ExpeER

Distributed Infrastructure for EXPErimentation in Ecosystem Research

Grant Agreement Number: 262060

SEVENTH FRAMEWORK PROGRAMME

Capacities

Integrating activities: Networks of Research Infrastructures (RIs)

Theme: Environment and Earth Sciences

DELIVERABLE D3.1

Deliverable title:Data Policy and Guidelines

Abstract:

The report summarises the outcome of the data management and data policy questionnaire distributed in the ExpeER community and provides a proposal for the data sharing policy within ExpeER. The questionnaire presents the current state of data management in the ExpeER community as well as a vision to promote a data policy that will enable a more effective and sustainable environmental science large-scale infrastructure.

Due date of deliverable: M12 | Actual submission date: M14

Start date of the project: December 1st, 2010 | Duration: 48 months

Organisation name of lead contractor: UFZ

Contributors: Mark Frenzel, Johannes Peterseil, Nicolas Bertrand

Revision N°: V 7

Dissemination level:

PU Public (must be available on the website)	**[X]**
PP Restricted to other programme participants (including the Commission Services)	**[]**
RE Restricted to a group specified by the consortium (including the Commission Services) (precise to whom it should be addressed)	**[ ]**
CO Confidential, only for members of the consortium(including the Commission Services)	**[ ]**

Table of Content

Glossary

CSW –The OGC Catalogue Service defines common interfaces to discover, browse, and query metadata about data, services, and other potential resources (source: Wikipedia)

Data management - referring to all methods of storing, managing and archiving data being digital or analogue

Dataset - describing a concrete dataset of an observation/experiment or a sum of observations/experiments (e.g. vegetation releveés from permanent plots, soil temperature measurements from a plot, etc.)

OGC - Open Geospatial Consortium, an international voluntary consensus standards organization (source: Wikipedia)

Site - referring to the observation

SPARQL -SPARQL (pronounced "sparkle", a recursive acronym for SPARQL Protocol and RDF Query Language) is an RDF query language, that is, a query language for databases, able to retrieve and manipulate data stored in Resource Description Framework format (source: Wikipedia)

SOS - Sensor Observation Service (SOS) is a web service collecting real time sensor data and sensor data time series (source: Wikipedia)

SWE – OGC's Sensor Web Enablement (SWE) framework defines a suite of web service interfaces and communication protocols abstracting from the heterogeneity of sensor (network) communication (source: Wikipedia)

WCS –The Open Geospatial Consortium Web Coverage Service Interface Standard (WCS) provides an interface allowing requests for geographical coverage across the web using platform-independent calls (source: Wikipedia)

WFS - The Open Geospatial Consortium Web Feature Service Interface Standard (WFS) provides an interface allowing requests for geographical features across the web using platform-independent calls (source: Wikipedia)

WMS - Web Map Service (WMS) is a standard protocol for serving geo-referenced map images over the Internet that are generated by a map server using data from a GIS database (source: Wikipedia)

1 Executive summary

Additionally to data sharing policy issues this deliverable covers data management practices and expectations to data management of the ExpeER partners which was gathered by means of a questionnaire. The reason to include this aspect was the expected high variability and heterogeneity in respect to data management and data provision due to the history of the sites.

A common data sharing policy for ExpeER and its users (especially regarding the Transnational Access(TA) users) is proposed based on the outcomes of a questionnaire filled by ExpeER partners and related frameworks of other projects, networks and institutions. The data policy is mainly addressing the data generated within ExpeER. Most of the data will be related to TA, which covers data of the TA projects and data from TA sites they are building on.

2 Introduction

ExpeER brings together sites from the experimental network as well as from the European long term monitoring network (LTER-Europe) to form a new platform of highly instrumented sites in Europe. These new kind of integrated infrastructures build on the existing sets.

Already finished and ongoing projects at European scale relevant to ICT issues (e.g. LifeWatch, SANY, etc.) or thematic projects (e.g. EBONE, TERENO, etc.) provide good examples and best practises in data management related to ecosystem research. In the field of ecosystem modelling, monitoring and research projects like ExpeER or EnvEurope address this topic.

To face the challenge of data management in ExpeER a basic survey within the project partners on the current situation in data management and data policy was conducted. In addition, the expectations from the consortium regarding to the products of WP3 were asked. This questionnaire was the first step in the analysis of the current environment of data management options.

3 Method

The task has been tackled by using a (A) questionnaire which was sent to the ExpeER partners and (B) making use of data sharing policies used in other projects dealing with similar topics.

3.1 Questionnaire

The questionnaire aimed to collect the current situation on data management and data sharing policies within the infrastructures as well as the expectations towards the project result. The questionnaire was developed together with the Life+ project EnvEurope in order to get comparable information across the different thematic domains. All analysis and work was done in the respective projects. The questionnaire is attached as Annex to the internal report. It was separated in six thematic sections:

1. ReadMeFirst – gives a general introduction to the questionaire as well as the aknowledgement of related projects.

2. Data Management - deals with questions about the current Data management. This includes questions about the format as well as the tools and software used to manage the data.
3. Data Sharing Policy – deals with questions about the Data Sharing Policy. The questions aimed to get an overview how data sharing policy is implemented and which share and cost models are applied at the different sites.
4. Data Access – deals with questions about the Data Access and Distribution. The questions aimed to get an overview how data are technically shared and which were the main target user groups for the data.
5. Expectations – deals with questions about requirements. These questions aim to get an overview about the expectations and requirements from infrastructure coordinators to the results of the work package WP3 Data Management.
6. General Comments – allowed general comments about the questionnaire in plain text format.

In the questionnaire the questions were numbered and short explanations as well as examples for each question were given. This aimed to enhance the usability. Either single answers or multiple answers were possible.

Questionnaire adapted from EnvEurope DM Survey

1. QUESTIONS REGARDING THE MANAGEMENT OF DATA RELEVANT TO EXPEER FACILITIES

EXPEER Site: *DOK trial* Contact: *Paul Mäder*

NOTE: Ask for help of colleagues (e.g. data managers) in case you cannot answer the questions!

	Select from list	Comments
Are your datasets stored in digital format	yes	basic data such as soil and plant analyses

If yes - your data are stored in digital format		
In which format are the data stored? Please also specifiy the software used for the data management.	**Select from list**	**Software, please specify**
Database (relational) *e.g. Oracle, Access, etc.*		
XML Database (XML repository)		
Structured files (spreadsheets) *e.g. Excel*		export as excel sheet
Unstructured file *e.g. unstructured text file*		
Spatial data file (GIS) *e.g. shape file*		
Spatial database (geodatabase) *e.g. PostGIS, Personal GeoDatabase, ArcSDE*		
Other	yes	Widas Software (Agroscope software)
Comments		

	Select from list	Number of places, please specify
Where are your data stored	Central	Agroscope ART Reckenholz Tänikon
Comments		

Figure 1 Design of the questionnaire on data management and data policy

The questionnaires were collected centrally and transferred to a Microsoft Access Database to do the further analysis. During the import routine a consistency check of the answer was made and missing answers were inserted with “no” or “N/A”. The resulting database will be provided in the internal area of the ExpeER web page and is part of the internal report.

The analysis is based on the counts per answer. For questions with multiple options only positive answers (=yes) were used. For the graphs percentage values were calculated.

3.1.1 Recipients and feedback

The questionnaire was distributed to the infrastructure coordinators. Table 1 gives an overview on the list of infrastructures and the response to the questionnaire until 6 February 2012. The results are based on n=26 infrastructures who participated in the questionnaire.

Table 1 List of ExpeER Infrastructures. Column "Done" refers to feedback of infrastructures to the questionnaire ((X): response after evaluation of the questionnaires; - no response).

DONE	SITE	Country	CONTACT1
X	Achenkirch	Austria	Barbara KITZLER
X	Apelsvoll	Norway	Annbjørg Øverli Kristoffersen
X	Beano	Italy	Giorgio ALBERTI
(X)	Biodiversity exploratories	Germany	Markus FISCHER
X	Biogeochemistry Lab	France	Cornelia RUMPEL
X	Braila Islands	Romania	Sergiu Cristofor
X	DOK Trial	Switzerland	Paul MÄDER
X	Donana	Spain	Ricardo DIAZ-DELGADO
-	Ecosylve	France	Denis LOUSTAU
X	Fruska gora	Serbia	Ante VUJIC
X	Hartz	Germany	Steffen ZACHARIAS
X	Hesse	France	Bernard LONGDOZ
X	Höglwald Forest	Germany	Rainer Gasche
X	Hyytiala	Finland	Jaana Bäck
X	JE Experiment	Germany	Anne EBELING
X	Klausenleopoldsdorf	Austria	Barbara KITZLER
X	Long-Term Experiments	UK	Andy MACDONALD
(X)	Lusignan	France	Abad CHABBI
(X)	Molecular Ecology Lab	Italy	Rita BARALDI
X	Montpellier Ecotron	France	Jacques ROY
X	Moor House	UK	Don MONTEITH
-	Negev	Israel	Moshe SHACHAK
X	Puechabon	France	Jean-Marc Ourcival
X	Roma-Lecceto	Italy	Francesco VACCARI

DONE	SITE	Country	CONTACT1
X	Seehornwald	Switzerland	Marcus SCHAUB
X	Silwood Park Ecotron	UK	Alexandru MILCU
X	Tatra Windstorm	Slovakia	Peter FLEISCHER
X	TERENO	Germany	Heye BOGENA
X	TF-LTEP (Tetto Frati)	Italy	Laura ZAVATTARO
X	Tolfa-Allumiere site	Italy	Franco MIGLIETTA
X	Upper Severn	UK	Inma Robinson
X	Whim	UK	Lucy SHEPPARD
X	Zöbelboden	Austria	Thomas DIRNBOECK

4 Results

4.1 Questionnaire: Current situation in ExpeER

4.1.1 Data management

Nearly all data are stored in digital format. 92% of the infrastructures answered “yes”. Only for some old data non-digital formats are still used. Figure 2 gives an overview on the formats and tools used to store the data. The main data storing formats for observational data are relational databases and Microsoft Excel files. The main software packages used for relational databases are commercial products like Microsoft Access (n=5) and Oracle (n=4), but also a high share of open source products like MySQL (n=5) and PostgreSQL (n=4). For structured files mainly Microsoft Excel (n=20) and ASCII text files (n=4) are used.

Spatial data are managed either as spatial data files or stored in geo-databases. The main software used are the ESRI products like ArcGIS and ArcView. As geo-databases either the ESRI Personal Geodatabase (n=2), PostGIS (n=2) or ArcSDE (n=1) are used.

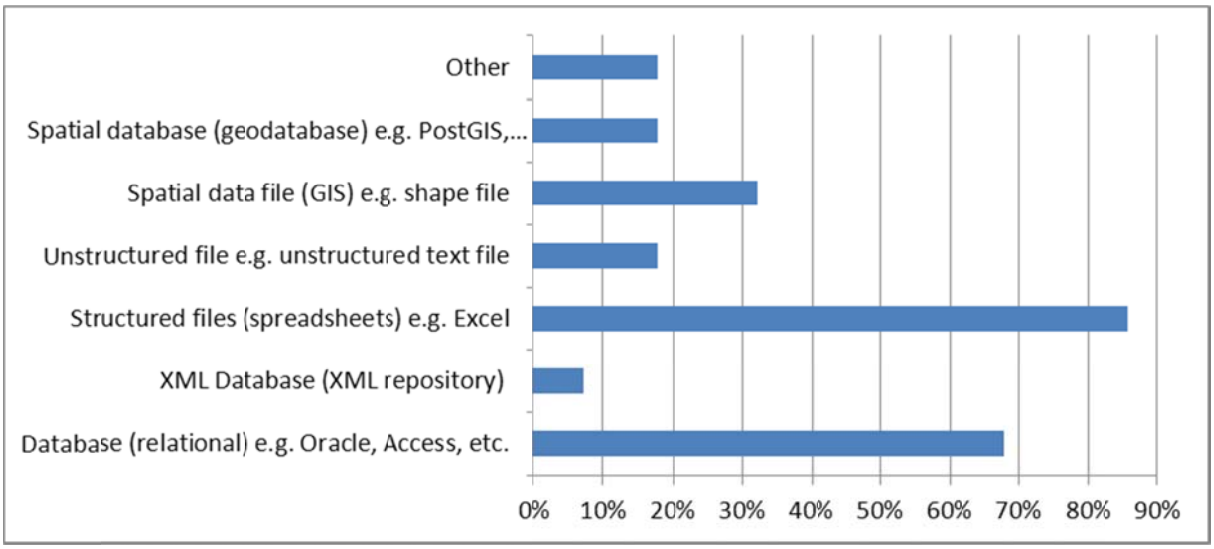

Figure 2 Use of digital or standardised data formats. Bars indicate the proportion of data formats used by ExpeER partners (multiple answers possible).

Despite the wide distribution of databases, nearly 20% of the infrastructures replied that data are provided as unstructured files, which is a challenge for data integration and harmonisation due to the fragmented data storage. These are mainly ASCII files or raw data from loggers.

Other file formats are proprietary binary or ASCII formats of measurement devices. In total nearly 30% of the infrastructures are reporting this file format. Figure 3 gives an overview about the categories. About 15% of the infrastructures provide data still in paper format. These are mainly historic data. The attempts to digitise these data sources were not investigated in the questionnaire.

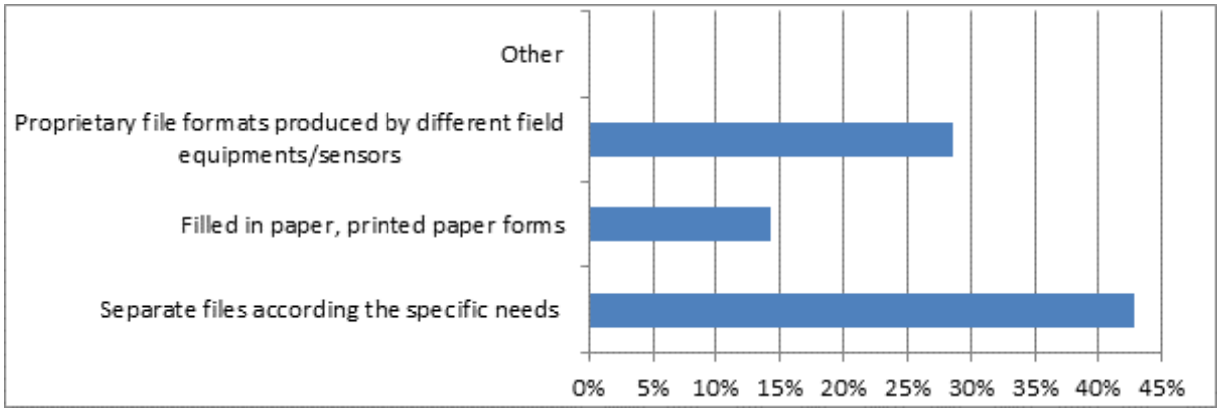

Figure 3 Non-digital or non-standardised data formats used

The distribution of data storage places reveals a large amount of centralised data management (figure 4). Data management is done either in a central manner or the data sources are distributed within one organisation. Only about 10% of infrastructures indicated distributed data management across multiple institutions. The number of places for distributed data management ranges from 2 to 4 showing a low degree of fragmentation.

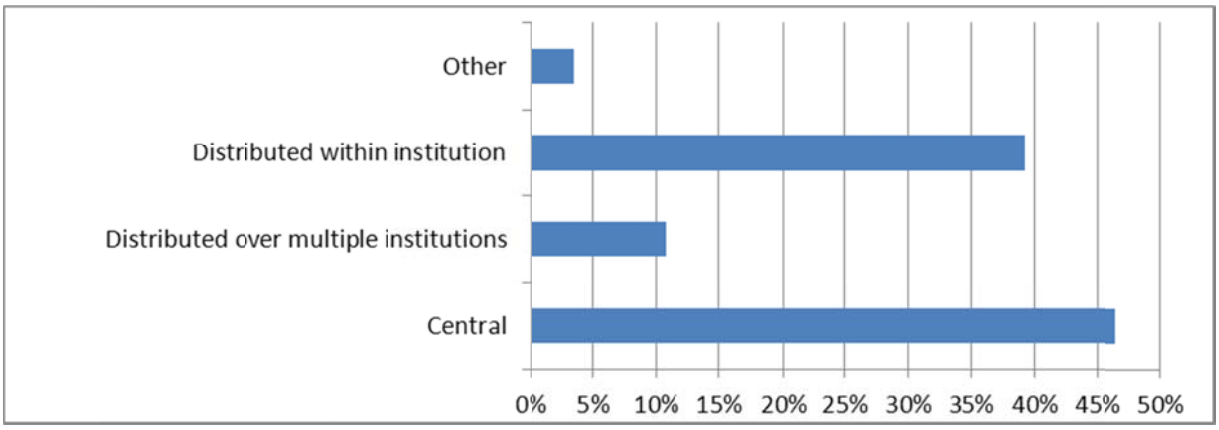

Figure 4 Places of data storage

Data management is mainly done within the scientific departments, either by the scientific staff itself or by technical staff. Nearly one third of the infrastructures replied that a separate data management / IT department is dealing with this subject (figure 5).

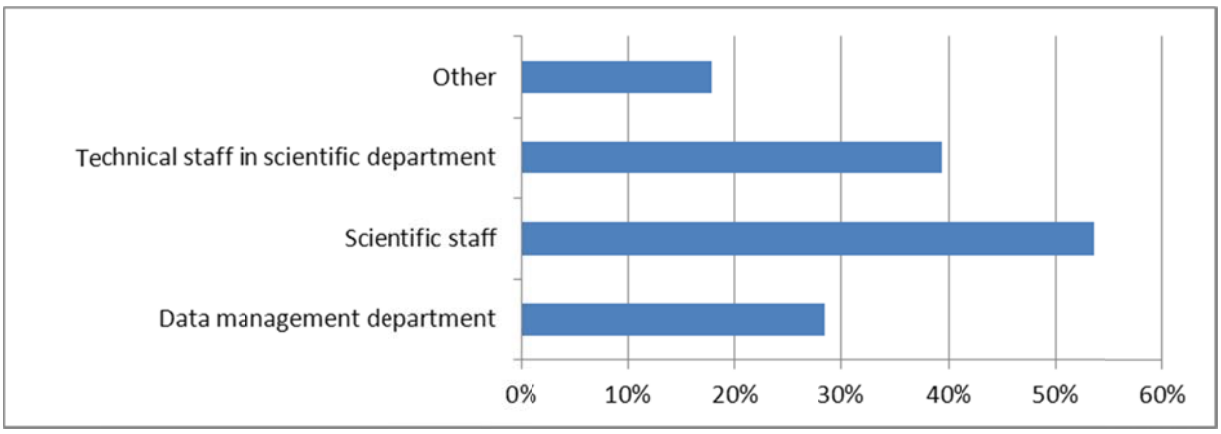

Figure 5 Staff for Data Management

4.1.2 Data sharing policy

The section on data sharing policy tried to collect information from Expeer partners about their expectations and willingness to share data, the licensing of data and the cost model for the data sharing. The following questions related to data sharing policy have been asked:

a) Conditions of data availability for different user groups
b) Foreseen issues of the ExpeER site data with regards to data dissemination and re-use
c) Expeer institution's provision in terms of data management planning, and data curation
d) Expeer institution's expectations with regards to availability of the data generated by Transnational Access users at your site
e) Expectations about reimbursement for use of site-related data gained outside ExpeER, but related/useful for Transnational Access Projects
f) Main user groups of the data (e.g. universities, municipalities, etc.)

g) Procedure of requesting data: online availability (data portal / data service), to be requested by direct contact
h) Existence of "sensitive" data (e.g. data about rare protected species, personal data, etc), which can be shared only under certain circumstances

The results are shown in the following sub-chapters.

4.1.2.1 Data policy model: Availability of data

The questionnaire provided the following data license categories:

- "Free" access means that the data can be used by everybody under specified terms of use (e.g. notification) but no major restrictions were applied for the use.
- "Free upon request" means that the terms of use are negotiated case by case, but the data are in principal free to use.
- "Restricted" means that the use of the data is only for either a restricted group or a restricted set of purposes; the terms of use are specified.
- "No access" means that the data are not free to use.
- "Other": If other terms of use are existing it was possible also to specify this.

The main data sharing model is "free upon request". Here on a case wise evaluation of the situation the transfer of the data is negotiated. Overall, not much difference between the user groups can be seen. An overview of the results is given in Figure 6.

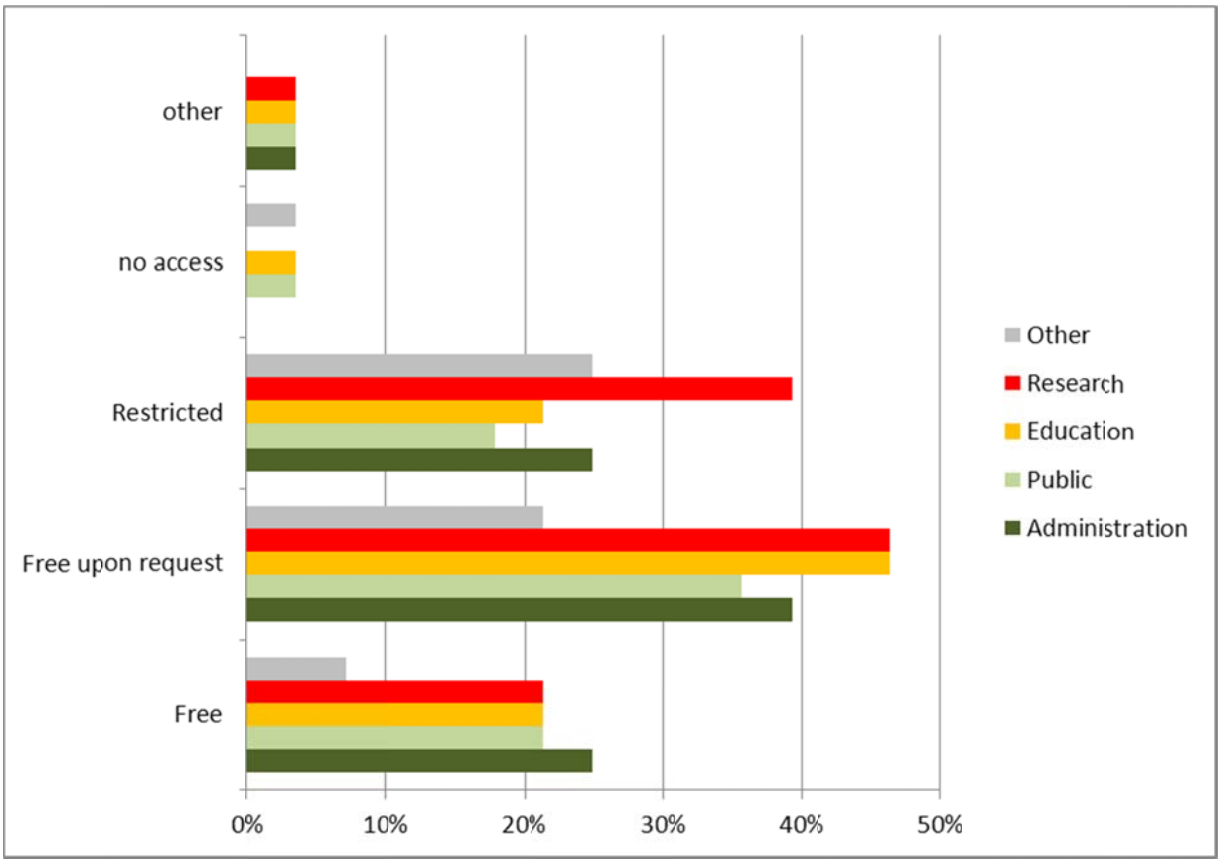

Figure 6 Data policy model for sharing data

The overall picture shows that most of the data are available "upon request" or "restricted", which reflects the common practice in science. The data sharing model "free" shows a slightly higher value for the user group administration than for the others. But the values are very similar and differences are not pronounced. "No access" and "other" not specified access models are about 4%, representing just a single infrastructure. The questionnaire shows that no clear data access policy is implemented across the infrastructures. Usually each infrastructure has its own model to share the data with the main user groups.

Further questions tried to specify the data policy model in the ExpeER context. Actually the specification of the data which should be available in ExpeER is still on going. To overcome IPR issues within ongoing projects the data policy model "free upon request" seems to be appropriate. Sharing data within cooperation – even if the data and/or results are not published in a peer-reviewed journal – seems to be possible for most of the infrastructures. Summarising the comments, three different data policy strategies can be defined:

a) No access to data (e.g. for security reasons)
b) Access to data upon request (to be decided by data owner; e.g. in case of ongoing projects or PhD thesis)
c) Free access to published data (like e.g. meteorological data provided on web pages)

Regarding the conditions for sharing the data several options were mentioned (two thirds of the infrastructures gave feedback). The comments indicate either case by case negotiations or specified terms of use:

a) Institutional agreements, e.g. Memorandum of Cooperation, which apply only to the signing parties and allow data exchange even before finalisation of projects
b) General terms of use, e.g. Data policy, which apply to all users of the data in the same way.
c) Request and acknowledgement, which apply for the personal level.

Data release usually follows two ways: a) an automatic release of data (under specified terms of use) after a certain period after publication or finalisation of the project or b) a negotiation based data exchange (based on negotiated terms of use) even within the runtime of a project.

Acknowledgement of data used in a publication or co-authorship was also mentioned as a condition for data sharing and exchange.

Provisions for data management, data curation and data policy are mainly defined on a network or institutional level (e.g. ECN, TERENO; answered by about half of infrastructures).

This result could be the basis for further discussions on the data sharing and access policy within the ExpeER project.

4.1.2.2 Data cost model – reimbursement of costs

The cost model for data sharing allowed choices between

- "no cost": no additional costs are charged for the data sharing
- "data manipulation" costs: only costs for the data manipulation, e.g. query time, transfer time, extract time, etc., were charged in the data sharing process, but no costs for the data creation
- "data creation" costs: costs for the data creation are charged

In addition it was possible to list other cost models if necessary.

The main cost model for sharing the data is either "no costs" or charging "data manipulation costs". This question resulted in a low response from the infrastructures. Figure 7 provides an overview of the results.

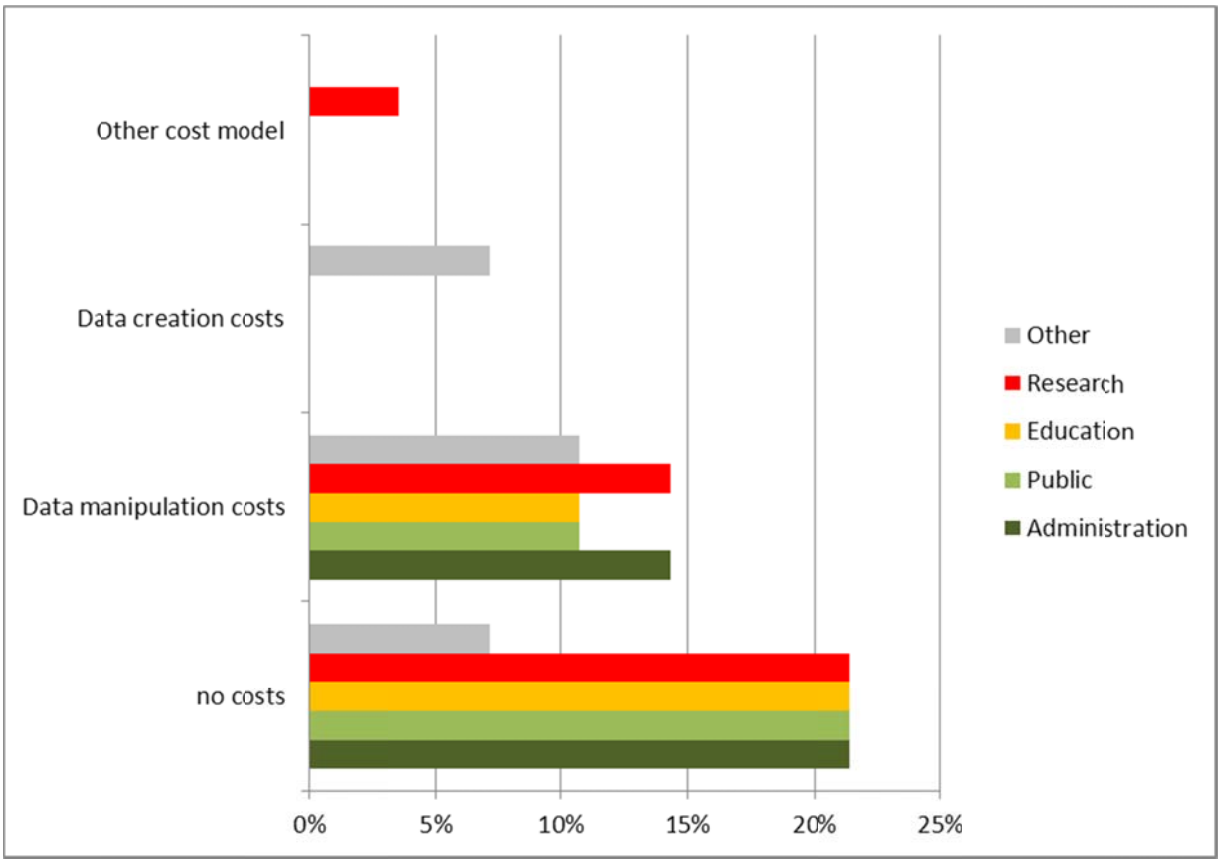

Figure 7 Cost model for sharing data

There is no real difference between the user groups. Research, education, public and administration are treated in the same way.

In total the replies showed that the main data cost model for all user groups was the model "free upon request" which means a case to case negotiation. This hinders an automatic seamless access to data on a "data market square", as negotiated terms of access for bigger user groups are needed. And these agreed terms need to be accepted in the course of the data download. The main cost model, except for commercial use of the data, is the "data manipulation cost model". This would be compliant to the INSPIRE directive which allows to charge costs for the manipulation of the data.

4.1.2.3 Main user groups

The main user groups for data resulting from ExpeER related infrastructures are researchers and people using data for educational purposes. The results are shown in Figure 9. As the main user groups for the category "research", universities and other research institutes were named. For education mainly university related Master PhD courses were listed in the comments.

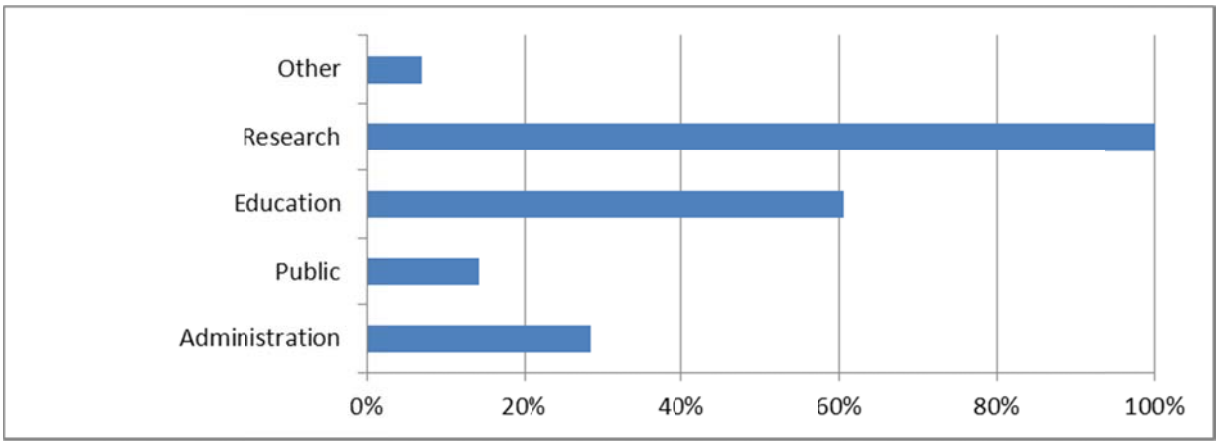

Figure 8 Main user groups

The public and administration are seen as user groups of lower importance. Local, regional and national administration was named for "Administration" and media and interested public for "Public".

The category "Other" comprises the general scientific community as well as NGOs which are not further specified.

4.1.2.4 Sensitive data

The existence of "sensitive data" which can normally not be distributed without special considerations is an issue: about 43% of the ExpeER infrastructures are holding sensitive data. Sensitive data are regarded as

a) Data which can not be shared because of national regulations, e.g. personal data
b) Locations which can not be shared because of forseen threats, e.g. nature protection
c) Data which can not be shared, because of pending publications

The ExpeER data policy has to deal with this aspect.

4.1.3 Data access

The questions about the data access covered the technical as well as the organisational aspect as well. The questionnaire was divided into two sections – one about the services provided and one about the user groups served.

About 38% of infrastructures provide a data portal on the institutions or network web site. Examples for these are:

- European Eddy Fluxes databases cluster: http://www.europe-fluxdata.eu/newtcdc2/IMECC-TCDC_home/login.aspx
- SOERE FORET: http://www.gip-ecofor.org/f-ore-t/(under construction)
- ECN Data Portal - http://data.ecn.ac.uk/
- CEH Information Gateway - https://gateway.ceh.ac.uk/
- e-RA: the electronic Rothamsted Archive:http://www.era.rothamsted.ac.uk/
- Canopy flux data: http://www.europe-fluxdata.eu/newtcdc2/p_Home.aspx

- TERENO: http://teodoor.icg.kfa-juelich.de/tereno-online-portal-folder/Tereno%20Online%20Data%20Portal
- LTER-Europe (eMORIS): https://secure.umweltbundesamt.at/eMORIS/
- Estación Biológica de Doñana: http://icts.ebd.csic.es

In addition about one fourth of the infrastructures also provide metadata and/or data services. Figure 9 gives an overview on the results. About 15% provide a metadata service (CSW) for their data. GeoNetwork is the main application used in the consortium.

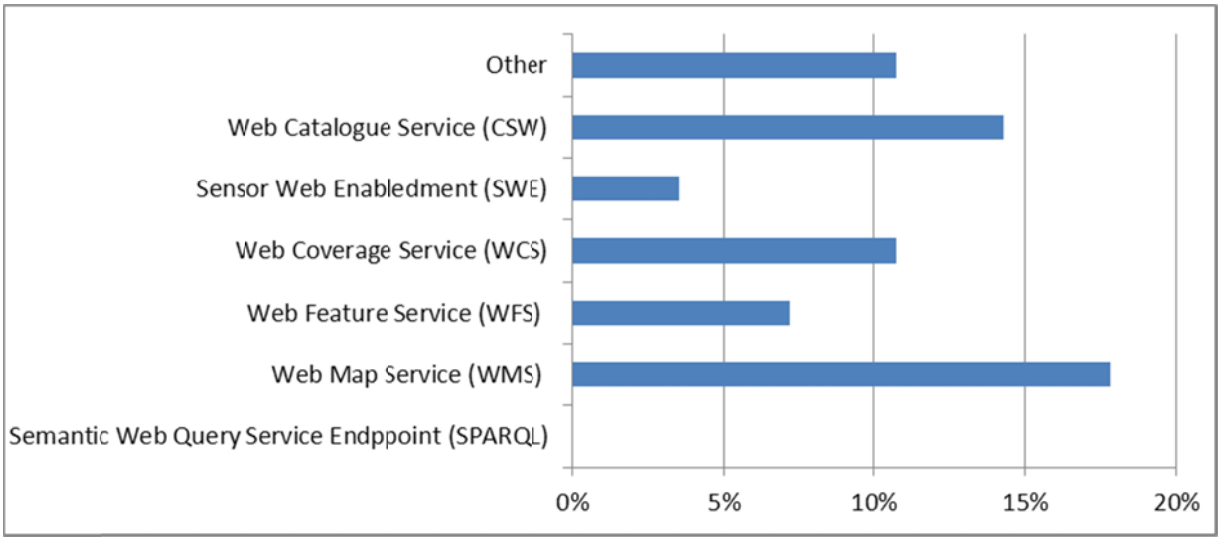

Figure 9 Metadata and Data Services

For OGC WMS, WFS, WCS mainly ArcGIS-Server (n=7) and GeoServer (n=3) are used. In addition MiraMon and MapServer are listed. Sensor Web Enablement (SWE) is realised by using the 52°North software (n=3).

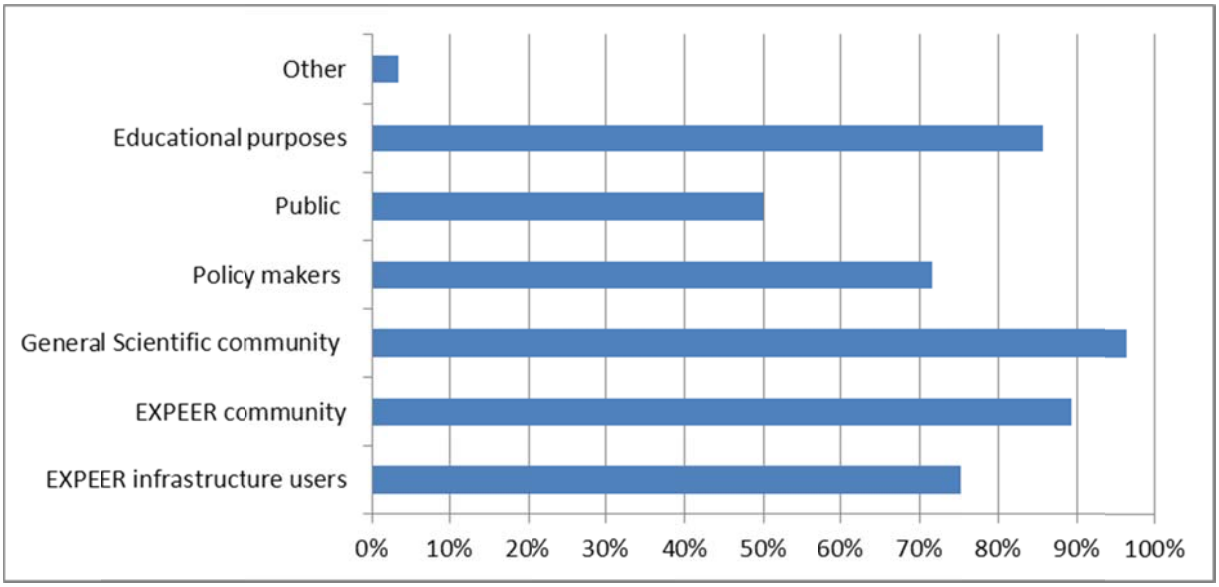

Figure 90 Main user groups

The main way to access the data is offline using mail or telephone. This reflects also the data policy model which makes the decision on “free upon request” easy, because every use of the data needs to be discussed and can be negotiated between the data owner and the data user. About one third of the infrastructures are also using online tools to distribute and share data. These are mainly data portals. The results are shown in Figure 11.

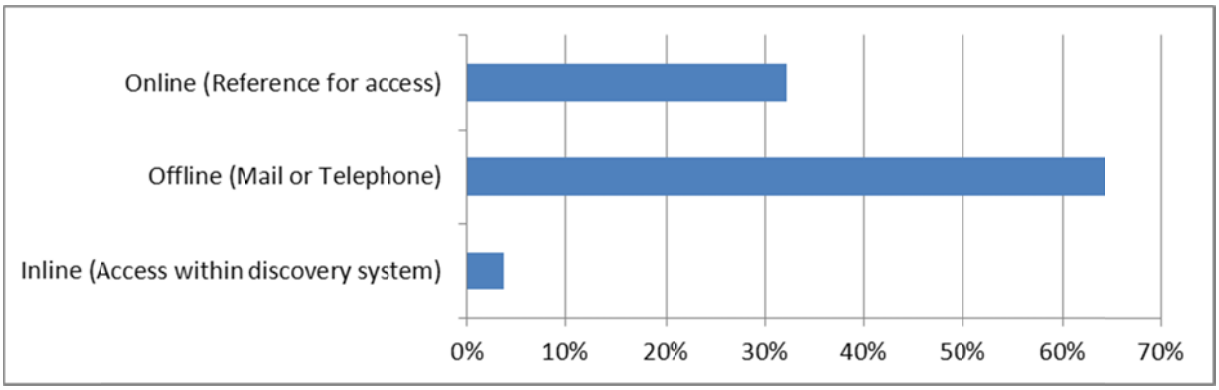

Figure 11 Type of data request

4.1.4 Expectations

The questionnaire also tried to collect expectations from the ExpeER infrastructures with regard to data management and data policy.

4.1.4.1 Data management

A central database, a data portal or web services to share the commonly used data within the project was equally expected by the ExpeER infrastructures (Figure 12).

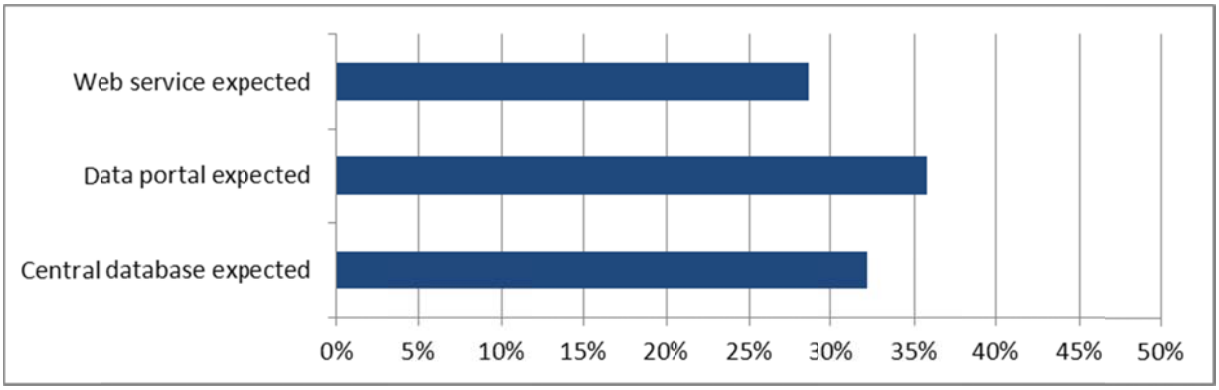

Figure 12 Expected functionalities (multiple answers possible)

The share of the different service types showed that the OGC metadata (CSW) and data services (WMS, WFS, WCS) were the main services expected. These services are mainly used to share simple

structured spatial information. Figure 103 gives an overview about the expected services according to the questionnaire.

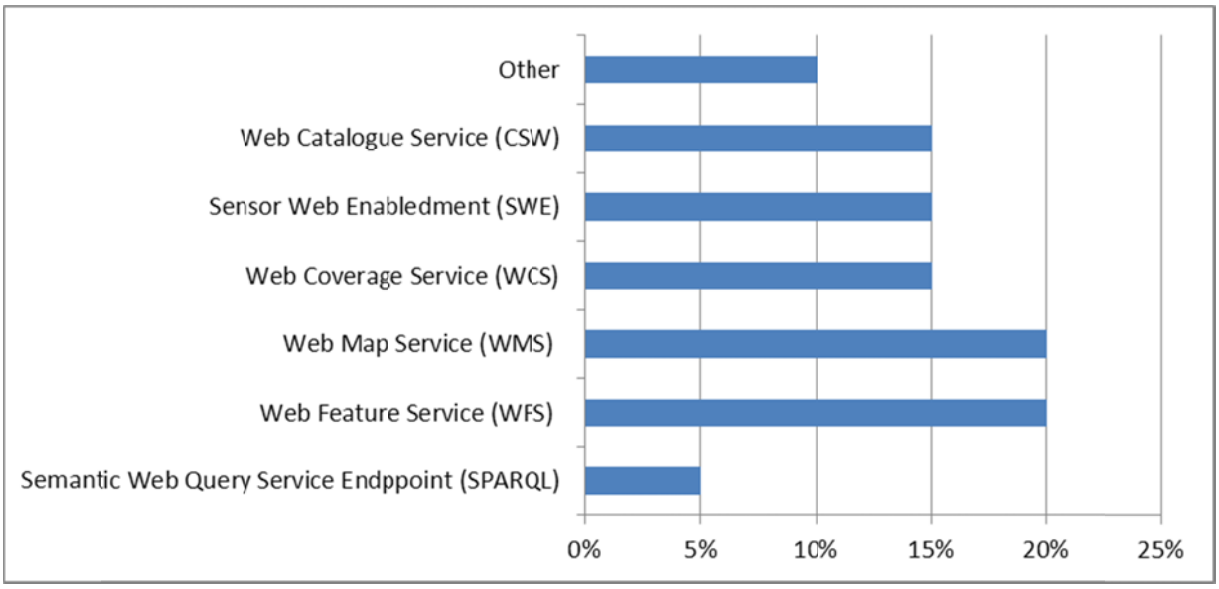

Figure 10 Type of service expected (Share of yes)

Sensor web enablement (SWE, SOS) for observation data was expected by 11%, Semantic Web Query Service Endpoints (e.g. SPARQL interface) by 4% of the infrastructures.

The expected user groups were the same as the current user groups, but with much lower levels (Figure 14). The main user groups for ExpeER data are the general scientific community, ExpeER partners and users, as well as scientific education.

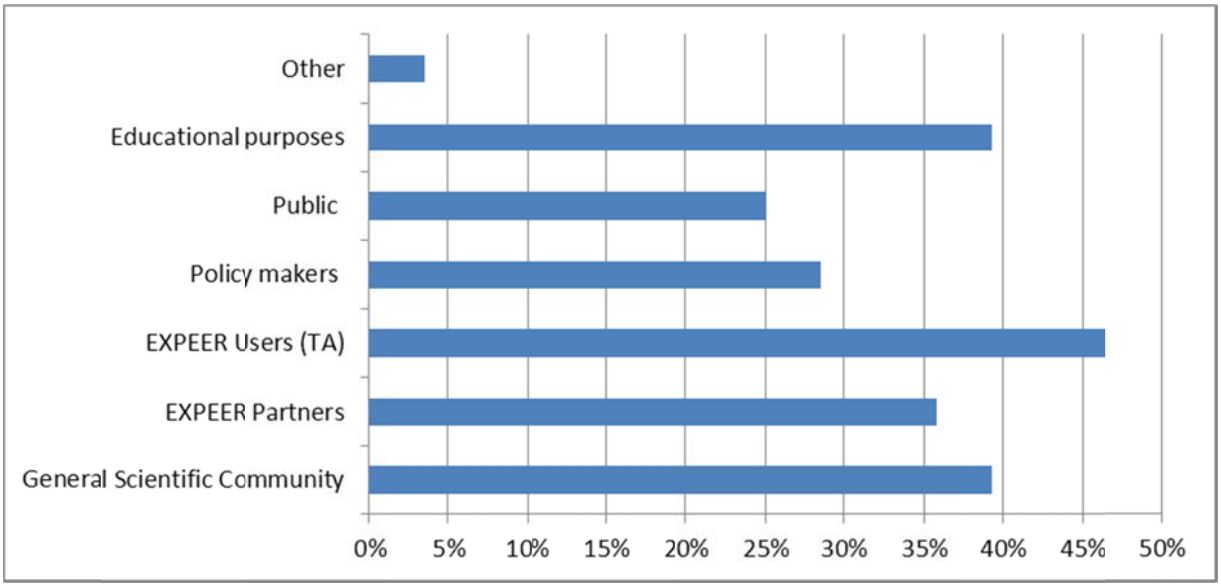

Figure 114 Expected user groups for the ExpeER project

Only 14 % of the infrastructures were willing to participate as demonstration infrastructure for data management developments in the ExpeER project. This indicates a low willingness to participate in the work on further enhancements of the data management within the network of infrastructures.

4.1.4.2 Data policy

The expectations from transnational access (TA) within the ExpeER project are aiming mainly at the enhancement of data availability for modelling and analysis. The free exchange and use of data is one of the main statements. This applies for

a) Commonly used data in the project context
b) Data generated within the project context in TA from partners

4.2 Existing Data sharing policies

There are several data sharing policies available from other projects operating in a similar field of research. Some are quite general. Others are going more into detail. It seems that large communities (composed of different institutions) tend to be less explicit with data policies than smaller communities (e.g. within one institution). Moreover, it is often preferred to rather express the commitment based on a memorandum of understanding than as a legal contract, as institutions are quite careful in signing legally binding data policies. However, it may happen that sometimes these data policies are overruled by e.g. national regulations or regulations from the EC.

The most often used pragmatic way of sharing data regardless of regulations behind seems to be the direct contact to the data owner, asking for data and offering something like e.g. co-authorship of a paper based on these data. Perhaps this was and will work in most cases related to data sharing between scientists.

The following data policies were taken into consideration to serve as templates for the ExpeER data sharing policy: TERENO, NERC, Rothamsted, NEON, CarboEurope, IMMEC (see Annex 2)

4.2.1 ExpeER Consortium Agreement

A basic and agreed document providing basic regulations and rules is the consortium agreement which is actually in the process of beeing signed by all Expeer partners. The following citations have been extracted from the ExpeER consortium agreement based on the file *ExpeER - Consortium Agreement (DRAFT 3).docx*:

SECTION 1: DEFINITIONS

"**Data Base(s)"** means the database(s) provided for use in/developed under the Project by the Parties, which may be accessible/made available as a downloaded file or via any other medium (CD, floppy disk etc.). Data Base shall include:

- -the presentation of the Data therein and the structure of the database;
- -the thesaurus and the indexing system used for the Data Base;
- -the associated documentation (explanatory files or other written documents);

- -all updates, modified versions, additions and copies of the Data Base.

SECTION 8: FOREGROUND

Regarding Foreground, EC-GA Article II.26. - Article II.29. shall apply with the following additions:

8.1 ownership

Foreground shall be the property of the Party carrying out the work generating that Foreground. Each Partner shall remain the sole owner of its Background.

8.2 Joint ownership

Where several Parties have jointly carried out work generating Foreground and where their respective share of the work cannot be ascertained, they shall have joint ownership of such Foreground.

The Parties concerned (the joint-owners) shall make their best efforts to establish a joint-ownership agreement regarding the allocation and terms of exercising such joint ownership, within 10 months from the date of creation of the jointly owned Foreground.

The share of each of the Joint -Owners to the development of such joint-Foreground shall be defined between the Joint -Owners proportionally to the intellectual contribution by each respective Joint -Owner.

Each of the Joint -Owners shall be entitled to use their jointly owned Foreground for commercial purposes and to grant non-exclusive licenses to third parties, without any right to sub-license, subject to the following conditions: at least 45 calendar days prior notice must be given to the other Joint -Owner(s); and fair and reasonable compensation must be provided to the other Joint -Owner(s); and written consent must be provided by the other Joint -Owner(s) (In any case Joint -Owners may freely use their jointly owned Foreground for the purposes of academic research, teaching and education without the prior consent of the other Joint Owners).

8.3 Transfer of Foreground

8.3.1 Each Party may transfer ownership of its own Foreground following the procedures of the EC-GA Article II 27.

8.3.2 It may identify specific third parties it intends to transfer the ownership of its Foreground to in Attachment (7) to this Consortium Agreement. The other Parties hereby waive their right to object to a transfer to listed third parties according to the EC-GA Article II.27.3.

8.3.3 The transferring Party shall, however, notify the other Parties of such transfer and shall ensure that the rights of the other Parties will not be affected by such transfer.

Any addition to Attachment (7) after signature of this Agreement requires a decision of the General Assembly.

- Page 24, section 8.3.3: Please add following words: "... includes their Foreground or Background in application of the Access rights as specified in this Consortium Agreement. However, confidentiality and publication clauses have to be respected."

8.3.4 The Parties recognize that in the framework of a merger or an acquisition of an important part of its assets, a Party may be subject to confidentiality obligations which prevent it from giving the full 45 days prior notice for the transfer as foreseen in the EC-GA, Article II 27.2.

5 Proposal for ExpeER data sharing policy

The backbone for the proposal for a data sharing policy within ExpeER is the recently elaborated data policy for the German TERENO (Terrestrial Environmental Observatories) initiative. This seems to be a suitable template for ExpeER, as the kind of data gathered (from device-based to modelling) is quite similar to ExpeER, it is developed as a frame for different institutions and several TERENO institutions are ExpeER partners. The data policies in the annex were used to amend and complete the backbone.

5.1 Scope

The ExpeER Data Policy addresses issues related to the provision, exchange, availability, maintenance and use of data produced by the ExpeER research activities (mainly based on transnational access projects).

5.2 Purpose

The purpose of the ExpeER Data Policy is to set up fundamental principles in view of

- facilitating collaboration among the participants to the ExpeER project;
- release of metadata as soon as data is collected.
- ensuring timely submission of data for the use within the consortium without preventing fair use;
- protecting the researchers' Intellectual Property Right(s) (IPR) and rights to publish their results;
- providing rules for use of the data within the consortium and by third parties;
- building an homogeneous body of valuable scientific information on environmental data and experimental results gathered at ExpeER facilities.;
- providing the broader scientific community with an easy access to the above integrated knowledge.

5.3 Data definition and data access

Data are always being regarded as a combination of observation data and its describing metadata. Metadata have to be provided by the data originators for all kinds of data. Metadata standards recommendations are defined within ExpeER.

Dataset documentation is a critical step in the data management planning process.

Metadata release according to the metadata standard recommendations is to be done as soon as data is collected.

Three observation data levels depending on the degree of underlying pre-processing and evaluation steps are defined:

- Level 0: raw data without any pre-processing
- Level 1: pre-processed data, quality checked
- Level 2: derived data, spatially and/or temporally aggregated

Four different observation/experiment data types gained within or in the context of ExpeER are defined:

- Standard monitoring data produced by ExpeER installed/financed devices: data from monitoring networks, which have been especially setup for and dedicated to data online access
- Experiment data
- Project data: data acquired within or necessary for a certain ExpeER transnational access project
- Third party data: data from external institutions outside the ExpeER project which have been provided to ExpeER partners for special purposes and/or internal use.

Supplementary data: this covers all data gathered outside the context of ExpeER. This may be e.g. historic data provided and arranged for specific purposes within ExpeER. This type of data is mainly of level 2 being temporally and sometime spatially aggregated.

5.4 Availability and exchange of data

The access to ExpeER data may be either free to the public or restricted to certain users, e.g. to the consortium partners or upon request according to the following matrix:

	Standard-monitoring data	Experiment data	Project data	Third party data	Supplementary data
Level 0	-	-	-	-	-
Level 1	Free	Restricted	Restricted	Restricted	Restricted
Level 2	Free	Free	Free	Restricted	Free within ExpeER

5.5 Data delivery

In order to warrant continuity to the research activities within ExpeER, it is essential that data produced by one ExpeER team are made available as soon as possible to the other ExpeER groups (e.g. via accessible databases of the dedicated Data Centre). In producing ExpeER datasets, site operators need to refer to the Data Management Plan and the ExpeER conditions of use (chapter 5.11).

Standard monitoring data are to be made available online within three months after installation. Project data are to be made available by database access within three months after project lifetime or according to other allocated submission deadlines.

It will be the site managers' responsibility to ensure that submission deadlines are observed and to chase possibly late or missing data.

Failure to deliver data on time may be reported to the Executive Committee for action to be taken.

5.6 Sensitive data

There may exist "sensitive data" which are not to be distributed without special considerations. In this case data holders are not committed to deliver these data. However it is recommended for measures to be put in place as part of the data management planning to maximise release of information without releasing the sensitive aspects of the data. Sensitive data are regarded as

d) Data which can not be shared because of national regulations, e.g. personal data
e) Locations which can not be shared because of forseen threats, e.g. nature protection
f) Data which can not be shared, because of pending publications within the retention time (see 5.7)

5.7 Data access within the ExpeER community

The term ExpeER community includes all partner institutions being members of the project and the transnational access users as well.

All data will be freely available within the ExpeER community as soon as they have been finalized and inserted into the database of the partner institution.

Requests for data will be dealt with by email notification to data owners, and permission will include IPR owners' disclaimer.

In order to allow investigators to validate and analyse their data and to publish their results, access to ExpeER data will be restricted for a certain amount of time (=retention time). This retention time period will be five years after the submission due date or two years after the end of the project, depending on whichever occurs first.

5.8 Data access from outside the ExpeER community

While the retention period applies, data may be exceptionally communicated to external collaborators who have applied as above, after approval by the Data Originator(s) and Data Owner(s).

After the retention period (maximum of 5 years, sooner if appropriate), the ExpeER data will be released to the public but conditions of use will continue to apply regarding publications based on the data.

Metadata are free for use to any party within and outside the ExpeER community.

5.9 Intellectual property right (IPR) issues

IPRs belong to the Data Originator or their respective institutes. During the retention period, the originators will have the priority to publish the analysis of their own data. In case of shared ownership analysis of the data will be published collectively by the originators.

Furthermore, at any time (i.e., during or after the retention period), investigators who have collected and/or done the initial analysis of the data will be invited to be co-authors of any publication or communication based on or using ExpeER data, whether undertaken by scientists within the ExpeER project or outside of it.

In the event that the offer is declined, investigators who have collected and analysed the data must be duly acknowledged.

5.10 Security and longevity

Observational / experimental data will be kept for the long term in decentralized data bases operated by each partner. Data bases will be maintained and security will be ensured by making regular backups.

Model data are more sensitive to theoretical progress and code updates, and these data will be reviewed in interaction with the originators.

5.11 ExpeER Conditions of Use

The following conditions will apply to any use of ExpeER data, at any time. These data were generated or collected within the framework of the ExpeER partners. Metadata are free for use to any party within and outside the ExpeER community.
The user of other ExpeER data besides metadata agrees:

- to contact the Data Originator prior to any use of the data;
- not to disclose the data to other parties;
- to restrict the use of the data to the context of the research topic specified at the time of application, when this application was made to access data still restricted from the public domain;
- that the Intellectual Property Right remains with the Data Originator;
- not to use the data for commercial purposes;
- to offer the Data Originator(s) co-authorship of any publication or communication based on these data; in the event that the offer is declined, Data Originators must be duly acknowledged.

6 Discussion

The results of the questionnaire showed a high level of organisation in terms of data management and data policy for infrastructures participating in the ExpeER project. The infrastructures are HIES (highly instrumented experimental sites) and HIOS (highly instrumented observation sites). Both are often characterised by a centralised data management and a higher share of online available data.

Despite approaches to a standardised data policy within the infrastructures the main data policy model still is "free upon request". Direct negotiation between the data user and the data owner based on the specific use case is state of the art.

In this respect the low expectations according to the results of data management within the ExpeER project are not surprising. Existing data infrastructures are implemented or are planned on the level of the infrastructures and are supported by them.

Nevertheless the subject of enhancing the free availability of data is addressed in the results of the questionnaire. Topics of data harmonisation and data integration are not really in the focus of the infrastructures of the ExpeER projects. Nevertheless, at a European scale perspective these are the main topics to be solved in the near future.

The policy recommendations outlined in this report take into consideration the current information management landscape, a fairly diverse and heterogeneous community with a high focus on scientific outputs and diverse and relatively low expectations for what an infrastructure project like ExpeER can and should implement in terms of data management.

However, it is essential for a research infrastructure like ExpeER to invest and implement a strong data policy that enables discovery and promotes reuse of the data and information generated by those facilities and its users.

7 References

This deliverable is based on data derived from a questionnaire circulated within the ExpeER consortium and on examples of data policies available in the web. Therefore only some web links of mentioned examples for data policies are referenced in this section.

1. CarboEurope: http://www.carboeurope.org/ (see "About us" section)
2. IMMEC (Infrastructure for Measurements of the European Carbon Cycle): http://www.europe-fluxdata.eu/newtcdc2/IMECC-TCDC_home/Guidelines_Copy1/Obtaining/datapol.aspx
3. NERC / CEH: http://www.nerc.ac.uk/research/sites/data/policy.asp
4. ECN: http://data.ecn.ac.uk/datapolicy.asp
5. Rothamsted: http://www.era.rothamsted.ac.uk/index.php?area=home&page=index
6. NEON: http://www.neoninc.org/sites/default/files/NEON%20Data%20Policy%20Spring%202010.pdf
7. TERENO: http://www.tereno.net
8. UK Environmental Change Network: http://data.ecn.ac.uk/datapolicy.asp

Annexes

8 Questionnaire

Questionnaire adapted from EnvEurope DM Survey

1. QUESTIONS REGARDING THE MANAGEMENT OF DATA RELEVANT TO EXPEER FACILITIES

EXPEER Site: *Please fill in site name!* Contact: *Please fill in name/email of contact person!*

NOTE: Ask for help of colleagues (e.g. data managers) in case you cannot answer the questions!

	Select from list	Comments
Are your datasets stored in digital format		

If yes - your data are stored in digital format		
In which format are the data stored? Please also specifiy the software used for the data management.	**Select from list**	**Software, please specify**
Database (relational) *e.g. Oracle, Access, etc.*		
XML Database (XML repository)		
Structured files (spreadsheets) *e.g. Excel*		
Unstructured file *e.g. unstructured text file*		
Spatial data file (GIS) *e.g. shape file*		
Spatial database (geodatabase) *e.g. PostGIS, Personal GeoDatabase, ArcSDE*		
Other		
Comments		

	Select from list	Number of places, please specify
Where are your data stored		
Comments		

If no - your data are not stored in digital format or proprietary digital format	**Select from list**	**Please specify**
Please specify the way your data are managed		
Separate files according the specific needs		
Filled in paper, printed paper forms		
Proprietary file formats produced by different field equipments/sensors		
Other		
Comments		

Do you have special staff – data managers employed or the databases are managed by scientific or technical staff of your institute?	**Select from list**	**Contacts**
Data management department		
Scientific staff		
Technical staff in scientific department		
Other		

2. QUESTIONS ABOUT DATA SHARING POLICY OF DATA RELEVANT TO EXPEER FACILITIES

NOTE: Ask for help of colleagues (e.g. data managers) in case you cannot answer the questions!

Are these data freely available?

	Free	Free upon request	Restricted	No	Other	Definition of user groups, please specify
Administration						
Public						
Education						
Research						
Other (e.g. Commercial...)						
Comments (e.g. define "others" if appropriate)						
Could any specific subset of the EXPEER site data pose issues with regards to data dissemination and re-use?						
Under what conditions would that data mentioned before be made available?						
What are your institution's provision in terms of data management planning, and data curation?						
What are your institution's expectations with regards to availability of the data generated by Transnational Access users at your site?						

Access options

Free … *can be used by everybody under the specified terms of use*
Free upon request … *the terms of use are negotiated case by case, but the data are in principal free to use*
Restricted … *the use of the data is only for either a restricted group or for restricted purposes. The terms of use are specified.*
No … *data are not free to use or fees are charged*
Other … *other terms of use, please specify*

Do you expect reimbursement for use of site-related data gained outside EXPEER, but related/useful for Transnational Access Projects?

	No	Data manipulation costs	Data creation costs	Other	Definition of user groups, please specify
Administration					
Public					
Education					
Research					
Other					
Comments					

Cost options:

No … *no costs are charged for the data exchange*
Charge Data manipulation cost … *only data manipulation costs are charged to the user*
Charge Data creation costs … *also costs of the data generation are charged to the user*
Other … *other cost model, please specify*

What are the main user groups? Please also indicate the users more specifically (e.g. universities, municipalities, etc.).	**Select from list**	**Please specify**
Administration		
Public		
Education		
Research		
Other		
Comments		

	Select from list	**Link or contact, please specify**
How can the data be requested? Are the data available online (data portal / data service) or must it be requested by direct contact?		
Comments		

	Select from list	**Comments**
Do you have "sensitive" data (e.g. data about rare protected species, personal data, etc), which can be shared only under certain circumstances?		

3. QUESTIONS ABOUT DATA ACCESS AND DATA DISTRIBUTION

NOTE: Ask for help of colleagues (e.g. data managers) in case you cannot answer the questions!

	Select from list	URL/Link	Comment
Are you providing a data portal?			

	Select from list	Comments
Are you providing a web service to provide data online?		

If yes: what kind of web services is used (ask your IT experts)?	**Select from list**	**Software, please specify**	**URL/link**
Web Feature Service (WFS)			
Web Map Service (WMS)			
Web Coverage Service (WCS)			
Sensor Web Enabledment (SWE)			
Web Catalogue Service (CSW)			
Semantic Web Query Service Endppoint (SPARQL)			
Other			
Please indicate whether any of these services could be made available through an "ExpeER" portal.			
Comments			

To whom it might your data be useful? (multiple selection can be made)	Select from list	Please specify
EXPEER community		
EXPEER Infrastructure users		
General Scientific community		
Policy makers		
Public		
Educational purposes		
Other		
Comments		

4. EXPECTATIONS AND REQUIREMENTS FOR EXPEER DATA MANAGEMENT

NOTE: Ask for help of colleagues (e.g. data managers) in case you cannot answer the questions!

	Select from list	Requirements, please specify
Are you expecting a data portal providing the EXPEER data for query and download? (Please consult)		

	Select from list	Requirements, please specify
Are you expecting a central database providing the EXPEER data for query and download?		

	Select from list	Requirements, please specify
Are you expecting a web service to provide data online?		
If yes: What kind of web services are you expecting (maybe you have to ask your IT experts)?	**Select from list**	**Please specify**
Web Feature Service (WFS)		
Web Map Service (WMS)		
Web Coverage Service (WCS)		
Sensor Web Enabledment (SWE)		
Web Catalogue Service (CSW)		
Semantic Web Query Service Endppoint (SPARQL)		
Other		
Comments		
If answered yes to any of the above questions, would you be willing to be part of a pilot team to demonstrate EXPEER web service capabilities?		
If yes: Who should be able to access and use the data pool of EXPEER? (multiple selection can be made)	**Select from list**	**Please specify**
EXPEER Partners only		
EXPEER Users (Transnational Access)		
General Scientific community		
Policy makers		
Public		
Educational purposes		
Other		
Comments		
Which data are you interested in to get from OTHER ENVIRONMENTAL DATA Networks?		

5. COMMENTS AND SUGGESTIONS

Please, indicate any comments, suggestions and/or information you judge necessary to be integrated in this questionnaire that are not previously included.	

9 Examples. Data sharing policy

9.1 TERENO Data Policy

Introduction

This document outlines the principles adopted by the TERENO project regarding its data policy. The detailed implementation of these principles will be further described in the TERENO Data Management Plan, to be developed in the course of the first 6 months of the project and updated throughout its lifetime. This Data Policy was developed by the Data Management Coordination Team (DMCT) and its principles have been agreed by the TERENO Scientific Steering Committee.

9.1.1 Scope

The TERENO Data Policy addresses issues related to the provision, exchange, availability, maintenance and use of data produced by the TERENO research activities or acquired from third parties as a support to this research.

9.1.2 Purpose

The purpose of the TERENO Data Policy is to set up fundamental principles in view of

- easing collaboration among the participants to the TERENO project;
- ensuring timely submission of data for the use within the consortium;
- protecting the researchers' Intellectual Property Right(s) (IPR) and rights to publish their results;
- providing rules for use of the data within the consortium and by third parties;
- building an homogeneous body of valuable scientific information on results observed within TERENO network;
- providing the broader scientific community with an easy access to the above integrated knowledge.

9.1.3 Data definition and data access

Data are always being regarded as a combination of the observation data and its describing metadata. Metadata have to be provided by the data originators for all kinds of data. Applicable standards are specified by the Data Management Plan.

Three data levels depending on the degree of underlying pre-processing and evaluation steps are defined:

- Level 0: raw data without any pre-processing
- Level1: pre-processed data, quality checked
- Level2: derived data, spatially and/or temporally aggregated

Three different data types are defined:

- Standard monitoring data: data from monitoring networks, which have been especially setup for and dedicated to data online access
- Project data: data acquired within a certain TERENO observation project

- Third party data: data from external institutions outside the TERENO project which have been provided to the TERNO partners for special purposes and/or internal use.

9.1.4 Availability and exchange of data

The access to data may be either free to the public (see section 5.4) or restricted to certain users, e.g. to the consortium partners or upon request according to the following matrix:

	Standard-monitoring	Project data	Third party data
Level 0	-	-	-
Level 1	Free	Restricted	restricted
Level 2	Free	Free	restricted

9.1.5 Data delivery

In order to warrant continuity to the research activities within TERENO, it is essential that data produced by one TERENO team are made available as soon as possible to the other TERENO groups (e.g. via accessible databases of the dedicated Data Centre). In producing TERENO datasets, site operators need to refer to the Data Management Plan and the TERENO conditions of use (chapter 10).

Standard monitoring data are to be made available online within three months after installation. Project data are to be made available by database access within three months after project lifetime or according to other allocated submission deadlines.

It will be the Observatory's Data Managers' responsibility to ensure that submission deadlines are observed and to chase possibly late or missing data.

An initial time chart considering the internal dynamics and interactions of all TERENO components in terms of data deliverables and needs is included in the Data Management Plan.

It is likely that this schedule will evolve as a function of the work progress and the possible unforeseen obstacles or successes met in the course of the TERENO project by each of its components.

Failure to deliver data on time may be reported to the Scientific Steering Committee (SSC) for action to be taken.

9.1.6 Data access within the TERENO community

TERENO community includes all institutions contributing to one of the TERENO observatories.

All data will be immediately freely available within the TERENO community as soon as they have been finalized and inserted into the database.

Requests for data will be dealt with by email notification to data owners, and permission will include IPR owners' disclaimer.

In order to allow investigators to validate and analyse their data and to publish their results, access to TERENO data will be restricted for a certain amount of time (=retention time). This retention time period will be five years after the submission due date or two years after the end of the project, depending on whichever occurs first.

9.1.7 Data access from outside the TERENO community

While the retention period applies, data may be exceptionally communicated to external collaborators who have applied as above, after approval by the Data Originator(s) and Data Owner(s).

On implementation of the Central Database (Web Portal), a public report will be available to indicate what data are available. Thereafter, interested parties may apply for datasets as described above.

After the retention period (maximum of 5 years, sooner if appropriate), the TERENO data will be released to the public but conditions of use will continue to apply regarding publications based on the data. The way this is enforced will be defined in the Data Management Plan (DMP).

In the event that third-party data are lodged at one of the DC to support TERENO research, access to these data will be ruled by the provider's data protocol. It is recommended that confidential data, that is, data not intended to be made public due to their potential commercial value or their private nature, be archived at one of the TERENO DC only if they are a potential source of derived products that may be publicly distributed (e.g. subsets, averages, statistics,…).

However, as it is likely that deriving by-products will require additional resources, it is left to the DC to make a decision on a case per case basis.

9.1.8 Intellectual property right (IPR) issues

IPRs belong to the Data Originator or their respective institutes (observatory). During the retention period, the originators will have the priority to publish the analysis of their own data. In case of shared ownership analysis of the data will be published collectively by the originators.

Furthermore, at any time (i.e., during or after the retention period), investigators who have collected and/or done the initial analysis of the data will be invited to be co-authors of any publication or communication based on or using TERENO data, whether undertaken by scientists within the TERENO project or outside of it.

In the event that the offer is declined, investigators who have collected and analysed the data must be duly acknowledged.

9.1.9 Security and longevity

Observational data will be kept for the long term in a decentralized data base operated for each observatory by the designated data centres. The designated DC will maintain the TERENO datasets and ensure the security of the database by making regular backups.

Model data are more sensitive to theoretical progress and code updates, and these data will be reviewed in interaction with the originators, according to the DC's established procedures.

Options for long term archiving, in particular beyond the retention period and the project run time, will be explored in the course of TERENO.

9.1.10 TERENO Conditions of Use

The following conditions will apply to any use of TERENO data, at any time. These data were generated or collected within the framework of the TERENO consortia. The user agrees:

- to contact the Data Originator prior to any use of the data;
- not to disclose the data to other parties;

- to restrict the use of the data to the context of the research topic specified at the time of application, when this application was made to access data still restricted from the public domain;
- that the Intellectual Property Right remains with the Data Originator;
- not to use the data for commercial purposes;
- to offer the Data Originator(s) co-authorship of any publication or communication based on these data; in the event that the offer is declined, Data Originators must be duly acknowledged."

9.1.11 Links with a relevance to the Data Policy

TERENO website: http://www.tereno.net

TERENO Data Online Repository (TEODOOR): https://teodoor.icg.kfa-juelich.de

9.1.12 List of acronyms and definitions

ODM	Observatory Data Manager
Data Originator	Investigator who has collected and / or conducted the initial analysis of the data
Data Owner	Institution owning the collected data
DC	Data Centre(s)
DMCT	Data Management Coordination Team
DMP	Data Management Plan
DP	Data Policy
External Collaborator	Third Party (see below) working in a similar field who has entered a (formal or informal) two-way collaboration agreement with participants of TERENO consortium
IPR	Intellectual Property Right(s)
PI	Principal Investigator. Person responsible for the execution of the project or part of the project at a partner institute as defined in the Consortium Agreement
SSC	Scientific Steering Committee
Third Party	Provider or user of data who is not part of the TERENO consortium

9.2 NERC Data Policy

This new version of the NERC Data Policy was approved by the NERC Executive Board in September 2010, and comes into force in January 2011; however, the requirement for data management plans will not be implemented until 2012, to allow NERC time to implement new grant application and review processes fully as part of the migration of grant processing to the RCUK Shared Service Centre.

9.2.1 NERC Data Policy Statement

NERC has a policy on data in order to:

1. Ensure the continuing availability of environmental data of long-term value for research, teaching, and for wider exploitation for the public good, by individuals, government, business and other organisations.
2. Support the integrity, transparency and openness of the research it supports.
3. Help in the formal publication of data sets, as well as enabling the tracking of their usage to be tracked through citation and data licences.
4. Meet relevant legislation and government guidance on the management and distribution of environmental information.

NERC defines environmental data as individual items or records (both digital and analogue) usually obtained by measurement, observation or modelling of the natural world and the impact of humans upon it. This includes data generated through complex systems, such as information retrieval algorithms, data assimilation techniques and the application of models.

This policy covers environmental data acquired, assembled or created through research, survey and monitoring activities that are either fully or partially funded by NERC. It also applies to environmental data managed by NERC where NERC was not the original funder. This policy does not cover NERC's information products*.

This policy will be reviewed at regular intervals to ensure it keeps pace with scientific requirements and data management best practice.

9.2.2 Key principles

The environmental data produced by the activities funded by NERC are considered a public good and they will be made openly available for others to use. NERC is committed to supporting long-term environmental data management to enable continuing access to these data.

NERC will supply the environmental data it holds for free, apart from a few special cases as detailed in the policy.

NERC requires that all environmental data of long-term value generated through NERC-funded activities must be submitted to NERC for long-term management and dissemination.

9.2.3 Access to data

It is NERC's policy that:

1. All the environmental data held by the NERC Environmental Data Centres will normally be made openly available to any person or any organisation who requests them.
2. The only restrictions on access which we will apply are those supported by the exceptions on disclosure in the Environmental Information Regulations (2004). If it is proposed to restrict access to any data we will explain why.
3. To protect the research process NERC will allow those who undertake NERC-funded work a period to work exclusively on, and publish the results of, the data they have collected. This period will normally be a maximum of two years from the end of data collection.
4. All data held by the NERC Environmental Data Centres will be supplied for free except for large or complex requests where we may charge the cost of supply, or where third-party licence conditions either prevent such free supply, or require us to make specific charges.
5. All environmental data made available by the NERC Environmental Data Centres will be accompanied by a data licence. Data originally provided to NERC by a third-party may have their own access and licence conditions which restrict how or when we can make data available to others, in which case our data licence conditions will reflect these.
6. All those who use data provided by NERC are required to acknowledge the source of the data.

9.2.4 NERC's Environmental Data Centres

Successful long-term data management requires both specialist data curation skills and an understanding of the science behind the data. NERC achieves this by supporting Environmental Data Centres and co-locating these within its research and collaborative centres to take advantage of the scientific expertise they possess.

It is NERC's policy that:

1. NERC will maintain Environmental Data Centres for the management and dissemination of environmental data of long-term value generated through NERC funding or deposited by third-parties.
2. The data centres will act impartially towards all data producers, regardless of whether they are based within or outside of NERC. The environmental data within the data centres will be open to all on the same basis.
3. Working with the environmental science community NERC will maintain criteria to identify environmental data of long-term value (a Data Value Checklist). These criteria will be used to inform all decisions that NERC makes on the acceptance and disposal of data by its data centres.

4. Information on all data held within the data centres will be made available through the NERC Data Discovery Service.

9.2.5 Data collection

NERC expects everyone that it funds to manage the data they produce in an effective manner for the lifetime of their project, and for these data to be made available for others to use with as few restrictions as possible, and in a timely manner.

It is NERC's policy that:

1. All applications for NERC funding must include an outline Data Management Plan, which must identify which of the data sets being produced are considered to be of long-term value, based on the criteria in NERC's Data Value Checklist. The funding application must also identify all resources needed to implement the Data Management Plan.
2. The outline data management plan will be evaluated as part of the standard NERC grant assessment process. All successful applications will be required to produce a detailed data management plan in conjunction with the appropriate NERC data centre.
3. All NERC-funded projects must work with the appropriate NERC data centre to implement the data management plan, ensuring that data of long-term value are submitted to the data centre in an agreed format and accompanied by all necessary metadata.
4. Data from NERC-funded activities are provided to the data centres on a non-exclusive basis without prejudice to any intellectual property rights. This is to enable NERC to manage and make openly available publicly funded research data.
5. Those funded by NERC who do not meet these requirements risk having award payments withheld or becoming ineligible for future funding from NERC.

9.2.6 Open access to data underpinning research publications

NERC considers that long-term, open access to the data that underpin research publications will help to ensure the integrity, transparency and robustness of the research record. Access to these data supports the fundamental scientific requirement of allowing others to confirm or challenge research results.

It is NERC's policy that:

1. All research publications arising from NERC funding must include a statement on how the supporting data and any other relevant research materials can be accessed.
2. For all research publications produced by NERC's own staff, the supporting data will be made available through the NERC data centres.

* In line with UK Government policy, NERC distinguishes between data and information products. NERC defines environmental data as individual items or records (both digital and analogue) usually obtained by measurement, observation or modelling of the natural world and human impacts upon it, including all necessary calibration and quality control. This includes data generated through complex systems, such as information retrieval algorithms, data assimilation techniques and the application of models. Whereas, information products are created by adding a level of intellectual input that refines or adds value to data through interpretation and/or combination with other data. Model codes are not covered by this policy.

If you have any comments or questions, or require further information on the NERC Data Policy, please contact NERC's data management co-ordinator, Mark Thorley.

9.3 ECN Data Centre Data Policy

9.3.1 Ownership

The Natural Environment Research Council (NERC) holds Intellectual Property Rights (IPR) for all ECN data jointly with each participant, participants having equal IPR with NERC for their own sites' data. Data originators retain exclusive rights to ECN data collected at their own sites for applications outside the remit of ECN.

ECN has adopted the principles of the NERC data policy for long-term management and dissemination of data.

9.3.2 Data Access and Licensing

ECN Core Data

Data collected by the ECN Network are for the public good and will be made openly available for others to use. ECN will supply data for free (unless there are specific third-party licensing arrangements in place or if a request is particularly complex). The policy allows ECN researchers a period (a maximum of two years) to work on the data exclusively before data are made available to others. Data are made available under license.

ECN Summary Data

ECN summary data are freely available, but subject to the conditions of use stated in the ECN license agreement. These conditions are repeated on the relevant internet pages. Users must pass through this page to progress with the database access system, and they accept the conditions by the act of pressing the 'Accept Conditions and access the Data' button. Users and use are logged by email and IP address.

9.3.3 Privacy and Cookies

Privacy - ECN does occasionally need to hold personal information about users – for example to administer data requests. This will be clearly indicated on the web page when we need to do this. This information will be held in the ECN Database, in accordance with the Data Protection Act, and will only be used for ECN purposes.

Cookies - When we provide services, we want to make them easy, useful and reliable. Where services are delivered on the internet, this sometimes involves placing small amounts of information on your device, for example, computer or mobile phone. These include small files known as cookies. They cannot be used to identify you personally.

These pieces of information are used to improve services for you through, for example:

- enabling a service to recognize your device so you don't have to give the same information several times during one task
- recognizing that you may already have given a username and password so you don't need to do it for every web page requested
- measuring how many people are using services, so they can be made easier to use and there's enough capacity to ensure they are fast

You can manage these small files yourself and learn more about them through "Internet browser cookies - what they are and how to manage them" www.direct.gov.uk/managingcookies

The ECN Data Centre uses Google Analytics to measure website traffic and help improve the effectiveness of this website. Google Analytics sets cookies to help us accurately estimate the number of visitors to the website and volumes of usage. This allows us to ensure that the service is available when you want it and fast. For further details on the cookies set by Google Analytics, please refer to the Google Code website.

The ECN Data Centre also uses cookies to temporarily store information on your computer to enable session state between Active Server Pages. Session state is a way of remembering you between web pages and we use it, for example, to ensure that you only have to log in once when viewing protected areas of the website. We do not collect any information about you except that required for system administration of the web server. These cookies automatically expire when you exit your browser.

9.4 e-Rothamsted Archive: Data Access Policy – Classical and Long Term Experiments

http://www.era.rothamsted.ac.uk/index.php?area=home&page=index

9.4.1 Principles

The Rothamsted collections of long-term experimental data represent a considerable investment of scientific expertise over many years and are owned by the Lawes Agricultural Trust. The data and associated documentation are a unique historical record from experiments that have measured the effects of agricultural practices on soils, crops and associated ecosystems continuously for over 150 years. Correct interpretation of much of these data, together with the accompanying meteorological records, requires extensive knowledge of the experimental design and data collection methods employed, many of which have evolved over their long lifetime and in many cases only documented in the original paper records. Rothamsted Research is committed to sharing these data and associated sample collections for bona fide scientific research. For the reasons stated, and associated costs of data and sample access, however, we require that potential users first make contact with us to establish the nature of the research to be undertaken so that we can (at our discretion) nominate a Rothamsted scientist as designated collaborator.

Past experience has demonstrated that dialogue with Rothamsted scientists able to direct collaborators to appropriate subsets of the data and with knowledge of the shortcomings will save our collaborators time and help prevent inadvertent misinterpretations of the data. The level of interaction required will, of course depend on the complexity of the research question and the range of data required. Where a designated (Rothamsted) collaborator has been nominated, reflecting a complex or time consuming interaction with a particular data access, then we reserve the right to approve the content of scientific publications using Rothamsted data or samples. Where the level of effort or volume of data is modest, then at our discretion, corresponding reduced recognition of the Rothamsted would be expected. The minimum recognition would be an acknowledgement.

Where access to Rothamsted data and samples will form part of a research grant application, early discussions should be had about the costs of data or sample preparation and distribution.

9.4.2 Conditions

Access by research scientists with a bona fide requirement for access to the data, intending to publish their research in peer reviewed scientific journals is never (or rarely) denied.

Access by bona fide students should be made through their course tutor or research supervisor who will be considered the official collaborator.

In the case of large or complex data or sample requests Rothamsted reserves the right to recover the full economic cost from the collaborator.

Access for the use in commercial research is encouraged, but will be subject to special license agreements negotiated with our contracts office.

Data obtained from Rothamsted Research should not be passed to a third party without written permission from either the designated Rothamsted collaborator or the e-RA curators.

9.4.3 Process

1. Access by e-mail from the e-RA website is the most common first method of contact. Direct e-mail to the e-RA curators (Claudia Underwood or Margaret Glendining) is equally acceptable.

2. An email discussion will establish the scale and nature of the data access request and usually, a designated Rothamsted Collaborator will be assigned. Further discussions as to the data required would then take place.

3. The applicant requesting access would be expected to complete a formal request document (see below) submitted either on paper or electronically.

4. Once the data subsets required have been identified, and formal approval agreed, either the applicant will be given their own password to the e-RA database and then allowed to run queries and download the agreed datasets at their convenience, or a dataset will be prepared for them.

5. Questions relating to data analysis and interpretation during the research would generally be directed to the designated Rothamsted collaborator. Technical questions would be directed to the e-RA curators.

6. When the research was ready for publication, the designated Rothamsted collaborator/co-author would be sent the manuscript for review and approval in the normal manner of collaborative publications.

7. Upon publication, the designated Rothamsted collaborator would be sent a reprint of the article for the e-RA citations database.

9.4.4 Request for access to e-RA

Applicant Name (Block Capitals)……..

Email address……..

Name and address of Institute/Organisation/Company…….

Brief description of Institute/Organisation/Company…….

Dataset(s) for which access is requested

……

Brief description of intended use:

I request access to data from the e-Rothamsted Archive 'e-RA'. I agree that the copyright and database rights for this information are vested in the owner as specified on the e-RA web site. I agree not make the data available to any third party, or publish any part of the data in any form either physical or electronic without first obtaining the approval of Rothamsted Research. I agree that I will acknowledge Rothamsted Research as the owner of the data in any publication that arises from their use. I will take all precautions necessary to ensure that the data are safeguarded against unauthorised access and/or use by any other person. My authority to use the data will terminate on a date four years from the date of my signature of this declaration and I will not retain any copy of the data after this date. Subsequent use of this data shall be conditional upon a request to renew access to the data. This declaration shall be subject to and construed and interpreted according to the Law of England.

Applicant Signature …… Date ……

Name of contact at Rothamsted Research……..

Claudia Underwood and Margaret Glendining

e-Rothamsted Archive (e-RA) Curators

Biomathematics and Bioinformatics Division (Bawden)

Rothamsted Research

Harpenden, Herts AL5 2JQ UK

9.5 NEON Data Policy March 2010

http://www.neoninc.org/sites/default/files/NEON%20Data%20Policy%20Spring%202010.pdf

9.5.1 INTRODUCTION

The National Ecological Observatory Network is an observatory comprising more than 60 atmospheric, environmental and biological monitoring locations distributed throughout twenty domains across the United States, Hawaii, Alaska & the Caribbean, which are monitored and controlled in real-time from our headquarters in Boulder, Colorado. A leading-edge, flexible cyberinfrastructure (CI) is being designed to support current and future needs of the Observatory, including acquisition, storage, calibration and distribution of the ecological data.

It is the responsibility of NEON, Inc., in consultation with the National Science Foundation and the scientific community, to implement procedures to effectively manage and provide necessary access to NEON data, while at the same time ensuring the quality, integrity and availability of the information. As a matter of principle, NEON is committed to the free and open scientific exchange of data, and NEON's data policies uphold this principle to the maximum extent allowed by applicable law and regulation. The purpose of this policy is to provide a structured and consistent process to maintain data integrity and to allow necessary data access by researchers, researchers, educators, students, policymakers and citizens.

The functions of the NEON CI are not described by the data policy, although the NEON CI system is responsible for meeting the obligations imposed by the data policy. A separate NEON Sample Use Policy document (RD[02]) outlines procedures to effectively manage and provide necessary access to NEON biological samples, while at the same time ensuring the quality, integrity and availability of the samples.

9.5.2 STATEMENT OF POLICIES

Note added September 2009: During fourth-quarter 2009, the System Engineering Product Team is transferring all Observatory requirements (including several stated or implied by this document) into the Observatory Requirements Database (RD[03]). This document remains in the project configuration to provide additional description and context for the NEON data policies.

9.5.2.1 Archiving

- All data, whether from core sites, relocatable sites, airborne platform or mobile deployment systems, or experimental programs, will be securely archived within the NEON CI.
- The basic NEON data streams from the AQU, FIU, FSU, LUAP, AOP and experiments, and selected data taken as part of instrument calibration, engineering or software testing, will be open-access and available as soon as possible once basic QA/QC procedures have been applied. Data may be available prior to QA/QC checks by special request.
- Data archives shall include easily accessible information about data holdings; high-level data products, including quality assessments and supporting relevant information Guidance for locating and obtaining the data, will be provided on the NEON Data Portal.

9.5.2.2 Accessibility

- Data will be made publicly available to users through an open-access general web interface.
- NEON will utilize community standards for the provisioning of datasets e.g. EML, NetCDF.
- NEON will enforce a principle of non-discriminatory access so that all users will be treated equally.
- Open-access means that data will be freely-distributed without charge; there may be charges for the cost of reproduction and delivery when access is not web-based. Data distribution will be unrestricted (unless for reasons of national security).
- NEON will provide an annual report (online) on use of NEON data. The NEON CI will register and record users of data and this information will be available to investigators seeking to avoid duplication of effort.
- Certain types of data have legal access restrictions, for example, information related to threatened and endangered species. If such data are held in the NEON CI, all legal restrictions will be enforced and access to such data may be limited to forms (e.g. statistical summaries) that do not violate applicable laws and regulations.
- During the construction of the NEON Observatory, all data released to the community will be considered "Pre-Release" data. These data will be offered on a best-effort basis. Pre-release data will not be reprocessed during the period of NEON construction except in extraordinary cases.

9.5.2.3 Metadata

- Metadata associated with NEON data must meet or exceed applicable national requirements (e.g., Federal Geographic Data Committee; www.fgdc.gov) and be closely associated with the archived data.
- NEON metadata includes information about data provenance, description, quality, maturity level (raw data streams, automated quality control, processed, derivative products), and collection context, and support interoperability with other observatories, archives, and databases.

NEON metadata standards will be published and openly-available.

9.5.2.4 Experiments

It is planned that the NEON infrastructure will support a wide variety of Principal Investigator (PI)-driven experiment and observational systems at the NEON domains. These systems may return data to the NEON Archive via NEON CI, and be accessible through the NEON web processes. The following policies are identified:

- NEON encourages investigators working in experimental-set aside areas to make their research data freely and openly available as soon as possible and requires data sets so collected to be registered with the NEON data archive.
- NEON will permit restricted data access for a reasonable period not exceeding two years from the end-of-experiment date for data collected by investigators using instruments

connected to NEON cyberinfrastructure or through PI-requested deployments of NEON facilities. At the end of a restricted access period, all data will be freely and openly available.

- The restricted period may be extended under exceptional circumstances, but only by agreement between the Principal Investigator and the NEON Director. With the PI's permission, data can be released to a third-party requesting user before it is publicly accessible. Data that has already become public cannot be made proprietary under any circumstances. After the exclusivity limit, the historical data as well as the current stream from the individual PI data source will be made publicly available under the policy described for core and community data. Data which has been published in whole or part becomes publicly available at the time of publication.

9.5.2.5 Business systems

The internal computers, networks, application software and data repositories of NEON, Inc. are critical resources of NEON and must be protected against inappropriate access and/or disruption of service. Active measures are necessary to ensure data integrity and reduce the risk of system compromise, especially when sensitive information may be at risk. Established procedures for protection and release of sensitive information must be followed regardless of the platform used to store that data.

9.5.2.6 Satellite Sites

It is anticipated that members of the external community may implement some or all of the NEON infrastructure at sites not directly connected to NEON. In those cases, these organizations may choose to provide their data to the NEON archive. These sites are referred to as NEON Satellite Sites. The following policies can be identified:

- NEON Satellite Sites will establish an archival and metadata plan which will be approved by NEON Observatory management.
- Data providers are responsible for insuring that data files for their sites meet NEON metadata standards, including information about data provenance, description, quality, maturity level (raw data streams, automated quality control, processed, derivative products), and collection context, and support interoperability with other observatories, archives, and databases.

9.5.2.7 Acknowledgement

Any published papers that use data from the NEON Archive should carry the following footnote or acknowledgement in the text: "The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON, Inc."

9.5.2.8 Public License

All data products published by the NEON Observatory will include the following licensing information:

Copyright © 2010 National Ecological Observatory Network, Inc.

Permission is hereby granted, free of charge, to any person or organization obtaining a copy of the NEON Observatory data products, associated software and accompanying documentation covered by this license (the "NEON Data Products") to use, reproduce, display, distribute, publish, execute, and transmit the NEON Data Products, and to prepare derivative works of the NEON Data Products, and to permit third-parties to whom the NEON Data Products are furnished to do so, all subject to the following:

The copyright notices in the NEON Data Products and this entire statement, including the above copyright statement, license grant, this restriction and the following disclaimer, must be included in all copies of the NEON Data Products, in whole or in part, and all derivative works of the Data Products.

THE NEON DATA PRODUCTS ARE PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. IN NO EVENT SHALL THE COPYRIGHT HOLDERS OR ANYONE DISTRIBUTING THE NEON DATA PRODUCTS BE LIABLE FOR ANY DAMAGES OR OTHER LIABILITY, WHETHER IN CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE NEON DATA PRODUCTS.

9.5.3 IMPLEMENTATION

The following implementation policies can be identified:

- The NEON data policies will be implemented in the NEON CI so that they are to the maximum extent possible self-enforcing.
- NEON may allow for exceptions to the guidance contained in this Statement on a case-by-case basis where permitted by law and in the furtherance of the public interest. Request for exceptions to this policy should be sent to the NEON Director.
- NEON shall review and, if warranted, update this data policy as a part of the regular reviews of NEON operations.

These principles are derived from discussions with the community, responses to the survey conducted by the Ecological Society of America, data policies followed by Federal agencies and other NSF facilities.

A subcommittee of the NEON Science, Technology and Education Advisory Committee, with representation from professional societies, will review the NEON Data Policies on a regular basis. A final NEON data policy will be approved by the NEON Board and NSF.

A separate Sample Use Policy document (RD[02]) outlines procedures to effectively manage and provide necessary access to NEON biological samples, while at the same time ensuring the quality, integrity and availability of the samples.

9.6 CarboEurope IPData Policy

9.6.1 Rights and responsibilities

Each participant (institution) in the CarboEurope-IP who handles the CarboEurope database, will produce data from measurements and models to be used within the CarboEurope-IP. Rights and responsibilities of participants are defined at the level of principle investigators (PIs). A participant may be represented by several PIs with different rights and responsibilities.

Each PI owns his/her own data and may make these data available at any time to anyone.

Each PI is responsible for making data available through the database, according to the rules in this policy document, to other PIs within the CarboEurope-IP.

Each PI is responsible for making data available directly, according to the rules of the CarboEurope-IP, to authorised participants in the CarboEurope-IP.

The CarboEurope Data Management Committee determines questions related to dataset documentation, dataset format, and quality level. The Data Management Committee also adjudicates possible disputes relating to this data policy. The Data Management Committee is composed of the PI of the central CarboEurope database and the four PIs of the CarboEurope data centres for ecosystem data, atmosphere data, regional experiment data, and auxilliary data. It is chaired by the PI of the central CarboEurope database.

9.6.2 Documentation of datasets

All datasets should be documented according to a standard. This standard is an integral part of the database and consists of the following main elements: (i) Title, (ii) Authors/owners, (iii) Introduction (incl. objectives), (iv) Theory of measurements, (v) Equipment (incl. instrument description, calibration, procedures), (vi) Observations, (vii) Parameter/variable description, (viii) Data manipulations, (ix) Errors, (x) Notes, (xi) References, and (xii) Glossary of acronyms.

Documentation of a given measurement site or region is done separately from the database. This documentation should be available on the web site of the CarboEurope-IP.

9.6.3 Access to data in the CarboEurope-IP

Access to data from version 1.0 and higher (cf. paragraph 8.d) is granted to each PI.

Data access is initially restricted to participants of the CarboEurope-IP. All data from version 1.0 and higher (cf. paragraph 8.d) are immediately available as soon as they have been inserted into the database.

All datasets delivered to the CarboEurope database may eventually be published on CD (or another suitable medium existing at the time of publication). Such publication will open free

access to the published CarboEurope-IP dat a. A possible publication will take place in certain intervals, e.g. at mid-term and after the project is finished, *i.e.*, after 1 st of January, 2009. The publication of data during the running time of the CarboEurope-IP requires the written agreement of the data owner.

The CarboEurope-IP data policy follows the subsidiary principle, *i.e.*, all datasets are organized in databases at the level of the scientific Components, called "CarboEurope data centres", with a reference to and access from the central CarboEurope-IP database, which includes the common meta-database. Participants requesting data from Components in which they are not involved will be served by the central database.

The manager of the central database will be responsible for the collection of metadata from each data centre. MPI-BGC is responsible for the management of this central CarboEurope-IP database.

The Data Management Committee, in agreement with the Executive Board and the IP Co-ordinator, is responsible for publishing, and making available on the Internet, a list of PI names in the CarboEurope-IP who are authorised to request data from the CarboEurope database.

The Data Management Committee shall facilitate the publication of data from FP5 projects of the CarboEurope cluster and provide access to these data via the CarboEurope database.

9.6.4 Internal use of data from CarboEurope-IP

Data which have been accessed by a PI must only be used for purposes necessary to carry out his/her own work in the CarboEurope-IP.

Data which have been accessed must only be used for scientific purposes, *i.e.*, commercial use of data is not allowed.

Use of data, to any significant degree, belonging to another PI for the purpose of scientific publication must always be based on an agreement between the PI and the data requester. The CarboEurope database shall offer the possibility to trace all downloads of a given dataset.

It is prohibited to distribute other PI's data to a third party without the written consent of the PI.

The termination of the participation of a participant shall in no way affects its obligation to grant access rights to the other participants pursuant to paragraph 4.a until the end of the CarboEurope-IP.

9.6.5 External use of data from CarboEurope-IP

A meta-database located at the central database will contain information about all datasets in the CarboEurope database and about the PIs responsible for them. The meta-database will be accessible via the internet for external users.

Access to data for external users is restricted to published data from version 2.0 or 3.0 (paragraph 8.d) and higher.

Access to unpublished data for external users can be granted if there is a written consent of the data owner (PI). Access to data in the CarboEurope database requires a written request to the Data Management Committee, which explains the purpose of using the data.

Special case: Joint EU/US assessments ¼

External use of data is subject to the rules set in paragraphs 4.b, 4.c, and 4.d.

9.6.6 Delivery of data to CarboEurope-IP

Data delivery implies delivery of (i) a documentation of a given dataset according to the prescribed format, and (ii) a dataset according to the documentation. Data will be inserted into the database when both parts are available.

Datasets must be specified in advance such that the database managers at the data centres, the Data Management Committee, and the External Advisory Board can determine whether a delivery is fulfilled or not. This information is also necessary for the time schedule of the database managers and for identification of hardware requirements. The Data Management Committee is responsible for collection of information about data intended for submission to the database. Even if the PI does not know exactly what the dataset will look like, she/he should submit tentative information about (i) type and number of variables, (ii) frequency of collection, (iii) time period covered, and (iv) approximate size of the dataset file. This information will be relatively straightforward to assemble for continuous time-series data but must be subject to scientific discussions within the review panel when it comes to model output and airborne measurements. This information should be gathered as early as possible. Datasets will be registred by the database manager when there are tentative descriptions of them, accepted by the Data Management Committee. Registration dates for data deliveries from the respective PIs, together with acceptance dates by the Data Managemen t Committee, will be documented and available as part of the database.

Data from continuous and regular measurements such as fluxes shall be delivered to the database every half year with a month delay after the start of the project, *i.e.*, on 1 August 2004, 1 February 2005, 1 August 2005, 1 February 2006, 1 August 2006, 1 February 2007, 1 August 2007, 1 February 2008, 1 August 2008. The datasets shall cover the time period up to one month before data delivery.

Data from time-limited field campaigns, irregular measurements, and model results shall be delivered to the database within one month after the end of the campaign, or finalization of the measurements or modelling.

9.6.7 Composite and external datasets

- Pursuant of paragraph II.35.1.c), CarboEurope-IP participants are encouraged to enter pre-existing know-how into the database. As minimum, participants shall

provide metadata to the database containing information about all available datasets potentially useful for the activities within the CarboEurope-IP and about the PIs responsible for them so that data can be requested directly from the PI.

- Data provided by organisations, external to the CarboEurope-IP and not bound by this data policy, can be entered into CarboEurope database once the Data Management Committee has appointed a specific PI for those data. This PI has the same duties towards the database for such data as a PI delivering data from his/her own activities.
- Datasets used for various modelling and analyses may be composed of a mixture of data from external and internal sources. The review panel can appoint a specific PI for such datasets. This PI has the same duties towards the database for such data as a PI delivering data from his/her own activities.
- Delivery deadlines for composite and external datasets should be set by the Data Management Committee and will be specified in the 18-months implementation plan.

9.6.8 Quality assurance

- Quality control of data in CarboEurope-IP relies on a careful review of different components submitted to the database. The Data Management Committee determine when submitted components are acceptable. Data will be inserted into the database after such acceptance.
- The Data Management Committee determines what should be accepted as a dataset. Technical and scientific reasons may require that several datasets be merged or a given dataset split to achieve functionality of the database.
- Datasets must be submitted to the database before deadline, independent of the quality level of the data. The PI should suggest, and the Data Management Committee determine, a suitable quality rating for each dataset. The PI has the chance to submit improved versions of a given dataset until the 55 th month of the project.
- Each dataset will have a version number assigned to it where the first digit signifies the quality level, and the decimal number signifies successively improved versions at a given quality level:

Version	Quality requirements	Documentation requirements
0.x	Raw data (it is up to each PI to define the meaning of "raw") which have no physical meaning to an end user (*e.g.*, mV values)	A short, simple text file is sufficient
1.x	Raw data expressed in physically meaningful units (*e.g.*, W m -2)	A documentation according to the standard format where only information about PI,

		equipment, and variable/parameter description need to be complete
2.x	Removal of erroneous data caused by obvious measurement problems (electronic spikes, *etc*.), and physically impossible or extremely rare values. Correction of data by standard procedures.	A complete documentation according to the standard format. The quality control measures must be documented.
3.x	Removal of erroneous data after comparison with other, independently measured variables (e.g., clear-sky radiation is not accepted during heavy rain), or model-derived variables. Correction of data by novel procedures.	High-level documentation of calibration and quality-control procedures
4.x	Closure of gaps in the dataset with scientifically well-defined interpolation methods.	High-level documentation of gap-filling procedures. Documentation of the quality level as a function of time.

- The decimal "x" in the version number starts with 1 for the first delivery and is followed by 2, 3, *etc*. for successive deliveries. Successive deliveries could be caused by new variables added to a dataset, additional time periods, measurement errors being corrected, *etc*. Each new version should be followed by a documentation stating what has been changed since the previous one.
- Delivery of datasets with version 0.x is primarily meant as a backup service and to simplify data exchange within a project
- Delivery of datasets labelled 1.x are intended as a quick way to exchange data for various modelling and analysis purposes within the CarboEurope-IP.

It is the goal of the CarboEurope-IP to publish datasets after peer review in an integral and referable form together with traditional scolarly papers. Datasets will only be accepted for such a publication when they have reached version 2.x or higher.

9.7 IMMEC (Infrastructure for Measurements of the European Carbon Cycle)

9.7.1 Data policy

These data are made freely available to the public and the scientific community in the belief that their wide dissemination will lead to greater understanding and new scientific insights. Every effort is made to produce the most accurate and precise measurements possible. However, the data owners represented by the project PI reserve the right to make corrections to the data for reasons deemed scientifically justified. Neither the data owners nor the IMECC consortium as a whole can be made responsible for results and conclusions based on the use of these data without regard to this warning.

The availability of these data does not constitute publication of the data. The IMECC consortium relies on the ethics and integrity of the users to assure that the data providers and IMECC as a whole receive fair credit for their work. If the data are obtained for potential use in a publication or presentation, the data owner and the Scientific Executive Board of IMECC should be informed early in this work. If the IMECC data are essential to the work, or if an important result or conclusion depends on data from IMECC, co-authorship of the data owners may be appropriate. This should be discussed at an early stage in the work. Manuscripts using data from the IMECC should be sent to the data owner and to the IMECC coordination office for review before they are submitted for publication so that it is insured that the quality and limitations of the data are accurately represented and that IMECC is properly referenced.

Data sets should be cited by the versioning number and the referring paper(s) as listed in the attached documentation.

The IMECC database provides access also the CarboEuropeIP data that since September 2011 are available to all the users directly and with an open data policy, reported below:

9.7.2 CarboEuropeIP data policy

The data downloaded have been acquired during past or ongoing European Research projects or furnished by individual scientists who encourage their use under an open data policy that emphasizes the free and open exchange of scientific information.

The data are made freely available to the public and the scientific community in the belief that their wide dissemination will lead to greater understanding and new scientific insights and that global scientific problems require international cooperation. Data download is unrestricted and requires only a free registration needed for web security reasons.

The data owners and EU projects that decided to share these data openly rely on the ethics and integrity of the users to assure that the data providers and funding projects and agencies receive fair credit for their work through inclusion of the texts provided below in the acknowledgment.

An email has been automatically sent to the contact person of each dataset downloaded to inform them about your interest in the data. We kindly ask you to send accepted papers or links to them to the database staff and PIs of the sites used in the paper. It is also recommended to contact the site PIs prior to publication to prevent potential misuse or misinterpretation of the data; if the work is based on only a few sites, this contact is strongly recommended.

Downloaded data cannot be redistributed to others and must not be redistributed via other websites, databases r any other storage system to prevent circulation of different versions of the datasets.

10 Validation by ExpeER partners

Final version of the Deliverable 3.1 Data policy and guidelines and D3.2 ExpeER metadata standard for dataset level have been uploaded to the ExpeER website in April 2012, the links to the deliverables in the website have been sent to all ExpeER participant via the internal bulletin in April 2012 as well. A deadline (17 May) was set to receive feedback and comments. No comments were received. To ensure the validation of these deliverables by each ExpeER partner, an individual email was sent to all ExpeER partner to ask for feedback by 15 June. After this deadline, principle contacts of each ExpeER partner were asked to send an email of validation to the project manager by 27 June. They were asked to use the following format:

I have read and agree to the following ExpeER deliverables : D3.1 "ExpeER data policy and guidelines" and D3.2 "ExpeER Metadata standard for dataset level". I confirm that they comply with the data management of (name of the organization).

Here is the list of ExpeER partners whose confirmation email were received till the submission of this deliverable.

Partner n°	Short name	Country	Confirmation to validate D3.1 D3.2
2	BFW	Austria	yes
3	BGU	Israel	yes
4	Bioforsk	Norway	yes
7	CSIC	Spain	yes
8	DTU	Denmark	yes
10	ERCE	Poland	yes
13	FSU JENA	Germany	yes
14	MTA ÖK	Hungary	yes
15	Imperial	UK	only for D3.1 so far
17	Jülich	Germany	yes
18	KIT	Germany	yes
21	Rothamsted Research	UK	ask for amendment in D3.1
23	SOTON	UK	yes
24	SYKE	Finland	yes
25	TUM	Germany	yes
26	UA	Belgium	yes
27	UB	Romania	yes
28	UFZ	Germany	yes
29	UHEL	Finland	yes
30	ULUND	Sweden	yes
31	UNITO	Italy	yes
32	UNIUD	Italy	yes
33	UNIVLEEDS	UK	yes
34	UNS	Serbia	yes
36	VUA	The Netherlands	yes