F Climate Research

Ilse Hamann

1 Climate science

The general public awareness that Earth's climate is an important conditioning factor for the quality of human life has grown because earth science has in the past decades produced and publicised many new and fascinating facts about the dynamics of our complex climate system. Practically everybody experiences, reflects upon, communicates about and depends on the weather and climatic conditions in earning his/her livelihood. Access to often vital information is a prerequisite for strategic planning and taking economically sound decisions.

Climate change challenges the stability of human and environmental systems alike, therefore the interest in a reliable database of global and regional earth system data is great, since such data are needed to find out how resilient these systems are, how climate shifts impact different sectors of environment and society and how to optimally adapt to the increased variability of our climate.

In Germany the National Committee on Global Change Research (NKGCF)[1] has for the past 15 years coordinated German contributions to global change research, with the consideration of the global change-related decisions of the senate commissions of the German Science Foundation (DFG) and of the advisory board of the German Federal Ministry of Education and Research (BMBF). The NKGCF is also the national contact point for the international global environmental change programmes IGBP,[2] WCRP,[3] DI-

[1] http://www.nkgcf.org.

[2] International Geosphere-Biosphere Programme, http://igbp.sv.internetborder.se.

[3] World Climate Research Programme, http://www.wcrp-climate.org.

VERSITAS[4] and IHDP[5] as well as for the Earth System Science Partnership (ESSP).[6] The NKGCF's 15 voting members[7] are working in natural and social science areas with relevance to global change.

Climate science derives quantitative information from observations of key climate variables and from simulations using mathematical, numerical models. Many different types of data are useful for statistical and dynamical approaches in climate research. Observational data from in-situ measurements in the atmosphere, on land and in the oceans form one strong pillar of the database. They are, however, irregularly distributed in space and time, and the global coverage of such measurements of the key variables is insufficient. Remotely sensed data of near-surface properties at Earth's surface augment the in-situ data by highly improved spatial and sometimes temporal sampling density.

On the basis of these data and well-established physical principles, mathematical models of the dynamical behaviour of the atmosphere, the oceans and the cryosphere have been developed and numerical experiments are being carried out with coupled models that involve not only these physical realms, but also biogeochemical processes in and between them. A detailed discussion of the capacity of global climate models to reproduce observed features of the recent climate, investigate past climate changes and produce credible quantitative estimates of the climate development in the future, i.e. possible climatic "futures", is given, for example, by Randall, Wood, Bony et al. (2007) in chapter 8 of Solomon et al. (2007)[8]

The findings described in the following sections of this chapter are to a large extent based on an analysis of documents available to me at the in-

[4] Integrating biodiversity Science for human well-being, http://www.diversitas-international.org.

[5] International Human Dimensions Programme on Global Environmental Change, http://www.ihdp.unu.edu.

[6] http://www.essp.org/index.php?id=10.

[7] http://www.nkgcf.org/committee_members.php?year=2009-2011.

[8] Much of this chapter describes the research infrastructure in climate modelling, on the basis of which the last so-called "IPCC Report" or Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) was written. This Status Report was widely publicised when the IPCC together with Al Gore was awarded the Nobel Peace Prize 2007 "for their efforts to build up and disseminate greater knowledge about man-made climate change, and to lay the foundations for the measures that are needed to counteract such change". The contribution of Working Group I to this Assessment Report, i.e. "Climate Change 2007: The Physical Science Basis", provides extensive descriptions of the complex processes of Earth's climate system (Solomon et al. 2007). The IPCC was established by the World Meteorological Organization (WMO) and the United Nations Environmental Program (UNEP) to assess scientific information on climate change. The IPCC regularly publishes reports that summarise the state of climate science: http://www.ipcc.ch/publications_and_data/publications_and_data_reports.shtml.

stitution where I work, i.e. in the Data Management department (DM)[9] of the German Climate Computing Centre (DKRZ, Deutsches Klimarechenzentrum),[10] or are part of websites of other institutions in the climate research community. Additional information about aspects of the research infrastructure in climate science comes from what I could glean from the literature and by receiving answers to a questionnaire (Spohr, 2010, see Appendix 1) from senior scientists working at six different institutes in teams conducting research or providing service(s) in this field. In the next sections, frequent reference is made to these information sources, and many of the diagrams and figures shown have been prepared by my colleagues and by the principal investigators of large research projects for a variety of purposes. The figures include depictions of workflows, organisational diagrams highlighting collaborative aspects and easy-to-grasp displays of pathways of information, data and scientific results within the research cycle.

When planning the scope of my work for this study, i.e. writing the climate science subject-specific chapter, I felt that I needed to strike a balance between a broad survey of climate research programmes and institutions worldwide and a highly selective case study of the research infrastructure at DM/DKRZ and at the World Data Center for Climate (WDCC,[11] maintained by DM/DKRZ), for which I would be able to provide much more detail (WDCC is partner in this Task 7.1 in OpenAIRE). I opted for an eclectic analysis of what I think are important entities in Germany operating at different locations/stages of the climate research cycles and in the climate data web. Job titles of climate scientists working in these organisations and institutions have many facets as there are linkages of climate science with many other disciplines.

First, in section 2, descriptions of the workflows adopted for the presently worldwide largest international climate change research project illustrate relevant climate modelling research infrastructure. In section 3, an overview is given of the major data centres that curate data from observational climate science programmes in order to describe the infrastructure that is in place in Germany for this second large branch of climate science. Section 4 contains the results of a survey of selected climate researchers with differing scientific foci. The survey intended to determine current practices (workflows) at institutions in Germany regarding climate research, climate data management and the kinds of services provided on behalf of the climate science community for a variety of end users, as well as Open Access policies and incentives with respect to literature and data in the respective organisation.

[9] http://www.dkrz.de/daten-en.
[10] DKRZ is a non-profit and non-commercial limited company with four shareholders.
[11] http://www.dkrz.de/daten-en/wdcc?set_language=en.

Open Access aspects, practices and policies with respect to literature and data, whether already implemented or being presently developed in the climate research groups and institutions which I have surveyed, are described in sections 5 and 6, respectively. At the chapter's end, challenges for climate science research infrastructures are mentioned, and an outlook towards upcoming developments is given. Finally, some pointers are given that relate to the Open Access infrastructure in the field of climate science.

2 Current status of climate modelling research infrastructure

2.1 Data management and WDCC at the German Climate Computing Centre (DM/DKRZ and WDCC/DKRZ)

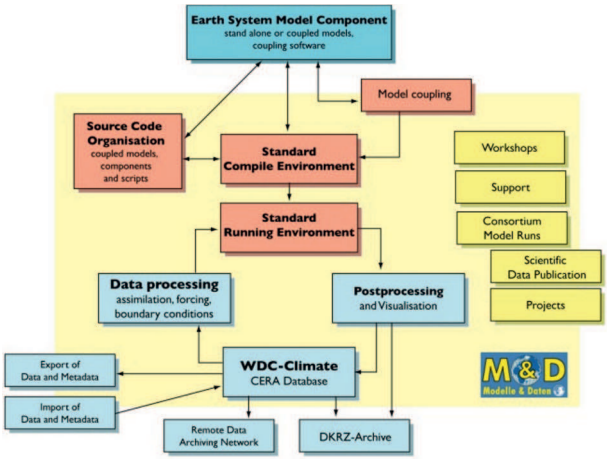

Figure F.1 M&D services (the diagram was taken from the static mirror of the original M&D website)[12]

Since the time when DM/DKRZ was still called "Model & Data" (M&D, during the decade 2000–2009[13]) and when it was part of the Max Planck Institute for Meteorology in Hamburg, it has availed itself of the highly developed research infrastructure characteristic for the Max Planck Society, possibly the most renowned research organisation in Germany.[14] The scien-

[12] http://www.mad.zmaw.de/service-support/index.html.
[13] http://www.mad.zmaw.de.
[14] http://www.mpg.de/183251/portrait.

tists at the Max Planck Institute for Meteorology in Hamburg (founded in 1975) "have been studying how physical, chemical and biological processes and human behaviour contribute to global and regional climate changes. The scientists develop numerical models and measurement methods to explain the natural variability of the atmosphere, the oceans and the biosphere and to assess the influence of land use changes, industrial development, urbanisation and other human influences. Together with the Max Planck Institutes for Biogeochemistry (Jena) and for Chemistry (Mainz), they strive to provide a better understanding of the chemical and biological factors that determine the concentrations of greenhouse and other trace gases in the atmosphere and how they interact with the terrestrial and marine biospheres".[15]

The scientists who worked in the M&D group and now belong to DM/DKRZ provide central support for the German and European climate research community, with an emphasis on development and implementation of best practice methods for Earth System modelling and related data management. Figure F.1 summarises the support services that M&D offered.

The DKRZ is a national service provider, providing high performance computing platforms, sophisticated and high capacity data management, and superior service for premium climate science.[16] DKRZ operates a fully scalable supercomputing system designed for and dedicated to earth system modelling.

Today the DKRZ supports the whole data life cycle of climate (model) data,[17] i.e. data creation, diagnostics, visualisation, archiving and dissemination to scientists all over the world. Therefore the day-to-day work of earth system scientists and the long-term archiving of scientific results is supported within a virtual research environment (see Figure F.2).

While this data cycle management corresponds in principle to the mission of the WDCC/DKRZ,[18] the WDCC restricts itself to curating mostly *data products* from climate modelling, since storage of *raw data* from satellites or climate models, for example, on a global basis is beyond the scope of the available facilities. In section 3, thematically corresponding data centres like Earth observation, meteorology, oceanography, paleoclimate and environment are described. WDCC is enhancing its cooperation with these centres with the aim to establish a complete network for climate data.

The WDCC's development during the past two decades:

- Scientific data management at DKRZ started in the 1990s, with the aim of collecting, scrutinising and disseminating data related to climate change on all time scales

[15] http://www.mpg.de/155345/meteorologie?section=all.
[16] http://www.dkrz.de/about-en/aufgaben.
[17] http://www.dkrz.de/daten-en.
[18] http://www.dkrz.de/daten-en/wdcc.

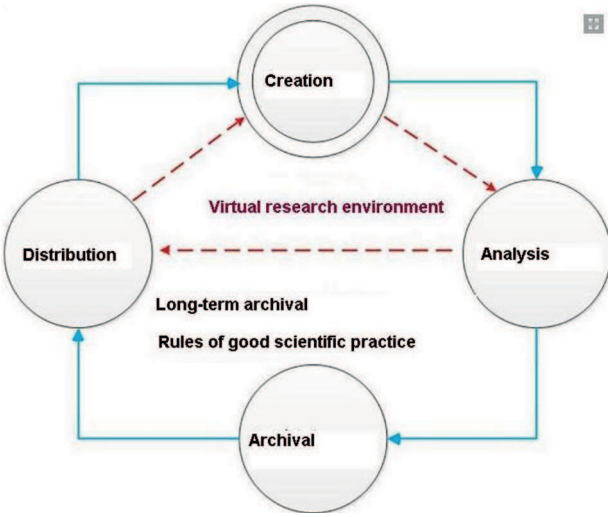

Figure F.2 Data life cycle and virtual research environment

- In 2001 DKRZ began the development of the Climate and Environmental Retrieval and Archiving (CERA-2) data model, which was a collaboration of M&D, the Alfred Wegener Institute for Polar and Marine Research (AWI)[19] and the Potsdam Institute for Climate Impact Research (PIK)[20]
- In 2003 the CERA database was accepted by the WDC for Climate (WDCC) of the International Council of Scientific Unions (ICSU, today called International Council for Science).[21] In August 2011 the WDCC was re-evaluated and accepted as a member into the new International Council for Science World Data System (WDS)
- In 2004/2005 the "consortium experiments" were established, i.e. "capacity computing" necessary for highly relevant national or international research.
- Since 2010 the WDCC has been integrated in the Data Management department of DKRZ.

The geographical distribution of WDCC users by continent and by country is shown on the WDCC "Statistics" web page[22] and in Figures F.3 and F.4.

[19] http://www.awi.de/en/institute.
[20] http://www.pik-potsdam.de.
[21] http://www.icsu.org.
[22] http://www.dkrz.de/daten-en/wdcc/statistics/wdcc-statistics-2010.
[23] Country Group 1 (1 user):

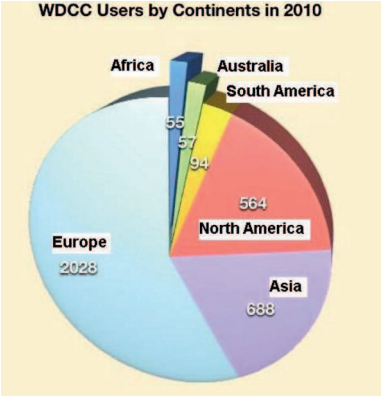

Figure F.3 Number of WDCC users by continent in 2010

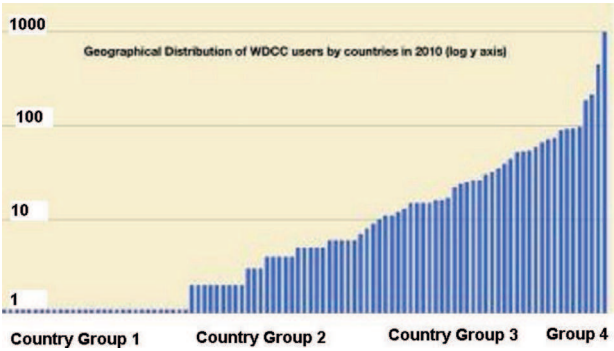

Figure F.4 Geographical distribution of WDCC users by country groups[23] in 2010 (logarithmic y-axis)

Tansania, Solomon Islands, Georgia, Jordan, Uganda, Cyprus, Venezuela, Madagascar, British Indian Ocean Territory, Latvia, Mali, Lesotho, Saudi Arabia, Serbia and Montenegro, Macao, Lebanon, Myanmar, Oman, Cuba, Estonia, Slovenia, Iceland, Armenia, Croatia, Luxembourg, South Georgia, Trinidad and Tobago, Paraguay, North Korea, Congo.
Country Group 2 (2–10 users):
Cameroon, Sudan, Senegal, Bulgaria, Zambia, Nicaragua, Ukraine, Mongolia, Costa Rica, Bangladesh, Bolivia, Jamaica, Nepal, Niger, Lithuania, Colombia, Hungary, Romania, Philippines, Hong Kong, Nigeria, Ethiopia, Egypt, Pakistan, Kenya, Singapore, Israel, Viet Nam, Czech Republic, Chile.
Country Group 3 (11–100 users):
Malaysia, Peru, Argentina, New Zealand, Mexico, Poland, South Africa, Portugal, Finland, Ireland, Turkey, Denmark, Norway, Greece, Russian Federation, Indonesia,

2.2 The Climate and Environmental Retrieval and Archive (CERA) database

The georeferenced data held at WDCC, i.e. climate model results from global and regional climate model experiments performed at DKRZ, are available from the WDCC via the operational CERA data and information system. Input is accepted in electronic form. This includes present-day climate, paleoclimate simulations and IPCC scenario[24] runs for the future. The WDCC archives and disseminates more than 341 TB climate model data and related observations. In Figure F.5, the increase of the volume of data content since the first version of CERA is shown.

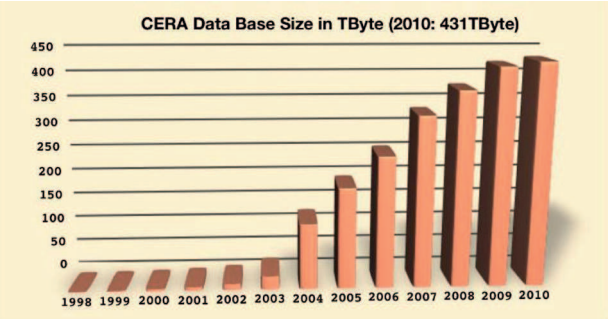

Figure F.5 Development of the CERA database size (terabytes) since 1998[25]

The graphical user interface of the CERA Portal guides the user to search functions for the data sets in the database, and to utilities and tools for data download, processing and visualisation[26] (Figure F.6). More information on the data models used in CERA, the modules of the catalogue, details about the technological implementation and metadata usage and access to the CERA database can be found at http://www.dkrz.de/daten-en/cera.

Belgium, Thailand, Taiwan, Austria, Sweden, Australia, Netherlands, Iran, Brazil, Switzerland, South Korea, Spain, Italy, Japan, Canada, France, India.
Country Group 4 (more than 100 to over 1000 users):
Great Britain, China, USA, Germany.

[24] http://www.ipcc.ch/pdf/special-reports/spm/sres-en.pdf.
[25] http://www.dkrz.de/daten-en/wdcc/statistics/wdcc-statistics-2010.
[26] http://cera-www.dkrz.de/WDCC/ui/Index.jsp.

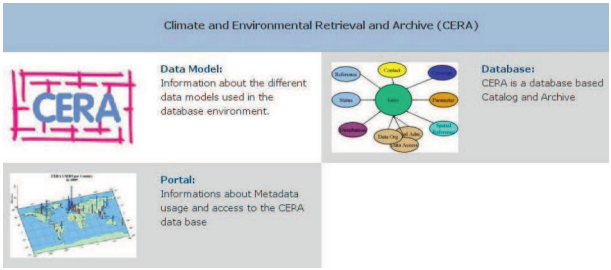

Figure F.6 Components of the CERA archive: data model, database and data portal

2.3 The Coupled Model Intercomparison Projects (CMIP)

M&D was and DM/DKRZ is now a vital component of the Coupled Model Comparison Projects CMIP3[27] and CMIP5,[28] respectively. In these projects, particularly large data volumes were/are generated, which require storage and archiving and need to be distributed among and maintained for (re)-use worldwide. WDCC holds data from model simulations for the Fourth Assessment Report (AR4, CMIP3),[29] and in CMIP5, WDCC/DKRZ is one of the four gateways to which a subset of the data that project participants are delivering to the Program for Climate Model Diagnosis and Intercomparison at the Lawrence Livermore National Laboratory in California (PCMDI, USA) is replicated. The data management for CMIP5 will be shared between the British Atmospheric Data Centre (BADC, UK), the PCMDI and WDCC/DKRZ.

2.3.1 The Coupled Model Intercomparison Project, phase 3 (CMIP3)

During the years 2005 and 2006, climate model output from simulations of the past, present and future climate was collected by the PCMDI. The joint Working Group on Coupled Modelling (WGCM) of the World Climate Research Programme (WCRP) and of Climate Variability and Predictability (CLIVAR), which are both projects of the World Meteorological Organization (WMO), organised this activity to enable researchers "outside the major modelling centres to perform research of relevance to climate scientists preparing the AR4 of IPCC".[30] This so-called "WCRP CMIP3 multimodel data set" was to serve IPCC's Working Group 1, which focuses on the physi-

[27] The Coupled Model Intercomparison Project, phase 3.
[28] The Coupled Model Intercomparison Project, phase 5.
[29] http://www.dkrz.de/daten-en/wdcc/projects_cooperations/ipcc-data-1.
[30] http://cmip-pcmdi.llnl.gov/cmip3_overview.html?submenuheader=1.

cal climate system (atmosphere, land surface, ocean and sea ice). The size of the data set amounts to 36 TB in 83,000 files. By the end of 30 January 2009, 2570 users had downloaded 536 TB in 1,781,000 files[31] More comprehensive sets of output for a specific model may be available from the modelling centre that produced it.

With the consent of participating climate modelling groups, the WGCM has declared the CMIP3 multimodel data set open and free for non-commercial purposes. After registering and agreeing to the "terms of use",[32] anyone can obtain model output[33]

2.3.2 The Coupled Model Intercomparison Project, phase 5 (CMIP5)

After publication of the AR4 in 2007, the WGCM had agreed on a new set of coordinated climate model experiments in phase 5 of the CMIP. In a Memorandum of Understanding (MoU) the PCMDI, BADC and WDCC outlined how they would deliver a model intercomparison archive for CMIP5 and the IPCC Data Distribution Centre (DDC)[34] (this MoU was signed in December 2008). Subsequently a description of the design of the experiments to be done in CMIP5 was given by Taylor, Stouffer and Meehl, in 2009[35] (updated in Taylor et al., 2011). The results of CMIP5 are expected to be useful not only to the IPCC's Working Group 1, but also to those considering possible consequences of climate change, e.g. the IPCC Working Groups 2 "Impacts, Adaptation, and Vulnerability"[36,37] (Parry, Canziani, Palutikof, van der Linden and Hanson, 2007) and 3 "Mitigation of Climate Change"[38,39] (Metz, Davidson, Bosch, Dave and Meyer, 2007).

According to the CMIP5 web page[40] (which includes a schedule) the objectives of the planned standard set of model simulations of CMIP5 are:

- to evaluate how realistic the models are in simulating the recent past,
- to provide projections of future climate change on two time scales (near term to about 2035 and long term to 2100 and beyond),

[31] http://www-pcmdi.llnl.gov/ipcc/usage_statistics.php.
[32] http://www-pcmdi.llnl.gov/ipcc/info_for_analysts.php.
[33] https://esg.llnl.gov:8443/index.jsp.
[34] http://home.badc.rl.ac.uk/lawrence/static/2008/12/03/cmip5_archive_mou_final.pdf.
[35] http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf.
[36] http://www.ipcc-wg2.gov/AR4/website/fi.pdf.
[37] http://www.ipcc-wg2.gov.
[38] http://www.ipcc.ch/publications_and_data/ar4/wg3/en/contents.html.
[39] http://www.ipcc-wg3.de/publications/assessment-reports/ar4.
[40] http://cmip-pcmdi.llnl.gov/cmip5/index.html.

– to reach a better understanding of some of the factors responsible for differences in model projections, including quantifying some key feedbacks such as those involving clouds and the carbon cycle

There are 21 modelling groups in 12 countries participating in CMIP5 (Table F.1).

In CMIP5, data archiving is again – as it was in CMIP3 – managed by the PCMDI, which also collects the multimodel output data and is responsible for authorisation and authentication. PCMDI and two more data centres constitute the Earth System Grid Federation[41] (ESGF), i.e. the BADC which organises the description of the data (metadata) and the replication of the data sets, and the WDCC which is responsible for the development of some quality control tools and data publication. Information on data access and availability is updated on a daily basis.[42] A check of the CMIP5 archive status page[43] shows that nine of the 21 modelling groups have delivered data so far (groups 1, 2, 4, 6, 12, 17, 18, 19 and 20 in Table F.1)

Compared with CMIP3, CMIP5 model documentation will be made more comprehensive and accessible by a standardised vocabulary for describing models and model simulations, i.e. the data reference syntax (DRS). Model metadata[44] will include global attributes (information about the experiment and the model which originate the data), variable attributes (names and units of output variables) and coordinate variables (bounds of the model region, grid axes; Taylor and Doutriaux, 2010).[45] Furthermore, an interactive web-based questionnaire that makes it easier for modelling groups to provide the model and simulation documentation, is being developed by the (mostly European) consortium of project METAFOR[46] (Common Metadata for Climate Modelling Digital Repositories). WDCC is a project partner in METAFOR, charged with the development of common information model (CIM) creation tools (workpackage 6 of METAFOR; Toussaint and Lautenschlager, 2008).[47] The primarily US Earth System Curator[48] team is providing tools for ingesting the information in the questionnaire, designing web-based discovery tools for interrogating the documentation and integrating these tools into

[41] http://esg-pcmdi.llnl.gov/esgf.
[42] http://cmip-pcmdi.llnl.gov/cmip5/availability.html?submenuheader=3.
[43] http://esgf.org/wiki/Cmip5Status/ArchiveView, 16 August 2011.
[44] The metadata shall be consistent with the CF (Climate and Forecast Metadata) Convention, while data files will be accepted in the Network Common Data Form (NetCDF).
[45] http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_output_metadata_requirements.pdf.
[46] http://metaforclimate.eu.
[47] http://colab.mpdl.mpg.de/mediawiki/images/2/20/ESci08_Sem_3_CERA-2_Toussaint.pdf.
[48] http://www.earthsystemcurator.org.

the ESG framework, whereby this meta-information is put in a searchable database linked to the model output.[49]

The Thematic Real-time Environmental Distributed Data Services[50] (THREDDS) data server is the central distribution unit at WDCC to deliver or publish CMIP5 data to those data centres that are members of the Earth System Grid Federation.[51]

Table F.1 List of groups participating in CMIP5 and terms of use of model output data[52]

No.	Primary group/acronym	Full name	Country	ToU*
1	NCC	Norwegian Climate Center	Norway	ns
2	MOHC	Met Office Hadley Centre	UK	a, ns
3	GFDL	Geophysical Fluid Dynamics Laboratory	USA	a
4	IPSL & LMD	Institut Pierre-Simon Laplace	France	a
5	NIES & U Tokyo	National Institute for Environmental Studies	Japan	nc
6	CCCMA	Canadian Centre for Climate Modelling and Analysis	Canada	nc
7	CSIRO & BMRC	Commonwealth Sci. and Industrial Research Org/Bureau of Meteorology Res. Centre	Australia	nc
8	MPI	Max Planck Institute for Meteorology	Germany	a
9	INGV, CEMCC	Istituto Nazionale di Geofisica e Vulcanologia	Italy	nc
10	EC-Earth Consortium	EC-Earth model based on European Centre for Medium-Range Weather Forecasting's seasonal forecasting system	Europe	nc
11	NASA GSFC	NASA Goddard Space Flight Center	USA	ns
12	CSIRO & QCCCE	CSIRO/Queensland Climate Change Centre of Excellence	Australia	ns
13	NCAR	National Center for Atmospheric Research	USA	ns
14	MRI	Meteorological Research Institute	Japan	nc
15	METRI (with MOHC)	National Institute of Meteorological Research	Korea	ns

[49] http://www.earthsystemcurator.org/projects/end2end.shtml.
[50] THREDDS is middleware facilitating the supply of data from provider to users, http://www.unidata.ucar.edu/publications/factsheets/2007sheets/threddsFactSheet-1.doc.
[51] http://esgf.org/wiki/ESGF%20Members.
[52] https://is.enes.org/documents/Taylor_CMIP5_update_pub.pdf.

16	LASG IAP	LASG, Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences (CAS)	China	ns
17	NASA GISS	NASA Goddard Institute for Space Studies	USA	ns
18	BCC	(National) Beijing Climate Center, China Meteorological Administration	China	ns
19	INM	Institute for Numerical Mathematics	Russia	a
20	CERFACS & CNRM	Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	France	ns
21	U. Reading	University of Reading	UK	ns

* ToU: Terms of use of output data: nc: non-commercial; a: unrestricted; ns: not specified

2.4 The Earth System Grid data infrastructure

A lesson from CMIP3 has been that there lies great value in archiving multi-model output in a structured and uniform way. The user community expects to be able to extract data efficiently and in a uniform way across all models. The modelling centres that are contributing the data are responsible for writing the data in that desired way.[53] For the extensive list of model output that is requested in CMIP5 specifications for writing this output are provided. In addition, a software library, the so-called Climate Model Output Rewriter (CMOR2)[54] has been written to facilitate writing model output that conforms to these requirements. CMOR2 fills the gap that ISO metadata standards leave, i.e. information about different modelling grids or the rotation of the earth's pole with respect to the model grid, for example.

Other software developed and distributed by PCMDI includes the ESG data node software for archiving and publishing[55] output and the ESG gateway software to deliver data to end users and provide portal services like registration, security, search and discovery, subsetting and server-side calculations, automated capability to inform users of database withdrawals/additions and use statistics (e.g. number of downloads categorized by model/expt/variable). For data management and analysis for the Earth System Grid, see Williams et al. (2008).

CMIP5 model output will be served by federated centres around the world using different storage architectures, which are as far as possible hidden from the user (Figure F.7).

[53] http://cmip-pcmdi.llnl.gov/cmip5/output_req.html.
[54] http://www2-pcmdi.llnl.gov/cmor/documentation.
[55] Here "publishing" means "provision of data sets to users".
[56] http://cmip-pcmdi.llnl.gov/cmip5/submit.html?submenuheader=2.

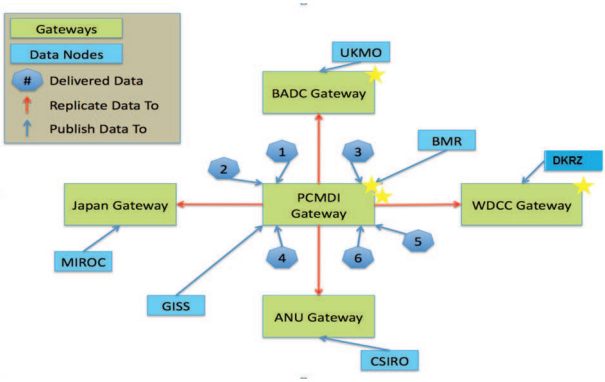

Figure F.7 The Earth System Grid data infrastructure (slightly modified from source)[56]

The CMIP5 archive will be distributed among several centres and will appear to be a single archive. The blue heptagons in Figure F.7 stand for several modelling groups/data nodes publishing data to the PCMDI, while the blue data nodes are publishing data to the respective green gateways. The data centres with the yellow stars will curate the complete CMIP5 data archive, and the second yellow star at PCMDI denotes that here access control is being exercised.

In an ESG Federation wiki on the CMIP5 status page, the operational data nodes and gateways are listed.[57] The size of the CMIP5 archive will be approximately 2 Petabytes published and 1 Petabyte replicated: see, for example, the WDCC gateway realised via the gateway portal software.[58]

The PCMDI also developed the climate data analysis tools (CDAT), whose utility grew from a model behaviour assessment tool to an open-source environment of software which facilitates the analysis of very large data volumes generated by model intercomparison projects and observational programmes that are widely dispersed among many international institutions[59] (Williams, Doutriaux, Drach and McCoy, 2009).

[57] http://esgf.org/wiki/Cmip5Status.
[58] http://ipcc-ar5.dkrz.de/home.htm;jsessionid=47933B7BA7DC201C98803F5D20FF57F2.
[59] http://www.ametsoc.org/meet/annual/annual90shortcourses/1.30pm%20Doutriaux%20II.pdf.

In his IS-ENES[60] Barcelona presentation, Karl Taylor outlined the path of CMIP5 model output through the data infrastructure as follows:[61]

- model output is produced and sent to PCMDI for quality control (QC) checks,
- model output produced and checked for compliance with some output requirements (by CMOR or CMOR-checker),[62]
- METAFOR questionnaire[63] completed generating model and simulation documentation,
- data are made available via the Earth System Grid Federation,
- digital object identifiers (DOIs) are assigned to model output for reference by published literature,[64]
- data are served by ESG gateways via web interfaces.[65]

At DM/DKRZ, the MPI data node (compare with Figure F.7) is being structured as shown in Figure F.8.

2.5 Quality control of CMIP5 model output data[66]

The ESGF partners at PCMDI, BADC and DKRZ which are hosting an ESG gateway and who are producing data replicates of subsets of CMIP5-Data for the AR5[67] carry out distributed quality control. QC occurs at different levels:

- Level 1: CMOR2 and ESG publisher conformance checks are performed at all ESGF partners during ESG publication. The QC1 metadata checks are testing for completeness and execute the technical validation of the questionnaire input.
- Level 2 is performed on requested subsets of CMIP5 data at all ESGF partners. With respect to data, QC2 involves consistency checks, i.e.

[60] "Infrastructure for the European Network for Earth System Modelling" is an FP7-Project funded by the European Commission under the Capacities Programme, Integrating Activities. The project has started on the 1st March 2009 and will finish on the 28th February 2013. IS-ENES promotes the development of a common distributed modelling research infrastructure in Europe in order to facilitate the development and exploitation of climate models and better fulfill the societal needs with regards to climate change issues.

[61] https://is.enes.org/documents/Taylor_CMIP5_update_pub.pdf.

[62] http://cmip-pcmdi.llnl.gov/cmip5/output_req.html?submenuheader=3#cmor.

[63] http://q.cmip5.ceda.ac.uk.

[64] Note that this "publishing" of scientific primary data has a different connotation than that described in section 2.1.5.

[65] ESG data gateways are located at BADC, DKRZ, NASA JPL, NCAR, NCI, NERSC, PCMDI, and ORNL (full names of these institutions can be found in the Acronym section at the end of this chapter).
http://www.earthsystemgrid.org/about/overview.htm.

[66] Description of quality control arrangements courtesy of Martina Stockhause, DM/DKRZ.

[67] Fifth Assessment Report of the IPCC.

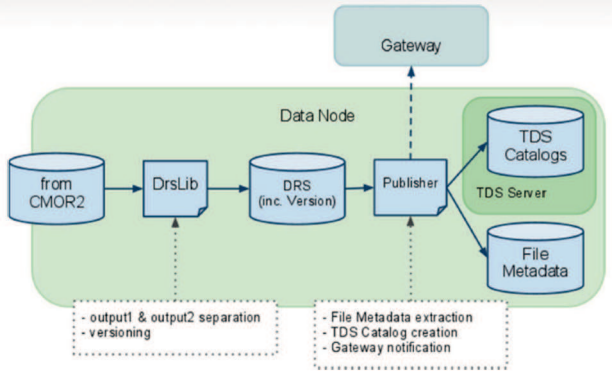

Figure F.8 Processing chain for model output data (courtesy of Estanislao Gonzalez, DM/DKRZ)

check of statistical global values and additional DRS checks (software developed at WDCC),[68] i.e. where exactly in the directory structure the data file is to be found (Taylor et al., 2011).[69] With respect to metadata a subjective QC2 is carried out by the scientist producing the model output data.

– Level 3: technical quality assurance implies double and cross-checking of data and metadata (approval needed from the author); QC Level 3 scientific quality assurance is a check of data and metadata (and approval) by the author.

2.6 Long-term archiving at WDCC

The numbers in circles in Figure F.9 denote steps 1–8 of the following sequence of actions:

1. information gathering and consulting with regards to a request for long-term archiving of data set(s),
2. project specification and cost estimate,
3. defining and including metadata into the WDCC archive,
4. integration and filling of data into the WDCC database,
5. quality assurance,
6. assignment of a DOI (optional),
7. completion of archiving assignment, activation of access permission,
8. maintenance of data archive, possibly adaptation of access right.

[68] QC Level 2 tool developed by Heinz-Dieter Hollweg, DM/DKRZ.
[69] http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax.pdf.

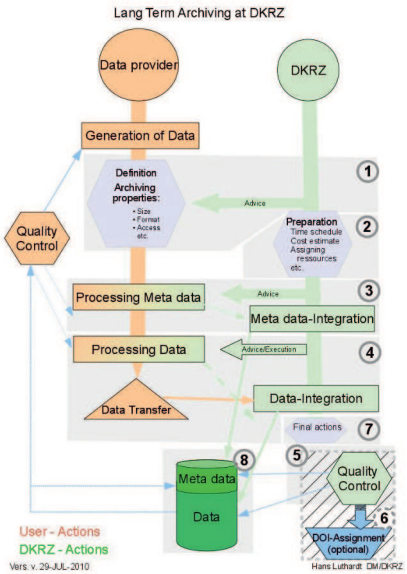

Figure F.9 Work flow of long-term archiving at WDCC (courtesy of Hans Luthardt, DM/DKRZ)

2.7 Publication and citation of scientific primary data

Weather and climate researchers need comprehensive and detailed data in order to arrive at reliable findings. The project Publication and Citation of Scientific Primary Data – Scientific and Technical Data (STD-DOI), for the development of a standard to publish and secure the valuable data for the long term, was supported by the DFG. The aim was to make primary scientific data citeable as publications. In this system, a data set is attributed to its investigators as authors like it is done for a work in the conventional scientific literature. Scientific primary data should therefore not exclusively be understood as part of a scientific publication, but may have its own identity. Since completion of the STD-DOI project, a production service for DOIs has been established. Figure F.10 gives a schematic overview of the network of data publication agents in Germany, i.e. data centres for scientific and technical data in the earth and environmental sciences, the registration agency Technische Informationsbibliothek (TIB Hannover),[70] i.e. the German National

[70] http://www.tib-hannover.de/en/the-tib/doi-registration-agency.

Library of Science and Technology, and the International DOI Foundation (IDF).[71]

Of the archives shown in Figure F.10, the World Data Center for Marine Environmental Sciences (WDC-MARE) was the first one to register data sets as part of a bibliographic citation using the DOI (see also section 3.3). Using the system, scientists worldwide gain access to a web-based platform that enables them to enter and find the data. The system ensures high data quality and long-term use with persistent identifiers (DOI/uniform resource name) for tomorrow's research. With the proposed publication process a method is given by which credit can properly be assigned for the data producers related to a defined citation.

In a new DFG-funded project, KomFor,[72] the TIB cooperates with the four data centres shown in Figure F.10 to establish a competence centre, i.e. the Centre of Expertise for Research Data from the Earth and Environment. In another DFG-funded project, Wikidora,[73] DKRZ and two partners[74] are preparing meteorological research data for persistent identifier registration, which is realised via the web-based workflow application Atarrabi. This new workflow system will be made available as open source software to scientists worldwide[75] (see also Hense, Hense and Lautenschlager, 2010).

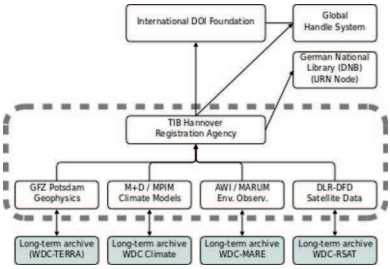

Figure F.10 Network of publication agents in Germany for scientific and technical data in the earth sciences, the registration agency and the International DOI Foundation (source: STD-DOI project homepage)[76]

Figure F.11 shows the process followed when the WDCC, as a publication agent, has been asked to publish environmental data. After mutual agreement for publication has been reached between the scientist and WDCC, the

[71] http://www.doi.org.
[72] http://www.tib-hannover.de/en/the-tib/projects/komfor.
[73] http://umwelt.wikidora.com/wikidora.
[74] Bonn-Rhein-Sieg University of Applied Sciences (Department of Computer Science) and the Meteorological Institute of Bonn University.
[75] http://sourceforge.net/projects/atarrabi.
[76] http://www.std-doi.de.

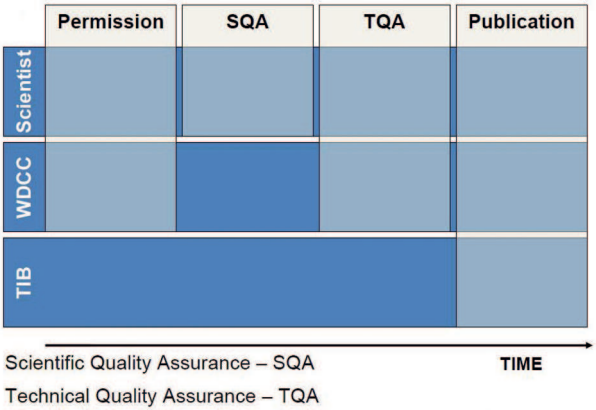

Figure F.11 Responsibilities (upper x-axis) of scientist, WDCC and TIB (y-axis) in the data publishing process (lower time x-axis), indicated as shaded cells in this tabular diagram (courtesy of Heinke Hoeck, DM/DKRZ; Hoeck, 2010)

former takes on the scientific quality assurance (SQA), whereupon WDCC gets involved with technical quality assurance (TQA). This includes checks whether:

- number of data sets is correct and not equal 0,
- size of every data set is not equal 0,
- data sets and corresponding metadata are all accessible via the internet,
- data size is controlled and correct,
- time description (metadata) and existence of data are consistent, complete, start/ stop date consistent, continuous time steps are correct,
- format is correct,
- variable description and data are consistent.

After TQA follows the quality control of the descriptive metadata set by the author and WDCC (Figure F.12), after which the registration agency TIB assigns a DOI (Figure F.13). The DOI consists of two parts. The prefix is assigned by the registration agency. The suffix is provided by the data centre, i.e. the agency which is responsible for the contents. Resolution occurs via the resolver[77] or directly.[78]

TIB was the first DOI registration agency for primary data worldwide (since 2005) and is one of the founding members of DataCite[79] (a not-for-

[77] http://dx.doi.org.
[78] http://dx.doi.org/10.1594/WDCC/CCSRNIES_SRES_B2.
[79] http://www.datacite.org/whatisdatacite.

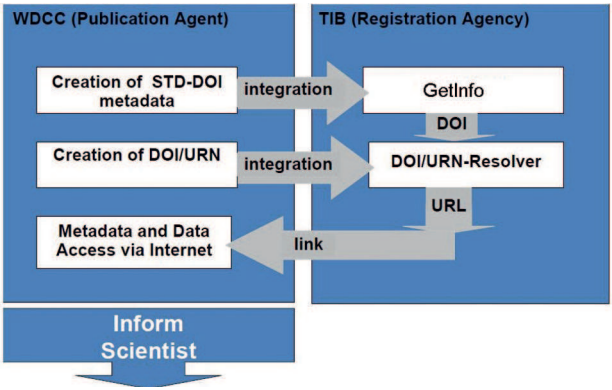

Figure F.12 Workflow for publication of environmental data sets (courtesy of Heinke Hoeck, DM/DKRZ; Hoeck, 2010)

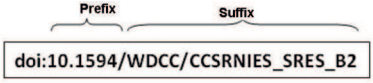

Figure F.13 Composition of the DOI (courtesy of Heinke Hoeck of DM/DKRZ; Hoeck, 2010)

profit organisation formed in London on 1 December 2009). In fact, not DataCite itself but its members are the national institutions who function as registration agencies. WDCC/DKRZ in Germany, for example, has a contract with the TIB as their registration agency. For a university or research institute in the USA, for example, the (contract) partner would be a US registration agency (like the California Digital Library as a member of DataCite).

Meanwhile three further registration agencies in Germany are offering DOI registration services, i.e. the German National Library of Medicine (ZB MED), the Leibniz Institute for Social Science (GESIS) and the German National Library of Economics (ZBW). By doing this, the registration agencies facilitate and promote non-profit online publications:[80]

[80] ZB MED lists these types of digital content for which it can assign DOIs to: publications such as journal articles, research reports, websites with scientific/academic contents, congress publications, posters, and Research data such as image data, videos, audio data, statistical data, sequence data, interview data.

- TIB for all German data centres in the fields of science and engineering, architecture, information technology, mathematics,
- ZB MED for the fields of medicine, health, nutrition, the environment and agriculture,[81]
- GESIS for social science data in Germany,[82]
- ZBW for economic literature and data.[83]

3 Current status of the research infrastructure of observational climate science programmes in Germany

In section 2, earth system and climate modelling projects and their data management arrangements have been described. The complex mathematical models producing the primary data need to be validated, however, by and their results compared with observations done either *in situ* within the compartments of the earth system or remotely from space, for example. Therefore observational earth science is indispensable in climate research. Since a comprehensive treatment of the research infrastructure for climate scientists developing sensors and deploying these in the field, for example, is beyond the scope of this study, merely the main data archives for climate-relevant observation programmes and projects are briefly described in the following six sections. Since the WDCC also holds some observational data sets, it is included (section 3.6).

3.1 German Weather Service (DWD, Deutscher Wetterdienst), Offenbach

The DWD, a public institution under the German Federal Ministry of Transport, Building and Urban Development,[84] is responsible for meeting meteorological requirements arising from all areas of economy and society in Germany, The DWD, founded in 1952 as National Meteorological Service of the Federal Republic of Germany, provides services in the form of weather and climate information. These include meteorological safeguarding of aviation and marine shipping as well as issuing warnings of meteorological events that could endanger public safety and order.

[81] http://www.zbmed.de/en/about-us/who-we-are/doi-service.html.
[82] http://www.gesis.org/dara/en/home/?lang=en.
[83] http://www.zbw.eu/e_services/e_publication_services.htm.
[84] http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=dwdwww_wir_ueberuns&_nfls=false.

DWD operates Germany's densest meteorological and climatological observing network, in which data have been collected for many decades for further processing and archiving. Approximately 100 billion climate data entries were gathered, some time series dating back to the 18th century. Data stem from surface weather stations, upper-air stations and ships, as recorded every day at the synoptic hours, and are disseminated or released in encoded form and archived and reusable in synoptic ordering. They are archived as collected during a single day in the various observing networks, e.g. the main synoptic-climatological network, the secondary climate and precipitation network etc., for climatological purposes, and are stored in various formats, stages of validation and sorting orders.[85] In the future, phenological data[86] will be used more and more for trend analyses in climate diagnosis, as the dates of the beginning of many phenological phases can be shown to correspond to trends in temperature. Besides being a curator of these observational data, the DWD is hosting several transnational and global data centres (Table F.2).

Table F.2 Data centres at the Deutscher Wetterdienst (DWD) (table modified from source)[87]

Acronym	Purpose/task
NKDZ	The National Climate Data Centre makes climatological data collected by the DWD available for users. Development of methods and applications for quality and data management
CM-SAF	Satellite Application Facility on Climate Monitoring
GCC	Global Collecting Center: international centre to receive and to distribute the non-real-time data under the Marine Climatological Summaries Scheme
GZS	Global Center for Weather Reports from Ships: national centre to archive the worldwide data of maritime meteorological platforms
ACD	Archive of the worldwide CLIMAT data: national archive of monthly and annual climate data (monthly and annual means or totals), which are provided by about 2500 stations worldwide on a monthly base

[85] http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_windowLabel=dwdwww_main_book&switchLang=en&_pageLabel=dwdwww_book.

[86] Phenology deals with the periodically recurring growth and development phenomena of plants during the course of a year. The starting time of characteristic vegetation stages (phases) is observed and recorded. These beginnings are closely connected to the weather and climate and are thus suited for the most varied areas of application and for manifold scientific studies.

[87] http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=_dwdwww_klima_umwelt_datenzentren&T21400353661157011331648gsbDocumentPath=BEA__Navigation%2FKlima__Umwelt%2FKlimadatenzentren.html%3F__nnn%3Dtrue&lastPageLabel=_dwdwww_klima_umwelt_klimadaten_deutschland.

GPCC	The Global Precipitation Climatology Centre analyses the monthly precipitation on earth's land surface based on rain gauge station data. It supports global and regional climate monitoring and research and is a German contribution to the World Climate Research Programme (WCRP) and to the Global Climate Observation System (GCOS)
GSNMC	The Global Climate Observing System Surface Network Monitoring Center monitors the availability and quality of CLIMAT reports from stations of the GCOS surface network exchanged via the Global Telecommunication System of WMO

To stay current with technological developments and ensure continuity in its services as National Meteorological Service for the protection of life, the DWD needs to replace its technical infrastructure at least every 10 years. By participating in projects such as VGISC, SIMDAT and C3-Grid,[88] the DWD contributes to the development of new tools and infrastructure for improved services.

In VGISC, the DWD together with Meteo France and the UK Met Office are creating global information system centres (GISCs), data collection and production centres (DCPCs) and national centres virtually tied together according to the WIS standard of the World Meteorological Organisation (WMO-Information System). In the EU-funded project SIMDAT (Data Grids for Process and Product Development using Numerical Simulation and Knowledge Discovery) and with the additional partners of the European Centre for Medium-Range Weather Forecasts (ECMWF) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), these three meteorological services are developing generic Grid technology for the solution of complex application problems.[89]

In the project Collaborative Climate Community Data and Processing Grid (C3-Grid), an effective grid-based environment for earth system research in Germany was created to enable distributed data processing and inter-institutional exchange of large-volume model and observational data.

[88] http://www.c3grid.de/index.php?id=44&L=1.

[89] http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=dwdwww_zusammenarbeit&T17401110631149743806488gsbDocumentPath=Navigation%2FOeffentlichkeit%2FZusammenarbeit%2FTechnikprojekte%2FHome_node.html%3F__nnn%3Dtrue.

3.2 The German Remote Sensing Data Center (DFD, Deutsches Fernerkundungsdatenzentrum), Oberpfaffenhofen[90]

DFD is part of Germany's national research centre for aeronautics and space (DLR, a chartered non-profit organisation). DFD and DLR's Remote Sensing Technology Institute (IMF) together comprise the Earth Observation Center (EOC),[91] whose institutional funding is governed by the research programme of the Helmholtz Association, Germany's largest scientific organisation.[92] From the Earth Observation Center website:[93] "IMF and DFD are the leading national earth observation research and development institutions with public funding. DFD's expertise is in operational tasks (Center for Satellite-based Crisis Information ZKI, National Remote Sensing Data Library NRSDL, international data reception facilities) as well as in the application of remote sensing to obtain information about the land surface, the atmosphere and civil crisis situations. IMF focuses on scientific research related to sensor specific algorithms and methodology development, image processing, and data product development."

While IMF develops methodologies for processing radar data and sophisticated image analysis, specifically for marine remote sensing, DFD's focus lies in the development of user-oriented products and services. With DFD's national and international receiving stations, direct access to data from earth observation missions is possible and information products from the raw data are being derived. Dissemination of these products to users and curating all data in the National Remote Sensing Data Library for long-term use are further tasks regularly performed by DFD. Applications focus on the land surface, civil security and the atmosphere. For core competences and seven departments of DFD, see the website.[94]

3.2.1 Atmosphere (AT)

Research and development in the AT department entails basic research, new applications and data products. The department combines satellite measurements with numerical models to develop innovative, demand-driven services for future operational implementation, as well as technologies, and data products relating to the atmosphere. Research and development extends also to aerosols, radiation, trace gases and atmospheric dynamics. By these activi-

[90] http://www.dlr.de/caf/en/desktopdefault.aspx/tabid-5278/8856_read-15911.
[91] http://www.dlr.de/caf.
[92] http://www.helmholtz.de/en.
[93] http://www.dlr.de/caf/en/desktopdefault.aspx/tabid-5277/8858_read-15912.
[94] http://www.dlr.de/caf/en/desktopdefault.aspx/tabid-5278/8856_read-15911.

ties, DLR contributes to the Global Earth Observation System of Systems[95] (GEOSS), offering access to remote sensing, geospatial static and in-situ data, information and services via the GEO Portal[96] (operated by the European Space Agency and the Food and Agriculture Organization of the United Nations).

3.2.2 Land surface (LS)

Natural processes and human activities are constantly influencing the Earth's surface. These changes can be detected and analysed using remote sensing methodologies. The LS department of the DFD defines systems and mission parameters for new optical earth observation missions, supervises missions and customises information derived from optical and synthetic aperture radar (SAR) sensor systems to meet the needs of investigators from geology, soil science, geography, agriculture and forestry. The LS department thus enables the applied geosciences to benefit from the engineering-related achievements of the EOC.

3.2.3 World Data Center for Remote Sensing of the Atmosphere (WDC-RSAT)[97]

DFD has hosted and operated the WDC-RSAT since 2003, which holds and offers free access to atmosphere-related satellite-based data sets (raw as well as value added data), information products and services.[98] Data on atmospheric trace gases, aerosols, dynamics, solar radiation, cloud physical parameters and surface parameters (land and sea), such as the vegetation index for the northern and southern hemisphere and surface temperatures are available. Data may either be directly accessed if they are stored at the WDC-RSAT or found through the WDC-RSAT portal if they are safeguarded by other providers (compare with section 6.2).

The WDC-RSAT is a member of the WDC cluster Earth System Research and is a publication agent for digital data (see section 6.2 and Figure F.10).

[95] http://www.earthobservations.org/geoss.shtml.
[96] http://www.geoportal.org/web/guest/geo_home.
[97] http://wdc.dlr.de.
[98] http://wdc.dlr.de/about/index.php.

3.3 World Data Center for Marine Environmental Sciences (WDC-MARE, Biogeochemistry, Circulation, and Life of Present and Past Oceans), Bremen[99]

WDC-MARE (founded in 2001) curates data from marine environmental research and facilitates the international collection and exchange of all forms of marine data. WDC-MARE collects, critically reviews and disseminates data related to global change and earth system research in the fields of environmental oceanography, marine geosciences, and marine biology. Its focus is on georeferenced data, and the PANGAEA information system is used as WDC-MARE's long-term archive and publication unit (Publishing Network for Geoscientific and Environmental Data).[100]

WDC-MARE is maintained by AWI, a research centre of the Helmholtz Association, and the Center for Marine Environmental Sciences (MARUM), University of Bremen, with additional support from the DFG Research Center Ocean Margins. On behalf of Germany's participation in the Integrated Ocean Drilling Program[101] (IODP), Bremen University operates the international core repository[102] (Bremen Core Repository, BCR), i.e. 1100 m^2 refrigerated storage area, Since 1994 about 142 km of deep-sea cores from 83 ocean drilling cruise legs in around 210,000 boxes have been collected. The core repository is visited by approximately 200 scientists per year for sampling and around 50,000 samples are taken by guests and by the repository staff. The BCR also houses 142 km of core taken in the North and South Atlantic and Arctic Oceans, and the Mediterranean and the Black Seas, while the other two core repositories in the world maintain cores from the Pacific Ocean plate, the Southern Ocean south of 60ŕS latitude (except Kerguelan Plateau), the Gulf of Mexico, the Caribbean Sea (at the Gulf Coast Repository in Texas, USA, more than 116 km of core), and from the Indian Ocean and marginal seas, the western and northern marginal seas of the Pacific region, defined by the plate boundaries that extend from the Aleutian trench to the Macquarie Ridge (Kochi Core Repository, Japan, 91 km of core; source: BCR brochure).[103]

WDC-MARE was the first publication agent in Germany centre using the DOI to automatically register scientific and technical data sets as part of a full bibliographic citation (see Figure F.10).

[99] http://www.wdc-mare.org.
[100] http://www.pangaea.de/about.
[101] http://www.oceandrilling.org.
[102] http://www.marum.de/en/IODP_Core_Repository.html.
[103] http://www.ecord.org/pub/BCR.pdf.

3.4 National Oceanographic Data Centre for Germany (NODC), Hamburg[104]

The NODC is the focal point of the national and international exchange of oceanographic data. It acquires the marine data sampled by German institutes and agencies, archives it and promotes data exchange on a national and international level. Both NODC and WDC-MARE participate in the IOC/UNESCO International Oceanographic Data and Information Exchange (IODE). NODC also curates the Baltic and North Sea/North-East Atlantic monitoring data according to the resolutions of the Oslo/Paris[105] and Helsinki Conventions,[106] respectively. The Marine Environmental Monitoring Network in the North Sea and Baltic Sea[107] (MARNET) presently comprises ten automated measuring stations. The NODC is hosted by the Federal Maritime and Hydrographic Agency (BSH) in Hamburg. Data are curated in the Marine Environmental Database[108] (Meeresumweltdatenbank MUDAB) which was developed jointly by BSH and the Federal Environmental Agency (Umweltbundesamt UBA), Dessau.

The oceanographic data, which generally are relevant for climate research, are based on in-situ hydrographic measurements from regular surveys as well as from several large oceanographic research programmes. The database covers about 5500 cruises, with data from 250,000 stations (more than 13 million records). The data are quantitative information about the environmental status of the North and Baltic Seas, i.e. values of physical variables like temperature and salinity, chemical variables like nutrients and organic, inorganic and radiochemical components and biological data (distribution of benthos species, for example).

Data originators, external experts and members of the public may access the MUDAB via a Web client.[109] The data are categorised into:

1. water and suspended matter: samples at individual or repeatedly visited stations well as from light vessels and buoys,
2. sediment and pore water: samples at individually or repeatedly visited stations,
3. biota: organisms living in the water body,
4. benthos: organisms living on the ocean bottom

[104] http://www.bsh.de/en/Marine_data/Observations/DOD_Data_Centre/index.jsp.
[105] http://www.ospar.org/content/content.asp?menu=00170301000000_000000_000000.
[106] http://www.itameriportaali.fi/en/tietoa/helcom_seuranta/en_GB/helcom_seuranta.
[107] http://www.bsh.de/en/Marine_data/Observations/MARNET_monitoring_network/index.jsp.
[108] http://www.informus.de:8080/mudab/welcome.faces.
[109] http://www.informus.de:8080/mudab/welcome.faces.

The inventory[110] for these categories of MUDAB (based on a query in May 2007) is given as:

1. vertical profiles: about 22 million data sets for 345 parameters,
2. time-series of weather data, water levels, temperature, salinity and sometimes nutrients and oxygen: about 13 million data sets for 20 parameters, and a historical climate time series for temperature and salinity from four stations in the Baltic and North Seas,[111]
3. about 619,000 data sets for 293 parameters,
4. about 33,000 data sets for 43 parameters,
5. about 1000 data sets for five parameters.

3.5 National Bathymetric Data Centre, Rostock

As part of the BSH, the Bathymetric Data Centre archives bathymetric data from German marine research missions at the BSH branch in Rostock, Germany. Bathymetric data sets from all oceans except the Southern Ocean are available. Sea depths are important marginal information in earth system and climate research. By delivering bathymetric data sets to the International Hydrographic Office[112] (specifically to the IHO Data Center for Digital Bathymetry, IHO-DCDB), BSH supports indirectly the development and updating of global and regional bathymetric charts. International marine research institutions have central access to worldwide bathymetric data through the IHO-DCDB.

A non-exhaustive list of marine geophysical data and information, data compilations and data holders was published by the Commission on the Limits of the Continental Shelf (CLCS) of the United Nations Division for Ocean Affairs and Law of the Sea[113]

3.6 World Data Center for Climate (WDCC)[114]

While the vast majority of the data in the CERA database at WDCC results from global or regional climate modelling experiments, some data sets from observational climate research programmes and projects have also been

[110] http://www.informus.de:8080/mudab/documents/070530_mudab_webclient_faltblatt.pdf.
[111] http://www.bsh.de/de/Meeresdaten/Beobachtungen/MARNET-Messnetz/Klima_MARNET/Klima.jsp.
[112] http://88.208.211.37/srv1.
[113] http://www.un.org/Depts/los/clcs_new/sources/data_portals_holders_websites.pdf.
[114] http://www.bsh.de/en/Marine_data/Hydrographic_surveys_and_wreck_search/Bathymetry/index.jsp.

archived and are being re-used. The high values for data set size and number of downloads per project are associated with data from numerical model experiments (Figure F.14).

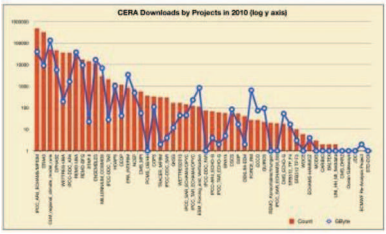

Figure F.14 Number of (red bars) and size of (blue line) data sets downloaded from the CERA database in 2010[115] (note the logarithmic scale of the y-axis). Along the x-axis the projects are sorted such that the project data sets which were downloaded most often come first (left side), followed by downloads of project data sets that were less frequent (towards the right).

Data sets from some important observational measurement campaigns aiming to better understand climate aspects of the earth system appear on the right of the x-axis of Figure F.14. These and other observational projects are listed in Table F.3. Most are from field experiments and monitoring efforts of the last three decades, but there are also some historical compilations, such as the "Global Land Cover Reconstruction AD 800 to 1992" covering over 1000 years and observational set ups that are still ongoing (e.g. at the Wettermast Hamburg).[116] The land-based observation areas are located in Germany and Austria, the oceanic ones in the North Atlantic and Arctic Ocean and adjacent seas. Satellite data sets often have a global coverage, as do those from the WOCE (World Ocean Circulation Experiment) Hydrographic Programme. The experiments listed in the first column of Table F.3 as well data sets from numerical experiments can be found, for example, by "browsing by experiment" after accessing the database through the CERA portal[117] (cf. lower left in Figure F.6).

Table F.3 Data from observational programmes at the WDCC

Project name, purpose	Geographic region	Variables measured, processes observed	Instruments, carriers, platforms

[115] http://www.dkrz.de/daten-en/wdcc/statistics/wdcc-statistics-2010.
[116] http://wettermast-hamburg.zmaw.de.
[117] http://cera-www.dkrz.de/WDCC/ui/Index.jsp.

ACSYS, air mass modification in on-ice air flows (1998), arctic atmospheric boundary layer and sea ice interaction study (2003)	Arctic	Turbulent fluxes, standard meteorological parameters, radiosonde measurements[118]	Research aircraft, buoys, ship
AQUA_AMSRE:[119] to better understand the Earth's water cycle and determine if the water cycle is accelerating as a result of climate change	Global ocean	Geophysical parameters, including SST, wind speed, atmospheric water vapour, cloud water, and rain rate, local copy of NASA data	Advanced Microwave Scanning Radiometer (AMSRE), Satellite Aqua
AVHRR Pathfinder SST v5,[120] a more accurate, consistent land mask, higher spatial resolution, inclusion of sea ice information, better flagging of aerosol-contaminated data retrieval	global	4 km AVHRR Pathfinder version 5, SST monthly means data set (daytime measurements), local copy of NASA data	Advanced Very High Resolution Radiometer (AVHRR), NOAA satellite
ALKOR, BASIS, eight field experiments (1998, 2000, 2001) to collect a comprehensive data set to validate the coupled model system BALTIMOS[121] for the Baltic	Central Baltic Sea	Atmospheric boundary layer structure and processes and air-sea-ice interaction over areas with inhomogeneous sea ice cover; atmospheric boundary layer structure over open water under different synoptic conditions such as cold-air advection, warm-air advection or frontal passages, radiosonde	Various, RV *Alkor*

[118] http://www.erh.noaa.gov/gyx/weather_balloons.htm.
[119] http://ssmi.com/amsr/amsr_data_description.html.
[120] http://podaac.jpl.nasa.gov/SeaSurfaceTemperature/AVHRR-Pathfinder.
[121] http://www.borenv.net/BER/pdfs/ber7/ber7-371.pdf.

ARKTIS 1988, atmospheric boundary layer in the marginal ice zone, investigation of boundary layer modification and certain cloud structures in cases of off-ice and on-ice air flows	Fram Strait	Mean structures, variances and covariances at different distances from the ice edge	ships, aircraft, Icebreaker *Polarstern*, RV *Valdivia*, several aircraft operating from the airport at Longyearbyen on Spitsbergen
ARKTIS 1991, cellular convection, investigation of cold air outbreaks from the surrounding Arctic ice sheets	Norwegian Sea	Observation of newly formed boundary layer: its depth, mean temperature, moisture with increasing distance from the ice edge	RV *Valdivia*, research aircraft FALCON-20 and DO-128
ARKTIS 1993, air mass modification in off-ice air flows, investigation of cold air outbreaks from the Arctic sea ice onto the open water	West Spitsbergen current	Aerological data collected at three land stations (Bear Island, Danmarkshavn, NyAlesund), radiosonde	RV *Polarstern*, RV *Valdivia*, RV *Prof. Multanovsky*, aircraft Falcon and DO-128
ASTEX 1992, Atlantic Stratocumulus Transition Experiment, observations with modelling activities to investigate the consequences to the atmosphere and ocean of marine stratocumulus clouds and their life-cycle variations, including the important broken cloud regimes	Azores and Madeira Islands, north-eastern Atlantic	156 radiosonde ascents	Satellite, airborne, island, buoy, RV *Valdivia*

BALTEX[122] Baltic Sea Experiment, meteorological, hydrological and oceanographic research to explore and model the various mechanisms determining the space and time variability of energy and water budgets of the BALTEX region and this region's interactions with surrounding regions	Baltic Sea, Danish Straits	Lateral exchange with the atmosphere outside the BALTEX region, wind stress at the sea surface, evaporation and precipitation over land and sea, heat and energy flux at the air–sea and air–land interfaces, including radiation, river runoff, in- and outflow through the Danish Straits (each country providing its own set of meteorological parameters)[123]	Radiances, OVS atmospheric temperature-humidity and ice/snow correlative data sets,[124] satellites
BASIS 1998, 2001, Baltic Air Sea Ice Study	Gulf of Bothnia, Baltic Sea	Standard meteorological measurements, surface measurements at four land stations (Marjaniemi, Oulu, Kuivaniemi, Haparanda), radiosonde ascents	RV *Aranda*, research aircraft DO-128
COPS[125] (2007), Convective and Orographically induced Precipitation Study, to advance the quality of forecasts of orographically induced convective precipitation by 4D observations and modelling of its life cycle	Black Forest, Germany	Atmospheric Radiation Measurement (ARM), tropospheric profiles of water vapour and wind, and many more,[126] radiosonde ascents	Soil moisture network, research aircraft, mobile meteorological masts

[122] http://www.baltex-research.eu/background/bp1.html.
[123] http://www.baltex-research.eu/data.
[124] http://eosweb.larc.nasa.gov/GUIDE/dataset_documents/base_isccp_d1_d2_dataset.html#overview.
[125] http://www.cops2007.de.
[126] http://www.meteo.uni-bonn.de/messdaten/passive-microwave-radiometer-admirari/cops-measurements-2.

AVISO,[127] Archiving, Validation and Interpretation of Satellite Oceanographic data DUACS, Data Unification and Altimeter Combination System, processing data from altimeter missions to provide a consistent and homogeneous catalogue of products for varied applications, both for near-real-time applications and offline studies	Global	Altimeter, monthly gridded sea surface heights computed with respect to a 7-year mean (averaged sea surface heights averaging month by month), monthly means created from weekly sea level anomaly maps	Satellites Jason1, Topex/ Poseidon, Envisat, GFO, ERS1 and 2, Geosat
DAMOCLES 2007–2008, Developing Arctic Modeling and Observing Capabilities for Long-term Environmental Studies, Hamburg Arctic Ocean Buoy Drift Experiment	Central Arctic Ocean	Ice drift, sea ice	Array of 16 drifting autonomous buoys
FGGE 1979, First GARP Global Experiment	Central equato-rial Atlantic Ocean	Near-surface oceanographic and surface meteorological data,[128] 291 radiosondes	RV *Meteor* and 40 other ships
CARIBIC (1997–2002) and CARIBIC–LH (2007), Civil Aircraft for the Regular Investigation of the atmosphere Based on an Instrumented Container, to study and monitor important chemical and physical processes in the Earth's atmosphere	Along inter-conti-nental flight tracks	Suite of variables collected during each flight and analysed in-flight on board or later in the laboratory[129]	Commercial aircraft

[127] http://www.aviso.oceanobs.com/en/data/products/sea-surface-height-products/global/index.html.
[128] http://www.sciencedirect.com/science/article/pii/007966118690008X.
[129] http://www.caribic-atmospheric.com.

FRAMZY (1999, 2002, 2007, 2008, 2009), five field experiments to investigate the properties of Fram Strait cyclones, their cyclogenetic conditions on the large- and meso-scale, and their local effects on sea ice drift and sea ice distribution and, thus, on the freshwater flow through the Fram Strait. The data were used for validation of cyclone simulations with coupled mesoscale models of the atmosphere-ice-ocean system	Fram Strait, Greenland Sea	Meteorological data, ice, ice drift	14 autonomous ice buoys, RV *Aranda*, research aircraft Falcon, satellites (NOAA-AVHRR, RADARSAT, DMSP-SSM/I)
FRONTEX 1989, atmospheric fronts, to investigate cold fronts moving in from the North Sea and reaching the coastal area with high temporal and spatial resolution	Coastal area of northern Germany, Heligoland, Schleswig, Hanover, Emden, Berlin	Ground-based remote sensing and in-situ measurements, physical properties of sea and land surface (roughness, humidity, temperature, heat conduction and heat capacity), radiosonde ascents	Research vessel, research aircraft POLAR-2, POLAR-4, DO-128
GEBCO,[130] General Bathymetric Chart of the Oceans	Global	Depth soundings	Ships, various others
Glacier monitoring data of Austria[131]	Austria[132]	Hydrological parameters, glacier mass balance	

[130] http://www.bodc.ac.uk.
[131] http://imgi.uibk.ac.at.
[132] http://imgi.uibk.ac.at/iceclim/glacierinventory.

GOP (2007), General Observation Period of Priority Program on Quantitative Precipitation Forecasting	German Weather Service net-works	Rain gauges, weather radar, Light Detection And Ranging (Raman Lidar), ground-based Global Positioning System (GPS), lightning, satellite data,[133] radiation, microwave radiometer, ceilometer, cloud radar, wind profiler, radiosonde ascents	Ground stations, satellites (Meteosat Second Generation MSG, MODIS and MERIS)
HADEX,[134] global climate extremes indices	Global	Land-based climate extremes data set, 27 indices of temperature and precipitation on a 2.5×3.75řgrid from 1951 to 2003. Indices represent seasonal and/or annual values derived from daily station data	

[133] http://www.geo.fu-berlin.de/met/ag/sat/satdaten/index.html.
[134] http://www.hadobs.org.

KONTROL (1984, 1985), experiment on convection and turbulence with the objectives (1) to observe the formation and time variation of regularly organised convection in the lower troposphere as a function of the mean atmospheric flow and the lower boundary condition and to quantify the dependence of the vertical transports of momentum, heat and water mass on various scales of motion, and (2) to determine the mean and turbulent quantities within the marine atmospheric boundary layer, including the large-scale horizontal and vertical advection of momentum, heat and water vapour, cloud microphysics and the radiation field. Goal: to test existing convection models and to provide an observational background for the extension of theoretical concepts	German Bight, south-eastern North Sea	Continuous aerological and surface observations at fixed stations (island Heligoland, Borkumriff, moored platform, ships), detailed observations during special periods done by aircraft, supporting observations, such as satellite images, cloud photography, surface and upper air large-scale fields from routine data	RV *Valdivia*, RV *Meteor*, RV *Gauss*/*Poseidon*, research platforms Nordsee and Elbe 1, research aircraft (Falcon 20, DO-28 Sky-servant, Hercules C-130)

LOFZY 2005, first field experiment on cyclones over the Norwegian Sea, low-pressure systems (cyclones) and the climate system of the North Atlantic	Lofoten archi-pelago	From ship: meteorological observations (radiosondes, standard parameters), oceanographic CTD measurements. From aircraft: observation of synoptic conditions with high spatial and temporal resolution. Additionally deployment of 23 autonomous marine buoys in advance of the campaign to measure drift, air-temperature and -pressure and water-temperature	RV *Celtic Explorer*, research aircraft Falcon
MODIS_ACDNC, adiabatic cloud droplet number concentration daily value	Global	Cloud droplet number concentration is derived from MODerate Resolution Imager Spectroradiometer (MODIS)[135]	NASA satellite Terra (EOS AM)[136]
Reconstruction of global land use and land cover AD 800 to 1992[137]	Global, 30 min-ute resolu-tion	Population data,[138] three human land use types (crop, pasture) and 11 natural vegetation types[139]	
SeaWinds on QuikSCAT[140] Level 3 Daily Gridded Ocean Wind Vectors[141]	Global	Gridded values of scalar wind speed, meridional and zonal components of wind velocity, wind speed squared and time given in fraction of a day	NASA satellite QuikSCAT (Quick Scat-terometer)

[135] http://modis.gsfc.nasa.gov/about.
[136] http://terra.nasa.gov.
[137] http://www.mpimet.mpg.de/fileadmin/publikationen/Reports/WEB_Bze_51.pdf.
[138] Atlas of World Population History (McEvedy and Jones 1978)
[139] http://www.sage.wisc.edu/pubs/abstracts/ramankuttyGBC1999.html.
[140] http://winds.jpl.nasa.gov/missions/quikscat/index.cfm.
[141] http://idn.ceos.org/KeywordSearch/Metadata.do?Portal=idn_daacs&KeywordPath=[Source_Name%3A+Short_Name%3D%27QUIKSCAT%27]&NumericId=27759&MetadataView=Text&MetadataType=0&lbnode=mdlb2.

Wettermast Hamburg	Hamburg-Billwerder	Ground-based continuous measurement of weather data since 1995 at several height levels: 2, 10, 50, 70, 110, 175, 250 m above the ground	Broadcasting tower of 300 m height
WOCE Hydrographic data, Onetime and Repeat Survey,[142] carried out mostly between 1990 and 1998	World ocean along trans-oceanic sections	Full-depth CTD profiles of temperature, salinity, oxygen, from water bottle samples chemical properties were analysed, including nutrients, chemical oxygen demand, chlorofluorocarbons, tritium, helium and other tracers	Many research vessels and volunteer observing ships worldwide

CTD: conductivity-temperature-depth; SSS: sea surface temperature.

[142] http://woce.nodc.noaa.gov/wdiu/diu_summaries/whp/index.htm.

4 Results of a survey concerning climate research practices in six German institutions

In addition to the climate modelling and climate data resources available in German data centres that have been already described in sections 2 and 3, respectively, I used a sample questionnaire to learn directly perceptions of the infrastructure from eight researchers working in representative climate research departments of major research institutions in Germany (see Table 4). The questionnaire was designed by Dennis Spohr of the Centre of Excellence Cognitive Interaction Technology (CITEC, Bielefeld) and supplied to all subject-specific chapter authors (Spohr, 2010).

The aim of the survey was to hear first-hand from climate scientists the details regarding the infrastructure with respect to data and literature which they experience on a day-to-day basis. After explaining first the group's work focus, they described the characteristic data lifecycles, data processing data formats, data management, access to data and the ways that publication and exchange of research data are customary in their institution. The third set of questions dealt with the organisation of literature, to what extent publication of literature and research data jointly was established and whether Open Access was customary. Finally the group's specific outlook was queried for the future developments in data and literature infrastructure in climate science. The interviewed scientists are affiliated with institutions/organisations and departments focusing on either research, service providing and infrastructure development (Table F.4).

Table F.4 German research institutions participating in the survey regarding climate research practices

	Names of institution, department and/or group of researchers interviewed	Personnel and equipment		
		People	PCs	Other
1 r	Max Planck Institute for Meteorology MPI-M, Hamburg (a) Dept. "Land in the Earth System", group "Terrestrial Remote Sensing" (b) Dept. "Atmosphere in the Earth System", group "Observations & Process Studies"	(a) 8 (b) 15	35 for both groups	(a) 1 data server (b) 1 computer server
2 sp (r)	Climate Service Center (CSC), Hamburg, (an Institution at the Helmholtz-Zentrum Geesthacht),[143] Dept. "Climate System"	8	12	–
3 r	Karlsruhe Institute of Technology/Institute for Meteorology and Climate Research, Atmospheric Environmental Research (KIT/IMK-IFU), Garmisch-Partenkirchen a) Regional climate and hydrological modelling b) Collection and analysis of observational data	18	30	1 data server 1 powerful computer (Linux cluster)
4 id	Collaborative Climate Community, Data and Processing Grid (C3-Grid) at Alfred Wegener Institute for Polar and Marine Research (AWI), Bremerhaven	30		Distributed data processing capacity at various WDCs
5 r	Helmholtz-Zentrum Geesthacht (HZG), Geesthacht, Department "System Analysis", Paleoclimatology group	4	4	–
6 sp	Federal Maritime and Hydrographic Agency (BSH), Hamburg, Dept. "Marine Sciences", Division "Data and Interpretation Systems"	28	28	6 servers

id: infrastructure development ; r: research; sp: service providing.

[143] http://www.flyhy.eu/HZG.html.

4.1 Max Planck Institute for Meteorology (MPI-M), Hamburg

4.1.1 General information

Two researchers were interviewed, one (R1) speaking for the group "Terrestrial Remote Sensing" (department "Land in the Earth System" of MPI-Met), the other (R2) for the group "Observations & Process Studies" (department "Atmosphere in the Earth System" of MPI-Met). R2 is responsible for a new working group "observational data", installed to organise the whole suite of observations done by scientists at MPI-M. The institute has a third department, i.e. "Ocean in the Earth System".

In R1's group, the primary research objective is atmospheric observation by airplane, with instruments at the earth surface, e.g. at the Cloud Observatory on Barbados,[144] and remotely sensed data from satellites. Inside MPI-M there are numerous cooperations with the model developers, and outside with colleagues at the Meteorological Institute of the University of Hamburg and other national and international teams.

4.1.2 Data infrastructure

Both R1 and R2 collect a variety of atmospheric observational data and receive the level-1 satellite data. Processing of data includes quality control and the derivation of level-2 satellite data. The data are enriched by the calculation of geophysical parameters from the level-1 satellite data. The data are archived and re-used. The data types include those taken with a camera from an airplane or collected using a stationary webcam (GB range). Collected binary and ASCII data are in the GB to TB range. Some proprietary software is used in order to collect primary data (the instrumentation manufacturer may require this). Data products, i.e. "exploited" level-2 data are distributed, for example the HOAPS Climatology[145] (Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data).

The data are further annotated with metadata, and frequently re-formatted according to the CF-convention (NetCDF Climate and Forecast (CF) Metadata Convention).[146] In some rare cases the scientists have to deviate from standard metadata formats, for example where there is no representation possible for soil moisture. As satellite data formats are quite heterogeneous, the scientists in this group cannot very often rely on conventional formats and software for data representation and processing. Developed software as

[144] http://www.mpimet.mpg.de/en/wissenschaft/atmosphaere-im-erdsystem/initiativen/barbadosstation.html.
[145] http://www.hoaps.zmaw.de.
[146] http://cf-pcmdi.llnl.gov.

well as primary and secondary data are stored and archived in a "repository" (definition: versioning is possible, using, for example, SVN[147]) within the group on a central file server. In addition copies of the software are kept on own storage devices on the hard disk of the office PC. R1 and R2 believe this to be a representative practice in climate science. Having seen the need, the MPI-M recently created the "Observation Steering Group" for observational data which is responsible for data management issues.

Data access is ensured by dissemination via the CERA database maintained at WDCC (see section 2.2), and also via the Integrated Climate Data Center[148] (ICDC) of CLISAP[149] (Integrated Climate System Analysis and Prediction), a "Cluster of Excellence" at the University of Hamburg which is funded by the German Research Foundation (Stockhause and Hoeck;[150] in Curdt and Bareth, 2010).

Research data and software are made available to close colleagues and other research projects, e.g. for use in publications. The general public may receive secondary data only. This restriction exists in order to give the group and its collaborators priority for publishing first results based on the data they collected and processed. Furthermore, there are only limited resources available in the group to guarantee maintenance and user support (there is, for example, no help desk). What is said here for observational data is transferable to modelling software: without additional support not all earth system models are usable as "community models", and this is generally true in climate science. Sporadically, however, software is made available to other institutions. This typically includes source code as software/models need to be compiled locally. The principal investigators in the group delivering the data must have priority with respect to analysis and publication, however. Data exchange happens via file transfer protocol or by shipping of a hard disk. For internal exchange within the MPI-M, the data server is used.

4.1.3 Literature

Online access is available to most subject-specific journals (cf. section 5). Tools that are used include JabRef, an open source bibliography reference manager using the file format BibTeX, a standard LaTeX bibliography format and others that individual researchers choose. Similar to other researchers in the climate community, publications both as print medium (e.g. article or book) and as electronic publication online or offline are preferred and

[147] http://svnbook.red-bean.com.
[148] http://icdc.zmaw.de/icdc.html?&L=1.
[149] http://www.klimacampus.de/clisap0.html?&L=1.
[150] http://icdc.zmaw.de/397+M59fd2f2bea8.html.

established. Publishing via a combination (e.g. book/CD-ROM or proceedings/website) is also preferable but is done rather infrequently.

The following scientific journals are often chosen for publication:

- *Hydrology and Earth System Sciences:* interactive Open Access journal (European Geosciences Union)
- *Remote Sensing of the Environment:* interdisciplinary journal for results on theory, science, applications and technology of remote sensing of Earth resources and environment (Elsevier)
- *Journal of Climate:* online journal (American Meteorological Society)
- *Remote Sensing:* online journal (Yale's Center for Earth Observation)
- *Biogeosciences:* interactive Open Access journal (European Geosciences Union)
- *Journal of Geophysical Research (Atmospheres):* journal (American Geophysical Union)
- *Geophysical Research Letters:* journal (American Geophysical Union).

Some publishers enable the exchange of data and/or literature and it is currently possible to publish both together. It is/would be desirable to be able to also publish a movie together with an online accessible reference. This would happen if, for example, the funding agency required this. However, a data server would be required to guarantee long-term storage.

As for other climate research groups, these interviewees confirm that Open Access practices are supported by having a dedicated database and data server. Access to literature is well established at large research facilities through, for example, the national licences of the German Research Foundation.

4.1.4 Outlook

Satellite data sets already now have large volumes. In the coming 5–20 years, satellite data volumes will increase exponentially. It will not be possible any longer to move such large data sets from A to B via ftp (TB to PB range). The EU funds sentinel satellites,[151] for which 1 TB of data per sensor per day are expected! The data to be returned from this mission is tractable for high spatial resolution of a small area only for a short time of interest. However, climate researchers who are more interested in a global coverage need to go elsewhere, i.e. to hosting data centres with computing facilities like ESA/ESRIN (European Space Agency/European Space Research Institute of the European Space Agency, near Rome), ECMWF and DKRZ. The user participates via cloud computing in the analysis of such data sets. For example, it would take 8 weeks to download 3-hourly global data at a horizontal

[151] http://www.esa.int/esaLP/SEMZHM0DU8E_LPgmes_0.html.

resolution of 6 km for the past 25 years. A workshop took place in Hamburg from 30 March to 1 April 2011 on Climate Knowledge Discovery. The goal is to find new fields of application for new technologies (e.g. pattern recognition software).

4.2 Climate Service Center (CSC, Helmholtz-Zentrum Geesthacht), Hamburg

The CSC is predominantly a service providing institution, i.e. a national agent brokering climate information to aid the dialogue between climate science and politics. The five departments of the CSC all focus their work on four sectors, i.e. agriculture, forestry, energy and health. The CSC prepares the knowledge derived from climate research in a practice-oriented way and conveys it to decision makers. Besides from the information on its own website,[152] a brief profile of the CSC can be found on the website of the Regional Science Service Center in the Southern African subregion (RSSC).[153] The CSC is one of the German institutions involved in this joint initiative of Angola, Botswana, Namibia, South Africa, Zambia and Germany, responding to the challenges of global change.

4.2.1 General information

The two senior scientists interviewed are meteorologists and belong to the CSC's "Climate Science" department.

4.2.2 Data infrastructure

Only binary data in the terabyte range are collected from the CERA database, processed and enriched, i.e. new quantities like, for example, precipitation are computed. Data are archived, re-used and distributed. No proprietary software is used, but software that was developed at the DKRZ.

Regarding data processing, metadata standards exist, but specification of all important information is not yet possible with them, e.g. the type of model grid that a used. Some software is being developed but also proprietary software is used (e.g. Aquacis, ARCGIS). Software and secondary data are mainly stored and archived on own storage devices. No person in group is specifically dealing with data management issues, although this is deemed necessary in order to make secondary data available and (re)usable for customers. Access to data is offered via a ftp server.

[152] http://www.climate-service-center.de/index.html.en.
[153] http://www.sasscal.org/.

Software and primary data are made available only among close colleagues, whereas secondary data are available also for members of the general public. The scientist should have priority for publishing using the data of which he/she was the originator, and the user/customer may be unable to use software necessary for data visualisation, analysis etc. due to lack of technical possibilities. If the group had the appropriate technological equipment, data could be exchanged with other groups within the centre. No software is given to other institutions, because the mission of the Climate Service Center is to deliver only products to customers. At present, general terms and conditions for data delivery are under development at the centre. Data exchange is supported by maintaining a ftp server.

4.2.3 Literature

As far as internal and external publications are concerned, the rules of the HZG (i.e. the umbrella organisation of the Climate Service Center) apply: all publications and lectures need to be registered before submission.

Publications as print medium as well as electronic publication online or offline are preferred and established at the CSC. Scientists use a broad spectrum of journals to publish papers, because climate and climate change impacts occur in many other disciplines. Some publishers enable the exchange of data and/or literature together. Open Access is to some extent established in the group, which is common in the discipline of climate research. Good networks facilitate this practice.

The interviewees envision for the future the creation of a network of climate service centres at the national and the international level and implementation of cooperation contracts.

4.3 Karlsruhe Institute of Technology/Institute for Meteorology and Climate Research, Atmospheric Environmental Research (KIT/IMK-IFU), Garmisch-Partenkirchen

4.3.1 General information

The interviewees are a senior scientist, whose research objective is regional climate and hydrological modelling, and a PhD candidate who concentrates on the collection and analysis of observational data.

4.3.2 Data infrastructure

Data collected include the forcing data of/for a global climate model, e.g. a data set generated by a modern, consistent and invariant data assimilation system such as the ERA set of reanalysis data.[154] The climate model output data then may serve, for example, as input data for regional model experiments. The data volume is in the terabyte range.

Observational data may come from X band radars detecting small particles in the atmosphere. Precipitation intensity can be measured with this device over areas of 50 km extent, data are in binary format and volume is. Another instrument, with which the observational group at this institution collects data, is a disdrometer, i.e. an instrument used to measure the drop size distribution and velocity of falling hydrometeors. Some disdrometers can distinguish between rain, graupel and hail. Thirdly, microwave transmission line integrated precipitation and humidity observations are carried out and deliver data. Data are stored on a universal mobile telecommunications system (UMTS) data server.

Software is developed to derive statistical quantities, such as occurrence frequencies for specific climate indicators. In the numerical modelling group, the software has been extended to some extent, e.g. the coupling software enabling a hydrological model to be linked to an atmospheric climate model.

Annotation of the data sets with metadata happens mostly on a bilateral basis, although metadata standards exist. Sometimes software is developed in-house (e.g. after completion of a numerical experiment) to derive further secondary data. Usage of proprietary software is not required for the formats of data sets, but when analysing data sets, such software products like ARCGIS and MATLAB are used.

Data sets (primary and secondary) as well as software are generally stored in a repository within the group and in a repository shared with other institutes or institute wide. This situation is comparable with another regional modelling group at the Max Planck Institute for Meteorology (the REMO group) in Hamburg. No particular person in the group at KIT/IFM-IFU is dealing with data management issues, but it would be desirable. Somebody is needed who has a combined IT knowledge and also knows what the climate scientists need.

In the project DEKLIM-QUIRCS[155] (Quantification of uncertainties in regional climate and climate change simulations), the partners decided on an exchange formats for model comparisons, the data exchange being achieved on a bilateral basis via shell access to a ftp data server.

[154] http://www.ecmwf.int/research/era/do/get/index.
[155] http://imk-ifu.fzk.de/441.php.

As far as publication and exchange of research data is concerned, the group makes software available only among close colleagues, and primary and secondary data also to other research projects. This restriction (no free delivery to the general public) is done for the well-known reasons that scientific judgment is required to handle data and software reasonably. Incentives for data exchange are good agreement within research projects, for example, functioning on a give-and-take basis. This is considered to reflect the general attitude within the discipline of climate science. If one should decide to also make software available to other institutions, this would, in most cases, also include the source code.

There are no special rules within the group regarding time frames for exchanging data. Raw data are kept as long as possible (several years). The time frame for archiving is guided by users needs. The ftp tool is used for model input data, but for model-produced data a special selection is agreed on and the subset is placed on the ftp server for distribution.

4.3.3 Literature

The group uses Zotero[156] as a free tool, i.e. a plug-in to the Firefox web browser, to collect, organise, cite and share their research resources. The preferred and established modi of publication are the print medium (e.g. article or book) and electronic publication online or offline. At times, but rather infrequently, a combination (e.g. book/CD-ROM or proceedings/website) is chosen for publication of research results. Again, this practice is felt to be representative for climate science.

Scientific journals that members of the group use are:

- *Hydrology and Earth System Sciences*: interactive Open Access journal (European Geosciences Union)
- Journal of Geophysical Research (Atmospheres): journal (American Geophysical Union)
- *ScienceDirect*:[157] SciVerse ScienceDirect scientific database contains more than 10 million journal articles and book chapters. Peer-reviewed full-text articles can be accessed.
- Comptes Rendus Geoscience: journal (Elsevier)
- *IEEE Geoscience and Remote Sensing*: journal (IEEE Geoscience and Remote Sensing Society).

The answer to the question as to whether there are publishers which enable the exchange of data and/or literature is affirmative.

Open Access according to the Berlin Declaration is to some extent established in the group, but is implemented after communication and not

[156] http://www.zotero.org.
[157] http://www.sciencedirect.com.

automatic. This is felt to be similar to the policies of other climate science institutions. The interviewees point out that Open Access practices need the technological backing of good and fast networks (internet, ftp, mailing via the post office).

4.3.4 Outlook

Metadata are found to be extremely important. For literature, abstracts need to be preserved and a good indexing scheme needs to be in place.

4.4 Collaborative Climate Community Data and Processing Grid (C3-Grid), AWI, Bremerhaven[158]

4.4.1 General information

The interviewee is coordinator of the project C3-Grid which has created a unified and transparent access to several large geographically distributed data archives.

C3-Grid's objectives are to provide a service mainly to the German climate science community and to new and emerging scientific communities such as the one primarily concerned with "climate impact", agencies concerned with strategies to adapt to climate change, but also biostatisticians, photon physicists, and others. C3-Grid is supported by the German Federal Ministry of Education and Research (BMBF).

An examples of the data management facilities adopted by C3-Grid and innovative developments in the climate community to alleviate the metadata generation, extraction and management is Fedora Enabled Repository with Cocoon (Federico),[159] which is a state-of-the-art AJAX front end for the Fedora Commons Repository developed in the scope of the Work Package 3 of WissGrid,[160] for the long-term preservation of research archives.

In their presentation "A Collaborative Environment for Climate Data Handling"[161] at the "Geoinformatics 2008 – Data to Knowledge" conference in Potsdam, Germany, Kindermann and Stockhause described some problems encountered in the C3-Grid project. The infrastructure for tracking the whole data cycle from discovery of input data to publication and archiving of the results is designed in three layers: a common data discovery layer, a data

[158] http://www.c3grid.de/index.php?id=32&L=1.
[159] http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-D3.5.3-grid-repository-Federico.pdf.
[160] http://www.wissgrid.de/index_en.html.
[161] http://www.c3grid.de/fileadmin/c3outreach/material/Kindermann_C3-geoinf.pdf.

access layer, and a data manipulation layer. In the extended abstract published on page 31 in the proceedings of this conference[162] (Brady, Sinha and Gundersen, 2008), the authors summarise the challenges:

"In general, a major challenge in the project is to find or develop legal agreements that reflect an elaborate balance between technical progress and manageable effort. The established data and computing-service providers want to re-use their current implementations in order to minimise the maintenance of their software and the labor required to adapt to changes that are necessary when building the infrastructure. Yet, integrating collaborative environments always requires the creation of prototypes and the adoption of not-yet-established technologies. Different technological pathways have to be merged with respect to the specific needs of the existing scientific community and the future needs of intercommunity cyber infrastructures."

4.5 Coastal Research of the Helmholtz-Zentrum Geesthacht (HZG), Geesthacht

4.5.1 General information

The interviewed researcher is a senior scientist and his group's objective is "Climate simulation of the past millennia". His role in the group includes climate research, supervision of PhD students, and responsibilities as a member of the editorial board in some journals. He collaborates with one other group within the institute and about ten external groups.

4.5.2 Data infrastructure

Research data are generated with the help of climate models, typically within one calendar year. Archiving occurs almost simultaneously to data generation, while data analysis may take between 2 and 5 years. The data are collected in binary form and in the TB range. No own software development takes place in primary data generation nor is proprietary software being used (community models produce the data). On the basis of the primary model, output data summaries of the data and spatial and temporal coverage are stored. The data are annotated with metadata, whereby all steps in post processing are automatically generated and attached to the resulting secondary data. Almost always are the standard metadata used. For the derivation of secondary data, software is being developed; it is not necessary to use proprietary formats or software for secondary data representation.

Software and secondary data are stored and archived in a repository within the group while primary data are stored and archived on a supercomputer at

[162] http://pubs.usgs.gov/sir/2008/5172/sir2008-5172.pdf.

DKRZ. In addition, copies of the software are kept on own storage devices on the hard disk of the office PC. This situation is thought to be representative for climate research. There is no person specifically responsible in the group for data management issues, which the researcher finds satisfactory, because data management is an integral and closely related part of the scientific activity. It should be integrated in the research projects proper. The data are stored on servers accessible to every member of the group. The local computer network is the support tool for this.

Software and secondary data of this group are made available to the general public, but because of their large volume the primary data is only available to other research projects. Another reason for this restriction is that some sort of guidelines and interpretations from the originating group are needed, but is true for secondary data as well. It seems that the situation is representative in climate science and no further incentives for data exchange of research data and software seem to be necessary. Also as a matter of principle, software including its source code is made available to other institutions. The reason is that when software is requested to exactly understand the data post processing, software sharing is the most accurate and convenient way. Only when data sharing conflicts with current research projects occur, an exception/restriction is made. Data exchange is enabled by writing the data in a commonly used format and placing them on a public server from where they may be downloaded via ftp.

4.5.3 Literature

For this climate science group, publication of results in print media is still preferred but infrequent, whereas publication online or offline is the preferred and established practice, typical for climate science today. The combined publishing as book/CD-ROM or proceedings plus website is not preferred. There are publishers which enable the exchange of data and/or literature, but for this group it is currently not possible to publish data and literature together. However, this is not viewed as a problem. The reason is that the volume of data may be very large and the same data may be used in several publications, e.g. showing different types of analysis for recent research targets.

Open Access is to some extent established in the group, and this seems to be representative for the discipline, Open Access journals offering their own technologies for online publication.

4.5.4 Outlook

Data infrastructure: It is envisioned that centralised sites would be used as repositories and would manage the data sharing, taking into account proprietary issues and the need of scientific replication. These sites would accept data from different groups and make them available to other groups conditional on certain use (e.g. replication of published results), or extended analysis where coauthorship should be required.

Literature infrastructure: This researcher does not foresee a lot of further developments. Some journals will remain restricted and others would go for full Open Access (Golden Route). Journals will increasingly offer the possibility to publish graphic material, such as videos, and perhaps some journals will adopt wiki technologies. But since journals are commercial products after all, their success depends on marketing factors among scientists which are very difficult to predict.

4.6 Federal Maritime and Hydrographic Agency (BSH), Hamburg

4.6.1 General information

The interviewed researcher is a senior scientist in the group "Marine Data and Interpretation Systems", which is mainly concerned with national and international data management projects. The 28 employees in this group have 28 PCs at their disposal and access to six servers. The members of the group cooperate with six other groups within the institution and 72 external groups.

4.6.2 Data infrastructure

Data collection focuses on oceanographic real-time data, e.g. hydrographic, ocean current and wave height measurements, from the North West European Shelf and the Baltic Sea. The gathered data are undergoing real-time quality control, and are enriched by adding metadata and combining the data sets with data from other disciplines. Data are archived in the German National Oceanographic Data Centre, which is part of the BSH. The principal (re-)use of these data occurs within the spatial data infrastructure of the institution. The storage requirements for such oceanographic real-time data sets lie in the gigabyte range (ASCII and binary data).

From the primary data the following secondary data are typically derived: oceanographic products, maps, statistics, trends and some others, using also software that was developed within the institution. The secondary data are published in the internet and in the scientific literature. For metadata annotation, the standard ISO19115 is mostly used. If some partners in European

projects do not use ISO19115, however, deviations from this standard may occur. When collecting, representing and processing the primary data, the group relies on proprietary formats and software to a great extent.

Primary and secondary data as well as developed software are all archived on an in-house data server. The interviewee feels that having an in-house data server is typical or "standard" for comparable groups in climate/ocean science and also that certain staff members are responsible for data management.

To assure access to their data, this BSH group enters into so-called service level agreements, for example in the EU project MyOcean[163] (ocean monitoring and forecasting), a 3-year project that started on 1 April 2009. In the module "Weather, seasonal forecasting and climate", interested parties may register to view the full catalogue of products and services. A non-exhaustive list is shown in Figure F.15.

This interviewee's partner's group has also signed memoranda of understanding with the Baltic Operational Oceanographic System (BOOS)[164] and the North West European Shelf Operational Oceanographic System (NOOS).[165] The principal vision of both organisations is to develop and implement online operational marine data and information services. One of the objectives of BOOS is to contribute to ocean climate variability studies and seasonal climate prediction, and NOOS wants to establish a marine database from which time series and statistical analyses can be obtained, including trends and changes in the marine environment and the economic, environmental and social impacts.

PRODUCTS AND SERVICES

WEATHER, SEASONAL FORECASTING & CLIMATE

DOMAINS OF APPLICATION	FREQUENTLY REQUESTED PARAMETERS	PRODUCT FAMILIES FREQUENTLY IN USE	USERS INVOLVED IN "*Users Requirement Definition*" (*)
MyOcean delivers reliable and robust data to the European and national meteorological services. Physical parameters of the ocean's surface are used as boundary conditions for atmospheric models. Changes in sea ice extent, concentration and volume are signals used to detect global warming for instance.	Temperature Salinity Currents Sea Level Sea Ice	Reanalysis of physical parameters at various temporal resolutions (monthly, seasonal, yearly) Long time-series of in-situ (physical parameters) and remote sensing (SST, SLA) products Analysis and forecasts of hydrodynamic models at global and regional scales	ECMWF (European Centre for Medium-Range Weather Forecasts) is a MyOcean key-user. National Weather Services Climate Research centres

Figure F.15 List of parameters, their domains of application, products and user groups that typically request data for the module "Weather, seasonal forecasting and climate" via the MyOcean website[166]

[163] http://www.myocean.eu.org.
[164] http://www.boos.org.
[165] http://www.noos.cc.
[166] http://www.myocean.eu/web/19-products-and-services.php?domain=forecast.

The BSH group makes available research products, i.e. software, to close colleagues and other research projects, and primary data (some of which are output data from numerical models) both to these users and to the general public as well. Sporadically, software is also made available to other institutions (including the source code).

The main rule regarding the use of data provided by this group is that real-time data are handled as fast as possible. Data exchange is supported by ftp, http and web portals. Apart from these tools, spatial data infrastructures enable the re-use of data, while foreign data policies may hamper data re-use.

4.6.3 Literature

For the preparation of internal or external publications, software like MS-Office, LateX, FrameMaker and Adobe programs are used. Publications are done preferably on- and offline, albeit rather infrequently. Print media are not preferred as publication medium. However, this is viewed as being atypical for the discipline, where the preferred publishing medium is printed papers. This group does not foster Open Access practices with respect to literature.

4.6.4 Outlook

The way that the EU project INSPIRE is developing a data infrastructure is seen as a promising way for the future.[167]

5 Current status of Open Access in climate research literature

The implementation of Open Access publishing in science in general is contingent on the success of new ways of financing the system[168] and on a common understanding of intellectual property[169] rights.[170] Besides these aspects, in climate science the strong international interdependence on each other's data calls for agreements between partners which may be subjected to differing legal frameworks. Such agreements should be formulated at the beginning of the scientific workflows between potential users of the research results. The utilisation of climate science's large data sets, in particular, should be and is being backed by appropriate data policies and the development of technolo-

[167] http://www.inspire-geoportal.eu.
[168] http://www.ercim.eu/publication/Ercim_News/enw64/velterop.html.
[169] http://en.wikipedia.org/wiki/Intellectual_property.
[170] http://www.w3.org/IPR.

gies that enable transfer and access of the voluminous data sets in climate research. This aspect is dealt with in section 6.

In this section, first the organisation of scholarly information in general is described, followed by an overview of library resources in Germany that are of specific interest to climate scientists. As has been described in the previous sections, climate researchers operate in the given infrastructure their institutions, data repositories and data centres are embedded in. In order to make the outcome of their work known, to collaborate with colleagues and to advance their careers, scientists publish write ups of their research results and also, increasingly, data that their work is based on or that may be a useful starting point for further research by others. Some aspects concerning journals in which climate researchers frequently publish results of their work and some information on Open Access are given in section 5.3

5.1 Library management in Germany

Libraries as part of "Education, Culture and Science" are subject to regulations at the state level (*Länder*), and library legislation is a political statement in concrete terms that a state intends to guide, configure, cultivate and fund libraries. However, there exists no law in Germany in any of its 16 states that makes the operation and maintenance of public libraries mandatory. Discussions have been going on for over 50 years how this can be achieved, and in recent years the German Library Association (DeutscherBibliotheksverband DBV) has designed a "sample national library law",[171] which the states in Germany's federal system are encouraged to use, adjust and incorporate into their *Länder* laws. The DBV was inspired by the best practice examples observed in Denmark, Finland and Great Britain which all include mandatory library services, highly topical holdings taking new media and information technology developments into account, free-of-charge usage by anybody, sufficient funding by the municipality, financial provision for infrastructures and networks by the government and integration of the libraries into educational concepts

This lack of a firm legislative backbone notwithstanding, library services are supported financially by the DFG in its "Nationwide Library Services and National Licenses" programme to "facilitate the provision of a comprehensive range of highly specialised literature collections and digital sources of information for use in scientific research in Germany",[172] with the aim to allow access to specialised scientific information that cannot be had at/through

[171] http://www.bibliotheksportal.de/bibliotheken/bibliotheken-in-deutschland/bibliotheksgesetz.html.

[172] http://www.dfg.de/en/research_funding/programmes/infrastructure/lis/digital_information/library_licenses/index.html.

individual university libraries. All scientists and academics in Germany may use "internet-based services for bibliographical research, interlibrary loans, document delivery" and access digital collections directly online.

The DFG also offers funding opportunities for the acquisition of national licences for nationwide access to literature in digital form. All members of universities, (technical) colleges and research institutions located in Germany which have secured a national licence for certain publications may access these free of charge from the campus networks and the catalogues of German state and university libraries, while others may register for free individual use of many databases and text collections at www.nationallizenzen.de.

5.2 Libraries, literature databases and search tools

The DFG maintains a system of literature supply called "special subject collections"[173] (SSG, Sondersammelgebiete). Besides the libraries listed there, other specialist libraries allow scientists to find literature from the field of geosciences and specifically about climate topics.

5.2.1 Library of the Center for Marine and Atmospheric Sciences (ZMAW), Hamburg

The ZMAW[174] is a cooperative centre of several institutes of the University of Hamburg and the Max Planck Institute for Meteorology promoting research in the fields of marine, climate and earth system science in Hamburg. The Institute of Coastal Research of the Helmholtz Centre Geesthacht (HZG) has been an associated member of the ZMAW since 2005.

The ZMAW Library[175] is a special library for the earth sciences. It is also a member of the German Association of Marine Science Libraries and Information Centers (GAMSLIC), whose catalogue[176] contains the holdings of the libraries of 12 other institutions:

- Alfred Wegener Institute Foundation for Polar and Marine Research,[177] Bremerhaven (including the library of the former Biologische Anstalt Helgoland)

[173] http://dispatch.opac.ddb.de/DB=1.1/LNG=EN/SID=a4a010ab-17/SSG.
[174] http://www.zmaw.de/The-ZMAW.4.0.html?&L=1.
[175] http://www.zmaw.de/index.php?id=5&L=1.
[176] http://gso.gbv.de/DB=2.910/LNG=EN/?COOKIE=U999,K999,D2.910,E874afaef-192,I0,B9994++++++,SY,A\delimiter"026E30F9008+*,H13-15,,17-23,,30,,73-78,,88-90,NGAST,R136.172.96.229,FN&UCLOAD=Y&COOKIE=U999,K999,D2.910,E874afaef-192,I0,B9994++++++,SY,A\delimiter"026E30F9008+*,H13-15,,17-23,,30,,50,,60-61,,73-78,,88-90,NGAST,R136.172.122.54,FN.
[177] http://www.awi.de/en/infrastructure/library.

- Bundesamt für Seeschifffahrt und Hydrographie,[178] Hamburg
- Forschungsanstalt der Bundeswehr für Wasserschall und Geophysik,[179] Kiel
- German Maritime Museum,[180] Bremerhaven
- HZG: Institute of Coastal Research[181] (journals only), Geesthacht
- Johann Heinrich von Thünen-Institute, Federal Research Institute for Rural Areas, Forestry and Fisheries,[182] Hamburg
- Leibniz Institute for Baltic Sea Research,[183] Warnemünde
- Leibniz Institute of Marine Sciences[184] (IFM-GEOMAR), Kiel
- Leibniz Center for Tropical Marine Ecology,[185] Bremen
- Max Planck Institute for Evolutionary Biology,[186] Plön
- Deutsches Meeresmuseum,[187] Stralsund
- Terramare Research Centre,[188] Wilhelmshaven.

While these libraries have a marine focus, the ZMAW's Library and Information Service (LIS) has its roots in the Department of Earth Sciences of the University of Hamburg and the Max Planck-Institute for Meteorology and therefore has a special focus on these research fields. The LIS on its website[189] provides entry points to catalogues, databases, journals, the LIS service and ZMAW publications.

A search often starts with a click on "Journals", resulting in the display shown in Figure F.16.[190] Some Open Access journal groups are indicated directly, while many more may be found in the Electronics Journal Library[191] (Elektronische Zeitschriftenbibliothek; EZB).

The EZB provides information about electronic journals (not the article texts themselves). Partner institutions of the EZB are University Libraries in Germany, institutes belonging to the Max Planck Society or the Helmholtz Association, State and Regional Libraries, some university libraries in Austria and Switzerland, and other institutions. Researchers at these institutions have free full-text access to those journals/articles that are marked green

[178] http://www.bsh.de/en/The_BSH/Organisation/Library/index.jsp.
[179] http://www.fwg-kiel.de.
[180] http://www.dsm.museum/bibliothek.33.de.html.
[181] http://www.hzg.de/central_departments/library/index.html.en.
[182] http://vzopc4.gbv.de:8080/DB=19.2/LNG=DU.
[183] http://www.io-warnemuende.de/library-and-it-group.html.
[184] http://www.ifm-geomar.de/index.php?id=bibliothek_home&L=1.
[185] http://www.zmt-bremen.de/en/Library.html.
[186] http://www.evolbio.mpg.de/english/bibliothek/index.html.
[187] http://www.meeresmuseum.de/wissenschaft/bibliothek.html.
[188] http://www.icbm.de/32114.html.
[189] http://www.zmaw.de/index.php?id=5&L=1.
[190] http://www.zmaw.de/Journals.46.0.html?&L=1.
[191] http://rzblx1.uni-regensburg.de/ezeit/index.phtml?bibid=DM&colors=7&lang=en.

Figure F.16 The "Journals" web page of ZMAW's Library and Information Service

(indication of complete Open Access) or yellow (indication that a licence fee has been paid) at the URL that is produced when doing an alphabetic journal title search on the EZB page, for example "Journal of Geophysical Research".[192]

The EZB, which used to be funded by the BMBF, the DFG and the Bavarian Ministry of Science, Research and Art, provides links to journals, newspapers and databases in German and Austrian libraries and is a cooperative effort of 4300 libraries. The EZB is one of the largest databases worldwide for finding journals, newspapers, reports and other periodicals from any country and language in electronic format.

The result of a search for all journals that are available online from the American Meteorological Society, for example, is shown in Figure F.17.[193] All titles shown are Open Access, some with restrictions as to the actuality, i.e. only older issues are fully Open Access. A search in the EZB by subject "geosciences" lists journal titles in alphabetical order, as shown in Figure F.18.

For each journal, the access mode is indicated by the (coloured) dots along the right hand margin: Green (OXX): full text is freely accessible (Open Access); Yellow (XOX): full text can be accessed within the Campus-Net and for university members also from outside the campus; Yellow/red (XOO): full text access only for parts of all published issues; Red (XXO: no free access

[192] http://rzblx1.uni-regensburg.de/ezeit/fl.phtml?bibid=MPIM&colors=7&lang=de&notation=ALL&sc=J&lc=K&sindex=3650#jumpto.

[193] http://rzblx1.uni-regensburg.de/ezeit/searchres.phtml?bibid=MPIM&colors=7&lang=de&jq_type1=KT&jq_term1=&jq_bool2=AND&jq_not2=+&jq_type2=KS&jq_term2=&jq_bool3=AND&jq_not3=+&jq_type3=PU&jq_term3=american+meteorological+society&jq_bool4=AND&%20jq_not4=+&jq_type4=IS&jq_term4=&offset=1&hits_per_page=50&search_journal=Suche*starten&Notations[]=all&selected_colors[]=1&selected_colors[]=2&selected_colors[]=4.

possible for the location. Sometimes access is possible to abstracts or tables of content.

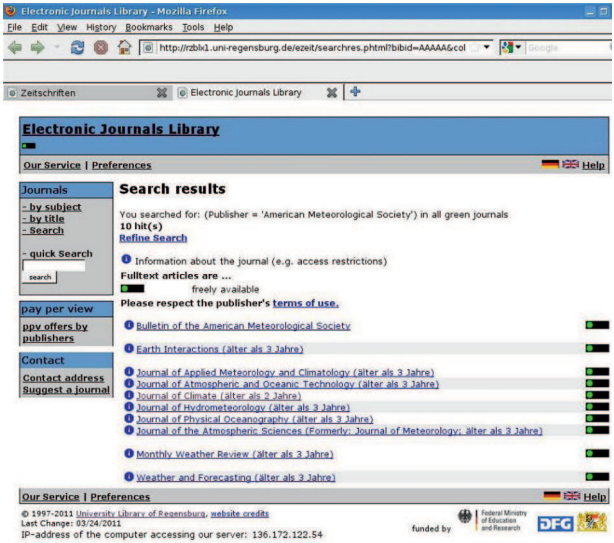

Figure F.17 Journals of the American Meteorological Society, available within the library system of the University of Hamburg

Figure F.18 EZB listing of journal titles in the field of "geoscience" with access information for University of Hamburg users

5.2.2 National Meteorological Library hosted by the German Weather Service[194]

The stock of this library reaches back to the 15th century, documenting the development of meteorology as an independent science from its origins up to now. This library holds 180,000 volumes and approximately 800 current journal and periodical titles. It is the official special collecting library for meteorology, climatology and meteorological maps and charts.

The library offers various search methods through the Meteorological Literature Information System (METLIS),[195] which draws from entries about practically all specialist publications since the introduction of electronic data processing,[196] and referring also to external databases such as the internationally recognised Meteorological and Geoastrophysical Abstracts (MGA).[197]

5.2.3 Library at the Federal Maritime and Hydrographic Agency, Hamburg and Rostock[198]

This library is the central maritime library in Germany. The initial stock dates back to 1868 and began with nautical charts and books. The collection grew steadily, but in World War II parts of it were irretrievably lost. After the German reunion the library collections of the BSH and the Seehydrographischer Dienst (SHD) of the German Democratic Republic (GDR) were combined. The collection includes approximately 170,000 media and 50,000 nautical charts, openly accessible by all scientists, researchers, historians, students and the general public. The holdings cover the subjects interesting for climate scientists like oceanography (excluding marine biology), marine physics, marine chemistry, marine geophysics, marine geology, marine meteorology, marine environmental protection and various other nautical subjects.

5.3 Climate science literature management

In sections 5.1 and 5.2, the infrastructure of literature resources that researchers have at their disposal in Germany was described. In this section,

[194] http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_windowLabel=dwdwww_main_book&T34202240811665321680 92gsbDocumentPath=&switchLang=en&_pageLabel=dwdwww_menu2_bibliothek.

[195] http://oflsd45.dwd.de:8060/alipac/LGDONSVLIKSRQCMRGYBQ-00001/form/find-simple.

[196] http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/PB/PBFB/Bibliothek/Allgemein/en_Bibliotheksflyer,templateId=raw,property=publicationFile.pdf/en_Bibliotheksflyer.pdf.

[197] http://www.csa.com/factsheets/mga-set-c.php.

[198] http://www.bsh.de/en/The_BSH/Organisation/Library/index.jsp.

some aspects of literature management from the viewpoint of the scientists are discussed.

5.3.1 Journal Citation Report (JCR, ISI) and Open Access in climate science journals

From the survey of researchers (section 4), journal titles emerged which were preferred for publishing scientific results. I used the Journal Citation Report[199] (JCR, ISI) to determine the most frequently cited journals in the field of climate science[200] and to see how many of those are Open Access, i.e. are also listed in the Directory of Open Access Journals[201] (DOAJ).

In the JCR (ISI) subject category "Meteorology & Atmospherics Sciences", an impact factor[202] (and other journal metrical quantities) is given for 68 journals. The journal with the highest impact factor, "Atmospheric Chemistry and Physics" is, in fact, also an Open Access title. The other four Open Access journals have satisfactory impact factors because the median impact factor is 1.6.[203]

In the DOAJ, one finds journals by "browsing by subject".[204] For example, in subject "Earth and Environmental Sciences" 26 journals are listed in the subcategory "Meteorology and Climatology"[205]. For five of these 26 journal titles, journal metrics are given in the JCR (ISI) subject category "Meteorology & Atmospheric Sciences".[206] This means that 19% of the relevant journal titles in this subject area are Open Access. Their titles, ISSN and impact factor (IF) are:

- *Atmospheric Chemistry and Physics:* ISSN 1680-7316, IF 5.5
- *Atmospheric Measurement Techniques:* ISSN 1867-1381, IF 2.6
- *Natural Hazards and Earth System Sciences:* ISSN 1561-8633, IF 1.8
- *Journal of the Meteorological Society of Japan:* ISSN 0026-1165, IF 1.1
- *SOLA: Scientific Online Letters on the Atmosphere:* ISSN 1349-6476, IF 1.0.

199 http://admin-apps.webofknowledge.com/JCR/JCR?SID=U2699KNOGA%406h4HepF1.
200 http://admin-apps.webofknowledge.com/JCR/help/h_jcrabout.htm.
201 http://www.doaj.org.
202 The annual Journal Citation Reports impact factor is a ratio between citations and recent citable items published: a journal's impact factor is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years. http://thomsonreuters.com/products_services/science/academic/impact_factor.
203 http://admin-apps.webofknowledge.com/JCR/JCR?RQ=LIST_SUMMARY_CATEGORY&category_sort_by=cat_title&cursor=1.
204 http://www.doaj.org/doaj?func=subject&cpid=78&uiLanguage=en.
205 http://www.doaj.org/doaj?func=subject&cpid=86&uiLanguage=en.
206 http://admin-apps.webofknowledge.com/JCR/JCR?RQ=LIST_SUMMARY_JOURNAL&cursor=1.

5.3.2 Open Access mandates of science organisations

In the survey (section 4), the interviewed scientists described how they as authors of papers and data sets organise the publication of the results of their research. There are established rules, either through the agencies funding projects or the institutional policies where climate science produces results.

Via the OpenAIRE Portal, one finds descriptions of the two policies presently in place for Open Access in Europe,[207] i.e. mandates for scientists funded through projects of the European Research Council (ERC)[208] and the Seventh Framework Programme of the European Commission (FP7),[209] and what, where and when should be deposited.

The main research organisations supporting climate science in their institutes provide information on their Open Access policies and requirements on special websites, e.g. the Max Planck Society[210] Helmholtz Association[211] and Fraunhofer-Gesellschaft.[212] The Leibniz Association's Open Access Working Group has been developing a basic position on Open Access since 2005[213] and the Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (i.e. the Leibniz Association) maintains a distributed Open Access repository[214] More general information and Open Access implementation advice is available through the Open Access information platform[215]

5.3.3 Tools used for literature management

Scientists use a variety of tools that support their literature management und publication in practice. One is Zotero,[216] a plug in to the Firefox web browser as a free tool that is used to collect, organise, cite and share their research resources. Another one is JabRef,[217] an open source bibliography reference manager using the file format BibTeX, a standard LaTeX bibliography format, or other bibliographic software. A widely used tool for literature searches and sorting the discovered information is Elsevier's platform, SciVerse,[218]

[207] http://www.openaire.eu/en/open-access/open-access-in-fp7.
[208] http://www.openaire.eu/en/component/attachments/download/3.
[209] http://www.openaire.eu/en/component/attachments/download/4.html.
[210] http://oa.mpg.de.
[211] http://oa.helmholtz.de/index.php?id=137.
[212] http://www.fraunhofer.de/content/dam/zv/de/publikationen/Fraunhofer_OpenAccessPolicy.pdf.
[213] http://www.wgl.de/?nid=akroa.
[214] http://www.tib-hannover.de/en/the-tib/wgl-repository.
[215] http://open-access.net/de_en/homepage.
[216] http://www.zotero.org.
[217] http://jabref.sourceforge.net.
[218] http://www.info.sciverse.com/UserFiles/resource_library_brochures/sciverse-brochure.pdf.

which allows access to the ScienceDirect database of peer-reviewed articles and books and to the Scopus citation database, among other applications.

5.3.4 Preferred journals for publication in climate research

From the survey of climate scientists, the following journals[219] emerged as the main publications where the research results are being published. Most of them charge a (often moderate) publication fee and offer various Open Access options.

1. *Hydrology and Earth System Sciences*: ISSN 1027-5606, IF 2.5), interactive Open Access journal of the European Geosciences Union, Open Access, public peer-review and interactive public discussion, personalised copyright under a Creative Commons Licence, moderate service charges.[220,221]
2. *Remote Sensing of the Environment:* ISSN 0034-4257, IF 3.95, interdisciplinary journal of Elsevier. Open Access journal offering authors the option of making their article freely available to all via the ScienceDirect platform. The fee of $3,000 excludes taxes and other potential author fees such as colour charges.[222,223]
3. *Journal of Climate:* ISSN 0894-8755, IF 3.5, online journal of the American Meteorological Society (at the Library and Information Service of the ZMAW in Hamburg there is free access to full text for issues older than 2–3 years).[224,225]
4. *Remote Sensing:* ISSN 2072-4292, online journal of Yale's Center for Earth Observation, Open Access journal, is published by the Multidisciplinary Digital Publishing Institute (MDPI) online monthly. Free for readers, with low publishing fees paid by authors or their institutions. Rapid publication: accepted papers are immediately published online.

[219] Where available from the 2010 JCR Science Edition, the ISSN and (impact factor) are also given.
[220] http://www.hydrology-and-earth-system-sciences.net/home.html.
[221] http://www.hydrology-and-earth-system-sciences.net/submission/service_charges.html.
[222] http://www.elsevier.com/wps/find/journaldescription.cws_home/505733/description#description.
[223] http://www.elsevier.com/wps/find/journaldescription.cws_home/505733/authorinstructions.
[224] http://journals.ametsoc.org/toc/clim/current.
[225] http://rzblx1.uni-regensburg.de/ezeit/searchres.phtml?bibid=AAAAA&colors=1&lang=en&jq_type1=KT&jq_term1=&jq_bool2=AND&jq_not2=+&jq_type2=KS&jq_term2=&jq_bool3=AND&jq_not3=+&jq_type3=PU&jq_term3=American+Meteorological+Society&jq_bool4=AND&jq_not4=+&jq_type4=IS&jq_term4=&offset=-1&hits_per_page=50&search_journal=Start+search&Notations[]=all&selected_colors[]=1.

MDPI is a publisher of peer-reviewed, Open Access journals since its establishment in 1996.[226,227]

5. *Biogeosciences:* ISSN 1726-4170, IF 3.6, interactive Open Access journal of the European Geosciences Union. Public peer review and interactive public discussion, personalised copyright under a Creative Commons Licence.[228]
6. *Journal of Geophysical Research (Atmospheres):* ISSN 0148-0227, IF 3.3, journal of the American Geophysical Union.[229]
7. *Geophysical Research Letters:* ISSN 0094-8276, IF 3.5, journal of the American Geophysical Union.[230]
8. *Journal of Hydrology:* ISSN 0022-1694, IF 2.5, journal of Elsevier, with Open Access solutions.[231,232]
9. *Global Environmental Change:* journal by Elsevier, with Open Access solutions.[233,234]
10. *Comptes Rendus Geoscience:* ISSN 1631-0713, IF 1.7, journal by Elsevier.[235]
11. *IEEE Geoscience and Remote Sensing:* ISSN 0196-2892, publication(s) of the IEEE Geoscience and Remote Sensing Society.[236]
12. *ISPRS Journal of Photogrammetry and Remote Sensing:* ISSN 0924-2716, IF 2.2, journal of Elsevier.[237]
13. *Earth-Science Reviews:* ISSN 0012-8252, IF 5.8, journal of Elsevier.[238]

More journals relevant for climate research can be found in a Geographic Information System and Remote Sensing journal list which has been compiled by Prof Giorgos Mountrakis of the State University of New York, College of Environmental Science and Forestry.[239]

[226] http://www.mdpi.com/journal/remotesensing.
[227] http://www.mdpi.com/about.
[228] http://www.biogeosciences.net.
[229] http://www.agu.org/journals/jd.
[230] http://www.agu.org/journals/gl.
[231] http://www.elsevier.com/wps/find/journaldescription.cws_home/503343/description#description.
[232] http://www.elsevier.com/wps/find/authorsview.authors/openaccess.
[233] http://www.elsevier.com/wps/find/journaldescription.cws_home/30425/description#description.
[234] http://www.sciencedirect.com/science/journal/09593780.
[235] http://www.sciencedirect.com/science/journal/16310713.
[236] http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=36.
[237] http://www.elsevier.com/wps/find/journaldescription.cws_home/503340/description#description.
[238] http://www.elsevier.com/wps/find/journaldescription.cws_home/503329/description#description.
[239] http://www.aboutgis.com/gis-and-remote-sensing-journal-list-with-impact-factors.

6 Current status of Open Access in climate research data

The different data collections that climate research produces or uses were the subject of section 3. In the following sections, additional information is given regarding the access to the data held in those centres, including policies and tools for data retrieval and sharing.

A large step forward in the development of climate data infrastructure was the creation of the World Data Center System[240] which came into existence as a result of the International Geophysical Year (IGY) of 1957/58. The purpose of World Data Centers was to ensure that observational data from the IGY programme would be readily available to scientists of all countries. The arrangement became permanent under the auspices of the International Council of Sciences (ICSU) and has remained so. Today the WDC system includes 52 centres in 12 different countries, 15 of which operate in the USA, 7 in Russia, 11 in Europe, 10 in Australia, India and Japan and 9 in China.[241] WDCs are funded and maintained by their host countries. Details regarding rules, responsibilities, data acquisition and usage of the WDC system are described in the World Data Center System guide.[242]

As a major funding agency of climate research in Germany, the DFG, more exactly its Committee on Scientific Library Services and Information Systems, Subcommittee on Information Management, in 2009 published seven recommendations for secure storage and availability of digital primary research data.[243] In the field of climate science these recommendations have already been successfully implemented to a large extent.

6.1 German Weather Service (DWD Deutscher Wetterdienst), Offenbach[244]

This is the DWD's contact point for the provision of data and products (e.g. numerical products or radar data) to special users, i.e. its clients are typically meteorological service providers, universities and research institutions, and authorities of the German Federal Government or the *Länder* (16 German states), who may receive real-time data and products as well as archived data (called "climate data" by the DWD) and products.

[240] http://www.icsu-wds.org/.
[241] http://www.icsu-wds.org/wds-members/wds-members.
[242] http://www.wdc.rl.ac.uk/wdc/guide/wdcguide.html.
[243] http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901_en.pdf.
[244] http://www.dwd.de.

A portion of the data are offered for "free", that means free of charge and mostly without any restrictions for the use of these data (where applicable, restrictions of use are specified next to the data set concerned). The list of data that are available online includes climate data for Germany, climate data from satellites, climate data worldwide and precipitation data worldwide. The broad spectrum of data is displayed on the "Climate and Environment" web page of DWD[245] (note: not all descriptions are available on the English version of this page).

Various web applications of the DWD contain extensive information about the climate data and available (metadata), which facilitate access to the data sets. One such application is WebWerdis (Web-based Weather Request and Distribution System) which provides access to free online climate data without the necessity of registration. WebWerdis is aimed at users with some prior subject knowledge, in particular from research and educational institutions and public authorities. Proper registration with WebWerdis, however, allows full access to all contents and many usage options.

Additional metadata are contained in the Climate Catalogue of the DWD's Climate Data Center (CDC). This data catalogue is currently still being extended and will in the end contain nationally and internationally standardised metadata (ISO19115[246] and ISO19139[247]) for all data available at the DWD. Direct access will be possible in many cases.

The Global Data Set (GDS) comprises freely available data and products relating to current weather and weather forecasts as well as freely available climate data. The data and products are usually presented in the same form that is used for their exchange within and between the national meteorological services. Access is possible via ftp. The use of the GDS is free of charge but requires registration. To obtain any of a large list of other data, products, and services, DWD asks potential customers to contact staff directly.

For special data collections in some international data centres hosted by DWD, see also Table F.2 in section 3.1.

[245] http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_windowLabel=dwdwww_main_book&T82002gsbDocumentPath=Navigation%2FOeffentlichkeit%2FKlima__Umwelt%2FKlimadaten%2Fkldaten__kostenfrei%2FAbrufsysteme__Daten__node.html%3F__nnn%3Dtrue&switchLang=de&_pageLabel=_dwdwww_klima_umwelt_klimadaten_deutschland.

[246] ISO 19115:2003 defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020.

[247] ISO/TS 19139:2007 defines Geographic MetaData XML (gmd) encoding, an XML Schema implementation derived from ISO 19115, http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557.

6.2 World Data Center for Remote Sensing of the Atmosphere (WDC-RSAT), Oberpfaffenhofen[248]

Data and products available through WDC-RSAT cover many subjects and span a wide range of processing levels. Direct access online is provided to atmosphere-related satellite-based data sets if they are either stored at the WDC-RSAT or found through the WDC-RSAT portal, i.e. when are safeguarded by other providers (for data and variable types compare with section 3.2). Furthermore a table of services regarding air quality forecasting and monitoring, Antarctic ozone hole monitoring, solar energy, virtual lab and sunburn time is on display.[249]

The policy of data use[250] requires that acknowledgement/reference is made to the ICSU World Data Center for Remote Sensing of the Atmosphere or to cite specific references where these are provided.

6.3 World Data Center for Marine Environmental Sciences (WDC-MARE), Bremen[251]

Data contained in WDC-MARE are discovered via the information system Publishing Network for Geoscientific and Environmental Data (PANGAEA), the search engine for which online guidance is provided.[252] PANGAEA operates as an Open Access library to archive, publish and distribute georeferenced data from earth system research. The operating institutions have committed themselves to make the content of WDC-MARE available for the long term.[253]

The majority of the data are freely available, usable under the terms of the licence expressed in the data set description. Some data sets are under moratorium from ongoing projects, but metadata are visible, as is the contact information of the principal investigator who may be asked for access.

Data sets at WDC-MARE can be identified, shared, published and cited via a DOI. Data may be archived as supplements to publications or as citable data collections. The German National Library of Science and Technology has a portal through which citations may be obtained.[254]

Data management and archiving at WDC-MARE follows the principles and responsibilities of ICSU World Data Centers[255] and the OECD principles

[248] http://wdc.dlr.de/data_products.
[249] http://wdc.dlr.de/data_products/SERVICES.
[250] http://wdc.dlr.de/data_products/data_use_policy.php.
[251] http://www.wdc-mare.org/data.
[252] http://www.wdc-mare.org/shared/help/help.php/search/index.html.
[253] http://www.pangaea.de/about.
[254] https://getinfo.de/app/?lang=en.
[255] http://www.wdc.rl.ac.uk/wdc/guide/gdsystema.html.

and guidelines for access to research data from public funding.[256] Authors submitting data to the PANGAEA data library for archiving agree that all data are provided under a Creative Commons Licence.[257]

For data set discovery, the data mining tool Advanced Retrieval Tool (ART)[258] and an internet mapper for georeferenced data may be used.[259] Documentation on the usage of these tools is available through a wiki.[260]

6.4 The Bremen Core Repository (BCR)[261]

The procedure for obtaining sediment core data/samples from any ocean drilling programme or deep sea drilling project is to fill out the Integrated Ocean Drilling Program (IODP) online sample request form[262] (for samples from any of the three repositories, see section 3.3). The same holds for samples from the IODP from expeditions with the non-riser vessel (JOIDES Resolution). More information on data access, policies, guidelines, procedures and obligations and a variety of sample request forms is provided on the corresponding IODP web page.[263] Samples can be either taken by repository staff, or by the scientists themselves. MARUM provides a checklist for planning a visit to the repository.[264]

6.5 National Oceanographic Datacentre for Germany (NODC), Hamburg[265]

The NODC of Germany is partner of the SeaDataNet[266] consortium of 35 countries forming a a unique virtual data management system network providing integrated or raw data sets of standardised quality online. This is achieved by the Common Data Index (CDI) service, that gives users detailed insight in the availability and geographical spread of marine data across the different data centres in Europe. The CDI provides an ISO19115 based index (metadatabase) to individual data sets – which may be samples, time series, profiles or trajectories – and it is the interface to online data access. Direct

[256] http://www.oecd.org/document/2/0,3746,en_2649_34293_38500791_1_1_1_1,00.html
[257] http://creativecommons.org/licenses.
[258] http://www.pangaea.de/advanced/ART.php.
[259] http://mapserver.pangaea.de.
[260] http://wiki.pangaea.de/wiki/Main_Page.
[261] http://www.marum.de/en/IODP_Core_Repository.html.
[262] http://www.iodp.tamu.edu/curation/samples.html.
[263] http://www.iodp.org/access-data.
[264] http://www.marum.de/en/Information_for_visitors_to_the_BCR.html.
[265] http://www.bsh.de/en/Marine_data/Observations/DOD_Data_Centre/index.jsp.
[266] http://www.seadatanet.org.

access to all NODC data is possible via this CDI. For submitting data requests and for downloading data, registration is required[267] (see Figure F.19). Such registration requests are managed per country by the specific National Oceanographic Data Centre/Marine Data Centre, which checks the authenticity of the user, and if ok, lets the central user register provide the user with access information.

In addition to offering this direct data access, the NODC of Germany contributes to SeaDataNet's five metadata catalogues and has the leading role for one of them, i.e. the Cruise Summary Reports Database. NODC therefore gathers all Cruise summary reports from the 42 SeaDataNet partners. It includes cruises from 1873 until today from more than 2000 different research vessels amounting to a total of more than 40000 cruises, in all European waters and global oceans. NODC is also integrated in the GeoSeaPortal[268] thereby implementing a national infrastructure for all marine data.

As was said in section 3.4, data may be accessed through the MUDAB web client,[269] for which online information in a brochure[270] and the MUDAB handbook[271] is offered (both texts are only available in German at present). When searching for data, users and customers of NODC are guided by a menu-based interface to the desired information. As shown in Figure F.20, starting points are either at the cruise, at the station or at the data level.

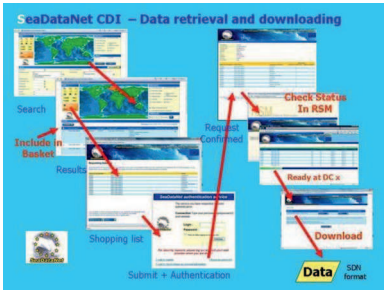

Figure F.19 SeaDataNet's Common Data Index: data retrieval and downloading (source: SeaDataNet newsletter no. 6, March 2011)[272]

[267] http://www.seadatanet.org/Data-Access/User-registration.
[268] http://www.bsh.de/de/Meeresdaten/Geodaten/index.jsp.
[269] http://www.mudab.de.
[270] http://www.informus.de:8080/mudab/documents/070530_mudab_webclient_faltblatt.pdf.
[271] http://www.informus.de:8080/mudab/documents/mudabws-usage.html.
[272] http://www.seadatanet.org/News/Seadatanet-Newsletter-n-6-March-2011.
[273] http://www.bsh.de/de/Meeresdaten/Umweltschutz/MUDAB-Datenbank/_1493.gif.

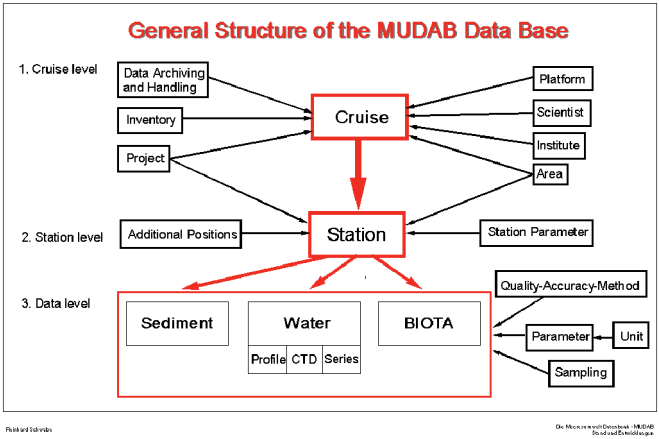

Figure F.20 MUDAB database scheme (courtesy of Reinhard Schwabe, NODC)[273]

It is also possible to retrieve information by SQL. The criteria for database retrieval are dependents of space and time, measured variables and sampling and analysis methods. The retrieved information can be exported for further data processing in a format readable by many PC-based tools. In the near future, the data export facilities will be implemented either into standard formats or into an interface for biological data.

Whereas data requests to the NODC are generally free of costs, for dedicated products and other demands, a price list for digital data including some usage conditions (annexes 2 and 3) is available on the BSH website.[274]

The general terms of use for the MUDAB[275] are based on Open Access principles (commercial companies may be charged for retrieval costs):

- For all datasets, access is granted free of on condition that the user agrees:
  - data are for your scientific use, in particular it is not allowed to use them for any commercial purpose, and shall not be forwarded to third party without our notice;

[274] http://www.bsh.de/de/Produkte/Preise/Entgeltverzeichnis_digitale_Daten.pdf.

[275] http://www.informus.de:8080/mudab/termsofuse.faces.

- to acknowledge the source of the data in all publications and applications;
- to help improve the quality of the data by noting and reporting any errors or omissions discovered;
- to help improve the quality of the Data Service by giving feed back on functionalities and data packaging;
- to help improve the efficiency of our reporting by supplying us with documented digital copies of data and information derived from the data so that it can be re-used by the Agency with reference to the source;
- to supply us with a copy of/URL to all publications and other products based on the datasets.

6.6 World Data Center for Climate (WDCC), Hamburg[276]

WDCC collects, stores and disseminates data related to climate research. It restricts itself to climate data products. Emphasis is spent on climate modelling and related data products. DKRZ as the operating institution of WDCC has committed itself as running WDCC as a long-term archive.

Data can be addressed and cited using a DOI. Metadata collected for data sets are available for the public. Data itself are available under a Creative Commons Licence unless not stated otherwise. To download data, registration at WDCC is required. Most of the data is available for immediate download.

Figure F.21 shows the gateway to the CERA database[277] at WDCC through which data sets can be located and retrieved. Several search options are offered, e.g. finding data sets by the name of an experiment. Entering one of the acronyms of observational projects listed in Table F.3 in section 3.6, will immediately lead to a display of the size of the downloadable data set. However, as most of the data sets in CERA are from numerical experiments, one may first want to browse the experiments and select, for example, IPCC_AR4_ECHAM5/MPIOM,[278] and from the ensuing list further choose the numerical simulation EH5-T63L31_AMIP_1_MM.[279]

Before downloading any digital data, which often have quite large volumes, it is possible to display selected post-processed variables of the model output as shown in Figure F.22. There the variable "EH5_AMIP_1_MM_STP1000", i.e. monthly means of air temperature at 1000 hectopascal" was selected

[276] http://www.dkrz.de/daten-en/wdcc.
[277] http://cera-www.dkrz.de/WDCC/ui/Index.jsp.
[278] Coupled numerical experiments carried out during CMIP3 with atmospheric model ECHAM5 and MPIOM of the Max Planck Institute for Meteorology, Hamburg.
[279] http://www-pcmdi.llnl.gov/projects/amip/NEWS/overview.php.
[280] http://cera-www.dkrz.de/WDCC/ui/Index.jsp

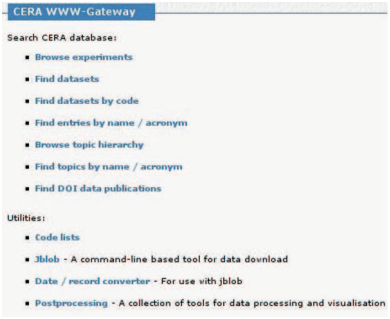

Figure F.21 The CERA WWW gateway[280]

and all 264 values for the runtime of the numerical simulation plotted (January 1978 through December 1999). Actually, since the monthly means of temperature had not been calculated and are therefore not contained in the database, the relative global maxima and minima were plotted instead (upper and lower curve, respectively).

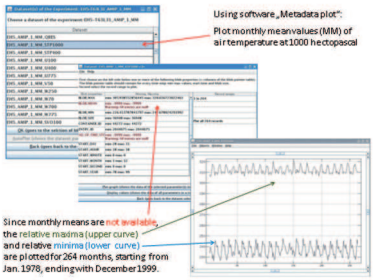

Figure F.22 Plotting selected model output variables before data set download

Apart from general information about the project in the context of which this coupled climate model experiment was carried out, the following information is provided:

citation specification of the originator of the data, storage details of the downloadable output, the computing environment in which the data were produced, a list of related experiments, the model domain in three spatial dimensions, the temporal coverage, data formats and data set size, and whether/that download permission exists.

Access to CMIP5 data Much of section 2 has focused on the currently ongoing CMIP5 project that has been designed around Open Access princi-

ples. Ideally data will be released immediately after they have been quality-checked. The adoption of Open Access principles of the CMIP5 Federation is motivated by the highly risen interest in climate research results and data by a wide spectrum of users. These include members of the natural and social sciences that work on climate (change) impacts and options for adaptation who respond to increasing public pressure and demand for sustainable development.

According to PCMDI's website,[281] all model output in the CMIP5 archive is available for "non-commercial research and educational purposes", and a subset of the data has also been released for "unrestricted" use, i.e. users registering to access CMIP5 output will be granted access to some or all of the data, depending on which terms of use have been agreed on. For users seeking CMIP5 model output, a "getting started" tutorial is provided.[282]

According to information which we had received from PCMDI earlier this year and which was recompiled in Table F.1 (see section 2.3.2), five of the 21 modelling groups had announced that they would make their data available to all users, six groups would release their data for non-commercial use and ten groups had not (yet) specified by whom the data from their experiments may be obtained and used.

A check of the status of the CMIP5 archive[283] showed that data sets are now available from nine of the participating Centres, i.e. group numbers 1, 2, 4, 6, 12, 17, 18, 19 and 20 in Table F.1 (status on 31 August 2011). More details about available data sets can be seen in a CMIP5 wiki[284]

7 Challenges for Open Access e-Infrastructures in climate research

Concurring with *Science* staff writers, who point out in the introduction to *Science*'s Special Online Collection "Dealing with Data"[285] that "most scientific disciplines are finding the data deluge to be extremely challenging, and tremendous opportunities can be realised if we can better organise and access the data", the research that went into this climate science chapter has clearly shown this to be true. The organisational effort invested into planning the data management and assuring future re-usability in the CMIP projects was described in sections 2.3–2.7. Nonetheless, bottlenecks still exist in data

[281] http://cmip-pcmdi.llnl.gov/cmip5/terms.html.
[282] http://cmip-pcmdi.llnl.gov/cmip5/data_getting_started.html?submenuheader=3.
[283] http://cmip-pcmdi.llnl.gov/cmip5/availability.html?submenuheader=3.
[284] http://esgf.org/wiki/Cmip5Status/ArchiveView.
[285] http://www.sciencemag.org/site/special/data.

storage capacities, access and (re-)usage by users from a wider community, i.e. non-experts in climate research (or from the public), common metadata and securing sufficient funding to support archiving.

Scientists need to meet their responsibilities toward transparency, standardisation and data archiving because large integrated data sets can potentially provide a better understanding of nature and society, thereby allowing new approaches in research that address key societal problems like, for example, coping with climate change. The associated challenges are to:

1. equip data repositories with the necessary hard-, middle- and software to manage large data sets from climate and earth system research,
2. ensure blockage-free data flow through networks,
3. staff the data centres/repositories adequately to guarantee an effective re-usage of data.

The "wares" of the first challenge can be met by securing funding for appropriate shopping lists of corresponding products and technical support for them.

With regard to the second challenge, Kostas Glinos, head of the European Commission's GÉANT and e-Infrastructures Unit, highlighted some European e-Infrastructure mainstays[286] at a conference on the "Role of e-infrastructures for Climate Change Research"[287] in May 2011 at the Abdus Salam International Centre for Theoretical Physics (ICTP) in Trieste.[288] While access to and connectivity with a global community of data terminals via the internet is a given today, the switching of ever-increasing data streams across (academic) networks is a challenge that the GÉANT2 Joint Research Programme[289,290] has taken on (Figure F.23). The European Commission and the National Research and Education Networks (NRENs) are funding this first international hybrid research and education network.[291] While still active (until April 2010) the EU project EGEE[292] (Enabling Grids for E-sciencE) managed the GÉANT network. More information on GÉANT2 can be found at EUROPA, the gateway to the European Union[293]

[286] http://cdsagenda5.ictp.trieste.it/askArchive.php?subtalk=1&base=agenda&categ=a10141&id=a10141s2t9/slides.

[287] The main themes of the conference were climate change modelling and adaptation/mitigation policies and the role of e-Infrastructures in climate change studies. Pertinent challenges and ways forward both from the organisational and policy perspective and from the enabling technology vantage point were illustrated. Final programme version: http://users.ictp.it/~smr2238/Program_einfrastructures.pdf.

[288] http://users.ictp.it/~smr2238.

[289] http://www.geant2.net.

[290] http://www.geant2.net/upload/pdf/PUB-06-014_point_to_point_leaflet.pdf.

[291] http://www.geant2.net/upload/pdf/GN2_Topology_Feb_09.pdf.

[292] http://www.eu-egee.org.

[293] http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/08/133&format=HTML&aged=1&language=EN&guiLanguage=en.

The third challenge is an especially important one because the provision of tools for accessing, downloading, visualising and using data sets for new analyses requires communication and contact to data producers and/or data curators.

In an editorial in *Nature*, entitled "Data's shameful neglect",[294] the appeal is made to research funding agencies to "recognise that preservation of and access to digital data are central to their mission, and need to be supported accordingly". Data provenance, giving credit to "data contributors" also with respect to their career opportunities, and decisions on a "data library infrastructure" are critical issues that need to be taken care of and require a dedicated effort at the highest levels. To this end the editorial cites as a progressive example the establishment of the Joint Information Systems Committee by seven UK research councils in 1993, which made data-sharing a priority, and helped to establish a digital curation centre.

The commentary closes by recommending/demanding that "information management" should be a mandatory subject and that "data management should be woven into every course in science, as one of the foundations of knowledge".

Some optimistic correspondence[295] to this editorial appeal ensued, i.e. from members of the natural history research community, propagating the Mammal Networked Information System[296] (Guralnick, Constable, Wieczorek, Moritz and Peterson, 2009).

In Germany, the Alliance of German Science Organisations[297] has created the priority initiative "Digital Information"[298] in 2008 with the focal areas of:

- national licensing
- Open Access
- national hosting strategy
- primary research data
- virtual research environments
- legal frameworks.

[294] http://www.nature.com/nature/journal/v461/n7261/full/461145a.html.
[295] http://www.nature.com/nature/journal/v462/n7269/full/462034a.html.
[296] http://manisnet.org.
[297] Members are: Alexander von Humboldt Foundation, German Academic Exchange Service, German Research Foundation, Fraunhofer Society, Helmholtz Association of German Research Centers, Association of Universities and other higher Education Institutions in Germany, Leibniz Association, Max Planck Society, Germany Science Council.
[298] http://www.wissenschaftsrat.de/download/archiv/Allianz-digitale%20Info_engl.pdf.
[299] http://www.geant2.net/upload/pdf/GN2_Topology_Feb_09.pdf.

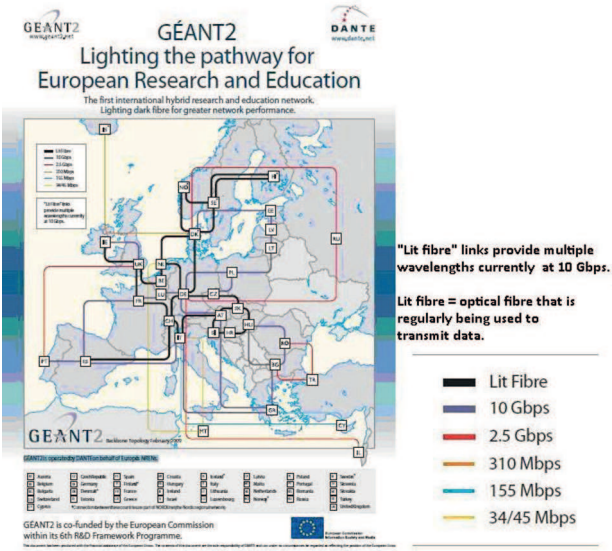

Figure F.23 GÉANT2, the first international hybrid research and education network (modified from source)[299]

Kommission Zukunft der Informationsinfrastruktur The Leibniz Association was mandated in 2009 to develop a concept for the technical information infrastructure in Germany (delivered in April 2011). A commission "Future of information infrastructure" (*Zukunft der Informationsinfrastruktur KII*),[300] consisting of approximately 135 persons from 60 institutions contributed to the concept, which included the topics licensing, hosting and longterm archiving, non-textual material, retro-digitalisation/cultural heritage, virtual research environments, Open Access/electronic publishing, research data and information competency/education. They took the work of the above-mentioned priority initiative "Digital Information" into account.

Furthermore, the DFG on behalf of the Alliance of German Science Organisations also funded a special study on the establishment of a federated strategy on perpetual access and hosting electronic resources for Germany[301] At the European level the High level Expert Group on Scientific Data drew up their terms of reference in 2010.[302]

[300] http://www.wgl.de/?nid=kiikom&nidap=&print=0.
[301] http://www.allianzinitiative.de/fileadmin/hosting_studie_e.pdf.
[302] http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/tor.pdf.

8 Future directions and summary

The data that are being generated in climate or earth system science are already and will be re-used not only by scientists but should continue to be beneficially exploited by other societal groups. The scientific community, politicians and stakeholders concerned with climate change (and adaption to and mitigation of it) are to a varying degree already familiar with the existence of data sets useful for their work. There are mechanisms in place that let them tap successfully into the climate data resource.

In Hamburg, for example, the CSC,[303] specifically former members of the "Service Group Adaption" (SGA), assist participants of the project KLIMZUG[304] by providing them – in cooperation with the German Weather Service – with a common database of regional climate model and climate monitoring data and guidance concerning methods of climate data analysis.[305] Climate indices like precipitation days, snow, frost, ice, summer days, tropical nights, days with strong winds, length of the growing season, hot days, wet days, and diagrams and animations of simulated data are also available.[306] This information is not only valuable economically, but may also help to anticipate direct and indirect effects to human health.[307]

Before KLIMZUG, the BMBF-funded research about ways to deal with climate change in its programme "klimazwei".[308] Cooperative projects were carried out in which business and societal challenges and opportunities that exist due to climate change were addressed in transdisciplinary approaches. The need to manage both mitigation measures, e.g. reduction of greenhouse gas emissions, and adaptation strategies spawned a number of projects whose results were presented in a brochure[309] in 2009. Innovative ways to minimise CO2 and other greenhouse gas emissions in various industrial processing chains were derived and several possible solutions to the heat and water stresses brought about by climate change were suggested for several metropolitan areas. Particularly in the second category (adaptation) effective ways of communicating uncertainties and risks were central to some projects.

[303] http://www.climate-service-center.de.
[304] http://www.klimzug.de/en/211.php, KLIMZUG is presently funded by BMBF in the category "Managing climate change in the regions for the future".
[305] http://mud.dkrz.de/projects-at-md/sg-adaptation/sga/sga-introduction-english/index.html.
[306] http://www.klimazwei.de/Portals/0/SGA_flyer_engl_jul09.pdf.
[307] http://wiki.bildungsserver.de/klimawandel/index.php/Klimawandel_und_Gesundheit.
[308] SGA was created during BMBF programme "klimazwei – research for climate protection and protection from climate impacts", which started in 2006.
[309] http://www.klimazwei.de/Portals/0/klimazwei-Ergebnisbrosch%C3%BCre.pdf.

The CSC's priority is thus to provide decision takers from politics, business, administration and society with pertinent climate information that lets them do an improved service to their customers, which may be, for example, users from forestry, farmers, tourism managers and insurance companies. As a new instrument supporting the CSC's principal mission, the Climate Navigator[310] was launched in July 2011. The information that is accessible through this online platform is provided by more than 30 research organisations/units.

The CSC also contributes to the City of Hamburg's Educational Server (Hamburger Bildungsserver) with the maintenance of a climate wiki[311] which informs about general facts and processes concerning climate (change).

As the effects of climate change vary regionally, the Helmholtz Association has created four regional climate offices in Germany,[312] each focusing on a different region and having slightly different priorities: (1) the North German Climate Office,[313] at the Geesthacht Centre for Materials and Coastal Research, concentrates on changes in storms, storm surges, ocean waves, and coastal climate; (2) the Climate Office for Central Germany in Leipzig "offers information on adaptation strategies and on the impact of climate change on the environment, land use and society" (host: Centre for Environmental Research, UFZ);[314] (3) the South German Climate Office in Karlsruhe scores with its expertise on regional climate modelling and informs about extreme weather events such as heavy precipitation and flooding; and (4) the Climate Office for Polar Regions and Sea Level Rise, hosted by AWI in Bremerhaven, concentrates on relaying exactly this information to the public. All climate offices summarise their information in an online regional climate atlas.[315]

At the European level, ESA's Group on Earth Observations maintains the GeoPortal,[316] an entry point to access remote sensing, geospatial static and in-situ data, information and services. The Global Earth Observation System of Systems[317] (GEOSS) in particular promises to provide decision-support tools to a wide variety of users. Browsing for "climate"[318] shows the following services as being available: early warning, monitoring, analysing, mapping,

[310] http://www.hzg.de/science_and_industrie/klimaberatung/csc_web/012225/index_0012225.html.de, in German.
[311] http://www.klimawiki.org/, in German only.
[312] http://www.klimabuero.de/index_en.html.
[313] http://www.norddeutsches-klimabuero.de.
[314] http://www.mitteldeutsches-klimabuero.de/.
[315] http://www.regionaler-klimaatlas.de.
[316] http://www.geoportal.org/web/guest/geo_home?cache_control=0.
[317] http://www.earthobservations.org/geoss.shtml.
[318] http://www.geoportal.org/web/guest/geo_search_overview?p_p_id=srgPortlet_WAR_geoportal&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_count=4&_srgPortlet_WAR_geoportal_searchType=browse&_srgPortlet_WAR_geoportal_sbaId=4.

assessment, alerting, geospatial web service, data processing and data provision. The services are based on global and regional resources and are intended to help understand and evaluate key indicators of climate change.

A potential interest in climate data exists, however, also outside of the large earth system science network and in new communities recruiting themselves from as-yet-unknown areas of the general public. To raise awareness about the existence of and foster the re-use of high-quality earth system science data sets, the international and interdisciplinary journal *Earth System Science Data* (ESSD)[319] was established in 2008, and is published by Copernicus (Copernicus Publications). Another important objective of this endeavour is to reward "data authors" with the recognition of their peer-reviewed and appropriately described data set as an academic achievement. The chief editors of ESSD, David Carlson[320] and Hans Pfeiffenberger,[321] describe the peer-review process that they envision in their paper introducing this journal,[322] and distinguish between a-priori and a-posteriori quality assurance. The former may already exist in researcher communities that utilise established methods of documentation, validation and explicit quality control levels when generating or processing primary and secondary data. For the most part, this may be assumed for the data discussed in the previous sections of this chapter. Guidance for the review of and a-posteriori quality assessment of potentially valuable data resulting from less tested methods is provided by the journal.[323]

The evaluation of data sets based on a peer review like, for example, the one practiced by ESSD, represents an added value, as the recognition or impact of a data publication is a function of the quality control/assurance, since issuing a DOI for a data set does not render it a publication in a sense that is comparable to a usual (peer-reviewed) scientific publication.

Besides quality stamping data sets, data repositories as well may earn recognition for trustworthy data management and stewardship. The Data Seal of Approval (DSA)[324] may be awarded to repositories holding digital research data if they score highly in the DSA's 16 guidelines[325]. While not being one of the six recipients of this seal, the WDCC may claim adher-

[319] http://www.earth-system-science-data.net.

[320] Science Communication Director for the non-profit geodesy consortium UNAVCO, Boulder, Colorado, USA, http://www.unavco.org.

[321] Head of IT infrastructure at AWI, and speaker of the Helmholtz Association's Open Access working group.

[322] http://www.dlib.org/dlib/january11/pfeiffenberger/01pfeiffenberger.html.

[323] http://www.earth-system-science-data.net/review/ms_evaluation_criteria.html.

[324] http://www.datasealofapproval.org.

[325] http://www.datasealofapproval.org/sites/default/files/DSA_informationfolder_web_0.pdf.

ence to all those guidelines, as it has the technical, institutional and cultural frameworks required to support such open data access.

The founding of the WDC cluster Earth System Research[326] in 2004 implies a commitment to long-term archiving facilities and data libraries with funding to meet the data management infrastructure, expertise and manpower needs. Members are WDC-MARE, WDC-Climate, WDC-RSAT and WDC-TERRA (WDC of the Lithosphere, GeoForschungsZentrum, Potsdam). The WDC cluster guides the peer-review process for scientific data, acts as a publication agent (compare with section 2.7), checks that metadata are complete, that the methods used have been validated and that data values are quality controlled with respect to their precision, sequence and ranges. After that, the independent data entities suitable for publication can be identified and data publication is initiated, i.e. the data entities with persistent identifiers (DOI) are then citable.

DKRZ is also a partner in several national and international projects which further develop a distributed data handling and processing infrastructure. At the conference "Grid Computing: a new tool for Science and Innovation" (VeliLošinj, Croatia, August 2009) Stephan Kindermann presented experiences from the German C3-Grid project and the prototype C3-Grid/EGEE integration and the current data-handling effort for the next Intergovernmental Panel of Climate Change (IPCC) assessment report. The abstract of his lecture "Distributed Data Handling Infrastructures in Climatology and the Grid" has been published on page 20 of the proceedings of this event[327] (compare also with sections 2.2–2.4).

More future developments in climate science e-Infrastructure were voiced at the ICTP conference mentioned in section 7. Kostas Glinos pointed to project METAFOR as an "integrated and coordinated approach to capture the comprehensive metadata requirements of the Climate Research community to create the Common Information Model (CIM) and the tools that will exploit it" and emphasised that the European Grid Initiative (EGI)[328] is the world's largest multiscience grid with over 350 sites including more than 200,000 CPUs and 100 petabytes of storage, which are being used to complete approximately 150,000 jobs per day.

At present, from July to November 2011, the 10th FP7-Infrastructures call is open for applications, and support is intended for the third implementation phase of Partnership for Advanced Computing (PRACE).[329] The roadmap,

[326] http://www.dkrz.de/daten-en/wdcc/wdc-cluster-earth-system-research?set_language=en.
[327] http://indico.cern.ch/getFile.py/access?resId=0&materialId=paper&confId=60021.
[328] http://www.egi.eu.
[329] http://www.prace-project.eu.

services and creation of a European High performance Computing (HPC) ecosystem were outlined by Sergio Bernardi,[330] representing CINECA,[331] the Italian "ConsorzioInteruniversitario" which is one of PRACE's 21 members.[332]

The contribution of project METAFOR to creating a basic information infrastructure in support of climate science was illustrated in a comprehensive way (39 slides)[333] by Bryan Lawrence[334]

Steven Newhouse, another keynote speaker at the ICTP conference and project director of Integrated Sustainable Pan-European Infrastructure for Researchers in Europe (EGI-INSPIRE) described the EGI as indispensible for a sustainable production e-Infrastructure in the support of structured international research.[335]

Venkatramani Balaji[336] showed his and Dean N Williams' presentation[337] on the Earth System Grid Center for Enabling Technologies (ESG-CET) and how data from multiple sources are being integrated (compare with sections 2.3 and 2.4). Taking CMIP5 as an example, Balaji summed up how a number of challenges in integration are being tackled. In a second presentation Balaji showed that in the project "Climate analytics on distributed exascale data archives" (ExArch),[338] investigations would be carried out on how to satisfy the need for policy-scale information from global-scale research and gave an example for removing uncertainty at regional scales (slides 25 and 26).[339]

Before listing the implications for OpenAIRE in the following section, we point to the various grid initiatives that are presently active in climate research such as WissGrid[340] and C3-Grid,[341] and support the call for an organised effort and leadership from funders, societies, journals, educators, individual scientists and the society at large to make measurable progress in

[330] http://cdsagenda5.ictp.trieste.it/askArchive.php?subtalk=1&base=agenda&categ=a10141&id=a10141s2t11/slides.
[331] http://www.cineca.it/en.
[332] http://www.prace-ri.eu/Members.
[333] http://cdsagenda5.ictp.trieste.it/askArchive.php?subtalk=1&base=agenda&categ=a10141&id=a10141s2t25/slides.
[334] Director of Environmental Data Curation at the UK Science and Technology Facilities Council, and head of the British Atmospheric Data Centre.
[335] http://cdsagenda5.ictp.trieste.it/askArchive.php?subtalk=1&base=agenda&categ=a10141&id=a10141s2t10/slides.
[336] Director of the Modeling Systems Group at NOAA/GFDL and Princeton University, USA.
[337] http://cdsagenda5.ictp.it//askArchive.php?categ=a10141&id=a10141s2t16&ifd=37961&down=1&type=slides.
[338] http://proj.badc.rl.ac.uk/exarch.
[339] http://cdsagenda5.ictp.trieste.it/askArchive.php?subtalk=1&base=agenda&categ=a10141&id=a10141s2t27/slides.
[340] http://www.wissgrid.de/index_en.html.
[341] http://www.c3grid.de/index.php?id=44&L=1.

handling and beneficial exploitation of the deluge of data that was made in the introduction to *Science*'s special issue on data handling[342]

9 Implications for OpenAIRE

In the foregoing sections, the infrastructure in climate research, climate data management and organisational and technical implications of some Open Access e-Infrastructures were described. From these and bearing in mind the four objectives stated in the OpenAIRE Annex I[343], the development of the following information services appears to be particularly important:

- climate science data need to be citeable
- the provenance of the growing amount of data needs to be trackable
- climate data sets and publications need to be assigned persistent identifiers
- the quality control procedure which data and publications have undergone needs to be documented and published together with the data
- it should be possible to publish data at a variety of publication levels: peer reviewed, quality checked, unchecked and at some intermediate stages
- support for harvesting the growing number of climate data archives needs to be established
- due to the international nature of climate science cooperation with the Open Access initiatives in other parts of the world needs to be fostered
- consistent e.g. OAIS-based[344] interfaces to long term climate data preservation repositories are needed to enable Open Access.

10 List of figures

[342] http://www.sciencemag.org/content/331/6018/692.short.

[343] Objective 1: Building Support Structures for Researchers in Depositing FP7 Research Publication (Networking); Objective 2: Establishment and Operation of the OpenAIRE e-Infrastructure for Peer-Reviewed Articles (Services); Objective 3: Exploration of and experimentation with Scientific Data Management Services (Research); Objective 4: Sustainability of the OpenAIRE e-Infrastructure and Supporting Structures, Exploitation and Promotion.

[344] Open Archival Information System, http://public.ccsds.org/publications/archive/650x0b1.PDF
http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683

11 List of tables

12 Bibliography

Brady, SR, Sinha, AK, & Gundersen, LC. (eds.) Proceedings from Geoinformatics 2008 – Data to Knowledge. Scientific Investigations Report 2008-5172. US Department of the Interior, US Geological Survey, 2008. http://pubs.usgs.gov/sir/2008/5172.

Curdt, C, & Bareth, G. (eds.) Proceedings from the Data Management Workshop, 9 30 Oct. 2009. Germany, University of Cologne, 2010.

Guralnick, R, Constable, H, Wieczorek, J, Moritz, C, & Peterson, AT. Sharing: lessons from natural history's success story. *Nature* 2009, 462, 34. doi:10.1038/462034a. http://www.nature.com/nature/journal/v462/n7269/full/462034a.html.

Hense, AV, Hense, AN, & Lautenschlager, M. Acquired, analysed, archived – climate data for our future. Presentation given at the DACH Conference in Bonn, 21 September 2010, 29 slides. 2010. http://umwelt.wikidora.com/wikidora/attach/Documents/PU_talk_DACH_20100924.pdf.

Hoeck, H. std-doi Publication of Climate Data at WDCC. Presentation given at the DataCite Summer Meeting in Hanover, 7–8 June 2010. http://datacite.org/slides/DataCite2010_Hoeck.ppt

Metz, B, Davidson, OR, Bosch, PR, Dave, R, & Meyer, LA. (eds.) Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. New York, Cambridge University Press, 2007. http://www.ipcc-wg3.de.

Parry, ML, Canziani, OF, Palutikof, JP, van der Linden, PJ, & Hanson, CE. (eds.) Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. New York, Cambridge University Press, 2007. http://www.ipcc-wg2.gov/publications/AR4/index.html.

Randall, DA, Wood, RA, Bony, S, Colman, R, Fichefet, T, Fyfe, J, Kattsov, V, Pitman, A, Shukla, J, Srinivasan, J, Stouffer, RJ, Sumi, A, & Taylor, KE. Climate models and their evaluation. In Solomon et al., 2007. pp. 589–662. http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-chapter8.pdf.

Science Special online collection: Dealing with data. *Science* 2011, 331, 649.

Solomon, S, Qin, D, Manning, M, Chen, Z, Marquis, M, Averyt, KB, Tignor, M, & Miller, HL. (eds.) Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, New York, Cambridge University Press, 2007. http://www.ipcc.ch/publications_and_data/ar4/wg1/en/contents.html.

Spohr, D. Survey on the research infrastructure. 2010, Sample Questionary (see Appendix 1)

Taylor, KE, Stouffer, RJ, & Meehl, GA. A summary of the CMIP5 experiment design. 2009. http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf (updated 22 January 2011).

Taylor, KE, & Doutriaux, CM. CMIP5 Model Output Requirements: file contents and format, data structure and metadata. Presentation, 7 January 2010. http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_output_metadata_requirements.pdf.

Taylor, KE, Balaji, V, Hankin, S, Juckes, M, Lawrence, B, & Pascoe, S. CMIP5 Data Reference syntax (DRS) and Controlled Vocabularies, version 1.2. 9 March 2011. http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax.pdf.

Toussaint, F, & Lautenschlager, M. The CERA-2 Meta Database and needs for a common information structure. Presentation given in Berlin, 14–15 October 2008. http://colab.mpdl.mpg.de/mediawiki/images/2/20/ESci08_Sem_3_CERA-2_Toussaint.pdf.

Williams, DN, Ananthakrishnan, R, Bernholdt, DE, Bharathi, S, Brown, D, Chen, M, Chervenak, AL, Cinquini, L, Drach, R, Foster, IT, Fox, P, Hankin, S, Henson, VE, Jones, P, Middleton, DE, Schwidder, J, Schweitzer, R, Schuler, R, Shoshani, A Siebenlist, F, Sim, A, Strand, WG, Wilhelmi, N, S& u, M. Data management and analysis for the Earth System Grid. *Journal of Physics: Conference Series* 2008, 125, 1, 01272.

Williams, DN, Doutriaux, CM, Drach, RS, & McCoy, RB. The flexible Climate Data Analysis Tools (CDAT) for multi-model climate simulation data. Proceedings of ICDM Workshops, 2009. pp. 254–261.

13 Appendices

Appendix 1: Survey on the research infrastructure at a specific institute, Sample Questionary

Survey on the research infrastructure at

NAME OF INSTITUT

This survey is part of a study which is being carried out within an EC-funded project. This study aims to derive discipline-specific requirements on research infrastructure.

The participation in this survey is voluntary. Your answers will be treated anonymously and will only be used for the purposes mentioned above. This research underlies the rules and regulations of the data protection act. Your answers are not linked with your person in any way.

After completing the questionnaire in your favourite PDF viewer, please print it as a PS or PDF file and send it to **NAME OF RESEARCHER (E-MAIL, address)**. In case you have any questions regarding the project, the present study or this survey, please contact **NAME OF RESEARCHER**.

We would like to thank you for your participation in this survey.

A. General information

A.1 Personal information

Title Ms/Mrs/Mr, Dr/PhD, Prof. Dr, Jun. Prof.

Last name, first name

Position

Role and duties

A.2 Information on the working group

Name of the group	
Primary research objects	
Size of the group	
People	ca.
PCs	ca.
Servers	ca.
Other research instruments	
Number of cooperations with other groups	
inside **INSTITUTE**	ca.
outside **INSTITUTE**	ca.

B. Data infrastructure

B.1 Data lifecycle

Please use the menus to sketch the different stages that research data typically pass through in your group, and explain them briefly wrt. the time frame, used instruments, metadata and possibly involved people. Please add further stages if necessary.

	Stage	Explanation (e.g. time frame, instruments, people)
1	Choose from: Data collection Processing Enrichment Archiving Reuse Distribution	
2	Choose from: See above	
3	Choose from: See above	
4	Choose from: See above	
5	Choose from: See above	

B.2 Data collection

Which types of primary data are typically collected in a series of experiments concerning a particular object of research, and what are the storage requirements in terms of size?

	Audio data	Choose from: kB range MB range GB range TB range
	Video data	Choose from: See above
	Textual data	Choose from: See above
	Other Data:	Choose from: See above
Do you develop software in order to collect primary data?		Y / N
To what extent do you rely on proprietary software in order to collect primary data?		Choose from: not at all hardly sometimes considerably (almost) entirely

B.3 Data processing and data formats

Which types of secondary data are typically derived on the basis of the primary data, and how are they represented?

Are the data further annotated with metadata?	Y / N
If so, which kinds of information are typically expressed?	
Are there established metadata standards in your field?	Y / N
If so, do you use them?	Choose from: not at all infrequently quite frequently (almost) always
Are there kinds of information for which you have to deviate from standard formats (e.g. because they cannot be represented)?	Y / N
If so, which are these kinds of information?	
Do you develop software to derive secondary data?	Y / N
To what extent do you rely on proprietary formats and software for data representation and processing? Formats Software	 Choose from: not at all hardly sometimes considerably (almost) entirely Choose from: See above
Further explanations (e.g. information on proprietary formats and software):	

B.4 Data management

How are developed software as well as primary and secondary data stored and archived?

	Software	Primary	Secondary
In a repository within the group or institute (e.g. hard disk on a central file server of the group)			
In a repository shared with other groups or institutes (e.g. faculty-wide or university-wide)			
In an external repository			
Mainly on your own storage devices (e.g. on the hard disk of your office PC)			
Would you consider the situation in your group representative for your research discipline?	Y / N		
If not, what is common practice in this respect?			
Are there members of your group which deal with data management issues?	Y / N		
If not, would you consider it desirable or even necessary?	Choose from: yes, would be necessary desirable, but not necessary no, not desirable		
Why?			

B.5 Access to data

Which practices exist in your group in order to ensure the access to produced data?

Which tools support these practices?

B.6 Publication and exchange of research data

Would it be conceivable in your group to make research data available to others (in anonymised form where necessary)?	Software	Primary data	Secondary data
only among your close colleagues also to other research projects to the general public			
If not or only restricted, which disadvantages or boundaries do you see?			
Which incentives for data exchange are mentioned in your group, and are there technologies supporting these incentives?			
Would you consider the situation in your group representative for your research discipline?	Y / N		
If not, what is common practice in this respect?			

Do you make software available to other institutions?	Choose from: yes, as a matter of principle yes, sporadically no, but is conceivable no, and is not conceivable
If so, does this typically include source code?	Choose from: yes, as a matter of principle mostly frequently rarely never
Explanations and reasons	

Which rules exist in your group wrt. use, exclusiveness and time frames for exchanging data?

How is data exchange supported and which tools are used?

Which practices and tools enable data reuse, and which boundaries exist?

C. Literature

C.1 Organisation

Which tools do you use to organise internal and external publications?

Which forms of publication are preferred in your group?
(Check first column and choose from options given in third column)

	Publication as print medium (e.g. article or book)	Choose from: preferred and established preferred, but rather infrequent not preferred
	Electronic publication online or offline	Choose from: See above
	Combination (e.g. book/CD-ROM or proceedings/website)	Choose from: See above

Would you consider the situation in your group representative for your research discipline?	Y / N
If not, what is common practice in this respect?	

Which scientific journals, publishers etc. are preferred in your group/discipline?

C.2 Combination of literature and research data

Are there publishers which enable the exchange of data and/or literature?	Y / N
Is it currently possible to publish data and literature together?	Y / N
If not, would you consider this reasonable or desirable in your discipline?	Y / N
Which developments would be necessary in order to achieve this?	
Which (dis)advantages do you see?	

C.3 Open Access

Is Open Access established in your group?	Chose from: yes a bit no
Would you consider the situation in your group representative for your research discipline?	Y / N
If not, what is common practice in this respect?	
Which technologies and practices support Open Access publications?	

D. Outlook

Which future developments for research infrastructures do you see in your discipline? If possible, please comment on developments wrt. an Open Access infrastructure as well.

Data infrastructure

Literature infrastructure

Thank you very much for your participation!

DATE

Appendix 2: Terms of use for data from WDCC

Conditions for using the WDCC database[345]

1 Creative Commons Licence All data and pages available from WDCC are licensed under a Creative Commons Licence (http://creativecommons.org/licenses/by-nc-sa/2.0/de/deed.en) as far as those conditions are

[345] http://cera-www.dkrz.de/WDCC/ui/docs/TermsOfUse.html.

not in any way modified by the following conditions or by any conditions specific to data or pages.

2 Special data owner conditions You must agree with the special data owner conditions (CERA WWW-Gateway » Browse experiments » Show datasets » Click on dataset name » Distribution). If there is a conflict between the Creative Commons Licence and the special data owner conditions, the latter shall have precedence.

3 Always quote reference Articles, papers or written scientific works of any form, based in whole or in part on data supplied by WDCC, will contain an acknowledgment concerning the supplied data. Always quote reference of the experiment in the citation index when using WDCC data (CERA WWW-Gateway » Browse experiments » Experiment information » CERA UI Compact » Citation.

In addition to your comments and suggestions, we are VERY interested in how you use the data we distribute. We would be most appreciative if you would send us a line or two describing your application of these data to data@dkrz.de.

4 Used freely for research only Data from the projects available on the WDCC server may be used freely for research only.

Non-research use of WDCC data: If you do not feel able to accept the conditions for access to the WDCC research data service, you may still be able to acquire data via WDCC Data Services. Please contact data@dkrz.de.

5 Personal account The WDCC will provide you with a personal account and password. If you copy, distribute, display and perform the data, you may do it only under the conditions identical to this one. You are responsible for maintaining the confidentiality of any password(s) you are given to access the WDCC site and are fully responsible for all activities that occur under your password(s). You agree to notify WDCC immediately of any unauthorised use of your password(s).

6 Audit All downloads from WDC-Climate are audited.

7 Share alike Any person extracting data from this server will accept responsibility for informing all data users of these conditions.

8 Liability/warranty

(1) The data are delivered to the user without a warranty of any kind. The user is aware that the data were generated in keeping with the current state of science and technology.
(2) The WDCC assumes no obligations on the basis of this agreement towards third parties. There is no liability of the WDCC for any harm arising from the delivery and subsequent processing of the data products. The user exempts the WDCC from any liability towards third parties.

The disclaimer under 8 (1) and (2) does not apply if and in so far as the WDCC has acted with gross negligence or with intent.

9 Other provisions

- If the user is a merchant, a legal person under public law or a special asset under public law, Hamburg shall be the place of jurisdiction for all disputes arising from this use agreement.
- Should there be one or more provisions in this agreement that are void, in whole or in part, then this will not affect the validity of the other provisions. The invalid provisions will be replaced retroactively by an arrangement as similar as possible in content and coming closest to fulfilling the purpose of the intended provision.