# Universities Facility for Atmospheric Measurement (UFAM)
# Data Management Plan

**British Atmospheric Data Centre (BADC)**

## Contents



## 1. Introduction and Scope

It is NERC policy to ensure as wide as practical archiving of data for future use. In fact, the NERC Data Policy Handbook states: "NERC grant-holders in academia are required to offer to lodge with NERC a copy of the data resulting from the supported research when it is completed, together with documentation/metadata describing these data" – see http://www.nerc.ac.uk/data/policy.shtml for a copy of the handbook.

Successful archiving is difficult unless the task and the significant effort involved are recognised at the outset of activity, and are fully built into plans. Typically this discipline will be undertaken for all continuous measurement data and for field studies where generation of data for shared use or long-term reference is an aim. In planning the archive work a sensible judgement of the level of data to be archived is needed.

This Data Management Plan is the result of discussions between the:

- UFAM Facility Manager
- Lead Scientists
- Instrument Scientists (attendants of the UFAM Utilization Committee).
- BADC.

These discussions have also resulted in the production of the UFAM Data Protocol (see Annex 1). Ongoing issues of data management will be agreed by the above parties.

The purpose of the UFAM Data Management Plan is to set up a coherent approach to data issues relating to data produced by UFAM instruments and associated datasets. The objective is to ensure that:

- Appropriate data support is provided to the scientists involved with UFAM.
- UFAM datasets[1] are made available to collaborators in a timely fashion.
- Distribution conditions and data usage do not infringe on the individuals' rights to publish their own work.
- Potentially scientifically valuable data are kept for the long-term.
- A high quality documented UFAM data archive is created.
- Data and documents are eventually distributed to the scientific community.

## 2. Access restrictions to UFAM-related datasets

### 2.1 Datasets covered by this Data Management Plan

Since UFAM is a collection of instruments the term 'UFAM datasets' needs to be clarified. There are 5 categories of data that can be considered as 'UFAM datasets' due to their direct involvement or association with UFAM instruments. These are presented in table 1.

**Table 1.** Definition of 'UFAM datasets'.

| | | |
|---|---|---|
| **i** | **UFAM campaigns** | Measurements by UFAM instruments as part of research campaigns (typically involving many UFAM instruments). |
| **ii** | **Research projects** | Measurements by UFAM instruments as part of other research projects (typically involving only one UFAM instrument). |
| **iii** | **Routine monitoring** | Measurements by UFAM instruments when not deployed as part of a particular research campaign or project (often for the general interest of the PI or Instrument Scientist). This type of measurement is referred to from here onwards as 'routine monitoring'. |
| **iv** | **Non-UFAM instruments** | Measurements by non-UFAM instruments and data generated by models during research campaigns and projects that have UFAM involvement. |
| **v** | **Third-party data** | Third-party data gathered to support UFAM-related campaigns and projects (such as numerical weather prediction data and observations from the UK Met Office). |

### 2.2 Restrictions on Data

For all NERC-funded campaigns and projects (items (i), (ii) and (iv) in table 1) the BADC is the agreed designated data centre for UFAM-related datasets. UFAM Instrument Scientists are responsible for the long-term archival of data produced by UFAM instruments. The usual route is for this data to be archived at the BADC in the agreed format (see section 4.5). If the data for a particular campaign, project or routine monitoring is not to be held at the BADC then the Instrument Scientist should provide a detailed description of the dataset and metadata to the BADC. Restrictions on this data will follow those set out in the UFAM Data Protocol (see Annex 1). The usual case is that access to all data submitted to the designated data centres will be restricted to the UFAM campaign/project participants during one year following the concerned project end date.

For data produced by routine monitoring (item (iii)), the Instrument Scientist and Instrument PI should discuss with the BADC the merits of copying the data to the BADC. Restrictions on these datasets will depend on the wishes of the data owner (usually the Instrument PI), who will typically be asked to authorise each individual application via e-mail.

[1] See section 2.1 for a definition of *UFAM datasets*.

For third-party data (see section 3) in support of UFAM campaigns and projects (item (v)) the restrictions on the data will be determined by the standard BADC license agreement with the source organisation (usually the Met Office or European Centre for Medium-Range Weather Forecasts (ECMWF)). This issue will be dealt with on a case-by-case basis, as it may be appropriate to store the data in the standard BADC archive for that data type or not to archive the third-party data at all.

### 2.3 Data restrictions in the case of NERC Thematic Programmes

On some occasions UFAM instruments will be deployed as part of a NERC Thematic Programme which will have its own data management rules. In such cases the data restrictions for the Thematic Programme will apply to the data unless specific exception is made by the PIs of the project, instrument and the UFAM Facility Manager.

### 2.4 Data restrictions in the case of non-NERC data

Some data from UFAM instruments will be generated under funding from non-NERC sources (such as the EU). In such cases the funding body may prohibit archival and distribution by the BADC. Such datasets will be addressed on a case-by-case basis in consultation with the instrument PI and UFAM Facility Manager.

### 2.5 Access rules to increase research use of UFAM instrument data

Some users will be interested in accessing data from one instrument across a number of campaigns and projects. The BADC will therefore endeavour to create a 'dual access' system for UFAM instrument datasets. This will allow access to be granted for individual campaigns as well as for a particular instrument only (spanning more than one campaign/project). Where restrictions exist for a limited time after a campaign the data will not be available to non-participants in the campaign until after the restrictions have been removed.

## 3. Access to third-party data

### 3.1 Third-party data support for UFAM campaigns and projects

Third-party data required for the support of UFAM-related campaigns and projects which is held at the BADC, such as ECMWF and Met Office datasets, will be made available to the participants, subject to current access conditions. Key datasets identified of interest to UFAM participants are:

- Met Office surface station meteorological observations (in near real-time[2]) such as temperature, rainfall and wind parameters.
- Met Office radiosonde ascent data for the UK (in near real-time[2]).
- Met Office Unified Model (UM) numerical weather prediction analysis and forecast data at global and mesoscale resolutions[3] (in near real-time[2]).
- Trajectory data from the BADC web trajectory service.
- ECMWF Integrated Forecasting System (IFS) numerical weather prediction analysis and forecast data (in near real-time[2], special permission required[4]).

If required, the BADC will endeavour to retrieve datasets from other sources at no cost or will negotiate their acquisition at the best possible cost.

[2] Note that the BADC currently extracts each of its Met Office data types from a different location within the Office. Access to data in 'near real-time' is taken to mean "within 24 hours of production/measurement".
[3] Note that the Unified Model data can only be provided by the BADC in the standard configurations used to create the current UM data archive.
[4] Note that ECMWF analysis and forecast data require special permission to be requested for each UFAM campaign or project if near real-time data is sought.

# 4. UFAM data archive

## 4.1 Archive location and long-term credibility

The central UFAM archive will be located at the BADC which will act as the archive of last resort. Datasets will be routinely backed up in case of damage or technical failures.

## 4.2 Archiving policy

In recognition that validated raw data (i.e. QA/QC'ed data prior to additional processing) potentially represent an invaluable source of information for the future, UFAM participants will archive them in a way that guarantees longevity and accessibility. Although not necessarily located at the BADC, validated raw databases and their access should be fully documented at the BADC. Processed (final) data should be archived at the BADC. In addition, investigators are encouraged to submit model results which would be the basis of theoretical studies or would illustrate the use of their models. Model results generated by UFAM-related campaigns and projects that are required by other participants during the archived at the BADC for provision in near real-time.

## 4.3 CAST – A secure workspace for preliminary data and document sharing

Preliminary datasets and associated documents generated by UFAM-related campaigns and projects can be shared amongst participants using the CAST (Collaboratory for Atmospheric Science and Technology) system. CAST, hosted by the BADC, provides a secure (password protected) web-interface to workspaces that can be used for file sharing and discussion in the post-measurement analysis and interpretation stages. See http://badc.nerc.ac.uk/community/workspaces for more information on shared workspaces.

## 4.4 Data types in the UFAM archive

The key datasets to be included in the UFAM archive will be produced by instruments measuring a range of atmospheric variables. Data will also include model output, trajectory runs, and third-party data such as Met Office surface station measurements.

## 4.5 File Formats

All data should be submitted in the agreed file format of NetCDF (following the Climate and Forecasts (CF) Metadata Convention[5]). Compliance with the CF convention enables the inclusion of valuable metadata within the data files. Data may also be accepted in the NASA Ames (ASCII) format. Documentation on both formats is available from the BADC (http://badc.nerc.ac.uk/help/formats/), as well as links to downloadable free software packages to produce and read NetCDF files.

## 4.6 Archive directory structures

In order to maximise future access to instrumental datasets the BADC will group data under instrument type as well as by campaign or project. The restrictions on each data subset (such as each campaign or project an instrument is involved in) will still govern access to the data as mentioned in section 2.

The directory structure for the data will follow the convention:

**/badc/UFAM_CAMPAIGN/data/INSTRUMENT_NAME/YYYY/[MM]/[DD]/[hh]/[mm]/[ss]/filename**

and:

**/badc/ufam/INSTRUMENT_NAME/data/YYYY/MM/[DD]/[hh]/[mm]/[ss]/filename**

Where:

**UFAM_CAMPAIGN** is the main name of the research campaign or project.

**INSTRUMENT_NAME** is the agreed name for the instrument (or model, trajectory model etc.,).

**YYYY** is the year (4 digits).

**MM** is the month (2 digits). Note that the **[MM]** directory is optional in the **UFAM_CAMPAIGN** directory structure but not the **INSTRUMENT_NAME** directory structure.

[5] The CF convention is fully documented at: http://www.cgd.ucar.edu/cms/eaton/cf-metadata/index.html

**[DD]/[hh]/[mm]/[ss]/** are optional directories for day, hour, minute and second. The depth of directories included to describe the time field will depend on the amount of files present. A rule of thumb is that a new level should be included if more than 100 files exist in the directory.

**filename** is the name of the data file itself following the file-naming convention (see section 4.7).

For example, the archive location for the UFAM 1290MHz Wind Profiler (radar-1290mhz) data collected during the Summer 2002 NAMBLEX field campaign would be:

**/badc/namblex/data/radar-1290mhz/2002/08/radar-1290mhz_macehead_20020801_low-res-1h-1.na**

Once the initial data restriction period had elapsed the data would also be available to researchers interested only in the wind profiler instrument. The data would then also be visible at:

**/badc/ufam/radar-1290mhz/data/2002/08/ radar-1290mhz_macehead_20020801_low-res-1h-1.na**

### 4.7 The UFAM file-naming convention

All UFAM-related data submitted to the BADC must follow the filenaming convention described here (see also: http://badc.nerc.ac.uk/data/ufam/file_naming.html). The file-naming convention uses long file names to provide useful information without the user having to read the file header or refer to the directory structure. Important attributes in the file name include *instrument*, *location* and *time*.

The chosen convention is as follows:

**instrument_location_YYYYMMDD[hh][mm][ss][_extra].ext**

Where:

**instrument** is the instrument name (full or shortened). When the same instrument is used by a number of groups, the instrument name should be prefixed with the institute name/code and a hyphen, for example **leeds-fssp** and **umist-fssp**. This field can also provide the model name in the case of model data.

**location** is the location name (full or shortened). This refers to the location of the observation and not the institute or location of the participating scientist/group. This field could be used for a range of items such as a site, a station, a platform or a place name.

**YYYYMMDD** is the date on which measurements were taken. If a data file spans more than one day then this field should represent the first day during which data was recorded. The year is given as four digits with month and day as two digits each.

**[hh][mm][ss]** is the time of day (optional). If included, hours, minutes and seconds must be represented as two digits each. Hours can be used alone, only hours and minutes used or all three fields can be included. However, minutes or seconds cannot be used without the preceding time unit (i.e. no minute field allowed without the hour field).

**[_extra]** - this section allows additional code to define such things as different range resolutions and so forth. It could also be used for version numbers, chemical species etc,.

**.ext** - will normally be **.nc** (NetCDF) or **.na** (NASA Ames) although occasionally other formats will be used, in particular **.png** and **.gif** for Image files.

Filename fields should contain only the characters "**-**", "**0-9**", "**a-z**". Spaces are forbidden and upper case characters should be avoided. The underscore "**_**" character can only be used as a separator between fields.

There is also a list of instrument, location and other related names commonly used in file names at:
http://badc.nerc.ac.uk/data/ufam/common_names.html

### 4.8 Data submission

The BADC provides an automatic web-based file uploader accessible via the *Data* and *My BADC* tabs on the web site (or directly from: http://badc.nerc.ac.uk/data/submit.html). Online assistance is provided for this service. Alternatively, files can be submitted by FTP. Both methods are fully documented on the BADC web site.

When required by other participants in UFAM research, preliminary data should be made available as soon as possible, via one of the designated data centres (if appropriate). Processed data and model results should be supplied to the relevant data centre as soon as they are ready, and no later than the project end date. Individual project archives should be completed by the end date of the project.

### 4.9 Updates to datasets

Occasionally a data provider will process a new version of a dataset or subset. Files must contain the version number or history in the metadata so that newer versions can be distinguished from previous versions. When a new version of a dataset is received the BADC will send a message to all known users of the data to inform them of the change.

### 4.10 Documentation

Metadata are a crucial part of any data archive since they ensure the readability of the data. It is therefore essential that metadata are submitted at the same time as the datasets to which they pertain. Metadata pertaining to all UFAM-related data archived elsewhere should also be supplied to the BADC.

To guarantee data archive quality, full documentation on all validated raw and processed data, as well as on models and model results, should be provided to the BADC.

Standard metadata will be archived within the NetCDF (and NASA Ames) data files. More information on providing metadata is detailed at:

http://badc.nerc.ac.uk/data/ufam/metadata.html

In addition to the standard metadata, investigators are encouraged to archive at BADC all relevant information, including references, papers, reports etc. Designated directories will be created in the UFAM archive for this purpose.

## 5. Data distribution

Where other restrictions do not apply (see section 2), access to all data submitted to the designated data centres will be restricted to the UFAM campaign/project participants during one year following the concerned project end date. After this period the data will be released into the public domain (exceptions may be made for third-party data).

A password-protected access system will be set up at the BADC to reflect the defined permissions. Whilst the data are restricted from the public domain, participants will be prompted to agree with the UFAM Data Protocol (see Annex 1) in order to access the UFAM archive.

After release of the data to the public domain, anonymous users will be requested to contact the relevant data providers before using the data and to acknowledge the owners of the data and the data suppliers in any publication using UFAM-related data.

Distribution of UFAM data held at the BADC will take place via the Internet and FTP. During the validation period, entitled UFAM participants who have applied for access to the data will be allocated an account at the BADC allowing them to directly download the data from the archive. This facility will be extended to external collaborators who will have been personally authorised to access the data by the

project PI. A UFAM web front page has been set up at http://www.badc.rl.ac.uk/data/ufam/. This will be the gateway to all UFAM data and metadata, and to all relevant information and links.

UFAM data held at the BADC will benefit from future development of access technology. Facilities currently under development include the e-Science NERC DataGrid, and a *Live Access Server*[6] allowing data subsetting, visualisation and conversion via a web-interface.

## 6. Publication

Results coming out of the UFAM-related research will be published in the usual way. During the data validation period of a particular campaign or project, each investigator will have the right to refuse the use of his/her results in a publication or a presentation prior to the investigator's own publication of that work. If measurements or model results from other UFAM-related research are used in a UFAM participant's publication, joint authorship should be offered. This will not necessarily have to be accepted, particularly in cases where due credit and acknowledgement can be given in other, possibly more appropriate, ways. References of publications should be communicated to the BADC where a list of published works will be held.

[6] Live Access Server (LAS) provides a data manipulation interface to data via the web, see http://ferret.wrc.noaa.gov/Ferret/LAS/ferret_LAS.html for more details.

# Annex 1 - UFAM Data Protocol

The scope of this Data Protocol is intended to be datasets generated by UFAM instruments. This may include research campaigns and projects using those instruments as well as 'background' data recorded when the instrument is not deployed on field research. Where the BADC is the archive of last resort for a particular campaign this protocol covers data generated by that campaign including non-UFAM sources.

The aims of the Data Protocol are:

- to encourage rapid dissemination of scientific results from UFAM;
- to protect the rights of the individual scientists making use of UFAM instruments;
- to ensure that all involved researchers are treated equitably;
- to ensure the quality of the data in the UFAM data archive.

These aims conflict at times, and it is hoped that the provisions of this protocol will resolve these conflicts fairly. It is recognised that this cannot always be achieved to everyone's complete satisfaction. There are likely to be cases where individual interests clash with those of UFAM. Therefore to try to meet these aims, all PIs, Co-Is and Instrument Scientists involved in UFAM must agree to abide by the following conditions:

1. UFAM data produced under the auspices of a NERC Thematic Programme will be subject to the Data Protocol of that programme.
2. UFAM data and model results produced during campaigns and projects will be made available to all relevant participants, and to those participants only, during a *restricted access period* ending one year after the concerned project end date, after which data and model results will be released to the public domain. At an investigator's request, access may be extended to personally authorised collaborators.
3. The designated data centre for UFAM-derived data is the BADC.
4. The longevity of validated raw data must be ensured in a secure archive, possibly, but not necessarily at the BADC. Details pertaining to the validated raw data (i.e. metadata), whether or not archived at BADC, must be sent to the BADC, as well as information on how to access the data.
5. Preliminary datasets will normally be made available to other UFAM collaborators as soon as possible (usually via the Collaborative workspaces on CAST*, hosted by the BADC). Any corrections or amendments to the preliminary data should be announced as soon as possible.
6. Validated processed data (i.e. datasets in their final form) must be archived at the designated UFAM data centre with the required metadata. The data providers and the BADC should arrange a data submission date. Archival should take place no later than this agreed date.
7. Results of model studies feeding or accompanying UFAM campaign/project data analysis can be made available via the BADC.
8. Data produced outside of the auspices of NERC programmes and campaigns should be provided to the BADC where possible. The data owner may stipulate data access restrictions on such occasions.
9. Data submitted to the BADC must be in the data format agreed between UFAM investigators and the BADC (*NetCDF* (following the CF convention) preferred, *NASA Ames* also accepted). All agreed metadata describing data, models and model results, regardless of their archival location, must be supplied to the BADC. Format and metadata are documented at BADC.
10. It is the responsibility of UFAM Instrument Scientists to provide data, metadata and appropriate documentation to the BADC for all campaigns, projects and routine measurements.
11. It is the responsibility of each Investigator to ensure that the data used in publications are the best available at that time.
12. If measurements or model results from other UFAM-related research is used in a publication by a UFAM project participant, joint authorship must be offered. This does not necessarily have to be accepted, particularly in cases where due credit and acknowledgement can be given in other, possibly more appropriate, ways.
13. Whilst the data are restricted from the public domain (see Clause 2), each investigator has the right to refuse to allow his/her work, whether measurement or calculation, to be used in a publication or presentation prior to the Principal Investigator's own publication of that work.
14. Whilst the data are restricted from the public domain, no data should be transferred to a third party without the originator's consent.
15. In the event of dispute, a scientific steering committee will be set up consisting of the UFAM Lead Scientists and/or the Principal Investigators of the specific campaign or project, who will make a final decision.

* CAST – Collaboratory for Atmospheric Science and Technology (see: http://badc.nerc.ac.uk/community/about_cast.html).