



Rapporti Tecnici INAF INAF Technical Reports

Number	183
Publication Year	2022
Acceptance in OA@INAF	2022-10-03T10:44:15Z
Title	VOSpace user guide
Authors	URBAN, Cristiano; BERTOCCO, SARA; CALABRIA, NICOLA FULVIO; KNAPIC, Cristina; SPONZA, Massimo; ZORBA, SONIA
Affiliation of first author	O.A. Trieste
Handle	http://hdl.handle.net/20.500.12386/32678 ; https://doi.org/10.20371/INAF/TechRep/183



VOSpace user guide

Issue/Rev. No.	1.0
Date	Sep 14, 2022
Page	1 of 17

VOSpace user guide

Issue/Rev. No.: 1.0

Date: September 14, 2022

Authors: Cristiano Urban^a, Sara Bertocco^a, Nicola Fulvio Calabria^a, Cristina Knapic^a, Massimo Sponza^a and Sonia Zorba

Affiliations: ^aINAF - Istituto Nazionale di Astrofisica

Approved by: Cristina Knapic

Contents

1	Introduction	3
2	The VOSpace User Interface	3
2.1	Anonymous users	3
2.2	Authentication & authorization	4
2.3	Authenticated users	4
2.3.1	Main screen	4
2.3.2	Jobs screen	5
3	Operations on nodes	6
3.1	New folder	7
3.2	Create link	7
3.3	Move	8
3.4	Copy	9
3.5	Rename	9
3.6	Delete	10
3.7	Other functionalities	10
3.7.1	Archives creation	10
3.7.2	Sharing	11
4	Transferring data	12
4.1	Upload	12
4.1.1	Synchronous upload	12
4.1.2	Asynchronous upload	13
4.2	Download	13
4.2.1	Synchronous download	13
4.2.2	Asynchronous download	14
5	Frequently Asked Questions (FAQ)	14
6	Final remarks and future developments	16
	References	17

1 Introduction

The initial idea from which we started was to provide users with a tool to be able to easily recover their data previously stored into a tape library. The project structure then evolved into something more complex and led to the development of a VOSpace implementation compliant with the IVOA VOSpace specification[1]. The current VOSpace allows users to store and retrieve data using local, hot and cold storage points:

- a local storage point stores data uploaded through the VOSpace User Interface
- hot and cold storage points store data uploaded asynchronously using a transfer node.

Both local and hot storage points rely on the Lustre[2] parallel file system. Retrieving data from a local storage point is immediate since users just have to launch a download, so this operation is synchronous. Conversely, data stored into hot or cold storage points must be retrieved asynchronously. More precisely, the VOSpace creates symbolic links to make data available from hot storage points. In this way there is no need to perform a copy. Data stored into cold storage points (e.g. a tape library), instead, must be recalled. This report describes all the operations and the functionalities of the current INAF VOSpace implementation from the user point of view.

2 The VOSpace User Interface

The VOSpace User Interface frontend is a single-page application¹. Unlike other implementations, the VOSpace developed at INAF relies on an authentication system for accessing private data. This allows us to choose whether the data should be public or private.

In the following sections we will give a brief overview of the VOSpace UI distinguishing between anonymous users and authenticated users.

2.1 Anonymous users

Anonymous users cannot perform any action on data. They can only browse public data and download it, if available in place².

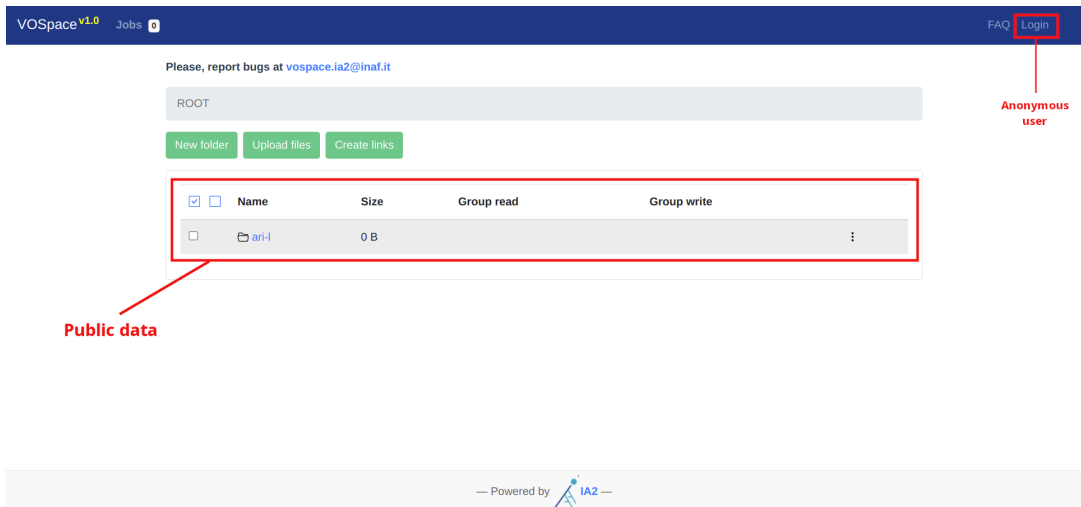


Figure 1: VOSpace user interface seen by anonymous users: the **ari-I** public folder is shown.

¹https://en.wikipedia.org/wiki/Single-page_application
²Data stored on disk or data stored on tape, but already recalled by an authenticated user

	<h1>VOSpace user guide</h1>	<table border="0"> <tr> <td>Issue/Rev. No.</td> <td>1.0</td> </tr> <tr> <td>Date</td> <td>Sep 14, 2022</td> </tr> <tr> <td>Page</td> <td>4 of 17</td> </tr> </table>	Issue/Rev. No.	1.0	Date	Sep 14, 2022	Page	4 of 17
Issue/Rev. No.	1.0							
Date	Sep 14, 2022							
Page	4 of 17							

2.2 Authentication & authorization

Users can login by clicking on the **Login** menu voice situated on the top navigation bar. The authentication procedure is handled by the Remote Authentication Portal (RAP)[3]. User permissions are managed through the Group Membership Service (GMS)[4]. For the first login, users must request registration via email to the VOSpace administrators.

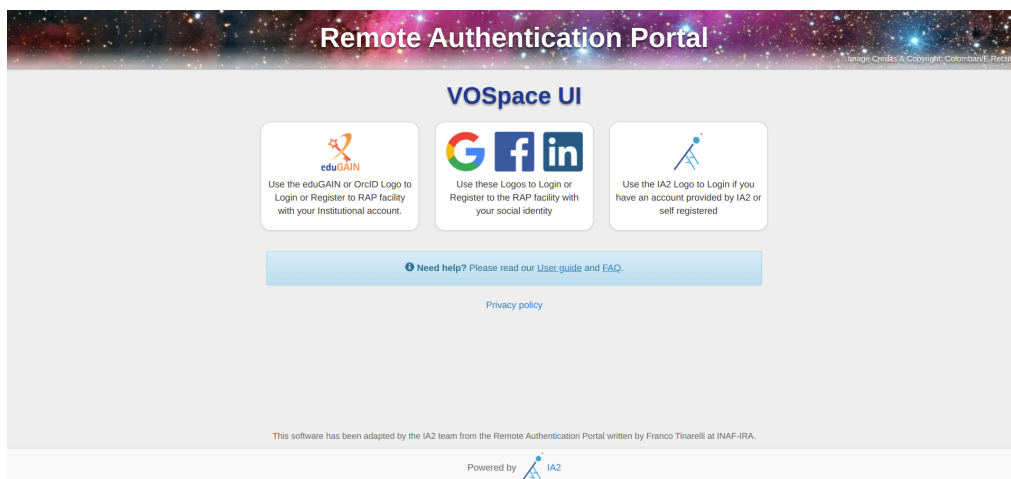


Figure 2: The Remote Authentication Portal (RAP).

2.3 Authenticated users

Logged in users have full access to the VOSpace web interface. They can essentially switch between two screens: the main screen and the jobs screen.

2.3.1 Main screen

Figure 3 shows the main VOSpace screen for an authenticated user. In the navigation bar we find, from right to left:

- the drop-down menu for user login/logout operations
- the **FAQ** (Frequently Asked Questions) menu item that opens the respective modal window
- the **Jobs** menu item which takes us to the user jobs screen.

Continuing down, we find the current VOSpace path (`ROOT/cristiano.urban`), some green buttons for generic actions (**New folder**, **Upload files**, **Create links**) and a blue button containing a drop-down menu that appears only if one or more nodes³ are selected.

Finally, the properties of the nodes present within the current VOSpace path are shown:

- **Name:** the node name
- **Size:** the node size in Bytes
- **Group read:** the list of groups which have read-only permissions on the node
- **Group write:** the list of groups which have read-write permissions on the node.

³In the VOSpace domain everything is a node, in the same way as in Unix everything is a file

The node name cannot contain the following characters: < > ? \ / " : | ' ` *`

The **size** is meaningful only for data nodes (files). For container nodes (folders) and link nodes (links to files or folders) the size is set to 0 Bytes.

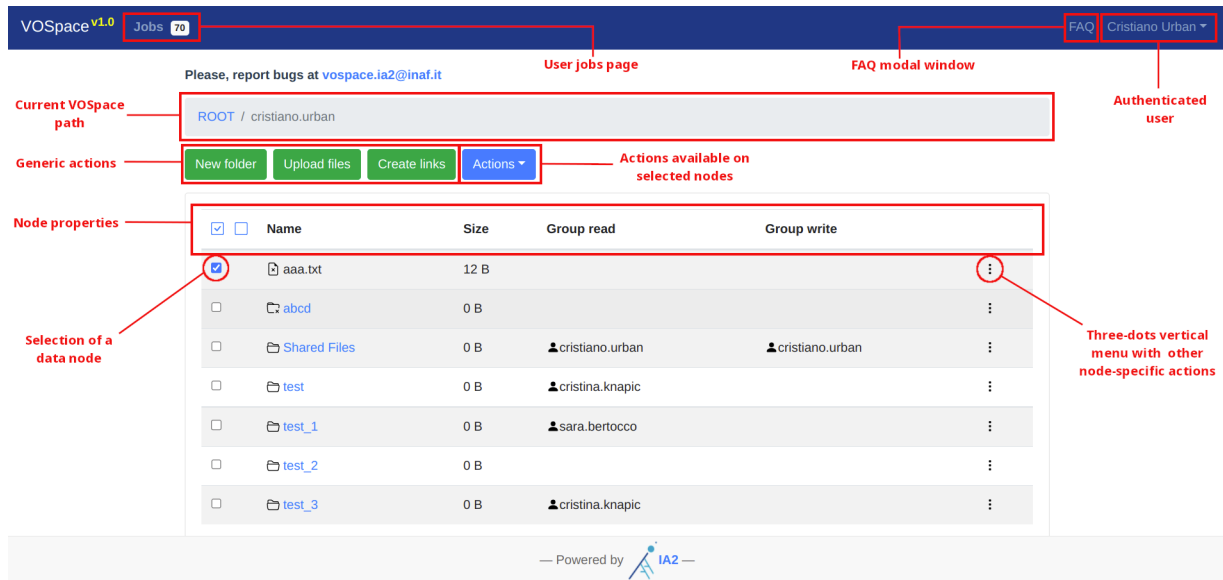


Figure 3: VOSpace main screen for authenticated users.

2.3.2 Jobs screen

By clicking on the **Jobs** item on the top navigation bar, we can reach the user jobs page. Close to the **Jobs** menu voice we can notice the presence of a number ('71' in this specific case). This number represents the job counter, namely, the total number of jobs launched by the user over time.

The **Reload** button allows the user to refresh the page results.

The page shows the complete list of the user jobs with some properties:

- **Type:** the job type
- **Creation time:** a timestamp showing when the job was created by the user
- **Id:** an alphanumeric string that uniquely identifies the job
- **Link:** a download link for previously generated tar/zip archives
- **Phase:** a string representing the job status (QUEUED, EXECUTING, COMPLETED, ERROR).

We have two types of job:

- **ASYNC_RECALL:** retrieves data from a storage point (hot or cold)
- **ARCHIVE:** generates a tar/zip archive for a one or more files/folders.

All these jobs are asynchronous, namely, they run in background.

A full description on how to build asynchronous services is provided by the Universal Worker Service pattern (UWS)[5].

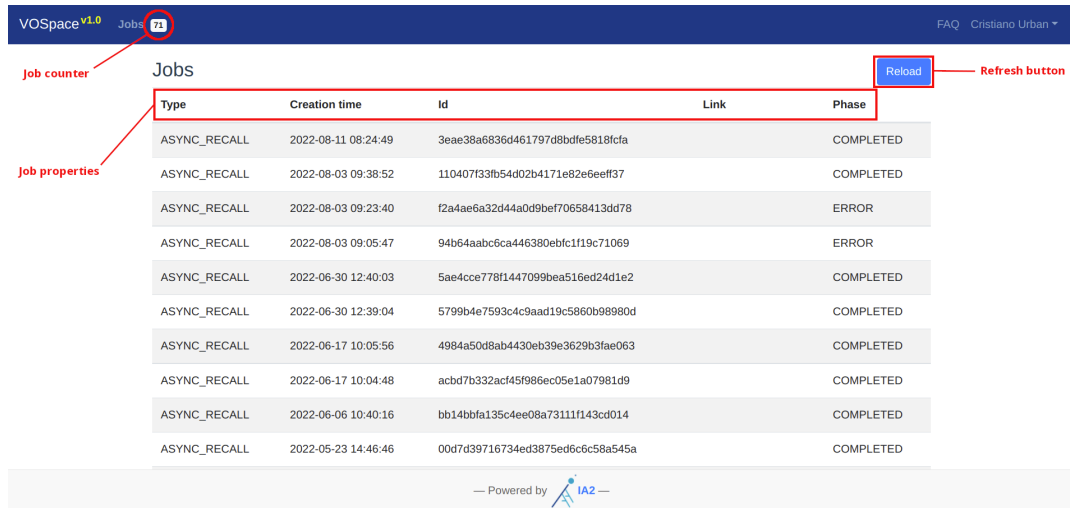


Figure 4: VOSpace user jobs screen.

3 Operations on nodes

Each object within the VOSpace is represented by a node, which in turn is identified by a URI. We can distinguish between three node types:

- **container node**: a node that can contain other nodes
- **data node**: a data object stored in the VOSpace
- **link node**: a pointer to another node.

According to the storage point we can classify two node types:

- **synchronous node**: a node stored on a local storage point, which can be directly downloaded by the user
- **asynchronous node**: a node stored on a hot or cold storage point, which has to be asynchronously recalled before to be downloaded by the user.

Data associated to asynchronous nodes can be dowloaded either through the web interface or through a transfer node⁴. Which one to choose depends on the volume of data involved.

Asynchronous nodes can be easily distinguished from synchronous nodes on the web interface, as the former have an icon marked with an **X**.



Figure 5: Asynchronous container node and asynchronous data node icons.

⁴more details are available in chapter 4 of this report

3.1 New folder

To create a new folder, just click on the green button **New folder**. A small modal window asking for the name of the folder will appear: write the name of the folder and then click on the blue button **Create**. At this point you should see the new container node on the current VOSpace path.

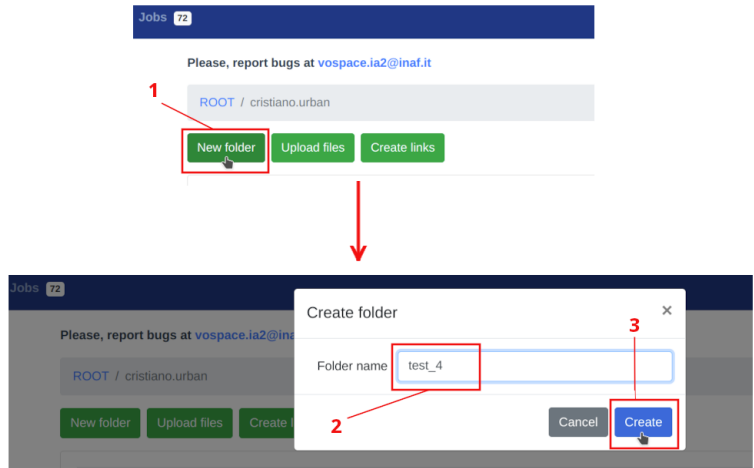


Figure 6: Creating a new folder.

3.2 Create link

Link creation allows users to choose between creating a single link or creating multiple links. A single link can be created by filling in the **URL** and **Node name** fields. If we need to create multiple links at a time, we have to upload a text file containing a list of URLs, each one separated by a new line. A limit of 1000 URLs per file is set by default. Finally, we need to click on the blue button **Create** and then the link(s) will be available on the current VOSpace path.

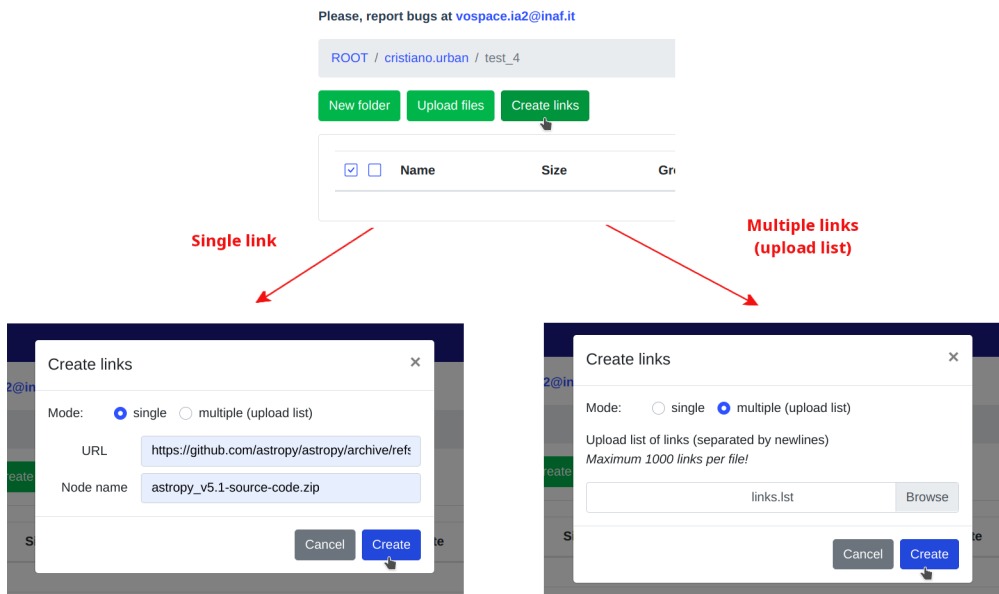


Figure 7: Creating links.



3.3 Move

The **Move** (or **moveNode**) operation allows the user to move one or multiple nodes from a VOSpace path to another.

This operation does not touch any file, it operates only at database level and this is the reason why it is very efficient.

Let's see how it works with a simple example. Suppose we have a data node called `VPN_OATs_nm_linux.pdf` located at the VOSpace path `ROOT/cristiano.urban/test_3`.

We want to move the data node from `ROOT/cristiano.urban/test_3` to `ROOT/cristiano.urban/test_4`. Figure 8 shows all the steps to perform:

1. select the data node by clicking on the checkbox
2. click on **Actions** → **Move**
3. a modal window appears: select the VOSpace destination path
4. click on the blue button **Move nodes**.



The **Move** action cannot be performed on asynchronous nodes.

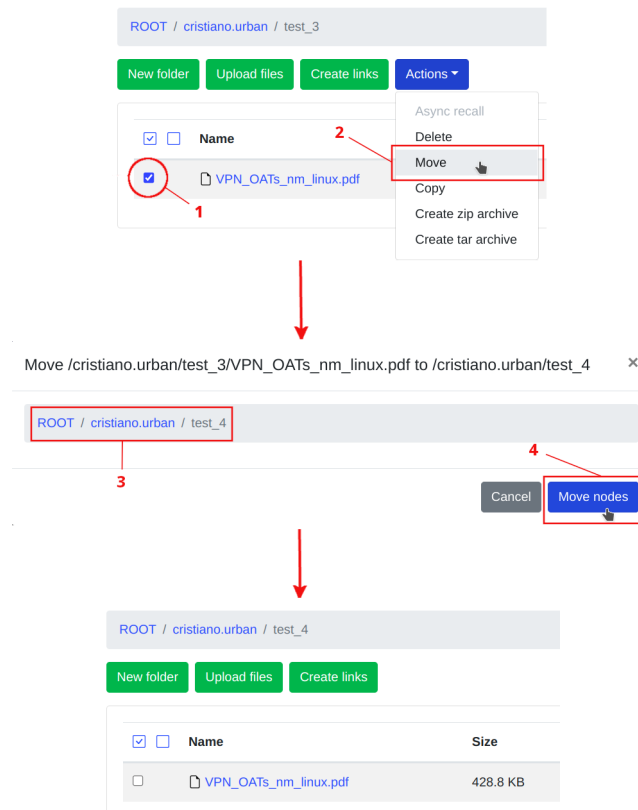


Figure 8: Moving a data node.





If we need to move a single node, another way is to click on the three-dots vertical menu associated to that node and then click on **Move**.

3.4 Copy

The **Copy** (or **copyNode**) operation allows the user to create a copy of one or multiple nodes to a given VOSpace path. The steps to follow are very similar to those described for the **Move** operation, with very small differences:

1. select one or multiple nodes by clicking on the respective checkboxes
2. click on **Actions** → **Copy**
3. a modal window appears: select the VOSpace destination path
4. click on the blue button **Copy nodes**.

 If we need to copy a single node, another way is to click on the three-dots vertical menu associated to that node and then click on **Copy**.

 The **Copy** action cannot be performed on asynchronous nodes.

3.5 Rename

A node can be renamed in a few steps:

1. click on the three-dots vertical menu of the node
2. a drop-down menu appears, click on **Rename**
3. a new modal window appears, insert the new node name
4. click on the blue button **Rename**.

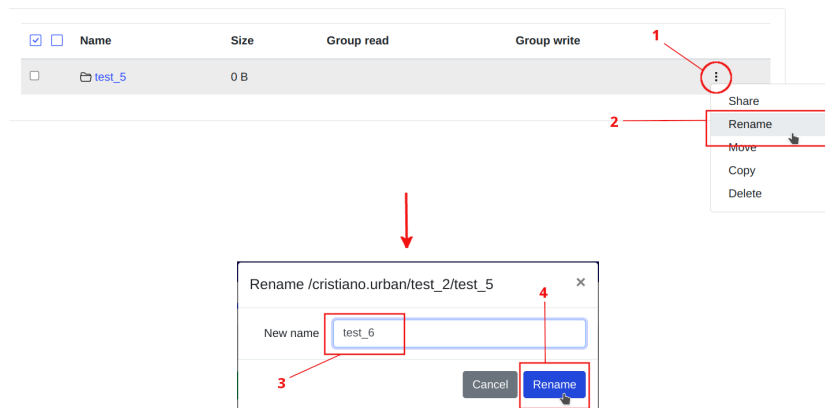



Figure 9: Renaming a container node.

 Asynchronous nodes cannot be renamed.



3.6 Delete

The user can delete one or multiple nodes at a time with a few steps:

1. select one or multiple nodes by clicking on the respective checkboxes
2. click on **Actions** → **Delete**
3. a modal asking for confirmation appears: click on the red button **Yes, delete**.



If we need to delete a single node, another way is to click on the three-dots vertical menu associated to that node and then click on **Delete**.



Deleting a container node involves deleting all of its contents. Deleting a link node does not delete the linked node, but only the pointer to that node. Also, asynchronous nodes cannot be deleted.

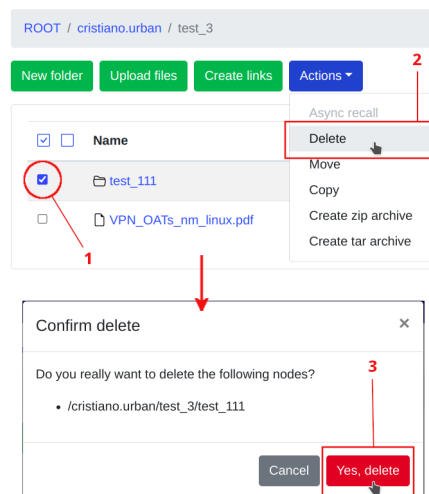


Figure 10: Deleting a container node.

3.7 Other functionalities

In these sections we briefly describe the following additional features:

- archives creation
- sharing data with users and groups.

3.7.1 Archives creation

A user can create and download a tar/zip from single or multiple nodes with five simple steps:

1. select one or multiple nodes by clicking on the respective checkboxes
2. click on **Actions** → **Create zip archive** (or **Create tar archive**)
3. at this point the tar/zip creation procedure will start in background and some pop-up messages will notify the status of the job

- once the job has finished, we switch to the Jobs screen and go down until we find the first job of type **ARCHIVE**
- now we can download the archive by clicking on **Download archive**.

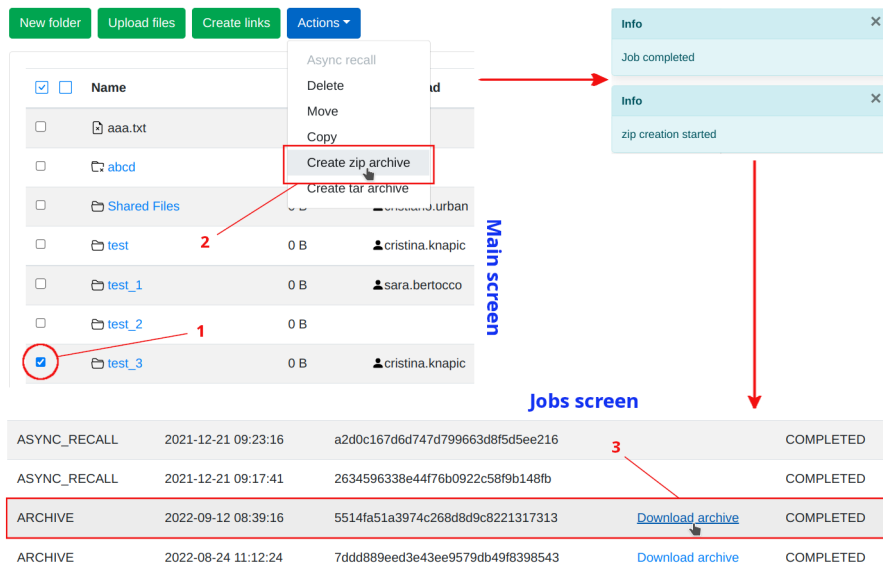




Figure 11: Creating a .zip archive for a container node.

 The whole data size involving the tar/zip creation must be less than 10 GB.

 Tar/zip archive creation is allowed only on data nodes and container nodes. If the user tries to perform this action on link nodes or on container nodes containing link nodes, the job will fail.

3.7.2 Sharing

The VOSpace specification implements sharing functionalities by using the properties `groupread` and `groupwrite`, so there is no concept of sharing a node with a specific user: there are always groups involved. The INAF VOSpace does not simply lie on `groupread` and `groupwrite` properties, but implements something that is very similar to the sharing feature present in well-known cloud platforms like Google Drive or OwnCloud. When a user shares a file with another user, a single-user group is used. Single-user groups are managed through the GMS.

Figure 12 shows how to share a container node with another user with read-only permissions. First of all we have to click on the three-dots vertical menu associated to that node and then on **Share**. At this point a modal window appears: here we can choose to share the node with a single user or with a group. We can also set the type of permissions: read permissions (read-only) or write permissions (read and write). Finally, we select the **Apply to children** flag and then we click to the blue button **Save**. At this point, the user with whom we shared the container node will find a link node pointing to that node within a special container node called "Shared Files".

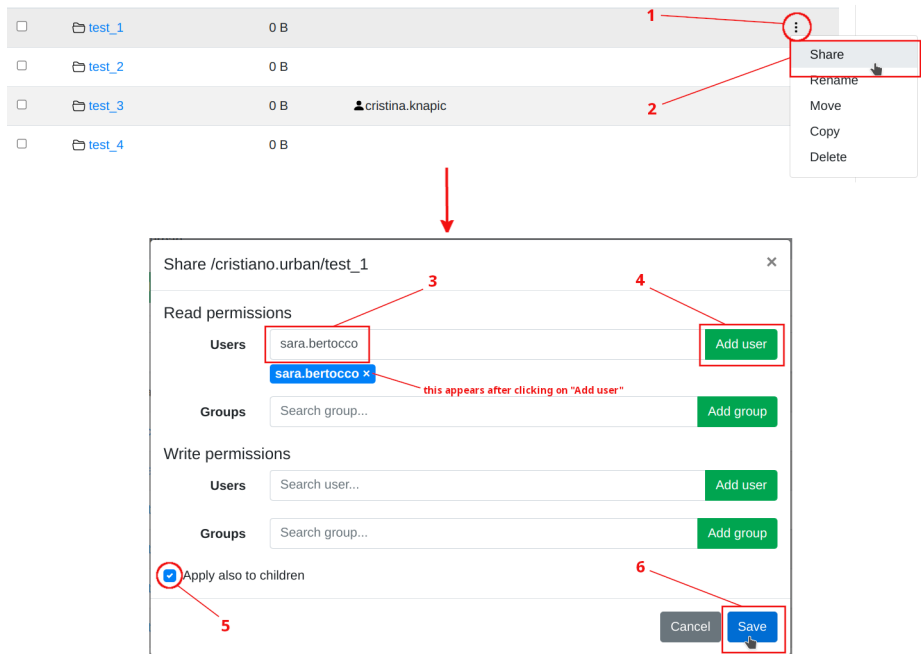


Figure 12: Sharing a container node with a user (read-only permissions).

4 Transferring data

In the following sections we describe how users can transfer data from and to the VOSpace service. The synchronous upload and download methods are made available to all the users and are used through the VOSpace UI.

The asynchronous upload and download methods are reserved for a limited number of users who need to store a huge amount of data in the order of terabytes and who want to recover portions of that data over time. This kind of users can ask to the VOSpace administrators for a SSH access to a transfer node.

4.1 Upload

According to the type of node and the storage point involved, we can distinguish between two different upload methods:

- synchronous upload
- asynchronous upload.

4.1.1 Synchronous upload

A synchronous upload operation consists of uploading files using the VOSpace User Interface. A user can upload one or more files at a time, but a single file must not exceed the size limit of 10 GB.

To upload files through the VOSpace web interface we just need to click on the green button **Upload files**, then click on **Browse** to select the files to upload and finally click on the blue button **Upload**. At this point the upload procedure will start and a progress bar will show the upload status.

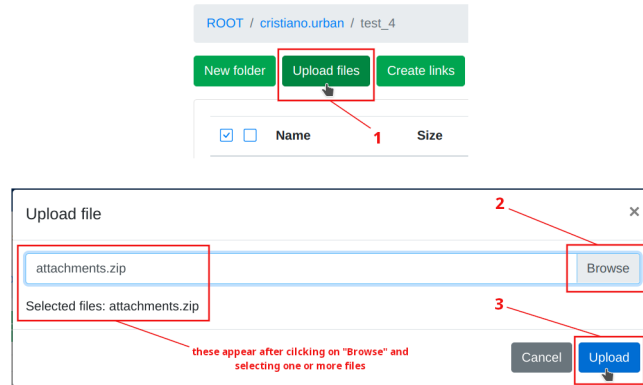



Figure 13: Uploading a file synchronously through the VOSpace UI.

4.1.2 Asynchronous upload

The asynchronous upload operation allows users to store data into hot or cold storage points by copying that data from an external storage point to a dedicated transfer node.

This is a semiautomatic operation: once data is copied on the transfer node the user must contact the VOSpace admins to ask them to perform the upload. At this point the admins can proceed to freeze and store the data into the right storage point, according to the user needs, and finally import metadata into the VOSpace database. At the end of these operations the user will be able to see the nodes on the VOSpace UI.

 **Important:** during the storage procedure, leaf directories containing more than 2000 files are converted in .tar files.

To copy data to the transfer node we can use `scp`^[6] or `rsync`^[7] over SSH. Here below are shown some examples:

```
# Copy 'foo.zip' file to the transfer node using scp
scp foo.zip name.surname@tn-bastion.ia2.inaf.it:~/store/

# Copy 'bar' folder to the transfer node using scp
scp -r bar name.surname@tn-bastion.ia2.inaf.it:~/store/

# Copy 'foo.zip' to the transfer node using rsync over SSH
rsync -arvh --progress --stats foo.zip name.surname@tn-bastion.ia2.inaf.it:~/store/

# Copy 'bar' folder to the transfer node using rsync over SSH
rsync -arvh --progress --stats bar name.surname@tn-bastion.ia2.inaf.it:~/store/
```

Listing 1: Using `scp` and `rsync` to copy data to the VOSpace transfer node.

In most cases the amount of data to copy is large and the transfer may take some time, so a good practice is to use `scp` or `rsync` in combination with the `screen`^[8] command.

4.2 Download

As with the upload action, also here we can distinguish between two different download methods:

- synchronous download
- asynchronous download.

4.2.1 Synchronous download

Downloading a file synchronously from the VOSpace is very simple and can be done by clicking on the name of the respective node. Obviously, this is true only for data nodes and link nodes, because clicking on the name of a container node means accessing the list of the contents of that node.

4.2.2 Asynchronous download

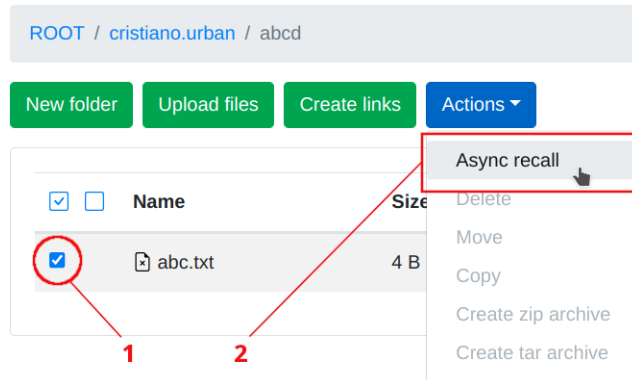


Figure 14: Recalling a file (data node) asynchronously through the VOSpace UI.

Data associated to asynchronous VOSpace nodes is not immediately available and must be recalled first. Here below are the steps to take:

1. select the nodes to recall by clicking on the respective checkboxes
2. click on **Actions** → **Async recall**
3. at this point a job starts in background and the user will be notified both via pop-up messages and via email on the job status
4. once the job is completed, refresh the VOSpace main screen on the browser and download the nodes by simply clicking on their name (for files that are not too large, in the order of tens of gigabytes)
5. for huge data, please, consider to retrieve the data using a transfer node, as explained with some examples here below.

```
# Retrieve 'foo.zip' file from the transfer node using scp
scp name.surname@tn-bastion.ia2.inaf.it:~/retrieve/foo.zip .

# Retrieve 'bar' folder from the transfer node using scp
scp -r name.surname@tn-bastion.ia2.inaf.it:~/retrieve/bar .

# Retrieve 'foo.zip' from the transfer node using rsync over SSH
rsync -arvh --progress --stats name.surname@tn-bastion.ia2.inaf.it:~/retrieve/foo.zip .

# Retrieve 'bar' folder from the transfer node using rsync over SSH
rsync -arvh --progress --stats name.surname@tn-bastion.ia2.inaf.it:~/retrieve/bar .
```

Listing 2: Using `scp` and `rsync` to retrieve data from the VOSpace transfer node: data is retrieved on the current directory (note that the dot is used as destination path).

As with the asynchronous upload, if the amount of data to retrieve is large, we can use `scp` or `rsync` inside a screen session.

5 Frequently Asked Questions (FAQ)

Q: Can I use the VOSpace UI as an anonymous user?

A: No, you must be logged in. Without login you can only browse and download public data, if available in place.

Q: How can I obtain an account?



A: Just send an email to the administrators at vospace.ia2@inaf.it.

Q: I don't see my home folder (`ROOT/name.surname`) after login and I cannot perform some actions on the VOSpace UI. What can I do?

A: Send an email to the administrators and ask for registration.

Q: Are there any restrictions on allowed characters for file/folder names?

A: Yes, the following characters are not allowed for file/folder names: `< > ? \ / " : | ' ` *`

Q: Is there a size limit to a file upload through the web UI?

A: Yes, the limit is 10 GB. If you need to upload huge files you have to ask the admins for SSH access to a dedicated transfer node. The admins will start the so-called storage procedure at the end of which the data will be available via web interface. Bear in mind that, during the storage procedure, leaf directories containing more than 2000 files are converted in .tar files. You can use the following Bash script to perform a check on your files before uploading them.

```
#!/bin/bash
#
# Prints path, num of files and dimension of a folder
# if the num of files is greater than a certain threshold
#
# Usage
# $ bash <script_name> <target_dir> <threshold>
#
# Example of usage:
# $ bash numfiles.sh mydir 1000
#
# Exit on errors
set -e

DIRS_LIST=$(du "$1" | awk '{ $1="" ; print substr($0,2) }')
TMP_OUTPUT="tmp_crowded_dirs.txt"
OUTPUT="crowded_dirs.txt"

touch $TMP_OUTPUT
touch $OUTPUT

echo "$DIRS_LIST" | while read i
do
    NUM_FILES=$(find "$i" -maxdepth 1 -type f | wc -l)
    if [ "$NUM_FILES" -gt "$2" ]; then
        DISK_USAGE=$(du -S -h -d 0 "$i" | awk '{ print $1 }')
        echo -ne "$i\x20" >> "$TMP_OUTPUT"
        echo -ne "$NUM_FILES\x20" >> "$TMP_OUTPUT"
        echo "$DISK_USAGE" >> "$TMP_OUTPUT"
    fi
done

# Creates a better view...
cat "$TMP_OUTPUT" | column -t > "$OUTPUT" && rm -f "$TMP_OUTPUT"
```

Listing 3: The numfiles.sh Bash script.


Q: Is there a limit to VOSpace number of nodes and storage space (hot or cold) per user?

A: Yes, obviously these limits are set by our available resources and can be agreed with IA2.

Q: Is there a size limit to a file download through the web UI?

A: Generally speaking the answer is no, but if you are going to download huge files, please ask the admins for a SSH access to a transfer node.

Q: Is there a size limit to a tar/zip creation through the web UI?

	VOSpace user guide	<table border="1"> <tr> <td>Issue/Rev. No.</td> <td>1.0</td> </tr> <tr> <td>Date</td> <td>Sep 14, 2022</td> </tr> <tr> <td>Page</td> <td>16 of 17</td> </tr> </table>	Issue/Rev. No.	1.0	Date	Sep 14, 2022	Page	16 of 17
Issue/Rev. No.	1.0							
Date	Sep 14, 2022							
Page	16 of 17							

A: Yes, each user can create a tar/zip from single or multiple files/folders if the whole data size is less than 10 GB.

Q: I have just created a tar/zip archive, how can I download it?

A: Go to your Jobs section, look for the last **COMPLETED** job of type **ARCHIVE** and click on **Download archive**.

Q: A file/folder icon is marked with an **X** and there is no download link. What does this mean?

A: It means that the file/folder is not in place and cannot be directly downloaded, it has to be "recalled" first. This can be done by selecting the file/folder checkbox, clicking on the **Actions** dropdown menu at the top of the page and then on **Async recall**. This action will start an asynchronous job to recall the file/folder and make it available for download. An email notification is sent to the job owner when the job is queued and when it completes.

Q: I have recalled some files/folders via Async recall, how long will the data be in place for download?

A: The data will be available for download for a few days (e.g. two days).

Q: How can I abort the execution of an asynchronous job?

A: Asynchronous jobs cannot be stopped.

Q: Can I share a file/folder with other users?

A: VOSpace specification implements file sharing by using the properties **groupread** and **groupwrite**, so there is no concept of sharing a file with a specific user, there are always groups involved. Anyway, file sharing services usually provide a mechanism to share a file with single users, so our UI implements it. Just click on the 3-dots dropdown menu on the right of your file/folder and then on **Share**. On the modal window you can choose and add the users/groups to share the file/folder with.

Q: Where can I find the VOSpace specification?

A: The VOSpace specification v2.1 can be found here: <https://www.ivoa.net/documents/VOSpace/>.

Q: How can I ask for support?

A: You can send an email to the administrators at vospace.ia2@inaf.it.

6 Final remarks and future developments

In this document we have recalled some basic concepts concerning the VOSpace specification and we have described the current INAF VOSpace implementation from the user perspective.

We have seen how we can operate on nodes and how to transfer data from and to the VOSpace service.

Possible future developments may include DOI[9] integration and exploring more efficient upload/download options for asynchronous nodes (e.g. SFTP⁵ and GridFTP[10]).

⁵https://it.wikipedia.org/wiki/SSH_File_Transfer_Protocol



References

- [1] *VOSpace specification v2.1*. June 2018. URL: <https://www.ivoa.net/documents/VOSpace/20180620/REC-VOSpace-2.1.html>.
- [2] *The Lustre file system*. URL: <https://www.lustre.org/>.
- [3] *Remote Authentication Portal (RAP)*. URL: <https://sso.ia2.inaf.it/rap-ia2/>.
- [4] *Group Membership Service (GMS)*. URL: <https://sso.ia2.inaf.it/gms/>.
- [5] *UWS specification v1.1*. Oct. 2016. URL: <https://www.ivoa.net/documents/UWS/20161024/REC-UWS-1.1-20161024.html>.
- [6] *OpenSSH*. URL: <https://www.openssh.com/>.
- [7] *Rsync*. URL: <https://rsync.samba.org/>.
- [8] *GNU Screen*. URL: <https://www.gnu.org/software/screen/>.
- [9] *Digital Object Identifier (DOI)*. URL: <https://www.doi.org/>.
- [10] *GridFTP*. URL: <https://gridcf.org/gct-docs/latest/gridftp/index.html>.