

ESA–VOSpace Update IVOA Interop 2017

Sara Nieto

ESAC Science Data Centre (ESDC)
05/16/2017

Issue/Revision: 1.0

Reference: Interop 17

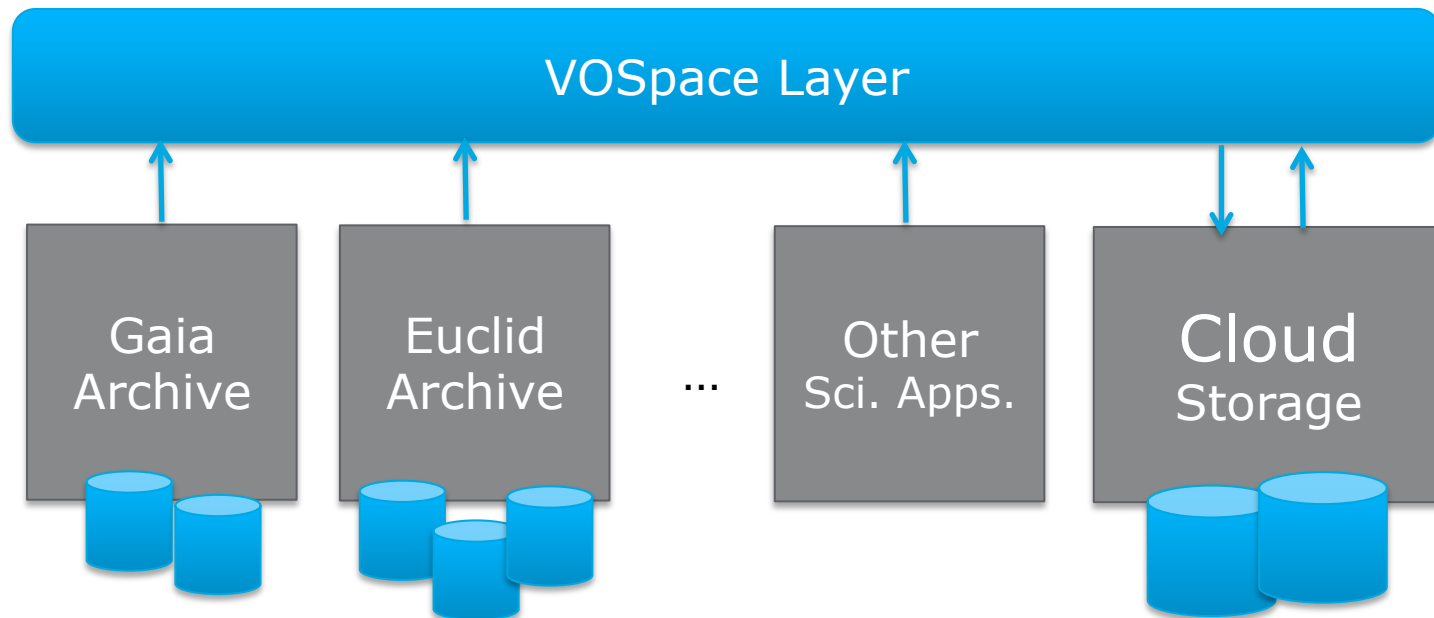
Status: Issued

ESA UNCLASSIFIED - Releasable to the Public

- 1. Euclid case: Registration of Link Nodes in VOspace**
2. Search capabilities
3. Operations that become compliant with spec v2.1
 - a. Sharing links
 - b. Transfer parameters
4. What's next?

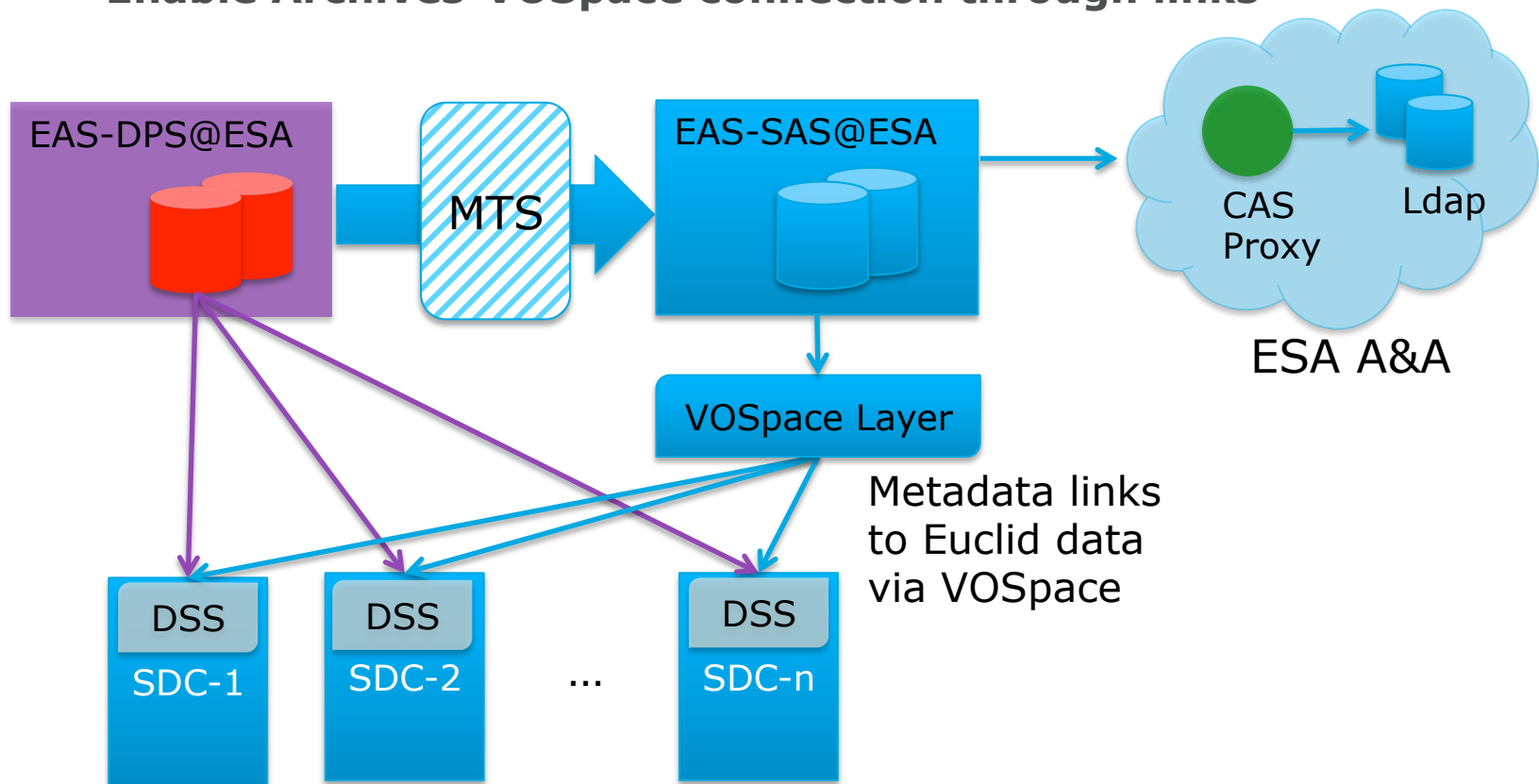
VOSpace Links to data sets

- **VOSpace links pointing to remote data sets** located at archive repositories, external Data Centers, etc. This functionality avoids transferring data sets from original repositories to VOSpace and allows being accessed/downloaded through VOSpace...



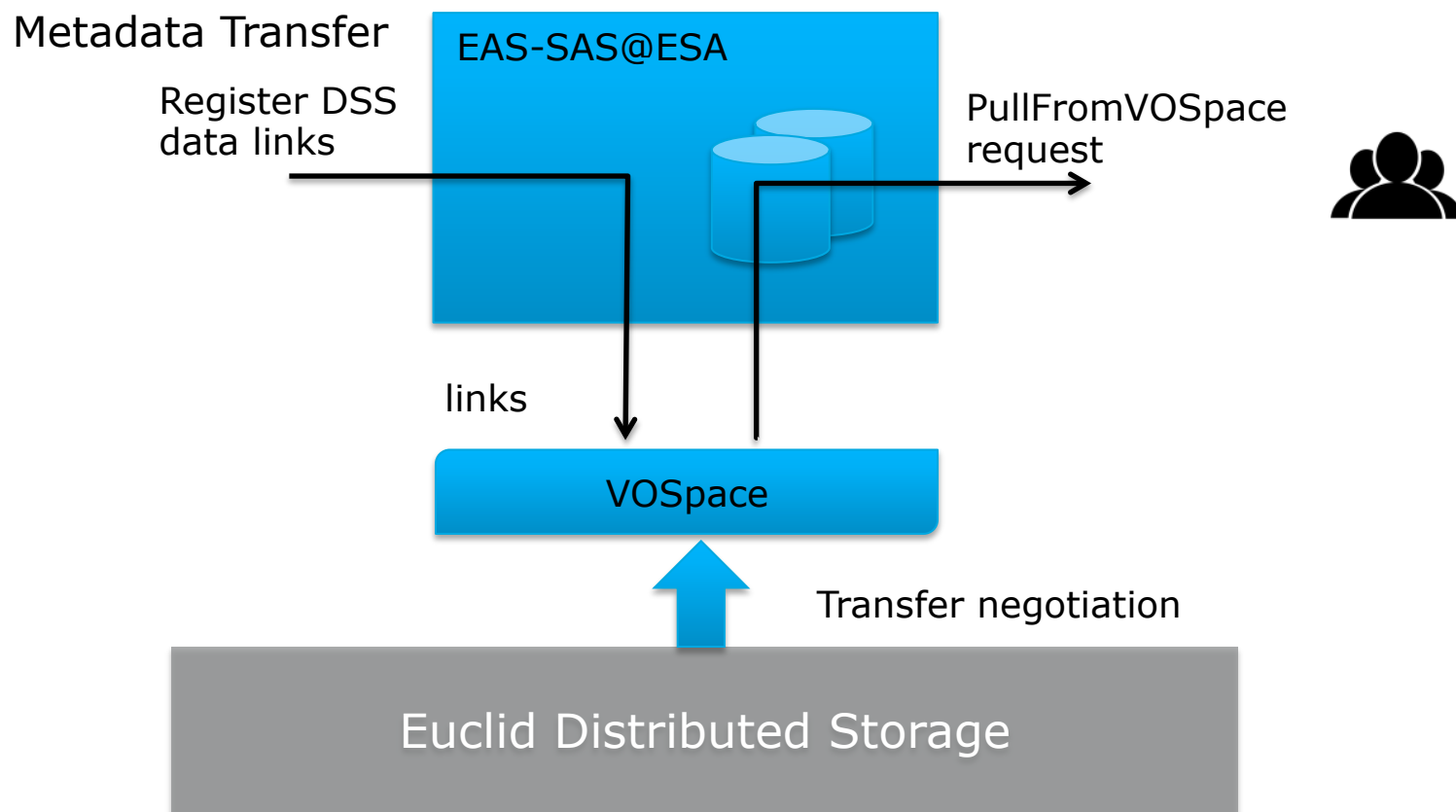
Euclid – VOSpace Use Case

- Use Case: Use VOSpace as an entry point to Euclid data (via SAS)
 - Register VOSpace links (URI) pointing to nodes at SDCs
 - **Enable Archives-VOSpace connection through links**



Euclid – VOSpace Use Case

- Registration of links in VOSpace and downloading process request



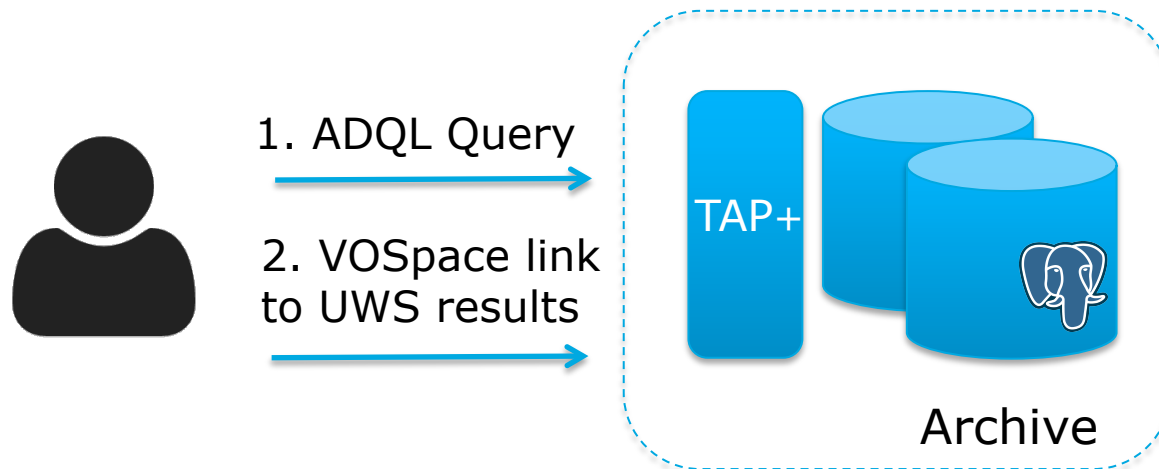
LinkNodes From EAS to VOSpace Workflow

1. Link registration process from EAS -> VOSpace

a. Create link in VOSpace user account

- Link URI include UWS result with <jobId>
.../vospace/<uname>/<projectName>/<jobId>
- Target: endpoint to query result (UWS result)
.../server/tap/async/1493676605966L/results/result

2. Link resolution process from VOSpace -> EAS



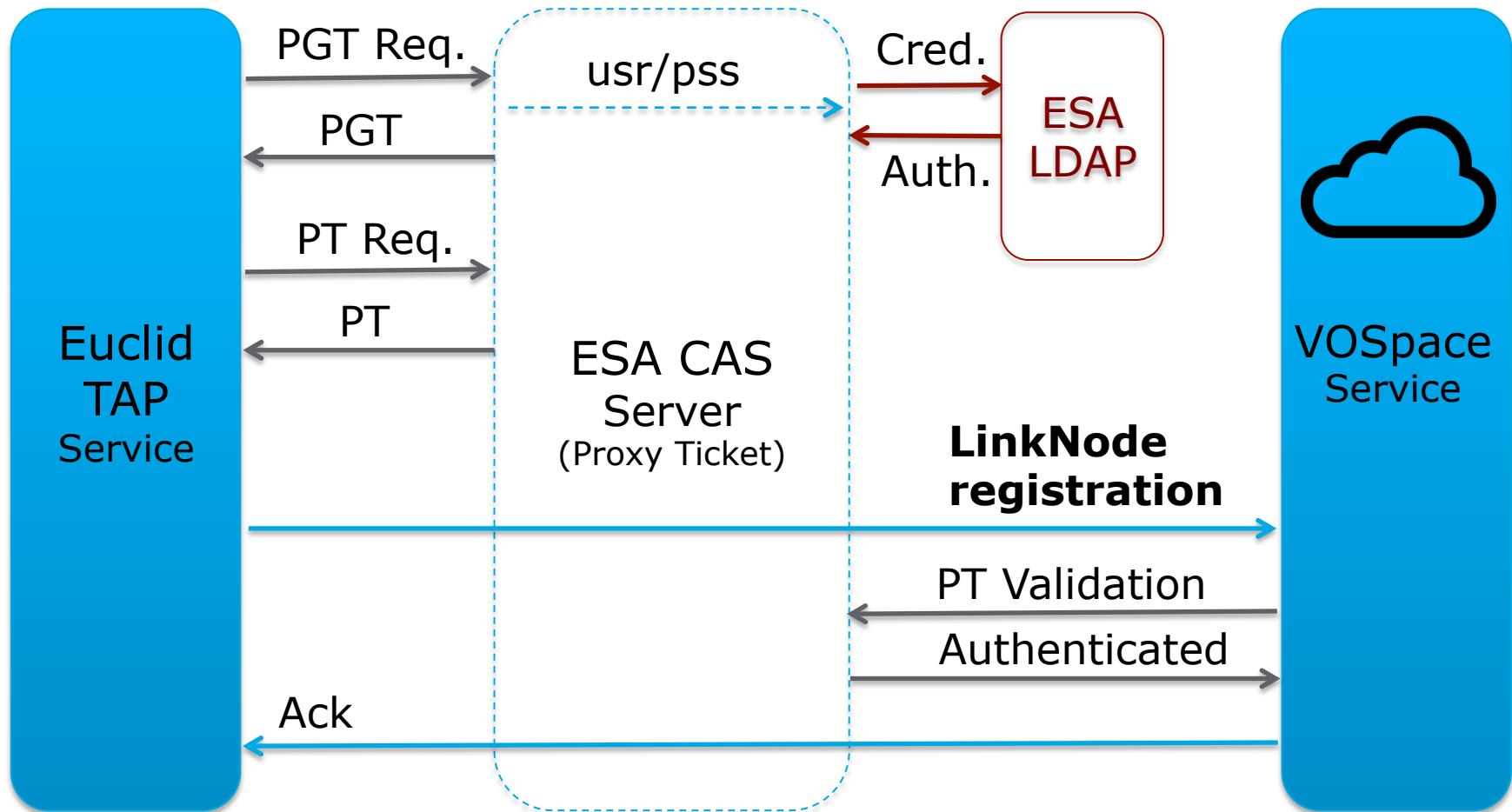
LinkNodes From EAS to VOspace

General considerations



- Linking query results **from ESDC Archives into VOspace**
 - How do we resolve URIs into URLs?
 - Identifiers resolution process
 - <uname>/**euclid**/*<nodePath>**
 - <uname>/**archive**/*<nodePath>**
 - What do we need for data retrieval?
 - Identifier resolution from URI -> URL resource
 - Get URI of the service e.g. Euclid archive, SDC...
 - Get URL of the data set from the service e.g. TAP+
 - Endpoint service, protocol, security method
 - Persistence?
 - Availability of linked results is under user responsibility?
 - Persistent queries **on archive data**, DOIs?
 - Non persistent queries **on temporary data results**

LinkNodes From EAS to VOSpace Authorization schema



LinkNodes From EAS to VOspace Links to UWS results



Euclid Science Archive System



Access Euclid Scientific Data Model v0.3 and the following catalogues: XMM DR#5, KiDS and Euclid SC#2.



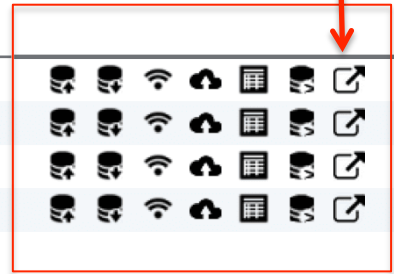
1

Output format: csv

Submit Clear

Link registration

Status	Job	Creation date	Num. rows	Size	
✓	<input type="checkbox"/> 1493636554652L	01-May-2017, 13:02:34	23707	1 MB	
✓	<input type="checkbox"/> 1474380545208L	20-Sep-2016, 16:09:05	49	64 KB	
✓	<input type="checkbox"/> 1474380409883L	20-Sep-2016, 16:06:49	9	10 KB	
✓	<input type="checkbox"/> 1473152650693L	06-Sep-2016, 11:04:10	9	10 KB	



LinkNodes From EAS to VOspace Request a new link (to query results)




Euclid Science Archive System



Access Euclid Scientific Data Model v0.3 and the following catalogues: XMM DR#5, KiDS and Euclid SC#2.

Link URI->.../vospace/<uname>/euclid/**1493676605966L**
Target URL -> .../euclid/tap/async/**1493676605966L**/results/result

 EAS-SAS link to VOspace

Job 1493636554652L
Destination folder/snieto/EAS/
Link name
 Overwrite link

Cancel Send

Output format: csv

Status	Job			Size	
✓	<input type="checkbox"/> 1493636554652L	01-May-2017, 13:02:34	23707	1 MB	
✓	<input type="checkbox"/> 1474380545208L	20-Sep-2016, 16:09:05	49	64 KB	
✓	<input type="checkbox"/> 1474380409883L	20-Sep-2016, 16:06:49	9	10 KB	
✓	<input type="checkbox"/> 1473152650693L	06-Sep-2016, 11:04:10	9	10 KB	

LinkNodes From EAS to VOspace Links in VOspace



VOspace Browser



44.7 GB (44%) of 100GB [snieto](#) [Help](#) [Sign Out](#)



▼ My VOspace

- B
- ContainerTest2
- EAS**
- Euclid
- GACS
- GAVIP

▼ Shared with me

- MMI
- OU_presentations_2015
- Bern2017
- Challenge6
- hola

▶ My VOspace ▶ EAS

	NAME	SHARED WITH	OWNER	SIZE	LAST MODIFIED
<input type="checkbox"/>	1470134484787D.		Sara Nieto	48 KB	03/15/2017 08:48:09
<input type="checkbox"/>	1473091040483D.		Sara Nieto	14.7 KB	09/05/2016 17:57:38
<input checked="" type="checkbox"/>	1473152650693L.		Sara Nieto	0 Bytes	05/02/2017 00:04:11
<input type="checkbox"/>	1473341959598D.		Sara Nieto	14.7 KB	09/09/2016 16:51:33
<input checked="" type="checkbox"/>	1474380409883L.		Sara Nieto	0 Bytes	05/02/2017 00:04:18
<input checked="" type="checkbox"/>	1474380545208L.		Sara Nieto	0 Bytes	05/01/2017 23:41:09
<input type="checkbox"/>	1474531757673D.		Sara Nieto	286 KB	09/22/2016 10:09:46
<input type="checkbox"/>	1474987516601D.		Sara Nieto	3.1 MB	03/15/2017 08:47:29
<input type="checkbox"/>	1489410621065D.		Sara Nieto	48.2 KB	03/15/2017 10:15:05
<input type="checkbox"/>	1489569046133D.		Sara Nieto	19.9 MB	04/14/2017 11:29:50
<input checked="" type="checkbox"/>	1493636554652L.		Sara Nieto	0 Bytes	05/01/2017 23:38:22

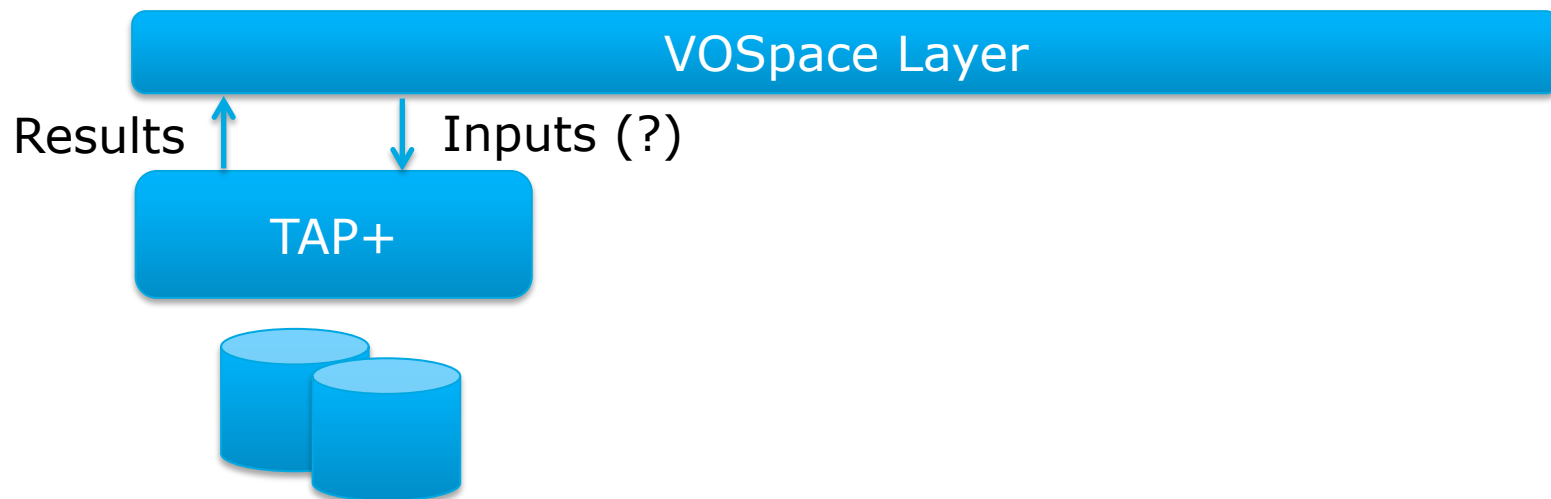
VOSpace Link Nodes

Open points

- **Use VOSpace as data input for archives**

In case of archives, to be able to use VOSpace data (in table format, e.g. VO table) to do x-match with data located in the archive. The same concept can be applied to visualization interfaces that use HiPS maps as inputs.

- In this context, could we use VOSpace links to...
 - ...reference external TAP results? for x-match?



1. Current on-going work:

- a. **Search capabilities: Apache Solr to index VOSpace content**
- b. Operations compliant with v2.1
 - Sharing links
 - Transfer parameters
- c. Improvements on GUI: Sharing features, etc.

- Users demand search capabilities in VOSpace
 - At metadata level -> Solr for metadata **indexing**
 - At data content level (TBD) -> Tika for content **parsing** (content analysis toolkit). Long list of formats supported: [View list](#)
 - Scientific: fits, ...
 - XML and derived
 - Text and PDF formats



Indexing VOSpace metadata Solr – Simple example



Use original UI



- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- vospace
- Overview
- Analysis
- Dataimport
- Documents
- Files
- Ping
- Plugins / Stats
- Query
- Replication
- Schema
- Segments info

/dataimport

Command: full-import

Verbose
 Clean
 Commit
 Optimize
 Debug

Entity: []

Start, Rows: 0 to 4444

Custom Parameters: key1=val1&key2=val2

Auto-Refresh Status

Last Update: 00:37:11

Indexing completed. Added/Updated: 3862 documents. Deleted 0 documents. (Duration: 01s)
Requests: 1 1/s, Fetched: 3,862 3,862/s, Skipped: 0 , Processed: 3,862 3,862/s
Started: less than a minute ago

Raw Status-Output

Configuration Debug-Mode Reload

```
<?xml version="1.0" encoding="UTF-8"?>
<dataConfig>
  <dataSource type="JdbcDataSource" driver="org.postgresql.Driver" url="jdbc:postgresql://localhost:5432/postgres" />
  <document name="nodeDocument">
    <entity name="nodeDocument" query="SELECT * from vos.node">
      <field column="node_oid" name="id" type="int"/>
      <field column="parent_container_id" name="parent" type="int"/>
      <field column="name" name="name" type="string" indexed="true" stored="true" required="true"/>
    </entity>
  </document>
</dataConfig>
```

- Documentation
- Issue Tracker
- IRC Channel
- Community forum
- Solr Query Syntax



Dashboard

Logging

Core Admin

Java Properties

Thread Dump

vospace

Overview

Analysis

Dataimport

Documents

Files

Ping

Plugins / Stats

Query

Replication

Schema

Segments info

Request-Handler (qt)

/select

— common —

q

name:*test*

fq

sort

start, rows

0 10

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

indent

debugQuery

dismax

edismax

hl

facet

spatial

spellcheck

Execute Query

http://localhost:8984/solr/vospace/select?indent=on&q=name:*test*&wt=json

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 29,
    "params": {
      "q": "name:*test*",
      "indent": "on",
      "wt": "json",
      "_": "1493228301621"
    }
  },
  "response": {
    "numFound": 89,
    "start": 0,
    "docs": [
      {
        "id": "47450",
        "name": ["testing11"],
        "_version_": 1565764482762801153
      },
      {
        "id": "52973",
        "name": ["Test5"],
        "_version_": 1565764482775384070
      },
      {
        "id": "52843",
        "name": ["Test1"],
        "_version_": 1565764482776432640
      },
      {
        "id": "49315",
        "name": ["test10062016.vot"],
        "_version_": 1565764482785869824
      },
      {
        "id": "52883",
        "name": ["Test3"],
        "_version_": 1565764482805792769
      },
      {
        "id": "125",
        "name": ["testFile"],
        "_version_": 1565764482809987086
      },
      {
        "id": "132",
        "name": ["Testing"]
      }
    ]
  }
}
```

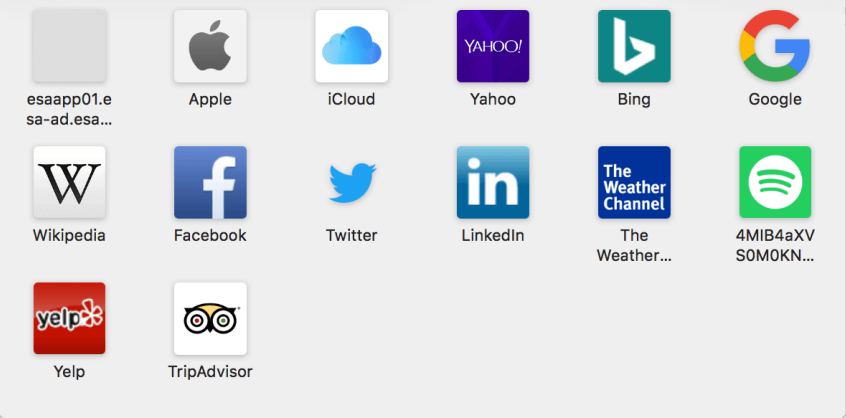
localhost:8984/solr/vospace/select?indent=on&q=name:*test*&wt=json

Solr Admin

```

{
  "responseHeader": {
    "status": 0,
    "QTime": 4,
    "params": {
      "q": "name:*test*",
      "indent": "on",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 12,
    "start": 0,
    "docs": [
      {
        "id": "47450",
        "name": ["testir"],
        "_version_": 156
      },
      {
        "id": "52973",
        "name": ["Test5"],
        "_version_": 156
      },
      {
        "id": "52843",
        "name": ["Test1"],
        "_version_": 1565764482776432640,
      },
      {
        "id": "49315",
        "name": ["test10062016.vot"],
        "_version_": 1565764482785869824,
      },
      {
        "id": "52883",
        "name": ["Test3"],
        "_version_": 1565764482805792769,
      },
      {
        "id": "125",
        "name": ["testFile"],
        "_version_": 1565764482809987086,
      },
      {
        "id": "132",
        "name": ["Testing"],
        "_version_": 1565764482811035653,
      },
      {
        "id": "152",
        "name": ["test"],
        "_version_": 1565764482812084227,
      },
      {
        "id": "261",
        "name": ["test"],
        "_version_": 1565764482812084236,
      },
      {
        "id": "262",
        "name": ["test3"],
        "_version_": 1565764482812084237,
      }
    ]
  }
}

```



After VOSpace operations:

- Update Solr index
- Performance (TBD)

URL Parameter Transfers – VO Compliant v2.1



1. </vospace-local/service/synctrans?TARGET=vos://esavo!vospace/snieto/1.png&DIRECTION=pullFromVoSpace&PROTOCOL=httpget&REQUEST=redirect>

Internal redirection to...

2. </vospace-local/sh/9167d25ab65aa3ad4a1cb093cb9a486ee2ead8?dl=1>

Same case for Sharing links

- In VOSpace v1.5:
 - Euclid – VOSpace interoperability (based on Link Nodes)
 - Search capabilities based on Apache Lucene
 - Bring Public Share mechanism compliant with specification v2.1
- Future work:
 - Java, Python VOSpace CLI clients
 - PoC of Federated Authentication for VOSpace (TBD)
 - Technical VOSpace forum with regular meetings?
 - Transfer protocol between VOSpace and external data providers
 - Interoperability testing

The background of the central slide is a dark blue field with glowing white circuit traces. Several translucent blue folder icons are scattered across the scene, and a blue laptop icon is visible in the lower right quadrant. The text 'Thanks' and the URL 'https://vospace.esac.esa.int' are centered in white.

Thanks
<https://vospace.esac.esa.int>