Discussion topic notes

Allow CRS WKT to represent the CRS without requiring reader to compare with grid mapping Parameters (Alan Snow)
https://github.com/cf-convention/cf-conventions/issues/222

CF guide reference:
http://cfconventions.org/cf-conventions/cf-conventions.html#use-of-the-crs-well-known-text-format

Proposed change:

There will be occasions when a given CRS property value is duplicated in both a single-property grid mapping attribute and the crs_wkt attribute. In such cases the onus is on data producers to ensure that the property values are consistent. *If both crs_wkt and grid mapping attributes exist, the attributes must be the same and grid mapping parameters should always be completed as fully as possible. As such, information from either one (or both) may be read in by the user without needing to check both.* However, in those situations where the two values of a given property are different, *the CRS information cannot be interpreted accurately and users should inform the provider so the issue can be addressed.* , then the value specified by the single-property attribute shall take precedence. For example, if the semi-major axis length of the ellipsoid is defined by the grid mapping attribute semi_major_axis and also by the crs_wkt attribute (via the WKT SPHEROID[…] element), *the value of this attribute cannot be interpreted accurately.* then the former, being the more specific attribute, takes precedence. Naturally if the two values are equal then no ambiguity arises.

Benefits:

1. The CRS could originate from several different formats such as WKT, PROJ, or SRS Authority Code. If there are errors in the conversion process to the CF or WKT representation, only the provider would have the original CRS representation. As such, if there are conflicts, the provider would be the best source to go to in order to resolve the conflicts.
2. Making this change will simplify the lives of software developers so they can just read in the WKT or grid mapping CF parameters for the CRS without a need to compare the two.

Jonathan Gregory: The current position is that the grid_mapping takes precedence over WKT if both are present. To know whether this is the case, the data-user would have to interpret both and compare them, and take the grid_mapping as correct if the two are inconsistent. If I understand correctly, the idea of the proposal is to remove the requirement of precedence, so that the data-user can read the WKT alone and not consider the grid_mapping. In fact, I think the data-user can behave in just that way as things stand. The data-user is not obliged to check that the dataset is self-consistent. They are entitled to assume that it is, because the onus is on the data-producer to ensure this. I believe that's what we had in mind when we decided the current convention, which was a sort of compromise. The statement of precedence is intended to resolve problems if they are detected. Hence I don't think we need to change the convention to achieve the aim stated (Allow CRS WKT to represent the CRS without requiring comparison with grid mapping Parameters), if I've understood it correctly. It's already OK.

The discussion in https://github.com/cf-convention/cf-conventions/issues/222 has gone further than that, however. A lot of it is about whether the grid_mapping should be able to represent the information in the WKT. I think this is a much more difficult question. As I've commented in the issue, I have serious concerns about this. I realise that other people have good reasons for suggesting it, and I hope we can find a consensus, as usual - honestly, I do! I have two kinds of concern, which are related.

- If grid_mapping isn't required to represent everything, data-writers may choose to omit it. Some people would like that, because it would permit them to write WKT in the CF-netCDF file and allow it to be read and used as a "black box" required by certain software. I can see the practical advantage in that, but I think that's contrary to the general intention of CF, which has made it successful and popular. CF metadata is intended to be self-explanatory and intelligible to humans, designed to be read and written with minimal possibilities for mistakes. WKT looks to me more like code, it's not so self-explanatory because it doesn't label all its parts, the order in which things appear determines what they mean. It's not like CF-like.

- I suspect that WKT may be inconsistent with the CF data model in some ways. I can't be sure unless we understand it thoroughly and how it relates to CF metadata. For instance, @snowman2 noted in the issue that, "The `coordinate system` and `area of use` currently don't have an equivalent in the CF conventions." I don't know what they are exactly, but I think that coordinate system is an idea which is inconsistent with the CF data model, as an explicitly recognised part of the metadata. I recall also (but perhaps incorrectly) that there are defaults about directions and units, which may be inconsistent with CF.

Because of these concerns, I continue to think that we should require data-producers to describe everything they want to in grid_mapping, unless we can write down explicitly the mappings between

CF metadata and WKT. We don't have to be able to carry out the conversion in software, but we should write down how to do it. When that has been done, we would have much greater clarity. We could be confident in stating that certain parts of WKT were equivalent to CF metadata. It's possible that there are vocabularies in WKT which could be adopted by CF, and thus not maintained by CF. That idea is often suggested (e.g. by Jim in this issue) and it makes sense provided the vocabulary is one which can exactly "slot" into CF metadata. That is, it must suit our data model.

It seems to me that this approach of mapping is what is being followed with the discussions about standard names and ontologies. It is not being suggested that standard names should be replaced by vocabularies from elsewhere, for the same kind of reason that there isn't a one-to-one correspondence. However, the mappings can be established to help with interoperability.

But having written down the mapping between CF metadata and WKT, we would be able to write software that can do the conversion completely (@snowman2 already has such software, and probably others do). That could be used in the cf-checker to check that the grid_mapping and WKT are consistent, which is the problem we started with in this issue.

We don't want to have files that only contain CF-WKT. What needs to be changed to fully support conversion?

- Have to make it programmatically possible to achieve
- Add units to UDUNITS?
    - https://github.com/cf-convention/cf-conventions/issues/248
    - https://github.com/Unidata/UDUNITS-2/issues
- Add supplementary/replacement units to CF?
- Direction of projection coordinates
    - https://github.com/cf-convention/cf-conventions/issues/247
- Identify grid mappings that are not included
    - Add tripolar ocean projection?
        - Can add it if there is a formula
        - Need new coordinate names (dimensionless)
    - Add notes in the documentation about making an issue on GitHub if a grid mapping does not exist.
- Improve the structure of section 5.6 to make the crs_wkt property pop out more
- Remove erroneous unit attribute from one of the grid_mapping variables in an example in the convention text.