

RESEARCH FOR GRAND CHALLENGES

HELMHOLTZ FEDERATED IT SERVICES

Integrated dataset placement service for scientists

<u>Tim Wetzel</u>, Michael Schuh, Patrick Fuhrmann, Thomas Beermann, Paul Millar, Johannes Reppin, Muhammad Aleem Sarwar

@ Deutsches Elektronen-Synchrotron DESY on behalf of HIFIS

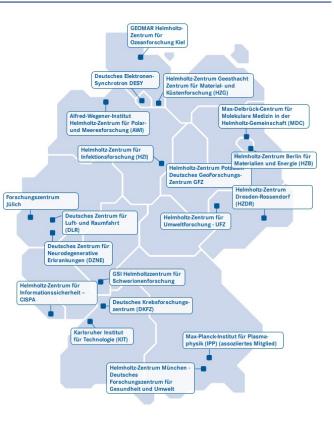
IBERGRID2022, Oct 12th, 2022

www.helmholtz.de

Helmholtz Association



- Founded in 1995 to formalise relationships between research centres
- Members: 19 autonomous research centres in Germany
- Mission: Contributions to grand challenges facing society, science and industry
- Increasing importance of cloud access to common data treasure and -services
- Rapidly growing data exchange from research instruments requires excellent data networking
- Growing connections between Helmholtz, EOSC and FAIR



Helmholtz Association



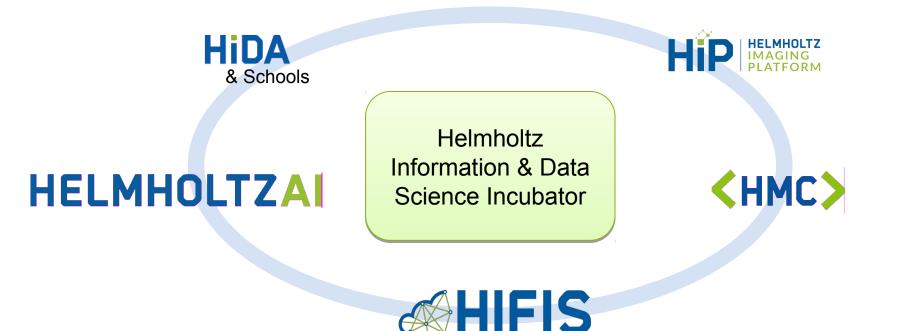
- Founded in 1995 to formalise relationships between research centres
- Members: 19 autonomous research centres in Germany
- Mission: Contributions to grand challenges facing society, science and industry
- Increasing importance of cloud access to common data treasure and -services
- Rapidly growing data exchange from research instruments requires excellent data networking
- Growing connections between Helmholtz, EOSC and FAIR



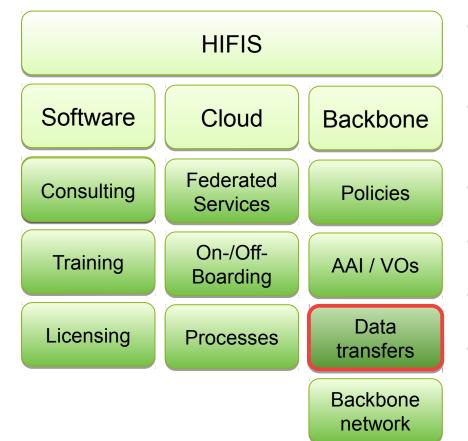
Helmholtz Incubator



Helmholtz aims for joint research & information environment for all Research Fields



HIFIS - Helmholtz Federated IT Services

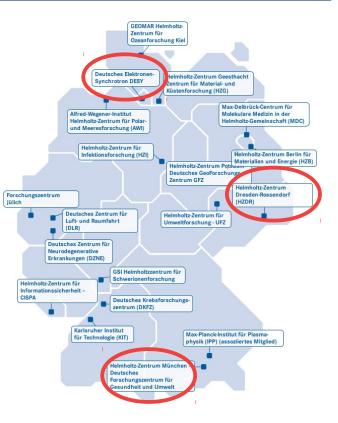


- Joint use of IT services provided by Helmholtz centres
- Exploitation of synergy effects by reducing the service provisioning efforts for each centre
- simplifying collaboration processes for scientists
- Team of ~120 people from 11 Helmholtz centres
- **support and coordination platform** for federation in Helmholtz
- Close connections with other
 incubator platforms and all centres

Why do we need data transfers?



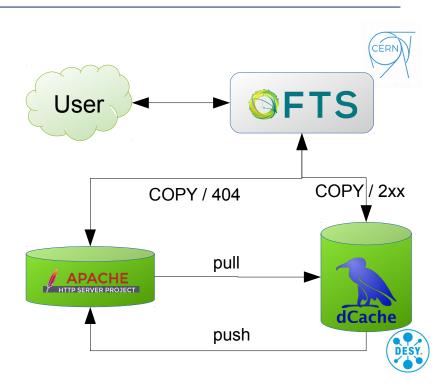
- Large data sets in collaborative research projects
- Data analysis often sensitive to latency
- Data locality is important!
- HI and HelmholtzAI projects use data storage at the computing sites and their own locations
- Collaborating centres distributed over Germany
- Reliable, comfortable and robust transfer endpoints and service needed



HIFIS Transfer Service (HTS)

- Core service in HIFIS backbone
- CERN's FTS3 as backend
- webFTS as comfortable WebUI
- **FTS3-REST** as CLI for scripted transfers
- Next: Ru RUCIO policy driven transfers

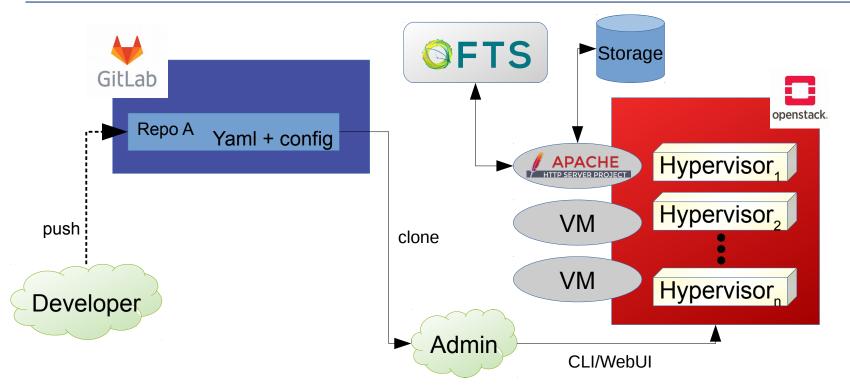
- Apache httpd as passive endpoint
- Lightweight solution for ad-hoc transfers





Rollout on Openstack VM

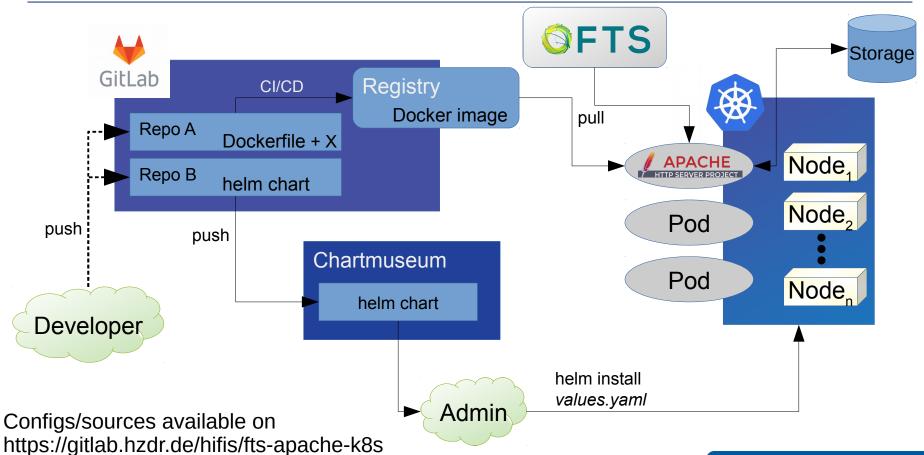




Configs/sources available on https://gitlab.hzdr.de/hifis/hifis-transfer-service

Rollout on Kubernetes via Helm charts

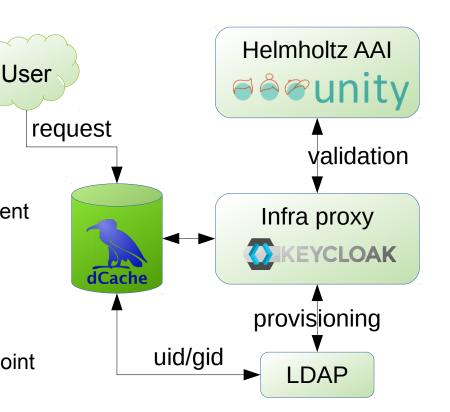




IBERGRID2022

Helmholtz Imaging storage

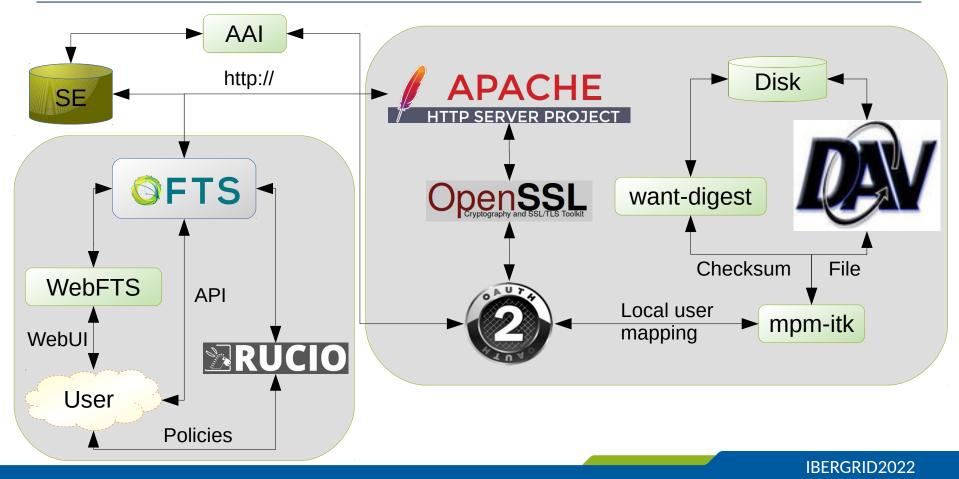
- Central data repository
 - Requested by HI community
 - → dCache @ DESY
 - AuthN/Z via OIDC
 - Automated user provisioning in development
 - Connection between Keycloak, HelmholtzAAI and DESY-LDAP
- Setup shows how to connect a community
- Process to be integrated into Apache endpoint





Detail view on interaction





WebUI WebFTS



s Tim Wetzel, via Helmholtz AAI												
TS Simplifying power												
🕇 Home				🗟 My jobs		Submit a transfer						
Grid SE Grid Storage Element			•	>	Grid SE Grid Storag			•				
				<								
https://os-234-apache-fts.desy.	.de		Load	Overwrite Files	https://dcache-der	mo.desy.de:2443/Users/wetzelti		Load				
(Create Folder X Delete	Rename		Over write thes		Create Folder X De	lete Rename					
				Compare Checksums	_							
Select Al	Il Files None 📿 Refre	esh Show filters		LFC registration		Select All Files None 🗯	Refresh Show filters					
				lfc://				1				
Name	Mode	Date	Size		Name	Mode	Date	Size				
SSozaRulAf.txt	-rwxrwxrwx	14 Jun 12:02	56.1 MB		0							
file.100mb_via_fts	-rwxrwxrwx	17 Dec 20	95.4 MB		file.100mb_via	a_ftsrwxrwxrwx	22 Mar 14:40	95.4 MB				
hewsecret.txt	-rwxrwxrwx	04 Aug 20	22 B		file.png	-rwxrwxrwx	25 Feb 07:32	10.9 kB				
paul-test-1	drwxrwxrwx	23 Jul 20	-		fts-test-small-	fil drwxrwxrwx	07 Oct 20	2 2)				
secret.txt	-rwxrwxrwx	09 Jul 20	22 B		fts-test	drwxrwxrwx	09 Nov 20					
testfile.txt	-rwxrwxrwx	28 Sep 20	5.9 kB		hifis-service-te	est drwxrwxrwx	22 Oct 20	-				
verysecret.txt	-rwxrwxrwx	22 Jul 20	28 B		mod_speling.	c -rwxrwxrwx	30 Sep 20	18.5 kB				
verysecret2.txt	-rwxrwxrwx	16 Sep 20	26 B		random.txt	-rwxrwxrwx	26 Feb 09:15	100.0 MB				
o File (=) O de stad					randomfile.txt	-rwxrwxrwx	02 Oct 20	95.4 MB				
0 File(s) Selected					randomtest.tx	t -rwxrwxrwx	28 Sep 20	249 B				
					secret.txt	-rwxrwxrwx	28 Aug 20	22 B				
					iestbla	drwxrwxrwx	04 Aug 20					
					testfile.txt	-rwxrwxrwx	22 Sep 20	5.9 kB				



IBERGRID2022

CLI bindings:

fts-rest-whoami -s <ENDPOINT> --access-token \$TOKEN fts-rest-transfer-submit -s <ENDPOINT> --access-token \$TOKEN <SRC> <DEST> fts-rest-transfer-status -s <ENDPOINT> --access-token \$TOKEN <JOB ID>

Direct API calls with cURL:

curl -s --capath /etc/grid-security/certificates/ -X POST -H "Content-Type: application/json" -H "Authorization: Bearer \${TOKEN}" --data @submission.json https://fts-public-002.cern.ch:8446/jobs

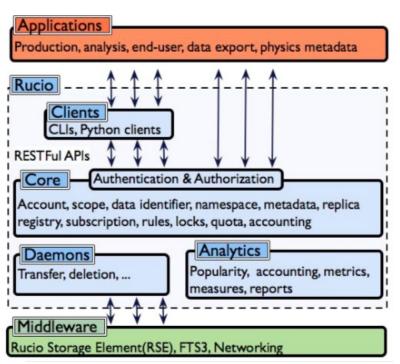
Documentation:https://fts3-docs.web.cern.ch/fts3-docs/fts-rest/docs/api.html

(Rough) Introduction to Rucio

- Orchestration layer on top of FTS3
- Policy- & rule-based organisation of data transfers
- Multi-VO capable
- Token-based AuthN/Z
- Access via WebUI or CLI

Integration tests with JupyterLab-plugin have been shown by colleagues in the context of the ESCAPE datalake







Rucio integration with Jupyter



💭 File Edit View Run Kernel Tabs Settings Help

	2RUCIO	
	ZROCIO	
0	EXPLORE NOTEBOOK	± \$
=	LOFAR_ASTRON_GRANGE:*	a Q
_	Search Everything -	
	SEARCH RESULTS	
	LUFAR_ASTRUN_GRANGE.L12340	•
*	LOFAR_ASTRON_GRANGE:L27183_cal	0
	LOFAR_ASTRON_GRANGE:L31415	0
	LOFAR_ASTRON_GRANGE:L557202_morning	0
	LOFAR_ASTRON_GRANGE:L557202_SB133_uv.MS_helloCERN.tar	0
	LOFAR_ASTRON_GRANGE:mies	0
	COPAR_ASTRON_GRANGE:datalist.dat	24.6KiB
	COFAR_ASTRON_GRANGE:DLaaS_shopping_file1.dat	200MiB
	S Available	Add to Notebook
	LOFAR_ASTRON_GRANGE:DLaaS_shopping_file2.dat	200MiB
	COPAR_ASTRON_GRANGE:DLaaS_shopping_file3.dat	200MiB

List existing datasets in scope \rightarrow make available \rightarrow use file path from mounted storage element in notebook

Rucio integration with Jupyter



💭 File Edit View Run Kernel Tabs Settings Help

	≥ RUCIO			- 1	📃 Unt	titled2	.ipynb			×						
-	ZRUCIU				8 -	+ 9	< 0	Ĉ	►	- 1	C	▶ ▶ (ode	~		
0	EXPLORE NOTEBOOK		± ¢	•		[]:	:									
≔	LOFAR_ASTRON_GRANGE:*															
	Search Everything -															
'n	SEARCH RESULTS															
+	LUFAR_ASTRUN_GRANGE:L12340		U													
*	LOFAR_ASTRON_GRANGE:L27183_cal		0													
	LOFAR_ASTRON_GRANGE:L31415		0													
	LOFAR_ASTRON_GRANGE:L557202_morning		RSE		State	Activity			Interval			· ©			New request	
	LOFAR_ASTRON_GRANGE:L557202_SB133_uv.MS_helloCERN.tar	Show 100 - entries	RSE			User S	lebscriptions		14		days	. 0		Search	e l	
	LOFAR_ASTRON_GRANGE:mies	Name LOFAR_ASTRON_GRANGE:datalist.dat		Account	RSE Expre			Creation	Date 10709:48:49.0	Remainin Lifetime		ate EPLICATING	Locks OK	Locks Replicating	Locks Stuck 0	•
	COFAR_ASTRON_GRANGE:datalist.dat	Name Showing 1 to 1 of 1 entries		Account	RSE Expro	ression		Creation		Remainir Lifetime	10 S	ate	Lecks OK		uting Locks Stuck	
	COFAR_ASTRON_GRANGE:DLaaS_shopping_file1.dat	Select all													Debte rule(s)	6
	Available	download as JSON														
	Contraction LOFAR_ASTRON_GRANGE:DLaaS_shopping_file2.dat		200MiB													
	COPAR_ASTRON_GRANGE:DLaaS_shopping_file3.dat		200MiB													

List existing datasets in scope \rightarrow make available \rightarrow use file path from mounted storage element in notebook

Rucio integration with Jupyter



IBERGRID2022

File Edit View Run Kernel Tabs Settings Help

1

RUCIO		Untitled2.lpynb										
ZRUCIU		B + X 10 🖞 ▶ ■ C → Code ∨ ØReady Python 3 ●										
EXPLORE NOTEBOOK	± ¢	<pre>[1]: print(shopping_list1)</pre>										
LOFAR_ASTRON_GRANGE:*	БQ	/dcache/escape/wp2_rucio_testbed/desy_dcache/LOFAR_ASTRON_GRANGE/f4/ec/DLaaS_shopping_file1.dat										
Search Everything +		<pre>[*]: with open(shopping_list1, 'rb') as fHandle: for line in fHandle:</pre>										
SEARCH RESULTS		<pre>print(line) b'\xec\x8e\x890# g\x95\xc1Ewg\xee\xf40\xcb\x96\xa1\xbchJg\x66f\xe4\xa3D4P\xd0F\r\xfd\x92\xbdo\x7f\xf8\xe7\xb8\xb6\x11\xdb\xcc]L\x</pre>										
LOFAR_ASTRON_GRANGE:L12345	•	b \Xec\Xee\Xee\Xee\Xee\Xee\Xee\Xee\Xee\Xee										
LOFAR_ASTRON_GRANGE:L27183_cal	Ø	b'\xf4\xec\xa6\xb6}Y\xea\x87\x84\n' b'\x1b\x1c\x86\xb6}Y\xea\x87\x84\n' b'\x1b\x1c\x86\t\x94 \x83\xd9\xa2L\xa3\xd9\xd6\xf6\$\xc6(\xca\xff\xa8\xeb\xe9\x83\xd0H\xbf\xb6\x04X\x65H'\x6n(d\x8d\x15C\x80Y										
LOFAR_ASTRON_GRANGE:L31415	0	b \x1b\x1c\xeb\x1xbol\x94 xk55\xb0\xb0\xb0\xb0\xb0\xb0\xb0\xb0\xb0\xb0										
LOFAR_ASTRON_GRANGE:L557202_morning	Θ	2W\x8c(\x80\xee\x81\x95\xd5\xbb\xdb;\x86\x86\xffp\x0c\x16\x06e\xec8-S\xfe\xdc\x23\xbeG\xacC\xc0F\xa2\xbe\x17\xd6\x0										
LOFAR_ASTRON_GRANGE:L557202_SB133_uv.MS_helloCERN.tar	0	5\x1ar \xe5r\x0b\x80_\x1c\xdf\xfe\xe2\x81\xbe\x0c\xc9\x16U}\xecTeiz\xb2\x082\x9e\xe7K\xce7\r\xac\x63t\x04\\\xaa=6 <cz\x98\x97\xc 3\x96}\xbb5\x19g}\x89\xef\xc4\xbc\xaf0\xb4\xbe\x8e0\xe0\xbd6\xdctX\xeej\xe1\xd8\xf9:W\x99\xa0%g\xc1\\\\78W\xa1#y\x92'\xfe=14\xdb</cz\x98\x97\xc 										
LOFAR_ASTRON_GRANGE:mies	0	\xb2zv\xc5ob(l\xeaw\xf9\xed\xbc\x89\xfb;\x80\x65\x06\xa9\xc7\x1f6\xb4\xae\xd7\x8dB\xe6b\xe1\x82\x91\t\xad\x06\xba\xcf,\x10\xf3\x ec\xba\xbd3T9\xa3l\x0e\x8e\xf6\xf3Qq4\x8d)\xf5\x1f\xeeM\xb9\xf5\xf8hfcklK\x01\xfa\xc1\x89\x1c\xa9*,\x1e\xff\x8c\xb4\xc6\xabGa\\x										
COFAR_ASTRON_GRANGE:datalist.dat	24.6KiB	18DT\x01\xf3\xbea\x19tb\x8e\xf2\x9e8,\xb3\xcbaQ\xa9u\n" b'\x10\xcb\x01\xe9\x04\xc9\xaa\xa6\x13#\xb8y\xe9;6\x96HVJL\x8e>\xe6\xb5F\x8c\xab\xdf\xc1\xc6\xea\x87C\xd6\x8c\x97\xc4[962\xdfuv7f										
LOFAR_ASTRON_GRANGE:DLaaS_shopping_file1.dat	200MIB	<pre>qoBE\xa4ke\xx3\xe57E\xa7i\xf2+y\x1d1\x9eL\r\x85\xc36v\xb7o\x94\xb8\xef\x13\x91\xb7)\x84\x01\x80Q\xea\x7f\xe9\x13?y[.\x14\x80\x0c \x101h\xcfU\xc1\x64\x81\x80\x62\x37y[.\x14\x80\x0c</pre>										
Available		<pre>b'\xf8\x81\x81\x81\x81\x84\x84\x84\x84\x8d\x8d\x82\x15<,\x8d\x81\x85\x81\x81\x11\x12\x15<,\x81\x11\x12\x12\x15<,\x81\x11\x12\x12\x11\x12\x12\x11\x12\x12\x1</pre>										
LOFAR_ASTRON_GRANGE:DLaaS_shopping_file2.dat	200MiB	c7?\xa2W\x87h% \x13\x07\\\x88\x18\xca\xeb \n'										
Available	Add to Notebook	b"_\x881\xa2\x10?\x88\xd0\x9e\x02a\x867\xf2D]\xde\x13\x89\x80\xf10\x1e\x8d\xc0\xd1\x13\xf3\x14\xe5\xbe3\xcb\xa0\xc1\xb2\x ba\xf4c\xf4c\xad\x81\xb85\x1d\x81\yg'\x13\xe7\xbf\x9fHH\xb44\xf4\xac-\x14g\x1b\x8f4\x86\xc0\xd3\xcb\xf9\xa0e\x96 U_3\xf4c\xf4c\xad\x81\xb65\x61\x80\x76\x60\xf8\x61\x61\x82\x										
LOFAR_ASTRON_GRANGE:DLaaS_shopping_file3.dat	200MiB	H;3\xb3R\xaf\xla\xbf%\x87\x07\x96e3\xc4\xb8\xf5\xfc\x8d=`\x82x\xf7\n" b'\xf9v\xab\x8f\x9b!\xfe\xbe\xc3[\x9d\x0e}\x14]\xb6x\xf5\xf7\xb9\xdb]\xe9Ki<\xdf\x93So\x16,\xc4\xa9\x11\x80\xa1N\xe9\xee5\xa3cr\x										
S Available	Add to Notebook	cc\xbb\x1f\x1cF\xebixf%\x90z6\t\xcd:`\x07w\xc3\xec\xa2\xe6\xb5\x01q\x00\xae<\x9c\xa2fE.\x90fc\xc3'\xab\xfbMw\xee\xe2\xe1z\xc5\x88 \xfa\x07\xc5\x8f\xe8\x17\x96\x90H\xda\xd2\x83\x1av\x85\x04u\xc5\xcd\xeb\xba\xf4\xa5\x0f\xdc\xd01\xbaqM,\x1a\x90										
LOFAR_ASTRON_GRANGE:DLaaS_shopping_file4.dat	200MIB	05\$?\x02m\xf8\xab\x92\x1f\xa6r\xa6&T0\x95\xfb\xf0\xddg\x14o{4\xb4\$L\xe2\xcb _a\x80\x82\xc7\x8d\x8e\xaa\xebR\x13\xcb\x86\xef\xbd M.\xde\x85u\xae\xc5\x95\xb5,X\x87NX\xdf\xdc\xf2\xbd\xfa\x10\xf9\x02\t\x90\xb3>\x08\xcd\x80\x9c\xd1T\x06C\xd3\xaa										
LOFAR_ASTRON_GRANGE:DLaaS_shopping_file5.dat	200MiB	0+V3W\xd1\xce\x87#\x1b\x89\xd1\xae\x3#\xd1+\x94a\x02F\x64\xb)x51\x65\x850\xfdY\xa7\xceH\x89\x42\x65\xb8\xd1\xd2\x7by\x62\x7b8\xd1\xce\x79\xd Q-\x0b\x02\x1a^\x7b2\x7b3\x61\x7b3\x7b3\x7b3\x7b3\x7b3\x7b3\x7b3\x7b3										

List existing datasets in scope \rightarrow make available \rightarrow use file path from mounted storage element in notebook

Summary & Outlook



- HIFIS Transfer Service
 - Starting to be established in Helmholtz
 - Endpoint installation to be simplified
 - Outreach is increasing
- RUCIO
 - Testing instance set up
 - AuthN/Z token integration next
 - Endpoint integration afterwards





RESEARCH FOR GRAND CHALLENGES



Thank you! Questions?

www.helmholtz.de