This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

D2.2 – Urban flood and Water & Food Insecurity Design

Deliverable #	**D2.2**
WP#	**WP2**
WP Leader:	**e-GEOS**
Revision Date:	28.08.2023
Submission Date:	31.08.2023
Document Type:	Report
Author(s):	Luisa Bettili (Project Coordinator, e-GEOS) Vincenzo Scotti (e-GEOS) Adriano Michelangeli (e-GEOS)
Contributor(s):	CENTAUR Consortium
Verified by:	CENTAUR Consortium Board and CENTAUR Steering Committee
Approved by:	Luisa Bettili (Project Coordinator, e-GEOS) and shared with the Consortium
Version	1.0
Dissemination Level:	Public (PU)

Disclaimer

This document is issued within the frame and for the purpose of the CENTAUR project. This project has received funding from the European Union's HORIZON-CL4-2021-SPACE-01 - Strategic Autonomy in Developing, Deploying and Using Global Space-based Infrastructures, Services, Application and Data *Framework Programme under Grant Agreement No. 101082720. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.*

This document and its content are the property of the CENTAUR Consortium. All rights relevant to this document are determined by the applicable laws. Access to this document does not grant any right or license on the document or its contents. This document or its contents are not to be used or treated in any manner inconsistent with the rights or interests of the CENTAUR Consortium or the Partners detriment and are not to be disclosed externally without prior written consent from the CENTAUR Partners. Each Partner may use this document in conformity with the CENTAUR Consortium Grant Agreement provisions.

HISTORY OF CHANGES

Date	Version	Author	Change Description
17.05.2023	0.1	Luisa Bettili	Initial version with ToC
31.08.2023	1.0	CENTAUR Consortium	First version of the document

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

TABLE OF CONTENTS

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

LIST OF FIGURES

LIST OF TABLES

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

1 EXECUTIVE SUMMARY

The present document represents the deliverable D2.2 – Urban Flod and Water & Food Security Design that will be developed under the WP2 – Thematic Product engineering, particularly under Task 2.4, 2.5, and 2.6 carried out in parallel, which have respectively the aim to design and implement innovative indicators in the fields of socio-economic and political, urban flood and water & food security. Then, in Task 2.7, that runs in parallel of the tasks above, with a shift of three months) an integrated multi criteria analysis and synthetic indexes design and implementation will be performed.

The document provides an overview and details on the following elements of CENTAUR project:

- whether applicable, detailed processing flow that will be carried out to generate specific input data needed for the development of the innovative indicators identified.
- the design of each innovative indicator within the thematic areas of urban flood, water & food security, and socioeconomic and political related matters. Particularly, each indicator is structured providing the following content: description, workflow, workflow related steps description, input data needed and the output produced.

Contributes collection has been centralized using this report: each partner in charge of a specific input data processing flow and innovative indicators development have provided technical detail for the design.

Particularly, the input data processing flow is reported for each data type with respect to those tracked in the D2.1 - Catalogue of CENTAUR data and related specifications:

- geospatial data domain.
- open-source socio-economic data domain.
- meteorological data modelling.

Moreover, for each innovative indicators a common template for the workflow (customized for each indicator) was shared to have a standardized view of the technical steps that will be performed in the development phase. With regards to the water and food security and socio-economic indicators, a general It has to be noted that with Regarding as well as a type, specifications relevant to the context of the project have been tracked and mapped with reference to each innovative indicator under design.

2 INTRODUCTION

2.1 CENTAUR PROJECT DESCRIPTION

Climate change is a fact and its impact on human lives and security is continuously growing. The EU understood the importance and consequences of climate change a long time ago, adopting ambitious legislation in different policy areas. The Green Deal recognises that tackling climate change and striving for climate neutrality should be placed at the centre of societal and economic transformation. Over the last 50 years, more than 11.000 reported disasters related to extreme weather and climate conditions have caused over 2 million deaths and US$ 3,64 trillion in losses. The number of disasters has multiplied by a factor of five during that period, mainly driven by climate and more weather extremes[1]. In particular, the last twenty years have seen the number of major floods more than double, from 1.389 to 3.254, while the incidence of storms grew from 1.457 to 2.034[2]. Floods and

---

[1] World Meteorological Organization (2021). WMO atlas of mortality and economic losses from weather, climate and water extremes (1970–2019).

[2] UNDRR report: The human cost of disasters: an overview of the last 20 years (2000-2019).

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

storms were the most prevalent events and floods are the most common type of disaster worldwide, accounting for 44% of total events registered in the last twenty years. A global temperature increase of the global climate is estimated to increase the frequency of potentially high impact natural hazard events across the world. This could render current national and local strategies for disaster risk reduction and climate change adaptation obsolete in many countries. In total, between 2000 and 2019, there were 3,068 disaster events in Asia, 1,756 events in the Americas and 1,192 events in Africa. Climate change is increasingly acknowledged within the EU's integrated approach to security. The related environmental degradation is recognized as a threat multiplier and an aggravating factor for political instability with serious implications for peace and security across the globe[3]. Nowadays, climate change is already causing people to migrate, and while migration should not be directly labelled as a security problem, implicitly the link with pressures on society and increased competition for resources are often made[4]. People living in places affected by violent conflict are particularly vulnerable to climate change and it is agreed that some of the factors that increase the risk of violent conflict are sensitive to climate change[5]. This way, it is estimated that 95 % of new displacements by conflicts in 2020 happened in countries that have high or very high vulnerability to climate change[6]. From 2008 to 2016, this represents over 20 million people per year that have been forced to migrate due to climate change effects[7]. Within Copernicus Security and Emergency Services evolution, the objective of CENTAUR is to respond to societal challenges deriving from Climate Change threats by developing and demonstrating new service components for the Copernicus Emergency Management Service (CEMS) and Copernicus Security Service - Support to EU External Action service (CSS-SEA), aiming to:

1. Improve situational awareness and preparedness around climate change and its impact on complex emergencies and multi-dimensional (security) crises.
2. Anticipate the occurrence and possible knock-on effects of crisis events, in particular those triggered by climatic extremes, thus contributing to resilience and effective adaptation.

In the emergency domain, CENTAUR will address the flood-related threats to population, assets and infrastructures in urban areas. In the Security domain, CENTAUR will address water & food insecurity. The two work streams will be connected via a cross-cutting component focusing on exposure and vulnerability to climate change, as well as resilience and societal capacity for managing environmental risks and social conflict. Across work streams, indicators and models will be validated by different methods. CENTAUR will integrate data coming from multiple heterogeneous sources, with a specific focus on those generated by other Copernicus services, and in particular, those of the Climate Change Service). It will combine these with meteorological data, socio-economic data, and data coming from new sensors (e.g. traditional and social media). Thus, it will enhance current capacities to produce composite risk indexes and to perform multi-criteria analyses in the emergency and security domains.

---

[3] Meyer, C., Vantaggiato, F. P., & Youngs, R. (2021). Preparing the CSDP for the new security environment created by climate change. European Union.

[4] Schaik, L., Bakker, T. (2017). Climate-migration-security: Policy Brief Making the most of a contested relationship. Planetary Security.

[5] W.N., J.M. Pulhin, J. Barnett, G.D. Dabelko, G.K. Hovelsrud, M. Levy, Ú. Oswald Spring, and C.H. Vogel (2014). Human security. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken,
P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 755-791.

[6] University of Notre Dame. (n.d.). Country index // Notre Dame Global Adaptation Initiative // University of Notre Dame. Notre Dame Global Adaptation Initiative. Retrieved January 23, 2022, from https://gain.nd.edu/our-work/country-index/.

[7] WEF (2020). *The Global Risks Report 2020*, Insight Report 15th Edition. World Economic Forum, Geneva Switzerland, p. 102. https://www.weforum.org/reports/the-global-risks-report-2020.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

2.2 SCOPE OF THE DOCUMENT

This document is produced under WP2 – Thematic product engineering, which has the objective to generate workflows for collecting necessary data for the development of risk indicators and crisis indexes as well as for their implementation accounting for the user requirements collected in WP1.

Tasks 2.1, 2.2 and 2.3 have the objective of harvesting and pre-processing data collected from several repositories, which were catalogued, tracked, and mapped to the innovative indicators identified in the D2.1 – CENTAUR data catalogue related deliverable. In the D2.1 deliverable [RD04] the input data type where clustered into Geospatial domain (including Earth-Observation and georeferenced ancillary data, agriculture and vegetation data, elevation data and settlements data), Opensource and Socio-Economic domain (including social and traditional media, census data and population), Meteorological domain (including the high-resolution precipitation data and the seamless forecast data). On the above three data domains identified, specific workflows for collecting and generating required input data for specific innovative indicators are described in the present document (whether applicable).

Following the CEMS RM/RRM and CSS-SEA services assessment conducted in WP1, as well as the User Requirements consolidation through the URR meeting (M6), the innovative indicators identified in D1.1[RD05] deliverable [RD05] (in the context of Urban Flood, Water & Food Security domain and related Socio-economic and Political thematic areas), to fill the existing gaps in the above mentioned services and to meet the user needs, where refined in terms of their effective feasibility and designed from an operational perspective (workflow, input data selection, output data description and status of development with respect to the state of the art.

To cover the above objectives, the document has been structured into the following sections:

- Section 1. Summary of the document's contents.
- Section 2. Introduction, scope of the document, definitions, abbreviations, acronyms, and reference documents.
- Section 3. Indicators innovative design. The section includes four main sub-sections to focus on specific processing flow that will be performed to gather the input data needed for the development of the innovative indicators in scope of CENTAUR (subsection 3.1). In addition, the design step-by-step workflow, as well as details of the input data needed, and the output data produced including the process with respect to the state of the art described in D1.1 report is described [RD05]. The main subsections 3.2, 3.3 and 3.4 the above-described content per innovative indictors grouped per thematic areas as followed: Urban Flood, Water & Food Security, and related Socio-Economic and political related matters.
- Section 4. Main conclusions.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

2.3 DEFINITIONS, ABBREVIATIONS AND ACRONYMS

Acronym	Description
ACD	Automated Change Detection
ACLED	Armed Conflict Location & Event Data Project
ACLED CAST	Armed Conflict Location & Event Data Project Conflict Alert System
AdBPo	Autorità di Bacino Distrettuale del Fiume Po
AoI	Area of Interest
API	Application programming interface
CDS	Copernicus Data Store
CEMS RM	Copernicus Emergency Management Service Rapid Mapping
CEMS RRM	Copernicus Emergency Management Service Risk & Recovery Mapping
CIASIN	Center for International Earth Science Information Network
CNN	Convolutional Neural Network
CSS-SEA	Copernicus Security Service in Support to EU External Action
DNB	Day Night Band
DTM	Digital Terrain Model
DWH	Data Ware House
ECMWF	European Centre for Medium-Range Weather Forecasts
ECA&D	European Climate Assessment & Dataset project
FAO DIEM	Food and Agriculture Organization - Data in Emergencies Monitoring
FIES	Food Insecurity Experience Scale
IOM DTM	International Organization for Migration Displacement Tracking Matrix
EC	European Commission
EEAS	European External Action Service
EO	Earth Observation
E-OBS	ENSEMBLES daily gridded observational dataset
EU	European Union
FAO	Food and Agriculture Organisation of the United Nations
GDACS	Global Disaster Alert and Coordination System
GEV	Generalized Extreme Value
GIS	Geographic Information System

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Acronym	Description
GHSL	Global Human Settlement Layer
GHS-POP	Global Human Settlement – Population
GPCC	Global Precipitation Climatology Centre
HDDS	Household Dietary Diversity Score
HOT	Humanitarian Open Street Map
H-RES	High-RESolution forecast
HR/VP	High Representative of the Union for Foreign Affairs and Security Policy/Vice-President of the European Commission
IDP	Internally Displaced Persons
JPSS	Joint Polar-orbiting Satellite System
JRC	Joint Research Centre
LAS	LASer file format
LCSI	Livelihood-based Coping Strategies Index
LST	Land Surface Temperature
MASE	Ministry of the Environment and Energy Security
MMS	Media Mining System
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
NLP	Natural Language Processing
NMAD	Normalized Median Absolute Deviation
NSO	National Statistical Offices
NTL	Night-Time Light
OSM	Open Street Map
REST	Representational State Transfer
RMSE	Root Mean Square Error
RZSM	Root Zone Soil Moisture
SAR	Synthetic-Aperture Radar
SOTA	State of the Art
SPI	Standardized Precipitation Index
SPIDI	Standardized Precipitation Index and Drought Indices
UF	Urban Flood

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Acronym	Description
UNHCR	United Nations High Commissioner for Refugees
VH	Cross-polarization (Vertical – Horizontal)
VHR	Very High Resolution
VIIRS	Visible and Infrared Imaging Suite
VV	Co-polarization (Vertical – Vertical)
WFS	Water & Food Security
WMO	World Meteorological Organisation
WSF	World Settlement Footprint

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

2.4 APPLICABLE AND REFERENCE DOCUMENTS

Table 1: Applicable and reference documents

ID	Document name
[RD01]	Copernicus Service in Support to EU External Action: https://sea.security.copernicus.eu/
[RD02]	Disaster Risk Reduction in EU external action - Council conclusions (28 November 2022): https://data.consilium.europa.eu/doc/document/ST-14463-2022-INIT/en/pdf
[RD03]	D1.1 - Report on Urban Flood and Water & Food security indicators
[RD04]	D2.1 - Catalogue of CENTAUR data and related specifications
[RD05]	D3.1 - Platform Design Document v1
[RD06]	Copernicus Emergency Management Service – Rapid Mapping and Risk & Recovery: https://emergency.copernicus.eu/

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3 INNOVATIVE INDICATORS DESIGN

3.1 INPUT DATA PROCESSING FLOW

In the present chapter, specific processing flow that are put in place for data harvesting and generation (whether applicable) are reported for each input data domain (i.e. geospatial, opensource and meteorological) in alignment with the D2.1 – Catalogue of CENTAUR data and related specifications [RD04].

3.1.1 Geospatial data domain

3.1.1.1 Earth-Observation and georeferenced ancillary data

Concerning geospatial data utilized as input for various indicators, we can focus on earth observation-related and georeferenced ancillary data.

For data originating from earth observation, they play a role as input across different models to generate indicators. For example, in UF-ID 2, ERA5 data is applied. The ERA5 dataset involves retroactively analysing hourly meteorological conditions from as far back as 1979. It amalgamates a weather model with observational data from satellites and ground sensors to construct a consistent, long-term climate archive. In indicators UF-ID 3, 4, 5, and 6, data from earth observation, particularly SAR sensors, are harnessed to create flood masks.

The availability of interferometrically compatible pre and co-event SAR data is the main condition of ID-4 production. Pre-event and co-event amplitudes (depending on the available polarizations) and coherences must be generated and stacked. Ancillary data such as the Global Human Settlement Layer (GHSL) or DLR World Settlement Footprint (WSF) are also needed to focus the analysis on urban areas.

In WFS-ID-5, geolocated ancillary datasets are also brought into play. Specifically, for different case studies conducted during the project, national datasets or those drawn from OpenStreetMap are used to identify detailed information about geographic features, such: building types, roads, facilities, types of industries, and other data that best delineate the relevant urban elements.

Lastly, in WFS-ID-7, geospatial ancillary data regarding population distribution within the urban context of interest is utilized. In detail, very high-resolution data produced by SAR sensors are used to observe and monitor the sites that host refugees who come from climate crisis linked also to social and political event.

3.1.1.2 Agriculture and Vegetation

In the Agriculture and Vegetation domain, the main goal of the CENTAUR project is to develop a new set of innovative indicators related to the potential occurrence and impact of agricultural droughts (see Sections 3.3.4 – 3.3.6 for further description of the end products). Agricultural droughts are defined as prolonged periods of time with insufficient soil moisture, causing significant impact to agricultural production. Hence, in order to detect and predict occurrence/impact of agricultural droughts, we need information on vegetation condition, soil moisture levels, land cover and meteorological conditions. To this end, a number of Earth Observation datasets are harvested: time series of the Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), Root Zone Soil Moisture (RZSM), air temperature, precipitation and land cover information (see CENTAUR deliverable D2.1 - Catalogue of CENTAUR data and related specifications ([RD04]) for more information regarding these datasets). Figure 1 summarizes the processing workflow to be adopted for preparing these input datasets, while Table further specifies the individual processing steps.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

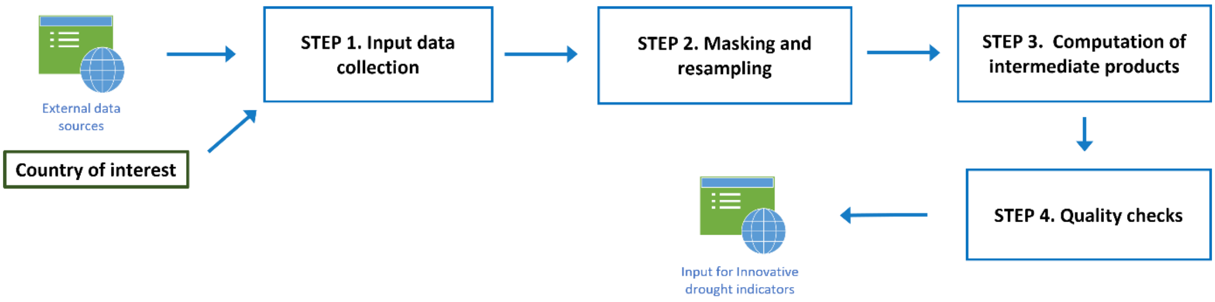


Figure 1: Input data processing workflow for agriculture and vegetation related data.

The Table 2 below is tailored to outline the specific process flow for this area and the associated data sources.

Table 2: Description of the input data processing workflow related to agriculture and vegetation.

Step #	Name	Description
1	Input data collection	All required input datasets will be acquired for the specific country of interest. For tile-based datasets (NDVI, LST), this step includes the selection and download of the appropriate tiles. For global-scale products (land cover, soil moisture datasets), the country of interest is extracted from the global extent.
2	Masking and resampling	Each input dataset employed in the context of agricultural drought monitoring is available at different spatial resolutions and projected according to different geographical projection systems. In order to ease the combination of all these datasets further down the processing pipeline, all datasets are first reprojected to a common projection system and resampled to 1 km resolution. All pixels outside the country borders (adopted from FAO GAUL dataset) are masked for further analysis.
3	Computation of intermediate products	Several intermediate products will be derived from the raw input time series: - Phenology (start and end of growing seasons) will be derived from the NDVI time series according to a threshold-based algorithm. Each pixel will be allowed to contain a maximum of 2 growing seasons per calendar year. - The Normalized Difference Moisture Index (NDMI) will be computed based on optical data from MODIS/Sentinel-3, representing the actual water content of the vegetation. - A thermal drought stress indicator will be derived by combining Land Surface Temperature with ambient air temperature. A low difference between these two variables point to stressed vegetation. - NDVI anomalies (comparison of current value with value during similar time in other years) will be computed to determine how exceptional the vegetation currently is behaving. - Likewise, soil moisture anomalies will be computed as well. The intermediate products will be combined with the original time series in a multi-criteria decision framework to determine

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		the occurrence and impact of agricultural droughts (see sections 3.3.4 – 3.3.6).
4	Quality checks	A series of automated quality checks will be built into this processing pipeline to ensure the values of produced time series and intermediate indicators are within the ranges typically expected. Additionally, random visual quality checks will be undertaken to identify any production artefact that might occasionally take place due to e.g. missing input data.

3.1.1.3 Elevation

A high-resolution Digital Terrain Model (DTM) plays a vital role in precise and efficient flood modelling by accurately capturing intricate topographic details, enabling more realistic representation of these features. This, in turn, helps to obtain improved accuracy in predicting flood extents, flow patterns, and inundation depths.

The role of ITHACA in Task 2.1 was to harvest elevation data for generating a high-resolution Digital Terrain Model (DTM), which serves as one of the inputs for Urban Flood Indicators. The workflow was designed to utilize LiDAR data processing with the aim of producing a detailed DTM. The Meisino area of city of Turin in Italy was identified as one of the Areas of Interest (AoIs) for the urban flood use cases and this area was considered for the DTM generation. The main input data for this task was provided by the Polytechnic University of Turin, which conducted the data survey in January 2022. While the workflow presented in Figure 2 is replicable for any data source of similar characteristics.

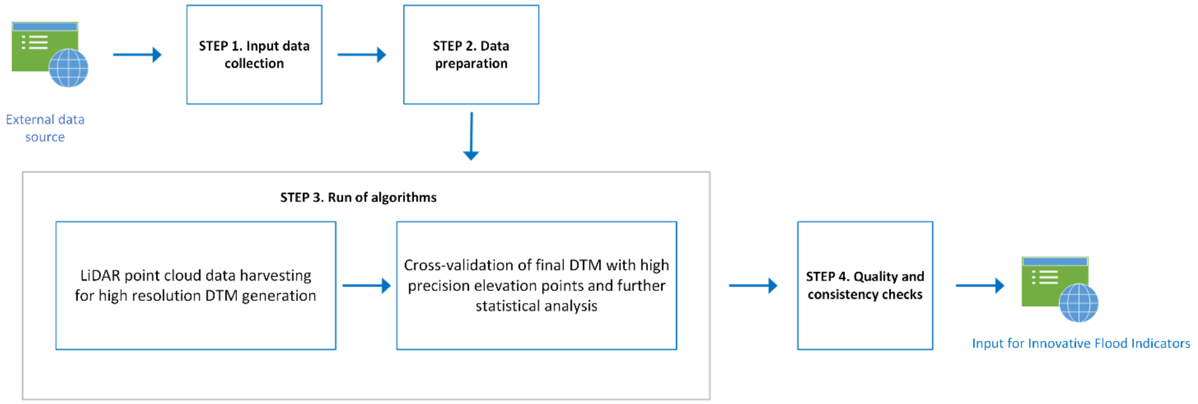


Figure 2: Workflow for Elevation Data Harvesting

The Table 3 below is tailored to outline the specific process flow for this area and the associated data sources.

Table 3: Description of the Elevation Data Harvesting Workflow

Step #	Name	Description
1	Input data collection	The main input data utilized for the processing was the LiDAR point cloud, which was generated from a survey conducted by the Polytechnic University of Turin[8] in January 2022. The survey was focused on the geographic area encompassed by the administrative boundaries of the Municipality of Turin. However, for this specific

---

[8] With entire data property rights belonging to Polytechnic University of Turin (PoliTO).

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		Task, only the LiDAR data referring to the area of Meisino was considered. The cloud consists of 24,504,671 points in total and these data were distributed across 48 LAS files (tiles). The points generated by the acquisition was highly dense in urban areas, providing precise representations of buildings and infrastructure with high accuracy. On the contrary, due to the technical limitations of the instruments and constrains regarding the acquisition of water bodies through LiDAR sensors[9], the water courses were not acquired during this survey. For this reason, it was necessary to consider other data sources to extract elevation information on the Po, Dora Riparia and Stura di Lanza watercourses. To address this issue, various data sources available in the Italian national and regional data portals were involved.
2	Data preparation	For the complementary input data, initially, the Ministry of the Environment and Energy Security (MASE) data acquired during a LiDAR survey in 2009 was considered. This dataset offered a more detailed representation of watercourses compared to the 2022 survey. The obtained dataset consisted of both the LiDAR point cloud and an already processed Digital Terrain Model (DTM) with a resolution of 1 m/pixel, covering nearly the entire Area of Interest (AOI) under consideration. Moreover, for the Po River, additional data from the Po Basin Authority (AdBPo[10]) with DTMs (2 m/pixel resolution) generated through LiDAR acquisitions in 2004-2005, with a specific focus on the watercourse was also employed. By incorporating these different data sources, elevation information on the watercourses, particularly the Po, Dora Riparia, and Stura di Lanzo, was extracted to complement the limitations encountered in the 2022 LiDAR survey.
3	Run of algorithms – Elevation Data Harvesting	Using the diverse data sources, four DTMs were generated with varying levels of accuracy and resolution. In this specific case, the ENVI LiDAR software was utilized to create very high-resolution DTMs ($\text{DTM}_{\text{L, processed}}$). In the second part, the focus was to combine the reliable urban elevation information extracted from the $\text{DTM}_{\text{L, processed}}$ with other sources capable of representing watercourse information. To do this, water masks had to be created to divide the urban areas with respect to the rivers in the AoI. Using the vector data of the areal hydrography available both in the Municipality of Turin and Piedmont Regional databases, the boundaries of the watercourses were drawn, and the mask was created. The remaining part was considered as urban, and a second mask was created to include these areas removed from the watercourses. Once the masks were created, they were used to extract reliable information from the DTMs, considering both the DTMs processed by ENVI LiDAR

---

[9] D'Andrimont, Raphael & Marlier, Catherine & Defourny, Pierre. (2017). Hyperspatial and Multi-Source Water Body Mapping: A Framework to Handle Heterogeneities from Observations and Targets over Large Areas. Remote Sensing. 9. 211. 10.3390/rs9030211.

[10] Autorità di Bacino Distrettuale del Fiume Po.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		and external DTMs. The final stage of the processing part consisted in merging the two extracted DTMs, one representing the urban areas and the other the watercourses. In this way, the final very high-resolution model (DTM_f) was created as a combination of several reliable sources. To validate these DTMs and determine the most optimal one(s) for flood modelling input, high-precision elevation points were incorporated as reference data. These reference points were gathered from the Piedmont Regional Geoportal, which offers public access to various products acquired and processed over the years for the regional territory. The validation procedure involved a comparison between the elevation data of the high-precision reference points and the corresponding points extracted from the various final DTMs. By conducting this comparison, statistical parameters were calculated to assess the discrepancies between the datasets. This way, a thorough analysis of the errors between the datasets could be performed to evaluate the accuracy and reliability of the final outputs.
4	Quality and consistency checks	The main quality and consistency checks were conducted during the validation phase. When interpreting the results and assessing the accuracy of the final DTMs, it is essential to account for technological differences, systematic errors during data acquisition, temporal variations in the terrain, and the influence of processing parameters. Addressing these factors will contribute to a more comprehensive understanding of the accuracy and reliability of the generated DTMs. In terms of assessing the quality of a dataset against an accurate reference, the Root Mean Square Error (RMSE) and Normalized Median Absolute Deviation (NMAD) are generally considered a robust indicator for the assessment of elevation data[11]. The final outputs, yielding more performant parameters with lower RMSE and NMAD values, were then considered as the input data to test the feasibility with the flood detection algorithm.

The integration of a high-resolution DTM is crucial in the development of urban flood indicators as it enables the flood detection algorithm to harness precise elevation data, which is instrumental in simulating flood scenarios and accurately predicting water flow patterns in different flood conditions. This output generated particularly for the AoI Turin Center - Meisino will be used as an input data for the Piedmont Use Case.

3.1.1.4 Settlements

Large-scale and fine spatial resolution settlement layers describing the extent, location and characteristics of the built environment are an essential input needed across the development of different indicators and indices. In the framework of Urban Floods for example, these layers play an instrumental role in assessing damages arising from flood events (e.g. UF-ID-5, UF-ID-7), estimate financial losses (e.g. UF-ID-9) and devising effective evacuation strategies (e.g. UF-ID-13). As such, within CENTAUR, one of the tasks of DLR has been to identify and harness optimal sources of settlement layers that align seamlessly with design, implementation, and development of the aforementioned indicators (Task 2.1). Among the array of options, two primary sources have been found optimal, these include: The WSF suite produced by German Aerospace Center (DLR), and the GHSL produced by the Joint

---

[11] Jacobsen, K. Characteristics and Accuracy of Large Area Covering Height Models. Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci. 2013, XL-1/W1, 157–162.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Research Center (JRC). These datasets were selected on the basis of their high temporal and spatial resolution (e.g. 2016-present at 10m and 1975-2030 at 100m, respectively), and their extended thematic characterisation of the built environment which include binary delineations, density estimations (e.g. percent of impervious surface), and three-dimensional representations (e.g. building volume).

3.1.2 Opensource data domain

3.1.2.1 Social and traditional media

Information derived from traditional and social media provide complementary information to the other information sources within CENTAUR. One the one hand, they can provide immediate, unfiltered, and from-the-ground information - typically traditional and social media markers emerge as one of the earliest sources accessible during and after an event. On the other hand, they may be biased, manipulated and sometimes outright wrong (intentionally or unintentionally) and need to be treated with care. Media-based information will come in multiple modalities – text, audio, image, and video – and in multiple languages, dialects, and scripts. The processing capabilities need to reflect this variety and will have to be adjusted (fine-tuned) to the different cases at hand.

Within CENTAUR, data sources and technologies and models for enrichment are managed as part of an overall process as depicted in Figure 3 Workflow for media-based harvesting and processing below. A set of data harvesting components (set up and adjusted specifically per use-case) collects data on a continuous basis. Content and meta-data are gathered and enriched using a series of technologies from NLP and Computer Vision. The resulting information provides a comprehensive picture of a situation and serves as input for the generation of media-based indicators. Due to the nature of media-derived indicators, we view them rather as moderating/adjusting elements for other kinds of indicators (e.g. also can be used for calibration). However, the exact nature and use depends on the exact use-case and will be determined during indicator implementation and evaluation (Figure 3).

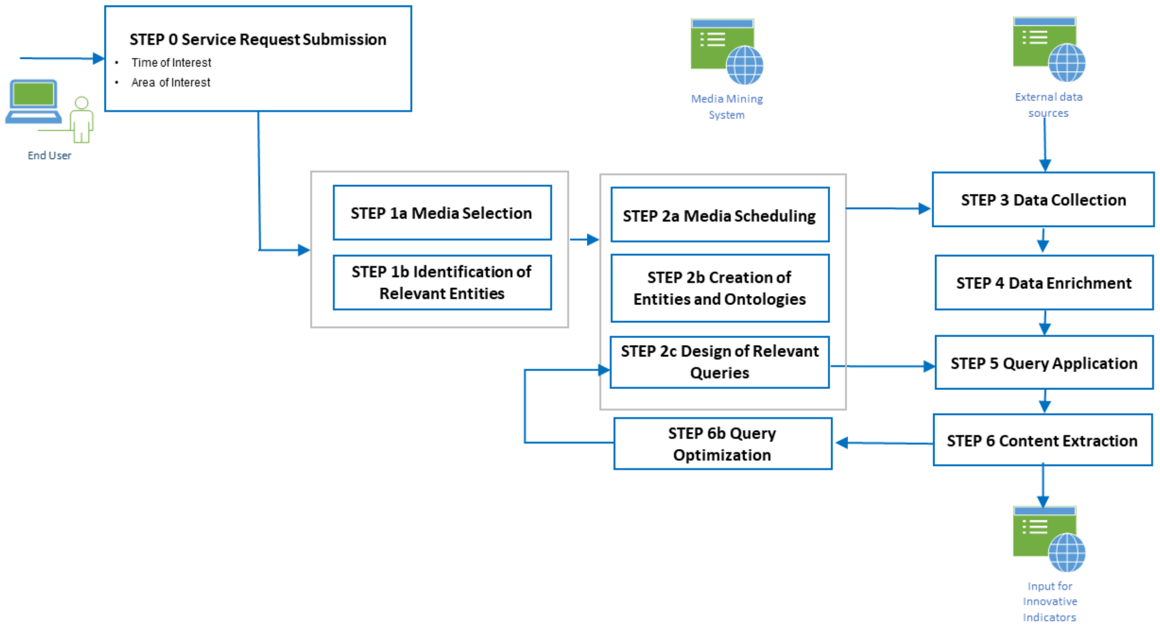


Figure 3 Workflow for media-based harvesting and processing

The Table 4 below is tailored to outline the specific process flow for this area and the associated data sources.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Table 4 Overview of steps for media-based data collection and processing

Step #	Name	Description
1	Media Selection and Identification of Relevant Entities	As media coverage depends on the AOI and timeframe, the portfolio needs to be adjusted and configured to best reflect the AOI and timeframe. Knowledge about the respective media-landscape and use is required. Entities, such as critical locations, names of actors (organizations) etc. in the respective language(s) need to be prepared to best capture their use and appearance in media.
2a and 2b	Media Scheduling, Creation of Ontologies	Sources are typically available in a variety of manners – unstructured, structured or via APIs. The exact setup (e.g. link to relevant feeds) needs to be investigated and set up. Access mechanisms requiring registration need to followed and respective be carried out. Entities are created and connected via links, forming ontologies. These in turn form the basis for concept-based retrieval. Once sources are scheduled within the Media Mining System, content will be gathered and enriched on a continuous basis.
2c	Query Design	Set up of queries yielding relevant contents for the computation of indicators. Queries tend to become more complex over time (iterative process) until the exact set of relevant content is retrieved.
3	Data Collection	Continuous collection (harvesting) of data according to the set schedules.
4	Data Enrichment	Enrichment of contents via Natural Language Programming (NLP) and CV models using entities.
5	Query Application	This step combines step 2c with the results of steps 3 and 4.
6	Content Extraction	Content is extracted from the set of collected and enriched documents using an API (of the Media Mining System - MMS). The resulting set of documents and associated meta-data can be used for inspection, visualization, and analysis. Queries (but also sources and entities) can be adjusted based on insights in an iterative manner. The output of this step – documents, meta-data and information inferred from them – form the basis and input for indicator computation.

3.1.2.2 Census data and population

For the design and implementation of different indicators, in particular those related to exposure to security risks and resources shortage indicators (e.g. WFS-ID-8, WFS-ID-9, WFS-ID-10) updated official census data (e.g. administrative boundaries and population counts) will be collected, when possible, from sources such as National Statistical Offices (NSO). For those AoIs where official census data are not openly available, population data will be derived from the raw global census data used for the Gridded Population of the World, version 4.11 (GPW4.11) produced by the Center for International Earth Science Information Network (CIESIN).

Upon assembling population data, a process of enrichment will follow, involving the application of advanced modeling techniques to produce high-quality, finely detailed population estimates with superior spatial resolution. For the years 2016-to present, gridded population datasets will be produced by DLR using a proprietary algorithm that relies on improved World Settlement Footprint products. There in, 10m spatial resolution gridded population datasets will be made available for all indicators that rely on population data. These datasets will represent population per pixel at given year and will be used mainly for monitoring tasks within the CENTAUR project.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

When the need arises for predictive analysis, particularly for forecasting upcoming events, population datasets from the Global Human Settlement – Population (GHS-POP) dataset will be employed. These will allow predictions for the years 2025 and 2030, respectively, at a spatial resolution of 100m. These datasets will be systematically collected and harmonized to facilitate seamless integration into the project's framework.

3.1.3 Meteorological data domain

3.1.3.1 High-resolution precipitation

The most important meteorological input data is precipitation, for both, urban flood and water and food security indicators. For any type of monitoring, and where available, in-situ observations are used. For example, the E-OBS data set from the European Climate Assessment & Dataset project (ECA&D) is used in the estimation of return periods of extreme precipitation events (UF-ID-1). This data set is a high-resolution, gridded observational data base, based on more than 23335 in-situ stations, that has a high reliability for capturing extreme events and a long record covering the years 1950–today. Where observations are unavailable or do not capture the severity of the event, state-of-the art reanalysis data from ECMWF's ERA5 are used (see short description in Section 3.1.1.1). For the drought indicators developed in the water- and food security part of this project (e.g., WFS-ID-1), reanalysis data are used, as they provide a coherent, high-resolution, and gap-free timeseries of precipitation – even at places where observations are unavailable.

For the forecasts of precipitation, in the context of providing an early-warning system for urban floods, 3-day forecasts from ECMWF's high-resolution ensemble forecast are used.

3.1.3.1 Seamless forecasts

For the water- and food security domain, high-resolution forecasts extending up to 6 months ahead, are required. Here, a suite of forecasts, from medium-range weather forecasts from ECMWF, covering the coming 15 days, to seasonal forecasts covering the coming 6 months, are concatenated to create a seamless forecast of precipitation. The creation of a seamless forecast bridges the gap between short-term weather forecasting and climate forecasts. Key benefits of seamless forecasts over seasonal forecasts are an improved accuracy and reliability and more frequent updates[12]. The same forecasts, but for additional variables (e.g., temperature and net radiation) are used as inputs for the agricultural drought indicators (WFS-ID-5). The workflow for this process is detailed in Figure 4 and Table 5. Note, however, that this input data workflow is performed on ECMWF premises.

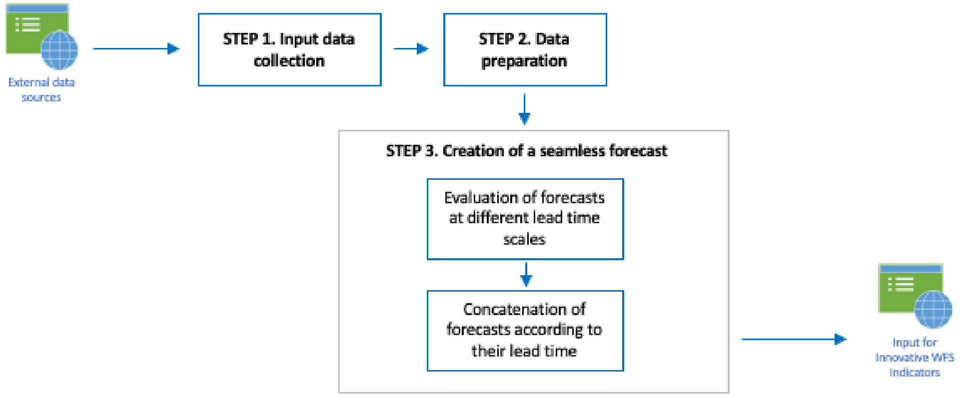


Figure 4: Workflow for the creation of a seamless forecast as input for the WFS Elevation Data Harvesting

---

[12] Wetterhall, F. and Di Giuseppe, F.: The benefit of seamless forecasts for hydrological predictions over Europe, Hydrol. Earth Syst. Sci., 22, 3409–3420, https://doi.org/10.5194/hess-22-3409-2018, 2018.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

The Table 5 below is tailored to outline the specific process flow for this area and the associated data sources.

Table 5: Description of workflow for the creation of a seamless forecast.

Step #	Name	Description
1	Input data collection	A range of forecasts, covering lead times from 0-15 days (medium-range ensemble forecast ENS), 16–42 days (extended-range medium-range ensemble forecast ENS-ER), and up to 6 months (seasonal forecast SEA) are collected from the ECMWF internal data base.
2	Data preparation	The forecasts are aggregated to the spatial scales of interest, e.g., a county of country.
3	Creation of a seamless forecast	The forecasts are concatenated according to their lead times, and the resulting seamless forecast serves as input for the innovative indicators WFS-ID-2 and WFS-ID-5.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3.2 URBAN FLOOD INDICATORS

3.2.1 UF-ID-1: Static map of precipitation associated to return period

Description

This indicator estimates the return periods of extreme precipitation over urban areas in Europe and Mozambique using historical observations. The resulting maps of 1-, 10-, 20-, 50-, 100-, 500-year return periods will be used in conjunction with the speed-flood hydraulic model to derive inundation maps. The output of this indicator is a static catalogue of precipitation totals associated with return periods for all urban areas in Europe and Mozambique that enables an ad-hoc identification of extreme events.

Workflow

In Figure 5, the workflow schematic and a detailed description of the step-by-step design of the indicator is given.

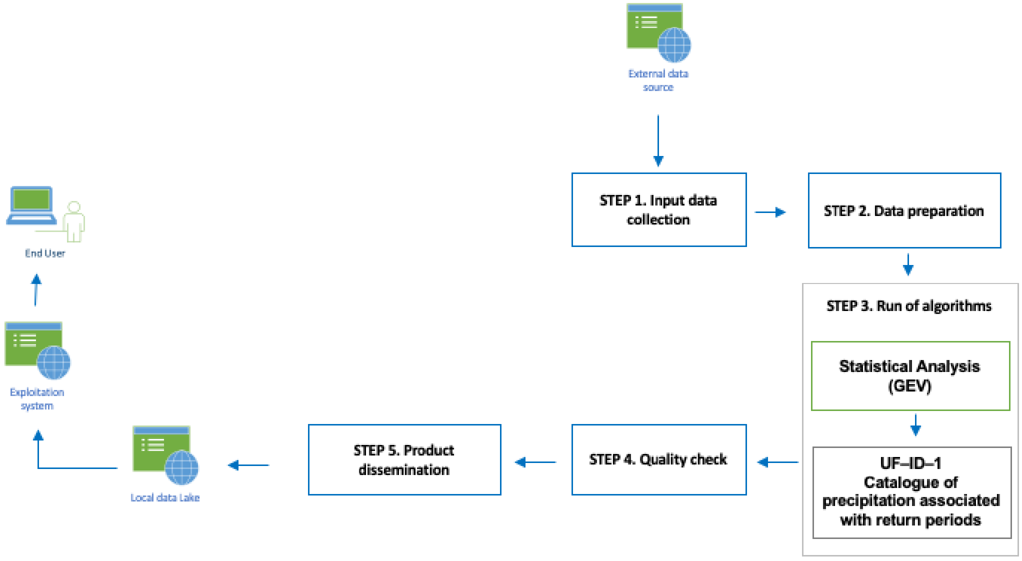


Figure 5: Workflow schematic of the UF-ID-1 Historical 6-hourly return period (pre-event phase)

The Table 6 below is tailored to outline the specific process flow for this area and the associated data sources.

Table 6: Description of the steps part of the workflow for UF-ID-1

Step #	Name	Description
1	Input data collection	The longest record of observed, gridded observation products with a high reliability for capturing extreme events, are downloaded from the web. For Europe, daily precipitation from E-OBS is downloaded for the period 1950–2022.
2	Data preparation	The input data (i.e., precipitation) is quality-checked, downscaled to 1.5 km and aggregated to 6-hourly precipitation totals. Ocean areas are masked out. Grid points with more than 30% missing data are neglected from the analysis.
3	Run of algorithms – statistical analysis (GEV) to estimate return periods	A Generalized Extreme Value distribution (GEV) is fitted to the full observational timeseries of precipitation at each grid point and enables the estimation of return periods. The resulting return periods are further chunked into 1-, 10-, 20-, 50-, 100-, 500-year

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		periods and corresponding precipitation totals can be extracted. A catalogue detailing the total precipitation, the corresponding continuous return period for all dates available and all grid points considered is created. This catalogue serves as a reference for the forecast, enabling the extraction of precipitation totals from forecasted return periods for each region of interest. The resulting return periods can be casted onto a regional map for potential flooding events, highlighting the severity of the observed precipitation.
4	**Quality check**	An automated quality check of the generated product will be performed. Products not fulfilling quality standards or with limited data availability will be discarded.
5	**Product dissemination**	This catalogue of return periods will serve as an input for other indicators and will be delivered through the Service Data Lake.

Input data

The input data used for the indicator generation are derived from:

- E-OBS
- MSWEP https://www.gloh2o.org/mswep/, this specific data will be adopted for Mozambique use case.

Below in Table 7, the details of the input data used for the indicator calculation.

Table 7: Input data required for UF-ID-1

Input data Name	Description	Format
Precipitation from E-OBS	E-OBS (ENSEMBLES daily gridded observational dataset) the observational data base from the European Climate Assessment & Dataset project (ECA&D) that monitors changes in weather and climate extremes. ECA&D receives meteorological observations from 65 countries and E-OBS contains observations from 23335 stations in these countries. These observations are distributed heterogeneously but casted onto a regular grid. Daily precipitation from E-OBS is available on a regular 0.1-degree grid, covering 25N – 71.5N and 25W – 45E, for the period 1950–2022.	NetCDF

Indicator output

This urban flood indicator enables to identify rare precipitation events in the historical records using return periods (Table 8). The output of this analysis is a catalogue of precipitation associated with a set of return periods (i.e., 1-, 10-, 20-, 50-, 100-, 500-years) for each grid point. The generated catalogue will be used as a training dataset for a machine learning model that aims to predict the return period of precipitation events (see UF-ID-2 in section 3.2.2).

Table 8: Output data description and specifications for UF-ID-1

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Return period	Return period of precipitation events, categorized in 1, 10, 20,	NetCDF / Tif	1.5 km over urban areas in Europe and	Novel, machine-learning-based emulator that predicts return periods of

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
	50, 100, 500-year periods.		Mozambique, static (no temporal resolution).	extreme precipitation events.
Total precipitation	Total precipitation associated with the set of return periods defined above.	NetCDF / Tif	1.5 km over urban areas in Europe and Mozambique, static (no temporal resolution).	–

3.2.2 UF-ID-2: Forecast of return period

Description

This indicator presents a novel forecasting of extreme precipitation events over urban areas. Using the catalogue of observed extreme precipitation events and estimated return periods presented above (see UF-ID-1 in section 3.2.1), we train a convolutional neural network model to predict return periods of precipitation events up to three days in advance.

Workflow

In Figure 6, the workflow schematic and a detailed description of the step-by-step design of the indicator is given.

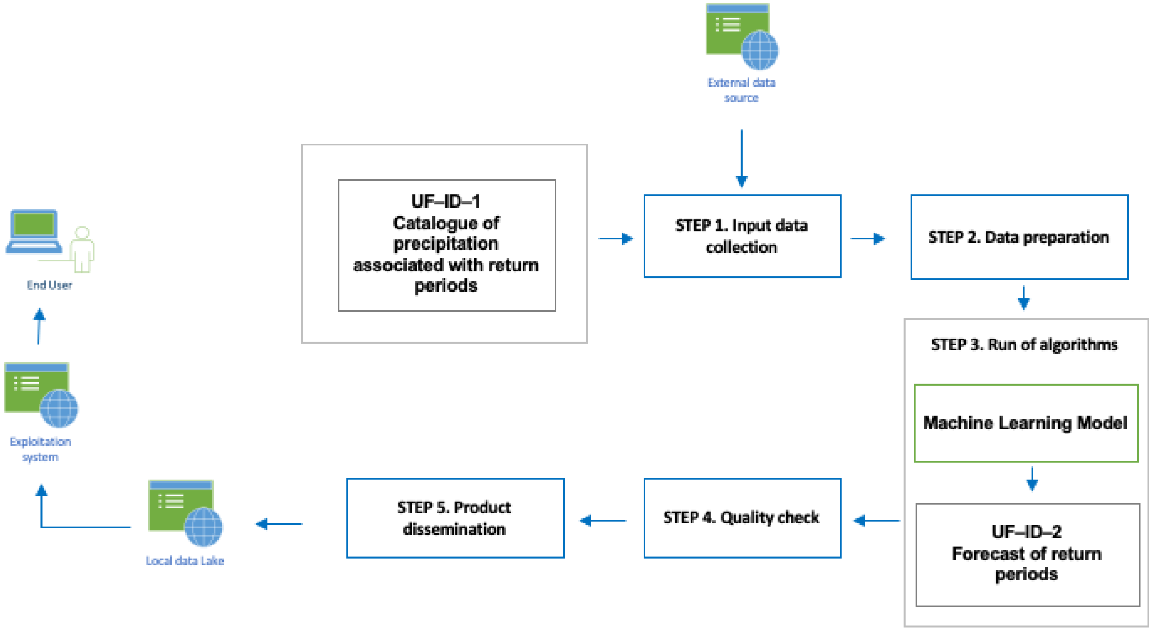


Figure 6: Workflow schematic of the UF-ID-2 ML Data Driven Forecast of return period-based precipitation events in urban areas (early warning-phase)

The Table 9 below is tailored to outline the specific process flow for this area and the associated data sources.

Table 9: Description of the steps part of the workflow for UF-ID-2

Step #	Name	Description
1	Input data collection	Two main input data sets are used. For the training of the Convolutional Neural Network (CNN), the entire history of ERA5 reanalyses is used (once). For the daily forecasts of return periods

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		with this trained CNN, the latest forecasts of the ECMWF high-resolution simulation at 9 km are collected; both are available online and inhouse at ECMWF. In addition, the catalogue of return-periods from UF-ID-1 (in section 3.2.1) are used; this catalogue is also available online and inhouse at ECMWF.
2	Data preparation	All variables used as predictors in the CNN are aggregated to 6-hourly values, which is the temporal resolution of the target variable (return periods from UF-ID-1).
3	Run of algorithms – Training and predictions of return periods using a CNN	In a first step, the full record of ERA5 data (variables detailed below) is used to train a CNN. In a second step, high-resolution forecasts of 3 days lead time of the same variables are used as input to the CNN to forecast return periods of precipitation. The predicted return periods present 1-, 10-, 20-, 50-, 100-, 500-year periods and are indicative of extreme precipitation events in the coming days, asserting a warning level. Using the catalogue from UF-ID-1 (see ch. 3.2.1). This return period can be translated to a total precipitation quantity, which in turn can be used as forcing for the speed-model.
4	Quality check	An automated quality check of the generated product will be performed. Products not fulfilling quality standards or with limited data availability will be discarded.
5	Product dissemination	The forecasted return periods are disseminated to CENTAUR partners through the Service Data Lake and made available to the user through the Exploitation System.

Input data

The input data used for the indicator generation are derived from:

- UF-ID-1 (see Section 3.2.1)
- ECMWF reanalysis ERA5
- ECMWF high-resolution forecast

Below in Table 10, we report the details of the input data used for the indicator calculation.

Table 10: Input data required for UF-ID-2

Input data Name	Description	Format
UF-ID-1	This indicator estimates the return periods of extreme precipitation over urban areas in Europe and Mozambique using historical observations. The output of this indicator is a static catalogue of precipitation totals associated with categorized return periods (1-, 10-, 20-, 50-, 100-, 500-years) for all urban areas in Europe and Mozambique that enables an ad-hoc identification of extreme events.	NetCDF / Tif
ECMWF reanalysis ERA5	ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather. The ERA5 reanalysis combines observations and models to derive a spatio-temporal complete and consistent dataset following the laws of physics. Data from ERA5 is available on a 0.25 grid from 1940 to today at hourly resolution. ERA5 is updated with a latency of 5 days, but an early release (ERA5T) is available near real-time monitoring. The following atmospheric and environmental predictors are used to train the CNN: horizontal and vertical	Grib/NetCDF

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	wind, specific humidity, geopotential, temperature, potential vorticity, convective available potential energy, total column water vapor, total column rainwater, orography, cloud cover, land-sea mask, precipitation	
ECMWF high-resolution forecast	The high-resolution forecast (H-RES) from ECMWF is a global weather forecast with a lead time of 15 days. This forecast is run on a global 9 km grid and is updated 4 times daily (00, 06, 12, and 18 UTC). The following atmospheric and environmental predictors are used to train the CNN: horizontal and vertical wind, specific humidity, geopotential, temperature, potential vorticity, convective available potential energy, total column water vapor, total column rainwater, orography, cloud cover, land-sea mask, precipitation	Grib/netCDF

Output

The output of this forecast indicator (Table 11) is a timeseries of the predicted return period for grid point and each 6-hourly interval in the coming 3 days. To filter for potential extreme events, return periods smaller than x years may be masked.

Table 11: Output data description and specifications for UF-ID-2

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Forecasted return period	Return period of precipitation events, categorized in 1-, 10-, 20-, 50-, 100-, 500-year periods	netCDF	1.5 km over urban areas in Europe and Mozambique, static, 6 hourly resolutions up to 3 days ahead (updated daily)	To our knowledge, this is the first machine-learning based emulator of a precipitation parameterisation.

3.2.3 UF-ID-3: High-Resolution urban flood risk maps for various return periods

Description

The flood indicator is derived from the Speedy-flood model, incorporating precipitation intensity maps generated through return period analysis. This integration enables the projection of potential future flood scenarios. This tool serves a dual purpose: primarily, it aids in comprehending floods from a management standpoint, and secondarily, it offers insights into the high-risk flood zones within a given urban context. The indicator is meticulously crafted to delineate flooding across various magnitudes of rainfall events.

Workflow

In Figure 7 the workflow schema and a detailed description of the step-by-step design of the indicator is provided.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

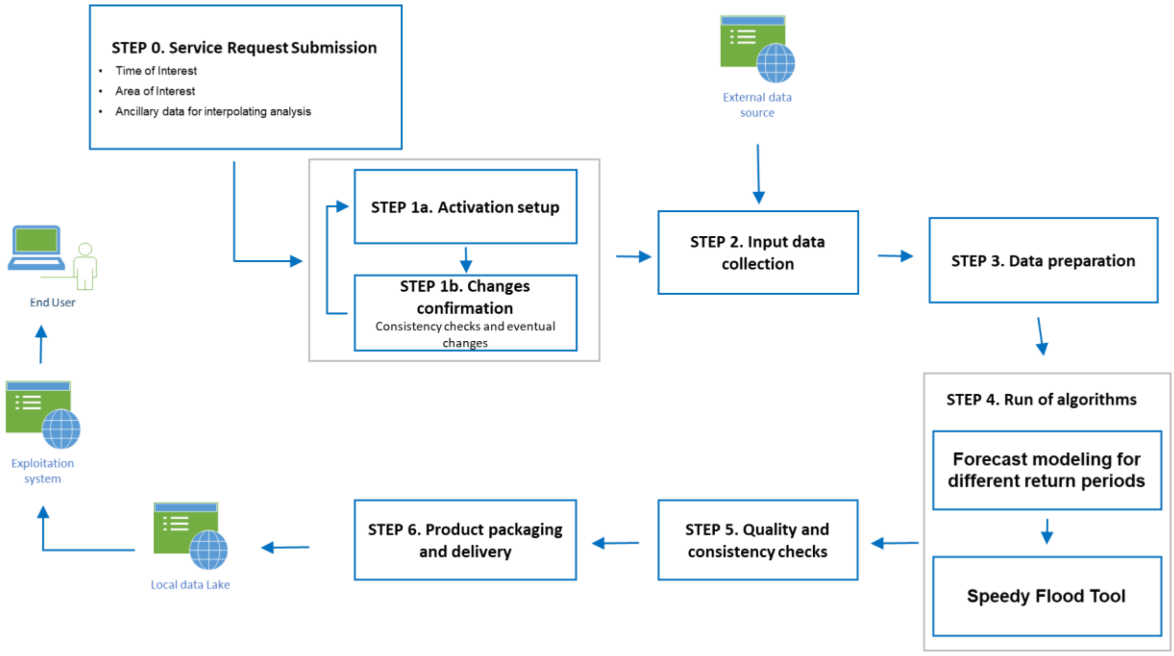


Figure 7: Workflow schema of the UF-ID-3- High-Resolution urban flood risk maps for various return periods

The Table 12 below is tailored to outline the specific process flow for this area and the associated data sources.

Table 12: Description of the steps part of the workflow for UF-ID-3

Step #	Name	Description
0	Service Request Submission	The user activates the indicator providing, through the specific Dashboard, information like: • Time of Interest • Area of Interest • Upload of ancillary data for interpolating analysis • Information about the flood of interest
1a	Activation setup	Depending on available data and processing options set up by the end user (frequency, AoI, etc), processing is activated.
1b	Changes confirmation	This step performs additional checks on End User request, e.g. consistency with requested products and spatial and temporal coverage, limitations due to availability of input data, etc. Eventual changes or suggested modifications of the initial service request are proposed to the user. If rejected, interaction with the user allows to agree a new activation setup.
2	Input data collection	Acquisition of all essential input data necessary for the algorithm's execution includes: - ECMWF precipitation map layers for each return period scenario. - DTM. - Land Use layer. - Flooding mask corresponding to the event of interest (if obtainable for a previous event that shares the same return period scenarios of interest).

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		- Social and traditional media markers (if obtainable for a previous event that corresponds to the same return period scenarios of interest).
3	Data preparation	The data essential for the algorithms within this ID undergo preprocessing to facilitate seamless ingestion during execution. This involves preparing the input data sourced from the ECMWF dataset, crucial for generating return period maps. Similarly, for the Speedy Flood tool, all necessary input data is acquired, processed, and made ready for utilization during runtime.
4	Run algorithm	The ECMWF algorithm is employed to generate return period precipitation maps, which then serve as inputs for the Speedy Flood tool to generate corresponding flooding maps. A flooding map is produced using the Speedy Flood tool for each return period precipitation map.
5	Quality check & consistency check	A consistency check of the generated products is automatically done. The quality check consists of automatic steps to verify the quality level of the generated products. Products with not enough quality and/or consistence are discarded.
6	Product Packaging & Delivery	This phase includes the preparation of the product package to be delivered to the end user through the Service Data Lake and to be made available to the user through the Exploitation System.

Input data

The input data used for the indicator generation are derive from:

- ECMWF reanalysis ERA5.
- ECMWF high-resolution forecast.
- Flood extent delineated using satellite-derived Synthetic Aperture Radar (SAR) data.
- Land Use.
- DTM.
- Social/traditional media markers.

Below, in Table 13 the details of the input data used for the indicator calculation.

Table 13: Input data required for UF-ID-3

Input data Name	Description	Format
ECMWF reanalysis ERA5	ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather. The ERA5 reanalysis combines observations and models to derive a spatio-temporal complete and consistent dataset following the laws of physics. Data from ERA5 is available on a 0.25 grid from 1940 to today at hourly resolution. ERA5 is updated with a latency of 5 days, but an early release (ERA5T) is available near real-time monitoring. The following atmospheric and environmental predictors are used to train the CNN: horizontal and vertical wind, specific humidity, geopotential, temperature, potential vorticity, convective available potential energy, total column water vapor, total column	Geojson

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	rainwater, orography, cloud cover, land-sea mask, precipitation.	
ECMWF high-resolution forecast	The high-resolution forecast (H-RES) from ECMWF is a global weather forecast with a lead time of 15 days. This forecast is run on a global 9 km grid and is updated 4 times daily (00, 06, 12, and 18 UTC). The following atmospheric and environmental predictors are used to train the CNN: horizontal and vertical wind, specific humidity, geopotential, temperature, potential vorticity, convective available potential energy, total column water vapor, total column rainwater, orography, cloud cover, land-sea mask, precipitation.	Raster (.tif)
Flood extent delineated using satellite-derived Synthetic Aperture Radar (SAR) data.	Creating a flood map using Synthetic Aperture Radar (SAR) data involves leveraging advanced remote sensing technology to detect changes in the Earth's surface caused by flooding. SAR uses microwave signals that are transmitted from a satellite and reflected to the sensor. By analysing the returned signals, SAR can provide valuable information about the extent of flooding, even in cloudy or nighttime conditions. For the different use cases, in this project flood masks deriving from past CEMS Rapid Mapping activations are used.	Gdb (vector and raster layers)
Land Use	A land use layer refers to a thematic map or dataset that categorizes and illustrates the different ways in which land is utilized within a specific geographic area. It provides information about the various human activities and natural features present on the land, such as residential areas, commercial zones, agricultural fields, forests, water bodies, and more. If the use case area is located in Europe a CLC+ layer is used. CLC+ layer typically refers to an enhanced version or an extension of the Corine Land Cover (CLC) dataset. Corine Land Cover is a European program that provides detailed information about land cover and land use across the continent. The "+" sign in CLC+ indicates additional or improved features beyond the standard CLC dataset. In other worldwide areas, alternative datasets may be employed. These can originate from national sources or the Global Land Cover 30 dataset, which boasts comprehensive global coverage.	Depending on the data source (could be vector layer and/or tabular data)
Digital Terrain Model (DTM)	A DTM is a digital representation of the Earth's surface that includes elevation data. It provides a detailed and accurate depiction of the topography of a specific area, showing variations in height and shape of the land surface. In order to be used as input for the Speedy Flood tool and to facilitate the reconstruction of flood maps with a satisfactory level of accuracy, it is essential to have a DTM layer characterized by a high spatial resolution. A minimum spatial resolution of 10	Raster (tif)

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	meters is required to effectively reconstruct acceptable flood maps.	
Social/traditional media markers	"Social and traditional media markers" refer to indicators or elements that are used to track and analyse the presence, impact, or influence of certain topics, events, or information in both social media platforms and traditional media outlets. In this project, this element is used in the Speedy Flood results validations and depth calculation.	Vector (geojson, shapefile)

Output

This outcome (Table 14), particularly valuable in the context of water management, enables the swift identification of urban areas highly susceptible to flooding. It provides more accurate details, including the extent and depth of the flood. Such information facilitates the rapid assessment of urban areas at the greatest risk of flooding, aiding in the prompt recognition of high-exposure zones within the urban landscape.

Table 14: Output data description and specifications for UF-ID-3

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
High-Resolution urban flood risk maps for various return periods	Indicator that clearly shows the flood extent and depth for different return period scenarios.	Raster (tif)	Strongly related to the DTM spatial resolution	Currently, there are no flooding products available with such fine spatial resolutions. Leveraging high-resolution data enables more precise delineation of potential flooding scenarios within urban areas. This enhanced capability significantly improves the accuracy of depicting potential flood situations compared to what has been achievable thus far. As a result, it offers a tangible and substantial contribution to water management by providing invaluable insights into the urban flood risk landscape.

3.2.4 UF-ID-4: Inferred INSAR urban flood extent

Description

Floodwater detection over urban areas using Radar and artificial intelligence. This indicator is purely based on available SAR data and uses other types of information available as a validation source. The goal of this indicator is to give an information with respect to the flood extent in urban areas at a time as close as possible to the time of interest, depending on the acquisition time of the SAR data.

Workflow

In Figure 8 the workflow schema and a detailed description of the step-by-step design of the indicator is given.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

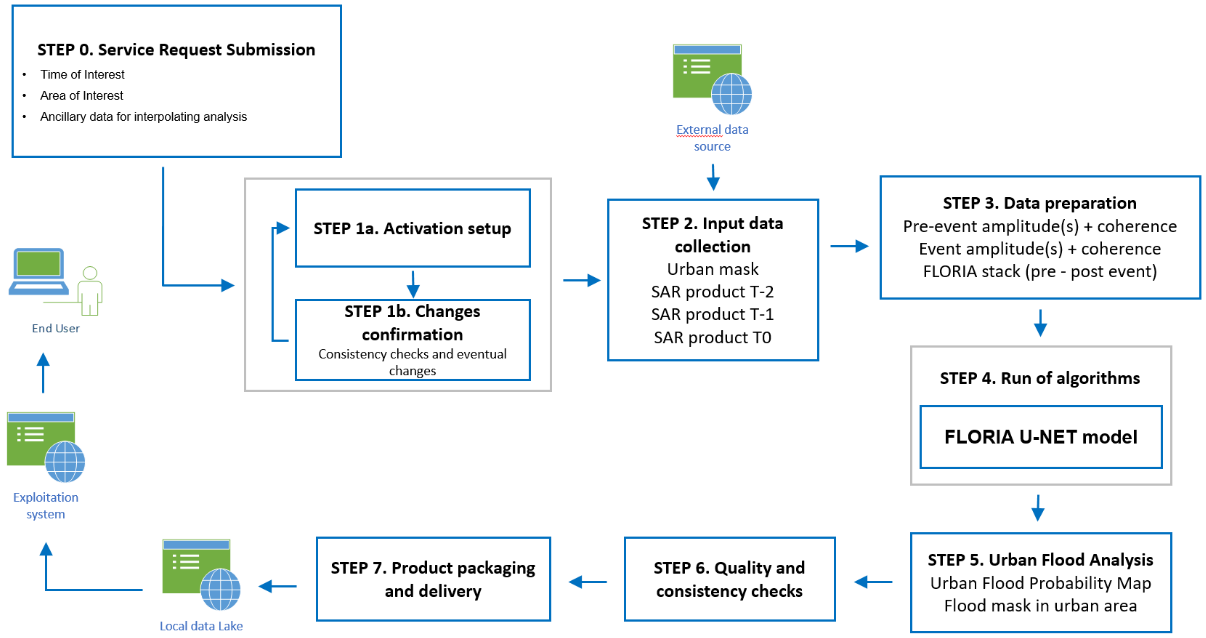


Figure 8:Inferred INSAR urban flood extent

The Table 15 below is tailored to outline the specific process flow for this area and the associated data sources.

Table 15: Description of the steps part of the workflow for UF-ID-4

Step #	Name	Description
0	Service Request Submission	The user activates the indicator providing, through the specific dashboard, information like: • Time of Interest. • Area of Interest. • Upload of ancillary data (ground observations of urban floods …). • Information about the flood of interest.
1a	Activation setup	Depending on available data and processing options set up by the end user (frequency, AoI, etc), processing is activated.
1b	Changes confirmation	This step performs additional checks on End User request, e.g. consistency with requested products and spatial and temporal coverage, limitations due to availability of input data, etc. Eventual changes or suggested modifications of the initial service request are proposed to the user. If rejected, interaction with the user allows to agree a new activation setup.
2	Input data collection	Acquisition of all essential input data necessary for the algorithm's execution includes: - Two pre-event SAR products acquired at different times. - One co/post-event SAR product. - Urban mask. Ancillary data related to the flood extent and impact
3	Data preparation	This indicator requires 3 SAR products that are interferometrically compatible: two pre-event images and one acquired during the flood. To be considered compatible, they need to have a small

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		temporal baseline to ensure the pertinence of the computed coherence and a small perpendicular baseline. The information extracted from those 3 products are combined in one stack: two bands for the coherence of the pre-event pair and the co-event pair, two bands for the pre-event amplitude based on respectively VV and VH polarizations and two for the co-event amplitude based on VV and VH polarizations.
4	Run algorithm	The previously generated stack is ingested in a U-Net model. This model is trained on a database of previous flood events where ground truth was available. The algorithm generates a confidence map of floods from the ingested stack and masks the results in non-urban areas to minimize errors. Each urban pixel gets a confidence score between 0 and 1. The urban flood mask is generated as a vector file through adaptive thresholding. Another mask is computed by classifying the confidence scores in five intervals based on the minimum and maximum values of the results. This is also outputted as a vector file.
5	Urban flood analysis	
6	Quality check & consistency check	A consistency check of the generated products is carried out manually. The quality check consists of manual steps in order to verify the quality level of the generated products. Products with not enough quality and/or consistency are discarded.
7	Product Packaging & Delivery	This phase includes the preparation of the product package to be delivered to the end user through the Service Data Lake and to be made available to the user through the Exploitation System.

Input data

The input data used for the indicator generation derives from:

- SAR pre and co-event products.
- Urban mask.
- European Union Copernicus portal where it is possible to see in detail the mapping and the damages due to the specific event, if available.
- On-line ground-based sources, if available.
- WSF-World Settlement Footprint.

Below, in Table 16 the details of the input data used for the indicator calculation.

Table 16: Input data required for UF-ID-4

Input data Name	Description	Format
SAR products	3 SAR products: two acquired before the flood event, at two different times, and one acquired during the flood. It is important to note that the 3 SAR products must have a small temporal baseline (few days or at the very least they must have been acquired in the same season as the flood event) and a small perpendicular baseline (inferior to 150 m).	Data provider dependent
Urban mask	Binary raster delimitating urban areas, used by the algorithm to discard results in non-urban areas.	GeoTIFF

Funded by the European Union
This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	This mask is for now derived from the GHSL (Global Human Settlement Layer) built areas product, released in 2019.	
CEMS RM and/or RRM activation info in terms of affected population and size of the flooded area	CEMS RM: This service consists of the on-demand and fast provision (hours-days) of geospatial information in support of emergency management activities immediately following disaster. The service is based on the acquisition, processing, and analysis, in rapid mode, of satellite imagery and other geospatial raster and vector data sources, and social media when relevant. CEMS RRM: This service provides on-demand geospatial information to support emergency management activities outside the immediate response phase. It covers the prevention, preparedness, disaster risk reduction, and recovery phases and is divided into two sub-categories: ▪ Risk and Recovery STANDARD for a predefined set of standardized products, ▪ Risk and Recovery FLEX for tailor-made studies. Information related to the event extent and damage assessment provided within this framework may be useful to validate the FLORIA results and help the thresholding process.	Gdb (vector and raster layers)
Ground based sources	Geocoded pictures, newspapers, geocoded testimony, etc. Any information that gives the state of a given place (flooded or not) at the time of interest, to serve as validation source and help the thresholding process.	Various
WFS - World Settlement Footprint	The dataset depicts the extent and location of human settlements derived from the analyses of multi-temporal statistics in radar an optical data from Sentinel 1 and Sentinel-2 imagery.	Raster layer (.tif)
WSF-3D	The dataset provides quantification of the building height, building area, building fraction, and building volume, within the built-up environment as described by the WSF2019. The units of the dataset are in dm, m, percent, and cubic meters, respectively.	Raster layer (.tif)

Output

This output (Table 17), useful especially during the flood management phase, allows to quickly identify the inferred urban flood areas.

Table 17: Output data description and specifications for UF-ID-4

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Inferred INSAR urban flood extent	Floodwater detection over urban areas using Radar and artificial intelligence (FLORIA).	Vector shapefile	Event related	Mapping of urban flooding using SAR satellite images with a water-extraction workflow not yet exploited in the CEMS RM - RRM. Furthermore, this indicator requires only SAR products, therefore is theoretically applicable anywhere (if SAR data is available).

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3.2.5 UF-ID-5: Enhanced Urban Flood Damage Assessment

Description

The flood indicator is a product of the synergistic combination of the Speedy-flood and FLORIA models. This integration harnesses the capabilities of Speedy Flood for precise flood delineation and depth calculation, while the FLORIA model employs AI techniques to enhance flood delineation within urban regions. By amalgamating the outcomes of these two methodologies, an exceptionally accurate flooding map can be reconstructed for urban environments. Utilizing this amalgamated flooding map, it becomes feasible to conduct a comprehensive assessment of flood-related damages, focusing on urban infrastructure and key components.

Workflow

In Figure 9 the workflow schema and a detailed description of the step-by-step design of the indicator is provided.

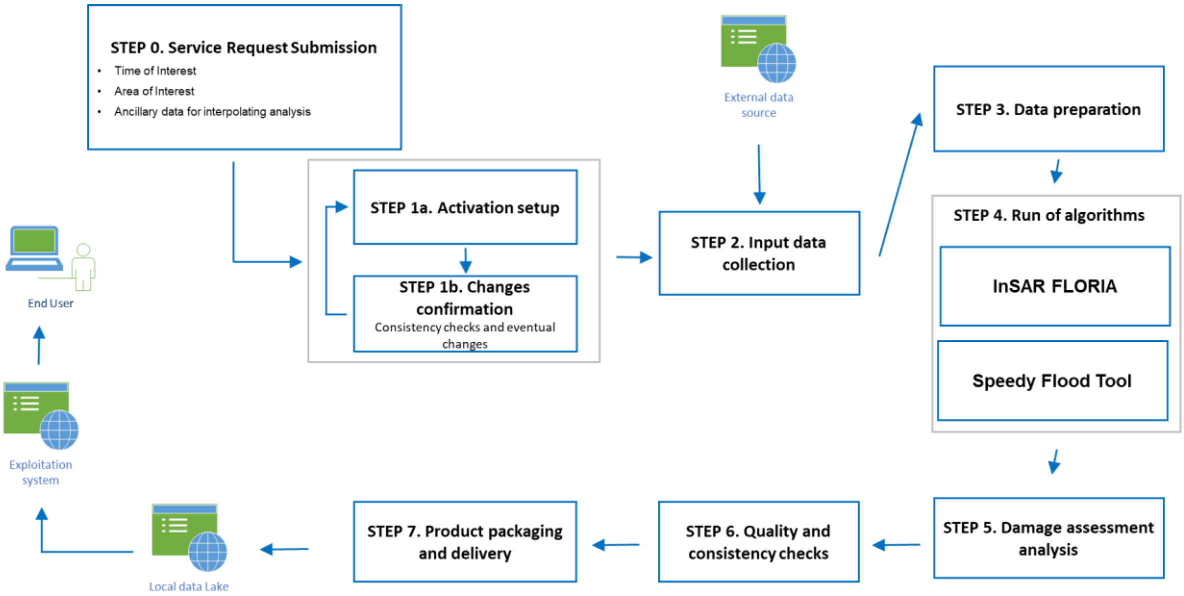


Figure 9: Workflow schema of the UF-ID-5- Enhanced Urban Flood Damage Assessment

The Table 18 below is tailored to outline the specific process flow for this area and the associated data sources.

Table 18: Description of the steps part of the workflow for UF-ID-5

Step #	Name	Description
0	Service Request Submission	The user activates the indicator providing, through the specific Dashboard, information like: • Time of Interest • Area of Interest • Upload of ancillary data for interpolating analysis • Information about the flood of interest
1a	Activation setup	Depending on available data and processing options set up by the end user (frequency, AoI, etc), processing is activated.
1b	Changes confirmation	This step performs additional checks on End User request, e.g. consistency with requested products and spatial and temporal coverage, limitations due to availability of input data, etc.

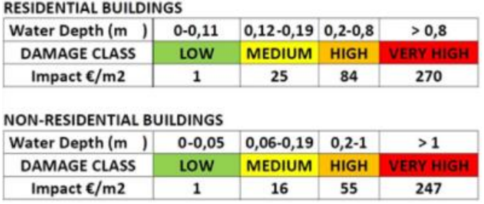
Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		Eventual changes or suggested modifications of the initial service request are proposed to the user. If rejected, interaction with the user allows to agree a new activation setup.
2	Input data collection	Acquisition of all essential input data necessary for the algorithm's execution includes: - ID-4 Inferred INSAR urban flood extent indicator. - Digital Terrain Model (DTM). - Land Use layer. - Flooding mask corresponding to the event of interest. - Social and traditional media markers. - Hydrography layer for DTM pre-processing. - Ancillary data layers useful for the flooding damage assessment like buildings, transportations, facilities, etc.
3	Data preparation	During this phase, the primary actions involve downloading and extracting clips of the input layers within the designated Area of Interest (AOI). These steps are essential for both the execution of the Speedy Flood model and the subsequent damage assessment analysis. Additionally, a preprocessing step is carried out, focusing on enhancements to the river network. This includes tasks such as river channel excavation and the refinement of secondary river networks, including channel replication. These measures collectively contribute to the refinement and preparation of the data before further analysis and modelling can be conducted.
4	Run algorithm	This phase involves executing both the FLORIA model and the Speedy Flood model. Initially, the FLORIA model is run to produce an output that serves as input for the subsequent Speedy Flood analysis. The FLORIA model generates the ID-4 indicator, which identifies flooded areas within the urban landscape. This ID-4 indicator output is integrated with the flood extent calculated by the Speedy Flood model. Importantly, the Speedy Flood model exhibits enhanced performance in sub-urban contexts. To capitalize on its strengths, a fusion of these two models is performed. This integration is pivotal as it facilitates the creation of an exceptionally accurate flood map that covers all contexts within the Area of Interest (AoI). Following this, the flood extent obtained through the fusion of these models is utilized for precise calculations of flooding depths. This is achieved using a module within the Speedy Flood tool. By adopting this comprehensive approach, a detailed comprehension of both the spatial distribution and the depth of flooding across the AOI is ensured.
5	Damage assessment analysis	Utilizing the flooding depth outcomes derived from step 4, a thorough flooding damage assessment is conducted. By adhering to predefined rules, it becomes feasible to determine the damage grade for various elements within the urban environment. A specific illustration is provided for building elements, specifically residential and non-residential structures, in the table presented below.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

<table>
<tr><th>Step #</th><th>Name</th><th>Description</th></tr>
<tr><td></td><td></td><td>
RESIDENTIAL BUILDINGS
<table>
<tr><td>Water Depth (m)</td><td>0-0,11</td><td>0,12-0,19</td><td>0,2-0,8</td><td>> 0,8</td></tr>
<tr><td>DAMAGE CLASS</td><td>LOW</td><td>MEDIUM</td><td>HIGH</td><td>VERY HIGH</td></tr>
<tr><td>Impact €/m2</td><td>1</td><td>25</td><td>84</td><td>270</td></tr>
</table>
NON-RESIDENTIAL BUILDINGS
<table>
<tr><td>Water Depth (m)</td><td>0-0,05</td><td>0,06-0,19</td><td>0,2-1</td><td>> 1</td></tr>
<tr><td>DAMAGE CLASS</td><td>LOW</td><td>MEDIUM</td><td>HIGH</td><td>VERY HIGH</td></tr>
<tr><td>Impact €/m2</td><td>1</td><td>16</td><td>55</td><td>247</td></tr>
</table></td></tr>
<tr><td>6</td><td>Quality check & consistency check</td><td>A consistency check of the generated products is automatically done. The quality check consists of automatic steps to verify the quality level of the generated products. Products with not enough quality and/or consistence are discarded.</td></tr>
<tr><td>7</td><td>Product Packaging & Delivery</td><td>This phase includes the preparation of the product package to be delivered to the end user through the Service Data Lake and to be made available to the user through the Exploitation System.</td></tr>
</table>

Input data

The input data used for the indicator generation are derived from Table 19:

- UF-ID-4 Inferred INSAR urban flood extent indicator.
- Digital Terrain Model (DTM);
- Land Use layer.
- Flooding mask corresponding to the event of interest.
- Social and traditional media markers.
- Hydrography layer for DTM preprocessing.
- WFS - Population
- WSF - World Settlement Footprint
- Ancillary data layers useful for the flooding damage assessment like buildings, transportations, facilities, etc.

- WSF-3D

Below, in Table 19 the details of the input data used for the indicator calculation.

Table 19: Input data required for UF-ID-5

Input data Name	Description	Format
UF-ID-4 Inferred INSAR urban flood extent indicator	A flooding mask is created using SAR satellite sensor data, with a specific focus on urban areas. This resulting output from the FLORIA model is subsequently employed as an input for the Speedy Flood tool.	Vector (shapefile)
Flood extent delineated using satellite-derived Synthetic Aperture Radar (SAR) data.	Creating a flood map using Synthetic Aperture Radar (SAR) data involves leveraging advanced remote sensing technology to detect changes in the Earth's surface caused by flooding. SAR uses microwave signals that are transmitted from a satellite and reflected back to the sensor. By analysing the returned signals, SAR can provide valuable information about the extent of flooding, even in cloudy or nighttime conditions. For the different use cases, in this project flood masks deriving from past CEMS Rapid Mapping activations are used.	Gdb (vector and raster layers)

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
Land Use	A land use layer refers to a thematic map or dataset that categorizes and illustrates the different ways in which land is utilized within a specific geographic area. It provides information about the various human activities and natural features present on the land, such as residential areas, commercial zones, agricultural fields, forests, water bodies, and more. If the use case area is located in Europe a CLC+ layer is used. CLC+ layer typically refers to an enhanced version or an extension of the Corine Land Cover (CLC) dataset. Corine Land Cover is a European program that provides detailed information about land cover and land use across the continent. The "+" sign in CLC+ indicates additional or improved features beyond the standard CLC dataset. In other worldwide areas, alternative datasets may be employed. These can originate from national sources or the Global Land Cover 30 dataset, which boasts comprehensive global coverage.	Depending on the data source (could be vector layer and/or tabular data)
Digital Terrain Model (DTM)	A Digital Terrain Model (DTM) is a digital representation of the Earth's surface that includes elevation data. It provides a detailed and accurate depiction of the topography of a specific area, showing variations in height and shape of the land surface. In order to be used as input for the Speedy Flood tool and to facilitate the reconstruction of flood maps with a satisfactory level of accuracy, it is essential to have a DTM layer characterized by a high spatial resolution. A minimum spatial resolution of 10 meters is required to effectively reconstruct acceptable flood maps.	Raster (tif)
Social/traditional media markers	"Social and traditional media markers" refer to indicators or elements that are used to track and analyze the presence, impact, or influence of certain topics, events, or information in both social media platforms and traditional media outlets. In this project, this element is used in the Speedy Flood results validations and depth calculation.	Vector (geojson, shapefile)
Hydrography layer for DTM preprocessing	This layer primarily finds application in the preprocessing phase of the Digital Terrain Model (DTM), specifically for the excavation of river channels. Often, the DTM may lack complete representation of all river networks within the Area of Interest (AOI). Hence, there's a need to recreate the entire river network. The hydrography layer serves the purpose of identifying all rivers within the relevant context. By leveraging bathymetry data, this layer facilitates the reconstruction of rivers using appropriate tools.	Vector (shapefile)
WFS-Population	Residential population estimates derived from CIESIN GPW4.11 are disaggregated from census or administrative units to grid cells, informed by the distribution, height and functional use of settlements derived from the WSF-datasets.	Raster layer (.tif)

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	Each dataset has a spatial resolution of 10m and temporal coverage of 2016-to present.	
WSF - World Settlement Footprint	The dataset depicts the extent and location of human settlements derived from the analyses of multi-temporal statistics in radar an optical data from Sentinel 1 and Sentinel-2 imagery.	Raster layer (.tif)
WSF-3D	The dataset provides quantification of the building height, building area, building fraction, and building volume, within the built-up environment as described by the WSF2019. The units of the dataset are in dm, m, percent, and cubic meters, respectively.	Raster layer (.tif)
Ancillary data layers useful for the flooding damage assessment	In a post-event scenario, this layers are crucial for conducting damage assessment within the Area of Interest (AOI). Various ancillary data sources come into play in this context, including information on buildings, transportation networks, land use, facilities, and more. These data sources are integrated with the flooding depth data obtained from the Speedy Flood model. This integration allows for a comprehensive flood damage assessment to be carried out, enabling a thorough understanding of the impact on the area's infrastructure and assets.	Various

Output

This indicator provides valuable insights into the extent of damages sustained within the specified urban context. It illustrates the severity of damage inflicted, as determined by the flooding depth computed through the Speedy Flood model (Table 20). To ensure accuracy in delineating flood extent, the FLORIA model, empowered by AI techniques, is also employed. The precision of these models, coupled with the utilization of high-quality input data, ensures a meticulous damage assessment.

The indicator classifies the damage grade into distinct categories, each correlated with the recorded flooding depth. These categories encompass: "very low," "low," "medium," and "high." Such a comprehensive classification scheme helps provide a nuanced understanding of the range and magnitude of damages inflicted due to flooding.

Table 20: Output data description and specifications for UF-ID-5

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Enhanced Urban Flood Damage Assessment	Indicator shows the flood damage assessment based on flooding depth.	Raster (tif)	Strongly related to the DTM spatial resolution and the quality of the ancillary data used for the damage assessment analysis.	In general, and in the current CEMS product, flood damage assessment relies on labor-intensive photointerpretation involving human analysis, which demands substantial time investment. However, our innovative approach within CENTAUR revolutionizes this process. We can now expedite damage assessment by utilizing flooding depth outcomes, maintaining an equivalent level of accuracy. This advancement allows us to

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
				achieve swift results without compromising precision, marking a significant leap forward in flood damage assessment methodologies.

3.2.6 UF-ID-6: Social/Traditional media indicators for Urban Flooding Map

Description

This indicator is exclusively constructed from traditional and social media markers linked to the specific event of interest. These markers stand as crucial components, vividly capturing the developments within the urban landscape during an extreme occurrence. By leveraging geo-tagged photos, videos, and comments, a comprehensive understanding of the situation within the urban area is attainable. The outcome of this indicator manifests as a map that presents the social/traditional media markers that surfaced during the event of interest. Each social/traditional media marker viewed as a point in the map is accompanied by a link to relevant news, alongside details such as the hour, date, and time of occurrence.

Workflow

In Figure 10 the workflow schema and a detailed description of the step-by-step design of the indicator is provided.

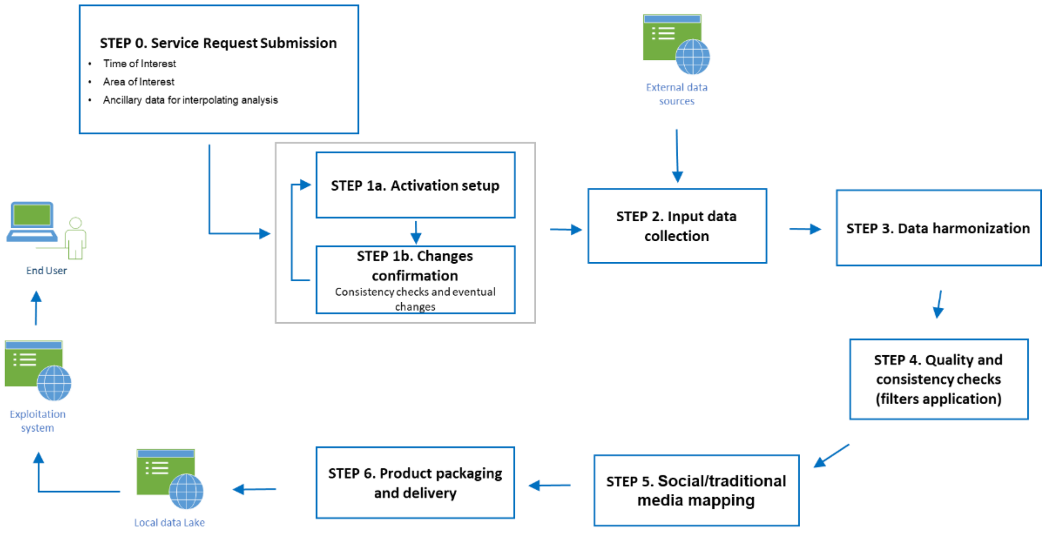


Figure 10: Workflow schema of the UF-ID-6- Social/Traditional media indicators for Urban Flooding Map

In particular, in Table 21 below are reported the input data used for the indicator calculation.

Table 21: Description of the steps part of the workflow for UF-ID-6

Step #	Name	Description
0	Service Request Submission	The user activates the indicator providing, through the specific Dashboard, information like: • Time of Interest. • Area of Interest. • Upload of ancillary data for interpolating analysis.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		• Information about the flood of interest.
1a	Activation setup	Depending on available data and processing options set up by the end user (frequency, AoI, etc), processing is activated.
1b	Changes confirmation	This step performs additional checks on End User request, e.g. consistency with requested products and spatial and temporal coverage, limitations due to availability of input data, etc. Eventual changes or suggested modifications of the initial service request are proposed to the user. If rejected, interaction with the user allows to agree on a new activation setup.
2	Input data collection	Download of all the social and traditional media emerged during all the days of the extreme event of interest on the AOI. An additional input required during this phase is the collection of the UF-ID-4 indicator output.
3	Data harmonization	In this stage, all the social and traditional media collected for the event of interest undergo preprocessing to ensure harmonization and readiness for subsequent steps. These media can originate from diverse sources and encompass various formats (e.g., videos, comments). To enable proper mapping, each marker undergoes verification in terms of location accuracy and information quality. Subsequently, they are transformed into a vector layer for further analysis.
4	Quality check & consistency check	A validation process is carried out to ensure the integrity of the generated products, encompassing both automated and manual checks. The automatic validation involves employing keyword filters to sift through the data and eliminate any potential fake news or misinformation. Subsequently, a manual validation is executed, wherein an operator visually assesses the products, gauging their quality and coherence. Any products that do not meet the required quality and consistency standards are excluded from further consideration.
5	Social/traditional media mapping	In this stage, the traditional and social media markers identified in the previous step are geographically positioned. Each marker is depicted on the map as a vector point. This mapping process is linked to each day of the urban event of interest. For those days where InSAR data is available, the mapping also includes the flood delineation obtained from ID-4, alongside the vector representation of the media markers.
6	Product Packaging & Delivery	This phase includes the preparation of the product package to be delivered to the end user through the Service Data Lake and to be made available to the user through the Exploitation System.

Input data

The input data used for the indicator generation are derived from:

- Social media containing markers.
- Traditional media containing markers.
- UF-ID-4 Inferred INSAR urban flood extent indicator.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Below, in Table 22 the details of the input data used for the indicator calculation.

Table 22: Input data required for UF-ID-6

Input data Name	Description	Format
Social media containing markers	Social media markers refer to specific pieces of information or content shared on social media platforms, which can provide insights into events, incidents, or trends. These markers include various types of content, such as: 1. **Geotagged Posts:** Posts on platforms like Twitter, Instagram, or Facebook that include location information, allowing them to be associated with a specific geographic point. 2. **Photos and Videos:** Visual content shared on social media that can provide real-time documentation of events or situations. 3. **Comments and Descriptions:** Textual descriptions, comments, or captions that accompany posts and provide context or details about an event. 4. **Hashtags and Keywords:** Specific keywords or hashtags that are used to label and categorize posts related to a particular topic or event. Social media markers can be particularly valuable for understanding real-time developments, public sentiment, and local conditions during various events, such as natural disasters, protests, emergencies, or other significant occurrences. They offer a way to gather and analyse information from a wide range of sources, often providing on-the-ground perspectives that may not be captured by traditional news outlets.	Source related
Traditional media containing markers	Traditional media markers refer to specific elements or cues found within traditional media sources, such as newspapers, television broadcasts, radio reports, and other established forms of media. These markers help identify and track certain events, trends, or topics. Examples of traditional media markers include: 1. **News Articles:** Articles in newspapers or online news platforms that report on a particular event, issue, or development. 2. **TV News Coverage:** News segments or reports aired on television news programs that cover specific incidents, stories, or occurrences. 3. **Radio Reports:** News updates or features broadcasted on radio stations that discuss noteworthy subjects. 4. **Headlines and Coverage Frequency:** The prominence of a particular story, indicated by its placement on the front page of a newspaper or its frequency of coverage across different media outlets. 5. **Editorials and Opinions:** Editorial pieces, op-eds, or commentary articles that express viewpoints and opinions on certain matters. 6. **Photographs and Visuals:** Images or visual content accompanying news articles or broadcasts that provide visual context to the story. Traditional media markers help researchers and analysts track how events are reported, the level of attention they receive, and how they are perceived by the public. These markers contribute to understanding the narrative and impact of events in a broader societal context.	Source related
UF-ID-4	A flooding mask is created using SAR satellite sensor data, with a specific focus on urban areas. This resulting output from the FLORIA model is subsequently employed as an input for the Speedy Flood tool.	Vector (shapefile)

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output

This indicator effectively addresses the constraints posed by other tools used in flood mapping, such as satellite sensor revisit times and computational modelling time steps (Table 23). It stands out as highly efficient in terms of information generation speed, making it one of the fastest options available. Traditional and social media markers emerge as one of the earliest sources accessible during and after an event. With a set of maps featuring geolocated markers for each day of the extreme event, a rapid and lucid understanding of the areas most impacted by flooding becomes readily attainable.

Table 23: Output data description and specifications for UF-ID-6

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Social/Traditional media indicators for Urban Flooding Map	Indicator that shows the urban context hit by floods using social and traditional media markers.	vector (geojson/shapefile)	NA	This indicator is currently absent in operational modes and services. Swift product availability is crucial, particularly for emergency support. This indicator provides a clear visualization of flood-affected areas, enabling timely intervention and response measures.

3.2.7 UF-ID-7: Hazard web sources indicator

Description

Indicator based on several input data available on the web that allow to characterize an extreme flood event in an area of interest in terms of hazard. This flood indicator could generally be used to understand floods from management perspective. The indicator is designed to recognize the flood intensities including their magnitude and clustering features.

Workflow

In Figure 11 the workflow schema and a detailed description of the step-by-step design of the indicator is given.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

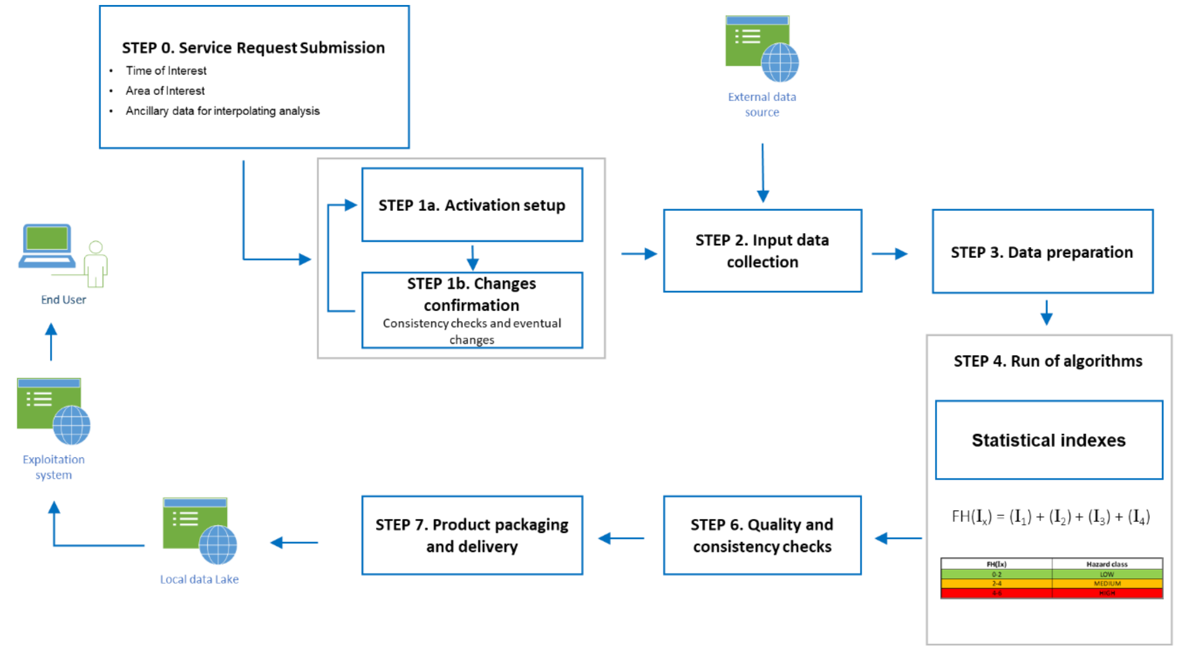


Figure 11: Workflow schema of the UF-ID-7-Hazard web sources indicator

In particular, in Table 24 below are reported the input data used for the indicator calculation.

Table 24: Description of the steps part of the workflow for UF-ID-7

Step #	Name	Description
0	Service Request Submission	The user activates the indicator providing, through the specific Dashboard, information like: • Time of Interest. • Area of Interest. • Upload of ancillary data for interpolating analysis. • Information about the flood of interest.
1a	Activation setup	Depending on available data and processing options set up by the end user (frequency, AoI, etc), processing is activated.
1b	Changes confirmation	This step performs additional checks on End User request, e.g. consistency with requested products and spatial and temporal coverage, limitations due to availability of input data, etc. Eventual changes or suggested modifications of the initial service request are proposed to the user. If rejected, interaction with the user allows to agree a new activation setup.
3	Data preparation	All data are pre-processed and resampled in order to match the temporal and spatial resolution needed for the next steps. The data from the different indicators have different formats. For example, in the case of GDACS, we have tabular information, in the case of the precipitation that occurred which is useful for defining the hazard of the event is a raster, while for population and poverty distribution we have vector layers. These different formats, to be considered in the algorithm, must be unified. The final format that is given as input to the model, is tabular. The input files properly harmonized, are then aggregated- if there is more than one reporting information on

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

<table>
<tr><th>Step #</th><th>Name</th><th>Description</th></tr>
<tr><td></td><td></td><td>the same issue- so that they are ready for the run of the extreme event hazard assessment model.</td></tr>
<tr><td>4</td><td>Run algorithm</td><td>Each indicator considered is represented in the model as a weight (Ix) and intervenes in the following equation useful for defining hydraulic hazard (main output of the model): FH(I_x) = (I_1) + (I_2) + (I_3) + (I_4). In quantitative terms, the sum of the proposed indices will result in 3 possible hazard classes:
Low
Medium
High
<table>
<tr><th>FH(Ix)</th><th>Hazard class</th></tr>
<tr><td>0-2</td><td>LOW</td></tr>
<tr><td>2-4</td><td>MEDIUM</td></tr>
<tr><td>4-6</td><td>HIGH</td></tr>
</table></td></tr>
<tr><td>3</td><td>Quality check & consistency check</td><td>A consistency check of the generated products is automatically done. The quality check consists of automatic steps to verify the quality level of the generated products. Products with not enough quality and/or consistence are discarded.</td></tr>
<tr><td>4</td><td>Product Packaging & Delivery</td><td>This phase includes the preparation of the product package to be delivered to the end user through the Service Data Lake and to be made available to the user through the Exploitation System.</td></tr>
</table>

Input data

The input data used for the indicator generation are derived from:

- European Union portal Copernicus where it is possible to see in detail the cartographies and the damages of the specific event.
- Global Disaster Alert and Coordination System (GDACS) portal which is a cooperation framework between the United Nations, the European Commission and disaster managers worldwide to improve warnings in case of risky events.
- CEMS RM and/or RRM activations info in terms of affected population and size of the flooded area
- WFS-Population data.
- GHS-Pop R2023A.
- WSF - World Settlement Footprint.
- GHS-Built-S R2023A.

Below, in Table 25 the details of the input data used for the indicator calculation.

Table 25: Input data required for UF-ID-7

Input data Name	Description	Format
GDACS hazard indicator	GDACS is a cooperation framework between the United Nations and the European Commission. It includes disaster managers and disaster information systems worldwide and aims at filling the information and coordination gaps in the first phase after major disasters. GDACS provides real-time access to web-based disaster information systems and related coordination tools.	Geojson

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
WFS-Population	Residential population estimates derived from CIESIN GPW4.11 are disaggregated from census or administrative units to grid cells, informed by the distribution, height and functional use of settlements derived from the WSF-datasets. Each dataset has a spatial resolution of 10m and temporal coverage of 2016-to present.	Raster (.tif)
GHS-Pop-R2023A	Residential population estimates between 1975 and 2020 in 5-year intervals and projections to 2025 and 2030 derived from CIESIN GPWv4.11 were disaggregated from census or administrative units to grid cells, informed by the distribution, volume, and classification of built-up as mapped in the Global Human Settlement Layer (GHSL) global layer per corresponding epoch. Each dataset has a resolution of 100m	Raster Layer
WSF - World Settlement Footprint	The dataset depicts the extent and location of human settlements derived from the analyses of multi-temporal statistics in radar an optical data from Sentinel 1 and Sentinel-2 imagery.	Raster layer (.tif)
GHS-Built-S R2023A	Global Human Settlement- Built-up surface grid. The dataset depicts the distribution of built-up surfaces, expressed as number of square metres. The data is made by spatial-temporal interpolation of 5 observed collections. Landsat data supports 1975,1990,2000 and 2014 epochs, while S2 supports 2018 epoch.	Raster layer (.tif)
CEMS RM and/or RRM activations info in terms of affected population and size of the flooded area	CEMS RM: This service consists of the on-demand and fast provision (hours-days) of geospatial information in support of emergency management activities immediately following disaster. The service is based on the acquisition, processing, and analysis, in rapid mode, of satellite imagery and other geospatial raster and vector data sources, and social media when relevant. CEMS RRM: This service provides on-demand geospatial information to support emergency management activities outside the immediate response phase. It covers the prevention, preparedness, disaster risk reduction, and recovery phases and is divided into two sub-categories: ▪ Risk and Recovery STANDARD for a predefined set of standardized products, ▪ Risk and Recovery FLEX for tailor-made studies.	Gdb (vector and raster layers)
Web Sources	Data available from web that allow to obtain information on the scenario involved by flood and that is object of the simulation. These data describe the scenario from different point of views: facilities and industries present in the context of interest, economic damages estimated; poverty distribution, elements of particular interest hit by flood, etc.	Depending on the data source (could be vector layer and/or tabular data)

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output

This output (Table 26), useful especially during the water management phase, allow to identify quickly the urban areas most vulnerable by flood and to give more precise information on the situation from different point of view: population distribution, population poverty, event magnitude, particular facilities and industries present in the area of interest.

Table 26: Output data description and specifications for UF-ID-7

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Hazard web sources indicator	Indicator that clearly shows the flood hazards emerged by the event of interest. The flood hazard is characterized by three different classes: low, medium, and high.	Table	Event related	To date, there is no indicator in the operational flood services that allows to consider different sources such as those considered in this algorithm. This type of indicator makes it possible to clearly highlight flood hazard by considering different aspects such as: extreme events from the meteorological point of view, population composition, poverty classes and hazard elements present in the urban context.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3.3 WATER & FOOD SECURITY INDICATORS

BIOPHYSICAL PARAMETERS AND AGRICULTURE, WATER AND METEOROLOGICAL VARIABLES.

3.3.1 WFS-ID-1: Meteorological drought indicator (Monitoring)

Description

This indicator is based on observation-based meteorological input data, and it identifies the occurrence and severity of meteorological droughts at the global scale in near real-time. While it focuses on droughts, it can also identify unusually wet periods such as floods, and thus enables us to better understand the impact of natural hazards on water and food security from a monitoring perspective. The indicator is produced at high-resolution (0.25) and will be updated frequently (approx. every 5 days, depending on data availability).

Workflow

In Figure 12, the workflow schematic and a detailed description of the step-by-step design of the indicator is provided.

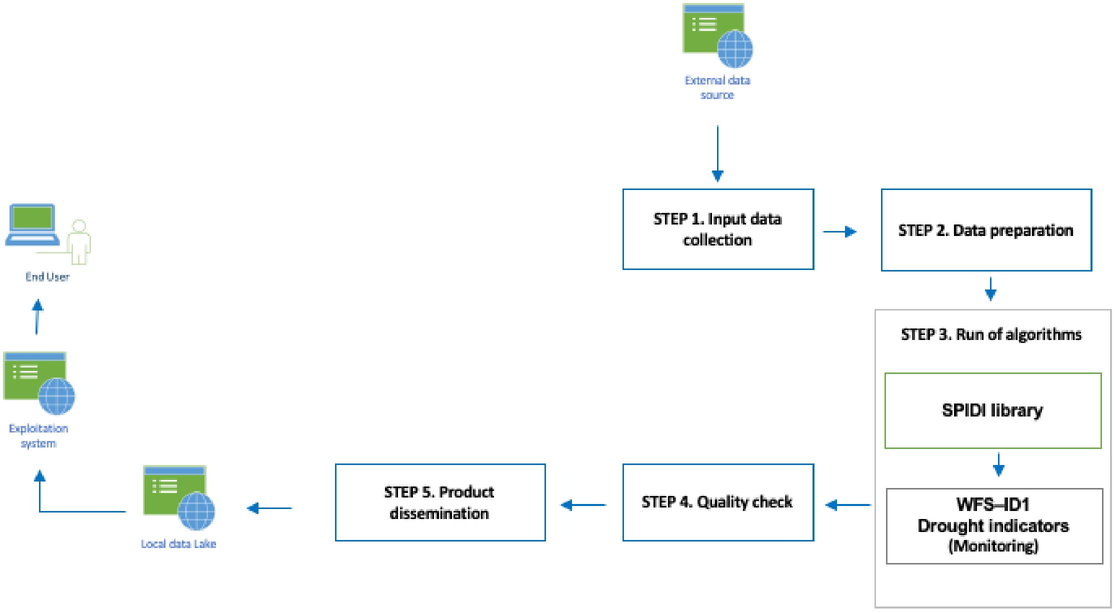


Figure 12: Workflow schematic of the WFS-ID-1-Meteorological drought indicator (Monitoring)

In particular, in Table 27 below are reported the input data used for the indicator calculation.

Table 27: Description of the steps part of the workflow for WFS-ID-1

Step #	Name	Description
1	Input data collection	The most recently available observed precipitation from GPCC and ERA5 are obtained, either through ECMWF or the web. The data sets are then concatenated to the historical observation period.
2	Data preparation	The input data (i.e., precipitation) is aggregated to the corresponding time scale of the drought indicator (i.e., monthly accumulation periods). The final format that is given as input to the model, is grib/netCDF. The input files properly harmonized before

Step #	Name	Description
		they are passed to the SPIDI library that evaluates the drought indicator.
3	**Run of algorithms (SPIDI library) to identify ongoing drought events**	Using the timeseries of precipitation at each grid point, ongoing drought events are identified using the 'Standardized Precipitation Index and other Drought Indices' (SPIDI) library. Therefore, the Standardized Precipitation Index (SPI) is calculated, using the full historical record of observations from each data set. To calculate the SPI, a gamma distribution is fitted to the full time series of precipitation, and then transformed to a normal distribution. The resulting SPI values can then be interpreted as the number of standard deviations by which the observed precipitation anomaly deviates from the long-term mean. Positive values indicate positive precipitation anomalies, i.e., wetter conditions compared to the long-term mean. Negative values indicate dry anomalies. The exceedance of certain thresholds, i.e., 1, 1.5, and 2 indicates moderate, severe, and extreme events. For example, an SPI lower than –2 indicates an extreme drought event. The SPI is calculated for different accumulation time periods, i.e., 1, 3, 6 and 12 months, using the cumulative sum of precipitation over these periods. The most recent SPI value is then indicative of ongoing drought events and can be illustrated as global maps that show the regions of the world experiencing moderate, severe, or extreme drought conditions. The most recent SPI values are always concatenated to the historical values, generating a drought archive near real-time.
4	**Quality check**	An automated quality check of the generated product will be performed. Products not fulfilling quality standards or with limited data availability will be discarded.
5	**Product dissemination**	The drought monitoring products will be delivered to the end user through the Service Data Lake and will be made available to the user through the Exploitation System and the Copernicus Data Store (CDS).

Input data

The input data used for the indicator generation are derived from:

- GPCC
- ECMWF ERA5 reanalysis

Below, in Table 28 the details of the input data used for the indicator calculation.

Table 28: Input data required for WFS-ID-1

Input data Name	Description	Format
Precipitation from GPCC	The Global Precipitation Climatology Centre (GPCC) is a service operated by the German Meteorological Service (Deutscher Wetterdienst, DWD) under the auspices of the World Meteorological Organisation (WMO). GPCC provides gridded precipitation data based on quality-controlled measurements from approximately 64.400 rain gauges worldwide. For near real-time monitoring,	Grib / NetCDF

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	the GPCC First Guess Product will be used. Precipitation from GPCC is available on a 1 grid from 1891 to today.	
Precipitation from ECMWF ERA5 reanalysis	ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather. The ERA5 reanalysis combines observations and models to derive a spatio-temporal complete and consistent dataset following the laws of physics. Data from ERA5 is available on a 0.25 grid from 1940 to today at hourly resolution. ERA5 is updated with a latency of 5 days, but an early release (ERA5T) is available near real-time monitoring.	Grib / NetCDF

<u>Output</u>

This monitoring drought indicator enables us to identify ongoing droughts worldwide near real-time (Table 29). To rate the severity of those droughts and its impact on the environment and society, several time ranges are considered, i.e., 1, 3, 6 and 12 months. The monitoring drought indicator will be provided for each time range. Using, for example, the standardised precipitation index (SPI), the four indicators SPI–1, SPI–3, SPI–6, SPI–12 will be outputted. The indicators over short time scales (SPI–1 and SPI–3) quantify the accumulated precipitation deficits over 1 and 3 months respectively and the SPI rates the corresponding precipitation anomaly with respect to the long-term historical rainfall at each location. Large anomalies of the SPI–1 and SPI–3 are typically used to indicate immediate impacts, such as reduced soil moisture or snowpack and decreasing flow in smaller creeks. The indicators aggregated over longer time scales (SPI–6 and SPI–12) are commonly used to indicate reduced stream flow and reservoir storage. For each time range, global snapshots of the occurrence of moderate, severe and extreme droughts will be generated.

Table 29: Output data description and specifications for WFS-ID-1

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
SPI–1	Indicator that rates the accumulated precipitation over 1 month with respect to the long-term historical record. Drought events are identified using thresholds of the SPI, i.e. moderate, severe and extreme drought events occur if the SPI is smaller than −1, −1.5, and −2, respectively. The SPI–1 is indicative of reduced soil moisture and snowpack availability and decreasing streamflow in smaller creeks.	Grib / netCDF	0.25, monthly (update frequency every 5 days)	There exist several operational drought monitors, such as the European and the Global Drought Observatory (EDO and GDO). However, to date, these are produced at coarser resolutions and updated less frequently.
SPI–3	Same as SPI–1, but accumulated precipitation over 3 months.	Grib / netCDF	Same as SPI–1	Same as SPI–1
SPI–6	Same as SPI–1, but accumulated precipitation over 6 months. The SPI–6 is indicative of reduced stream flow and reservoir storage.	Grib / netCDF	Same as SPI–1	Same as SPI–1
SPI–12	Same as SPI–1, but accumulated precipitation over 12 months. The SPI–12	Grib / netCDF	Same as SPI–1	Same as SPI–1

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
	is indicative of reduced stream flow and reservoir storage.			

3.3.2 WFS-ID-2: Meteorological drought indicator (Forecast)

Description

This indicator is based on meteorological forecasts from ECMWF, and it forecasts the occurrence and severity of meteorological droughts at the global scale at lead times spanning from day 1 to 6 months ahead. While it focuses on droughts, it can also forecast unusually wet periods that may cause flooding, and thus enables us to forecast potential impacts of those natural hazards on water and food security up to 6 months ahead. The indicator is produced at variable resolutions — from 9 km to 35 km, depending on lead time — and will be updated monthly.

Workflow

In Figure 13 the workflow schema and a detailed description of the step-by-step design of the indicator is provided.

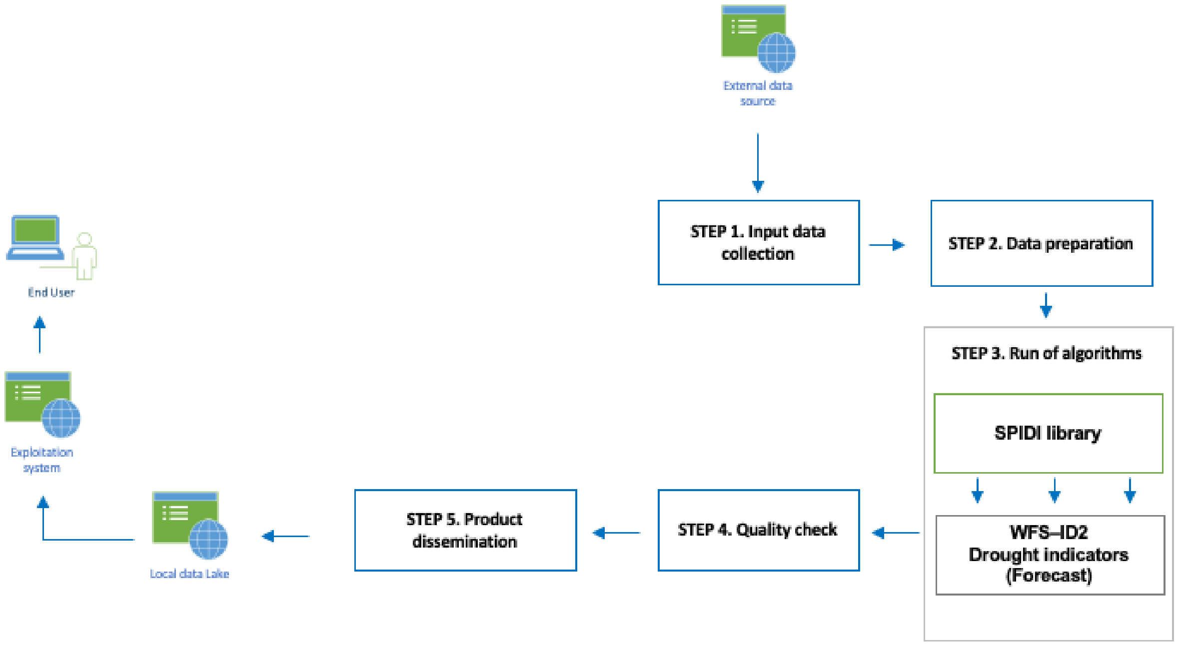


Figure 13: Workflow schematic of the WFS-ID-2-Meteorological drought indicator (Forecast)

In particular, in Table 30 below are reported the input data used for the indicator calculation.

Table 30: Description of the steps part of the workflow for WFS-ID-2

Step #	Name	Description
1	Input data collection	The three forecasts used for the calculation of this indicator are retrieved through the MARS archive at ECMWF.
2	Data preparation	All forecasts are pre-processed and concatenated according to their lead time, creating a seamless prediction of precipitation from day 1 to 6 months ahead that benefits from the skill of each forecast product. I.e., from day 1 – day 14, the medium-range ensemble

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		forecast (ENS) will be used. From day 15 – day 31, the extended-range ensemble forecast will be used (ENS–ER), and the remaining months (month 2 – month 6) will be based on the seasonal forecast. To create a long-time series covering at least 40 years, corresponding reforecasts are used and concatenated analogously. Precipitation is then aggregated to the corresponding time scale of the drought indicator (i.e., monthly accumulation periods). The final format that is given as input to the model, is grib/netCDF. The input files properly harmonized before they are passed to the SPIDI library that evaluates the drought indicator.
3	Run of algorithms (SPIDI library) to identify forecasted drought events and estimate the probability of drought occurrence and severity	For each forecast product at the native resolution and precipitation at monthly time scales, the SPI (see ch. 3.3.1, WFS-ID-1), is calculated. The forecasts are evaluated for the exceedance of certain thresholds, i.e., –1, –1.5, and –2 to identify moderate, severe, and extreme events, respectively. This is done for every ensemble member of each forecast class (ENS, ENS–ER, SEA) and a probability of drought occurrence for each severity category (moderate, severe, extreme) is calculated using this ensemble. Analogously to the WFS–ID–1, the SPI is calculated for different accumulation time periods, i.e., 1, 3, 6 and 12 months, using the cumulative sum of precipitation over these periods, indicating the duration of the forecasted drought events. The probability of drought occurrence and severity can then be illustrated as global maps that show the regions of the world that might experience moderate, severe, or extreme drought conditions in the coming 6 months.
4	Quality check	An automated quality check of the generated product will be performed. Products not fulfilling quality standards or with limited data availability will be discarded.
5	Product Delivery	The drought monitoring products will be delivered to the end user through the Service Data Lake and will be made available to the user through the Exploitation System and the Copernicus Data Store (CDS).

Indicator input data

The input data used for the indicator are derived from:

- ECMWF ensemble forecast and reforecasts (ENS and ENS–ER)
- ECMWF seasonal forecast and reforecasts (SEA)
- ECMWF ERA5 reanalysis
- Below, in Table 28 the details of the input data used for the indicator calculation.

Below, in Table 31 the details of the input data used for the indicator calculation.

Table 31: Input data required for WFS-ID-2

Input data Name	Description	Format
Precipitation from ECMWF ensemble	The medium-range ensemble forecast (ENS) from ECMWF is a global weather forecast with a lead time of 15 days. From cycle 48r1 onwards (implemented in June 2023), this forecast will run on a 9km grid and will be	Grib / NetCDF

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
forecast (ENS and ENS–ER)	updated 4 times daily (00, 06, 12, and 18 UTC). It consists of 51 ensemble members. In addition, an extended-range ensemble forecast (ENS–ER), consisting of 101 members and spanning a lead time of 46 days, is issued every day at 00 UTC. For each forecast (ENS and ENS–ER), corresponding reforecasts are produced, enabling the construction of a long-term record.	
Precipitation from ECMWF seasonal forecast (SEA)	The seasonal forecast from ECMWF (SEA) is based on the SEAS5 system with a resolution of 35 km and predicts up to 7 months ahead. It is a 51-member ensemble that is on the first day of the month at 00 UTC.	Grib / NetCDF
Precipitation from ECMWF ERA5 reanalysis	ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather. The ERA5 reanalysis combines observations and models to derive a spatio-temporal complete and consistent dataset following the laws of physics. Data from ERA5 is available on a 0.25 grid from 1940 to today at hourly resolution. ERA5 is updated with a latency of 5 days, but an early release (ERA5T) is available near real-time monitoring.	Grib / NetCDF

Output

This forecasting drought indicator enables us to identify regions droughts worldwide near real-time (Table 32). To rate the severity of those droughts and its impact on the environment and society, several time ranges are considered, i.e., 1, 3, 6 and 12 months. The monitoring drought indicator will be provided for each time range. Using, for example, the standardised precipitation index (SPI), the four indicators SPI–1, SPI–3, SPI–6, SPI–12 will be outputted. The indicators over short time scales (SPI–1 and SPI–3) quantify the accumulated precipitation deficits over 1 and 3 months respectively and the SPI rates the corresponding precipitation anomaly with respect to the long-term historical rainfall at each location. Large anomalies of the SPI–1 and SPI–3 are typically used to indicate immediate impacts, such as reduced soil moisture or snowpack and decreasing flow in smaller creeks. The indicators aggregated over longer time scales (SPI–6 and SPI–12) are commonly used to indicate reduced stream flow and reservoir storage. For each time range, global snapshots of the occurrence of moderate, severe and extreme droughts will be generated.

Table 32: Output data description and specifications for WFS-ID-2

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
SPI–1	Indicator that rates the forecasted accumulated precipitation over 1 month with respect to the long-term historical record. Drought events are identified using thresholds of the SPI, i.e. moderate, severe and extreme drought events occur if the SPI is smaller than –1, –1.5, and –2, respectively. The SPI–1 is indicative of reduced soil moisture and snowpack availability and decreasing streamflow in smaller creeks.	Grib / netCDF	0.25, monthly (update frequency every day or every 5 days)	To date, there exists no operational global drought forecast.
SPI–3	Same as SPI–1, but accumulated precipitation over 3 months.	Grib / netCDF	Same as SPI–1	Same as SPI–1

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
SPI–6	Same as SPI–1, but accumulated precipitation over 6 months. The SPI–6 is indicative of reduced stream flow and reservoir storage.	Grib / netCDF	Same as SPI–1	Same as SPI–1
SPI–12	Same as SPI–1, but accumulated precipitation over 12 months. The SPI–12 is indicative of reduced stream flow and reservoir storage.	Grib / netCDF	Same as SPI–1	Same as SPI–1
Probability of moderate drought	Indicator that evaluates a large ensemble of SPI values up to 6 months ahead. Here, the threshold of SPI being smaller than –1 is evaluated for all time ranges (i.e., SPI–1, SPI–3, SPI–6, SPI–12).	Grib / netCDF	Same as SPI–1	Same as SPI–1
Probability of severe drought	Same as probability of moderate drought but using a threshold of SPI being smaller than –1.5.	Grib / netCDF	Same as SPI–1	Same as SPI–1
Probability of extreme drought	Same as probability of moderate drought but using a threshold of SPI being smaller than –2.	Grib / netCDF	Same as SPI–1	Same as SPI–1

3.3.3 WFS-ID-3: Meteorological drought indicator (danger levels)

Description

This indicator evaluates the meteorological drought forecast with respect to the monitoring state and estimates danger levels. It is based is based on the two previous meteorological drought indicators, i.e., WFS–ID1 and WFS–ID2. Over already existing drought regions, the probabilities of drought continuation, drought aggravation and drought recovery are estimated. Over regions that do not experience drought yet, the probability of drought occurrence and the corresponding severity are estimated. The indicator is produced at a resolution of 35 km and will be updated monthly.

Workflow

In Figure 14 the workflow schematic and a detailed description of the step-by-step design of the indicator is provided.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

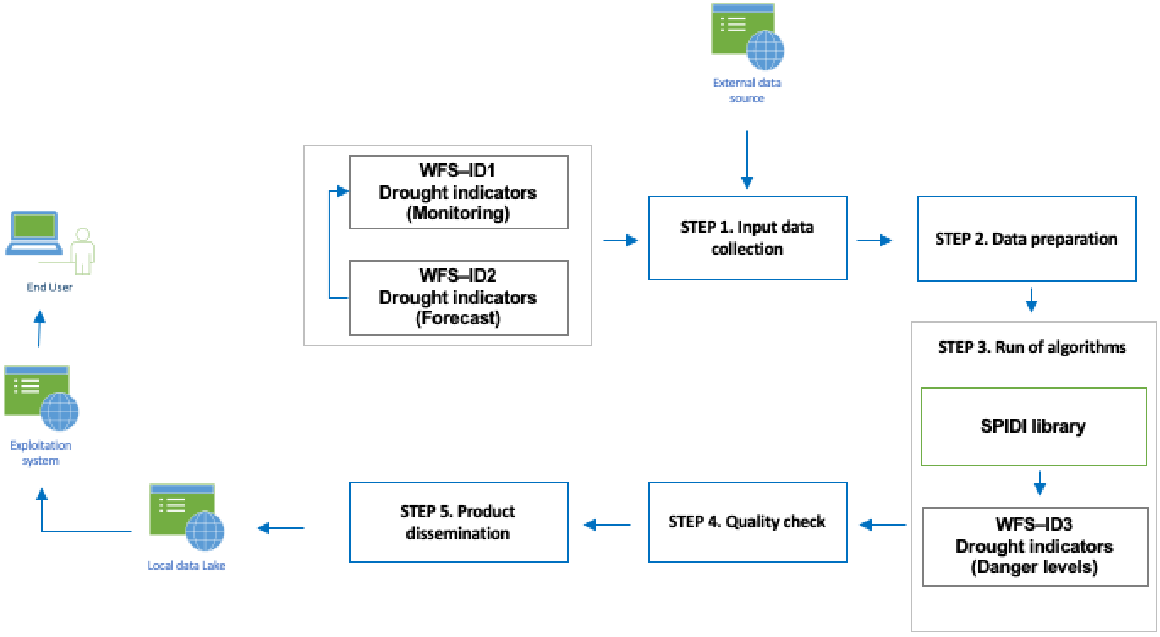


Figure 14: Workflow schematic of the WFS-ID-3-Meteorological drought indicator (Danger levels)

In particular, in Table 33 below are reported the input data used for the indicator calculation.

Table 33: Description of the steps part of the workflow for WFS-ID-3

Step #	Name	Description
1	Input data collection	The two previous meteorological drought indicators are used as input and are available inhouse. Additional external data sources, used to validate the monitoring status, may be used. The latest monitoring status as well as the latest forecast are retrieved.
2	Data preparation	Both indicator data sets are pre-processed and made comparable, i.e., time and grid are unified.
3	Run of algorithms (SPIDI library) to evaluate the forecasted drought with respect to the monitoring status	The evaluation of the forecast with respect to the monitoring is done using the SPIDI library. There exist three internal steps for this evaluation. First, regions already experiencing a drought (using WFS–ID–1) are delineated. For those regions, the forecast (WFS–ID–2) is evaluated and the probability of drought aggravation or drought recovery is estimated. Second, the forecasts for regions that are — according to the monitoring indicator — not in drought, are evaluated. For those regions, the probability of drought occurrence is estimated, and its estimated severity is provided using thresholds of the SPI at various time scales. Finally, these probabilities can then be translated into danger levels highlighting regions that are likely to experience continued or aggravating drought conditions, that are expected to impact water availability.
4	Quality check	An automated quality check of the generated product will be performed. Products not fulfilling quality standards or with limited data availability will be discarded.
5	Product Delivery	The drought monitoring products will be delivered to the end user

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		through the Service Data Lake and will be made available to the user through the Exploitation System and the Copernicus Data Store (CDS).

Input data

The input data used for the indicator are derived from:

- WFS–ID1
- WFS–ID2

Below, in Table 34 the details of the input data used for the indicator calculation.

Table 34: Input data required for WFS-ID-3

Input data Name	Description	Format
WFS–ID1	This indicator is based on observation-based meteorological input data, and it identifies the occurrence and severity of meteorological droughts at the global scale in near real-time. The indicator is produced at high-resolution (0.25) and will be updated frequently (approx. every 5 days, depending on data availability). See Chapter 3.3.1.	Grib / NetCDF
WFS–ID2	This indicator is based on meteorological forecasts from ECMWF, and it forecasts the occurrence and severity of meteorological droughts at the global scale at lead times spanning from day 1 to 6 months ahead. This indicator is produced at variable resolution, from 9 to 36 km. See Chapter 3.3.2	Grib / NetCDF

Output

This forecasting drought indicator enables us to identify regions droughts worldwide near real-time (Table 35). To rate the severity of those droughts and its impact on the environment and society, several time ranges are considered, i.e., 1, 3, 6 and 12 months. The monitoring drought indicator will be provided for each time range. Using, for example, the standardised precipitation index (SPI), the four indicators SPI–1, SPI–3, SPI–6, SPI–12 will be outputted. The indicators over short time scales (SPI–1 and SPI–3) quantify the accumulated precipitation deficits over 1 and 3 months respectively and the SPI rates the corresponding precipitation anomaly with respect to the long-term historical rainfall at each location. Large anomalies of the SPI–1 and SPI–3 are typically used to indicate immediate impacts, such as reduced soil moisture or snowpack and decreasing flow in smaller creeks. The indicators aggregated over longer time scales (SPI–6 and SPI–12) are commonly used to indicate reduced stream flow and reservoir storage. For each time range, global snapshots of the occurrence of moderate, severe and extreme droughts will be generated.

Table 35: Output data description and specifications for WFS-ID-3

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Probability of drought aggravation	Indicator that evaluates the drought forecast with respect to the monitoring state. Over areas of existing drought, the probability of drought	Grib / netCDF	0.25, monthly	To date, there exists no comparable drought forecast that evaluates the forecast with respect to the monitoring states and the expected impact

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
	continuation and aggravation is estimated.			
Probability of drought recovery	Indicator that evaluates the drought forecast with respect to the monitoring state. Over areas of existing drought, the probability of drought attenuation and recovery is estimated.	Grib / netCDF	Same as probability of drought aggravation	Same as probability of drought aggravation
Probability of drought occurrence	Indicator that evaluates the drought forecast with respect to the monitoring state. Over areas without an existing drought, the probability of drought occurrence is estimated.	Grib / netCDF	Same as probability of drought aggravation	Same as probability of drought aggravation

3.3.4 WFS-ID-4: Agricultural drought monitoring (near real-time)

Description

Indicator that expresses the relative impact on a 0-100 scale of drought events on the current vegetation condition and productivity. It compares the current situation in terms of plant condition (Normalized Difference Vegetation Index, NDVI), plant drought stress (thermal stress indicator) and soil moisture conditions to a “normal baseline” for the region under consideration, the latter being based on a historical archive of indicators. The spatial resolution of this product will be 1 km and it will be updated every 10 days.

Workflow

This indicator will be generated automatically and continuously at dekadal (10-daily) interval for each country supported in the Centaur data platform. In practice, this implies that no user interaction is required to initiate product generation for a given AOI. Figure 15 illustrates the workflow involved in generating indicator WFS-ID-4, with the different steps being further specified in the table below.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

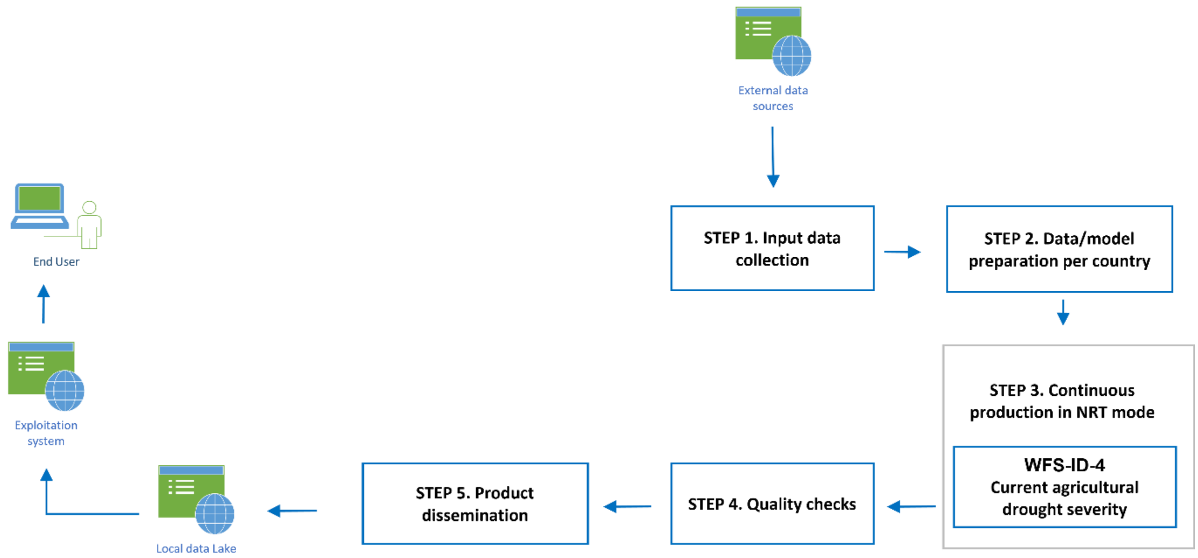


Figure 15: Workflow for generating agricultural drought indicator (WFS-ID-4)

In particular, in Table 36 below are reported the input data used for the indicator calculation.

Table 36: Description of the steps part of the workflow for WFS-ID-4

Step #	Name	Description
1	Input data collection	All input data required for computing the indicator (see next section) is gathered for the country of interest.
2	Data/model preparation per country	Before the indicator can be offered operationally for a given country, the “normal baseline” needs to be established for this country. This baseline will be constructed based on analysing an extensive historical archive of indicators related to vegetation condition (NDVI), thermal drought stress (calculated as Land surface temperature – air temperature), root zone soil moisture and meteorological information (mainly precipitation). This preparational analysis will result in a zonation of the country, each zone representing a region with uniform climatological conditions and uniform behaviour in terms of vegetation response to drought conditions. Recent land cover maps will be used to focus the analysis on vegetated areas only. For each of the resulting zones, the dominant phenological cycles will be derived as well. Next, zone-specific thresholds will be derived in terms of meteorological and vegetation conditions that mark the occurrence of an agricultural drought event.
3	Continuous production of indicator in NRT mode	Once the zonation along with its zone-specific thresholds and growing seasons has been established, the continuous monitoring part can start. For each dekad, an analysis will be made comparing recent and present meteorological and vegetation conditions to the normal baseline situation as expected based on the long-term archive of said indicators. In case the pixel is outside the expected growing seasons for its region, it is marked as “out of season”. For all “in season” pixels, we determine (based on the zone-specific

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		thresholds) whether an agricultural drought event is ongoing or not and what the expected relative impact will be on vegetation productivity. This indicator will also consider the specific stage of the growing season the pixel is currently at, as it is well known that drought conditions have a more significant impact on vegetation productivity during the early stages of crop/vegetation development.
4	**Quality check**	An automated quality check of the generated product will be performed. Pixels for which too few valid EO observations were available, will be flagged in the output products.
5	**Product dissemination**	The final indicator will be delivered to the end user through the Service Data Lake and will be made available to the user through the Exploitation System. The indicator will be delivered at the level of individual countries.

Input data

The input data used for the indicator are derived from:

- NDVI
- LST
- Root zone soil moisture (SMAP)
- Root zone soil moisture wetness (GRACE)
- Precipitation
- Air temperature
- Land cover

Below, in Table 37 the details of the input data used for the indicator calculation.

Table 37: Input data required for WFS-ID-4

Input data Name	Description	Format	Resolution
NDVI	The NDVI is an indicator of vegetation greenness. We will be using the harmonized NDVI collection offered through the Copernicus Global Land service. The data is derived from a combination of SPOT-VEGETATION, Proba-V and Sentinel-3 measurements.	GeoTIFF	1 km – 300 m Every 10 days
LST	Land Surface temperature derived from Sentinel-3 and/or MODIS thermal satellite measurements.	GeoTIFF	1 km Daily
Root zone soil moisture (SMAP)	L4 Root Zone Soil Moisture product as provided by NASA and based on the SMAP mission. Provides an indication of the amount of water stored in the first m of the soil.	HDF5	9 km Daily
Root zone soil moisture (GRACE)	Root zone soil moisture wetness percentile based on NASA GRACE gravimetric data. Provides a relative	GeoTIFF	25 km Every 7 days

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format	Resolution
	indication of the amount of water stored in the first m of the soil.		
Precipitation	Total precipitation, including rain and snow, that falls to the Earth's surface. The data will be retrieved from the ERA5-Land dataset, provided by ECMWF through the Copernicus Climate Data Store.	GRIB	9 km Hourly
Air temperature	Air temperature measured at 2m above the ground surface. The data will be retrieved from the ERA5-Land dataset, provided by ECMWF through the Copernicus Climate Data Store.	GRIB	9 km Hourly
Land cover	Yearly moderate-resolution land cover maps as provided through the Copernicus global land cover product. Land cover maps represent spatial information on different types (classes) of physical coverage of the Earth's surface, e.g. forests, grasslands, croplands, lakes, wetlands.	GeoTIFF	100 m Yearly

Output

The Table 38 below summarizes the main characteristics of the indicator to be produced.

Table 38: Output data description and specifications for WFS-ID-4

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Agricultural drought monitoring (near real-time)	Indicator that expresses the relative impact of drought events on the current vegetation condition and productivity. It integrates current and historical information on plant condition (NDVI), thermal drought stress and environmental conditions (precipitation, temperature, and soil moisture).	GeoTIFF	1 km Dekadal	Current drought early warning systems are mostly based on pixel-by-pixel temporal anomaly analysis of (primarily) NDVI data. However, an important lag time exists between the onset of impact of droughts on crop production and a visible drop in indicators such as NDVI. By combining NDVI with thermal drought stress and soil moisture indicators, we will provide more timely estimates of drought impacts on vegetation productivity. In addition, our indicator will be calibrated at regional level, as such considering important variations in spatial context, in turn mainly driven by climate, land cover/use and elevation.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3.3.5 WFS-ID-5: Agricultural drought forecast

Description

Indicator that expresses the relative impact, on a 0-100 scale of future drought events on the expected vegetation condition and productivity for different lead times ranging from a few days up to several months. It starts from the current situation in terms of agricultural drought conditions and looks at forecasts of meteorological and soil moisture conditions to determine the expected impact of drought conditions on vegetation productivity weeks up to months in advance. The spatial resolution of this product will be 1 km and it will be updated every 10 days.

Workflow

This indicator will be generated automatically and continuously at dekadal (10-daily) interval for each country supported in the Centaur data platform. In practice, this implies that no user interaction is required to initiate product generation for a given AOI. Figure 16 illustrates the workflow involved in generating indicator WFS-ID-5, with the different steps being further specified in the table below.

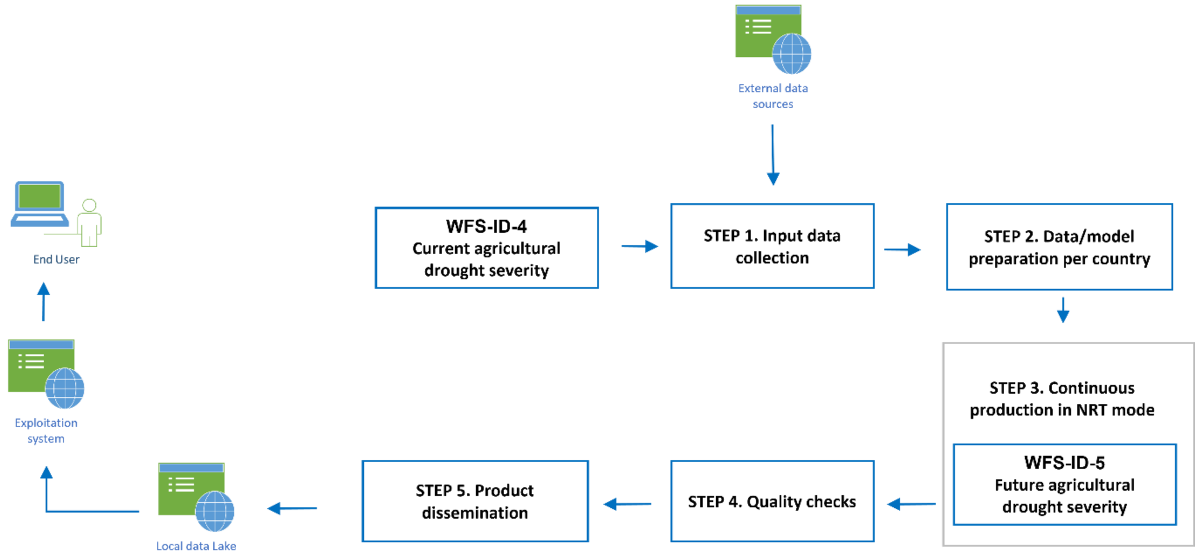


Figure 16: Workflow for generating agricultural drought forecasts indicator WFS-ID-5

In particular, in Table 39 below are reported the input data used for the indicator calculation.

Table 39: Description of the steps part of the workflow for WFS-ID-5

Step #	Name	Description
1	Input data collection	All input data required for computing the indicator (see next section) is gathered for the country of interest. This input data consists of one hand of the WFS-ID-4 indicator and on the other hand of meteorological forecasts as produced by ECMWF.
2	Data/model preparation per country	A threshold-based decision support system will be calibrated for each individual country based on a large historical archive of the agricultural drought indicator (WFS-ID-4) and historical forecasts of meteorological conditions and soil moisture anomalies. This will result in zone specific thresholds (same zones will be used as were developed for WFS-ID-4 indicator) indicating the relative impact of

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
		future drought events given current drought conditions and future meteorological conditions.
3	**Continuous production of indicator in NRT mode**	Once established, the decision support system can be run in near-real time, where for each dekad several forecasts will be produced based on the latest results from indicator WFS-ID-4 and latest forecasts of meteorological conditions and expected soil moisture anomalies. As a result, for each dekad the indicator will estimate the likelihood of occurrence and the impact (in relative terms) of future agricultural drought events for several lead times, ranging from a few days to several months in advance.
4	**Quality check**	An automated quality check of the generated product will be performed. Pixels for which too few valid EO observations were available, will be flagged in the output products.
5	**Product dissemination**	The final indicator will be delivered to the end user through the Service Data Lake and will be made available to the user through the Exploitation System. The indicator will be delivered at the level of individual countries.

Input data

This indicator will be heavily based on WFS-ID-4, which means that all inputs required for that indicator are relevant here as well. For brevity, they will not be included in the table below, where we only list new inputs used specifically for WFS-ID-5.

The input data used for the indicator are derived from:

- WFS-ID-4.
- Precipitation and air temperature forecast.
- Soil moisture anomaly forecast.

Below, in Table 40 the details of the input data used for the indicator calculation.

Table 40: Input data required for WFS-ID-5

Input data Name	Description	Format	Resolution
Precipitation and air temperature forecasts	Precipitation and air temperature from (i) ECMWFs ensemble forecast (ENS) up to 15 days ahead, (ii) ECMWFs extended-range ensemble forecast (ENS-ER) up to 46 days ahead, and (iii) ECMWFs seasonal forecast up to 6 months ahead (SEA)	Grib/NetCDF	variable (9 km (ENS), 36 km (ENS–ER), 35 km (SEA))
Soil moisture anomaly forecasts	Forecasts on soil moisture anomalies from ECMWF, using the same suite of forecasts as for precipitation and air temperature (i.e., ENS, ENS-ER, SEA)	Grib/NetCDF	variable (9 km (ENS), 36 km (ENS–ER), 35 km (SEA))

Output

The Table 41 below summarizes the main characteristics of the indicator to be produced.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Table 41: Output data description and specifications for WFS-ID-5

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Agricultural drought forecast	Indicator that expresses the likelihood of occurrence of future agricultural drought events and their relative impact on vegetation condition and productivity. It combines knowledge on the current drought conditions (cf.WFS-ID-4) with meteorological forecasts to predict the likelihood of adverse impacts on vegetation productivity potentially occurring in the future, up to several months in advance.	GeoTIFF	1 km Dekadal	Most existing agricultural drought early-warning systems only focus on monitoring the impact of the droughts in near real-time, but do not look into the future (i.e. several months ahead).

3.3.6 WFS-ID-6: Agricultural drought risk zone map

Description

Risk of occurrence of adverse drought impacts on vegetation productivity, expressed at a categorical scale (high risk – risk – no risk), considering both the present-day situation and future occurrence of agricultural drought events. Based on the outputs of indicators WFS-ID-4 and WFS-ID-5, each zone in a country will be classified into a risk category. The indicator will be updated every dekad based on the latest information on current vegetation condition and current/future meteorological conditions.

Workflow

This indicator will be generated automatically and continuously at dekadal (10-daily) interval for each country supported in the Centaur data platform. In practice, this implies that no user interaction is required to initiate product generation for a given AOI. Figure 17 illustrates the workflow involved in generating indicator WFS-ID-6, with the different steps being further specified in the table below.

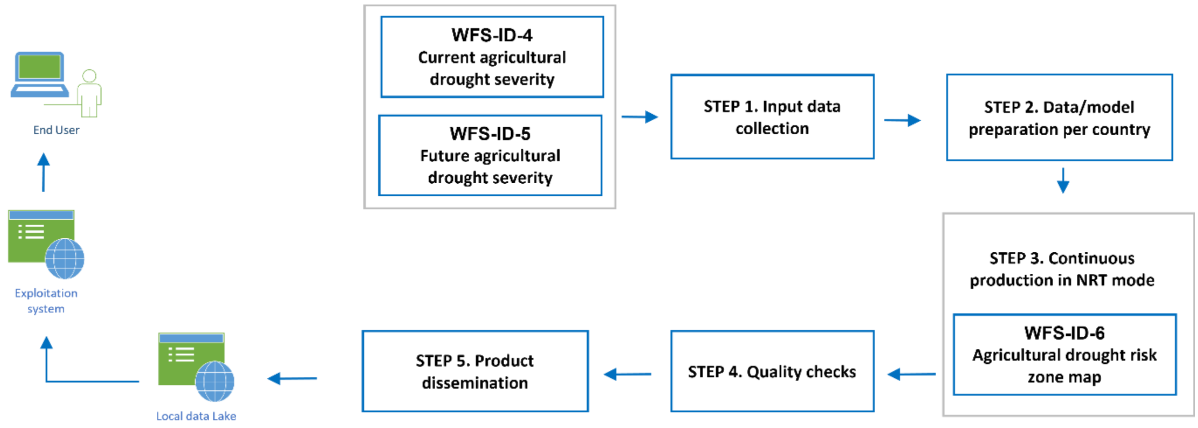


Figure 17: Workflow for generating agricultural drought risk zone map (WFS-ID-6).

In particular, in Table 42 below are reported the input data used for the indicator calculation.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Table 42: Description of the steps part of the workflow for WFS-ID-6

Step #	Name	Description
1	Input data collection	This indicator will be based on indicators WFS-ID-4 and WFS-ID-5.
2	Data/model preparation per country	A threshold-based decision support system will be calibrated for each individual country based on a large historical archive of the agricultural drought indicator (WFS-ID-4) and agricultural drought forecasts (WFS-ID-5). This will result in zone specific thresholds (same zones will be used as were developed for WFS-ID-4 indicator) indicating the risk (expressed in risk categories) in terms of occurrence and impact of agricultural droughts, both present-day and near future.
3	Continuous production of indicator in NRT mode	Once established, the decision support system can be run in near-real time, where for each dekad a risk zone map will be produced based on the latest results from indicators WFS-ID-4 and WFS-ID-5.
4	Quality check	An automated quality check of the generated product will be performed. Zones for which no conclusions can be drawn due to a lack of input data, will be flagged accordingly.
5	Product dissemination	The final indicator will be delivered to the end user through the Service Data Lake and will be made available to the user through the Exploitation System. The indicator will be delivered at the level of individual countries.

Input data

This indicator is purely based on the outcomes of indicators WFS-ID-4 and WFS-ID-5 and does not require additional input data (see ch. 3.3.5, Table 40).

Output

The Table 43 below summarizes the main characteristics of the indicator to be produced.

Table 43: Output data description and specifications for WFS-ID-6

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Agricultural drought risk zone map	Risk of occurrence of adverse drought impacts on vegetation productivity, expressed at a categorical scale (high risk – risk – no risk), considering both the present-day situation and future occurrence of agricultural drought events. A country will be split up into uniform zones with respect to meteorological conditions and vegetation response to droughts, after which every zone is categorized into one	GeoPackage	Dekadal	This indicator addresses the needs for an easily interpretable product capable of displaying visually and efficiently the risk of agricultural droughts at regional, national, and subnational scales, several months ahead in time. It is anticipated that the product will provide critical information in a manner that is immediately intelligible to the end user, who is likely

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
	of the abovementioned risk categories.			unfamiliar with the underlying science and technology.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3.4 SOCIO-ECONOMIC AND POLITICAL INDICATORS

Socio-economic and political indicators will be designed and implemented to capture possible socioeconomic and political impacts of climate-related shocks, as well as relevant socioeconomic and political scope conditions. Partly, these data will also serve for validating (i.e. ground truth) innovative indicators, as well as the results of predictive models developed in CENTAUR [T2.4, T.2.7].

Moreover, social/traditional media will be used to guide and support data collection over areas of interest, e.g. to identify initial areas of interest before satellite images become available. Furthermore, insights gained from social/traditional media will be used to trigger and complement remote sensing activities by providing immediate, from-the-ground information and contributing to the calibration of remote-sensing based estimates (such as water depth). The information provided by the multi-modal nature of social/traditional media - yielding text, images, videos, and audio – provides richer and complementary information to traditional remote-sensing based approaches and will, thus, enhance the scope of current CEMS and SEA activities.

Moreover, by integrating weather prediction data, CENTAUR platform will be able to identify the sources of remote sensing data that are most suitable to the type of event previously identified through the use media information (images, videos, TV, radio, press, etc.).

For the design and implementation of different indicators, in particular those related to exposure to security risks and resources shortage, updated official census data (e.g. administrative boundaries and population counts) will be collected, when possible, from sources such as National Statistical Offices (NSO). For those AOI where official census data are not openly available, population data will be derived from the raw global census data used for the Gridded Population of the World, version 4.11 (GPW4.11) produced by CIESIN.

Once population data is collected, it will be further enriched through the modelling of high-quality, high-spatial resolution population estimates. For the years 2016-to present, gridded population datasets will be produced by DLR, using a proprietary algorithm that relies on improved World Settlement Footprint products. There in, 10m spatial resolution gridded population datasets will be made available for all indicators that rely on population data. These datasets will represent population per pixel at given year and will be used mainly for monitoring tasks within the CENTAUR project. For forecasting events, 100m spatial resolution gridded population datasets from GHS-POP for the years 2025-2030 will be collected and harmonized accordingly.

Overall, the ability to assess risk with reference to specific areas of interest, combined with event detection, evaluation, and real-time information processing capabilities position CENTAUR as a project that focuses on actionable insights that support preparedness up to the actual event management phase. This seamless integration of these two phases of emergency management seems crucial to cope with the increased risk and occurrence of events due to climate change, which demand both awareness and rapid operational decisions.

Socio-economic and political indicators will be generated according to the general workflow outlined in Figure 18. More specifically, (sub)indicators based on social/traditional media will be generated following the workflow outlined in Figure 19.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

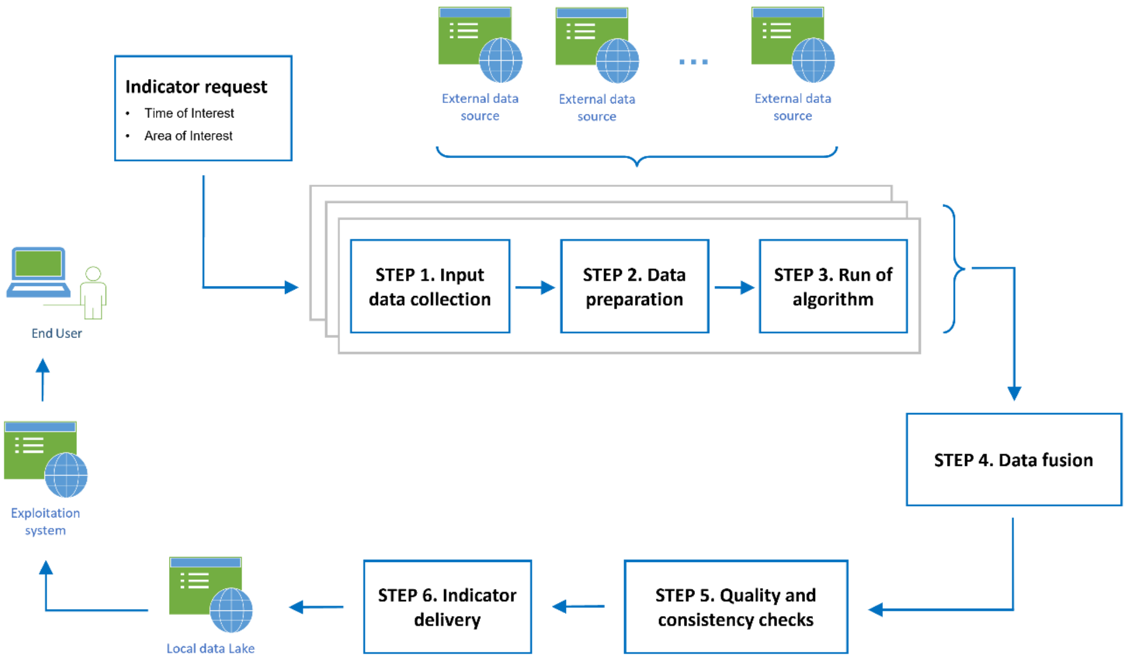


Figure 18: General workflow schema for the socio-economic and political indicators.

In particular, in Table 44 below are reported the input data used for the indicator calculation.

Table 44: Description of the steps part of the workflow for socio-economic and political indicators

Step #	Name	Description
0	Indicator request	Each request for any of the socio-economic and political indicators must provide information, such as: • Time of Interest. • Area of Interest.
1a	Input data collection: Activation setup	Depending on available data and processing options set up by the end user (time, area of interest etc.), processing is activated.
1b	Input data collection: Changes confirmation	This step performs additional checks on End User request (e.g. consistency with requested products and spatial and temporal coverage, limitations due to availability of input data, etc.). Eventual changes or suggested modifications of the initial service request are proposed to the user. If rejected, interaction with the user allows to agree a new activation setup.
2	Data preparation	All data is pre-processed and filtered for relevance. This can include, for example, the geocoding and matching of satellite images acquired for different time periods.
3	Run of algorithm	Calculation of each (sub-)indicator. For example, and automatic Change Detection Algorithm (ACD) could be run to identify changes over a time series of satellite images for an Area of Interest.
4	Data fusion	Results of different (sub-)indicators are combined into a single measure, where necessary.
5	Quality and consistency check	An automated quality and sanity check of the generated results. Products with not enough quality and/or consistence are discarded.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
6	Product packaging & Indicator delivery	Results delivered to the requesting service. This phase includes the preparation of the product package to be delivered to the end user through the Service Data Lake and to be made available to the user through the Exploitation System.

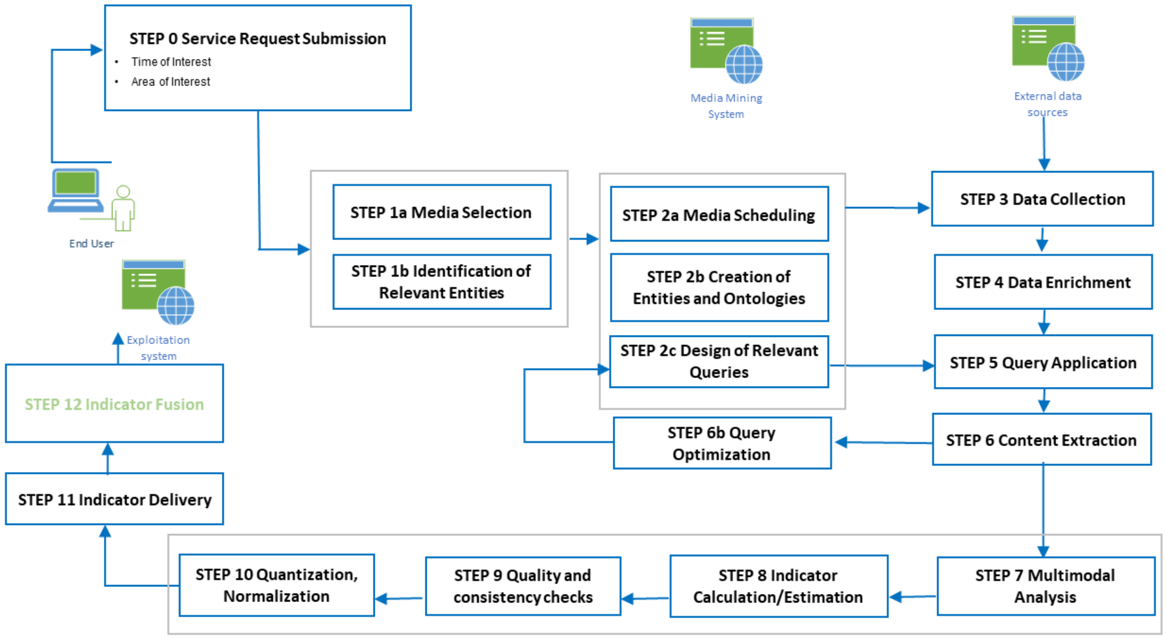


Figure 19: Workflow schema for (sub)indicators based on social/traditional media data.

Figure 19 presents a schematic overview for the generation of (sub)indicators based on data from traditional and social media. The scheme itself presents the generic approach which will be parameterized for the generation of all media-based indicators (identical mechanism but different parameters for each one).

- The workflow comprises initial steps to lay the foundations in terms of data and contents. Contents are then collected continuously, enriched, filtered and used for indicator calculation. In cases where data is already available (e.g. for cold-cases) steps 3 and 4 are not required but existing contents are used instead.
- Indicators are generated based on contents and their characteristics (such as volume, tone, etc).
- In order to allow fusion with other indicators, normalization and standardization steps are foreseen. Furthermore, values may be provided in discrete intervals to allow combination with further indicators. Step 12 will address fusion explicitly. The exact mechanisms will be developed as part of T2.4.

For the CENTAUR project, the following indicators will be considered with regards to socio-economic impact and vulnerability:

Socio-economic impact and vulnerability to urban floods

UF-ID-8:_Robustness and quality of the built environment

UF-ID-9:_Assets and financial resources

UF-ID-10: Public services and government support

UF-ID-11: Social networks and community support

Funded by
the European Union
This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

UF-ID-12: Timely access to information

UF-ID-13: Ability to evacuate

UF-ID-14: Economic impact of floods

Fine-scale population distribution and exposure to security risks and resources shortage

WFS-ID-7: IDP camps status indicator

WFS-ID-8: Populations at risk of food insecurity

WFS-ID-9: Populations at risk of water insecurity

WFS-ID-10: Number of people living in conflict-affected areas

Demographic and socio-economic stress, vulnerability

WFS-ID-11: Food security

WFS-ID-12: Economic security

WFS-ID-13: Displaced persons

WFS-ID-14: Violent conflict

WFS-ID-15: Radicalisation and polarisation

WFS-ID-16: Disruptions in food supply chains

WFS-ID-17: Humanitarian aid

WFS-ID-18: Resource capture

WFS-ID-19: Climate sensitivity of agri-food systems

WFS-ID-20: Obstacles to mobility

WFS-ID-21: Public services and infrastructures

WFS-ID-22: Strength of armed groups

WFS-ID-23: State-citizen relations

WFS-ID-24: Dispute resolution mechanisms

WFS-ID-25: Social cohesion and trust

Socio-economic impact and vulnerability to urban floods

3.4.1 UF-ID-8: Robustness and quality of the built environment

Indicator discarded. A decision was made by the partners to prioritise quality over quantity and focus available resources on more in-depth work on a limited number of indicator with a) high relevance in view of user needs and developed applications in CENTAUR b) availability of ancillary input data (to limit the number of socio-economic indicators entirely based on social/traditional media sources, which have important caveats) and c) comparably lower cost/effort of acquisition and/or processing of input data. That said, the partners keep the option to still include those indicators at a later stage, depending on available resources and time.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3.4.2 UF-ID-9: Assets and financial resources

Description

Indicator for the distribution of wealth/financial capacity across neighbourhoods. Based on information on household density, proximity to amenities and infrastructure, as well as social/traditional media information disparities in financial means, poverty, unemployment etc.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media.
- WSF-3D.
- WSF-Population.
- GHS-Built-V R2023A.
- GHS-Pop R2023A.
- WFS – Imperviousness.
- Travel time and distance to infrastructure and services.
- Restaurant prices.

Below, in Table 45 the details of the input data used for the indicator calculation.

Table 45: Input data required forUF-ID-9

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
WSF-3D	The dataset provides quantification of the building height, building area, building fraction, and building volume, within the built-up environment as described by the WSF2019. The units of the dataset are in dm, m, percent, and cubic meters, respectively.	GeoTIFF
WSF-Population	Residential population estimates derived from CIESIN GPW4.11 are disaggregated from census or administrative units to grid cells, informed by the distribution, height and functional use of settlements derived from the WSF-datasets.	GeoTIFF
GHS-Built-V R2023A	The dataset depicts the distribution of the built-up volumes, express as the number of cubic meters.	GeoTIFF
GHS-Pop R2023A	Residential population estimates between 1975 and 2020 in 5-year intervals and projections to 2025 and 2030 derived from CIESIN GPWv4.11 were disaggregated from census or administrative units to grid cells, informed by the distribution, volume, and classification of built-up as mapped in the Global Human Settlement Layer (GHSL) global layer per corresponding epoch.	GeoTIFF
WSF - Imperviousness	The dataset depicts the percent of impervious surface within the 10m pixels within the identified built-up pixels of the WSF layer. The PIS is	GeoTIFF

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	derived as the inverse of the NDVI, describing the "level of greenness" within the settlement pixels	
Travel time and distance to infrastructure and services	Proximity to high through roads and railways as proxy for +/- wealthy neighbourhoods (UF-ID-9), using OSM (for outside Europe) or national data (e.g. IGN for France).	GeoTIFF
Restaurant prices	Raster layer based on interpolated restaurant prices from google maps as proxy for wealth disparities between neighbourhoods.	GeoTIFF

Output

The Table 46 below summarizes the main characteristics of the indicator to be produced.

Table 46: Output data description and specifications for UF-ID-9

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Assets and financial resources	Indicator for the distribution of wealth/financial capacity across neighbourhoods	GeoTIFF	10-100m; according to update frequency of input data	Generating measures of socio-economic vulnerability to urban floods with broader coverage (including in developing countries) and greater spatial precision at comparably lower costs (i.e. bypassing tedious and expensive data collection techniques like surveys).

3.4.3 UF-ID-10: Public services and government support

Description

Indicator of general access to public services and help in an area (by government and civil society actors) as proxy for protection against floods (risk reduction/prevention) and support in times of floods (response).

Workflow

As outlined at the beginning of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media
- Travel time and distance to infrastructure and services

Below, in Table 47 the details of the input data used for the indicator calculation.

Table 47: Input data required for UF-ID-10

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images,

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
		video, audio
Travel time and distance to infrastructure and services	Proximity to essential services, e.g. distance to hospitals, fire stations etc. and density of transport networks such as roads and railways can be used as proxies for government efforts in investing in urban planning, using OSM (for outside Europe) or national data (e.g. IGN for France).	GeoTIFF

Output

The Table 48 below summarizes the main characteristics of the indicator to be produced.

Table 48: Output data description and specifications for UF-ID-10

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Public services and government support	Indicator of general access to public services and help in an area (by government and civil society actors) as proxy for protection against floods (risk reduction/prevention) and support in times of floods (response)	GeoTIFF	10-100m; according to update frequency of input data	Generating measures of socio-economic vulnerability to urban floods with broader coverage (including in developing countries) and greater spatial precision at comparably lower costs (i.e. bypassing tedious and expensive data collection techniques like surveys).

3.4.4 UF-ID-11: Social networks and community support

Indicator discarded. A decision was made by the partners to prioritise quality over quantity and focus available resources on more in-depth work on a limited number of indicator with a) high relevance in view of user needs and developed applications in CENTAUR b) availability of ancillary input data (to limit the number of socio-economic indicators entirely based on social/traditional media sources, which have important caveats) and c) comparably lower cost/effort of acquisition and/or processing of input data. That said, the partners keep the option to still include those indicators at a later stage, depending on available resources and time.

3.4.5 UF-ID-12: Timely access to information

Indicator discarded. A decision was made by the partners to prioritise quality over quantity and focus available resources on more in-depth work on a limited number of indicator with a) high relevance in view of user needs and developed applications in CENTAUR b) availability of ancillary input data (to limit the number of socio-economic indicators entirely based on social/traditional media sources, which have important caveats) and c) comparably lower cost/effort of acquisition and/or processing of input data. That said, the partners keep the option to still include those indicators at a later stage, depending on available resources and time.

3.4.6 UF-ID-13: Ability to evacuate

Description

Indicator that measures people’s abilities to move out of harm's way, in terms of available infrastructure, physical conditions, and social context (e.g. presence of vulnerable and dependent groups such as children, elderly, and disabled persons).

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Workflow

As outlined at the beginning of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media.
- Travel time and distance to infrastructure and services.
- WSF-Population.
- GHS-Pop R2023A.

Below, in Table 49 the details of the input data used for the indicator calculation.

Table 49: Input data required for UF-ID-13

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
Travel time and distance to infrastructure and services	Access to transport networks (e.g. highways), which in combination with population density and elevation can be used as proxy to estimate ability to evacuate, using OSM (for outside Europe) or national data (e.g. IGN for France).	GeoTIFF
WSF-Population	Residential population estimates derived from CIESIN GPW4.11 are disaggregated from census or administrative units to grid cells, informed by the distribution, height and functional use of settlements derived from the WSF-datasets.	GeoTIFF
GHS-Pop R2023A	Residential population estimates between 1975 and 2020 in 5-year intervals and projections to 2025 and 2030 derived from CIESIN GPWv4.11 were disaggregated from census or administrative units to grid cells, informed by the distribution, volume, and classification of built-up as mapped in the Global Human Settlement Layer (GHSL) global layer per corresponding epoch.	GeoTIFF

Output

The Table 50 below summarizes the main characteristics of the indicator to be produced.

Table 50: Output data description and specifications for UF-ID-13

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Ability to evacuate	Indicator that measures people's abilities to move out of harm's way, in terms of available infrastructure, physical conditions, and social context (e.g. presence of vulnerable and dependent groups such as children, elderly, and disabled	GeoTIFF	10-100m; according to update frequency of input data	Generating measures of socio-economic vulnerability to urban floods with broader coverage (including in developing countries) and greater spatial precision at comparably lower costs (i.e. bypassing tedious and

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
	persons) and support in times of floods (response).			expensive data collection techniques like surveys).

3.4.7 UF-ID-14: Economic impact of floods

Description

Indicators of estimated economic damage of floods in urban areas.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media.
- UF-ID-3.
- UF-ID-5.

Below, in Table 51 the details of the input data used for the indicator calculation.

Table 51: Input data required for UF-ID-14

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
UF-ID-3	High-Resolution urban flood risk maps for various return periods. Indicator that clearly shows the flood extent and depth for different return period scenarios.	Raster (tif)
UF-ID-5	Urban flooding map based on geomorphological and InSAR approach for an enhanced damage assessment. Indicator that clearly shows the flood damage assessment based on flooding depth.	Raster (tif)

Output

The Table 52 below summarizes the main characteristics of the indicator to be produced.

Table 52: Output data description and specifications for UF-ID-14

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Economic impact of floods	Indicators of estimated economic damage of floods in urban areas.	GeoTIFF	10-100m; according to update frequency of input data	New data on economic impact of floods ibn urban areas with higher resolution for both, post-event assessment and pre-event risk analysis

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Fine-scale population distribution and exposure to security risks and resources shortage

3.4.8 WFS-ID-7: IDP camps status indicator

Description

The analysis of the development of camps can be used as a connection with other indicators related to food insecurity events, conflicts, or political disorders. Indeed, it gives an overview about the status of the IDP camps, highlighting a temporal change within a certain period of time. Relying on IOM, UNHCR and other sources the precise location of refugee camps can be obtained. With respect to what was included in the GRANT it will be implemented just the section about accumulation points monitoring and not people movement as the workflow for automatic people movement detection wouldn't have been developed in time for the CENTAUR project.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18. In particular, in Table 53 below are reported the input data used for the indicator calculation.

Table 53: Description of the steps part of the workflow for WFS-ID-7

Step #	Name	Description
0	**Service Request Submission**	The user activates the indicator providing, through the specific Dashboard, information like: • Time of Interest • Area of Interest
1a	**Activation setup**	Depending on available data and processing options set up by the end user (frequency, AOI, etc.), processing is activated. Based on UNHCR, IOM and other sources of information, the location of interest at local scale are derived.
1b	**Changes confirmation**	This step performs additional checks on End User request, e.g. consistency with requested products and spatial and temporal coverage, limitations due to availability of input data, etc. Eventual changes or suggested modifications of the initial service request are proposed to the user. If rejected, interaction with the user allows to agree a new activation setup.
3	**Data preparation**	According to the user demand, data covering the location of interest are requested through the project platform. In order to realize a change detection product to detect changes in the radiometric amplitude response of the same area at two different times, a first stage of geocoding and co-registration is needed to perfectly match the two images acquired (pre and post event of interest).
4	**Run algorithm**	Firstly, a RGB composite product is created, in which the red channel is the amplitude of the master image and green and blue channels are the amplitudes of slave images. Then, an Automatic Change Detection algorithm (ACD) is applied to detect the changes occurred between pre and post event collections according to a threshold working on the colour space of the RGB false colour composite stacked product generated. The amplitude change detection map is generated as a raster highlighting in red the points where changes, e.g. different features, in the second image are present with respect to the first reference image, and in green the points that had a stronger response in the first image.

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Step #	Name	Description
3	Quality check & consistency check	A consistency check of the generated products is visually done. The quality check consists of operator-based steps to verify the quality level of the generated products. Products with not enough quality and/or consistence are discarded.
4	Product Packaging & Delivery	This phase includes the preparation of the product package to be delivered to the end user through the Service Data Lake and to be made available to the user through the Exploitation System.

Input data

The input data used for the indicator generation derive from:

- IOM Displacement Tracking Matrix (DTM) Report.
- UNHCR CCCM (Camp Coordination and Camp Management).
- SAR imagery from ESA DataWarehouse (DWH) Contributing Missions.

Below, in Table 54 the details of the input data used for the indicator calculation.

Table 54: Input data required for WFS-ID-7

Input data Name	Description	Format
IOM DTM	The Displacement Tracking Matrix (DTM) is a system for collecting and analysing data about the mobility, vulnerabilities, and needs of displaced and mobile populations. Data may be collected at displacement locations or locations with high human mobility. Examples of these locations include internal displacement camps, informal settlement, and communities hosting displaced people, among others.	Report
UNHCR CCCM	To respond to the growing displacement and protection concerns of IDPs in Somalia, the CCCM cluster was activated in May 2017. The aim of the Cluster was to improve the living conditions and protection of IDPs in sites and settlements and ensure equitable access to services and the assistance of all persons in need.	Report
HR-VHR SAR EO	High resolution SAR imagery from ESA DWH Contributing Missions covering the areas of interest	Raster

Output

The output of change detection in the form of a raster map is generated highlighting the points where changes, e.g. different features, in the second image are present with respect to the first reference image (Table 55).

Table 55: Output data description and specifications for WFS-ID-7

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
IDP camps status indicator	The output of change detection in the form of a raster map is generated highlighting the points where changes, e.g. different features, in the second image	Raster / vectorial	The spatial resolution is constrained by the input spatial resolution; depends	Usually, refugee camp monitoring is pursued with optical images and not high-resolution SAR data linked with information of different nature, like socio-

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
	are present with respect to the first reference image.		on data temporal availability	economic or political events.

3.4.9 WFS-ID-8: Populations at risk of food insecurity

Description

Compound-indicator that estimates the total number (proportion or percentage) of people at risk of food insecurity derived from level or status of food security within a given area.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18.

Input data

The input data used for the indicator generation are derived from:

- WFS-Population.
- GHS-Pop-R2023A.
- WFS-ID-11.

Below, in Table 56 the details of the input data used for the indicator calculation.

Table 56: Input data needed for WFS-ID-18

Input data Name	Description	Format
WFS-Population data	Residential population estimates derived from CIESIN GPW4.11 are disaggregated from census or administrative units to grid cells, informed by the distribution, height and functional use of settlements derived from the WSF-datasets. Each dataset has a spatial resolution of 10m and temporal coverage of 2016-to present.	Raster layer
GHS-Pop-R2023A	Residential population estimates between 1975 and 2020 in 5-year intervals and projections to 2025 and 2030 derived from CIESIN GPWv4.11 were disaggregated from census or administrative units to grid cells, informed by the distribution, volume, and classification of built-up as mapped in the Global Human Settlement Layer (GHSL) global layer per corresponding epoch. Each dataset has a resolution of 100m	Raster Layer
WFS-ID-11	Level of food insecurity of the population in a given area. Can contain several components (e.g. available/produced food, food prices, ease of access, etc.).	Shapefile dataset at the admin1 or admin2 level

Output

The Table 57 below summarizes the main characteristics of the indicator to be produced.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Table 57: Output data description and related specifications for WFS-ID-8

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Populations at risk of food insecurity	This indicator will show the estimate of people living with different levels of in food insecurity (e.g. minimal, stress, crisis, emergency or famine) Results are aggregated to the administrative unit 1 or 2 level following the format of the food security data provided by indicator WFS-ID-11.	Shapefile or tabular	Monthly or quarterly. Following the update frequency of indicator WSF-ID-11	Since gridded population datasets used for design/implementation of this indicator are modelled using proxies that integrate height and settlement use information, the accuracy of population estimates at risk of food (in)security can be expected to be more accurate than previously reported. Furthermore, the improved spatial resolution of the WSF-Population guarantees a better integration with other data sources of fine-spatial resolution (e.g. AOI as small as 10m x 10m).

3.4.10 WFS-ID-9: Populations at risk of water insecurity

Description

Compound-indicator that measures the potentially available physical water per person, allowing to derive the degree of water (in) sufficiency, and estimate the total number (proportion or percentage) of people at risk of water insecurity.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18.

Input data

The input data used for the indicator generation are derived from:

- WSF-Population data.
- ERA5-Land monthly average run-off.

Below, in Table 58 the details of the input data used for the indicator calculation.

Table 58: Input data required for WFS-ID-9

Input data Name	Description	Format
WSF-Population data	Residential population estimates derived from CIESIN GPW4.11 are disaggregated from census or administrative units to grid cells, informed by the distribution, height and functional use of settlements derived from the WSF-datasets. Each dataset has a spatial resolution of 10m and temporal coverage of 2016-to present.	GeoTIFF
ERA5-Land monthly average run-off	This parameter is the total amount of water accumulated over a particular time which depends on the data extracted. The units of run-off are depth in meters. The data is pre-processed for the calculation of m3day-1	Raster Layer

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output

The Table 59 below summarizes the main characteristics of the indicator to be produced.

Table 59: Output data description and specifications for WFS-ID-9

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Populations at risk of water insecurity	The output of this indicator is two-fold: • Annual or monthly, static maps showing the water availability per person. • Annual or monthly static maps depicting water stress levels based on the Falkenmark indicator. • Statistics on the estimated number of people living in water stress for a given AOI.	Raster, Shapefile and tabular	Annual, Monthly	Since gridded population datasets used for design/implementation of this indicator are modelled using proxies that integrate height and settlement use information, the accuracy of population estimates at risk of food (in)security can be expected to be more accurate than previously reported. Furthermore, the improved spatial resolution of the WSF-Population guaranties a more consistent

3.4.11 WFS-ID-10: Number of people living in conflict-affected areas

Description

Compound-indicator that estimates the total number (proportion or percentage) of people living in (or in the proximity of) conflict-affected areas.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18.

Input data

The input data used for the indicator generation are derived from:

- WFS-Population data.
- GHS-Pop-R2023.
- WFS-ID-14.

Below, in Table 60 the details of the input data used for the indicator calculation.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Table 60: Input data required for WFS-ID-10

Input data Name	Description	Format
WFS-Population data	Residential population estimates derived from CIESIN GPW4.11 are disaggregated from census or administrative units to grid cells, informed by the distribution, height and functional use of settlements derived from the WSF-datasets. Each dataset has a spatial resolution of 10m and temporal coverage of 2016-to present.	Raster layer
GHS-Pop-R2023	Residential population estimates between 1975 and 2020 in 5-year intervals and projections to 2025 and 2030 derived from CIESIN GPWv4.11 were disaggregated from census or administrative units to grid cells, informed by the distribution, volume, and classification of built-up as mapped in the Global Human Settlement Layer (GHSL) global layer per corresponding epoch. Each dataset has a resolution of 100m.	Raster Layer
WFS-ID-14	Number/frequency of violent conflict events between 1990 and 2023 derived from ACLED. Point data on conflict events is aggregated to the admin1 or 2 level.	Shapefile dataset

Output

The Table 61 below summarizes the main characteristics of the indicator to be produced.

Table 61: Output data description and specifications for WFS-ID-10

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Number of people living in conflict-affected areas	This indicator will provide an estimate of the no. of people living in the proximity of conflict events. The results can be aggregated at the administrative level.	Shapefile or tabular	Monthly or quarterly. Following the update frequency of indicator WSF-ID-14	Since gridded population datasets used for design/implementation of this indicator are modelled using proxies that integrate height and settlement use information, the accuracy of population estimates at risk of food (in)security can be expected to be more accurate than previously reported. Furthermore, the improved spatial resolution of the WSF-Population guarantees a better integration with other data sources of fine-spatial resolution (e.g. AOI as small as 10m x 10m).

Demographic and socio-economic stress, vulnerability

3.4.12 WFS-ID-11: Food security

Description

Scale for level of food (in)security experienced by the population in a given area. Can contain several components (e.g. available/produced food, food prices, ease of access, etc.). Disaggregating by groups (e.g. IDPs, vulnerable groups) would be useful.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data

The input data used for the indicator generation are derived from:

- Social-/traditional media
- Afrobarometer.
- FAO DIEM (Data in Emergencies Monitoring).
- FEWS NET.

Below, in Table 62 the details of the input data used for the indicator calculation.

Table 62: Input data required for WFS-ID-11

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
Afrobarometer	This sub nationally geocoded dataset covers Rounds 1-6 of Afrobarometer's surveys in 37 African countries between 1999 and 2015. It provides, time-varying information about the priorities, preferences, experiences, and opinions of more than 200,000 African citizens in 28,000 localities. More specifically, it provides local data on levels of food insecurity experienced by survey respondents.	sav
FAO DIEM (Data in Emergencies Monitoring)	DIEM survey data is aggregated at the first administrative level and organized into four thematic areas: (1) Incomes, Shocks and Needs; (2) Crop production; (3) Livestock production; (4) Food Security. Specifically, it includes survey data on: Food Insecurity Experience Scale (FIES), Household Dietary Diversity Score (HDDS) and Livelihood-based Coping Strategies Index (LCSI) scores;	xlsx
FEWS NET	Data on acute food insecurity, based on IPC3.0 and considers first level outcomes (food consumption levels, energy intake, livelihood change), second level outcomes (nutritional status, mortality), and contributing factors (food availability, access, utilisation, stability, hazards, vulnerability). More information: https://www.ipcinfo.org/ipcinfo-website/ipc-overview-and-classification-system/ipc-acute-food-insecurity-classification/en/.	Shapefiles, geojson

Output

The Table 63 below summarizes the main characteristics of the indicator to be produced.

Table 63: Output data description and related specifications for WFS-ID-11

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Food security	Scale for level of food (in)security of the population in a given area.	Shapefile (polygon)	Admin area; according to update frequency of input data	Drawing on several data sources for a cross-validated measure of food (in)security and for filling data gaps.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3.4.13 WFS-ID-12: Economic security

Description

Scale for level of economic (in)security experienced by the population in a given area. Can contain several components (e.g. available/produced food, food prices, ease of access, etc.). Disaggregating by groups (e.g. IDPs, vulnerable groups) would be useful.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media
- Afrobarometer
- FAO DIEM (Data in Emergencies Monitoring)
- Livestock heat stress index

Below, in Table 64 the details of the input data used for the indicator calculation.

Table 64: Input data required for WFS-ID-12

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
Afrobarometer	This sub nationally geocoded dataset covers Rounds 1-6 of Afrobarometer's surveys in 37 African countries between 1999 and 2015. It provides, time-varying information about the priorities, preferences, experiences, and opinions of more than 200,000 African citizens in 28,000 localities. More specifically, it provides local data on levels of economic insecurity experienced by survey respondents.	sav
FAO DIEM (Data in Emergencies Monitoring)	DIEM survey data is aggregated at the first administrative level and organized into four thematic areas: (1) Incomes, Shocks and Needs; (2) Crop production; (3) Livestock production; (4) Food Security. Specifically, it includes survey data on: • Income shock, i.e. type/magnitude of income change (agricultural and non-agricultural), shocks that likely affect household capacity to produce food or generate income. • Crop module focuses on difficulties encountered by crop producers and their impacts on production and sales. They may include constrained access to crop inputs and marketing challenges. Production is appraised as relative change in the area planted and harvested (achieved or projected) compared to a recent typical year. • Livestock module focuses on difficulties encountered by livestock breeders and their effects on production and sales. These can include a lack of access to feed and other inputs, or veterinary services, as	xlsx

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	well as marketing issues. Production is measured as relative change in herd or flock size compared to the same time in the previous year.	
Livestock heat stress index	Index for anticipated and historic heat stress by livestock type. Computed as frequency, degree, or risk. This measure can be used to capture economic challenges for herders more precisely.	GeoTIFF

Output

The Table 65 below summarizes the main characteristics of the indicator to be produced.

Table 65: Output data description and specifications for WFS-ID-12

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Economic security	Scale for level of economic (in)security experienced by the population in a given area.	Shapefile (polygon)	Admin area; according to update frequency of input data	Drawing on several data sources for a cross-validated measure of economic (in)security and for filling data gaps.

3.4.14 WFS-ID-13: Displaced persons

Description

Estimate of the number of displaced persons in a given area, including IDP and refugees in neighbouring countries. Disaggregation by mobility type would be useful.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media
- IOM DTM Flow monitoring.
- IOM DTM mobility tracking.

Below, in Table 66 the details of the input data used for the indicator calculation.

Table 66: Input data required for WFS-ID-13

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
IOM DTM Flow monitoring	Quantifies the presence of population categories of interest (IDP, refugees…), reasons for displacement, length of displacement and needs (e.g. food, water, shelter etc.) within defined geographical areas and camp	xlsx

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	locations, with a frequency that captures mobility dynamics (typically monthly).	
IOM DTM Mobility tracking	Quantitative estimates of flows of individuals of interest (IDP, refugees…) moving from location A to location B. Update frequency typically monthly	xlsx

Output

The Table 67 below summarizes the main characteristics of the indicator to be produced.

Table 67: Output data description and specifications for WFS-ID-13

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Displaced persons	Estimate of the number of displaced persons in a given area, including IDP and refugees in neighbouring countries.	Shapefile (polygon)	Admin area; monthly	Facilitating access to compiled and aggregated numbers of persons of interest (IDP, refugees…).

3.4.15 WFS-ID-14: Violent conflict

Description

Indicator of the number/frequency of violent conflict events. Indicator captures historical events, as well as predicted violence up to six months in the future.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media
- Armed Conflict Location & Event Data Project (ACLED).
- ACLED Conflict Alert System (CLAST).

Below, in Table 68 the details of the input data used for the indicator calculation.

Table 68: Input data required for WFS-ID-14

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
ACLED	ACLED collects reported information on the type, agents, location, date, and other characteristics of political violence events, demonstration events, and other select non-violent, politically relevant developments in every country and territory in the world. ACLED focuses on tracking a range of violent and non-violent actions by or affecting political agents, including governments,	xlsx

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	rebels, militias, identity groups, political parties, external forces, rioters, protesters, and civilians.	
ACLED cast	The ACLED Conflict Alert System (CAST) is a new conflict forecasting tool that predicts political violence events up to six months in the future for every country in the world. Updated predictions are released each month for the following six months, alongside accuracy metrics for previous forecasts.	xlsx

Output

The Table 69 below summarizes the main characteristics of the indicator to be produced.

Table 69: Output data description and specifications for WFS-ID-14

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Violent conflict	Indicator of the number/frequency of violent conflict events. Indicator captures historical events, as well as predicted violence up to six months in the future.	Shapefile (polygon)	Admin area; daily	N/A

3.4.16 WFS-ID-15: Radicalisation and polarisation

Description

Indicator that provides an estimate for the level of radicalization, polarization, and animosity in the population that is conductive to periodic outbursts of violence (e.g., riots, clashes) and/or recruitment into armed groups. If necessary, possibility to distinguish sub-classes for this indicator (e.g., anti-government sentiment, ethnic/communal tensions, ideological/religious radicalization) will be considered.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media.
- Afrobarometer.
- ACLED.
- WFS-Population data.

Below, in Table 70 the details of the input data used for the indicator calculation.

Table 70: Input data required for WFS-ID-15

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
Afrobarometer	This sub nationally geocoded dataset covers Rounds 1-6 of Afrobarometer's surveys in 37 African countries between 1999 and 2015. It provides hyperlocal, time-varying information about the priorities, preferences, experiences, and opinions of more than 200,000 African citizens in 28,000 localities. More specifically, it provides local level data on: (1) the most important problems that citizens would like to see their governments address; (2) the perceived performance of local and national institutions in solving specific problems, such as the provision of basic education and health services; (3) access to information via newspapers, radios, televisions, and mobile phones; (4) levels of civic engagement and political participation; (5) engagement with and confidence in local and central government institutions; (6) the physical presence or absence of state institutions (e.g. police stations, health clinics, schools, water and sanitation systems); (7) the reported quality of local public services; and (8) local perceptions of and experiences with various forms of corruption. Examples of questions from codebook: (1) how often without food, without cooking fuel; (2) most important problems in country (incl. food shortage, famine as response); (3) discrimination experience; (4) ethnic group treatment; (5) trust in gov, authorities; (6) gov performance in ensuring everyone has enough to eat; (7) gov performance in keeping prices stable; (8) what type of shelter; (9) what was roof of home or shelter made of; (10) electric connection, water source, toilet.	sav
ACLED	ACLED collects reported information on the type, agents, location, date, and other characteristics of political violence events, demonstration events, and other select non-violent, politically relevant developments in every country and territory in the world. ACLED focuses on tracking a range of violent and non-violent actions by or affecting political agents, including governments, rebels, militias, identity groups, political parties, external forces, rioters, protesters, and civilians.	xlsx
WFS-Population data	Residential population estimates derived from CIESIN GPW4.11 are disaggregated from census or administrative units to grid cells, informed by the distribution, height and functional use of settlements derived from the WSF-datasets. Each dataset will be generated by DLR based on the WSF-Suite, with a spatial resolution of 10m and temporal coverage of 2016-to present.	Raster layer

Output

The Table 71 below summarizes the main characteristics of the indicator to be produced.

Table 71: Output data description and specification for WFS-ID-15

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Radicalisation and	This indicator will provide an estimate for the degree (or level) of radicalisation	Table / csv	Event related and admin	Conflict prediction models usually rely on data on manifest violent events and are hence not able to capture earlier stages of conflict escalation like growing

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
polarisation indicator	and polarisation in the population in a given area.	Shapefile (polygon)	ideally monthly	grievances and increasingly radical attitudes that generally precede and/or enable manifest violent behaviour. However, from an early warning and conflict prevention perspective, this would make sense. Survey data on radical attitudes exist but are incomplete and collected in wide time intervals (e.g., every three years), which makes monitoring difficult.

3.4.17 WFS-ID-16: Disruptions in food supply chains

Research on input data sources showed that information on food supply chains is often already included or accounted for in the input data used to measure food (in)security in CENTAUR. Therefore, a dedicated measure of disruptions in supply chains does not seem necessary anymore. Also, this frees resources for more in-depth work on other innovative indicators that rely on social/traditional media data input.

3.4.18 WFS-ID-17: Humanitarian aid

Description

Indicator of aid provided by national and international actors (e.g. government, international organisations, NGOs) to cushion the effect of extreme climatic conditions.

Workflow

As outlined at the beginning of Section 3.4 in Figure 18.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media
- FAO DIEM (Data in Emergencies Monitoring)

Below, in Table 72 the details of the input data used for the indicator calculation.

Table 72: Input data required for WFS-ID-17

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
FAO DIEM (Data in Emergencies Monitoring)	DIEM survey data is aggregated at the first administrative level and organized into four thematic areas: (1) Incomes, Shocks and Needs; (2) Crop production; (3) Livestock production; (4) Food Security. Specifically, it includes survey data on: • Needs for short- to mid-term assistance in relation to agricultural livelihoods.	xlsx

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	• Information about food security and livelihoods assistance received during the recall period.	

Output

The Table 73 below summarizes the main characteristics of the indicator to be produced.

Table 73: Output data description and specifications for WFS-ID-17

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Humanitarian aid	Indicator of aid provided by national and international actors (e.g. government, international organisations, NGOs) to cushion the effect of extreme climatic conditions.	Shapefile (polygon)	Admin area; 3-4 months period	N/A

3.4.19 WFS-ID-18: Resource capture

Description

Indicator for the appropriation of natural resources essential for food production and/or sustaining agricultural and pastoralist livelihoods by powerful actors (e.g. land grabbing by large companies) or specific groups (e.g. communal groups).

Workflow

As outlined at the start of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media.
- Rangeland land cover change.

Below, in Table 74 the details of the input data used for the indicator calculation.

Table 74: Input data required for WFS-ID-18

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
Rangeland land cover change	Long-term land cover transitions over rangelands.	GeoTIFF

Output

The Table 75 below summarizes the main characteristics of the indicator to be produced.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Table 75: Output data description and specifications for WFS-ID-18

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Resource capture	Indicator for the appropriation of natural resources essential for food production and rural livelihoods	Shapefile (polygon)	Admin area; Yearly	Resource capture reduces access to essential resources for less powerful actors and groups and can aggravate competition over scarce resources. Yet, indicators for the distribution of resources or unequal access to resources are usually not available in quantitative assessments.

3.4.20 WFS-ID-19: Climate sensitivity of agri-food systems

Description

Indicator for the sensitivity of rural livelihoods and food production systems to erratic climatic/weather conditions.

Workflow

As outlined at the start of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media
- FAO WaPOR Land Cover Classification 2021.

Below, in Table 76 the details of the input data used for the indicator calculation.

Table 76: Input data required for WFS-ID-19

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
FAO WaPOR: Land Cover Classification 2021	This land cover dataset at continental scale is based on the Copernicus Global Land cover map. WaPOR data is used to distinguish between irrigated and rainfed areas (and rangeland), assuming that rainfed areas are more vulnerable to the effect of climatic/weather shocks. The data are published on a yearly basis.	GeoTIFF

Output

The Table 77 below summarizes the main characteristics of the indicator to be produced.

Table 77: Output data description and specifications for WFS-ID-19

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Climate sensitivity of	Indicator for the sensitivity of rural livelihoods and food production systems to erratic climatic/weather conditions.	Shapefile (polygon)	Admin area; Yearly	N/A

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
agri-food systems				

3.4.21 WFS-ID-20: Obstacles to mobility

As part of the evolution of the CENTAUR project a decision was made to emphasise conflict over displacement as possible outcome of climatic stress. It has also been decided, for indicators strongly relying on social/traditional media, to prioritise quality over quantity in a first step and limit the number of such indicators to be pursued by the relevant partners. WFS-ID-20 (strongly connected to migration) is thus discarded in favour of freeing resources for more in-depth work on indicators that are more directly relevant from a conflict perspective.

3.4.22 WFS-ID-21: Public services and infrastructures

Description

Indicator that provides an estimate for the degree to which the government effectively and inclusively delivers services that are essential for withstanding extreme climatic conditions.

Workflow

As outlined at the start of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media.
- Afrobarometer.
- EOG Night-time light.
- Main roads.
- HOT (Humanitarian Open Street Map).

Below, in Table 78 the details of the input data used for the indicator calculation.

Table 78: Input data required for WFS-ID-21

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
Afrobarometer	This sub nationally geocoded dataset covers Rounds 1-6 of Afrobarometer's surveys in 37 African countries between 1999 and 2015. It provides time-varying information about the priorities, preferences, experiences, and opinions of more than 200,000 African citizens in 28,000 localities. More specifically, it provides local-level data on: (1) the most important problems that citizens would like to see their governments address; (2) the perceived performance of local and national institutions in solving specific problems, such as the provision of basic services; (3) the physical presence or absence of state institutions & services (e.g. police stations, health clinics, schools, water and sanitation systems); (4) the reported quality of	sav

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Input data Name	Description	Format
	local public services; and (5) local perceptions of and experiences with various forms of corruption.	
EOG Night-time Light	Data on global night-time light (NTL), provided by the Visible and Infrared Imaging Suite (VIIRS) Day Night Band (DNB) on board of the Joint Polar-orbiting Satellite System (JPSS). NTL is frequently used as a proxy measure for local levels of economic development and access to services.	GeoTIFF
Main roads	Main roads compiled from OpenStreetMap (OSM) and other available data sources. Serve as a proxy measure for public service provision; political marginalisation; ease of access (e.g. for government troops).	Raster (e.g. GeoTIFF) or Vector (e.g. geopackage, kml, shp)
HOT (Humanitarian OpenStreetMap)	Humanitarian OpenStreetMap Team (HOT) acts as a bridge between the traditional humanitarian responders and the OpenStreetMap Community. HOT works both remotely and physically in countries to assist the collection of geographic data, usage of that information and training others in OpenStreetMap. For the use cases Mali, Mozambique, Somalia and Sudan, data is available for roads, railways, airports, financial services (e.g. banks, atm, post office), ports, waterways (e.g. rivers, streams, canals), buildings, health facilities (e.g. clinics, hospitals), populated places (e.g. villages, isolated dwellings), education facilities (e.g. schools), points of interest (including previous infrastructure types + miscellaneous e.g. marketplace, camp site, tower, water well).	Vector (available as geopackage, kml, shp)

Output

The Table 79 below summarizes the main characteristics of the indicator to be produced.

Table 79: Output data description and specifications for WFS-ID-21

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Public services and infra-structure	Indicator that provides an estimate for the degree to which the government effectively and inclusively delivers services that are essential for withstanding extreme climatic conditions	shapefile (polygon)	Admin area; Yearly	Indicator commonly mentioned in the qualitative climate-conflict literature but rarely (if at all) included in models and assessments at the subnational level

3.4.23 WFS-ID-22: Strength of armed groups

Research on possible input data sources showed that EO data for WFS-ID-22 (e.g. location of military vehicles and camps) would be very resource-intensive and costly to acquire at a scale necessary for the CENTAUR project; in addition to concerns about limited capacities in the consortium to accommodate for a large number of indicators relying heavily on social/traditional media sources (see also discussion of WFS-ID-20 above). Collecting such data would also have raised considerable ethical and practical questions around the use of such sensitive information. On the other hand, WFS-ID-22 was judged quite important from the perspective of conflict analysis. A decision

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

was made to prioritize other indicators for the time being but with the option of coming back to this indicator, should resources and project schedule allow it.

3.4.24 WFS-ID-23: State-citizen relations

Description

Indicator that provides an estimate for the degree to which citizens trust public officials and feel included in political decision making. High levels of trust and inclusion are expected to motivate people to seek non-violent means of addressing concerns and grievances.

Workflow

As outlined at the start of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media.
- Afrobarometer.

Below, in Table 80 the details of the input data used for the indicator calculation.

Table 80: Input data required for WFS-ID-23

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
Afrobarometer	This sub nationally geocoded dataset covers Rounds 1-6 of Afrobarometer's surveys in 37 African countries between 1999 and 2015. It provides hyperlocal, time-varying information about the priorities, preferences, experiences, and opinions of more than 200,000 African citizens in 28,000 localities. More specifically, it provides local level data on: (1) the most important problems that citizens would like to see their governments address; (2) the perceived performance of local and national institutions in solving specific problems, such as the provision of basic education and health services; (3) access to information via newspapers, radios, televisions, and mobile phones; (4) levels of civic engagement and political participation; (5) engagement with and confidence in local and central government institutions; (6) the physical presence or absence of state institutions (e.g. police stations, health clinics, schools, water and sanitation systems); (7) the reported quality of local public services; and (8) local perceptions of and experiences with various forms of corruption. Examples of questions from codebook: (1) how often without food, without cooking fuel; (2) most important problems in country (incl. food shortage, famine as response); (3) discrimination experience; (4) ethnic group treatment; (5) trust in gov, authorities; (6) gov performance in ensuring everyone has enough to eat; (7) gov performance in keeping prices stable.	sav

Output

The Table 81 below summarizes the main characteristics of the indicator to be produced.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Table 81: Output data description and specifications for WFS-ID-23

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
State-citizens relations indicator	This indicator will provide an estimate for the degree to which citizens trust public officials and feel included in political decision making. High levels of trust and inclusion are expected to motivate people to seek non-violent means of addressing concerns and grievances.	Table/csv	Event related Admin area ideally monthly	State-citizens relations are frequently mentioned in qualitative climate-security literature, but they are rarely included in quantitative assessments. CENTAUR test the usability of the indicator with a large-scale dataset and in the context of a range of use cases.

3.4.25 WFS-ID-24: Dispute resolution mechanisms

Description

Indicator that provides an estimate for the Presence of trusted formal and informal mechanisms (e.g. legal recourse, inter- and intra-community dialogue) to address disputes, e.g. over access to and usage of natural resources.

Workflow

As outlined at the start of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media

Below, in Table 82 the details of the input data used for the indicator calculation.

Table 82: Input data required for WFS-ID-24

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio

Output

The Table 83 below summarizes the main characteristics of the indicator to be produced.

Table 83: Output data description and specifications for WFS-ID-24

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Dispute resolution mechanisms	Indicator that provides an estimate for the Presence of trusted formal and informal mechanisms to address disputes.	shapefile (polygon)	Admin area; Ideally monthly	Frequently mentioned in qualitative climate-security literature but rarely included in models and quantitative assessments, with some exceptions.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

3.4.26 WFS-ID-25: Social cohesion and trust

Description

Indicator that provides an estimate for the degree to which people feel connected and eager to stick together in the face of major challenges. If required in the context of the project, perceived social (in)equality might be considered as a sub-component or a separate indicator.

Workflow

As outlined at the start of Section 3.4 in Figure 18 and Figure 19.

Input data

The input data used for the indicator generation are derived from:

- Social/traditional media.
- Afrobarometer.

Below, in Table 84 the details of the input data used for the indicator calculation.

Table 84: Input data required for WFS-ID-25

Input data Name	Description	Format
Social/traditional media	Traditional and social media relevant for the corresponding use cases and areas of interest.	Text, images, video, audio
Afrobarometer	This sub nationally geocoded dataset covers Rounds 1-6 of Afrobarometer’s surveys in 37 African countries between 1999 and 2015. It provides hyperlocal, time-varying information about the priorities, preferences, experiences, and opinions of more than 200,000 African citizens in 28,000 localities. More specifically, it provides local level data on: (1) the most important problems that citizens would like to see their governments address; (2) the perceived performance of local and national institutions in solving specific problems, such as the provision of basic education and health services; (3) access to information via newspapers, radios, televisions, and mobile phones; (4) levels of civic engagement and political participation; (5) engagement with and confidence in local and central government institutions; (6) the physical presence or absence of state institutions (e.g. police stations, health clinics, schools, water and sanitation systems); (7) the reported quality of local public services; and (8) local perceptions of and experiences with various forms of corruption. Examples of questions from codebook: (1) how often without food, without cooking fuel; (2) most important problems in country (incl. food shortage, famine as response); (3) discrimination experience; (4) ethnic group treatment; (5) trust in gov, authorities; (6) gov performance in ensuring everyone has enough to eat; (7) gov performance in keeping prices stable.	sav

Output

The Table 85 below summarizes the main characteristics of the indicator to be produced.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Table 85: Output data description and specifications for WFS-ID-25

Output data Name	Description	Format	Spatial and temporal resolution	Progress beyond the SOTA
Social cohesion and trust indicator	Indicator that provides an estimate for the degree to which people feel connected and eager to stick together in the face of major challenges. If required in the context of the project, perceived social (in)equality might be considered as a sub-component or a separate indicator.	Table / csv	Event related Admin areas ideally monthly	Social cohesion and trust are frequently mentioned in qualitative climate-security literature. However, to our knowledge, they have not been considered in predictive models for conflict so far.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

4 CONCLUSIONS

On the basis of the outcomes produced by the review of Copernicus EMS operations for Urban Flood detection and monitoring (Task 1.1) and the review of Copernicus SEA for Water & Food Insecurity impacts (Task 1.2), as well as the user requirements analysis provided in [RD03] report, the set of innovative indicators identified were analysed in details from a technical perspective and related design reported in the present document.

Innovative indicators were clustered in the 3 following domains:

- Urban flood: from UF-ID-1 to UF-ID-7.
- Socio-economic and political: from UF-ID-8 to UF-ID-14 and from WFS-ID-7 to WFS-ID-25.
- Water & Food Security: from WFS-ID-1 to WFS-ID-6.

Data processing flow were reported for specific input data, in the domains of geospatial, opensource and meteorological, that requires a customized workflow for their harvesting and generation.

For each of the indicators mentioned above, the design was structured in the following points:

- Indicator description.
- Step-by-step workflow provision, with details of each step.
- List of input data, related description, and format.
- Output description, related technical specifications (i.e. spatial and temporal resolution, format, etc) and progress beyond the SOTA [RD03].

In the Urban Flood domain, all the 7 innovative indicators identified in T1.1 will be developed in the next activities (Task 2.5), as well as in the Water and Food Security domain all the 6 initial innovative indicators identified in T1.2 will be developed in the next activities (Task 2.6).

In the Socio-economic and political domain, among the initial set of 25 innovative indicators identified in T1.1 and T1.2, the following 3 were discarded in agreement with the consortium, to prioritise quality over quantity, high relevance for users and CENTAUR application, the availability of ancillary input data and the lower cost/effort of acquisition and/or processing of input data.

- UF-ID-8 - Robustness and quality of the built environment.
- UF-ID-11 - Social networks and community support.
- UF-ID-12 - Timely access to information.

In addition, the following 3 innovative indicators were discarded due to specific reasons mentioned below:

- WFS-ID-16 - Disruptions in food supply chains: research on input data showed that the information on food supply chains is often already accounted in the input data used for measuring food (in)security in CENTAUR, therefore a dedicated indicator does not seem necessary anymore.
- WFS-ID-20 – Obstacles to mobility: indicator discarded in favour of freeing resources for more in-depth work on indicators that are more directly relevant from a conflict perspective, as it was decided to emphasise conflict over displacement as possible outcome of climatic stress.
- WFS-ID-22 – Obstacles to mobility: indicator discarded for the time being, with the option of developing in a second phase if project resources will allow it, due to the very resource-intensive and costly input data to acquire at a scale necessary for the CENTAUR project.

The design and implementation of the service pipelines for the generation of complex Urban Flood, Water & Food Insecurity and Socio-economic/political indicators, integrating dataset for supporting monitoring services, crisis situation and impact assessment, will occur in Task 2.4, 2.5 ad 2.6 in the upcoming months (M10-M31).

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR