Deliverable 1.1

Specifications of the modular software architecture based on the end-user demand

Dissemination level: Public

WP1
Specification of the Online Security Management Toolkit

31st December 2012

SMaRT-Online WDN
Online Security Management and Reliability Toolkit for Water Distribution Networks

ANR reference project: ANR-11-SECU-006
BMBF reference project: 13N12180

Contact persons
Fereshte SEDEHIZADE Fereshte.Sedehizade@bwb.de
Olivier PILLER Olivier.Piller@irstea.fr

WP 1–Specification of the Online Security Management Toolkit

D1.2 Fine Specifications of the Demand of the End Users of the Online Security Management Toolkit

List of Deliverable 1.1 contributors:

From VERI

Marie Maurel (marie.maurel@veolia.com) [WP1 leader]
Damien Chenu (damien.chenu@veolia.com)
Pierre Mandel (pierre.mandel@veolia.com)

From Irstea

Olivier Piller (olivier.piller@irstea.fr)

Work package number	1	**Start date:**		17/04/2012				
Contributors id	VERI	BWB	CUS	3S Consult	IOSB	TZW	ENGEES	Irstea
Person-months participant	3	4	2	3	2	2	2.5	2

Keywords
System architecture – Data flows – Data format – End users

Objectives
Specifications of the data flows to be considered in order to reach the objectives of the online security management toolkit. In particular, the type (frequency, volume, format etc.) of flows should be determined, based on the already existing system (hardware and software) architecture of the end users.

Summary

The SMaRT-Online^{WDN} project aims at developing an online security management toolkit for drinking water distribution systems (DWDS) based on the combination monitoring/modelling.

Many modelling software tools for DWDS have been available for more than twenty years. However, very few of them are able to process all the real-time data, that are increasingly being generated by sensors. Handling these data and making sensors *smart* raises the issues of: (i) transferring/storing the output data they generate; (ii) processing continuously these data in order to provide operators with reliable and valuable information. The SMaRT-Online^{WDN} project shall deal with them.

The objective of this deliverable is to specify the data flows to be considered in order to reach the objectives of the online security management toolkit. In particular, the type (frequency, volume, format etc.) of flows should be determined, based on the already existing system (hardware and software) architecture of the end users.

Here are the main conclusions:

1) A general system architecture is proposed with the following modules (hardware or software tools): Sensors, Supervisory Control And Data Acquisition (SCADA) system, Data treatment, Model (hydraulics and quality), Online calibration, Database, Geographic Information System (GIS), Event detection algorithm, Statistical model, Optimisation algorithm and Source Identification;
2) The new modules to be developed are: Data treatment, Model (hydraulics and quality), online calibration, Database, Event detection algorithm, Statistical model, Optimisation algorithm, Source Identification;
3) Aside from the development of the new modules, the following modules, particular attention should be paid to the input/output data flows of the following elements: GIS and model (hydraulics and quality).

List of Figures:

List of Tables:

Contents:

1. Background

While multiple modelling software tools for drinking water distribution systems (DWDS) have been available for more than twenty years, very few of them are able to work with real-time data. However, recent developments in water quality monitoring are quickly popularising the use of sensors, which generate large amounts of real-time data. Handling these data and making sensors *smart* raises the issues of: (i) transferring/storing the output data they generate; (ii) processing continuously these data in order to provide operators with reliable and valuable information.

Modelling software tools for DWDS have first been developed to reproduce average hydraulic phenomena (pressure and flow). This coincided with the advent of the first computers affordable by research institutions in the early 1960s (Ormsbee, 2006). In the beginning of the 1980s, researchers started modelling water quality in DWDS and proposed several approaches, generally combined with the already existing hydraulic models (Graymann and Clarck, 1990). Recently, there has been a renewed interest in quality modelling since DWDS may be potential targets for terrorists. A review of recent water quality modelling methods can be found in (Ostfeld, 2005). Nowadays, while hydraulic simulation software is in widespread use within the water industry for strategic supply analysis, design of control strategies, network extensions, and maintenance planning (Machell *et al.*, 2010), water quality simulation software remains much less used.

Sensors for water quality monitoring in DWDS – small in size and able to communicate in short distance – have been developed recently, benefiting from the advances in wireless communications and electronics (Akyildiz *et al.*, 2005). These sensors differ according to the type of parameter they measure: On one hand, low-cost, low-power and multi-functional measure physicochemical parameter such as conductivity, pH, redox potential, diffusion, chlorine concentration, turbidity, UV absorbance (Storey et *al.*, 2010). On the other hand, broadband biosensor systems measure the toxicity of water by monitoring the agility of microorganisms highly sensitive to the presence of toxic ingredients in water (Tecon *et al.*, 2008). Due to their high cost, biosensor systems will not be considered within the SMaRT-Online^{WDN} project.

Data mining in sensor networks has lately received considerable attention because of the increasing amounts of data stored in databases, data warehouses or other repository information (Han and Kamber, 2001). Data mining may involve integration of techniques from multidiscipline such as database technology, statistics, machine learning, neural networks, information retrieval etc. (Hand *et al.*, 2001). Aside from these techniques, model and inference considerations, visualisation and online updating (for models, GIS etc.) are topics of particular relevance to the field of DWDS modelling.

2. Aims of the Online Security Management Toolkit

According to (Piller *et al.*, 2013) and for online management, the main innovative tasks of the SMaRT-Online^{WDN} Project are listed below:

1) Detection of abnormal events with a binary classifier of high accuracy. How to differentiate from non reported but normal change in the system?
2) Generation of real-time, reliable (i) flow and pressure values, (ii) water quality parameter values of the whole water network; which are physical phenomena to consider?

3) Consideration of water quality measurement for online calibration of the hydraulic model;
4) Semi-automatic aggregation of complex networks;
5) Semi-automatic update and online-calibration of the model;

Another important task for the project but to consider in the offline concept is using risk analysis methods for risk assessment and impact evaluation. It will no longer discuss in this report.

3. System Architecture and Workflow of the Online Security Management Toolkit

4.1. General View

In order to detail the data flows that have to be considered to reach the aims set in Part 2, the general system architecture has been designed and is schematically presented in Figure 1.
Various components, existing or not, hardware or software, have been assembled and the main data flows passing from one to another have been represented. This figure does not represent the data entered by the end-users to control the online security management tool (via a GUI for example). This will be presented in Part 4.
The components taken into account in the system architecture are listed below. The components to be developed are written in bold characters:

1) Sensors: considered here as completely independent, they are only sending regular data at a – relatively to all other data flows of the considered architecture – high frequency;
2) Supervisory Control And Data Acquisition (SCADA) system: receives and dispatches data possibly to other sites. The effective control of the network (actions on valve, pump and other devices) is not represented as it is beyond the scope of the SMaRT-Online^{WDN} project;
3) **Data treatment**: Software function, which (i) "cleans up" the raw data retrieved by the SCADA system and (ii) delivers the appropriate (select, recalculate etc.) data for subsequent tasks;
4) **Online Model (hydraulics and quality):** Software tool, which reproduces the behaviour of the DWDS. As its position in Figure 1 suggests, this component plays a central role and its accuracy will be crucial; slow transient model will be considered for hydraulic prediction in place of extended period simulation model. For the transport of usual physico-chemical products the dispersion phenomenon will be added and non-perfect mixing laws at nodes will apply for T and cross junctions. This should be regularly (e.g. 5 min) updated from some data measurements and the demand should be calibrated online.
5) **Database**: Structured collection of data that are stored at different frequency rates according to their type (treated measurements from the sensors, events, model settings etc.). As the database may not grow indefinitely, a regular purge should be scheduled. The corresponding data flow has not been represented on;
6) Geographic Information System (GIS): Software tool already existing mainly used in the proposed architecture as a display tool, possibly offering GUI possibilities;
7) **Event detection algorithm**: Software function, which determines if an event has occurred in the DWDS. Basically, the algorithm compares treated data series

(representing the current state and previous states) issued from the sensors to the results given by a statistical model. Previous data shall also be extracted from the database (previous states on a larger timescale, previous events etc.).

8) **Source Identification**: this module will use velocity recorded in the database and a reverse transport model to determine the mix and the spread of a detected event (contamination of water quality for example) and its evolution in the future.
9) **Statistical metamodel (for water quantity and quality at sensor network)**: Software function, which reproduces specific behaviours of the DWDS at sensor locations. It shall be designed to be less generic than the model presented before, targeting only the some parameters that may be affected by one of the event to be detected by the management toolkit. It shall be continuously "fed" with fresh data extracted from the database. For example, an Artificial Neural Network (ANN) may be used;
10) **Optimisation algorithm**: Software function, which finds optimal model settings in the sense of minimising the gap between observations (data issued from sensors) and predictions (data issued from the model) to calibrate the model.

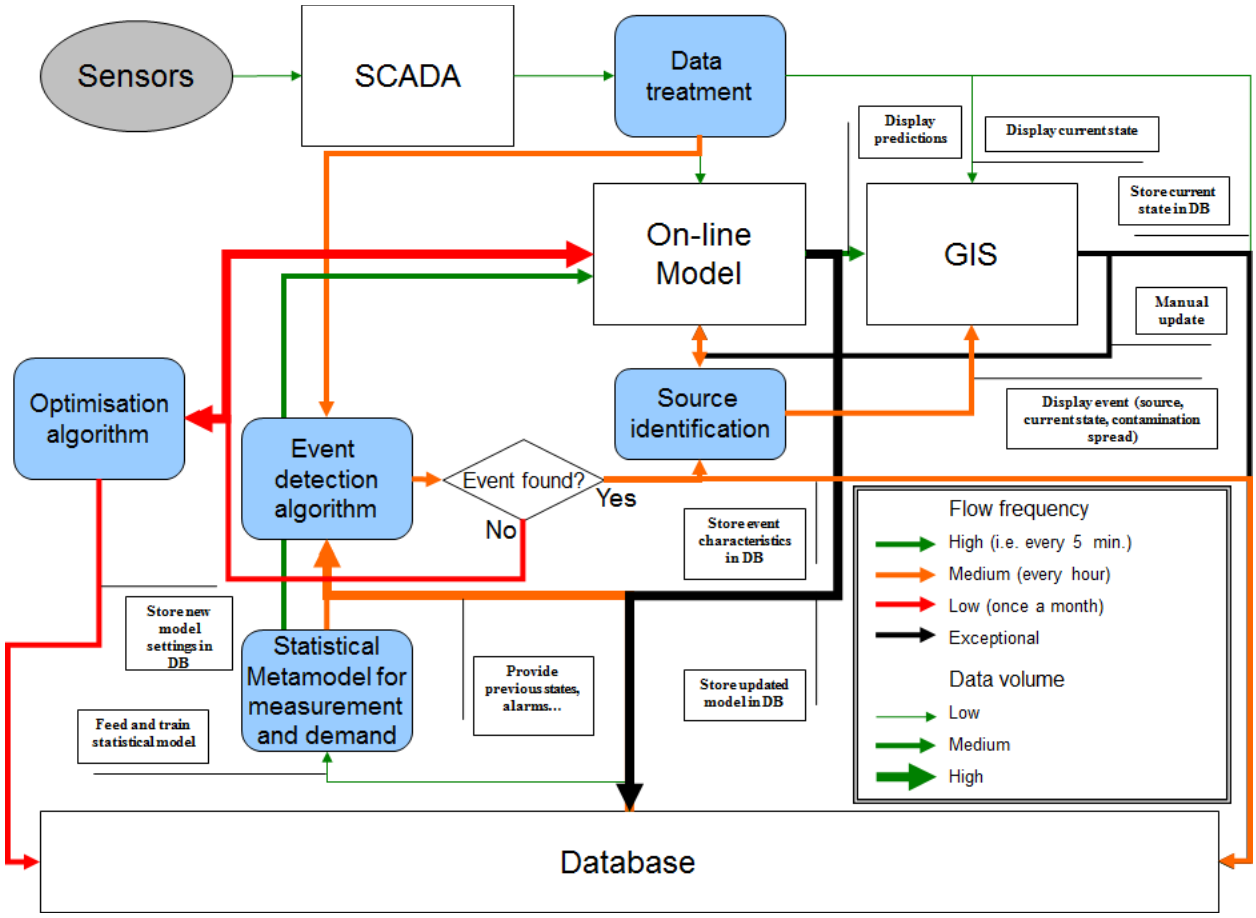

Figure 1: General system architecture and workflows

4.2. Detailed Workflows

We review in the following section the main characteristics of each workflow (and give some methods that could be used) and detail the data flows (content, type, size, frequency). Also, the data inputs from the operators via GUI will be mentioned.

High-frequency workflows

Figure 2 represents the workflows that have to be routinely applied to every data batch. The frequency shall depend upon the sensors settings and eventually on the frequency of acquisition.

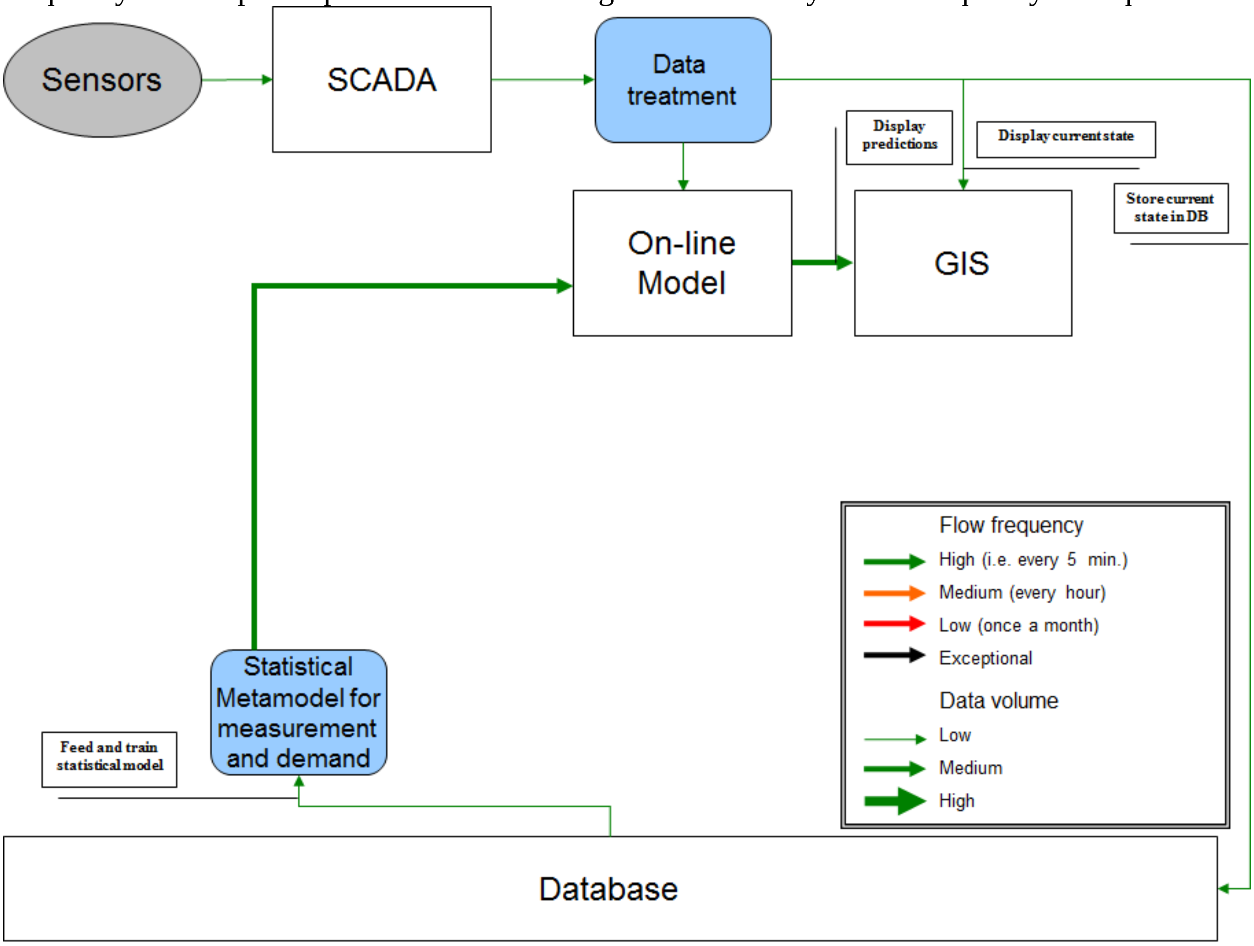

Figure 2: System architecture for high-frequency workflows

Data Treatment

The software tool to be developed should fulfil three tasks:

1) Basic data treatment:
    a. eliminate gross errors (such as negative concentrations etc.)
    b. filter (removing high frequency inputs, which may alter the signal)
    c. interpolate (recreating missing data)
2) Data reconciliation: some data should be used as inputs to the model in order to generate up-to-date hydraulic and water quality predictions every 5 minutes (for the current state and for the coming hours for example). If some data are inconsistent, for example water levels in the tanks and pressures in the DWDS, they should be corrected in some way not to lead to a crash of the model;
3) Possible intermediate calculation if the model needs inputs that differ from treated data; this may not be the case for display purposes in the GIS.

The treated data should be directly readable by the model and the GIS, so the data format must be specified.

Online Model: Hydraulic and quality

Hydraulic and water quality models should be fed by treated data at high frequency (for instance every 5 minutes). Water demands in the hydraulic model will be obtained directly by online calibration to match measurements in the system. Some measurements will substitute the predictions while other will be used for demand calibration. The overall defines the online model.

The online hydraulic and quality model may be run in several modes. The first one, or reference mode, represents the current state of the systems and include a look back in the past. A second one, or look ahead mode, will give point and confidence interval predictions for the next hours. It will use the demand forecast (from the statistical metamodel). A last one, or what if mode, will permit testing different control actions and their impacts onto the system.

Statistical metamodel: water quantity and quality

This software function will use some machine learning method to model an updated “normal” behaviour of the DWDS. In particular, the outputs of this model should be particularly impacted by the events sought to be detected (for example, conductivity if saline intrusion should be detected etc.). Data coming from the database should regularly “train” (calibrate) the metamodel; other data may be used for validation purposes. The frequency rate, at which data is brought to the model, may vary, according to the type of model selected, but it should be relatively high if both calibration and validation are desired. The model should continuously be based on a “moving” amount of data (i.e., last current month). This amount should be: (i) relatively large in order not to be influenced by a recent event not being detected yet; (ii) and relatively small to be representative of the current state of the DWDS. A similar software tool will be developed for water demand prognosis purpose.

Medium-frequency workflows

Figure 3 represents the workflows that have to be routinely applied to data batch at a “medium” frequency (every hour, for example). The frequency shall depend upon the frequency of acquisition and the end-users.

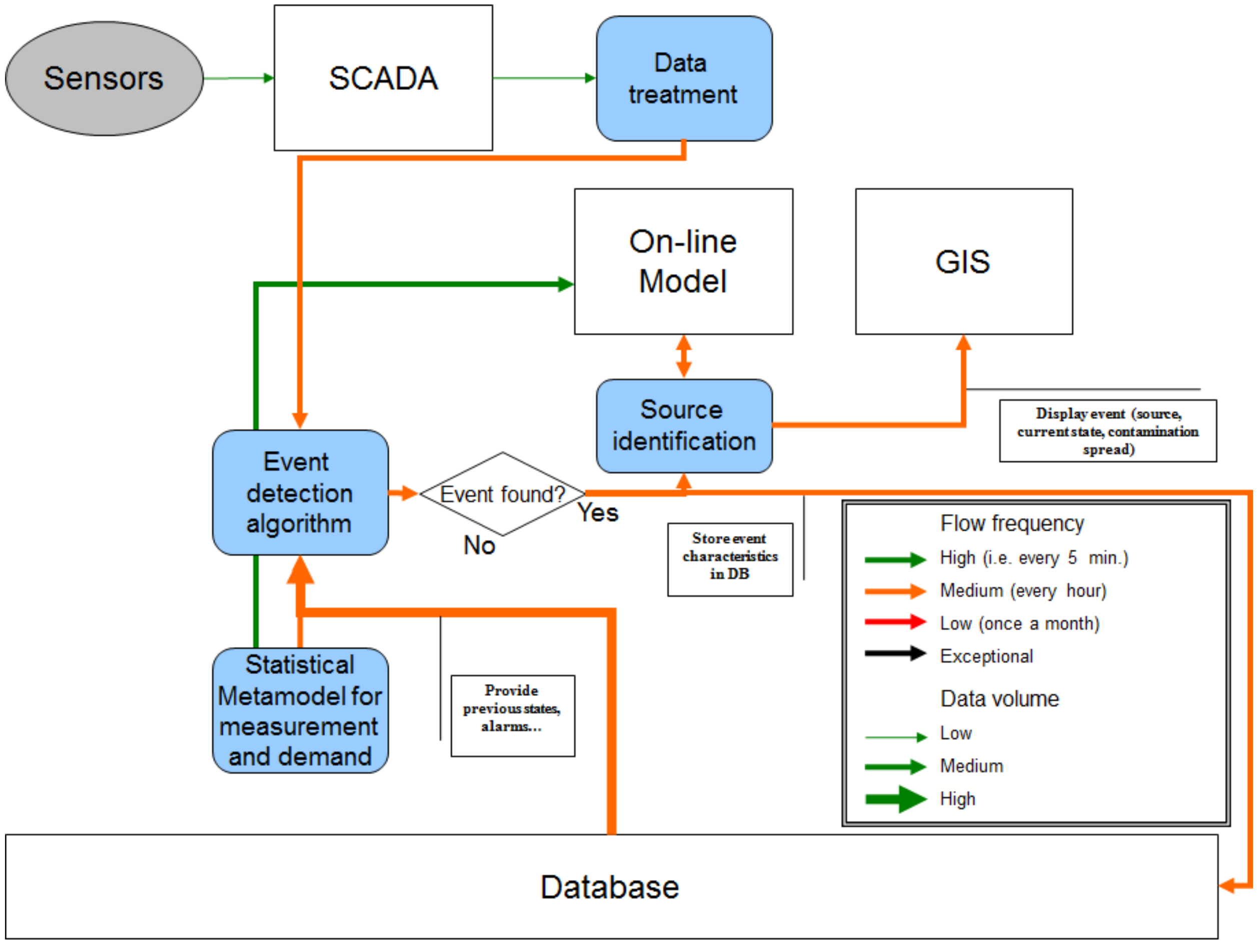

Figure 3: System architecture for medium-frequency workflows

Event detection algorithm
Based on the comparison of treated data with the predictions from the statistical metamodel, the detection algorithm should decide whether an event has occurred on the DWDS. The amount of treated data transmitted should represent a longer period than the acquisition period in order to be representative of an event. Many algorithms in the field of error detection might be used for that purpose. The statistical metamodel should also deliver results with uncertainty levels for the event detection algorithm to distinguish between natural variations in the data and variations related to an event.

Source Identification
Once an event has been detected, a specific module should be launched to evaluate the potential contaminant sources by backtracking in reverse time. Thanks to online modelling and storing of the velocities every 5 min source identification will be more robust and less sensitive to demand uncertainty.

The future impact of the event shall also be simulated by using the statistical metamodel for the demand forecast and the transport model in look ahead mode. All the output data regarding the event, its possible past (source) and future (impact) should be exportable to the GIS (through data tables) and finally stored into the database.

Low-frequency workflows

Figure 4 represents the workflows that have to be routinely applied to data batch at a "low" frequency (once a month, for example). The frequency shall depend upon the end-users.

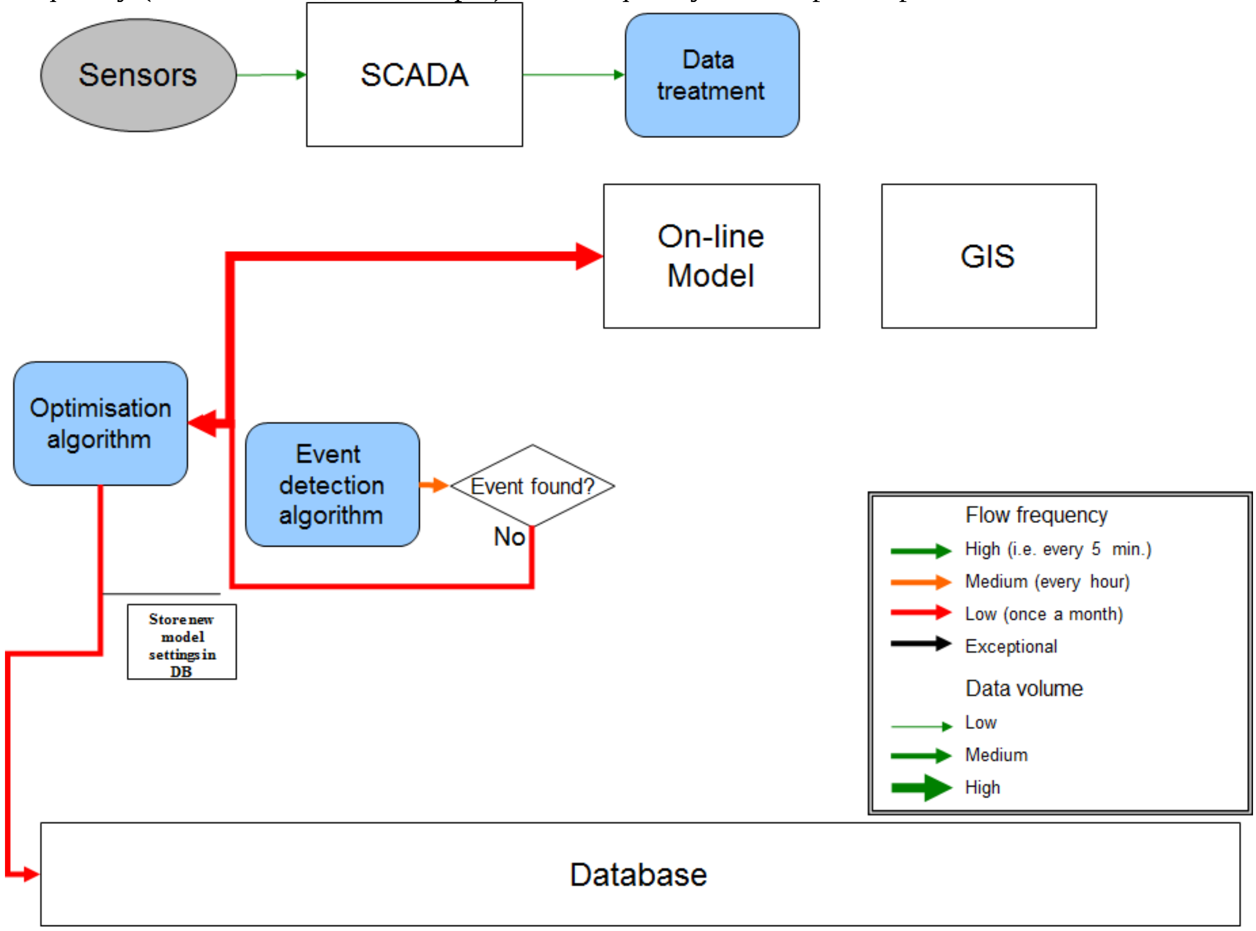

Figure 4: System architecture for low-frequency workflows

In case of no event on treated data found by the event detection algorithm, these data shall be used to recalibrate the models (hydraulics and or quality) and to feed the optimisation algorithm.

Optimisation algorithm

Based on observed data considered as "safe" (not having been considered as representative of an event), and with a determined frequency (once a month, for example), the model will be automatically recalibrated. Communication function shall be developed in order to: (i) modify the input files of the model according to the variable values determined by the optimisation algorithm; (ii) get the model output values to calculate the objectives and constraints. Despite its routine character, the optimisation procedure might require some user inputs, such as:

- The definition of the optimisation problem (and especially the objective or particular constraints);
- The selection of a configuration found optimal by the algorithm (in case of a multi-objective problem, one of the points belonging to the Pareto frontier for example).

Exceptional workflows

Exceptional workflows are manual updates. As such, they will not be detailed in this section. Automatic back-ups to the database should however be scheduled with every manual update.

4. Proposal for Software Implementation and Interfaces

4.1. Characteristics of existing architecture modules

The answers given by three end users (CUS, BWB and VEDIF) to a questionnaire on the existing elements of the online security management toolkit are summarised in Table 1.

Table 1: Main characteristics of the existing elements of the online security management toolkit for the end users

	End user		
	CUS	**BWB**	**Vedif**
Sensors			
Measurement frequency	30 s?	1 min?	5 min
SCADA			
Online data format	OPC-DA OPC-HDA	OPC	OPC Modbus
Data available			
Hydraulic data	Yes	Yes	Yes
Water quality	Yes	Yes	Yes
State control devices	Yes	Yes	Yes
Pump system	Yes	Yes	Yes
GIS			
Software	Elyx Oracle	MAP-Info	ArcGis Desktop
Communication tool	Elyx-WEB FME	Arc-GIS	Vitria FME
Data format			
Raster	Yes: ECW, TIFF	Yes	Yes: Ortho photo plan IGN France Raster
Vector	Yes: ASC, SHP, MIF, MID, DWG, DXF	Yes	Yes: Navteq Data Atlas
Model			
Name	Porteau 3.7	Epanet 7.5 / Stanet	Synergee 3 & 4
Online calibration	No	Yes	No

Some existing characteristics are shared by all the end users, they have been specified in green cells in Table 1; the characteristics that differ from each other have been specified in red cells.

4.2. Characteristics of the data flows of the online security management toolkit

Based on the proposed system architecture presented in Part 3 and on the answers to the questionnaire presented in the previous section, a summary of the characteristics of the data flows within the system architecture is proposed in Table 2. If the characteristic is already that of the existing structure, the cell background is green. Particular attention should be paid to the items written in cells with yellow backgrounds (those related to the GIS and the model).

Table 2: Main characteristics of the data flows

Data flow	**Characteristics**			
	#1	**#2**	**#3**	**#4**
Data treatment				
Input	OPC format			
Output to model	Treated data	"Reconciliated" data	Possible intermediate calculation	Model-specific format
Output to GIS	Treated data	GIS-specific or compatible format		
Output to Database	Treated data			
Statistical model				
Input from database	Regular and formatted	Frequency dependent upon type of statistical model		
Output to event detection algorithm	Formatted			
Event detection algorithm				
Input from data treatment	Treated data	"Reconciliated" data	More than a single batch	
Input from database	Formatted	Relative to event history		
Input from statistical model	Most likely simulated data	confidence intervals		
Output to GIS	Characteristics: past, present, future	GIS-specific or compatible format		
Output to database	Formatted			
Optimisation algorithm				
Input from event detection algorithm	"Save" data (no event)			
Input from model	Output files of model	Model-specific format		
Output to model	Input files of model	Model-specific format		
Output to database	Formatted			

5. Conclusion

Here are the main conclusions:

1) A general system architecture is proposed with the following modules (hardware or software tools): Sensors, Supervisory Control And Data Acquisition (SCADA) system, Data treatment, Model (hydraulics and quality), Database, Geographic Information System (GIS), Event detection algorithm, Statistical model (transport and quality), Optimisation algorithm;
2) The new modules to be developed are: Data treatment, Online transport model (it include the online demand calibration), Database, Event detection algorithm, Source identification, Statistical model, Optimisation algorithms for offline calibration;
3) Aside from the development of the new modules, the following modules, particular attention should be paid to the input/output data flows of the following elements: GIS and model (hydraulics and quality).

Further developments include: All data flows related to user inputs should be specified; in particular their interconnection with the data flows presented herein should be specified.

6. Referred bibliography

Akyildiz, I. F., Wang, X., Wang, W., (2005). „Wireless mesh networks: a survey" Computer Networks Journal, 47, (445–487).

Graymann, W. M. and Clarck, R. M., (1990). "An overview of water quality modeling in a distribution system", Proceedings of the 1990 Annual Conference of the American Water Works Association, Cincinnati, Ohio, USA (81-92).

Han, J. and M. Kamber, (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann, USA.

Hand, D. J.. Mannila, H. and Smyth, P., (2001). Principles of Data Mining. MIT Press, Cambridge, MA, USA.

Machell, J., Mounce, S. R. and Boxall, J. B., (2010). "Online modelling of water distribution systems: a UK case study" Drinking Water Engineering, 3, (21-27).

Ormsbee, L. E., (2006). "The History of Water Distribution Network Analysis: The Computer Age", Proceedings of the 8^{th} Annual Water Distribution Systems Analysis Symposium, Cincinnati, Ohio, USA, August 27-30.

Ostfeld, A., (2005). "A review of modeling water quality in distribution systems" Urban Water J., 2, (107–114).

Piller, O., Gilbert, D., Sedehizade F. , Lemoine C., Sandraz A.C., Werey C., Weber J.M., Deuerlein J., Korth and Bernard T. (2013). " SMaRT-Online^{WDN}: Online Security Management and Reliability Toolkit for Water Distribution Networks", Proceedings of the WISG 2013 Conference, Troyes, France, January 22-23.

Storey M. V., Van Der Gaag B., Burns B. P., Advances in on-line drinking water quality monitoring and early warning systems, Water Research, doi 10.1016/i.watres.2010.08.049

Tecon, R., van der Meer, J. R., (2008). "Bacterial biosensors for measuring availability of environmental pollutants", Sensors, 8, (4062–4080).