

International Journal of Spatial Data Infrastructures Research, 2018, Vol.13, 24-37  
Special Section: Citizen Data Science

## Citclops: Data Interpretation and Knowledge-based Systems Integration\*

Luigi Ceccaroni<sup>1</sup>, Filip Velickovski<sup>2</sup>, Alexander Steblin<sup>3</sup> and Laia Subirats<sup>4</sup>

<sup>1</sup> Earthwatch, UK, [lceccaroni@earthwatch.org.uk](mailto:lceccaroni@earthwatch.org.uk)

<sup>2</sup> Eurecat, Spain, [filip.velickovski@eurecat.org](mailto:filip.velickovski@eurecat.org)

<sup>3</sup> Eurecat, Spain, [alexander.stebelin@eurecat.org](mailto:alexander.stebelin@eurecat.org)

<sup>4</sup> Eurecat, Spain, [laia.subirats@eurecat.org](mailto:laia.subirats@eurecat.org)

### Abstract

Measuring the optical properties of water bodies (as indicators of, e.g., sewage impact, dissolved organic matter, sediment load or gross biological activity) is a way to assess their environmental status. The Citclops European project, in 2012-2015, developed systems to retrieve and use data on natural-water colour, transparency and fluorescence, using low-cost sensors combined with contextual information, taking into account existing experiences. This paper describes the general interpretation of data and delivery of information as carried out via the development of a decision support system named 'Citclops Data Explorer' and available from the main portal of the project.

**Keywords:** Citizen science, decision support system, data interpretation, knowledge-based systems integration, Citclops Data Explorer

---

\*This work is licensed under the Creative Commons Attribution-Non commercial Works 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/> or send a letter to Creative Commons, 543 Howard Street, 5<sup>th</sup> Floor, San Francisco, California, 94105, USA.

## 1. INTRODUCTION

Measuring the optical properties of water bodies (as indicators of, e.g., sewage impact, dissolved organic matter, sediment load or gross biological activity) is a way to assess their environmental status. The Citclops European project, in 2012-2015, developed systems to retrieve and use data on natural-water **colour**, **transparency** and **fluorescence**, using low-cost sensors combined with contextual information, taking into account existing experiences (e.g., Coastwatch Europe, ECSA's *Projects, data, tools & technology* working group and CSA's *Data and metadata* working group).

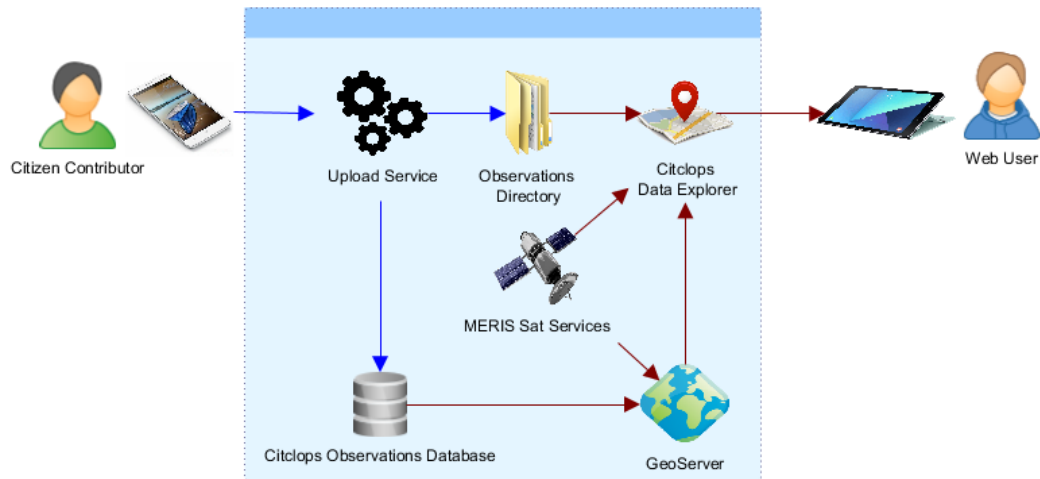
A decision support system named 'Citclops Data Explorer' has been developed and is available from the main portal of the European project Citclops, at [<http://www.citclops.eu/participate/citclops-data-explorer>]. The knowledge-based system behind the Citclops Data Explorer includes rules which relate sensor data streams, archived data sets and ecological status. EC directives that apply to the domain of study of Citclops have been surveyed. The local context conditions have also been surveyed and taken into account via the integration of available results of regular monitoring by governmental organisations. The basic methodology that couples Citclops measurements to the ecological status (like colour to bloom phenomena) has been described.

Different platforms have been necessary in order to develop the components of the Citclops system (e.g., water-quality sensor technology, mobile and Web applications) and in this paper the interfaces defined among the different platforms and technologies are described. Of particular importance to the users has been the integration among the end-user apps for data collection, the user interface and the data server used as a repository. The final integration of all components into one single system can be accessed by the users via the project portal [<http://www.citclops.eu>] (see Figure 1).

### 1.1. Data model for the captured data/observations

For archiving all the collected Citclops data, a data management system has been used, which includes the metadata presented in Table 1, and is interoperable with the SeaDataNet-EMODnet infrastructure to support two-way exchange. Tools for validating the gathered data, also using long-term environmental datasets as available through SeaDataNet and EMODnet, have been developed too, together with a service for discovery, access, and visualisation of Citclops data.

**Figure 1: Citclops's high-level architecture (source: Citclops)**



## 2. REQUIREMENT SPECIFICATIONS

The requirement specification included the definition of use case scenarios, functionality of prototypes, performance evaluation and test cases. At the end of each project period, the prototype being developed has been published and evaluated. Data *quality control* (QC) methods have been deployed during the project at various stages of development of the data interpretation. The QC consists of a protocol with knowledge rules that can analyse the status of an observation of colour. Additionally, information is provided to the observer citizen via the app interface with directions to improve the measurements.

## 3. SPECIFICATION OF DEVELOPMENT PLATFORMS

Different technologies have been used and finally integrated into development platforms. A careful choice of the development platforms has reduced integration problems. A comparison of real-time systems for data processing has been carried out and resulted in the definition of the development platforms and technologies used in Citclops's back-end (see

Figure 2) and front-end (see Figure 3).

All Citclops components shown in Figure 2 run on a single Debian Linux 6.0.8 server. The Citclops citizen observations are stored on a PostgreSQL 8.3 with PostGIS extension 2.0 to support geo-locations. These observations are accessible through a set of map-based web services through Geoserver 2.4.4 running on Apache Tomcat 7.0.2 to provide the satellite- and map-based services. The map-based user interface is provided through the python 2.7.9 Flask framework running on an Apache 2.4 web server.

**Table 1: Metadata for the captured data/observations (source: Citclops)**

Field name	Definition	Database file type	Multiplicity
Dataset_name	Short name of observation (e.g. ocean colour via citclops app)	text	1
Observation_ID	Unique id of measurement/observation (primary key)	text	1
Date_time	Date and time of measurement in ISO8601 notation: YYYY-MM-DDThh:mm:ss. Local time	date_time	1
Measuring_area_type	Geographical type, possible values: Point/Track/Area	text	1
Location_lat_long	Latitude/Longitude - decimal degrees, Mercator projection (for tracks/areas GML object with string of coordinate pairs). From time and position we can calculate the SZA. Make sure that we have sufficient spatial resolution (1 meter?), because this information is needed to the very local scale where (beach, harbour etc).	decimal	1-n
Datum_coordinate_system	Datum of the coordinate system. Default/Fixed = WGS84 (use SDN code list L101)	text	1
Parameters_measured	Via code used from vocabulary P02 - BODC Parameter Discovery Vocabulary (New codes might be needed.) (units will be part of the datafile!)	text	1-n
Flagging_reason_human	Output of humans saying why they are flagging something	text [1-16]	0-1
FU_value_human_app	Value of FU colour reported by a human, via an app, using colour bars superposed to a picture	integer [1-21]	0-1
FU_value_human_plastic_scale	Value of FU colour reported by a human using the colours of a plastic FU scale	integer [1-21]	0-1
FU_value_machine_DQC	Value of FU colour reported by a machine using the the DQC algorithm applied to a picture	integer [1-21]	0-1
Abstract	Short description of measurement. Default inserted by application. Specific details for measurement (other than strictly necessary for Data Quality check), such as aperture, exposure time etc. should be included in the abstract.**	text	1
Platform_type	Type of platform - Code from L061 - Platform type on which the sensor is mounted/carried	text	1
device_type	Sensor/device/instrument category - Code from L05 - Device categories	text	1
device_name	Specific name / brand of the device/instrument or sensor/device id (free text)	text	0-1
Station name	Name of measurement location = Field for name or id of observer. Although there might be problems with privacy, it might be interesting for specific groups to be coupled to instruments and observations. For example the Scuba club might check easily on the results of members and small communities could be proud to see the selection of 'their' effort to monitor the waters.	text	1
PI	For scientific data or for Citclops crowd the user name, mobile phone ID (privacy important!)	text	1
Station_start_date_time	Derived from first measurement date/time in case multiple observations/images are taken. ISO8601 notation: YYYY-MM-DDThh:mm:ss (converted to UTC)	date_time	1
Station_end_date_time	Derived from first measurement date/time in case multiple observations/images are taken. ISO8601 notation: YYYY-MM-DDThh:mm:ss (converted to UTC)	date_time	0-1
Language	Language of metadata in file (default = en)	text	1
Data_format	SDN data transport format via code from L241.	text	1
Data_Format_Version	Version number of data format	text	1
Data_file	Link to datafile via filename. A file can be of various types, either image, point measurement, or timeseries, or other.	text	1
viewing_angle	Specific for camera image: Angle towards the ground in which image has been taken (derived by appl). This is the angle in the vertical (up down) direction that is most important. For example, all traditional holiday pictures of a setting sun (angle 90 degrees, at the horizon) are useless.	decimal	0-1 (optional for other types)
azimuth_angle	A second angle is the azimuth angle, the horizontal angle between observation and the position of the sun. This is less important. Might be asked in the app (is your camera not pointing to the sun?)	decimal	0-1
Remark	Field for remark about e.g. data quality in free text	text	0-1 (optional)

Figure 2: Citclops's back-end technologies (source: Citclops)



Figure 3: Citclops's front-end technologies (source: Citclops)



The Citclops Data Explorer web pages utilise the Bootstrap framework to provide a responsive interface that can be accessible through PC, mobile phone or tablet. The maps are rendered through a JavaScript framework Leaflet and Bing maps are used for the satellite map layer. A combination of jQuery and Data-Drive Documents are used for plotting the charts of the Marine Data Analyser described in section 4.2.

#### 4. DEFINITION OF INTERFACES

The communication infrastructure for the different components has been investigated and defined (see Figure 1). The interfaces among components of this communication infrastructure allow Citclops's architecture to be easily extendable so that future applications for monitoring and decision support can be added freely and without much effort. These component interfaces also allow for updates with new services or modified contextual elements.

With respect to user interfaces, the Citclops Data Explorer presents three interfaces based on the content presented:

- Citizen Observations: <http://webmap.citclops.eu/citizen-observations>  
(see Section 4.1)
- Marine Data Analyser: <http://webmap.citclops.eu/marine-data-analyser>  
(see Section 4.2)
- Marine Data Repository: <http://www.citclops.eu/search/welcome.php>  
(see Section 4.3)

##### 4.1. Citizen Observation

The *Citizen Observations* part of the Citclops Data Explorer is used for navigation and exploration through all the observation samples of water taken by users of the “Citclops – Citizen water colour monitoring” app for smartphones or uploaded manually using the upload web-interface.

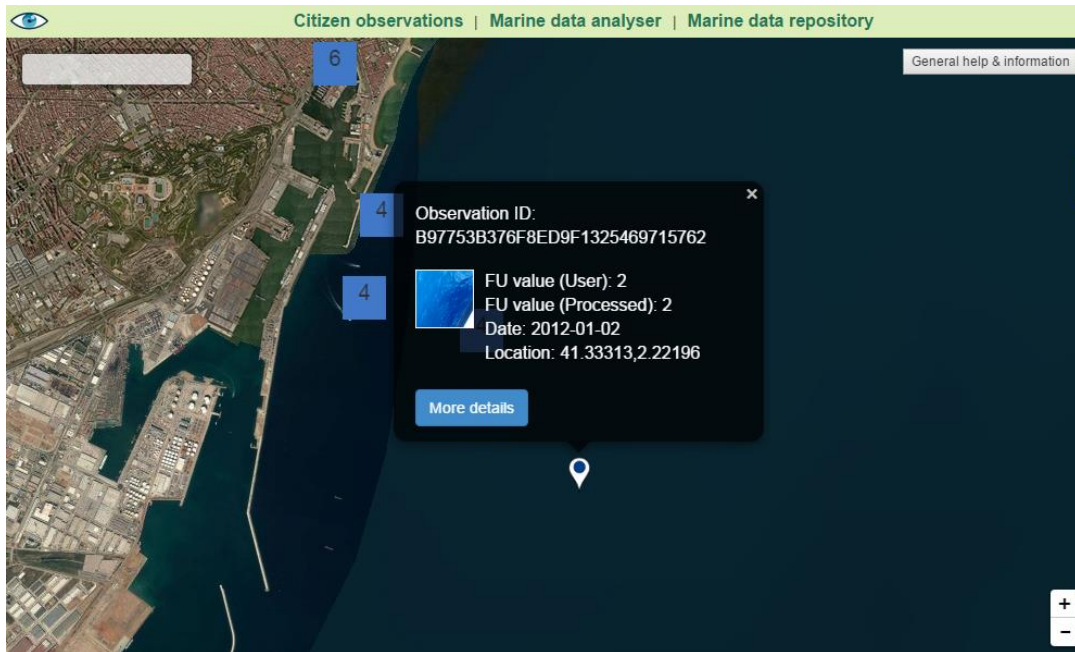
Once the user accesses the Citizen Observations interface, the screen shown in Figure 4 is presented. At the top, there is a menu with three options for the different views of the Citclops Data Explorer. On the map area there are markers indicating colour samples that have been collected by citizens during the project.

The user can click on these markers to consult detailed information about a particular sample. These colour samples are grouped by clusters or shown separately depending on the zoom level.

Upon clicking on one of the markers, a pop-up window will open with a brief description of the corresponding sample (its ID, date, two FU colour values, one reported by the user and the other calculated by the automatic algorithm, and coordinates). Pressing the *More details* button opens another pop-up window with more detailed information on the colour sample.

Clustering polygons are defined by a radius of 80 pixels from the central marker and are therefore relative to the zoom level.

Figure 4: 'Citizen Observations' interface (source: Citclops)



#### 4.2. Marine Data Analyser

The *Marine Data Analyser* (MDA) is a specific web map-based interface or view of the Citclops Data Explorer where an advanced user (such as a decision maker or marine researcher) can observe more in-depth information on the state of seawaters such as the historical state of bathing waters, seawater colour obtained from the MERIS satellite data, and historic weather data. This view allows data collected from different sources to be graphically visualised through time series plots in order to allow analysis and reveal possible patterns or connections among variables.

The main map view of the MDA consists of optional layers (Bathing waters, Satellite...) that may be enabled or disabled in the bottom right corner of the screen. By default, no layer is activated, and the view only shows a satellite map (provided by Microsoft Bing), centred on initialization at the geo-location of the user (if the location is available and if the user gives consent).

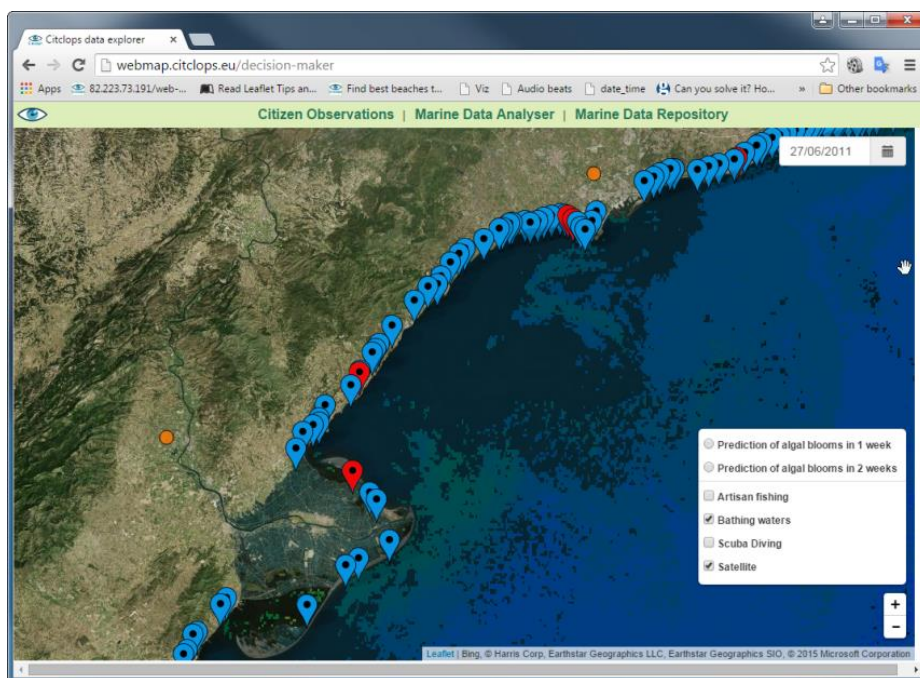
Since the 1970s, the EU has had rules in place to safeguard public health and clean bathing waters. The revised *Bathing Water Directive* (BWD) of 2006 updated and simplified these rules. It requires *member states* to monitor and



assess their bathing waters for at least two parameters of faecal bacteria: Enterococcus, and E. coli (Directive 2006/7/EC). In the MDA, the summer bathing period (June – September) of the year 2011, and the bathing waters of the Spanish autonomous communities of Catalonia, Valencia, and the Basque Country were selected as prototype test zones and period, due to the availability of data from multiple sources.

When the 'Bathing waters' layer is enabled, as shown in Figure 5, the MDA interface places markers in geo-locations corresponding to beaches that have been officially designated as bathing waters. Figure 5 is zoomed on the Ebro Delta coastal zone, showing bathing waters marked with blue and red markers. The blue markers represent indicators of normal levels of faecal bacteria as measured in the 2011 bathing period; red markers indicate that high or dangerous levels were measured. The orange circular markers are weather station locations, where data on rainfall were obtained.

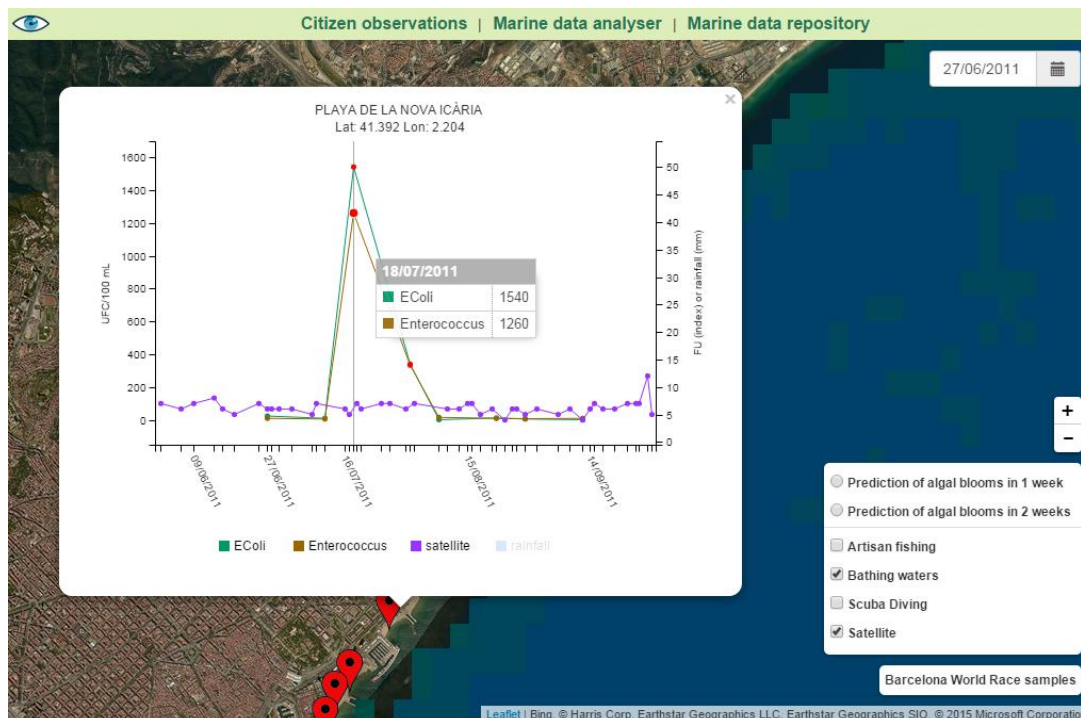
**Figure 5: The Marine Data Analyser interface with the 'bathing waters' layer selected, showing the bathing waters and weather stations close to the Ebro Delta (source: Citclops)**



When the 'Bathing waters' layer is enabled, a graph of multiple parameters is displayed as shown in Figure 6. The user can enable and disable different marine data parameters of interest (E. coli, Enterococcus, satellite Forel-Ule (FU)

colour, rainfall) and the interface will generate superimposed graphs so that patterns may be observed and investigated. Additionally, specific data points are marked in red where the faecal bacteria have been recorded to have a level above the safe limit.

**Figure 6: The Marine Data Analyser with the 'Bathing waters' layer and Nova Icaria beach in Barcelona selected, showing time series of the following parameters: E. coli, Enterococcus, satellite FU colour (source: Citclops)**

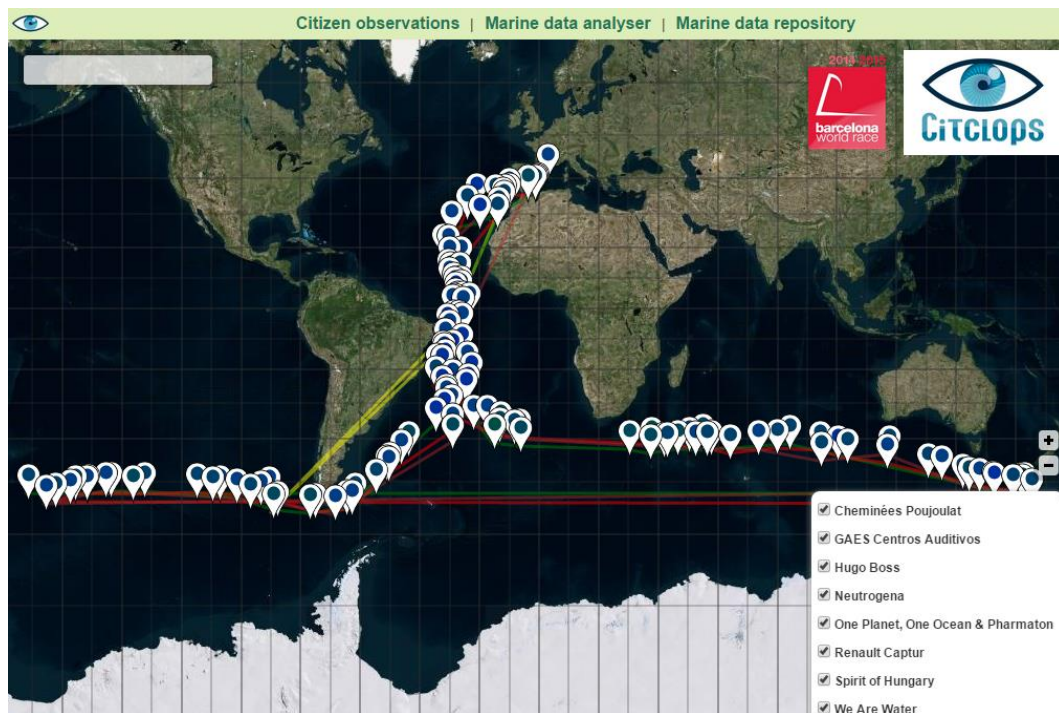


Apart from the graphics and satellite photographs, the MDA includes an additional view, which can be accessed through the *Barcelona World Race samples* button. This complementary view contains all the water-colour samples that were collected during the Barcelona World Race 2014-2015 (see Figure 7). One contribution to the project results was the collaboration with the *Barcelona World Race* (BWR).

During the almost four months of the race, 141 samples of the water colour of the ocean were collected. Many of the samples were obtained from areas with little or no direct information on water quality, especially from the Southern Ocean areas. All samples were then stored on the Citclops server and compared with historical data by the project researchers. During the comparison process, the

historical value which was taken into account was the one at the closest location to the sample, within a maximum radius of 300 km from the BWR's sample location. If there was no historical value in the radius of 300 km, the colour was considered unknown.

**Figure 7: Water-colour samples collected during the Barcelona World Race 2014-2015 (source: Citclops)**



For more information on the collaboration between Citclops and the BWR, the following dedicated website can be used: <http://citiclops-barcelonaworldrace.weebly.com>

#### **4.3. Marine Data Repository**

A repository has been developed to provide scientists with access to all the data collected via the “Citclops – Citizen water colour monitoring” smartphone app, together with other relevant datasets for doing research on the optical properties of the sea. Specifically, this repository, the Marine Data Repository, includes:

- Citclops/EyeOnWater apps observation data
- Historical ocean colour data (more than 100 years of data)
- EMODnet/SeaDataNet data: subset of optical properties

The interface can be accessed via the following URL:  
<http://www.citclops.eu/search/welcome.php>

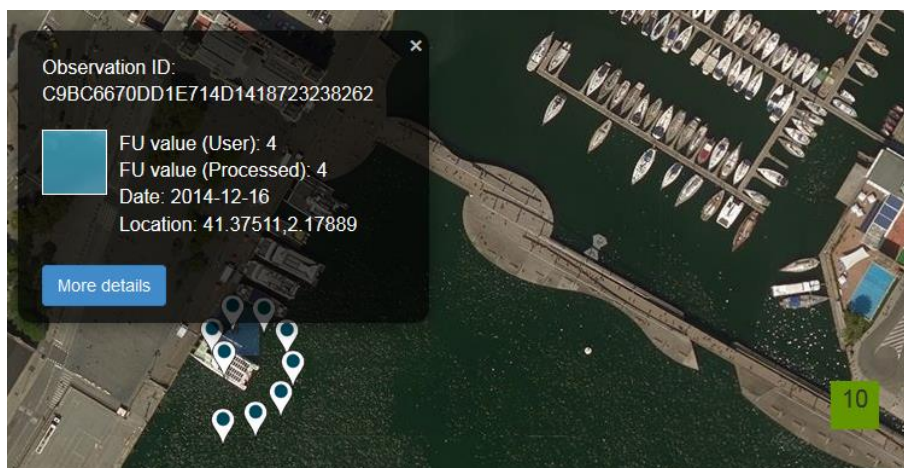
## 5. INTEGRATION INTO ONE PACKAGE

Sensor prototypes have been developed on different development platforms. Their integration consisted of making it possible to manage the sensors via apps. In this way all monitoring can be carried out by a mobile device. In this sense, all the different data-collection technologies have been integrated into one system and controlled via one interface.

With respect to information delivery, any citizen can navigate through all the observations of seawater, possibly grouped in clusters (see Figure 8), taken by users of the Citclops smartphone apps or uploaded manually using a dedicated web interface [<http://www.citclops.eu/participate/upload-your-image>].

In addition, to ensure the quality of the colour samples, two quality-control processes have been implemented and integrated to filter and augment the citizen observation data. *Data quality control* (DQC) is an important aspect when dealing with samples collected within citizen-science initiatives, and the "Citclops - Citizen water colour monitoring" app allows any user to upload measurements without registrations, or any experience in the field.

**Figure 8: Detail of the 'Citizen Observations' interface (source: Citclops)**

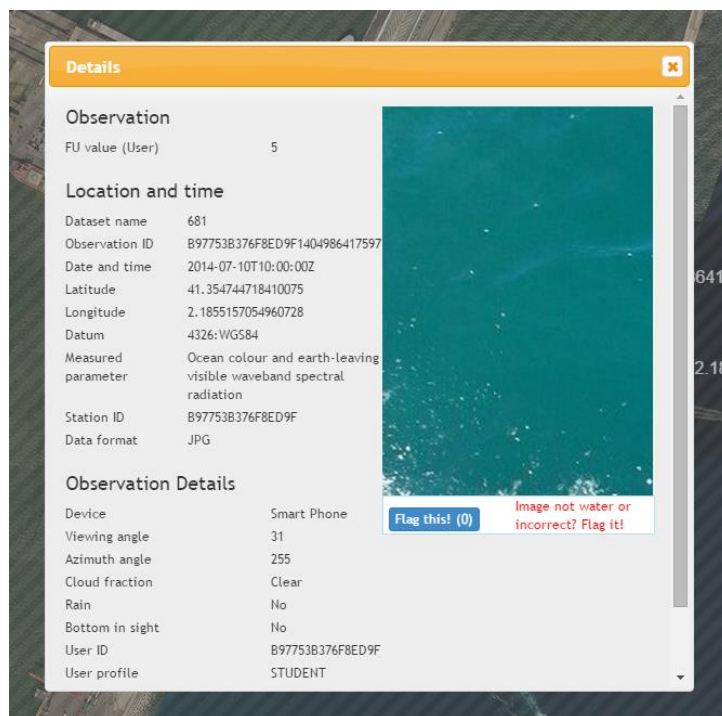


One of these processes is performed by a software module which computes the water colour from observation images collected via an app (Novoa et al., 2014 and Novoa et al., 2015). It works by locating, in the citizen-supplied image, a sub-

section that contains water, then computing the closest FU colour value of the sub-section. (Both FU colour values are stored: the one indicated by the user via the app and the one automatically calculated by the software.)

Another process to detect and remove invalid colour samples that do not correspond to a valid image consists in the use of collaborative techniques, which have been implemented and integrated into the user interface. As can be seen in Figure 9, under the image there is a “Flag this!” button, which provides the possibility for the user to flag the image if she thinks that the image is irrelevant within Citclops Data Explorer. Each user can vote only once per image. In brackets, the number of flags accumulated by each image is shown. Once the image has been flagged, the Citclops Data Explorer administrator will receive an email and will decide whether or not to proceed with the removal of the image. Thus the collective collaboration of the users of the tool will contribute to maintaining the quality of the samples.

**Figure 9: Collaborative, quality-control technique in the Citclops Data Explorer (source: Citclops)**



## 6. CONCLUSIONS

Raising environmental awareness and convincing citizens of the importance of environment measurements has been key to the success of Citclops. Web-based information is therefore presented in a way to attract a wide audience. A prototype of a decision support system, the Citclops Data Explorer, based on citizen science (Tulloch et al., 2013) in the domain of water quality has been defined and implemented. In particular, the Marine Data Analyser (part of the Citclops Data Explorer) allows the deployment and execution of computer-based, multi-scale models related to retrieved data on seawater colour. Intelligent ICT tools have been deployed to predict the evolution of environmental variables and to present them via the graphical visualisation environment of the Marine Data Analyser, which also includes a modular and formal knowledge base and an inference engine.

As a first step, an interface with a fixed set of interaction options has been developed. The interfaces between the sensors output, the Citclops Data Explorer and the interface module of mobile devices have been described. Using a modular structure, context awareness was added to the prototype. The context consists of time, environmental conditions, the user profile and the availability of other sources of information. The interfaces among the prototype modules have been explained. The prototype contributes a more flexible and citizen-based system to the monitoring community, which eases the access of end-users (citizens, decision makers and researchers) to acquired information. Also, the Citclops Data Explorer allows the establishment of a tie-in of citizen-science information and satellite mapping to provide citizens and decision makers with: an overall GIS impression of an area; post-processing of collected data (filtration and integration); derivation of information on the ecological status of an area; and production of key figures (maps and histograms) and their interpretation. Because of the relative availability of data in interoperable formats in the data server, the Citclops Data Explorer for water quality is mainly based on colour data only.

Nevertheless, the Citclops Data Explorer has been designed to be general enough to be applied to other domains and to include data on transparency and fluorescence as soon as they are available in the data server. The Marine Data Analyser, for example, is a modular system designed to be applied to the prediction of any data represented with the same constraints as the data retrieved in the Citclops project. However, in practice, domain-specific knowledge is necessary, and the Citclops Data Explorer uses this knowledge for reasoning, to handle typically occurring cases in the relatively narrow area of colour monitoring, in which the system has been validated (Ceccaroni et al., 2015a and Ceccaroni et al., 2015b). The user can access in-depth information on the state of seawater, such as the past state of bathing waters or the seawater colour

obtained from satellite observations. These data are collected from different sources and analysed by data mining algorithms. The results are shown by graphs or specific marks on the map (see Figure 6). Furthermore, in general, all the collected information on colour, transparency and fluorescence is converted to knowledge and maps on the ecological status of the waters. This knowledge is compared (close to the coast) to the status that is required by the Bathing Waters Directive. The results can be used to provide recommendations to the multiple sectors that make use of these waters, like local policy makers, tourists (regarding water quality and health) and fisheries.

## REFERENCES

- Ceccaroni, L., M. Blaas, M. R. Wernand, F. Velickovski, A. Blauw, and L. Subirats (2015). *A decision support system for water quality in the Wadden Sea*, Proceedings of the 47th International Liege Colloquium on Ocean Dynamics. Liège, Belgium.
- Ceccaroni, L., F. Velickovski, M. Blaas, M. R. Wernand, A. Blauw, and L. Subirats (2015). *Citclops Data Explorer: exploring water quality in the Wadden Sea*, Proceedings of the Environmental Information Infrastructures and Platforms Workshop. Barcelona, Spain.
- Novoa, S., M. R. Wernand, H. J. Van der Woerd (2014). The modern Forel-Ule scale: a 'do-it-yourself' colour comparator scale for water monitoring, *J. Eur. Opt. Soc. Rapid Publ.*, 9, 14025–1: 10.
- Novoa, S., M. R. Wernand, and H. J. Van der Woerd (2015). WACODI: a generic algorithm to derive the intrinsic color of natural waters from digital images. *Limnol. Oceanogr.: Methods*, 13: 697-711.
- Tulloch, A. I. T., H. P. Possingham, L. N. Joseph, J. Szabo, and T. G. Martin (2013). Realising the full potential of citizen science monitoring programs. *Biol. Conserv.*, 165: 128-138.
- Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC at <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006L0007>, [accessed 1 October 2015].