

Review of Mapping HDF5 to DAP2 - Technical Note

The goals of the Standards Process Group (SPG) of NASA's Earth Science Data Systems Working Groups are to:

1. Enable data and service providers to easily join NASA's Earth Science network of data systems through use of standards.
2. Facilitate interoperability between components of NASA's Earth Science network of data systems through use of standards.
3. Facilitate data stewardship and preservation through use of standards and adoption of best practices.
4. Develop and manage effective standards recommendation, adoption, and approval processes to guide the evolution of ESDS standards. Support the evolving strategies and goals of NASA's Earth Science activities through use of standards.

One of the ways we do this is by publishing Technical Notes relevant to Earth Science Data Systems. An SPG Technical Note is a document that contains important and useful information that is relevant to the domain of NASA Earth Science Data Systems, and does not necessarily describe a "standard" which has additional operational requirements. In order to assure a high level of technical quality, we conduct public reviews of Technical Notes that have been submitted to us for consideration.

We are asking you to review the Technical Note referenced below. Your assistance will help us to decide whether each it should be endorsed by the SPG.

- ESDS-RFC-017 - Mapping HDF5 to DAP2

<http://www.esdswg.net/spg/rfc/esds-rfc-017/ESDS-RFC-017v0.1.pdf>

You are invited to review this technical note and provide feedback that might make the document more useful (see review questions below).

1. Please provide your name, organization and contact information including e-mail address. (*This information will not be shared.*)
2. Are you answering for your entire organization, for a smaller group, or individually?
 a) Entire organization
 b) Smaller group (please specify) _____
 c) Individual response

3. Describe in a sentence or two your overall experience related to HDF5 or DAP:

(e.g., specification developer, specification implementer, systems architecture; tools developer, scientific analysis; etc.)

I am currently a netCDF developer responsible for the integration of the DAP protocol client into the netCDF library. This also entails defining the conversion of DAP to both the netCDF-3 (classic model) and the netCDF-4 model, which is closely patterned to HDF5.

4. Do any of your systems currently use (or expect to use) HDF5 or DAP?

Yes

5. Does the technical note contain internal inconsistencies? If so, please provide details.

No obvious inconsistencies.

6. Are any parts of any of the technical note ambiguous or poorly explained? If so, please provide details.

I am not sure if this is appropriate for this kind of technical note, but I would have been happier if the authors had discussed alternative mappings and why they chose the ones in this report.

Inherently, prose is ambiguous. I would suggest for the future (not this report), that you consider having such mapping documents be accompanied with an operational definition of some sort. By this I mean a program in some high level language that can carry out the mapping. For this report, for example, it would be appropriate to have a program that takes an HDF5 schema as input and produces the corresponding DAP DDS.

7. Did you find the technical note useful and would you like to see more such technical notes processed by the SPG?

Yes.

8. Should the SPG endorse this document as a Technical Note – why or why not?

Note: The SPG has already endorsed DAP2 and HDF 5 as standards. This question pertains strictly to whether this proposed Technical Note should be endorsed.

It would be ok to endorse it, although the comments attached might suggest some desirable modifications.

Additional technical comments.

This mapping makes very little use of attributes to help the translation. Given that DAP has the ability to represent nested attributes, it seems a pity to not exploit them to provide significant extra information as part of the translation.

It might be appropriate to note in the report that the chosen naming scheme for encoding groups disallows cyclic group containment. It effectively also converts acyclic group graphs with shared subgroups into a group tree.

The heart of the mapping is described in section 3.1. I am bound to say that I disagree with the list of unmappable HDF5 types; the following mapping would seem to be doable with not too much effort.

- Opaque – the obvious mapping of an opaque of size N is to a byte array of size N.
- Time – I realize that the external formats for time are bewilderingly complicated, but for the purposes of mapping, I would have thought that mapping to one fixed format would work just fine; am I missing something?
- Enumeration – enumerations can be converted to the corresponding integer value (of whatever size integer type) and attributes can be used to define the mapping of enumeration names to the integer values. Of course this can be cumbersome if one has a very large number of enumeration constants, but it is still better than not having enumerations at all.
- Array – I must confess I am confused about this. The comment is made in the document that “... the Array data type in HDF5 is not equivalent to the DAP2 Array type...” I do not understand why this is so and it seems to me the authors need to elaborate on this.
- Variable Length – in many cases, variable length data can be mapped to DAP2 Sequences. The issue here is the degree to which the translation must be “comprehensive”. Is it better to do no mapping at all or to do mapping when possible, even if not all mappings are possible.

Section 3.2.2.2 on object and region references provides a usable mapping using URLs. However, there is another possible mapping that the author’s might consider. It should be possible to map and graft the schema of the object being pointed to into the DAP2 translation.