Provenance within Data Interoperability Standards

Status of this RFC

This RFC Technical Note provides information on a proof of concept for transmitting provenance information between the Open Geospatial Consortium (OGC) Web Coverage Service (WCS) and the Open Source Project for a Network Data Access Protocol (OPeNDAP). This RFC does not specify an Earth Science Data Systems (ESDS) standard. Distribution of this memo is unlimited.

This RFC has been deprecated and is published in its current form for informational purposes only.

Change Explanation

This document is not a revision to an earlier version.

Abstract

This document presents a data provenance proof of concept developed for the CEOP Satellite Data Server, a NASA ACCESS project that provides a gateway between the OPeNDAP and WCS protocols. The gateway allows a user to use an OPeNDAP enabled client to access satellite data held at a WCS server. This project was done as part of the Committee on Earth Observation Satellites (CEOS) Working Group and Systems and Services (WGISS) project activity, the WGISS Test Facility for CEOP. A key request by users was to know the provenance (i.e., original data and subsequent processing history) of the results from an OPeNDAP request that were in turn served by a WCS server. This proof of concept presents one way of transmitting that information over the WCS protocol, through the gateway and then over the OPeNDAP protocol.

The provenance prototype we developed demonstrates that provenance can be transmitted to legacy clients, allowing them to make effective use of provenance information for virtual products resulting from workflow chains. Our approach of embedding provenance in data payloads also has the benefit of persisting along with the output product, making data management more tractable. However, it suffers from length and structure limitations imposed by the data payload within which it rides. The ideal solution would marry the richness and scalability benefits of the more elegant out-of-band solution with the usability of the embedded solution.

ESDS-RFC-015v0.0   Lynnes, Yang, Holloway, Enloe and Min
Category: Technical Note (DEPRECATED)   1/12/2009
Updates: n/a   Provenance within Data Interoperability Standards

1   Introduction

This paper describes a proof of concept of data provenance from a NASA ACCESS project to provide a gateway between the Open-source Project for a Network Data Access Protocol (OPeNDAP) and Web Coverage Service (WCS) protocols. The project developed a data handler for the OPeNDAP server that enabled serving of data obtained from a WCS server (Fig 1). The aim was to allow access to the data by analysis clients that have OPeNDAP support but not WCS support such as the Gridded Analysis and Display System (GrADS). The initial gateway implementation allowed a user to access data through an OPeNDAP gateway at one site which was serving data obtained from a WCS server at another site. In turn, that WCS server was often applying processing (such as a Coordinate Reference Service Transformation) to the original data (which may even be from yet another site). It was thus a challenge (or even an impossibility) for the user to investigate the original data or the possible effects of the transformation. In response, the CEOP Satellite Data Server project developed a data provenance proof of concept to present this information to the user.
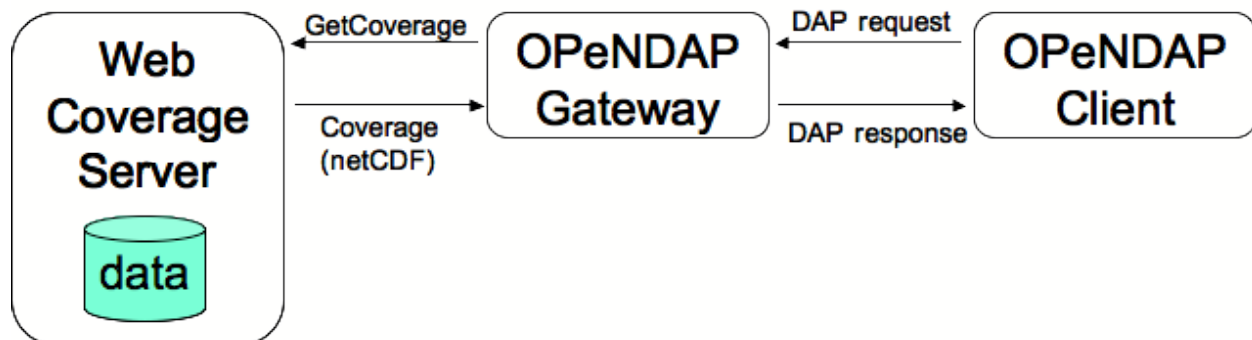

Figure 1. High-level architectural concept for OPeNDAP / WCS Gateway, designed to enable OPeNDAP clients to access data in Web Coverage Servers. The OPeNDAP client submits a DAP request to the gateway, which in turn submits a GetCoverage to the WCS Server. The response is a Coverage with the NetCDF profile.

In the proof of concept, the original data URLs and a description of the processing were stored in the netCDF data package. This was chosen in preference to "out-of-band" measures such as a separate provenance server [1] in order to ensure that the provenance stays with the data. Although less architecturally elegant than the alternative solutions, this principle of embedding the provenance within the data ensures persistence (whereas provenance servers may move or be shut down over time) and also eases the data management burden on the end-user.

As more data become available as service-based virtual data products, the issue of distributed provenance will grow in importance. This technical note does not seek to promote this as the ultimate solution to tracking provenance across a service chain, but rather to point out additional considerations that such a provenance solution should address, such as the ability of the user to access and use the provenance information from within his/her analysis client. However, the embedded provenance approach has its own issues, which are described in this technical note. An ideal solution would combine the usability benefits of the embedded provenance with the richness and scalability of the out-of-band approach.

2   CEOP Satellite Data Server Project Background

The CEOP Satellite Data Server was designed to provide OPeNDAP clients with access to Web Coverage Service servers. It does this through an OPeNDAP Server "format handler" that makes a WCS request to the WCS server. The WCS response is saved as a local file, and then served from cache to the client by the OPeNDAP server. As a result, a client on machine A can access an OPeNDAP server on machine B, which in turn acquires data from a WCS server on machine C where the original data reside. The data are (potentially) modified at each step, though most often at the WCS Server, which reprojects, resamples and subsets the data in order to return the requested coverage. (Further data modifications at the OPeNDAP gateway are possible, but not implemented in the current version of the CEOP Satellite Data Server). Because the original data may be modified, it is often important for the end user to be able to see what modifications have been done and what the original data were. This is the role served by data provenance, also often called lineage. This project was done as part of the Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services (WGISS) project activity, the WGISS Test Facility for CEOP.

3   Provenance Use Case

End users may use provenance for a number of different purposes. In order to make the scope manageable, we begin with a use case that is relatively simple:

(1) The user notices potential anomalies in the data returned by the CEOP Satellite Data Server and wants to determine if they are an artifact of the WCS or OPeNDAP servers.

(2) In order to do that, the user must see what form of processing was done, and must be able to obtain the original files for his/her own comparison.

4   Design Options

4.1   Retrospective vs. Prospective

Many discussions of provenance divide provenance systems into two types: retrospective and prospective. In retrospective provenance, the operations on the data are documented as they occur, often using alternative APIs or toolkits built into the processing system[1]. In prospective provenance, the provenance graph can be predicted for any given request, and so can be generated even for processing that has not yet occurred (e.g., through generation of a directed acyclic graph). For the CEOP Satellite Data Server, which crosses two different legacy systems, prospective provenance is considerably more difficult to achieve; accordingly we have adopted the more easily implemented retrospective mode.

4.2   Embedded vs. Out of Band

One option for transmitting provenance information is directly from server to client through an out of band mechanism (Fig. 2), either through a dedicated provenance protocol/URL, or as an addition to existing protocol standards, such as a hypothetical GetProvenance extension to OGC protocols. This out of band mechanism is flexible in the richness of the provenance information that can be accommodated, and has several architectural advantages[1]. However, it relies on a

persistent framework and system to provide provenance information after the fact. Also, it represents an extra level of complexity on the user's side to retrieve the lineage for a given result. Should the user choose to save it on his/her machine for preservation purposes, a data management problem arises in trying to keep the provenance information with the data.
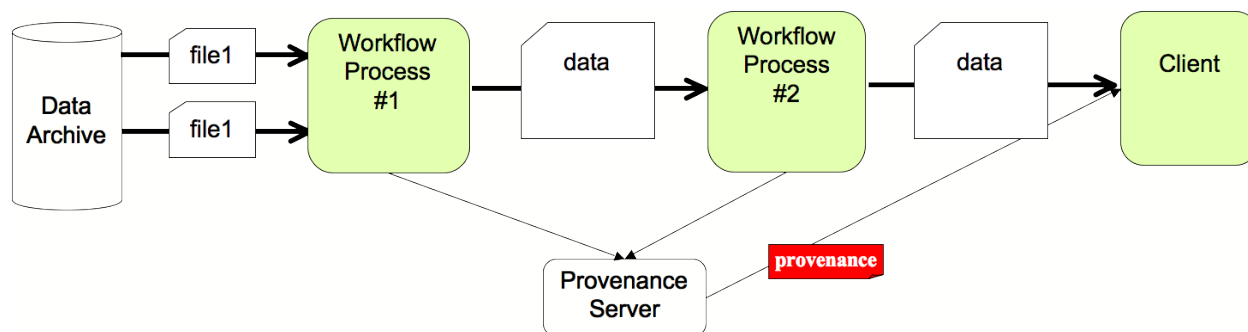

Figure 2. Out-of-band approach to provenance using a separate provenance server. Provenance is acquired by the client as a separate request.

An alternate approach is to embed the provenance within the data payload approach itself. While less flexible, by reusing existing data structures, the provenance is more likely to be accessible to legacy clients.
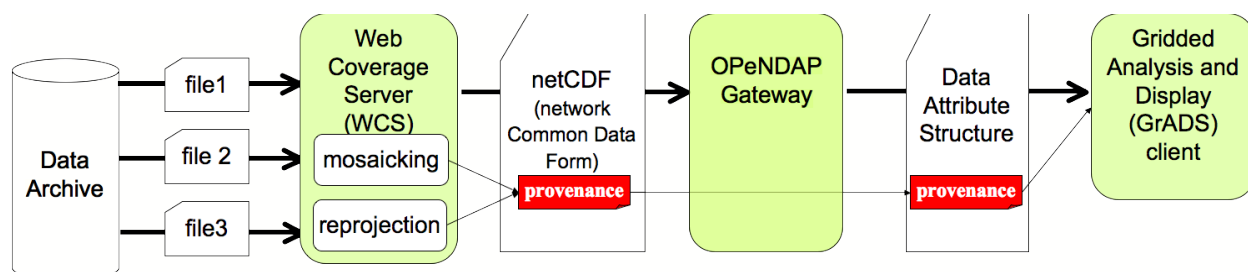

Figure 2. Embedded approach to provenance transmission. The provenance is embedded within the data payload.

As a result, we have chosen the embedded model. In this case, the lineage information is embedded within metadata objects within the data payload. As a result, it always remains with the data, no matter where the user moves them. However, the downside is that this limits both the size and the richness of the provenance metadata that we can include.

4.3   Standards

A number of applicable standards are possible. For instance, the Proof Markup Language has been used in many cases to record provenance information in a semantic web context. However, PML is not especially user friendly in the absence of sophisticated clients. Alternatively, we can use the ISO 19115 standard, which has some rudimentary placeholders for provenance information. The other applicable standard to consider is the netCDF/CF-1 format in which the data are transmitted by the CEOP Satellite Data Server. CF-1 specifies a History element, but

without a specific structure. Accordingly, we have chosen to format the lineage information using the ISO 19115 schema, but include it within the History element. One disadvantage of this with respect to our use case is that the URLs of the original data are deeply buried in the structure. This point has been surfaced with the ISO 19115 editors and may be addressed in the future. Because there is a lack of a single, definitive standard for provenance information, interoperability of provenance information among independent implementations is currently nascent. This lack of a definitive standard also presents a challenge for client developers in using provenance information.

5   Provenance Implementation

5.1   WCS Server

Current lineage information documents three major processing steps involved in the WCS server and the input data sources. In 19115, the processing step is recorded using */MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep/LI_ProcessStep/*. The input data source is recorded using */MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/source/*. The History attribute in the current implementation contains an ASCII string which is the *LI_Lineage* element in the ISO19139, XML Implementation Specification for ISO19115 and 19119. The attribute is formatted as:

```
History = "
<?xml version="1.0" encoding="UTF-8"?>
<lineage xmlns:smXML="http://www.isotc211.org/smXML"
 xmlns:xlink="http://www.w3.org/1999/xlink"
 xmlns:gml="http://www.opengis.net/gml">
<LI_Lineage>
Provenance information including processing steps and data sources
</LI_Lineage>
</lineage>"
```

The three processing steps recorded in the processing steps include 1) swath scan/frame coordinate reference system (CRS) to WGS84 latitude/longitude CRS transformation; 2) interpolation; and 3) per-pixel quality screening. The on-demand mosaicking is not listed as an independent step because it is implicitly included in CRS transformation and can also be reflected in the number of input data sources. For a specific output coverage, if more than one granule file is listed in data sources, mosaicking is performed in producing this coverage.

Input data sources are listed in the *LI_Source* element, i.e., *LI_Lineage/source/ LI_Source*. This element contains information such as input data description, CRS, spatial and temporal extents, responsible party, access protocol and access linkage. The spatial/temporal extents are extracted from input granule file's metadata and the access linkage is the FTP address(es) from where input data file(s) can be downloaded. As previously mentioned, the "linkage" attribute is currently embedded deep in the data source's *CI_OnlineResources* element, as: *LI_Lineage/source/LI_Source/sourceCitation/CI_Citation/citedResponsibleParty/ CI_ResponsibleParty/contactInfo/CI_Contact/onlineResource/CI_OnlineResource/linkage*. This design may be revised in updates of the ISO19115 standard. An example of actual lineage XML is included in Appendix B in this document.

Finally, this history information is then stored in the global attributes (i.e., "NC_GLOBAL") of the netCDF file[2] before transmission via the WCS protocol.

5.2   Provenance within the OPeNDAP Gateway

The OPeNDAP Gateway fetches the data from the WCS Server using the NetCDF profile. However, it does not simply turn around and serve up the data file as is to the end user. Rather it presents parts of the data file to the client as individual variables and attributes. The global attributes, including the History element, are available via the Dataset Attribute Structure (DAS), which most clients query in order to obtain important metadata.

6   Use Case Testing

6.1   GrADS Access

The test for the provenance proof of concept was based on the Use Case outlined above. In order to test the lineage inclusion, we developed a GrADS script that would read the lineage element and output it to a file and an external script to parse the external history file for the original data URLs. (We did not have enough time to add logic to parse the processing history as well, but the use of the external script is general enough to support virtually any parsing.) These scripts would be distributed to the community as an example of how the lineage data might be accessed and then used.

The actual scripts follow: a GrADS script is used to write the attributes to a file, which is then parsed by an external Perl/Python script to extract the URLs of the original data. (Note that the URL used as an example in the GrADS script is subject to change.)

```
'sdfopen http://dev1.opendap.org:8080/opendap/wcs/CEOP/BRA/ceopL2AIRSTimeOffset
Lineage/TSurfAir/2002-10-02'
'q attr'
undef = write('out.txt',result)
```

The following Perl script parses the above output file ("out.txt") to list the URLs of the original files:

```
while (<> ) {
    if (m#^global.*</lineage>#i) {  # Find global lineage attr.
        while (m#<URL>(.*?)</URL>#g) {  # Find URLs in lineage
            print "$1\n";
        }
    }
}
```

with a result in this case of:

```
ftp://airspar1u.ecs.nasa.gov/data/s4pa/Aqua_AIRS_Level2/AIRX2RET.005/2002/275/
AIRS.2002.10.02.041.L2.RetStd.v5.0.14.0.G07228184728.hdf
ftp://airspar1u.ecs.nasa.gov/data/s4pa/Aqua_AIRS_Level2/AIRX2RET.005/2002/275/
AIRS.2002.10.02.042.L2.RetStd.v5.0.14.0.G07228184855.hdf
ftp://airspar1u.ecs.nasa.gov/data/s4pa/Aqua_AIRS_Level2/AIRX2RET.005/2002/275/
AIRS.2002.10.02.164.L2.RetStd.v5.0.14.0.G07229011310.hdf
```

This allows the user to easily acquire the original input data.

6.2   THREDDS Aggregation and Provenance

Ideally, CEOP users often prefer to obtain a time series of data around a reference site, without having to write a GrADS script to loop through individual time slices. This functionality can be provided through a THREDDS aggregation server. However, although the THREDDS aggregation server can provide a time-ordered aggregation of the data, it assumes that all of the History elements are the same, and so does not aggregate those attributes at the same time. This is an object lesson in the potential difficulties in propagating provenance information through a complex chain of servers. Ideally, we would want the THREDDS aggregation server to aggregate those elements of the History element that are different for the input components, in this case, the source URLs.

7   Disadvantages of the Embedded Provenance Approach

In the course of our prototype, we ran into two key disadvantages of embedding provenance within the netCDF file. The first difficulty encountered was the fact that neither the netCDF attribute structure nor the CF-1 convention have an existing structure for recording provenance information. While there is a History attribute, which we ultimately used, this is a simple text string. We chose to use XML markup following the ISO 19115 standard within this text string, but this is essentially an arbitrary choice.

The second issue is the danger of overrunning the attribute size. We consciously avoided this problem by limiting the provenance information we embedded in the attributes, and further by limiting the scope of the WCS requests themselves through our OPeNDAP gateway configuration. However, it is conceivable that WCS requests touching many more data files could cause the attribute to overflow, particularly if more detailed information about the processing were recorded.

8   Conclusion

The field of provenance is currently quite active. However, much of the current emphasis is on the integration of the provenance capture with the workflow mechanisms, e.g., [2]. In this proof of concept, our emphasis has been on end-user usability. This necessitates viewing the problem from the point of view of the anticipated client applications, many of which are limited in their ability to interact with provenance systems. In order to circumvent this constraint, we have dynamically embedded the provenance information within metadata that are already accessed by the clients. This has the added virtue of keeping the provenance information for a particular output file with the file itself, so that any subsequent movement or transfer of the file will entrain the provenance information. Of course, this somewhat simplistic treatment does not take full advantage of the progress in provenance and workflow systems. Furthermore, there is no standard (yet) for representing the provenance within files, and the file formats may impose limits on the size of provenance attributes, leading to a scalability issue. An ideal solution might marry these more advanced approaches, such as provenance servers, with the embedding approach that we have demonstrated here.

9   References

[1]   Bose R. and J. Frew, (2005) "Lineage Retrieval for Scientific Data Processing: A Survey," ACM Computing Surveys, Vol. 37, No. 1, March 2005, pp. 1–28.

[2]   Moreau, L., Groth, P., Miles, S., Vazquez-Salceda, J., Ibbotson, J., Jiang, S., Munroe, S., Rana, O., Schreiber, A., Tan, V. and L. Varga, (2008) "The Provenance of Electronic Data", Communications of the ACM, vol. 51, No. 4, April 2008, p. 52-58.

10   Authors

The CEOP Satellite Data Server Project Team:

Christopher Lynnes, NASA GSFC, chris.lynnes@nasa.gov

Wenli Yang, George Mason University, wyang1@gmu.edu

Daniel Holloway, OPeNDAP.org, d.holloway@opendap.org

Yonsook Enloe, SGT, yonsook@mindspring

Min Min, George Mason University, mmin1@gmu.edu

Appendix A - Glossary

ACCESS - Advancing Collaborative Connections for Earth System Science

API – Application Programming Interface

CEOP – Coordinated Energy and Water Cycle Observations Project

CEOS – Committee on Earth Observation Satellites

CF-1 – Climate and Forecast convention 1

CRS – Coordinate Reference System

DAS – Dataset Attribute Structure

ESDS – Earth Science Data Systems

FTP – File Transfer Protocol

GES DISC – Goddard Earth Sciences Data and Information Services Center

GrADS – Gridded Analysis and Display System

NetCDF – Network Common Data Form

OGC – Open Geospatial Consortium

OPeNDAP – Open-source Project for a Network Data Access Protocol

PML – Proof Markup Language

THREDDS – Thematic Realtime Environmental Distributed Data Services

URL – Universal Reference Locator

WCS – Web Coverage Service

WGISS – Working Group on Information Systems and Services

XML – eXtensible Markup Language

ESDS-RFC-015v0.0
Category: Technical Note (DEPRECATED)
Updates: n/a

Lynnes, Yang, Holloway, Enloe and Min
1/12/2009
Provenance within Data Interoperability Standards

Appendix B - Example of the History Attribute in CF-1.0 NetCDF File

```xml
<?xml version=\"1.0\" encoding=\"UTF-8\"?>
<lineage xmlns:smXML=\"http://www.isotc211.org/smXML\"
 xmlns:xlink=\"http://www.w3.org/1999/xlink\"
 xmlns:gml=\"http://www.opengis.net/gml\">
 <LI_Lineage>
  <statement>
   <CharacterString>
    This lineage information documents the source and processing performed by the WCS server
in the CEOP satellite data server.
   </CharacterString>
  </statement>
  <processStep>
   <LI_ProcessStep>
    <description>
     <CharacterString>
      <Reprojection>
       <Name>SatelliteSwathToLatitudeLongitude</Name>
       <Method>ForwardMapping</Method>
      </Reprojection>
     </CharacterString>
    </description>
    <rationale>
     <CharacterString>
     Perform georectification by transforming satellite swath coordinate system to Latitude-
Longitude coordinate system.
     </CharacterString>
    </rationale>
    <dateTime>
     <DateTime>2008-12-29T17:59:02Z</DateTime>
    </dateTime>
    <processor>
     <CI_ResponsibleParty>
      <individualName>
       <CharacterString>Wenli Yang</CharacterString>
      </individualName>
      <organisationName>
       <CharacterString>NASA GES DISC</CharacterString>
      </organisationName>
      <contactInfo>
       <CI_Contact>
        <phone>
         <CI_Telephone>
          <voice>
           <CharacterString>301-614-5312</CharacterString>
          </voice>
         </CI_Telephone>
        </phone>
        <electronicMailAddress>
         <CharacterString>Wenli.Yang@nasa.gov</CharacterString>
        </electronicMailAddress>
       </CI_Contact>
      </contactInfo>
     </CI_ResponsibleParty>
    </processor>
   </LI_ProcessStep>
   <LI_ProcessStep>
    <description>
     <CharacterString>
      <Interpolation>
```

ESDS-RFC-015v0.0
Category: Technical Note (DEPRECATED)
Updates: n/a

Lynnes, Yang, Holloway, Enloe and Min
1/12/2009
Provenance within Data Interoperability Standards

```
    <Name>Interpolation</Name>
    <Method> NearestNeighbor Interpolation </Method>
   </Interpolation>
  </CharacterString>
 </description>
 <rationale>
  <CharacterString>Interpolatiion is needed for output cells containing no valid input
values.</CharacterString>
 </rationale>
<dateTime>
 <DateTime>2008-12-29T17:59:02Z</DateTime>
</dateTime>
<processor>
 <CI_ResponsibleParty>
  <individualName>
   <CharacterString>Wenli Yang</CharacterString>
  </individualName>
  <organisationName>
   <CharacterString>NASA GES DISC</CharacterString>
  </organisationName>
  <contactInfo>
   <CI_Contact>
    <phone>
     <CI_Telephone>
      <voice>
       <CharacterString>301-614-5312</CharacterString>
      </voice>
     </CI_Telephone>
    </phone>
    <electronicMailAddress>
     <CharacterString>Wenli.Yang@nasa.gov</CharacterString>
    </electronicMailAddress>
   </CI_Contact>
  </contactInfo>
 </CI_ResponsibleParty>
</processor>
</LI_ProcessStep>
<LI_ProcessStep>
 <description>
  <CharacterString>
   <QualityScreen>
    <Name>QualityScreen</Name>
    <Method>AIRSPerPixelQAFiltering</Method>
   </QualityScreen>
   </CharacterString>
 </description>
 <rationale>
  <CharacterString>Input cells not meeting QA criteria must be excluded in the resultant
coverage.</CharacterString>
 </rationale>
<dateTime>
 <DateTime>2008-12-29T17:59:02Z</DateTime>
</dateTime>
<processor>
 <CI_ResponsibleParty>
  <individualName>
   <CharacterString>Wenli Yang</CharacterString>
  </individualName>
  <organisationName>
   <CharacterString>NASA GES DISC</CharacterString>
  </organisationName>
  <contactInfo>
```

ESDS-RFC-015v0.0   Lynnes, Yang, Holloway, Enloe and Min
Category: Technical Note (DEPRECATED)   1/12/2009
Updates: n/a   Provenance within Data Interoperability Standards

```
    <CI_Contact>
     <phone>
      <CI_Telephone>
       <voice>
        <CharacterString>301-614-5312</CharacterString>
       </voice>
      </CI_Telephone>
     </phone>
     <electronicMailAddress>
      <CharacterString>Wenli.Yang@nasa.gov</CharacterString>
     </electronicMailAddress>
    </CI_Contact>
   </contactInfo>
  </CI_ResponsibleParty>
 </processor>
</LI_ProcessStep>
</processStep>
<source>
 <LI_Source>
  <description>
   <CharacterString>Input AIRS Release 5 Level 2 granule</CharacterString>
  </description>
  <sourceReferenceSystem>
   <MD_ReferenceSystem>
    <referenceSystemIdentifier>
     <RS_Identifier>
      <code>
       <CharacterString>OGC:Swath</CharacterString>
      </code>
      <codeSpace>
       <CharacterString>OGC</CharacterString>
      </codeSpace>
     </RS_Identifier>
    </referenceSystemIdentifier>
   </MD_ReferenceSystem>
  </sourceReferenceSystem>
  <sourceCitation>
   <CI_Citation>
    <title>
     <CharacterString>Input AIRS Release 5 Level 2 granule</CharacterString>
    </title>
    <date>
     <CI_Date>
      <date>
       <Date>2002-10-02</Date>
      </date>
      <dateType>
       <CI_DateTypeCode>publication</CI_DateTypeCode>
      </dateType>
     </CI_Date>
    </date>
    <identifier>
     <MD_Identifier>
      <authority />
      <code>
       <CharacterString>AIRX2RET</CharacterString>
      </code>
     </MD_Identifier>
    </identifier>
    <citedResponsibleParty>
     <CI_ResponsibleParty>
      <organisationName>
```

```xml
                  <CharacterString>NASA GES DISC</CharacterString>
              </organisationName>
              <contactInfo>
               <CI_Contact>
                <onlineResource>
                 <CI_OnlineResource>
                  <linkage>

 <URL>ftp://airspar1u.ecs.nasa.gov/data/s4pa/Aqua_AIRS_Level2/AIRX2RET.005/2002/275/AIRS.2002.
10.02.041.L2.RetStd.v5.0.14.0.G07228184728.hdf</URL>
                  </linkage>
                  <protocol>
                   <CharacterString>ftp</CharacterString>
                  </protocol>
                  <function>
                   <CI_OnLineFunctionCode>download</CI_OnLineFunctionCode>
                  </function>
                 </CI_OnlineResource>
                </onlineResource>
               </CI_Contact>
              </contactInfo>
              <role>
               <CI_RoleCode>originator</CI_RoleCode>
              </role>
            </CI_ResponsibleParty>
          </citedResponsibleParty>
         <presentationForm>
          <CI_PresentationFormCode codeSpace=\"TBD\" codeList=\"TBD\" codeListValue=\"TBD\" />
         </presentationForm>
       </CI_Citation>
      </sourceCitation>
      <sourceExtent>
       <EX_Extent>
        <description>
         <CharacterString>Granule Bounding Rectangle</CharacterString>
        </description>
        <geographicElement>
         <EX_BoundingPolygon>
          <polygon>
           <gml:Polygon srsName=\"urn:EPSG:geographicCRS:4326\">
            <gml:outerBoundaryIs>
             <gml:LinearRing>
              <gml:pos>-49.826756 5.287550</gml:pos>
              <gml:pos>-29.903366 5.287550</gml:pos>
              <gml:pos>-29.903366 -17.897720</gml:pos>
              <gml:pos>-49.826756 -17.897720</gml:pos>
             </gml:LinearRing>
            </gml:outerBoundaryIs>
           </gml:Polygon>
          </polygon>
         </EX_BoundingPolygon>
        </geographicElement>
        <temporalElement>
         <EX_TemporalExtent>
          <extent>
           <gml:TimePeriod>
            <gml:begin>
             <gml:TimeInstant>
              <gml:timePosition>2002-10-02T04:05:26.000000Z</gml:timePosition>
             </gml:TimeInstant>
            </gml:begin>
            <gml:end>
```

```xml
           <gml:TimeInstant>
            <gml:timePosition>2002-10-02T04:11:26.000000Z</gml:timePosition>
           </gml:TimeInstant>
          </gml:end>
         </gml:TimePeriod>
        </extent>
       </EX_TemporalExtent>
      </temporalElement>
     </EX_Extent>
    </sourceExtent>
   </LI_Source>
   <LI_Source>
    <description>
     <CharacterString>Input AIRS Release 5 Level 2 granule</CharacterString>
    </description>
    <sourceReferenceSystem>
     <MD_ReferenceSystem>
      <referenceSystemIdentifier>
       <RS_Identifier>
        <code>
         <CharacterString>OGC:Swath</CharacterString>
        </code>
        <codeSpace>
         <CharacterString>OGC</CharacterString>
        </codeSpace>
       </RS_Identifier>
      </referenceSystemIdentifier>
     </MD_ReferenceSystem>
    </sourceReferenceSystem>
    <sourceCitation>
     <CI_Citation>
      <title>
       <CharacterString>Input AIRS Release 5 Level 2 granule</CharacterString>
      </title>
      <date>
       <CI_Date>
        <date>
         <Date>2002-10-02</Date>
        </date>
        <dateType>
         <CI_DateTypeCode>publication</CI_DateTypeCode>
        </dateType>
       </CI_Date>
      </date>
      <identifier>
       <MD_Identifier>
        <authority />
        <code>
         <CharacterString>AIRX2RET</CharacterString>
        </code>
       </MD_Identifier>
      </identifier>
      <citedResponsibleParty>
       <CI_ResponsibleParty>
        <organisationName>
         <CharacterString>NASA GES DISC</CharacterString>
        </organisationName>
        <contactInfo>
         <CI_Contact>
          <onlineResource>
           <CI_OnlineResource>
            <linkage>
```

```
<URL>ftp://airspar1u.ecs.nasa.gov/data/s4pa/Aqua_AIRS_Level2/AIRX2RET.005/2002/275/AIRS.2002.
10.02.042.L2.RetStd.v5.0.14.0.G07228184855.hdf</URL>
          </linkage>
          <protocol>
           <CharacterString>ftp</CharacterString>
          </protocol>
          <function>
           <CI_OnLineFunctionCode>download</CI_OnLineFunctionCode>
          </function>
         </CI_OnlineResource>
        </onlineResource>
       </CI_Contact>
      </contactInfo>
      <role>
       <CI_RoleCode>originator</CI_RoleCode>
      </role>
     </CI_ResponsibleParty>
    </citedResponsibleParty>
   <presentationForm>
    <CI_PresentationFormCode codeSpace=\"TBD\" codeList=\"TBD\" codeListValue=\"TBD\" />
   </presentationForm>
  </CI_Citation>
 </sourceCitation>
 <sourceExtent>
  <EX_Extent>
   <description>
    <CharacterString>Granule Bounding Rectangle</CharacterString>
   </description>
   <geographicElement>
    <EX_BoundingPolygon>
     <polygon>
      <gml:Polygon srsName=\"urn:EPSG:geographicCRS:4326\">
       <gml:outerBoundaryIs>
        <gml:LinearRing>
         <gml:pos>-57.038685 -16.227081</gml:pos>
         <gml:pos>-34.146679 -16.227081</gml:pos>
         <gml:pos>-34.146679 -39.524109</gml:pos>
         <gml:pos>-57.038685 -39.524109</gml:pos>
        </gml:LinearRing>
       </gml:outerBoundaryIs>
      </gml:Polygon>
     </polygon>
    </EX_BoundingPolygon>
   </geographicElement>
   <temporalElement>
    <EX_TemporalExtent>
     <extent>
      <gml:TimePeriod>
       <gml:begin>
        <gml:TimeInstant>
         <gml:timePosition>2002-10-02T04:11:26.000000Z</gml:timePosition>
        </gml:TimeInstant>
       </gml:begin>
       <gml:end>
        <gml:TimeInstant>
         <gml:timePosition>2002-10-02T04:17:26.000000Z</gml:timePosition>
        </gml:TimeInstant>
       </gml:end>
      </gml:TimePeriod>
     </extent>
    </EX_TemporalExtent>
```

ESDS-RFC-015v0.0
Category: Technical Note (DEPRECATED)
Updates: n/a

Lynnes, Yang, Holloway, Enloe and Min
1/12/2009
Provenance within Data Interoperability Standards

```xml
        </temporalElement>
       </EX_Extent>
      </sourceExtent>
     </LI_Source>
     <LI_Source>
      <description>
       <CharacterString>Input AIRS Release 5 Level 2 granule</CharacterString>
      </description>
      <sourceReferenceSystem>
       <MD_ReferenceSystem>
        <referenceSystemIdentifier>
         <RS_Identifier>
          <code>
           <CharacterString>OGC:Swath</CharacterString>
          </code>
          <codeSpace>
           <CharacterString>OGC</CharacterString>
          </codeSpace>
         </RS_Identifier>
        </referenceSystemIdentifier>
       </MD_ReferenceSystem>
      </sourceReferenceSystem>
      <sourceCitation>
       <CI_Citation>
        <title>
         <CharacterString>Input AIRS Release 5 Level 2 granule</CharacterString>
        </title>
        <date>
         <CI_Date>
          <date>
           <Date>2002-10-02</Date>
          </date>
          <dateType>
           <CI_DateTypeCode>publication</CI_DateTypeCode>
          </dateType>
         </CI_Date>
        </date>
        <identifier>
         <MD_Identifier>
          <authority />
          <code>
           <CharacterString>AIRX2RET</CharacterString>
          </code>
         </MD_Identifier>
        </identifier>
        <citedResponsibleParty>
         <CI_ResponsibleParty>
          <organisationName>
           <CharacterString>NASA GES DISC</CharacterString>
          </organisationName>
          <contactInfo>
           <CI_Contact>
            <onlineResource>
             <CI_OnlineResource>
              <linkage>

<URL>ftp://airspar1u.ecs.nasa.gov/data/s4pa/Aqua_AIRS_Level2/AIRX2RET.005/2002/275/AIRS.2002.
10.02.164.L2.RetStd.v5.0.14.0.G07229011310.hdf</URL>
              </linkage>
              <protocol>
               <CharacterString>ftp</CharacterString>
              </protocol>
```

```xml
              <function>
               <CI_OnLineFunctionCode>download</CI_OnLineFunctionCode>
               </function>
              </CI_OnlineResource>
             </onlineResource>
            </CI_Contact>
           </contactInfo>
           <role>
            <CI_RoleCode>originator</CI_RoleCode>
           </role>
          </CI_ResponsibleParty>
         </citedResponsibleParty>
        <presentationForm>
         <CI_PresentationFormCode codeSpace=\"TBD\" codeList=\"TBD\" codeListValue=\"TBD\" />
        </presentationForm>
       </CI_Citation>
      </sourceCitation>
      <sourceExtent>
       <EX_Extent>
        <description>
         <CharacterString>Granule Bounding Rectangle</CharacterString>
        </description>
        <geographicElement>
         <EX_BoundingPolygon>
          <polygon>
           <gml:Polygon srsName=\"urn:EPSG:geographicCRS:4326\">
            <gml:outerBoundaryIs>
             <gml:LinearRing>
              <gml:pos>-52.027290 4.617317</gml:pos>
              <gml:pos>-32.070404 4.617317</gml:pos>
              <gml:pos>-32.070404 -18.795895</gml:pos>
              <gml:pos>-52.027290 -18.795895</gml:pos>
             </gml:LinearRing>
            </gml:outerBoundaryIs>
           </gml:Polygon>
          </polygon>
         </EX_BoundingPolygon>
        </geographicElement>
        <temporalElement>
         <EX_TemporalExtent>
          <extent>
           <gml:TimePeriod>
            <gml:begin>
             <gml:TimeInstant>
              <gml:timePosition>2002-10-02T16:23:26.000000Z</gml:timePosition>
             </gml:TimeInstant>
            </gml:begin>
            <gml:end>
             <gml:TimeInstant>
              <gml:timePosition>2002-10-02T16:29:26.000000Z</gml:timePosition>
             </gml:TimeInstant>
            </gml:end>
           </gml:TimePeriod>
          </extent>
         </EX_TemporalExtent>
        </temporalElement>
       </EX_Extent>
      </sourceExtent>
     </LI_Source>
    </source>
   </LI_Lineage>
  </lineage>
```