

Announcement of Opportunity Invitation to Tender:

Definition, tools and auxiliary data for the UK hi-res climate-data format of the Earth Observation Climate Information Service

1 INTRODUCTION

The National Centre for Earth Observation (NCEO) announces an invitation to tender (ITT) to develop a UK high-resolution (hi-res) data format (and associated tools and auxiliary information) for the UK Earth Observation Climate Information Service (EOCIS).

EOCIS is a collaborative project led by NCEO, involving over a dozen UK universities and research centres. Project leadership is placed at NCEO University of Reading, with administrative and investment leadership at NCEO University of Leicester.

EOCIS will demonstrate the UK's first national supply chain for a range of essential climate variables obtained by Earth Observation (EO) and building on UK research excellence. EOCIS will invest in transformations of climate data into higher-level climate information that advances both climate science and informed practical action in response to the challenge of climate change. EOCIS has been funded by the National Environment Research Council, part of UK Research and Innovation, with the funding provided by the Department for Science, Innovation and Technology (formerly BEIS)^{1 2}.

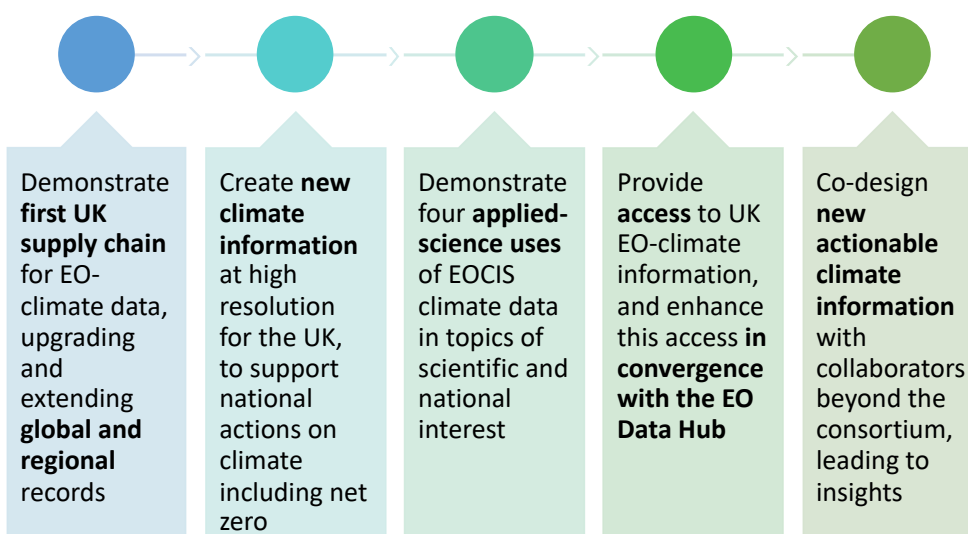
¹ <https://www.gov.uk/government/publications/earth-observation-investment>

² <https://www.nceo.ac.uk/article/uk-government-commits-close-to-400-million-to-earth-observation-research-and-industry-projects/>

1.1 EOCIS mission and activity

The UK Earth Observation Climate Information Service exists to make a positive difference to responses to the climate emergency. Climate information is an important tool in responding to climate change, in science, policy-making and business. EOCIS brings together UK research-community expertise to create and make available high quality, trustworthy climate information based on measurements of Earth’s environments from space. EOCIS will complement and collaborate with others meeting the need for climate information: nationally and internationally, and across science, commerce and policy.

Five programmatic elements will address the above scope shown in the figure below. The appendix lists the activities within these programmatic elements and briefly describes their main ambitions.



This AO relates to the second and fifth of these elements, and is a directed procurement to further the creation of new climate information at hi-res for the UK. This activity falls within EOCIS WP 3, which is led by the Project Lead, Chris Merchant, and is an “actionable information project” within EOCIS, intended to support the effective translation of climate data into new forms of actionable information.

2 TENDER DESCRIPTION

2.1 Bid scope

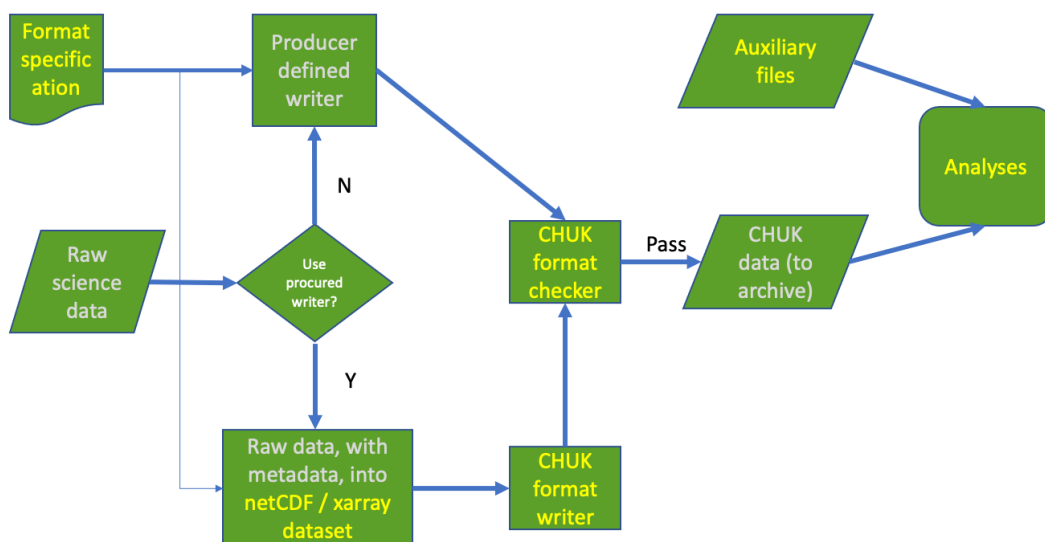
This announcement is to procure the following for the EOCIS project:

- A **comprehensive data and metadata definition** for datasets within the category of “climate-data at high resolution for the UK” (CHUK) activities in EOCIS. A **format specification document** fully describing this format.
- **Auxiliary files** of data transformed to the defined “CHUK format” with **documented descriptions** of how they were creation.

- **A CHUK-format data writer**, with appropriate data scaling / compression for efficient data volume.
- An effective **format checker tool** for the CHUK format, with associated test input and output data.
- **Code for efficient masking and subsetting** of CHUK-format data, e.g., against criteria that interrogate auxiliary files.
- **A methodology for estimating data volumes** given the format.

Bidders shall prepare bids that address all the above elements, in the light of the additional requirements set out below. For convenience of reference, requirements are numbered in the order they appear in this AO (e.g. [Req. X]).

The documents, files and code being procured will enable EOCIS CHUK-data producers to implement workflows such as that illustrated below:



Key: Yellow text indicates there is an associated deliverable to be supplied.

2.1.1 Requirements for CHUK format

Nature of CHUK data

EOCIS will create new climate data and information at hi res for the UK, to support national actions on climate including net zero.

[Req. 1] All the planned climate data and information (hereafter, “CHUK data”) will consist of 2-d rasters of quantities in space (no vertical dimensions) applicable at a stated time (or within a stated time window). The intervals between time slices may be regular or irregular, according to the nature of different CHUK data. The CHUK data for a given time slice may be gappy (the usual case, because a variable is relevant across only some of the domain, and because of cloud) or may be gap-filled.

[Req. 2] For storage efficiency, guidance on writing files with appropriate numerical precision of data and effective compression of unfilled data areas shall be part of the CHUK-format specification document.

CHUK domain (to be consolidated)

The provisional domain definition for the UK covers the land areas of the United Kingdom and most of the associated exclusive economic zones of the surrounding seas, corresponding to these limits:

Min. Lat 47.5°N

Min. Long 12.6°W (-12.6°)

Max. Lat 61.3°N

Max. Long 3.4°E

[Req. 3] The exact domain shall be defined: approximately in line with the bounds stated above; in the light of additional requirements on the domain's geographical projection (below); and in the light of any further key requirements that emerge by the time of kick-off.

Resolution and projection

All CHUK data will be generated on one common grid (i.e., raster definition) and projection (hereafter, the "CHUK grid").

[Req. 4] The exact CHUK grid specification shall be fully defined. The projection for the grid shall be that of the British National Grid (BNG)³. The domain (as noted above) shall be defined precisely, as a "rectangular" domain in that projection, and extrapolated over the ocean as necessary to address approximately the domain given above.

[Req. 5] The CHUK grid (raster) cell-size (or "resolution") shall be of order 100 m. The exact CHUK grid resolution and definition will take optimum ease of alignment with the following data sets, e.g., by defining grid corners that coincide with grid-corners other grids, facilitating robust averaging of CHUK data to other grids:

- HadUK-Grid 1 km interpolated weather observations⁴;
- CEH land cover data (see [Req. 11]); and
- the "2.2 km grid" used for local-model UK Climate Projections⁵ (UKCP18).

[Req. 6] Auxiliary files in CHUK-format shall be provided containing for each grid cell, the northwest, southwest, northeast and southeast grid corners and the grid cell centre as WGS84 latitude-longitude coordinates accurate to five decimal places (~1 m).

³ <https://britishnationalgrid.uk/>

⁴ <https://www.metoffice.gov.uk/research/climate/maps-and-data/data/haduk-grid/haduk-grid>

⁵ https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18_data_availability_jul-2021.pdf and links therein.

Metadata

Each CHUK-data file will require appropriate, comprehensive metadata.

[Req. 7] A CHUK file metadata format shall be specified and documented that is optimised with respect to the following considerations:

- Appropriate time-stamping of the data contents
- Open Geospatial Consortium standards for metadata
- Climate-Forecasting naming conventions including long names and units
- ESA Climate Change Initiative (meta)data standards
- Traceability of files to their origin, and to lower-level input files used
- Propagation of licence conditions and citation requirements to users

General CHUK-format requirements

[Req. 8] The CHUK format shall be a raster-based format that is optimised for ease of use and readability with tools and software that are widespread among the users of geospatial information who typically employ Geographical Information Systems for interacting with data. For example, cloud-optimised geoTIFF is a baseline choice; this and other options (e.g., Zarr with a GIS interface tool) shall be reviewed, leading to a documented, reasoned selection.

[Req. 9] In defining the CHUK format, the following implementation aspects shall be optimised with respect to the foreseen architecture of facilities at Jasmin/CEDA and the EO Data Hub. A reasoned justification for the optimum choice shall be documented in appendices to the format specification document.

- Data scaling and representation, avoiding spurious precision and minimising data volume
- Compression, especially with respect to data voids
- Data chunking with respect to different usages: reading and processing of a large-area time slice (“time-slice”); reading and processing of data for a limited location across many time slices (e.g., to create a time series; “time-line”); efficient masking of data with respect to boolean criteria applied on auxiliary files (e.g., to process only data in a particular county, or a particular land classification within a nation)

It is expected that most “on the fly” analyses via the EO Data Hub interfaces will be time-slice analyses.

[Req. 10] Reading and processing performance (time-slice and time-line) of the CHUK-format files within the context of an object-store shall be considered, this being a foreseen EO Data Hub facility.

2.1.2 Requirements for auxiliary information

Auxiliary CHUK variables

Auxiliary files will be created in addition to those mentioned above under [Req. 6], as follows.

[Req. 11] CHUK-format files containing the following variables shall be created:

- [D3.1] A classification variable distinguishing (with bidder-proposed definitions to be agreed):
 - Permanent land
 - Permanent inland water (distinguishing lakes and rivers)
 - Permanent sea and estuary
 - Mixed or variable land/water
- A classification variable that identifies every grid cell with tags, based on the belonging of the majority area of the cell, for each the following based on Office for National Statistics digital boundaries⁶.
 - [D3.2] The devolved nation
 - [D3.3] The county / council / unitary authority / metropolitan or London borough
 - [D3.4] The parish / community / town council
 - [D3.5] The UK postcode sector (i.e., the level ABxy z., with ~12000 sectors)
 - [D3.6] Other boundaries of potential interest such as health and fire service districts (bidder to propose)
- [D3.7] The dominant land classification from the Centre for Ecology and Hydrology⁷ 2021
- [D3.8] The nearest valid UN-system land classification from the ESA Land Cover CCI medium-resolution land cover product, for each available year (2001 to 2020)
- [D3.9] The built and paved fractions for a recent year or years (bidder to propose source)

[Req. 12] Where no source for auxiliary data is indicated in [Req. 11], the bidder shall identify and justify in their proposal a credible source as part of their bid, which will be used in the bid assessment.

[Req. 13] Auxiliary files will be delivered in CHUK-format.

2.1.3 Requirements for code

[Req. 14] The CHUK software package shall be written in Python. It shall support all modern Python versions from 3.7 upwards, as demonstrated through CI testing (see below). The code shall be built on existing mature open-source software libraries (such as, but not necessarily, python-netCDF4, Numpy and Xarray). There shall be no proprietary software dependencies. The code shall be compatible with Linux operating systems and does not need to be tested on a Windows environment.

The CHUK software package shall be constructed following modern Python conventions, with dependent packages available as versioned releases that are published on PyPI and/or conda-forge. For reproducibility, a conda environment yml file shall be provided to facilitate installation of the dependencies.

⁶ <https://www.ons.gov.uk/methodology/geography/geographicalproducts/digitalboundaries>

⁷ I.e., based on CEH Land Cover Map 2021, choosing between the 10 m or 25 m products for reasons that shall be justified: <https://catalogue.ceh.ac.uk/documents/017313c6-954b-4343-8784-3d61aa6e44da>

[Req. 15] Code shall use intuitive variable naming and a modular structure. Code shall be well commented. Code shall be delivered in the form of a private GitHub repository. Unit tests shall be included in the software package, these should all pass successfully by sourcing, generating and testing with real example data. The unit tests should cover all main aspects of the functionality provided in the software package. The unit tests shall be integrated with a Continuous Integration (CI) system such as GitHub Actions that ensures they are run before code is merged into the "main" branch.

[Req. 16] A data writer that creates fully compliant CHUK-format files shall be supplied, along with test input data (e.g. a raster variable in netCDF with full corresponding metadata) and corresponding output data for verification (in CHUK-format). (The input and output of a suitable auxiliary file may be used.)

[Req. 17] The CHUK-format data writer shall be callable with options enabling the writing of a complete new file, and enabling overwriting of a subset of data within an existing file.

[Req. 18] A CHUK-format-checker tool shall be supplied. This tool will verify format compliance comprehensively, including:

- Typing and scaling of fields
- Presence and validity of all metadata, including presence of units and variable definition
- Correct representation of null/no-data values
- All numerical values within valid limits (as specified in metadata)
- Readability by a raster driver within the Python GDAL library

[Req. 19] An example Python module shall be supplied that exemplifies efficient masking of a CHUK-format data file against Boolean logic applied to a combination of two auxiliary data files (e.g., an operation to be applied to all CHUK grid cells within a stated county with a built/paved fraction of less than a threshold).

[Req. 20] Code delivered and accepted shall be copyright of University of Reading, held on behalf of NCEO and the EOCIS partners. No restriction on use for any application by any NCEO or EOCIS partner shall apply. The copyright holder will have the right to distribute the code at any later time as open-source code on a licence that allows anyone to reuse the code for any purpose. The supplier shall have a free, permanent, irrevocable licence to exploit the delivered code for any purpose.

2.1.4 Requirements on documentation

[Req. 21] The solutions adopted for the CHUK format shall be explained and justified in relation to requirements identified in this AO and in relation to other evidence and insight available to and/or gather by the contractor regarding community needs and drivers.

[Req. 22] A concise document shall be supplied describing a methodology for forecasting data volumes for CHUK-format products, accounting for:

- Numbers of files
- Data scaling option used
- Average fraction of raster populated with non-null data

- Any compression and its efficiency (with a means to establish efficiency from a sample of outputs)

[Req. 23] A product specification document fully describing the CHUK format shall be supplied. This will be of sufficient level of detail that the format could be created from the document without reference to the data writing code.

[Req. 24] A User Guide and Developer Guide shall be provided that explain how to install, test and run the checker and writer components of the code. The User Guide shall have demonstration code to get a user started. The Developer Guide shall include information about running the unit tests and how to fork the repository and make Pull Requests in order to suggest updates to the code. For all code supplied, inputs/outputs, installation and usage shall be clearly documented.

2.1.5 Requirements on technical approach

[Req. 25] An iterative approach to format definition, code development and data creation shall be adopted, with frequent interactions with EOCIS to discuss prototypes. (Bidders are invited to elaborate on their approach in their proposal.)

2.1.6 Requirements on review and acceptance

[Req. 26] All code, documents and auxiliary files will be subject to review by reviewers within EOCIS and NCEO-associated staff at Jasmin/CEDA. Review items raised will be resolved to the satisfaction of the EOCIS reviewer(s) to obtain acceptance of the deliverable.

2.2 Bid assessment

Bids shall be scored on the following measures of quality:

1. Compliance to the scope and general understanding of context (weighting, 25%)
2. Rigour, clarity and appropriateness of proposed approaches to meeting each requirement (40%)
3. Named personnel's experience and expertise for their respective tasks (25%)
4. Timeliness and credibility of proposed delivery schedule (10%)

The score obtained as above will then be considered in conjunction with the proposed firm fixed price to determine overall value for money, leading to bid selection.

2.3 Deliverables summary

D0	Explanation and justification of approach (refer to [Req. 21]; Report)
D1	CHUK format specification document (Report)
D2	Guidance for estimating CHUK data volumes (Report)
D3	Data package of auxiliary CHUK-format files (Data for download) (component deliverables D3.1 to D3.8 are identified in [Req. 11])
D4	CHUK format data writer (Software)
D5	CHUK format checker (Software)
D6	Sample code for efficient conditional masking (Software)

3 GUIDELINES FOR PREPARING A PROPOSAL

Applicants shall submit their bid as a single PDF file by email to eocis@reading.ac.uk using 'EOCIS CHUK ITT' in the subject line, with a copy to the EOCIS project lead (c.j.merchant@reading.ac.uk). An emailed acknowledgement of receipt will be sent.

The deadline for receiving proposals as above is 17.00 BST on 26th May 2023.

The bid package should consist of:

1. A cover letter including:
 - a committing offer to the University of Leicester, who will contractually administer the contract on behalf of the UK EOCIS consortium and on behalf of NCEO
 - the email address to which acknowledgement of receipt should be sent
2. The completed application table below.
3. A main proposal of no more than 30 pages⁸ of A4 (12 point, Arial), explaining the proposed approach in terms of
 - Work organisation (work packages or tasks)
 - Proposed schedule of delivery, including justification of feasibility
 - Proposed schedule of review and coordination meetings
 - Proposed technical solutions, where these can be identified at bid stage
 - Options and considerations the bidder proposes to explore with regards to technical solutions that are open
 - Compliance statement for every requirement
4. A complete statement of the financial proposal including:
 - hourly rates for named personnel
 - total hours proposed for each named person, on identified work packages or tasks
 - the budgetary allowance made for travel, meeting and any other costs, with explicit assumptions
 - any other charges
5. A 2-page track record for each staff member, focussed on expertise in the work packages or tasks they will work on

The financial proposal should:

- Be succinct.
- Include a table of personnel costs including hourly rates and total hours.

⁸ 30 pages is an upper limit, not a target. Shorter but complete proposals are welcome.

- Define all other costs by item, e.g., cost of meetings (estimated number of people, estimated number of meetings, assumed travel etc.).

Questions for clarification may be submitted by email to the project lead with the subject line, “EOCIS CHUK format AO clarification”. Questions posed along with their answers will be made publicly available. To preserve anonymity, the identity of the originator of the query will not be made public.

4 CONTRACTUAL INFORMATION

Award(s) will take the form of a contract between the University of Leicester, NCEO and the project manager’s organisation. The organisation will be expected to place contracts, if required, to flow down to any collaborating partners.

The award will be made on a firm fixed price basis. The guideline price is £100,000 exc. VAT. All bids will be considered irrespective of price relative to that guideline and will be judged on overall value for money (see 2.2).

30% of the payment will be made on KO and 70% on acceptance of all deliverables.

A draft contract may be requested from eocis@reading.ac.uk.

5 ASSESSMENT OF PROPOSALS

The review panel will consist of NCEO personnel. Proposals will be evaluated against the criteria in section 2.2.

6 ELIGIBILITY

Bids are invited from UK organisations who are members of Space for Climate and any others to whom this Announcement of Opportunity has been specifically directed by the EOCIS project lead.

1. APPLICATION FORM

The Application Form below should be completed and submitted with the Cover Letter.

Title of Project	UK EOCIS
Proposing (Lead) organisation	
Lead contact name and email	
Other organisations involved	
Cost for lead	£
Cost for each partner organisation (if any, add rows as necessary)	£
Total proposed budget	£
Address of lead organisation	