(*rough*) **Compilation of the inputs collected from the Marine Data Managers** (*June 2003*)
prepared by Bernard Avril (*JGOFS IPO*)

There is a need for:

- defining the goals of the present action: work out a business model of how data is published, and how users will (not just want) access or exchange these data, and how this transaction can be managed and tracked
- whatever message delivered to the parent bodies IGBP and SCOR to fully represent the views of the "real" data managers, those who did the job and will do it again
- defining the practicalities of the data management and its relations with funding agencies…
- defining requirements, resources and benefits of the data management framework…
- listing the obstacles to a good data management: e.g., not a priority for scientists; different time and protocols for data delivery, and different efforts to acquire data, which as a whole create a sense of data possession, especially for biological data

---

- a strong message to the (funding / political) authorities about the importance of "basic" science and the related data management, recognizing that oceanography is an expensive science but it is worthwhile and that data management is an essential part of the science process
- the highest quality in data management to support the best science
- recognizing that data management should be fully integrated and funded, proportionally to the science cost, in order to properly take care of those expensive but cost-worthy data, in any new projects or parts of the new project
- the limitation of the scope of the funded data management in relation to the core science of the project
- some strong connections of "basic" science with the economic sector, on the side of the funding agencies / political arena
- a strong and continuous commitment for data management, independently from any political agenda and political "swing"

- identifying the defaults / drawbacks in the JGOFS data management system: e.g., separated, national initiatives that prevented a good international, integrated data synthesis
- identifying the defaults / drawbacks in the WOCE data management system: e.g., some DACs were not maintained during the whole duration of the project. The know-how and techniques in data management developed in the DAC was not transmitted to the WDC
- recognizing that there is some feeling that JGOFS data management was derived of WOCE data management, while GLOBEC data management is related somehow to CLIVAR and GLOSS data managements (but this is not true!)
- acknowledging that any new oceanographic project in biogeochemistry and ecology will still be considered as a "small and complicated" data project when compared with other physical, observational projects, which produce very large but simple datasets, so that a specific, proper, adequate data management be set up in the future
- an agreement, a "permit" (with evaluation?) prior to participate in the international project, so that commitment to a proper, national data management framework is maintained and efficient

- a full-time paid Data Manager for each national contribution as a mandatory condition for the national participation to the project
- the identification of all funding agencies involved nationally and internationally in the project, so that a minimum agreement is widely reached
- a full agreement between national data managers and funding agencies and project parent bodies about data policy enforcement
- international data management guidelines and framework, for protocols of data collection and submission, for data policy with a clear timeframe and with continuity during the whole duration of the project, and for data management standards, such as use of XML, data dictionary…

- an IPO as a international focal point of activities and intelligence, in charge of careful and continuouss monitoring of the setup and progress of projects and field operations, in each participating countries
- a Data and Information Management Committee (DIMC) to be set up from the start, with an clear executive duty and mandate and to be science-driven, with at least 50% members from the scientific community (data producers, data users)… scientists being both end-users and specifiers / co-designers of the database objectives
- a national focal point in each participating countries
- a semi-distributed data management structure to be designed and funded accordingly to data types with national or international funding and with reporting to DIMC
- reconciling both approaches (national focus and data-type focus)
- the use of the existing RNODC for specific datasets, when relevant
- some national / regional funding for international initiatives especially for the DEUs
- the clarification of the disciplines covered by the DEUs
- the data producers to have full control of their data, provided that it is in accordance with the project rules of funding and project requirements
- a promotion of the moral obligation for data submission among the scientists

---

- some increase in "education" / awareness for the institutions managers (in science and data management) so they know that oceanographic data are not all easy to manage (collect, process, quality-control, document, organise, disseminate, archive…)
- a specific top-down strategy so "basic" data managers have support and recognition from the data management institution managers
- some clear direction and procedures at each level, for each actors (data managers, scientists, funding agencies, operation managers…), so that data management is a success in the project
- a listing of all services provided by data managers for each community, with specific goals and infrastructure (long-term stewardship, data management and publication for data users, and data user requirements…)
- data managers to work fully with, and for, the scientists and using their "language" in order to encourage and facilitate communication, for a full, better and positive cooperation
- distinguishing between "exclusive" and "inclusive" data management plans: exclusive means "you dictate the rules, and if we don't agree, we don't play". Inclusive means "you seek out from us what we think we can do, and together, we work out how to meet both of our needs". An "exclusive" will tend to generate competition among organizations rather than symbiotic cooperation

- a strong and clear enforcement procedure, if a submission deadline is set up
- a clarification of the conflict between data submission deadline and interdependency of measurements that slow down the whole data submission process
- distinguishing between "live data" (available after 6 months / 1 year / 3 years to the "operation" participants) and "archived data" (available after 1 year / 3 years / 5 years to all project participants)
- a clarification of the protocols and timeline for data delivery especially to the DEUs, and especially internationally
- the data submission policy to be enforced by the funding agencies or its representatives (focus points, operation managers…)
- an easy access to data for all, not only for the modellers and analysts, and for an even more easy and fast access within one field operation
- tools of data management, data analysis and data visualisation to be openly shared and to be supported through a open-community framework. Examples are the ODV software (www.awi-bremerhaven.de/GEO/ODV/), the GMT software (gmt.soest.hawaii.edu/) or the JOA software (odf.ucsd.edu/joa/), which are (great) successes. Is it better to have / promote individual initiatives (see above)…, or to setup an official initiative led by a task force for tool development (e.g., DODS (www.unidata.ucar.edu/packages/dods/index.html)…?

- the development of capacity building in data management, at the national and international levels, and of technology transfer
- recognizing the marked interest in the Digital Identification Object (DOI) development
- identifying any specific conditions and requirements for the Less Developed Countries?
- establishing a clear view of history and of local constraints in the development of national data management framework.