
CERN

Opencast

Integrating the CERN multimedia production

Our goals:

1. Independent (non-intrusive) encoding and processing system.
2. SSO authentication integration for each media resource. Remove DFS ICACLs dependency.
3. Indico and CDS plug & play integration.
4. Historical repository transformation/adaptation.
5. OpenSource Video Player for multiple video streams, captioning, HLS support for language audio streams, etc.

How to achieve our goals?

1. Opencast Free OpenSource encoding and publishing platform.
2. Apache balanced web servers: OpenID Connect integration & CEPH fs storage.
3. Python Tools for OpenCast (PyCast): Our own Python library (opencast, indico, media, openid, transform, transcription … classes)
4. Transform scripts to new player data.json format (legacy).
5. Paella Player.

Implementation:

- Every service is Puppet managed (CERN GitLab) & OpenStack IaaS hosted.
- PyCast is a python package for easy deployment and fast upgrading.
- It also uses the Opencast https API. Avoid direct access to files and raw data.

Additional services:

- Rundeck automation tool. (CERN puppet module).
- Let’s Encrypt Certificates with auto-renewal.
- S3 backup.
- Matomo analytics.

Other tools:

- Telegraf for system metrics monitoring.
- Filebeat for Elasticsearch/Kibana log search.

1. Opencast ecosystem.

- Opencast is a free open-source enterprise level lecture recording system. The core of the system delivers functionality for scheduling, media encoding, editing and content delivery.
- Opencast contains everything you need for scheduling captures, "**trimming, captioning (external service), and conversion of output media to several formats**".
- Opencast does not mean to replace publishing services at CERN (Indico and CDS).

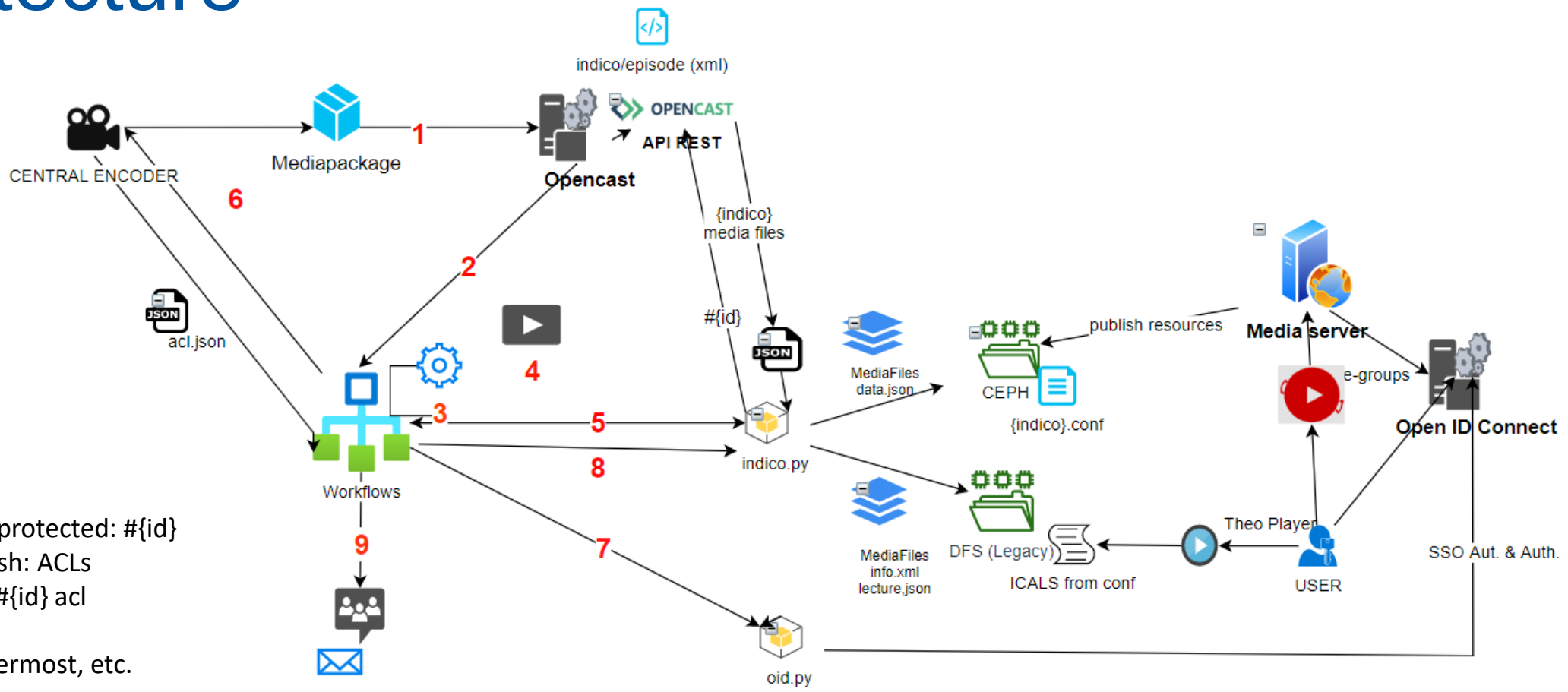
Concepts

- ***Mediapackage***:
  - A mediapackage consists on a XML definition for a Opencast Event and related assets (multimedia files and metadata files).
- ***Asset****:*
  - Mediapackage element (files): such as .mp4 video, dublincore xml, vtt.
- ***Workflow***:
  - Workflows are the central element to define how a media package is being processed by the Opencast services. Their definitions consist of a list of workflow operations
- ***Operation***:
  - It is basically map a piece of configuration to Opencast code. XML definition.
  - Around 80 operations available.

Infrastructure

- **Admin node**: activemq messaging service, elasticsearch for internal and resource indexing, mySQL database (hosted externally).
- **Worker nodes**: run tasks (mainly ffmpeg encoding), pyCast Tools.
- **Engage node**: For internal video player. Strong dependency on Opencast System. Not required: We use external video-player.

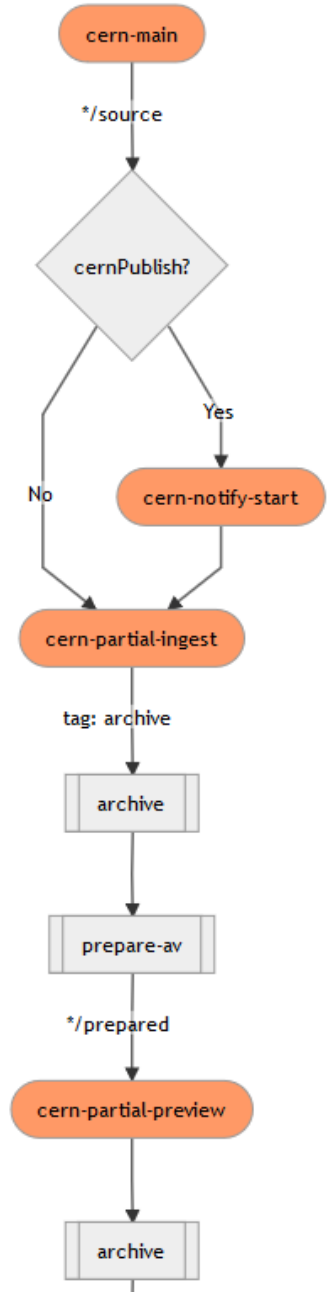
Architecture



1.- API: Ingest
2.- WF: cern-main
3.- OP: Encode
4.- Publish Internal
5.- Export & Publish protected: #{id}
6.- Notify: CES; Publish: ACLs
7.- Create OID Role: #{id} acl
8.- Propagate ACL
9.- Notify: CES, Mattermost, etc.

Main workflow: cern-main

```
<?xml version="1.0" encoding="UTF-8"?>
<definition xmlns="http://workflow.opencastproject.org">
  <id>cern-main</id>
  <title>Encode and export to CEPH File System</title>
  <description>Prepare Indico assets, encode in different qualities, generate the slides and copy all the
media to CEPH file system. The files will be accesible only for opencast-
admins. It also generates the data.json for PaellaPlayer.</description>
  <operations>
    <operation
        id="defaults"
        description="Applying default configuration values">
      <configurations>
        <configuration key="flagForCutting">false</configuration>
        <configuration key="flagForReview">false</configuration>
        <configuration key="flagForTranscription">false</configuration>
      …
  <operation
      id="include"
      if="${cernPublish}"
      fail-on-error="false"
      exception-handler-workflow="cern-partial-error"
      description="Notification of workflow starting">
    <configurations>
      <configuration key="workflow-id">cern-notify-start</configuration>
    </configurations>
   </operation>
```



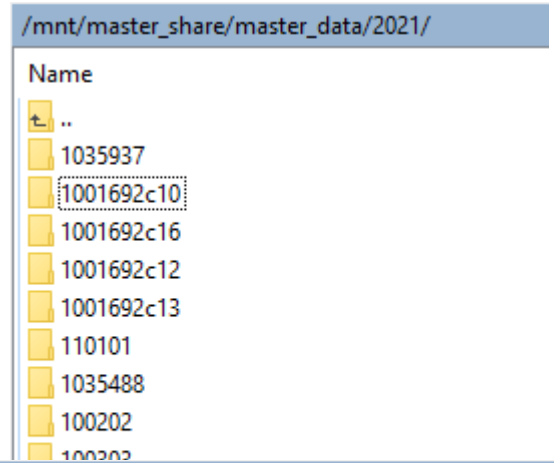
Workflows: cern-**master**-export

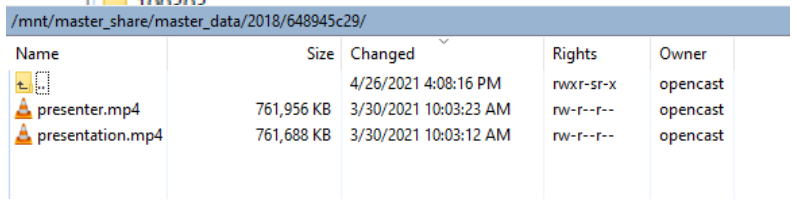
- Preserving source files.

Succeeded	snapshot	Archive preview information
Succeeded	include	Media: Export mediaPackage source files to CEPH File System.
Succeeded	mattermost-notify	Notify Mattermost workflow starting

Succeeded	segment-video	Detecting slide transitions in presentation track
Succeeded	segmentpreviews	Creating preview images for presentation segments
Succeeded	extract-text	Extracting text from presentation segments
Succeeded	publish-engage	Publishing to Opencast Media Module

Succeeded	mattermost-notify	Notify Mattermost: metadata and files export
Succeeded	execute-once	Media: Export mediaPackage to CEPH File System.
Skipped	execute-once	Media: Export other stored event to CEPH File System.
Succeeded	execute-once	Create ACL .conf file for this event. It will make the resources accesible only for the opencast-admins role.



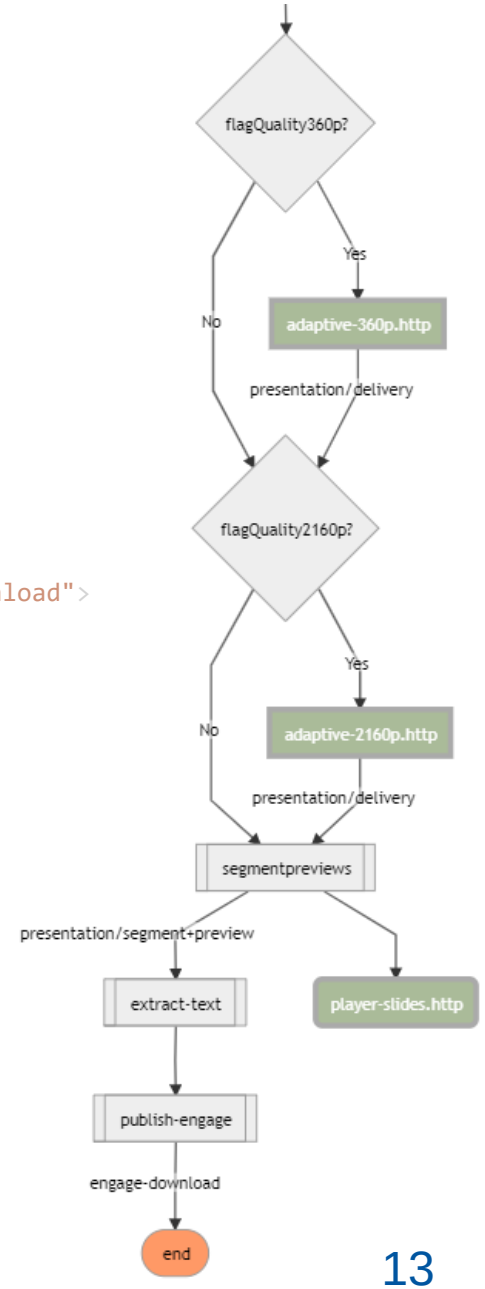


	Files	Size
Conferences (1976-2020)	6481	22Tb
Weblectures (1999-2020)	932828	33Tb

Workflows: cern-partial-publish

- Encoding operations.

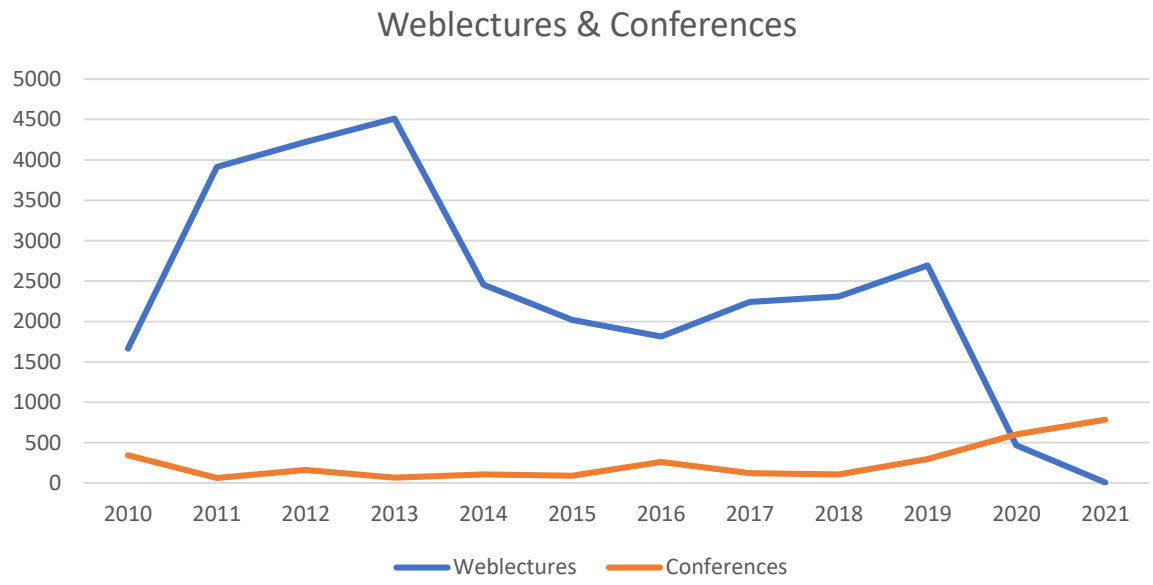
```
<operation
  if="${flagQuality360p} AND NOT ${flagQualityMultiple} AND NOT ${flagQualityParallel}"
  id="encode“  exception-handler-workflow="cern-encoding-error“ description="Encoding 360p video to MP4 download">
  <configurations>
    <configuration key="source-flavor">*/themed</configuration>
    <configuration key="target-flavor">*/delivery</configuration>
    <configuration key="target-tags">engage-download,engage-streaming,360p-quality</configuration>
    <configuration key="encoding-profile">adaptive-360p.http</configuration>
  </configurations>
</operation>
```



Workflows: cern-media-export

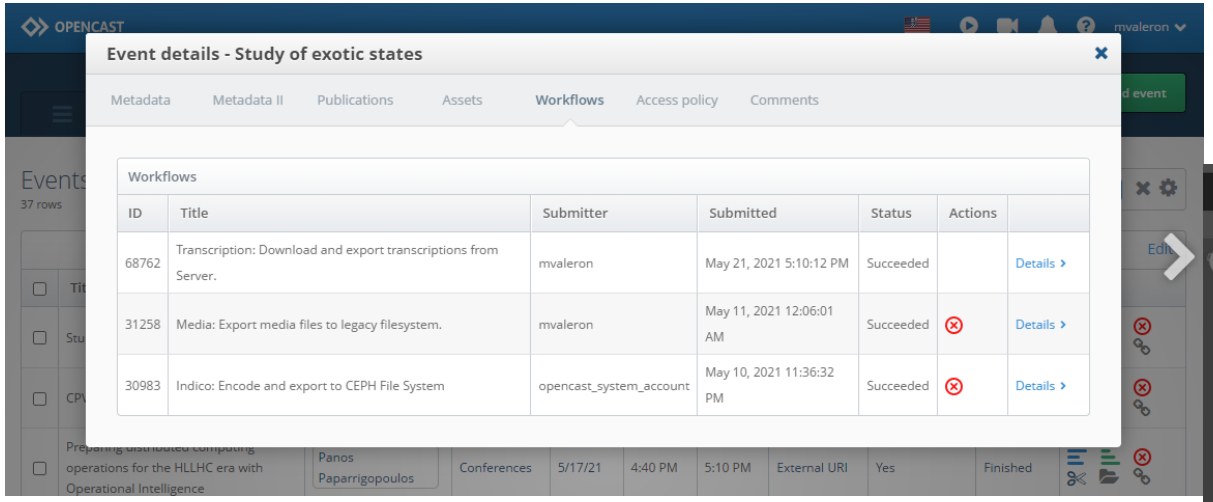
- Content published.

	Folders	Files	Size
Conferences (1976-2020)	6614	195670	9Tb
Weblectures (1999-2020)	30376	760054	30Tb



Workflows: cern-transcription-ingest

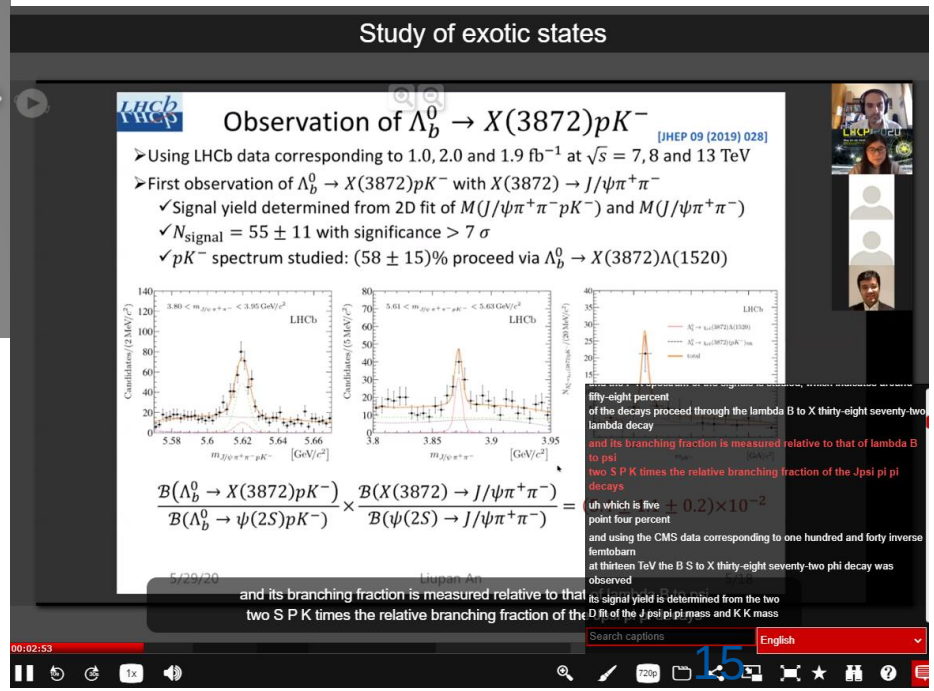
- Custom Workflow for external transcription tools.



When the transcription and translations jobs are finished, the .vtt files are downloaded and data.json is rebuild with the new captions.

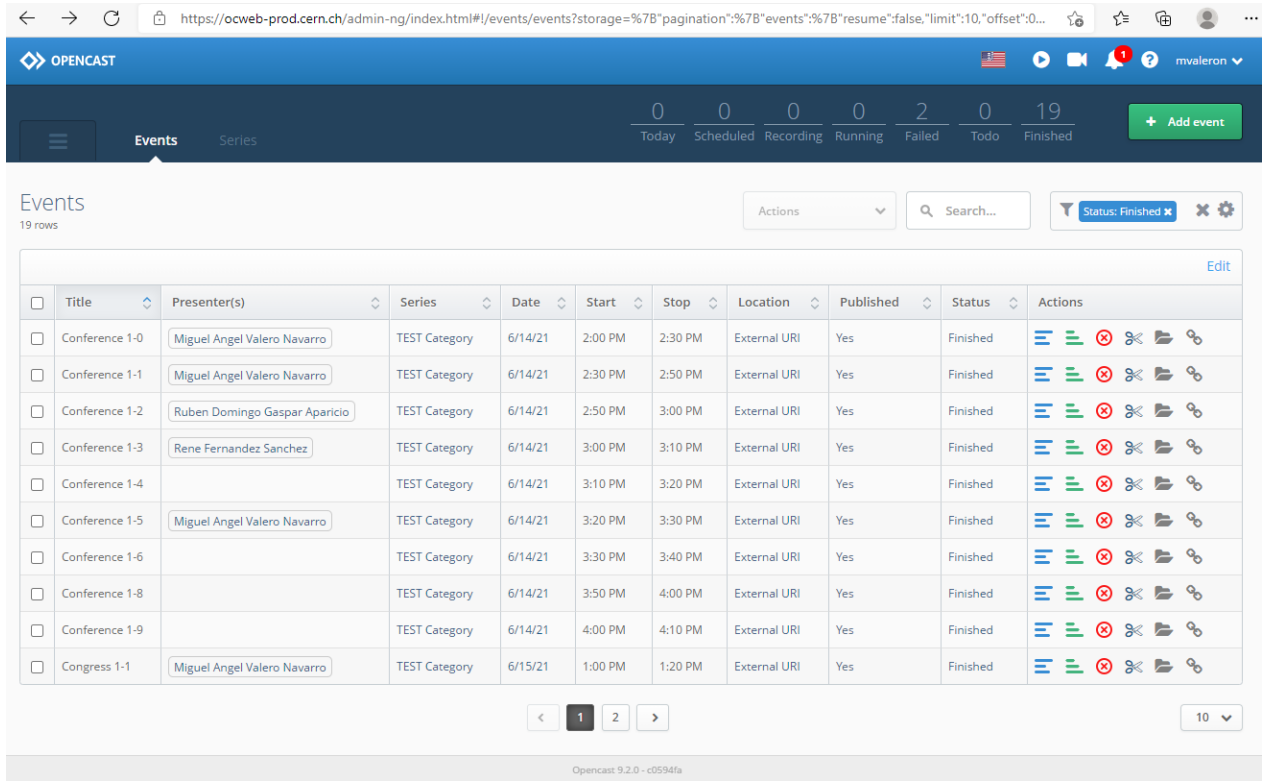
00:02:14.820 --> 00:02:21.150 the lambda B to X thirty-eight seventy-two P K decay was observed for the first time

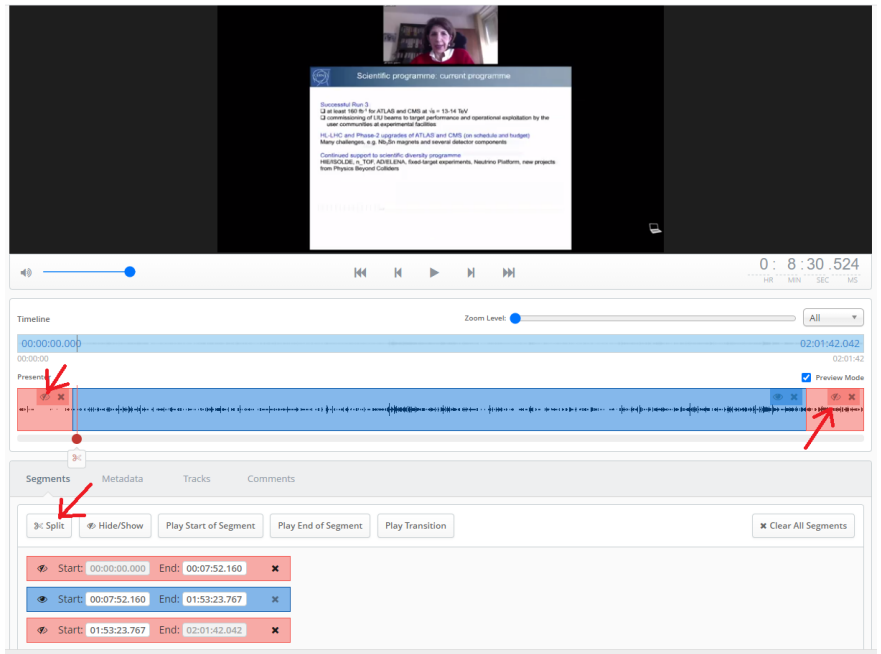
21 00:02:21.350 --> 00:02:29.180 its signal yield is determined from a two dimensional fit of the J psi pi pi P K mass and J psi pi pi mass



Opencast admin web UI.

- Everything is automated but it also offers a user management interface.





- Themes (intro/outro).
- Series (indico category).
- Events.
- Workflow.
- Video editing.
- Monitoring

2. Apache web servers

- SSO: OpenID Connect integration.
- LBaaS: Openstack pool: balanced servers based on the year.
  - 1975-2020 apo.cern.ch
  - 2021- bakony.cern.ch
- Let’s Encrypt free certificate.

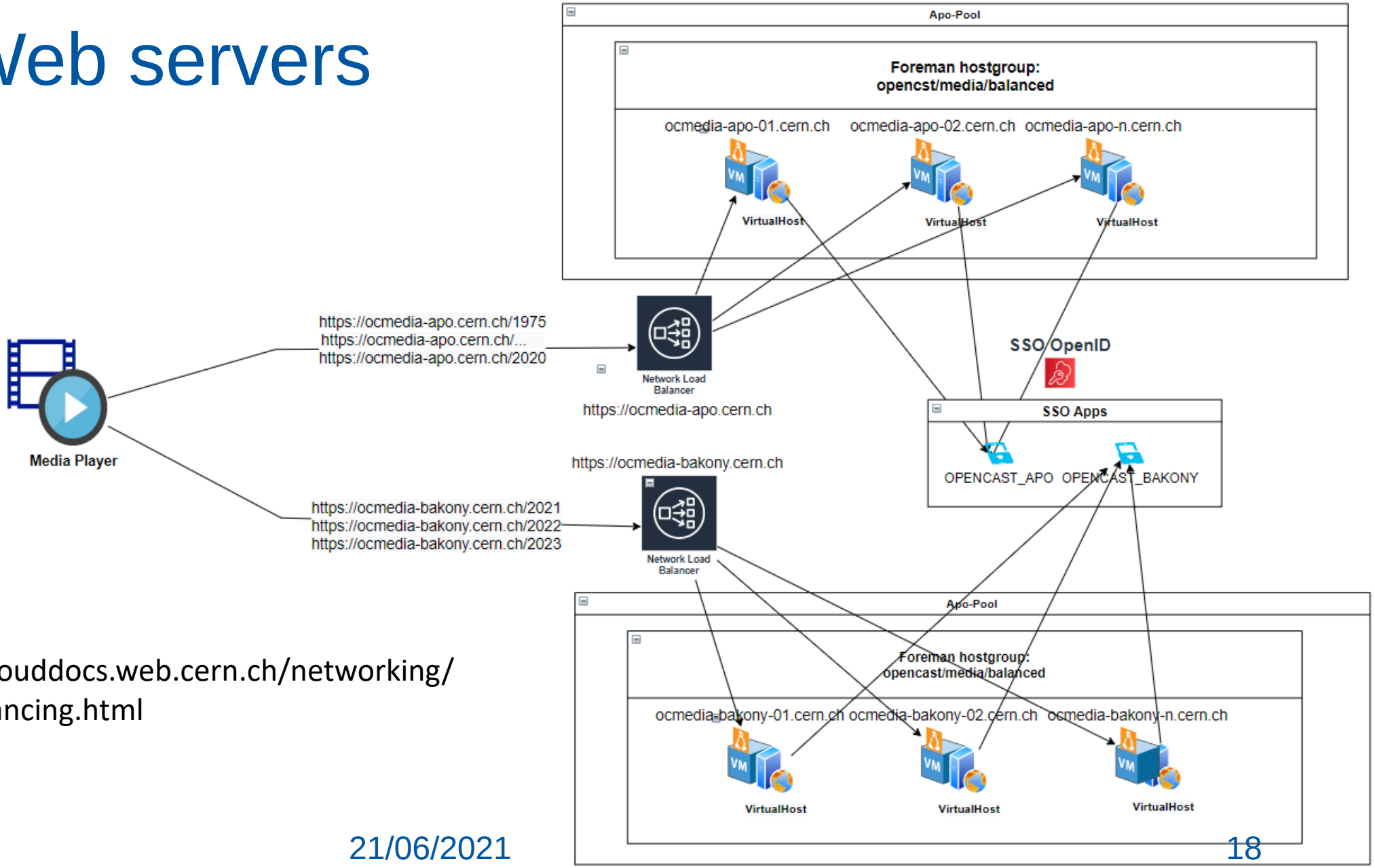
```
<VirtualHost *:443>
  ServerName <%= @balancer %>
   …
  # Protect any directory not explicitily exposed
  <Directory <%= @web_path %>/* >
    Require all denied
  </Directory>

  Include "/etc/httpd/conf/oidc_sec_<%= @app %>.conf"

  OIDCProviderMetadataURL
       https://auth.cern.ch/auth/realms/cern/.well-known/openid-configuration
  OIDCClientID <%= @app %>
  OIDCRedirectURI /redirect_uri
  OIDCProviderTokenEndpointAuth client_secret_basic
  OIDCRemoteUserClaim sub
  OIDCRefreshAccessTokenBeforeExpiry 30
   …

</VirtualHost>
```

2. Web servers



https://clouddocs.web.cern.ch/networking/load_balancing.html

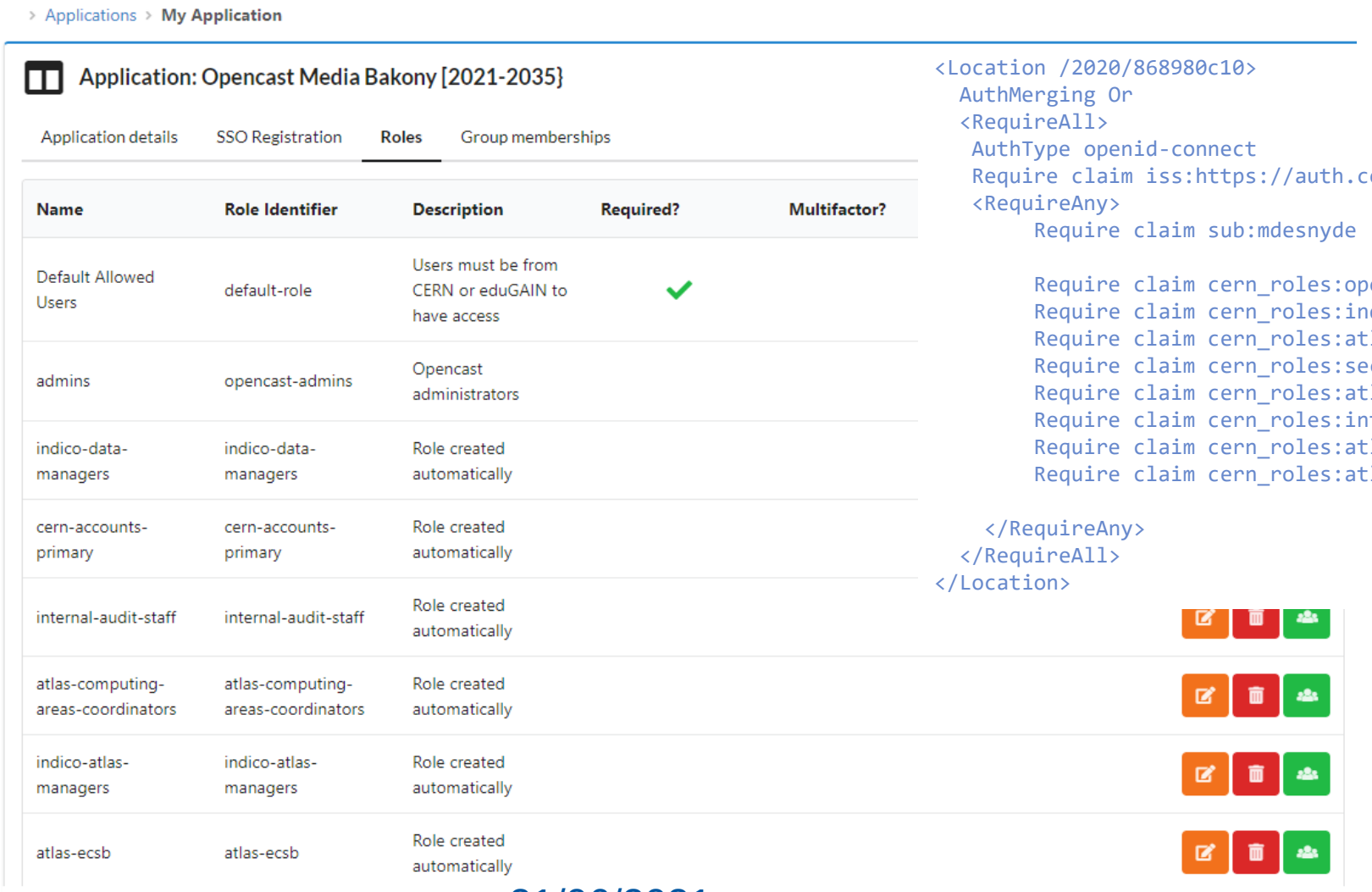
2. Web servers. Authorization

- Roles created dynamically: authorization service API
  - https://auth.docs.cern.ch/authzsvc/overview/
  - https://authorization-service-api.web.cern.ch/api/v1.0
- One eGroup -> One App Role
- No user-> Role needed.
- OpenID Connect. Role mapping.

Thanks Hannah Short (IT-CDA-IC) and Pablo Saiz (IT-CM-LCS)

	Roles	Users
1976-2020	104	352
2021	39	

2. Applications portal.

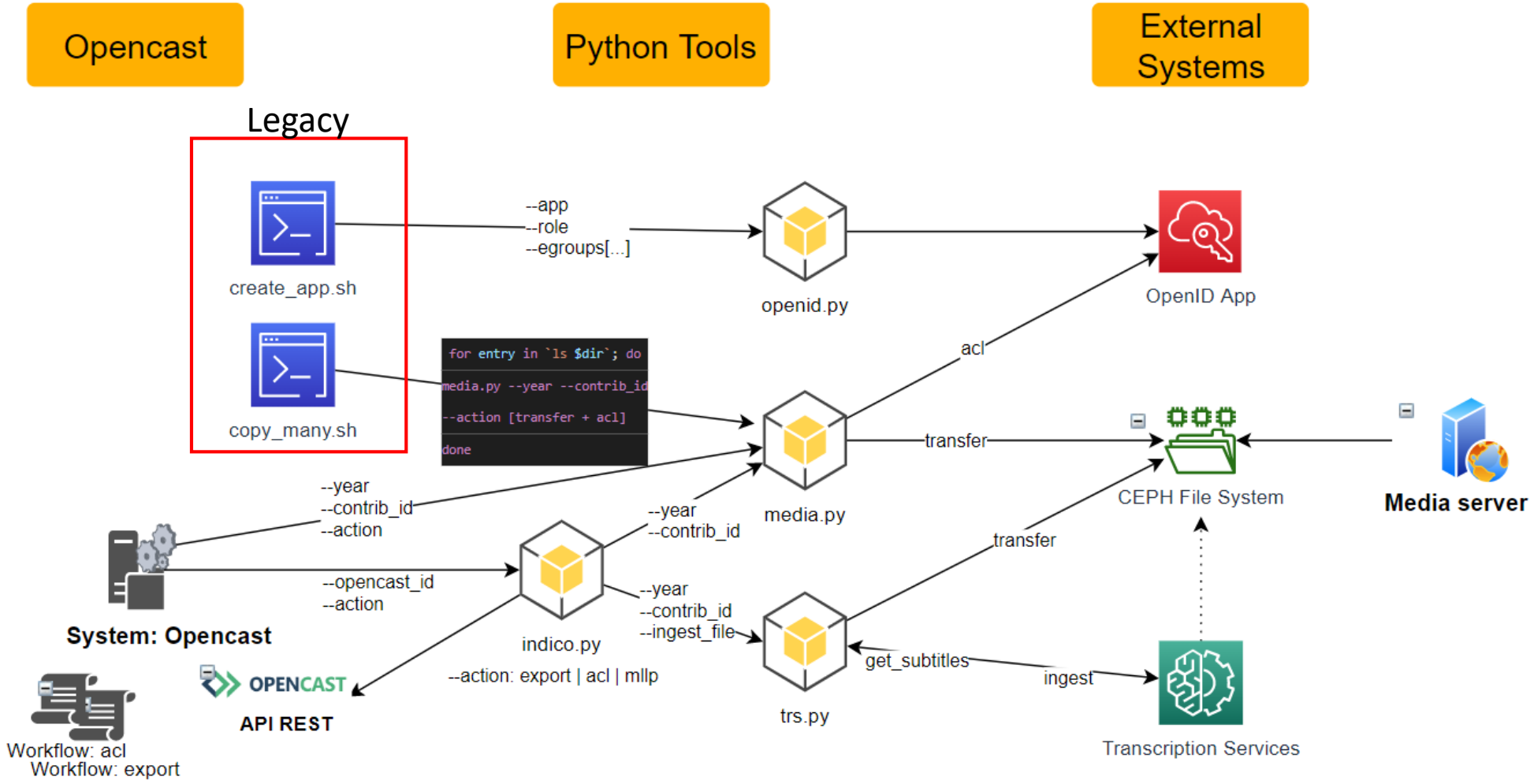



```
<Location /2020/868980c10>
  AuthMerging Or
  <RequireAll>
   AuthType openid-connect
   Require claim iss:https://auth.cern.ch/auth/realms/cern
   <RequireAny>
        Require claim sub:mdesnyde

        Require claim cern_roles:opencast-admins
        Require claim cern_roles:indico-atlas-managers
        Require claim cern_roles:atlas-mgt-members
        Require claim cern_roles:secretariat-atlas
        Require claim cern_roles:atlas-students-group
        Require claim cern_roles:internal-audit-staff
        Require claim cern_roles:atlas-cb-chair
        Require claim cern_roles:atlas-readaccess-active-members

    </RequireAny>
  </RequireAll>
</Location>
```

3. Python Tools (pyCast)



5. Servers and Services

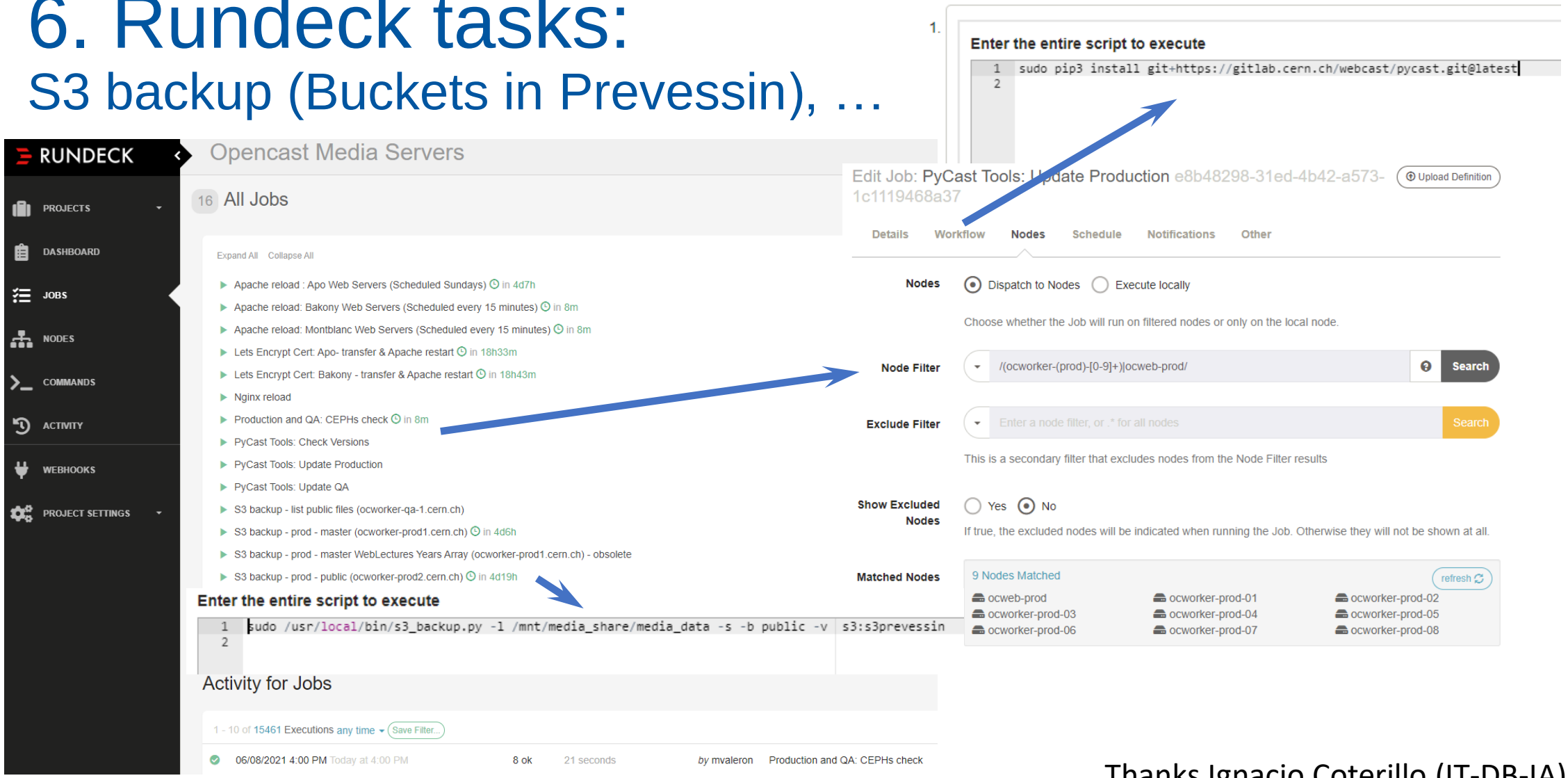
- Production:
  - 1 admin node, 8 workers, 1 mySQL database (central services)
    - Flavor=2xlarge: 16 cores, 16GBytes.
  - 4 balanced apache web servers (2 for 1975-2020 (apo), 2 for - (bakony)).
  - 1 Rundeck automation server, 1 mySQL database (servers reload, S3 backup, letsencrypt certs renewal, …)
- QA:
  - 1 admin node, 3 workers, 1 mySQL database
  - 1 apache web server (for 1975- (montblanc))
  - 1 Rundeck automation server
  - 1 Matomo server

6. Additional services

- Rundeck
- Matomo
- Telegraf
- Elasticsearch

6. Rundeck tasks:

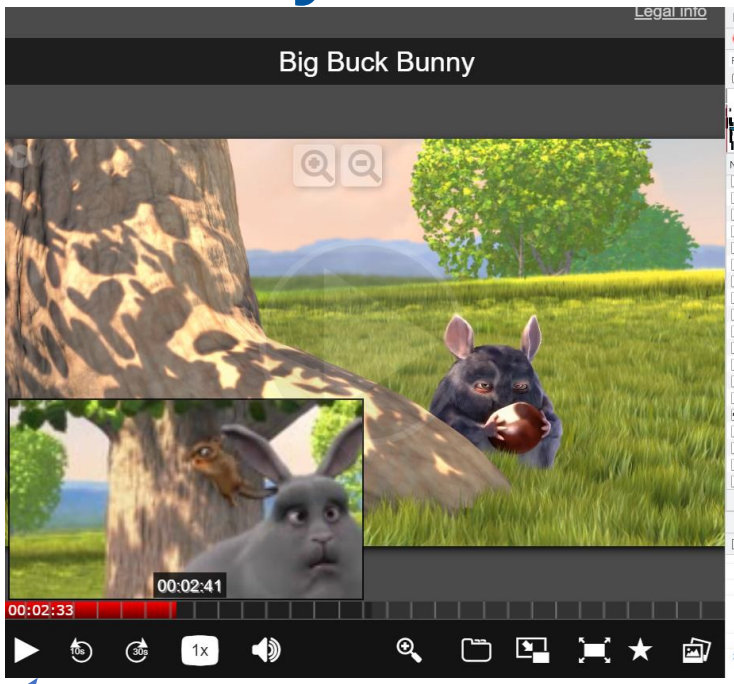
S3 backup (Buckets in Prevessin), …

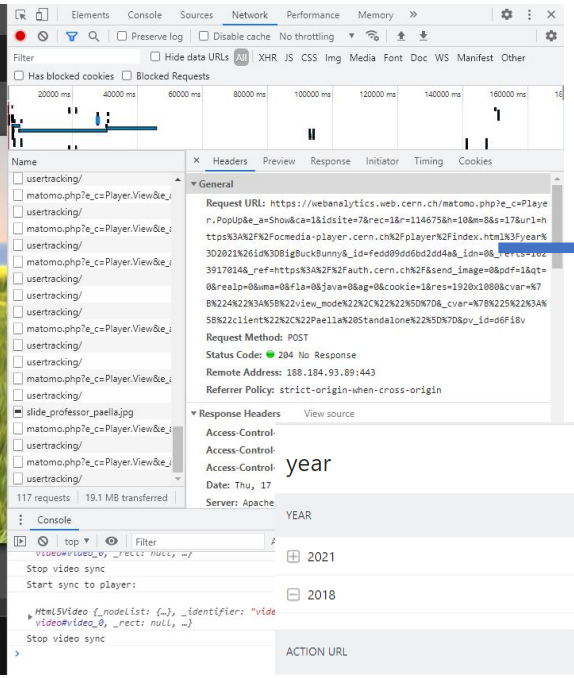


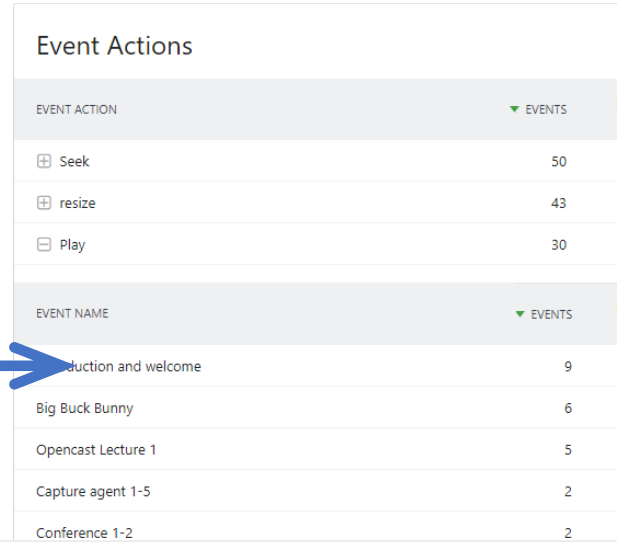
https://clouddocs.web.cern.ch/object_store/s3cmd.html

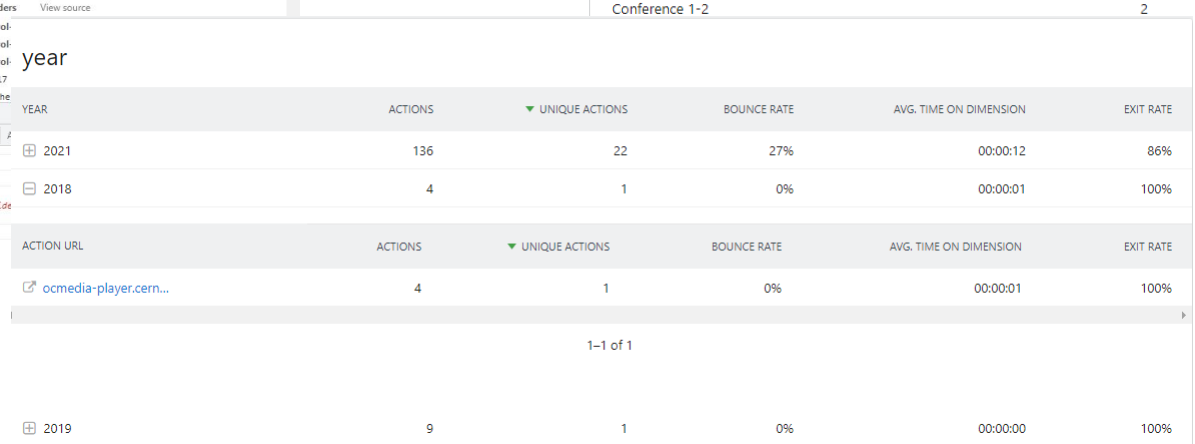
Thanks Ignacio Coterillo (IT-DB-IA) for sharing his implentation code.

6. Matomo analytics: Visits and anonymous user actions.





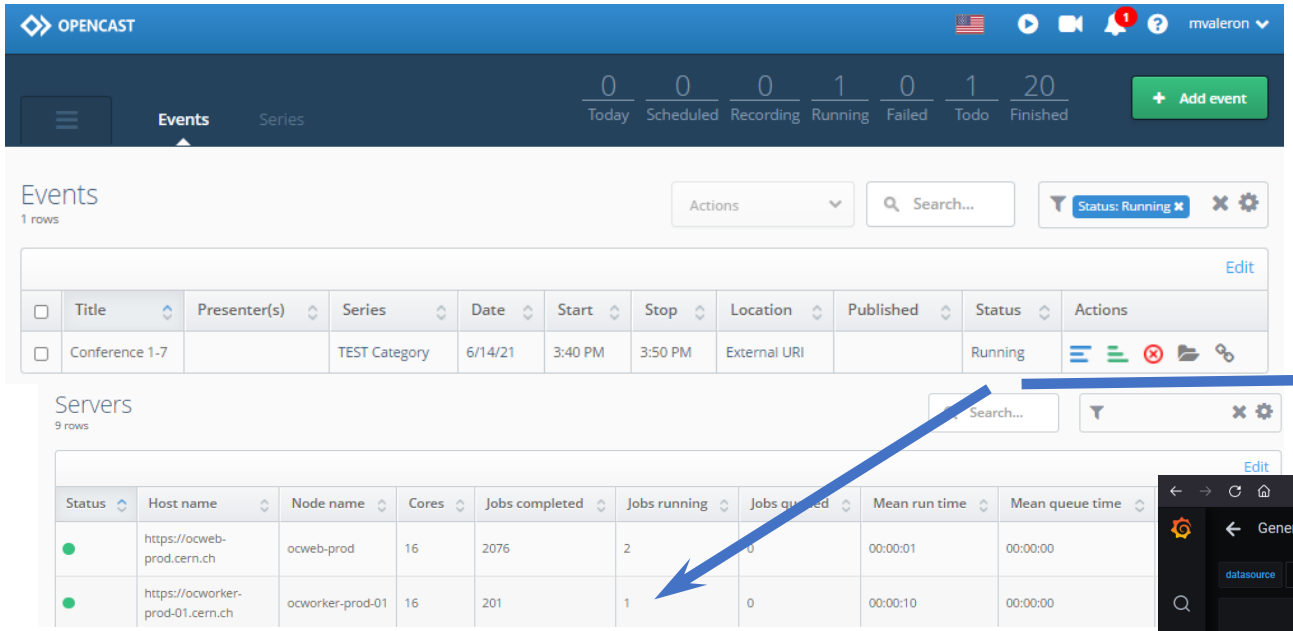


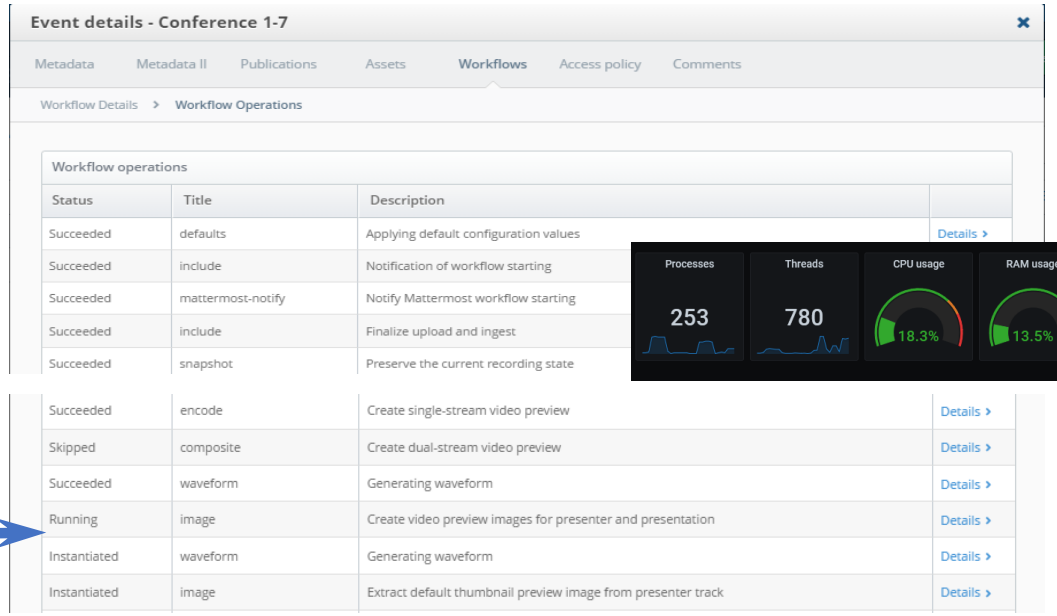


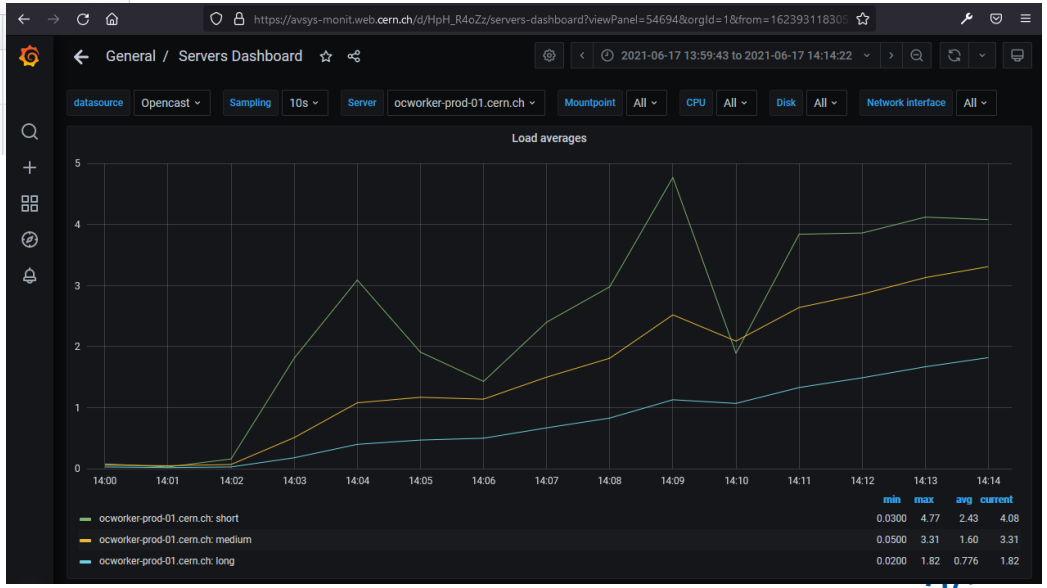

```
"org.opencast.usertracking.MatomoSaverPlugIn": {
  "enabled": true,
  "server": "https://webanalytics.web.cern.ch/",
  "site_id": 7,
  "tracking_client_name": "matomo",
  "heartbeat": 30,
  "ask_for_concent": true,
  "privacy_policy_url": "https://matomo.org/blog/2018/04/
          how-should-i-write-my-privacy-notice-for-matomo-analytics-under-gdpr/",
  "cookieconsent_base_color": "#1d8a8a",
  "cookieconsent_highlight_color": "#62ffaa"
},
```

Thanks Pete Jones (IT-CDA-WF) for registering the Sites.

Other tools:Telegraf.








```
package { 'telegraf':
  ensure  => latest,
  require => Yumrepo['influxdata'],
}
```

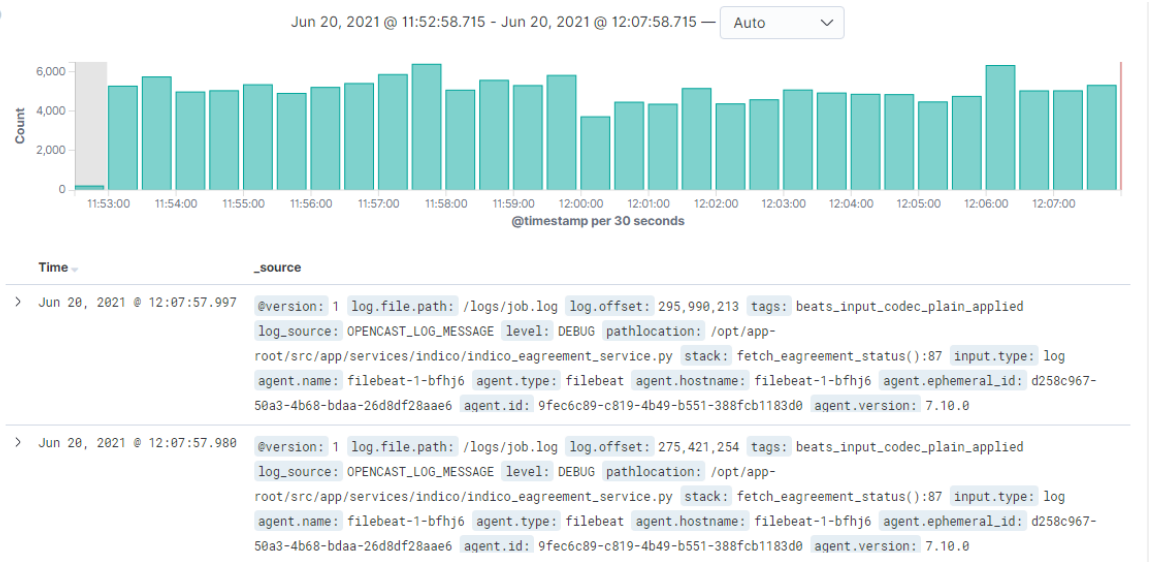
http://linuxsoft.cern.ch/mirror/repos.influxdata.com/

https://avsys-monit.web.cern.ch/

Other tools: Kibana.

Opencast logs

2021-06-17T14:13:52,276 | INFO | (ComposerServiceImpl:543) | Starting parallel encode with profile adaptive-720p.http with job load 2
2021-06-17T14:13:55,554 | INFO | (EncoderEngine:217) | Executing encoding command: [ffmpeg, -nostdin, -nostats, -i,
/mnt/opencast_share/opencast_data/9/workspace/https_ocweb-prod.cern.ch/files/mediapacl
1adf201368db/camera.mp4, -c:v, libx264, -crf, 23, -maxrate, 1200k, -bufsize, 2400k, -max_mux
x264opts, keyint=25:min-keyint=25:no-scenecut, -movflags, +faststart, -filter:v, scale=1280:-2:f
-b:a, 64k, -ac, 1, -f, mp4, /mnt/opencast_share/opencast_data/9/workspace/https_ocweb-proc
c1ea42ab572a/e473f953-3d83-4508-8b1b-1adf201368db/camera_5cea50e0-9f24-4833-83e8-
2021-06-17T14:13:55,766 | INFO | (EncoderEngine:460) | Identified output file /mnt/opencas
prod.cern.ch/files/mediapackage/f8e21072-7c1e-47e2-a7d7-c1ea42ab572a/e473f953-3d83-45
1368db/camera_5cea50e0-9f24-4833-83e8-e5274c564499-preview.mp4



PyCast logs

021-05-08 04:06:43,862 | INFO | (main:2996799) | Contribution indico_id format: 1001692c21
2021-05-08 04:06:43,875 | INFO | (main:2996799) | Get ACL for: 2021/1001692c21.conf
2021-05-08 04:06:43,881 | INFO | (main:2996799) | Gen file: /mnt/media_share/media_data/
2021-05-08 04:06:43,905 | INFO | (main:2996799) | Saving file: 1001692c21.conf
2021-05-08 04:06:43,908 | DEBUG | (log:2996799) | Media call: media.py -c /etc/pycast/pycas
/mnt/media_share/media_data/2021/config -conf_type closed
ir
2021-05-08 04:06:43,940 | DEBUG | (log:2996790) | Opencast call to: indico.py -c /etc/pycast/pycast.cfg -action acl -ev_id be0700ba-df0d-42d5-a012-a9b2de7d0b93 -conf_type closed -s acl
2021-05-08 04:07:12,213 | INFO | (main:2996819) | Indico ID:1001692 Contribution ID:1001692c21 & Series ID: through provided event_id and /api/events/#{id}/metadata endpoint
2021-05-08 04:07:12,213 | INFO | (main:2996819) | Notification of Opencast action: processing-finished to CES
2021-05-08 04:07:12,266 | DEBUG | (log:2996819) | Opencast call to: indico.py -c /etc/pycast/pycast.cfg -action notify -ev_id be0700ba-df0d-42d5-a012-a9b2de7d0b93 -s processing-finished

CES logs

2021-06-20 12:11:40,873 | DEBUG | job.speaker_releases | Contribution type is conference | /opt/app-root/src/app/services/indico/indico_eagreement_service.py | fetch_eagreement():87

https://es-collaborativeapps7.cern.ch/kibana/

Thank you!

Questions?

www.cern.ch