

# Underwater noise levels raw data database

---

Final Version of 15/10/2021

Deliverable Number D.3.3.2.



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	3.2
Activity Title	Field surveys for data collection
Partner in Charge	IOF
Partners Involved	CNR, BWI, ARPA FVG
Authors	Damir Ivanković (IOF), Stipe Muslim (IOF), Vlado Dadić (IOF), Hrvoje Mihanović (IOF), Ivan Rojnica (IOF)
Status	Final
Distribution	Public
Citation	Ivanković D., Muslim S., Dadić V., Mihanović H., Rojnica I. Underwater noise levels raw data database. SOUNDSCAPE project, WP3, 20pp, 2021

## Summary

1	Abstract.....	4
2	Introduction.....	4
3	Data exchange protocols – basic requirements .....	4
4	General requirements for storing raw data – data storage devices (DSD).....	5
5	General requirements for processed data storage – data sharing platform (DSP).....	6
	APPENDIX I: Synology Cross-Office File Syncing & Sharing technical specifications.....	9
	Storage Management .....	9
	Btrfs File System .....	9
	Directory Service.....	9
	Synology Directory Server.....	10
	Drive.....	11
	Drive PC client.....	12
	Snapshot .....	13
	Replication .....	13
	Recovery .....	13
	Security .....	14
	Log Center.....	15

## 1 Abstract

Document describes general requirements for storing raw and processed data in database, their validation and protocols for safely data exchange among project partners to be implemented in the Soundscape project

## 2 Introduction

Underwater acoustic measurements generate relatively large amount of data. For instance, an instrument that is continuously recording with a sample rate of 48 kHz and 16-bit depth will generate about 500 GB of raw data per month. Given that several sensors will be used within the SOUNDSCAPE project and that measurements will last about a year; large amount of data will be generated. This in turn creates special requirements for the data exchange protocols and data storage, compared to other types of environmental monitoring measurements.

The raw data will be processed by applying signal processing tools, and this will generate relevant statistical measures depending on monitoring requirements. Therefore, there are three main goals for data storage: (i) to store raw data and (ii) to provide access to data for signal processing use (SPL and (iii) to share results and output products. Usage of protocol/encryption and available authentication methods for accessing data online should be discussed. Another matter that needs to be discussed is legal classification of raw acoustic data and its availability for public. Thus, we need to define which kind of averages can be publicly available (1 second, 20 seconds, 1 minute, 1 hour averages?).

## 3 Data exchange protocols – basic requirements

Two harmonized systems should be deployed, one at IOF and one at CNR. On the same system storage will be organized for all three data levels:

1. Raw data files containing uncompressed sound (WAV files)
2. SPL processed data
3. Averages and processed data.

The simplest solution is to use Network Attached Storage with synchronization option as is Synology Cross-Office File Syncing & Sharing. Technical specifications of the system are in the Appendix I.

The data exchange protocols and database should be focused on three components:

- 1) Raw data (WAV files): Majority of disk usage will be used for storage of wav files. Disc capacity needs to be defined according estimated size of wav files, once modellers' needs and data processing procedures are agreed. Data exchange between partners could have two options: Option A: through network file system over Internet (NAS protocols (S) FTP, WEB(NAS), SMB, ...). Partners could have the same file system. Suggestion is to have equipment (NAS) of the same manufacturer, to be able to easily establish synchronization between partners, i.e., NAS computers. It is not necessary to have the same equipment, but it simplifies solutions. Option B: partners could send data on portable storage by post to the master storage system (i.e., at IOF address). The delegated person at IOF would copy data to the intended directory. Other partners would be able to download the data immediately. Automatic synchronization between partners (between NAS computers) can also be initiated immediately.
- 2) SPL spectra (single frequencies or 1/3 octave): this should be defined in coordination with modelling team.
- 3) Presentation of data, statistics, averages, maps etc. that needs to be agreed between the partners.

Besides data exchange protocols and data storage, data related to deployment and recovery of instruments need to be carefully recorded and metadata need to be stored.

## 4 General requirements for storing raw data – data storage devices (DSD)

The archiving structure of data storage devices (DSD) should reflect the organization of the measurements as well as the requirements of the users.

The basic requirements for DSD are listed below:

- The DSD should be easily accessible
- The DSD should be protected from malicious intrusion
- The DSD should offer a safe back-up facility (in case of synchronized IOF-CNR devices each one is backup for another one)
- The archiving/folder structure should be organized in intuitive way and provide access by GUI interface
- The data should be stored in raw (.wav) format. Particular file size/length should be discussed (6 or 8 hours output from instrument). Smaller size is more convenient for finding particular event and harder to manage (big number of files), and bigger files are more convenient for storage, but harder to find particular event. In case of database usage, database can provide "address" of event: file name and relative offset inside file.

## 5 General requirements for processed data storage – data sharing platform (DSP)

Data sharing platform can be divided to two parts:

- File system for storage of SPL Spectra files
- Relational database for storage and processing of SPL Spectra files

Size of SPL Spectra file in case of averaging of 1 second and 1/3 octave processing is approx. 0.5GB/1GB (10 kHz/20kHz) per station per month (text format) or zipped by average compression rate of 2.6 aprox. 190MB/360MB. For database storage requires storage for 2592000 rows in 30/60 columns (time, total and 28 freq.) per station per month.

Database should have web interface and allow the user to upload and download data. Also database should be able to load SPL Spectra files semi-automatically. All files from designated folder should be loaded by automated process/script without user interventions.

For purposes of automatic database data load for SPL Files we propose text format with metadata stored in file header.

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.developic	#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	#Processing 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	63.8200 60.0676 65.5164 86.2278 90.6028 94.0508 103.001 106.021 98.0571 90.3310 80.9	

Figure 1. Example of SPL file

For lowering possibility of processing errors in the SPL data files data should be in decibels. Inside database data are transform to linear scale, averaged and returned to decibels. Various averages should be calculated inside database:

- Minutes
- Hours
- Days
- Months

### File naming rules

SPL file name example: N1\_190723162704.spl

### Explanation

- N1 station ID (chars until \_ or predefined length 2 chars)
- 190723162704 date and time in format YYMMDDhh24miss (year 2 digit, month, day, hour 2 digit i 24hrs format, minute 2 digit and seconds 2 digits)

Same naming rule should 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, ...)

Proposed file format for SPL and processed data is text file with header and tab separated columns. This file type can be directly loaded to and produced by database without and transformations. Other formats can be discussed if it is necessary (netcdf, hdf5). In case of usage of other formats, appropriate software for file creation should be find and used.

### Security aspects

Both the DSP and the DSD have to offer protection that prevents non-authorized persons to access the units. NAS devices and database should be connectable using Internet. The users can only login by using username and password. To prevent transport of username and password thought unencrypted connection, secured layer should be used (SFTP, https, ssh).

### Synchronization

Two main NAS storage units (IOF and CNR) should be synchronized automatically on daily basis. Synchronization should be scheduled during the night, when network and system usage is lower.

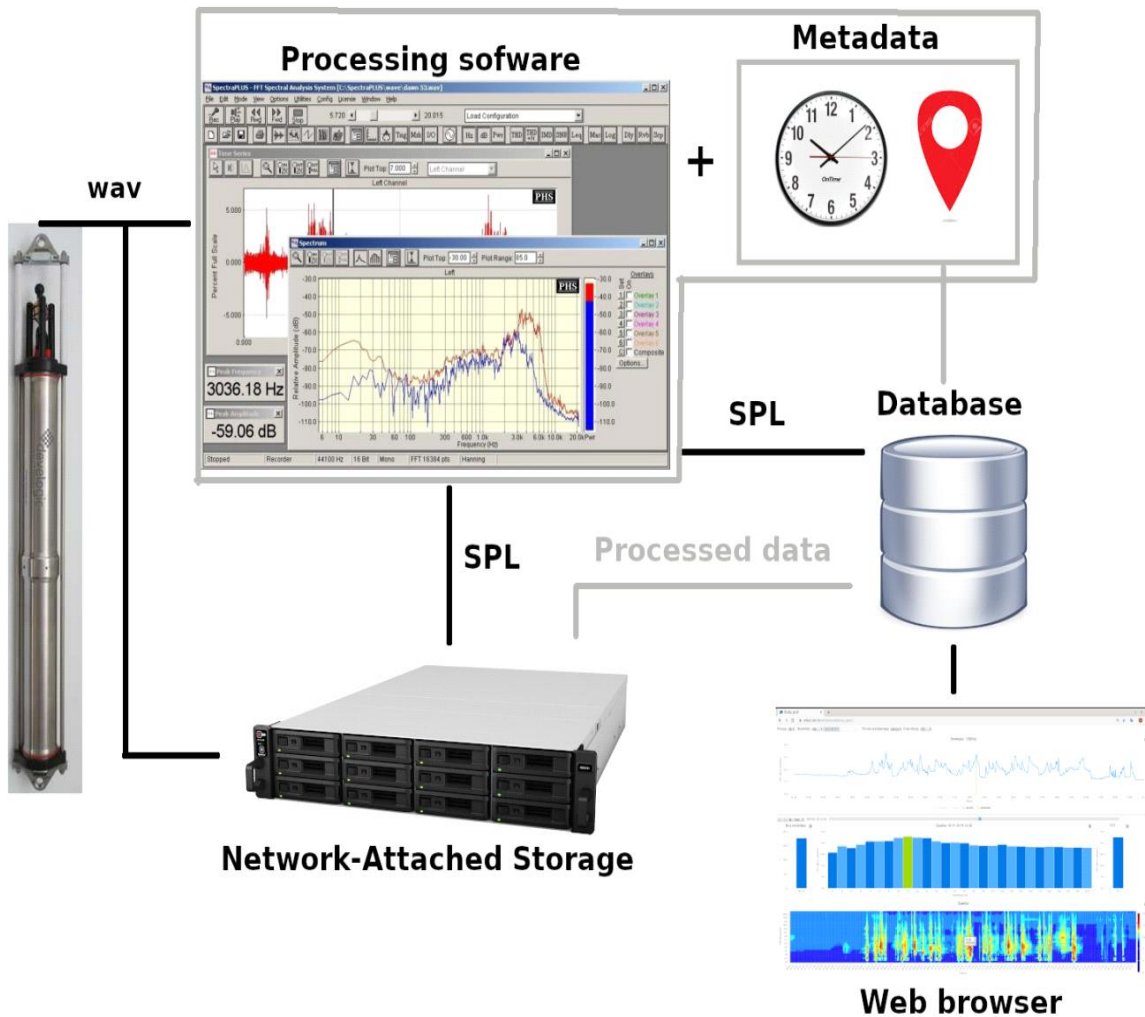


Figure 2. Data flow schema



## APPENDIX I: Synology Cross-Office File Syncing & Sharing technical specifications.

### Storage Management

- Supported raw capacity with expansion units: up to 1344 TB with RS18017xs+
- Support up to 512 shared folders\*
- Synology Hybrid RAID, Basic, JBOD, RAID 0, RAID 1, RAID 5, RAID 6, RAID 10, RAID F1 (Please check [RAID calculator](#) for the supported models of different RAID types)
- Global Hot Spare for RAID 1, RAID 5, RAID 6, RAID 10, RAID F1 and SHR (check the availability on [product comparison](#) page)
- Supports scheduling Data Scrubbing for RAID 5, RAID 6, RAID F1, SHR (composed of three or more disks), and BTRFS volume
- Supports single volume or multiple volumes on RAID
- Supports RAID Group (check the compatibility on [product comparison](#) page)
- File system for internal disks: ext4, Btrfs (check the availability on our [product comparison](#) page)
- File system for external disks: ext4, ext3, FAT, NTFS, Btrfs, exFAT and HFS+ (check the availability on our [product comparison](#) page)
- HDD Hibernation
- Scheduled S.M.A.R.T. tests
- Dynamic Bad Sector Mapping
- SSD cache support (check the availability [here](#))
- Requires 2 SSDs teamed for read-write caching

### Btrfs File System

- Set usage/user quota for shared folders
- Enable file self-healing to automatically detect silent data corruption, and to recover silent data error on RAID disks (RAID 1, RAID 5, RAID 6, RAID 10 and RAID F1)
- Enable file compression to increase the flexibility of volume usage
- Perform defragmentation on a volume to increase system performance
- Supports shared folder snapshots and replications
- Supports instant SMB server-side copy
- Supports browsing read-only snapshots

### Directory Service

- Supported local user accounts: up to 16,000 with RS18017xs+
- Supported local user groups: up to 512 with RS18017xs+
- Provider-consumer architecture

- Password strength rules for LDAP users
- Active Directory
  - Trusted Domain Support
- LDAP Server
  - Google Apps Single Sign-On Integration
- LDAP client
  - OpenLDAP
  - IBM domino
  - Customized Profile

#### **Apps supporting LDAP/AD users**

- SMB
- FTP
- WebDAV
- File Station
- Network Backup
- Cloud Station
- Cloud Sync
- Audio Station
- Video Station
- Mail Service
- Surveillance Station
- Personal Web Station
- Photo Station
- VPN Server
- Note Station
- Drive
- Moments

#### **System Requirements**

- Windows Domain: Windows Server 2003 Active Directory or later
- LDAP server: LDAP standard version 3 (RFC2251)

#### **Synology Directory Server**

- Support for Windows RSAT
- Support for TLS domain controller certification
- Support for custom NetBIOS domain names
- DNS auto registration
- Support for single domain controllers

### Account & Privileges

- Group membership and policies
- Roaming user profiles
- Account single sign-on
- Home folder support

### Security & Access Control

- Kerberos-based authentication
- Password reset via email
- Password strength policies
- Account lockout policies

### Domain Clients

- Microsoft Windows 7 and above
- Linux

### Drive

#### General

- Recommended number of hosted files: up to 5,000,000\*
- Recommended number of concurrent connections for PC clients: up to 2,000\*

#### File management

- Supports accessing and managing files via web browsers, desktop clients, and mobile platforms
- Supports creating folders and adding files of any types from the local end and Synology NAS
- Supports managing files with folders, labels and stars
- Supports moving and copying items to a different location including Team Folders and folders shared by others
- Supports sorting items by title, modified time, type, owner and size
- Supports various view modes including list view, detail view, thumbnail view and tile view
- Supports various hotkeys for efficient file management
- Supports searching for files using various criteria including keyword, location, type, owner, labels, size and date
- Customizable permission settings for folders and files to be set as private, internal or public
- Supports advanced protection links to add password and validity period for public file sharing
- The viewing, editing and managing permission can be customized for each account
- Supports Team Folder, which any account with permission can access directly without extra sharing settings required
- Support grasping major updates via notifications
- Supports operation logs and log export
- Supports downloading and restoring historical versions of file

- Link any folder to Synology Chat channels to notify users of important updates
- Administrators can set up user's My Drive and Team Folder from the Drive admin console
- Administrators can restrict non-admin users' privilege to share files publicly or with specific accounts

### Synchronization

- Supports syncing files and folders between PCs and Synology Drive with [Drive PC Client](#)
- All files in My Drive, Team Folder, and Shared with me can be synced to users' desktop
- Administrators can configure different sync profiles for each user to limit the file size and types that can be synced
- Provides client lists for viewing connected devices

### Administration

- Keeps up to 32 historical file versions with incremental data
- Folder-level restoration with timeline navigation is supported for administrators in Drive Admin Console
- Administrators can adjust content indexing service's scope for Drive folders
- Administrators can customize the maximum number of file versions for each Team Folder or user's My Drive
- Administrators can disable the versioning function

### Integration

- Synology documents, spreadsheets and slides can be created in Synology Drive by installing the add-on package, Synology Office 3.0
- Images and videos in Synology Moments can be managed using Synology Drive
- Supports backing up the Drive package via Hyper Backup

### Drive PC client

- Drive PC Client is supported on Windows, Mac, and Ubuntu
- Supports creating connections and syncing with multiple Drive servers
- Supports syncing files and folders shared by other users
- Supports double clicking on the tray menu to open files of any event in web browsers
- Supports single clicking on the tray menu to open containing folders
- Supports creating sharing links and advanced protection links for synchronized files
- Provides minimalist tray icons
- Open Drive in web browsers from the system tray
- Supports instant pausing or deleting of established connections
- Supports filtering event and notification logs by server
- Notified files can be opened in containing folders or synced from the corresponding notification logs
- Icon overlays can be enabled or disabled according to user settings
- Customizable notification settings

- Context menu in local file browsers for adjusting sync settings or creating sharing links
- Customizable permission settings for sharing links to be set as public/internal and edit/view-only
- Supports proxy connections

#### **Snapshot**

- Local snapshots of shared folders on Btrfs file system and advanced file LUNs
- Schedule snapshot and customize retention rules
- Remove or lock versions
- Browse and restore snapshots in the #snapshot folder in all file services or File Station
- View snapshot data in the Previous Versions tab in the Properties dialogue on Windows PC.
- Support up to 1,024 snapshots per shared folder\*
- Support up to 256 snapshots per iSCSI LUN\*

#### **Replication**

- Replicate data to another destination by taking and sending snapshots to local or another server. The maximum number of replication tasks is determined by different modules of Synology NAS. For more information, please go to [Synology's official website](#)
- Schedule replication and customize retention rules
- Send the initial replication copy via the Internet or storage devices
- Replicate data to multiple destination servers from the same source server
- Statistics of transferred size and sync duration are recorded and displayed as a report
- Overview topology of data replication
- Supports port forwarding by customizing address and port of a replication task
- Encrypt replication transmission
- iSCSI LUN replication can be managed via Synology Snapshot Replication or VMware vCenter Site Recovery Manager. (Installing Synology SRA on VMware Site Recovery Manager server is required.)

#### **Recovery**

- Restore to a previous version via various methods including common protocols, File Station, and Windows File Explorer.
- Perform a failover on shared folders or iSCSI LUNs and make the replicas on the destination server writable when the source server is not healthy
- Perform a re-protect action to sync data on the two servers again after performing a failover
- Perform a test failover to make a clone and run all necessary tests in the event of disaster
- Perform a force failover when the source server is healthy.

## **Security**

### **Security Advisor**

- Malware Detection and removal
- Account and password, network-related, System-related security scan
- DSM and packages version check

### **Firewall & Traffic Control**

- Allow/deny access to selected ports or services for specific IP addresses
- Create firewall rules based on geographic region
- Organize firewall rules into different firewall profiles
- Traffic Control according to services and TCP/UDP ports
- DDoS protection
- VPN pass-through

### **AntiVirus Essential**

- AntiVirus Essential powered by ClamAV scanning engine
- Full system scan
- Scheduled scan
- White List to exclude files from being scanned
- Automatic update of virus definitions

### **AntiVirus by McAfee**

- Full system scan
- Scheduled scan
- White List to exclude files from being scanned
- Automatic update of virus definitions
- Available on specific models only. Please see this [product comparison page](#) to learn more

### **Auto Block**

- Block after specified number of attempts
- Custom expiration time
- Black and White List

### **DSM SSO Client**

- Supports Synology SSO Server
- Supports OpenID Connect SSO Servers:
  - Azure
  - Websphere

### **Certificate Management**

- Supports the import and management of multiple certificates
- Certificate encryption algorithm only supports RSA (ECC algorithm is not supported)
- Supports multiple certificates for different services:

- Web Apps (HTTPS) and WebDAV
- FTP SSL/TLS
- Mail Services
- RADIUS Server
- VPN Server
- Supports the creation and auto-renew of Let's Encrypt certificate.

#### **TLS/SSL Profile Level Management**

- Supports multiple TLS/SSL Profile Level for different services:
  - Web Apps (HTTPS) and WebDAV
  - FTP SSL/TLS
  - Mail Services
  - RADIUS Server
  - VPN Server

#### **2-step verification**

- Time-based One-Time Password (TOTP) protocol
- 2-step verification enforcement for admin groups or all users
- Manage trusted devices

#### **Log Center**

- Receive up to 3,000 syslog events per second\*
- Historical log activity chart
- Archiving logs by specified time, number of logs, the data size, and different hosts
- Download archives as HTML or CSV files
- Proactive email notification according to log severity or specified keywords
- Customize log transferring protocol
- Customized log formats
- Advanced log search engine
- SSL connection
- Log rotation: 1 month, 3 months, 6 months, 1 year.

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

All data collected from nine stations for 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). More information there is in deliverable D3.6.1.

General overview off the number of collected data by time at the all nine measuring stations is shown at the Tables 1, 2, 3 and 4.

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		

Total size of raw data wav database is: **92220 Total files count**  
**28.95 TB total files size**

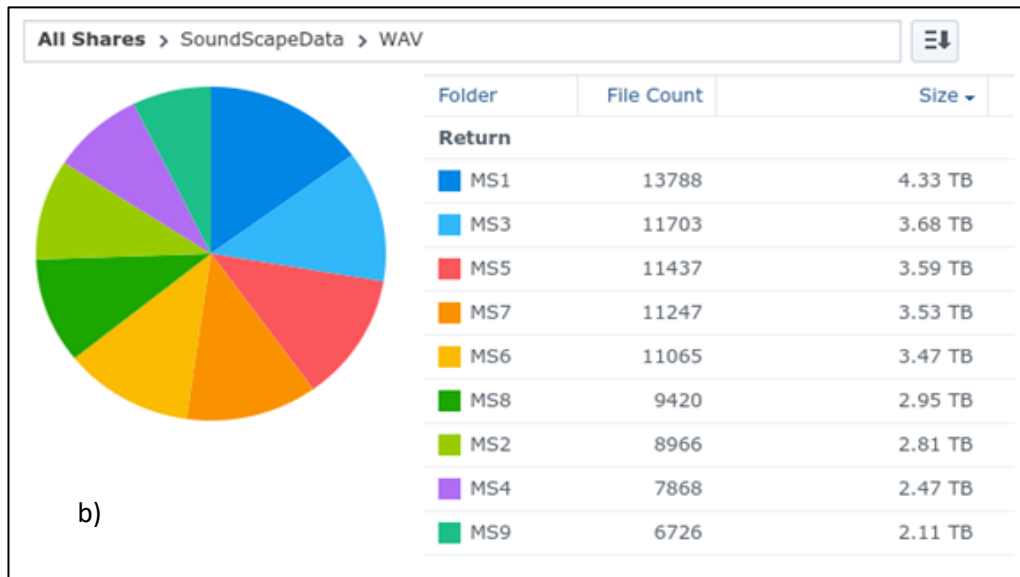
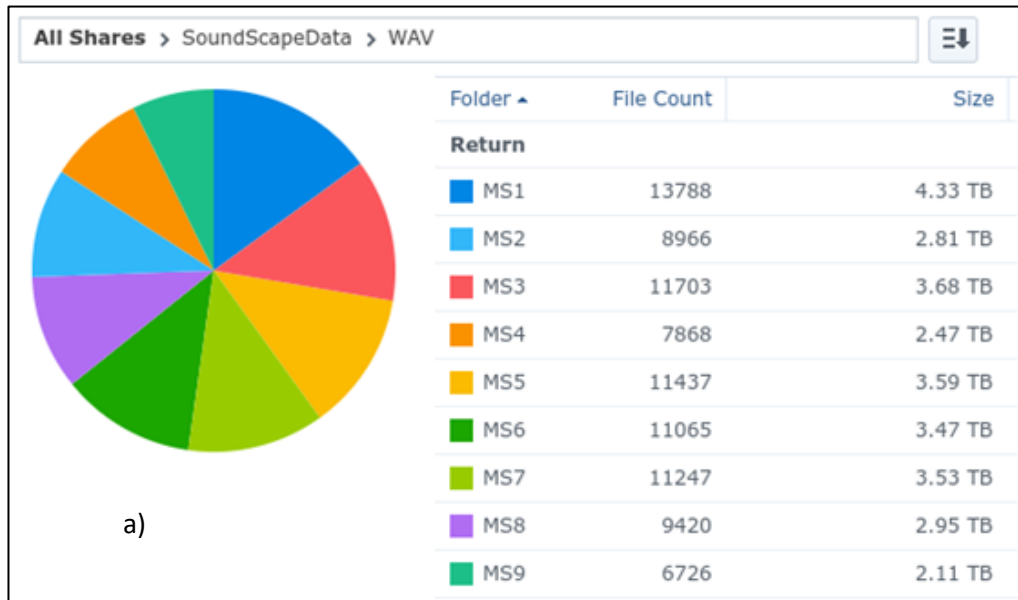


Figure 3 File count and size of wav database ordered by station (a) and file count and size of wav database ordered by total size (b)