ExpressWeb Admin guide

Introduction

This guide explain how to manage users and data.

ExpressWeb use an access management system. Users and Dataset(expression data) belong to groups who gives them access's rights. Only users members of group "admin" and account Admin (SuperAdmin) can import DataSet, create Organism's annotation and manage users .

SuperAdmin had additional rights (Reset Database for example).

Three Master Group (demo,members,admin) are defined and gives the following rights:

- **demo**
    - use Dataset in group 'demo'
    - can display clustering result.
    - min
        - no Profile management
        - no rights to launch cluster jobs or create sub-datasets
- **members**
    - use all Dataset in group 'demo' or 'members'
    - if user belong to another group, gains access to Dataset in this group
    - can launch cluster job, create sub-dataset
    - access to Profile and File management
    - min
        - no access to Dataset management
- **admin**
  same as members
    - plus
        - import annotation. renew Dataset annotation or create new one.
        - admin users . create users, groups.
        - admin tables. remove tables and dependancies.
        - cluster management and skill to kill running process

Principles

Before opening application to users, you need, as Admin, to execute different tasks:
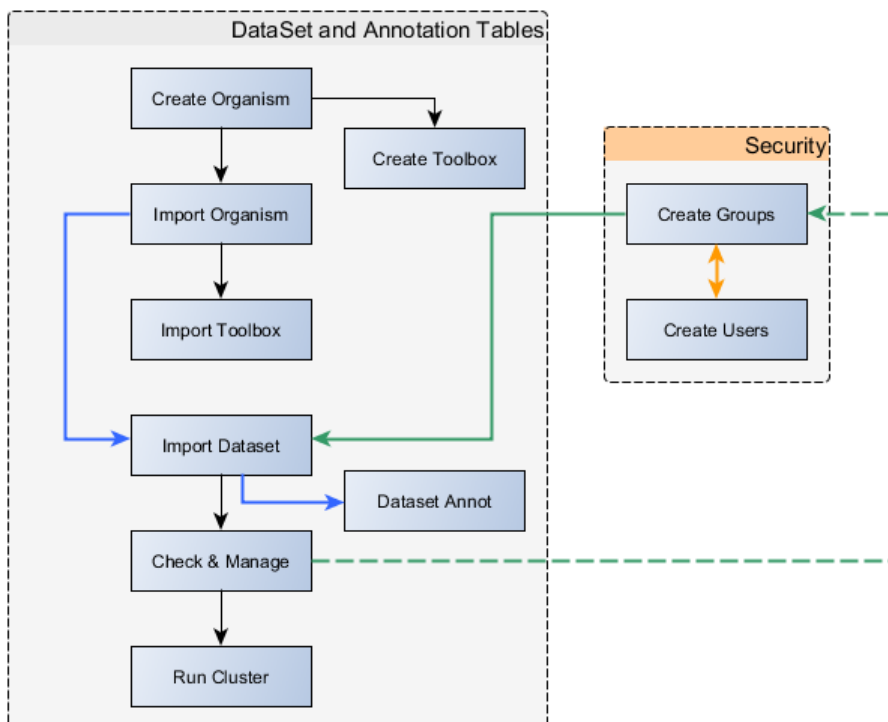
1. Create Organism:
   a. ExpressWeb contains one organisms (Arabidopsis thaliana) in the demo set.
   b. You can remove all demo dataset and table by using as 'Admin', the Reset Db command. Organisms table will be truncated and allow you to start from ground
   c. When you create an organism a Toolbox is automatically created. You may upload data to feed it or not (See Toolbox)
2. Import organism:
   a. Upload personal or Phytozom annotation about your organism. This data will be used to create DataSet annotation (limit number of records when cluster results are displayed)
3. Import Toolbox
   a. Allow additional granularity in genes expression analysis
4. Import Dataset
   a. Import your genes expression data. If successful, an annotation table is created using Organism annotation to extract information corresponding to dataSet's genes.
5. Check and Manage:
   a. Affect or change default group of DataSet.
   b. Create new group
   c. Create user and set group membership
6. Run cluster:
   a. Choose threshold value and launch clustering process

Users and Groups management

Users account Menu

Admin can create users and group via this menu. They can also monitor cluster activity.

User Activity

1. List users currently logged. Username, IP address, date and time of last login
2. Process Activity on cluster. Show running process on cluster
   a. Two instances of 'launch_cluster.sh' , script used to communicate with cluster
   b. One line with 'execute_sh ': cluster script.
      Launch jobs and manage cluster responses.
   c. You can kill a process by clicking on button in this column.
   d. If you want to stop clustering, click on line 3 (sh exexecute_bash.sh).
      Job will end with error status 20 , temporary tables Cluster and Order will be
      removed from database.
3. Show jobs in qsub process.
   a. Display info about job state.
   b. Click on button with qsub Job id (ex:232940) will stop jobs , create an
      EndJob_pid.txt file and add error code 20 at end of Job_pid.txt
4. Information about cluster state
   a. Two buttons will display informations on cluster state:
   b. 'qstat shrt' (A) : show number of lines return by command qstat | wc –l .
   c. 'qstat' (B): list all jobs on the cluster.

User accounts

Select 'Users Accounts' tab and click on 'Manage User Accounts'

Back to Users management

Users

Below is a list of the users.

First Name	Last Name	Login Name	Email	Groups	Status	Action
Admin	istrator	administrator	savelli@lrsv.ups-tlse.fr	admin members	Active	Edit
demo	demo	demo	demo@admin.org	Demo	Active	Edit
Clark	Gable	Rhett	c.gable@gonewwind.com	members Cell	Active	Edit

Create a new user | Create a new group

1. Edit - click on edit link.
   - **Note** Login Name is used to create users directories and store uploaded and saved files. If you change it , you may create several directories ... but user access only directory matching is login.
2. Activate
   - click on 'Status' link. Active or Inactive current login
   - user will receive an email to renew is password.
   - easy way to activate user or deny login. Inactive account can't log in
3. Create user
   - Fill all applicable fields and submit form.
   - Active user. An email will be send to user.

External users may ask for an account via 'Sign Up' menu.

Request will be send to Admin who can activate or not the account.

User Groups

Select 'Users Groups' tab and clik on 'Manage User Groups'

Back to Users Account management

Groups

Below is a list of the groups.

| Groups	Description:	Action
admin	Administrator	Edit
members	General User	Edit
Demo	Demo account	Edit
Cell	Cell team	Edit

Create a new group

1. Edit
   - click on edit link.
   - enter group name and description
   - **Note** The membership of a user to a group is referenced by an internal number. changing group name doesn't affect user membership.
2. Create group (Or from Manage User Accounts/Create a new group)
   - click on create link.
   - enter group name and description
3. Modify users or tables group membership
   - Go to Users account or in Db admin menu.

Database management

From **Db admin** menu, four tabs allow database and data management:
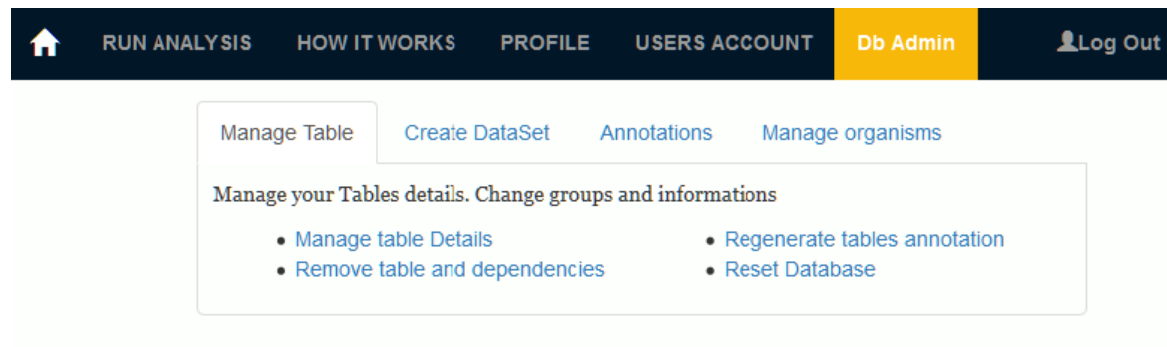
1. Manage Table:
   - Manage table Details: edit table information, Groups membership
   - Remove : delete table, annotations and files generated.
   - Regenerate annotation: update table annotation . Used after Organism update
   - Reset Database: remove all DataSet, annotation, clustering results.
2. Create DataSet : import data and convert to SQL table
3. Annotations: Import/Update Organisms annotation from Phytozom or formatted data
   - Import
     i. Import annotation from Phytozom
     ii. Create annotation: from tabular text file, create your own annotation
     iii. Import Toolbox: generate Gene Toolbox

   - Edit
     i. Upgrade annotation from Phytozom
     ii. Upgrade annotation: from tabular file
4. Manage Organisms: create Organism definition

Manage Table tab

Manage table Details

Display all tables (DataSet, clustering results (Cluster & Order) and Annotation) status.
Tables belong to group(s) (master group are between brackets [admin]) and a specific organism.

TableName: 4 different types of tables will be displayed:

1. <u>Annotation_Number</u> (**Annotation_3**): Annotation table for organism Number.
    - These tables contains annotations for **organisms** genes.
    - Number refer to idOrganisms in Organisms table.
    - These tables are used for create Dataset Annotation. Do not delete !!
    - These tables belong to admin group only.
    - Not visible by members or demo users.
    - Name can't be changed.
2. <u>Annotation_NameOfDataSet</u> (**Annotation_Myco_AnnotTest**): Annotation for NameOfDataSet
    - When user import a DataSet in ExpressWeb, an annotation companion is created from organism annotation.
        - If Dataset contains transcript and Annotation_N not, annotation with transcript Id is created. Same on reverse state.
        - With big organism, annotation file may be huge and may slow down display .
    - Name can't be changed.

3. DataSet(**Myco_AnnotTest) :** Experimental data imported in expressWeb Db.
    - Master group is group of user who upload data in Db.
    - DataSet are always at Root Level.
    - They may belong to different groups and be affected to others groups.
4. DataSet_Result (Myco_AnnotTest_0_9_Cluster,Myco_AnnotTest_0_9_Order):
    - When clustering is apply to a DataSet, two files are created.
        - Cluster : contains Gene_ID, cluster N° and group ref
        - Order: contains expression values for each genes
        - Threshold: _0_9_ indicate that clustering have been made with a threshold of 0.9.
    - DataSet Result are always a Child of DataSet.
    - They belong to DataSet group and can't be affected to others groups

Organism: Name of organism

Submitter: Name of user who create the table.

Groups: Groups affected to table.

- DataSet may have multiple groups.
- Group used for create the table is the Master Group (in brackets [group])

Level: identify relationship between tables. Myco_annotTest_0_9_cluster && Order are childs of Myco_AnnotTest.

Action: Edit

> **Warning:** changes will affect relation between tables.
>
> For a fresh imported DataSet or Organism Annotation , changing Organism or master group as no impact.
>
> But if you change DataSet name or organism after launching a job and create Cluster and Order tables, you have to modify ALL the tables in the same way… and recreate an Annotation_DataSet
>
> You can affect DataSet tables to new groups.
>
> Annotation tables can't be renamed!!

Remove table and dependencies

back to Admin Express Db

TableName	Dependencies	Engine	Nbr lines	Size	Created Date	Submitter	version	Action
Annotation_3	☐ **Annotation_3**	MyISAM	48	3056	2017-04-03	administrator	1	Delete table
Myco_AnnotTest	☐ Annotation_Myco_AnnotTest [A]	MyISAM	48	3132	2017-04-04	administrator	1	Delete table
		MyISAM	48	3080	2017-04-06			
	☐ **Myco_AnnotTest**	MyISAM	19	1140	2017-04-03			
	☐ Myco_AnnotTest_0_28_Cluster	MyISAM	19	475	2017-04-18			
	☐ Myco_AnnotTest_0_28_Order	MyISAM	19	931	2017-04-18			
	☐ Myco_AnnotTest_0_54_Cluster	MyISAM	19	475	2017-04-07			
	☐ Myco_AnnotTest_0_54_Order	MyISAM	19	931	2017-04-07			
	☐ Myco_AnnotTest_0_9_Cluster	MyISAM	19	475	2017-04-19			
	☐ Myco_AnnotTest_0_9_Order	MyISAM	19	931	2017-04-19			
		MyISAM	19	988	2017-04-06			
		MyISAM	19	475	2017-04-18			
		MyISAM	19	779	2017-04-18			
		MyISAM	19	475	2017-04-18			
		MyISAM	19	779	2017-04-18			

From this page, you can delete tables and dependencies.

Be aware that deleting Annotation organism will have a critical impact.

1. You can select one or more tables on the same "line".
2. Click on the delete button
3. All tables and dependencies (files generated while clustering) will be deleted

Regenerate tables annotation

From this page, you can generate DataSet annotation ('process' and Generate button) or re-generate existing Annotation (after Organisms annotation update for example).

 If you upload a new annotation set for an organism, use this page to upgrade DataSet annotation.

Note: DataSet annotation table will be truncated before upgrade.

Annotation for organism can't be regenerate from this page. Link ('Update') redirect to 'Upgrade annotation'.

Reset Database

From this link you can remove all DataSet, Annotation and Toolbox SQL tables , truncate Organisms,tables and tables_groups tables.

Users directories will be destroyed and all files generated by cluster callculations (in network and similarity directories) will be erased…

Only SuperAdmin (account 'admin') may access this link.

back to Admin Db

Reset Db

WARNING
Validating this form will:
- Erase all dataset, annotation, toolbox.
- truncate Organisms table
- remove all datas in assets/networks, assets/similarity, assets/users/* directories

TableName	Dependencies
Annotation_1	Annotation_1
Toolbox_1	Toolbox_1

Reset Db   Back to Admin Db

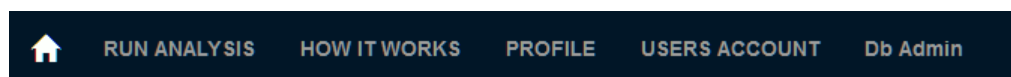Root tables will be displayed with dependencies (*Cluster and *Order tables)

Create DataSet tab

Upload your data in this form.

Upload –step 1

1. **Select master group**: from groups list select one group or create a new one
2. **Select Organism**: from organisms lists select one organism or create a new one.
3. **File Name**: select your data. Use text only file. Not excel, word or any proprietary file
4. **Header**: if your file contains a header on first line, check this box. It will be used for columns name.
5. **Separator**: choose one of the separator
6. **Post Processing**: Default None. If your data contains replicates for a same gene, check 'Define replicates'
7. **Type of Data**: Raw data only for now…
8. **Limit**: if you want to create a small dataset from a huge file, type number of lines to keep.
9. Submit: Submit form. See next page

Upload step 2

First part list file information and name of tables for selected Organism

Second part, display required fields to name and describe your DataSet.

Check 'Force DROP' to destroy previous table with the same name …

For long gene_name, you can add a supplementary comment field. Gene_Name will be truncated at max length size characters for Organism GeneName and stored in Gene_Name field, full gene name in comment field.

Third part, show SQL structures of your imported file.

Displayed values in this table match first line of submited file Demo_Arabido.csv

N.	Title	Value	Type	Option	Size	Is Index	Include ✔	Required ✔
1	Gene_Name	AT5G35935.1	VarChar ▼	▼	15	✔	✔	✔
2	Mean_M6_0h	377.36	Double ▼	SIGNED ▼	9	☐	✔	✔
3	Mean_M6_2h	367.15	Double ▼	SIGNED ▼	16	☐	✔	✔
4	Mean_M6_4h	464.433	Double ▼	SIGNED ▼	9	☐	✔	✔
5	Mean_M6_8h	470.186666666667	Double ▼	SIGNED ▼	16	☐	✔	✔
6	Mean_Col_0h	38.8837	Double ▼	SIGNED ▼	9	☐	✔	✔
7	Mean_Col_2h	39.1950666666667	Double ▼	SIGNED ▼	16	☐	✔	✔
8	Mean_Col_4h	46.7433	Double ▼	SIGNED ▼	9	☐	✔	✔
9	Mean_Col_8h	46.3613666666667	Double ▼	SIGNED ▼	16	☐	✔	✔

Submit   Reset

Note Value in column **Size** indicate max value found for this column in your whole file !!
This may not reflect value displayed in this table (first line of data)

Title: contains column's name take from header file. Gene_Name is required and can't be changed.

Value: value of the first line of data

Type: SQL type calculated from column analysis. Know what you are doing if you change these values…

Option: SQL option.

Size: size of fields calculated from column analysis. You may change these values. In this example, set size to 9 for all fields of type 'Double', option SIGNED (for negative values). Gene_Name size is set to size defined in Organisms definition.

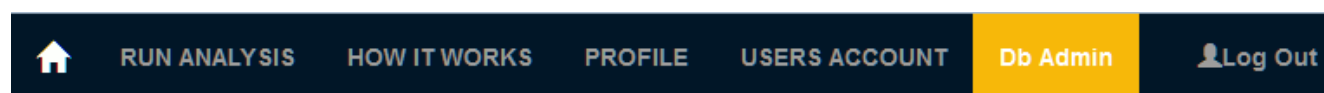**Is Index**: choose which field will be indexed. Gene_Name by default.

Include: check off fields you don't want to include in your DataSet.

Required: data must be present in each lines for checked fields

Submit your form…

.

Annotations tab

Import

1.1. Import annotation from Phytozom

1. Select Organism in list or create a new one
    a. If an organism had already an annotation, an alert message is displayed and Force update checkbox selected. See next figure
2. Select an annotation file from Phytozom (https://phytozome.jgi.doe.gov/pz/portal.html)
    a. Annotation are structured as this:

```
    1: Phytozome internal transcript ID (potentially useful to connect to
biomart datasets)
    2: Phytozome gene locus name
    3: Phytozome transcript name
```

```
        4: Phytozome protein name (often same as transcript name, but this can
     vary)
        5: PFAM
        6: Panther
        7: KOG
        8: KEGG ec
        9: KEGG Orthology
       10: Gene Ontology terms
       11: best Orthologous hit name
       12: best Orthologous hit symbol
       13: best Orthologous hit defline
```

3. Force update. By default, uploaded data are merged at end of existing annotation.
   a. Check checkbox to truncate existing table and upload new data
   b. Stay checkbox unselected to append data at end of the table.
4. Submit form
5. After processing , result page will be displayed:

Process Phytozom

Process a new file;

Your file Athaliana_167_TAIR10.annotation_info.txt have been processed.

File lenght: 5641039
Organism: 1
SQL table: Annotation_1
pid id_organism 1 Force_Update
Memory usage: 7651152
lines inserted 32699
lines with missings fields 2688 . Look log files

Existing annotation

With existing annotation, alert message will be displayed and Force update selected

Create annotation

Create annotation..

Provide an annotation file.
File will be converted to mysql table.
File must be formated as this:

GeneName	Analysis	Signature	Description	Misc
Ref12345	GO	GO:0008152	metabolic process	P
Ref12345	IPRSCAN	IPR016039	Thiolase-like	

WARNING
Organism annotation table are used by Toolbox and Dataset annotation !
Don't forget to regenerate existing annotation table ! (Regenerate Annotation)

Select Organism: *	Arabidopsis thaliana ▼ Add organism
File Name	Parcourir... Aucun fichier sélectionné.
Header	☐ check if header
Separator	○ ; ◉ tabulation
Force update	☐ Replace existing values

[Upload Annotation] [Reset]

*: required!
!! : annotation for this organism already exist !!. Force update selected by default. Otherwise, data will be add at end of table.

1. Select Organism in liste or create a new one
   If an organism had already an annotation, an alert message is displayed and Force update checkbox selected.
2. Select an annotation file
   a. Annotation must be structured as this:

   1: GeneName(15 char): gene locus name or transcript name
   2: Analysis(21 char): type of analysis GO, PFAM, PANTHER, IPRSCAN, KEGG,…
   3: Signature(18 char): reference for analysis (GO =>GO:0008152 PFAM=> PF00139)
   4: Description((255 char): comment about signature element
   5: Misc(15 char): for GO type of functional class (P:Molecular process; F: biological function, C: Cellular component)

3. **Header**: if check, first line will be removed
4. **Separator**: character used to separate fields. Semi-colon or tabulation
5. **Force update**: By default, uploaded data are merged at end of existing annotation.
   a. Check checkbox to truncate existing table and upload new data
   b. Stay checkbox unselected to append data at end of the table.
6. **Submit form**

Import Toolbox

Toolbox contains list of genes categorized by functional or biological activity. Only one toolbox per organisms.

Toolbox can use an additional field to indicate if genes belong to external reference. Database usage is defined during web installation. You can change this setting in application/config/ExpressWeb.php file . Set $config['ExtDb'] = true; and give a name to $config['ExtDbName']

toolbox_name(40)	gene_name	Annotation(100)	functional_class(255)	biological_activity(255)	WB_Db
Cell Wall	AT1G01300	Asp protease	proteases (P)	Asp protease (pepsin family) (Peptidase family A01.A05, MEROPS)	YES
Cell Wall	AT1G01540	Receptor-like kinase family	proteins possibly involved in signaling (PS)	TGF-beta-activated kinase-like	NO
Polysaccharides Biosynthesis	AT1G01570	Glycosyl Transferase Family	biosynthesis of cell wall polysaccharides (BCWP)		NO
Cell Wall	AT1G01900	AtSBT1.1 , Subtilase family	proteases (P)	subtilisin-like serine protease	YES
Cell Wall	AT1G01980	berberine-bridge oxido-re	oxido-reductases (OR)	berberine-bridge enzyme (S)-reticulin:oxygen oxido-reductase	YES
Polysaccharides Biosynthesis	AT1G02000	GAE2 , Nucleotide-sugar	biosynthesis of cell wall polysaccharides (BCWP)	UDP-D-glucuronate 4-epimerase 2	NO
Cell Wall	AT1G02305	Cys protease	proteases (P)	Cys protease (cathepsin B family) (Peptidase family C01.049, MEROPS)	YES
Cell Wall	AT1G02360	glycoside hydrolase (GH)	proteins acting on cell wall polysaccharides (PAC)	glycoside hydrolase family 19 - GH19	YES
Cell Wall	AT1G02460	Glycoside Hydrolase Family	proteins acting on cell wall polysaccharides (PAC)	polygalacturonase	NO

Database constrains between parenthesis (toolbox_name(40) means max characters in field).
GeneName size will use information from Organisms table.
WB_Db (10) . Used to indicate if gene is included in external reference database. (YES/NO)

If you want to use different values for toolbox fields, change it in models/generic.php lines 1017 in function create_toolbox-table().

Know what you do !!!

Upload your Toolbox

File will be converted to mysql table and will upgrade the organism Toolbox Table.

Choose Organism ▼ | Parcourir... Aucun fichier sélectionné.

☐ check if header
☐ Replace existing values **(Force update)**

UPLOAD

Please structure your Toolbox file this way before upload

Use tabulation as separator's field.

Toolbox Name	Gene Name	Annotation	Functional Class	Biological Activity	Presence in WallProt Database
Flavonoid biosynthesis	AT2G37040	PAL1, AtPAL1	Flavonoid metabolism	Phe ammonia lyase	NO

1. Select organism
2. Select toolbox file formatted as required.
3. If header, check box to remove first line
4. Replace existing values will truncate table before uploading your data
5. Upload
6. Toolbox will be available in cluster analysis

Note: on the same toolbox you can upload more than one file. Data will be merged at end of the table.

Edit

Upgrade annotation from Phytozom

Same process as create a new annotation. Only existing organism annotation are displayed in Organisms list

back to Admin Db

Update Phytozom annotation

Enter Phytozom annotation file

File will be converted to mysql table
File must contains 13 fields:
#pacid locusName transcriptName peptideName Pfam Panther KOG KEGG KO GO Best-hit symbol defline

WARNING
Organism annotation table are used by Toolbox and Dataset annotation !
Don't forget to regenerate existing annotation table ! (Regenerate Annotation)

Select Organism: *	Eucalyptus grandis ▾ Add organism
File Name	Parcourir... Aucun fichier sélectionné.
Force update	☑ Replace existing values
Submit Reset	

*: required!
Force update selected by default. Otherwise, data will be add at end of table.

Upgrade annotation

Same process as create a new annotation. Only existing organism annotation are displayed in Organisms list

Manage organisms

Manage organisms

back to Admin Express Db

Organism	Action
Eucalyptus grandis	Ed.
Arabidopsis thaliana	Ed.
Rhizophagus irregularis	Ed.
Oriza sativa	Ed.

Add organism

You can edit or create a new organism

Edit

back to Manage organisms

Organism Eucalyptus grandis

Max transcript lenght: 15

[update organism info] [Reset]

Add

back to Admin Express Db

New Organism:

Max transcript lenght:

[create organism] [Reset]