The Added Value of Seasonal Climate Forecasts for Integrated Risk Management Decisions (SECLI-FIRM)

EU H2020 Project (ref. n. 776868)

D6.8: Data Management Plan Update (v2)

[Dissemination level: Public]

Version Table

Name/Party	Description	Date
A Troccoli (UEA)	Updated deliverable v1	14/05/19
A Troccoli (UEA)	Review/Edit of Final draft and finalise deliverable	31/07/19
A Troccoli (UEA), L Sanger (WEMC), K Nielsen (UL)	Updated deliverable v2	24/07/20
A Troccoli (UEA)	Review/Edit of Final v2 draft and finalise deliverable	31/07/20

Internal Review Table

Name/Party	Description	Date
P. Bett (MO)	Review of content/edits	6/08/18
L. Sanger (WEMC)	Edits to updated deliverable v1	23/05/19

Contributors (Consortium Party/Person)

UEA	A. Troccoli
ENEL	
ENEA	
MET Office	P. Bett
UL – AWS Truepower	K Nielsen
KNMI	
WEMC	L Sanger
Eurac	
Alperia	
Météo France	

This document has been produced within the scope of the SECLI-FIRM project. The utilisation and release of this document is subject to the conditions of the grant agreement no. 776868 within the H2020 Framework Programme and to the conditions of the SECLI-FIRM Consortium Agreement.

Image on front cover courtesy of rawpixel.com from Pexels

Table of Contents

1 Introduction

This Data Management Plan (DMP) describes the life cycle of the data utilised by the SECLI-FIRM project. It includes the data that the project plans to generate, collect, process, store, archive, keep safe and possibly make available to others. It will be used as a standard data guideline for the project consortium to ensure interoperability and possible ultimate sharing for public use of the datasets generated by the project. It also identifies issues of data risk or governance that need particular action during the execution of the SECLI-FIRM project. Treatment of personal information related to stakeholders, such as telephone numbers, is also included in this DMP.

One of the main goals of the DMP is to assist the SECLI-FIRM consortium in ensuring that the research data produced by the project is findable, accessible, interoperable and reusable (FAIR), and that it is soundly managed, in accordance with EU Horizon 2020 (H2020) guidelines introduced with the Open Research Data Pilot (ORD pilot). The ORD pilot aims to improve and maximise access to and re-use of research data generated by H2020 projects and takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions. The ORD pilot applies primarily to the data needed to validate the results presented in scientific publications. The SECLI-FIRM project therefore plans to adhere to the ORD pilot approach as closely as possible.

More specifically, the SECLI-FIRM DMP contains information on:

- The handling of research data during and after the end of the SECLI-FIRM project
- The type of data collected, processed and/or generated by SECLI-FIRM
- The methodology and standards applied to the SECLI-FIRM data
- SECLI-FIRM data sharing and procedures
- The storage, security and long-term preservation of the SECLI-FIRM data.

At this stage of the project, the main technical data being considered are the output of seasonal climate forecast models and of the ERA5 reanalysis, which are centrally retrieved onto the project server. The conventions set out in this DMP will also assist with the generation and handling of other data that will be generated as part of the implementation of the technical work, such as statistics, skill assessments, multi-model fields, etc. Conventions for new data (possibly including software) will also provide a useful reference implementation (likely in R or Python), complete with the lookup tables for the different fields, in a version-controlled repository accessible to the project partners (e.g. on github).

This DMP will evolve over the course of the SECLI-FIRM project to reflect changes such as:

- new data produced by the project
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

With this update, aside from reviewing the entire DMP, a new section on *homogenization procedures* (3.1.5), in line with the work reported under Task 2.1, D2.1, has been introduced.

Additional changes to the DMP will be documented accordingly in future official releases (the next and final one is currently due at month 42, namely July 2021).

1.1 DMP – This Update

This DMP is an update of the original DMP – deliverable report D6.4, submitted in Month 6, July 2018. The key updates relate to:

- the reference models resolution (section 3.1.1)
- the use of more standard abbreviations for climate and energy variables (section 3.1.3)
- the inclusion of new variables that have been added to the SECLI-FIRM database since the previous DMP (section 3.1.3)
- the filename conventions for the model output (section 3.1.3)
- the directory structure conventions for the model output (section 3.1.3)

2 Overview of the SECLI-FIRM Project

The central objective of The Added Value of Seasonal Climate Forecasts for Integrated Risk Management Decisions (SECLI-FIRM) project (Feb 2018 – July 2021) is to demonstrate how the use of improved climate forecasts, out to several months ahead, can add practical and economic value to decision-making processes and outcomes, primarily in the energy sector, but also in the water sector. Specifically, for the energy sector, SECLI-FIRM is assessing the impact on operational planning and portfolio management, such as hedging and asset optimization, thus enabling quantification of the value-add provided by seasonal forecasts which have been calibrated, evaluated and tailored for each specific application. Improvements in management decisions will ultimately lead to an improved supply-demand balance and therefore to a more efficient energy system, particularly with respect to renewable energy, with corresponding benefits for climate change mitigation.

A simple, but effective, methodology will be used to assess value added. A control case will typically utilise climatological conditions based on historical averaged values – currently the most common approach – while a test case will also consider individually optimised and tailored state-of-the-art probabilistic seasonal forecasts. This will be done for nine case studies

for Europe and South America: recent seasons with anomalous/extreme climate conditions leading to problematic and quantifiable impacts for the energy and/or water industry. Crucially for success, the case studies will be co-designed by industrial and research partners.

3 Data produced by SECLI-FIRM

The purpose of the data collected and generated by SECLI-FIRM is in the first place to fulfil the objective of the project, namely the assessment of the added value of the use of seasonal climate forecasts for the decision making in the nine case studies. To this end, access to the seasonal climate forecasts should be as practical as possible for project partners and industrial stakeholders contributing to the case studies. Collection, generation and distribution of data will therefore be as uniform across originators as possible as well as practical, which often means also dispensing with highly technical formats and/or specifications of the source data.

From a scientific/technical point of view, the data utilised by the SECLI-FIRM project is related to seasonal climate forecasts as well as historical climate data – typically re-analyses – for the calibration of the seasonal forecasts. It is also possible that SECLI-FIRM will develop some station/reanalysis-based observed climatologies for benchmarking the value added, although in the majority of cases these data sets are already held by the respective industrial partners (and are not encompassed by this DMP). Energy and water data will also be utilised to the extent that they are needed to tune the forecasts and to perform the assessment of the seasonal forecast value-add.

Seasonal climate forecast model data are accessible from a number of distribution portals around the world, which pertain to different producing centres (e.g. http://www.wmolc.org/). However, in general each producing centre adopts different conventions, including different data retrieval methodology (whether by FTP or by retrieval through a script or by a visual interface), spatial and temporal resolution of the forecast output, duration of the retrospective forecasts, number of ensemble members, etc. – details of the different features will be described in deliverable report D2.1, due January 2019. For the purpose of this deliverable, we mainly cover the general approach to data management.

In terms of metadata, and filename conventions, we also refer to relevant references, such as conventions developed by the INSPIRE Directive (2007/2/EC)[1], or the Copernicus Climate Change Service (C3S), Group of Earth Observations (GEO), or Climate and Forecasts (CF). The latter are principally used for NetCDF data files:

- CF conventions general definitions http://cfconventions.org
- Standard name tables from CF conventions: http://cfconventions.org/Data/cf-standard-names/37/build/cf-standard-name-table.html

[1] "INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119" available at: http://inspire.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf

- NetCDF4 Common data model[2]
- Unidata Common data model[3]

3.1 Data Summary

In this section we identify the main characteristics of the data to be used by the SECLI-FIRM project. These characteristics include:

- Domain, resolution, and reference periods
- List and description of variables
- Filename conventions
- File structure and format
- Quality assurance
- Title, summary and keywords
- Author(s), project name and date of creation
- Rights/license
- Lineage
- Link to related datasets

3.1.1 Domain, resolution and reference periods

While various regional geographical domains will be considered in SECLI-FIRM based on the specific case studies, whenever possible climate variables will be provided for the entire globe. This is also to allow the possibility for some of the research investigations to produce indicators such as large-scale patterns. The definition of climate variable sub-domains for specific research tasks and/or case studies will be discussed once this work gets underway, but it is expected that this will be guided by the domain covered by each case study. In the case of sectoral data (energy and water), the choice is clearer as these will by necessity refer to the case study domains.

There will be two reference geographical resolutions for climate variables:

- The highest available resolution for each seasonal climate forecast data set, which is typically 1° by 1°, but for some producing centre it is lower, namely 2.5° by 2.5°; we will normally only consider the higher 1° by 1°resolution;
- A common resolution of 1° by 1°, whereby all seasonal model output is mapped onto a common grid, as defined for the Copernicus Climate Change Service (C3S) data or for the North American Multi-Model Ensemble (NMME) – the two are very similar, typically at 1° by 1°, even if they can differ on the number of latitudinal points (180 or 181).

[2] http://www.unidata.ucar.edu/software/netcdf/workshops/2008/netcdf4/Nc4DataModel.html
[3] http://www.unidata.ucar.edu/software/thredds/current/netcdf- java/CDM/

The historical reference data used to calibrate seasonal forecasts will be taken from re-analyses products, and in most cases from the latest ERA5 dataset. As this has a resolution higher than the highest seasonal climate forecast output (in the case of ERA5 it is about 0.25° by 0.25°), the geographical resolution for the historical reference period will be adjusted to match that of the forecast data. Therefore two versions of the historical dataset will be produced – one at 1° by 1° and another at 2.5° by 2.5°. The same approach would work for other reanalyses such as ERA-Interim: this has a resolution of about 0.8° by 0.8°. Use of ERA-Interim would be required in the case the ERA5 data which are currently being produced did not extend as far back as the seasonal climate reforecast data.

For sectoral data, the spatial resolution will be the highest available, while also ensuring representativeness issues are minimised. For instance, in case of electricity demand, the highest geographical resolution is sub-country zones, which in the case of the Italian power market is six zones.

As with the spatial resolution, the temporal resolution of climate variables will be collected and provided at both the highest resolution (6 hours for some variables) and the common resolution, which is daily in many cases, although some producing centres only provide output as monthly averages. Analogous considerations apply to the historical data as well as to the sectoral data.

The reference period typically used for both the calibration and the assessment of seasonal climate forecasts, through the use of retrospective forecasts (sometimes referred to as hindcasts), is the 30-year period 1981-2010. Shorter periods might also possible when retrospective forecasts do not cover the full 30-year period, although the robustness of the calibration may reduce and the assessment statistics become noisier. The latter also applies to the number of available ensemble members for the retrospective forecasts: assessment statistics can even be biased if ensemble sizes are too small.

3.1.2 List and description of variables

In SECLI-FIRM, variables are divided into climate and sectoral (energy or water) data categories. These variables have been selected following consultations with SECLI-FIRM industrial partners and stakeholders. Note however that the exact list of variables may be refined, depending, amongst other things, on the actual availability of output variables from the various seasonal forecast models.

Table 1 provides the list of climate variables, along with their main specifications. Note that some variables are indicated as critical, as they are both most useful for our case studies and most common across the global seasonal forecast models. Note that in Table 1 only the basic meteorological variables are listed, but it is possible that new variables may be added to account for indices such as weather typing and teleconnection indices, in case they are derived from multiple meteorological variables. Whenever there are multiple standard names, because

of different possible conventions used by the different forecast producing centres, more than one name may be found. In SECLI-FIRM, we will attempt to use a single name for a specific climate (or sectoral) variable.

The sectoral energy and water variables are listed in Table 2. Note that in terms of solar power, only PhotoVoltaic (PV) power will be considered in SECLI-FIRM as current penetration of concentrating solar power is still marginal in the overall energy mix, except perhaps in Spain.

3.1.3 Filename conventions

In order to assist with the exchange of information within the SECLI-FIRM project, a detailed filename convention is required. Table 3 shows a list of descriptors that will form the filename structure for climate and sectoral variables, alongside an example for each descriptor. The aim is to capture as many possible instances of data structures so as to reduce ambiguity, while trying to limit the number of components that form a filename. Having fixed fields (number of characters) per component, and separated by an underscore, will help with the writing of robust software code. To ensure all file names have the same length, leading minus sign (i.e. -) are also used when the specific component has fewer characters than the allocated ones for the descriptor being considered.

Look-up tables will be used to access the many combinations of filename components. A full catalogue of entries for each filename descriptor will be constructed during the execution of SECLI-FIRM. Ideally a database containing all the descriptors, and the relative Uniform Resource Identifier (URI) pointing to the data file, should also be constructed.

Table 1: Climate variables used by SECLI-FIRM. Variables denoted with an asterisk (*) are critical to the objectives of SECLI-FIRM. Others are useful and in some cases they can be derived (e.g. wind speed at 100 m height).

Climate Variable	Unit	Abbreviation
*Air temperature (at 2 m height)	°C	TA
*Total Precipitation	mm	TP
*Global Horizontal Irradiance	W m^{-2}	GHI
*Wind Speed (at 10 m height)	m s^{-1}	WS
*Geopotential height (at 500 hPa)	m	GPH
*Wave height	m	WH
*Snow depth	m	SDP
Snow fall	m	SF
Snow density	kg m^{-3}	SND
Sea Surface Temperature	°C	SST

Air temperature (at 850 hPa)	°C	TA
Mean Sea Level Pressure	hPa	SLP
Sunshine hours	hr	SH
Relative humidity	%	RH
Zonal wind component	m s^{-1}	U
Meridional wind component	m s^{-1}	V

Table 2: Sectoral variables used by SECLI-FIRM.

Sectoral Variable	Unit	Abbreviation
Power price	€ or £	PP
Gas price	€ or £	GP
Power/Electricity demand	Wh	EDM
Hydro power run-of-river	Wh	HRO
Hydro power reservoir	Wh	HRE
Wind power onshore	Wh	WON
Wind power offshore	Wh	WOF
Solar power (PV)	Wh	SPV
Thermal power	Wh	THP
Water demand	m^3	WAD
Water supply	m^3	WAS
Infrastructure maintenance duration	day	IMD
Outage/Shortage duration	day	OSD

Table 3: Descriptors of filenames for both climate variables. Only one entry per descriptor is shown here.

Filename Descriptor	n char	Historical Data	Seasonal Forecast
Category	1	Historical: H	Seasonal: S
Originator	5	ECMWF: ECMWF	ECMWF: ECMWF
Generation	3	ERA5: ER5	System 5: S05
Variable	3	Air T: TA-	As for historical
Level	5	2 m: 0002m	As for historical
Region	4	Global: Glob	As for historical

Spatial Resolution	4	1.00 deg: 100d	As for historical
Latitudinal points	4	181 points: i181	As for historical
Temporal Resolution	3	6 hours: 06h	As for historical
Valid Time	10	---	DD_MM_YYYY
Start Date	8	YYYYMMDD	---
End Date	8	YYYYMMDD	---

An example of filename for historical data (ERA5 reanalysis) is thus:

S_ECMWF_ER5_U--_0010m_Glob_100d_i181_06h_19930101_20161231.nc

and for a seasonal climate forecast model output it is:

S_ECMWF_S05_U--_500hP_Glob_100d_i181_12h_01_04_1993.nc

Additional descriptors for filenames, for future use, are provided in Table 4.

Table 4: Additional descriptors of filenames for both derived climate variables or sectoral variables. Only one entry per descriptor is shown here.

Filename Descriptor	n char	Historical Data	Seasonal Forecast
Bias Adjustment	3	No adjustment: noa	Mean bias adj: mba
Stats	3	Original: org	Rank Prob. Skill Score: RPS
Ensemble number	2	N/A: NA	1: 01
Transfer function	5	Wind power curve: wpc01	As for historical

A balance needs to be found between sub-directory levels and number of files in each sub-directory. Accordingly, the directory tree structure will use the descriptors listed in Table 5.

Table 5: Descriptors for directory names.

	Max n. char	Descriptor
Sector	4	CLIM: Climate ENER: Energy WATE: Water
Category	4	HIST: historical SEAS: seasonal forecast
Model/System	10	ERA5: C3S ERA5 reanalysis C3S_ECMWF CMC-CanCM4

Variable	3	See Tables 1 and 2
Temporal Resolution (Seasonal forecasts only)	3	MON: Monthly means ORIG: 3-hourly/6-hourly/12-hourly/daily (depending on the variable)

For instance, the directory structure for air temperature would be 'CLIM/HIST/ERA5/TA' for historical data or 'CLIM/SEAS/C3S_ECMWF/TA/MON' for seasonal forecasts.

3.1.4 File structure and format

The SECLI-FIRM platform manages alpha-numerical and geographical datasets. The main data formats used for SECLI-FIRM implementation are shown in Table 6.

For the research component of SECLI-FIRM, the data format most commonly used will be NetCDF, and occasionally GRIB, although there are softwares that can easily convert the less common GRIB format into netCDF. While (gridded) climate data are usually available in NetCDF format, SECLI-FIRM industry stakeholders often prefer files that are easy to handle with standard software such as Excel. As such, the file structure should be as simple as possible to allow users to open and process the data with minimal effort. Time series or relatively small gridded data are therefore preferred. As a consequence, the most common file format used to exchange data with the industry stakeholders will be Comma Separated Values (CSV) (or equivalent). However, databases like SQL or Access as support to share metadata may be considered too. Even with simple files like CSV, metadata other than those contained in the filename structure will be provided, also as a redundancy check. An example of metadata descriptor to be inserted at the start of the CSV (or similar) files is provided in Appendix A.

Also, as a way to enhance data inter-operability, use of the XML format will be investigated. Critical information, such as forecast start and valid time (hence a double time-axis) for seasonal forecasts, should be included in the file.

Table 6: Data and file standards used in SECLI-FIRM.

Data Type	Standards
Climate variables (re-analysis and seasonal climate forecasts)	NetCDF (or GRIB)
User geographical layers	Shape files (SHP)
Sectoral data	CSV or equivalent
Coordinate system	WGS84
Date & time	ISO 8160
Internal geographical layers	GeoJSON

3.1.5 Homogenization procedures

Spatial/temporal homogenization of the seasonal forecast data is a valuable procedure to ensure the SECLI-FIRM dataset which is created from different forecasting systems are organised in the same way, and following our DMP conventions. For instance, seemingly simple differences such as different order in the parameters in the arrays (within the NetCDF files) representing the forecasts, could lead to unnecessarily elaborate and time consuming pre-processing. Indeed, the purpose of constructing a homogenous SECLI-FIRM dataset is to remove the complexity in the handling the data and instead fully focus on the analysis of the data.

Datasets are retrieved from multiple centres, each with their own associated data conventions. For example, ECMWF, UK Met Office and NASA all have differing practices for file naming, NetCDF metadata organisation and distributing their datasets.

Retrieved datasets are initially processed to homogenise the filename according to the conventions described in section 3.1.3. This will retain the source centre name, date and variable names from the input filenames as identifiers for the output filename. Python and native packages (os), string slicing and manipulation are utilised for this procedure and the script can be easily updated to allow for new data sources and their associated file naming conventions.

Climate Data Operators (CDO) and NetCDF Operators (NCO) are toolsets of efficient command line operations for manipulating climate data. CDO is utilised in the homogenisation procedure to remove unnecessary dimensions from the NetCDF cube data. For SECLI-FIRM, we are only concerned with the variable and its associated time, number, longitude and latitude dimensions. First is to remove all additional dimensions such as the date the forecast was runned (S). The function below shows an example of a way of doing this by simply averaging over that single unit:

ncwa -a S infile.nc outfile.nc

NCO can then be used to rename the remaining dimensions, utilising the **ncrename** function. For example a data centre may label latitude/longitude as *lat/lon* or time as *date* etc. and require renaming for homogenisation to the rest of the dataset. The example below shows the function used to rename the dimension level to number in the file.nc.

ncrename -h -O -v level,number file.nc

Finally, below is an example of the NCO function used to re-order the dimensions inside the NetCDF file.

ncpdq -a time,number,latitude,longitude infile.nc outfile.nc

This results in the following structure.

Var(time,number,longitude,latitude)

time describing the lead time of the forecast in units of hours since 1900-01-01, *number* being the different ensemble members and *latitude* and *longitude* in degrees.

After the homogenisation procedure, NetCDF metadata structure should look like the following (using ncdump):

```
dimensions:
        longitude = 361 ;
        latitude = 181 ;
        number = 25 ;
        time = 6 ;
variables:
        float longitude(longitude) ;
                longitude:units = "degrees_east" ;
                longitude:long_name = "longitude" ;
        float latitude(latitude) ;
                latitude:units = "degrees_north" ;
                latitude:long_name = "latitude" ;
        int number(number) ;
                number:long_name = "ensemble_member" ;
        int time(time) ;
                time:units = "hours since 1900-01-01 00:00:00.0" ;
                time:long_name = "time" ;
                time:calendar = "gregorian" ;
        short TA(time, number, latitude, longitude) ;
                TA:scale_factor = 0.00478337249935147 ;
                TA:add_offset = 156.73198331375 ;
                TA:_FillValue = -32767s ;
                TA:missing_value = -32767s ;
                TA:units = "K" ;
                        TA:long_name = "2 metre temperature" ;
```

3.1.6 Quality assurance

For Quality Assurance (QA) of climate data we produce standard plots to ensure no obvious non-physical values are present through visual inspection. We also compare the seasonal forecast output with ERA5 reanalysis to screen for evident outliers. As a pragmatic approach, station data or even reanalysis products are assumed to be error free. However, we ensure that the data source is of adequate documented quality. Also, if a sufficiently high number of

observations is available, cross-comparison is applied to check data quality. The gridded products are likely poorer in quality in data sparse regions and this potentially also applies to the reanalyses. For sectoral variables, analogous procedures to those for the climate data are applied.

3.1.7 Title, summary and keywords

The title of the dataset produced by SECLI-FIRM, a brief description (one paragraph) and a comma separated list of keywords relevant to the dataset will also be provided. Reference will be made to the NetCDF attribute convention or other metadata formats (e.g. ISO 19115).

3.1.8 Author(s), project name and date of creation

The name of the people involved in the creation of each dataset and their contact details and institution are typically provided, except when the author of the data does not authorise the project to publish all this information (see also section 3.6 on personal data management). The metadata also includes the name of the project (SECLI-FIRM) and the date when the dataset has been created.

3.1.9 Rights/license

The description of the restrictions about data access, re-use and distribution adhere to good-practice as with the European Environment Agency metadata http://www.eea.europa.eu/data-and-maps/figures/share-of-renewable-electricity-in#tab-metadata example or the Eurostat one http://ec.europa.eu/eurostat/cache/metadata/en/med_esms.htm

3.1.10 Lineage

Information about lineage will be recorded for all data used by SECLI-FIRM to ensure traceability. Where available, we also note the version number of the product, but not all datasets incorporate version numbering systems. Similarly to the information contained within NetCDF headers, information about the processing performed on datasets (e.g. interpolation or time averaging), or information about the software/algorithm that has been used to generate it, will be reported.

3.1.11 Link to related datasets

Data produced by SECLI-FIRM can be connected to other datasets, for example as in the case of indices that are created using multiple meteorological variables (e.g. a link to the temperature dataset that has been used to compute a Heating/Cooling Degree Day index). A link or a reference (when a URL is not available) pointing to all the related datasets is provided.

3.2 Findable, accessible, interoperable and reusable (FAIR) data

3.2.1 Making data findable, including provisions for metadata

The metadata used in SECLI-FIRM are based on international conventions such as the INSPIRE Directive (2007/2/EC) and the ISO 19115/ISO 19119 metadata standard[4], as well as to other international data standards, such as the World Wide Web Consortium (W3C) recommendations[5] or the CF convention for NetCDF files. Ensuring that the metadata follow international standards allows for a better interfacing with scientific and industrial data users in SECLI-FIRM. In this context, interoperability is another important aspect to be taken into consideration.

Variable characteristics will be made available for individual climate and sectoral variables, as indicated by the metadata template presented in Table 7, which is based on that developed by the Copernicus Climate Change Service (C3S) European Climatic Energy Mixes (ECEM)[6] and by the EU project Atopica[7], will be used as the basis for the SECLI-FIRM variable description document, which is provided to assist stakeholders with the selection and interpretation of the data collected and produced by SECLI-FIRM. This information also provides the basis of a metadata catalogue (see Section 3.2.2). Metadata descriptors will also be provided within the data file, in line with but also as a complement to the filename descriptors presented earlier.

3.2.2 Making data openly accessible

Potential open access to some data collected or produced by SECLI-FIRM will be considered by the project Steering Committee, based on various factors including competitive issues especially for industrial partners. Amongst the data to be provided in open access, the output of the SECLI-FIRM research on seasonal climate forecasts is likely to be made available.

If open access is provided, the selected datasets generated during the project, possibly including relevant software, will be archived on a dedicated server. They will be provided under the Creative Commons Licences CC-BY[8]. In that case, access to these datasets will be provided from the public interface of the SECLI-FIRM platform during the life time of the project, allowing access to public metadata and data sets stored in the system; this also includes a metadata catalogue to assist with data finding. In addition, the SECLI-FIRM open data

---

[4] The structure of the ISO 19115/ISO 19119 metadata is included and explained in the document "INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119", available at: http://inspire.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf
[5] https://www.w3.org/TR/tabular-data-model/#embedded-metadata
[6] http://ecem.climate.copernicus.eu/
[7] https://www.atopica.eu/
[8] https://creativecommons.org/licenses/

repository will be deposited in the Registry of Research Data Repositories https://www.re3data.org/.

After the end of the project, access to the data sets produced will likely be offered as a service, under contract and/or licence. This will be further defined as part of the SECLI-FIRM exploitation plan.

Table 7: Metadata descriptors.

1. General	
1.1 Description	How this variable is computed and what it represents for the two periods considered: historical and seasonal forecasts
1.2 Reference date	Date of data publication/revision/creation in format dd/mm/yyyy
1.3 Variable type	Observation/Model output/Derived/Other (specify)
1.4 Unit	Unit of measure of variable and whether it is absolute or a relative change (e.g. anomaly compared to historical period)
1.5 URL	Link in the SECLI-FIRM portal or dataset
1.6 Data format	Downloadable data in e.g. CSV
1.7 Keywords	E.g., Temperature, air
1.8 Contact	Name and email address of person providing the data
2. Dataset coverage	
2.1 Geographic area	Coordinates of region, location, etc. Provide lat/long boundaries / co-ordinates.
2.2 Spatial resolution	Grid resolution or point representation (for obs)
2.3 Temporal resolution	Annual/monthly/seasonal/daily. Time series / long-term averages.
2.4 Time period	Start and end dates
3. Usage	
3.1 License conditions	Specify conditions or else state ‘none’
3.2 Citation(s)	List of relevant references
4. Lineage statement	
4.1 Original data source	Brief description of process history of dataset (e.g. post-processing of original 'raw' data). Note if official / legal dataset
4.2 Linked datasets	Datasets that are ‘linked’ to this one (e.g. datasets that have been used to compute a sectoral indicator)

3.2.3 Making data interoperable

Interoperability describes the extent to which systems and devices can exchange data and interpret that shared data. For two systems to be interoperable, they must be able to exchange data and subsequently present that data such that it can be understood by a user regardless of their physical architecture and operating systems. Interoperability within SECLI-FIRM will be achieved through the use of Open Standards (ISO/W3C/OGC) and the use of Open-Source software. This will allow further development of downstream services based on the output produced by SECLI-FIRM. For instance, OPeNDAP (Open-source Project for a Network Data Access Protocol) is one such standard. It supports the DAP2 protocol that provides a discipline-neutral means of requesting and providing data across the World Wide Web. The goal is to allow project partners and case-study industry stakeholders to immediately access whatever data they require in a form they can use, all the while using applications they already possess and are familiar with. This will facilitate the exploitation of the results of SECLI-FIRM.

3.2.4 Data Sharing

SECLI-FIRM will create a datastore for data sharing between project partners and industrial stakeholders involved in the nine case studies. The datastore is likely going to be based on the OPeNDAP standard via a Thematic Real-time Environmental Distributed Data Services (THREDDS) server, although databases like SQL are being considered too. The role of THREDDS is to provide a coherent access to a large collection of real-time and archived datasets from a variety of environmental data sources at a number of distributed server sites. The THREDDS server is a web server that provides metadata and data access for scientific datasets, using a variety of remote data access protocols[9]. Scripts will be provided to the SECLI-FIRM consortium to download data from the SECLI-FIRM THREDDS server as well as to produce simple plots to perform an initial QA.

There is no other form of data sharing with external third party in the framework of the SECLI-FIRM platform, in particular because the system is managing sensitive data on energy and water management in a very competitive environment.

3.3 Allocation of resources

The cost of making data FAIR will be covered by the SECLI-FIRM budget. However, not all information is currently available to compute a precise estimate of the cost. It is expected that this will be provided with the next update of the DMP. As an indication, the main costs are related to data storage (current estimate is between 500 and 1,000 TB), machines which allow fast access to a THREDDS server as well as computation, and technical expertise.

[9] https://www.unidata.ucar.edu/software/thredds/current/tds/

The data will be made accessible for a period of five years after the completion of the project. However, as part of the exploitation of the project results, options will be investigated during the lifetime of the project to extend the access to the sharable SECLI-FIRM data beyond this period.

The decision on the long-term preservation and access to the data will be taken by the project Steering Committee, with input from the project Advisory Board, but it will also depend on the success of raising interest and additional funds beyond the duration of SECLI-FIRM. Individual partners may also decide to take responsibility to make components of the SECLI-FIRM data accessible. The responsibility of the overall project data management and finance assessment ultimately rests with the SECLI-FIRM coordinator, Prof. Alberto Troccoli.

3.4 Data security

Security of data will be regularly monitored by screening information and statistics in the control panel and messages received by the host service for e.g. the THREDDS server. To tackle both security and reliability of service, backup copies of sensitive and difficult-to-reproduce data will be maintained at a site separate from where the hosting resides.

3.5 Ethical aspects

SECLI-FIRM project does not cover any action related to ethical issues identified by H2020 programme, and therefore there is no data related to such ethical issues managed by the platform.

All the personal data management in the framework of SECLI-FIRM project is implemented according to the latest EU General Data Protection Regulation (GDPR): see next section.

3.6 Personal data management

Security of personal information, such as stakeholder telephone numbers, is paramount for SECLI-FIRM. Thus, handling and storage of personal information will comply with data protection acts (DPAs), specifically taking the General Data Protection Regulation[10] (GDPR) as our main reference. Designed to help safeguard data protection rights for individuals, the GDPR introduces a single set of rules across the EU when it comes to how organisations handle data relating to identifiable individuals. The GDPR controls how one's personal information is used by organisations, businesses or the government and everyone responsible for using data has to follow strict rules called 'data protection principles'. Personal information includes people's names, addresses, HR records, customer lists and even online identifiers such as a computer's IP address.

[10] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations_en

Personal data, typically email address and telephone number, will be stored for stakeholders attending SECLI-FIRM events (e.g. workshops, webinars), signing up to the SECLI-FIRM newsletter or producing publicly available data. These data will be administered by Project Manager, Ms Lesley Penny, and stored in a password protected place; they may also be copied in the project protected area on OneDrive. This data will only be used for SECLI-FIRM communications.

4 Appendix A – Template for metadata in file header

The header of the file, typically a CSV format file, will contain the following metadata descriptors. Here an example for the solar radiation variable is presented, which is also provided as a csv file, with filename:

H_ERA5_ECMW_T511_GHI_0000m_GLOB_CTRY_AC_TIM_12m_NA_nba_org_NA_NA-_19790101_200171231.csv

General

Title

Global Horizontal Irradiance (or solar radiation)

Abstract

ERA5 reanalysis taking 1-hour forecasts from 06:00 and 18:00 UTC of averaged values over 1-hour intervals

Date

2018-06-15

Date type

Publication: Date identifies when the data was issued

Unit

W/m2

URL

ftp://secli-firm.eu/CLIM/HIST/GHI

Data format

CSV (according to Request for Comments [RFC] 4180, see: https://tools.ietf.org/html/rfc4180)

Keywords

solar radiation, reanalysis

Point of contact

Individual name

John Smith

#####Electronic mail address

john.smith@organisation.org

Organisation name

Centre for the Creation of Data Management Plans (CCDMP)

Role

Owner: Party that owns the resource

##Usage

Access constraints

Intellectual property rights: The IP of these data belongs to the EU H2020

Use constraints

[this is an example, you may use the Creative commons licence] **Copyright**: Exclusive right to the publication, production, or sale of the rights to a literary, dramatic, musical, or artistic work, or to the use of a commercial print or label, granted by law for a specified period of time to an author, composer, artist, distributor

Citation(s)

Full report citation, or of a relevant publication if available

Temporal extent

Begin date

1979-01-01

End date

2017-12-31

Temporal resolution

hourly

Geographic bounding box

westBoundLongitude 00.00

eastBoundLongitude 360.00

southBoundLatitude -90.00

northBoundLatitude 90.00

Spatial resolution

1 deg

Lineage Statement

Original Data Source

Original (uncalibrated) ERA5 forecast data

Statement

The original data source is ECMWF ERA5 Reanalysis (available at: http://www.ecmwf.int/en/research/climate-reanalysis/era-interim)

The Added Value of Seasonal Climate Forecasting for Integrated Risk Management (SECLI-FIRM)

For more information visit

www.secli-firm.eu

or contact the SECLI-FIRM team at

info@secli-firm.eu

Grant Agreement n. 776868