FAIR Storage Service

DRAFT: Doc in preparation
Comments, suggestions and corrections are really appreciated

Lorentz Institute offers its members a data storage option that complies with the ***FAIR*** *Guiding Principles for scientific data management and stewardship.*[1] The goal is to offer a standard platform that enables, facilitates, and promotes the implementation of robust data-management plans complying with the most stringent professional requirements (scientific publications, research grants applications, etc..). Our FAIR storage server is powered by the *iRODS* Data Management Software[2], the *de facto* standard in scientific data management. Below you will find brief instructions on how to interact (store/view/retrieve data) with our storage, nonetheless you are strongly advised to read the official iRODS documentation before operating on our server.

The IL *FAIR* storage system is fault tolerant to prevent any catastrophic data loss. Data-loss protection occurs on two levels: hardware and software.

The first protection layer against data loss is provided by *ad hoc* enterprise-grade storage servers that constitute the backend of the iRODS system. Our main storage system is a TrueNAS Enterprise R50 with disks arranged in a RAIDZ2 array with a fault tolerance of two disks.

The second layer of protection against data loss is provided by iRODS replicas (copies of data across multiple storage backends). We offer a storage resource in which two copies of each data object are kept at any time across two different enterprise-graded storage systems essentially increasing fault tolerance.

Research data older than 10 years could be moved to different storage systems.

FAIR Storage Access

Access to this server must be requested by sending an email to support@lorentz.leidenuniv.nl.

Once access has been granted, you can interact with the server via any of the supported iRODS clients. For the sake of simplicity, we have set up all the IL GNU/Linux workstations and the xmaris cluster with the following iRODS clients so that you can start using our storage system immediately

- iCommands
- DavRODS

iCommands is the most flexible and powerful client, but it might require a basic prior knowledge of the GNU/Linux command line.

At the Lorentz Institute we also offer several automated data-ingestion rules that post-process submitted data to our storage system using iCommands, for instance

- automatic storage of metadata of jupyter notebooks
- automatic storage of EXIF metadata of JPG, PNG images
- automatic addition of metadata to any data object from simple templates
- custom ingestion rules for your data? Contact support@lorentz.leidenuniv.nl

 Please note that the automatic post-processing functionality has currently been tested only with iCommands and DavRODS.

iCommands Setup

Create an iRODS configuration file at `${HOME}/.irods/irods_environment.json`. If the directory `${HOME}/.irods` does not exist, then create it

```
mkdir -p ${HOME}/.irods
```

Create/edit in the directory mentioned above a file named `irods_environment.json` starting from the following simple template

```
{
    "irods_default_number_of_transfer_threads": 4,
    "irods_host": "<HOSTNAME>",
    "irods_maximum_size_for_single_buffer_in_megabytes": 32,
    "irods_port": <PORT>,
    "irods_default_resource": "<RESOURCE_NAME>",
    "irods_transfer_buffer_size_for_parallel_transfer_in_megabytes": 4,
    "irods_user_name": "<USERNAME>",
    "irods_zone_name": "<ZONE_NAME>",
    "irods_authentication_scheme": "PAM_password"
}
```

Edit the lines above at your convenience by looking at the available options and their meaning in the docs. The advice here is to keep your conf file simple at the beginning by just copying the template above to the specified location and by appropriately replacing the tags denoted by < > with the desired values. To connect to the IL *FAIR* storage you have the following possible options

Tag	**Values**	**Notes**
HOSTNAME	icat.lorentz.leidenuniv.nl	Access limited to IL subnet
PORT	1247	
RESOURCE_NAME	ilorentzNoReplicaResource	135 TB. Quota 20 GB. Use: Archiving
	ilorentzResource	20 TB, Quota: 20 GB. Use: Archiving Important Data
USERNAME	*Your_IL_username*	Request access to support@lorentz.leidenuniv.nl
	anonymous	Anonymous access to public share
ZONE_NAME	ilZone	

Please not that even if you specify a default resource name in your iRODS configuration file, it is always possible to overwrite the resource destination during iCommands operations by specifying the

option `-R`.

In a terminal window of your GNU/Linux workstation or the xmaris cluster, load the iCommands module

```
module load icommands
```

and initiate a connection to our storage server by typing `iinit`

```
remote [878] $ iinit
Enter your current PAM password:
remote [879] $
```

Upon a successful authentication you will be able to interact with our *FAIR* storage server. Many iRODS commands have the same names and use of common GNU/Linux terminal commands with the only difference that they are prefixed by the letter `i`. For example to list the contents of your home collection (that is your home directory on the storage server), just type `ils`

```
remote [879] $ ils
/ilZone/home/bongo:
  irodsfs_amd64_linux_v0.7.6.tar
  QT_7b.mp4
  test1
  test2
  test3
  C- /ilZone/home/bongo/TEST1
```

A more detailed output is obtained by using the options `-l` (long output) or `-L` (super long output) and an overview of the available options is obtained by passing the option `-h`.

A list of available commands is given by typing `ihelp` or by browsing to the iRODS online documentation.

Common IRODS Operations

 In what follows iRODS objects are either *data* objects (e.g. a file) or *collection* objects (e.g. a directory).

Session Management and Info

Login	iinit
Logout	iexit
List icommands Available	ihelp
List Client Settings	ienv
List User Info	iuserinfo
List Available Storage Resources	ilsresc

Listing

More Info:	**ils -h**
Collection Listing	ils
Collection Listing including Replicas Information	ils -l
Collection Listing with Replicas Information and Actual Object Location on Server	ils -L
Collection Listing including ACL Information	ils -A

Objects Upload

More Info:	**iput -h, irsync -h**
Upload New Data Object	iput source_data
Upload and Overwrite Existing Data Object on Server	iput -f source_data
Add metadata to Data Object when Uploading it	iput [-f] object_name –metadata="A;V;[U];A;V;[U]"
Local Directory Upload	iput -r source_directory
Synchronize local Data to remote Data and *viceversa*	irsync source destination_object
Upload local file to Specific Collection	iput source_data destination_collection
Upload Data Object and Store its Checksum	iput source_data -k

Files Removal from Server

More Info:	**irm -h**
Remove Data Object (moves it to Trash)	irm data_object
Permanently Remove Data Object	irm -f data_object
Permanently Remove Collection and its Contents	irm -rf collection
Purge Trash Bin	irmtrash

Object Organisation

More Info:	**icp -h, imv -h, ilocate -h**
Change Working Collection on Server	icd collection
Print Working Collection	ipwd
Locate Object	ilocate search_pattern
Copy Objects on Server (No Metadata)	icp source_object destination_object
Move Object	imv source_object object_destination

Files Download From Server

More Info:	**get -h**
Display Remote File Contents	iget data_object -
Save Remote File to Local Disk	iget data_object local_destination
Save Remote File to Local Disk even if it Exists	iget -f data_object local_destination

Metadata

Metadata are one of the most important features of iRODS that makes this software compliant with the *FAIR* principles. We advice you spend some time familiarize yourself with the concept of metadata and you get used to the idea that data without metadata is in practice incomplete data.

iRODS metadata are defined by Attribute-Value-Unit (AVU) triplets, for instance Length 10 meters

More Info:	**imeta -h, iquest -h**
Add Metadata to [Data, Collection] Object	imeta add [-d -C] object AttName AttValue [AttUnits]
Add metadata to Object when Uploading it	iput [-f] object_name –metadata="A;V;[U];A;V;[U]"
List Metadata Associate to Object [Data, Collection]	imeta ls [-d -C] object
List Available Metadata Attributes	iquest "select META_DATA_ATTR_NAME"
Search Object from Metadata	iquest "select DATA_NAME, COLL_NAME, META_DATA_ATTR_VALUE where META_DATA_ATTR_VALUE like '%pattern%'"
List Available iquest select Attributes	iquest attrs

Object Permissions: ACLs

More Info:	**ichmod -h**
List Object ACLs	ils -A object
Grant Other IL User READ Access to Data Object and its Metadata	ichmod read IL_USERNAME data_object_path
Grant Other IL User READ/WRITE Access to Data Object and its Metadata	ichmod write IL_USERNAME data_object_path
Remove Access for IL User to Object	ichmod null IL_USERNAME object_path
Grant group READ Access to Data Object and its Metadata	ichmod read group_name data_object_path
Grant Other IL User Recursive READ Access to Collection	ichmod -r read IL_USERNAME collection_path
List Existing Groups	iquest "select USER_GROUP_NAME"

Integrity: Checksums

More Info:	**iput -h, irsync -h, ichksum -h**
Check Object Integrity during Transfer	iput -[r]K object
Check Object Integrity during Transfer	irsync -[r]K source i:dest_object

External Collaborators

iRODS lets you easily share your data with external collaborators (users unknown to the IL systems). The only requirement is that they have access to iCommands or have a web browser. See the examples below

iCommands: Tickets

iRODS tickets constitute a powerful and flexible way to share your data with external collaborators who have access to iCommands at their institutions. In the example session below an IL user creates

a *read-only* access ticket for a data object called *results.dat* and shares this unique alphanumeric code with his collaborator who will use it to gain access to the data.

```
# IL user iRODS session
iticket create read results.dat
ticket:68CK4jheDK924Jz

# Take note of the ticket id and pass it to your collaborator with
# the full pat to your data object

iticket ls 68CK4jheDK924Jz
...
ticket type: read
obj type: data
owner name: bongo
owner zone: ilZone
...
expire time: none
data-object name: results.dat
data collection: /ilZone/home/bongo
No host restrictions
No user restrictions
No group restrictions
```

```
# External collaborator iRODS session
# Must first login as anonymous on the IL iRODS system (see above)
iget -t 68CK4jheDK924Jz /ilZone/home/bongo/results.dat
```

```
# IL user iRODS session
# delete the ticket if no longer needed
iticket delete 68CK4jheDK924Jz
```

If your collaborator has no access to icommands, then place the object you would like to share in the folder `/ilZone/home/anonymous` and *ichmod* to give the user anonymous reading access to your object which will be shared using the web browser at https://access.lorentz.leidenuniv.nl/anon .

Web Browsers: webdav

Without creating a ticket for a data object, you could login to our *FAIR* storage system place any object you would like to share in a special collection called `/ilZone/home/anonymous`. Any external collaborator can then access it (READ ONLY) by browsing to https://access.lorentz.leidenuniv.nl/anon in a similar fashion to the example here.

Custom Lorentz Institute Ingestion Rules

To facilitate the uploading of objects metadata – an important component of any *FAIR* storage system – the IL storage system is programmed to automatically add the most basic piece of metadata upon each data object upload making sure that each stored data object has a minimal metadata

consinstency. The metadata added to each data object is summarised in the table below

ATTRIBUTE NAME	DESCRIPTION	EXAMPLE
displayName	IL Employee Directory Listing: Employee Full Name	John Smith
uid	IL Employee Directory Listing: Employee Username	smith

Further, the uploading of *.ipynb*, *.png* and *.jpg* files will automatically add to the storage metadata information found in any jupyter notebook and in the EXIF of *.png* and *.jpg* images.

Another powerful data ingestion rule allows you to upload metadata to an arbitrary data object by uploading a file template whose name follows a certain convention. Let us suppose you have just uploaded a data object whose iRODS path is `/zoneName/home/username/collection_name/results.dat`, then to add metadata to this object create locally a template file with name `results.dat.metadata` where each line has the format `Attribute Name = Attribute Value`

```
#cat results.dat.metadata
cluster = xmaris
node = maris048
cmd = script input1 input2
date = Aug 2022
```

and upload it to the IL storage server via DevRODS or iCommands, for instance

```
iput results.dat.metadata
/zoneName/home/username/collection_name/results.dat.metadata
```

A suggested (but by no means exhaustive) list of metadata attribute names that you should add to any of your data files is given below

Attribute	Description	Example
title	The title of your dataset	*Average Gas Consumption*
description	Concise Description of what your data Represent	*Average Gas Consumption per month in Year 2022 across Europe*
field	Research Field	*Physics Astronomy Mathematics*
version	Data Versioning Number	*0.0.2*
tags	Keywords to your Data	*Blackhole Gravity Quantum Computers*
dois	If related to published material, Digital object Identifier[3)]	*10.1103/PhysRevD.97.043511*
pi	Name of Principal Investigator	*John B. Smith*
funder	Name organization funding this research	*NWO*
grant	Grant number/ID as issued by the funding organization	*SFT625344*
authors	Authors Names	*M.U.M. Prigles, Jack Smith, Leo Leon*
affiliation	Authors Affiliation	*Leiden University, Oxford*

WebDAV Access

This method lets you access the IL *FAIR* Storage with the comfort of a GUI. Note though that differently than when using iCommands, it is not possible to select a different destination resource than `ilorentzNoReplicaResource` if you use WebDAV.

Web Browser Access

URL	https://access.lorentz.leidenuniv.nl
Username	IL Username
Password	IL Password
Acessibility	IL Subnets Only
Operation Mode	READ ONLY

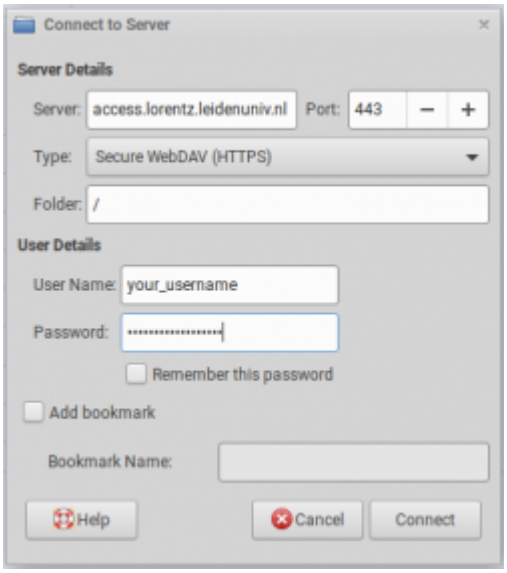
Anonymous access

URL	https://access.lorentz.leidenuniv.nl/anon

File Explorer Access

Acessibility	IL Subnets Only
Operation Mode	READ + WRITE

GNU/Linux OS



Menu → Places → Connect to Server[4)]

Server	access.lorentz.leidenuniv.nl
Port	443
Type	Secure WebDAV (HTTPS)
Folder	/
Username	IL Username
Password	IL Password

Mac OS

Finder → Go → Connect to Server

Server Address	https://access.lorentz.leidenuniv.nl

Specify your IL credentials when prompted and click on Connect.

1)

https://www.go-fair.org/fair-principles/

2)

https://irods.org/

3)

https://www.doi.org/

4)

This may vary depending on the OS. If in trouble search for Connect To Server.