AQUAculture USEr driven operational Remote Sensing information services

Deliverable 2.6

Data policy, data communication and access protocols

NIVA, WI, PML, FFCUL, GRAS, DHI, SGM

2014-04

AQUA-USERS is funded under the European Community's 7th Framework Program (Theme SPA.2013.1.1-06: Stimulating development of downstream services and service evolution, Grant Agreement N° 607325)

Task 2.6:
Data policy, data communication and access protocols

Deliverable 2.6:
Data policy guidelines report

Lead beneficiary	NIVA
Contributors	NIVA (5), WI (1), PML (3), FFCUL (4), GRAS (6), DHI (7), SGM (8)
Due date	30/04/2014
Actual submission date	30/04/2014
Dissemination level	PU

Change record

Issue	Date	Change record	Authors
0.1	07/04/2014	Initial outline	TDA (NIVA)
0.2	10/04/2014	Initial draft	TDA (NIVA), KP (WI)
0.3.	14/04/2014	Draft included safe data exchange protocols	NR (WI)
0.4	21/04/2014	Updated second draft included issues addressed in separate telecom	TDA (NIVA)
0.5	29/04/2014	Third draft with inputs from partners	TDA (NIVA), PML, GRAS, VU/VUmc
1.0	30/04/2014	Final version included inputs from remaining partners	TDA (NIVA), WI, FFCUL

Consortium

No	Name	Short Name
1	Water Insight BV	WI
2	Stichting VU-VUMC	VU/VUmc
3	Plymouth Marine Laboratory	PML
4	Fundação da Faculdade de Ciências da Universidade de Lisboa	FFCUL
5	Norsk institutt for vannforskning	NIVA
6	DHI GRAS	GRAS
7	DHI	DHI
8	Sagremarisco-Viveiros de Marisco Lda	SGM

To be cited as

Dale, T., Huber, S., Poser, K., De Reus, N. (2014) “Data policy guidelines report”, AQUA-USERS deliverable D2.6, EC FP7 grant agreement no: 607325, 13p.

Task objective (from DoW)

Establish data policy that ensure protection of sensitive information from each of the involved users, but also allows partners and users to cooperate to define and implement methods and new sciences.

Scope of this document

This document describes the general data policy within the AQUA-USERS. The document furthermore provides a data management plan, a protocol for safe exchange of data within the project, and guidelines for distribution of data to third party.

Abstract

AQUA-USERS is strongly user driven to ensure sustainable and user-relevant services after the project. A pivotal part of the AQUA-USERS project is the collection and integration of in-situ data into the database and application. It is important to identify the level of sensitivity associated with each type of data and information that might be sensitive and what kind of data and information that can be exchanged without restrictions within the consortium. It is therefore important to define a data policy that allows for partners and users to efficiently cooperate in defining and implementing methods and new services within the project while at the same time ensure the protection of sensitive information. This document contains a brief description of the process towards a data policy, where a main part of the process was to ask users (online questionnaire) to classify their data into one of three levels of sensitivity. The document states the data policy for AQUA-USERS which is a guide to help generate, maintain and safeguard high-quality data, and to share and gain access to data within the project according to levels of sensitivity agreed upon by project consortium. Finally this document includes a data management plan that facilitates the implementation of the data policy.

List of abbreviations

Abbreviation	Description
DoW Part A	Annex I. Description of Work
D	Deliverable
HAB	Harmful algal bloom

List of related documents

Short	Description	Date
D 7.3	SLA s with users	28/02/2014
D 7.4	Initial exploitation & business plan	28/02/2014
DoW	Grant Agreement Annex I (Description of Work)	01/07/2013
CA	Consortium agreement	24/05/2013

Table of contents

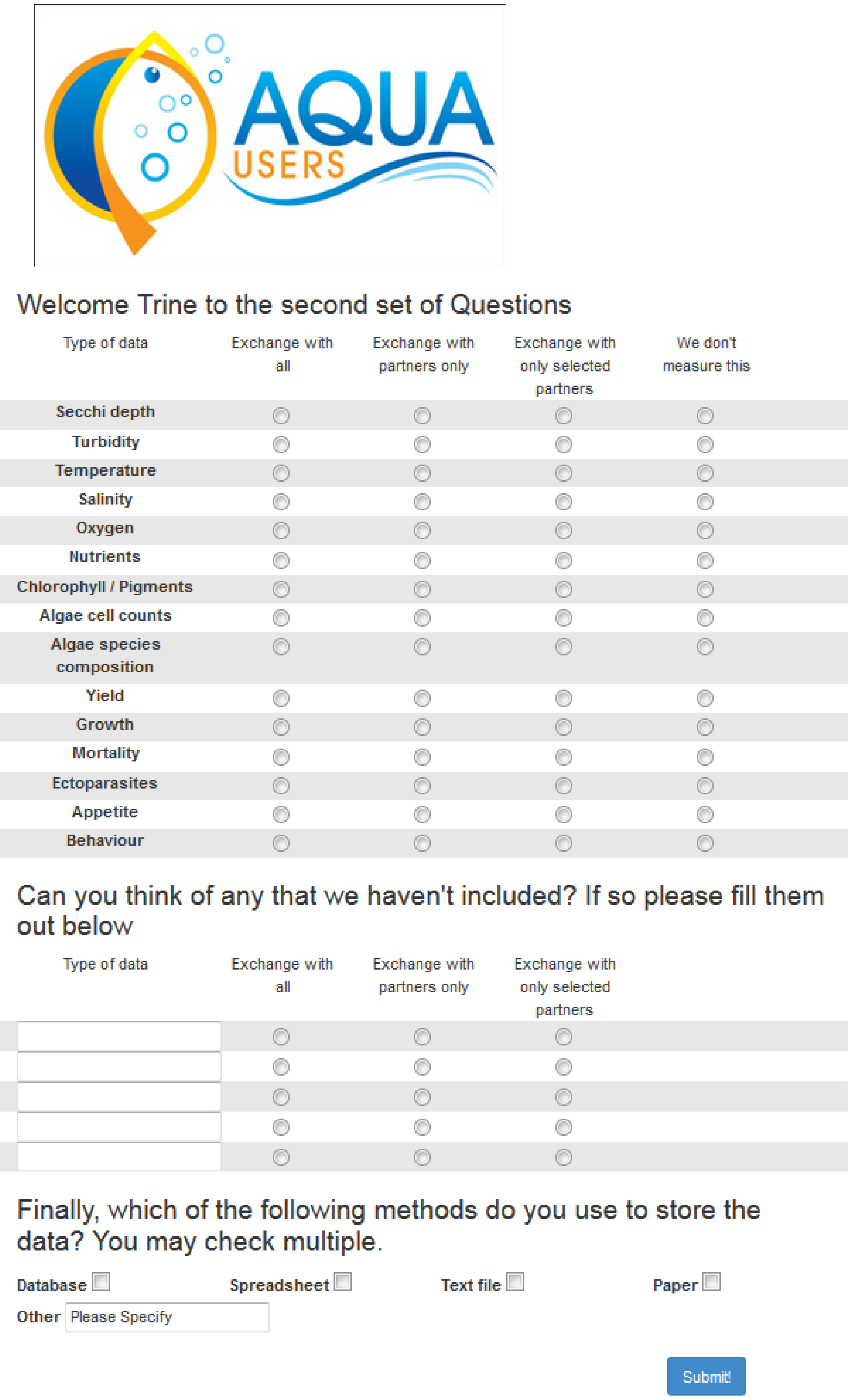
1 Introduction

AQUA-USERS will be strongly user driven to ensure sustainable and user-relevant services after the project. A pivotal part of the AQUA-USERS project is the collection and integration of in-situ data into the database and application. In close collaboration with the users, in-situ data will be collected at the users' production sites during the project period. These data include WISP-3 measurements, Secchi disc depth, cell counts, concentrations of pigments, solids and colored dissolved organic matter, data on phytoplankton composition, data on environmental physical conditions (temperature, oxygen levels etc.) as well as the actual response of the aquaculture species (e.g. mortality, growth, yield, and fish behaviour) produced. Furthermore, whenever available, historical data from the users’ sites will be used to develop indicators for e.g. HABs. Finally, partners may submit data either previously collected or collected during the project. It is important to identify the level of sensitivity associated with each type of data and information that might be sensitive and what kind of data and information that can be exchanged without restrictions within the consortium.

2 General data policy

2.1 Process towards a data policy

The data policy of AQUA-USERS is developed through an open process within the consortium. Data policy was addressed already at the kick-off meeting, where a general discussion took place. Issues raised by partners at the kick-off meeting has been followed up in two of the monthly telecons, and in one separate telecon specially assigned to data policy discussions. It is very important for the success of the projects to establish policies and protocols that ensure sufficient protection of sensitive information while at the same time ensure smooth cooperation and exchange of data in the project. Regarding user collected data, the general discussion at the kick-off meeting was followed up by a questionnaire distributed among the users (Fig 1). The purpose of the questionnaire was to identify the levels of sensitivity associated with the different types of data, and the users were asked to identify which types of information and data they consider sensitive, and what kind of information and data can be openly exchanged within the consortium. In the questionnaire every type of data to be collected/used in the project were classified into three categories indicating their level of sensitivity; “exchange with all consortium”, “exchange with partners only” and “exchange with only selected partners”. The classification will follow the data though the project.



Welcome Trine to the second set of Questions

Type of data	Exchange with all	Exchange with partners only	Exchange with only selected partners	We don't measure this
Secchi depth	○	○	○	○
Turbidity	○	○	○	○
Temperature	○	○	○	○
Salinity	○	○	○	○
Oxygen	○	○	○	○
Nutrients	○	○	○	○
Chlorophyll / Pigments	○	○	○	○
Algae cell counts	○	○	○	○
Algae species composition	○	○	○	○
Yield	○	○	○	○
Growth	○	○	○	○
Mortality	○	○	○	○
Ectoparasites	○	○	○	○
Appetite	○	○	○	○
Behaviour	○	○	○	○

Can you think of any that we haven't included? If so please fill them out below

Type of data	Exchange with all	Exchange with partners only	Exchange with only selected partners
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○

Finally, which of the following methods do you use to store the data? You may check multiple.

Database □ Spreadsheet □ Text file □ Paper □

Other Please Specify

Figure 1: Questionnaire used to identify levels of sensitivity of different type of user collected data

2.2 Data policy

Purpose

The purpose of the AQUA-USERS data policy is to arrange for that partner and users can efficiently cooperate to define and implement methods and new services in the project while at the same time ensure the protection of sensitive information. The AQUA-USERS data policy is a guide to help generate, maintain and safeguard high-quality data, and to share and gain access to data within the project according to levels of sensitivity agreed upon by project consortium.

Definitions and clarifications

This data policy concerns data collected during the AQUA-USERS project. The data policy also concerns already existing data that are made available for AQUA-USERS but not listed as Background in the Consortium agreement Attachment 1.

For the purpose of this policy “data” is interpreted as observational data, and data derived from analysis independent of format.

Data management and sharing of data within AQUA-USERS

- Project partners shall be responsible for the quality (according to D 5.1 in DoW), and completeness of data to be submitted to the AQUA-USERS database.

- Both project partners and users are responsible for description of the data, metadata, and associated products submitted to the AQUA-USERS database (see 2.3 Data Management Plan).

- Both project partners and users are responsible to transfer data to the AQUA-USERS database according to methods of transfer developed in D.5.1. in DoW.

- When data is being transferred to the AQUA-USERS database, data deemed sensitive or privileged must *be identified and appropriately labelled* (see 2.3. Data Management Plan).

- After the data and metadata are deposited with the AQUA-USERS database, WI assumes responsibility for the archiving, persistence of, and access to, the data, metadata, and ancillary holdings. For data security all data should be backed up with the users or on partner institution servers.

Sharing of data outside AQUA-USERS

- During the project period data is only intended for in-house use by the AQUA-USERS consortium and shall not be distributed, disclosed, or transferred to other parties without permission from data originator.

- For publications and other products based on the AQUA-USERS data and information, acknowledgements of the source shall be made and the service provider shall be supplied with a copy of /URL of these.

- At the termination of the project, data collected during the project will after a re-assessment of sensitivity level, be released into the public domain by submission into a relevant open-access database. The release of data will have a time delay of 3 years after termination of the project.

2.3 Data management plan

A data management plan facilitates the implementation of the data policy. When data is well managed, they are easier to find, to understand and to access. Data loss is prevented though security and backup procedures, and protection of sensitive information are facilitated. The here presented data management plan is not a fixed document but will evolve during the project as the work with other closely related deliverables concerning quality control (D5.1) and ingestion of data into the database (D5.4) are in progress.

Each data input should have its own table. For already existing datasets the template shown in table 2 should be used, while for data collected during the project table 3 should be used.

Table 1. Data inputs-Existing data collections.

1	**[Name of collection]**
Description:	Describe the data
Format:	Identify the formats in which the data are maintained and made available
Quality checks:	Specify procedures used to evaluate the existing data (e.g. verification, validation and assessment of usability)
Source	Identify the source of the data
Metadata:	Identify metadata standard that are used to describe the document
Volume estimate:	Volume of information
Backup & Storage:	Describe the approach to backup and storage of the information during the project
Access & Sharing within the project:	Specify who should have access to data and what type of access; • **Users**-specify one of the following levels of sensitivity; “exchange with all consortium”, “exchange with partners only”, “exchange with partner NN”. • **Partners**-specify “exchange with all consortium” Specify limitations on type of use; • **Partners and users**- specify one of the following; “no limitation”, “use in publications only after explicit permission of the data originator”
Citation	Specify how the data should be cited

Table 2. Data inputs-New data collections within AQUA-USERS.

2	[Provide a brief name to describe new data collection]
Description:	Describe the data
Format:	Identify the formats in which the data will be generated and made available
Protocols:	Identify any standard protocols or methodologies that are used to collect the data if possible
Quality checks:	Specify procedures for ensuring data quality (D 5.1)
Metadata:	Identify metadata standard that will be used to describe the document
Volume estimate:	Volume of information
Backup & storage:	Describe the approach to backup and storage of the information during the project
Access & Sharing within the project:	Specify who should have access to data and what type of access; • **Users**-specify one of the following levels of sensitivity; “exchange with all consortium”, “exchange with partners only”, “exchange with partner NN”. • **Partners**-specify “exchange with all consortium” Specify limitations on type of use • **Partners and users**- specify one of the following; “no limitation”, “use in publications only after explicit permission of the data originator”
Citation	Specify how the data should be cited

3 Protocols for safe data exchange

The resources which together constitute the infrastructure of the AQUA-USERS are distributed among a number of partners, and there is an undetermined number of client endpoints. To enable safe data exchange between the resources inter se and between the resources and the clients, we must have in place some measures that address each of the following five points:

1) Establish that the server is indeed the resource provider it claims to be.

2) Establish that the client is a user with permission to access the requested resource.

3) Protect against eavesdropping on login details or data en route.

4) Protect client authentication credentials at the endpoint.

5) Protect against loss of traffic such as might lead to data corruption between sender and receiver.

Ad 1.

The identity of servers must be guaranteed by using digitally signed certificates. By certificate in this context, we refer to a document encoded according to the X.509 certificate standard as specified in RFC 5280 of the IETF. Each certificate should be traceable through its signatures up to a trusted root certificate, e.g. Thawte, VeriSign, DigiCert, or other major internet authority. Self-signed certificates will be acceptable during testing, but should be replaced with proper certificates for the production environment.

Ad 2.

To ensure that clients can access only resources to which they are allowed, a login mechanism should be implemented wherein the user can provide an identity and associated password (or key challenge) to the server, after which the service should evaluate whether to accept the user's request. This form of authentication may be implemented at either the session level, or for each individual request, as is deemed appropriate for the particular service.

The option of using client X.509 certificates as well so that servers can limit their service to a select number of known and identity-assured clients shall remain at the discretion of each implementing party. While this technique would be more secure than password or key authentication, the process of generating such certificates is tedious, and few places in the infrastructure were identified where one-to-one or one-to-few relationships exist in which such an arrangement remains manageable.

Ad 3.

All services which are by design presented as web services following the server-client model can be secured with the standard Secure Sockets Layer (SSL) technology. This technology establishes endpoint to endpoint encryption between the web server and the client, thus protecting data such as login credentials, requests and response data from being readable to a third party observers on any network node along the route. The current design schematic of the AQUA-USERS architecture does not feature services of any other type, so the first statement covers all planned services. Should the need arise, any services that are designed to use raw data connections over an outside line too can encrypt the connection between their sockets using similar, though less standardized means. Transporting raw datastreams over anything but an internal connection is not recommended.

Ad 4.

Data which is important to authentication, such as passwords, should never be in plain text. They should be stored instead as encoded data-hashes using at least the SHA-1 algorithm. If the application owner requires even more security, SHA-256 may be considered a safe option, as to date there has not even in theory been any potential security space collision found for this encoding.

Ad 5.

Loss of data en route is countered by relying on the Transmission Control Protocol (TCP) rather than the User Datagram Protocol (UDP) at the transport layer, which removes the need for transport error checking and the validation of packet delivery (at the cost of increase of overhead traffic and latency). The prior work done by the IETF work group in designing the characteristics of the TCP we consider sufficient for the purpose of this project.

We recognize that even with all of this will not be possible to make the infrastructure entirely safe -- many services will need to run automated, which means that there must either exist unconditional trust relationships between parts of the infrastructure that need each other's data or services, or that servers must have locally stored script-useable keys and passwords to access each other. Either of these options presents a minor compromise in terms of security, but this is generally considered acceptable.