ECMWF COPERNICUS | REPORT

Copernicus
Europe's eyes on Earth

Copernicus Climate Change Service

Urban SIS
D 7.1 Storage of Urban SIS data

Issued by: Swedish Meteorological and Hydrological Institute

Date: 29/04/2016

REF.: C3S_441 Lot3 Urban SIS D7.1 Storage of Urban SIS data

European Commission

Copernicus
Europe's eyes on Earth

IMPLEMENTED BY
ECMWF

Urban SIS
D 7.1 Storage of Urban SIS data

Edited by: SMHI (Lena Strömbäck, WP leader)
SMHI (Per Lewau)

Date: 29/04/2016
SIS REF.: C3S_441 Lot3 Urban SIS D7.1 Storage of Urban SIS data

Contents:

Contents: 3
Introduction 2
General overview 3
Content and look and feel 3
Monthly statistics 4
Requirements and architecture for the data portal 5
Data properties 5
Data representation 6
Functional requirements 7
Technical architecture 8
Current implementation 9
Plans for further development 12
Contact persons 12
New functionality 12
Time plan 13
Conclusions 14
Appendix 1: Sample Google analytics report 15
Appendix 2: Sample NetCDF metadata for percipitation 17
Appendix 3: Example of meta-data from D5.4.1 19

Introduction

This report describes the technical architecture of the Urban SIS web pages and data portal to be used making Urban SIS data available for the end users. We have made a technical design of the storage solution and a first version has been implemented and is available at http://urbansis.climate.copernicus.eu. This site contains a few preliminary datasets that can be downloaded to test the functionality of the portal.

The report starts with a general overview of the web solution for Urban SIS including the web based information pages, and the portal to be used for the data storage and visualization. After this it gives more details about the technical solution for the data portal and motivates choices made. We conclude by defining a plan for continued work towards the final technical solution.

General overview

This section gives a general overview of the Urban SIS project web site together with some general decisions made for the design. Technically the architecture consists of two parts, the general project information and the data portal. The focus for the general information pages is to describe the proof-of-concept project, its objectives, use-cases and partners. An important aim is that the information should be easily editable during the progress of the project and for that purpose we use the Wordpress (https://wordpress.org/) content management system for this part. The data portal will contain the data storage and functionality for visualisation and data download. The produced data will be represented and stored in NetCDF format. We have chosen to base the data portal on the THREDDs (http://www.unidata.ucar.edu/software/thredds/current/tds/) data server. Table 1 gives a general overview of the solution.

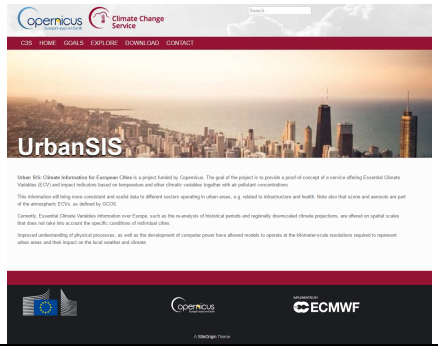
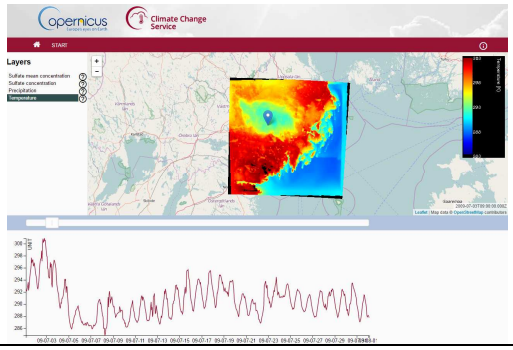
	Information pages	**Data portal**
Start page		
Content	Project information pages. Main entrance to the project pages which contain links to the data portal. Can be extended with blogs and other user community information when needed.	Visualisation and download of the Urban SIS data. Data is stored as NetCDF following convention CF-1.6.
Technical solution	Implemented using the Wordpress content management system.	The data portal is implemented with a server side based on THREADDS Data Server and a web client for visualization and download based on HTML and Java Script

Table 1: Overview of the Urban SIS web pages

Content and look and feel

As requested, the look and feel of the web pages follows the style of the Copernicus Climate site (http://climate.copernicus.eu/). As an example of the resulting design we show the first page of the Urban SIS pages in figure 2. Currently the web pages contain a general presentation of Urban SIS which will later be complemented with more detailed information about the use cases and instruction for how to use the site.

In addition the home page contains links to the portal. One is in the EXPLORE menu item which allows visualisation of a few test data sets and the second in the DOWNLOAD menu item where test data can be downloaded.

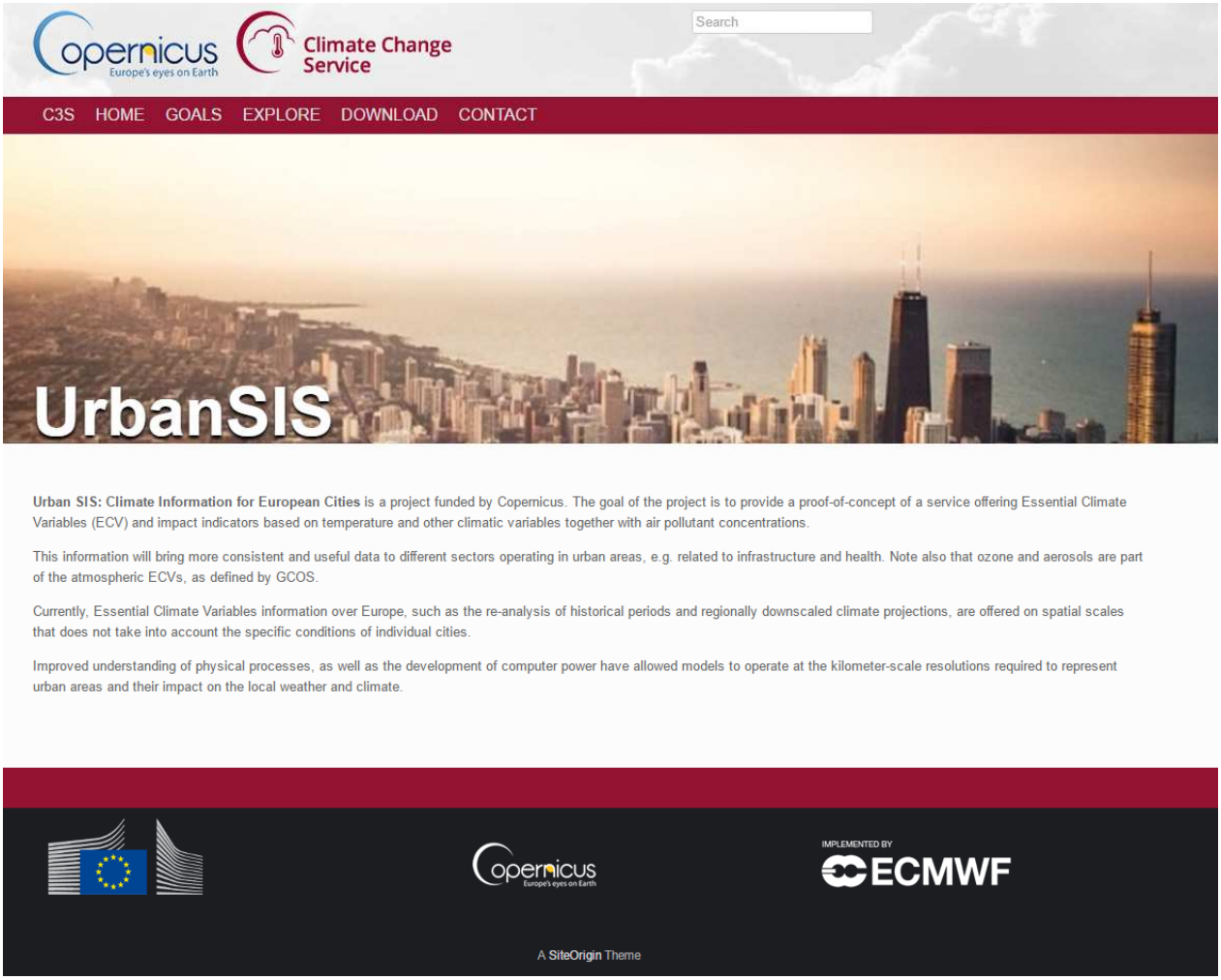


Figure 1: Design of the Urban SIS main page

Monthly statistics

The web pages have been set up so that it produces statistics over visitors via Google analytics. A sample report is shown in Appendix 1. Currently the report only covers visitors to the main Urban SIS site and do not include data and document downloads. This will be added when the final Urban SIS data are delivered to the portal from the other work packages.

Requirements and architecture for the data portal

In this section we focus on the data portal and the description of the data sets we plan for the future solution, as well as the main functional requirements. We also describe the technical architecture of the portal.

Data properties

The work in WP4 (Deliverable 4.1 and Deliverable 4.2) has defined Essential Climate Variables (ECVs) and a preliminary list of Impact Indicators which will be made available via the data portal according to the scheme in table 2. The final Impact Indicators will be defined in Deliverable 4.3, to be published in the end of October 2016.

	Essential Climate Variables	**Impact Indicators**
Model areas	3 pilot cities, the architecture should be able to handle all major cities in Europe.	
Identified variables	Around 23	Around 44 indicators identified
Domain size	Defined for each city; typical size 200x200 km^2	
Spatial resolution	1x1 km^2	
Coordinate system	Selected for each city	
Time frame	Three periods of interest identified where a 5-10 year time frame will be selected; Historical (Early 2000's); Present; and Future (around 2050)	
Time resolution	Hourly time series (for precipitation quarterly)	Varied from one value per period to time aggregated time series.
Estimated size per variable	~50 GB	Ranging from 1 MB for single value indicators to the size of ECVs for time series.
Estimated size per city	~1 TB	Ranging 50 MB to around 1 TB dependent of type and number of Impact Indicators.

Table 2: Properties of data that will be available in Urban SIS. For Impact indicators we have estimated the size

During the project we will work with data for the three defined pilot areas. However, the technical solution should be extendible to a future full scale system covering all major cities in Europe. The work in WP4 has currently defined 23 essential climate variables and 44 impact indicators of interest. We expect the number of ECVs to be

relatively stable, but due to different user needs, the number of impact indicator may increase.

For each variable we will make data available for a historical, present and future period. For each period we will produce data for a five to ten year time frame. For ECVs hourly time series will be made available for the selected five to ten year time frames.

At the end of the table we have made an estimation of the total size of data per variable and city. For these calculations we have assumed single precision storage (4 byte) for the indicator variable data values and double precision (8 byte) for the essential climate variable data values. The total size of Impact Indicators for each city is highly dependent on final number and type of indicators to be defined in D4.3 and may vary between cities.

Data representation

Each ECV or impact indicator will be stored as NetCDF files. The projection will be defined for each city, so that all data for the same area follows this projection. Note that the selected projection may differ between cities as many countries have a preferred projection that they usually work with. Also, the projections will not be the same as is used for the HARMONIE simulations, which means that data will be reprojected before being added to the portal. This will cause minor loss of information, but is considered an important requirement from the end-users.

Basic meta-data will be included in the NetCDF files and represented using the convention CF-1.6. (http://cfconventions.org/Data/cf-conventions/cf-conventions-1.6/build/cf-conventions.html) This defines the most essential information for each dataset. An example of how the meta-data for describing the geographic resolution has been used for one of the sample data files available for the Stockholm area is shown in figure 2. The example uses the national Swedish coordinate system SWEREF99 for the grid. In Appendix 2 we give the full meta-data description currently used for the dataset, including some global attributes to exemplify how NetCDF can be used for general meta-data information about the data.

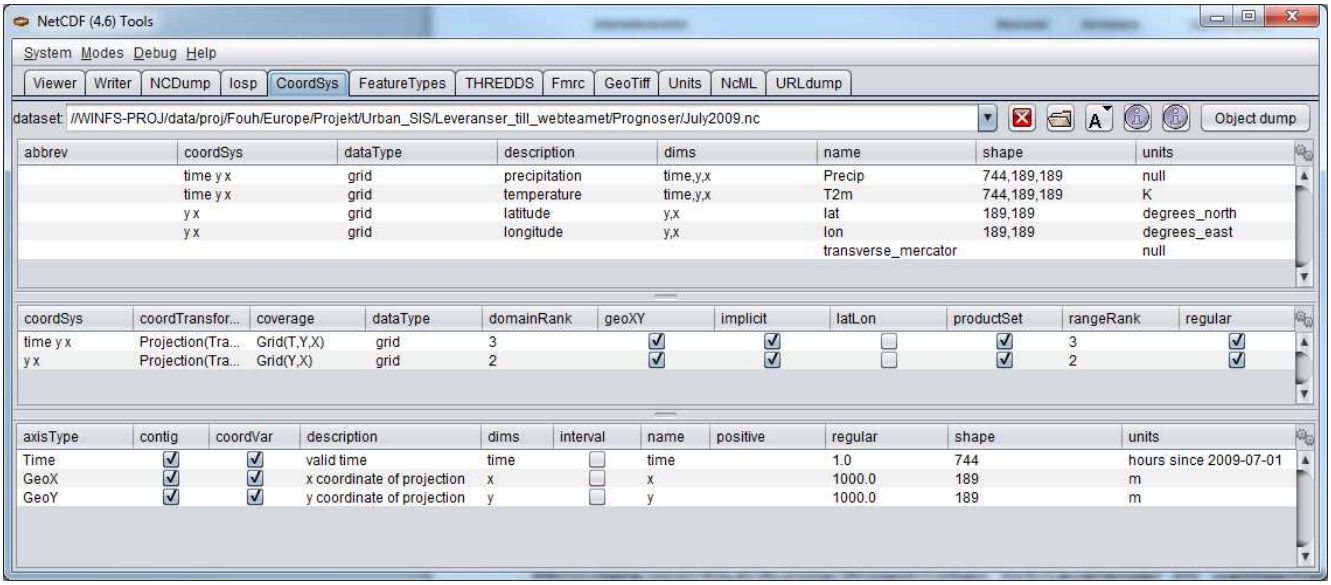


Figure 2: NetCDF basic meta data for describing the geographic resolution for one example data file containing meteorological data (precipitation and temperature) over Stockholm. For this presentation we use the tool available at ftp://ftp.unidata.ucar.edu/pub/netcdf-java/v4.6/toolsUI-4.6.jar, which allows easy inspection of the meta-data of a NetCDF file.

In addition to this basic meta-data, it is crucial to provide additional information to support the use of the datasets. This includes information needed for compatibility with GDAL (Geospatial Data Abstraction Library) and ESRI© ArcGIS when it comes to spatial referencing system, estimating the quality of the data as well as recommendations and instructions for further use of downloaded data. The baseline for provided meta-data will be the definition made when defining the KPIs in deliverable 5.4.1 (see example in appendix 3 of the present report). In case additional meta-data need are defined by the user studies planned for WP6 these will be added. It is likely that this information to be too extensive to be incorporated fully in the NetCDF file and alternatives will be investigated when needed. We expect the full dataset to consist of the NetCDF data file, meta-data information and links to the portal with information about the data.

Functional requirements

The table below gives an overview of the planned functionality of the Urban SIS Data portal. We define:

Visualisation	**Basic:** Visualisation of the data on a map. Visualisation of time series for a selected area. **Extended:** Comparison of data from several datasets. Visualisation of events, e.g. as an animation.
Download	**Basic:** Select a ECV or impact indicator for download. Specification of a subarea for download of data.

	Specification of time frame for download of data. **Extended:** The downloaded data should be an easy to use package containing data, meta-data and instructions on how to use the data.
User created indicators	**Extended:** Possibility to create new indicators or design rains based on modified threshold, combination of available indicators or ECVs, or combination of indicators or ECVs and user provided data.

Table 3: Functional requirements

The functionality will be further defined in WP4 and in the user studies in WP 6. Based on these studies WP4, WP6 and WP7 will together define the functional requirements of the final interface.

Technical architecture

An overview of the technical architecture is shown in figure 3. The main parts of the architecture consist of a platform providing data access and an application side providing the user interface and functionality. The application side consist of a web client and additional software that might be needed for extended functionality, such as data transformations. The platform side provides the basic for data storage and exposes a public API which makes it possible for other interesting partners to build their own applications on the available Urban SIS data. The dark blue parts in figure 3 is the baseline for the architecture whereas the light blue parts will be needed for extended functionality and will be further detailed out based on user studies in cooperation with WP4 and WP6.

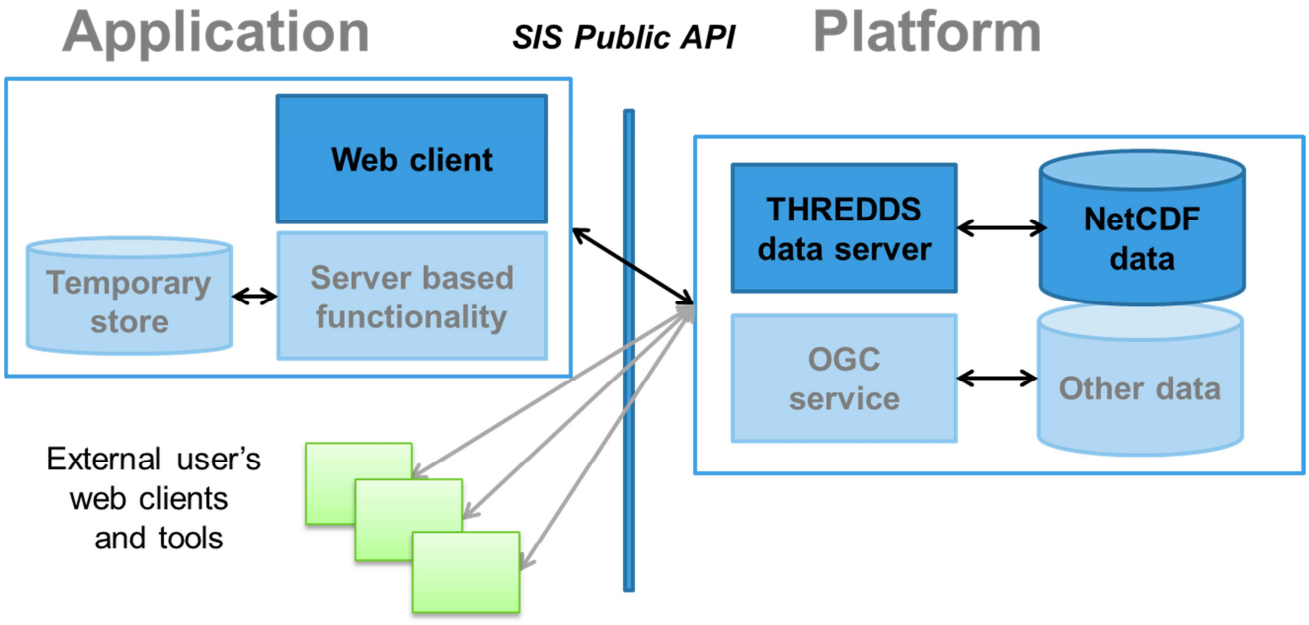


Figure 3: Technical architecture of the Urban SIS data portal

The platform will be based on THREDDS Data Server (http://www.unidata.ucar.edu/software/thredds/current/tds/) which enables efficient access and manipulation of NetCDF files following convention CF-1.6. In particular it provides functionality for selecting subsets of the NetCDF files and reformatting and downloads of data to a large number of specified formats, including OPENDAP, WMS and HTTP. It has been developed for Geo-scientific applications and applies to relevant OGC standardization. As indicated in figure 3, if user studies or work with data production within the project requires other kind of data than NetCDF to be managed, the platform can easily be extended with other server packages complying to OGC standards. A possibility which we have used in other SMHI services is using MapServer to manage GIS data.

The Urban SIS web client will be implemented in JavaScript and HTML and will demonstrate functionality for visualization and download of the data. For the extended functionality, such as data conversions and tailoring of indicators, we will use an additional server based implementation. The application side contains a temporary store to allow user to upload datasets to be used for tailoring of indicators and short term storage of the result before download. On this store, data will be kept for a short period (days) to allow the user to download the result. We do not plan to include any user registration on the platform, however, for requests with long processing times the user will be able to ask for a notification, e.g. via e-mail, when the requested data is ready.

Current implementation

A first version of the portal including home page, data server, a few test data sets, and a web client for visualization of data has been implemented. The current visualisation is shown in figure 4. The interface is a relatively simple test of the most important functionality. Available datasets can be selected in the list to the left. Once a dataset is selected the user can move in time using the slider at the bottom. If clicking at a location in the map, the time series corresponding to this point will be shown at the bottom of the page. More functionalities following table 3 will be added. These are further discussed in the section about new functionality below..

The interface are available for test on http://urbansis.climate.copernicus.eu/explore or http://urban-sis.smhi.se

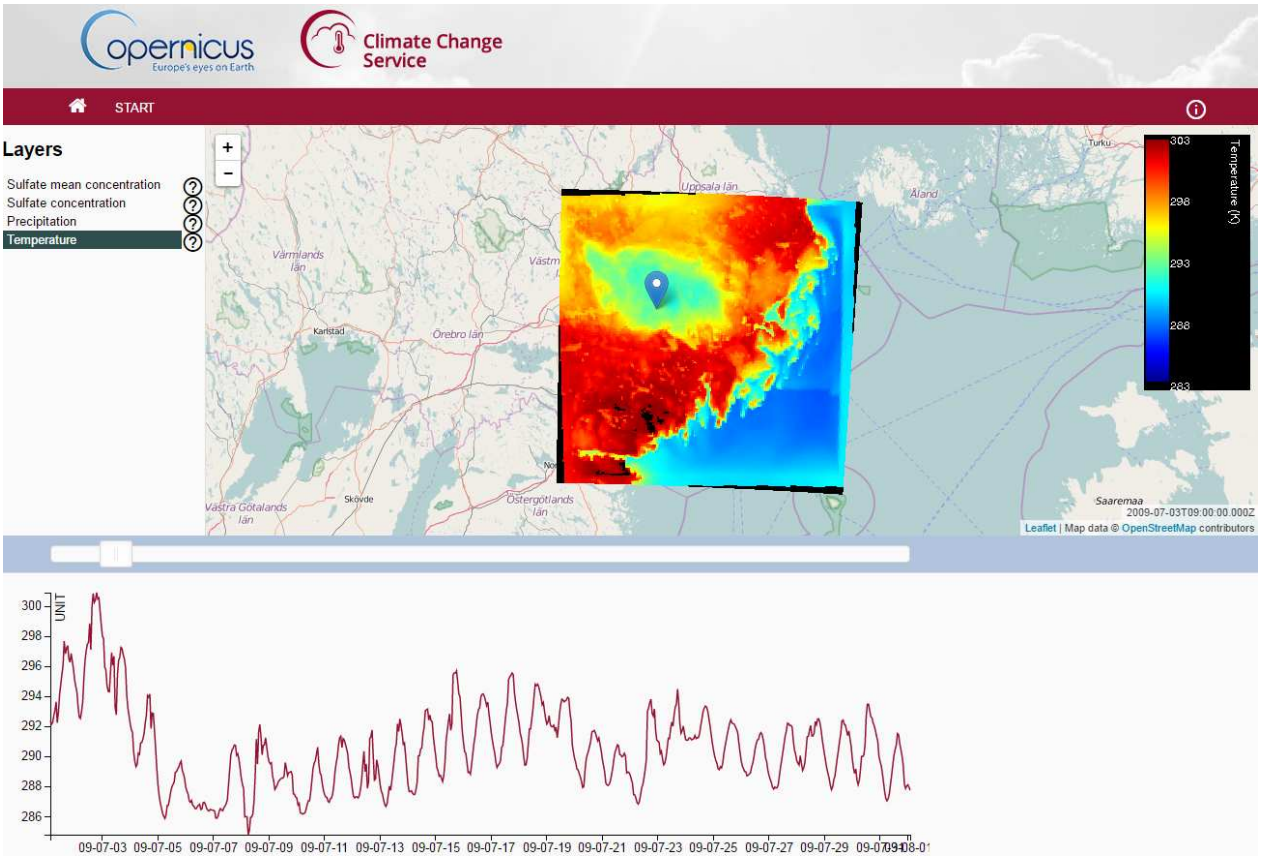


Figure 4: The web page for data exploration.

Functionality for data download will be developed later in the project when more datasets have been produced. However, for tests of available formats we provide a link to a simple download page at http://urbansis.climate.copernicus.eu/download or http://urban-sis.smhi.se/thredds/catalog/Urban-SIS/catalog.html. Figure 5 shows the list of formats currently available for download and test.

Our tests so far show that the solution gives a good performance for the data sizes needed within Urban SIS. The same architecture is used within the SWICCA (http://swicca.climate.copernicus.eu) project. In this project a number of datasets is available which shows the feasibility of the solution from a performance perspective.

Urban-SIS

THREDDS Data Server

Catalog http://urban-sis.smhi.se/thredds/catalog/Urban-SIS/catalog.html

Dataset: Urban-SIS Dataset/timeCPRC_resampled_2009-07-31.nc

- *Data size:* 3.818 Mbytes
- *Data type:* GRID
- *ID:* urban-sis-dataset/timeCPRC_resampled_2009-07-31.nc

Access:

1. **OPENDAP:** /thredds/dodsC/Urban-SIS/timeCPRC_resampled_2009-07-31.nc
2. **DAP4:** /thredds/dap4/Urban-SIS/timeCPRC_resampled_2009-07-31.nc
3. **HTTPServer:** /thredds/fileServer/Urban-SIS/timeCPRC_resampled_2009-07-31.nc
4. **WCS:** /thredds/wcs/Urban-SIS/timeCPRC_resampled_2009-07-31.nc
5. **WMS:** /thredds/wms/Urban-SIS/timeCPRC_resampled_2009-07-31.nc
6. **NetcdfSubset:** /thredds/ncss/Urban-SIS/timeCPRC_resampled_2009-07-31.nc
7. **ISO:** /thredds/iso/Urban-SIS/timeCPRC_resampled_2009-07-31.nc

Dates:

- 2015-12-22T11:23:01Z **(modified)**

Viewers:

- Godiva2 (browser-based)
- NetCDF-Java ToolsUI (webstart)
- Integrated Data Viewer (IDV) (webstart)

Figure 5: Available formats for download of the datasets.

Plans for further development

Contact persons

The following contact persons have been defined for discussion around technical issues in Urban SIS. When the Climate Store project has started, one representative from this project will be added to our contact list for discussions about storage solutions.

	Person		**Role**
SMHI	Lena Strömbäck	lena.stromback@smhi.se	WP Lead and main responsibility
	Per Lewau	per.lewau@smhi.se	Technical architecture and lead programmer
ECMWF	Eva Remete	eva.remete@ecmwf.int	Web content officer
	David Armstrong	david.armstrong@ecmwf.int	Copernicus Communication Team Leader

Table 4: Contact persons

New functionality

The continued development of the web pages and data portal will first be guided by input from the WP2, WP3 and WP4. Later in the project the input from the user studies in WP6 will be in focus for the design of the services. We expect that the continued development will be focused on the following functionalities:

- Further development of the user interface to allow the user to explore the datasets before download. This work will be based on the end-user dialogue in WP4 and include functionality for visualizing and comparing datasets in various ways. It also includes a tighter integration between the information pages and the data portal to allow easy access to relevant information on the datasets and how to use them for the end user.
- We will define a method for tailoring ECVs and impact indicators based on the end users need and demonstrate it on the portal. This is an additional functionality compared to the application but it follows the conclusions of D4.2 and 4.3. The functionality will include a possibility to make statistical or mathematical operations on time series or selected fields, for instance count days with temperature under a threshold or compute means or make operations based on data provided by the user. The design of the functionality will be done in cooperation with the user dialogue in WP4 and through end-user input from the WP6 use-cases.
- We have selected NetCDF for the storage base as this provides a consistent storage of all kinds of data for a city. However, based on the workshops performed by WP4, we foresee an end-user need to work with data in other data formats than NetCDF and we will explore methods for how this can be done. These methods range from automatic conversion of datasets, through dedicated downloadable tools to instructions and tips on how to apply commonly used tools on the datasets.
- As important as the actual format for downloaded data is the availability of relevant meta-data. The aim is to work towards that the downloaded package

includes data, meta-data and other supportive information that will make it easy for the end user to use the downloaded data package.
- As more data is added to the portal we will further develop the Google Analytics report to also include download of documents and data.

Time plan

The timeline for the continued development will be as follows:

May 2016–Dec 2016	Work together with WP4 to define the most important features and user needs for the platform. D 4.1-4.4 will give important requirements for further design of the web portal. Based on need from WP4, WP7 will support WP4 work with prototype development for testing a specific functionality
Jan 2017–Jun 2017	Development of the second prototype for the data portal and web client including Urban SIS datasets for the historical period (D3.1 Feb 2017) and a user interface based on results from WP4 work. Summarised as D7.2.
Oct 2017-Dec 2017.	Inclusion of climate datasets and possibly extended functionality.
Dec 2017-Jan 2018	Definition of deliveries and format for the operational phase. Summarized as D7.3.

Table 4: Timeline for future development

Conclusions

This report describes the technical architecture of the Urban SIS web pages and data portal to be used for project information and visualization and download of data. The web pages consist of webpages based on the Wordpress content management system and a data portal consisting of a platform based on the THREDDS data server and application software providing the user interface.

We have made a technical design and a first version has been implemented and is available for test at http://urbansis.climate.copernicus.eu. This site contains a few preliminary datasets that can be downloaded to test the functionality of the portal. Based on the foreseen type and amount of data in the project this technical solution is expected to meet the technical requirements in the project and for an operational setting. It can also be extended to meet additional user and data requirements that we can foresee during the project.

Appendix 1: Sample Google analytics report

Google Analytics

urbansis - http://www.urbansis.eu
All webbplatsdata

Öppna rapporten

Urban SIS report

5 mars 2016 - 4 apr. 2016

Alla användare
100,00 % Sessioner

Sessions

Sessioner

Sessions and Users by Country

Land	Sessioner	Användare
Sweden	46	17
France	9	7
Russia	7	1
United Kingdom	6	6
United States	5	5
Germany	3	2
(not set)	3	3
Greece	2	1
Italy	2	2
Spain	1	1

New Widge

Användartyp	Sessioner	Användare
Returning Visitor	51	18
New Visitor	36	36

Sessions

87
% av totalsumman: 100,00 % (87)

Users

48
% av totalsumman: 100,00 % (48)

Bounce rate

55,17 %
Ge. för vy: 55,17 % (0,00 %)

Sessioner

Sessioner by User Type

Users by Traffic Type

Trafiktyp	Användare
referral	33
direct	10
organic	9

Sessions and Usersby Internet Provider

Internetleverantör	Sessioner	Användare
swedish meteorological and hydrological institute	41	12
cjsc er-telecom company samara	4	1
digital ocean inc.	4	4
voxility.net	4	4
commissariat a l energie atomique	3	3
deutsches klimarechenzentrum (dkrz) hamburg	3	2
ip blocs for individual adsl accesses via	3	2
Instl656 st lambert bloc 2	3	2
cjsc er-telecom holding samara branch	2	1
enea - ente per le nuove tecnologie energia ambiente	2	2

Users by Operating System

Operativsystem	Användare
Windows	23
Linux	12
Macintosh	7
(not set)	5
iOS	1

Appendix 2: Sample NetCDF metadata for percipitation

```
netcdf //WINFS-
PROJ/data/proj/Fouh/Europe/Projekt/Urban_SIS/Leveranser_till_webteamet/Prognoser
/July2009.nc {
  dimensions:
    time = UNLIMITED;   // (744 currently)
    x = 189;
    y = 189;
  variables:
    double time(time=744);
      :long_name = "valid time";
      :units = "hours since 2009-07-01";
      :calendar = "julian";
      :_ChunkSizes = 524288; // int

    float x(x=189);
      :standard_name = "projection_x_coordinate";
      :long_name = "x coordinate of projection";
      :units = "m";

    float Precip(time=744, y=189, x=189);
      :long_name = "precipitation";
      :untis = "kg/(m2*h)";
      :grid_mapping = "transverse_mercator";
      :_ChunkSizes = 1, 189, 189; // int

    float T2m(time=744, y=189, x=189);
      :grid_mapping = "transverse_mercator";
      :long_name = "temperature";
      :units = "K";
      :_ChunkSizes = 1, 189, 189; // int

    float lat(y=189, x=189);
      :standard_name = "latitude";
      :units = "degrees_north";

    float lon(y=189, x=189);
      :standard_name = "longitude";
      :units = "degrees_east";

    float y(y=189);
      :standard_name = "projection_y_coordinate";
      :long_name = "y coordinate of projection";
      :units = "m";

    int transverse_mercator;
      :false_easting = 500000.0; // double
      :false_northing = 0.0; // double
      :grid_mapping_name = "transverse_mercator";
```

:latitude_of_projection_origin = 0.0; // double
:longitude_of_central_meridian = 15.0; // double
:proj4 = "+proj=utm +zone=33 +ellps=GRS80 +no_defs +datum=WGS84";
:scale_factor_at_central_meridian = 0.996; // double

// global attributes:
:Conventions = "CF-1.6";
:Originating_center = "Swedish Meteorological and Hydrological Institute (SMHI)";
:history = "Created on Tue 01 Dec 12:08 2015";
:title = "Downscaled historic climate";
:Comment = "NOTE! This file is assembled from multiple forecast runs. The original forecasts are run at 00, 06, 12 and 18 UTC during each day for July 2009. The gathered data in this file is created from the +1 to +7 hours from these runs creating a non-overlapping array of data. This does however lead to some inconsistensies in the seams between forecast runs. For example will the 1-hour precipitation between hours 7 and 8 (and so on for each seam) be an accumulation from an \"invisible\" time step.";
}

Appendix 3: Example of meta-data from D5.4.1

Characteristics	Example (Bologna)
domain	165x180 km^2
spatial / temporal resolution	1x1 km^2 / hourly
period	Jan 2006 – Dec 2010
downscaling model	HARMONIE cy40h1.1
forcing/boundary conditions	UERRA re-analysis over Europe with HARMONIE (11x11 km^2) (*add possible link to this data set*)
urban canopy model	SURFEX 7.3
land use	ECOCLIMAP II updated with Urban Atlas 2012
downscaling details	Urban SIS D3.1 Urban climate ECV and impact indicator data for historical conditions
validation	Urban SIS D5.1 Validation of climate variables

Metadata for Meteorological ECVs (downscaled historical period)

Characteristics	Example (Bologna)
domain	185x190 km2
spatial / temporal resolution	1x1 km2 / hourly
period 1 ("present climate")	2006-2010
period 2 ("future climate")	2055-2060
downscaling model	HARMONIE cy40 (climate mode)
forcing/boundary conditions	HARMONIE regional climate model over Europe (6x6 km2), forced by GCM EC_Earth output (T511 ~40x40km2). The GCM future scenario 2055-2085 is centered around a temperature increase of 4 degrees (SWL4) and driven by emission scenario RCP8.5 (add possible links to the RCM and GCM data sets)
urban canopy model	SURFEX 7.3
land use period 1 and 2	ECOCLIMAP II updated with Urban Atlas 2012 (held constant also for the "future" period 2)
downscaling details	Urban SIS D3.4 Urban climate ECV and impact indicator data for future conditions

Metadata for Meteorological ECVs (downscaled climate scenario)

impact indicators	Description (examples)
Valid for periods	2006-2010 (historical), 2006-2010 (“present climate”) and 2055-2060 (“future climate”)
Dry days per year	Mean and maximum number of consecutive dry days on yearly/ seasonally basis
Days with intense precipitation (per year)	Number of events when a given threshold is exceeded. Thresholds are considered as rainfall intensities computed as the 90th percentile of a climatological period. Significant durations for urban drainage/ flooding purposes are 1h, 3h, 6h, 12h, 24h. NOTE: Any information regarding sub-hourly intensities would be extremely valuable.
….	
….	
details	Urban SIS D4.3 Impact indicators for use in urban assessments

Metadata for Meterology/infrastructure impact indicators

ECMWF Shinfield Park Reading RG2 9AX UK