# ÉCLAIRE Data Management Plan

#### 

### Contents

* 1. Introduction
  2. Types of data generated by ÉCLAIRE
  3. Data management infrastructure, data centres, web portal and operations
  4. Database contents
  5. Formats
  6. Metadata
  7. Data file names
  8. Data submission
  9. Data validation and quality checking
  10. How to access data
  11. Support to the ÉCLAIRE researchers
  12. List of acronyms

Annex 1. Time-sequenced table of deliverables from work packages

Annex 2. The Data Exchange Support Group

Annex 3. Data Quality Assurance and Data Quality Control Protocol

Annex 4. ÉCLAIRE Conditions of Use

**1. Introduction**

This document sets the way to implement principles developed in the ÉCLAIRE Data Policy, to which it refers. It is intended to be a working document and some of its sections will be updated in the course of the project.

**2. Types of data generated by ÉCLAIRE**

Data gathered, compiled, collected or produced by the science components of the project are listed in the Measurement and Modelling Protocols which are being compiled during the first 12 months of the project. Each science component will be involved in some of the data activities summarised in Table 1, and will deliver data to one of the two DCs as listed in the last column of Table 1.

Data collected under Activities (1) and (2) (see Table 1) can present different levels of processing. Raw data, i.e. source measurements in the form that they have when they are first produced (more detailed *ad hoc* descriptions of raw data are given in the Measurement Protocols), will in general not be stored at the DCs but it is each PI’s responsibility to ensure that they are stored safely with the relevant processing software or, alternatively, with documentation on retrieval algorithms, at least for the retention period as defined in the ÉCLAIRE Data Policy.

Processed data, i.e. observation data that have been subject to some treatment or formatting, or data derived from these, will be stored at the DCs and made available by them to the ÉCLAIRE community.

TABLE 1 – TYPES OF DATA GENERATED BY ÉCLAIRE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Science component/work package | | Gathering of historic and external data; literature reviews | ÉCLAIRE measurements | Data synthesis/ processing | Model output | DC |
| C1 | WP1 | ✔ | ✔ | ✔ |  | CEH |
| WP2 |  | ✔ | ✔ |  | CEH |
| WP3 | ✔ |  | ✔ |  | JRC |
| WP4 | ✔ |  | ✔ |  | JRC |
| C2 | WP5 |  |  |  | ✔ | JRC |
| WP6 | ✔ |  | ✔ | ✔ | JRC |
| WP7 |  |  |  | ✔ | JRC |
| WP8 | ✔ |  |  | ✔ | JRC |
| C3 | WP9 | ✔ |  | ✔ |  | CEH |
| WP10 | ✔ | ✔ | ✔ |  | CEH |
| WP11 |  | ✔ | ✔ |  | CEH |
| WP12 |  |  | ✔ | ✔ | CEH |
| WP13 | ✔ |  |  | ✔ | CEH |
| C4 | WP14 |  |  |  | ✔ | JRC/  external |
| WP15 |  |  |  | ✔ | JRC |
| WP16 |  |  | ✔ | ✔ | JRC |
| WP17 | ✔ |  | ✔ | ✔ | JRC |
| C5 | WP18 | ✔ |  | ✔ | ✔ | From JRC |
| WP19 |  |  | ✔ | ✔ | From JRC |
| WP20 |  |  | ✔ | ✔ | From JRC |

**3. Data management infrastructure, data centres, web portal and operations**

The ÉCLAIRE science activity is divided into six components that will deliver different types of data to the ÉCLAIRE distributed database located at two dedicated data centres (DCs), namely CEH Edinburgh UK (managed by the CEH Environmental Information Data centre (EIDC) at Lancaster):

<http://eclairedata.ceh.ac.uk/page/login.aspx>

and the Joint Research Centre (JRC) at Ispra, Italy (AFOLU DATA Portal): <http://afoludata.jrc.ec.europa.eu/index.php/dataset>).

Table 1 shows which DC is allocated to each science component. The DC’s responsibilities include storing, checking for completeness, maintenance and distribution of ÉCLAIRE data. Data centres liaise with the different science components through the Work Package Data Managers (DMs), who organise the data handling and supervise data submission within their component, but data can be submitted to the DCs directly by the individual investigators.

Each science work package has a Data Manager (DM), whose responsibilities include liaison with the scientists and the Data Management Committee (see below), overseeing data collection and data quality checks, providing support to investigators in issues related to data formatting and submission.

The ÉCLAIRE Data Management Committee (DMC) consists of the work package DMs plus IT support and consultancy participants. The DMC co-ordinates and supervises all data management activities, ensures that the Data Policy is applied and makes decisions regarding its implementation. The DMC produces the ÉCLAIRE Data Management Plan where details of this implementation are given. The DMC members confer at regular intervals. The DMC is composed of

* + - seven data managers,
    - a representative of each DC,
    - the web portal manager,
    - the ÉCLAIRE Scientific Project Manager,
    - the ÉCLAIRE Co-ordinator or his representative.

The DMC issues a formal written annual report to the ESG. The DMC may form Task Forces to aid its work.

A central ÉCLAIRE web portal has been developed at CEH ([http://www.ECLAIRE.eu/](http://www.nitroeurope.eu/)). The portal includes a **Data** thumbnail leading to a page that links to the data centres, where submission deadlines and data related news will be posted. There are links to related sites and to the online support offered by the data centres. The website is the primary source of information on ÉCLAIRE data management issues.

**4. Database contents**

The ÉCLAIRE databases will host all processed observational data produced by the project, together with any documentation pertaining to the data.

The ÉCLAIRE CEH database will also host data resulting from plot scale modelling (C3). Model development is documented by individual partners using version control tools like Subversion. This tool also allows “freezing” of model versions. Frozen model source code, documentation, selected significant model simulation results as well as corresponding model input data and drivers will be stored first by the individual partners and later in the appropriate database.

In recognition that raw unprocessed data represent a potentially valuable source of future science developments, including possible revision of the processed data, the Principal Investigators (PIs) agree to ensure that they are stored safely with the relevant processing software or, alternatively, with documentation on retrieval algorithms, at least for the retention period as defined below. Although not necessarily stored at one of the DCs, they should be documented on the web portal. The DMC will examine the fate of raw data on a case per case basis and will ensure that in all cases they are kept for the long term in a way allowing future access and advise PIs on questions of raw data storage. A short definition of ‘raw data’ is in section 2 above, and will be detailed in the Measurement and Modelling Protocols.

**5. Formats**

Data hosted by CEH are stored in the form of relational databases (Oracle 11g). Data are supplied in templates and will be uploaded to the CEH database by WP managers, prior to checking and validating through the database web interface by DMs.

Data hosted by JRC will be file-based and no specific restrictions are made on the file format. However, geographic data in Component 5 data files should be based on the projection system Lambert Azimuthal Equal Area (ETRSLAEA, centre of projection: 52N, 10E) as proposed under INSPIRE. Images, text files and model output can be stored in the original format.

**6. Metadata**

Metadata (i.e. data about the data) are a crucial element of a data archive. They allow the data to be searched (“discovery” metadata), read by humans or software, understood, interpreted and used. Metadata includes supporting documentation (collection methods, algorithms, model parameterisations, references, plots, pictures, etc.) which will be stored alongside the data. Source codes will be stored (i) to help the understanding of the stored model output: in this case, it is preferable to store also a set of standard model input; (ii) for transparency: in case of future discrepancies, source code of important output will facilitate understanding how results have been generated.

Metadata for data uploaded to the CEH database will be defined in the Excel data templates and will be stored on the CEH data centre database with the data.

Meta-Data for data uploaded to the JRC database (AFOLU) will be stored there. This database is designed specifically for spatial metadata. Upload of a dataset will be restricted to those data with complete metadata information (based on ISO 19155). To ensure completeness of the metadata, the data portal includes a built-in meta data editor. The metadata format is XML.

**7. Data file names**

A common file name convention provides the database some homogeneity and ease of use.

1. **Excel data files uploaded to the CEH data centre.**

WPNN\_ countrycode\_sitename\_YYYYMMDD\_DESCRIPTOR20\_VNN.xls

Where

WPNN is the 4 characters WP number, eg WP02, WP15

Countrycode is a 2-character code for the country of origin of the data

Sitemname is a 3-character code for the site/lab where the measurements were made/or simulated

YYYYMMDD is the date of the first observation in the dataset

DESCRIPTOR20 is a 20-character bit of text to describe the experiment/measurement/model etc

VNN is a 3-character version number for that dataset, eg V01, V02, V15

Some activities may require a different file-naming convention. This should be discussed with the WP DM.

1. **Files uploaded to the JRC data centre**

No specific rules for data file names exist. File names should be (relatively) short and not contain special characters (blanks, ampersand, …).

**8. Data submission (largely for measurement data)**

**CEH data centre**

For WPs 1,2, 9,10,11, data and the associated meta-data are generated and collated by site/lab managers, and will be entered onto the Excel templates specific for each component. The Excel data templates will be uploaded to the CEH database according to a schedule to be compiled as the project progresses. The Schedule will be included in this Data Management Plan, and will be posted on the ÉCLAIRE data web pages. The Data Managers will support the site managers in this activity, and will check the data further before making it available as “validated data” on the CEH ÉCLAIRE database.

The exchanges of data between work packages are shown in Figure 1. Details of this figure are taken from the Description of Work (DoW) and the PERT diagram, (Figure 2). A table of work-package deliverables with month of delivery is given in Annex 1. These diagrams and table will assist the DMC to support timely data submission.

**JRC data centre**

Data stored in at AFOLU DATA portal are uploaded according to their need (in case of input data) and their availability (in case of results). It is responsibility of activity/work package/component leaders to make sure that all required input data and all relevant results of their activity/work package/component are included at the AFOLU DATA portal in a timely and appropriate manner.

**The Data Exchange Support Group**

Proposed by the Data management Committee and accepted by the Executive Steering Group, a Data Exchange Support Group (DESG) has been formed within WP21 to facilitate the exchange of data between work-packages, in a timely way. Annex 2 in the Data Policy sets out roles and responsibilities of ÉCLAIRE PIs, WP leaders, the DESG and the DMC for timely exchange of data between work-packages. The full DESG declaration of roles, responsibilities and actions is set out in Annex 2

**FIGURE 1 Exchange of data between work-packages.**

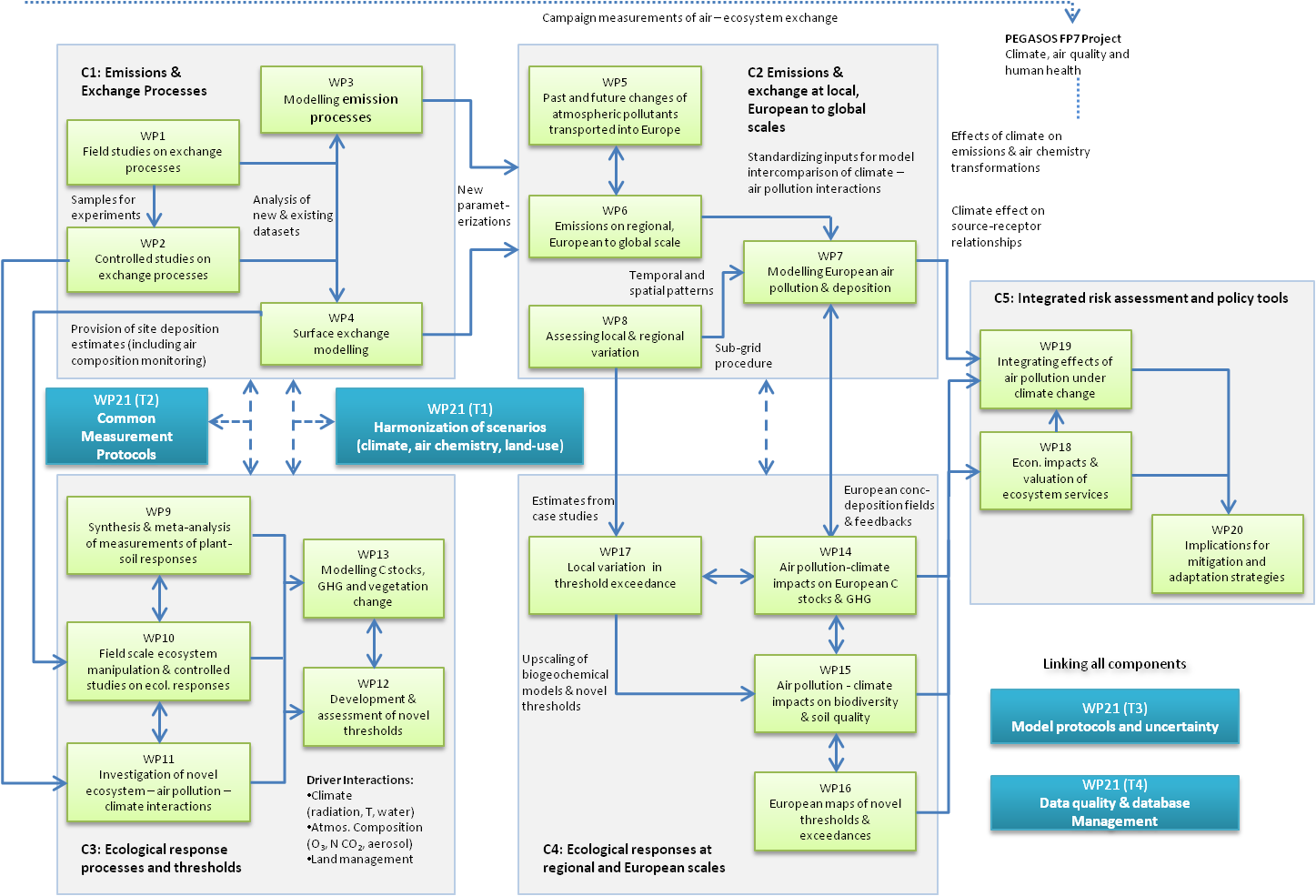
Yellow shading indicates data exchanges illustrated in the PERT diagram (Figure 2). Orange shading indicates data exchanges inferred from WP descriptions in the Description of Work (DoW)

The data are presented by row, eg reading down the left column: WP1 supplies data to WP2, WP3 and WP4 (according to the PERT diagram), and to WP8 (inferred from the DoW).

The numbers in the shaded squares refer to the project months for the WP deliverables.

****

**Figure 2 PERT diagram**

****

**9. Data validation and quality checking**

The Data Quality Assurance and Data Quality Control protocol is appended (Annex 3).

It is the PIs’ responsibility to perform required calibration and validation prior to data submission to the ÉCLAIRE database, and ensure that the data are of the best possible quality and include error estimates and/or flags as defined in the Measurement and Modelling Protocols. It will be one of the roles of the Work Package Data Managers to support and supervise quality checks.

The following details of the software functionality for data checks on upload are adapted from the “Modality Solutions Specification” for the CEH database:

## Excel upload

Each data entry form will support the upload of data from an Excel document, as long as it is in the same format as a provided template.

Excel documents will be uploaded to the web database application, saved and processed.

The Excel import procedure will:

1. Validate the workbook against the expected formed.

2. Work through every form field, reading the data from the uploaded Excel sheet

3. Each field will be checked for field format (numeric, date/time, etc).

4. Each field will then be checked against any validation rules (required / range).

5. Each field will then be checked for uniqueness – the same value uploaded for the same date for the same site

## 6. No Value - Excel documents with values in a field, or within a range of fields, will be able to specify “no value” – the system will allow the customization of the “no value” entry. Proposed options are: “NaN” or “No Value”. When a no value field is uploaded, it is not validated, and is not available for reporting.

7. Finally, the field value will be saved

An automatically generated report will be presented to the user, showing and validation errors (distinquishing between errors that require action and warnings), or successes on import. Finally, a web-based version of the form will be shown to review the data. All Excel documents uploaded will be saved by the application, so the original data will be stored.

Completeness of the data centres’ databases, in terms of completed templates and predefined aspects of template contents, will be checked by the designated DC at the time of submission. Completeness checking will be automated where appropriate. A submission report will be generated automatically, whereby defective datasets, accompanied by error diagnostics, will be provided to the data originators for improvement, and missing data will be signalled.

**The ultimate responsibility for data quality lies with the originators of the data.**

No data checks on the format or content of data sets will be performed on the spatial data stored in the DC at JRC. However, metadata will be checked on completeness. Only datasets whose metadata pass this check will be published.

**10. Access to data**

Access to ÉCLAIRE datasets in general will be restricted to ÉCLAIRE participants and particular collaborators during a retention period of 5 years after the submission due date or 2 years after the project end date, whichever occurs first. However, spatial datasets stored in the JRC data centre will not be restricted.

The ÉCLAIRE web portal ([http://www.ECLAIRE.eu](http://www.nitroeurope.eu)) is the central entry point to access all ÉCLAIRE databases hosted by the data centres (see Section 3). A special “DATA” tab leads to a centralised page providing information about the DMC and data management related issues and news. A web file manager will be built into the portal to enable sharing of documents between ÉCLAIRE partners, templates and other information based on this DMP.

For the general public, information about which data is collected and held in the data centres will be made available through the option to browse metadata describing the database content. In addition to that, contact information for data originators and IPR holders will be provided, in case ÉCLAIRE external parties are interested in accessing specific datasets. Access to ÉCLAIRE datasets will only be granted to external parties on a by-case basis and by the IPR holder of the specific dataset.

After the retention period, the ÉCLAIRE data will be released to the public domain, but the ÉCLAIRE Conditions of Use (see Annex 2 below) will continue to apply. The CEH database has a facility to define different groups of users and individuals for different access rights.

**11. Support to the ÉCLAIRE researchers**

The DCs and the ÉCLAIRE website administrator, advised and assisted by the DMC, will endeavour to provide support to the ÉCLAIRE scientists in all data related issues. These may include the following.

* + - Negotiation, retrieval, provision of 3rd party data (e.g. NRT data to support field campaigns).
    - Set up of data uploading/downloading system (ftp, web).
    - Online documentation (scientific and technical).This will includes a Cookbook, and guides for using the databases. [http://www.ECLAIRE.eu/](http://www.nitroeurope.eu/)
    - Data catalogue and search engine, including links to the database.
    - Data extraction, comparison, visualisation tools.
    - Format conversion (e.g. into Excel spreadsheets).
    - Web based protected workspace that would provide a forum for discussions, collaboration, exchange of preliminary data, etc.

**12. List of acronyms**

CEH Centre for Ecology and Hydrology

CDM Component data manager

DC Data centre(s)

DMC Data management committee

EIP Environmental Informatics Programme

JRC Joint Research Centre

ÉCLAIRE ÉCLAIRE

WP Work Package

|  |  |  |
| --- | --- | --- |
| **ANNEX 1 Table of deliverables from Work Packages, adapted from the**  **ÉCLAIRE secretariat.** | | |
| 1 to 12 month | | |
| 25 to 36 month | | |
| 37 to 48 month | | |
|  |  |  |
| Deliverable number | Deliverable Title | Delivery date mth |
| D9.1 | Progress report on availability of data for use in Activities 3.4 and 3.5 | 6 |
| D13.1 | Finalised list of models for use in C3, and list of data requirements for each model | 6 |
| D14.1 | Synthesis of applicable data on impacts of ozone on photosynthesis, stomatal conductance and plant functioning | 6 |
| D6.1 | Initial dynamic biogenic emissions, based on synthesis of existing work and mainly for test and set-up of ÉCLAIRE atmospheric model experiments WP2.3 and in 4.1. Test for compatibility of file format & establish appropriate resolution for use in atmospheric models | 8 |
| D20.1 | Report from stakeholder workshop | 9 |
| D8.1 | Synthesis report on the different local scale models dealing with atmosphere-biosphere exchange and their relevance for describing the climate change / air pollution interactions | 12 |
| D9.2 | First phase database for use in initial modelling and identification of data gaps for experiments being conducted in WP3.2 and WP3.3 | 12 |
| D10.1 | Ecosystem and plant characteristic data for model application | 12 |
| D12.1 | Summary report describing key response parameters derived from empirical studies and suitable for use in the first phase of the ecosystem valuation work | 12 |
| D16.1 | Indicators for geo-chemical and biological endpoints | 12 |
| D17.1 | Database of soil and vegetation data for the regional (5 x 5 km and 1 x 1 km) and landscape (~ 50 x 50 m) domains | 12 |
| D18.1 | Report on existing applications of the ESA in Europe and prioritisation of ecosystems and ecosystem services for detailed assessment | 12 |
| D19.1 | progress report on the implementation of new effect indicators and critical thresholds in the GAINS modelling system | 12 |
| D20.2 | Detailed description of model integration to establish 2050 scenarios | 14 |
| D1.4 | NH3 fluxes over Mediterranean semi-natural surfaces | 15 |
| D1.5 | Integrated dataset of canopy scale flux and in-canopy gradient measurements at a forest site | 16 |
| D8.3 | Concentration and deposition maps, for the regions mentioned above, at 5 x 5 km2, 1 x 1 km2, down to 50 x 50 m2 resolution for different components (e.g. NH3, NOx, O3) | 16 |
| D1.1 | First 6 months of continuous flux data of CO2, H2O, O3 and meteorological variables at 9 sites | 18 |
| D5.1 | Assessment of current GCMs and CTMS to reproduced recent trends models by comparison with selected observations | 18 |
| D7.1 | Maps of current air pollution metrics (APMs) across Europe, from the EMEP model and five other CTMs in order to provide a best-estimate and uncertainty range on vegetation effects metrics | 18 |
| D10.2 | One year ecosystem response data on plant responses to experimental changes | 18 |
| D11.1 | Parameterization of the impact of reduced and oxidised wet and dry N deposition on GHG and NOx fluxes, N immobilisation, natural vegetation types, species physiology, soil chemistry, and losses and allocation of C and N | 18 |
| D13.2 | New version of DO3SE model to simulate the combined effects of O3, N, S, diffuse radiation and climate on plant CO2 uptake | 18 |
| D14.2 | Updated versions of DGVMs and DSVMs that include O3 uptake model and N deposition on carbon uptake | 18 |
| D17.2 | Database of ammonia concentration and nitrogen deposition data (from A2.4) for the regional (5 x 5 km and 1 x 1 km) and landscape (~ 50 x 50 m) domains, where available | 18 |
| D20.3 | Detailed description of modelling system beyond 2050 | 19 |
| D11.5 | Quantification of minimum epidermal conductance under different loads of particles (*Month 12*) and estimation of transpiration increase by specified amounts of aerosols | 20 |
| D2.3 | Assessment of primary and secondary BVOC exchange rates in controlled conditions under simulated climate change and pollution scenarios | 22 |
| D20.4 | Description of the consequences of management change as an adaptation strategy on the scenarios investigated | 22 |
| D1.2 | Final 9 months of continuous flux data of CO2, H2O, O3 and meteorological variables at 9 sites | 24 |
| D1.3 | 2 x 6 weeks of campaign-based fluxes of VOCs, NH3 and NOx at selected sites | 24 |
| D2.1 | Initial database of controlled emission measurements on soil and litter | 24 |
| D2.2 | Data on microbial N turnover and NO (N2O) and CO2 emissions from soils following re-wetting of dried soils for improving parameterization of models | 24 |
| D5.4 | Boundary conditions for regional conditions | 24 |
| D9.3 | Completed database, response-functions and results of meta-analysis handed to WPs 3.4 and 3.5. | 24 |
| D10.3 | One year response data on ecosystem carbon balance responses to experimental changes and interactions with air pollution factors | 24 |
| D11.2 | Predictive modelling of GHG fluxes, especially CO2 under different N deposition regimes | 24 |
| D11.3 | Quantification and parameterization of foliar O3 deposition under progressing drought and temperature stress (mo. 16) and during leaf development and seasonal metabolic changes | 24 |
| D12.2 | Documentation of the DO3SE\_C model | 24 |
| D13.3 | Report on performance of site-based and regional-scale models in tests against experimental site data, to inform large-scale model application in C4 | 24 |
| D14.3 | Validated and evaluated version of models ( DGVMs and DSVMs) using databases on plant productivity | 24 |
| D15.1 | The model EUMOVE | 24 |
| D15.2 | Collated dataset of European soil 14C data used to define soil turnover times as a function of soil/vegetation type, for model parameterisation | 24 |
| D16.2 | Map of critical ozone uptake thresholds at European scale | 24 |
| D18.2 | Description of data for quantifying ecosystem effects and for valuation, including protocols for handling uncertainties | 24 |
| D19.2 | Reporton the modelling system for the impacts assessment under ÉCLAIRE | 24 |
| D20.5 | Preliminary report on cost optimization for 2050 scenarios | 28 |
| D1.6 | 4 publications on integrated forest campaign | 30 |
| D3.1 | NH3 emission model for agricultural management (improved with regard to its sensitivity to predicted changes in environmental conditions and management) | 30 |
| D3.2 | Background bi-directional NH3 exchange with soil/vegetation module (updated parameterization and inclusion of co-deposition effects) | 30 |
| D3.3 | Soil NO emission model (improved parameterization with regard to responses to changes in environmental conditions) | 30 |
| D3.4 | BVOC modelling framework allowing to integrate effects of climate and atmospheric composition change | 30 |
| D6.2 | Improved terrestrial (semi-) natural and agricultural emissions in response to integrated effects of climate change, change in atmospheric CO2 and N burden and land use/land management change | 30 |
| D6.3 | Sectoral emission profiles for selected source sectors and countries for application in local-to-regional scale models | 30 |
| D8.2 | Report on local scale interactions (and their variability) between air quality and climate change, based on modelling studies for selected regions in Europe | 30 |
| D8.4 | Sub-Grid module for inclusion in the EMEP model, enabling the description of sub-grid variability, based on activities from Tasks 2.4.1 and 2.4.2 | 30 |
| D11.6 | Parameterization of water use efficiency for model use (WP3.5) under conditions of particle pollution | 30 |
| D12.3 | Delivery of novel thresholds for key dose-response relationships for use in regional scale modelling and mapping relevant for ecosystem service assessment | 30 |
| D14.4 | Model runs (DGVMs and DSVMS) using the ÉCLAIRE scenarios of future emissions and climate change | 30 |
| D15.3 | The VSD+-EUMOVE and MADOC-EUMOVE models linked to European databases | 30 |
| D18.3 | Elaboration of the modelling approach, to include illustrative applications | 30 |
| D11.4 | Measurement and parameterization of the fraction of O3 that is taken up by leaves due to detoxification by constitutive BVOC, under associated environmental constraints and during leaf development | 34 |
| D16.3 | Map of critical N loads based on an inverse VSD+-EUMOVE approach at European scale | 34 |
| D4.1 | Improved pollution- and climate-sensitive exchange parameterisations for the main inorganic Nr compounds, suitable for inclusion in CTMs and underpinning mapping of novel thresholds and exceedances | 36 |
| D4.2 | Ozone dry deposition parameterisations, improved with respect to changes in climate and environmental conditions, suitable for inclusion in CTMs and underpinning mapping of novel thresholds and exceedances | 36 |
| D4.3 | A coupled pollutant deposition and photosynthesis model (DOSE\_C), based on the existing DO3SE model for O3 deposition | 36 |
| D4.4 | A vertically-resolved, multi-layer in-canopy chemical processing model of NO-NO2-O3-VOCs and NH3-HNO3-NH4NO3 exchange | 36 |
| D5.2 | Report describing the range of future evolutions of global, hemispheric and European O3, O3 precursors, and aerosol using a range of anthropogenic and natural emissions | 36 |
| D5.3 | Report describing the contributions of regions and processes on key environmental variables under future conditions | 36 |
| D7.5 | Source-receptor matrices of APMs for current and future conditions for use in effects and integrated assessment components | 36 |
| D13.4 | Report on assessment of the effects of combined air pollution and climate change scenarios on ecosystem C/GHG balance, soil quality and vegetation change at all experimental sites, based on integrated models | 36 |
| D14.5 | Ensemble dataset of model runs to assess the impact of combined air pollution and climate change scenarios on ecosystem C/GHG balance | 36 |
| D16.5 | Feedback from the GAINS model on the applicability of the newly acquired critical thresholds in scenario analysis" | 36 |
| D19.3 | Report on magnitude, location and robustness of assessments of adverse effects of GAINS scenarios | 36 |
| D20.6 | Assessment of sensitivities and uncertainties of the scenarios, considering the effects of management changes | 36 |
| D7.2 | Improved EMEP model with climate-change and canopy-chemistry capabilities, able to predict the impacts of changing climate and emissions on APMs | 40 |
| D12.4 | Final Report describing new dose-response relationships and novel thresholds | 40 |
| D14.6 | Report on the comparison of regional-scale models applied on test sites in C3 with large-scale model runs | 40 |
| D16.4 | Map of critical N load and critical ozone uptake exceedances based on a comparison with EMEP model results | 40 |
| D2.4 | Definition and improved parameterization of fluxes of BVOC under new environmental constraints and in relation to pollutants and endogenous induced emissions of NO and reactive oxygen species | 42 |
| D4.5 | Inferential model current and future estimates of Nr and O3 deposition at the ÉCLAIRE effect study sites of Component 3 | 42 |
| D14.7 | Report on ensemble application of DGVMs and DSVMs and intercomparison of model results on plant production/carbon sink strength in response to scenarios for the period 1900-2100 | 42 |
| D15.4 | Assessments of the effects of combined air pollution and climate change scenarios on plant species diversity and soil quality and derive thresholds based on criteria for soil quality and plant species diversity | 42 |
| D7.3 | Report on effects of in-canopy BVOC and NO emissions on in-canopy O3 and POD estimates | 44 |
| D14.8 | Report on the impacts of historic and future changes (period 1900-2100) in climate, air quality and agricultural management on crop production | 44 |
| D17.3 | Assessments of the uncertainty of critical thresholds for N and their exceedances at the European scale (50 x 50 km) due to model resolution for the present period and for simple future climate/emission/land cover scenarios (where available) | 44 |
| D19.4 | Final report on the development, implementation and scenario application of methods and data to assess adverse effects | 44 |
| D20.7 | Final cost optimization scenarios for 2050 and beyond | 44 |
| D7.4 | Report on effects of changes in global climate, chemistry, emissions and landcover changes on APMs | 48 |
| D18.4 | Scenario analysis to include policy recommendations and advice to other interest groups (*Month 48*, joint report with D5.3.9) | 48 |
| D20.8 | Policy recommendations and advice to other interest groups – joint report with D5.1.4 | 48 |

**ANNEX 2 The Data Exchange Support Group**

During the first 6 months of the ÉCLAIRE project, it has become apparent that there are many key items of data (such as measurements feeding into models, model output feeding into other models, among others) which need to be passed between workpackages and components to ensure the integration of the science and overall success of the project. Whilst aspects of these are covered by tasks (and therefore persons) within workpackage 21, it is felt that there is a need to have an overview on the project level, of whether these data exchange links are being made, and to make clear the roles and responsibilities for achieving these links.

To this end, we:

1. propose to establish a “WP21 Data Exchange Support Group” (DESG)
2. ask the following question:

“Who is ultimately responsible for ensuring that data exchanges between work packages and components are harmonised and synchronised properly?” By “data exchanges”, we mean ANY necessary information flow between work packages.

To avoid misunderstandings of responsibilities, and to scope the roles of WP21 task leaders in this process:

* ***We suggest that ultimately it is the responsibility of the Component leaders to ensure that timely and sufficient data exchange occurs between work-packages within their components, and between work-packages of different components***
* ***We suggest that Monitoring and Support can be provided from the proposed new “WP21 Data Exchange Support Group” (DESG)***

Within this framework, the DESG propose that -

**The Component Leader duties are:**

1. to ensure that communication on data needs and supply (measurements and models) are clear within components. This will involve communication between the component leader and the constituent WP leaders,
2. to ensure that communication on data needs and supply (measurements and models) are clear between components. This will involve communication between relevant component leaders,
3. to ensure that these data exchanges occur, and occur on time,
4. to provide a brief overview of communications and data exchange activities at the ESG 6-monthly meeting,
5. to actively seek support from the WP21 DESG when required.

It is important to note that the Component Leaders will draw on the support of WP leaders within their component, members of the Data Management Committee (users and providers) within their component, as well as the DESG team, to achieve the above, but that data exchange activities will be directed by the Component leaders themselves using whatever approach is most relevant.

**The WP21 Data Exchange Support Group duties are:**

1. to be alert for potential data exchange issues and problems as ÉCLAIRE progresses, and communicate these to relevant Component and WP leaders,
2. provide advice and support for intra- and inter-Component data exchanges when Component leaders request it,
3. communicate every ~3 months with Component leaders to check on data exchange links and issues, requesting feedback and alerting us to new problems arising,
4. monitor the risks relevant to data exchange (*listed below from the DoW*) (and any further ones as identified by the community)
5. review the 6-monthly reports provided to the ESG, and take/advise action as appropriate.

The DESG will draw upon Component Data Managers to assist and support them in this role.

***ÉCLAIRE project risks relevant to data exchange, to be monitored by the DESG (from the ÉCLAIRE DoW):***

-Slow delivery of verified flux datasets for modelling

-Slow delivery of climate dependence in model parametrizations for upscaling

-Slow delivery of verified effects datasets for modelling and threshold assessment

-Slow delivery of site-scale C, GHG and species modelling outputs for upscaling

-Slow delivery of regional emission and hemispheric background estimates

-Slow delivery of upscaled estimates of climate dependence on source-receptor relationships

-Inconsistent / incomplete measurements (WP21-2)

-Inconsistent model views & uncertainties (WP21-3)

-Inaccessible datasets and inadequate quality assurance by individual partners (WP21-4)

DESG will keep an overview of the data exchange through the overarching component links, i.e. C1-C2 (measurements to models), C3-C4 (measurements to models), C4-C5 (models to models).

The WP21 DESG will communicate/meet every ~3 months (or as necessary) to review data exchange status, and decide upon any necessary actions arising from this.

**Annex 3**

****

**ÉCLAIRE QUALITY ASSURANCE and QUALITY CONTROL PROTOCOL**

***QUALITY ASSURANCE***

**Data collection**

Quality Assurance of raw data collection lies outside the remit of the ÉCLAIRE DMC Data Quality Assurance Plan. However, the DMC issues the following guidelines:

*“During data collection researchers must ensure that the data recorded reflect the actual facts, responses, observations and events.”*

*“The quality of data collection methods used strongly influences data quality, and documenting in detail how data are collected provides evidence of such quality.”*

*“Calibrations, inter-laboratory comparisons, field notes, instrument performance and detection limits should be documented.”*

**Data entry and upload**

“*When data are digitised, transcribed, entered in a database or spreadsheet, or coded, quality is ensured by standardised and consistent procedures for data entry with clear instructions.”*

**Data checking**

*“Data checking is when data are edited, cleaned, verified, cross-checked and validated. Checking typically involves both automated and manual procedures.”*

***QUALITY CONTROL***

**Data collection**

Details of quality control measures during data collection can include:

* calibration of instruments to check the precision, bias and/or scale of measurement
* taking multiple measurements, observations or samples
* using standardised methods and protocols for capturing observations

**Data entry and upload**

For ÉCLAIRE, quality control at the stage of completing Excel templates and uploading to the CEH database includes:

* setting up validation rules or input masks in data entry software where appropriate
* for a limited number of data fields, using code lists and choice lists to minimise manual data entry
* setting up controlled vocabularies, which will be implemented during the first 18 months
* creating INSPIRE-compliant meta-data
* using the purpose-built database structures to organise data and datasets
* accompanying notes and documentation about the data

**Data checking**

For the ÉCLAIRE project, this will include:

* checking for out-of-range values, duplicates, and data of the wrong format when data are uploaded to the CEH database. This is performed by the database software, and generates an automatic message to the uploader, prompting a double check of out-of-range values and wrong formats. Rows of data with duplicate key fields abort the upload and the uploader is requested to address the duplicates issues.
* checking data completeness: This will be done via (i) the built-in database software functionality “Quick Site Summary”, (ii) off-line checks using FME, and (iii) where appropriate, via test runs of models,
* verifying random samples of the digital data against the original data: this may be set as a routine exercise every 12 months: small subsets of data will be downloaded from the database and sent to the originator to check against raw data files,
* statistical analyses such as frequencies, means, ranges or clustering to detect errors and anomalous values will be constructed and performed in R, using appropriate statistical techniques such as Times Series Analysis,
* encourage early writing of synthesis papers, as this brings remaining data defects to the attention of the research community

**Annex 4**

## **Conditions of Use of ÉCLAIRE IP Data**

These data were generated or collected within the framework of the ÉCLAIRE IP. The user agrees.

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