Ocean Sci. Discuss., 12, C1519–C1524, 2016
www.ocean-sci-discuss.net/12/C1519/2016/

Ocean Science Discussions

Interactive comment on "Technical note: Harmonizing met-ocean model data via standard web services within small research groups" *by* R. P. Signell and E. Camossi

R. P. Signell and E. Camossi

rsignell@usgs.gov

Received and published: 26 January 2016

Reviewer comment: The paper proposes a practical approach for aggregating, storing, searching and previewing model data for small research groups. The solution is basically a workflow based on a combination of some FOSS packages (THREDDS Data Server, pycsw, Iris, NCTOOLBOX), a proprietary software (Matlab) and some python scripts (e.g. scripts to connect THREDDS and pycsw). The implementation of interoperable web services is a specific requirement. This paper tackles an important issue which is make simple and sustainable (also for small research groups) the setup of an interoperable model data sharing system. The paper's structure seems quite clear, however there are some issues to solve before it can be published.

I think the main issue is the incompleteness of some parts. If I have properly understood the paper's objective, the section 3 should be the main part of the paper which should completely describe the authors' approach. On the contrary, it seems incomplete and not entirely developed. For instance, the integration between software components might be an interesting topic but the description is limited to what has been introduced in 3.3 (script to crawl THREDDS catalogs) and a few other details are scattered around the paper (e.g. 4.1, 4.2). In this case a more accurate description of the python scripts and their publication as Open Source Software would be very useful. The sections 3.2, 3.3, 3.4 are not fully explained, it's quite difficult to understand features and strengths of the proposed solution. E.g. the "Data preview" part doesn't describe the features of Godiva2 and the comparison with other WMS Clients. Other topics that could be addressed to complete the discussion: address possible security issues; description of the real effort to implement and maintain the server; description of the hardware and software requirements; insert a new figure for the section 3 (e.g. schema architecture) in order to highlight interactions between components and between the proposed solution and the users (e.g. researchers).

Author Reply:The "incomplete" nature of this paper was by design – we wanted to avoid details of the framework implementation.[f] We seek to convince small research groups that: (1) a standardized framework is useful, and (2) a standardized framework can be implemented with modest effort using free software components. Our approach was to give an overview of the benefits of standardization at a high level (the IOOS example) then describe specific components that can implement this standardization with minimal effort, then describe two specific situations where the framework has been implemented, demonstrating specific small group successes. We intentionally just touched the surface of many issues, to give a flavour of the entire framework without the details. We hope to convince research groups to become interested in implementing the framework, and if interested, they can find more detailed information which is readily available on the web. We will modify the

introduction to make it more clear that this is the intent, and have added web site reference for framework implementation details.

Some more specific points:

a) The two case studies present differences in some details (e. g. YAML file and python scripts to produce XML files). I think that in this way the proposed solution lacks of generality. The authors could integrate these differences in the section 3 in order to present a more complete solution which would adaptable at both (and more) case studies.

The USGS use case came after the NATO test case and the YAML file approach developed *could* have been applied to the NATO test case, as there were also ROMS results. We will add text to make it clear that this is a general approach that could be extended to other models to simplify the generation of the NcML files.

b) The same consideration applies to Ipython Notebook (part 4.1). If it is really convenient (for several reasons), it should be introduced in the section 3 to complete the proposed solution.

We considered this, and the reviewer is right that the IPython Notebook is a general tool, but so is the construction of a web portal using TerriaJS. These are just two examples of the many types of tools that are accessible once data has been standardized. We therefore prefer to reference them as actually used by the groups in the use cases.

c) The sentence ".. and Ocean Geospatial Consortium Web Map Service for data preview ..." in the Abstract is not clear. The acronym should be Open (not Ocean) Geospatial Consortium. Furthermore the OGC-WMS is not sufficient as "data preview" solution.

Agreed. We will fix the OGC acronym error and also mention that a WMS client

is required for data preview, not just a WMS server.

d) Many acronyms are introduced without the specification of their meaning

We specifically chose which ones to spell out. In some situations, like "pycsw", it's not actually an acronym, but a package name. In other situations, like "THREDDS", the acronym spelled out (Thematic Real-time Environmental Distributed Data Services) does not give additional insight, and in fact is never used in practice (it's effectively just a name).

e) It is not clear why is needed a different approach for "forecast models" (part 3.1)

Forecast models have overlapping data in time, thus construction of a continuous time series is more challenging. Traditionally users would have to cut out pieces of each forecast and join them together along the time dimension. The TDS does this automatically (and virtually). We will add additional clarification to the text.

f) Fix the wrong reference number for CKAN - Fig. 5 It should be Fig. 2

We will remove this figure and replace it with a figure that shows the flow of the standardized framework described in this paper, something like slide 7 here (https://speakerdeck.com/rsignell/catalog-driven-workflows-using-csw).

g) In the first case study I don't understand the differences between the approaches (CKAN, Geoserver, GeoNetwork vs. THREDDS, pycsw) and why the second is better.

We have removed the description of the data approach used by NATO before the approach described here.

h) Sometimes it seems the authors structured the article as if the readers all know how TDS works.

We will add the figure described above (and shown below) with additional text that highlights and further explains the central role of the TDS.

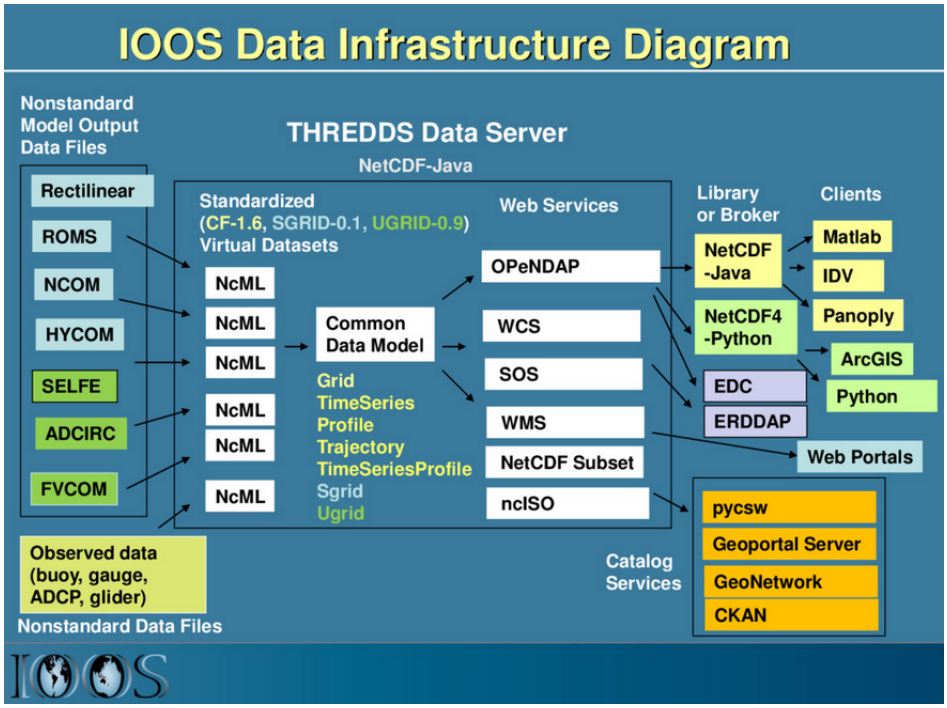

Fig. 1. Replacement for Figure 2.