Institute of Mathematics and Computer Science European Bioinformatics Institute

SIMS User Guide

Version 2.14

Accessing SIMS system

To access Sample Information Management System you need a valid login name and a password. If you think that you should have access to this page, but don't have login name/password you should contact the System Administrator.

Different users might have different access rights. Currently access rights are managed at *technology level* (the user can access only these Technologies access to which have been granted to him) and at *entry ownership* level - depending from ownership particular entries could be available either fully or partially or be hidden for a particular user. Also the access rights determine whether the user will be able to create new entries.

There are several pages or other functionality of the system that might be unavailable or partially available to users without sufficient access rights.

Welcome page

Welcome page can be displayed as a first page after login in the system. Whether this is the case depends from the settings of user account and can be modified by user by clicking on *User settings* tab. In addition, the page will be unconditionally displayed to users with administrator rights if there are some external tasks that need to be performed.

The main functionality of this page is the option to see and change if necessary the *default group* that will be used in current session, and consequently, what access rights other users will have to entries created within this particular session. The access rights specific to the selected *default group* are shown in table below.

User settings

User settings page allows the user change settings of several values associated with this user. The changes are permanent and will be preserved after the termination of the current session. There are some values (e.g. *Page size*) of user settings that can be configured only by a particular user on this page and some values (e.g. *Password*) that can be changed also by System Administrator in **Administrator tables** page.

Currently the settings of the following values can be changed by the user:

- Page size of lists. Number of data entries that will be shown in one page in list format. The default initial setting is 1000; a user might consider setting a lower value if using a low-speed Internet connection.
- Show full ids. Entry Ids are being generated automatically and usually are in the form "[prefix]-[number]". This setting determines whether the number part of Id will be shown with or without leading zeroes (in current version when zeroes are shown numbers are expanded up to 6 digits).
- Default group. The user can change his default group, selecting it from a list of groups he is a member (the membership to groups is assigned by System Administrator in **Administrator tables** page). The value of default group determines the group ownership that will be assigned to new "View" level (Studies) or First level (Experiments) entries created by this user. Changing this value doesn't affect group ownership of existing entries that have been created by this user.
- Change password. Allows the user to change his password.

First level (Persons)

List of Persons page allows to view/edit/delete dall Persons currently in database. If your access rights don't allow to access a particular Person even in *view only* mode, only the value of Person ID will be shown, all other values will be replaced by **.

Creating/viewing/editing/deleting Persons

- To add a new Person to database click on New Person link.
- To edit data about an Person click on the corresponding *edit* icon . If you have limited access rights one of the following icons will be shown instead: (the Edit dialog is available in *view only* mode), (you don't have access to Person).
 To create a new Person with the parameters copied from the existing one click on the
- To create a new Person with the parameters copied from the existing one click on the corresponding *clone* icon . If you don't have access rights to a particular Person, the icon will be replaced by .
- To delete entry about a particular Person click on the corresponding *edit* icon and then press *Delete* button.
- To change simultaneously the values of some properties for several Persons click on *Edit batch* link. You will have the option to select the Persons for editing and then, after pressing *Edit batch* link, you will be directed to edit dialog which allows to change Person properties.

The field **Person Id** depending from SIMS configuration can have either an arbitrary string value (up to slightly less than 200 characters - keeping in mind that derived Ids for Samples and Aliquots have to satisfy the same limit), or a control digit (character) might be used, in which case checksum of all digits must give the right value mod 10. In the latest case generally it is assumed that PMT tool is used for ID generation.

All Person Ids should be be unique.

Relative fields allow you to define relations ("parent", "child", "twin" etc) between persons. This is possible only between persons attributed to the same data source, the possible canditates for relatives can be selected from a list.

As many relatives as you wish can be defined, but to do this, after adding a new Relative you should press "Update" button and then enter Edit person dialog again. If you chose a person Y as a relative for person X, a reverse relation is added for person Y (e.g. if you select Y as a "child" of X, X automatically becomes "parent" of Y). However these relations are strictly binary with no further implicit assumptions. E.g. even if X is defined as a "sibling" of "Y" and "Y" as a "sibling" of "Z", "X" will not automatically become a sibling of "Z". Similarly, you are free to define that "X" is a "child" of "Y" and "Y" is a child of "X".

To delete a relation, change it to empty field in list of relative candidates.

Be careful - if you attribute a person to different data source, all relations for this person will be deleted!

The meaning of other fields hopefully should be self-apparent.

Second level (Visits)

List of Visits page allows to view/edit/delete data either for all Visits currently in database or Visits attributed to a particular Person. If your access rights don't allow to access a particular Visit even in *view only* mode, only the value of Visit ID will be shown, all other values will be replaced by **.

Creating/viewing/editing/deleting Samples

- To add a new Visit to database click on *New Visit* link. To be able to do this you must be viewing visits attributed to one particular person and new visit will be defined for that person. This includes also the option of adding of several Visits simultaneously (see Batch upload help).
- To edit data about a Visit click on the corresponding *edit* icon . If you have limited access rights one of the following icons will be shown instead: (the Edit dialog is available in *view only* mode), (you don't have access to Visit).
- To create a new Visit with the parameters copied from the existing one click on the corresponding *clone* icon . If you don't have access rights to a particular Visit, the icon will be replaced by .
- To delete a particular Visit entry click on the corresponding *edit* icon and then press *Delete* button.
- To change simultaneously the values of some properties for several Samples click on *Batch* link. You will have the option to select the Samples for editing and then, after pressing *Edit selected* link, you will be directed to edit dialog which allows to change Visit properties (see Batch edit help).

The field **Visit Id** in current version is not editable and is assigned automatically. The id will look like this:

"Person Id (without "-" and control digit)" + "-" + "CV dependent substring (if used in current SIMS configuration)" + "the number (starting from 1) of sample belonging to the same CV dependent substring".

All Sample Ids attributed to a particular Person will be unique.

The meaning of other fields hopefully should be self-apparent.

Third level (Samples)

List of Samples page allows to view/edit/delete data either for all Samples currently in database, Samples attributed to a particular Person, or Samples attributed to particular Visit. If you don't have access rights to a particular Sample even in *view only* mode, only the value of Sample ID will be shown, all other values will be replaced by **.

Creating/viewing/editing/deleting Samples

- To add a new Sample to database click on *New Sample* link. To be able to do this you must be viewing Samples attributed to one particular Visit and new Sample will be defined for that Visit. This includes also the option of adding several Aliquots simultaneously (see Batch upload help).
- To edit data about an Aliquot click on the corresponding *edit* icon . If you have limited access rights one of the following icons will be shown instead: (the Edit dialog is available in *view only* mode), (you don't have access to Sample).
- To create a new Sample with the parameters copied from the existing one click on the corresponding *clone* icon . If you don't have access rights to a particular Samples, the icon will be replaced by .
- To delete a particular Sample entry click on the corresponding edit icon \square and then press Delete button.
- To change simultaneously the values of some properties for several Samples click on *Batch* link. You will have the option to select the Samples for editing and then, after pressing *Edit selected* link, you will be directed to edit dialog which allows to change Sample properties.

The field **Sample Id** in current version is not editable and is assigned automatically. The id will look like this:

"Visit Id" + "-" + "string of letters (starting from A) corresponding to the current number of Sample belonging to this Visit".

All Sample Ids attributed to a particular Visit will be unique.

The meaning of other fields should be self-apparent.

There are two options available from **Add Sample** dialog, if you wish to add several Samples with the same characteristics:

- After defining the Sample press *Create* button. The currently entered Sample will be added to database.
- After defining the Sample press *Create multiple* button, after selecting a number (2-20) from the list. The specified number of Samples with the same properties (apart from Sample Ids) will be added to database.

Batch edit and upload

Batch edit

In *Batch edit* mode there is a checkbox for each of entries' fields. Check the ones you wish to change for all entries in the batch. The fields left unchecked (displayed gray) can hold different values for each entry and the checked ones (displayed white) will hold the changed value for all entries.

Batch upload

Upload [1st level (Persons) / 2nd level (Visits) / 3rd level (Samples)] page allows to upload multiple entries from a single tab-delimited file. The file must contain a header row with recognized column names (these consist of a level prefix followed by "." and the value corresponding to "Upload name" field in metadata configuration tables - e.g. *Person.Gender*) followed by rows corresponding to uploadable entries. Upload names of several "special" fields are not configurable in metadata tables, but given in "systemConst.xml" file (also level prefixes are defined there). Also, there is extended syntax options for columns describing uploadable files (see below).

To obtain a correct header for a particular level, you can download tab-delimited files containing only this header by clicking on *Download ... header* or *Download ... header* (*minimal*) links; these will provide correspondingly either a header containing all recognized column names, or a "minimal" header containing only names of mandatory columns.

There are three "checkbox" options that you can select for batch uploads at all levels:

- *Generate identifiers automatically*. If checked, automatic id generation of IDs will be used and the corresponding column, e.g. Studies.ID, either ignored or disallowed (depending from *Allow unrecognized columns* setting. ID generation will not work at 1st level (View), if the ID generation at this level is not configured for a particular Technology. If unchecked, column with IDs (e.g. Studies.ID) is mandatory and it is required that IDs must be unique (but see the next option).
- **Allow duplicate rows**. If checked, then entries with identical IDs are allowed in upload file, however only a single (assume that randomly selected) entry will be uploaded in system. Generally this option could be useful only in combination with unselected *Generate identifiers automatically* and selected *Allow unrecognized columns*. The expected use case is uploading a single file containing entries of all three levels in this case the "whole collection" can be uploaded by uploading the same file correspondingly at 1st, 2nd and 3rd level.
- **Allow unrecognized columns**. If checked, the upload file can contain arbitrary column names that are not recognized by the system. The most typical usage could be uploading a file with columns for all three SIMS levels (most of them of recognized only at their own level, apart from parent level ID, which could be used also at child level).

In addition, at 2nd and 3rd levels there is a radio button allowing to specify one of the following options:

- uploading all entries as child entries for parent entry shown (and selectable) in listbox;
- uploading each entry by specifying its own parent. In this case a column with parent ID must be present in upload file (e.g. when uploading 2nd level entries (Visits) file must contain also column *Person.ID*.

Locations of upload files.

In all situations the page will contain the following two *Browse* buttons:

- **Parameter file**. This allows you to browse your local file system and select tab-delimited file for upload.

- Local zipped data. This allows to provide a zipped archive containing all the files that are referenced in parameter file (if any). This ZIP archive is also uploaded form your local file system. Whilst technically this is a simplest option for file upload, http protocol might impose severe limitation of archive size that can be uploaded in this way.

If the system *and* you access rights are configured for pre-uploading files via FTP, two additional options will be provided:

- *FTP zipped data*. Instead of uploading ZIP archive from your local system, you can browse for archives that are already preuploaded on the server.
- *FTP directory*. This provides a similar pre-uploading option, however in this case you have to specify a pre-uploaded directory containing all the referenced files. This option may be more convenient for very large files that are hard to pack in a single archive. A notable difference from the previous option is that in this case you *shouldn't reference the same file more than once*. Ignoring this will result in a partially corrupt upload, requiring substantial manual correction!

If data/supplementary files are pre-uploaded via FTP, there will be some delay before Assays will be available for editing (will be shown instead of edit icon). (Edit dialogs still will be available for users with administrator rights, although note that data files still will not be copied into repository and changing file references are likely to cause errors.) If icon shown, this means that there have been some problems with copying data and supplementary files.

Referencing data and supplementary files.

The column names in header for referencing files that have to be uploaded are defined in "systemConst.xml" file and will be shown in complete header obtainable via *Download ... header* link. E.g., the column names might be *Person.DATA_F*, *Person.DF_TYPE* for data files, and *Person.SUPPL_F*, *Person.SF_TYPE* for supplementary (protocol) files. The following features have to be taken into account:

- columns for data or supplementary files will only be available (or recognized), if linking of files to entries is configured for particular Collection and particular level;
- the file name (given e.g. in *Person.DATA_F*) must contain a full path to file either in ZIP archive or in pre-uploaded directory.
- the file type (given e.g. in *Person.DF_TYPE*) is a memo field describing the file. The list of recognized values can be checked in Edit dialog that corresponds to uploaded entries (there could be none).
- there is extended syntax allowing to upload multiple files for a single entry. The names of columns that reference files can contain additional suffix of the form #file_number, where "file_number" is an arbitrary string which distinguishes different columns. E.g. to upload two files for a single entry two columns either headers Person.DATA_F#1, Person.DATA_F#2, or headers Person.DATA_F#SS, Person.DATA_F#17 could be used. The same suffixes can be used to specify corresponding file types, e.g. Person.DF_TYPE#1, Person.DF_TYPE#2.

Data filtering

Filter settings page allows to search for Study Groups/Assays with values of selected fields equal to the values defined in search page filters.

The fields, which are available for filtering, are defined in **parameters** and **common_parameters** tables. To define a filter for a particular field, enter a filtering expression for this field. The syntax of filtering expressions is as follows:

- symbol "%" matches any substring (including empty substrings),
- symbol "_" matches any single symbol,
- all other symbols match themselves.

The filtering is done by SQL queries and for some particular datatypes (boolean, lists of numbers, etc.) the syntax of "right" filtering expressions may not be intuitive. For fields of fixed vocabulary types a drop-down list is provided instead of filtering expression.

Filters are conjunctive, i.e. if a field is included in search, database entries satisfying filter values for this field **AND** filter values for all other fields will be found.

To display all database entries press *Cancel filter* button either on Study Groups or on Assays page.

Defining tabular views and reports

Create report

You can create a tab delimited report file in one of two ways.

By clicking the *Report* link directly you will get the values of all visible entries.

On 2nd (Studies) and 3rd (Assays) levels You can determine which entries are visible by choosing to view only entries corresponding to a particular 1st (View) and 2nd (Study) level entry.

Alternately you can press *Batch* first and only click *Report* after picking out the entries you want.

To determine what columns and in what order will appear in the report use **Define View Table** page.

Configure visible columns

Define ... **view** page allows to select which columns and in which order will be displayed in View, Study or Assay level pages in list (tabular) format. It also defines which fields will be exported by **Report** function (which actually might be the main reason for changing these settings). The level for which column display options can be changed corresponds to the level from which **Define** ... **view** page has been entered (by using **Columns** function).

To change the settings, enter numbers of columns in which the fields of interest should be displayed. Columns are numbered by rational positive numbers which should be entered in two adjacent textboxes containing correspondingly integer and fractional parts. On view pages columns will be shown from left to right according to increasing ordering of column numbers. Column number 0 means that the corresponding field will not be displayed.

To apply view settings press *OK* button. Pressing of *Tune* button will "normalize" the column numbering, i.e. columns will be assigned consecutive integer numbers starting from 1. Also, they will be re-ordered according to their numbers. Normalizing and re-ordering occurs also on re-entering **Define** ... **view** page.

Settings are preserved only within the current session. After the next login the default settings will be restored (the later can be configured from **Configuration tables** page by users having sufficient access rights to do so).

Vocabulary tables

Vocabulary tables page allows to view/add/edit/delete entries in fixed vocabularies.

Vocabularies that are available for editing depends from currently selected technology. Shown are vocabularies that are defined for the particular technology (this is done in **Configuration tables** page) as well as *common* vocabularies that are available for all technologies (these are defined separately in **Administrator tables** page).

Editing options include adding new entries to the vocabulary, deleting existing entries, renaming existing entries and changing the order in which entries will be available for selection. To add a new entry click on a *new* link for a specific vocabulary. To edit/delete entry click on *edit* icon for a specific vocabulary. To change the display order use correspondingly vand cons.

Editing of vocabularies should be done with some care - deleting of an existing entry and subsequent adding of new entry with the same name may lead to wrong values being associated with data that are already stored in the system. However changing the order in which entries are displayed for selection is safe operation in such context and doesn't affect the data already stored in the system.

The length of vocabulary tables is limited by a 4000 character long encoding of vocabulary in database.

Depending from your access rights you might have only view only access to **Vocabulary tables** page or the page might be unavailable for you.

System administration

Administrative tables page allows to access and (depending from your access rights) modify configuration tables of the system. Generally these tables can be divided in two groups: user and their access rights management and settings that define how data is displayed at Person, Sample and Aliquot level.

At the top of the page there are links: Export matadata and vocabularies and Import matadata and vocabularies. Export matadata and vocabularies allows exporting in XML format the content of all controlled vocabularies as well as Person metadata, Sample metadata and Aliquot metadata configuration tables. Import matadata and vocabularies allows importing of such XML file in SIMS that is unconfigured as yet. One should take care of importing XML file that compatible with a particular SIMS instance. Generally metadata import should be done just once for newly installed SIMS instances, however re-importing the same file shouldn't make any changes to the system.

The tables shown on this page and their meaning are described below.

Table Person metadata

Table specifies how data are displayed in **List of Persons** page and the corresponding Add/Edit/View dialogs.

The configurable fields always include the following: *visibleName* (Person Id), *relations1* ("meta-field", used to display information about relatives), *comment* (string of up to 4000 characters), *creator* (FK to *Users* table), *modifier* (FK to *Users* table), *createDate* (datatime) and *modifDate* (datatime).

In addition there could be a variable number of the following types of fields present: vtxx (FK to controlled vocabulary), ftxx (number of type float), itxx (number of type integer or boolean (displayed as checkbox)), stxx (string of either up to 200 characters or up to 4000 characters, this depends from database configuration), dtxx (date and/or time). The number of these fields depends from configuration from of a particular SIMS instance. The available range (values for xx) are from 0 to 99.

There are notable difference between *vtxx* and other types of fields: controlled vocabularies are shared between all levels (Person, Sample and Aliquot), whilst other types are individual for each level.

For SIMS instance to work properly the table need to contain **exactly one** entry for each of the following fields: *visibleName*, *relations1*, *comment*, all available fields of types *ftxx*, *itxx*, *stxx* and *dtxx*, and these fields of type *vtxx* which are being used at current level. The configuration of fields *creator*, *modifier*, *createDate* and *modifDate* is optional. The table contains the following configurable fields (in this order):

- Property name the field (selectable from drop-down list) being configured,
- Short name shown as column name in List of Persons page,
- Long name name shown on the left side of the field in Add/Edit/View dialogs,
- Upload name name used in header of batch upload file,
- View column number column number in List of Persons page (not shown, if 0),
- Sortable if true, the entries in **List of Persons** can be shorted by the values of this parameter,
- Editable currently unused,
- Show in report currently unused,
- Comment informal comment.

Table Sample metadata

Table specifies how data are displayed in **List of Samples** page and the corresponding Add/Edit/View dialogs.

The configurable fields always include the following: *visibleName* (Sample Id), *comment* (string of up to 4000 characters), *creator* (FK to *Users* table), *modifier* (FK to *Users* table),

createDate (datatime) and modifDate (datatime).

In addition there could be a variable number of the following types of fields present: vtxx (FK to controlled vocabulary), ftxx (number of type float), itxx (number of type integer or boolean (displayed as checkbox)), stxx (string of either up to 200 characters or up to 4000 characters, this depends from database configuration), dtxx (date and/or time). The number of these fields depends from configuration from of a particular SIMS instance. The available range (values for xx) are from 0 to 99.

There are notable difference between *vtxx* and other types of fields: controlled vocabularies are shared between all levels (Person, Sample and Aliquot), whilst other types are individual for each level.

For SIMS instance to work properly the table need to contain **exactly one** entry for each of the following fields: *visibleName, comment,* all available fields of types *ftxx, itxx, stxx* and *dtxx,* and these fields of type *vtxx* which are being used at current level. The configuration of fields *creator, modifier, createDate* and *modifDate* is optional.

The table contains the following configurable fields (in this order):

- Property name the field (selectable from drop-down list) being configured,
- Short name shown as column name in List of Samples page,
- Long name name shown on the left side of the field in Add/Edit/View dialogs,
- Upload name name used in header of batch upload file,
- View column number column number in List of Samples page (not shown, if 0),
- Sortable if true, the entries in **List of Samples** can be shorted by the values of this parameter,
- Editable currently unused,
- Show in report currently unused,
- Comment informal comment.

Table Aliquot metadata

Table specifies how data are displayed in **List of Aliquots** page and the corresponding Add/Edit/View dialogs.

The configurable fields always include the following: *visibleName* (Aliquot Id), *comment* (string of up to 4000 characters), *creator* (FK to *Users* table), *modifier* (FK to *Users* table), *createDate* (datatime) and *modifDate* (datatime).

In addition there could be a variable number of the following types of fields present: vtxx (FK to controlled vocabulary), ftxx (number of type float), itxx (number of type integer or boolean (displayed as checkbox)), stxx (string of either up to 200 characters or up to 4000 characters, this depends from database configuration), dtxx (date and/or time). The number of these fields depends from configuration from of a particular SIMS instance. The available range (values for xx) are from 0 to 99.

There are notable difference between *vtxx* and other types of fields: controlled vocabularies are shared between all levels (Person, Sample and Aliquot), whilst other types are individual for each level.

For SIMS instance to work properly the table need to contain **exactly one** entry for each of the following fields: *visibleName, comment,* all available fields of types *ftxx, itxx, stxx* and *dtxx,* and these fields of type *vtxx* which are being used at current level. The configuration of fields *creator, modifier, createDate* and *modifDate* is optional.

The table contains the following configurable fields (in this order):

- Property name the field (selectable from drop-down list) being configured,
- Short name shown as column name in List of Aliquots page,
- Long name name shown on the left side of the field in Add/Edit/View dialogs,
- Upload name name used in header of batch upload file,
- View column number column number in List of Aliquots page (not shown, if 0),
- Sortable if true, the entries in **List of Aliquots** can be shorted by the values of this parameter,
- Editable currently unused,
- Show in report currently unused,
- Comment informal comment.

Table Users

Defines Users of the system. The table contains the following fields:

- User description informal User description,
- Login name login name,
- Administrator rights administrator level rights for this user, the options are are non-accessible (no access to **AdminTables** page), view only, limited editing (actually meaning almost full access, some configuration printouts and **Pending tasks** table not shown) and full editing,
- User password password (shown in MD5 hashed form, but can be reset from this dialog),
- *Id prefix* a string which will be used as part of Id when creating new Experiments or Studies,
- Export is allowed if true, the User can export data in XML or tab-delimited form based on XML configuration templates (currently not used in SIMS),
- FTP is allowed if true, User can access files preuploaded via FTP (if FTP option is enabled for this AIMS configuration),
- FTP directory subdirectory from which data and supplementary files will be listed (if FTP option is enabled for this AIMS configuration),
- Default user group the User group, which owns Studies and Experiments created by this user. When a new user is created all User groups is available in drop-down list and initially default User group is selected. When data about already existing user is edited this list contains only User groups the User is member of,
- Comment informal comment.

At least one User group must be present (and one User is created automatically when newly installed AIMS is used for the first time). Users having *Administrator rights* at *limited editing* or higher level automatically have full access to all entries and **Vocabularies** and **Configuration** pages regardless of access rights configurations.

Table User rights

Defines granting of access rights at user level. The table contains the following fields:

- Granted by User to Studies and Experiments owned by whom the access is granted,
- Granted to User to which the access is granted,
- User rights the level of rights granted, options are non-accessible, view only, data files non-accessible, view only, data files accessible and full access,
- Comment informal comment.

Generally User has full access to entries owned by him, access to other entries depend from configuration in *User rights* and *User group rights* tables. User doesn't automatically have access to entries of user group he belongs to - this has to be configured in *User group rights* table.

Table User groups

Defines User groups.

The table contains the following fields:

- Group name name of User group,
- Is default if true this is the User group the newly created user is assigned to as to his default User group,
- Comment informal comment.

At least one User group must be present (and one User group is created automatically when

newly installed AIMS is used for the first time). Exactly one User group should be default - to facilitate this Is default checkbox can only be checked, not un-checked (when another User group is made default, the previous default User group loses that status).

Table User group rights

Defines granting of access rights at user group level. The table contains the following fields:

- Granted by User group to Studies and Experiments owned by whom the access is granted, Granted to User group to which the access is granted,
- User rights the level of rights granted, options are non-accessible, view only, data files non-accessible, view only, data files accessible and full access,
- Comment informal comment.

Generally User has full access to entries owned by him, access to other entries depend from configuration in *User rights* and *User group rights* tables. User doesn't automatically have access to entries of user group he belongs to - this has to be configured in *User group rights* table.

Table User group membership

Describes User groups to which User is a member. The table contains the following fields:

- User User for which membership is defined,
- User group User group to which user belongs,
- Comment informal comment.

When a new user is created, he is automatically assigned to default User group and the corresponding entry is placed in this table. The user can't be deleted from his default group.