White Paper #13 – Data and information delivery: communication, assembly and uptake[1]

Smith, N.R.[1], and Hankin, S.[2]

[1] Retired Scientist, Melbourne, Australia
[2] NOAA Pacific Marine Environmental Laboratory, United States

1. Introduction and history

This paper is principally addressing the 8th Term of Reference for the Tropical Pacific Observing System (TPOS) 2020 Workshop and Review:

> *Evaluate requirements for delivery of data, and derived products and information, in real time and delayed mode (e.g. availability, quality, latency, integration/interoperability); evaluate the existing data systems for fitness for purpose.*

The view has been taken that recent community consultation, findings and recommendations on this topic remain largely relevant and appropriate for TPOS and that as a result further extensive review and evaluation was not warranted. As important as data and information management and delivery may be, it was also reasoned that the current systems did not contribute a major risk to the operation and impact of the TPOS now, or its evolution toward 2020. The TPOS 2020 Workshop will determine whether these judgments were sound.

The paper will largely focus on high-level aspects and general requirements rather than those for specific facilities such as TAO/TRITON or Argo which are largely covered elsewhere, and will be more in the context of the global systems than focused on TPOS.

Just as the Tropical Oceans-Global Atmosphere Experiment and OceanObs '99 introduced paradigm shifts in our approach to data and information systems and exchange, this paper will argue that it is timely to consider yet another major change in our approach, with greater consolidation and integration of information management across the component parts of the Observing System, recognizing the data process and information systems as a service, both for the TPOS facilities/networks and for the ultimate users (scientists, operational agencies, policy and decision makers).

The tropical Pacific Ocean has long been a focus for innovation in ocean observation and science, from the early deployment of XBTs and monitoring of sea level, through to the major network and model development undertaken in TOGA (e.g., the initial TAO array; the first coupled model predictions of El Nino; the first free and open exchange of data in real-time) and WOCE (the development of floats and Argo; high-resolution models of the global ocean; innovations in scientific quality control, assembly and management of data).

---

[1] This white paper has not undergone the extensive community consultation of other white papers for this Review. Instead, it draws on material previously published as community papers, particularly for OceanObs09, and on relevant data and information findings from the other draft white papers. It also draws on input from Sylvie Pouliquen, Ken Casey, Scott Woodruff, Jim Cumming and Steve Worley, among others.

McPhaden et al. (1998) provide background on the TOGA observing system and the importance of providing real-time, quality measurement; both were lacking for the 1982-83 event and precipitated a major rethink in the approach. TAO introduced a new paradigm and systematic approach for ocean data and information delivery, a system that still underpins the current TAO/TRITON array. McPhaden et al. (2010) and TPOS WP01 highlighted the legacy of TOGA and the data and information systems established for the tropical Pacific, concluding they remain strong, particularly through the TRITON-TAO array and its pioneering role in open data exchange, real-time provision of subsurface data, and innovation in the assembly and serving of data and gridded data sets.

The community papers prepared for OceanObs ’09 cover much of the background and evaluation needed for TPOS 2020 (e.g., Pouliquen et al., 2010; Blanc et al. 2010; Hankin et al., 2010; de la Beaujardière et al., 2010; Beegle-Krause et al., 2010; Snowden et al., 2010). It is clear from those papers that data and information management has evolved and matured significantly since OceanObs ’99 and that the community of practice has also grown. In later sections we draw on these papers for conclusions and findings and, as appropriate place their recommendations into the current context.

The paper first examines some of the implications drawn from drafts of other papers contributing to the TPOS 2020 Workshop and Review. Next we draw out relevant findings and conclusions from previous community papers, and document implications for TPOS. The fourth section provides an evaluation and findings against the Terms of Reference, while the last section focuses on conclusions and recommendations.

2. Findings from other white papers

While at the time of writing a number of white papers remain in draft or outline form, a number of themes are already clear. The Global Ocean Observing System Framework for Ocean Observations provides the overarching context for TPOS 2020, and this is true for data and information management as much as it is for the observations process. At the most fundamental level the Framework consists of (a) scientific requirements, (b) the observation process, and (c) a data and products/output layer which we will term the data process. Part (a) influences priorities for (b) and (c); the solutions available for (c) influence choices in (a) and (b), and so on. Just as there are essential ocean variables, there are essential information system elements, and the feasibility and impact (utility) of possible information system contributions guide assessments of readiness and priorities for investment.

The Framework background document does address data and products but is less explicit around the definition of the data process or the way ‘readiness’ is assessed. Given the fact that TPOS will likely remain at the head of innovation, it may be timely to highlight essential ocean information system functions and promote assessments and metrics for these functions just as we do for ECVs.

The second theme running through many of the papers is the demand from science/research for efficient access to even more data; TPOS WP 3, WP 4 and 5, among other papers highlight the need for even more detailed and comprehensive data in order to understand processes that are limiting advances in climate prediction and data assimilation. In the words of TPOS WP03 “We believe that this progress will come from observing, diagnosing, understanding and teaching models to simulate the physical processes that underlie ocean-atmosphere coupling, and that this will have further benefits to much other science.” Similar sentiments are expressed in several other papers.

The take home message here is that the scientific requirements will continue to be a key driver of information systems, not just for managing data and information from the sustained observation networks, but also from the several underway/planned and *ad hoc* process experiments that are targeting the tropical Pacific Ocean. In an ideal world, the information system would be designed to accommodate information services to such initiatives, but we are some distance from that at present. TAO/TRITON and Argo track the use of their data in scientific publications but we lack more general metrics to measure impact. Figure 2.1 provides an example of metrics maintained by IMOS for impact within the research community.

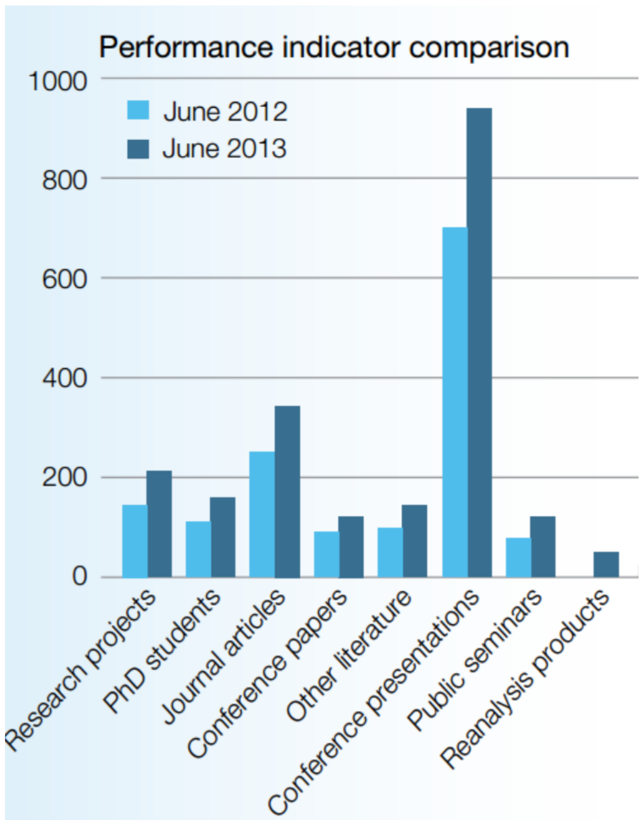

Figure 2.1 – Some metrics from 2012-13 from the Australian Integrated Marine Observing System (IMOS, 2013).

Another common theme was around the impact of modern information systems on the uptake of ocean data, products and information (see TPOS WP02 and WP08a, b in particular). This aligns directly with the concepts espoused in the Framework. We perhaps forget just how much change has occurred in the way the community accesses information, including in areas like the Pacific Islands, and that the notion “we want it all, and we want it now” is impacting information systems at a fundamental level.

TPOS WP 3, WP 9, WP 10 and WP 11 among other papers highlighted the importance of being able to bring diverse data to bear in both research and operational environments. The uses included validation of models, quality control, cross-calibration (satellites, Argo salinity), design and derivation of flux ECVs. In meteorology, Numerical Weather Prediction (NWP) systems play a dominant role in this area, however for TPOS we are still mainly operating at the lower levels of the data hierarchy (e.g., Level 2; see TPOS WP 3 and WP 11). Such requirements remain a challenge for our information systems. The effectiveness of both automated and subjective scientific quality control is directly dependent on our ability to place/view the data in the context of related data streams and scientific knowledge; information systems are a key enabler. In this context it is also evident that information systems provide services to the observation process.

Paper TPOS WP10 discusses data and information managements systems relevant to TPOS. The paper notes, among other things, that for "the global broad-scale in situ networks, data and information systems have developed around the individual networks" and that "if the primary utility of the high-resolution data is for regional observations including multiple data types, then consideration is needed for joint distribution of all datasets needed for a particular boundary current or other high-resolution phenomenon". Our interpretation of these findings is that consideration should now be given to taking the data and information process up a level where it is feasible and practical to do so, a theme we will pick up later.

Finally, a number of papers highlighted data and information services that were needed but not provided, usually because of lack of resources. For example, in TPOS WP 10, they note that there has been "no funding for Shipboard ADCP data processing within TAO since the 1990s" and that the onus often falls to individuals. Other papers note the challenges for managing underway SSS and SST data, and in Section 3 we note the difficulties in managing delayed-mode ocean-atmosphere data. At the edges of the core networks of TPOS support for data and information management is often thin or absent, in part because there is not the necessary scale of activity to justify investment. This does invite consideration of the concept of a (data and) information service, with an architecture capable of embracing *ad hoc*, experimental or small scale activities. If TPOS was to embrace the concept of an integrated approach to data and information management, built from the existing distributed network-based systems, then the data process would be seen as both a set of information systems and a suite of information services, to users within and beyond the TPOS.

3. Findings from previous community papers

TAO/TRITON and Argo have been exemplars for the development and implementation of effective and efficient data and information management systems. Pouliquen et al. (2010) describe the Argo data system (Figure 3.1 below) and highlight its ability to deliver to multiple requirements: fast data for real-time and near-real-time applications; delayed-mode data with higher levels of scientific quality control for climate applications and reanalysis; and data streams for general *ad hoc* scientific use. The use of selected hydrographic data to improve the quality of the larger Argo datasets exemplifies the power of the whole compared with individual datasets. Guinehot et al. (2009) provides yet another example of the whole system being used to add value to individual elements; *WP 9* provides other examples for satellite data.

The key message here is that the data processing phase can and does add value to the observations process (observations of the ECVs) and vice versa.

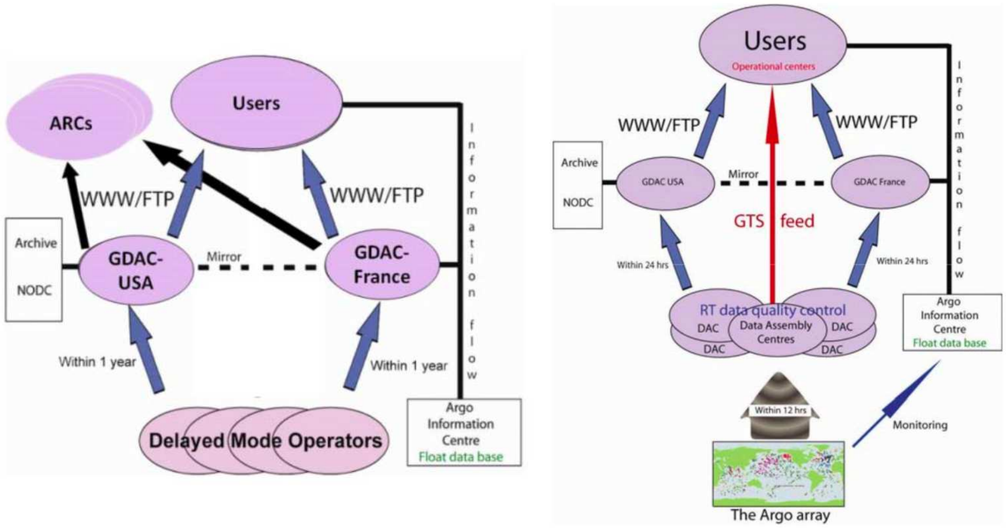

Figure 3.1 – (right) Panel showing the flow of real-time data from the network to users; (left) showing the value-add of delayed mode processing (after Pouliquen et al., 2010).

Blanc et al. (2010) discuss the contributions of GODAE to harmonising the production and exchange of various outputs and the development of essential and generic functionality to allow users access to products. Hankin et al. (2010) pursue the discussion in the context of standards and exchange protocols. What both papers highlight is the importance of agreeing on a system architecture that serves of geospatial data and products. Agencies are often encouraged to adopt enterprise architecture approaches to ICT planning and development (Figure 3.2 shows the Australian Government model; you will note significant similarities to the architecture approach for ocean information management).

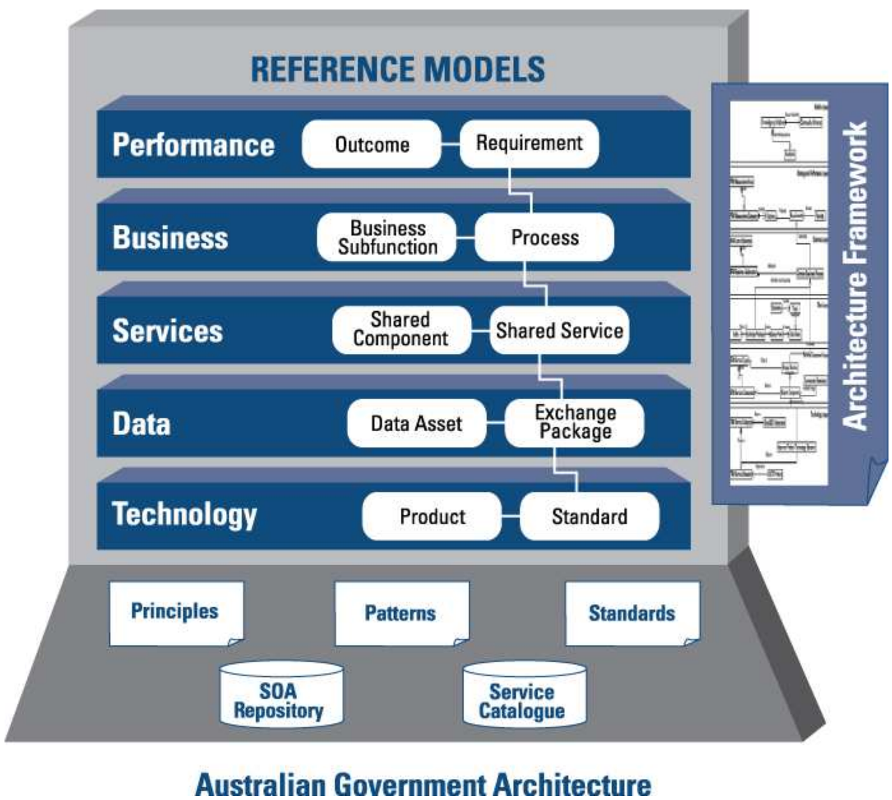

Figure 3.2 - Australian Government architecture model.

Blanc et al. (2010) note the importance of joining up the global community and perhaps it is timely, in the context of an integrated approach to data management across not just TPOS but the whole of GOOS, to agree on an architecture for that (data and information processing) system. The importance of a community of practice is implicit in the discussions of both Hankin et al. (2010) and Blanc et al. (2010), among other papers. As the latter notes progress "is dependent upon the [the ability of the] community to work together, to maintain a network of experts and to agree upon common approaches … concrete implementation is done through large integrated projects or programmes". TOGA and WOCE both fostered communities of practice and with consideration of TPOS 2020 it is timely to be more explicit about the role of a community of practice, to ensure an efficient approach to the essential ocean data management functions.

Belbeoch et al. (2010), and de la Beaujardière et al. (2010) refer to the scope of activities now being undertaken and the former highlights the value of a consolidated central information capability (JCOMMOPS) serving a broad range of users. De la Beaujardière notes "we have come to expect instantly available data and information" and as previously highlighted, the 'we want it all, and we want it now' culture is inevitably changing the architecture of our approach, with greater emphasis on the 'now'. De la Beaujardière et al. (2010) also pursue a theme around standards and recommend "existing open-standard approaches be used in preference to purpose-built or proprietary technologies", but emphasise the need for any standards to be purposeful and feasible (practical) (a pragmatic and sceptical approach; Hankin et al., 2010). We wish the community to move away from a culture where the investment is focused on IT innovation and focus more on using what we already have.

These same papers, as well as other white papers in this review, note that we have matured to the point where customised approaches for individual networks perhaps should be phased out in favour of more generic approaches (including for metadata), using industry-wide standards wherever practical and useful (the thinking behind Figure 3 is similar). Hankin et al conclude "We believe that a higher level of thoughtful awareness by the scientists, program managers and technology experts of the vital role of standards and the merits of alternative standards processes can help us as a community to reach our interoperability goals faster." It would be timely for TPOS 2020 to make similar findings/recommendations.

De la Beaujardière et al. (2010) make reference to levels of maturity and the need to inform the user community about these levels. Such 'levels of maturity' are often used in areas such as project management and ICT implementation and such an idea is worthy of deeper exploration here. JCOMM, GODAE-OceanView and Argo, among others do provide a level of oversight and review of their data and information capabilities but there is an opportunity to take this to a new level in keeping with the risks associated with weaknesses/failures of the system. One such idea is a more formal approach to data publishing whereby there are agreed (formalised mechanisms) such as DOIs, perhaps even licences to publish certain data streams. The idea is not so much top down imposition of criteria but ensuring that the maturity levels are recognised and respected. Consideration of the handling of delayed-mode surface data has raised a number of issues (K. Casey, S. Worley, S. Woodruff, personal; communication). The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) is usually regarded as the definitive holding of ocean-atmosphere data. For the

Global Tropical Moored Buoy Array (GTMBA) ICOADS usually relies on real-time transmissions over the GTS; this has worked reasonably well though the ability to produce high-quality data sets is constrained by the limitation on metadata transmission.

Delayed-mode data management and archiving is rather more problematic, partly due to the PMEL-NDBC transition; PMEL were good stewards of the data in the past. It is also partly because the governance around delayed mode GTMBA data is less well defined compared with real-time transmissions. The split of responsibilities within the US – NCDC, NDBC (previously PMEL), NODC, and then ICOADS probably does not help. NODC has been working with NDBC to improve the ocean data feed to the NODC archive. While the historical collection is not as clean and homogeneous as they would desire, data from January 2011 onward are now in excellent, interoperable netCDF formats. While not all data are coming in via this new modernized feed yet, NODC is confident the groundwork is now in place to enable the full and comprehensive datasets from NDBC to be archived at NODC with good metadata. The new data are already available to users by all of NODC's modern web services (not just ftp and http, but also via the THREDDS Data Server/OPeNDAP).

There are issues of substance with respect to the (historical) management and archive of GTMBA data, of which TPOS is a major subset:

- a high degree of format, resolution, and availability fragmentation
- The ideal of a single dataset in a single format, of uniform quality is some way off.

There may be value in an effort to harmonise the metadata and data quality. There is at least potential in having someone look into this in more detail, to see the degree to which the above mentioned fragmentation has harmed the high quality delayed-mode data set.

4. Findings relative to the Terms of Reference

It is clear the TPOS and its predecessor forms have sponsored many of the more significant achievements in data and information management (ToR 1); the TAO/TRITON system is one such case (Figure 4.1), but the we should also acknowledge the significant achievements in the way data and products are now made accessible and usable, as highlighted in TPOS WP 2 and TPOS WP 8a, 8b (ToR 2).

Terms of Reference 3 and 4 are couched in terms of ECVs but, as noted in Section 2, we should consider essential data and information management capability in parallel since the ECVs cannot achieve the desired impact without such capability. A similar remark applies to Terms of Reference 5 and 6, particularly as we are now considering feasibility and impact (readiness) within the envelope of available resources. Are the information services in place to accommodate extensions and/or innovations?

Modern ICT systems and their broad availability generally mean that logistics for information systems and services are becoming more readily available, not scarcer. Telecommunication still limits the exchange of some data (see TPOS WP10, 6 and 7, 11) and metadata; arguable this remains the biggest risk to an effective and efficient TPOS in 2020 from an information perspective.

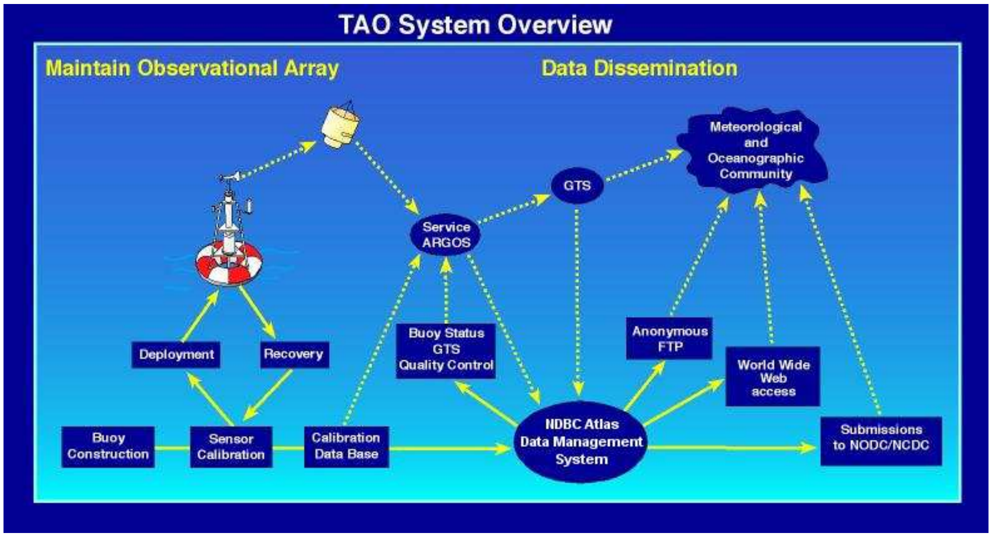

Figure 4.1 – The TAO system, including the data and information system.

Term of Reference 8 has been the main focus of this paper and we have relied heavily on more general evaluations performed for OceanObs '09 as well as a preliminary assessment of implications from other papers. Our main findings were:

1) It would appear timely to further elaborate the GOOS Framework to include explicit recognition of a data process, sitting alongside and working with the observations process. Figure 5 provides one depiction of such a separation, in this particular case to provide a clear separation of concerns in terms of governance for Australian Bureau of Meteorology observations and data. Arguable this is not the key issue here but such a concept also aids clarity in terms of implementation and encourages an integrated approach to information systems rather than a network-by-network approach. We should be testing and evaluating potential elements of the data process in the same way we use ECVs for the observations process.

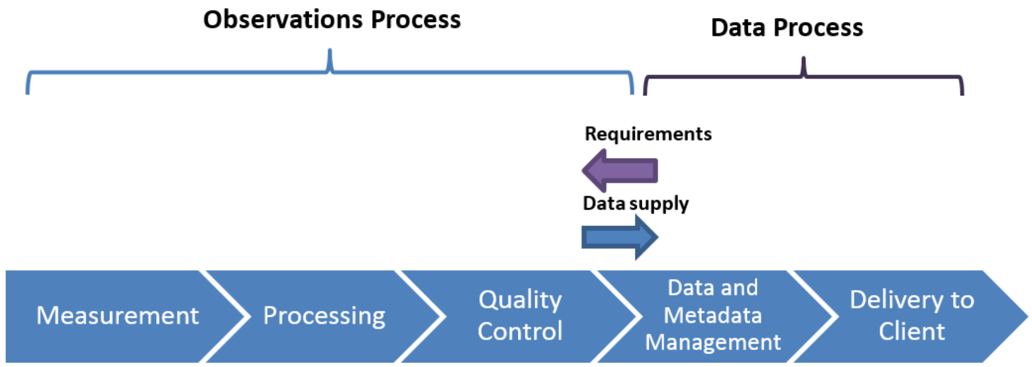

Figure 4.2 - One depiction of the separation of the data process from the observation process (Bureau of Meteorology, personal communication).

2) A second conceptual change would be to recommend greater attention to the architecture of the data process, to guide and promote effective implementation. A (GOOS) enterprise architecture approach would allow better articulation of the dependencies and risks and allow greater advantage to be taken of frameworks being developed elsewhere

(see for example Figure 3). This architecture should take account of the 'want it now' culture that characterises the new generation of users.

3) The first two findings should underpin a renewed drive toward integration and a culture of information systems and services, serving both the players internal to TPOS but also the external user community. Such an approach for TPOS might serve as a pilot for greater integration across the observing system as a whole (of course, working with and within JCOMM, IODE and the WMO Information Systems). In this way, we should avoid over investment in new IT solutions in favour of greater utilisation of already established systems and services.

4) From (3), it follows that we should focus on greater use of standards and protocols that have already been established rather than creating new ones. We should be testing all standards for their readiness (fitness for purpose) just as we would any proposed solution for ECVs.

5) The OceanObs '09 papers strongly emphasised the positive impact of what we have termed here a 'Community of Practice' for observing system information systems and services. TOGA and WOCE promoted such concepts and TPOS can provide leadership by promoting and supporting such communities; it is important that such communities are able to work within an agreed Framework and have agreed positions on approaches such as outlined in (1) – (4) above.

6) The requirements of research and science continue to be a high priority, perhaps higher than we might have anticipated at the last Review of the tropical observing system in 2001. Progress has been significant but models and data assimilation have yet to mature to the point where they add significant value in the data process (c.f. to the situation for NWP). The information systems therefore must continue to give priority to needs aligned with inquiry and investigation, often from small teams or individuals within universities. The value add from such engagement has been significant for both TAO/TRITON and Argo.

7) Operational requirements (by which we mean regular, routine needs, either from operational agencies or from policy/decisions makers) are now better known and better characterised. Data and products are required for levels 1 through 4 (see WP 4 discussion), for immediate applications (real-time and behind real-time) and for reanalysis and climate assessments and monitoring. Ease of access and use remains a key characteristic (WP 4 and 5) as does the need for high quality for climate change monitoring and detection.

8) For all users, we need better metrics on the way the information systems and services are being used (Figure 1 provided one example). Taking a line from improved knowledge of the architecture, we should be following uptake and use of services in all their forms. The Framework notes (in the context of the change in societal use of ocean information) that "We cannot manage what we do not measure"; this is also true of information management.

9) The importance of data and information services for improving the quality of outputs from the TPOS has been highlighted numerous times. Quality assurance and control happens within the observations process (at the point of measurement, in processing and in producing level 2 datasets; see Figure 5) but also within the data process as different data streams are brought to bear. TPOS 2020 might consider building on this strength with a clearer articulation within the enterprise architecture and in the service offerings.

10) There are issues of substance with respect to the (historical) management and archive of GTMBA data, of which TPOS is a major subset. *WP 11* explains in some detail the challenges of producing high-quality surface wind and flux datasets. It may be timely to get more precise guidance on the integrity and quality of data and metadata holdings for TPOS ocean-atmosphere data.

11) Finally, we noted levels of maturity vary across the information systems and services and that it may be timely to consider ways to formalise/capture this in the data process.

5. Conclusions

The tropical Pacific Ocean has hosted some of most innovative ocean data and information management initiatives over the last 30 years, beginning with sea level monitoring and TOGA and WOCE, and continuing through initiatives such as TAO/TRITON and Argo. The community is justifiably proud of these achievements.

The assessment and analysis of this paper does not point to any major risks for TPOS 2020 from the approach to information systems but it has identified a number of areas that may need improvement. Contemplation of the evolution of TPOS through to 2020 provides opportunities to consolidate and rationalise our overall approach, including through further integration and adoption of a community approach to the data and information process; there is no longer a strong case for platform specific approaches, and potentially much to be gained from joining up existing efforts. Some believe such a transformation and paradigm change is realisable prior to 2020.

For all users, there is strength in supporting an information system that delivers services through multiple channels, and with different offerings in terms of integration and quality. Acquire once and serve in multiple ways. Enabling improvements in quality, most of which are only possible with off-line scientific interventions, is an important function of the system. Further harmonisation of metadata management, versioning and publishing is possible.

The record of the data and information management community for TPOS is strong but can be improved, in terms of efficiency, robustness and effectiveness. TPOS 2020 provides a perfect opportunity to do this. In the event a major project around is supported, an underlying principle should be that around 10% of the total effort should be directed towards data and information management. This is particularly important for emerging and prototype technologies that will provide the basis for a future TPOS.

References

de la Beaujardière, J, Beegle-Krause, C. J., Bermudez, L., Hankin, S., Hazard, L., Howlett, E., Le, S., Proctor, R., Signell, R.P., Snowden, D., and Thomas, J. (2010): “Ocean and coastal data management”, in Proceedings of the "OceanObs’09: Sustained Ocean Observations and Information for Society" Conference (Annex), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306.

Beegle-Krause, C.J., Allen, A., Bub, F.L., Howlett, E., Glenn, S., Kohut, J., Schofield, O., Terrill,E., Thomas, J., Tomlinson, M.C., and Tintoré, J. (2010): “Observations as Assets in Decision Support”, in Proceedings of the "OceanObs’09 - Sustained Ocean Observations and Information for Society" Conference (Annex), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306.

Belbeoch M. & Co-authors (2010): "The JCOMM in situ Observing Platform Support Centre: A decade of progress and remaining challenges", in Proceedings of the "OceanObs’09: Sustained Ocean Observations and Information for Society" Conference (Vol. 2), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306.

Blanc, F., Baralle, V., Blower, J.D., Bronner, E., Cornillon, P., de La Beaujardiere, J., Donlon, C., Gemmel, A., Hankin, S.C., Keeley, R., Lauret, O., Loubrieu, T., Petiteville, I., Pouliquen, S., Price, M., Pugh, T., and Srinavasan, A. (2010): “Evolution in data and product management for serving operational oceanography, a GODAE feedback”, in Proceedings of the "OceanObs’09: Sustained Ocean Observations and Information for Society" Conference (Annex), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306, 2010.

Guinehut S., Coatanoan, C., Dhomps, A.L., Le Traon, P.Y., and Larnicol, G. (2009): "On the use of satellite altimeter data in Argo quality control". J. Atmos. Oceanic Technol. 26: pp. 395-402.

McPhaden, M. J., et al. (1998): The tropical ocean global atmosphere observing system: A decade of progress, Journal of Geophysical Research – Oceans, 103(C7), pp. 14169-14240, (doi:10.1029/97jc02906).

Integrated Marine Observing System (IMOS), (2013): “Annual Highlights Document 2012-13”, available from http://imos.org.au/, edited by T.Moltmann et al, 2013.

McPhaden, M. J., Ando, K., Bourlès, B., Freitag, H.P., Lumpkin, R., Masumoto, Y., Murty, V.S.N., Nobre, P., Ravichandran, M., Vialard, J., Vousden, D., and Yu, W. (2010): “The Global Tropical Moored Buoy Array”, in Proceedings of the "OceanObs’09: Sustained Ocean Observations and Information for Society" Conference (Annex), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306, 2010.

Pouliquen, S., Schmid, C., Wong, A., Guinehut, S., and Belbeoch, M. (2010): “Argo Data Management”, in Proceedings of the "OceanObs’09: Sustained Ocean Observations and Information for Society" Conference (Annex), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306, 2010.

Snowden, D., Belbeoch, M., Burnett, B., Carval, T., Graybeal, J., Habermann, T., Snaith, H., Viola, H., and Woodruff, S. (2010): “Metadata Management in Global Distributed Ocean Observing Networks”, in Proceedings of the "OceanObs’09: Sustained Ocean Observations and Information for Society" Conference (Annex), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306, 2010.