



# AQUAEXCEL

Aquaculture Infrastructures for Excellence in European Fish Research

Project number: 262336

Combination of CP & CSA  
Seventh Framework Programme  
Capacities

## ***Deliverable D3.5***

### **Experimental data repository.**

**Due date of deliverable:** M48

**Actual submission date:** M48

**Start date of the project:** March 1<sup>st</sup>, 2011

**Duration:** 48 months

**Organization name of the lead contractor:** VURH

**Revision:** V1

Project co-funded by the European Commission within the Seventh Framework Program(2007-2013)	
Dissemination Level	
<b>PU</b> Public	<b>X</b>
<b>PP</b> Restricted to other program- participants (including the Commission Services)	
<b>RE</b> Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b> Confidential, only for members of the consortium (including the Commission Services)	

## Table of Contents

Definitions .....	3
Glossary.....	3
Context .....	4
Summary .....	5
<b>1 Objectives.....</b>	<b>5</b>
<b>2 Rationale:.....</b>	<b>5</b>
<b>3 Teams involved:.....</b>	<b>5</b>
<b>4 Geographical areas covered:.....</b>	<b>5</b>
<b>5 Results:.....</b>	<b>5</b>
System description.....	6
<b>1 Repository description.....</b>	<b>6</b>
1.1 Overall scheme.....	6
1.2 Repository development .....	7
1.3 Data workflow .....	8
<b>2 Database.....</b>	<b>11</b>
<b>3 Access points.....</b>	<b>11</b>
3.1 BioWes – a fat client .....	12
3.2 Application framework .....	13
3.3 Protocol manager .....	14
3.4 Web interface.....	14
<b>4 Protocol designer .....</b>	<b>17</b>
4.1 Standardization.....	18
<b>5 Protocol generator .....</b>	<b>20</b>
5.1 External information.....	22
<b>6 Data processing modules .....</b>	<b>22</b>
<b>7 Technologies.....</b>	<b>23</b>
<b>8 Security.....</b>	<b>24</b>
Use cases .....	25
<b>1 Fish color analysis.....</b>	<b>25</b>
<b>2 Water quality – crayfish.....</b>	<b>25</b>
<b>3 Fish tracking.....</b>	<b>26</b>
<b>4 Bio-compatibility.....</b>	<b>26</b>
<b>5 TNA form.....</b>	<b>27</b>
Documentation .....	28
Conclusion .....	29
Annex 1.....	30

## Definitions

**AQUAEXCEL repository** – the system which consists of the server with Microsoft SQL Server, web interface to access the database, and stand-alone application *Protocol manager* for access of the database from personal computer

**Database** – Microsoft SQL Server installed on the server for storage of protocol templates and protocols

**Protocol template** – electronic version of empty protocol created using *Protocol manager*. It is used for filling information about experiment and attachment of experiment data

**Protocol** – electronic protocol. Filled protocol template with experimental data stored in the database.

**Protocol manager** – module of *BioWes* – fat client for the management of protocols and protocol templates

**Web interface** – webpage for management of protocols and protocol templates stored in the database.

**BioWes** – project for experimental data repository development.

**BioWes – fat client** - stand-alone application that provides the interface to the Database for creating and modification of protocols and protocol templates extendable by specialized modules

## Glossary

**Fat client** – is a computer program that provides rich functionality independent on the server.

**Thin client** – is a computer program that depends heavily on the server.

## Context

AQUAEXCEL has the ambition to be a platform for improvement of the aquaculture infrastructures. One of the aims is to establishing cooperation of research activities between its members who are producing a variety of experimental data and experiments description. The work between and within individual work packages is interconnected to fulfill the networking ambition of the project.

The cooperation between the partners requires exchange (sharing) of the experimental data and description of the datasets (description of experiment - metadata). The project also provides the opportunity to realize the experimental work at the infrastructures under the financing of TNA projects that also produce experimental data. The effective sharing of experimental results needs a standardized terminology, proper description of experiments and the tool for experimental data and metadata management.

The outcomes of the survey (2011) about the common procedures about experimental work management show that preparation and realization of the experiment is at a very good level but the support for experimental data and metadata storage, sharing and searching is at a low level.

The AQUAEXCEL repository is therefore designed as experimental data and metadata storage supported with the tool for data management, standardization, and processing and users cooperation. The repository is available for any AQUAEXCEL member through the *web interface* or *Protocol manager*. The repository incorporates the results produced by the AQUAEXCEL project in the field of ontologies development and uses it for standardization of the terminology in the description of experiments. The repository can be used for cooperation between partners (processing or sharing the experiment results) or just for optimization of the work of an individual researcher because it provides the tools for experimental data and metadata management.

# Summary

## 1 Objectives

The general objective of the Deliverable 3.3 is to develop a repository for storage of experimental data and description of experiments and other tools for the data management. The repository will be available for all AQUAEXCEL members and will be able to store all data types and structures produced by the project. The tools will provide the functionality to create an *electronic protocol*, attach experimental data and manage (organize, share, search) it in user friendly way. The repository will use the outcomes of the WP3 (ontologies) for the support of terminology standardization in experiment description. The repository will be tested on real experiments; and the training, guides and documentation will be created to support the usage of the repository by AQUAEXCEL members.

## 2 Rationale:

The approach, hardware and software chosen to reach the objective was chosen to develop a secured and stable repository able to store any type of data structures with a user friendly support for the repository content management. The survey was realized to identify the needs of the users and create a pilot version of the system for testing. The budget allocated to the task 3.3 was very limited and therefore the financing of the main development was financed by the grant BioWes number TA01010214 of the task leader supported by the Technological agency of the Czech Republic. This support allowed us to develop the system based on the Microsoft SQL Server and operate the server dedicated just for the AQUAEXCEL consortium. The pilot version of the system was tested by the involved partners and by the tester of VURH to meet the required functionality of the system. Re-implementation of the selected parts of the system improved the convenience of the system for daily work. The protocols and data from real experiments were used for the demonstration of the repository functionality.

## 3 Teams involved:

- VURH – the development of the repository
- IMARES
- NOFIMA
- ULPGC
- WU
- SINTEF

## 4 Geographical areas covered:

All Europe – The system is accessible by the tool independent of the operating system (Web interface) or designed for Windows operating system. The system can deposit any type of data and therefore is independent on the particular experiment, institute, and location.

## 5 Results:

The database for data storage is installed on the server operated by VURH. The tools for accessing the database are delivered in the form of software (*Protocol manager*) and web service (*Web interface*) and can be installed from the *BioWes* web page in download section. The examples of different protocol and experimental data were provided by the partners involved in the task. The documentation, video guides are available on the AQUAEXCEL webpage in the section Results-AQUAEXCEL repository.

# System description

## 1 Repository description

This section will describe the AQUAEXCEL repository from a schematic point of view to define system parts and functionality to understand how the experimental data and metadata are managed.

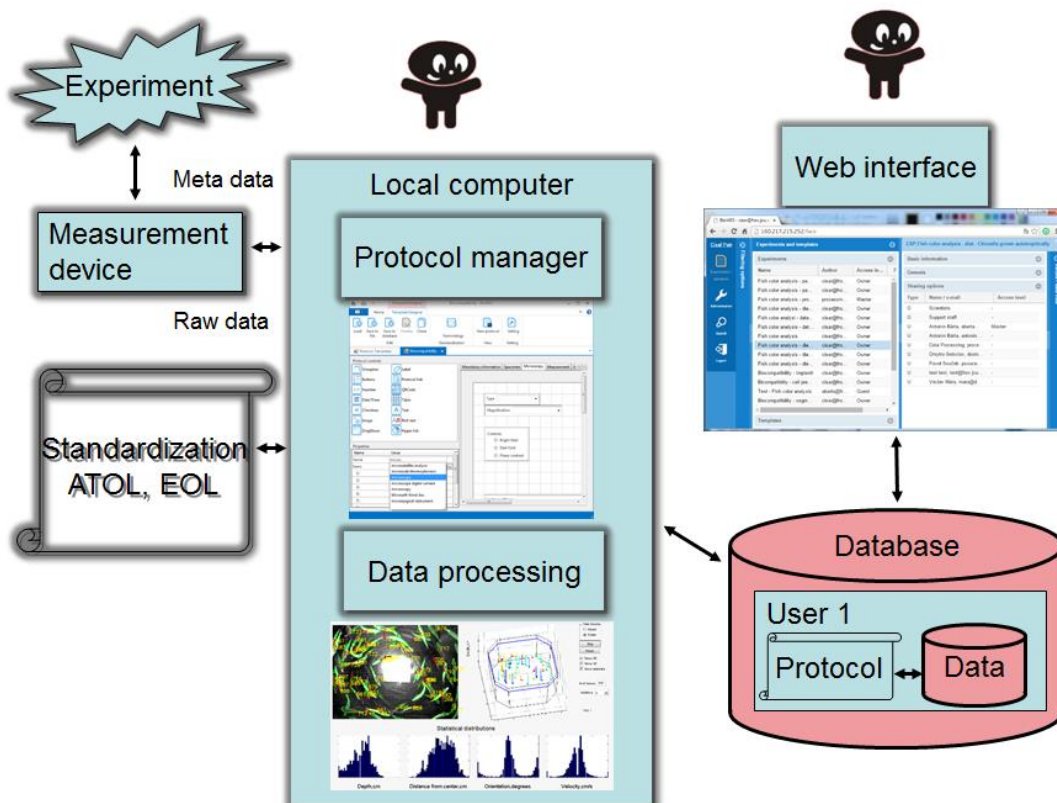
### 1.1 Overall scheme

The system is based on three main pillars. The core of the system is an Microsoft SQL Server installed on the server dedicated just for the AQUAEXCEL repository. The database stores the protocol templates and protocols with the metadata and experimental data. The database also contains the information about user accounts and access rights to the particular protocols.

There are two main access points to the database: A web interface and the stand-alone application Protocol manager. *Protocol manager* is a software implemented in C# .net which is installed on local computer and provides the access to the database plus the functionality for protocol template design, filling protocols and data processing as well as visualization. Protocol manager is designed just for Windows operating system.

*Web interface* is a webpage running on the server with the database that allows the user to create user accounts and manage protocols and protocol templates. The functionality provides the searches, sharing, visualization of metadata, access to experimental data and visualization of protocol genesis (chain of protocols).

Other functionalities, like reading information from measurement devices, standardization and data processing, are realized as modules of Protocol manager, Protocol designer or Protocol generator. Protocol manager has an open framework for plug-ins with the defined interface. Therefore, any third-party modules for data processing or visualization can be used as an extension of the AQUAEXCEL repository.



**Figure 1.** Scheme of the AQUAEXCEL repository

## 1.2 Repository development

Because the repository should serve for the experimental data and metadata storage for the AQUAEXCEL members, it was necessary to realize the survey at the beginning of the project (2011) focused on the need in the different research areas. Eighteen institutes all over Europe (6 AQUAEXCEL members) participated in the survey. The survey was focused on the identification of the practice in experimental data measurement, storage, sharing and technical solutions used to support these activities.

The issues and requirements for the system functionality were identified from the survey.

The development of the repository started in 2012 with the design of the database and tools for filing protocols. The protocols themselves were defined by using *xml* language.

The system for protocol design was found too complicated for the daily usage and the concept was modified (end of 2012). The *Protocol manager* was developed as the tool for protocols management at personal computers. The software had two modules: *Protocol designer* for graphical design of protocol templates and *Protocol generator* for filing the protocols.

*Protocol designer* was extended with a standardization module in 2013 that is able to read the standardized terminology from ontology (*owl*) files. This concept interconnected the activities in the WP3 and allows the users to use standardized terminology.

At the end of 2013, the first protocols were stored in the repository and the system was tested based on real experiments.

The training focused on the use of the repository by an AQUAEXCEL member was realized during the 4th AQUAEXCEL meeting.

The overall system was completely tested in the middle of 2014 by the participants of a summer school organized by VURH. More than 50 people participated at the testing. Recommendations resulting from the tester were collected and the main parts of the system were completely redesigned to improve the user interfaces to be more user friendly.

Modules for data processing and visualization that were implemented (end of 2014) for the specific datasets show the possibilities of the system.

Documentation and examples (testing account) were created to support the usage of the system.

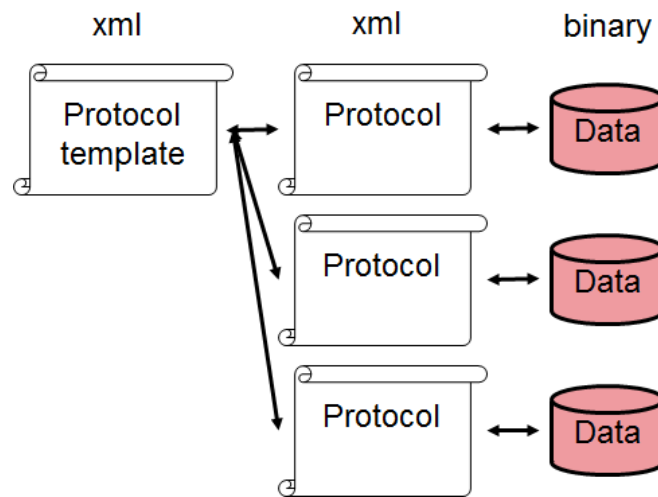
### 1.3 Data workflow

The repository is designed to store all the information about the experiment required for its reproducibility. Therefore, it has to be able to store description of the experiment and experimental data.

The description of experiment should consist of the information about experimental conditions, setting of devices and procedures used for experiment realization (all the parameters which can influence the results of the experiment). The description of the experiment should be understandable to the other researchers to reproduce the experiment. The usage of standardized terminology and standardized form of the information representation improve sharing of the description between cooperators. The description of the experiment is included in the experimental protocol. AQUAEXCEL repository uses two structures for representation of the experimental protocol to separate two kinds of information: protocol design and protocol values. The first structure is “protocol template” and contains description of the protocol appearance. The template is an empty protocol without any values (similar to pre-prepared form for filling) - see Figure 3. It describes the general representation of the experiment (what should be filled to describe experiment properly). The protocol template is represented as an *xml* file with the information about individual components (numbers, text arrays, tables, links), and their placement in the protocol - see Figure 5. This representation allows the system to visualize the protocol in many different ways because the description does not restrict the appearance. The protocol appearance is based only on the particular visualization tool the repository uses (*Protocol generator* in this case). To be able to keep this level of abstraction, the system has to use predefined components for definition of the protocol template. See more about the components in section 4.

The protocol template is designed before experiment realization by the experts in the field. It can be used as a guide through the experiment to identify the method that should be used and information that should or has to be recorded.

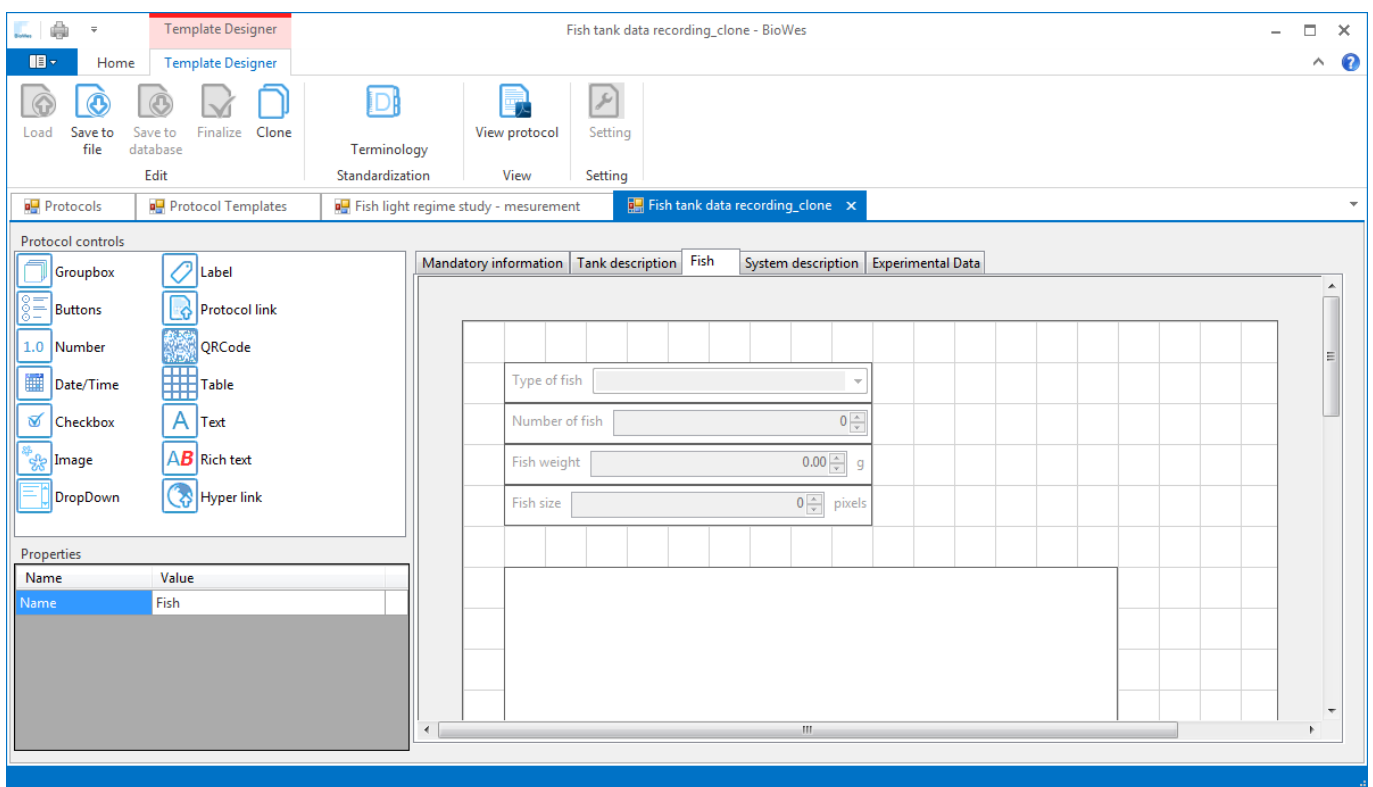




**Figure 2.** Data representation in the repository

The second structure used for storage of experiment description is “protocol” that contains the values describing the particular realization of the experiment - see Figure 4. The protocol has a link to the protocol template for the visualization - see Figure 2. When the user starts the experiment realization, the system uses protocol template for visualization of the protocol. The user fills the protocol by values and these values are stored within the protocol. The protocol is represented by data structure of defined data types given by the protocol template. Each component of the protocol template has defined a data type and unique identifier (unique for given protocol template). The identifier is used for a link between the value and the particular component. Only the information concerning values is stored in the protocol.

This system of protocol template and protocol allows the repository to reuse the protocol template for repetition of the experiments under the same or different conditions. There is only one protocol template, which describes the experiment and many protocols containing information about particular realizations.



**Figure 3.** Examples of protocol template design visualized in *protocol designer*. Protocol template contains the components characterizing the experiment

The screenshot shows the 'Protocol Generator Tools' window. The title bar indicates the current protocol is 'Fish light regime study - measurement - BioWes'. The menu bar includes 'Home', 'Protocol Generator', and 'Edit'. The 'Protocol Generator' menu has sub-items: 'Save', 'Finalize', 'Clear', 'Settings', 'Generate PDF', and 'Fill protocol'. The 'Edit' menu has 'Generate PDF' and 'External'. The main area has tabs for 'Protocols', 'Protocol Templates', 'Fish light regime study - measurement', and 'Fish tank data recording\_clone'. The 'Fish light regime study - measurement' tab is selected, showing a form with the following fields:

- Type of fish: Atlantic salmon (dropdown)
- Number of fish: 100 (spin box)
- Fish weight: 423.00 g (spin box)
- Fish size: 952 pixels (spin box)

Below the form is a large image of a fish. At the bottom is a 'Note' field.

**Figure 4.** Example of protocol template visualized in *Protocol generator* with filled values and images for a particular realization of experiment = protocol.

The components designed for experiment description contained in the protocol template, ensure the standard description of an experiment that can be visualized by any software tool, which can read the *xml* structure. Because of this standardization, the system is able to do the searches based on the description of the experiment. The information contained in the protocol template and the protocol (metadata about experiment), can be used to find a specific experiment with specific conditions (for example an experiment with sea bass under a given temperature).

The second kind of information provided by experimental work is the experimental data itself. The research areas of AQUAEXCEL members cover very broad areas of aquaculture research. The data produced by very different experiments are of a broad variety of data types from numbers to images. The data values are then represented by a variety of data structures (formats) which differ between different measurement devices, devices manufactures or often different versions of the same device of the same manufacture (for example the data format of LC-MS devices). It is impossible to create a general structure for data storage, which is able to represent all types of data. Therefore, the AQUAEXCEL repository stores the experimental data as binary data. The structure converts the experimental data to the block of binary data that are linked to the protocol structure in the database. This allows the system to link the description of experiment and experimental data of any type (see Figure 2). The experimental data are processed as black-box and the system cannot read the values because it does not know the structure of the experimental data. The disadvantage of the solution is that the experimental data cannot be used for queries to search the specific protocol for specific experimental data.

To read the specific experimental data, the system allows the user to use specialized data processing modules, which can be plugged-in into the *Protocol manager* to process or visualize the experimental data. The modules focus on the data with specific structure (images, *xls* files, LC-MS files).

```

<?xml version="1.0" encoding="UTF-8"?>
- <Protocol xsi:schemaLocation="http://Biowes/namespace file://ProtocolGenerator.x:
85b8-f504d0fc295a" Creator="">
  - <GeneralPage DPI="96" Units="mm" Orientation="Portrait" Height="297" Width="211
    <Margins Bottom="10" Right="10" Top="10" Left="10"> </Margins>
  </GeneralPage>
  <Terminology> </Terminology>
  - <Tab>
    <Name>Mandatory information</Name>
    - <Page>
      - <Object Version="1.1" Type="PureTextItem">
        <Name>Name of experimentalist:</Name>
        <LabelPosition>Left</LabelPosition>
        <ID>71de6316-77ee-43a7-b525-bb3194809c67</ID>
        <Text>{\rtf1\ansi\ansicpg1250\deff0\deflang1029{\fonttbl{\f0\fn
        <ReadOnly>>false</ReadOnly>
        <Optional>False</Optional>
        <Width>684</Width>
        <Height>36</Height>
        <External/>
        <Left>36</Left>
        <LeftText>144</LeftText>
        <Up>396</Up>
      </Object>
      - <Object Version="1.1" Type="PureTextItem">
        <Name>Name of experiment:</Name>
        <LabelPosition>Left</LabelPosition>
        <ID>ba6552e7-3c02-4f32-90a5-f8ce30b0ed36</ID>
        <Text>{\rtf1\ansi\ansicpg1250\deff0\deflang1029{\fonttbl{\f0\fn
        <ReadOnly>>false</ReadOnly>
        <Optional>False</Optional>
        <Width>684</Width>
        <Height>72</Height>
        <External/>
        <Left>36</Left>
        <LeftText>124</LeftText>
        <Up>324</Up>
      </Object>
    </Page>
  </Tab>
</Protocol>

```

Figure 5. Example of representation of protocol template as *xml* file. The unique identifier is marked red.

## 2 Database

The database is the central point of the repository, which serves for the storage of experimental data, description of experiments, relationships between protocols, user accounts and sharing rights. The relational database was chosen for the realization of the repository. The database model is described in the form of an SQL script and is applied on the database server Microsoft SQL Server 2012 Enterprise Edition with the named instance SQLEXPRESS as the database DB\_BIOWES for the storage of experiment description and DB\_BIOWES\_STORAGE for the storage of the experimental data in the form of binary blocks.

The structure and the tables of the database are documented in the Attachment 1. – Database structure of this document. The software and hardware specification of the server can be found in chapter 7 - Technologies.

## 3 Access points

The database is just the relational database for storage of the experimental data, metadata, relations between protocols and user accounts. To access the data or upload or download the data from the database, additional tools are necessary.

The repository has two access tools to the database: *Protocol manager* and *Web interface*. *Protocol manager* is a fat client installed on the local computer dedicated for protocol design, filling and data processing. *Web interface* is a thin client running on the server with the database and is dedicated for the account management, protocol organization, sharing and searching. Both access tools are able to read and write the data stored in the database - see Figure 1. Scheme of the AQUAEXCEL repository.

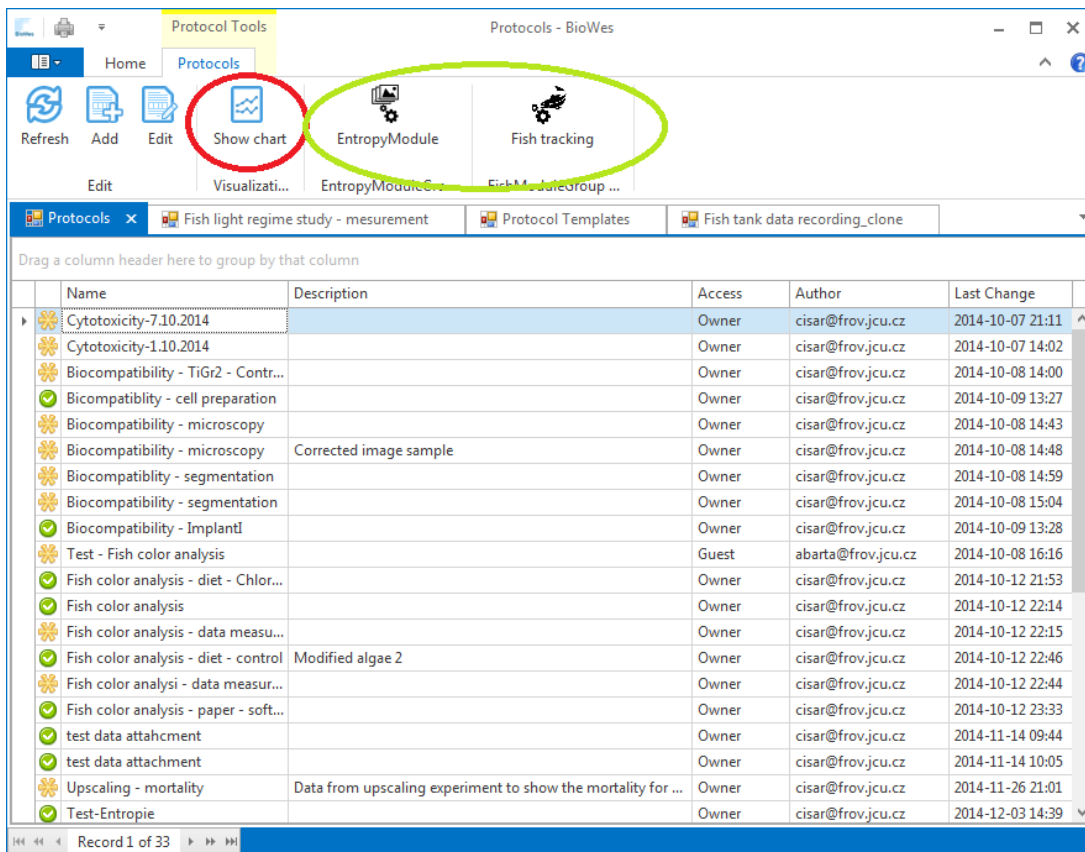
### 3.1 BioWes – a fat client

*BioWes* is a fat client running on Windows operating system for the creation and modification of the protocols and protocol templates. It is possible to extend or modify the functionality of the software by the specialized modules. The software allows:

- Management of protocols and protocol templates (using the module *Protocol manager*) see Figure 6.
- Graphical design of protocol templates (using the modules *Protocol designer*)
- Filling the protocols (using the module *Protocol generator*)
- Visualization and data processing (using the specialized modules for specific data types) see Figure 7.

	Name	Description	Access	Author	Last Change
▶	Cytotoxicity-7.10.2014		Owner	cisar@frov.jcu.cz	2014-10-07 21:11
✿	Cytotoxicity-1.10.2014		Owner	cisar@frov.jcu.cz	2014-10-07 14:02
✿	Biocompatibility - TiGr2 - Contr...		Owner	cisar@frov.jcu.cz	2014-10-08 14:00
✓	Biocompatibility - cell preparation		Owner	cisar@frov.jcu.cz	2014-10-09 13:27
✿	Biocompatibility - microscopy		Owner	cisar@frov.jcu.cz	2014-10-08 14:43
✿	Biocompatibility - microscopy	Corrected image sample	Owner	cisar@frov.jcu.cz	2014-10-08 14:48
✿	Biocompatibility - segmentation		Owner	cisar@frov.jcu.cz	2014-10-08 14:59
✿	Biocompatibility - segmentation		Owner	cisar@frov.jcu.cz	2014-10-08 15:04
✓	Biocompatibility - Implant		Owner	cisar@frov.jcu.cz	2014-10-09 13:28
✿	Test - Fish color analysis		Guest	abarta@frov.jcu.cz	2014-10-08 16:16
✓	Fish color analysis - diet - Chlor...		Owner	cisar@frov.jcu.cz	2014-10-12 21:53
✓	Fish color analysis		Owner	cisar@frov.jcu.cz	2014-10-12 22:14
✿	Fish color analysis - data measu...		Owner	cisar@frov.jcu.cz	2014-10-12 22:15
✓	Fish color analysis - diet - control	Modified algae 2	Owner	cisar@frov.jcu.cz	2014-10-12 22:46
✿	Fish color analysi - data measur...		Owner	cisar@frov.jcu.cz	2014-10-12 22:44
✓	Fish color analysis - paper - soft...		Owner	cisar@frov.jcu.cz	2014-10-12 23:33
✓	test data attachment		Owner	cisar@frov.jcu.cz	2014-11-14 09:44
✓	test data attachment		Owner	cisar@frov.jcu.cz	2014-11-14 10:05
✿	Upscaling - mortality	Data from upscaling experiment to show the mortality for ...	Owner	cisar@frov.jcu.cz	2014-11-26 21:01
✓	Test-Entropie		Owner	cisar@frov.jcu.cz	2014-12-03 14:39

**Figure 6.** BioWes – fat client with *Protocol manager* module. The module shows the list of protocol templates and protocols.



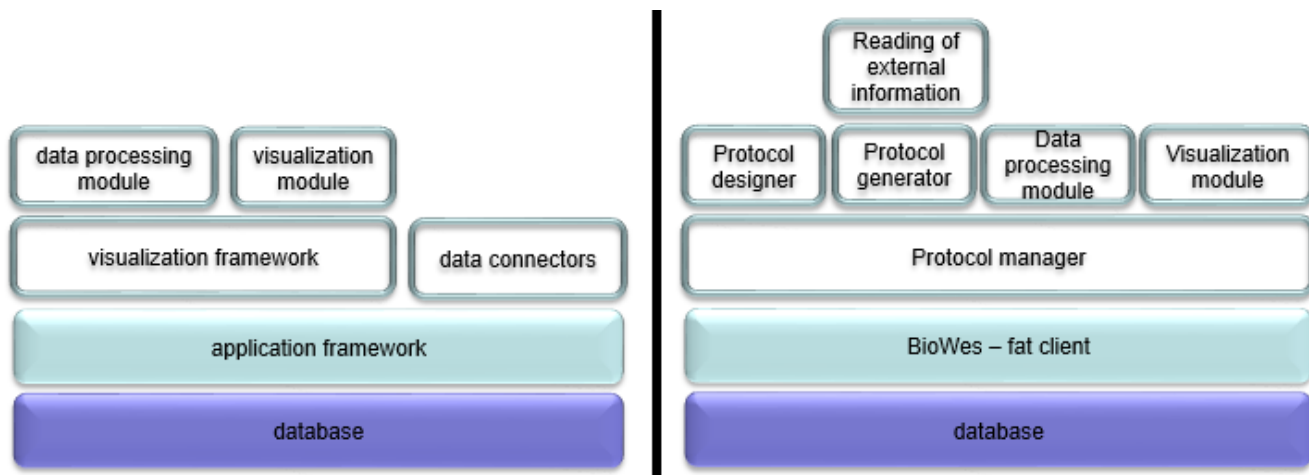
**Figure 7.** BioWes – fat client with *Protocol manager* module with plugged module for visualization (red) and for data processing (green).

### 3.2 Application framework

The *BioWes* – fat client uses the application framework as a basic platform for the development of the modular functionality. The framework is modular system that provides the services for the modules (see Figure 8 – left):

- Control of user access
- Authorization and authentication of users
- Creation of graphical user interfaces (windows, dialogs, lists,...)
- Control of the database connection

The main module of the application framework is *Protocol manager*. The graphical user interface of the module is activated immediately after the *BioWes* – fat client execution and valid login of the user. The other modules for protocol design, filling or data processing are executed using this module - see Figure 8 - right.



**Figure 8.** Left – schema of application layers of for *BioWes* – fat client. Right – schema of module hierarchy in *BioWes* – fat client.

### 3.3 Protocol manager

The module is the main module of *BioWes* - fat client. It provides the commands for visualization, creation and modification of protocols and protocol templates. The protocols and protocol templates are visualized in the form of list with the basic information (name, description, author, last modification date), see Figure 6 (list of protocols). Other modules are executed through this module. The other modules can add the commands to the menu of *Protocol manager* and provide their functions. For example the modules for data processing are executable from the menu of *Protocol manager*, see Figure 7 (green colour).

*Protocol manager* provides the access to the database of other modules. This module uses the client communication interface for the connection to the database and it is responsible for the connection live cycle. The other modules have the access to the created connection to the database and they can use the part of client communication interface for data upload and download through the services exported by *Protocol manager*.

### 3.4 Web interface

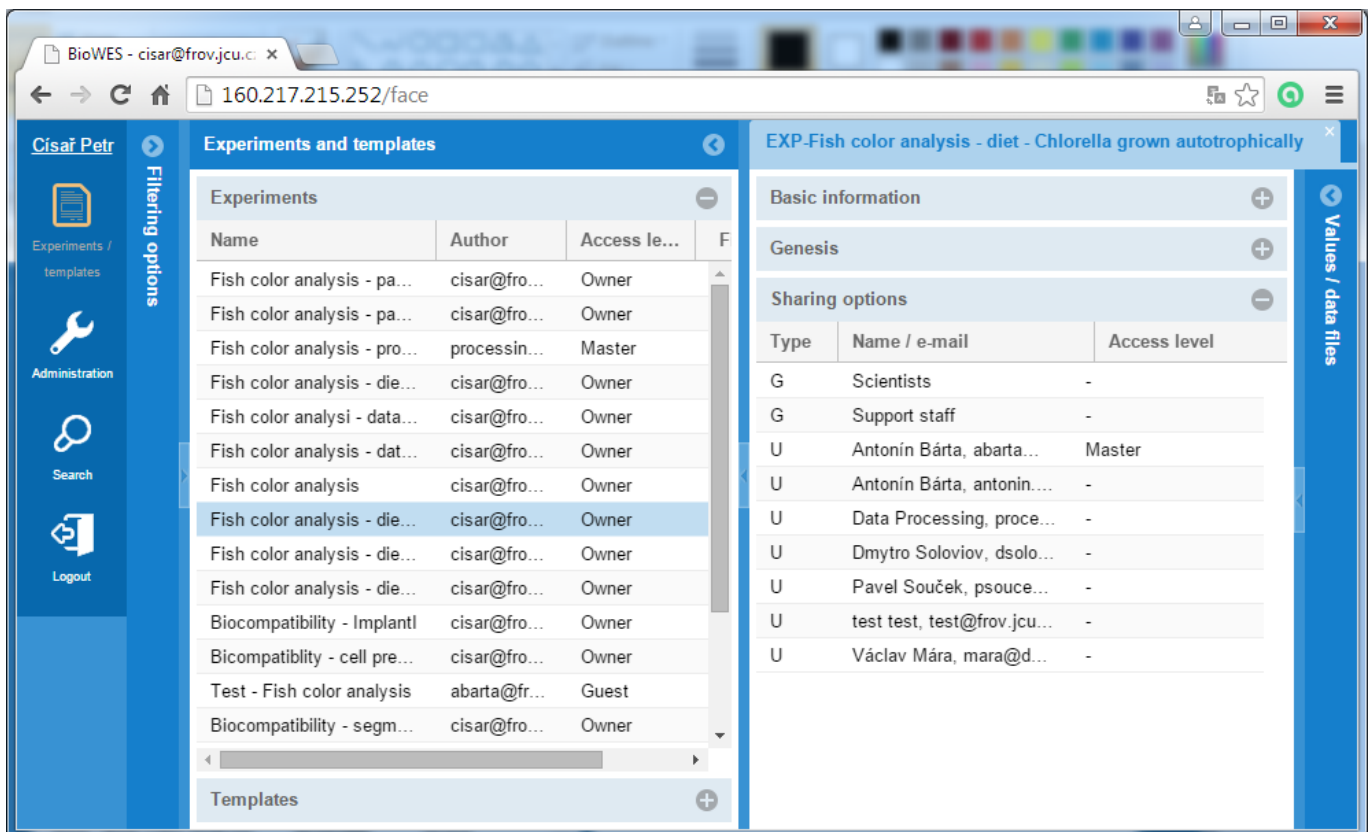
*Web interface* allows the user to access the repository without installing any fat client and provides the management of protocols and protocol templates stored in the database. It is implemented as a thin client (web application) using the services of web server APACHE version 2.4 and script language PHP version 5.3. The application on the server site is based on PHP framework NETTE version 2.

The main focus is on the user - friendly control, which is used for the visualization of information from protocols without the need to create or modify them.

The main features of the *Web interface*:

- Create user account in the system
- Log in to the system
- Show the list of protocols and protocol templates accessible by the user
- Show the details of protocols and protocol templates

- Show the values filled into the protocol
- Visualize the links between protocols and protocol templates
- Show the structure and content of the attached experimental data
- Define the access rights to the protocols, protocol templates and data files attached to protocols
- Full text search in the metadata of protocols and protocol templates
- Download data files attached to the protocols



**Figure 9.** Example of *Web interface*. Left window shows list of protocols. Right window shows sharing possibilities of the selected protocol.

### 3.4.1. Sharing

One of the ways to cooperate between the users on the processing of experimental data or realization of experiments is sharing of experimental data and metadata.

*Web interface* allows the user (owner of the protocol or protocol template) to share the protocols or protocol templates under defined access rights. Figure 9 shows the user interface for sharing of selected protocols. The user can share them only with other users registered in the system. The system provides 8 different levels of sharing. Each level contains the access right of the lower level. The lowest level is a "Guest" and highest one an "Owner".

Access right levels:

- L1: "Guest - 1" – user can read only basic information about protocol – protocol name and protocol information (key words, annotation, author)
- L2: "Guest - 2" – user can read experiment description



- L3: “Guest - 3” – user can read experimental data

The users with access right L1-L3 cannot share the protocol or protocol template with other users.

- L4: “Reviewer” - user can write only basic information about protocol – protocol name and protocol information (key words, annotation, author)
- L5: „Contributor“ – user can write experiment description
- L6: „Master“ - can write experimental data

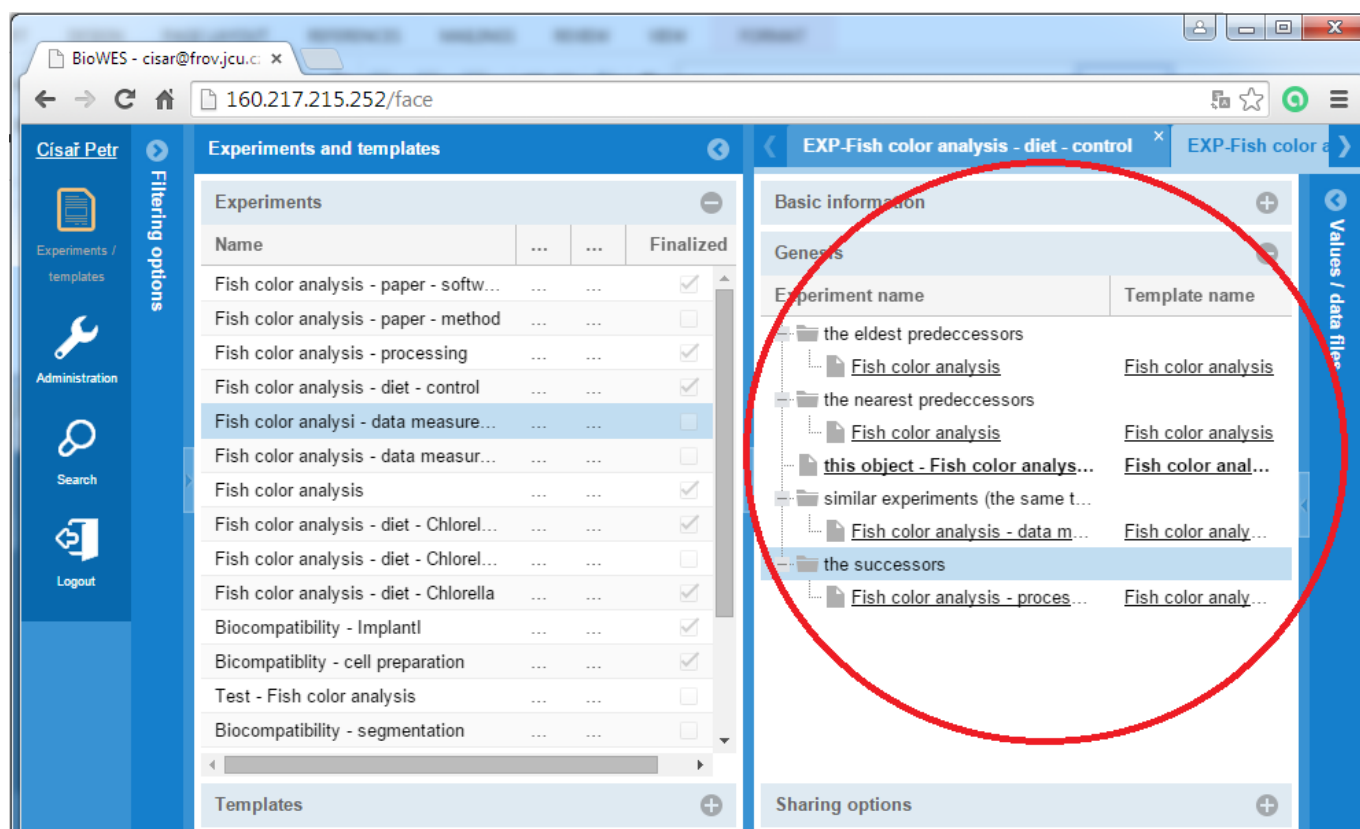
The users with the access right L4 – L6 can share the protocol or protocol template to other users but just those with the access rights L1

- L7: “Co-author” – full read/write access to protocol or protocol template. The user can share with others until level L6.
- L8: „Author“ (“Owner”) – full access to the protocol or protocol template. The user who created the protocol or protocol template.

### 3.4.2. Genesis

The advantage of AQUAEXCEL repository is the possibility to create links between protocols or between protocol templates. This mechanism allows the user to track the processing of the experimental data or the evolution of the protocols. For more information see section 5.

*Web interface* can be used for simple visualization of these relations. The tab Genesis in *Web interface* shows the eldest, nearest parent protocol, similar experiments (based on the same protocol template) and the successors of the selected protocol. The protocol templates used for the particular protocol are also listed.



**Figure 10.** Example is visualization of links between protocols for back-tracking of information on experimental data processing.

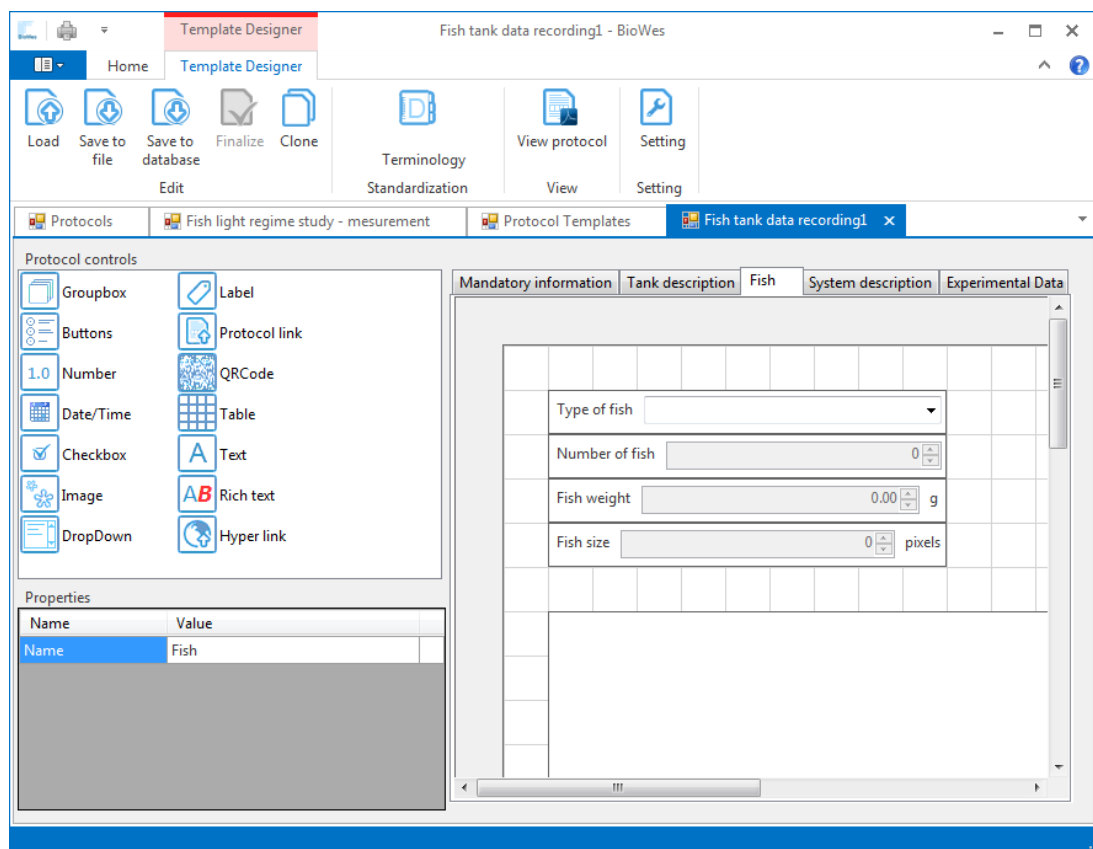
Detailed information about *Web interface* can be found in the Attachment no. 2 – *Web interface*



## 4 Protocol designer

*Protocol designer* is a module of *BioWes* – fat client for the graphical design of the protocol template. The module is executed by the *Protocol manager* in case that the user wants to create a new one, modify or finalize the protocol template. The module allows the user to create a new protocol template using 14 predefined components. The components were defined according to the survey of the experimental protocol, designed and later modified according to the recommendations based on the testing of the repository. The process of protocol design starts with the analysis of the experiment. The user should identify the information, which is necessary for a proper reproducibility of the experiment. The user divides the protocol description into N logical parts (sample preparation, measurement procedure, measurement...).

The *Protocol designer* visualizes the protocol template in the design window and the user can modify number tabs (see Figure 11) to the number of logical parts. The user can simply drag and drop the components from components list (Protocol controls) to the designing window, place it and resize depending on the purpose. Each component has a list of properties that are shown in Properties window. The common property of each component is a unique identifier, which identifies the component within the protocol template. After the user has created the protocol template to cover all necessary information about the experiment, the protocol can be finalized. Once the protocol has been finalized it cannot be changed. It can be only cloned. The cloning action creates a new protocol template, which is unlocked and is linked to original protocol template. This is used to simplify the design of a new protocol for similar experiments. The user can only reuse a part of the original protocol and add new components. Before the user finalizes the protocol template he can save it to the database. The protocol template is described by the *xml* file (see Figure 5) and stored to the database. The information about the components, component properties and component appearance is saved. The life cycle of a protocol template is show in Figure 13.



**Figure 11.** Example *Protocol designer* module with opened protocol template.

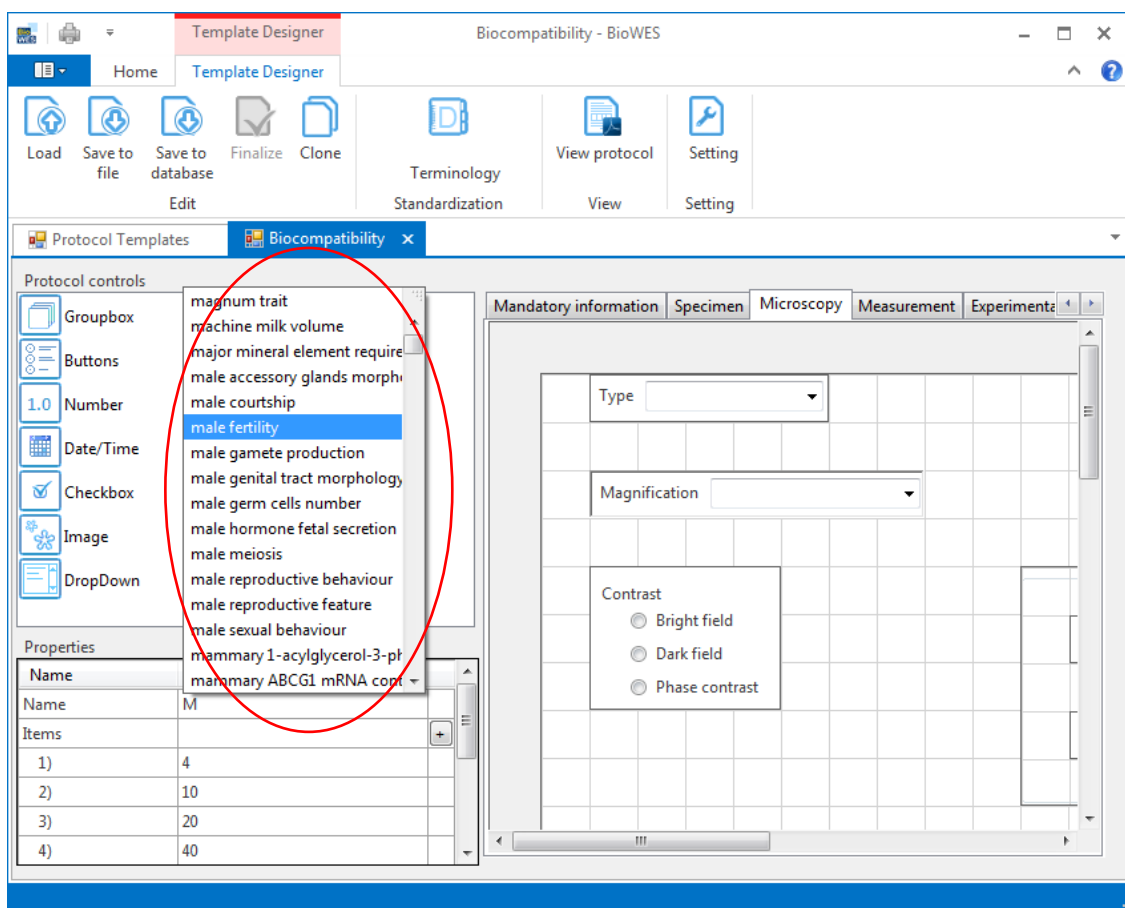
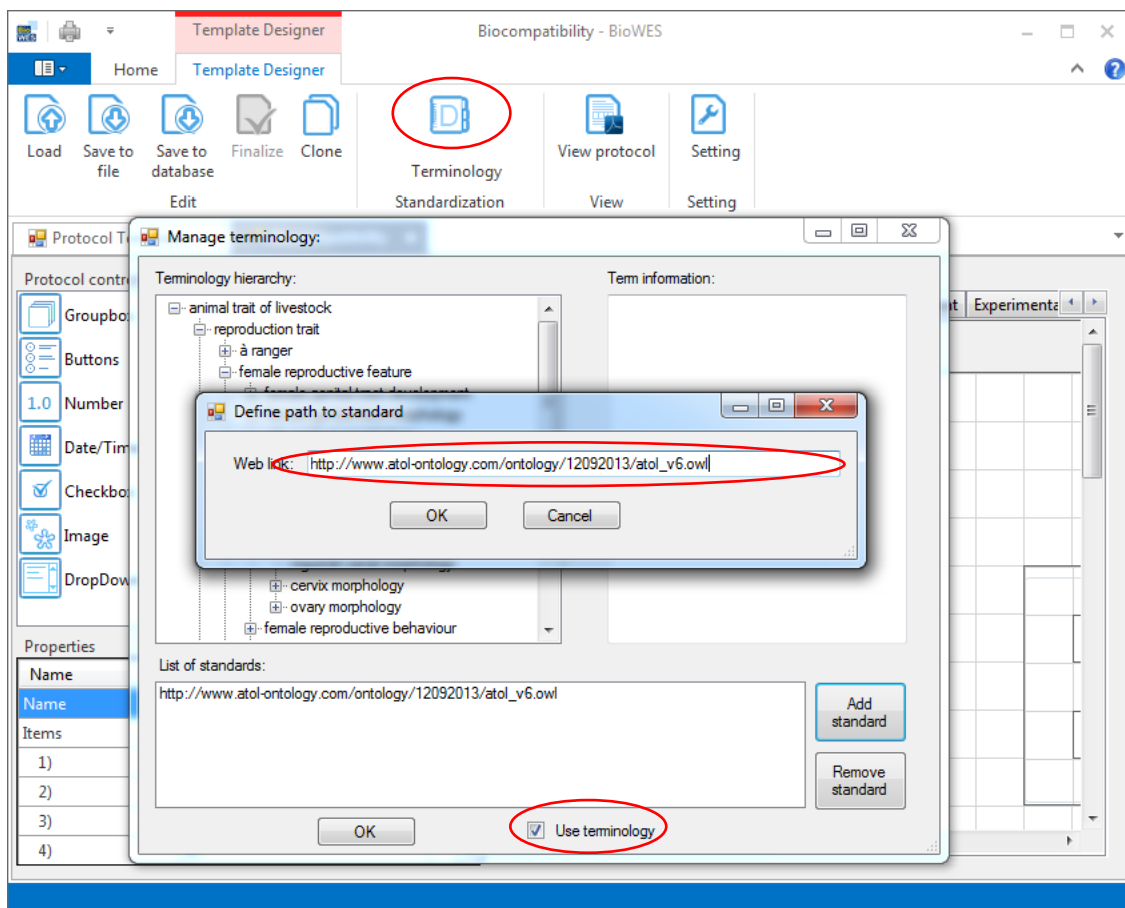
*Protocol designer* has the following components for design of the protocol:

- User input
  - Buttons
  - Number
  - Checkbox
  - Date/time
  - DropDown
  - Table
  - Text
  - Rich text
- Informative
  - Groupbox
  - Image
  - Label
  - QR code
- Links
  - Hyper link
  - Protocol link

The components are described in the Attachment no.3. – *Protocol designer* components.

#### 4.1 Standardization

Standardized terminology is one of the most important things for exchange and cooperation of more people on the same experimental data. Standardized terminology minimizes the possibility of misinterpretation of experiment descriptions. Several initiatives exist collecting the standardized terminology in many different areas of research (Open Biology Ontology <http://www.bioontology.org/>, Minimum Information for Biological and Biomedical Investigations Biomedical ontology <http://mibbi.sourceforge.net/portal.shtml>, BioSharing <http://BioSharing.org>, BioPortal <http://biportal.bioontology.org/>). *Protocol designer* has implemented the tools for support of the usage of terminology from standards available in the form of ontology OWL files on web pages. The user just defines the link to the OWL file and *Protocol designer* loads the ontology to the local disk. It parses the terminology from the ontology and offers the term to the user during the definition of the components - see Figure 12.



**Figure 12.** Usage of the ontologies for terminology standardization. Top – definition of the standard from web link. Bottom – offered terminology for the definition of component name.

The user can use the term from offered terminology or can define his own term. This is a simple way to use standardized terminology even if the user does not know the standard.

One of the outcomes of the AQUAEXCEL project is the ontology AQUAEXCEL-ATOL. It defines and organizes livestock traits, with a focus on the main types of fish production (meat, feed and fertility). The usage of the ontology for experiment description allows for automatic semantic searches. *Protocol designer* is able to read the ontology for the design of the protocol template because it is defined in the form of OWL file.

Once the protocol template has used the standard, the information about the standard is stored in the *xml* file and saved to the repository. Before the modification of the protocol (by the owner or by any other user to whom the protocol template was shared), *Protocol designer* warns the user that the protocol template uses the standard and offers to load the standard to keep modifications standardized.

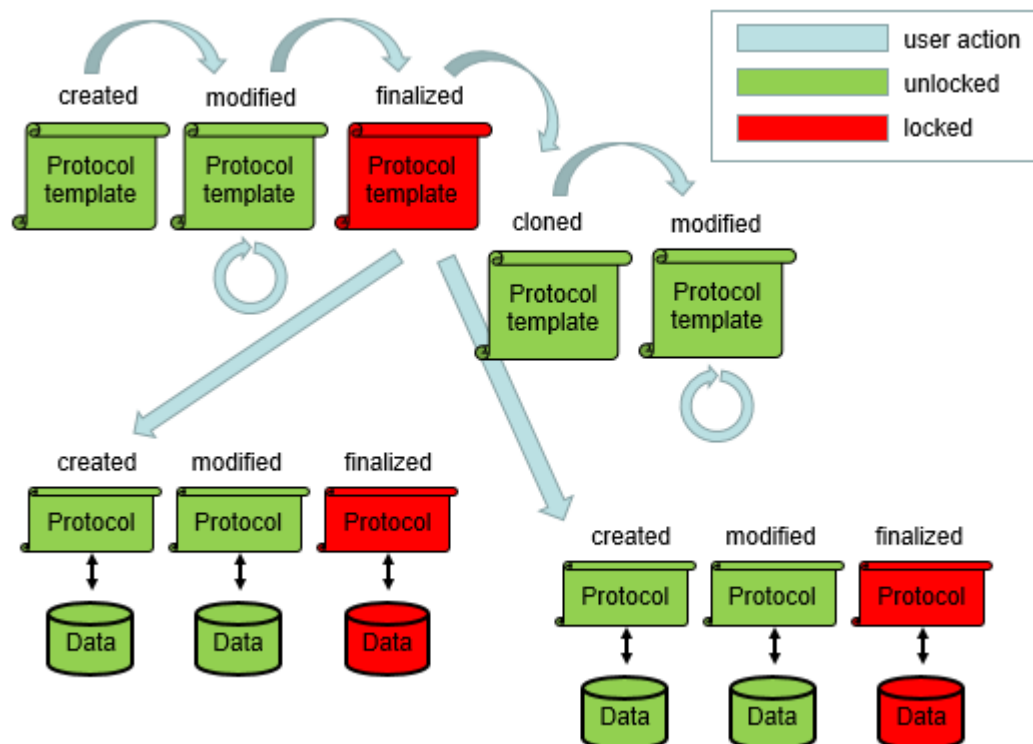
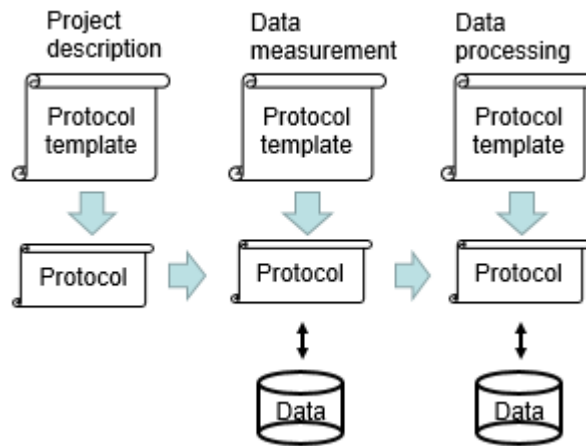


Figure 13. Protocol and protocol template life cycle.

## 5 Protocol generator

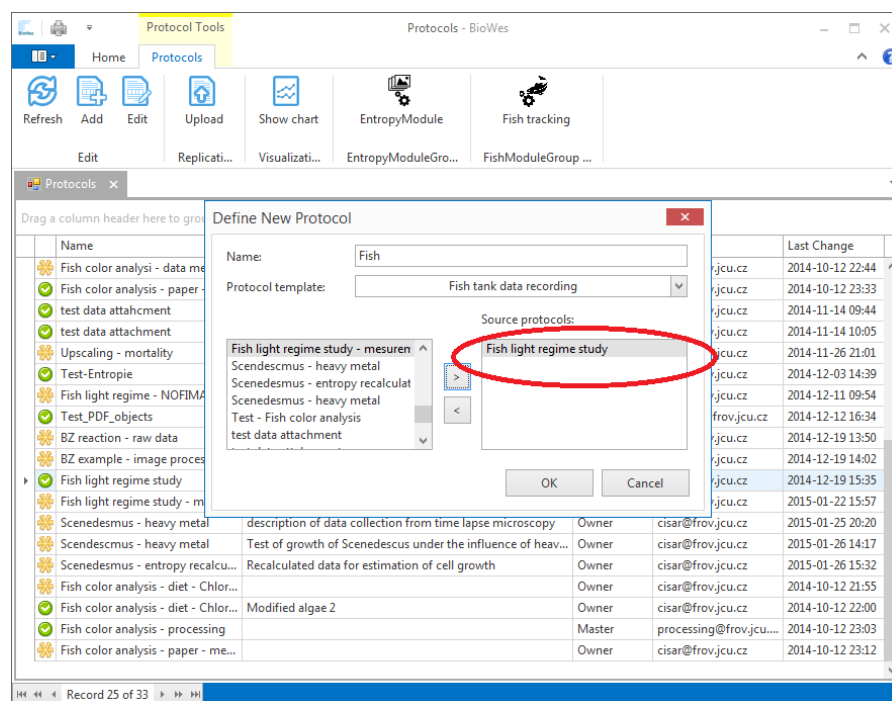
*Protocol generator* is a module of *BioWes* – fat client executed by the module *Protocol manager*, see Figure 8. The module allows the user to create, modify or finalize protocol. *Protocol generator* reads the protocol template from the database to realize specific experiment. The protocol is visualized in module's window and the user can insert the values, text, images and attached experimental data to the protocol, see Figure 4. The module can also serve as a guide through the experiment if the protocol template contains the information about realization of the experiment.

To create a new protocol, the user has to select the protocol template from the list of finalized protocol templates. The protocol template has to be finalized (locked for modifications) because it contains a prescription for the protocol visualization and definitions of data types for data entry. The user can fill the information into the protocol and save it to the database in unfinished version. The instance of protocol is created in the database and the information is stored and linked with the components (using the unique identifier of the component). The protocol is linked to the protocol template. The data (experimental data) attached to the protocol are stored to the database in the form of binary blocks. The link between protocol and the data is created. This system provides the overall information about relations between Protocol prescriptions (protocol template), filled information about experiment, and attached data.



**Figure 14.** Links between protocol templates and protocols for definition of data processing chain

The protocol itself can be linked to the parent protocol. This mechanism is used for the creation of data processing chain in the repository. The user can create one protocol for project description, the second protocol describes the data measurement and the third one describes the processing of measured data. *Web interface* provides the tool for visualization of the links between protocols for the possibility to backtrack the data processing steps of analyzing the evolution of the protocol template. The link to the parent protocol can be defined during the creation of the protocol or during the modification – see Figure 15.



**Figure 15.** Defining the link to parent protocol during protocol creation.

The second way how to create a link to another protocol is to use the component Protocol link. If the protocol template contains this component then the user can select the protocol from the list of existing protocols. If the user clicks to the linked protocol in *Protocol generator*, it is opened in a new window. This is the way to reduce the redundancy in the protocols. If the experiment needs some preparatory actions (cell preparation, mixing of cultivation medium) then it can be described in a separated protocol and just linked to the protocol of the experiment.

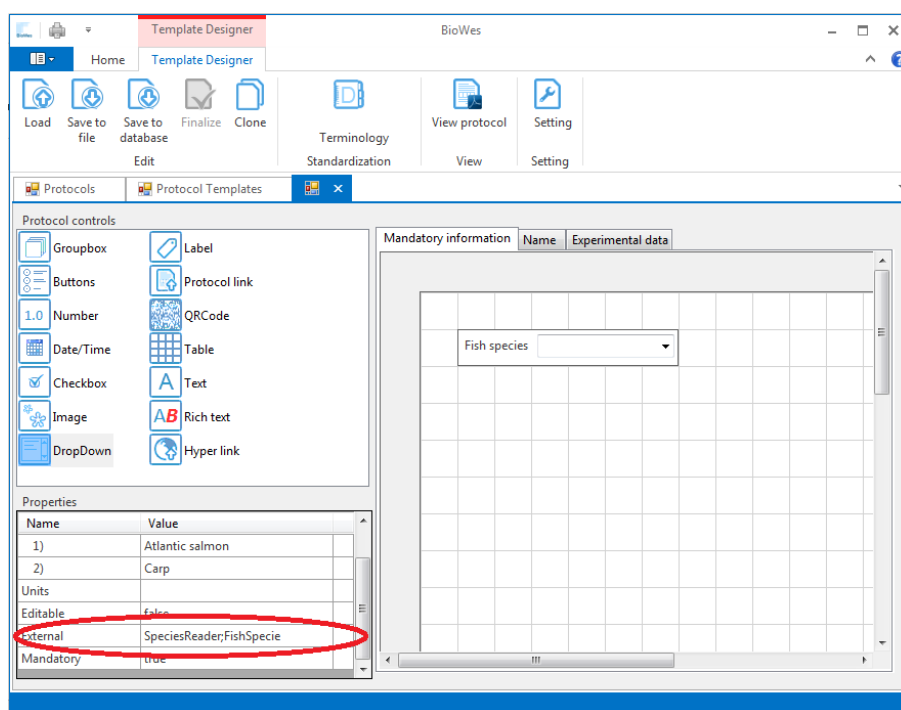
*Protocol generator* allows the user to print the filled protocol into a *pdf* file, which can be stored or printed. After the finalization of an experiment, the protocol can be finalized. Once the protocol is finalized, the user cannot change the description of the experiment, nor the attached data files. The user can still visualize the protocol and download the data from the database to a local computer depending on the access rights to the protocol.

## 5.1 External information

The measurement devices often provide the files with the configuration used for the measurement (LC MS files, images, etc..). In this case, the user does not need to manually fill the information about the setting of the measurement device into protocol because the *Protocol generator* allows executing the plug-ins for filling components by external information.

The plug-in for filling components is a specialized plug-in to read the information about specific measurement device setting from file. The plug-in sub Third party plug-ins in the form of *dll* files, which provide the predefined interface can be used by the *Protocol generator*.

The plugin and the values, which should be read from external file are defined during the protocol template definition in *Protocol designer* – see Figure 16. The components, which can be filled by plug-ins are: number, date/time, checkbox, image, dropdown, text, rich text, hyper link.



**Figure 16.** Definition of the plug-in and value user for automatic filling of the protocol from external file.

The user only clicks on the Fill protocol icon in the *Protocol generator* and selects the file with external information. If the plug-in is installed, it finds the value in selected file and fills it back to the protocol.

## 6 Data processing modules

After the protocol has been described and experimental data attached to the protocol and stored in the repository, the data should be processed and analyzed. The user (or the cooperator to whom the data were shared) can download the data using the *Protocol generator* or *Web interface* to a local computer and start the analysis.

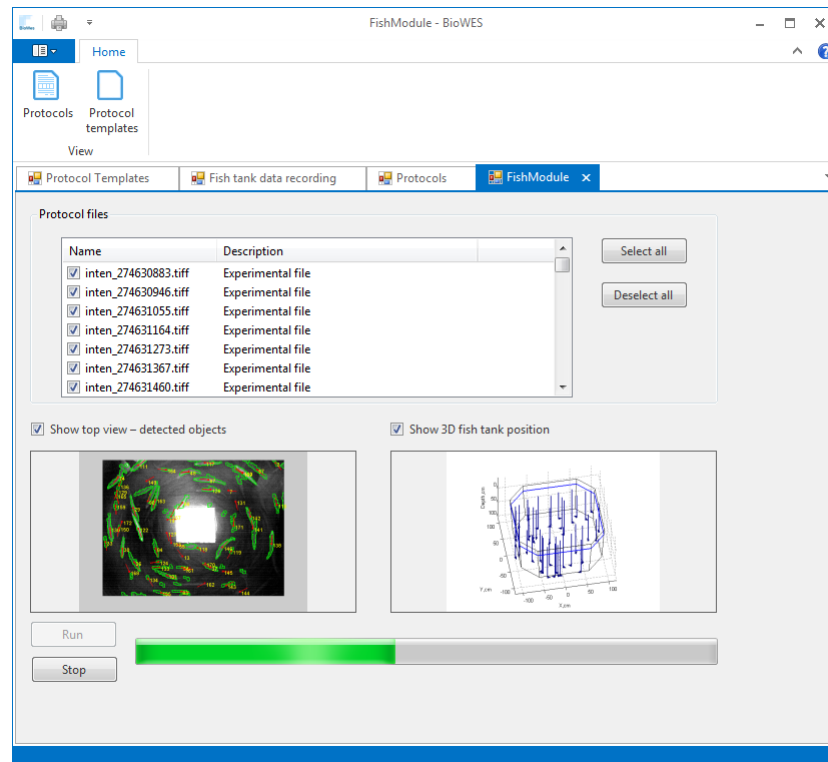
The user can also use one of the implemented modules for data processing or data visualization to analyze data directly using the AQUAEXCEL repository. Data processing or visualization modules are specialized plugins for BioWes – fat client, which can be executed by *Protocol manager* on the selected protocol, see Figure 8. The module can read specific data formats from the database through *Protocol manager* and do the processing. After the data are processed, the module can save the data to a local computer. The user can create new protocol describing the data processing and store the processed data to the repository. If the user creates the link between the original protocol and the protocol on data processing then the processing chain can be visualized on the *Web interface*. For more information about the processing chain see section 5.

Two visualization modules and three data processing modules are implemented and can be used by AQUAEXCEL members. The installations of the modules are available at this link <http://www.biowes.org/category/downloads/modules>.

Available modules:

- Fish tracking in tanks – see Figure 17.
- Image representation
- Object labeling
- Chart data visualization
- LC-MS data visualization

The SDK (software development kit) for implementation of the modules is available for the programmers and is described in the Attachment no.5 – Processing module SDK.



**Figure 17.** Example of data processing module. Fish tracking in the tank.

## 7 Technologies

This paragraph contains the description of the technologies used for implementation and running of the AQUAEXCEL repository:

Hardware configuration of the server:

- Processor Intel Xeon E3-1230Lv3
- Main board SuperMicro X10SLH-F
- RAM 2 x 4GB DDR3 Micron
- Drive controller SuperMicro sATA RAID 0/1/5/10
- Net interface Intel i210AT 2x GbE
- Hard drives WD Velociraptor 500 GB

## WD Raid Edition 4 x 2TB

Software configuration of the server:

- Operation system      Microsoft Windows Server 2012 Standard x64
- Database server      Microsoft SQL Server 2012 Enterprise Edition
- SW platform      Microsoft .NET Framework 4
- Web server      PHP 5.3.27 Non Thread-Safe + SqlSrv Driver 2.1  
Sencha ExtJS 4.2.1  
APACHE 2.4.7

Minimal requirements for the *BioWes* – fat client

Software:

- Operating system: Microsoft Windows 7 or 8
- Software platform: .NET Framework 4.5 or higher

Hardware:

- 64bit dual-core processor, 1 GHz
- 4 GB RAM
- 16GB free HDD space

Minimal requirements for the Web interface:

Software:

- Browsers: Google Chrome, Internet Explorer, Mozilla Firefox, Opera or Safari

## 8 Security

Because the repository can store sensitive data (experimental data which have not been published, personal data, etc..), the user has to be sure that the repository is secured.

The hardware security is realized by two methods. The hard drives used for server operation system and data storage are physically separated and the data storage is realized as the RAID 1 array (mirror). It means that if one disk fails the data are immediately recovered from the mirrored disk.

The software security (data access) is realized by several solutions. The user has the credentials for the access to the user account using *Web interface* and *BioWes* – fat client. *Protocol manager* of BioWes - fat client uses the encrypted connection to the database that uses internal login name and password created during the installation of *BioWes* – fat client (stored on a local computer in encrypted form) for database access and user credentials for user account access. The other modules of BioWes – fat client (*Protocol designer*, *Protocol generator*, *data processing modules*) use the established connection to the database and cannot create the connection itself.

Only the user decides about the accessibility of the protocol templates, protocols and data using the access right defined for sharing - see section 3.4.1.



## Use cases

The use cases are the particular experiments from different areas which demonstrate the functionality of protocol design, protocol filling and data and metadata management. The sample data together with the description of experiments are stored in the AQUAEXCEL repository as testing protocols. Each user can access the protocols using testing accounts:

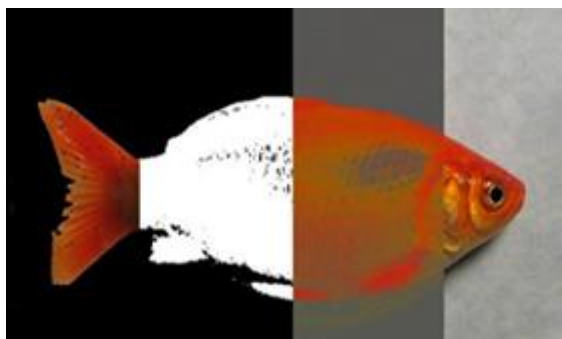
Login: [test@test.cz](mailto:test@test.cz)

Password: test

The user can use the login and password to access the account with the pre-prepared protocols, experimental data and processed data. The user can simply test the tools and functionality of the repository without the need of creation of the protocols and storage of user data.

### 1 Fish color analysis

The automatic fish color analysis is useful for the analysis of the influence of different fish diets. The example consists from three different protocols. The first protocol describes the diet itself. The second protocol describes the data measurement and the last protocol describes automatic fish detection and estimation of the fish color.



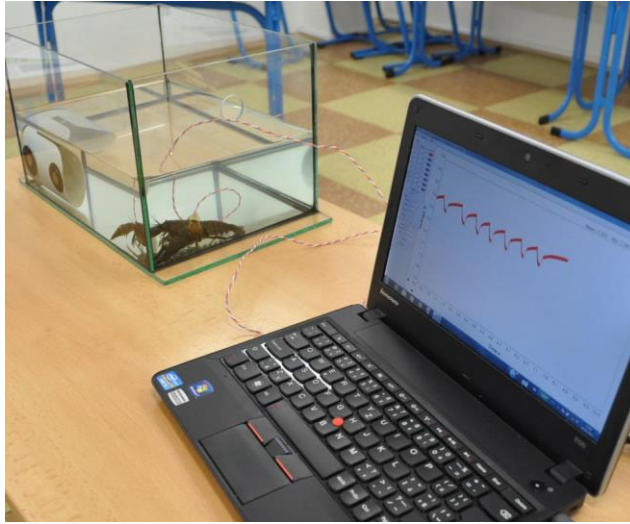
**Figure 18.** Illustrative example of automatic fish detection and representation in different color spaces.

The user can learn how to create and use the link to other protocols (data measurement links to the description of the specific diet).

More information about the example can be found in the attachment n.7 – Fish color.

### 2 Water quality – crayfish

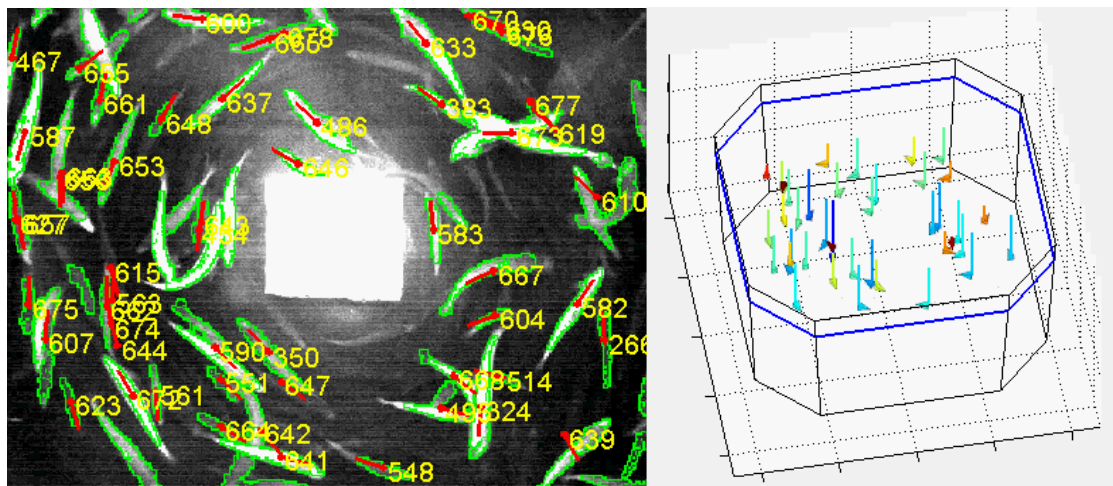
The use case describes the using of crayfish as a bio-indicator for the analysis of water quality (determination of the chemicals in the water). The experiment consist from two linked protocols which contains the original signal of crayfish heartbeat and processed data with signal features for detection of chemicals. See the description of experimental conditions in attachment n.8. – Water quality.



**Figure 19.** Crayfish as bio-indicator for water quality analysis

### 3 Fish tracking

The fish tracking example contain the data from small fish tank monitoring (2m tank) using the system of infra-red (IR) illuminator and IR camera. The camera is located above the tank and records the fish activity. The raw data is the series of IR images. The example can be used for the demonstration of the functionality of the data processing module for fish tracking. The user can simply select the protocol and execute the fish tracking module in the system and test the module.



**Figure 20.** Example of the fish tracks (left) and fish visualization in the 3D space in the tank.

### 4 Bio-compatibility

This particular example demonstrates individual steps of the system usage from the protocol template design until the sharing and searches of the protocols. It shows the advantages of the system in comparison with current systems used for experimental work and management of experimental data and metadata. The example introduces the basic functionality of the system. The detail description of individual functions and procedures is available in the help.

The aim of the biocompatibility experiment is to determine how the human body will react to the new material (dental implant). The first method of biocompatibility evaluation used before clinical tests is the tissue cells growth in the contact with the material. The method is based on the evolution of cell colony growth in defined time interval.(1-2 days). The cells are put directly on the material inserted into the leachate of the material. Time lapse microscopy is used for the monitoring of cell activity. It produces series of images in given time intervals (2 minutes).

Detail description of the protocols can be found in attachment n.6. – Particular example.

## 5 TNA form

This sample contains the TNA application form used by the applicants of TNA project. It is used for the demonstration that the system is prepared to store the data from TNA projects from the beginning (application) till the final paper.

The screenshot shows a web-based form titled "APPLICATION FORM FOR RESEARCH ACCESS" within the AQUAEXCEL repository. The form is part of a "Protocol Templates" system, as indicated by the "Template Designer" header and the "Protocol Templates" tab. The form is divided into several sections: "Application Reference Code (Leave Blank - will be filled by secretariat)", "Please read guidelines before completing this form", and three numbered sections for research infrastructure details. The "Properties" panel on the left shows the form's dimensions (Width: 152, Height: 40) and other settings. The "Form" panel on the right displays the form's content, including a "Research Infrastructure" dropdown menu and a "Lead Researcher" section with "Name" and "Position" fields.

Name	Value
Path	Internal image/C:/Users/Cisar/Desktop...
Label	
Label position	Below
ReadOnly	true
External	
Width	152
Height	40

**Figure 21.** Example of TNA application form realized as protocol in AQUAEXCEL repository.

## **Documentation**

To support the usage of the system the manual, video guides and testing protocols and datasets were prepared for AQUAEXCEL members. All materials can be found at the AQUAEXCEL webpage under the section Results/AQUAEXCEL repository.

The video guides show the individual steps of the system to teach the users how use the system for their work. The video guides also show the structure and data flow in the system.

## Conclusion

The objective of the deliverable D3.4 was to develop the repository suitable for the storage and management of experimental data and experiment description from a wide range of experiments. The AQUAEXCEL repository is therefore based on the model of digital protocol and black box model of the experimental data, which can store any data type and structure in the repository. The user can access the data by two different access tools: *Web interface* used for sharing, processing chain exploration and searches and *BioWes* – fat client used for protocol template design, protocol filling and data processing and visualization. The solution supports the usage of standardizations for protocol definition in the form of ontology and uses the results of WP3 (ATOL – ontology). The repository can be used not only for data storage but also for cooperation through sharing features, for data processing using the specialized modules and for reusability of the protocol templates. The solution was designed based on the survey of experimental data management practices, tested and modified to provide user friendly interfaces. The examples of different protocols were created for the practical training of AQUAEXCEL members supported by the documentation and video guides available on the AQUAEXCEL web page.

The AQUAEXCEL repository is available for any AQUAEXCEL member and can be used for the networking and joint research activities or the results of TNA projects.

## Annex 1

Deliverable Check list (to be completed by Deliverable leader)

	Check list	Comments
<b>BEFORE</b>	I have checked the due date and have planned completion in due time	<i>Please inform Management Team of any foreseen delays</i>
	The title corresponds to the title in the DOW	<i>If not please inform the Management Team with justification</i>
	The dissemination level corresponds to that indicated in the DOW	
	The contributors (authors) correspond to those indicated in the DOW	
	The Table of Contents has been validated with the Activity Leader	<i>Please validate the Table of Content with your Activity Leader before drafting the deliverable</i>
	I am using the AQUAEXCEL deliverable template (title page, styles etc)	<i>Available in "Useful Documents" on the collaborative workspace</i>
<b>The draft is ready</b>		
<b>AFTER</b>	I have written a good summary at the beginning of the Deliverable	<i>A 1-2 pages maximum summary is mandatory (not formal but really informative on the content of the Deliverable)</i>
	The deliverable has been reviewed by all contributors (authors)	<i>Make sure all contributors have reviewed and approved the final version of the deliverable. You should leave sufficient time for this validation.</i>
	I have done a spell check and had the English verified	<i>Ask a colleague with a good level of English to review the language of the text and do a spell-check too.</i>
	I have sent the final version to the Activity Leader and to the 2 <sup>nd</sup> Reviewer for approval	<i>Send the final draft to your Activity Leader and the 2<sup>nd</sup> Reviewer and leave 2 weeks for feedback and final changes before the due date. Once validated by the 2 reviewers, the draft is ready to be sent to the Management Team that will ask for the Coordinator validation and then transfer it to the EC.</i>

9