

A VOSpace deployment: interoperability and integration in big infrastructure

S. Bertocco

Trieste 13 -14 December 2017

Provide our users with a local data storage and computation infrastructure compliant with a world wide Virtual Observatory vision

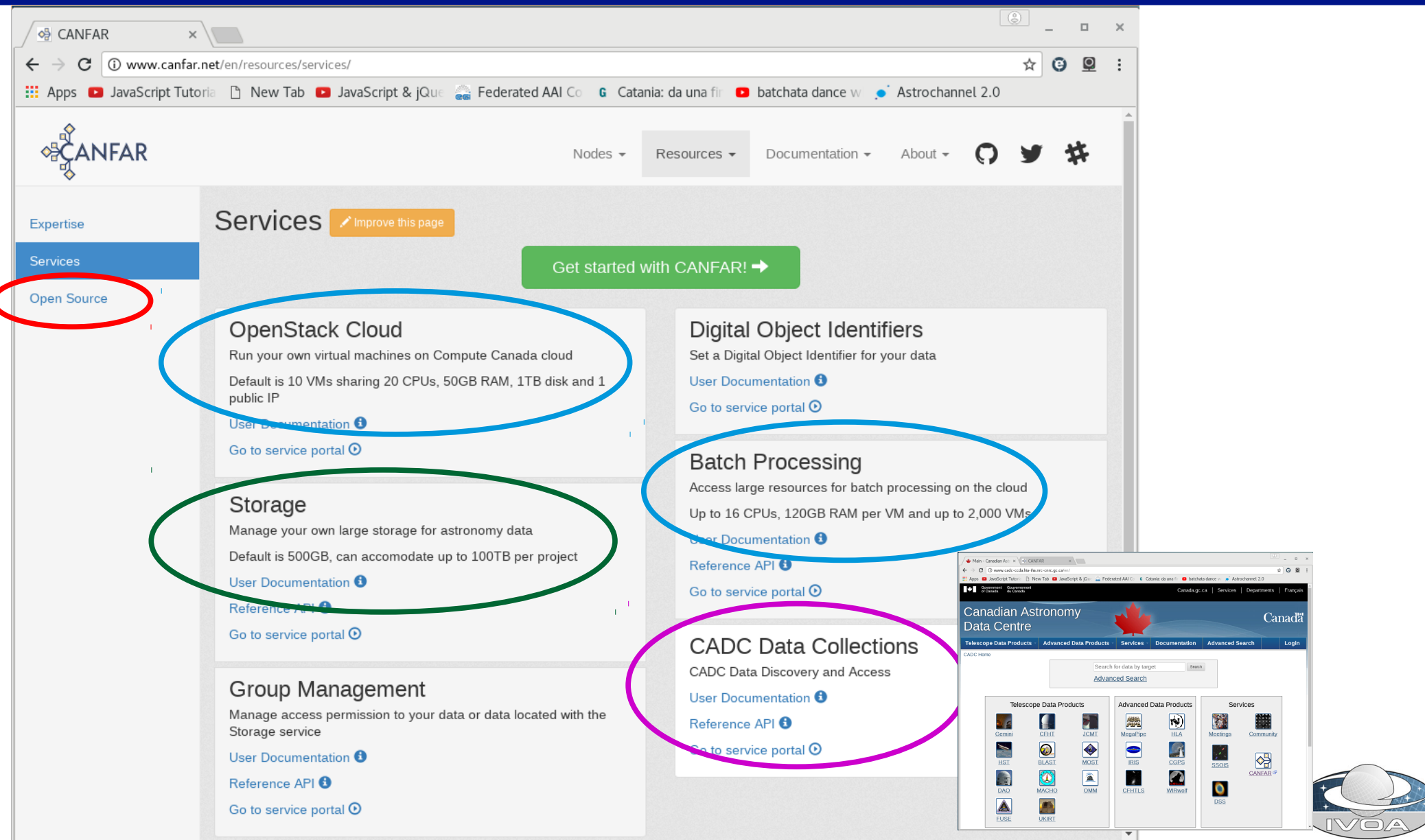
- **astronomical standards based**
to be widely accessible especially with astronomical tools
- **interoperable** with other similar storage services
to increase the accessible sets of data
- **integrated** with other big software infrastructures



CANFAR, the Canadian Advanced Network for Astronomical Research combines:

- the Canadian national research network (CANARIE),
- cloud processing and storage resources (Compute Canada)
- an astronomy data center
(Canadian Astronomy Data Center – CADC)
 - hosting a very large data set
 - specialized in data mining, data processing, data distribution and data transferring
 - providing a lot of sophisticated tools to support and enhance the research efforts of Canadian and international astronomers





The screenshot shows the CANFAR website interface. The sidebar on the left has a menu with 'Expertise', 'Services', and 'Open Source' (circled in red). The main content area is titled 'Services' and features a green button 'Get started with CANFAR!'. Several service cards are displayed, each with a title, description, and links for documentation and service portals. These cards are circled in different colors: OpenStack Cloud (blue), Storage (green), Digital Object Identifiers (blue), Batch Processing (blue), CADC Data Collections (magenta), and Group Management (green). An inset window in the bottom right corner shows the 'Canadian Astronomy Data Centre' website, which includes a search bar and a grid of data products and services.

Services [Improve this page](#)

[Get started with CANFAR!](#)

OpenStack Cloud
 Run your own virtual machines on Compute Canada cloud
 Default is 10 VMs sharing 20 CPUs, 50GB RAM, 1TB disk and 1 public IP
[User Documentation](#)
[Go to service portal](#)

Digital Object Identifiers
 Set a Digital Object Identifier for your data
[User Documentation](#)
[Go to service portal](#)

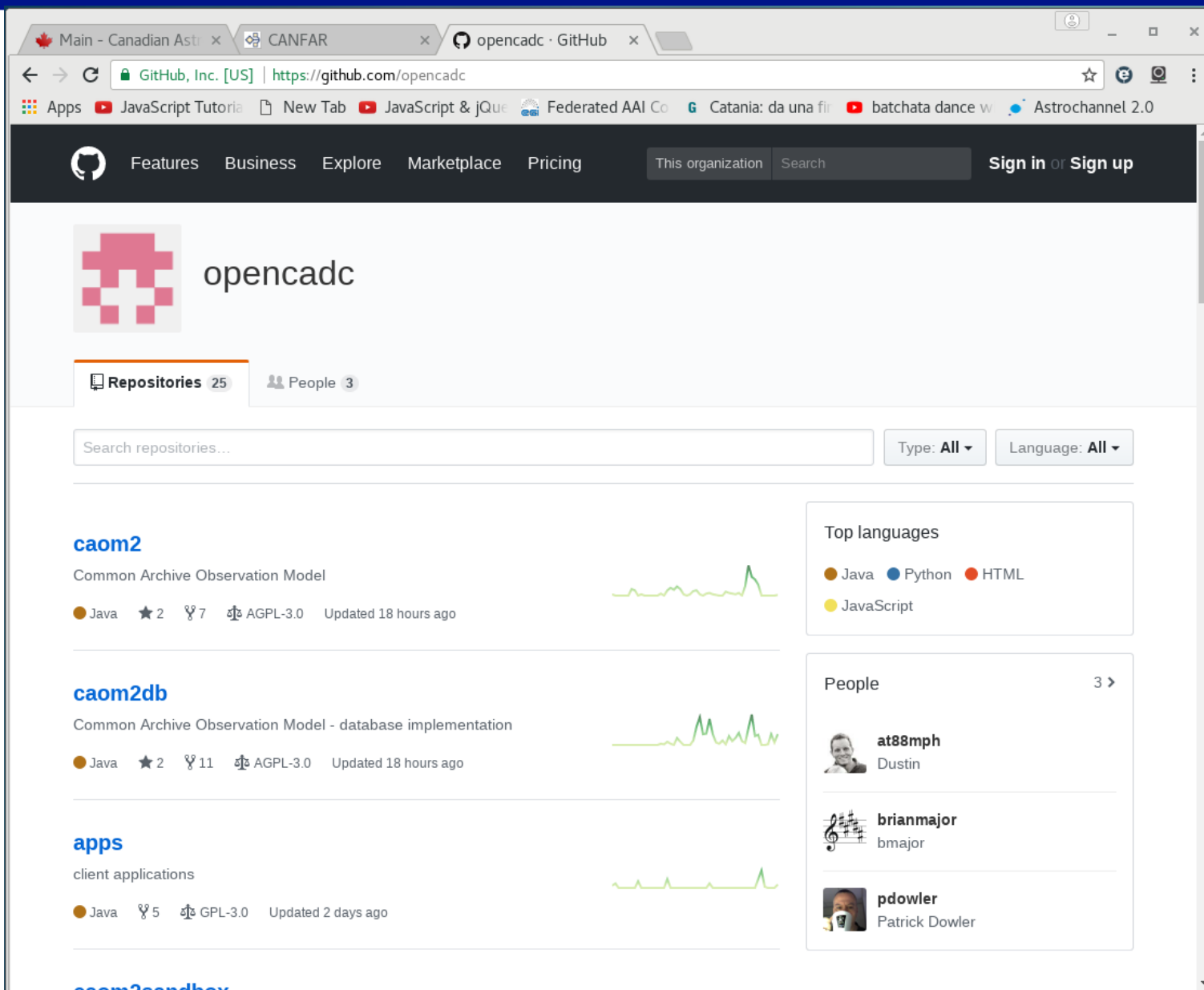
Batch Processing
 Access large resources for batch processing on the cloud
 Up to 16 CPUs, 120GB RAM per VM and up to 2,000 VMs
[User Documentation](#)
[Reference API](#)
[Go to service portal](#)

CADC Data Collections
 CADC Data Discovery and Access
[User Documentation](#)
[Reference API](#)
[Go to service portal](#)

Storage
 Manage your own large storage for astronomy data
 Default is 500GB, can accomodate up to 100TB per project
[User Documentation](#)
[Reference API](#)
[Go to service portal](#)

Group Management
 Manage access permission to your data or data located with the Storage service
[User Documentation](#)
[Reference API](#)
[Go to service portal](#)

Canadian Astronomy Data Centre
 Telescope Data Products | Advanced Data Products | Services | Documentation | Advanced Search | Login
 Search for data by target
[Advanced Search](#)
 Telescope Data Products: Gemini, CEHL, JCM1, HST, BLAST, MOST, DAO, MACHO, GMM, EUSE, UNIRL
 Advanced Data Products: MegaPipe, HLA, IRIS, CGPS, CFHTLS, WISE
 Services: Meetings, Community, SSOS, CANFAR, DSS



The screenshot shows the GitHub page for the OpenCADC organization. The browser tabs include 'Main - Canadian Astr...', 'CANFAR', and 'opencadc · GitHub'. The address bar shows 'https://github.com/opencadc'. The page header includes navigation links like 'Features', 'Business', 'Explore', 'Marketplace', and 'Pricing', along with a search bar and 'Sign in' or 'Sign up' buttons.

The main content area displays the OpenCADC organization profile with a pink pixelated logo. Below the profile, there are tabs for 'Repositories' (25) and 'People' (3). A search bar for repositories is present, along with filters for 'Type: All' and 'Language: All'.

Three repositories are listed:

- caom2**: Common Archive Observation Model. Languages: Java. Stars: 2. Forks: 7. License: AGPL-3.0. Updated 18 hours ago.
- caom2db**: Common Archive Observation Model - database implementation. Languages: Java. Stars: 2. Forks: 11. License: AGPL-3.0. Updated 18 hours ago.
- apps**: client applications. Languages: Java. Stars: 5. Forks: 5. License: GPL-3.0. Updated 2 days ago.

On the right side, there are two sections:

- Top languages**: A chart showing the distribution of languages used in the repositories. The legend indicates: Java (orange), Python (blue), HTML (red), and JavaScript (yellow).
- People**: A list of contributors with their avatars and names:
 - at88mph** (Dustin)
 - brianmajor** (bmajor)
 - pdowler** (Patrick Dowler)



➡ <https://github.com/opencadc>

CADC

Modules:

- **vos** VOSpace standard implementation
- **ac** Access Control (including GMS)
- **cdp** Credential Delegation Protocol implementation
- **reg** Registry Interface implementation (including VOSI)
- **uws** Universal Worker Service Pattern implementation
- **core** core utilities and logging

➡ IVOA Standards and recommendations
based
(<http://ivoa.net/>)



VOSpace recommendation:

“VOSpace is the IVOA interface to distributed storage. It specifies how VO agents and applications can use network attached data stores to persist and exchange data in a standard way.

A VOSpace web service is an access point for a distributed storage network. Through this access point, a client can:

- add or delete data objects in a tree data structure
- manipulate metadata for the data objects
- obtain URIs through which the content of the data objects can be accessed

VOSpace does not define how the data is stored or transferred, only the control messages to gain access. Thus, the VOSpace interface can readily be added to an existing storage system.

When we speak of “a VOSpace”, we mean the arrangement of data accessible through one particular VOSpace service.”



VOSpace:

- VOSpace front-end

<https://github.com/opencadc/vos>

<https://github.com/oats-cadc/oats-vospace-web>

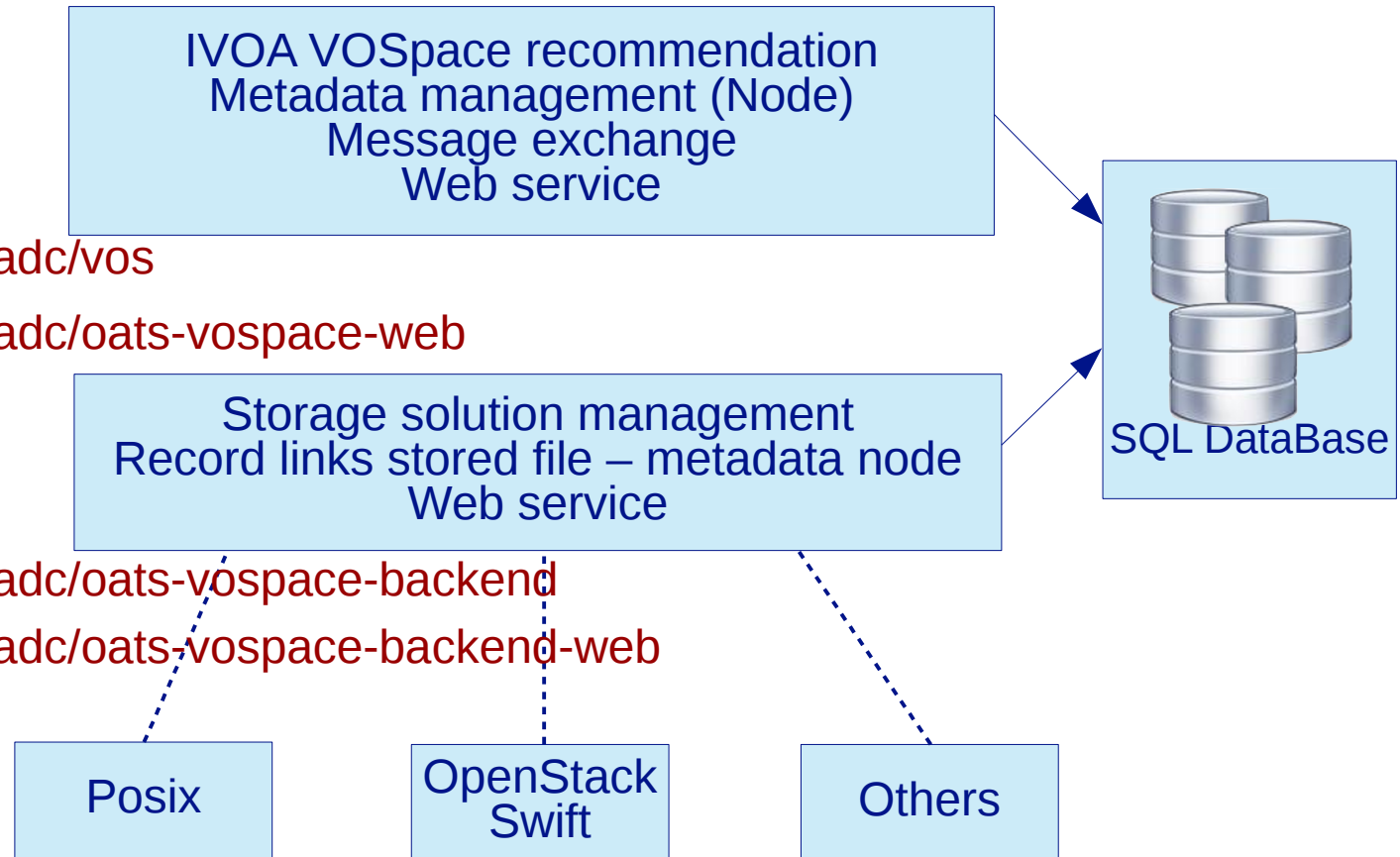
- VOSpace back-end

<https://github.com/oats-cadc/oats-vospace-backend>

<https://github.com/oats-cadc/oats-vospace-backend-web>

- Physical storage

<https://github.com/oats-cadc/oats-vospace-backend-developers-guide>



Access permissions stored in VOSpace database
VOSpace access policy based on group membership
Access Control Service (<https://github.com/opencadc/ac>)

- Manage users
- Manage groups
- Has info about group memberships
- Manage more user's identities:
 - username/password
 - cookies
 - numeric
 - X.509 certificates



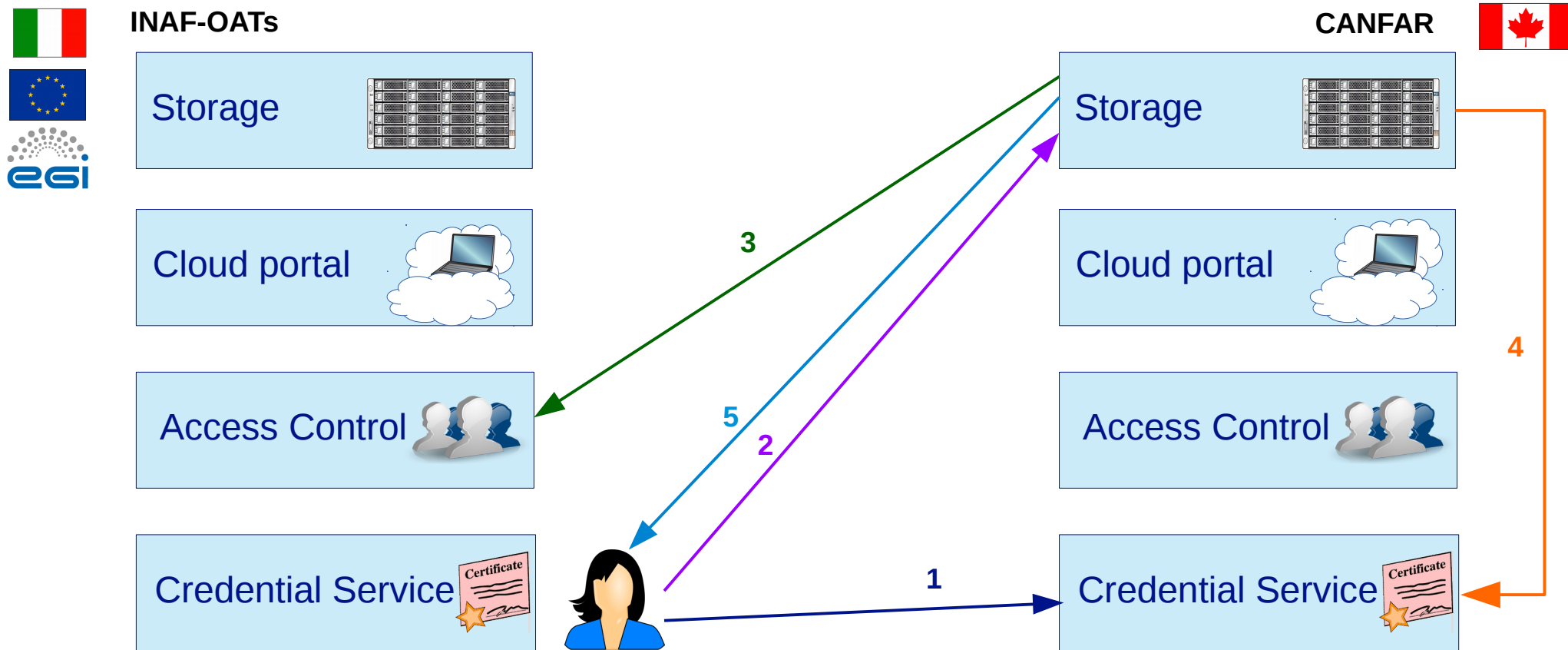
The credential delegation protocol allows a client program to delegate a user's credentials to a service such that the service may make requests of other services in the name of that user.

Credential Delegation Service

<https://github.com/opencadc/cdp>

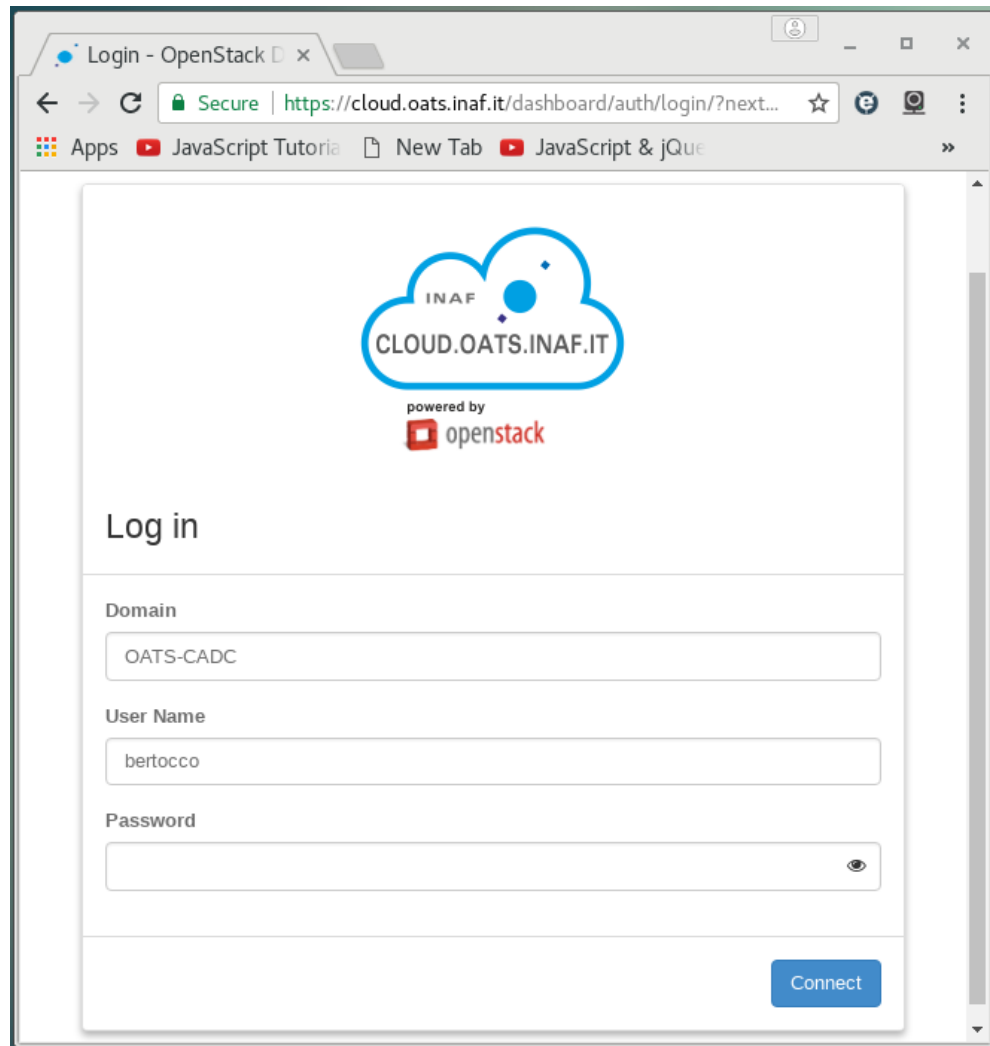
<https://github.com/oats-cadc/oats-cred-web>





- 1) INAF-OATs user Bertocco delegates her x509 credentials to CANFAR Credential Service
- 2) user Bertocco asks for data of her INAF-OATs group to CANFAR storage service
- 3) CANFAR storage service checks the group affiliation of the user in the INAF-OATs group management service
- 4) CANFAR storage service gets the user's delegated credentials from the CANFAR Credential Delegation Service to be able to make calls to each other service on behalf of the initial user
- 5) CANFAR storage service returns data to the INAF-OATs user Bertocco





OATs-INAF hosts a cloud site:

- OpenStack Mitaka based
- Storage:
 - 10TB VM storage (cinder)
 - 50 TB user's data storage
- Authentication: Keystone
 - backend Idap
 - Plus keystone-voms module



Cloud Management Framework: **OpenStack**
(compliant with EGI federated cloud architecture)

Common Authentication and interoperability:

◆ Using VOMS-proxy

- Request a cloud authorization token
- Connect to the OpenStack console
- Manage virtual machines
- Authenticate in VOSpace and other IVOA base services

◆ Using username/password

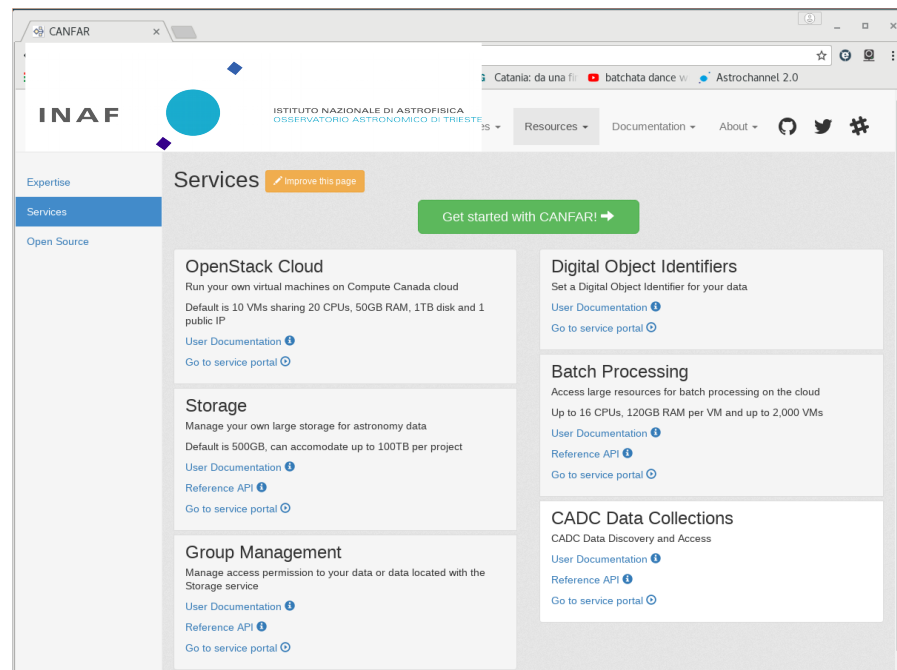
- log-in into VMs
- log-in into OATs cloud portal
- authenticate in VOSpace and other IVOA base services



- Services IVOA recommendation based
 - VOSpace, Access Control and Credential
 - Interoperable with CANFAR Services (VOSpace, Credential Service, Registry)
- Cloud resources
- EGI federation compliant and accessible by EGI users (X.509)
- Cloud access to data stored in VOSpace (both OATs and CANFAR hosted)



- An integrated approach to our services exploitation



- A common authentication (will be provided by EOSC Pilot ?)
and authorization model (will it be an IVOA recommendation on group-based auth?)



Thanks!

