DATA CENTRES AND DATA AVAILABILITY

Observations, measurements and models are an integral part of global change research and the resulting datasets form the basis for scientific publications. Data collection, curation and access are essential for reviewers and readers who wish to verify scientific findings. Future use of data not only depends on their availability, but also on the way they are archived. Issues addressing Earth system changes need to take global datasets into account. However, at present, research data are generated and stored by individual scientists and projects, which restricts the sharing of data with a broader research community. It is therefore highly desirable that such distributed data be made freely available to all users (Open Access) in a harmonised, machine-readable form that facilitates the exchange, compilation, processing and analysis of data.

Libraries are tasked with the long-term conservation and provision of access to printed publications. The Internet has added new possibilities for direct access to digital documents through publisher online catalogues, portals and search engines. However access to and distribution of digital objects is limited if the right infrastructure does not exist. The dynamic Internet and ever-changing technologies frequently lead to a "file not found (404)" error message when pages have been moved or deleted. Commercial publishers were among the first to introduce persistent identifiers in order to preserve access to electronic Web resources and ensure their availability over time. At present, more than 50 million Digital Object Identifiers (DOI®) are registered in a system operated by the International DOI Foundation. DOIs are currently mainly attributed to journal articles, and there are now global calls for primary data to also be given a persistent identifier. The "Publication and Citation of Scientific Primary Data" (STD-DOI) project, funded by the German Research Foundation (DFG), led, in 2009, to the establishment of DataCite, a partnership of leading research libraries and information providers set up to improve reliable access to research data on the Web by sustainable archiving, citation and identification of datasets by means of the DOI system (HTTP://WWW.DATACITE.ORG).

Data centres have been in existence for around the last 50 years, during which time the archiving of and access to digital objects has suffered from constantly changing storage media and formats. Many valuable data have been lost due to defective tapes and discs in cases where no backup was available. In addition, many disciplines did not take the archiving of research data very seriously. Nowadays, new electronic data handling tools prevent the loss of data by migration from one server to another. The capacity of storage systems has been increased to the petabyte range and the recent developments provide scientists with future-proof, long-term access to digital data from anywhere in the world through the combination of archives into science-specific portals.

During the International Geophysical Year 1957/58 the World Data Center System (WDC) was established by the International Council for Science (ICSU) in order to archive and distribute data from Earth system research. The WDC now includes 52 centres (German contributions see below) in numerous countries around the world that provide online access to global change research data. The 29th General Assembly of ICSU took the decision to create a new World Data System (WDS, http://icsu-wds.org) whose objectives were to move away from current stand-alone WDCs to a common globally interoperable distributed data system and become a global "community of excellence" for scientific data. A prototype data portal is operated in Germany as a common entry point for the broad range of existing data sources. Any organisation that is in possession of relevant data is encouraged to join the new WDS. Germany will contribute to WDS with the German WDC cluster for Earth System Research and the data publisher PANGAEA (HTTP://WWW.PANGAEA.DE).

Any discussion relating to the archiving of and access to research data always homes in on the huge quantity of data requiring enormous storage capacities. For example, a satellite like CryoSat,

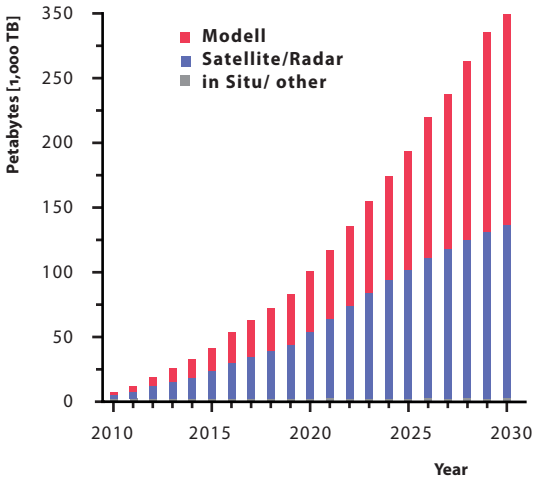

Fig. 1: Future data increase in climate research (Oberpeck et al. 2011, DOI:10.1126/science.1197869).

Since 2001 Germany contributes to the World Data Center system with three centres:

The ***World Data Center for Climate*** (WDCC) offers data management consulting for climate models over the lifetime of the data. In 2010, the quantity of model data available on the Web has grown to >400 TB. As a Data Collection or Production Centre (DCPC), the WDCC is part of the World Meteorological Organisation's information system. As part of the IPCC Assessment Reports, WDCC is one of three data nodes for climate model data collection that operates in cooperation with the Program for Climate Model Diagnosis and Intercomparison (PCMDI, USA) and the British Atmospheric Data Centre (BADC, UK).

HTTP://WWW.WDC-CLIMATE.DEW

The ***World Data Center for Marine Environmental Sciences*** (WDC-MARE) is aimed at collecting, scrutinising, and disseminating data related to Earth system research in all fields of marine sciences. The centre also publishes WDC-MARE reports of scientific results as provided by the document centres of research institutes. Projects that use WDC-MARE for data curation purposes, deal mainly with environmental, geological, biological, physical and chemical oceanography.

HTTP://WWW.WDC-MARE.ORG

The ***World Data Center for Remote Sensing of the Atmosphere*** (WDC-RSAT) provides a growing collection of atmosphere-related satellite-based datasets, information products and services. It focusses on atmospheric trace gases, aerosols, dynamics, radiation and cloud physical parameters along with complementary information. The dissemination of information is achieved either by giving free access to data stored at the centre or by acting as a portal that links up to external sources. WDC-RSAT is a member of the WMO-WDC group and serves as a management platform for the Network for the Detection of Mesopause Change (NDMC).

HTTP://WDC.DLR.DE

The ***German Research Centre for Geosciences*** (GFZ) offers geological data resulting from scientific drilling operations. Most of the data are accessible through a central portal and through discipline-specific portals, such as the World Stress Map (WSM), the Satellite Data Centre (ISDC), the Scientific Drilling Database (SDDB) and the GEOFON Seismological Network. Datasets are assigned a DOI and are citeable. The new portal and the publication of data are offered as a joint service of the GFZ Centre for Geoinformation Technology (CeGIT) and the GFZ library of "Wissenschaftspark Albert Einstein".

HTTP://WWW.GFZ-POTSDAM.DE:80/PORTAL/GFZ/SERVICES/FORSCHUNGSDATEN

CryoSat will produce 50 GB of data per day. (Image credit: ESA)

GERMAN WDC CLUSTER FOR EARTH SYSTEM RESEARCH

which was recently launched to monitor the behaviour of polar ice, will produce around 50 GB of raw data per day. Another major aspect that needs to be taken into account, but is rarely mentioned, is the variety of measurements taken in all parts of the geosphere. Data repositories are faced with the challenge of having to cope with a complex range of datasets with tens of thousands of variables produced by all the different academic disciplines involved in global change research. Smart data models are therefore required, such as the data library and publisher PANGAEA, which is able to handle this highly diverse output of geoscientific research.

Besides curiosity to find out how the Earth works, the recognition that researchers receive for their work is another major driver of scientific progress. It can therefore be safely assumed that standards for data citation that give credit to the author give researchers incentives for data sharing. Consequently, data archiving needs to become an integral part of the established scientific publication process, and most importantly needs to provide appropriate citable and reliable access to research data. To improve the current situation, scientists from Germany and the UK have initiated the journal Earth System Science Data (ESSD, **HTTP://WWW.EARTH-SYSTEM-SCIENCE-DATA.NET**), the first ever journal aimed at the publication of original research data. The first publication made available an eight year time series of ozone profiles from the Antarctic station of the former German Democratic Republic (*see Fig. 2*).

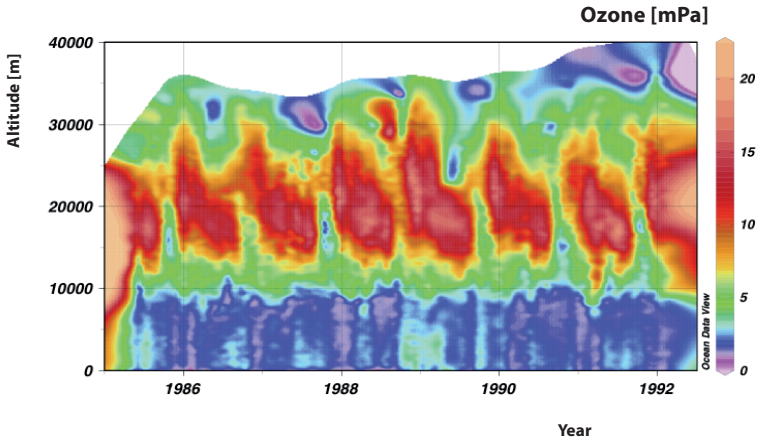

Fig. 2: Ozone time series from the Antarctic - the first publication in the data journal ESSD (König-Langlo & Gernandt 2009, DOI:10.5194/essd-1-1-2009).