
<input type="checkbox"/> Recommended	<input checked="" type="checkbox"/> Core Returnables	<input type="checkbox"/> Recommended	<input type="checkbox"/> Recommended if Applicable
<input type="checkbox"/> Suggested	<input checked="" type="checkbox"/> Additional Queryable Properties		

Select Dialects:

<input checked="" type="checkbox"/> ISO 19115 and ISO 19115-2	<input checked="" type="checkbox"/> Directory Interchange Format	<input type="checkbox"/> Metadata Interchange Standard (ECS)	<input type="checkbox"/> NASA Unified Metadata Model
<input type="checkbox"/> ISO 19115-1 / ISO 19115-3	<input checked="" type="checkbox"/> ECHO	<input type="checkbox"/> Metadata Interchange Standard	<input type="checkbox"/> Service Entry Resource Format

Submit

Showing 1 to 25 of 25 entries

Item	Score	ISO	DIF	ECHO
Abstract	3	X	X	X
Author / Originator	3	X	X	X
Bounding Box	3	X	X	X
Keyword	3	X	X	X
Modified Date	3	X	X	X
Organization Name	3	X	X	X
Publication Date	3	X	X	X
Resource Creation/Revision Date	3	X	X	X
Resource Format	3	X	X	X
Resource Identifier	3	X	X	X
Resource Title	3	X	X	X
Association	2			X
Coordinate Reference System (CRS)	2			X
Metadata Identifier	2		X	
Parent Identifier	2	X	X	
Publisher	2	X	X	
Security Constraints	2		X	
Contributor Name	1			
Keyword Type	1	X		
Metadata Language	1	X		
Related Resource Citation	1	X		
Resource Type	1	X		
Revision Date	1	X		
Rights	1	X		
Source	1	X		

CONCEPTS

Recommendation Max

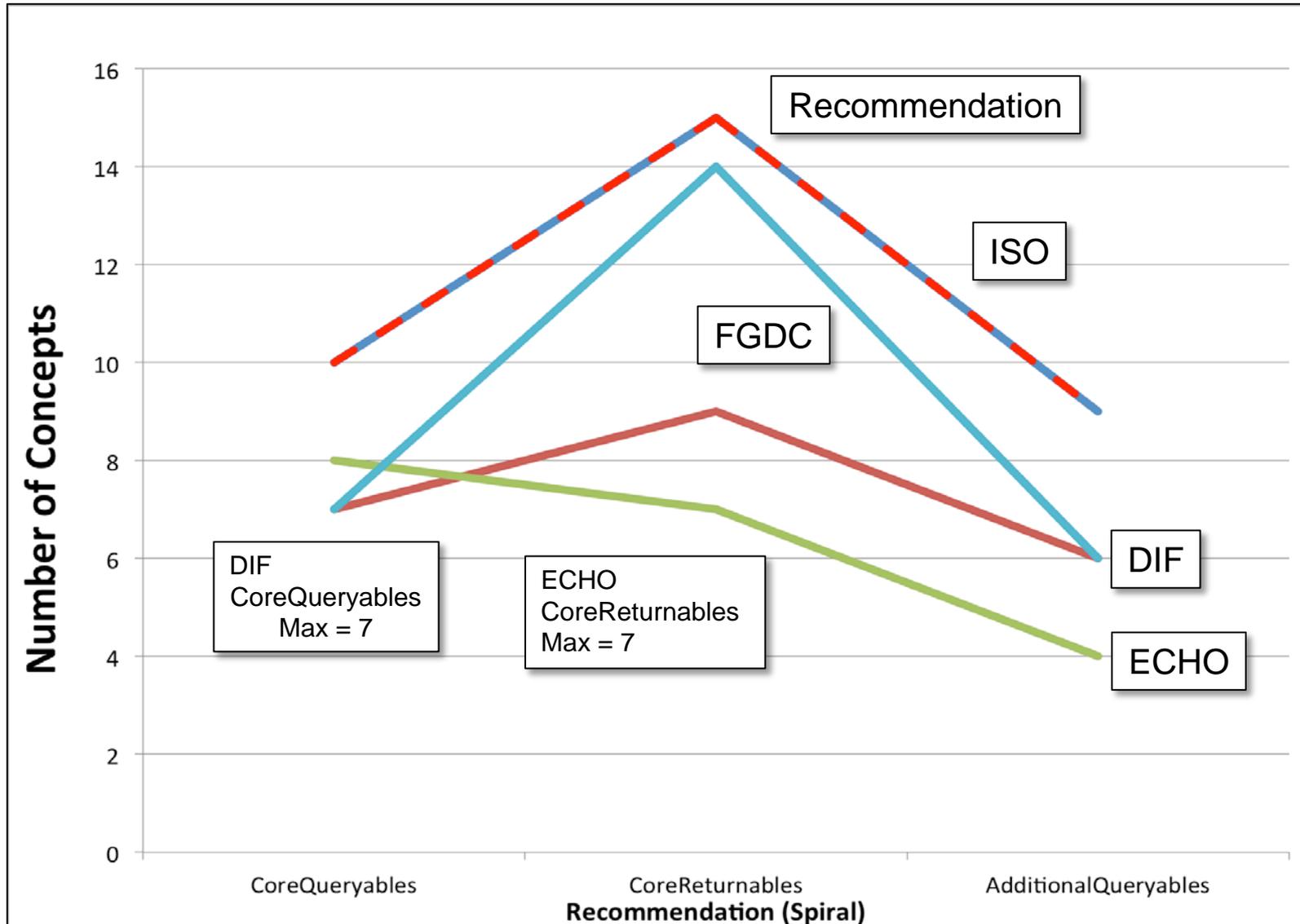
Dialect Max

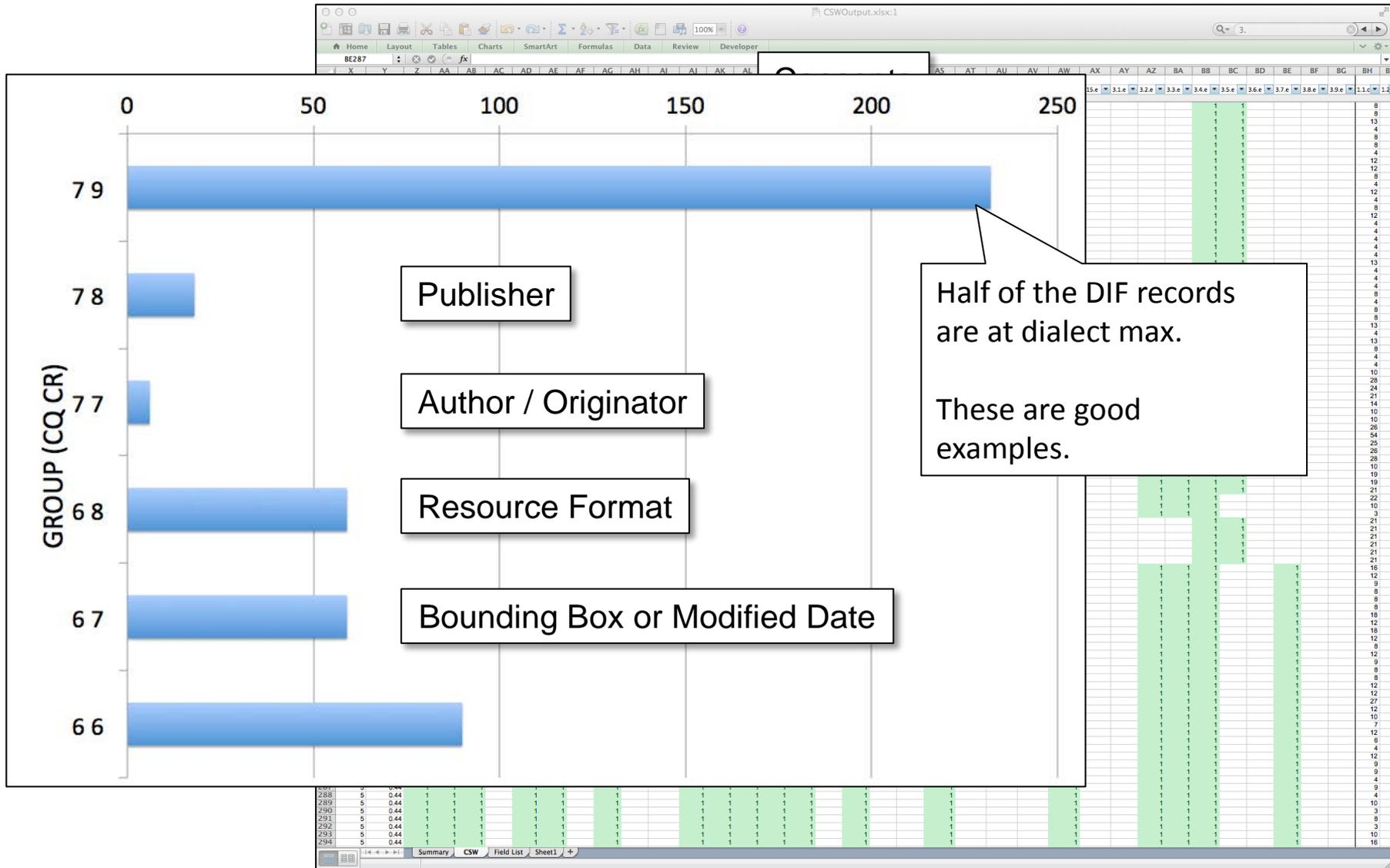
Dialect Max

Dialect Max



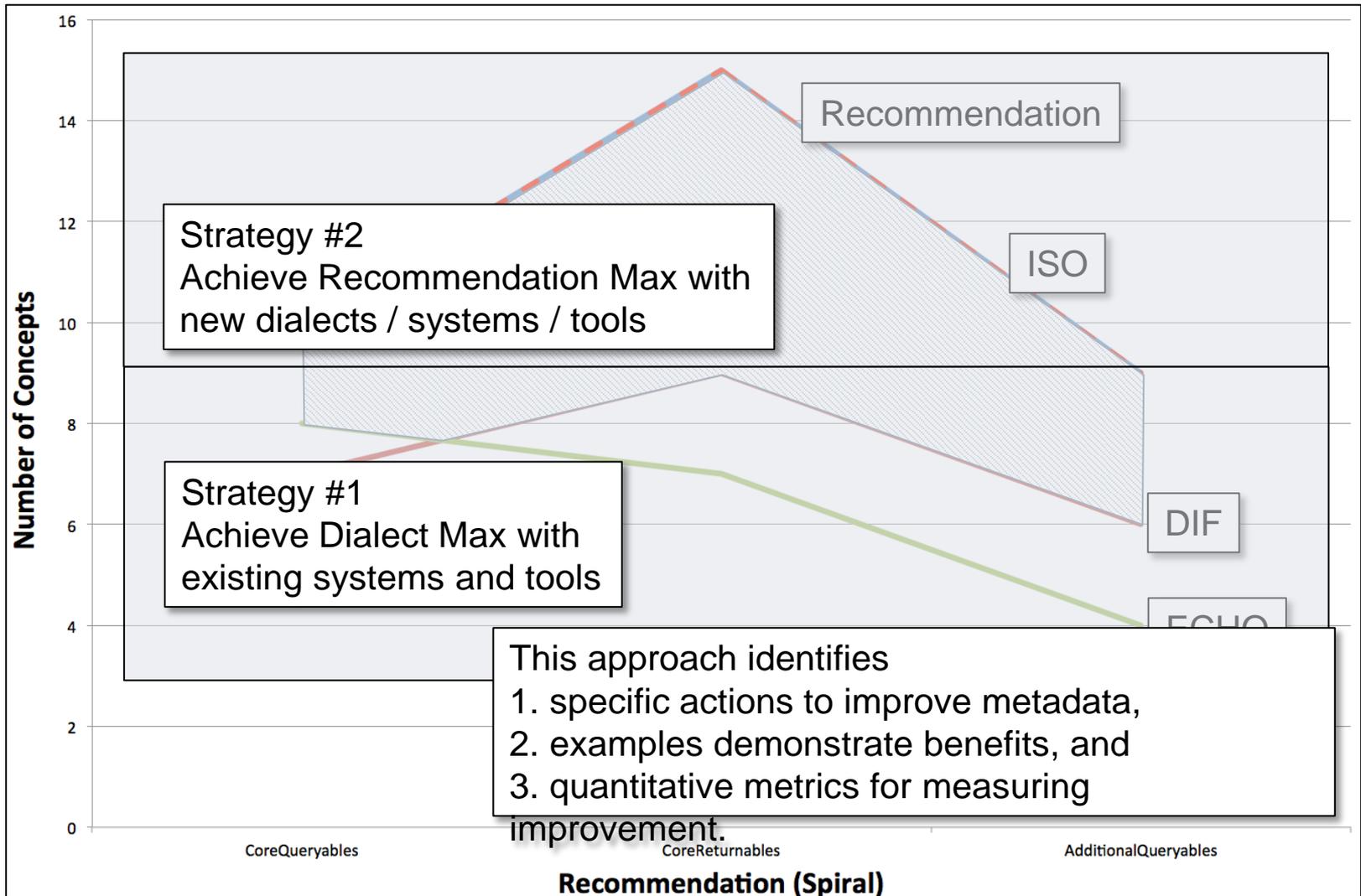
Recommendation and Dialect Maximums





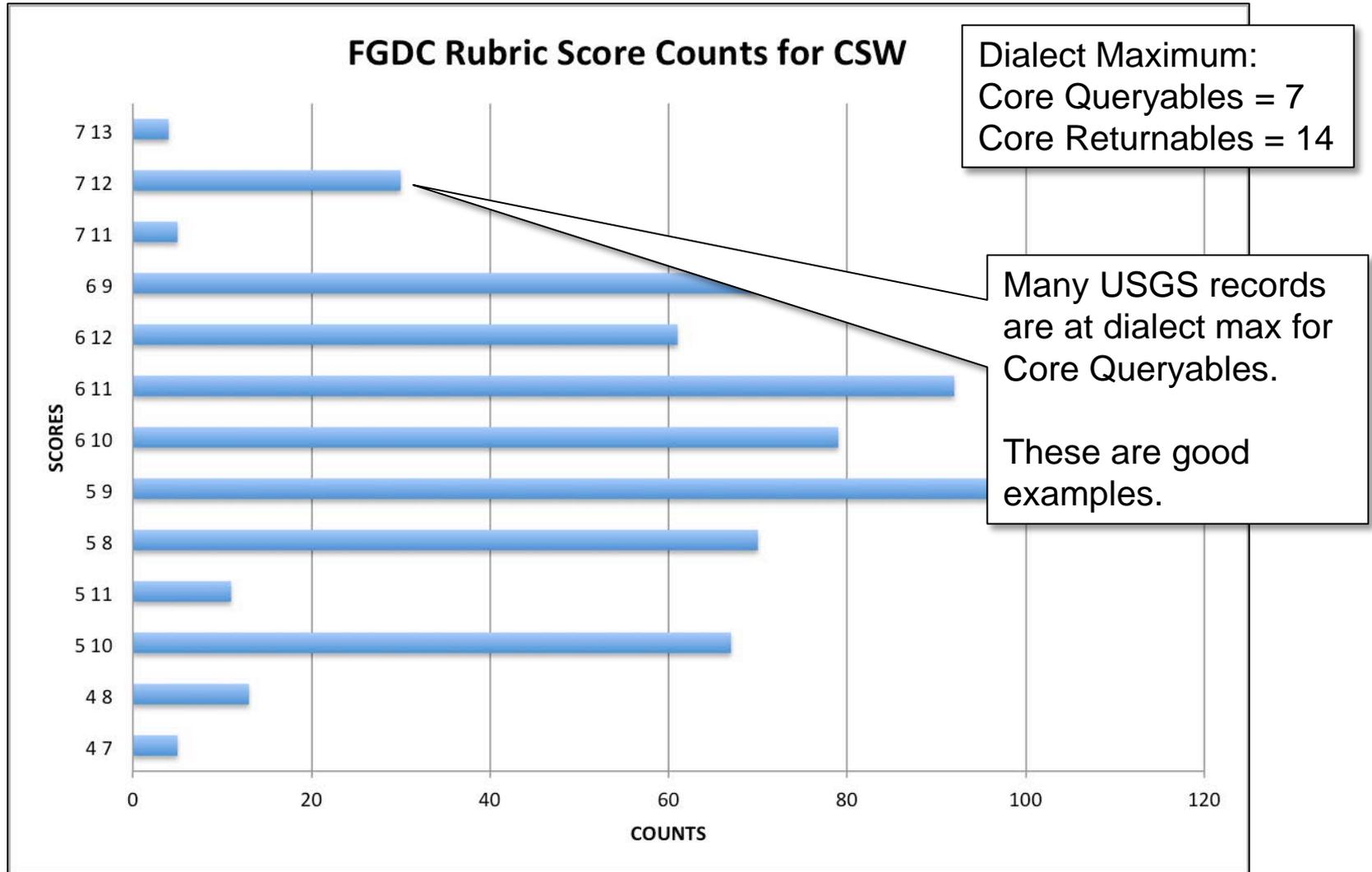


Two Improvement Strategies





Preliminary USGS Analysis (FGDC Dialect)



Wiki Updates

- Improved guidance (Elli)

- Updated technical content (John)

- Collection analytics results with connections to guidance

DOI – Two new recommendations

- ESDSWG

- NOAA

UMM – * Analytics

USGEO / Other Partners

ESIP Session