

Some comments about "HDF5 Groups and netCDF-4"

Mike Folk

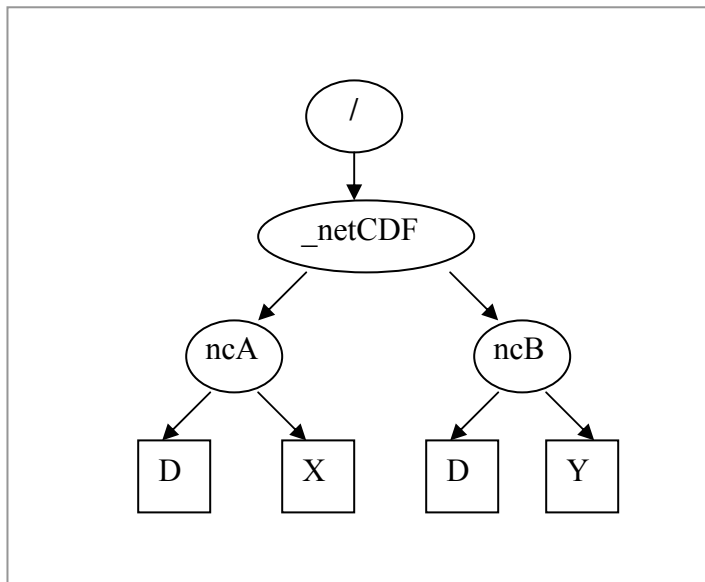
June 17, 2004

1. The paragraph that begins "Backward compatibility requires that each netCDF object have a single distinguished name, rather than multiple equivalent aliases...". I like this capability, not just for backward compatibility with netCDF 3, but because many other applications also like to have a single, unambiguous name. In our early thinking about HDF5, we said that each object would have an unambiguous name, but that never got implemented. Links got confused with names, and that confusion still persists, as you describe.

Two possible solutions to this would be either to create a convention that a certain attribute with a reserved name be used for specifying a name for groups and datasets. Alternatively, this could be a message in the header. The latter would require a format change. I was planning to recommend that we do this as part of the dimension scales implementation project, but don't know how it will be received.

2. Re recommendation #2. Didn't understand. How does it know that "x" is in G2 which is in G1? Wouldn't you need to call it "G1/G2/x"?

Also, won't netCDF applications typically expect a just file name, and assume that there's just one set of variables. What happens if a netCDF application opens a file with a hierarchy such as in the diagram? How does it know whether to use ncA or ncB? I guess there could be some sort of default, or perhaps it could return an error.



3. Re recommendation #3 – using a strict hierarchy. I'm sure you gave this a lot of thought. The only problem, which I imagine you know, is that this means that you can't use HDF5 links to have a group be shared by more than one group, but you could use other conventions for this, such as group attributes. Maybe this is what recommendation #10 speaks to.

Another question about this: Does this only apply to groups? Are you, for instance, allowing datasets to have more than one parent? I guess not, because of the same scoping requirement.

4. Re recommendation #5. Dimension scoping. I'm really sorry we haven't gotten further with our dimension scales work. I think this approach is consistent with where we are headed, though.

5. Re recommendation #7. Creation order. As I understand it, this says that "creation order" would be global, rather than by group. I'm probably confused, but I wonder if this might cause a problem with the desire (recommendation #4) to have each group to correspond to what is currently a netCDF file?

Also, Quincey mentioned to me that this might impact on the need (or lack thereof) for his current work on giving groups the ability to track creation order.

6. Re recommendation #9. I'm not sure I understand the details in this.

7. Re recommendation #11. I agree with your approach – it's good to go slow on this. That said, we've found with HDF-EOS (which takes the same approach) that almost every EOS science team creates hybrid files – files that have the HDF-EOS collection, then adds other things elsewhere. I think the reason is that HDF-EOS doesn't cover all of the types of things they want to store. For instance, it's really common to store a jpeg "browse image" in the file, separate from the HDF-EOS objects. In the netCDF case, this may be less of a problem because netCDF is already way more general than HDF-EOS.

8. I strongly agree with your final comments about backward compatibility and the need to keep features simple for vendors.