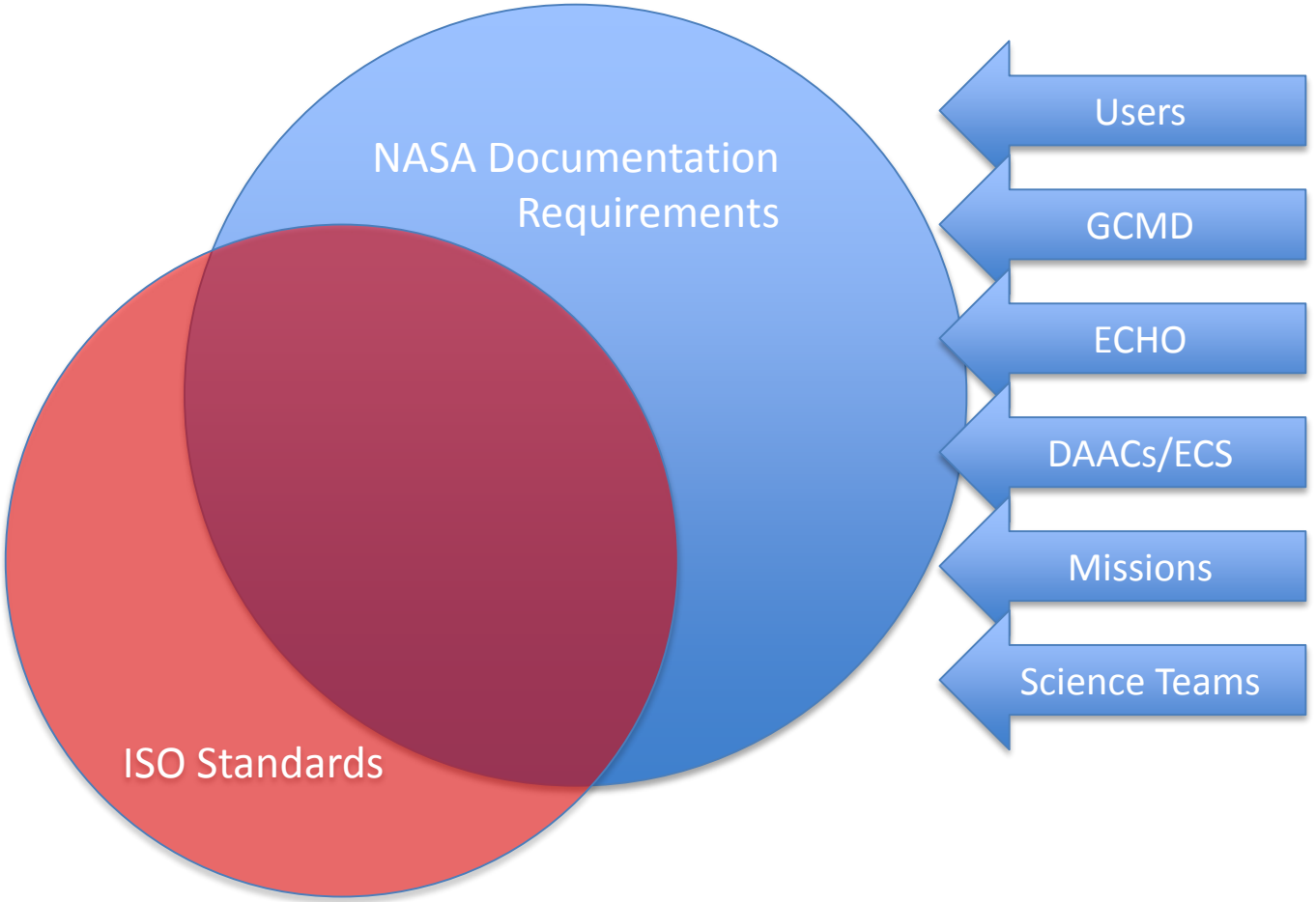
NASA Metadata Trends

Ted Habermann, GHRSST Data Assembly TAG

June, 2013



thabermann@hdfgroup.org

New Capabilities (ISO 19115-1)

1. Keyword Types – All GCMD Types included in standard codelist
2. Multiple names for measured or calculated parameters – local_name, long_name, standard_name,
3. Using NASA identifiers – added namespace, description fields (long names) to identifiers
4. Multiple metadata dates – metadata dates now have types (creation, revision, …)
5. References to documentation or other information that is published in the literature – science papers, user guides, metadata in other dialects…
6. Data usage and limitations identified by users – data provider responses
7. Special constraints or use limitations for different time periods – data use limitations in specific times
8. Describe coverages as physical measurements, quality information, auxiliary information, reference information, or model results.

Attribute Convention for Data Discovery

<attribute name="id" value="netcdf/attribute/@name=id"/>
<attribute name="title" value="netcdf/attribute/@name=title"/>
<attribute name="creator_email" value="netcdf/attribute/@name=creator_email"/>
<attribute name="creator_name" value="netcdf/attribute/@name=creator_name"/>
<attribute name="creator_url" value="netcdf/attribute/@name=creator_url"/>
<attribute name="institution" value="netcdf/attribute/@name=institution"/>
<attribute name="publisher_name" value="netcdf/attribute/@name=publisher_name"/>
<attribute name="publisher_url" value="netcdf/attribute/@name=publisher_url"/>
<attribute name="publisher_email" value="netcdf/attribute/@name=publisher_email"/>
<attribute name="date_created" value="netcdf/attribute/@name=date_created"/>
<attribute name="date_modified" value="netcdf/attribute/@name=date_modified"/>
<attribute name="date_issued" value="netcdf/attribute/@name=date_issued"/>

Object Convention for Data Discovery

Citation Group:

```
<group name="citation" uuid="UUID">
   <attribute name="objectType" value="acdd:Citation"/>
   <attribute name="title" value="netcdf/attribute/@name=title"/>
   <attribute name="identifier" value="netcdf/attribute/@name=id"/>
   <group name="creationDate" uuid="UUID">
  <group name="modificationDate" uuid="UUID">
  <group name="issuedDate" uuid="UUID">
  <group name="originator" uuid="UUID">
  <group name="publisher" uuid="UUID”>
</group>
```

Responsible Party Group:

```
<group name="originator" uuid="UUID">
     <attribute name="objectType" value="acdd:ResponsibleParty"/>
     <attribute name="role" value="originator"/>
     <attribute name="individualName" value="netcdf/attribute/@name=creator_name"/>
     <attribute name="organisationName" value="netcdf/attribute/@name=institution"/>
     <attribute name="electronicMailAddress" value="netcdf/attribute/@name=creator_email"/>
     <group name="onlineResource" uuid="UUID">
 </group>
```

NASA Metadata Dialects

ECHO:

1. ~150 Elements with content. Keywords, identifiers, and contact information are the most commonly occurring fields.
2. 99+% of granule and collection content translates to ISO (now in test at ECHO)
3. 324 AdditionalAttributes occur over 3000 times
4. Most Common: PROCESSVERSION QAPERCENTGOODQUALITY QAPERCENTOTHERQUALITY QAPERCENTNOTPRODUCEDCLOUD QAPERCENTNOTPRODUCEDOTHER

DIF:

1. ~130 elements with content. Keywords, identifiers, and contact information are the most commonly occurring fields.
2. 99% translation to ISO, ~79% translation to ECHO

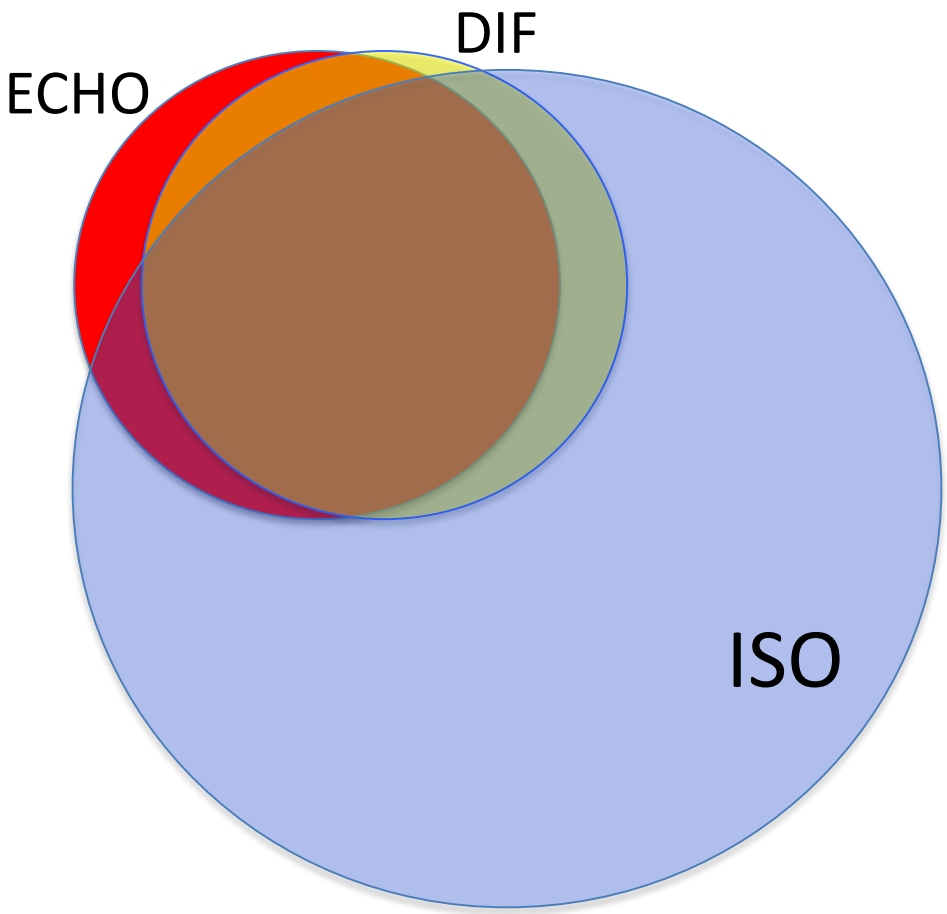
More Overlap Than Difference

The metadata dialects currently used by ESDIS have much more overlap than difference.

The mappings are generally well understood: we are in a tweaking stage.

The translations can be implemented using well-known, standard tools that are designed for XML processing.

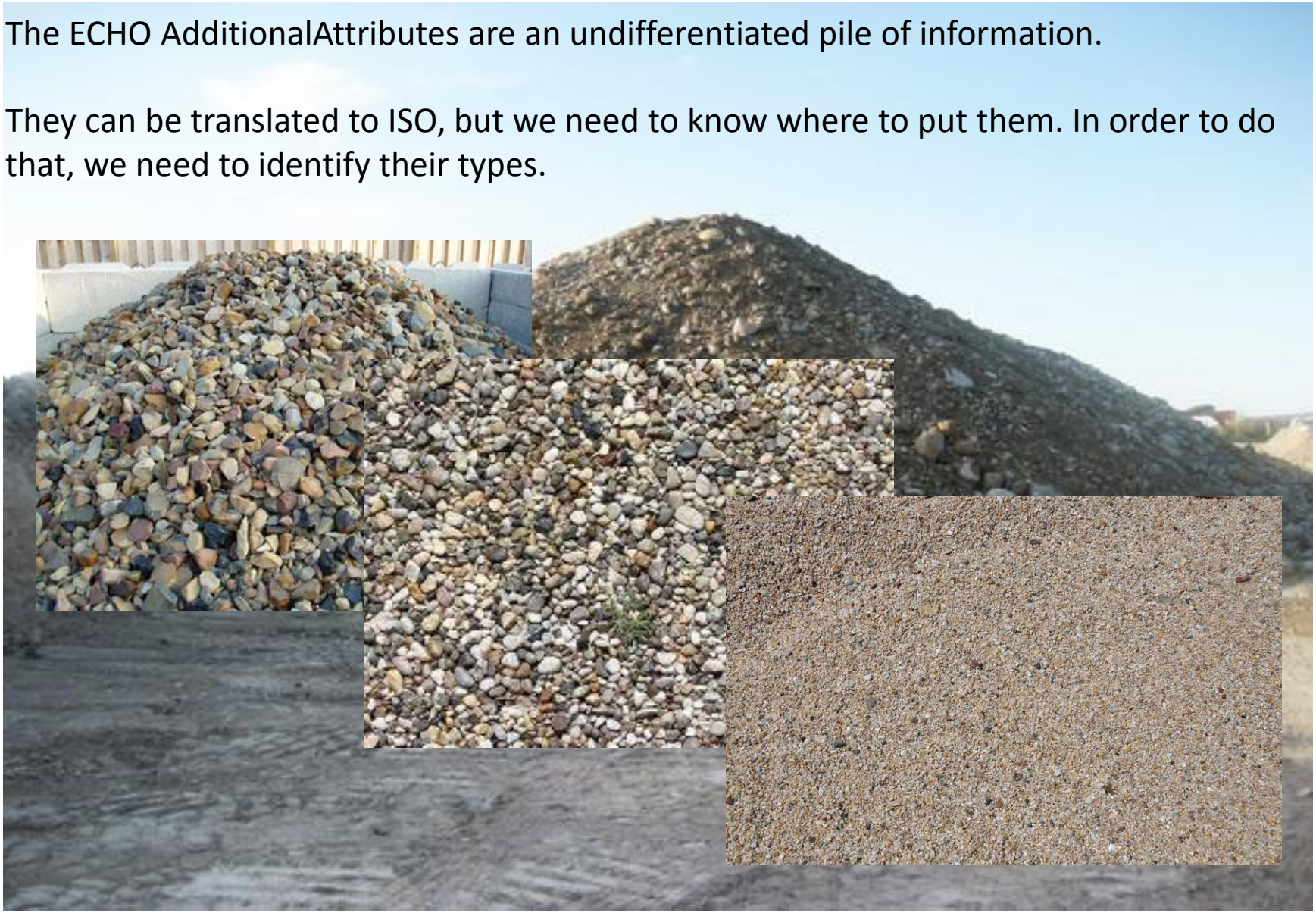
These are different from the programming languages generally used for scientific data processing.

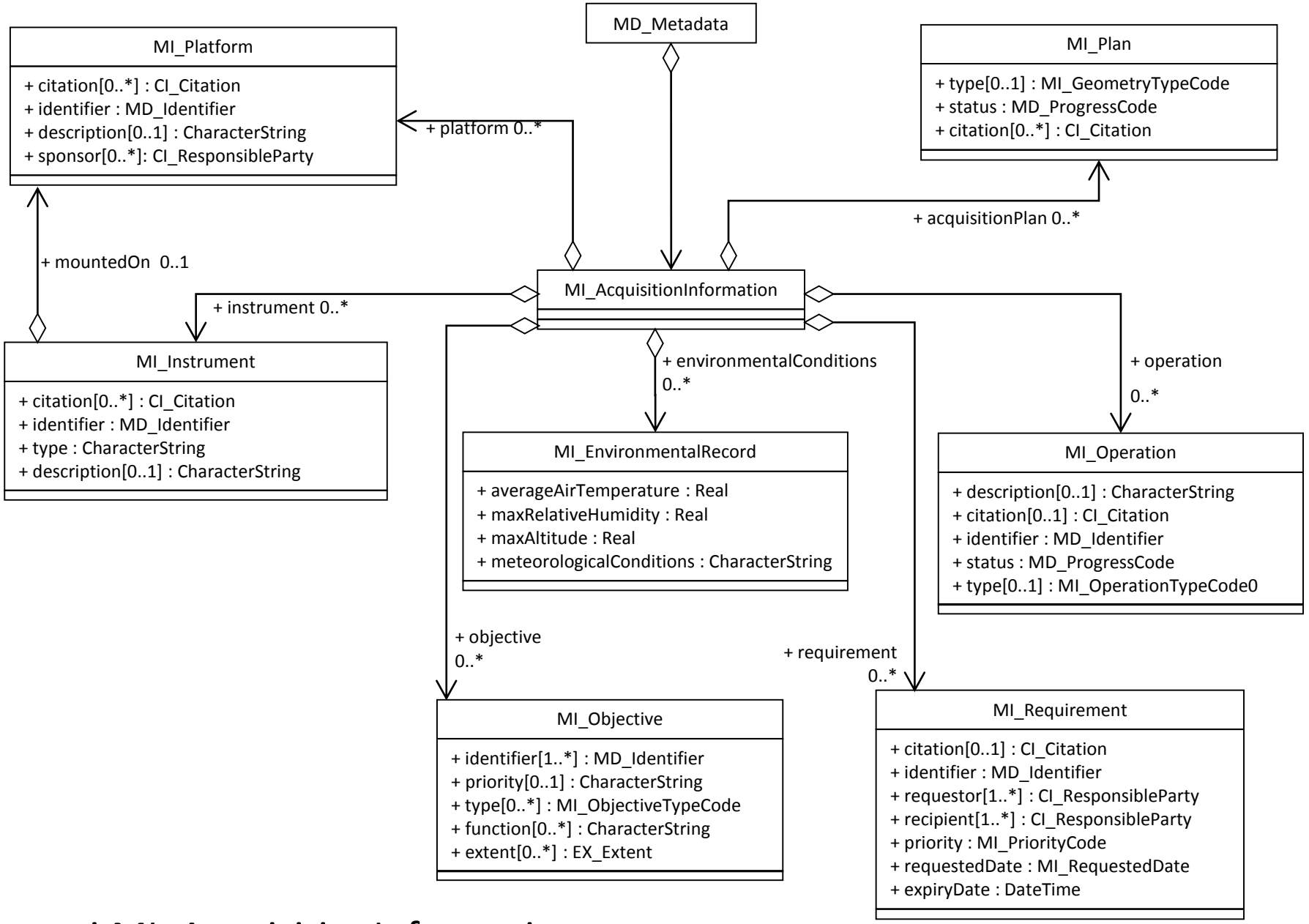


Additional Attributes

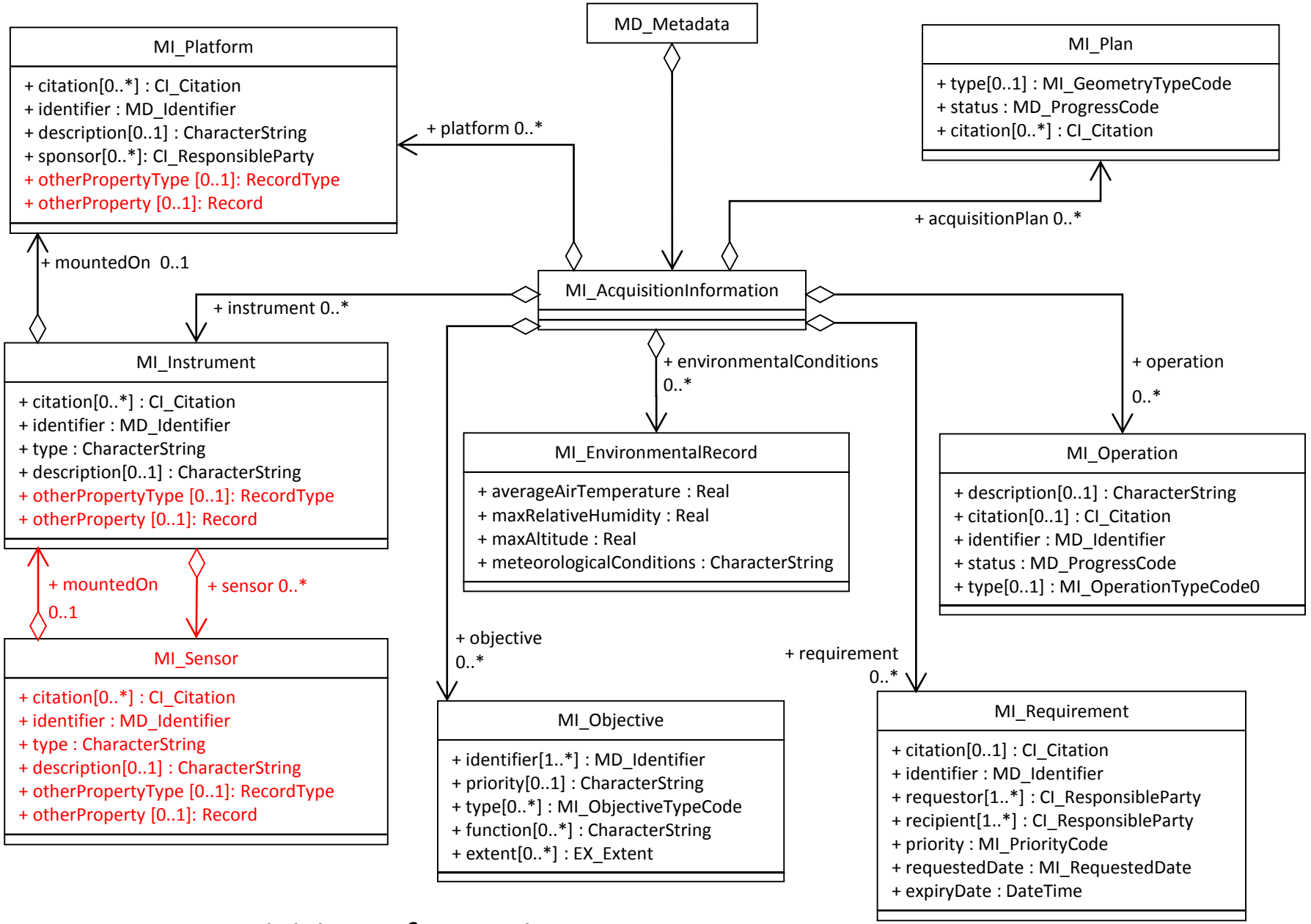
The ECHO AdditionalAttributes are an undifferentiated pile of information.

They can be translated to ISO, but we need to know where to put them. In order to do that, we need to identify their types.

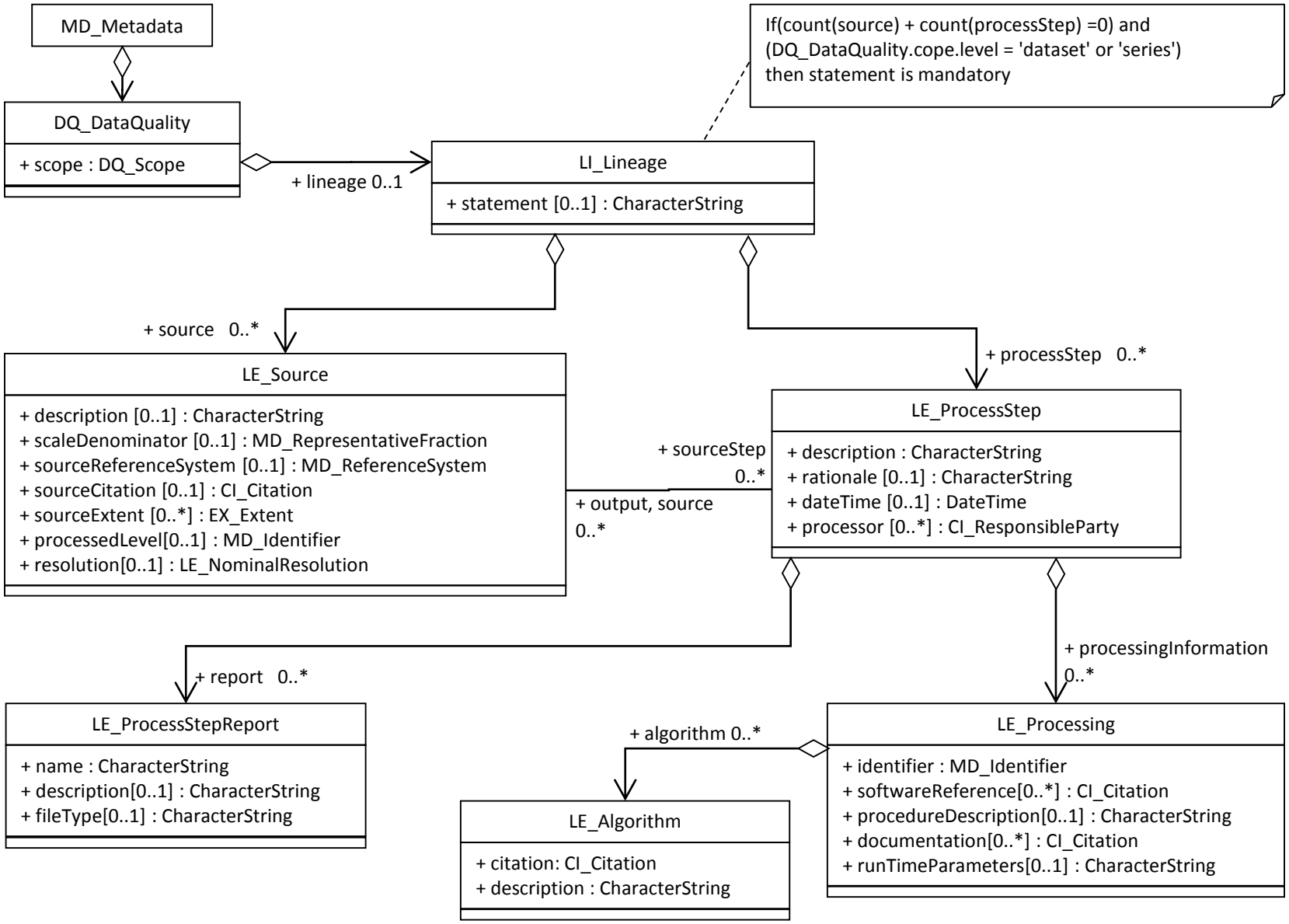




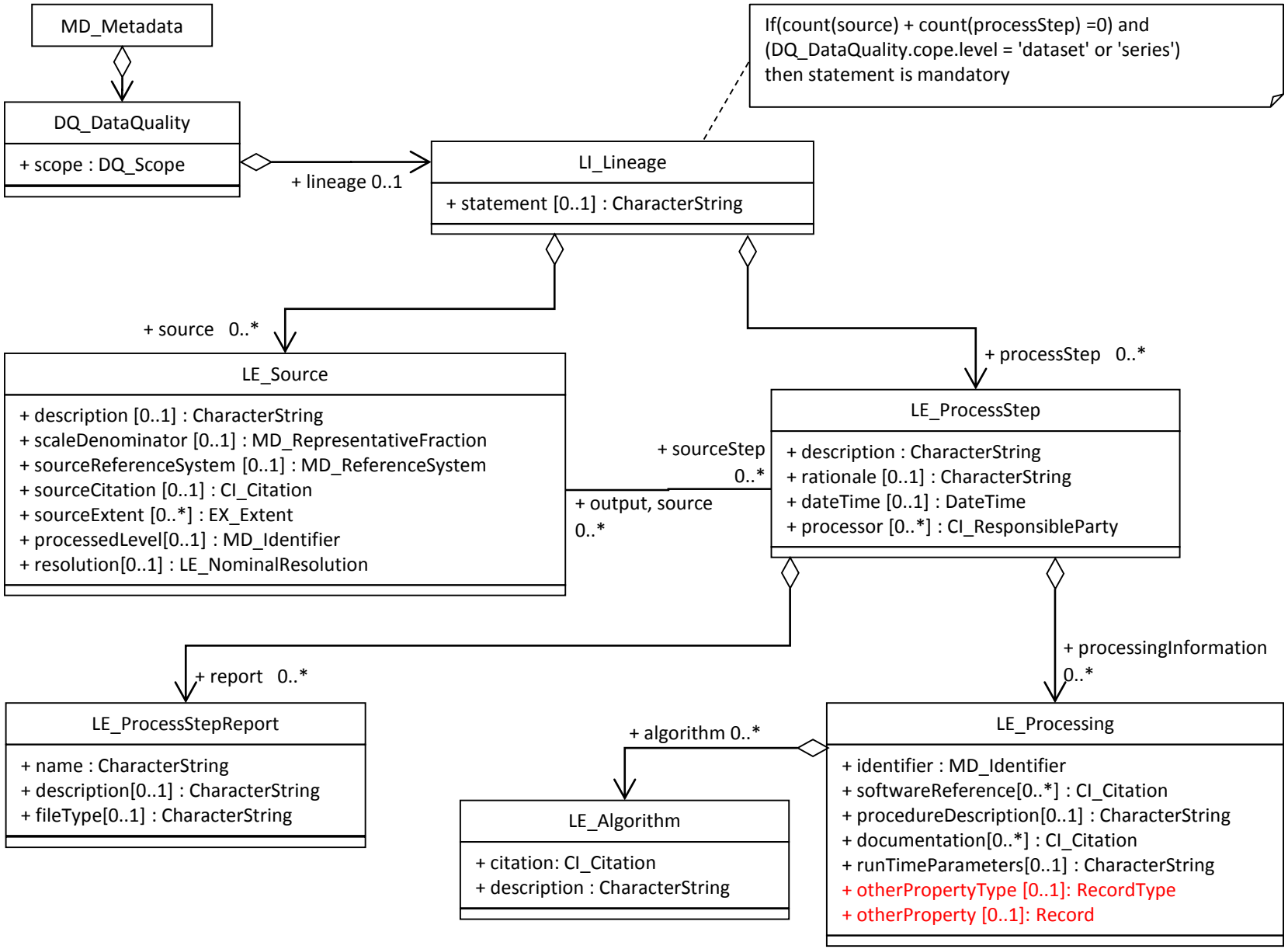
gmi:MI_AcquisitionInformation



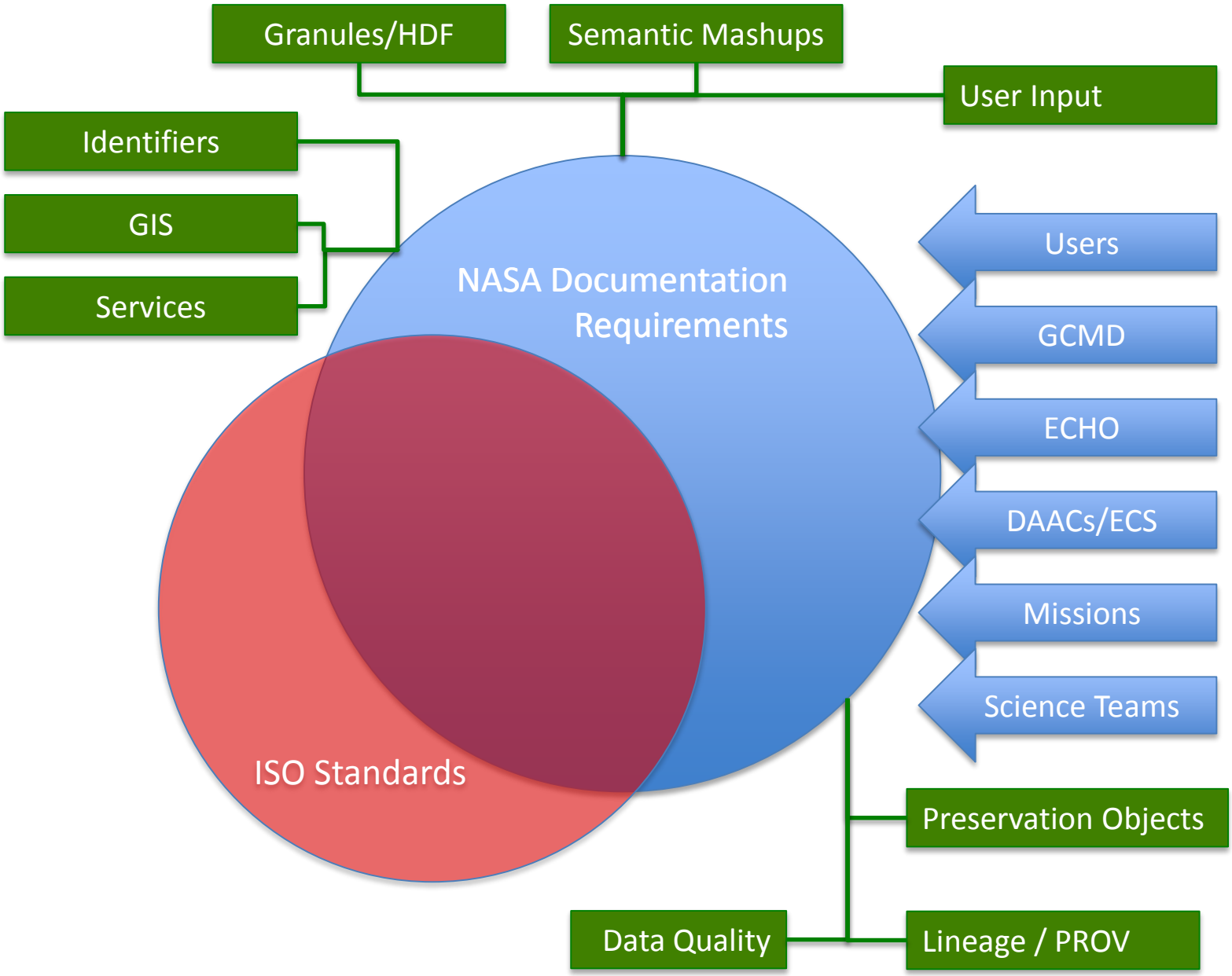
eos:MI_AcquisitionInformation



gmi:DQ_Lineage



eos:DQ_Lineage


Granules/HDF
Semantic Mashups
User Input
Identifiers
GIS
Services
NASA Documentation Requirements
ISO Standards
Users
GCMD
ECHO
DAACs/ECS
Missions
Science Teams
Preservation Objects
Data Quality
Lineage / PROV

Questions?

thabermann@hdfgroup.org