

OpenSource GeoSpatial Catalogue Platform-as-a-Service

Salvatore Pinto
Cloud Technologist (EGI.eu)

- **EGI & EGI Federated Cloud**
- Data Access and Dissemination PaaS service
- OpenSource GeoSpatial Catalogue

What is EGI

- **E**uropean
 - Over 35 countries
- **G**rid
 - Secure sharing
- **I**nfrastructure
 - Computers
 - Cloud
 - Data
 - Applications

Enabling more than 230 virtual organisations (>20K users) to perform digital research by pooling together their distributed resources

Coordinated by EGI.eu

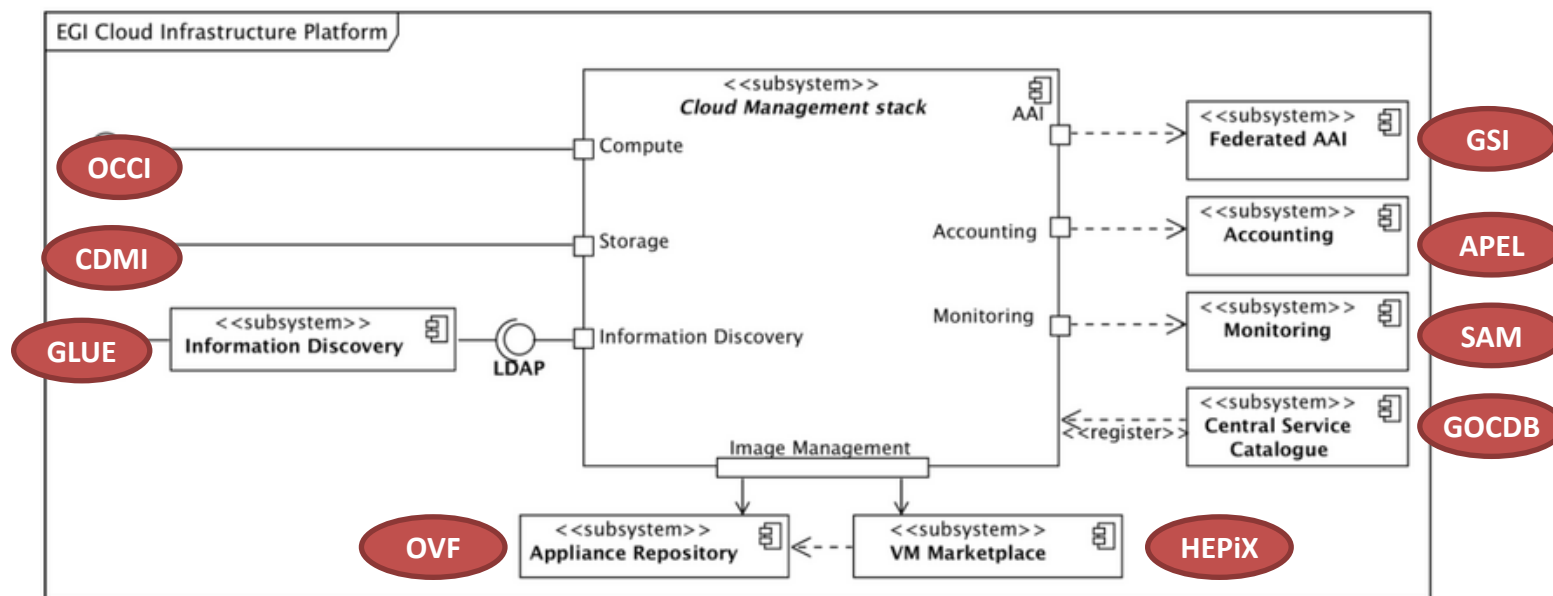
- Legal organisation based in Amsterdam



What is the EGI Federated Cloud

The EGI Federated Cloud is federation of institutional private Clouds, offering Cloud Infrastructure as a Service to scientists in Europe and worldwide.

- **Standards and validation:** federation is based on common **Open-Standards** – OCCI, CDMI, OVF, GLUE, etc...
- **Heterogeneous implementation:** no mandate on the cloud technology, the only condition is to expose the chosen interfaces and services.



Cyfronet

FZJ

OeRC

EGI.eu

CESNET

GWDG

INFN-Bari

CNRS

KTH

FCTSG

CETA

IGI

RADICAL

STFC

BSC

Imperial

LMU

IPHC

IISAS

SixSq

100% IT

CSC

IFAE

SRCE

IN2P3

Masaryk

INFN-CNAF

CESGA

SARA

IFCA

SZTAKI

GRNET

DANTE

Members

- ~70 individuals
- ~35 institutions
- >13 countries

Technologies

- OpenNebula.
- StratusLab.
- OpenStack.
- Synnefo.
- WNoDeS.
- PERUN
- SlipStream

Stakeholders

- 23 Resource Providers
 - 10 production
- 10 Technology Providers
- 8 User Communities
- 4 Liaisons

- EGI & EGI Federated Cloud
- **Data Access and Dissemination PaaS service**
- OpenSource GeoSpatial Catalogue

The problem:

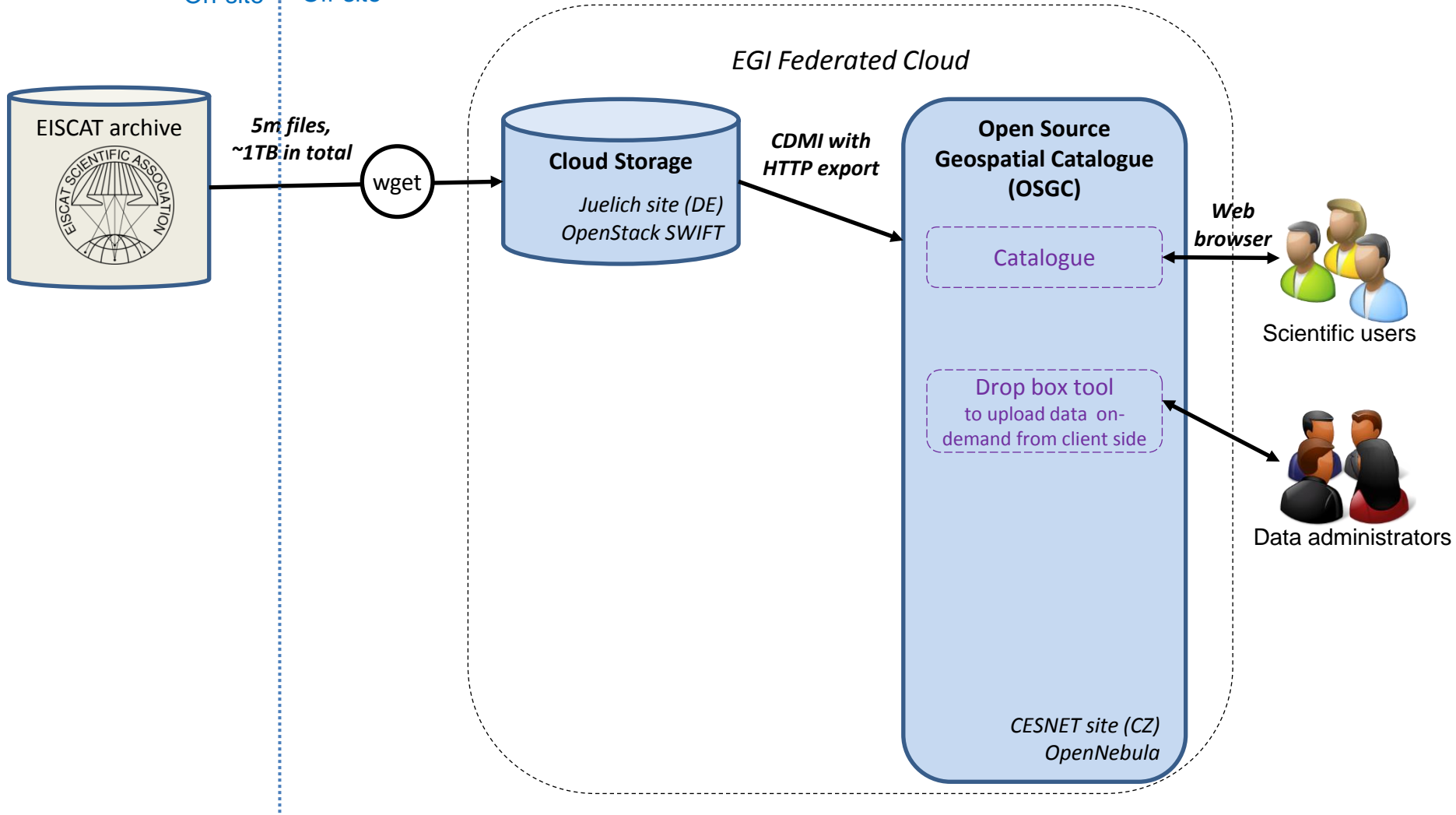
- Provide community with a ready-to-use service for collecting, cataloguing, searching and disseminating of huge amount of scientific data.
- Generic and customizable service (need to adapt to different scientific user communities)

The solution:

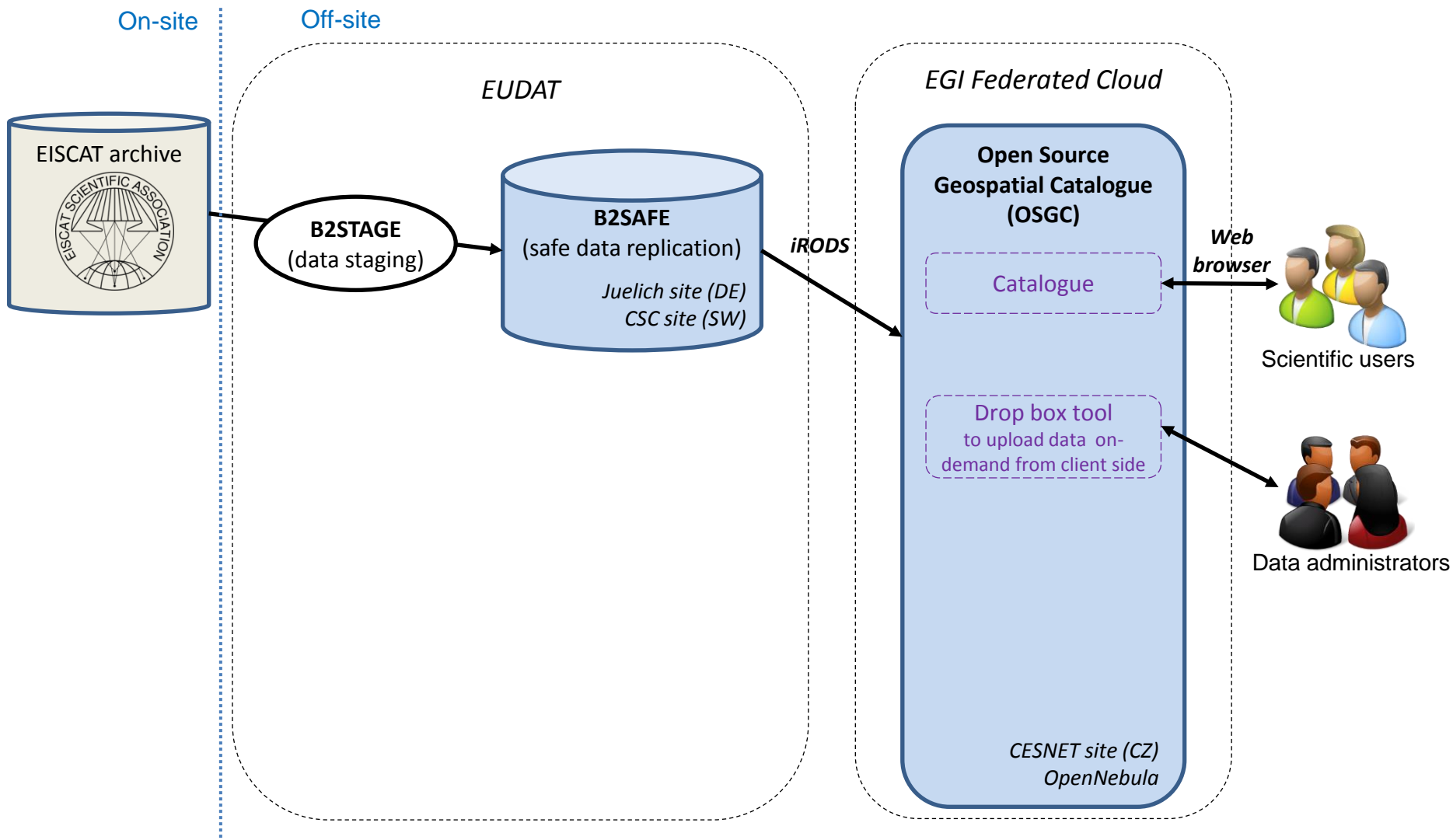
- Platform-as-a-Service built on top of Cloud IaaS and STaaS
- Service integrated with Cloud Storage, Long-term-data storage (ex. EUDAT safe replication) or local private storage.
- Based on the ENVRI project tools (OSGC catalogue)
- Pilot use case in collaboration with EISCAT 3D

ENVRI-EISCAT 3D Pilot (concept with EGI solutions)

On-site Off-site

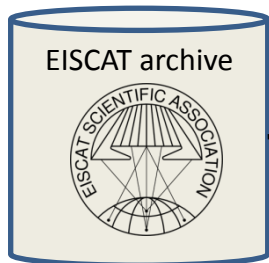


ENVRI-EISCAT 3D Pilot (integrating with EUDAT storage)



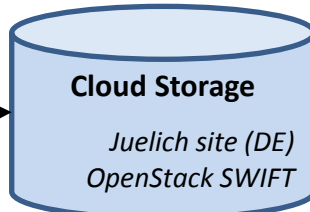
ENVRI-EISCAT 3D Pilot (future work)

On-site Off-site

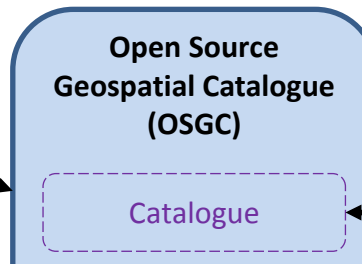


5m files,
~1TB in total

wget



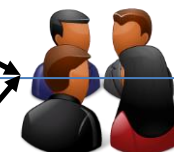
CDMI with
HTTP export



Web
browser



Scientific users



Data administrators

Further metadata have to be extracted from the data

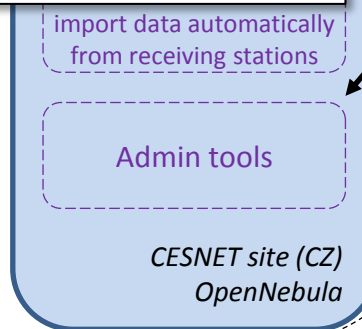
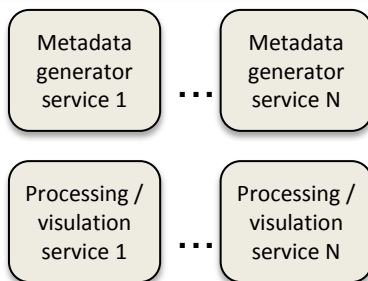
- E.g. Number of spikes, type of spikes

Complex search based on data patterns

- E.g. Particular shapes of the data FFT, etc...

Processing services (for metadata extraction and visualization)

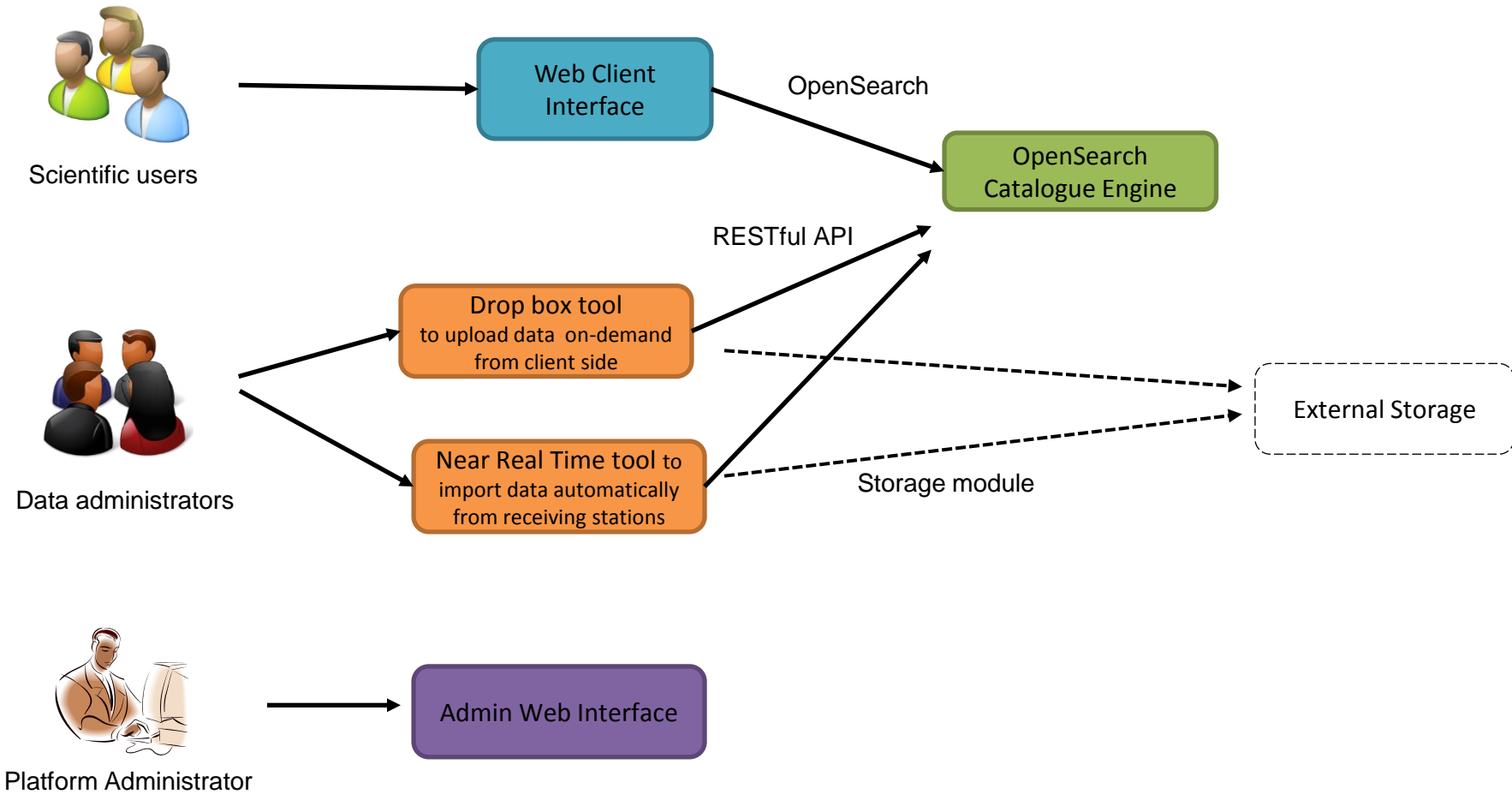
- Integrated analysis tools on the data (eg. With GUI SDAP) exposed to OSGC as services



- EGI & EGI Federated Cloud
- Data Access and Dissemination PaaS service
- **OpenSource GeoSpatial Catalogue**

- Based on OpenSearch and OGC OpenSearch GeoSpatial Extension
- Released under GPLv3 on SourceForge (<https://sourceforge.net/projects/osgcat/>)
- Built in PHP/PostgreSQL

- Timeline:
 - September 2013 : Defined requirements and started development
 - December 2013 : First Beta
 - January 2014 : Release Candidate
 - February 2014 : Start PoC with EISCAT 3D
 - June 2014 : Final release



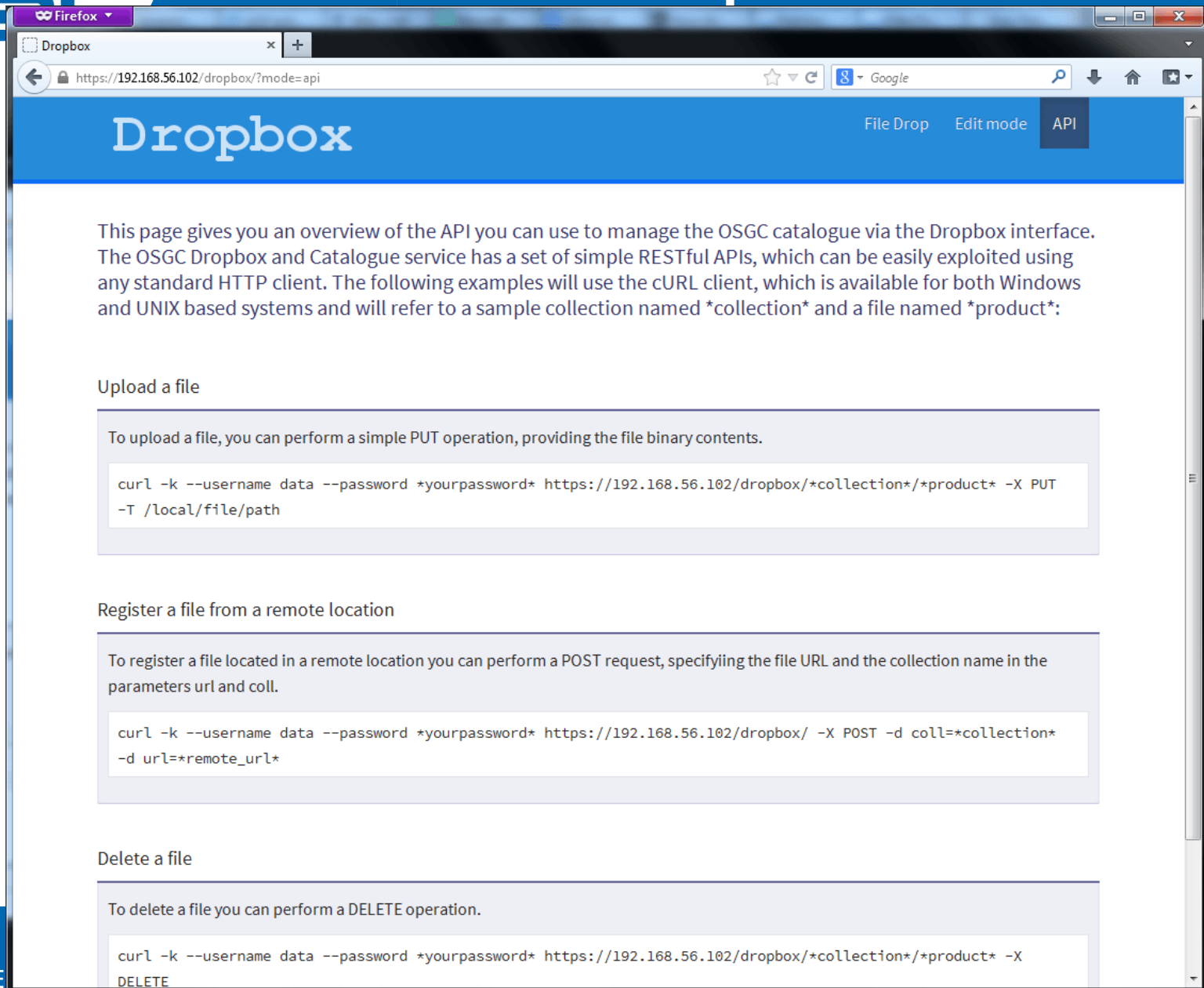
```

Firefox
http://192.168.56.101/c/SAMPLE/rdf?stop=2010-01-12T13:20:02.000Z&bbox=10,20,40,50
192.168.56.101/c/SAMPLE/rdf?stop=2010-01-12T13:20:02.000Z&bbox=10,20,40,50
This XML file does not appear to have any style information associated with it. The document tree is shown below.
- <rdf:RDF>
- <rdf:Description rdf:about="http://192.168.56.101/c/SAMPLE/rdf?stop=2010-01-12T13:20:02.000Z&bbox=10,20,40,50">
  <os:totalResults>0</os:totalResults>
  <atom:link atom:rel="self" atom:type="application/rdf+xml" atom:href="http://192.168.56.101/c/SAMPLE/rdf?stop=2010-01-12T13:20:02.000Z&bbox=10,20,40,50"/>
  <os:startIndex>0</os:startIndex>
  <os:itemsPerPage>400</os:itemsPerPage>
  <atom:link atom:rel="first" atom:type="application/rdf+xml" atom:href="http://192.168.56.101/c/SAMPLE/rdf?stop=2010-01-12T13:20:02.000Z&bbox=10%2C20%2C40%2C50&startPage=0"/>
  <atom:link atom:rel="last" atom:type="application/rdf+xml" atom:href="http://192.168.56.101/c/SAMPLE/rdf?stop=2010-01-12T13:20:02.000Z&bbox=10%2C20%2C40%2C50&startPage=0"/>
  <atom:link atom:rel="search" atom:type="application/opensearchdescription+xml" atom:href="http://192.168.56.101/c/SAMPLE/description"/>
  <dc:creator>OSGC Catalogue 1.0</dc:creator>
  <dc:date>2013-12-02T03:22:49.000Z</dc:date>
  <dc:SizeOrDuration>0.001 sec</dc:SizeOrDuration>
  <os:Query os:role="request" stop="2010-01-12T13:20:02.000Z" bbox="10,20,40,50"/>
</rdf:Description>
- <dclite4g:Series rdf:about="http://192.168.56.101/c/SAMPLE/SAMPLE/rdf">
  <dc:identifier>SAMPLE</dc:identifier>
  <atom:link atom:rel="search" atom:type="application/opensearchdescription+xml" atom:title="Search the SAMPLE Series" atom:href="http://192.168.56.101/c/SAMPLE/SAMPLE/description"/>
  <dc:title>Sample dataset</dc:title>
  <dc:subject>MYSAT</dc:subject>
  <eop:sensor>MYSENS</eop:sensor>
  <ical:dtstart>2013-10-25T06:10:50.000Z</ical:dtstart>
  <ical:dtend>2013-10-25T07:10:02.000Z</ical:dtend>
  <dct:spatial/>
  <dc:format>HDF-EOS</dc:format>
  <dct:created>2013-11-14T11:11:55.419Z</dct:created>
  <dct:modified>2013-11-17T01:11:24.193Z</dct:modified>
  <dct:extent>16</dct:extent>
  <dc:rights>Open</dc:rights>
</dclite4g:Series>
</rdf:RDF>

```

Features:

- **OGC 10-32r3 Compliant**
- **Custom product and series metadata definition**
(custom metadata schema per collection)
- **Custom output format** (Search output can be presented in custom formats, ex. RDF, Atom, KML)
- **Custom Search Fields** (OpenSearch search fields and rules are customizable)
- **REST Interface** (RESTful APIs, with GET, PUT, DELETE support)



Dropbox

File Drop Edit mode API

This page gives you an overview of the API you can use to manage the OSGC catalogue via the Dropbox interface. The OSGC Dropbox and Catalogue service has a set of simple RESTful APIs, which can be easily exploited using any standard HTTP client. The following examples will use the cURL client, which is available for both Windows and UNIX based systems and will refer to a sample collection named **collection** and a file named **product**:

Upload a file

To upload a file, you can perform a simple PUT operation, providing the file binary contents.

```
curl -k --username data --password *yourpassword* https://192.168.56.102/dropbox/*collection*/*product* -X PUT -T /local/file/path
```

Register a file from a remote location

To register a file located in a remote location you can perform a POST request, specifying the file URL and the collection name in the parameters url and coll.

```
curl -k --username data --password *yourpassword* https://192.168.56.102/dropbox/ -X POST -d coll=*collection* -d url=*remote_url*
```

Delete a file

To delete a file you can perform a DELETE operation.

```
curl -k --username data --password *yourpassword* https://192.168.56.102/dropbox/*collection*/*product* -X DELETE
```


Features:

- **Simple catalogue ingestion** (automatize metadata extraction, according to custom metadata extractor)
- **Simple storage ingestion** (automatize data storage)
- **Simple APIs** (RESTful APIs: PUT for file upload, DELETE for file removal)
- **Support for NRTService** metadata extractor

Firefox

OSGC Catalogue Browse

file:///C:/Users/Salvatore Pinto/Desktop/temp/OSCLIENT/index.html

Google

OPEN SOURCE GeoSpatial Catalogue

ASAR over India

Filter by: Collection Area Time Metadata

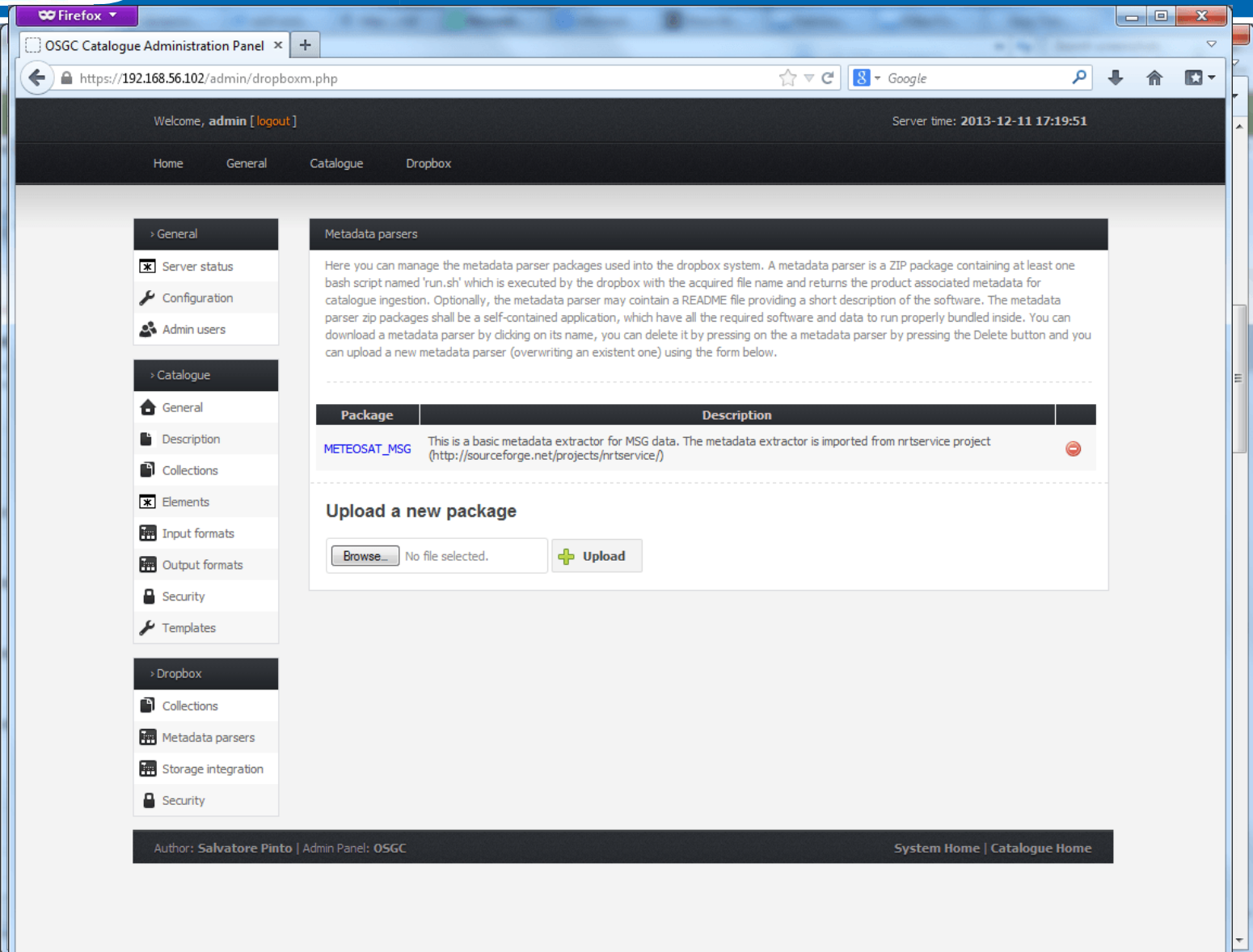
Results for collection: [ENVISAT ASAR Wide Swath Mode \(ASA\)](#)
Showing 20 results of 147704 (page 1)

ASA_WSM_1PNUPA20120405_111800_000000733113	Start: 2012-04-05T11:18:00.000Z	Stop: 2012-04-05T11:19:14.000Z	Size: 71406908	Orbit: 52780	Track: 180
ASA_WSM_1PNUPA20120402_094400_000000733113	Start: 2012-04-02T09:44:00.000Z	Stop: 2012-04-02T09:45:14.000Z	Size: 71406908	Orbit: 52780	Track: 180
ASA_WSM_1PNUPA20120402_094326_000000733113	Start: 2012-04-02T09:43:26.000Z	Stop: 2012-04-02T09:44:39.000Z	Size: 71406908	Orbit: 52780	Track: 180
ASA_WSM_1PNUPA20120402_094241_000000733113_00180_52780_1433.N1	Start: 2012-04-02T09:42:41.000Z	Stop: 2012-04-02T09:43:54.000Z	Size: 71406908	Orbit: 52780	Track: 180
ASA_WSM_1PNDPA20120408_074757_000000733113_00265_52865_2895.N1	Start: 2012-04-08T07:47:57.000Z	Stop: 2012-04-08T07:49:11.000Z	Size: 71088453	Orbit: 52865	Track: 265
ASA_WSM_1PNDPA20120408_074712_000000733113_00265_52865_2888.N1	Start: 2012-04-08T07:47:12.000Z	Stop: 2012-04-08T07:48:26.000Z	Size: 71056988	Orbit: 52865	Track: 265
ASA_WSM_1PNDPA20120402_080810_000000733113_00179_52779_2896.N1	Start: 2012-04-02T08:08:10.000Z	Stop: 2012-04-02T08:09:23.000Z	Size: 71127333	Orbit: 52779	Track: 179
ASA_WSM_1PNUPA20120330_095319_000000733113_00137_52737_1430.N1	Start: 2012-03-30T09:53:19.000Z	Stop: 2012-03-30T09:54:32.000Z	Size: 71477253	Orbit: 52737	Track: 137
ASA_WSM_1PNUPA20120330_095236_000000733113_00137_52737_1435.N1	Start: 2012-03-30T09:52:36.000Z	Stop: 2012-03-30T09:53:50.000Z	Size: 71523548	Orbit: 52737	Track: 137
ASA_WSM_1PNUPA20120311_095011_000000733112_00295_52464_1431.N1	Start: 2012-03-11T09:50:11.000Z	Stop: 2012-03-11T09:51:25.000Z	Size: 71555013	Orbit: 52464	Track: 295
ASA_WSM_1PNUPA20120311_094945_000000733112_00295_52464_1446.N1	Start: 2012-03-11T09:49:45.000Z	Stop: 2012-03-11T09:50:58.000Z	Size: 71393948	Orbit: 52464	Track: 295

85.94239, 12.02641

Features:

- **Easy to use** (users needs just a browser to query the catalogue)
- **Customizable** (you can customize query GUI and output metadata)
- **OGC OpenSearch support** (can query any catalogue, supposing that it exposes an OpenSearch RDF interface)
- **Integration into virtual laboratories** (can be integrated as standalone application to provide data access tools inside a virtual laboratory)



The screenshot shows a Firefox browser window displaying the OSGC Catalogue Administration Panel. The address bar shows the URL `https://192.168.56.102/admin/dropboxm.php`. The page header includes a welcome message for 'admin' and a server time of 2013-12-11 17:19:51. The navigation menu includes Home, General, Catalogue, and Dropbox. The left sidebar contains a tree view with categories: General (Server status, Configuration, Admin users), Catalogue (General, Description, Collections, Elements, Input formats, Output formats, Security, Templates), and Dropbox (Collections, Metadata parsers, Storage integration, Security). The main content area is titled 'Metadata parsers' and contains a text block explaining the metadata parser system. Below the text is a table with one entry: 'METEOSAT_MSG' with a description: 'This is a basic metadata extractor for MSG data. The metadata extractor is imported from nrtservice project (http://sourceforge.net/projects/nrtservice/)'. At the bottom of the main content area is an 'Upload a new package' section with a 'Browse...' button, the text 'No file selected.', and an 'Upload' button. The footer contains the text 'Author: Salvatore Pinto | Admin Panel: OSGC' and 'System Home | Catalogue Home'.

Features:

- **Full customization from the web**, with the possibility to setup and customize:
 - Metadata parsers
 - OpenSearch query fields and rules
 - Search output formats
 - Input metadata formats
 - Integration with storage

Thank you

Contacts:

Salvatore Pinto (salvatore.pinto@egi.eu)