

## NextData Project - Data submission and distribution policies

**1. General statement:** Instrumental measurements, field observations, paleoclimatic proxy data, numerical simulations and reanalyses hosted in the system of archives and portals of the NextData project, including the General Portal, **are open access**, along the lines indicated below. These specifications are in agreement with the recommended guidelines for the access to research data defined by [OECD](#) and in the framework of [GEO/GEOSS](#).

**2. Users' rights and requirements:** users are free to use, copy and distribute the data, to produce works from the database and to adapt and transform the data, provided that any public use of the database or works produced from the database is attributed in the specific ways indicated in the Portal (including **citation or acknowledgement of the data provider or owner**, and/or offering co-authorship to the data owners, and always including an attribution notice for redistribution). In order to download data the users are required to **register at the Portal**, providing their true identity, affiliation and contact information and explaining the main intended use of the data. Registered users will be asked to accept the terms of a specific data distribution licence following the conditions listed above.

**3. Providers' requirements:** Instrumental measurements, observational and proxy data, numerical simulations and reanalyses obtained thanks to the funding of the NextData project **are required** to be fully open, following the specifications in this document, to the scientific community, stakeholders and private users, after a suitable period of validation and control, not exceeding one year after the data/simulations/reanalyses have been obtained. Longer delays can be tolerated in specific cases but should be clearly motivated.

**4. Metadata:** All data hosted in the system of archives and portals of the NextData project, including the General Portal, **must be complemented by full metadata** information; **metadata must be maintained up-to-date in time**. Field data and gridded datasets (climate simulations and reconstructions/reanalyses) will have different metadata structure and will be stored in different archives. The format chosen for the description of field data should be coherent with the SHARE GeoNetwork and GAW-WMO metadata scheme. For the distribution of climate simulations the use of a metadata format consistent with CMIP5 and CORDEX specifications is encouraged.

**5. Data format:** All data should use **open data formats**, which are free and widely available to everybody. For the distribution of measured, observational and proxy data, **formats such as csv, text or netcdf will be required**. For the distribution of gridded datasets, the use of standard data formats, such as GRIB and netcdf, is required and compliance with international specifications (such as CMIP5 and CORDEX) is encouraged. For the distribution of geospatial vector data, the use of shape formats (.shp, .shx and .dbf files) is required. This is the *de facto* standard format for using data in a geographic information system software, webGIS publication and for sharing by web services (WMS, WFS, WCS).

**6. Instrumental measurements, observational and proxy data:** Only **validated data** will be directly accessible. However, NextData encourages the data providers to make available, upon specific request by users and when possible, also the raw data and/or to provide links to appropriate web sites and archives. The motivations of the requests by users will be evaluated by the Executive Committee of the NextData project and by the data owners.

**7. Gridded datasets:** In the case of reanalyses, reconstructions and gridded observation products, the raw data from which the final gridded products are built are not requested to be available in the database. However, NextData encourages the data providers to make available, whenever possible, also the raw data and/or provide links to appropriate web sites and archives. The same policy applies to intermediate raw outputs of numerical simulations, if available.

**8. Required output:** For numerical simulations, reanalyses and gridded datasets, at least monthly averages and daily values of key variables (e.g., 2m temperature, precipitation, etc) are requested to be available. We strongly encourage to allow full access to the widest possible range of variables with at least daily temporal resolution. For numerical simulations consistency with the priority lists of output variables and frequencies defined in international projects (such as CMIP5 and CORDEX) is encouraged.

**9. Previous data:** For field data/simulations/reanalyses used in the NextData project but produced prior to NextData, and/or by groups not involved in the project, and/or with significant funding from other sources, the availability of the data will be discussed with the owners of the data or the responsables of the relevant projects, with a clear encouragement to make them accessible and open. Conformity with the conditions specified in this document is essential for hosting and distributing these data in the system of archives and portals of NextData.

## Short glossary

**Data set:** identifiable collection of data (ISO 19115:2003).

**Field data:** Measurement data from station-based observations, instrumental measurements in the field, observational campaigns and surveys, including paleoclimatic measurements.

**Instrumental measurements:** data provided by in situ and remote measurement instruments from stations and ad-hoc campaigns.

**Metadata:** data about data. Information describing data sets and making it possible to discover, inventory and use them (ISO 19115:2003 and INSPIRE glossary). Information about meteorological and climatological data concerning how and when they were measured, their quality, known problems and other characteristics ([IPCC AR4](#)).

**Observational data:** data provided by field observations (eg, animal counts, species number).

**Numerical Climate Simulations:** output of a numerical representation of the climate system based on the physical, chemical and biological properties of its components, their interactions and feedback processes, and accounting for all or some of its known properties ([IPCC AR4](#)).

**Paleoclimate:** climate during periods prior to the development of measuring instruments, including historic and geologic time, for which only proxy climate records are available (IPCC AR4).

**Paleoclimate proxy data:** Climate-related data that are interpreted to represent some combination of climate-related variables in past time (modified from [IPCC AR4](#)).

**Reanalysis:** Reanalyses are atmospheric and oceanic analyses of temperature, wind, current, and other meteorological and oceanographic quantities, created by processing past meteorological and oceanographic data using fixed state-of-the-art ocean and atmospheric models and data assimilation techniques ([IPCC AR4](#)).