BASELINE

SELECTION AND IMPROVING OF FIT-FOR-PURPOSE SAMPLING PROCEDURES FOR SPECIFIC FOODS AND RISKS

# PROJECT DELIVERABLE

| | |
|---|---|
| **Collaborative project** | Large-scale integrating project |
| **Project acronym** | BASELINE |
| **Project title** | Selection and improving of fit-for-purpose sampling procedures for specific foods and risks |
| **Grant Agreement number** | 222738 |
| **Date of latest version of Annex I** | 01/04/2009 |

| Del. No. | Deliverable name | WP no. | Lead participant | Nature | Dissemination Level | Due delivery date from Annex I |
|---|---|---|---|---|---|---|
| 6.6 | Establish software (Tertiary modeling in SAS and/or R) for the updated or new survival and growth models of human pathogens | 6 | DKFZ | O | PU | M30 |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

**Delivery Date** 28/03/2012

**Project co-ordinator**

Prof. Gerardo Manfreda
Alma Mater Studiorum – Università di Bologna
Tel: +39 051 20 9 785
E-mail: gerardo.manfreda@unibo.it

**Project website address** www.baselineeurope.eu

# TABLE OF CONTENTS





*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

## Workpackage Collaborators

3. Centro Nacional de Tecnología y Seguridad Alimentaria (CNTA)
4. Deutsches Krebsforschungszentrum (DKFZ)
5. Université de Bretagne Occidentale (UBO)
6. Hungarian Food Safety Office (HFSO)
11. University of Cordoba (UCO)
14. Agence Française de Sécurité Sanitaire des Aliments (ANSES)
15. Agriculture and Food Development Authority (TEAGASC)

Subcontracted company: Optimum Quality SL

**Writing team:**

Francisco Mérida Quero- fjmerida@optimumquality.es
Computer engineer Optimum Quality

Supervised by UCO and DKFZ



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

## Executive summary

BASELINE Deliverable 6.6 presents a web-based software tool for predictive microbial models in foods, which relates to Task 6.1 “Quantitative prediction in microbiology”.

The purpose of this Deliverable is to describe the development and features of the application that will help end users to quantify microbial population for specific food and risk (pathogens) combinations as selected in the BASELINE project.

Quantitative data of the behaviour of microbiological pathogens in food play an important role within the framework of quantitative microbial risk assessment, where it contributes – when linked with data on human disease - to the estimation of the impact of contaminated food on public health.

Based on models and criteria collected and established in the BASELINE project, the software tool described in Deliverable 6.6 allows the standard user to input raw data, the modelling of data, the comparison and validation of models, the export of data and results in CSV, MS Excel or PDF. A personal space is provided for each user, allowing them to save and retrieve their own data for further use. By means of restricted local access, administrators will be able to further develop the software tool by including additional predictive models, microorganisms, food matrices or environmental factors.

Developed in PHP language, the web-based software uses the CakePHP framework, based on the MVC pattern together with XHTML 5.0 and UTF-8 encoding. Client and server based programming languages (PHP, JavaScript) is used to perform basic animations and controls prior to sending the information to the server. The use of the jQuery framework supported by Ajax and JSON technologies gives greater fluidity to the site and allows the consultations of the database without reloading the page. For data management, the software tool uses MySQL (database manager). It is hosted on an Apache Server running on Linux (according LAMP philosophy). Style and web-design are supported by CSS 3.0.

Furthermore, the application offers a multi-browser tool completely portable and operational in most common navigators: Firefox, Safari, Internet Explorer, Chrome and Opera.

## Software Tool Functions

### *Model Prediction*

The software tool described in Deliverable 6.6 provides a web-based platform for results communication within the BASELINE project regarding predictive modeling of selected microbiological pathogens in foods as selected in the project.

The structure or the application is shown in Figure 1.

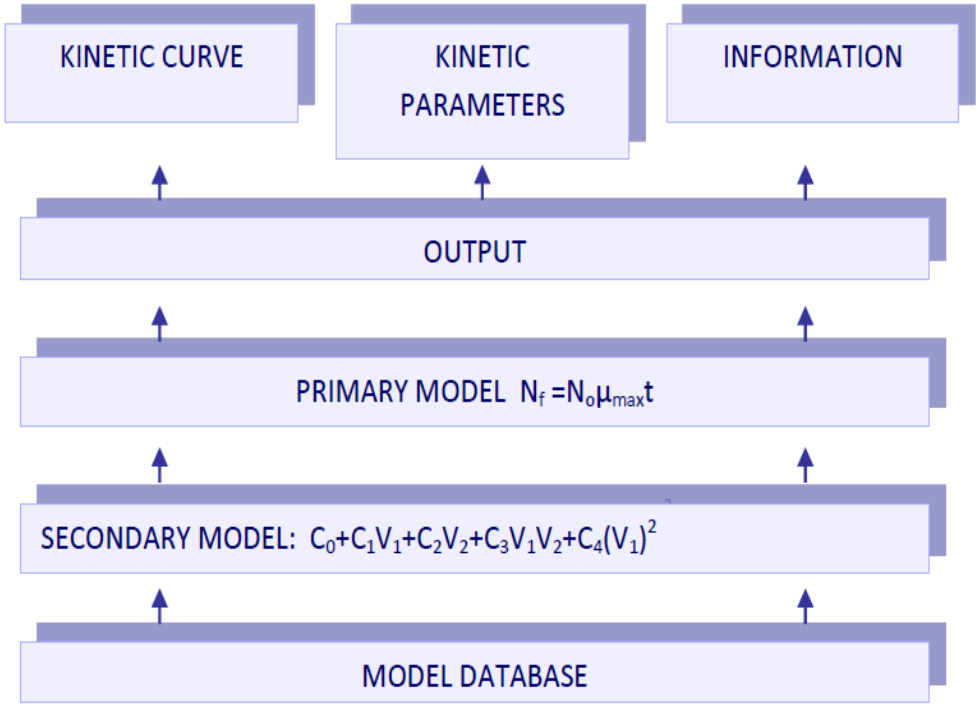


Figure 1. Structure of the application.

The software tool allows the user to carry out a comprehensive data analysis by assessing the behaviour of microbiological data by means of different models (growth, survival and inactivation) and varying parameters such as, e.g. initial contamination, level of the microbial hazard, environmental factors of the model, storage time, etc.

It enables model comparison, including cardinal, polynomial as well as stochastic models implemented in the model database, through a user-friendly interface.

Registered users can save their own settings, being accessible from their personal profile for further consultation and/or validation. They also have the option to generate a PDF report of the result of the model fitting or comparison, or simply export raw data in CSV and/or MS Excel format.



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

Figure 2 shows an example of the fitting of a square- root or Ratkowsky model to different microbiological pathogens and foods:

- The main menu, located on the right side of the page, lists all the predictive models implemented in the application, which may be sorted by food category (seafood, eggs and egg products, milk and dairy products, meat products and plant products) or microorganisms.
- The central right panel shows a table presenting the main results of the most recent fitted model. The first column lists the number of points or observations included in the graph, the next column the time (hours/days), followed by the predicted log concentration. This data can be easily exported into CSV or MS Excel format.
- The central left part provides a graphical representation of the predictive log concentration as a function of time and/or environmental factors (temperature, pH, water activity etc.).
- The top panel summarizes the most relevant information for each model, which is type of microbiological pathogen and food, primary and secondary model applied as well as predicted maximum growth rate, maximum population density, lag time and time to reach a hazardous level. The predicted values can be varied by the user to evaluate the effect of the environmental factors.

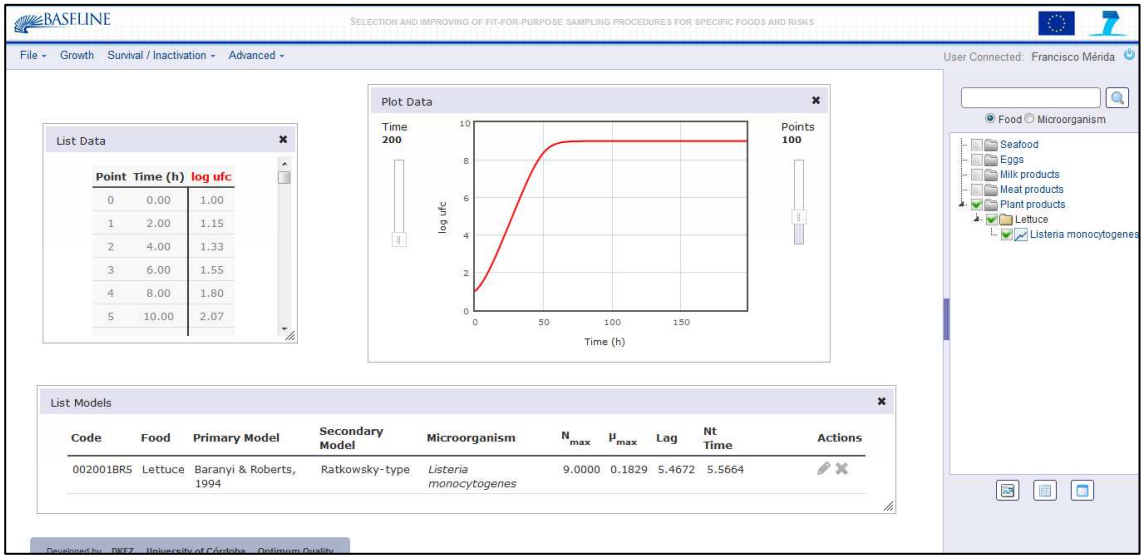


Figure 2. Example for model fitting.

## *Model Validation and evaluation criteria*

The implementation of a validation module allows the input and modelling of raw data and the comparison of the model output with the predictions generated by the established models implemented in the application.

Criteria to evaluate the difference of observed and predicted data implemented in the application are the Standard Error of Prediction (SEP) and the Root Mean Square Error (RMSE), as can be seen in Hervás et al. (2001).

$$SEP = \frac{100}{\bar{\mu}_{obs}} \sqrt{\frac{\sum_{i=1}^{n} (\mu_{obs} - \mu_{pred})^2}{n}} \qquad RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\mu_{pred} - \mu_{obs})^2}{n}}$$

Implemented validation indices are the accuracy factor ($A_f$) and bias factor ($B_f$) proposed by Ross (1996)

$$B_f = 10^{\left(\sum log\left(\frac{g_{pred}}{g_{obs}}\right)/n\right)} \qquad A_f = 10^{\left(\sum abs\left(log\left(\frac{g_{pred}}{g_{obs}}\right)/n\right)\right)}$$

In general, a model over- or under-predicts response time. $B_f$ is a measure for the degree of over- or under-prediction. E.g., a bias factor of 1.1 predicts longer generation times than are observed, exceeding the observed ones by an average of 10%. A bias factor of 0.5 indicates a poor model that predicts generation times, on average, of half of what has been observed.

The accuracy factor $A_f$ indicates the spread of observation about a model's prediction. E.g., a factor of two indicates that the prediction is, on the average, only half or twice as large as the observed data. Ideally, predictive models would have $A_f = B_f = 1$.

Additionally, the correlation between observed and predicted data can be visualized by means of a scatter plot with regression line (see Figure 3). The closer the data points are to the regression line, the stronger the correlation and the closer are the predicted and observed values.



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

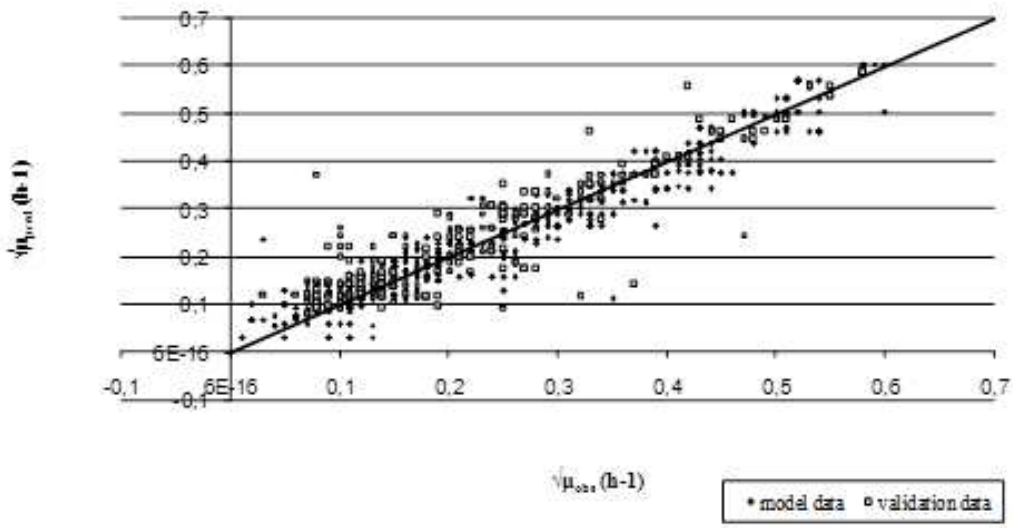


Figure 3. Scatter plot with regression line of observed and predicted values.

All parameter settings and results carried out can be saved on the website platform or exported into CSV or MS Excel format for further use. Furthermore, the application provides PDF reports of the performed analysis.

## *User Control*

The web application contains a user control system in order to restrict access to BASELINE members and/or other interested people. Registration of user will be initially supervised by Optimum Quality, being responsible to register new user accounts and setting privileges thereof. Upon fully development of the software tool and after extensive testing, standard users will be allowed to enter to the application. Administrators will take charge of supervising the correct use and maintenance of the software. The user control system allows two different roles:

Standard User

Standard users have access to the basic tools and setting options: import/export of data, model prediction, validation and generation of PDF reports. Own personal space is provided on the platform to store information.

Administrator

Administrators have full access to the software application. In addition to the standard user's options, administrators act as managers of the application within the administration portal. They are allowed to register new predictive models, pathogens, food matrices, and novel environmental factors to create a dynamic interface between users in order to upgrade the application with new information.

Administrators as well as standard users log in via the user account dialog box shown in Figure 4.

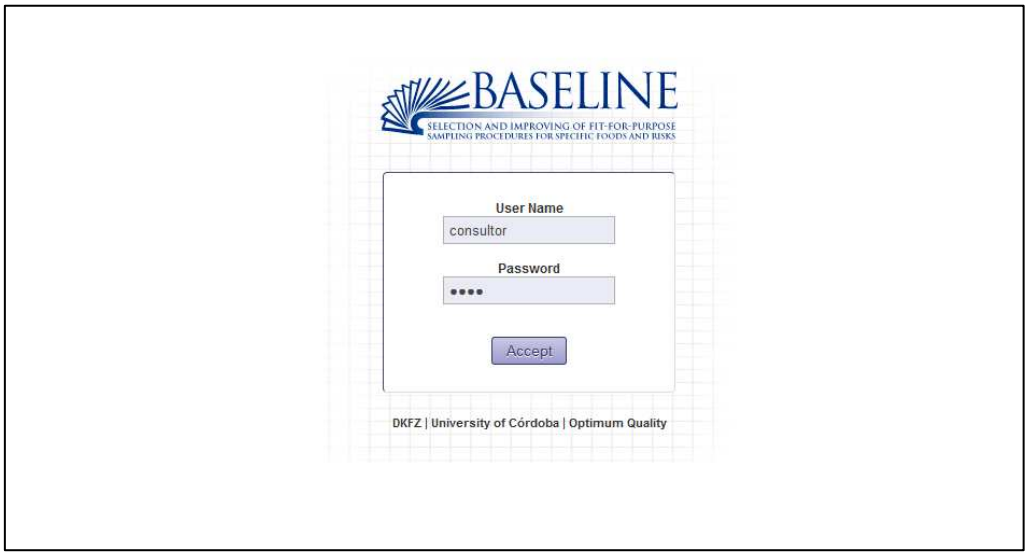


Figure 4. User account dialog box of the web application.

### *Admin Module*

Advanced use is restricted to administrators, thus enabling them to manage and further develop the content of the application. Administrators have access to the following features:

Models

Upgraded or new predictive models can be easily included in the application for further use, validation, and model comparison. The models included in the current version are:

- Growth:
  - Primary models: Baranyi, Gompertz and Linear Model (three linear phase, biphasic or linear).
  - Secondary models: Cardinal, Gamma and Ratkowsky type models.
- Survival / Inactivation:
  - Primary models: Weibull, Baranyi and Linear Model
  - Secondary models: Polynomial, Linear.

In general, model submission is realized in a simple and intuitive way by means of a dialog box (see Figure 5), followed by alert messages to the user.



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

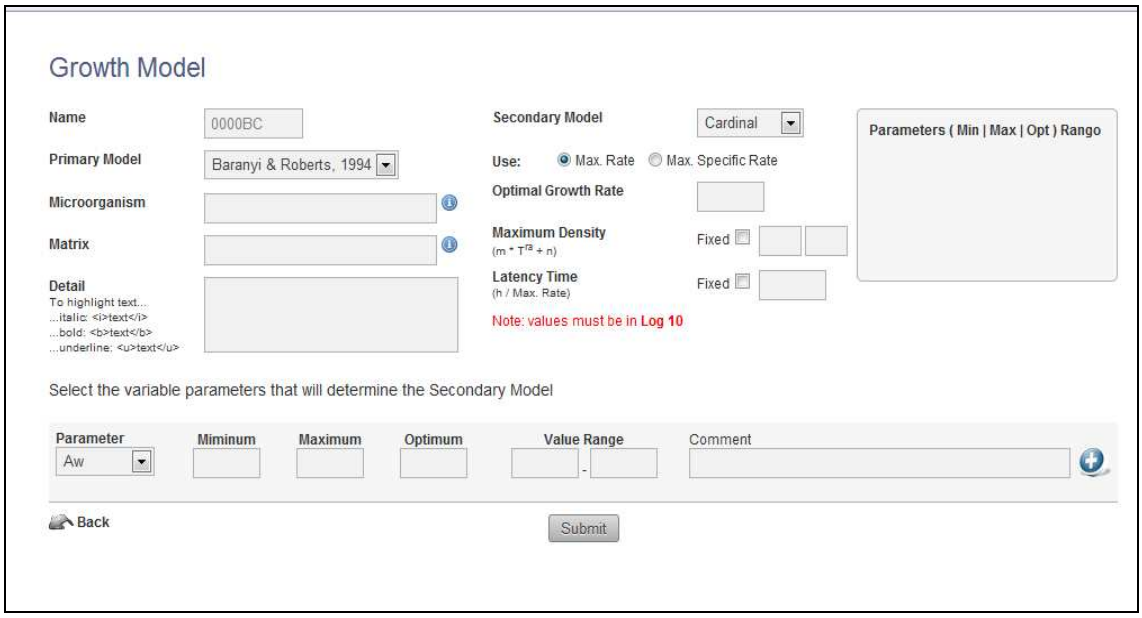


Figure 5. Dialog box for the inclusion of new models.

The field “name” in the dialog box references each model with a specific number, which is stored in the model database.

Details about the predictive model can be added in the field “Details”, e.g. publication reference, material and methods, pre-incubation conditions, etc.

The parameters of the secondary model are selected in the bottom row of the dialog box (minimum, maximum, optimum values, value range and additional comments).

If the maximum growth rate is to be expressed in ln, the maximum specific growth rate option has to be set; otherwise absolute values will be obtained.

Maximum density population and lag time can be fixed by the user, or be calculated through the use of linear equations.

Food

The software tool includes all food matrices selected in the BASELINE project. The food matrices are described in Deliverables 6.2 and 6.5.

Microorganisms

The software tool includes all microbial hazards selected in the BASELINE project. The hazards are described in Deliverables 6.2 and 6.5.

Environmental Factors

Environmental factors included in the software tool are temperature, pH, and water activity. New factors can be added in the case of additional effects or interactive terms in the model equation.

Figure 6 shows a list of foods selected by the BASELINE project, which are included in the model database of the software tool.



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

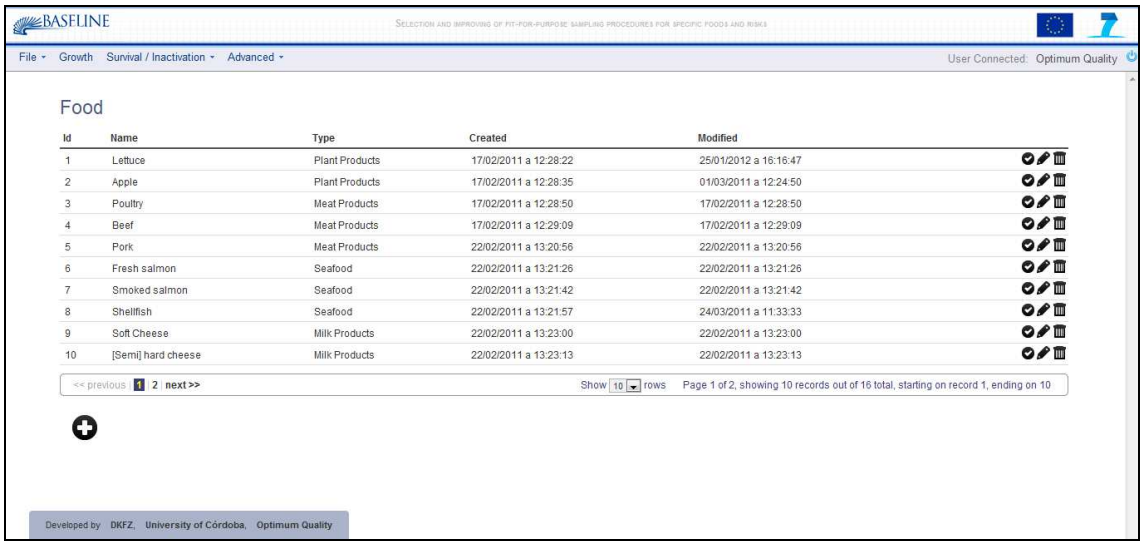

Figure 6. Example of foods selected by the BASELINE project, which are included in the software tool.

## Programming language

### *PHP*

The application is developed in PHP, which is a general-purpose server-side scripting language originally designed for web development to produce dynamic web pages. Among one of the first developed, it is embedded into an HTML source document, rather than calling an external file to process data. The code is interpreted by a web server with a PHP processor module, which generates the resulting web page.

### *CakePHP*

CakePHP is an open source framework for producing web applications, which is written in PHP. It is based on Model-View-Controller (MVC), an architectural pattern used in software engineering, which is often utilized by applications in need to maintain multiple views of the same data.

The MVC pattern hinges on a clear separation of objects in one of three categories - models for the maintenance of data, views for displaying all or part of the data, and controllers for handling events that affect the model or view(s).

Various programming languages (e.g. Pythons, Java, C++, etc) can be easily imported into the application.

### *XHTML*

Presented as a web platform, the interface (composition of views) will be conducted using XHTML (e**X**tensible **H**yper**T**ext **M**arkup **L**anguage). It is a family of XML markup languages that mirror or extend versions of the widely-used Hypertext Markup Language (HTML), the language in which web pages are written.

The software tool uses **XHTML 5.0**, which has both a regular text/html and XML serializations, which are known as XHTML5. In addition to the markup language, the specification includes a number of application programming interfaces. The Document Object Model is extended with APIs for editing, drag-and-drop, and data storage and network communication.

Use **UTF-8** (UCS Transformation Format 8-bit) is a variable-width encoding that can represent every character in the Unicode character set. It was designed for backward compatibility with ASCII to avoid the complications of endianness (ordering of individually addressable sub-components within the representation of a larger data item as stored in external memory) and byte order marks (compare UTF-16 and UTF-32). For these and other reasons, UTF-8 has become the dominant character encoding for the World-Wide Web.



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

## *JavaScript*

JavaScript is a prototype-based scripting language that is dynamic, weakly typed and has first-class functions. It is a multi-paradigm language, supporting object-oriented, imperative, and functional programming styles.

## *jQuery*

jQuery is a cross-browser JavaScript library designed to simplify the client-side scripting of HTML. Used by over 52% of the 10,000 most visited websites, jQuery is the most popular JavaScript library in use today. jQuery is a free, open source software.

Since web platforms depend heavily on connection and bandwidth, Ajax and JSON web development technologies were used to give greater fluidity to the site and carry out consultations in the database without reloading the page:

## *AJAX*

Ajax (Asynchronous JavaScript and XML) is a group of interrelated web development techniques used on the client-side to create asynchronous web applications. With Ajax, web applications can send data to, and retrieve data from a server asynchronously (in the background) without interfering with the display and behaviour of the existing page. Data is usually retrieved using the XMLHttpRequest object. Despite the name, the use of XML is not needed (JSON is often used instead), and the requests do not need to be asynchronous.

## *JSON*

JSON (JavaScript Object Notation) is a lightweight text-based open standard designed for human-readable data interchange. It is derived from the JavaScript scripting language for representing simple data structures and associative arrays, called objects. Despite its relationship to JavaScript, it is language-independent, with parsers available for most languages.



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

## Data Warehouse

### *Database*

Data management is based on MySQL, a relational database management system running as a server providing multi-user access to a number of databases. MySQL is named after the developer's Michal Widenius' daughter My and SQL (Structured Query Language).

The MySQL development project has made its source code available under the terms of the GNU General Public License as well as under a variety of proprietary agreements. MySQL was owned and sponsored by a single not-for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation.

MySQL is a popular choice of database for use in web applications, and is a central component of the widely used LAMP web application software stack – LAMP, which is an acronym for "Linux, Apache, MySQL, Perl / PHP / Python". MySQL is used in some of the most frequently visited websites on the Internet like Google, Wikipedia, Facebook, YouTube, etc.

### *Web Server*

The software tool is hosted on an Apache Server running on Linux.

The Apache HTTP Server, commonly referred to as Apache web server software, is notable for playing a key role in the initial growth of the World Wide Web, has since evolved to rival other web servers in terms of functionality and performance. In 2009, it became the first web server software to surpass the 100 million milestone website.

It is developed and maintained by an open community of developers under the auspices of the Apache Software Foundation. The application is available for a wide variety of operating systems, including UNIX, FreeBSD, Linux, Solaris, Novell NetWare, AmigaOS, Mac OS X, Microsoft Windows, OS/2, TPF, and eComStation. Typically Apache is run on a Unix-like operating system. Released under the Apache License, Apache is open-source software.



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

## Design of the software tool GUI

The General User Interface (GUI) will be easy-to-use. It is crucial that the tool helps the user to design and handle the different variables included.

### *Style*

In terms of style and design, the web portal supports CSS 3.0, the most recent and powerful CSS version to separate content from web style, thus fostering a potential addition of masks or designs without the need of structural changes.


### *CSS*

CSS (Cascading Style Sheets) is a style sheet language used to describe the presentation semantics (the look and formatting) of a document written in a markup language. Its most common application is to style web pages written in HTML and XHTML, but the language can also be applied to any kind of XML document, including plain XML, SVG and XUL. CSS is designed primarily to enable the separation of document content (written in HTML or a similar markup language) from document presentation, including elements such as layout, colours, and fonts. The separation improves content accessibility, provides more flexibility and control in the specification of presentation characteristics, enables multiple pages to share formatting, and reduces complexity and repetition in the structural content. It can also be used to allow the web page to display differently depending on the screen size or device on which it is being viewed. While the author of a document typically links that document to a CSS style sheet, readers can use a different style sheet, perhaps one on their own computer, to override the one the author has specified.

### *Navigators*

Multiple browser compatibility is an important quality feature of a website. It refers to the ability of a web application, web page, HTML construct, or client-side script to support all web browsers. The web tool described in Deliverable 6.6 is developed as a *multi-browser* tool completely portable and operational from the most famous Firefox, Safari, Internet Explorer, Chrome and Opera.



*Project funded under the Food, Agriculture and Fisheries and Biotechnology theme (KBBE)*

SEVENTH FRAMEWORK PROGRAMME

## References

Hervás, C., Zurera G., García, R.M., and Martinez, J. (2001). Optimization of computational neural network for its application to the prediction of microbial growth in foods. Food Science and Technology International 7:159–163.

Ross, T. (1996). Indices for performance evaluation of predictive models in food microbiology. Journal of Applied Bacteriology 81:501–508.