



Project no. 212196
COCOS

European data policies in view of other data policies worldwide

Martin Köchy*, Annette Freibauer

vTI-AK, Bundesallee 50, 38116 Braunschweig, Germany

*martin.koechy@vti.bund.de

Deliverable type: Report
File name: D1.2_data_policy_report.pdf
Deliverable reference num.: D1.2

Instrument:	Coordination and support action
Thematic Priority:	Earth Observation and assessment tools for sustainable development
Due date of deliverable:	month 12
Submission date:	22 June 2008
Start date of project:	1 May 2008
Duration:	36 months
Deliverable lead contractor:	vTI
Revision:	1.0
Work Package:	WP 1
Document ref number:	D1.2_v1.0

Abstract

Databases, and in some jurisdictions the contained data, are protected by intellectual property regulations. This serves the economic and intellectual interests of their creators. International, EU, and national bodies call for free and open access to scientific data, especially if it is funded by public money, to allow the unfettered advancement of scientific knowledge. Since the intellectual property protection of databases is automatic, it requires a specific release of data and database for free and public use. Our survey shows that a range of policies exist regarding database use. We recommend for new databases of publicly funded data that the data and catalogues of metadata be released into the public domain without restrictions to allow the least obstacles for data analysis and generation of scientific knowledge. In addition we recommend the adoption of international standards for metadata, data set identification, and database interoperability.

Table of Contents

Introduction	2
Intellectual Property Rights with regard to data and databases.....	3
Selected Open Access Declarations	3
Database policies in projects.....	5
Database amalgamation or portals	8
Metadata, links to other databases.....	9
Open Access Licences, Public Domain Dedication.....	9
Data set identificaton.....	10
Recommendation for a policy in COCOS	10
Acknowledgements	10
References.....	11

Introduction

Scientists are both creators and users of databases. Databases are polled for reference data, data are used for analyses of trends and patterns, databases are combined to create new insights. The data necessary for the study of regional topics, the tackling of global problems like climate change can no longer be provided by individual researchers and locally restricted databases but require the shared use of large collections of data provided by many researchers and organizations.

Databases and the contained data may be protected by legal and technical means that limit the electronic or physical access. The protection of data may be necessary to ensure the integrity of the data, to protect the priority of researchers for scientific discoveries based on the data, to charge for access to recover the costs for collecting the data or other not-for-profit activities, or to charge for access and use of the data for profit.

Physically, electronically or legally restricted access to data comes in the way of timely analysis within a larger scientific community to tackle regional to global issues. In the interest of the advancement of scientific research and ultimately for the public good, scientists, national, and international organizations have called

for full and open access to scientific databases. Copyright protection of databases does not prevent the owner to allow open access or to restrict unauthorized commercial use. Since, however, the intellectual property regulations protect databases and eventually data automatically, the free and open use of data and the database must be stated expressively by the data and database owners.

Reviews on European data exchange policies have already been presented in the ALTER-net project (van Daele et al. 2009). More detailed analyses of open data policies can be found in a special issue of the CODATA Data Science Journal (2007-07-29) In this report we provide a short overview of the legal status of intellectual property protection of data and data bases, contrast these with open access demands, and provide an overview of actual data policies in major projects. We end by suggesting a data policy for the COCOS project.

Intellectual Property Rights with regard to data and databases

Databases are protected legally by a copyright according to the Berne Convention for the Protection of Literary and Artistic Works (1886) in conjunction with the World Intellectual Property Organization (WIPO) Copyright Treaty (1996). The Berne convention has been enacted by 164 countries so far, the WIPO Treaty has been enacted by 70 countries so far. The EU and many of its member states have signed but not yet enacted the WIPO treaty.

The copyright is the right of an author (creator) to use, change, copy, and ensure the integrity of the original or copied works and to authorize others to execute certain rights with regard to the work (WIPO 909). In addition, the Berne Convention acknowledges non-transferable, moral rights of individual authors or creators independent from a copyright. These moral rights include the right to claim authority and the right to object to distortions of the author's work (WIPO 909). The copyright and moral rights are automatic and do not require any registration or labeling according to the WIPO Treaty and Berne Convention.

The copyright protects the form of a work but not its content. With regard to databases this is interpreted as that the structure of a database is protected by a copyright but not its content, the data. Data is protected in the EU by a so-called "sui generis" right of the creators of databases. EU directive 96/9/EC protects the database contents if a substantial investment (financial or otherwise) was necessary for the collection of the data and the creator is a EU resident or EU corporation. Similar to the copyright, the sui generis right prohibits the use or copying of the data except for private use in a non-electronic format, for illustration of educational or research material with acknowledgement of the source (Wikipedia 2009). The EU directive does not restrict the sui generis right for the benefit of scientific research. The sui generis right lasts for 15 years. Similar regulations for the protection of the content of databases have been proposed by WIPO but are still being discussed.

Selected Open Access Declarations

Global scientific problems require the international cooperation of scientists and sharing of scientific data in databases. Data and databases have been protected by national and international law to protect the commercial interests of data and database owners. To promote the generation of new scientific understanding to tackle international and global challenges, groups of scientists and organizations

have called for free and unrestricted access to the use of scientific data. These calls include the 'Berlin Declaration' and the 'OECD Principles and Guidelines'.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities – <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

The Berlin declaration (2003) building on the Budapest Open Access Initiative (<http://www.soros.org/openaccess>), the ECHO charter (<http://echo.mpiwg-berlin.mpg.de/home>) and the Bethesda Statement on Open Access Publishing (<http://www.earlham.edu/~peters>) provides a formal definition of open access, and extend the Budapest Open Access Initiative to comprise not only scholarly publications but "scientific knowledge and cultural heritage". These include "original scientific research results, **raw data and metadata** [our emphasis], source materials, digital representations of pictorial and graphical materials and scholarly multimedia material".

1. The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use.
2. A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, inter operability, and long-term archiving.

The Berlin declaration has been signed amongst others by the Royal Swedish Academy of Science, The Swiss National Science Foundation, Chinese Academy of Sciences, National Natural Science Foundation of China, Indian National Science Academy, Fonds National de la Recherche Scientifique (Belgium), Estonian Academy of Science, Deutsche Forschungsgemeinschaft (German Research Foundation), Centre National de la Recherche Scientifique (France), the Open Society Institute (U.S.A.), many universities, scientific organizations, and research institutes worldwide.

OECD Principles and Guidelines for Access to Research Data from Public Funding – <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

At the request of the European Community and the governments of Australia, Austria, Belgium, Canada, China, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Russian Federation, the Slovak Republic, the Republic of South Africa, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States at the meeting of science and technology ministers in 2004 the Organization for Economic Co-operation and Development prepared the OECD Principles and Guidelines for Access to Research Data from Public Funding (2007). The OECD Principles and Guidelines define the term "research data". They state that access to digital **research data from public funding** should be regulated according to thirteen principles including open access, protection of intellectual property, and quality.

The guidelines acknowledge that there may be a conflict between intellectual property and the interest of the scientific community for open access.

Database policies in projects

Existing data and database use policies reflect the balance between the interests of the public and research community to access the data as early as possible after their measurement on the one hand and, on the other hand, the individual researcher or organization that may have invested significant amount of physical resources, time, and money in the measurements and may be deprived of the priority of publication of the results or recovery of investments by free and unrestricted publication.

We compared the data policies of several international scientific projects or data depositories with respect to cost of access, right of copying, conditions for commercial use, conditions for attribution or acknowledgement, conditions for the redistribution of the data or amalgamation with other data, embargo periods for the protection of scientific priority, and other restrictions or conditions. We included in the comparison projects related to COCOS without any attempt to be comprehensive or systematic.

	Projects	main funding	URL to data policy, accessed 2009-06-03
AFOLU data	Agriculture, Forestry, and Other Land Use	EU	http://afoludata.jrc.ec.europa.eu
ALTER-Net	A Long-Term Biodiversity, Ecosystem and Awareness Research Network	EU	http://www.alter-net.info
CarboEurope	Integrated Project CarboEurope-IP	EU	http://www.carboeurope.org
CarboOcean	Assessment of the European Terrestrial Carbon Balance Integrated Project CARBOOCEAN - Marine carbon sources and sinks assessment	EU	http://www.carboocean.org/
CLIVAR	Climate Variability and Predictability	global	http://www.clivar.org
ENSEMBLES	Ensemble prediction system for climate change	EU	http://ensembles-eu.metoffice.com/
EMAN	Ecological Monitoring and Assessment Network	Canada	http://www.eman-rese.ca
EFI	European Forest Institute	EU	http://www.efi.int
ESA	European Space Agency	EU	http://www.esa.int
FAOSTAT	Food and Agriculture Organization of the UN	global	http://faostat.fao.org/
FLUXNET Synthesis	(Coordination regional and global analysis of observations from	global	http://www.fluxdata.org

GEMS	micrometeorological tower sites) Global and regional Earth-system (Atmosphere) Monitoring using Satellite and in-situ data	EU	http://gems.ecmwf.int/
G3OS (GCOS, GOOS, GTOS, GSN, GSNMC, GOSIC)	Global Climate/Terrestrial/Oceanic Observing System (Network), Global Observing System Information Center	global	http://gosic.org/
GCP	Global Carbon Project	global	http://www.globalcarbonproject.org/
GEOMON	Global Earth Observation and Monitoring of the Atmosphere	EU	http://www.geomon.eu/
GEOSS	Group on Earth Observations	global	http://www.earthobservations.org/
GLCF	Global Land Cover Facility	global	http://glcf.umiacs.umd.edu
GMES:geoland	Global Monitoring of Environment and Security; Integrated GMES Project on Land Cover and Vegetation	EU	http://www.landcover.org/
GMES:MERSEA	Global Monitoring of Environment and Security; Marine Environment and Security for the European Area	EU	http://www.mersea.eu.org
IOC IODE	Intergovernmental Oceanographic Commission of UNESCO, International Oceanographic Data and Information Exchange	global	http://www.iode.org
LBA	Large Scale Atmosphere Biosphere Experiment in Amazonia	BR	http://www.lbaeco.org
NASA:GES DISC	National Aeronautics and Space Administration: Goddard Earth Sciences Data and Information Center	US	http://daac.gsfc.nasa.gov
NCDC	National Climatic Data Center	US	http://www.ncdc.noaa.gov/oa/nndc
NEON	National Ecological Observatory Network	US	http://www.neoninc.org/
NERC projects	Natural Environment Research Council	UK	http://www.nerc.ac.uk

ORNL	Oak Ridge National Laboratory	USA	http://daac.ornl.gov
------	-------------------------------	-----	---

Among the compared projects one can distinguish four main policies with regard to the use of the data: 1. redirection to the original data provider, 2. free, unconditional use, 3. free use for scientific and educational purposes with some conditions, 4. only viewing. The data use policies are coupled with different access restrictions to the data: 1. immediate access to the data, 2. access after registration, or 3. access after authorization of the original data provider.

use \ access	immediate	registration	authorization
redirection, policy of provider	CarboOcean EFI, ECMWF (ENSEMBLE, GEMS) GOOS, GTOS GEOMON type II GLCF GMES:MERSEA IOCCP	AFOLU ALTER-net	
unconditional	NASA/GES DISC* NCDC (some data) NEON**		
conditional	CLIVAR ENSEMBLE GEOMON type I NCDC (some data) NERC FAOSTAT	EMAN GEMS CarboEurope GMES:geoland FLUXNET via ORNL LBA via ORNL	FLUXNET-Synthesis ESA
viewing	GCOS		

GCP: no links to data

GEOSS: not yet operational

* no copyright, but acknowledgement of portal requested

** except for reasons of national security, protection of endangered species, and other legal restrictions

A great part of projects include data portals, i.e. they do not provide direct access to the data but a link to an external database with individual data policies each. This avoids complicated harmonization of access and usage rights for the project but leaves the data user with a potentially unmanageable number of data policies. Only one data portal requires the authorization of the original data provider.

Where web sites store data directly, access is immediate or after on-line registration. Only one project requires the authorization of the original data provider.

All sites allowing unconditional use of the data, including commercial use, are funded by the United States government. In contrast, internationally, European, nationally, and some US funded projects allow data usage only conditionally. In most cases this condition is the acknowledgement of the data provider. Another common restriction is the limitation for scientific or educational use. Use of the data sometimes also requires that the project coordinator is provided with copies or references of resulting publications. In addition, IOC, GEMS and CLIVAR stipulate that the publication must be "in the open literature without delay or restriction". It is not specified whether this means that the results must be published in Open-Access journals. Commercial use may be granted with conditions imposed on a case-by-case basis in many projects. In some projects the data providers reserve the right to review and comment the conclusions based on the data. In case of significant intellectual input by the data provider, the provider may demand co-authorship. Most projects have no explicit policy for copying the data or including the data in other databases. Copying the data is presumably thought to be implicit when the data is viewable, but the EU database directive effectively prohibits this kind of use. The EU directive applies to EU-associated databases whenever there is no policy. This is why in the strict legal sense a substantial portion of the complete GCOS data may only be viewed and not copied although the data are freely available. Similarly, most projects have no policy for joining the data with other data. If such policies exist, they may stipulate that the original conditions (and metadata) must accompany the data or that the original source is identifiable or the data separable from the other data. Amalgamation may also require the authorization of the project coordinator or the original providers. The ENSEMBLES project prohibits data amalgamation.

The data is provided without cost for online access. If the user requires the data to be delivered in a different form, most policies stipulate that only shipping and handling may be charged.

Database amalgamation or portals

The European Commission launched the Shared Environment Information System (SEIS) in 2008 to promote open access to fulfill commitments to its own INSPIRE directive and the Global Monitoring for Environment and Security (GMES) initiative.

The many different conditions of data use prevent an effective amalgamation of databases. One solution to this obstacle is to refrain from an amalgamation of the data and to provide a data portal with links to individual databases. This would greatly slow the access to data and require the user to keep track of which data need to be cited or acknowledged in which form. This procedure would be no advance towards the project's aim of a 'one-stop-shop' for data. Alternatively, the existing licenses for individual data users could be replaced by a common license used by all data providers who agree to include their data in an amalgamated database.

As soon as a database and its content are created in the EU, they are automatically protected by restrictive copyright and sui generis rights. This includes metadata

catalogues. Therefore, a potential user needs the explicit permission or license of the data owner for each kind of use. The license should define

- the ownership of the data and the database
- the categories of usage (e.g. research, education, commercial)
- the rights to copy, change, aggregate, evaluate, amalgamate, and redistribute the data
- costs for access and usage
- obligations of acknowledgements
- embargo periods
- other restrictions regarding national security, rare species, personal data, entrepreneurial secrets.

In addition, the license must not be in conflict with national laws.

Metadata, links to other databases

The World Data System is a central portal and depository for data from national and international monitoring programs administrated by the International Council for Science in Paris, France. As of 2009 the WDS comprises the formerly thematically specialized World Data Centers, including the WDC for Biodiversity and Ecology at the Oak Ridge National Laboratory. Data hosted by the WDS are available for free via online access for non-commercial purposes. Submission of datasets or links to the portal for hosting or linking requires that metadata are supplied according the WDS standard. Metadata are collected according to the principles of the Directory Interchange Format (DIF, <http://gcmd.nasa.gov/User/difguide>) and the Service Entry Resource Format (SERF, <http://gcmd.nasa.gov/User/serfguide>). These formats are in line with the internationally agreed minimal metadata, the Dublin Core Set (<http://dublincore.org>).

The Open Archives Initiative (<http://www.openarchives.org>) is striving for the standardization of metadata access so that data sets may be aggregated with minimal effort.

Open Access Licences, Public Domain Dedication

Existing licenses that address these issues and follow the principles of Open Access include the Creative Commons license. The Creative Commons licenses are derived from copyright. Therefore, they do not regulate the use of the database contents protected by the EU sui generis right. Other unwelcome consequences related to the use of Creative Commons licenses on scientific data include the possible obligation of acknowledging hundreds of data providers after synthesizing data from many databases and the incompatibility of some license types when data are amalgamated. This has led to the "Protocol for Implementing Open Access Data" (Science Commons 2009). To achieve the intended effect of unrestricted access and the least hindrance to reuse of the data it was found to be necessary that the data owner waives all legal rights and relies on social and professional norms for acknowledgement and putting derived works in the public domain. One implementation of this protocol is the Public Domain Dedication and Licence

(PDDL; Open Data Commons 2009). A more restrictive licence is the Open Database Licence (ODbL) reserves some rights of the owner and does not extend to the content of the database. The data itself may be licensed according to the Database Contents Licence (DCL). Both ODbL and DCL are still under development.

Data set identificaton

Data released into the public domain may be changed, merged, and relabeled. For the control of scientific works, however, it is imperative that the data used for drawing conclusions is identifiable and traceable. The identification can be realized through a Universal Numerical Fingerprint (UNF; Altman and King 2007, Dataverse Network Project 2009) and a uniform resource identifier like the Digital Object Identifier (DOI, International DOI Foundation 2009) or Handles (Corporation for National Research Initiatives 2009). The UNF ensures the integrity of the data. It is promoted by the Dataverse Network Project whose partners include amongst others the United States National Archives, The Norwegian Social Science Data Services, the Public Knowledge Project, and United Nations University Merit. The DOI identifies individual data sets. It is commonly known in the scientific world for its reference to individual published articles. This policy is used by the Oak Ridge National Laboratory and envisioned by the NEON project. The World Data Centers recommend the use of Scientific and Technical Data DOIs (STD-DOI 2009), which are globally administered by the German National Library of Science and Technology.

Recommendation for a policy in COCOS

The GEOSS project (http://www.earthobservations.org/geoss_dsp.shtml) is currently (2009) revising the draft guidelines for data sharing principles based on the comments of national and international organizations. In the interest of efficient international cooperation and in agreement with the suggestions of the ALTER-Net project (van Daele et al. 2009) we suggest that new databases containing publicly funded data

- endorse the GEOSS guidelines (to be adopted formally in November 2010),
- implement it by using the Public Domain Dedication and Licence,
- apply uniform resource identifiers and universal numerical fingerprints to the data, and
- comply to the World Data Center System metadata standards for inclusion of the database in other networks and
- implement interoperability standards suggested by the Open Archive initiative
- provide free and open access even to the metadata

Acknowledgements

This work was supported by COCOS (Coordination Action Carbon Observing Systems, grant 212196) and ICOS (Integrated Carbon Observation System, grant 211574) funded by the 7. Framework Programme of the European Commission.

References

- Altman, M., and G. King. 2007. A proposed standard for the scholarly citation of quantitative data. *D-Lib Magazine* 13. doi: 10.1045/march2007-altman
- Corporation for National Research Initiatives. 2009. The Handle System. <http://www.handle.net/>, viewed 2009-06-04.
- van Daele, T., M. Kertez, K. Schleidt, and J. Peterseil. 2009. Data exchange policy on a distributed information system. Report: WPI6_2009-02. ALTER-Net, Project no. GOCE-CT-2003-505298.
- Dataverse Network Project. 2009. The standard. <http://thedata.org/citation/standard>, viewed 2009-06-04.
- International DOI Foundation. 2009. Welcome to the DOI® System. <http://www.doi.org/>, viewed 2009-06-04.
- Open Data Commons. 2009. ODC Public Domain Dedication and Licence (PDDL). <http://www.opendatacommons.org/licenses/pddl/1.0/>,
- Science Commons. 2009. Protocol for implementing open access data. <http://sciencecommons.org/projects/publishing/open-access-data-protocol/>, viewed 2009-06-04.
- STD-DOI. 2009. Publication and Citation of Scientific Primary Data. <http://www.std-doi.de/>, viewed 2009-06-04.
- Wikipedia. 2009. Database directive. http://en.wikipedia.org/wiki/Directive_on_the_legal_protection_of_databases, viewed 2009-05-19.