Data base definition document
Deliverable Number: D3.5

Mahfoudh Abed, Christian Renié, Réza Samadi
December 12, 2014

Contents

1. Project Vision

The objective of this chapter is to collect, analyze and define high level needs and characteristics of the future SSI (Stellar Seismic Indices) website. It focuses on the features required by users, and the purpose of these requirements.

1.1. Positioning

The stellar seismic Indices Website wants to be the web interface of interrogation and exploitation of the SSI database. It is intended for researchers in stellar field.

1.2. Functional Requirements

The SSI interface will gather all the necessary features related to search and exploitation of SSI data.

Search

The SSI website should provide to user a configurable research and take into consideration the following criteria:

- Stars Identification,
- Stars Properties (Alpha, Delta, Teff, log_g, V magnitude, R magnitude, Color (B-R for KEPLER, V-R for COROT), spectral class and luminosity class),
- Seismic indices (value and precision)

Exploitation

The SSI website should allow:

- Display: in paginated data-table,
- Download: in a CSV file.

1.3. Non-functional requirement

Quality Requirements

The use of SSI web interface must be simple and intuitive providing:

- A sober and effective ergonomics,
- A simple search form,
- An effective help support.

Performance requirement

The SSI database will contain at least 8,000 Corot stars, 15 000 Kepler stars and 12 indices per star. Thus, we have to manage over 276,000 indices providing a multi-criteria search. However, the search time should not exceed 3 seconds.

1.4. Conception constraints

Data update

SSI web site allows users to interrogate and exploit the SSI database. The database update can be realized with SQL[1] language. However, this requires advanced technical knowledge to ensure the security and consistency of data.

This operation must be carried out by researchers who do not necessarily have knowledge in database. This is why it's necessary to develop an intuitive administrator interface allowing quick, secure and simple data update in few clicks.

Reliability

To minimize the risk of an unfortunate data loss, we need to back up[2] the database to preserve modifications to data on a regular basis.

A backup and restore strategy helps to protect databases against data loss caused by a variety of failures in order to perform consistently without degradation.

Security

Administrator functions (update, backup, restore) must be carried out in a secure environment. The administrator's username and password are encrypted and stored in the database. Data transfer is also secured following HTTPS[3].

2. Requirements specification according use case

In software and systems engineering, a use case is a list of steps, typically defining interactions between a role (known in Unified Modeling Language (UML) as an "actor") and a system, to achieve a goal. The actor can be a human, an external system or time.

Actors

Human actors in SSI website are:

- Administrator: the actor who is responsible of maintaining the proper functioning of the site and data consistency.

---

[1] **SQL (Structured Query Language):** is a special-purpose programming language designed for managing data held in a relational database management system (RDBMS). The scope of SQL includes data insert, query, update and delete, schema creation and modification, and data access control.

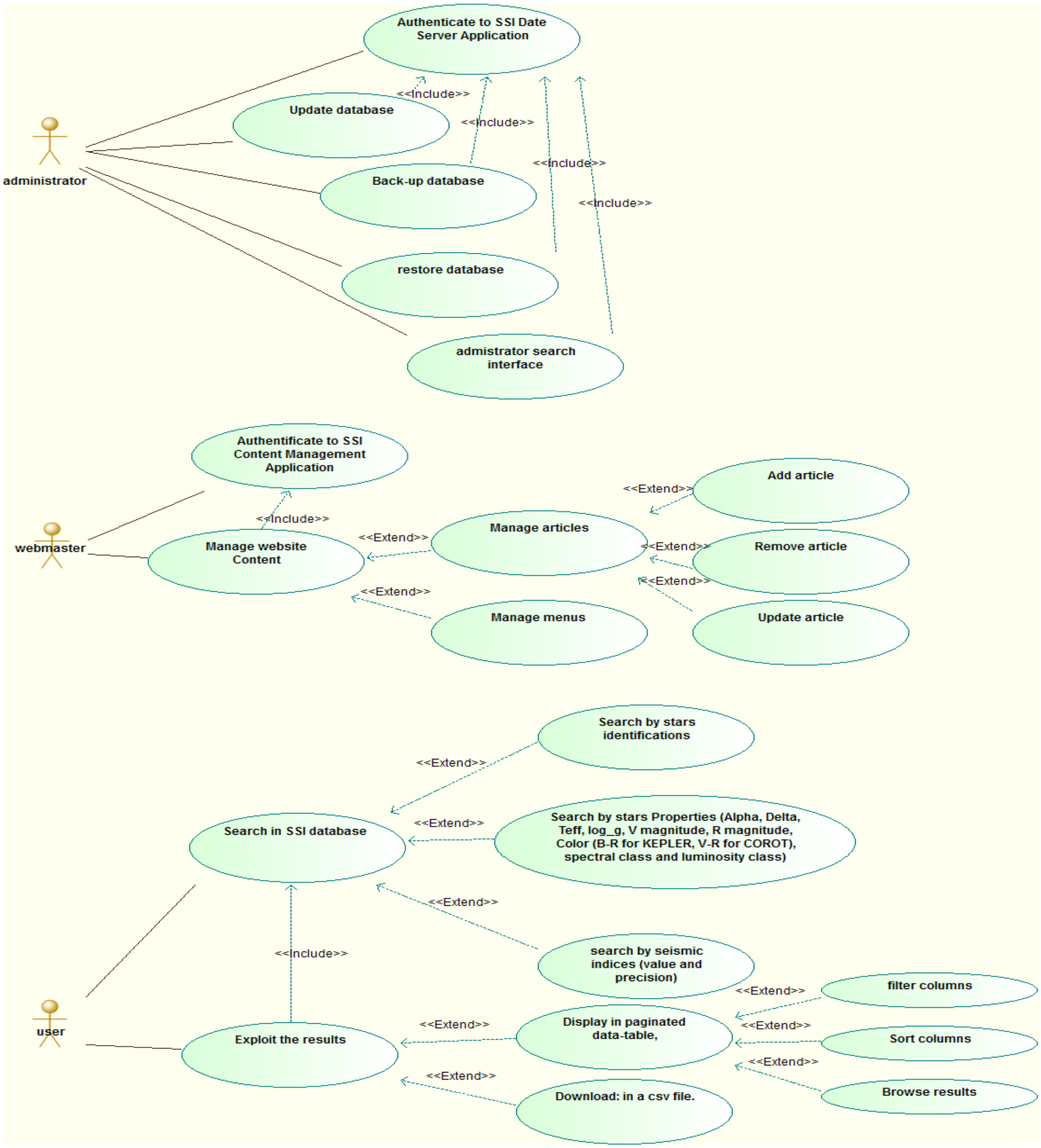
[2] **Back up:** Copies the data or log records from a SQL Server database or its transaction log to a backup device, such as a disk, to create a data backup or log backup.

[3] **HTTPS (Hypertext Transfer Protocol Secure)**: is a communications protocol for secure communication over a computer network. Technically, it is not a protocol in and of itself; rather, it is the result of simply layering the Hypertext Transfer Protocol (HTTP) on top of the SSL/TLS protocol, thus adding the security capabilities of SSL/TLS to standard HTTP communications.

- Webmaster: the actor who is responsible of maintaining the textual content of the site
- User: the person visiting the site to search. This is the most important actor, the one for which the website exists.

Use case diagram

The following diagram (use case) represents different interactions between actors and SSI website.



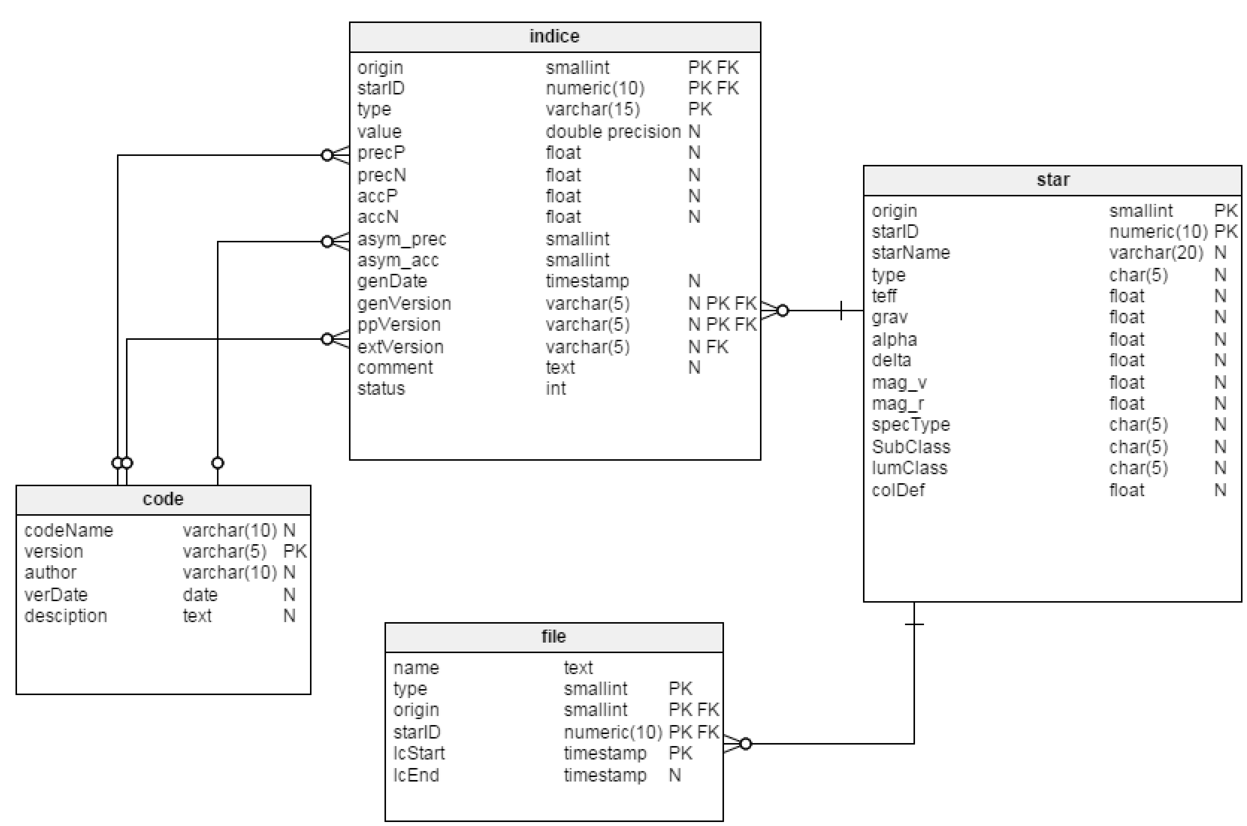
3. SSI database

3.1. Structure of the data base

The date base contains four tables (see Figure 2) and one dynamic table with the following meaning:

- Star: characteristics of the stars (e.g. name, origin, position, magnitude, spectral type ...) ;
- Indice: the seismic indices (e.g. value, precision) ;
- File: the set of files (light-curves) that have been merged together for a given target ;
- Code: versions of the different components of the pipeline.
- Request: dynamically constructed in order to optimize the time response (<2sec)

The data base is developed under postgresql, an open source data base system (http://www.postgresql.org/about/).



Star table

This table contains information related with the star:

- Origin: origin of the data ;

    = 0: for an unknown origin

= 1: for CoRoT data
= 2: for Kepler data
= 3: for OGGLE data (not yet implemented)

- starID: the star ID (CoRoTID in the case of a CoRoT target, and the KIC number in the case of a Kepler target) ;
- starName: Name of Star (HD number for exemple)
- Type: type of light-curve (depending of the instrument, the light-curves can be of different type. For instance for CoRoT: monochromatic or chromatic light-curves, or light-curves from the asteroseismology channel) ;
- teff: effective temperature (in K);
- grav: gravity of the star in log decimal (in cm/s^2) ;
- alpha: right ascension (equatorial coordinate in degree) ;
- delta: declination (equatorial coordinate in degree) ;
- mag_v: magnitude in the V band (depends on the origin and type of the light-curve) ;
- mag_r: magnitude in the R band (depends on the origin and type of the light-curve) ;
- specType: spectral type ;
- subClass: spectral type sub-class ;
- lumClass: luminosity class.
- colDef: Color (B-R for kepler, V-R for corot)

Indice table

This table contains the seismic indices (values and associated data):

- Origin: as in the Star table ;
- sta ID: as in the Star table ;
- type: name of the seismic indice;
- value: value of the seismic indice (Negative value if unknown) ;
- precP, precN: positve and negative deviation of precision (1- σ uncertainty) with which the seismic indice was measured (Negative value if unknown) ;
- accP, accN: positve and negative deviation of accuracy with which the seismic indice was measured (Negative value if unknown) ;
- ssym_prec: flag (true if prec is asymmetric)
- assym_acc: flag (true if acc is asymmetric)
- genDate: generation date ;
- genVersion: version of the generation pipeline ;
- ppVersion: version of the pre-processing ;
- extVersion: version of the extraction program
- comment: any comment associated with the indice.
- status: status of the indice, status=0 for a valid indice otherwise a non-zero value ;

The names of the currently available seismic indices are

- Deltanu: the mean large separation $\Delta\nu^2$ [in μHz] obtained by the auto-correlation function (ACF) method;
- Amax: the maximum of the AFC ;

- numax: the peak frequency v_{max}^2[in μHz] of the oscillation envelop obtained by the least square fitting;
- A_env: the envelop [in ppm^2/μHz] ;
- W_env: the width of the oscillation envelope [in μHz];
- numax_up: the peak frequency derived with the Universal Pattern (UP) method [in μHz];
- Tau_gran: the granulation timescale [in sec] ;
- P_gran: the granulation power [in ppm2/μHz] ;
- Alpha_gran: the slope of the granulation component ;
- Deltanu_up: the large separation derived with the UP method;
- taue_gran: the e-folding time of the granulation
- sigma2_gran: the mean-square of the granulation brightness fluctuations

Code table

This table contains versions of the different components of the pipeline

- codeName: the code of the component .
- version: the component version
- author: the component's author
- verDate: the component creation date.
- Description: the component description

File table

This table contains information related with the set of files (light-curves) that have been merged together for a given target (the table star)

- name: the name of the file
- type: type of light-curve (depending of the instrument, the light-curves can be of different type. For instance for CoRoT: monochromatic or chromatic light-curves, or light-curves from the asteroseismology channel) origin: origin of data

  = 0: for an unknown origin
  = 1: for CoRoT data
  = 2: for Kepler data
  = 3: for OGGLE data

- starID: the star ID (CoRoTID in the case of a CoRoT target, and the KIC number in the case of a Kepler target).
- IcStart : light corve start
- IcEnd: light corve end

Request table

Another table (request) is dynamically constructed in order to optimize the time response (<2 sec). It contains information extracted from three tables: star, indice and file. The list of their attributes is defined in the Annex.

3.2. Relationships between tables

Star 1→ ∞ indice: each star has many indices (calculated from the light curve) but each indice refers to one star

Indice 1→ 1 code: There are 3 "one to one" relationships between indice and code. For each indice, there is one code of extraction, one code of preprocessing and one code of generation

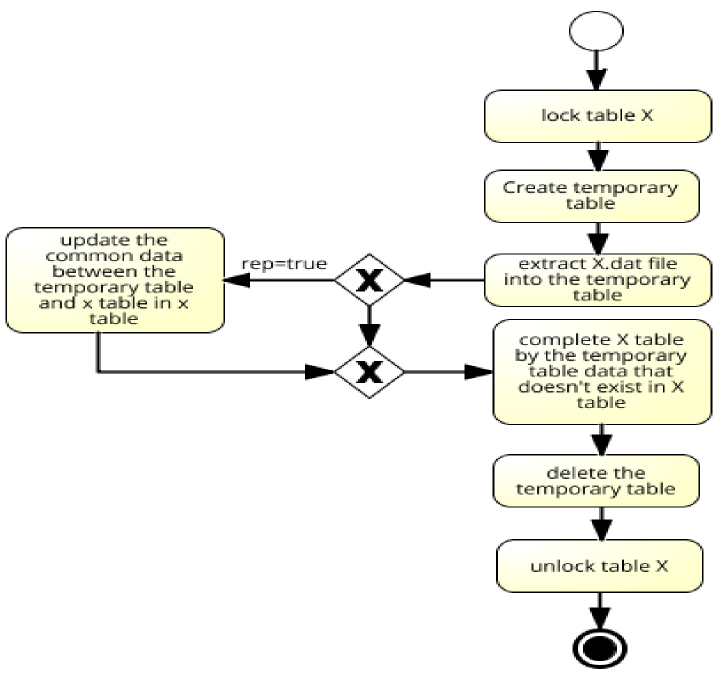
Star 1→ ∞ file: each file describes one star but information of each star come from many (one or more) files

3.3. Data Loading

The four tables (star, indice, code, file) are filled using an SQL function under plpgsql LOAD_DATA_X (source text, rep boolean) where

- X refers to the name of the table; it can be equal to star, indice, code or file.
- Source: is the source of data
- Rep: indicates if the common data between source and the table X must be updated

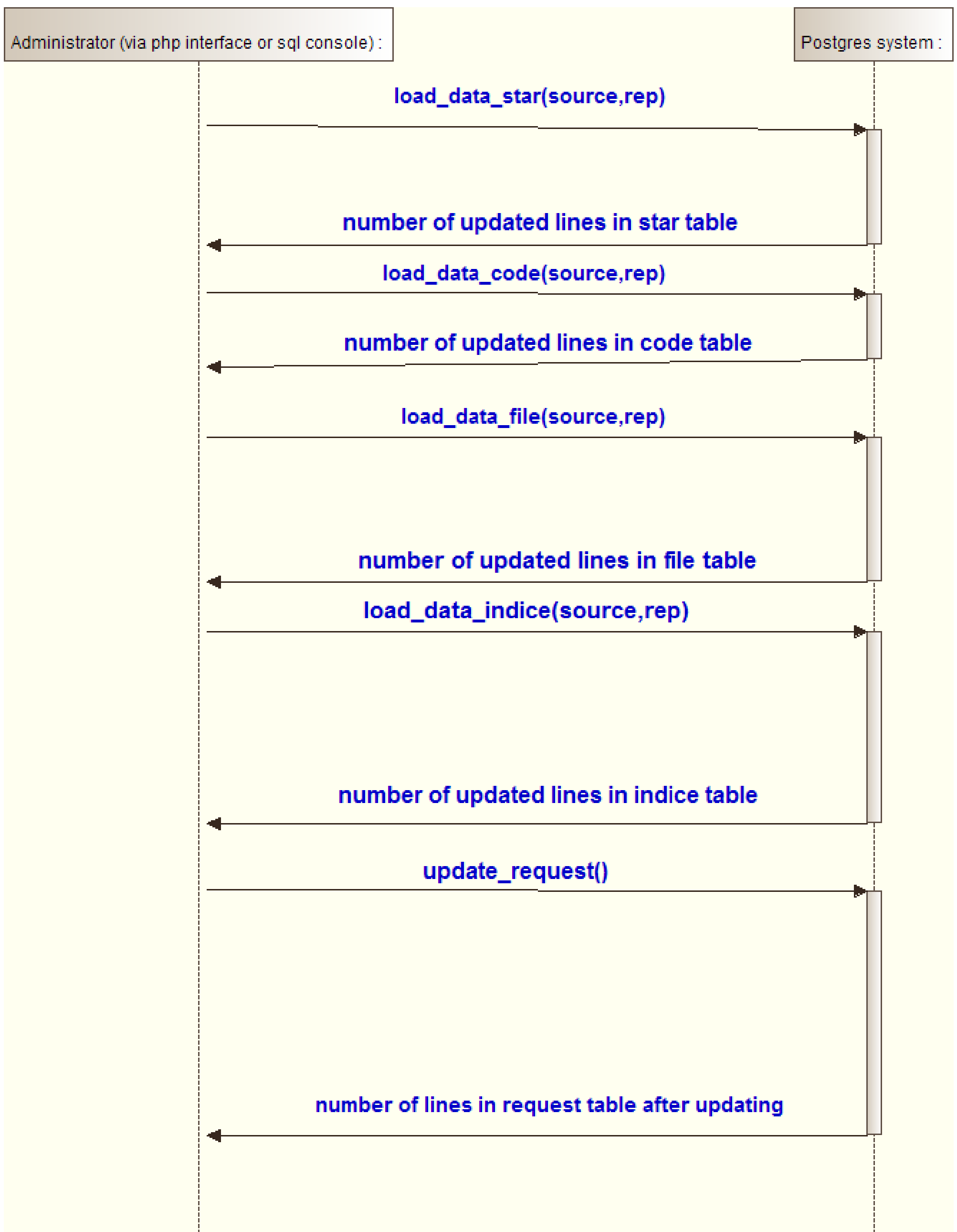
The following scheme illustrates the process of these functions:



The updates are performed by the administrator via the SSI administrator interface or using the Sql console.

Tables are updated in this order: star, code, file, indice and at last request using respectively LOAD_DATA_star, LOAD_DATA_code, LOAD_DATA_file, LOAD_DATA_indice and update_request functions. Each function returns the number of updated lines in the table.

The following diagram illustrates the processing order of update functions.



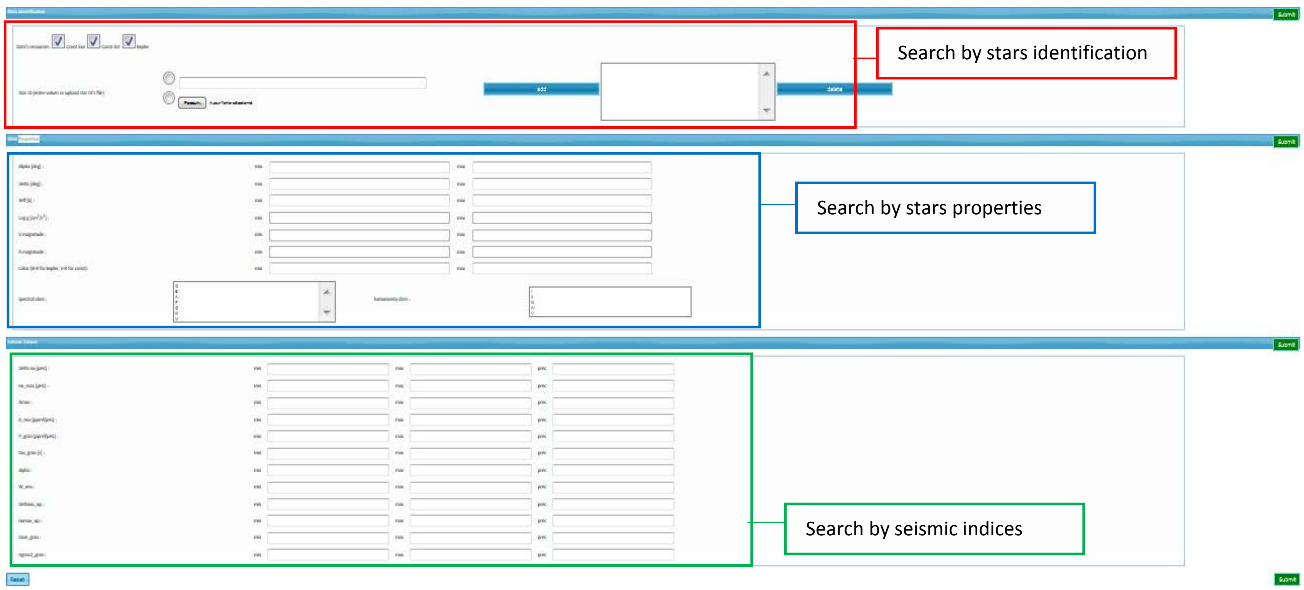
4. SSI web site

4.1. User interface

The home page contains a link to the search interface

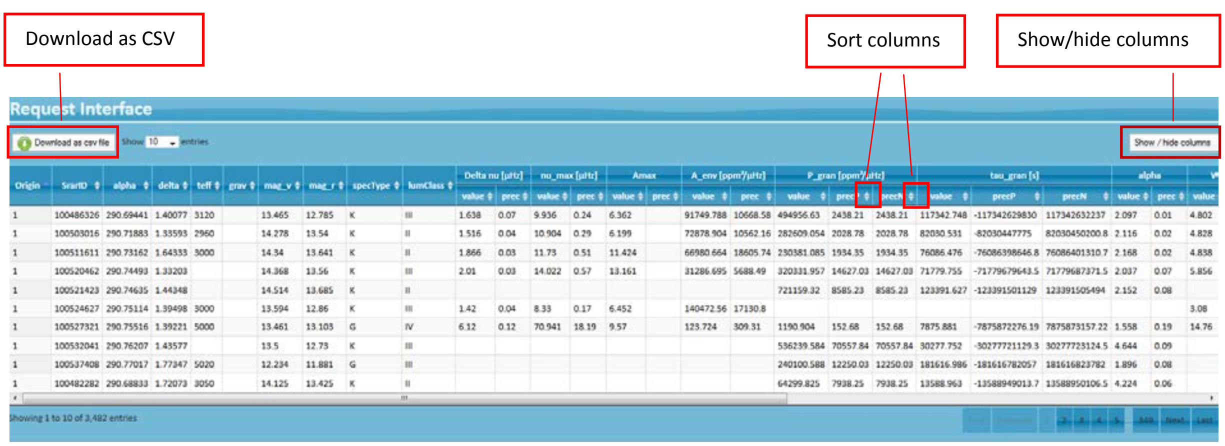
A user can search in SSI database:

- By stars identification ;
- By stars properties(alpha, delta, teff, log_g, V magnitude, R magnitude, color (B-R for kepler, V-R for COROT), spectral class and luminosity class) ;
- By seismic indices (range of values and precisions).



A user can also exploit the results of search

- Display results in paginated data tables, many options are available (filter columns, sort columns and browse results) ;
- Download results in CSV files.



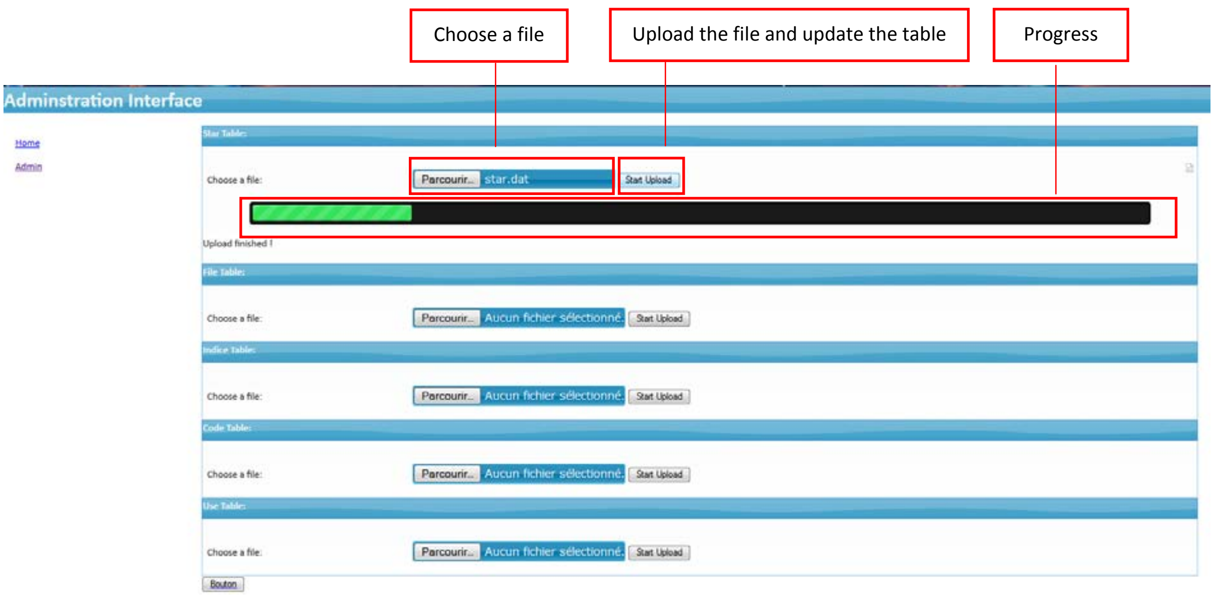
The display format of indices value and precision is released according the following algorithm:

- **V** is the value of the indice;

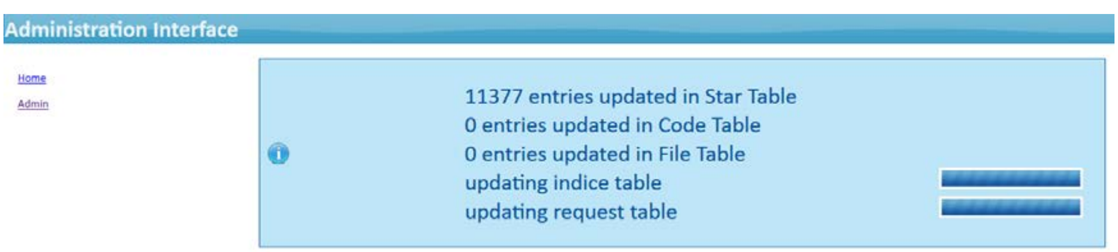
- P_+ and P_- are respectively the positive and negative deviation of precision;
- P_{min}=max (10^{-3},min (P_+, P_-));
- **T**=round (P_{min} /10^{-3})* 10^{-3};
- The displayed value **V'**=round (10***V**/**T**)***T**/10;
- The displayed precision $P'_{+/-}$ =round($P_{+/-}$/**T**)***T;**

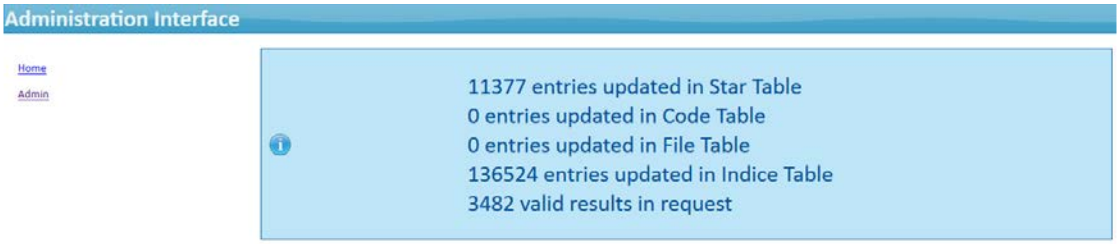
4.2. Administrator interface

After being connected, an administrator can update the data base. For this, he chooses a data file, uploads it and updates a table in SSI database



After uploading files, we can follow the progress of the update. Thereafter two pictures, the first is taken during update where the second is taken after update.





5. Virtual Observatory access:

The SSI database is also available through a Virtual Observatory (VO) Web service.

This service follows the VO TAP protocol and is implemented with the DACHS framework (http://vo.ari.uni-heidelberg.de/soft/dachs) developed at the German VO (GAVO).

The service publishes seismic valid indices (with status equals to zero) for a given or a list of star identifiers.

The ADQL language (http://www.ivoa.net/documents/latest/ADQL.html) is used to query the service and results can be retrieved in different formats (VOTable, CSV …).

The fields published through the TAP service are extracted from the table 'request' (cf. annex Tap fields).

6. Project progress

The following table represents the progress in the development of SSI website features (refer to use case diagram, Section 1).

Use case	Technology	Progress
Search in SSI data base	PHP, JQuery, Ajax	Completed
Exploiting of the results of search	PHP, JQuery, Ajax, Datatable.js, PostgreSQL	Test
Webmaster authentication	SPIP, LDAP	Completed
managing of Website content	SPIP	Completed
Administrator authentication	PHP, PostgreSQL, HTTPS	Under construction
Database update	PHP, PostgreSQL, Jquery, Ajax	Completed
Database back up	PHP, PostgreSQL, Jquery, Ajax	Under construction
Database restore	PHP, PostgreSQL, Jquery, Ajax	Under construction
Administrator search interface	PHP, PostgreSQL, Jquery, Ajax	Under construction
Virtual Observatory access	DACHS framework, TAP protocol	Completed

Annex

List of Request table attributes

The attributes of the request table are:

- origin: as in the Star table ;
- starID: as in the Star table ;
- alpha: as in the Star table ;
- delta: as in the Star table ;
- type: as in the Star table ;
- Teff: as in the Star table ;
- grav: as in the Star table ;
- mag_v: as in the Star table ;
- mag_r: as in the Star table ;
- specType: as in the Star table ;
- lumClass: as in the Star table ;
- alpha_gran_value: Slope of the granulation component
- alpha_gran_precP: Precision on alpha_gran_value
- Amax_value: Maximum of the Auto-Correlation Function (ACF)
- Amax_precP: Precision on a_max_value
- A_env_value: Maximum of the oscillation envelope
- A_env_precP: Precision on a_env_value
- W_env_value: Width of the oscillation envelope
- W_env_precP: Precision on w_env_value
- Deltanu_value: Large separation derived with the ACF method
- Deltanu_precP: Precision on deltanu_value
- Deltanu_up_value: Large separation derived with the Universal Pattern (UP) method
- Deltanu_up_precP: Precision on deltanu_up_value
- numax_value: Peak frequency of the oscillation envelope
- numax_precP: Precision on numax_value
- numax_up_value: Peak frequency derived with the Universal Pattern (UP) method
- numax_up_precP: Precision of numax_up_value
- P_gran_value: Granulation power
- P_gran_precP: Precision on p_gran_value (positive deviation)
- P_gran_precN: Precision on p_gran_value (negative deviation)
- tau_gran_value : Granulation time-scale
- tau_gran_precP : Precision on tau_gran_value (positive deviation)
- tau_gran_precN : Precision on tau_gran_value (negative deviation)
- taue_gran_value : e-folding time of the granulation

- taue_gran_precP : Precision on taue_gran_value (positive deviation)
- taue_gran_precN : Precision on taue_gran_value (negative deviation)
- sigma2_gran_value: Mean-square of the granulation brightness fluctuations
- sigma2_gran_precP: Precision on sigma2_gran_value (positive deviation)
- sigma2_gran_precN: Precision on sigma2_gran_value (negative deviation)
- preprocessing_code: Version of the preprocessing used to calculate the SI
- generation_code: Version of the generation used to calculate the SI
- time_min: Acquisition start time (in JD)
- time_max: Acquisition stop time (in JD)
- files_used: Files name used to calculate the SI (comma separated list)

TAP fields

Name	Description	Unit	UCD
origin	Source instrument (1=CoRot, 2=Kepler)	N/A	meta.id
star_id	Star identifier	N/A	meta.id
s_ra	RA of (center of) observation, ICRS	deg	pos.eq.ra
s_dec	DEC of (center of) observation, ICRS	deg	pos.eq.dec
alpha_gran_value	Slope of the granulation component	N/A	meta.id
alpha_gran_prec	Precision on alpha_gran_value	N/A	meta.id
a_max_value	Maximum of the Auto-Correlation Function (ACF)	N/A	meta.id
a_max_prec	Precision on a_max_value	N/A	meta.id
a_env_value	Maximum of the oscillation envelope	ppm2/uHz	meta.id
a_env_prec	Precision on a_env_value	ppm2/uHz	meta.id
w_env_value	Width of the oscillation envelope	uHz	meta.id
w_env_prec	Precision on w_env_value	uHz	meta.id
deltanu_value	Large separation derived with the ACF method	uHz	meta.id
deltanu_prec	Precision on deltanu_value	uHz	meta.id
deltanu_up_value	Large separation derived with the Universal Pattern (UP) method	uHz	meta.id
deltanu_up_prec	Precision on deltanu_up_value	uHz	meta.id
numax_value	Peak frequency of the oscillation enveloppe	uHz	meta.id
numax_prec	Precision on numax_value	uHz	meta.id
numax_up_value	Peak frequency derived derived with the Universal Patt ern (UP) method	uHz	meta.id
numax_up_prec	N/A	uHz	meta.id
p_gran_value	Granulation power	ppm2/uHz	meta.id
p_gran_precP	Precision on p_gran_value (positive deviation)	ppm2/uHz	meta.id
p_gran_precN	Precision on p_gran_value (negative deviation)	ppm2/uHz	meta.id
tau_gran_value	Granulation time-scale	s	meta.id
tau_gran_precP	Precision on tau_gran_value (positive deviation)	s	meta.id

tau_gran_precN	Precision on tau_gran_value (negative deviation)	s	meta.id
taue_gran_value	e-folding time of the granulation	s	meta.id
taue_gran_precP	Precision on taue_gran_value (positive deviation)	s	meta.id
taue_gran_precN	Precision on taue_gran_value (negative deviation)	s	meta.id
sigma2_gran_value	Mean-square of the granulation brigthness fluctuations	ppm2	meta.id
sigma2_gran_precP	Precision on sigma2_gran_value (positive deviation)	ppm2	meta.id
sigma2_gran_precN	Precision on sigma2_gran_value (negative deviation)	ppm2	meta.id
preprocessing_code	Version of the preprocessing used to calculate the SI	N/A	meta.id
generation_code	Version of the generation used to calculate the SI	N/A	meta.id
time_min	Acquisition start time (in JD)	d	time.start
time_max	Acquisition stop time (in JD)	d	time.stop
files_used	Files name used to caculate the SI (comma sparated list)	N/A	meta.id