Use case

A use case diagram in XMI format is available here, produced using Umbrello 4.0.3
(http://uml.sourceforge.net/)

Actors

Database owner is the owner of a biomedical database that can be used for research
purposes. Data analyst knows how to perform an analysis method that can potentially
be used on a biomedical database for research purposes. Statistical software
developer develops a graphical user interface (GUI) for a statistical package.

Database posting

Short description: This use case describes how a database owner posts data and meta-
data on a biomedical database.

Preconditions: A database owner has information on the biomedical database as whole
as well as information on each variable in the database.

Basic flow: The database owner goes to a Web site, say Dataverse (http://thedata.org/),
where data and meta-data can be uploaded through a Web interface. Data are
uploaded in a standard format (e.g., a replication data set produced through zelig -
http://gking.harvard.edu/zelig/docs/zelig.pdf), which contains variable-level information
on names, description, and data quality information. Database owner checks all the
variable-level information on a paginated display and makes corrections by editing the
document as needed. Variable-level information is saved. Information on the
biomedical database as a whole is entered through a web form, one by one.
Information is saved.

Alternative flow: The variable-level information is entered into a spreadsheet in a
standardized format with separate columns for variable names, descriptors and quality
information. The spreadsheet is uploaded to the system, the information checked and
modified by database owner as needed, and then saved

Postconditions: The database-level and variable-level information on the biomedical
database is available for display in the Web-based system.

Formulating a research question

Short description: This use case describes the database owner formulating a research
question that is attached to the biomedical database meta-data

Preconditions: The database owner knows from previous experience the research
question that are most unique about the biomedical database and can establish the
main outcomes and predictors involved in this question.

Basic flow: The database owner goes to the Web site containing the meta-data and
data about the biomedical database. The database owner adds a general description of
the research questions, one at a time. The outcome and main predictor variables is
then individually selected from the list of available database-variables.

Alternative flow: If the research question uses a transformation from the original format
in the biomedical database, e.g. if a continuous variable is to be transformed to an
ordinal variable, then the database owner makes a modification to the meta-data of that
variable.

Postconditions: The meta-data about the research questions is attached to the meta-
data of the biomedical database.

Posting an analysis method

Short description: This use case describes the data analyst posting information on a data analysis method.

Preconditions: The data analyst knows a data analysis method at five different levels: A general description of the data analysis method contained in a paragraph, the required input variables for the data analysis method, the resulting output from the data analysis method, previous publications that serve as examples of the data analysis method, and literature that explains the data analysis method in a progressive level of complexity.

Basic flow: The data analyst goes to the Web site containing a form, e.g. a semantic media wiki (http://www.semanticweb.org/wiki/Semantic_MediaWiki) with the semantic forms plugin (http://www.mediawiki.org/wiki/Extension:Semantic_Forms), with the five fields required for description of the data analysis method. The data analyst adds information on each of the fields in relation to the data analysis method of her expertise. The information is saved.

Postconditions: The meta-data about the data analysis method is stored.

Finding a data analyst

Short description: This use case describes the database owner finding a data analyst who can conduct the analysis for her research question

Preconditions: The information on the database has been posted along with the research question. The information on multiple data analysis methods have been posted.

Basic flow: The database owner receives an RSS feed connecting her to the data analysis method as well as all data analysts capable of conducting that data analysis

Alternative flow: The database owner receives the same information described in the basic flow, but through an e-mail.

Postconditions: The database owner has information on the data analysis method and the data analysts. She can then contact the data analyst of her choice.

Finding a database

Short description: This use case describes the data analyst finding a biomedical database that can be used for an analysis with the methods she knows

Preconditions: The information on the data analysis method has been posted. The information on multiple biomedical databases along with their respective research questions have been posted.

Basic flow: The data analyst, e.g. a graduate student in biostatistics or any other quantitative discipline in search of a suitable thesis or dissertation project, receives an RSS feed connecting her to the biomedical databases and respective database owners through which the data analysis can be executed.

Alternative flow: The data analyst receives the same information described in the basic flow, but through an e-mail.

Postconditions: The data analyst has information on multiple biomedical databases that would be suitable for use with her data analysis method. She also has information on the respective database owners, and can then contact the ones of her choice.

Creating a statistical GUI

Short description: This use case describes a statistical software developer creating a graphical user interface for a statistical package.

Preconditions: The statistical software developer has the code to receive the input data required to run each of the data analysis methods

Basic flow: The statistical software developer maps the input fields required by each of the data analysis methods to respective elements of the data transformation ontology. These same elements are then mapped to the fields in the graphical user interface that will receive the data input.

Postconditions: The fields provide the data required to run each of the data analysis methods.