

# Chapter 2

## Connecting Biodiversity Data During the IPY: The Path Towards e-Polar Science

**Bruno Danis, Anton Van de Putte, Sylvain Renaudier  
and Huw Griffiths**

### 2.1 IPY and the Need for Data Sharing

The International Polar Year (IPY) was a unprecedented effort in polar research committing thousands of participants from many nations (Carlson 2010), with a common objective: to describe and understand complex systems using multiple approaches, in an integrated fashion. IPY scientists have collected enormous amounts of data in many formats, ranging from multimedia, hourly to millennial time series, isotope ratios and fractions, energy and material fluxes, species identification and distribution patterns, disease types and rates, genetic sequences, samples identifiers, singular events and gradual processes, to sociological studies (Carlson 2011).

As a result of this approach, IPY scientists now have to deal with a “data deluge” which, if well utilized, has potential to considerably change our vision on

---

B. Danis (✉) · A. Van de Putte  
Royal Belgian Institute of Natural Sciences, Rue Vautier 29, 1000 Brussels, Belgium  
e-mail: bruno.danis@gmail.com

A. Van de Putte  
e-mail: antonarctica@gmail.com

B. Danis · A. Van de Putte · S. Renaudier  
Antarctic Biodiversity Information Facility (ANTABIF), Rue Vautier 29, 1000 Brussels,  
Belgium  
e-mail: s.renaudier@gmail.com

B. Danis  
Laboratoire de Biologie Marine, CP160/15, Université Libre de Bruxelles, Avenue FD  
Roosevelt, 50, 1050 Brussels, Belgium

H. Griffiths  
British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK  
e-mail: hjg@bas.ac.uk

many polar systems. More and more, simulations are based on these data, and in turn produce very large amounts of newly generated data. Even if data-intensive science emphasis has been on dealing with large quantities of data (Hey and Hey 2006; Hey et al. 2009; Newman and Ellisman 2003), their heterogeneity, together with the arising need to interconnect them to understand complex problems, is an even more overwhelming challenge.

The cultural diversity of the data collectors involved in the IPY presents yet another challenge for data managers. Convincing them of the usefulness of sharing their data requires a lot of imagination, and a well-balanced usage of sticks and carrots (the latter being far more efficient). For example, depending on the cultural background of the scientists, the tools to reward their efforts towards data publication, will not be the same. This approach requires rapid adaptation efforts from both parties.

IPY data managers have established a working vision to guide their efforts for a useful vision for science in general (Parsons et al. 2011). Data should be discoverable, open, linked, useful, and safe.

More specifically, data should be:

1. capable of being located, identified, and generally accessible using simple tools
2. openly accessible. We find the current Wikipedia definition describes our intentions well: “Open data is a philosophy and practice requiring that certain data be freely available to everyone, without restrictions from copyright, patents or other mechanisms of control”
3. citable. So scientists can be acknowledged for their effort to make these data available and to provide a mechanism of validation of these efforts
4. interrelated and connected. There are many ways to interconnect data and we use the term “linked data” rather generally
5. usable for a practical, advantageous purpose or in several ways by defined but possibly very different users
6. protected from risk, corruption, and loss; now and over the long term.

These are simple terms or concepts, but they sum up the primary objectives that should be sought in managing diverse scientific data.

As biologists, the perspective on data management displays certain similarities to that described by Star and Griesemer (1989): the huge amounts of data generated during the IPY can be referred to as a “data ecosystem”, including the people, technical infrastructures and technologies as well as the multiple interactions between them. These systems undergo selection processes and evolve by adapting to constantly changing needs from the communities they are serving. In this respect, there are many exciting aspects to be explored by IPY data managers. For example, IPY data ecosystems present large numbers of niches that emerge, co-evolve, and survive (or do not) depending on environmental context. If this context changes rapidly, components of the ecosystems need to cope with these variations, or they will go extinct. Ecosystems are considered most healthy when diverse, and this is why it is important that IPY data are discoverable and available in a large number of contexts. The system also includes new and legacy technologies and ever-growing

data along with their countless potential mutations, provoked or undesired. The system must be understood holistically in order to optimize not just the individual components, for the benefit of society.

To help understand the IPY data ecosystem and its components and interactions, we describe an example of information network, SCAR-MarBIN (Scientific Committee on Antarctic Research—Marine Biodiversity Information Network, [www.scarmarbin.be](http://www.scarmarbin.be)) and its successor (ANTABIF, the Antarctic Biodiversity Information Facility, [www.biodiversity.aq](http://www.biodiversity.aq)), which both focus on biodiversity data and which took up the challenge of managing data from an extremely ambitious IPY program, the Census of Antarctic Marine Life (CAML).

## 2.2 The Antarctic Biodiversity Data Ecosystem

Initially, the SCAR-MarBIN initiative was launched during the CAML and IPY timeframe to provide the community with the tools to establish a web-based inventory of the Antarctic marine biodiversity, from micro-organisms to whales, that would be openly accessible to all. In the course of the IPY, CAML coordinated international expeditions that collected an unprecedented mass of information on the diversity, occurrence and abundance of Antarctic marine species. During this effort, thousands of species new to science have been discovered, in particular in the Antarctic deep sea (Brandt et al. 2007).

At the beginning, the intention of SCAR-MarBIN was to bring in as many data types as possible, that made sense for biodiversity research. But given the extreme variability of data that could potentially be considered as biodiversity-relevant (ranging from molecular to ecosystems realms) our endeavour had to adjust its objectives to a more pragmatic approach, with a philosophy to focus on robust standards that would allow the system to expand in the future, jointly with new projects, and with other disciplines. Hence, SCAR-MarBIN focused on three main data types: taxonomy (the names and classification of living things), biogeographic data (where these living things are found) and metadata (descriptions of datasets).

Establishing a complete and accurate inventory of the presently described Antarctic marine fauna and flora had never been attempted so far, and required a strong commitment of the Antarctic research community, as well as a robust data management system. Thanks to a collective effort of a board of 64 taxonomic editors, each of them being a renown specialist, and the support of the World Register of Marine Species, the first Register of Antarctic Marine Species (RAMS) was compiled and published (De Broyer and Danis 2010), building upon a first block contributed by Clarke and Johnston (2003). RAMS includes an accurate list of more than 8,200 valid species with an up-to-date systematic classification comprising more than 17,500 taxon names. This taxonomic information is constantly completed by the experts. RAMS contributes to the “taxonomic backbone” of the SCAR-MarBIN, ANTABIF, OBIS and GBIF networks, to establish a dynamic, knowledge-based information system on Antarctic marine biodiversity for the future.

The biogeographic information component of SCAR-MarBIN (ANTOBIS) was progressively built by linking databases, distributed across the planet. A single access is now available through the SCAR-MarBIN data portal, which totals 198 datasets and 1.3 million records. This information can be visualized, mapped and downloaded from the portal, for free.

Although it is focused primarily on biogeographic data, SCAR-MarBIN also develops tools to improve its usefulness for conservation and sustainable management. Data can be processed and presented in a variety of ways, allowing the integration of biogeographic data into the development of policy for Antarctic conservation. Furthermore, current data can be used to construct a panel of indicators to detect changes in biodiversity and hence adapt conservation strategies.

The new ANTABIF project builds upon these grounds and has been building a completely new architecture, strongly relying on Open Source standards and technologies provided by the Global Biodiversity Information Facility (GBIF). This architecture includes a series of websites, all focusing on a specific data type or data-derived product. The philosophy is to continue providing free and open access to Antarctic biodiversity data, while at the same time building new projects to showcase the power of a linked, community-driven tool.

## 2.3 Findings and Motivations

CAML and its sister project, SCAR-MarBIN, have been instrumental in facilitating, supporting and highlighting the work of students and early-career Antarctic scientists. CAML enabled this new generation of polar scientists to attend meetings, workshops and even scientific voyages to Antarctica during IPY through direct funding and in-kind support. SCAR-MarBIN made new datasets and resources freely available to all, enabling students and scientists around the world to access information and data previously only available in a handful of large institutes. Perhaps most significantly, CAML gave this new generation of scientists a voice and presence in the international scientific arena and the global media. Many of the CAML highlights, online diaries and press releases focused on the work of young scientists at the cutting edge of polar science and helped to bring this work to the attention of the world.

The legacies of CAML and the ongoing development of SCAR-MarBIN have begun to include practical tools and resources for use in conservation, planning and management in the Antarctic and Southern Ocean. As a result of a single CAML cruise, the Commission for the Conservation of Antarctic Living Resources (CCAMLR) declared two Vulnerable Marine Ecosystems in late 2008, based on images and samples collected by the R/V Aurora Australis. The areas, each comprising around 400 km<sup>2</sup>, have a high diversity of marine life, in particular habitat-forming hydrocorals and sponges. They are also home to species previously unknown to science. This declaration ensures that these unique areas are not damaged by indiscriminate fishing practices. The wealth of biodiversity,

distribution and biogeographic data collected under the CAML umbrella form a unique insight into the distribution of marine life in the Southern Ocean. Such a resource, whilst incredibly valuable for scientific research, becomes the basis of a range of data-derived products including scientific synthesis papers, the Antarctic Field Guides and Biogeographic Atlas. These products feed into international efforts to create and classify new Marine Protected Areas and Vulnerable Marine Ecosystems within the Southern Ocean.

To attract more users and showcase the usefulness of the SCAR information networks, a series of data derived products are developed, also using a community-based approach. As examples, ANTABIF and SCAR-MarBIN are developing the Antarctic Field Guides (AFG) and the Biogeographic Atlas of the Southern Ocean (BASO) projects.

The objective of the AFG project is to develop a community-driven, online identification aid, which takes the form of an interactive, customizable field guide ([afg.biodiversity.aq](http://afg.biodiversity.aq)). The field guide offers access to dynamically generated information (built on the fly from the content of RAMS, SCAR-MarBIN, ANTABIF, and AFG custom databases), and allows the user to create his/her own field guide to the Antarctic wildlife, in function of the region he/she is interested in, or the taxa he/she wants to focus on. The user can download and share his custom field guide as a pdf file, and allow others to build upon it. As part of a community effort, a substantial amount of information has already been made available by A Clarke and D Barnes (British Antarctic Survey) and S Schiaparelli (University of Genova). The content of the AFG is progressively built and focuses on high-quality pictures, which are useful for the identification of Antarctic organisms. In the long run, we hope to couple the AFG with an online Identification Keys system to further help Antarctic scientists to identify what they find in the field using modern, interoperable tools.

The Antarctic Field Guides is an Open Source project.

The multi-authored Biogeographic Atlas of the Southern Ocean (BASO) will provide an up-to-date synthesis of Antarctic and sub-Antarctic biogeographic knowledge and a new comprehensive online resource for visualisation, analysis and modelling of species distribution. The major objectives of the Biogeographic Atlas of the Southern Ocean are to establish a new synthesis of the biogeography of the Southern Ocean (patterns and processes), covering benthos, zooplankton, nekton, birds and seals, to provide a benchmark of current biogeographic knowledge, to model predictive distribution of species and assemblages according to abiotic factors and to develop dynamic editions of the Atlas, including a dedicated webportal ([atlas.biodiversity.aq](http://atlas.biodiversity.aq)) and electronic books, connected to our network of interoperable databases, allowing easy access to the new biogeographic syntheses. It is envisioned to offer access to updated occurrence and environmental data, customized mapping and data visualisation and analysis, thereby facilitating the ecoregionalization process and the modelling of predictive distribution of species and assemblages in the context of environmental changes. So far a total of over 60 specialists (taxonomists, ecologists, modellers, developers) are contributing data

and expertise to this project. In the long term, it is hoped that this approach can be expanded to other realms.

The Biogeographic Atlas of the Southern Ocean is an Open Source project.

## 2.4 The Biodiversity Data Paper Concept

The new Data Paper concept is a new incentive for biodiversity researchers to publish their data (Chavan and Penev 2011). As discussed previously, free and open access to biodiversity data is essential to feed emerging data-intensive science, and also to develop informed, adaptative conservation of Antarctic biodiversity, currently challenged by rapid environmental change. Even if SCAR-MarBIN gives access to an unprecedented of expert-validated data, we suspect that a significant amount of primary data are not yet easily accessible nor discoverable. Among potential impediments, one is a lack of incentives to data publishers for publishing of their data resources. This activity is often considered as time-consuming and unrewarded in the short term. The Data Paper concept brings up conventional scholarly publication of enriched metadata (detailed description of datasets).

Integrated in the new ANTABIF workflow, the GBIF Integrated Publishing Toolkit ([ipt.biodiversity.aq](http://ipt.biodiversity.aq)) is an online informatics suite which allow users (individual scientists, institutions,...) to standardize, manage, document and publish their primary biodiversity data. Once the user decides, new datasets are exposed on the web, and become available to various data aggregators, which expose the data using dedicated data portals. The user can opt for the option of generating a manuscript (Data Paper), from the IPT metadata, and submit it for peer-reviewing and eventual publication in a journal accepting this type of contribution.

Using the IPT, the mechanism is in place for incentivizing data publishers efforts towards easy, efficient and enhanced publishing, dissemination, sharing and re-use of biodiversity data. The Data Paper is one possible mechanism to offer scholarly reward for efforts and investment by scientists in enriching metadata and publishing them as citable academic papers.

The pioneering Data Paper concept arises from a joint effort led by GBIF in collaboration with Pensoft Publishers through its journals *Zookeys*, *PhytoKeys*, *MycKeys*, *BioRisk*, *NeoBiota*, *Nature Conservation* and the forthcoming *Biodiversity Data Journal* which all accept IPT-generated manuscripts.

It is hoped that in addition to rewarding those publishing their biodiversity data, the Data Papers concept will also speed up the publishing of fit-for-use, well documented biodiversity data resources. However, to be proven, the concept requires a high degree of commitment by the SCAR communities, starting by the biodiversity researchers and expanding towards other scientific fields.

## 2.5 The Future: Towards True Integration

ANTABIF (Antarctic Biodiversity Information Facility), funded by the Belgian Science Policy Office, takes a step further and is building a dedicated Antarctic biodiversity data portal providing access to a distributed network of contributing database, according to the principles of the Global Biodiversity Information Facility. Building upon the SCAR-MarBIN experience, ANTABIF gives access to biodiversity data from marine, terrestrial and limnetic realms.

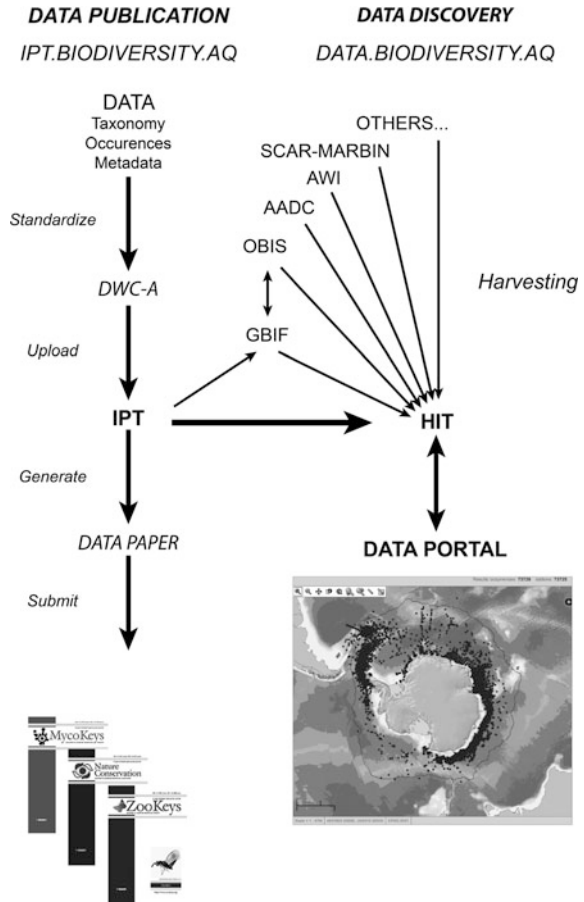
The design of ANTABIF architecture has been the object of special attention in order to offer an added value to the user and partners, by integrating the best available resources and standards, such as those developed in the frame of SCAR-MarBIN V2 ([data.scarmarbin.be](http://data.scarmarbin.be)), building upon webservice-oriented, dynamic technologies. Careful attention is also taken to make the best possible use of GBIF informatics tools. Using this approach, the Antarctic biodiversity networks will be interoperable with many entities, allowing the publication of the data in many different contexts, leveraging the power of Web2.0 and collective intelligence.

The ANTABIF technological ecosystem is largely based on the progress made in the framework of the SCAR-MarBIN V2 data portal. It is based on 100 % Open Source solutions, and integrates the latest GBIF informatics components (Harvesting and Indexing Toolkit (HIT) and Integrated Publishing Toolkit (IPT)). The philosophy is to try and see as far as possible by adopting robust standards, which have been widely accepted by their respective community. The main principles are details in Fig. 2.1.

Another important aspect in the design of the ANTABIF infrastructure is the use of advanced, highly customizable search engines. The search engines are schema-free and document based, which means that the user can enter a complex query in the search box. As a community-based initiative, we also thought that the data portal should be able to learn from the behaviour of its users, a feature that is embedded in the current design. As users query the data systems, the search engine is capable of refining the hits it will feedback to the user. ANTABIF uses customized GBIF's IT infrastructure tools, which are foreseen to become widely adopted by the biodiversity science community. For example, the Integrated Publishing Toolkit (IPT) allows ANTABIF to push newly published data to global scale networks such as GBIF and OBIS. GBIF's Harvesting and Indexing Toolkit (HIT) allows ANTABIF to harvest data from various sources, for example from institutions with relatively advanced IT infrastructure. To ensure no dataset is harvested twice, GBIF maintains a registry, which keeps a trace of all existing and new biodiversity datasets available for publication.

A similar approach is adopted for metadata, to maximize the probabilities to discover available datasets through search engines such as Google. The metadata is stored in a central repository, the Global Change Master Directory (GCMD), which also serves as a repository for other entities, such as the Australian Antarctic Data Center (AADC), OBIS, SCAR-MarBIN and SCAR (through the Antarctic Master Directory, AMD). Using this integrated approach, ANTABIF is in line with

**Fig. 2.1** How ANTABIF builds geospatial data content, and interacts with other components of the data ecosystem



the vision developed by IPY data managers, allowing primary biodiversity data to be discoverable, open, linked, useful and safe.

We have high hopes that certain elements of the architectural design adopted for ANTABIF is transposable to other disciplines than biodiversity: adopting open standards, and promoting them, developing a few open source tools which can be used in a wide variety of contexts, offering a simple, flexible and responsive platform to discover, visualize and access data openly, is a rather generic method. Our approach, once checked against reality (and extensively tested by its community in an iterative way), needs to be promoted and communicated to a wider audience, so our efforts can be rewarded by being adopted in other contexts, or networks. Again, this demands strong commitment from the community of users.

At the basis of SCAR-MarBIN’s success, we believe that personal communication is a key factor. In the context of ANTABIF, once the online products are launched, constant efforts should be sought between data providers and data managers, to ensure the latter do not get disconnected from the community they

serve. This need for communication is not new, and was already stressed by Vickery (1999), finding that “Scientists and technologists need to interact and communicate”. Data managers are not just technologists, and having them actually take part in field activities together with scientists, who should reversely also play an active role in data curation and tool developments, is an elegant way of maintaining mutual trust and understanding of each other’s reality. Only in this way will scientists feel they are part of their own data ecosystem, and take more responsibility in its management and get involved in their developments. This facet of scientific data management was already stated by Star and Ruhleder (1996): “We must ask users to meet designers halfway by learning their language and developing an understanding of the design domain. If designers are at fault for assuming that all user requirements can be formally captured and codified, users are often equally at fault for expecting “magic bullets”—technical systems that will solve social or organizational problems”.

Another idea is to ensure technologies adapt to the evolving behaviour of the users, by becoming more sophisticated and flexible. Giving access to new types of information including actual samples, digital data, documentation, images, multimedia, publications, etc., is a good example. These objects could be linked together across disciplines as in a data network. Collaborative research efforts are highly needed to explore how to capture and enrich this kind of contextual information. Another crucial (and complex) aspect is the capture of data uncertainties related to the newly available knowledge and assumptions of any data collector (Parsons et al. 2010).

The need for research and development in polar biodiversity informatics highlights another important issue in what we are facing in sustaining data systems: sponsorship of the data ecosystem. During the IPY, the final report of the Data Policy and Management Subcommittee includes the following note: “In the period leading up to the start of IPY, data stewardship was undervalued, despite robust data management plans within the IPY Framework Document, the strong recommendations of the ICSU Program Area Assessment, and telling examples from earlier international projects” (Parsons et al. 2011).

At the beginning of the IPY it was extremely difficult to obtain data management funding, especially for broad international or interdisciplinary support. SCAR-MarBIN was an exception to this rule, as the Alfred P. Sloan Foundation through seed money, and the Belgian Science Policy Office supporting the setup of the network in congruence with the Census of Antarctic Marine Life initiative. By the end of the IPY, more governments began to endorse information systems, which was in fact too late. It is time for funding agencies to realize that if they fund data collection, which in the case of polar research can be extremely costly, they also need to fund data stewardship, acknowledging their responsibility in enhancing the data availability, which is a key to improving Polar Science. In some cases, progressively shifting towards an internationally funded model makes a lot of sense, as information networks benefit the wide community, taking the forms of a common.

Ideally, sponsors should be actively engaged in the planning and execution of data systems for ambitious research programs, offering the possibility to have a very long planning period. The ideal situation would be one where the scientists personally know the data managers, know which data centre they can submit their data to, and are familiar with standards. This was not the case during the IPY during which data systems were thought of, when expeditions were ongoing, or were over.

We are facing many challenges, and the fact that the data are still highly distributed and stored in many different types of institution in many different formats is probably one of the most fundamental ones. The diversity of the needs of the data users are also extremely diverse and often unpredictable. Some short-term technical solutions to these challenges include the use of more open, cloud-based approaches of data casting, using simpler approaches in the architectural design of information systems, based on proven concepts and robust standards, and iterate to progressively interoperate with other systems and communities.

To be consistent with some of the norms guiding scientific research (such as rigor, transparency, integrity, and reproducibility), open access to data should actually be a central rule: scientific data should be viewed as a common good that is open (except for legitimate ethical reasons). Also, data should be used in a framework where data providers are given fair attribution for their contribution to these commons. The example of the Data Paper concept is one step in this direction. Data managers should ensure the integrity of the data that they are entrusted to and provide the necessary documentation and context to support robust science. This is of course a long-term vision, but a simple element, such as requiring research proposals to include detailed data management and archiving plans, would probably help.

In the course of the IPY, the rapid and dramatic changes in the polar regions environment has been extensively documented, and we can expect that the demand for rigorous, transparent and scrupulously documented data will increase dramatically in the near future. Many challenges remain, but it does not help to deny our collective responsibility to keep the data safe, to avoid vanishing of common knowledge before it even reaches future generations.

## 2.6 Summary

The International Polar Year was an unprecedented research effort, which involved thousands of participants from many nations. This effort generated massive amounts of extremely diverse data, which need to be interrelated to understand complex problems such as environmental change and its impact on Antarctic biodiversity. We use the experience of SCAR Marine Biodiversity Information Network (SCAR-MarBIN), and the new Antarctic Biodiversity Information Facility (ANTABIF) as a case study to examine an example of IPY data management approaches seeking to address issues around such complex scientific

questions. In the present chapter, we discuss the different facets of this data management experience, and the design used for ANTABIF to be transposed to other contexts.

To ensure that ANTABIF and its successors are successful in maintaining open access to Antarctic biodiversity data, the heterogeneity and the distributed nature of data will require technologies to be much simpler, well standardized and agile to follow the constantly changing needs of the community it is serving. The IPY has shown that the Polar Regions are changing rapidly and the need for rapid technical and cultural changes, towards true openness becomes urgent. We suggest several short and long-term strategies to facilitate the evolution of the overall SCAR data ecosystem towards universal access to high quality scientific data, which would otherwise be lost.

**Acknowledgments** This book chapter is contribution #78 to the Census of Antarctic Marine Life initiative. The authors wish to thank the vast community involved in the biodiversity information networks, and in particular the Scientific and Follow-up Committees of SCAR-MarBIN and ANTABIF. The two initiatives are funded by the Belgian Science Policy Office and are implemented within the Belgian Biodiversity Platform. We would also like to thank the different sponsors, including: the Alfred P. Sloan Foundation, The Total Foundation for biodiversity, the Census of Antarctic Marine Life, The Alfred Wegener Institute, the German Research Foundation, the Scientific Committee on Antarctic Research, the Australian Antarctic Division, the Netherlands Organization for Scientific Research.

## References

- Brandt A, Gooday AJ, Brandão SN, Brix S, Brökeland W, Cedhagen T, Choudhury M et al (2007) First insights into the biodiversity and biogeography of the Southern Ocean deep sea. *Nature* 447:307–311 doi:[10.1038/nature05827](https://doi.org/10.1038/nature05827)
- Carlson DJ (2010) Why do we have a 4th IPY? In: Barr S, Lüdecke C (eds) *The history of the international polar years (IPYS)*. Springer, Berlin, pp 299–312
- Carlson DJ (2011) IPY 2007–2008: where the threads of the double helix and Sputnik intertwine. In: Huettmann F (ed) *Protection of the three poles*. Springer, Japan
- Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12 (Suppl 15):S2. doi:[10.1186/1471-2105-12-S15-S2](https://doi.org/10.1186/1471-2105-12-S15-S2), <http://www.biomedcentral.com/1471-2105/12/S15/S2>
- Clarke A, Johnston NM (2003) Antarctic marine benthic diversity. *Oceanogr Mar Biol Ann Rev* 41:47–114
- De Broyer C, Danis B (2010) How many species in the Southern Ocean? Towards a dynamic inventory of the Antarctic marine species. *Deep-Sea Research II: Topical Studies in Oceanography*. doi:[10.1016/j.dsr2.2010.10.007](https://doi.org/10.1016/j.dsr2.2010.10.007) <http://dx.doi.org/10.1016/>
- Hey T, Hey J (2006) e-Science and its implications for the library community. *Libr Hi Tech* 24:515–528
- Hey T, Tansley S, Tolle K (eds) (2009) *The fourth paradigm: data-intensive scientific discovery*. Microsoft Research, USA
- Newman HB, Ellisman MH (2003) Orcutt JA (2003) Data-intensive e-science frontier research. *Commun ACM* 46:68–77
- Parsons MA, Duerr R, Minster JB (2010) Data citation and peer-review. *Eos, Transactions of the American Geophys Union* 91:297–298

- Parsons MA, Godoy O, LeDrew E, de Bruin TF, Danis B, Tomlinson S, Carlson D (2011) A conceptual framework for managing very diverse data for complex, interdisciplinary science. *J Inform Sci* (Oct). doi:[10.1177/0165551511412705](https://doi.org/10.1177/0165551511412705), <http://jis.sagepub.com/cgi>
- Star SL, Griesemer JR (1989) Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Soc Stud Sci* 19:387–420
- Star SL, Ruhleder K (1996) Steps toward an ecology of infrastructure: Design and access for large information spaces. *Inform Syst Res* 7:111
- Vickery B (1999) A century of scientific and technical information. *J Documentation* 55:476–527