Detecting changes in essential ecosystem and biodiversity properties- towards a Biosphere Atmosphere Change Index: BACI

Deliverable 1.2: Data Management Plan

Project title:	Detecting changes in essential ecosystem and biodiversity properties- towards a Biosphere Atmosphere Change Index
Project Acronym:	BACI
Grant Agreement number:	640176
Main pillar:	Industrial Leadership
Topic:	EO- 1- 2014: New ideas for Earth-relevant space applications
Start date of the project:	1st April 2015
Duration of the project:	48 months
Dissemination level:	Public
Responsible of the deliverable	Miguel Mahecha (MPG) Email: mmahecha@bgc-jena.mpg.de
Date of submission:	30.09.2015

Document History

Version	Date	Author	Description	Approved by Ex. Board	Approved by the GA
01	23.09.15	MPG, CEDA	Initial document	25.09.2015	29.09.2015

Table of Contents

1. Executive Summary

The BACI project (Grant Agreement N°640176) participates in the “pilot action on open access to research data” and is therefore required to develop a “Data Management Plan” in order to specify open data sets. In line with the “H2020 Guidelines on Data Management in Horizon 2020, Version 1.0, December 2013” the purpose of the BACI-Data Management Plan (DMP) is to provide documentation on:

- Types of data product that are expected/delivered,
- Data ownership,
- Intellectual property protection of output data,
- The main elements of the data management policy,
- Approached to curate and preserve data,
- Data standards.

The DMP evolves during the lifespan of the project and should be considered a “living document” (see versioning). This document should be considered in combination with “Section 3 of the BACI Grant Agreement N°640176: *Rights and Obligations related to Background and Results;* Articles *from 23 to 31”.*

The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

The purpose of this data management plan (DMP) is to set up a coherent approach to data issues pertaining to BACI. The data management objectives are to ensure that:

- A high quality and documented data archive is created.
- Appropriate data support is provided to the data users and creators.
- Data are made available to users in a timely fashion.
- Academic credit for data creation is given.
- Conditions of use, access and deposit are clearly stated and do not infringe on the data creators' rights nor contradict the participation of BACI in the H2020 Open Data Pilot.
- Potentially scientifically valuable data are kept for reuse in the long-term and by other disciplines.
- Results based on BACI data can be controlled and validated.
- Risks to project data are managed.
- Appropriate preservation strategies are employed.
- Status of data holdings are visible and support the scientific endeavours of the consortium.

This document is an agreed record of the data management needs and issues within the project. It defines **who is responsible for data management activities** both within data centres and by the data creators.

It **lists the expected data products** and provides a mechanism for recording and agreeing on potential changes. Other data needs and issues are also laid out so that problems can be identified early.

It includes **conditions of use and deposit to clearly express the ownership, responsibilities and rights associated with the data.**

This data management plan has been prepared by the Project Coordinator and the CEDA Data Centre, agreed by the Executive Board and proposed and approved by the General Assembly.

Changes to the document will be recorded and data Management tables will be updated on the internal pages of the BACI webpage to ensure visibility to all project partners and communicated by electronic means to the corresponding consortium bodies.

Version 1 of the Data Management Plan reflects the state of affairs at the early data acquisition phase of the project (month 6, corresponding to September 2015). Much information with the data management tables remains to be filled out in later phase of the project.

All downstream BACI data products of small to intermediate size will be available for download from the BACI data portal hosted by the partner MPG at the facilities of the Max-Planck-Institute for Biogeochemistry.

High Size GB Earth Observation data sets that will be engaged with by multiple international partners within a collaborative environment will be curated by CEDA and its structure for evolving data intensive projects.

Both data structures honours all constituent elements of the H2020 Open Data Pilot – Data Management Plan template, but also incorporates risk management, review procedure and exposes the acquisition/archival/preservation status of project data.

2. Definitions and Acronyms

Associated partner	A person, group, or project that agreed to join the project (without budget) because of common scientific interest in the general objective of BACI and sharing own data or expertise.
BACI	EU project: “Detecting changes in essential ecosystem and biodiversity properties- towards a Biosphere Atmosphere Change Index: BACI”.
Background	Any data, know-how or information, including any rights such as intellectual property rights, that is held by the beneficiaries before they acceded to the Agreement, and is needed to implement the action or exploit the results.
Carbo-Extreme	FP7 project presented at http://www.carbo-extreme.eu/
CCSDS	Consultative Committee on Space Data Standards: multi-national forum for the development of communications and data systems standards for spaceflight.
Data	Every digital object (i.e. observations and model/synthesis results) that is stored in one of the component databases or in the original database, when available, in case of external data and hence accessible to others people.
EBV	“Essential Biodiversity Variable” sensu Pereira et al. (2013).
ECV	“Essential Climate Variable”.
EEV	“Essential Ecosystem Variable” – is treated here as synonym for an “ecosystem functional property. The latter are “quantities that characterize ecosystem processes and responses in an integrated and comparable manner” sensu Reichstein et al. (2014) and Musavi et al. (2015). This product will contain maps of variables derived from other products that cannot be directly monitored, but are essential to understand ecosystem functioning. In particular, we consider that an EEV should integrate elements of EBVs and ECVs.
Data access	The possibility for a registered user to login into the database and get access to the data that can be downloaded.
Data assimilation tools	Incorporate observations into a computer model of a real system.
Data provider	A person (could be also a group or project) that is responsible for a dataset present in the database and that has full rights on these data.
Data use	The possibility for a user to present results based on data downloaded by the project database.
Digital Curation Centre	Centre that provides advice and help to store, manage, protect and share digital research data.
DMP	Data management plan.
DOI	Digital object identification-remains fixed over the lifetime of the document.
EO	Earth observations.
ESA	European Space Agency.
External contributor	A person, group, or project not directly interested to the project goals and activities but offering relevant data available to the project partners or part of the project partners.
FLUXNET data	Regional and global data of micrometeorological tower sites' observations.
GA	Grant Agreement.
General user	A person or agency interested in accessing the data and not included in the categories “data provider”, “project partner”, “associated partner”, “external contributor.”

GEOCARBON	FP7 project presented at http://www.geocarbon.net/
GHG-Europe	FP7 project presented at http://www.ghg-europe.eu/
GWS	Group Work Spaces are portions of storage allocated for particular projects to manage themselves, enabling collaborating scientists to share high-performance disk storage.
H2020 Open Data Pilot	European initiative to deposit research data in research data repositories where data can be accessed, mined, exploited, reproduced and disseminated, free of charge for the user.
HPC environment	High-performance computing environment for parallel processing data analytics or simulations.
LiDAR	Remote sensing technology that measures distance by illuminating a target with a laser and analysing the reflected light.
Metadata	Data that provides information about one or more aspects of the data, like time and date of creation, author, location, etc. The exact requirements metadata associated to BACI data is specified in the DMP
NERC	Natural Environment Research Council.
netCDF	Network Common Data Form- is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.
Open research data	Research data that can be accessed, mined, exploited, reproduced and disseminated, free of charge for the user.
Project partner	Scientific responsible listed in the agreement of the BACI project delegates from him. Project partners receive a specific account in the database.
RDA-Active Data Management Plans IG	Research Data Alliance' Group to support data management planning.
Project Results	Any (tangible or intangible) output of the action such as data, knowledge or information — whatever its form or nature, whether it can be protected or not — that is generated in the action.
TRY data base	Global compilation of 215 plant trait databases with coverage for 80.000 of the ca. 350. 000 existing plant species.

3. Project Information

Project Name	BACI
Abstract	The alarming rate of biodiversity loss and ecosystem transitions make it clear that new strategies are required to sustain functioning of the coupled ecological-societal system. Existing space data archives and data streams from the ESA Sentinels, offer unprecedented opportunities to provide rapid, high quality indicators necessary for informed management of key ecosystem services. Yet, it remains largely unclear how space and ground-based observations can be optimally integrated to generate products required by end user communities (Secretariat of the Convention on Biological Diversity, 2014). By fusing extensive expertise on optical and radar remote sensing, ground data on ecosystem state and function, “big data” scientists, and active participation of user groups, BACI will advance this integration. BACI will translate space data to new variables not directly observable from space but that encode ecosystem functional properties and status metrics. This will empower concepts of “essential biodiversity variables”. Advanced machine learning methods will be employed to reveal new and fundamental relationships between space observations and ecosystem status. BACI will incorporate a wide range of original and downstream data products specifically targeting needs for early-warning systems, including a novel “Biosphere-Atmosphere Change Index”. We will prioritize selected key European and African regions now undergoing massive societal-ecological transformations, offering perspective towards operational assessments. A formal attribution framework will disentangle climate-induced ecosystem changes and socio-economic/ecological transformation processes. Overall, BACI will advance usage of European space data to monitor relevant vegetation traits, status, and ecosystem functioning. By capitalizing on existing datasets, we will prototype new algorithms to rapidly implement these metrics and thus space-to-ground integration of the new ESA Sentinels.
Grant Reference(s)	640176
Coordinator	The Max Planck Society (MPG) via the Max Planck Institute for Biogeochemistry, Jena, Germany. Person in charge: Miguel Mahecha
Start/End Dates	01.04.2015 - 31.03.2019

4. Organisation

Nominated Data Centres[1]	• CEMS academic archive (formerly NERC Earth Observation Data Centre) and/or • Max Planck Gesellschaft zur Förderung der Wissenschaften e.V..
Data Centre Contacts	For CEMS: • Esther Conway (esther.conway@stfc.ac.uk), For MPG: • Gerhard Bönisch (gboenisch@bgc-jena.mpg.de) • Ulrich Weber (uweber@bgc-jena.mpg.de)
Project Data Contact	Miguel Mahecha (mmahecha@bgc-jena.mpg.de)
Any other team members with responsibility for data	Victoria Bennett (victoria.bennett@stfc.ac.uk) Mathias Disney (mathias.disney@ucl.ac.uk) David Frank (david.frank@wsl.ch) Jens Kattge (jkattge@bgc-jena.mpg.de) Miguel Mahecha (mmahecha@bgc-jena.mpg.de) Dario Papale (darpap@unitus.it)

[1] Depending on the data set properties and data access requirements.

5. Formal Commitments to Data Management

Ownership of results

Project's results (outputs of the action such as data, knowledge or information) will be owned by the beneficiary that generates them.

In case two or more beneficiaries will own results jointly because:

a) they generated them jointly and-or
b) it is not possible to:
   - establish the respective contribution of each beneficiary
   - separate them for the purpose of applying for, obtaining or maintaining their protection

The joint owners must agree (in writing) on the allocation and terms of exercise of their joint ownership ('**joint ownership agreement**') considering the conditions stated in Article 26.2 of the BACI GA.

Open access

Each beneficiary ensures open access (free of charge online access for any user) to all peer reviewed scientific publications relating to its results and *digital research data generated in the action*.

Associated with this, also research data are free: All data generated should be published under the most open policy possible i.e. the BACI data policy will be based on the **Data licence Germany – attribution – version 2.0: https://www.govdata.de/dl-de/by-2-0**

Data licence Germany – attribution – version 2.0

(1) Any use will be permitted provided it fulfils the requirements of this "Data licence Germany – attribution – Version 2.0".

The data and meta-data provided may, for commercial and non-commercial use, in particular

1. *be copied, printed, presented, altered, processed and transmitted to third parties;*
2. *be merged with own data and with the data of others and be combined to form new and independent datasets;*
3. *be integrated in internal and external business processes, products and applications in public and non-public electronic networks.*

(2) The user must ensure that the source note contains the following information:

1. *the name of the provider,*
2. *the annotation "Data licence Germany – attribution – Version 2.0" or "dl-de/by-2-0" referring to the licence text available at www.govdata.de/dl-de/by-2-0, and*
3. *a reference to the dataset (URI).*

This applies only if the entity keeping the data provides the pieces of information 1-3 for the source note.

(3) Changes, editing, new designs or other amendments must be marked as such in the source note.

6. BACI data use policy

BACI has developed a data use policy that needs to be accepted by all data users. In particular, this data policy takes the BACI participation in the H2020 pilot action on open access to research data into consideration. The following text will be available online; all data users are requested to approve this data policy previous to downloading any BACI data product; the text must be distributed along with the table containing the relevant definitions.

Preamble:

BACI partners aim to be as open as possible in terms of data access and data sharing as requested by the pilot action of on open access to research data by the EU. Hence the consortium has chosen the most open data policy possible (see below).

Data sharing approach trusts in the principle that sharing data opens more possibilities for collaborations and increases the potential for generating more interesting results thanks to the involvement of groups with different backgrounds, experiences and interests.

The BACI consortium may include data from project-external groups in its data repositories that could be interested or not in scientific collaborations. Data not generated in the context of BACI have to be treated according to their respective data policy (if no other explicit agreement with BACI has been established).

The BACI consortium aims at fostering “data providers” to have the opportunity to participate in scientific activities and discoveries based on their products. Ideally data providers have the option to be involved in the paper writing process wherever justified i.e. adhering to the common rules of good scientific practice. “Data providers” should have the priority in case of conflicting interests, when for example too similar analyses are planned or ongoing.

The BACI consortium follows the idea that “general users” rather subscribe to a data set than simply download it. The **registration is not intended to limit data use**, rather registration of General users guarantees:

- Having a means to contact general users in cases where data errors or processing bugs have been identified or important updates of the data have been released.
- “Data providers” can receive digest on data subscripts/downloads in order to understand the level of interest in their sites and the project coordinator and database manager evaluate the interest in the project data.
- “Data providers” have the opportunity to get in contact with “general users” to express their interest in becoming involved in the scientific explorations of their products.

This approach requires data download to be tracked and the user information to be stored by the BACI consortium.

Data access and use regulation:
The data provided by BACI are licensed by the *"Data licence Germany – attribution – Version 2.0"* **https://www.govdata.de/dl-de/by-2-0**

Persistent data identification:

BACI intends to guarantee a persistent identification of the data generated by the project. Hence, all project data contained in the data bases should be tagged by a digital object identifier (DOI). The partner MPG provides the DOIs.

7. Project Resources

BACI has two means for sharing data: The GWS at CEDA and the BACI data exchange portal. Section 8 specifies for each output data set a means to be distributed.

Collaborative Infrastructure

Group Work Spaces (GWS) are **portions of storage allocated for particular projects** to manage themselves, enabling collaborating scientists to share high-performance disk storage.

Users can pull data from external sites to shared storage, process and analyze their data, and where allowed, exploit data available from other group workspaces and from the CEDA archive.

GWSs are often provided in conjunction with project-specific computing resources, configured and deployed as virtual machines in the JASMIN infrastructure. The **BACI** virtual machine has been provided with read-only access to the CEDA Archive.

It is important to understand that these Group Work Spaces are not the same as the CEDA archive: GWS data are not curated for long-term storage. Currently this data set will be deleted when the GWS is closed post project. If under review data is considered a long term asset for transfer to the CEMS academic archive/Alternate Archive; additional data management tables will be added here

Data in a GWS are the responsibility of the designated manager **Mat Disney (UCL)**. A system is currently being deployed to enable GWS Managers to transfer data to/from tape and make best use of their allocation of online disk storage. The BACI project has been provided with **50 Tb of Elastic Tape.** Summary of resources supplied to BACI partners below

Account/Service	Academic Partner	Industrial Partner
CEDA Account	Yes	Yes
Login Account	CEMS account	Industrial Partner account
CEDA Data	Apply to CEDA subject to IP restrictions	Apply to CEDA subject to IP restrictions
JASMIN ANALYSIS Platform	Comes as standard with CEMS login	No (due to licensing Restrictions)

Account/Service	Academic Partner	Industrial Partner
LOTUS	Comes as standard with CEMS login and BACI VM	Available with BACI VM
BACI GWS	Upon Application; BACI set up	Upon Application; Dedicated BACI VM has been set up
Virtual Machine	Upon Application: Dedicated BACI VM has been set up	Upon Application: Dedicated BACI VM has been set up
High Performance Transfer	Upon Application	Upon Application

Data Management Resources and Scientist Support

The CEDA team will provide data scientist support and advice in following areas

- Data structures
- File formats
- Naming convention
- Metadata

Full details of partner man months are detailed in BACI grant agreement, post project resources for long term data archival will be released subject to review procedures against NERC data value checklist for data sets to be archived with CEDA.

BACI data exchange portal

In the context of the preceding FP7 projects CARBO-Extreme and GEOCARBON, the Max-Planck Institute for Biogeochemistry developed a web based data portal through which data can be shared project internally or externally by researchers from different locations.

https://www.bgc-jena.mpg.de/geodb/projects/Home.php

BACI builds on this experience and is currently in the process of constructing an analogous data platform for curating and sharing observations relevant to the internal project synthesis activities and relevant to the general public. Technically, the task includes generating an associated meta-dataset per dataset, and guaranteeing sharing via a visible data portal. The Metadata (i.e. the description of the dataset) are derived from the netCDF headers of the data and visible to the interested public through the data portal website and to a broader audience via GEOSS (the central global web portal for finding global geodata of any kind: http://www.geoportal.org). The latter is possible, given that we will export the metadata to an XML file that is automatically imported into GEOSS on a daily basis. The elaborated data portal allows implementing dataset specific usage terms, all fulfilling the requirements of the “guidelines on Data Management in Horizon 2020”. The data portal is structured such that the usage of the data is traceable, i.e. the concept is “subscribe to a dataset” that allows to reach the user community in case of updates, potential corrections, extensions etc. Hence, we can also guarantee a transparent versioning and respecting the intellectual properties of the data owners.

The Max-Planck Society guarantees to maintain the data platform for least 10 years (guarantee of reproducible science) with the intention to identify a permanent archive.

8. Output Data

Remarks on the formatting of the BACI output data files:

- Data should be provided as netCDF-Files.
- The file format has to stick to CMIP5 conventions wherever possible (cf. for a description http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_output_metadata_requirements.pdf ... but obviously we ignore the stuff that is specific to the CMIP5 experiments); this approach might not be perfectly suitable for all data sets and exceptions are normal. However, we should intend to minimize the risk of errors in the synthesis-phase of the components/project where typically, unit conversion etc. may causes trouble. Some basics requirements that must be fulfilled are e.g. to have
    - Global attributes (examples)
        - institution = "XY Institute for XYZ";
        - :institute_id = "XYZ";
        - :data_id = xyz_data";
        - :contact = "Data Producer (dummy@xxx.z)" ;
        - :history = "Output from original paper… / Data manger X converted the data to comply with BACI requirements." ;
        - :references = "Your citation " ;
        - :frequency = "mon" ;
        - :creation_date = "2010-04-21T21:05:23Z" ;
    - Ordering of dimensions
- Data should be ideally in CMIP5 conformal units and using the corresponding long and short names; at http://cmip-pcmdi.llnl.gov/cmip5/output_req.html you find a spread sheet called "standard_output" where a wide range of variables is described. Otherwise please use some clean and non-confoundable namings of your variables.
- The meta-information (ncdump -h) will be visible to everybody - but the actual data only to "General users" etc. ....see the data policy agreement on this issue.
- Please name the files as BACI_VariableName_institute_id_VersionNr.nc

Data Set: Ecosystem functional properties (Deliverable: 4.2)

Product Name	Ecosystem functional properties
Data Description	Maps of ecosystem functional properties sensu Reichstein et al. (2014) and Musavi et al. (2015) are "quantities that characterize ecosystem processes and responses in an integrated and comparable manner". This product will contain maps of variables derived from other products that cannot be directly monitored, but are essential to understand ecosystem functioning. In particular, we consider the variables • Maximum Light Use Efficiency • Inherent Water Use Efficiency • Basal Ecosystem Respiration at a given rate of photosynthesis • rate of evapotranspiration decay under drought • …*tbd*
Strategic Value and Target Community	Data of this kind are typically not provided to the scientific community. However, these data sets are very informative for assessing ecosystem functioning under current conditions and in relation to biodiversity patterns. We anticipate that variables of this kind could bridge the concepts of EBVs and ECVs by a new category, which is "Essential Ecosystem Variable".
Data Creator (Scientist in charge)	WP4: Martin Jung, Dario Papale, Markus Reichstein
Data of publishing…	*tbd*
Institution in charge	MPG
Location on GWS or MPG data server	*tbd*
Version	*tbd*
References to User Support Materials and Technical Specification Documents	Brief description of document and links to documents in **XXXXX** document repository
Relevant Software	none
Product format	netCDF
Data temporal resolution	Annually or integrated over multiple years
Data spatial gridding	Minimum 0.25, ideally at much higher resolution – which is, however, not the target
Data coverage: temporal	Level 1: 2000-2015
Data coverage: spatial	Level 1: Global
DOI	*tbd*
Size GB	Expected: min. 4 GB
Number of files	1
Licence for Distribution	BACI data policy
Related Data Sets	FLUXCOM products
Comments	*tbd*
Archival plan	*tbd*
Product Name	

Data Set D: Upscaled Tree Ring Width (Deliverable: 4.3)

Product Name	Upscaled Tree Ring Width
Data Description	Tree ring width are sampled locally. This product explores the spatiotemporal correlation with RS data to produce a coherent explicit space/time data stream of tree ring width data at annual scales.
Strategic Value and Target Community	Observed tree ring width integrates a substantial part of biosphere responses to environmental conditions. Hence, these data are typically used to reconstruct impacts of climate anomalies on land ecosystems. However, the data are scattered in space and typically don't allow for interpolated assessments. Modern machine learning will allow for deriving the hypothetical responses all over the place. The approach of strategic value for the community for assessing the impacts of extremes on net productivity and as independent reference for evaluating land surface models.
Data Creator (Scientist in charge)	WP4: David Frank, Flurin Babst, Martin Jung, Miguel Mahecha
Data of publishing…	*tbd*
Institution in charge	WSL
Location on GWS or MPG data server	*tbd*
Version	*tbd*
References to User Support Materials and Technical Specification Documents	Brief description of document and links to documents in **XXXXX** document repository
Relevant Software	none
Product format	netCDF
Data temporal resolution	Annual values
Data spatial gridding	Minimum 0.25 to 0.5, ideally at much higher resolution – which is, however, not the target
Data coverage: temporal	Level 1: 1980-2014; coverage towards present becomes increasingly limited; the significant potential to extend dataset back in time will be explored.
Data coverage: spatial	Level 1: Europe; Tests for NH extra-tropics will be performed
DOI	*tbd*
Size GB	Expected: min. 4 GB
Number of files	1
Licence for Distribution	BACI data policy
Related Data Sets	TRY
Comments	*tbd*
Archival plan	*tbd*
Product Name	

Data Set: WP4 Sub-daily Biosphere-Atmosphere Fluxes (Deliverable: 4.4)

Product Name	Sub-daily Biosphere-Atmosphere fluxes
Data Description	Information content of eddy-covariance stations is merged with remote sensing data. The final product is containing empirically predictions of land-atmosphere gross and net exchange fluxes of CO_2, H_2O and energy that resolve the diurnal cycle.
Strategic Value and Target Community	The data set is unprecedented in the scientific community. So far, upscaled fluxes were only available at monthly temporal resolutions. Latest efforts (i.e. the FLUXCOM initiative) is aiming at providing high spatial resolutions of the fluxes, but with temporal resolution of 8 days, or having a coarse spatial resolution of 0.5° where the temporal resolution is daily. However, no subdaily data set could be provided so far which would be crucial for further developing land-surface models and investigating e.g. effects of climate extremes on ecosystem functioning. High temporal resolutions of fluxes would also allow for deriving ecosystem functional properties more accurately, which could become valuable complements empowering the concept of "essential biodiversity variables".
Data Creator (Scientist in charge)	WP4: Martin Jung, Markus Reichstein, Paul Bodesheim, Dario Papale, Gianluca Tramontana
Data of publishing…	*tbd*
Institution in charge	MPG
Location on GWS or MPG data server	*tbd*
Version	*tbd*
References to User Support Materials and Technical Specification Documents	Brief description of document and links to documents in **XXXXX** document repository
Relevant Software	none
Product format	netCDF
Data temporal resolution	Subdaily, e.g. ½ hourly to 3-hourly (still under investigation)
Data spatial gridding	Minimum 0.25, ideally at much higher resolution – which is, however, not the target
Data coverage: temporal	Level 1: 2000-2015
Data coverage: spatial	Level 1: Global
DOI	*tbd*
Size GB	Expected: min. 46 TB
Number of files	1
Licence for Distribution	BACI data policy
Related Data Sets	FLUXCOM products
Comments	*tbd*
Archival plan	*tbd*

Data Set: Plant Trait Phenologies (Deliverable: 4.5)

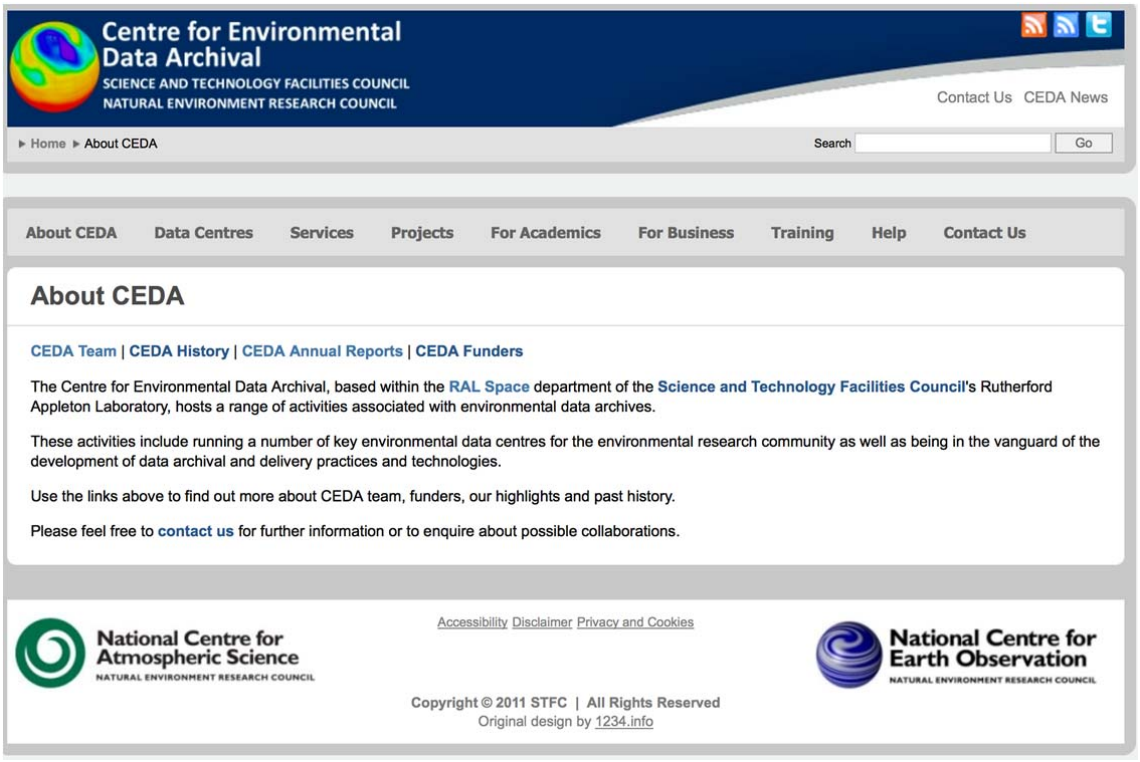
Product Name	Plant Trait Phenologies
Data Description	The plant functional traits: • Specific Leaf Area • Leaf nitrogen content per dry mass and per area (LeafN) are upscaled to regional and continental scales by intersecting in-situ data with RS products. The data should take the temporal dynamics of the spatial patterns into consideration.
Strategic Value and Target Community	Plant traits are considered EBVs and deserve much attention. Plant functional are key determinants of ecosystem functioning. However, the relevance of the trait values is depending on vegetation activity. Hence, one has to interlink plant traits with plant phenology to derive a time-dynamic map when certain plant traits are active and fully develop in order to understand effects on the land-surface interactions.
Data Creator (Scientist in charge)	WP4: Jens Kattge, Miguel Mahecha
Data of publishing…	*tbd*
Institution in charge	MPG
Location on GWS or MPG data server	*tbd*
Version	*tbd*
References to User Support Materials and Technical Specification Documents	Brief description of document and links to documents in **XXXXX** document repository
Relevant Software	none
Product format	netCDF
Data temporal resolution	Sub-annual values
Data spatial gridding	Minimum 0.25, ideally at much higher resolution – which is, however, not the target
Data coverage: temporal	Level 1: 2000-2015
Data coverage: spatial	Level 1: Europe
DOI	*tbd*
Size GB	Expected: min. 48 GB
Number of files	1
Licence for Distribution	BACI data policy
Related Data Sets	TRY
Comments	*tbd*
Archival plan	*tbd*
Product Name	

9. Appendices

Appendix 1: CEDA

The Centre for Environmental Data Archival, based within the RAL Space department of the Science and Technology Facilities Council's Rutherford Appleton Laboratory, hosts a range of activities associated with environmental data archives.

These activities include running a number of key environmental data centres for the environmental research community as well as being in the vanguard of the development of data archival and delivery practices and technologies.



The Centre for Environmental Data Archival was established primarily to facilitate and curate both datasets needed for academic research and datasets produced by such research.

CEDA is primarily funded by the Natural Environment Research Council, but it is also funded by DECC, and the missions differ slightly.

Further information on CEDA can be found here: www.ceda.ac.uk

Appendix 2: NERC data value check list

In order to preserve and maintain access to the BACI data product and maximise the impact of the impact of the project in the long term NEODC must release funds order to do this. Victoria Bennett head of Earth Observation will appraise BACI output in accordance with the NERC data value checklist to approve resources.

Data Value Checklist - Purpose and scope

The Data Value Checklist aims to identify which data should be considered for accession to the NERC Environmental Data Centres. The individual Data Centres' collections policies (both written and informal) will assist in deciding which Data Centre is the most appropriate place to deposit the data depending upon the science area and type of data collected.

The Data Value Checklist is intended to be used in the following circumstances:

a) When preparing a full Data Management Plan to assist Data Centres and Principal Investigators in determining the likely long term value of the data to be produced by a project.

b) Upon receipt of the data for deposit with the Data Centres, to assess their quality, integrity, originality and content

This will ensure that data included in the NERC Data Centre collections are of long term value to the scientific community.

Further information on the NERC data value checklist can be found here: http://www.nerc.ac.uk/research/sites/data/policy/data-value-checklist.pdf

Appendix 3:CF- NetCDF Standards

Data producers on the BACI project will be supported by CEDA staff to produce CF-NetCDF compliant data products.

NetCDF (network Common Data Form) is an interface for array-orientated data access and a library that provides an implementation of that interface. Many groups have adopted netCDF as a standard way to represent their scientific data. The netCDF software was developed at the Unidata Program Center in Boulder Colorado USA.

The NetCDF format has a wide range of reasons why it is one of CEDA's recommended formats, including:

- being extensively used within the terrestrial, atmospheric and oceanic science communities.
- being a portable self-describing binary data format.
- it is network-transparent, meaning that it can be accessed by computers that store integers, characters and floating-point numbers in different ways.
- it provides direct-access: a small subset of a large dataset may be accessed efficiently, without first reading through all the preceding data.
- it is appendable: data can be appended to a netCDF dataset along one dimension without copying the dataset or redefining its structure.
- datasets can be read and written in a number of languages, these include C, C++, FORTRAN, Julia, IDL, Matlab, R, Python, Perl, and Java.
- the different language implementations are freely available from the UNIDATA ftp area or from other mirror sites or as toolboxes from the respective language community repositories.
- several graphics packages support netCDF input, making it very easy to display and analyse netCDF datasets. For instance FERRET and CDAT provide both command line and graphical user interfaces for displaying and analysing gridded data.
- netCDF is completely and methodically documented in UNIDATA's NetCDF User's Guide.
- several groups have defined conventions for netCDF files, to enable the exchange of data. CEDA has adopted the Climate and Forecasting (CF) conventions for netCDF data.

CEDA supports and strongly recommends the compliance with the Climate and Forecast (CF) Metadata Convention.

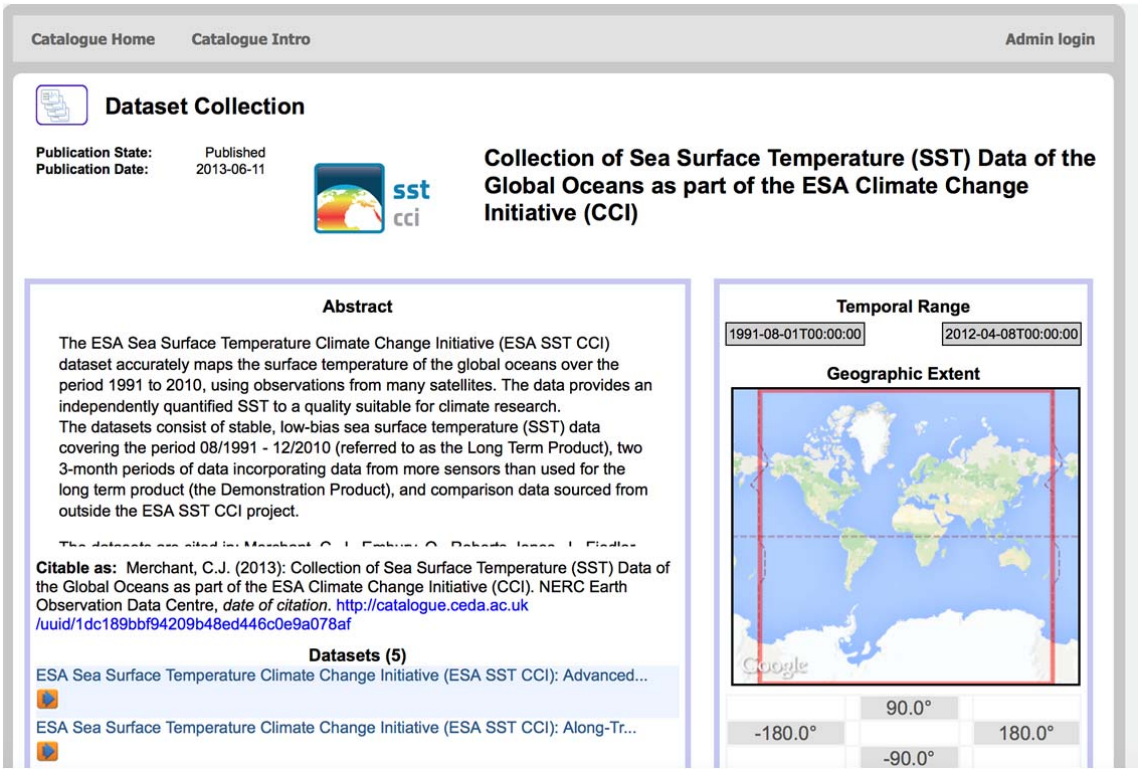
- CF conventions are guidelines and recommendations as to where to put information within a netCDF file, and they provide advice as to what type of information you might want to include. CF conventions allow the creator of the dataset to include information about the data and the dataset itself (metadata) in a structured way, which makes it easier for other users to retrieve the information. *Global attributes* describe the general properties and origins of the dataset while *local attributes* are used to characterise the recorded variables.
- CEDA provides Information on the CF Convention, illustrated by some examples.
- All data hosted by the MPI-BGC at the end will use the netCDF-metadata as descriptors of the files.

Further information on the CF – NetCDF standard can be found here: http://www.ceda.ac.uk/help/users-guide/file-formats/netcdf/

Appendix 4: CEDA - MOLES

The MOLES3 CEDA catalogue provides support for organising information about data, and for user navigation around data holdings. “CEDA-MOLES”, also supports data management functions for the Centre for Environmental Data Archival, CEDA.

The latest MOLES3 has enhanced data provenance, for further structured information to support ISO19115 discovery metadata export (for EU INSPIRE compliance), and provides appropriate fixed landing pages for Digital Object Identifiers (DOIs) in the presence of evolving datasets. It compliant with ISO19156 Observations and Measurements (O&M) and provides a standardised framework for organising information about BACI project data. Creating a MOLES 3 record and publishing data through CEDA also exposes BACI project data to other federated discovery mechanisms such as the NERC data portal. The MOLES CEDA catalogue can be accessed here: http://www.ceda.ac.uk/services/dataSearch/



ESA SST CCI catalogue record

Appendix 5: Citation and DOI

Data form the BACI project held in the CEDA archives to be assigned a Digital Object Identifier - or DOI. A DOI enables scientists to cite datasets in the same manner as a scientific journal article, thereby enabling credit to be assigned to the dataset creators, and ensuring the discoverability, permanence and stability of the dataset. For further details on DOIs please see the NERC DOI webpage, or refer to the "Getting a citation for your dat (DOIs)" page in the "Guide to data providers"

A typical citation to use for data with a DOI is:

Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [S. A. Callaghan, J. Waight, C. J. Walden, J. Agnew and S. Ventouras]. GBS 20.7GHz slant path radio propagation measurements, Sparsholt site, [Internet]. British Atmospheric Data Centre, 2003-2005, 1st April 2011, doi:10.5285/E8F43A51-0198-4323-A926-FE69225D57DD

The DOI landing page will additionally:

a) display the EU emblem and
b) include the following text:

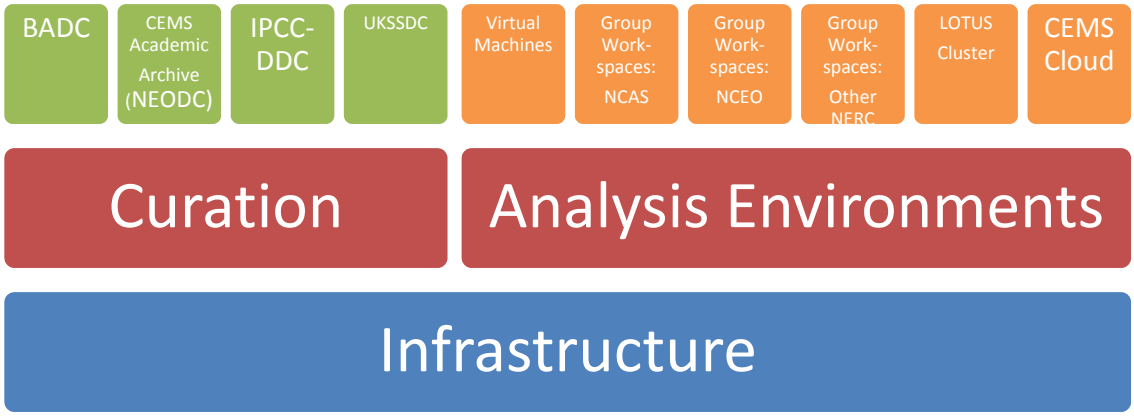
"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 640176".

Scientist citing utilising and citing data will additionally be encouraged to acknowledge H2020 funding under grant agreement No 640176.

Appendix 6: JASMIN/CEMS

JASMIN provides the UK and European climate and earth-system science communities with an efficient data analysis environment. Many datasets, particularly EO satellite data, are too big to be easily shipped around: JASMIN enables the BACI project to bring their processing to the data.

The BACI project will be provided with self-managing group workspaces and dedicated virtual machine, enabling models and algorithms to be evaluated alongside curated archive data, and for data to be shared and evaluated before being deposited in the permanent archive. Further detail on the JASMIN CEMS infrastructure can be found here www.jasmin.ac.uk



JASMIN Infrastructure

Appendix 7: Software and GNU GPL Licence

Software code produced on the project in a dedicate svn accessible at http://proj.badc.rl.ac.uk/XXXX/browser

This svn repository will be maintained 5 years post project after which it will be reviewed in terms of value to data archived at CEDA and the need of the broader scientific community.

At the consortiums discretion the software code will be made publically available under the GPL licence and can be associate with archived data through the CEDA catalogue. The GNU General Public License is a free, copyleft license for software and other kinds of works. More information on the GNU GPL licence can be found here http://www.gnu.org/licenses/gpl-3.0.en.html .

Appendix 8: OGL licence

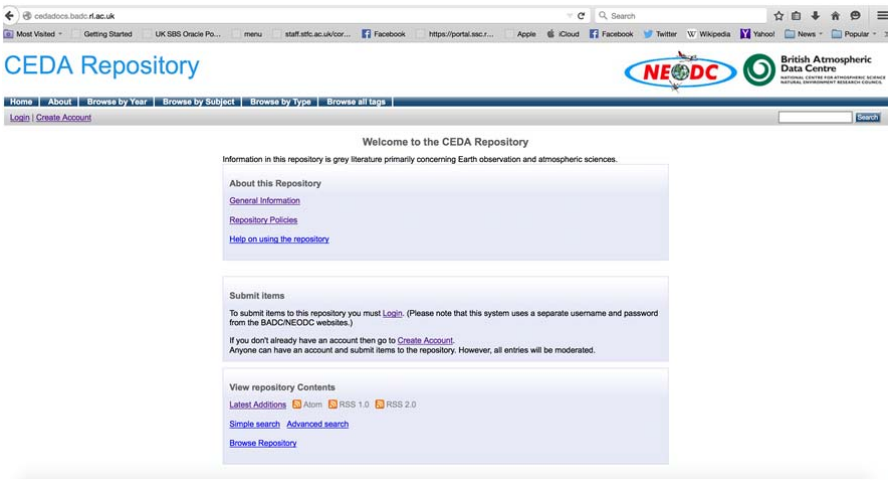
The BACI project will use version 3.0 of the Open Government Licence for data products approved for archival within the CEDA repository. The Controller of HMSO may, from time to time, issue new versions of the Open Government Licence. If you are already using Information under a previous version of the Open Government Licence, the terms of that licence will continue to apply.

These terms are compatible with the Creative Commons Attribution License 4.0 and the Open Data Commons Attribution License, both of which license copyright and database rights. This means that when the Information is adapted and licensed under either of those licences, you automatically satisfy the conditions of the OGL when you comply with the other licence. The OGLv3.0 is Open Definition compliant.

Further context, best practice and guidance can be found in the UK Government Licensing Framework section on The National Archives website.

Appendix 9: CEDA Document Repository

The CEDA Document Repository will act as the permanent repository for public facing documentation, scientific publications and presentations, which will also be linked to the BACI website. The CEDA repository is for grey literature primarily concerning Earth observation and the atmospheric sciences with all content being publically available. Project partners are encouraged to deposit documentation and presentation in PDF format where possible



CEDA Document Repository: http://cedadocs.badc.rl.ac.uk/

The Preservation Policy of the CEDA Documents Repository is as follows

1. Items will be retained indefinitely.
2. The repository will try to ensure continued readability and accessibility.
    - Items will be migrated to new file formats where necessary.
    - It may not be possible to guarantee the readability of some unusual file formats.
3. The repository regularly backs up its files according to current best practice.
4. The original bit stream is retained for all items, in addition to any upgraded formats.
5. Items may not normally be removed from the repository.
6. Acceptable reasons for withdrawal include:
    - Proven copyright violation or plagiarism
    - Legal requirements and proven violations
    - National Security
7. Withdrawn items are not deleted *per se*, but are removed from public view.
8. Withdrawn items' identifiers/URLs are retained indefinitely.
9. URLs will continue to point to 'tombstone' citations, to avoid broken links and to retain item histories.
10. The metadata of withdrawn items will not be searchable.
11. In the event of the repository being closed down, the database will be transferred to another appropriate archive.

Appendix 10: Data Deposit Agreement with CEDA

General Deposit Conditions for Data

Data creators or custodians of data held on project group workspaces will be required to agree to the following data deposit conditions before the data are added/transferred to the archive.

1. *The depositor confirms that he/she is the owner of the data and/or has the right to deposit the data in a NERC archive.*
2. *Ownership of the data remains with the data creator.*
3. *NERC reserves the right to store the data, and make the data available under appropriate Conditions of Use that are consistent with NERC data policy.*
4. *The depositor grants NERC permission to, without changing content, translate the data to any medium or format for the purpose of future preservation and accessibility.*

.

Appendix 11: BACI Data Exchange Portal at MPI-BGC

The Max-Planck Institute for Biogeochemistry operates a data portal for several projects. These include GEOCARBON and CARBO-Extreme.

The URL is https://www.bgc-jena.mpg.de/geodb/projects/Home.php.

The figure below shows the BACI project in the data portal.