Ref. Ares(2021)7175279 - 22/11/2021

GoNEXUS

Tools and solutions for governing the nexus

Deliverable D1.1: Initial Data Management Plan

November 2021, UPV

WP1

GoNEXUS has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 101003722.

www.gonexus.eu

Version 1.0
November, 2021

D1.1: Initial Data Management Plan

Lead by UPV
Hector Macian-Sorribes (UPV), Adria Rubio-Martin (UPV), Marc F.P. Bierkens (UU), Edwin H. Sutanudjaja (UU), Ioannis Mitsios (E3M), Alberto Garcia-Prats (UPV), Andrea Castelletti (POLIMI), Manuel Pulido-Velazquez (UPV)

Dissemination level of document

Public

Abstract

This deliverable describes the data management life cycle of GoNEXUS, including a summary of the data collected, post-processed and produced by the project, the provisions made to ensure FAIR (findable, accessible, interoperable and reproducible), the resources that GoNEXUS will allocate to ensure FAIR data, the data security guidelines followed by the project, the ethical aspects of GoNEXUS data and a table describing all the datasets that will be generated. This deliverable complies with the DMP template set by H2020, and it will be regularly updated during the project lifetime.

GoNEXUS has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 101003722.

Version History

Version	Date	Authors	Description
V1	27/10/2021	Hector Macian-Sorribes (UPV), Adria Rubio-Martin (UPV), Marc F.P. Bierkens (UU), Edwin H. Sutanudjaja (UU)**¡Error! No se encuentra el origen de la referencia.**	Generation of first draft
V1.1	8/11/2021	Ioannis Mitsios (E3M), Alberto Garcia-Prats (UPV), Hector Macian-Sorribes (UPV)	First round of internal review of V1 and comments to be addressed
V1.2	22/11/2021	Andrea Castelletti (POLIMI), Manuel Pulido-Velazquez (UPV), Hector Macian-Sorribes (UPV)	Second round of internal review of V1 and comments to be addressed

Table of contents

1. Data summary 4
1.1 Data collection and generation 4
1.2 Types and formats of data 5
1.3 Reuse of existing data 6
1.4 Origin of data 6
1.5 Size of data 7
2. FAIR data 7
2.1 Making data findable 7
2.1.1 Data Management Process 7
2.1.2 Naming conventions and keywords 9
2.1.3 Version numbers 10
2.1.2 Folder structure in the internal GoNEXUS server 10
2.2 Making data openly accessible 11
2.3 Making data interoperable 11
2.4 Increasing data reuse 12
3. Allocation of resources 12
4. Data security 13
4.1 Data made openly accessible through Zenodo 13
4.2 Data stored in an internal server 14
Hyrax 14
UU Yoda 14
UPV NAS 14
5. Ethical aspects 14
5.1 Informed consent and assent 15
5.2 Data privacy 16
6. GoNEXUS Data Tables 17
Table 1: Data collected and post-processed 17
Table 2: Data generated 17

1. Data summary

1.1 Data collection and generation

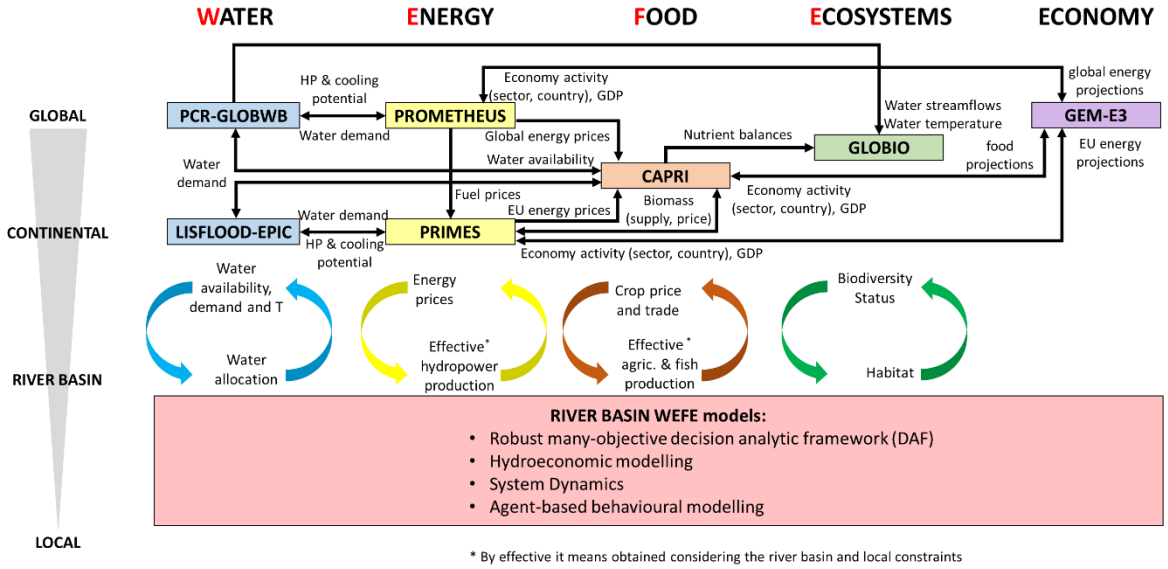
GoNEXUS will collect, generate, merge, summarize and analyze a significant amount of data. Most of the project datasets (in number) will be generated by the model toolbox, which implies several interconnected models covering from global to regional scales and offering results for alternative future scenarios and variables. Furthermore, GoNEXUS will collect data from two main sources: from external data services and through the Nexus Dialogues.

GoNEXUS will collect data (understanding data as datasets) to build the future scenarios (WP2), calibrate and validate the model toolbox (WP3, WP4), build evidence (WP5), and characterize the WEFE solutions (WP7). The datasets collected will mainly refer to:

- External data services: data from this source will be downloaded directly from the services (e.g. Copernicus Climate Change Service, EUROCORDEX, Data Portal from the Spanish Ministry for the Ecological Transition) and, if required, post-processed to ensure its applicability to the relevant spatiotemporal domains to be considered for each application. Collected datasets will mainly consist of data on future scenarios (e.g. climate change meteorological time series); observational data (e.g. time series of stored water resources, energy demands, agricultural production); and information on the performance of WEFE solutions (e.g. water and energy shifts from the adoption of drip irrigation technologies).
- Nexus Dialogues: data from this source will be facilitated by the participants in the Nexus Dialogues, either by direct communication or by granting access to restricted datasets, or both. These datasets will cover gaps of the external data service, and refine and contrast them (e.g. additional specifications on WEFE solutions, data from private gauging stations).

The datasets generated by GoNEXUS will mainly refer to the outputs of the WEFE models from the model toolbox, which will provide inputs for other parts of the model toolbox to improve its interconnection (WP3, WP4), build the evidence (WP5), and obtain the performance of the WEFE solutions analyzed (WP7). In order to ensure the interconnection between models, and between models and evidence, an input harmonization strategy will be set out to reconcile spatial and temporal scales, variables and units. This strategy will minimize the need for post-processing the datasets generated by any model before using them as inputs of further models. The datasets generated will refer to a wide range of variables, in particular, the ones that connect two or more elements of the WEFE nexus (e.g. water demands, which refer simultaneously to water and energy/food /ecosystems), as shown in the figure below; and the ones that will be used to build the evidence, which may depend on the spatial and temporal scale and the particular context of each case study.

Fig 1 : Model interconnections of the GoNEXUS model toolbox



1.2 Types and formats of data

GoNEXUS will collect and generate data of two different types:

- Point time series: they will be attached to a particular location or model calculation unit (e.g. gauging station, city, catchment, country) and cover a particular time span (between the beginning and the end of the recording period) with a certain periodicity (e.g. on daily, monthly, yearly basis).
- Gridded time series: they will refer to an area (e.g. sub-basin, irrigation district) that will be divided into grid points through a mesh, and cover a particular time span (between the beginning and the end of the recording period) with a certain periodicity (e.g. daily, monthly, yearly).

The data formats employed will depend on the type of data.

1. *Point time series*
   1.1. Text files (.txt, .csv) will be the standard format for these data. These formats are easily created and read by any programming language (e.g. Python, C++, Java, R, Visual Basic) or software (e.g. Notepad++, Microsoft Excel, OpenOffice).
   1.2. Microsoft Excel files (.xls, .xlsx, .xlsm) will be used in case that the text file format is not possible. It offers similar capabilities to text files, but the resulting files are usually larger in terms of size, and it may be bound to particular software products (Microsoft Excel, OpenOffice).
2. *Gridded time series:*
   2.1. NetCDF (.nc) will be the standard format for these data, in particular, its NetCDF4 version. This format efficiently packs all the gridded time series of a particular variable or collection of variables into a single archive, which can be easily generated and read using GIS servers (e.g. QGIS, ArcGIS) or programming languages (e.g. Python, R). It is a standard format for the provision of meteorological data and climate change scenarios.
   2.2. GRIB (.grib) will be used in case the NetCDF file format is not possible. This format is a standard format for the provision of meteorological data used by the WMO among others. GRIB files are usually more efficient than NetCDF in terms of size, but creating and reading a GRIB file is distinctly more difficult than NetCDF.

2.3. GeoTIFF (.tiff) will be used in case that the previous file formats are not possible. In this case, a collection of files will be generated, each one attached to a particular time (year, month and day). It offers the easiest alternative in terms of creation and reading, but the packing of the information is distinctly cumbersome than the previous ones and will demand more space.

1.3 Reuse of existing data

Some data sources that will be used by GoNEXUS (e.g. precipitation or streamflow records) are already in place and have produced data from years or decades ago. Some of these datasets will be used to calibrate and validate the models developed in GoNEXUS.

The project will also use outputs from models for which a stable version was developed in previous projects, whose results could be used as inputs for GoNEXUS (e.g. existing results from global models for CMIP5 scenarios). This data from the past will be used and combined for the development of model interconnections while new data produced by GoNEXUS is ready for use. Similarly, already-existing data on scenarios and Nexus Dialogues will be used in the early stages of the project, while GoNEXUS data is produced.

The treatment of existing data from these sources will follow the same process as done for the new data, outlined previously. It is expected that, at the end of GoNEXUS, all existing data reused, apart from data employed for calibration and validation, will be updated and replaced by data generated in the frame of GoNEXUS.

1.4 Origin of data

The data generated and collected by GoNEXUS will have three main origins:

- Scenarios: this origin includes all the data collected from external services that will be used to build the future scenarios (e.g. climate change projections, land use and land cover maps). These datasets are the product of distinct and continuous efforts done by the scientific community. The distinct feature of this origin is that GoNEXUS partners will have no control over which and how the datasets are produced and the scenarios to which they refer.
- Models: this origin refers to any result obtained by the model toolbox that will be used for modelling interconnection and/or evidence building. These datasets will be the product of the scientific research of previous projects as well as the activities of GoNEXUS. In opposition to the previous origin, GoNEXUS partners will be in control of the datasets generated.
- Nexus Dialogues: this origin refers to any dataset or information collected in the frame of the Nexus Dialogues, either to build scenarios, calibrate and validate models and build evidence. Some information will be the result of the recording or scientific efforts performed by the participants in the Nexus Dialogues, and some others will be the result of their own experience and points of view. For datasets recorded, GoNEXUS partners will have no control, as in the case of scenarios. For information referring to the experience or expert judgement of the participants in the Nexus Dialogues, GoNEXUS partners will have partial control: they will reflect the points of view of the participants, which do not depend on GoNEXUS; but the GoNEXUS partners will be able to ask for clarification or further information if necessary.

1.5 Size of data

The exact size of the data cannot be anticipated at the current stage, since it will depend on the particular configuration of each model and Nexus Dialogue. The expected size of the data is in the order of Tb, given the large amounts of data that are expected to be collected.

To provide an overview of the size of the data, the following examples of the size of an individual dataset are provided:

- 1 .txt file with 180 records on a particular point (15 years of monthly data): 3 KB.
- 1 .xls file with 180 records on a particular point (15 years of monthly data): 24 KB.
- 1 .xlsx file with 180 records on a particular point (15 years of monthly data): 19 KB.
- 1 .nc file with 65 years of daily precipitation and temperature records covering Spain with a regular mesh of ~10 km: 1.3 GB.
- 1 .tiff (geotiff) of an area of 10.000 square kilometers: 102MB

The size of the data collected will be regularly annotated and reported in the updated versions of the data management plan.

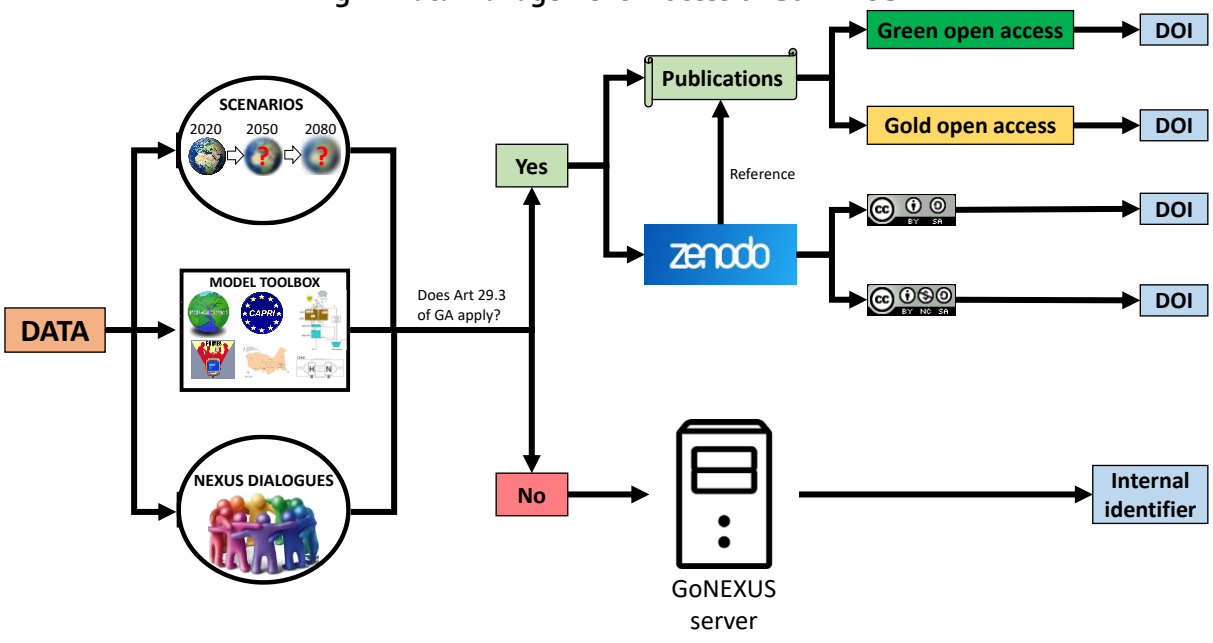
2. FAIR data

2.1 Making data findable

2.1.1 Data Management Process

In compliance with article 29.3 of the Grant Agreement, the management of the data collected and produced by GoNEXUS will follow the process presented in the figure below.

Fig 2 : Data Management Process of GoNEXUS



All data collected or generated by GoNEXUS will be used for at least one of the following purposes, which will together build the evidence:

- To develop the scenarios to be analyzed
- To serve as input for the WEFE models
- To be employed in the Nexus Dialogues

For each data set or subset, the key management decision will be the application or not of Article 29.3 of the Grant Agreement (GA). This article indicates that beneficiaries must, as soon as possible, deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the data, including associated metadata, needed to validate the results presented in scientific publications; and other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (this document). This additional data not used in scientific publication will refer to the compliance of the licenses that will apply to the data collected (e.g. data collected under a Creative Commons Attribution ShareAlike will require to openly share the results in which it was employed). For these datasets, beneficiaries should provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

The data that complies with the requirements of Art 29.3 of the GA will be deposited in the Zenodo public repository (an OpenAIRE and CERN collaboration, https://www.zenodo.org/) or equivalent. Zenodo assigns a Digital Object Identifier (DOI) and a Creative Commons (CC) 4.0 license (https://creativecommons.org/licenses/). This DOI will be properly referenced in the publication(s) that use the deposited data. Data deposited in Zenodo during GoNEXUS will be shared under the Creative Commons ShareAlike (CC-BY-SA) and Creative Commons NonCommercial ShareAlike (CC-BY-NC-SA). The ShareAlike licenses bound the user to share any product using the data following the same license as the data used, while NonCommercial licenses prevent any user to make commercial use of the shared data.

Furthermore, any dataset uploaded into Zenodo will include a description of the tools and instruments required to access the data, if any, together with links to the places in which those would be accessible.

If Art 29.3 is not applicable, data will be stored in the data server that GoNEXUS will develop for data-sharing. Data storage options to be considered and tested if necessary, include the Network Data Access Protocol (OPeNDAP) modular data server Hyrax; the Research Data Management Service Yoda from Universiteit Utrecht (https://www.uu.nl/en/research/yoda/what-is-yoda), and the Network Attached Storage (NAS) server of Universitat Politècnica de València. An internal identifier will be assigned to this data using the standards described below, based on ISIMIP naming conventions. Access to these datasets by the rest of the beneficiaries will be granted under the conditions set by the GoNEXUS Consortium Agreement.

Regardless of the Data Management Process, any dataset collected or generated by GoNEXUS will be properly annotated in the GoNEXUS Data Tables (section 6 of this document). In this section, Table 1 refers to data collected and post-processed by GoNEXUS, while Table 2 refers to data generated. Collected data not post-processed should be excluded from Table 1 and referred directly to the original source. Similarly, Table 2 should only include datasets with further use in GoNEXUS (e.g. to build the evidence or to evaluate the performance of solutions). Datasets referring to intermediate model variables or datasets not further employed should be excluded from this table.

2.1.2 Naming conventions and keywords

Naming conventions will be based on the ISIMIP naming conventions. In any case, all the specifications of any dataset used will be detailed in the GoNEXUS Data Tables (section 6) and in the deliverables using it. The following naming conventions will be applied in GoNEXUS, separated by underscores ("_"). Hyphens ("-") may be kept if they are part of the name of a particular model or scenario (e.g. pcr-glob, CMCC-ESM2). These conventions may be updated based on the experience and requirements identified during the life of the project.

1. *Model name*: name or abbreviation of the model used to produce the datasets (e.g. PCR-GLOBWB V2 could be abbreviated as pcr-glob).
2. *Bias correction*:
   2.1. "adj" if a bias correction method is used. Indication on the particular bias correction method would be possible (e.g. adj-qmap would mean adjusted by quantile mapping).
   2.2. In case of no bias correction method, no text will be added to the name.
3. *Forcing climate projection*: the climate model that was used to force the model producing the datasets (e.g. EC-EARTH). It should be preceded by the family of scenarios (e.g CMIP6, CORDEX) separated by a hyphen. In case of a chain of models, they would be separated by a hyphen (e.g EC-EARTH-RACMO22E).
4. *Forcing landuse model*: the land-use model used to force the model producing the datasets (e.g. IMAGE).
5. *SSP scenario*: the underlying socioeconomic pathway associated with the land use (and climate in the case of CMIP6) model (e.g. SSP1)
6. *RCP scenario*: the underlying CO_2 representative concentration pathway associated with the climate model without the point (e.g. RCP 1.9 would be represented as 19)
7. *Variable name*: If there is a standard variable name abbreviation in the CF Standard Name Table (https://cfconventions.org/Data/cf-standard-names/77/build/cf-standard-name-table.html), this abbreviation name will be chosen (e.g. "pr" for precipitation). Otherwise, an abbreviation will be adopted by GoNEXUS and added to this document.
8. *Region*: the region to which the results are applicable.
   8.1. "glob" (global)
   8.2. "EU" (European Union)
   8.3. "Dan" (Danube)
   8.4. "Como" (Lake Como – Adda)
   8.5. "Juc" (Jucar)
   8.6. "TS" (Tagus-Segura)
   8.7. "Zamb" (Zambezi)
   8.8. "Sen" (Senegal)
9. *Time step*: the time aggregation of the dataset provided (may not correspond to the timestep used by the model).
   9.1. "hour" (hourly)
   9.2. "day" (daily)
   9.3. "mon" (month)
   9.4. "year" (yearly)
   9.5. "5year" (5-year time scale)
10. *Start year*: e.g. 2041
11. *End year*: e.g. 2070
12. *GoNEXUS scenario or set of runs*: this will represent the particular GoNEXUS scenario to be simulated on top of the previous one. The abbreviation will be one of the following (the list will be updated as soon as new scenarios become available):
    12.1. Baseline scenario with no model interconnection (basetier1)

12.2. Baseline scenario with model interconnection (basetier2)
12.3. Scenario run with an updated version of the CAP (updateCAP)

For example, a NetCDF (.nc) file with global discharge (runoff, labelled as mrro in the current version of CF conventions) data from PCR-GLOBWB V2, forced by bias-adjusted climate data from the CMCC-ESM2 model of CMIP6 and the IMAGE land-use model, given a particular combination of RCPs and SSPs, for the tier 1 baseline scenario, would be:

pcr-glob_adj_CMIP6-CMCC-ESM2_IMAGE_SSP3_7.0_ mrro_glob_day_2041_2070_basetier1.nc

2.1.3 Version numbers

GoNEXUS datasets will be given version numbers with the following convention:

- In case of data collected, its version will correspond to the version given to the dataset by the data provider. If the data provided does not indicate any version, the year and month in which the dataset was collected will be indicated (e.g 2020_09).
- In case of data generated by a mathematical model of the GoNEXUS Model Toolbox for which a stable version already exists before the start of the project (e.g. CAPRI, LISFLOOD-EPIC), its version will correspond to the version of the model. If the model does not use versions, the date (year_month_day) in which the dataset was generated will be indicated (e.g 2020_09_01).
- In case of data generated by a mathematical model of the GoNEXUS Model Toolbox without a stable version before the start of the project (e.g. System Dynamics model for the Tagus-Segura basin), an explicit division between the results obtained before and after the release of the first stable mention will be made. In case the mathematical model uses numerical versions, the versions before the first stable release will start by zero (e.g 0.19). In case the mathematical model does not use numerical versions, the prefix "prel_" will be added to the date on which the dataset was generated. Once the first stable version of the model is released, the provisions set in the previous point will apply.

The version number will be given at the end of the name of the file described in section 2.1.2. For example, the previous PCR-GLOBWB V2 dataset would have the version number "v2" at its end:

pcr-glob_adj_CMIP6-CMCC-ESM2_IMAGE_SSP3_7.0_mrro_glob_day_2041_2070_basetier1_v2.nc

2.1.2 Folder structure in the internal GoNEXUS server

To facilitate the navigation through the internal server of GoNEXUS, the folder structure specified below will be adopted. This structure will be updated if found adequate depending on the

- Scenario data
    - Climate scenarios
        - CMIP5
        - CMIP6
    - Landuse and socioeconomic scenarios
        - SSP1
        - SSP2
        - SSP3
        - SSP4
        - SSP5

    - Policy scenarios
- Baseline tier 1 model runs (without model interconnections)
    - Global models
        - PCR-GLOBWB
        - CAPRI
        - PROMETHEUS
        - GEM-E3
        - GLOBIO
    - Continental models
        - LISFLOOD-EPIC
        - PRIMES
    - River basin models
        - High solution models
        - Many-objective models
        - Hydroeconomic models
        - System Dynamics models
        - Behavioural models
- Baseline tier 2 model runs (with model interconnections, with the same structure and baseline tier 1 run)
- GoNEXUS solution runs (1 folder per solution, each with the same structure as baseline runs)

2.2 Making data openly accessible

As indicated before, in compliance with Art 29.3 of the GA, the collected and generated data made openly available to the general public by default will be the data required to validate the results presented in the GoNEXUS scientific publications, as well as when licenses and regulations bound to make data openly accessible. This data will be accessed through the Zenodo repository or equivalent, being accessible by the internet with no additional tools than the ones required to visualize the datasets.

Apart from the previous situations, open access to the data collected and generated by GoNEXUS will be restricted to the conditions of not hindering GoNEXUS obligations on protection (Art. 27 GA) and exploitation (Art 28 GA) of results. The decision on granting open access to the datasets collected and generated will be made by the Steering Committee, after verifying that the access granted does not collide with the Initial Exploitation Plan (D8.7) and the Final Exploitation Plan (D8.8). The Steering Committee will grant open access to all the datasets generated or to a particular subset of data uploading the datasets into the Zenodo repository.

2.3 Making data interoperable

All the datasets collected and generated by GoNEXUS will be included in the Data Table (section 6), which will give information on the formats, units and conventions used. Unless specified otherwise, GoNEXUS will follow the International System of Units and the standard conventions related to the particular data formats if any (e.g. NETCDF .nc files will follow the NETCDF4 CF 1.8 convention, http://cfconventions.org/). In case of using its own conventions, the datasets affected by them will be indicated in the Data Table and the specifications of those conventions will be included in this document to guarantee the interoperability of the related datasets.

2.4 Increasing data reuse

In order to maximize the reuse of the openly accessible data of GoNEXUS, Creative Commons (CC) 4.0 licenses (https://creativecommons.org/licenses/) will be used in all the datasets shared unless agreed otherwise by the Steering Committee. In particular, the following CC licenses will be used:

- Creative Commons ShareAlike (CC-BY-SA): This license lets others freely share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) under the following terms: (i) attribution (any user must give appropriate credit, provide a link to the license, and indicate if changes were made; they may do so in any reasonable manner, but not in any way that suggests the licensor endorses they or their use); (ii) no additional restrictions (they may not apply legal terms or technological measures that legally restrict others from doing anything the license permits). This is the license used by Wikipedia and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.
- Creative Commons NonCommercial ShareAlike (CC-BY-NC-SA): This license lets others remix, adapt, and build upon your work non-commercially, as long as they credit you and license their new creations under identical terms.

The Zenodo repository allows the user to introduce the desired Creative Commons License and it is automatically shown to the users accessing these datasets. The data made openly accessible in compliance with Art 29.3 of the GA will be made accessible and available for reuse the day in which the first version of the related manuscript is sent to the scientific journal, and will be permanently available in Zenodo. In the unlikely event that Zenodo stops its operation, the corresponding author of the paper will upload the data into another repository with features comparable to Zenodo.

3. Allocation of resources

The costs for making data FAIR in GoNEXUS will consist mainly of the personnel costs associated with post-processing the data collected and generated to adapt it to the formats and conventions set in this Data Management Plan. Furthermore, there will be personnel costs associated with uploading data into the Zenodo repository in case of data complying with Art 29.3 of the GA or data that is shared openly by regulation or by decision of the Steering Committee.

The Direct Personnel Costs item of GoNEXUS budget already takes into account these costs. In any case, the staff effort required to make data FAIR is not considered significant compared to the global Direct Personnel Costs. Given that the Zenodo repository has no costs associated with depositing data, there will be no other costs associated with making data FAIR.

During the project, the Steering Committee will be the main responsible for data management, which will follow the guidelines of this Data Management Plan. In any case, decisions on data management will conflict with the exploitation of the project's results foreseen by the Initial Exploitation Plan (D8.7) and the Final Exploitation Plan (D8.8). Its members will be responsible for informing their respective research teams on the provisions set by this Data Management Plan and monitoring its compliance in their own institution.

After the project end, if not specified by the Final Exploitation Plan (D8.8), the Steering Committee will appoint, from the project PIs and the Exploitation Manager, and for each data set or subset, one

or several data curators that will have the same responsibilities as the Steering Committee for the project afterlife.

4. Data security

4.1 Data made openly accessible through Zenodo

The data shared openly using the Zenodo repository will be subject to the provisions made by Zenodo (https://about.zenodo.org/policies/ and https://about.zenodo.org/infrastructure/):

- Zenodo security guidelines:
  - *CERN Data Centre*: The Data Centre used by Zenodo is located on CERN premises and all physical access is restricted to a limited number of staff members with appropriate training and who have been granted access in line with their professional duties. Zenodo staff do not have physical access to the CERN Data Centre).
  - *Servers*: Zenodo servers are managed according to the CERN Security Baseline for Servers, meaning e.g. remote access to them are restricted to Zenodo staff with appropriate training, and the operating system and installed applications are kept updated with the latest security patches via our automatic configuration management system Puppet.
  - *Network*: CERN Security Team runs both host and network-based intrusion detection systems and monitors the traffic flow, pattern and contents into and out of CERN networks in order to detect attacks. Access to zenodo.org happens over HTTPS, except for static documentation pages which are hosted on GitHub Pages.
  - *Data*: Zenodo stores user passwords using strong cryptographic password hashing algorithms (currently PBKDF2+SHA512). Users' access tokens to GitHub and ORCID are stored encrypted and can only be decrypted with the application's secret key.
  - *Application*: they are employing a suite of techniques to protect the sessions from being stolen by an attacker when logged in and run vulnerability scans against the application.
  - *Staff*: CERN staff with access to user data operate under CERN Operational Circular no. 5, meaning among other things that:
    - Staff should not exchange among themselves information acquired unless it is expressly required for the execution of their duties.
    - Access to user data must always be consistent with the professional duties and only permitted for resolution of problems, detection of security issues, monitoring of resources and similar.
    - Staff are liable for damage resulting from any infringement and can have access withdrawn and/or be subject to disciplinary or legal proceedings depending on the seriousness of the infringement.
  - *Replicas*: All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Zenodo retention period: Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental program defined for the next 20 years at least.

4.2 Data stored in an internal server

Hyrax

The specifications on Hyrax security may be found in https://opendap.github.io/hyrax_guide/Master_Hyrax_Guide.html#_hyrax_security. GoNEXUS will follow all the Hyrax recommendations to maximize the security of the stored data such as:

- Keep the Hyrax server behind a firewall, configure the firewall to only forward requests to the appropriate port on the Hyrax system, and use the firewall to block direct access to the Back End Server (BES).
- Use separate machines for the OPeNDAP Lightweight Front end Server (OLFS) and the BES.
- Restrict access to the log files generated and the configuration files used.
- Hyrax will be run by restricted users, who should have restricted privileges and should only be allowed to write to the directories required by Tomcat and the BES. Tomcat will be configured to demand user authentication.

UU Yoda

The specifications on Yoda security may be found in https://www.uu.nl/en/research/yoda. Yoda complies with Utrecht University's Information Security policy for data classified as public, internal use, sensitive or critical.

UPV NAS

The allocation of space in the UPV NAS server will be done under UPV premises on security. Security issues will be handled by the Information and Communication Technologies (ICT) Area (ASIC) of UPV (http://www.upv.es/entidades/ASIC/index.html). The GoNEXUS team will follow the instructions of ASIC concerning security issues.

5. Ethical aspects

The main ethical issue of GONEXUS concerning data refers to the use and processing of personal data since the scenarios and Nexus Dialogues will collect data from stakeholders. Furthermore, part of this data will be exchanged between EU countries and third countries. Personal data is any information that relates to an identified or identifiable living individual, including pieces of information that collected together can lead to the identification of a particular person. Examples of personal data include a name and surname; a home address; an email address such as name.surname@company.com; an identification card number; and location data (for example a location of a borehole). Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible. For example, time series of pumped water from a particular location are considered personal data if information about the location is included, but they are not personal data if they are presented without any information on where is it from.

The GoNEXUS consortium will inform users on which data will be collected from them, how it will be processed and, in case the user or users to which each dataset refers could be identified, which

measures will be taken by the consortium to guarantee the privacy of all users involved unless we obtain their explicit prior informed consent.

The project will comply with the international and national regulations in order to observe strict ethical standards and guarantee participants rights referring privacy protection:

- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016); and its corrigendum on 23rd May 2018.
- In January 2012, The European Commission proposes a comprehensive reform of the EU's 1995 data protection rules to strengthen online privacy rights and boost Europe's digital economy, 2012/0011 (COD).
- Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communication sector.
- Handbook on European data protection law by the European Union Agency for Fundamental Rights and the Council of Europe (2013).
- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (Spain).

In case of conflict between EU and third country legislation, the strictest requirements between them will be applied for each particular situation. Personal data transfers between EU and third countries will be explicitly indicated.

5.1 Informed consent and assent

Informed consent and assent are required when research involves the participation of users and stakeholders and is meant to guarantee the voluntary participation and understanding of the research purposes. Informed consent and assent consist of three components: adequate information, voluntariness and competence. Users should be clearly informed of the research goals, methods used for data collection and processing, possible adverse events, possibilities to refuse participation or withdraw from the research, at any time, and without consequences. Information to the users should be provided in such a way that they fully understand it and are aware of the implications of their assent.

GoNEXUS users and stakeholders involved will be provided at least with the following information beforehand through proper Information Sheets:

- The purposes of the research and information about what will happen with the results of the research.
- The experimental procedures and a detailed description of the involvement of the participants.
- All benefits to the participants or to others that may reasonably be expected to occur.
- A description of the procedures adopted to guarantee the participants privacy.
- Contact details for researchers who can be contacted to answer pertinent questions about the research and the participant's rights.

- A clear statement that the participation is voluntary, that the refusal to participate will involve no penalty and that the participant may decide at any time, to discontinue participation without penalty.
- General information about the research project.

GoNEXUS will demand and collect Informed Consent Forms from users concerning their understanding of the content of the Information Sheets and their consent to participate in the project. These forms will be safely collected, handled and preserved by the partners, and will be available to the European Commission if required. UPV will have track of all the Informed Consent Forms collected by partners in each case study and will include the number of how many users signed the form in each case study in this Data Management Plan (D1.1).

5.2 Data privacy

Laid down as principles in the Charter of Fundamental Rights and the Treaty on the Functioning of the European Union, privacy and data protection are fundamental rights that need to be protected at all times. Data protection in GoNEXUS is meant to guarantee participants right to privacy and refers to the technical framework and security measures designed to guarantee that all personal data are safe from unforeseen, unintended or malevolent use. Users involved will be informed about the procedures used to collect, store and process data.

The principle of proportionality will be observed, and no more data than necessary for the research and validation purpose will be collected. Personal data collection will be adequate and relevant. Data collected will be anonymous and.

Storage, processing and communication of personal data will be undertaken in conformity with the EU General Data Protection Regulation (Regulation (EU) 2016/679 and its corrigendum on 23rd May 2018) and the relevant laws in each country. Each partner will identify the applicable local or national legal requirements in their countries and the competent authorities to provide the necessary authorizations for the handling of personal data.

Any research action involving collecting personal data will be previously evaluated and approved by the Steering Committee, and no personal data will be gathered without the informed explicit consent of the participant. All personal information will be coded or anonymized as early as possible in the data processing.

6. GoNEXUS Data Tables

Table 1: Data collected and post-processed

Do not include raw datasets (they should be referred to the original source)

DATASET NAME	MODEL NAME	VARIABLE	BIAS CORRECTED	FORCING CLIMATE MODEL	FORCING LANDUSE MODEL	SSP SCENARIO	RCP SCENARIO	VARIABLE NAME	REGION	TIME STEP	PERIOD	GoNEXUS SCENARIO	VERSION	FORMAT	DATA PRODUCER	OBSERVATIONS	CONTACT FOR QUERIES

Table 2: Data generated

Include only datasets further used in GoNEXUS (e.g. to build the evidence, or to analyse the performance of solutions)

DATASET NAME	MODEL NAME	VARIABLE	BIAS CORRECTED	FORCING CLIMATE MODEL	FORCING LANDUSE MODEL	SSP SCENARIO	RCP SCENARIO	VARIABLE NAME	REGION	TIME STEP	PERIOD	GoNEXUS SCENARIO	VERSION	FORMAT	DATA PRODUCER	OBSERVATIONS	CONTACT FOR QUERIES