Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 293-307
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

# Summary report of the panel discussion

In the context of ocean data management, scientists, data managers and decision-makers are all very much dependent on each other. Decision-makers will stimulate research topics with policy priority and hence guide researchers. Scientists need to provide data managers with reliable and first quality controlled data in such a way that the latter can translate and make them available for the decision-makers. But do they speak the same 'language'? Are they happy with the access they have to the data? And if not, can they learn from each other's expectations and experience?

The last time slot of the symposium was used for a panel discussion. The main objective of the panel was to bring about a discussion between data centre managers on the one hand, and users of the data, in the first place the scientists, on the other.

There were two panel members from each of the data management and the scientific communities, and from international organisations. The panel discussion was divided into two parts; the first part consisted of short opening statements by the panel members, based on the opening questions listed below; each panel member covered 6 questions: three general, three specific for the community (s)he represented. The second part was dedicated to open debate.

## Panel members

- Chair: Savi Narayanan, MEDS, Canada
- Representatives from data centres:
    - Lesley Rickards, BODC, UK
    - Catherine Maillard, IFREMER/SISMER, France
- Representatives from the science community:
    - Peter Herman, NIOO/CEME, The Netherlands
    - Neville Smith, Bureau of Meteorology, Australia
- Representatives from international organisations:
    - Alan Edwards, EU
    - Peter Pissierssens, IOC/IODE

## Synthesis

We have tried to summarise and synthesise the main points that were discussed during the conversations in this section. It should be noted that the debate was a very lively one, and much of the input that is now part of the synthesis below came actually from the members of the public.

### *Changing role of data centre*

Changes in technology have been leading to changes in the role of data centres. There is a trend to move away from the traditional data centre, with its main task of archiving datasets, to become more service-oriented.

Data centres can look towards libraries for inspiration to redefine their role; libraries provide expertise and guidance in cataloguing. Archives are grey and dusty, libraries are active and open; data centres should strive to resemble the latter rather than the former. Data management needs an equivalent to the 'Web of Science': a mechanism to bring up a list of relevant, available, quality controlled, peer reviewed datasets.

There is a need to create data and information products; not only towards other data managers and scientists, but also towards the policy makers and the society at large. These products will assist in increasing the visibility of the data centres, thus assisting in attracting both funding for further activities, and data submissions from scientists.

Some traditional roles of data centres remain important: long-term stewardship of data, integrating datasets, documenting and redistributing datasets, development of standards and standard operational procedures…

### *Bridging the gap between scientists and data managers*

Both data centres, and data and information management procedures are very poorly known by marine scientists. In most university programmes, there is no training on data management, no information on data centres, data management procedures… Data management is perceived too much as an IT topic. There is a need to investigate how to put data and information management on the curriculum of academic institutions. This would result in a better knowledge of the data centres, and an increased quantity and quality of data submissions.

Data managers should actively seek collaboration with scientists. If data managers have a background in science, it is possible to establish a relationship of trust with the scientists, a smoother collaboration, and a greater input of the data managers in the development of data collection. The involvement of the data managers in the planning of projects from a very early stage makes 'End to end data management' a reality.

EU has the mandate and the funds to support and improve training for scientists in data management, and could be playing a role in this.

### *Creating incentives for scientists to submit data to data centres*

To a large extent, data centres are dependent on scientists to submit data. Especially in view of the extent to which scientists are not aware of the role or even the existence of data centres, this is a potential problem. Several actions can be taken in this respect.

- Creating awareness about importance of data management, by *e.g.* including data and information management in the curriculum of universities;
- Requirement for data management written into project condition for funding – is already the case for EU proposals, and happens for short-term data management;

- Developing peer review and quality control procedures, to assess usefulness of a dataset, and making a dataset citeable, so that a scientist's contribution of data to a data centre can be measured, and taken into account for career advancement.

### *Need for long-term activities*

Datasets often result from projects, which usually have a limited time-span. Data management on short term, within the time span of the project, is usually no problem: scientists do need data management to produce the deliverables to the project; moreover, making provisions for data management is a prerequisite to have a proposal accepted in the first place. There is an obvious need for activities beyond the duration of the data-generating project, to assure continued availability of the data. This always has been, and probably should remain, one of the tasks of data centres.

Funds for long-term data management should not come from research budgets, but rather from operational networks or other mechanisms. Several initiatives of the EU are relevant in this respect. Within Framework 6, there is a possibility to fund the operations of large 'Networks of Excellence' that will operate on time spans much longer than a typical project. The Global Monitoring for Environment and Security (GMES) initiative is another potential mechanism.

### *Duplication of efforts*

A certain degree of duplication is unavoidable, and is a fundamental aspect of the scientific process. There has to be room for experimentation, different attempts at solving the same problem. After some time, however, experimenting should stop and be replaced with one or a couple of strategies.

Undesirable duplication can partly be stopped during the project proposal review process. One of the objectives of the Networks of Excellence, as proposed by the EU, is to increase communication between partners of the network, raising awareness of each other's activities, and hence decrease the probability of duplication.

### *Need for peer review of datasets, and for standard practices*

There has to be peer-review, as a way to measure and recognise progress, to recognise value and expertise, and as a foundation for standards and accepted procedures. Standards and audit procedures are needed to allow objective peer review. Developing these standards is a task for the data centres.

Peer review is a way to increase the compliance with standards. Countries, or even institutions or scientists, could be tempted to work along principles that were developed locally; obviously, these will fit local needs, and are usually much faster to develop. Doing so, however, can lead to fragmentation, and hamper data exchange.

### *Difference between biological and physical data management*

The problems of biological and physical data management are different: physics datasets are often high volume and low complexity; biology datasets are low volume but high complexity. Taxonomy brings a $5^{th}$ dimension to ocean data management.

The lower level of standardisation in biology makes importance of proper documentation with the datasets even greater.

Commonalities are more important than differences: both biological and physical data management need for long-term activities; quality control and peer review; creation of data products

### *Involving the developing countries*

Participation of developing countries in global programmes is the best way to transfer expertise. Global programmes can operate at several levels, so that they can serve both global and local needs.

Internet access is a problem in many third-world countries, and assisting with connectivity and basic telecommunications should be made a priority in any capacity-building programme. Where internet is available, the bandwidth is often very limited, making it virtually impossible to download large volumes of data. As long as this problem remains, data should also be distributed on alternative carriers such as CD ROM or DVD. Data warehousing and brokering can assist in locating and selecting relevant datasets, thus limiting the volumes of data to be downloaded.

Also funds to purchase hard- and software, and expertise to maintain the systems, are a factor that is more limiting in developing countries. The data management community should provide platform-independent software that is open source and runs on hardware that is compatible with technological expertise available. Reliable and stable standards should ensure that data are available in a form that can be handled by these tools. Capacity building programmes should be organised making use of these tools and standards.

### *Suggestions for actions*

Investigate how data and information management can be made part of the curriculum of marine sciences in academic institutions;

Develop standard operational procedures and a peer review process to allow an objective assessment the quality of datasets;

Guide the user community directly to relevant, and quality controlled datasets, by setting up portal sites;

Create integrated data products, to increase the visibility of the data centres;

Distribute data not only through the internet, but also on CD or DVD;

Assist third-world countries with basic telecommunications, Internet access, and data warehousing;

Create a collection of open source, platform-independent software for the benefit of third-world countries, and organise capacity building programmes around these.

# Panel starting questions

### *Data centre representatives:*

1. What do you see as the role for data centres in managing data from the global science programs?
2. What are the challenges you see that data centres need to face individually and collectively?
3. What added value comes from managing data in data centres, rather than in the originating institutions? What should a data centre have on offer to be more than just a convenient data archive?

### *Scientists:*

1. What are your expectations from the global network of data management systems?
2. Can the global network meet your expectations now, with some changes or with radical changes?
3. What governance structure would ensure effective and efficient management of global data, assuring and documenting data quality, securing data for future generations, and providing easy access to integrated multi-disciplinary data?

### *International organizations:*

1. What is the role of international organizations to address the data management requirements?
2. What do you think are the major challenges that the international organizations face in global data management?
3. What changes should be implemented at the international level to better deliver the global data management mandate?

### *All:*

1. What data management practices can be employed to reduce the impacts of technological differences between developing and developed countries?
2. What do you see as the main differences in data management practices between biological and physical oceanographers? What can be done to bridge these differences?
3. If you had three wishes to improve global data management, what would they be?

# Lesley Rickards

What do you see as the role for data centres in global science programmes?

- Partnership with scientists to create integrated project datasets
- (Scientific) quality control and documentation of data is crucial
- Long term stewardship and archival of datasets: archival; scientists move on to the next piece of work
- Infrastructure and IT expertise
- Development of standards

- Access to other datasets

What are the challenges you see that data centres face individually and collectively?
- Adequate funding: data management is left to last, when money has run out
- Working collaboratively whilst responding to national remit
- Increasing data diversity (many different parameters being measured)
- Working to common standards
- Adapting to new technology (distributed systems, XML, etc.)
- Developing systems to deliver (near) real-time data
- Effective data dissemination mechanisms

What added value comes from managing data in data centres? What should a data centre have to offer to be more than just a convenient archive?
- Integrated datasets and value-added products
- Long term stewardship of data
  - o End-to-end data management
  - o Pre-cruise to final data product
- Real-time through to delayed-mode
- Effective dissemination mechanisms
  - o *e.g.* CD-ROMs, DVDs, internet/web-based
- Advice on quality control & standards
- Use of new technology
  - o *e.g.* distributed systems, XML, etc.

What data management practices can be employed to reduce the impacts of technological differences between developing and developed countries?
- Ensure data and information are available in the most appropriate form for both – remembering that the same solution is not always appropriate (*e.g.* web vs CD-ROM)
- Sharing of expertise and working together
- Ensuring high-quality capacity building programmes
- Encouraging the participation of developing countries in global programmes (with local/regional applications) *E.g.* GLOSS: serves both local and global needs, at different levels
- Platform independent software

What do you see as the main differences in data management practices between biological and physical oceanographers? What can be done to bridge these differences?
- Physical oceanographers often more computer literate (*e.g.* develop their own software for processing, quality control and analysis)
- Some measurement techniques more standardised
- Biological datasets may be more complex and require much more supporting documentation to describe collection methods, etc.
- Development of flexible data storage mechanisms to accommodate both biological and physical data
- Provision of (appropriate) software tools
- Provision of integrated datasets

If you had three wishes to improve global data management, what would they be?

- Proper collaborative efforts, building on existing standards and practices – not reinventing the wheel each time a new project comes along – *e.g.* WOCE (continued into CLIVAR?) Danger of loosing expertise in the intervening period.
- Recognition of the value of data management and stewardship by the scientific (and operational?) community. Proper recognition of the value of data management and data stewardship long-term archival from the scientific world and the operational communities, so that it doesn't fall off the bottom of the list each time
- Adequate funding also after termination of short-term programmes

## Catherine Maillard

Role of the data Centres in managing data from the global science programs

- To offer a perennial (long term stewardship of the data) and high technology level infrastructure and professional staff (to make sure the hard- and software available is matched to the latest developments of technology) to contribute to the project data management structure
- To disseminate the existing international standards and expertises as a minimum basic requirement for data management, especially for the meta-data and the quality control assessments – avoid re-invention of standards
- To provide standardized complementary datasets to the projects – *e.g.* data produced by other projects, historical data

Challenges that the data centres need to face individually and collectively

- To be professional (first priority is data dissemination/redistribution; first priority of scientists is to do science, not sharing data) and operational to insure that all the data are
  - o safely archived
  - o checked for quality
  - o timely processed and delivered - disseminated by internet and other adapted communication systems
- To return products to the data providers and the other users
  - o integrated observations datasets from various sources
  - o value added synthetic products
  - o contributions to scientific papers (like happens in *e.g.* OCL) – data managers should not compete with scientists! Data managers have to contribute and be visible, but in a very specific way.
- To convince the data collectors and the supporting agencies that it is valuable to spend a low percentage of the overall cost of the data collection for the archiving. Comparison of cost of ship time with money for data management

What added value comes from managing data in data centres?

- Providing a perennial archiving system able to cope with the changes in the media technology
- Insuring that the data documentation is in accordance to the international standards and the data organization not 'person dependent' = main difference between data manager and scientist
- Avoiding the dispersion of the data in hundreds of laboratories or scientific teams with different data management methodologies

- Assembling long time series of data of the same type
- Providing timely access to data through dedicated data servicing teams
  - o providing data in real time
- Optimising the costs and manpower

What to do to reduce the impacts of technological differences between developing and developed countries?

- To insure one WWW link to all: that is possible, at least with limited speed
  - o also distribute data on alternative carriers
- To insure the participation of both developing/developed countries in the development of common databases
- The data policy should insure a wide data circulation
- The developing countries should also contribute to the international effort
- The job training on practical issues should enhance the mutual sharing of expertise

Differences in data management practices between biological and physical oceanographers?

- More complicated datasets in biology:
  - o a 5th (or even 6th) dimension with the taxonomy, genetics…
- Necessity of linking environmental parameters, bio-chemistry and biological observations in different compartments water column, SPM, sediment, biota ..
- Less practises of data exchange and computer technology among the biologists, including the documentation of the coordinates
- The bio observations are frequently too dependent on the measurement technology to be inter-comparable

Three wishes

- A common work on the standards for various parameters, not only for temperature and salinity
- A better acknowledgement of the data submission from the experimentators: the submission of a data set that pass the QC at a data centre should be recognized as a 'rank A' publication
- A better public awareness of the data legacy and the 'data librarian duty': how to follow the climate change without long time series of good data?

## Peter Herman

What are your expectations from the global network of data management systems?

- Easily accessible, highly resolved ecological data
  - o Geo-referenced
  - o Consistent taxonomy
  - o Auto-ecological information (as in FishBase, AlgaeBase…)
  - o Well-documented methods
  - o Physical and chemical data (depth, light, chlorophyll, nutrients, sediment composition, physical stress…) linked
- Spatio-temporal variation represented

Can the global network meet your expectations now, with some changes or with radical changes?

- More, especially coastal, datasets would need to be made available
- Emphasis should not only be on species occurrences, but also on density, biomass, functional types,...
- Digested physical, chemical and biogeochemical information would be very worthwhile (+ access to primary data)

What governance structure would ensure effective and efficient management of global data, assuring and documenting data quality, securing data for future generations, and providing easy access to integrated multi-disciplinary data?
- Distributed databases to avoid problems of (incomplete, faulty) duplication
  - o It is hard to transfer biological data from one database to another
- Citable databases? Or: storage of databases in conjunction with publications? In any case: an incentive for the researcher is needed
- Peer review of database quality
  - o Even more important for biological data than for physical data
- Reliable exchange protocols, including taxonomic standards

What data management practices can be employed to reduce the impacts of technological differences between developing and developed countries?
- Reliable, stable standards. Data managers are trying too hard to keep up with recent technological developments, at the expense of stability, and the opportunity to bring developing countries up to the level needed to contribute to and to take advantage of global data management
- A comprehensive, consistent set of open-source software according to these standards
- Time to develop systems and build capacity before everything changes again

What do you see as the main differences in data management practices between biological and physical oceanographers? What can be done to bridge these differences?
- The huge variety of variables (every species being one, plus all the measurements on these species) in biological studies
- What can be done: continue to decrease biodiversity?

If you had three wishes to improve global data management, what would they be?
- More involvement of scientists as data suppliers/users, through appropriate incentives
- Technology that promotes active participation of non-wizards, *e.g.* in distributed databases, instead of xml-ing them out
- Comprehensive portal sites that do not lead to metadata-basis nonsense, but directly to data and information

## Neville Smith

What are your expectations from the global network of data management systems?
- Efficient and effective communications
- Standard for marine metadata
- Leadership in quality control, quality assurance and data set integrity: assembly
- Data communications, transport

- Responsibility for archives
- Partnerships with science - data and information services

Can the global network meet your expectations now, with some changes or with radical changes?
- There are community attitude changes that are required
- Moving away from ocean data archiving, to ocean data services
- It requires collaboration and cooperation across a broad community
- Questions of strategy and tactics
- Do we trust in natural evolution? [No.]
- Will the sum meet expectations? [I don't know]
- Is it possible to productively harness initiatives

Governance?
- A disciplined approach
- Responsibility for infrastructure
- It seems a single structure can stretch across all of the ocean/marine disciplines.
- It must enjoy the same status and recognition as a science program and work seamlessly between science and operations/applications.

Developing and developed countries?
- Using practices and tools that are commonly available
- Data warehousing and brokering
  - Production of data products by data centres; interpretation of data, filtering of data
- Assisting in basic telecommunications

Differences between biological and physical data management
- Physical data has simple characteristics c.f. biological
  - Biology: complex structure but low volume
  - Physics: simpler structure, but high volume
- Physical community should regard biological community as an important client
- Marine XML a common tool
- Assembly standards: assurance
- Innovative transport and access methods
- Commonalities are more important than differences
  - Need for peer review system

Three wishes
- Community attitude: make scientists and data managers work as one
- Dynamicism and integration; moving towards a more service-based system
- Frontier projects – do things in a different way, not necessarily with new technology.

## Alan Edwards

What is the role of international organisations in addressing data management requirements?
- To provide a suitable framework for effective data management (framework is more than just policies, it also has to include at the very least funding)
  - This should acknowledge and protect the rights of ownership

- Take account of commercial considerations (not necessarily at the point of delivery, but somewhere in the loop)

What are the major challenges that international organisations face in global data management?

- Compliance
  - It is one thing to have a 'policy'; it is another to have everyone respect this.
  - The carrot and the stick: funding. But there should be more than this; it should be a partnership
- Rapid technological change
  - This can lead to policies/legislation being redundant before they are adopted.

What changes should be implemented at the international level to deliver better global data management?

- Proper investment in data management.
- Recognition that good data management serves multiple purposes, especially the provision of information products. Products should be seen to contribute to societal needs by politicians
- Actual mechanisms to allow users to properly define what information products they need and in what formatHow can international organisations help to advance Data Management?
- By establishing a proper legislative and / or operational framework to support data management. [*e.g.*, within the 6th Research Framework Programme and the GMES Initiative (Global Monitoring for Environment & Security)]
- By ensuring adequate resources for data management initiatives are available.

## Peter Pissierssens

What is the role of international organizations to address the data management requirements?

- To facilitate and promote the management and sharing of data at the global level
- To develop and promote standards
- To assist countries in acquiring the necessary capacity to manage their national data and to participate at equal level at the international level
- To promote cross-sectoriality

What do you think are the major challenges that the international organizations face in global data management?

- Countries preferring to use their own standards rather than adopting those internationally agreed upon
  - Problem of inertia: it takes too long to agree upon a standard; by the time the standard was agreed upon, it might be obsolete
- Countries placing national short-term gains over long-term global benefits
- Countries unable or unwilling to contribute resources for international initiatives

What changes should be implemented at the international level to better deliver the global data management mandate?

- Countries to actually implement internationally agreed-upon standards
- Countries to actively contribute to international programmes

What data management practices can be employed to reduce the impacts of technological differences between developing and developed countries?

- Wrong question: we need to close the divide between developed and developing in terms of technology availability. We cannot stop technological development so we need to ensure global access to the technology (problem of stability of the technology!!)
- Possibly concentrate on low-cost, platform-independent and open source technology that is therefore more easily accessible (see examples from COD)
- No science without library, no science without communication; spend money on internet access! This was one of the priorities of the Odinafrica project

What do you see as the main differences in data management between biological and physical oceanographers? What can be done to bridge the differences?

- The main difference is the time it takes to process the samples; moving from delayed to real-time is difficult
- To some extent technology could assist (image processing & recognition…)

If you had three wishes to improve global data management, what will they be?

- Real implementation of agreements and standards
- Real commitments and giving more attention to international over national interests
- Coordination and cooperation between all stakeholders and data centres
  - Do not give up on international initiatives; it is possibly easier to develop isolated solutions, but this would lead to a multitude of systems that would not communicate

## General discussion

Don Robertson: Data management left to last minute/dollar; remedy: price data management into contract; last money is only paid on submission of data; this is a simple contractual agreement

Lesley Rickards: legislation would be more of a problem in government departments; in the BODC, it is mainly a matter of intellectual property rights, which are acknowledged in the data centre's standard procedures (the originating scientist's name gets attached to the data set).

Catherine Maillard: Data from other country's EEZ does not get distributed; other types of data with restrictions are eg data relevant to national security

Franciscus Colijn: Too much reinvention of wheels; there should be mechanisms during project proposal evaluation to avoid duplication

Savi Narayanan: Are there plans to solve problems of duplication and of life cycle planning with the funding organisations?

Alan Edwards: Part of the FP 6 mechanisms is created to alleviate the problems of fragmentation and the need for long-term activities. Large Networks of excellence created to avoid duplication, increase communication, and increase integration

Peter Pissierssens: Problem that projects are limited in time, have to produce deliverables after 3-4 years; projects are seen as discrete events, with a beginning and an end. What to do with needs for long-term activities?

Alan Edwards: Research budgets are not appropriate mechanism for very long term activities. There is a tendency to increase project duration, but even this will not cover adequately the needs of long-term activities. This has to be done by funds for *e.g.* operational networks, or through activities like GMES.

Jerry Miller: Most innovative work has been done by a combination of a small amount of own data with large amounts of other people's data. No programmes are funded unless full sharing of data. Sharing implies data management, and this obviously requires funding. Proposals hardly ever include Principle Investigator for Data management.

Peter Herman: EU proposals have to include funds for data management, and during the review process this requirement is checked. Often during the negotiation phase the amount of resources allocated to data management is increased.

Peter Pissierssens: In how far is DM part of academic curriculum, of the scientific culture? Do scientists find their way to data centres? Do data managers visit the scientific institutions?

Savi Narayanan: Scientists do not know about data centres.
Murray Brown: 10 year ago, people did refuse to share data. Since then things have changed. There are now contractual obligations to submit data, and this involves data sharing and management. In this process, scientists have learned to work together with data managers. Duplication of efforts is not necessarily bad in its own, and unavoidable as part of the scientific process. It is not possible to make a big master plan to stamp out all duplication.

Neville Smith: We should not strive at a one-size-fits-all; 'experimenting' and duplication is indeed part of the scientific process. But there is a time that experimentation should stop, and replaced with one or a couple of strategies. Data sharing is a means to an end, not a goal in itself. Data sharing has to bring added value. Standards' sharing is on another level, and useful in its own right.

Savi Narayanan: Brings up Peter's point again.

Karin Stocks: It is necessary to train people in data management; but it is also necessary to have professional recognition for people spending time on data management task, *e.g.* reviewing quality of datasets and web sites, so that data managers can get professional recognition, can put data management activities on their CV.

Daniel Davis: There is not enough incentive for scientists to work together with technologists; ocean science is lagging behind in technology.

Catherine Maillard: Peer refereeing of datasets has drawback: it weighs on the time span on which the data can be delivered.

Edward Vanden Berghe: Standard operational procedures by which dataset is generated can be quality-controlled, rather than data set itself, thus avoiding delays in data delivery.

Bob Branton: Standard approaches are difficult; it is often easier/one is tempted to devise one's own solution.

Tim Deprez: No training on data management in the academic curriculum, no information given on data centres, data management procedures (cataloguing, archiving...). Scientists are computer literate, and working with ease with Excel... but not with RDBMSs like Access. Also, scientists don't know about 'big' data centres, but know about little ones.
Bernard Avril: Mathematics, statistics... are part of the academic curriculum, but not data management; it is perceived too much as an IT topic. Awareness of data centres: through libraries? Librarian is helpful as a paradigm to data manager.

Peter Pissierssens: Necessity to put data and information management on curriculum of academic institutions. By having students trained in Data and Information Management, they will realise the importance of data management. It will no longer take a stick to have them submit their data to data centres, but their awareness of the importance of data management, and the role data centres can play in this, will serve as a carrot.

## Closing remarks

Savi Narayanan: Data management is more than an IT problem, there are concepts involved; there is a need to inform scientists about these concepts, and create awareness of their importance. Scientists who attended the panel discussion should take these ideas to their home institutions, and discuss possible ways to bring data and information management of the curriculum of their institution.

Lesley Rickards: Scientists should make sure to organise their own data; this would make the work of data centres much lighter, and avoid the danger of data loss. Also: most senior BODC staff have a background in science, and hold PhDs; this allows to establish a relation of trust with the scientists, a smoother collaboration, and a more creative input of the data managers in the development of data collection.

Peter Herman: Data management needs an equivalent to the web of Science: a mechanism to bring up a list of relevant, available, quality controlled data. A portal site has to provide access to data, preferably using a common interface.

Neville Smith: Data centres have to move away from their traditional role of data archiving centre, towards a data service centre. They can take an example from libraries, which provide expertise and guidance in cataloguing. There has to be peer-review and citation for datasets, as a way to measure and recognise progress, to recognise value and expertise, and as a foundation for standards and accepted procedures. Peer-review and an audit system will allow enforcing standards and orderly progress in what we do.

Catherine Maillard: A new type of data centre, more service-orientated, is emerging.There should be more efforts from the side of the data centres to ensure that data centres have enough visibility. It is difficult to avoid overlapping tasks.

Peter Pissierssens: Data managers have to find new ways of communicating with scientists; this symposium has been a very good opportunity to start this dialog. Both scientists and data managers are making data and information products; a close collaboration between the two

groups will produce synergies, allowing the complementarity between the two skill sets to come into play.

Alan Edwards: Summarises the main points brought up during the discussion panel. Training: EU has money and the mandate to support ad improve training for scientists in data management. Library view of data centre: Archives are grey and dusty; libraries are active and open. Requirement for data management written into project condition for funding – is already the case for EU proposals, and happens for short-term data management. For long-term data management, often no funds are available anymore.

Savi Narayanan: Long-term data management is a challenge.