*International Environmental Modelling and Software Society (iEMSs)*
*7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,*
*Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)*
*http://www.iemss.org/society/index.php/iemss-2014-proceedings*

# OLES
# Online Laboratory for Environmental Sciences

**Sandrine Anquetin, Xavier Beaufils, Véronique Chaffard, Patrick Juen**

*Univ. Grenoble Alpes, LTHE, Domaine Universitaire, OSUG-B, F-38000 Grenoble Cedex 9, France*
*Univ. Grenoble Alpes, OSUG, Domaine Universitaire, OSUG-A, F-38000 Grenoble, France*
*sandrine.anquetin@ujf-grenoble.fr*

**Abstract:** The integrated cyber-infrastructure, OLES, provides an access to observation data and to tools and models built to enhance our understanding on the evolution of the Earth's water resources and climate. OLES aims at i) extracting the required data from database portals using OGC webservice (CSW, SOS), ii) building a specific process chain based on modules that use NetCDF format for exchanging data, iii) running the process in chosen computing facilities, OLES can connect outside on a private LAN and iv) visualizing the result of the process. Based on J2EE, the MMI of OLES is a web interface and interacts with EJB objects. OLES uses web services to communicate with a sequencer developed in C++.

**Keywords:** Cyber-infrastructure; database; water cycle modeling tools.

Acronyms are listed in ANNEX.

## 1. INTRODUCTION

One of the major scientific challenges in the 21$^{st}$ century is to improve our understanding on the evolution of the water cycle associated with the climate variability. Main issues concern the prediction of i) the water resource and the access to drinkable water and ii) the extreme events, both droughts and floods. Observation strategies covering a wide range of space and time scales must therefore be set up, while continuing advanced research on the involved mechanisms and developing integrated modeling approaches.

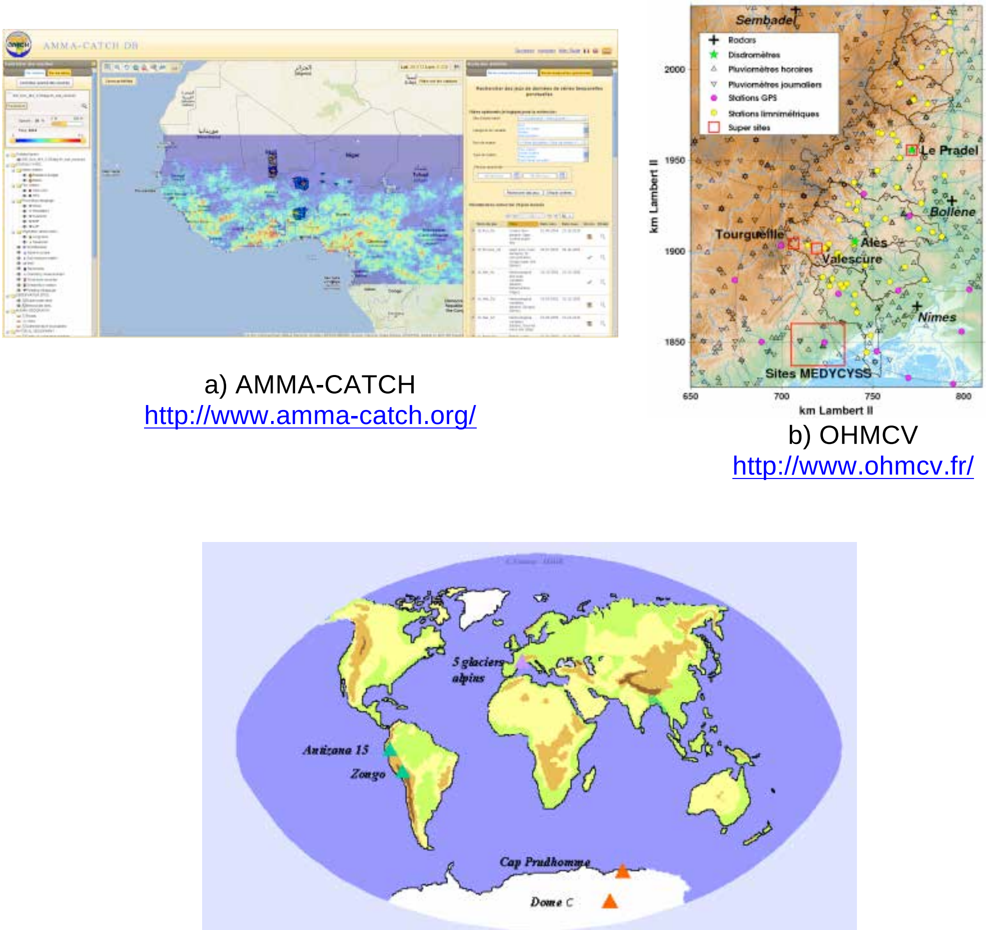
Within this general context, the present work relies on three natural observatories, located in West Africa, Worldwide Glaciers, and in Mediterranean region, managed at LTHE (Laboratoire d'étude des Transferts en Hydrologie et Environnement; Grenoble, France) and gathered at OSUG (Observatoire des Sciences de l'Univers; Grenoble, France). Their scientific objectives aim at improving the understanding of the water cycle functioning, providing water and mass balances for multi-scale basin sizes, and evaluating the hydrological impacts of the evolving climate.

Each of the observatories has its own database, and modeling tools were developed separately leading to important efforts often duplicated. Therefore, there was a need to build an integrated cyber-infrastructure to provide access to data, and to shared tools and models that enable the understanding of the water cycle. This is the project called OLES, for Online Laboratory for Environmental Sciences.

Focused on the understanding of the water cycle under contrasted climates, OLES will facilitate the work of the scientific community and then, help interactions between the research community and water agencies or diverse stakeholders. Long-term objective is to promote education centered in water science strongly connected with climatic issues.

This paper aims at introducing this ongoing project.

In section 2, we introduce the three natural observatories and the associated observation strategy based on their specific scientific question. The three databases are shortly described. In section 3, the concepts, the technical choices and the implementation of the cyber-infrastructure OLES are detailed.




a) AMMA-CATCH
http://www.amma-catch.org/

b) OHMCV
http://www.ohmcv.fr/

c) GLACIOCLIM http://www-lgge.ujf-grenoble.fr/ServiceObs/

**Figure 1.** Location of the three natural observatories

## 2. TOWARDS THE UNDERSTANDING OF THE WATER CYCLE UNDER CONTRASTED CLIMATES

To improve our understanding of the Earth's water and climate, our contribution deals with three regions (Fig.1) where we aim at investigating three different key questions (i.e. evolution of water resources; hydrological risk assessment and risks associated with glacier dynamics) that are strongly connected to the region of interest and for which social issues are addressed. Observations and modeling approaches are thus built to reach these objectives.

The hydrological and meteorological observatory on West Africa (AMMA-CATCH; Fig.1a) aims at **understanding the vegetation's dynamic, the water cycle and their interaction with the climate in West Africa**. The main objective is to study the hydrological impact of climate change. The observations involve i) documenting the spatial variability of precipitation fields associated with convective systems (internal variability and life cycle) and the impact of his variability on the various terms of the continental hydrological cycle; ii) identifying the characteristic times of different processes (runoff, infiltration, groundwater flow) and their role in the intra-seasonal and interannual variability; iii) characterizing the seasonal cycle of vegetation and its interaction with the water cycle.

The Mediterranean hydrometeorological observatory (OHMCV; Fig.1b) was set up to improve the **understanding and prediction of devastating flash flood**. The observatory gathers both physical (Boudevillain et al., 2011) and social (Ruin et al., 2014) observations. The collected data concern i) rainfall time series, ii) river water level and discharge time series, iii) radar data collected from operational radars, iv) postflood investigations (peak discharge; timing based on interviews; semistructured interviews to collect timing and spatial information related to the evolution of the environmental conditions and the individuals' location and pace of activities. Its objective is to document how individuals switch from routine activities to emergency coping behaviors).

GLACIOCLIM (Fig.1c) aims at gathering long-term glacial and meteorological observations i) to study the **link between climate and icecap**, ii) to predict the **water resource associated with future evolution of glaciers** and iii) to understand the dynamics of the glaciers (velocity, depth and length) and to **study the risks associated with glaciers evolution**. GLACIOCLIM is present in three contrasted climatic regions: i) the French Alps as representative of temperate climate, ii) the Andes (Bolivia; Ecuador) for tropical climate and iii) Antarctic for polar climate.

Variables associated with the water cycle (i.e. precipitation; soil moisture; snow cover; discharge; air and river temperatures; suspended materials) are observed and stored in three different databases built under specific technical constraints linked to the respective partnerships of the natural observatories.

## 2.1. Databases: data storage and metadata

Tab.1 synthesizes the variables stored in the three databases. For more details on the sampling strategy, refer to the respective websites. Each data producer already performs data critical analysis. However, a critical analysis carried out on the whole data set coming from different networks may significantly improve the detection of erroneous data. As an example, AMMA-CATCH and OHMCV use geostatiscal data quality control technique to identify any abnormal rain amounts with neighbors.

Each observatory has its own database from which time series and gridded data can be extracted associated with their metadata. Tab.2 gives the main specificities of these databases.

AMMA-CATCH information system architecture uses three OGC web services:
- The CSW web service allows querying and retrieving the metadata in ISO 19139 and DublinCore formats;
- The SOS allows querying and retrieving the observation data in OGC O&M (Observations and Measurements) format as well as the sensor metadata in OGC SensorML (Sensor Model Language format) format;
- The WMS web service allows querying and displaying the GIS layers on the map (sensor positions, different layers, etc.).

These standards facilitate system interoperability. OLES relies on them to interface with databases.

**Table 1**. Water cycle variables stored in the three databases. RG stands for rain gauge data, DSD for disdrometers observations, Accu. for snow accumulation, EB for "energy balance", LAS for "large aperture scintillometer", T for temperature, RH for relative humidity; WL for water level, Veget. for,vegetation, WT for water table and SM for soil moisture.

| | **AMMA-CATCH** | | | **OHMCV** | **GLACIOCLIM** | | |
|---|---|---|---|---|---|---|---|
| | Niger | Benin | Mali | | Alps | Andes | Antarctica |
| Surface (km$^2$) | 10000 | 15000 | 30000 | 33600 | 350 | 310 | 100000 |
| Rain | RG | RG; DSD<br>Radar X | RG | RG; DSD<br>Radar S | RG<br>Accu. | RG<br>Accu. | Accu.<br>DSD |
| Atmosphere | Flux<br>EB | Flux<br>LAS<br>EB | Flux<br>EB | | EB, surface wind, T, RH | | |
| Surface | Erosion<br>Veget.<br>SM | SM | Veget.<br>SM | | | | |
| Surface water | WL | WL | | WL | | | |
| Ground water | WT | WT | | | | | |
| Glacier | | | | | Ablation; Albedo; Accumulation | | |
| | | | | | Thickness; Ice velocities; Length fluctuations | | |

**Table 2.** Specificities of the three databases

| | AMMA-CATCH | OHMCV | GLACIOCLIM |
|---|---|---|---|
| | **Data access and format** | | |
| Data Access | Extraction on line through a web portal<br><br>http://bd.amma-catch.org/amma-catch2 | Extraction on line through a web portal<br><br>http://sevnol.ohmcv.fr/Sevnol2 | Today: Files accessible on line through a web site<br>Near future: Extraction on line through a web portal |
| Data format | CSV, netCDF, XML (OGC O&M) | XML, CSV, netCDF | Today: XLS<br>Near future: CSV, netCDF, XML (OGC O&M) |
| Metadata format | XML (ISO 19139 and DublinCore for dataset, OGC SensorML for sensors) | XML | Near future: XML (ISO 19139 and DublinCore for dataset, OGC SensorML for sensors) |
| | **Information System architecture** | | |
| Back end | DBMS PostgreSQL + PostGIS | Files | Today: Files<br>Near future: DBMS |
| Front end | Java JSF Web application | Java Web application | Today: web site<br>Near future: Java JSF Web application |
| Interoperability | By implementing OGC web services: CSW, SOS, WMS | | Near future: By implementing OGC web services: CSW, SOS, WMS |

### 2.2 Models and associated tools

Several tools and models are required to provide a comprehensive view of the evolution of the hydrosphere and/or the cryosphere under contrasted systems and glacial systems.
Hydrological models aim at testing hypotheses on the processes involved in the functioning of the studied catchment. In this context, hydrological modeling is based here on either DHSVM (Doten et aL, 2006; Naz et al;, 2014), CVN (Anquetin et al., 2010; Vannier, 2013) and TOPAmma (Lelay et al., 2008), depending on the scientific issues.
To complement field observations, several “external” data are also needed such as DEM, satellite observations, meteorological re-analysis, etc. These data are stored in “external” databases that have their own specificities.
Thus, running hydrological models require several pre-treatments of both field observations and “external” data.
Many tools are thus been developed at the scale of the community of each observatory (data quality control; statistics; krigging; etc); the programming language differs depending on the programmers and the constraints associated with the observatory.

Tab.3 summarizes some of these tools.
There was a need to find a way to share these efforts, and, thus, save time. This was one of the motivations to build OLES.

**Table 3.** List of the tools and models that will be implemented in OLES

| | Description | Format |
|---|---|---|
| **Conversion tools** | Change of format; Grid manipulation; … | FORTRAN; C++ |
| **Data treatment** | DSD treatment; Rain gauge quality treatment; Radar observation treatment (ground clutter; mask; etc); … | FORTRAN |
| **Utilities** | Simple statistics on 1D or 2D data | FORTRAN; R |
| **Spatial interpolation** | Variogram; krigging; … | Python; R |
| **DEM analysis** | DEM correction; Catchment extraction; … | C; FORTRAN |
| **Hydrological models** | ETo computation; DHSVM; CVN; TOPAmma; … | Excel; C++; FORTRAN |
| **Evaluation/performance** | Several scores (Nash; R2; BIAS; etc); … | FORTRAN; R |

## 3. CYBER-INFRASTRUCTURE FOR UNDERSTANDING THE EVOLUTION OF THE WATER CYCLE UNDER CONTRASTED CLIMATES

The cyber-infrastructure, OLES, is built to reach the following objectives:
- Being able to equally select data from the three different databases;

- Making easier the pre-treatment of the required data for modeling issues;
- Being able to share data, tools and models;
- Keeping the intellectual property of each of them; the owner can define the intellectual property of either data and modules: worldwide open, some restriction or private;
- Being able to add and/or to build new distant databases, new modules or meta-modules (sequence of several modules);
- Being able to use modules written with any programming languages;

A synthetic view of the general context of OLES is given in Fig.2.

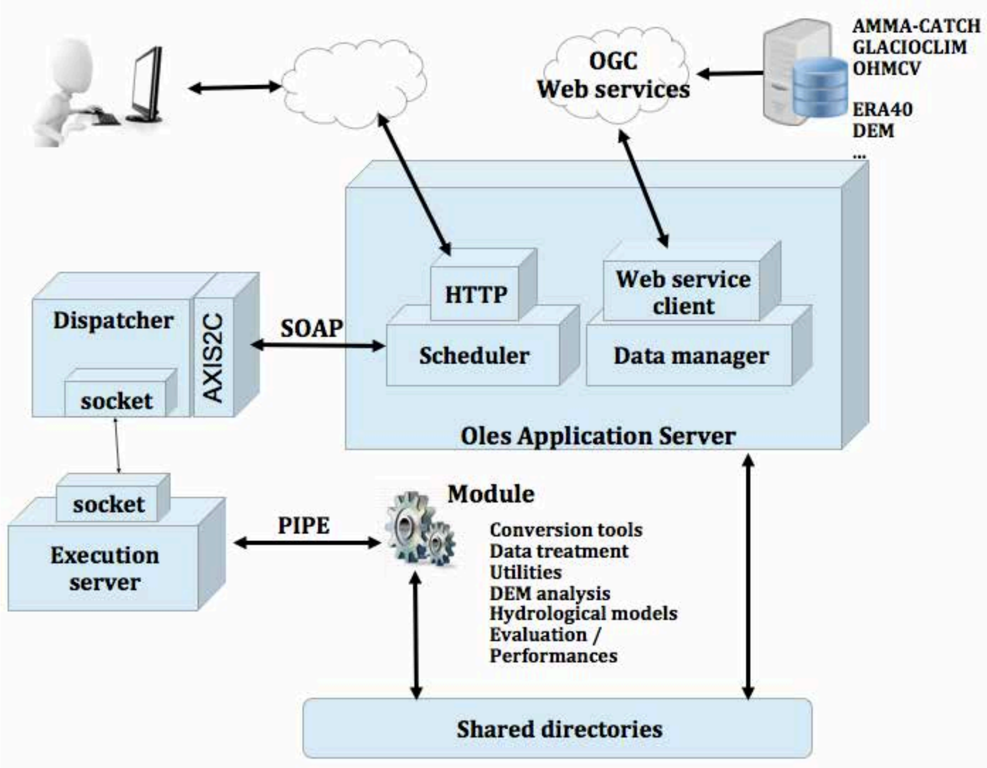



**Figure 2.** General concept of OLES

Based on J2EE, the MMI of OLES is a web interface and interacts with EJB objects. OLES uses web services to communicate with a dispatcher developed in C++.

Fig.3 presents screenshots of this web interface, from which the user can:
i) Extract the required data from a GIS server (Fig. 3a);
ii) Build a specific process chain based on modules (Fig.3b);
iii) Launch the process in a chosen local or distant computing facility (LINUX server; High computing center; etc …); OLES can connect outside on a private LAN (Fig.3c);
iv) Visualize the result of the process.

### 3.1 Data access and communication

The interface between OLES and databases is managed by standardized OGC web services to allow the access of distant databases.
The CSW GetRecords operation allows for metadata retrieving and the SOS GetObservation operation for data time series retrieving.
OLES displays the content of each registered database by calling the CSW GetRecords operation (Fig.3a). For each dataset, the user can then select part of the database based on either spatial or time filters.

To retrieve data from database, two SOS operations are implemented in the OLES workflow:

- SOS GetCapabilities operation is used to get information concerning the observation and the sensor metadata to build the final data request in a next operation;
- SOS GetOperation is used to extract data using several filters such as the list of observed variables (called procedure), the list of sensor system (called features of interest) and a temporal filter, according to the user's needs.

The SOS returns the observation time series in the O&M XML format

The collected data are then saved in the disk cache that is automatically emptied based on the updating frequency of each accessed sites. A shared storage capacity is also integrated to the OLES architecture.

### 3.2 Data interoperability

The use of NetCDF V4 format ensures the data interoperability within the system. It is based on the CF convention. Two different templates are defined to build the exchange files within OLES: station time series and grid time series.

### 3.3 Model implementation and execution

The implementation of a model is based on the execution of successive distributed modules in different distant computing facilities. This sequence is called meta-module; the outputs of a module thus become the inputs of the following modules.

For the construction phase (Fig.3b), the user has to specify the links between the modules and the data required for the execution.
Each files generated by the modules are stored in a shared file system. When the process uses several computing facilities, the modules access to the same data.

Based on the sequences described in the meta-module (Fig.3b), the cyber-infrastructure OLES organizes the order of the module execution and sends it to the dispatcher that forwards it to the appropriate computing facility, as previously defined in the construction phase.
A client OLES has been deployed on all the computing facilities, and is waiting for the order. When the client receives the order to execute the module, the module is then run with the parameters provided by the user in the construction phase.
When the process is finished, the client OLES sends back a completion code and the outputs to the dispatcher that forwards this information to OLES.
If the process is successful, then OLES goes on to the next modules (Fig.3c).

## 4. CONCLUSIONS

OLES, the cyber-infrastructure described in the paper is on development. A successful prototype allows collecting data from the AMMA-CATCH database and several modules for spatial interpolation are already implemented. They are run on distant distributed computing facilities.

## 5. ACKNOWLEDGMENTS

This work has been supported by a grant from Labex OSUG@2020 (Investissements d'avenir – ANR10 LABX56). Sandrine Anquetin, Véronique Chaffard and Patrick Juen (LTHE, Grenoble, France) and Xavier Beaufils (OSUG, Grenoble, France) are part of Labex OSUG@2020 (ANR10 LABX56). Moreover the authors deeply thank the contribution of the OLES user's committee that helps to precise the specifications required for OLES.

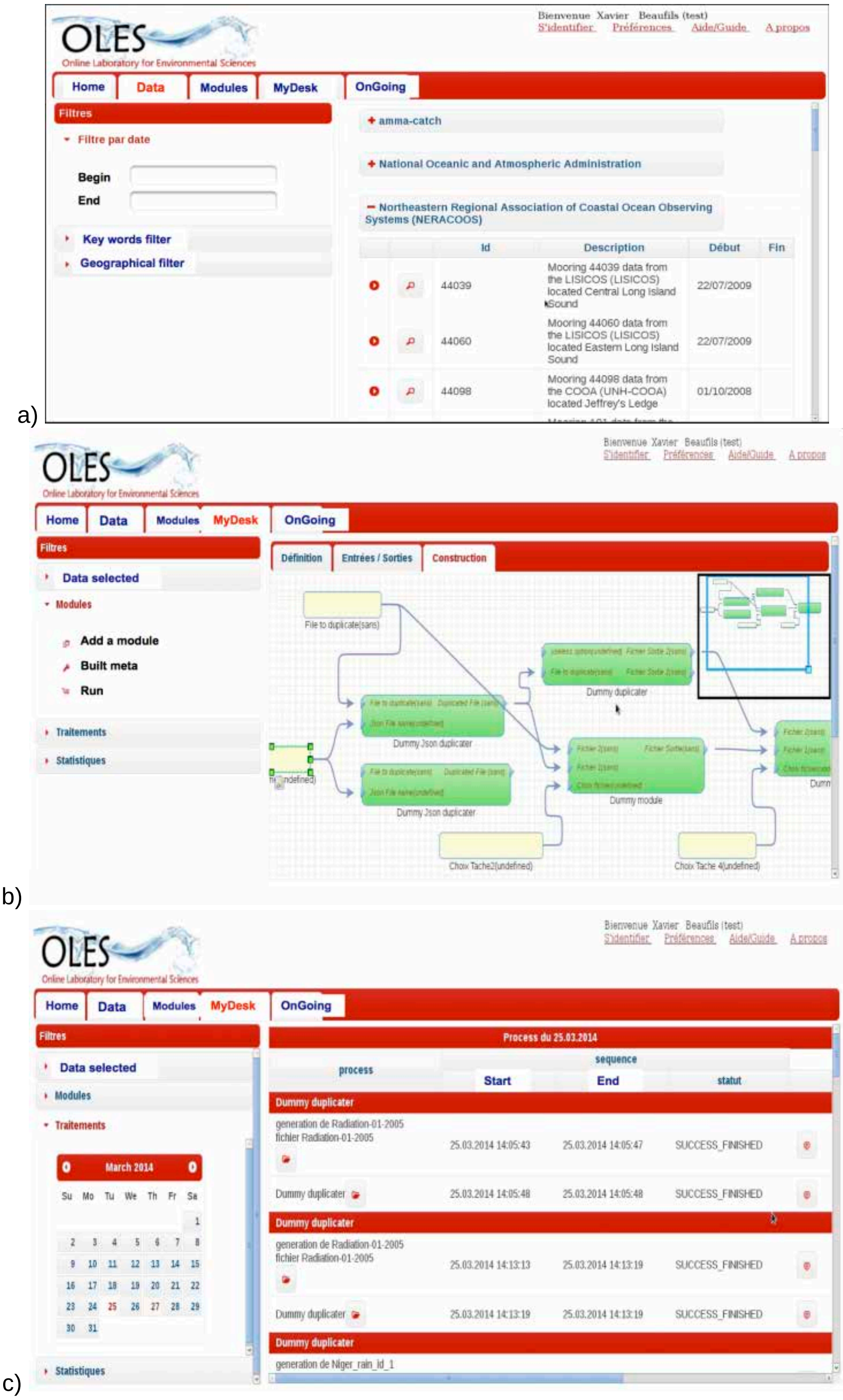



**Figure 3.** Screenshots of the web interface OLES. a) Data collection. In this example, the data are extracted from the AMMA-CATCH database; b) Construction of a chain of treatment (meta-module); c) Processing of the meta-module.

## 6. REFERENCES

Anquetin, S., Braud, I., Vannier, O., Viallet, P., Boudevillain, B., Creutin, J.-D., and Manus, C.: 2010. Sensitivity of the hydrological response to the variability of rainfall fields and soils for the Gard 2002 flash-flood event, J. of Hydrology, 394, 134 – 147.

Boudevillain, B., Delrieu, G., Bonnifait, L., Bouilloud, L., Kirstetter, P. E., and Mosini, M. L.: 2011. The Cévennes-Vivarais Mediterranean hydrometeorological observatory database, Water Resources Research, 47, 6.

Doten, C.O., L.C. Bowling, E.P. Maurer, J.S. Lanini, and D.P. Lettenmaier: 2006. A spatially distributed model for the dynamic prediction of sediment erosion and transport in mountainous forested watersheds, Water Res. Res., Vol. 42, N. 4.

Le Lay, M., Saulnier, G.-M., Galle, S., Séguis, L., Métadier, M. et Peugeot, C. (2008). "Model representation of the Sudanian hydrological processes : Application on the Donga catchment (Benin)". Journal of Hydrology. Vol. 363. no. 1-4, p. 32–41.

Naz, B.S., C.D. Frans, G.K. Clarke, P. Burns, and D.P. Lettenmaier: 2014. Modeling the effect of glacier recession on streamflow response using a coupled glacio-hydrological model, Hydrol. Earth Syst. Sci., 18, 787-802

Vannier, O.: 2013. Apport de la modélisation hydrologique régionale à la compréhension des processus de crue en zone méditerranéenne. Grenoble University (PhD in French*).*

Ruin, I., Boudevillain, B., Creutin, J. D., Lutoff, C., Anquetin, S., Bertran Rojo, M., Boissier, L., Bonnifait, L., Borga, M., Colbeau-Justin, L., Creton Cazanave, L., Delrieu, G., Douvinet, J., Gaume, E., Gruntfest, E., Naulin, J. P., Payrastre, O., and Vannier, O.: 2014. Post-event field investigation: A Socio-Hydro-Meteorological approach to the 15-16 June 2010 disastrous flash flood event in the area of Draguignan, Weather, Climate and Society, 6(1), 135-153.

## 7. ANNEX – LIST OF ACRONYMS

AMMA-CATCH: African Monsoon Multidisciplinary Analysis - Coupling the Tropical Atmosphere and the Hydrological Cycle
CSW: Catalog Service for the Web
DEM: Digital Elevation Model
EJB: Enterprise Java Bean
GIS: Geographical Information System
GLACIOCLIM: Les GLACIers, un Observatoire du CLIMat
J2EE: Java Enterprise Editor
LAN: Local Area Network
MMI: Man Machine Interface
OGC: Open Geospatial Consortium
OHMCV: Observatoire Hydrométéorologique Méditerranéen – Cévennes Vivarais
O&M: Observation and Measurement
SOS: Sensor Observation Service
WMS: Web Map Service