Database of processed data

Final Version of 31/10/2021

Deliverable Number D.3.6.1.

Project Acronym	SOUNDSCAPE
Project ID Number	10043643
Project Title	Soundscapes in the north Adriatic Sea and their impact on marine biological resources
Priority Axis	3
Specific Objective	3.2
Work Package Number	3
Work Package Title	Soundscape assessment
Activity Number	3.6
Activity Title	Processing of data input for modelling and uncertainty analysis
Partner in Charge	IOF
Partners Involved	CNR, BWI, ARPA FVG
Authors	Damir Ivanković (IOR), Stipe Muslim (IOF), Dalibor Jelavić (IOR), Ivan Rojnica (IOF), Vlado Dadić (IOF)
Status	Final
Distribution	Public
Citation	Ivanković D., Muslim S., Jelavić D., Rojnica I., Dadić V. Database of processed data. SOUNDSCAPE project, WP3, 19 pp, 2021

Contents

1 Summary 3

2 Introduction 3

3 FTP network storage 3

4 Data management and data flow 5

5 Database properties and design 6

6 Metadata database 6

7 Database for loading and process SPL data 8

8 Data processing 11

9 Data visualization 12

10 Archived data in IOF and CNR databases and their availability 14

1 Summary

The document describes database was developed for archiving underwater noise raw data in WAV format, and PSL data averaged at 20 second time intervals and corresponding referral data. In addition, agreed procedures for upload/download, validation, data processing, and presentation of output data were implemented in the database. The database was developed in the third reporting period, and all collected data were uploaded in the last reporting period (November 2021).

2 Introduction

Data collected in the frame of the project can be divided into four groups:

1. Metadata describing deployment, recovery and calibration
2. WAV files – SPL (Sound Pressure level) sampled with 48000Hz
3. 1/3 octave spectrum for every 20s (or lower) period of SPL (WAV) data
4. Averages, statistical and processed data

For three of those categories relational database storage should be applied. WAV files (RAW data) should be stored only on the file system because of external processing and big amount of sequential data. Although meta data about where WAV files are saved (including links to WAV data) should be also stored in the database.

3 FTP network storage

For raw WAV data network storage has been set. Storage is organized on the address data01.izor.hr, and 100 TB of storage is available.

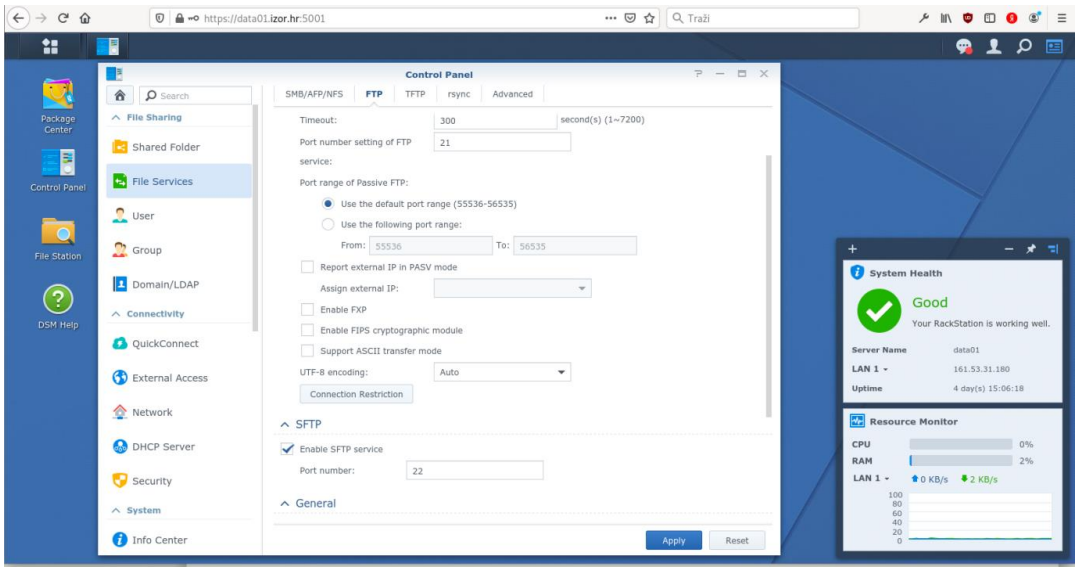


Figure 1 FTP properties from network storage web interfaces

Standard FTP software can be used for connection to network storage and file upload. Naming rule to be used for raw wav files: Files should be stored in folders with following structure:

- First level folder - data type (WAV, SPL, PROCESSED)
- Second level folder station name/ID (N1, N2, ...)
- Third level folder month of file beginning (2020_02, 2020_03, ...).

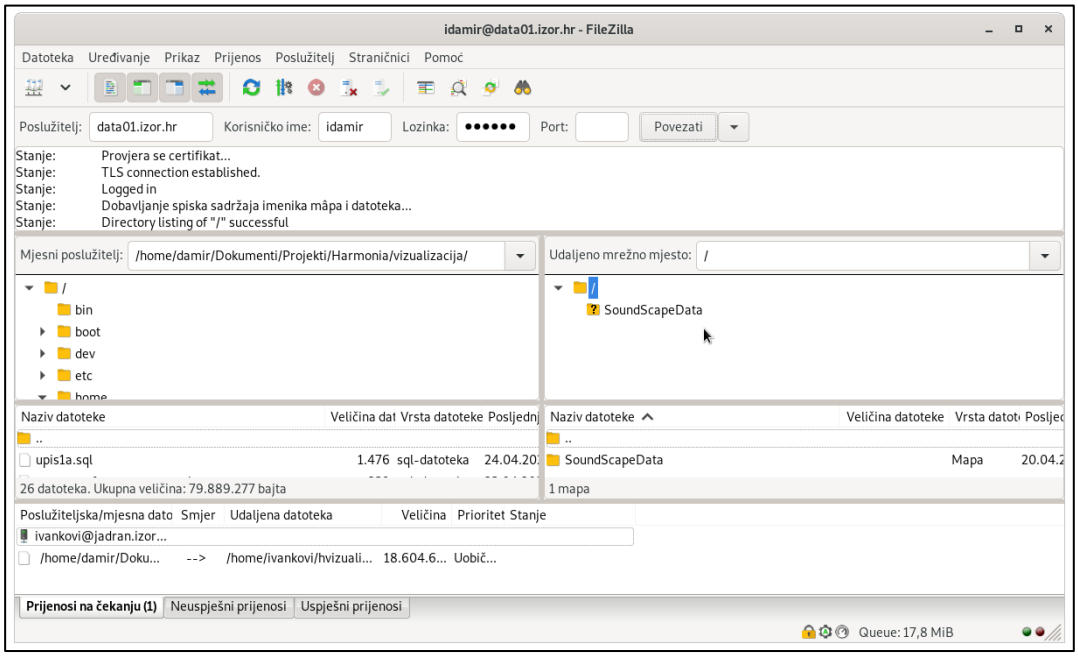


Figure 2 Connection to network storage using open source software FileZilla

4 Data management and data flow

Some steps that should be followed in different phases of data acquiring:

- Hydrophone deployment and recovery
  - Filling deployment and recovery sheets with metadata
  - Upload WAV files, calibration check WAV files, **SVR_CFG.XML** file, **SYSTEM.LOG** file, **CTD_data deployment_sheet.pdf** and **recovery_sheet.pdf** to the associated location at the **NAS** (network attached storage)
  - Inserting metadata to database from **deployment_sheet.pdf** and r**ecovery_sheet.pdf** for each station
- 1/3 octave spectrums for every 20s (or lower) period of SPL (WAV) data. Spectrums are generated using external processing software (SPL file extension or SPL files)
  - Uploading 1/3 octave spectrums data to the associated location at the **NAS**
  - Loading 1/3 octave spectrums files to database and joining with associated metadata
- Database processing procedures
  - Calculation of SPL averages for defined 1/3 octave band and time period
  - Data visualizations and first step validation.

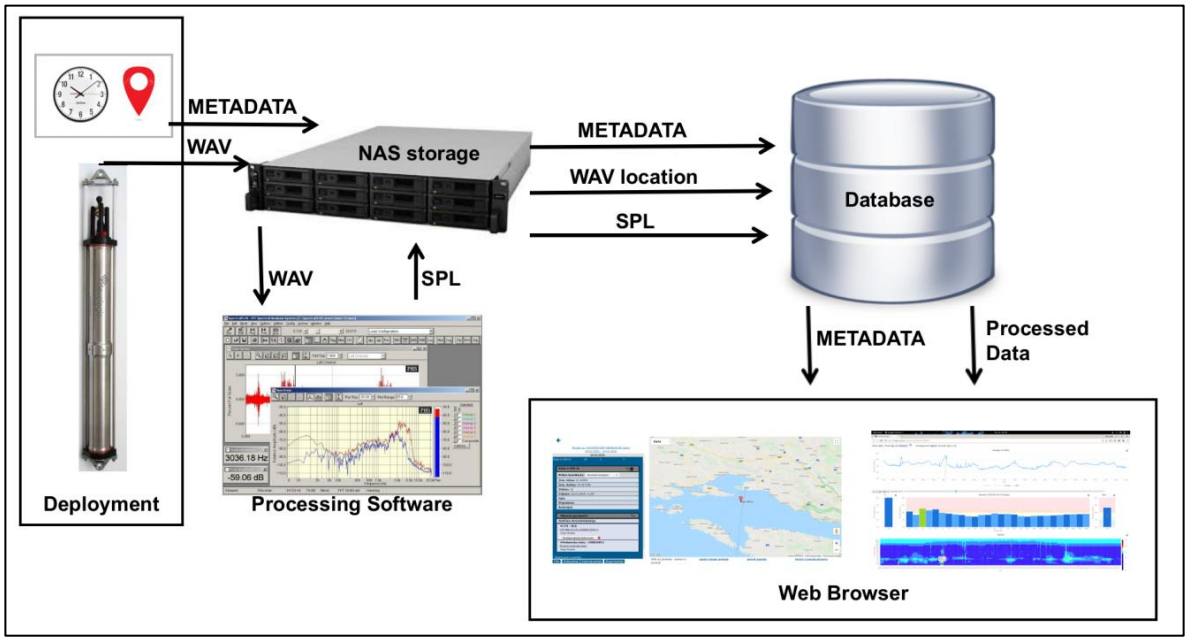


Figure 3 General data flow scheme

5 Database properties and design

Database for data management and processing is relational database with web interface.

General database properties are:

- Linux CentOS 7 (enterprise Linux operating system)
- Oracle 19.3 database, Standard edition 2
- Oracle ORDS for database web interface
- Tomcat 9 application server
- Apache 2 web server, proxy and traffic encryption
- Https protocol for web interface (secured http)
- 128GiB DDR4 Synchronous 2666 MHz
- 2x Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz (16 Cores, 32 Threads)
- RAID bus controller MegaRAID SAS-3 3108
- 2x logical volumes (Raid 50) 32TB - OS and Database dedicated
- 2x Ethernet Connection X722 for 10GBASE-T
- Permanent Internet link 100Mbps (CARNet)
- UPS – uninterruptible power supply

Database have two main parts: Metadata database and Database for loading and process SPL data.

6 Metadata database

This part of the database is designed to manage the monitoring station's metadata. The web interface is used with Google maps API 3 for spatial data visualization. All metadata can be inserted using web interface in particular for the station. Design of this part of database is classical with relations between main entities.

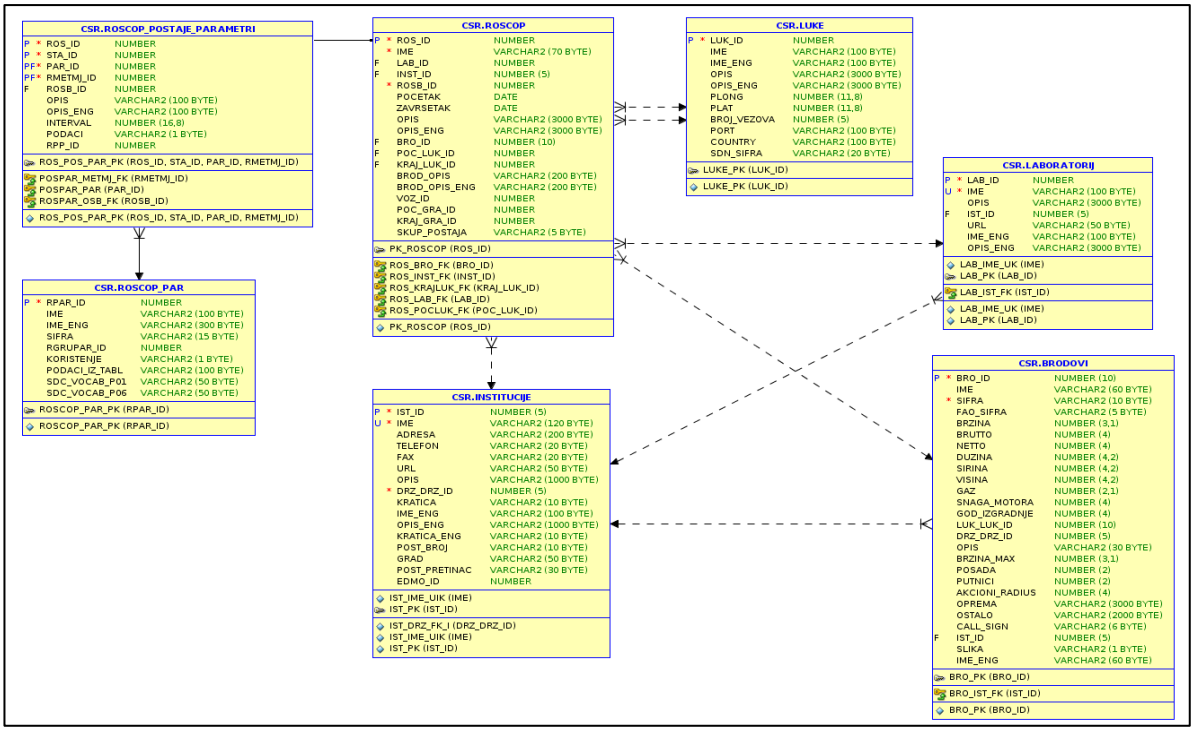


Figure 4 Relational diagram of main objects in metadata database

Web interface is used for inserting and data *view*. User can find and filter inserted cruises. For each cruise detailed metadata is available.

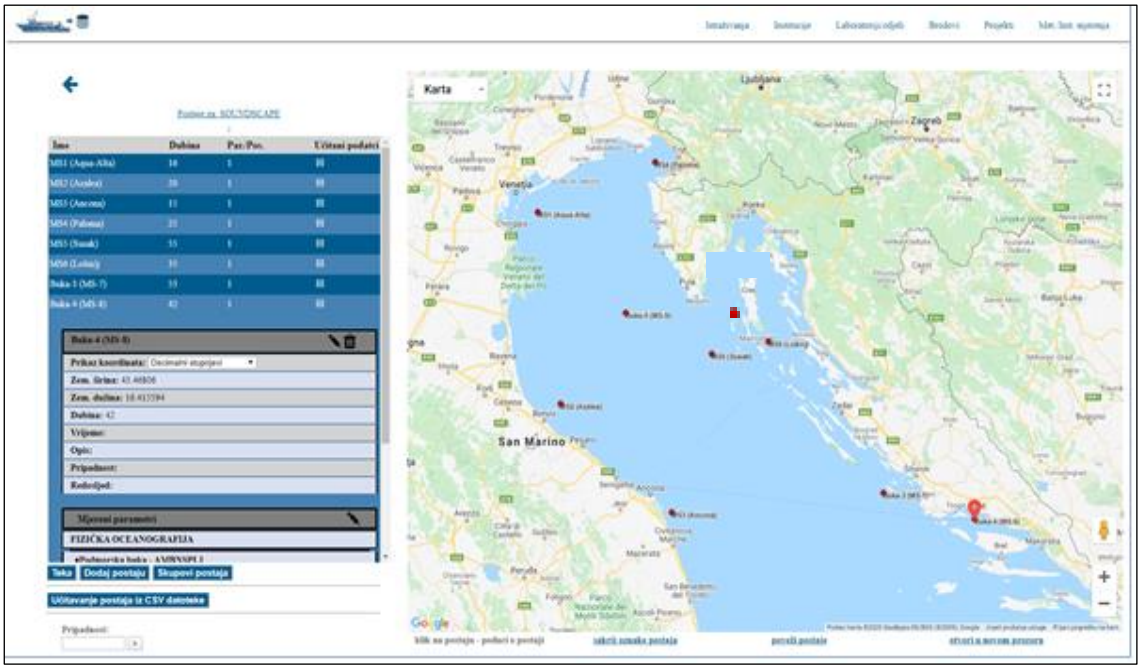
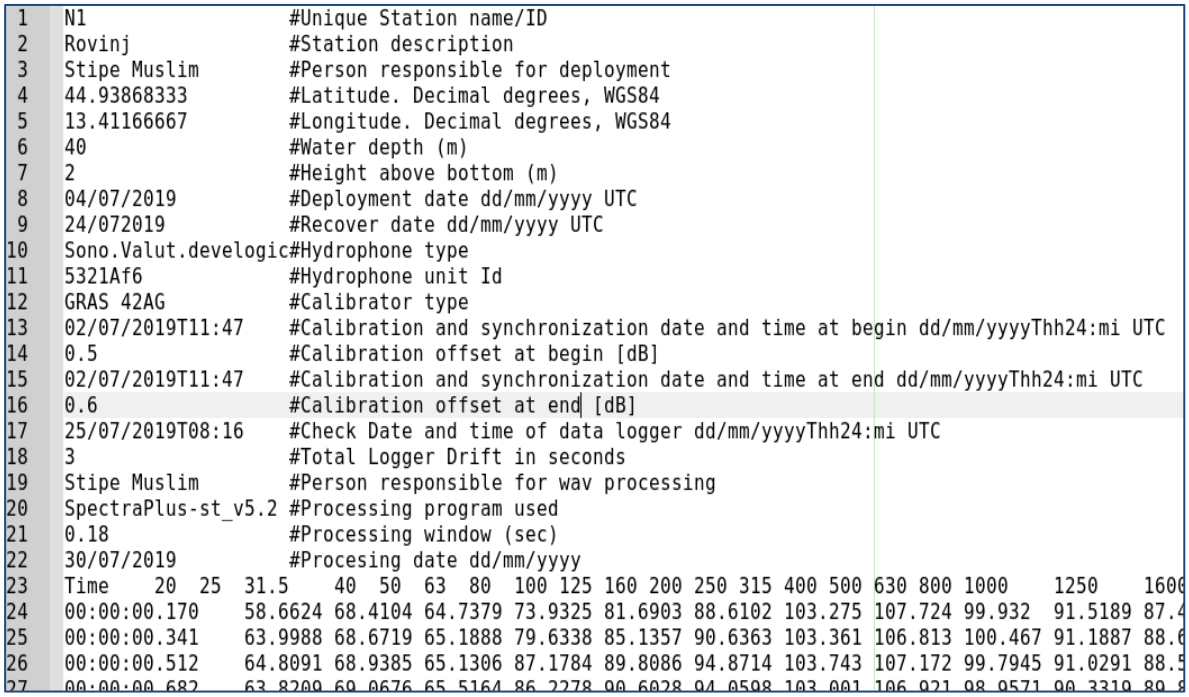


Figure 5 Metadata web interface example

7 Database for loading and process SPL data

It is specially designed for loading and processing 1/3 octave spectrums of SPL (Sound Pressure level) data. Because of that spectrum files (SPL files) are generated by external software (Blackbox) 1/3 octave spectrum files have to fulfil some requirements to be suitable for importing to the database and for further data processing:

- Black box output files should be in textual format (ASCII) with separated columns (tab or semicolon separation)
- File names have to contain station name to join data with the rest of metadata, and beginning referent time of 1/3 octave spectrums of SPL assuming that step for each 20s spectrum is time-step fixed
- They can contain some number of metadata rows at beginning
- They have to contain timestamp column containing sample offset time from beginning
- Except timestamp column, rest of columns should contain sound pressure levels for predefined 1/3 octave bands in dB (re 1µPA) $_{RMS}$ values.



```
1  N1                  #Unique Station name/ID
2  Rovinj              #Station description
3  Stipe Muslim        #Person responsible for deployment
4  44.93868333         #Latitude. Decimal degrees, WGS84
5  13.41166667         #Longitude. Decimal degrees, WGS84
6  40                  #Water depth (m)
7  2                   #Height above bottom (m)
8  04/07/2019          #Deployment date dd/mm/yyyy UTC
9  24/072019           #Recover date dd/mm/yyyy UTC
10 Sono.Valut.develogic#Hydrophone type
11 5321Af6             #Hydrophone unit Id
12 GRAS 42AG           #Calibrator type
13 02/07/2019T11:47    #Calibration and synchronization date and time at begin dd/mm/yyyyThh24:mi UTC
14 0.5                 #Calibration offset at begin [dB]
15 02/07/2019T11:47    #Calibration and synchronization date and time at end dd/mm/yyyyThh24:mi UTC
16 0.6                 #Calibration offset at end [dB]
17 25/07/2019T08:16    #Check Date and time of data logger dd/mm/yyyyThh24:mi UTC
18 3                   #Total Logger Drift in seconds
19 Stipe Muslim        #Person responsible for wav processing
20 SpectraPlus-st_v5.2 #Processing program used
21 0.18                #Processing window (sec)
22 30/07/2019          #Procesing date dd/mm/yyyy
23 Time    20  25  31.5    40  50  63  80  100 125 160 200 250 315 400 500 630 800 1000    1250    1600
24 00:00:00.170    58.6624 68.4104 64.7379 73.9325 81.6903 88.6102 103.275 107.724 99.932  91.5189 87.4
25 00:00:00.341    63.9988 68.6719 65.1888 79.6338 85.1357 90.6363 103.361 106.813 100.467 91.1887 88.6
26 00:00:00.512    64.8091 68.9385 65.1306 87.1784 89.8086 94.8714 103.743 107.172 99.7945 91.0291 88.5
27 00:00:00.682    63.8209 69.0676 65.5164 86.2278 90.6028 94.0598 103.001 106.921 98.0571 90.3319 89.8
```

Figure 6 Example of 1/3 octave spectrums of SPL file

SPL data loading process is executing by appropriate batch script. Database loads files directly from NAS storage. Option for uploading data using a web browser was skipped because of the big data problem which requires too much time to load data to the database. Data is loaded using external tables. For file containing 126.561 rows, loading process took approx. 1 min.

```
#!/bin/bash

ORACLE_HOME=/u01/app/oracle/product/19.3.0/dbhome_1
export ORACLE_HOME
#ORACLE_SID=MEDAS
#export ORACLE_SID;
LD_LIBRARY_PATH=/u01/app/oracle/product/19.3.0/dbhome_1/lib:/u01/app/oracle/product/19.3.0/dbhome_1/network/lib
export LD_LIBRARY_PATH
PATH=/usr/local/bin:/usr/bin:/usr/local/sbin:/usr/sbin:/home/oracle/.local/bin:/home/oracle/bin:/u01/app/oracle/product/
export PATH

export ORACLE_SID=medas
ORAENV_ASK=NO
export ORAENV_ASK
. /usr/local/bin/oraenv $ORACLE_SID


cd /u01/app/oracle/admin/medas/dpdump/9684D4231AE757FAE0534D1F35A1BC7D/zvuk/

shopt -s nullglob
for f in /u01/app/oracle/admin/medas/dpdump/9684D4231AE757FAE0534D1F35A1BC7D/zvuk/*.zvp
do
#mv $f /u01/app/oracle/admin/medas/dpdump/9684D4231AE757FAE0534D1F35A1BC7D/zvuk/zvuk.dat
sed -e 1,1d $f>/u01/app/oracle/admin/medas/dpdump/9684D4231AE757FAE0534D1F35A1BC7D/zvuk/zvuk.dat
/usr/bin/dos2unix /u01/app/oracle/admin/medas/dpdump/9684D4231AE757FAE0534D1F35A1BC7D/zvuk/zvuk.dat
mv $f /u01/app/oracle/admin/medas/dpdump/9684D4231AE757FAE0534D1F35A1BC7D/zvuk/obradjene
/u01/app/oracle/product/19.3.0/dbhome_1/bin/sqlplus -s [redacted]@v<<END > sqlplus.log
execute ZVUK_UCITAJ('$f');
exit;
END

rm /u01/app/oracle/admin/medas/dpdump/9684D4231AE757FAE0534D1F35A1BC7D/zvuk/TEMP_ZVUK_*.log
```

Figure 7 Example of batch script

Three group of tables are used for processing SPL over each 1/3 octave band:

- Tables containing metadata about already loaded files and log which file is currently loading (green)
- Tables containing SPL loaded data in dB and converted to linear version of SPL data (red).
- Tables containing averaged data for different time periods (yellow).

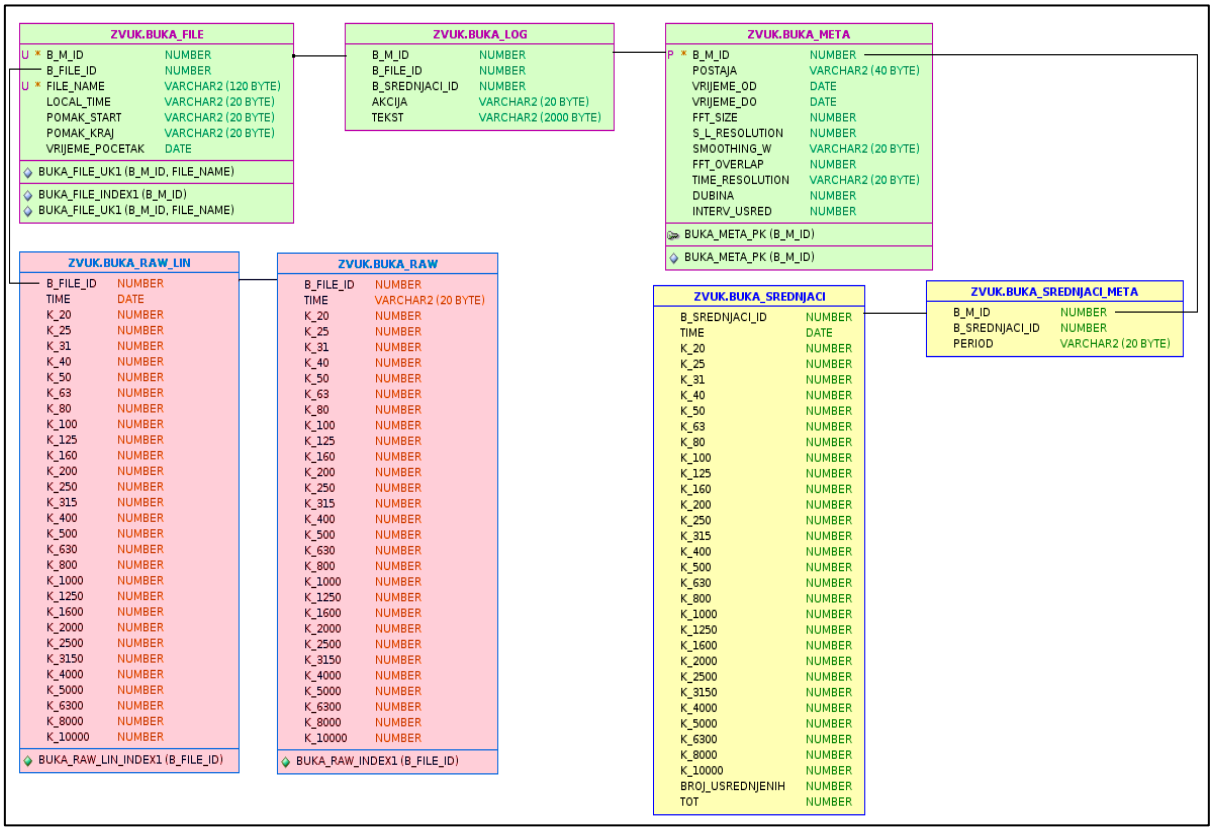


Figure 8 SPL spectrums processing tables schema

SPL averages for each 1/3 octave band over various time periods are calculated inside the database. Time periods are user-selected and can be defined as:

- Minutes
- Hours
- Days
- Month

Referent time is pointing to the beginning of the averaged period (e.g. 25.07.2019 12:00 in hourly averages is for period from 25.07.2019 12:00 to 25.07.2019 13:00). Time in database is in UTC. Input SPL files should have time also in UTC. Exceptionally they can use different time zone, but is necessary to mark offset from UTC in that case.

8 Data processing

After uploading SPL file user can insert some additional metadata about uploaded SPL spectrum files:

- Sea depth at station
- FFT size (samples)
- Spectral Line Resolution
- Smoothing Window
- FFT Overlap
- Time Resolution
- Local time.

List of the additional metadata can be easily changed or expanded. Uploaded SPL file sometime contains unwanted part at beginning or end of file (noise during deployment or during recovery of the instrument). Therefore, user can do trimming to uploaded data (using discard before and discard after parameters). After trimming SPL averages are recalculated. All loaded SPL data and SPL calculated averages are accessible via web browser.

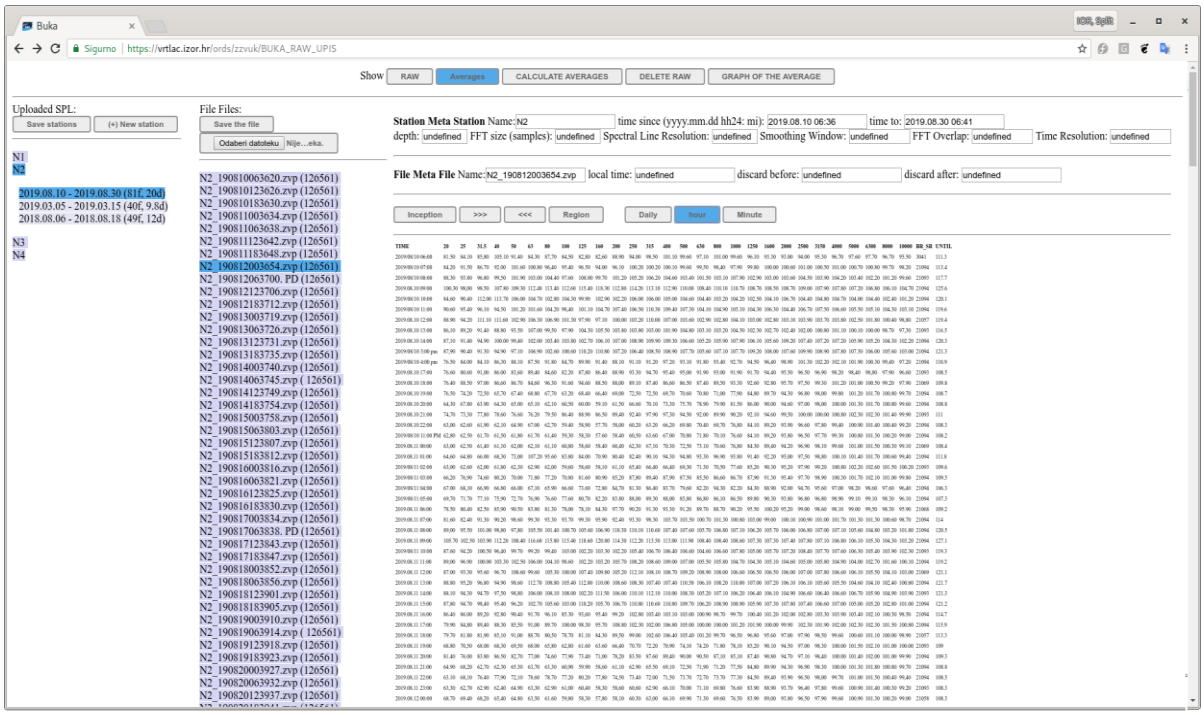


Figure 9 Web form for uploaded SPL files management

9 Data visualization

For data visualization, a special interactive view is developed using a Highcharts (Interactive JavaScript Charts) multi-platform charting library. Data retrieved to the client are in JSON (JavaScript Object Notation) format and they are processed on the client-side. User can select data in hierarchical way:

- Station Menu
- List of all available stations.

Data selection for an available period:

- Year
- Month
- Day

Show data using averaging period:

- Day
- Hour
- Minute.

The view is divided into three sections:

- Graph on top represents averaged SPL for appropriate 1/3 octave band over time or it represents a number of samples included in selected averaged period
- Graph in the middle represents averaged SPL spectrum for every 1/3 octave band for selected time period, with time slider and time controls
- The lowest graph represents the spectrogram for a whole loaded period. It is synchronized with the other two graphs. When a user clicks on some data in the spectrogram other two graphs automatically loads data for a clicked (selected) time and selected 1/3 octave band to show a more details.

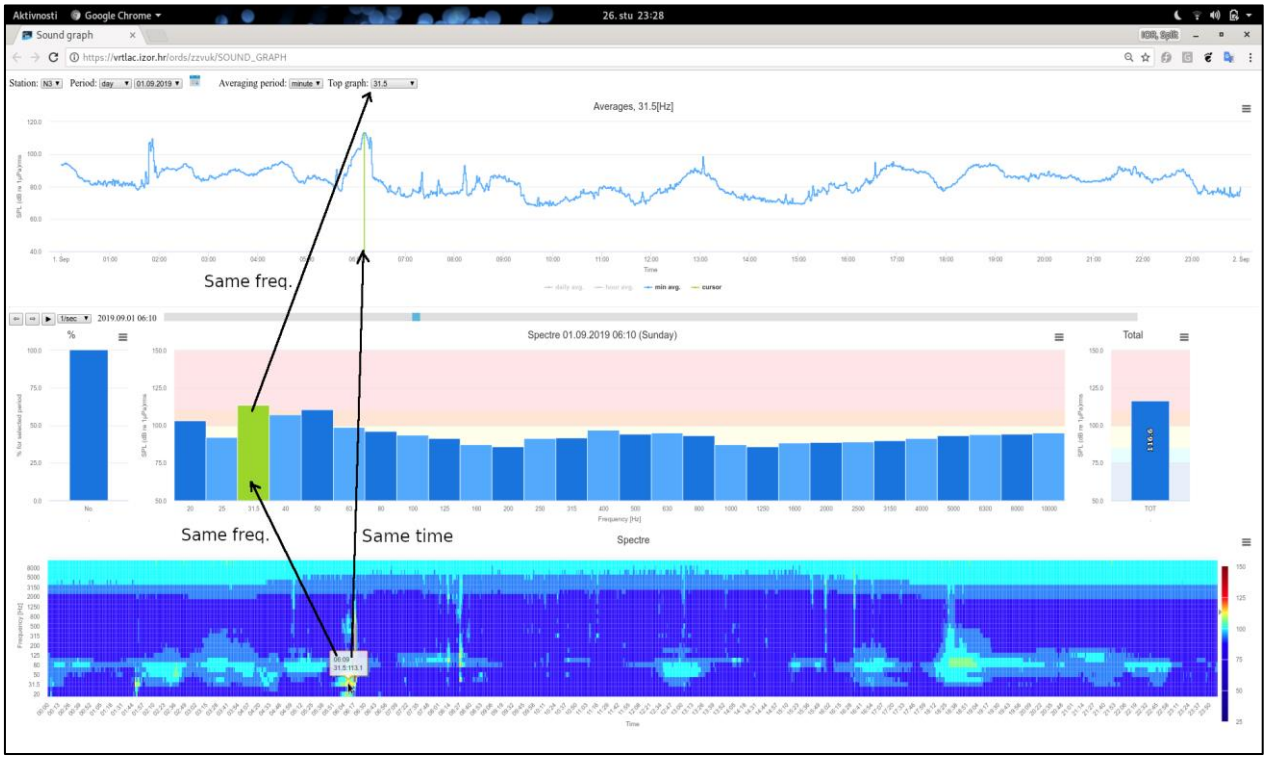


Figure 10 Visualizations page shows selected point in the SPL spectrogram which is synchronized with SPL spectrum and SPL graph for selected band and time.

10 Archived data in IOF and CNR databases and their availability

All data collected from nine stations during fifteen months' measurements have been archived in original "wav" data in two databases developed at IOF and CNR. According to the agreed data policy in the Soundscape project, "wav" data are classified, and 20 seconds' averaged SPL data are available for the broad community (End users).

Link to the IOF databases with the collection of available 20 seconds' averaged SPL data from all nine measuring stations collected during the recording period (March 1, 2020 – June 30, 2021) is:

FTP://data01.izor.hr/ (connect as: ***anonymous*** user).

At the Figure 11 there is an example of retrieved folders and data-files with 20 seconds averaged SPL archived t the NAS at the IOF server using *ftp* protocol.

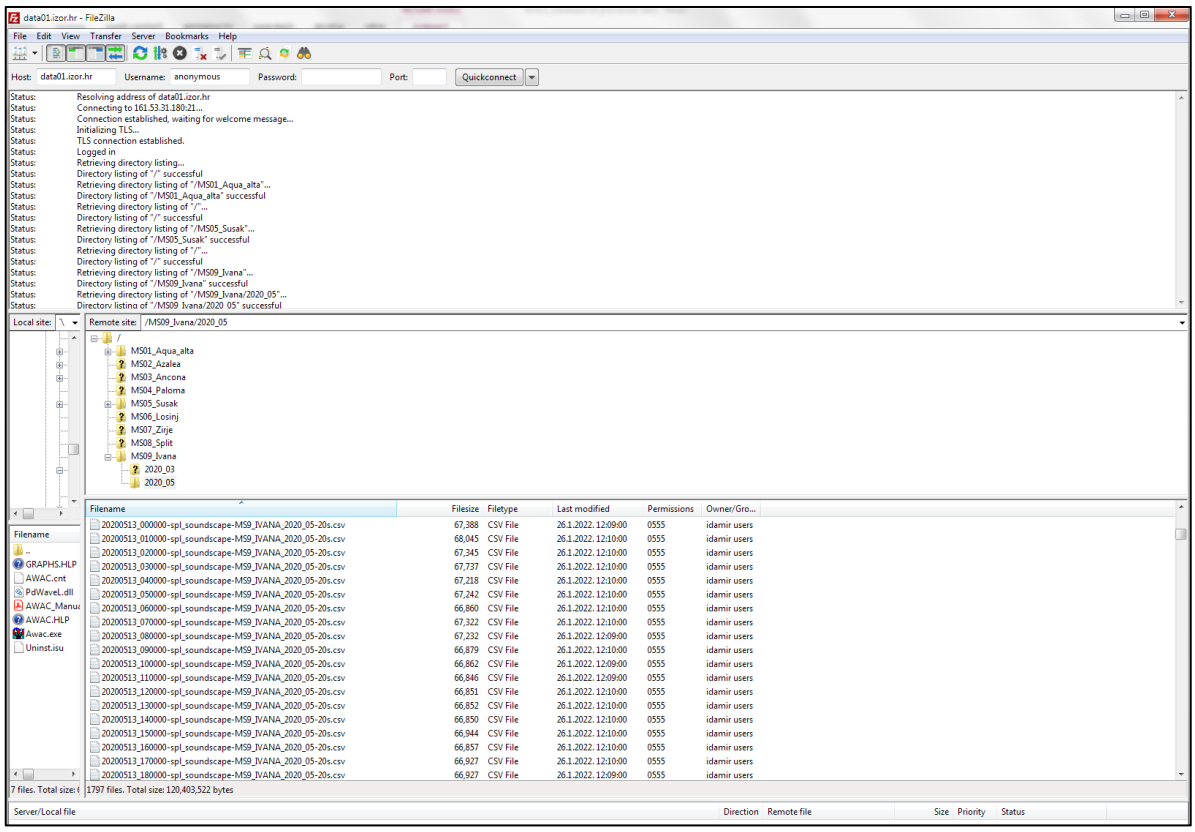


Figure 11 An example of selected folders and data-files with 20 seconds averaged SPL data using *ftp* protocol

General overview off the number of collected data by time at the all nine measuring stations is shown at the Figure 6 (data format) and Tables 1, 2, 3 and 4 (data sets by station and deployment).

Table 1 General overview of the number of collected data by time at the measuring stations MS1 Aqua Alta and MS2 Azalea

Measurement station	Data set No.	Measurement period	Data recording days	Total operate days per station
MS1 AQUA ALTA	1	2020/02/21 - 2020/03/11	19	448
	2	2020/04/09 - 2020/06/11	64	
	3	2020/06/15 – 2020/08/12	59	
	4	2020/08/12 – 2020/11/18	99	
	5	2020/11/18 – 2021/03/03	106	
	6	2021/03/11 – 2021/03/16	6	
	7	2021/05/04 – 2021/08/06	95	
MS2 AZALEA	1	2020/02/29 – 2020/04/30	62	375
	2	2020/05/31 – 2020/07/18	49	
	3	2020/08/01 – 2020/10/10	71	
	4	2020/10/24 – 2020/12/20	58	
	5	2021/01/30 – 2021/04/03	64	
	6	2021/04/25 – 2021/07/04	71	

Table 2 The general overview of the number of collected data by time at the measuring stations MS3 Ancona and MS4 Paloma

Measurement station	Data set No.	Measurement period	Data recording days	Total recording days per station
MS3 ANCONA	1	2020/02/21 – 2020/04/22	62	387
	2	2020/06/10 – 2020/09/10	93	
	3	2020/09/29 – 2021/01/11	105	
	4	2021/02/17 – 2021/05/05	78	
	5	2021/05/14 – 2021/07/11	49	
MS 4 PALOMA	1	2020/02/21 – 2020/03/05	14	336
	2	2020/03/11 – 2020/04/27	48	
	3	2020/04/27 – 2020/06/10	45	
	4	2020/06/10 – 2020/08/13	65	
	5	2020/08/13 – 2020/10/14	63	
	6	2021/02/05 – 2021/03/01	24	
	7	2021/03/18 – 2021/04/04	18	
	8	2021/04/08 – 2021/04/23	16	
	9	2021/05/01 – 2021/05/10	10	
	10	2021/06/17 – 2021/07/19	33	

Table 3 The general overview of the number of collected data by time at the measuring stations MS5 Susak and MS6 Lošinj

Measurement station	Data set No.	Measurement period	Data recording days	Total recording days per station
MS5 SUSAK	1	2020/03/05 – 2020/04/09	36	481
	2	2020/04/11 – 2020/06/11	62	
	3	2020/06/14 – 2020/08/09	57	
	4	2020/08/13 – 2020/10/18	67	
	5	2020/10/20 – 2021/01/15	88	
	6	2021/01/18 – 2021/03/17	59	
	7	2021/03/17 – 2021/06/01	73	
	8	2021/06/04 – 2021/07/12	39	
MS6 LOŠINJ	1	2020/02/22 – 2020/04/09	48	466
	2	2020/05/07 – 2020/06/11	36	
	3	2020/06/14 – 2020/08/09	57	
	4	2020/08/13 – 2020/10/18	67	
	5	2020/10/20 – 2021/01/15	88	
	6	2021/01/18 – 2021/03/17	59	
	7	2021/03/18 – 2021/06/01	76	
	8	2021/06/04 – 2021/07/08	35	

Table 4 The general overview of the number of collected data by time at the measuring stations MS7 Žirje, MS8 Split and MS9 Ivana

Measurement station	Data set. No.	Measurement period	Data recording days	Total recording days per station
MS7 ŽIRJE	1	2020/05/05 – 2020/08/01	89	383
	2	2020/08/01 – 2020/11/29	121	
	3	2020/12/04 – 2021/04/04	122	
	4	2021/05/19 – 2021/07/08	51	
MS8 SPLIT	1	2020/02/24 – 2020/05/06	73	397
	2	2020/05/06 – 2020/07/01	57	
	3	2020/07/01 – 2020/09/29	91	
	4	2020/11/25 – 2021/03/26	122	
	5	2021/05/14 – 2021/07/06	54	
MS9 IVANA D MS9* IVANA E	1	2020/03/10 – 2020/05/08	60	351
	2	2020/05/08 – 2020/07/22	76	
	3*	2020/12/14 – 2021/05/09	147	
	4*	2021/05/09 – 2021/07/15	68	

All documents uploaded at the Soundscape web portal and 20 sec averaged SPL data archived at IOF, and CNR databases are available to End users according to Open science principles, non-discriminatory, free of charge, without liability for consequences that the use of Data sets may result in. Available SPL data sets are archived in open Data format presented in Figure 6. Therefore, all data archived at the databases at IOF and CNR are considered freely available, and End Users should only be informed about the applicable data sharing License that reflects these principles. In addition, all data are open for usage to the general community with a Creative Commons Attribution version 4.0 (CC-BY 4.0).

Ž

Data sets are provided through a link at the Soundscape web portal “as is”. Therefore, soundscape project and project partners cannot be held responsible for the quality and accuracy of Data sets, misuse and misinterpretation of Data sets by the End Users, or any consequences arising from the use of Data sets by the End Users.

In order to maximize the usability of Data sets, data format has been explained in this document. Furthermore, more information about the result of processing data can be found in deliverable D3.6.3 and produced monthly spatial maps of underwater described, which uploaded at the Tools4MSP web portal (http://data.tools4msp.eu/tools4msp/) and described in deliverables D.5.3.1 and D5.3.2 published at the Soundscape web pages.