EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

International Metadata Standards and Enterprise Data Quality Metadata Systems

Ted Habermann

Director of Earth Science

The HDF Group

thabermann@hdfgroup.org

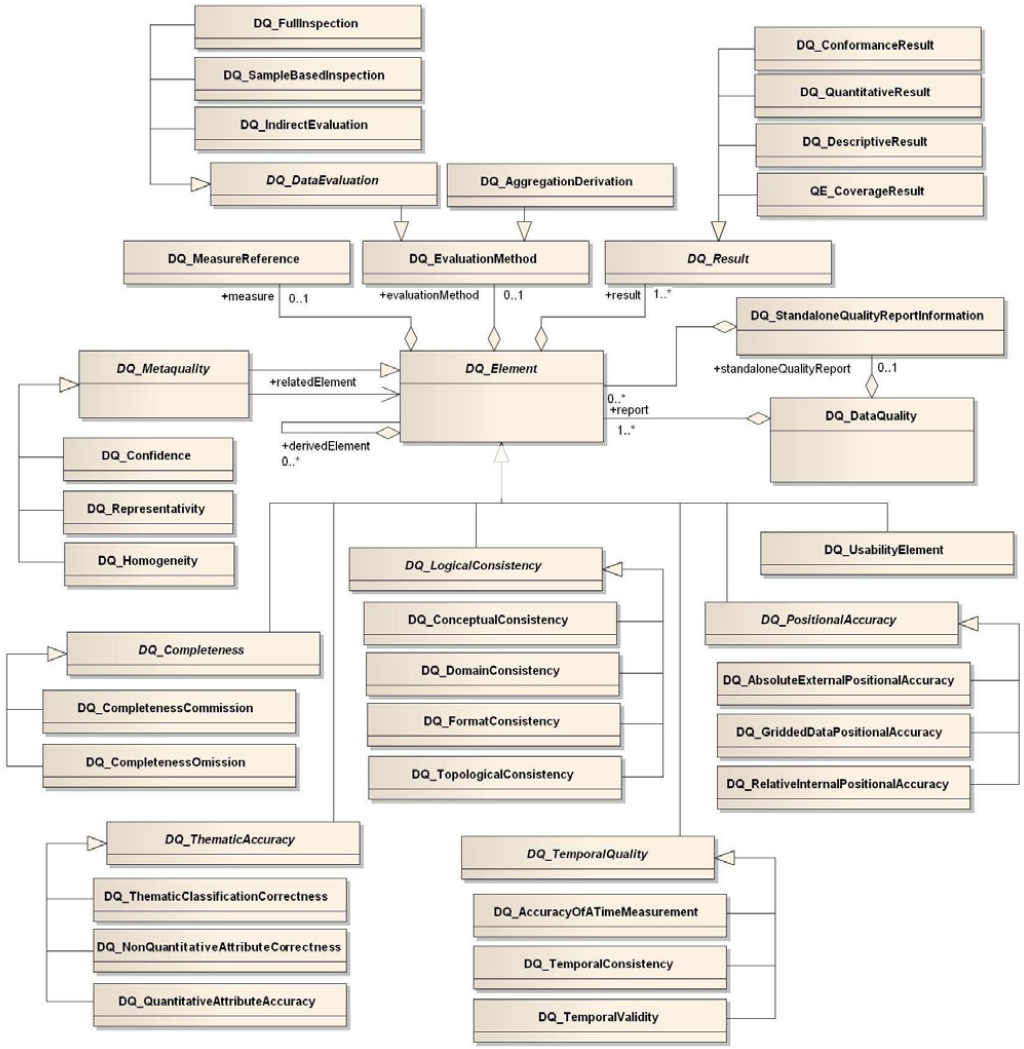
The HDF Group

This work was supported by NASA/GSFC under
Raytheon Co. contract number NNG15HZ39C

The Big Picture

ISO 19157 is a conceptual model of data quality metadata that was recently approved as an international standard. It combines three older standards into a unified model for describing data quality.

Many of the principle elements of this conceptual model are abstract, and can be implemented in several ways.



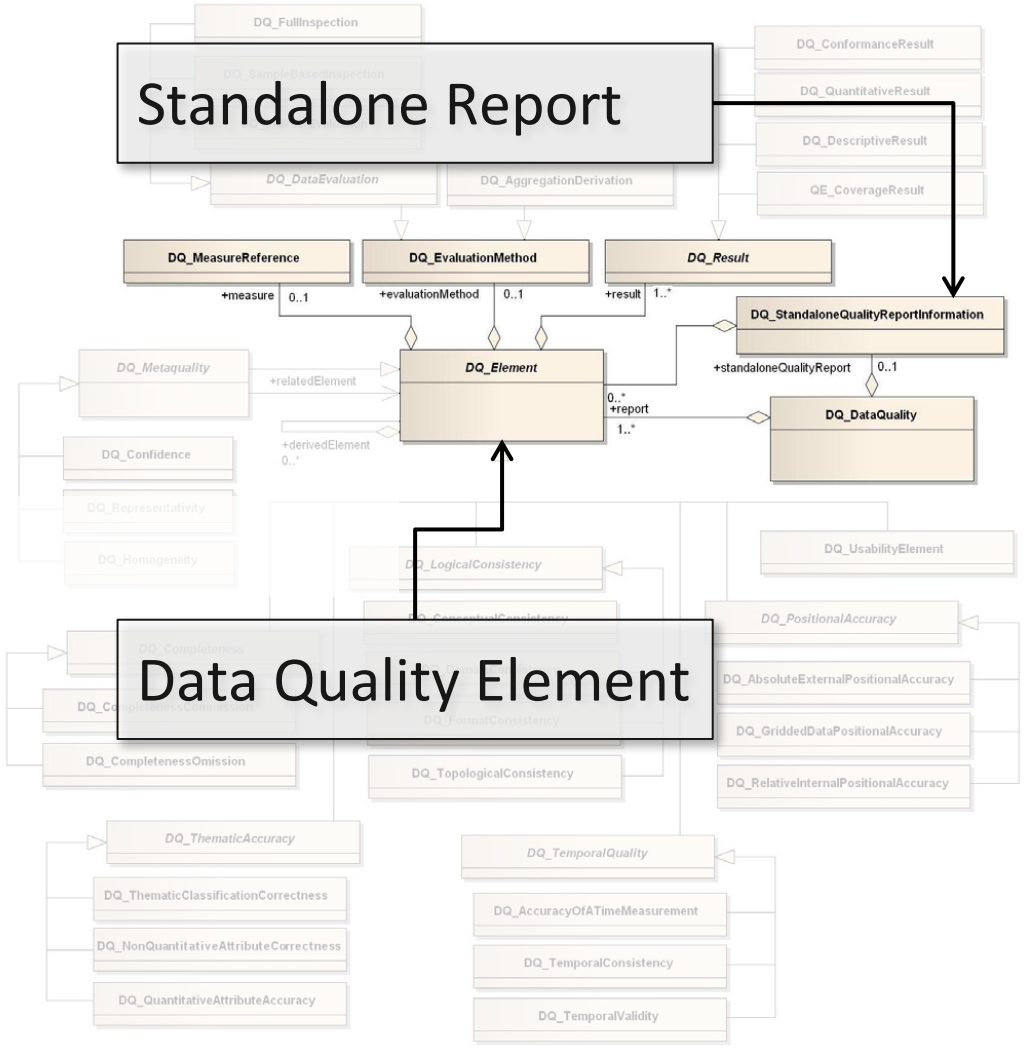
EOSDIS

The Big Picture

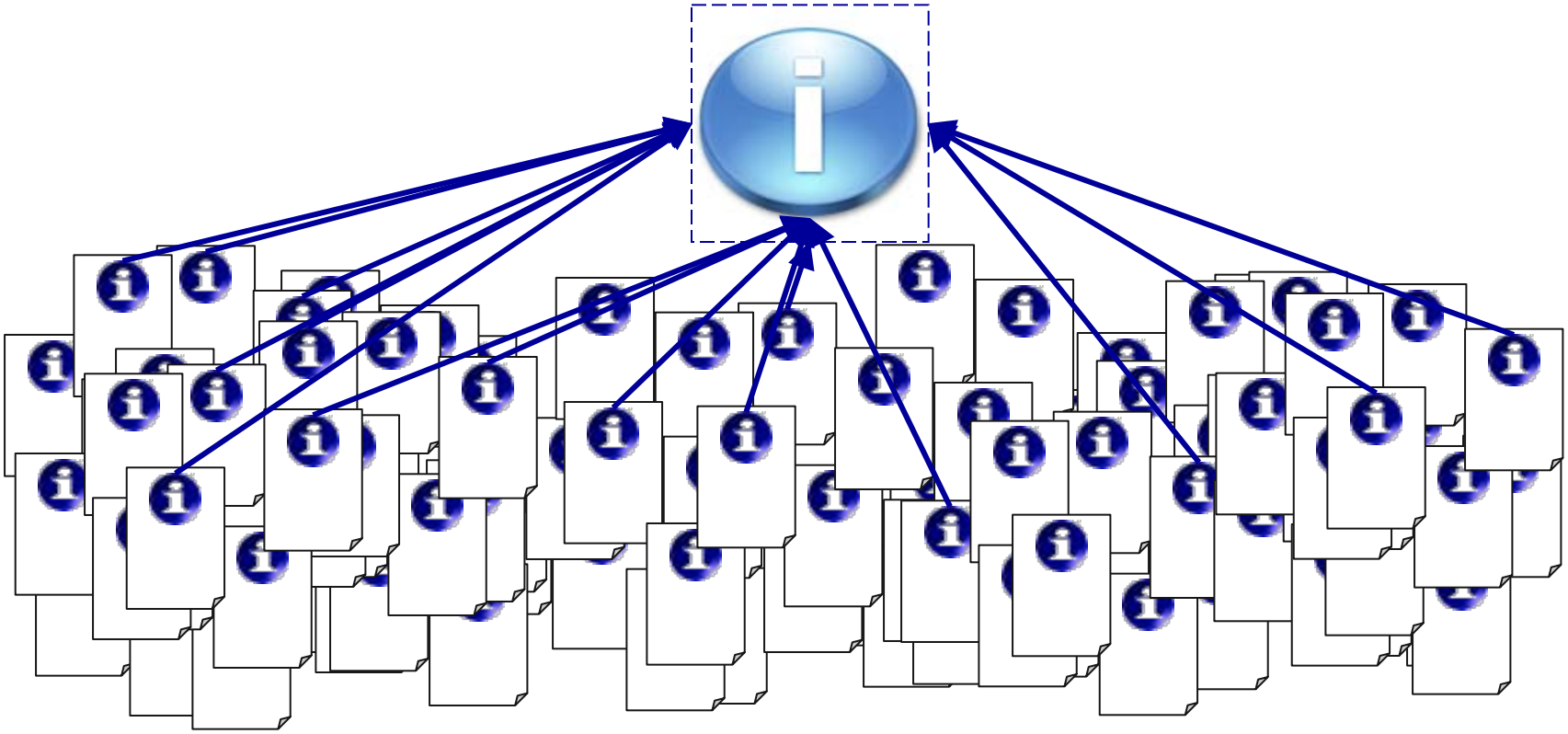
ISO 19157 is a conceptual model of data quality metadata that was recently approved as an international standard. It combines three older standards into a unified model for describing data quality.

Many of the principle elements of this conceptual model are abstract, they can be implemented in several ways.

When only the abstract concepts are considered, the model is very simple.



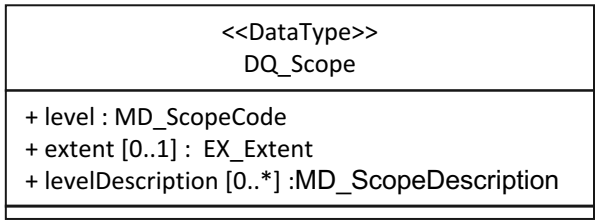
Enterprise Systems?



Data Quality Scope

"The quality of my data vary in time and space and different parameters have different quality measures and results."

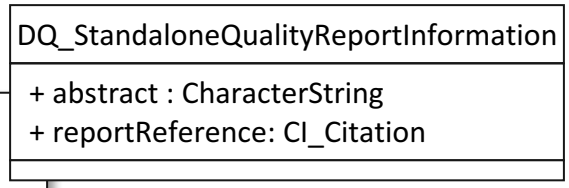
ISO quality reports all include descriptions of temporal and spatial extents and elements of the data set that they pertain to. You can say things like:
Between 2001 and 2002 the quality of the data in the northern hemisphere …
or
The data collected by this sensor degraded during June 2011 because…
or
Quality information for this parameter is in this variable…



Stand Alone Quality Reports

"There are papers and web pages that describe the quality of my data."

Papers and reports that describe data quality are StandAloneReports. Metadata can include brief descriptions of the results (abstracts) and references to any number of these (citations).



Abstract: The fire training-set may also have been biased against savanna and savanna woodland fires since their detection is more difficult than in humid, forest environments with cool background temperatures [Malingreau, 1990]. There may, therefore, be an under-sampling of warmer background environments.

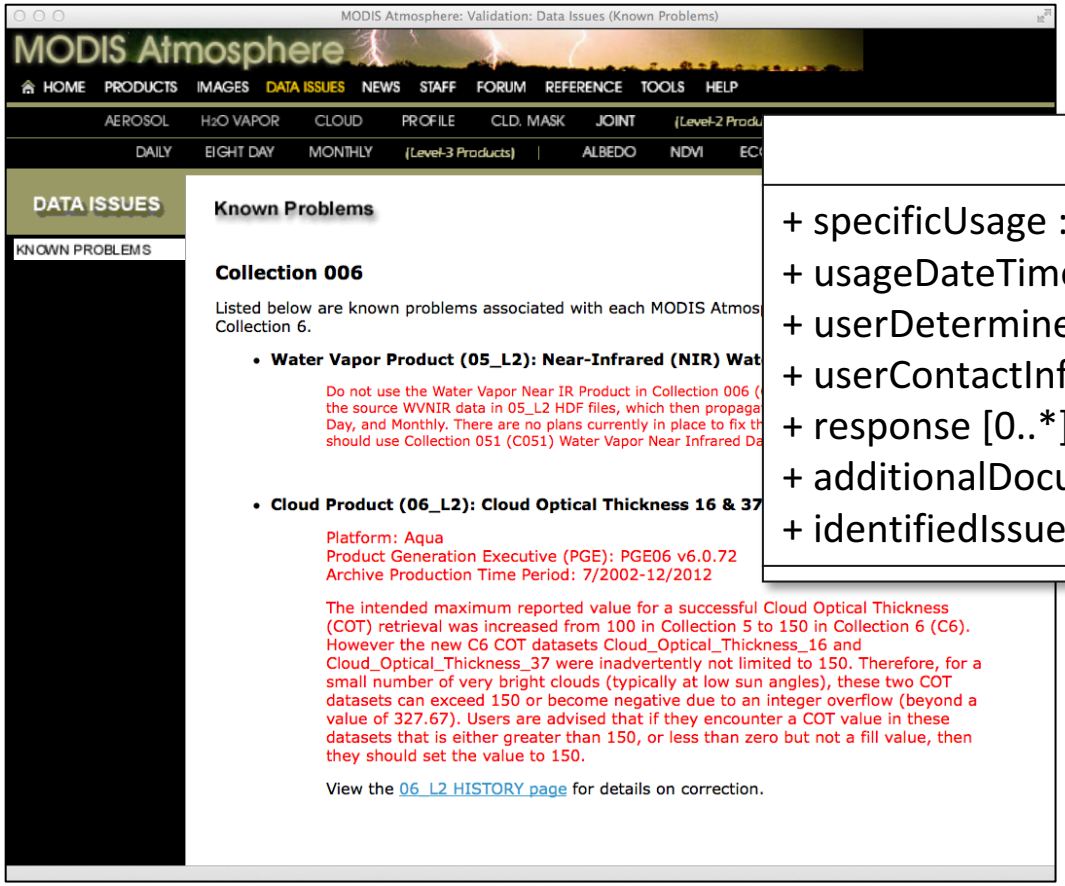
DOI

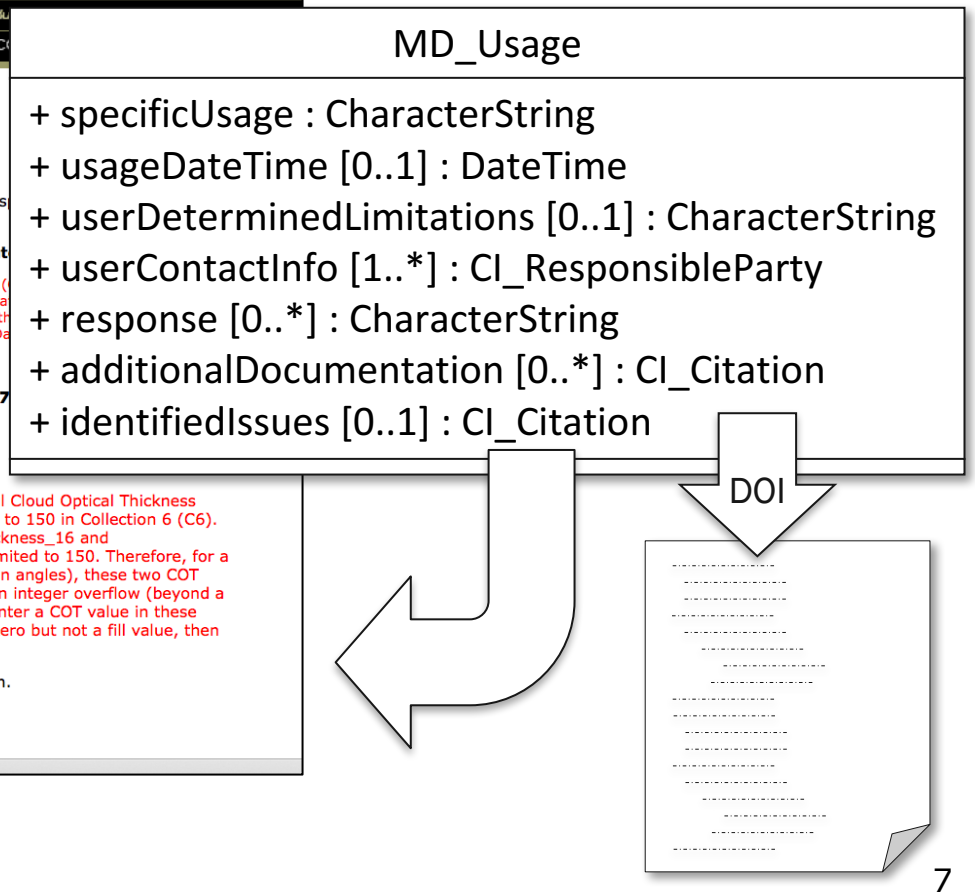
Citation: Malingreau J.P, 1990, The contribution of remote sensing to the global monitoring of fires in tropical and subtropical ecosystems. In: *Fire in Tropical Biota*, (J.G. Goldammer , editor), Springer Verlag , Berlin: 337-370.

EOSDIS

Data Usage (19115-1)

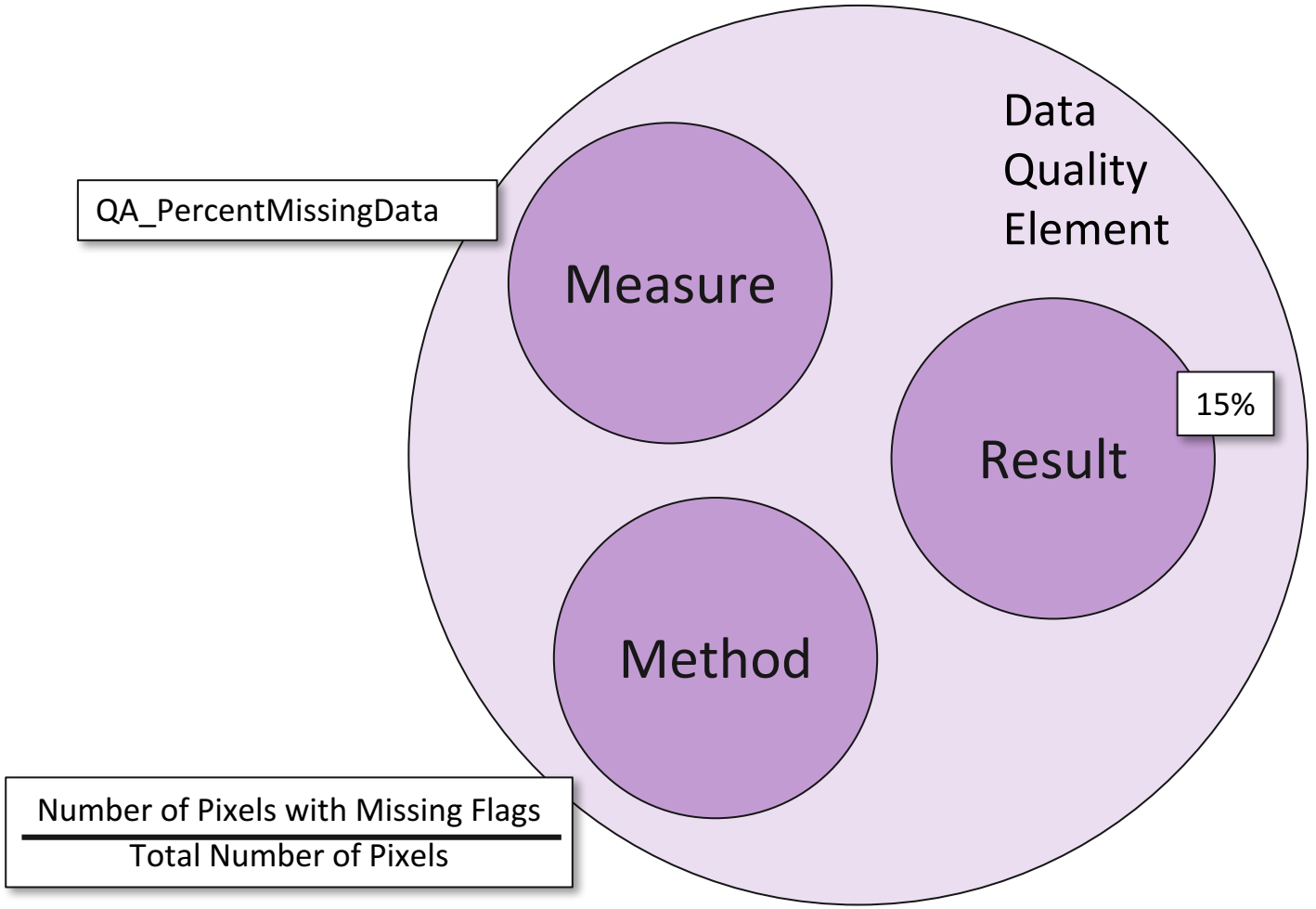
"Users increase our understanding of data quality. We need to keep them in the loop."





EOSDIS

What is a Data Quality Element?



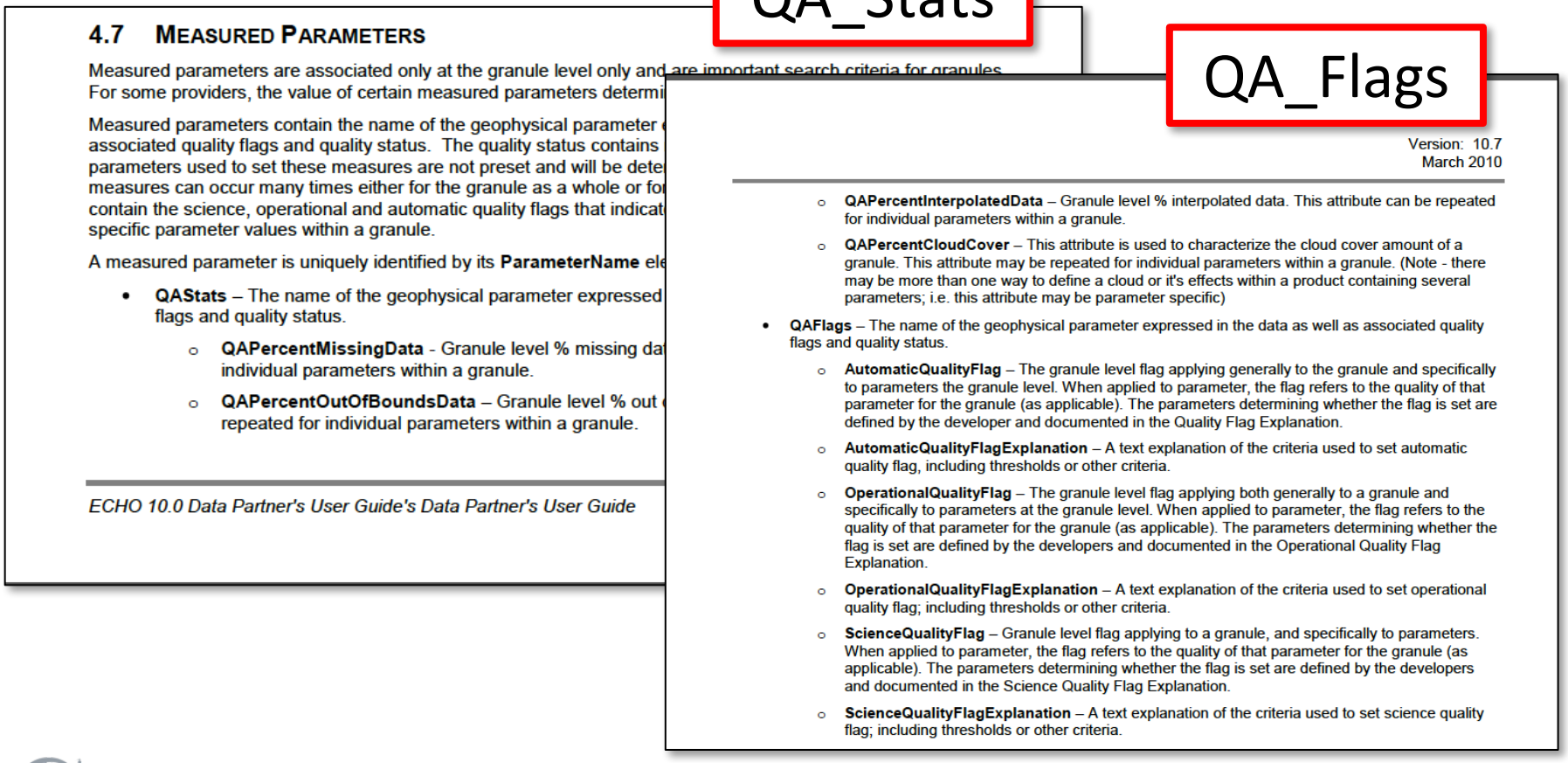
What Are Quality Measures?

"My metadata already include data quality measures ."

NASA EOSDIS metadata includes two types of quality measures.

QA_Stats

QA_Flags



4.7 MEASURED PARAMETERS

Measured parameters are associated only at the granule level only and are important search criteria for granules. For some providers, the value of certain measured parameters determi

Measured parameters contain the name of the geophysical parameter associated quality flags and quality status. The quality status contains parameters used to set these measures are not preset and will be dete measures can occur many times either for the granule as a whole or fo contain the science, operational and automatic quality flags that indicat specific parameter values within a granule.

A measured parameter is uniquely identified by its **ParameterName** ele

- **QAStats** – The name of the geophysical parameter expressed flags and quality status.
  - **QAPercentMissingData** - Granule level % missing da individual parameters within a granule.
  - **QAPercentOutOfBoundsData** – Granule level % out repeated for individual parameters within a granule.

ECHO 10.0 Data Partner's User Guide's Data Partner's User Guide

Version: 10.7
March 2010

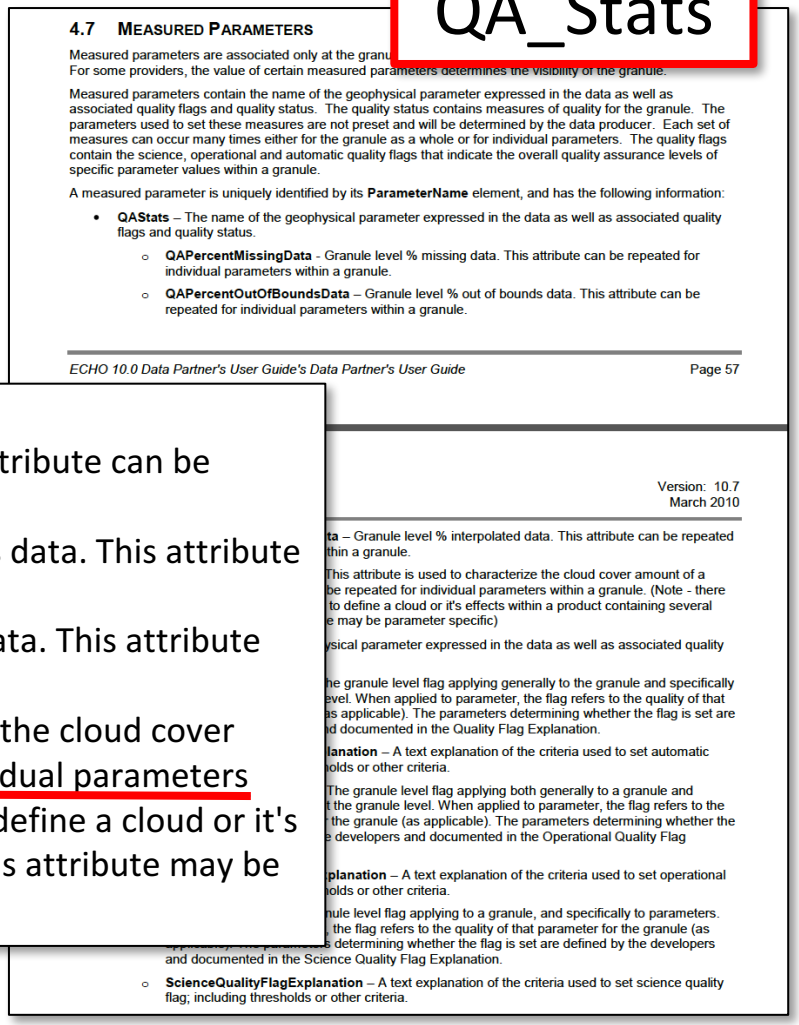
- **QAPercentInterpolatedData** – Granule level % interpolated data. This attribute can be repeated for individual parameters within a granule.
- **QAPercentCloudCover** – This attribute is used to characterize the cloud cover amount of a granule. This attribute may be repeated for individual parameters within a granule. (Note - there may be more than one way to define a cloud or it's effects within a product containing several parameters; i.e. this attribute may be parameter specific)

- **QAFlags** – The name of the geophysical parameter expressed in the data as well as associated quality flags and quality status.
  - **AutomaticQualityFlag** – The granule level flag applying generally to the granule and specifically to parameters the granule level. When applied to parameter, the flag refers to the quality of that parameter for the granule (as applicable). The parameters determining whether the flag is set are defined by the developer and documented in the Quality Flag Explanation.
  - **AutomaticQualityFlagExplanation** – A text explanation of the criteria used to set automatic quality flag, including thresholds or other criteria.
  - **OperationalQualityFlag** – The granule level flag applying both generally to a granule and specifically to parameters at the granule level. When applied to parameter, the flag refers to the quality of that parameter for the granule (as applicable). The parameters determining whether the flag is set are defined by the developers and documented in the Operational Quality Flag Explanation.
  - **OperationalQualityFlagExplanation** – A text explanation of the criteria used to set operational quality flag; including thresholds or other criteria.
  - **ScienceQualityFlag** – Granule level flag applying to a granule, and specifically to parameters. When applied to parameter, the flag refers to the quality of that parameter for the granule (as applicable). The parameters determining whether the flag is set are defined by the developers and documented in the Science Quality Flag Explanation.
  - **ScienceQualityFlagExplanation** – A text explanation of the criteria used to set science quality flag; including thresholds or other criteria.

What Are Quality Measures?

"I use consistent Quality Measures across many products."

QA_Stats



4.7 MEASURED PARAMETERS

Measured parameters are associated only at the granu
For some providers, the value of certain measured parameters determines the visibility of the granule.

Measured parameters contain the name of the geophysical parameter expressed in the data as well as associated quality flags and quality status. The quality status contains measures of quality for the granule. The parameters used to set these measures are not preset and will be determined by the data producer. Each set of measures can occur many times either for the granule as a whole or for individual parameters. The quality flags contain the science, operational and automatic quality flags that indicate the overall quality assurance levels of specific parameter values within a granule.

A measured parameter is uniquely identified by its **ParameterName** element, and has the following information:

- **QAStats** – The name of the geophysical parameter expressed in the data as well as associated quality flags and quality status.
  - **QAPercentMissingData** - Granule level % missing data. This attribute can be repeated for individual parameters within a granule.
  - **QAPercentOutOfBoundsData** – Granule level % out of bounds data. This attribute can be repeated for individual parameters within a granule.

ECHO 10.0 Data Partner's User Guide's Data Partner's User Guide Page 57

Version: 10.7
March 2010

ta – Granule level % interpolated data. This attribute can be repeated
thin a granule.

This attribute is used to characterize the cloud cover amount of a
be repeated for individual parameters within a granule. (Note - there
to define a cloud or it's effects within a product containing several
e may be parameter specific)

ysical parameter expressed in the data as well as associated quality

he granule level flag applying generally to the granule and specifically
evel. When applied to parameter, the flag refers to the quality of that
as applicable). The parameters determining whether the flag is set are
d documented in the Quality Flag Explanation.

lanation – A text explanation of the criteria used to set automatic
olds or other criteria.

The granule level flag applying both generally to a granule and
the granule level. When applied to parameter, the flag refers to the
the granule (as applicable). The parameters determining whether the
e developers and documented in the Operational Quality Flag

planation – A text explanation of the criteria used to set operational
olds or other criteria.

nule level flag applying to a granule, and specifically to parameters.
the flag refers to the quality of that parameter for the granule (as
determining whether the flag is set are defined by the developers
and documented in the Science Quality Flag Explanation.

- **ScienceQualityFlagExplanation** – A text explanation of the criteria used to set science quality flag; including thresholds or other criteria.

QAStats – **Standard measures for all products**
QAPercentMissingData - Granule level % missing data. This attribute can be repeated for individual parameters within a granule.
QAPercentOutOfBoundsData – Granule level % out of bounds data. This attribute can be repeated for individual parameters within a granule.
QAPercentInterpolatedData – Granule level % interpolated data. This attribute can be repeated for individual parameters within a granule.
QAPercentCloudCover – This attribute is used to characterize the cloud cover amount of a granule. This attribute may be repeated for individual parameters within a granule. (Note - there may be more than one way to define a cloud or it's effects within a product containing several parameters; i.e. this attribute may be parameter specific)

EOSDIS

What Are Quality Measures?

"I use consistent types of Quality Measure across many products."

QA_Flags

QAFlags – **Classes of quality measures with product specific implementations**

AutomaticQualityFlag – The granule level flag applying generally to the granule and specifically to parameters the granule level. When applied to parameter, the flag refers to the quality of that parameter for the granule (as applicable). The parameters determining whether the flag is set are defined by the developer and documented in the Quality Flag Explanation.

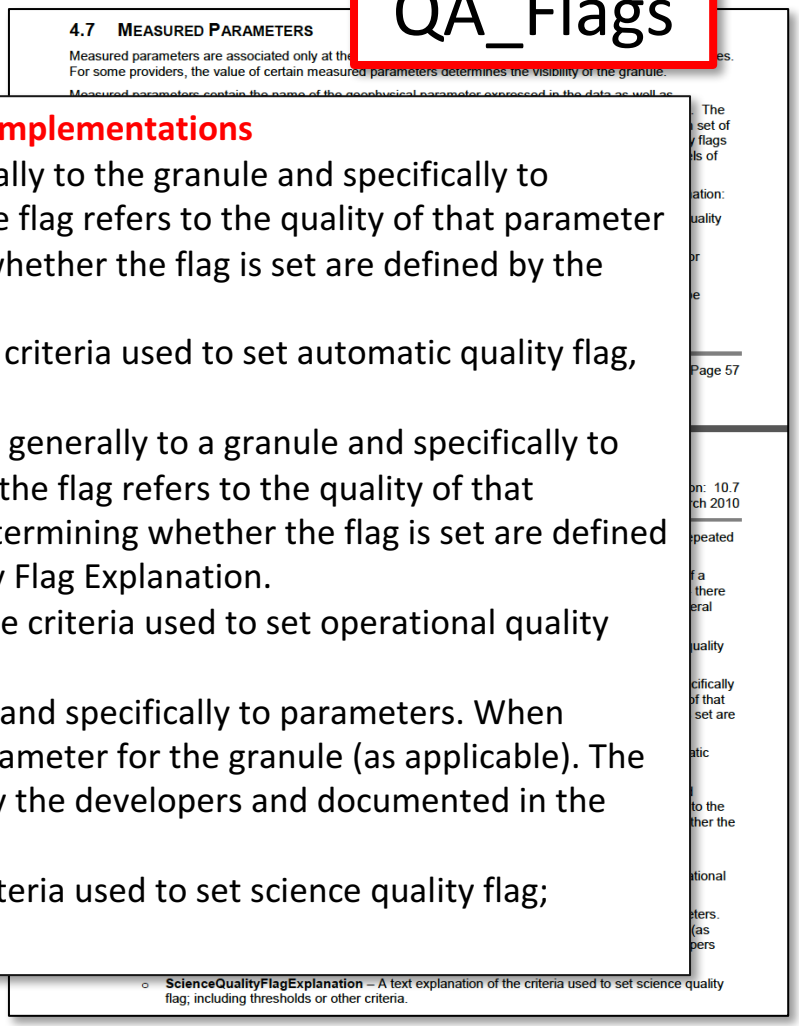
AutomaticQualityFlagExplanation – A text explanation of the criteria used to set automatic quality flag, including thresholds or other criteria.

OperationalQualityFlag – The granule level flag applying both generally to a granule and specifically to parameters at the granule level. When applied to parameter, the flag refers to the quality of that parameter for the granule (as applicable). The parameters determining whether the flag is set are defined by the developers and documented in the Operational Quality Flag Explanation.

OperationalQualityFlagExplanation – A text explanation of the criteria used to set operational quality flag; including thresholds or other criteria.

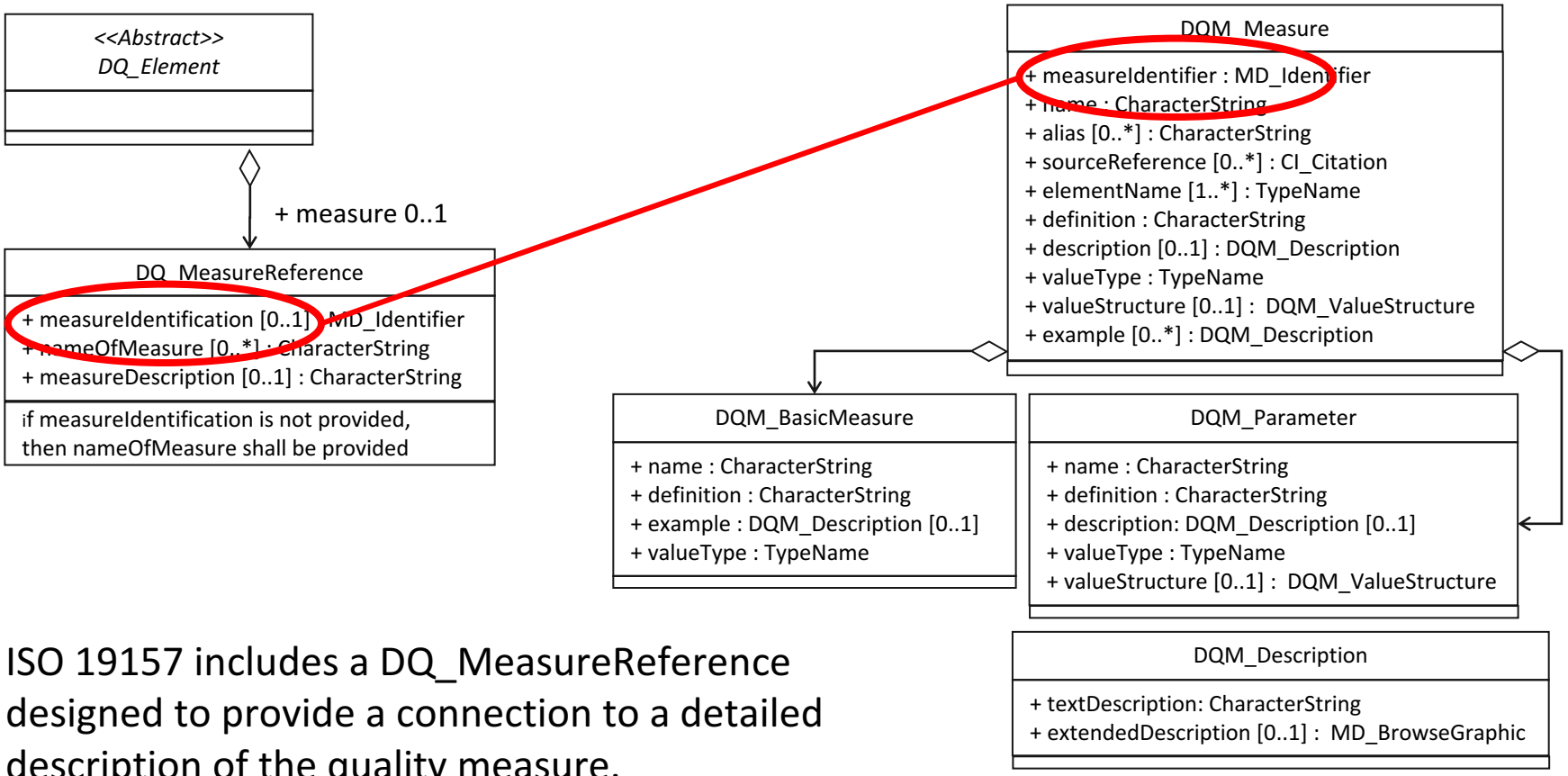
ScienceQualityFlag – Granule level flag applying to a granule, and specifically to parameters. When applied to parameter, the flag refers to the quality of that parameter for the granule (as applicable). The parameters determining whether the flag is set are defined by the developers and documented in the Science Quality Flag Explanation.

ScienceQualityFlagExplanation – A text explanation of the criteria used to set science quality flag; including thresholds or other criteria.



Data Quality Measures

"My data quality measures are consistently described in a database ."



ISO 19157 includes a DQ_MeasureReference designed to provide a connection to a detailed description of the quality measure.

Data Quality Measures

"I need to clearly and consistently explain how I measure quality."

The ISO model for quality measures includes identifiers, definitions, descriptions, references and illustrations.

Table D.28 — Number of invalid self-overlap errors

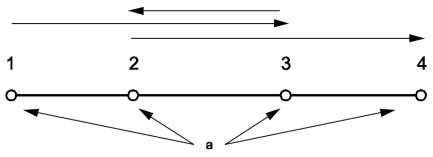
Line	Component	Description
1	Name	number of invalid self-overlap errors
2	Alias	kickbacks
3	Element name	topological consistency
4	Basic measure	error count
5	Definition	count of all items in the data that illegally self overlap
6	Description	–
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	 **Key** a Vertices
12	Identifier	27

Table D.31 — Mean value of positional uncertainties excluding outliers

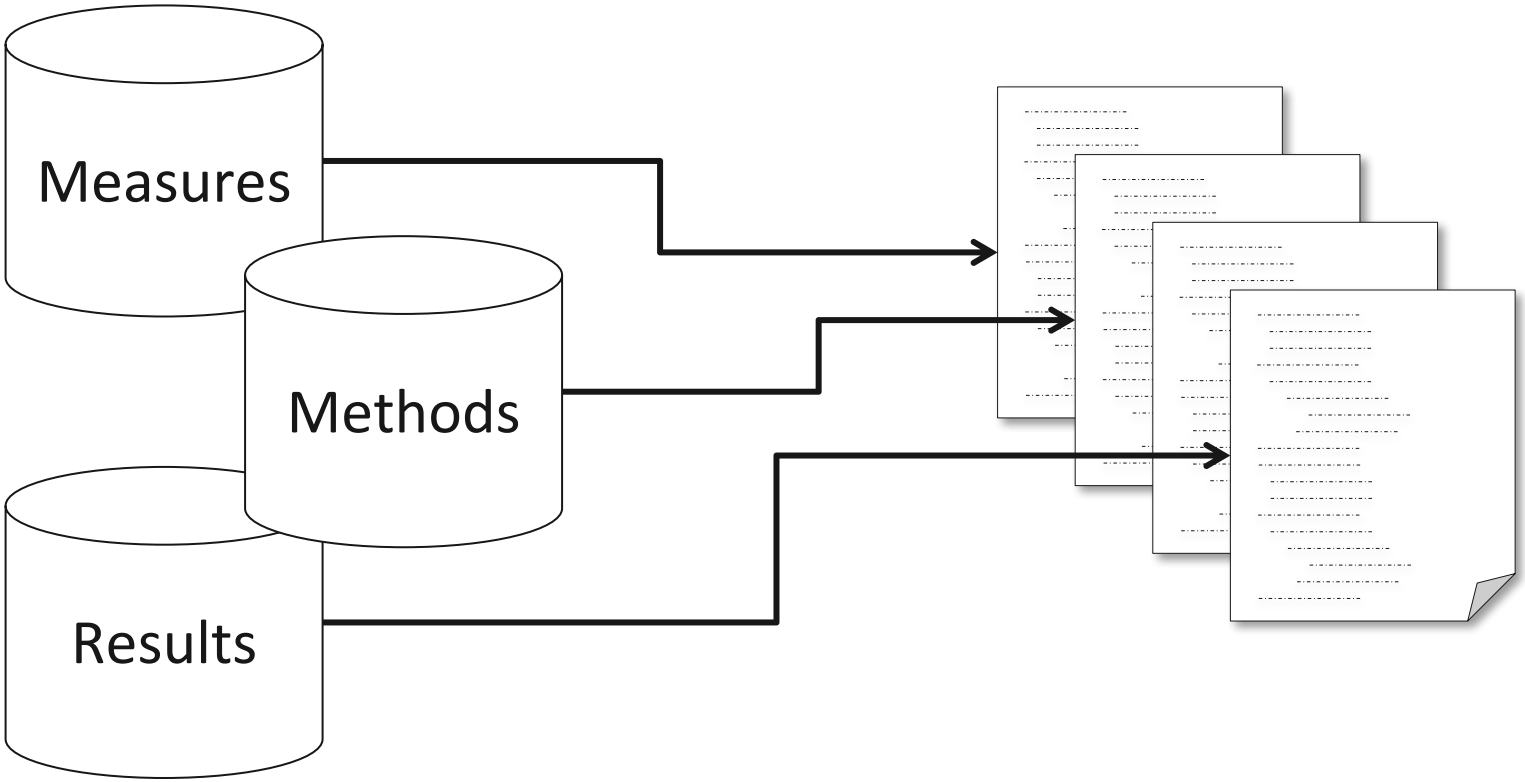
Line	Component	Description
1	Name	mean value of positional uncertainties excluding outliers (2D)
2	Alias	–
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	for a set of points where the distance does not exceed a defined threshold, the arithmetical average of distances between their measured positions and what is considered as the corresponding true positions
6	Description	For a number of points (N), the measured positions are given as x_{mi}, y_{mi} and z_{mi} coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, x_{ti}, y_{ti} and z_{ti}, are considered to represent the true positions. All positional uncertainties above a defined threshold e_{max} are then removed from the set. The positional uncertainties are calculated as $e_i' = \begin{cases} e_i, & if \quad e_i \le e_{max} \\ 0, & if \quad e_i > e_{max} \end{cases}$ The calculation of e_i is given by the data quality measure "mean value of positional uncertainties" in one, two and three dimensions. For the remaining number of errors (N_R), the mean of the horizontal absolute positions is calculated as $\bar{e}_{\text{excluding outliers}} = \frac{1}{N_R}\sum_{i=1}^{N} e_i'$ A criterion for the establishing of correspondence should also be stated (e.g. allowing for correspondence to the closest position, correspondence on vertices or along lines). The criteria for finding the corresponding points shall be reported with the data quality evaluation result.
7	Parameter	Name: e_{max} Definition: is the threshold for accepted positional uncertainties Value type: Number
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	29

EOSDIS

Modular DQ Information

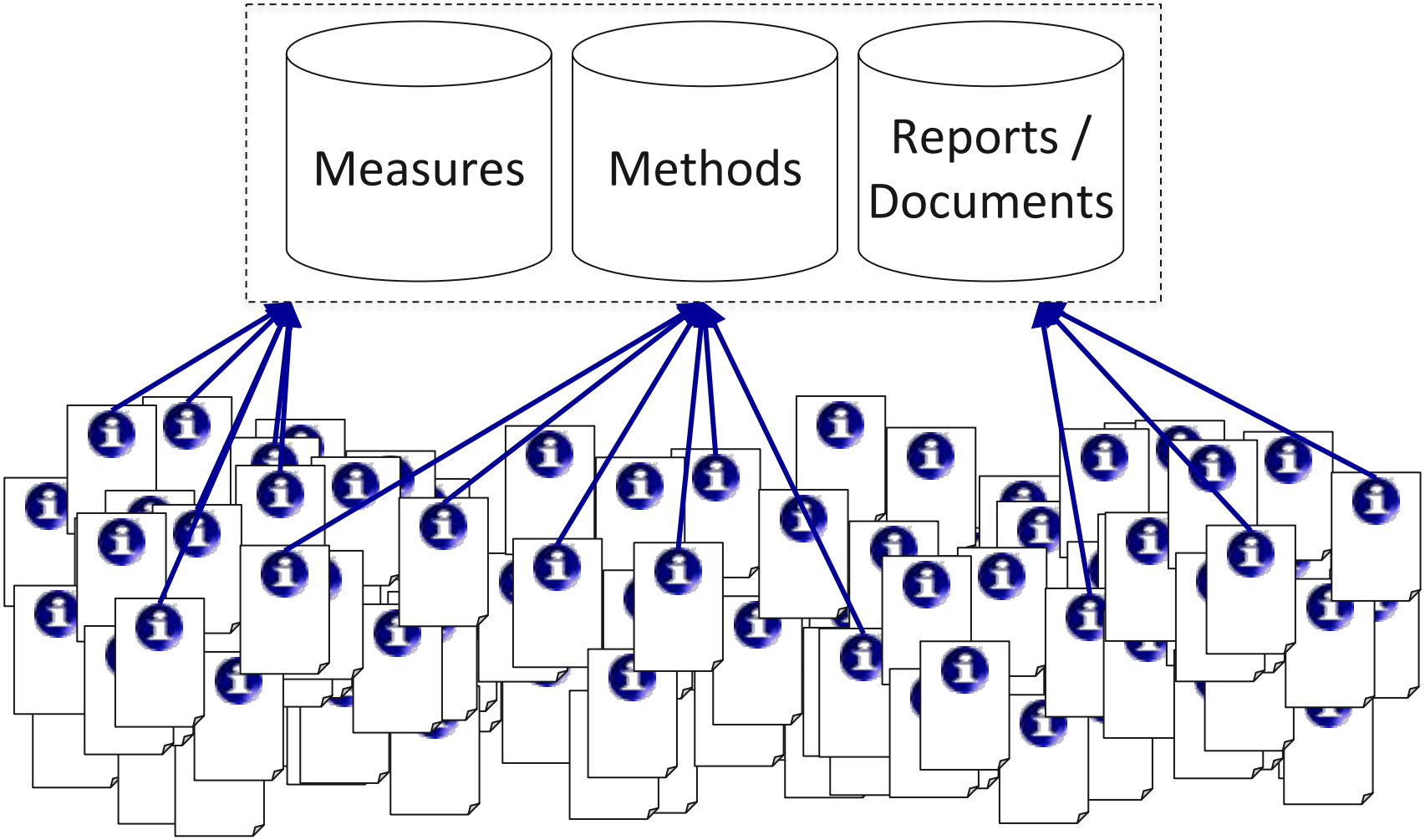
"My data quality information exists in databases or web services."

Major elements of the 19157 conceptual model are separate components that can be independently connected to the metadata and reused in multiple records.



EOSDIS

Enterprise Systems?



EOSDIS

Data Quality Results

"My metadata currently includes descriptions of the quality of my data."

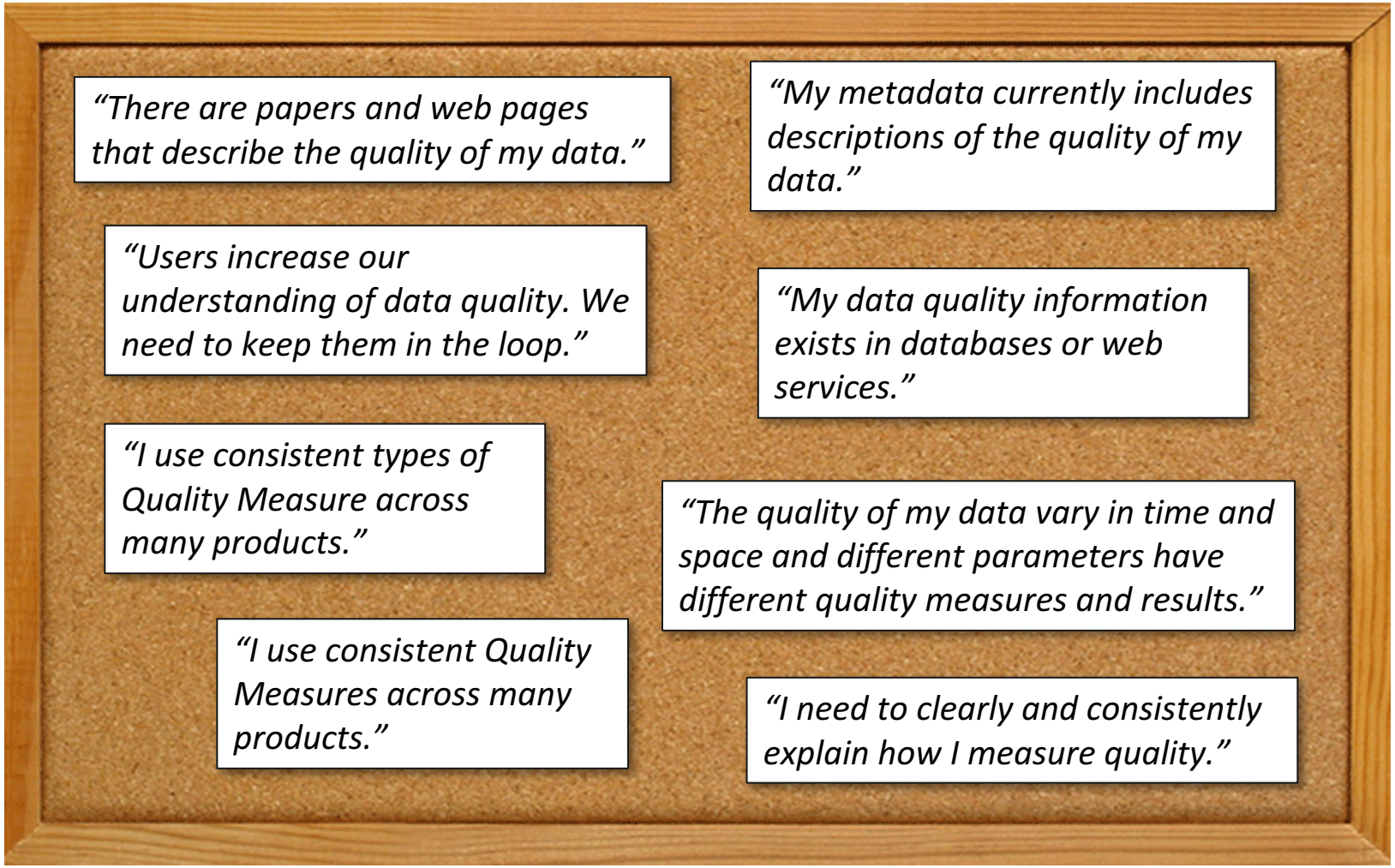
These descriptions can be included in 19157 metadata as descriptive reports.

```
<Quality>
        Due to the lack of high resolution data available over the region for
        1993-94, it has been hard to validate the product. However the maps of
        burnt areas correspond well with active fire maps for the
        region. Where large [>3km] scars are found, the detection is more
        reliable.  In areas of small scars more problems are involved. It is
        hoped that the 1994-95 data set will cover the whole of the study area
        and be calibrated by high resolution data.
  </Quality>
```

```
<gco:CharacterString>
        Due to the lack of high resolution data available over the region for
        1993-94, it has been hard to validate the product. However the maps of
        burnt areas correspond well with active fire maps for the
        region. Where large [>3km] scars are found, the detection is more
        reliable.  In areas of small scars more problems are involved. It is
        hoped that the 1994-95 data set will cover the whole of the study area
        and be calibrated by high resolution data.
  </gco:CharacterString>
```

DQ_DescriptiveResult
+ statement : CharacterString

Summary


"There are papers and web pages that describe the quality of my data."
"My metadata currently includes descriptions of the quality of my data."
"Users increase our understanding of data quality. We need to keep them in the loop."
"My data quality information exists in databases or web services."
"I use consistent types of Quality Measure across many products."
"The quality of my data vary in time and space and different parameters have different quality measures and results."
"I use consistent Quality Measures across many products."
"I need to clearly and consistently explain how I measure quality."

EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

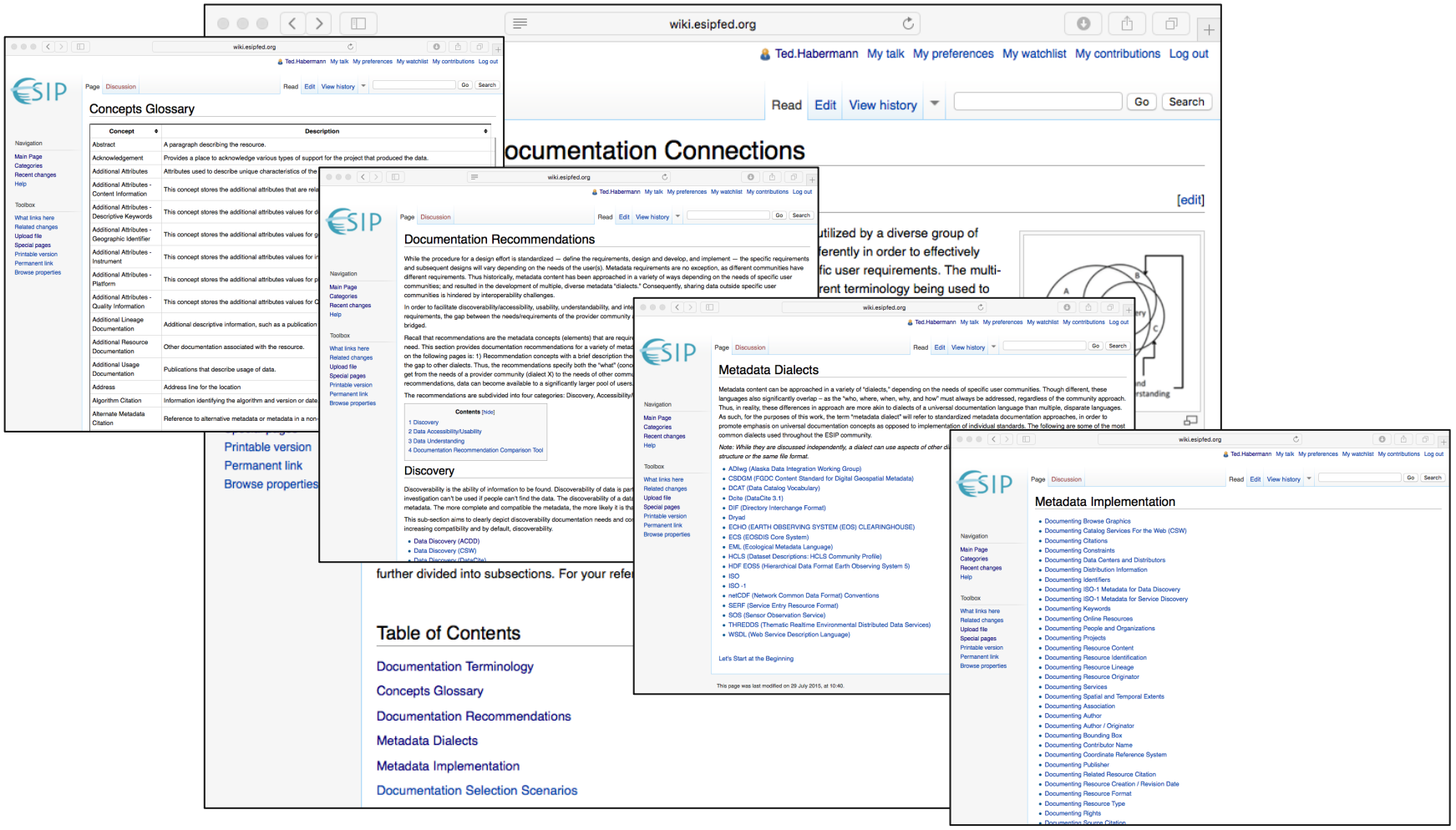
Documentation Resources on the ESIP Wiki

Ted Habermann, Sean Gordon, John Kozimor
The HDF Group
thabermann@hdfgroup.org

This work was supported by NASA/GSFC under
Raytheon Co. contract number NNG15HZ39C

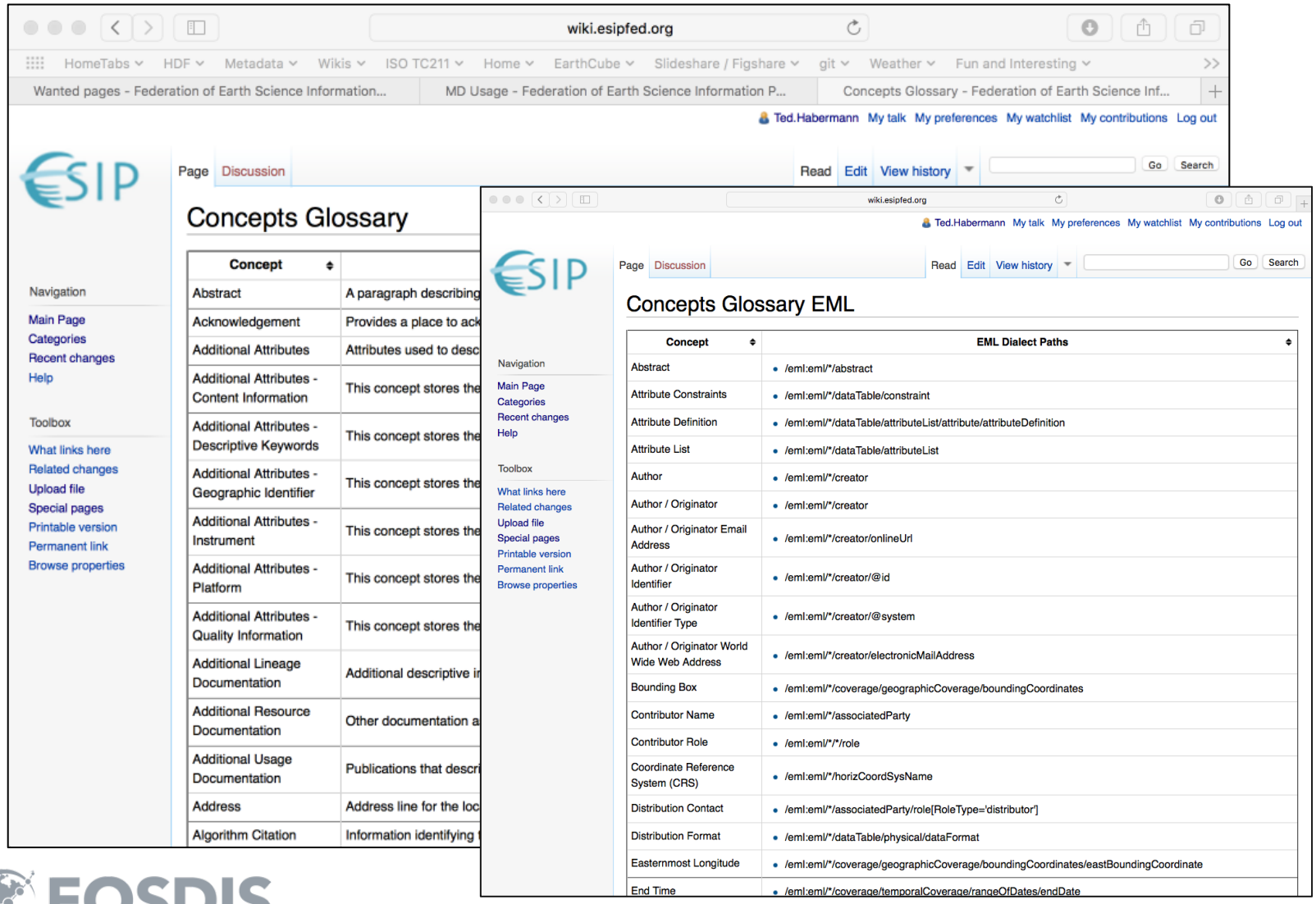
Documentation Connections

Documentation concepts, recommendations and implementations in multiple dialects

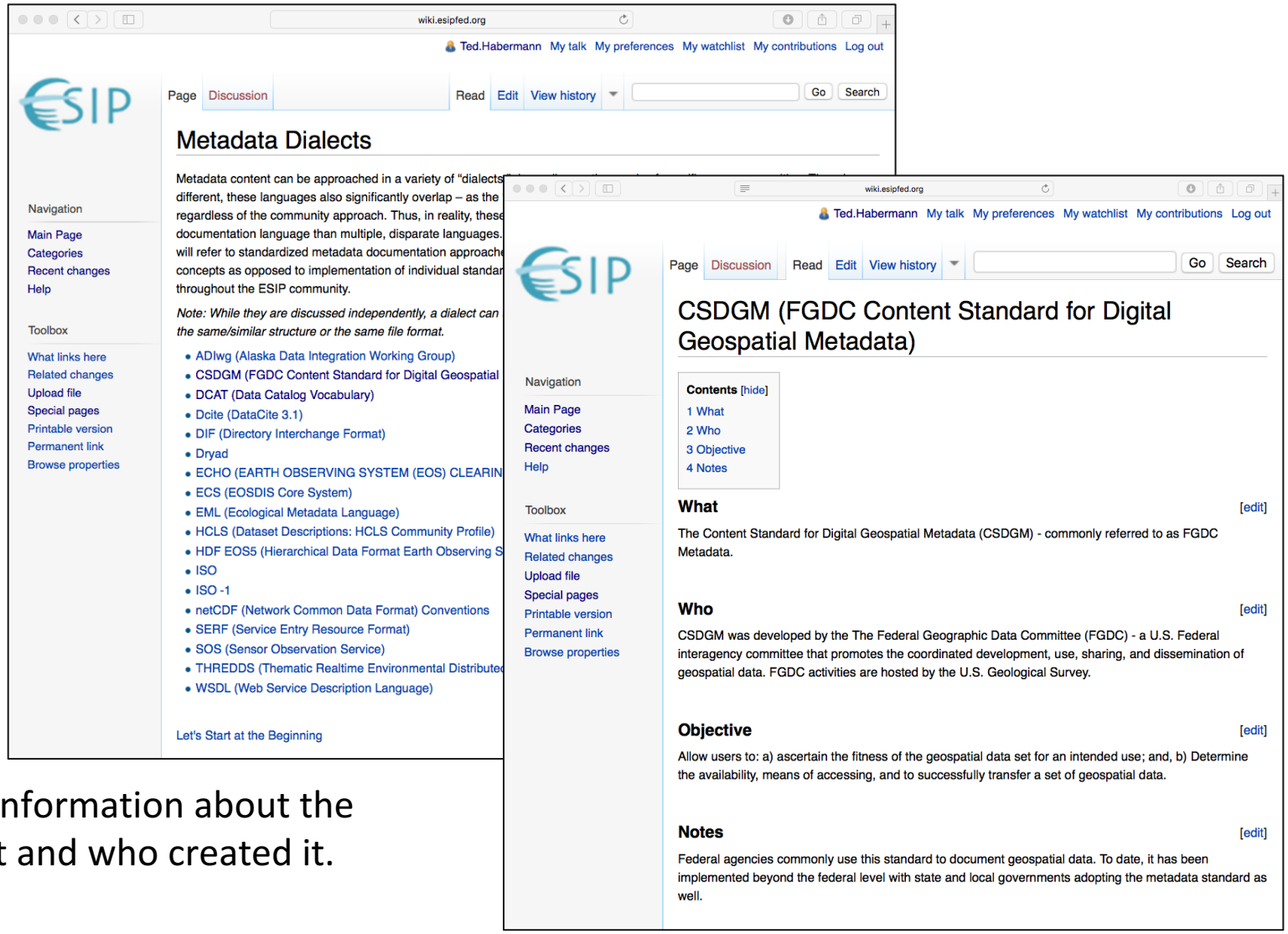


http://wiki.esipfed.org/index.php/Category:Documentation_Connections

Concept Glossary



Dialects



Basic information about the dialect and who created it.

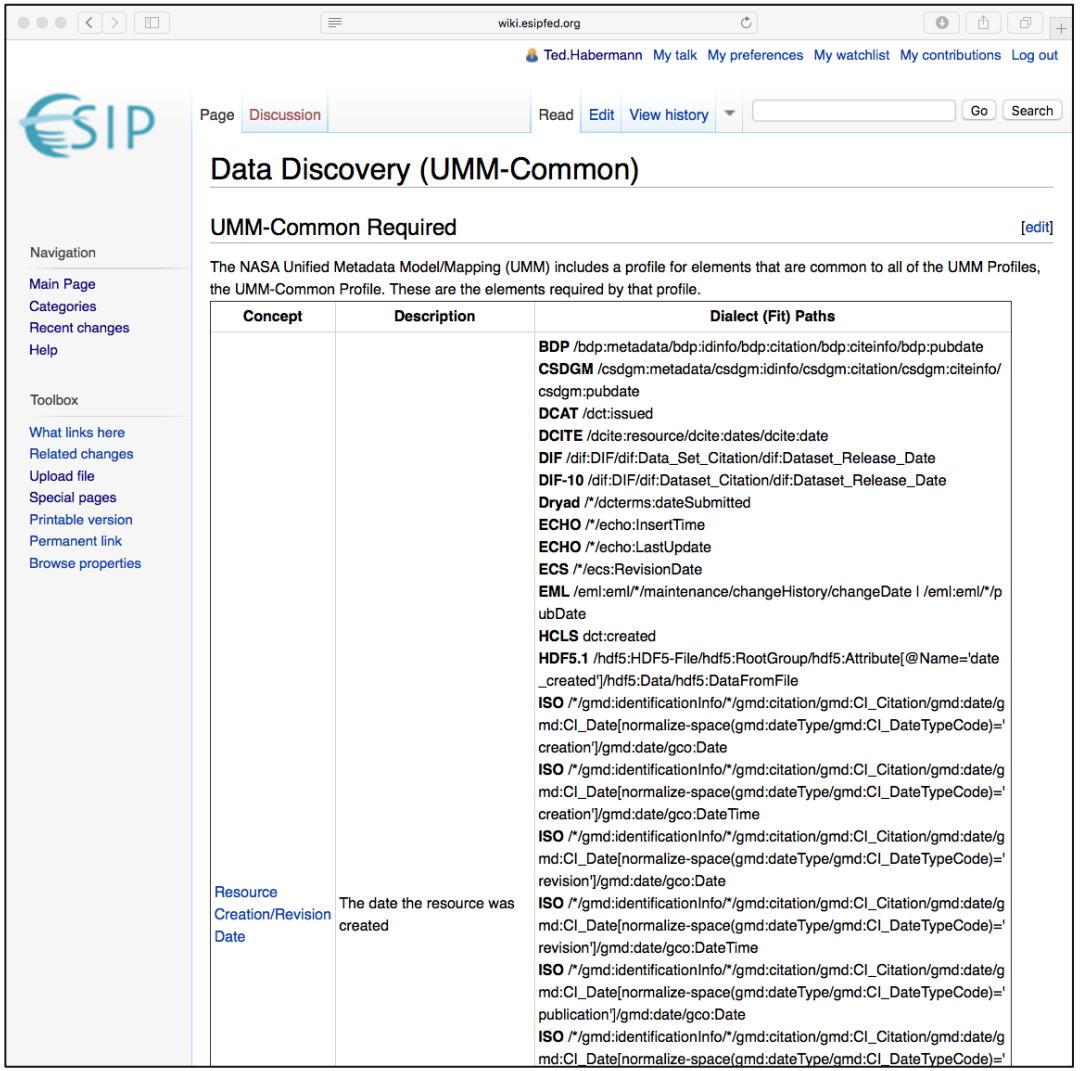
EOSDIS

Recommendations

Many recommendations include multiple levels (mandatory, recommended, optipnal). A recommendation page gives:

1. Concept Names
2. Concept Definitions and
3. Concept Implementations (multiple dialects)

For each recommendation level



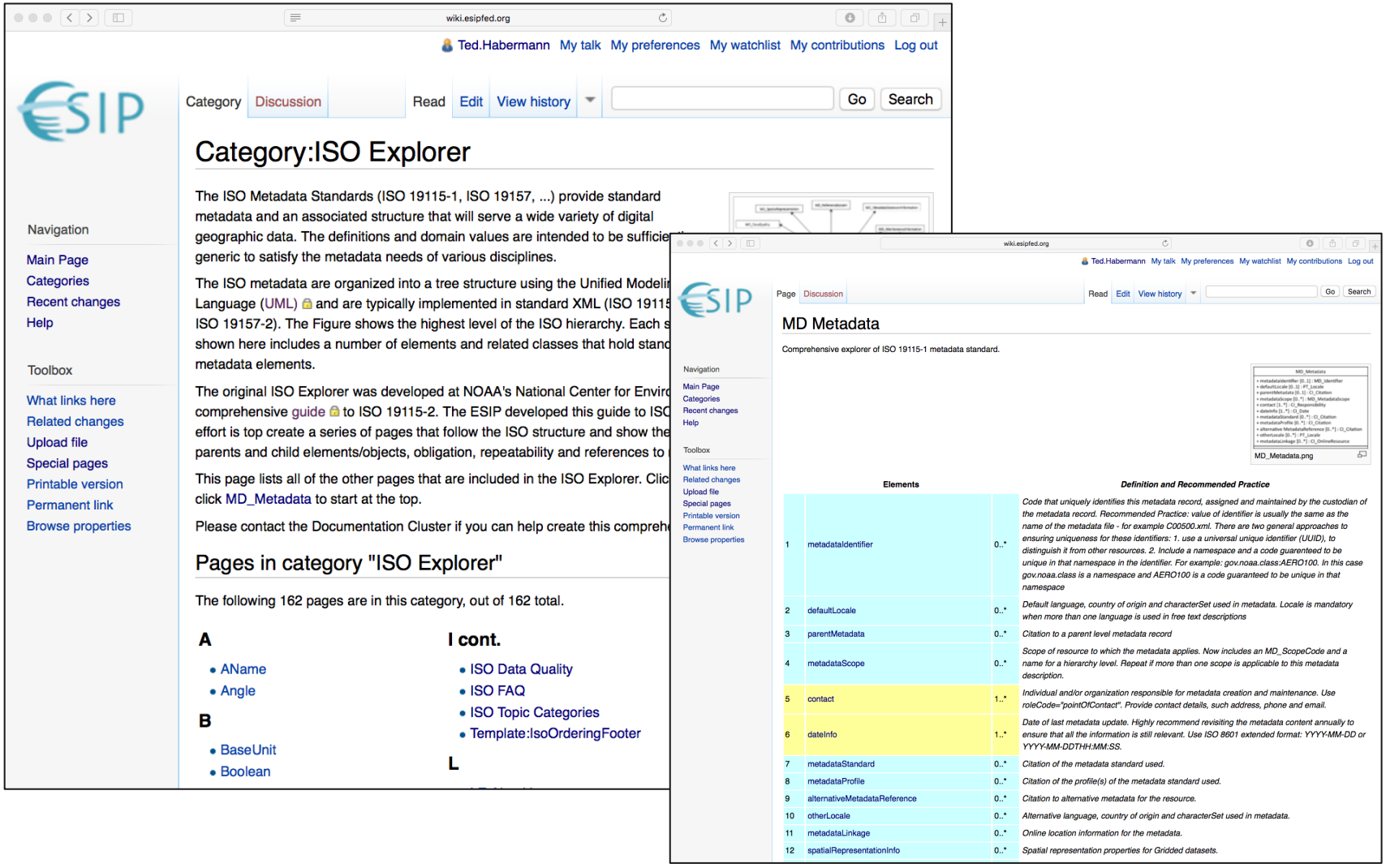
EOSDIS

Resource Title Concept

Concept	Description	Crosswalk
Resource Title	A short description of the resource. The title should be descriptive enough so that when a user is presented with a list of titles the general content of the data set can be determined.	**ADIwg** /adiwg:project/adiwg:idinfo/adiwg:citation/adiwg:citeinfo/adiwg:title **BDP** /bdp:metadata/bdp:idinfo/bdp:citation/bdp:citeinfo/bdp:title **CSDGM** /csdgm:metadata/csdgm:idinfo/csdgm:citation/csdgm:citeinfo/csdgm:title **DCAT** /dct:title **DCITE** /dcite:resource/dcite:titles/dcite:title **DIF** /dif:DIF/dif:Entry_Title **DIF** /dif:DIF/dif:Data_Set_Citation/dif:Dataset_Title **DIF-10** /dif:DIF/dif:Entry_Title **DIF-10** /dif:DIF/dif:Dataset_Citation/dif:Dataset_Title **Dryad** /*/dcterms:title **ECHO** /*/echo:ShortName | /*/echo:LongName **ECHO** /*/echo:DataSetId **ECS** /*/ecs:ShortName | /*/ecs:LongName **EML** /eml:eml/*/title **HCLS** dct:title **HDF5.1** /hdf5:HDF5-File/hdf5:RootGroup/hdf5:Attribute[@Name='title']/hdf5:Data/hdf5:DataFromFile **HDF5.1** /hdf5:HDF5-File/hdf5:RootGroup/hdf5:Group[@Name='METADATA']/hdf5:Group[@Name='COLLECTIONMETADATA']/hdf5:Attribute[@Name='LongName']/hdf5:Data/hdf5:DataFromFile **ISO** /*/gmd:identificationInfo/*/gmd:citation/gmd:CI_Citation/gmd:title//* **ISO-1** /mdb:MD_Metadata/mdb:identificationInfo/*/mri:citation/cit:CI_Citation/cit:title//* **MODS** //mods:mods/mods:titleInfo/mods:title **Mercury** /mercury:metadata/mercury:idinfo/mercury:citation/mercury:citeinfo/mercury:title **OGC-SOS** /sos:Capabilities/ows:ServiceIdentification/ows:Title **Onedcx** /onedcx:metadata/onedcx:simpleDc/dcterms:title **RDA-CISL** /rda:dsOverview/rda:title **SERF** /serf:SERF/serf:Entry_Title **THREDDS** /thredds:catalog/thredds:dataset/@name **THREDDS** /thredds:catalog/thredds:dataset/thredds:metadata/dc:title **THREDDS** //thredds:dataset[1]/@name **UMM** /umm:UMM/umm:CollectionCitation/umm:Title

EOSDIS

ISO Explorer



EOSDIS

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

ISO Explorer Pages

Class Name

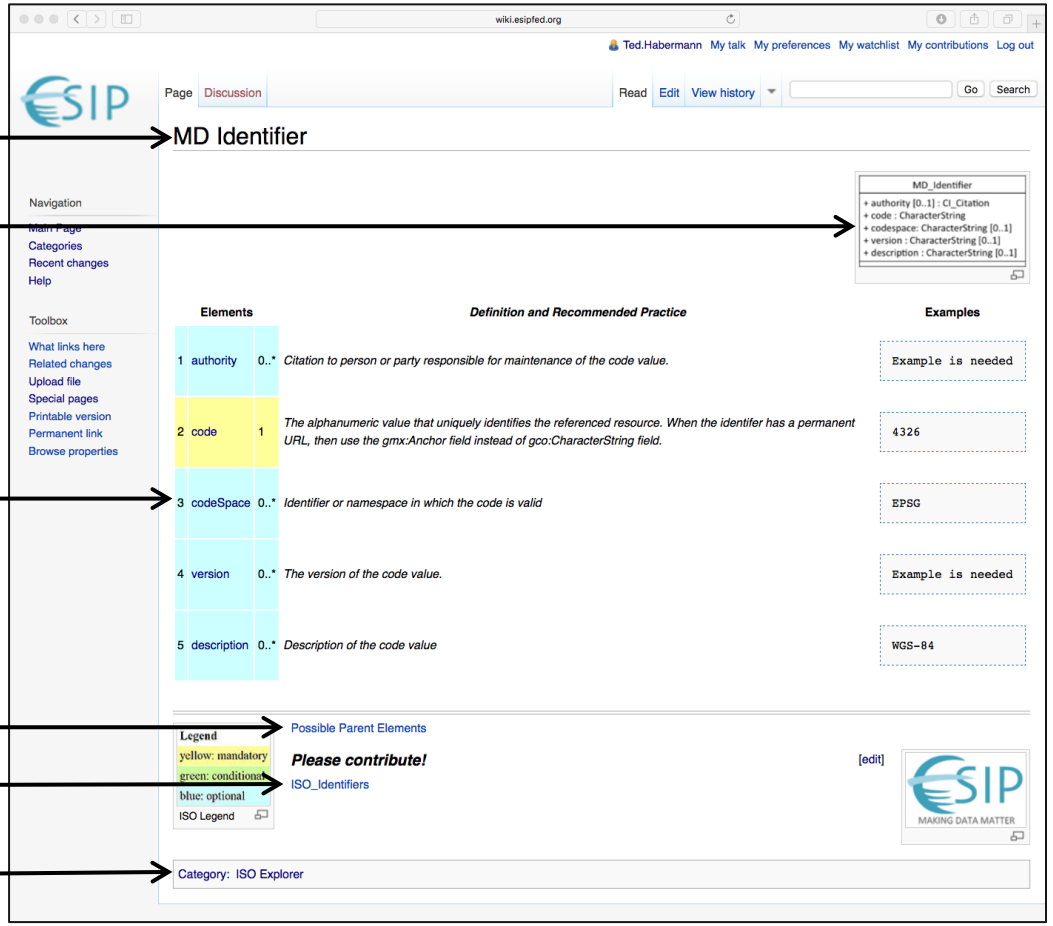
UML

Element Names, Definitions and Examples

Parents

Guidance

All Explorer Pages



Acknowledgements

This work was partially supported by contract number NNG15HZ39C from NASA.

Raytheon

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NASA or The HDF Group.

The HDF Group