



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

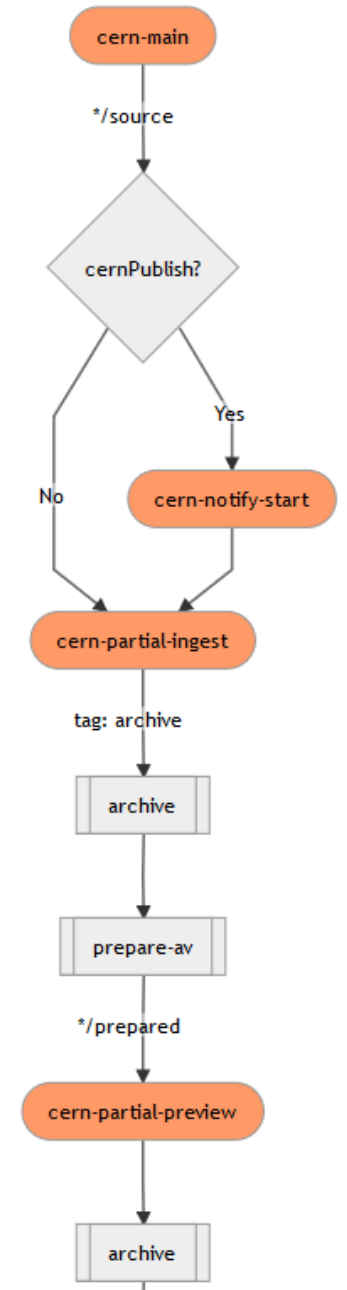
- **Mediapackage:**
 - A mediapackage consists on a XML definition for a Opencast Event and related assets (multimedia files and metadata files).
- **Asset:**
 - Mediapackage element (files): such as .mp4 video, dublincore xml, vtt.
- **Workflow:**
 - Workflows are the central element to define how a 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.

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>
      </configurations>
    </operation>
    ...
    <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>
  </operations>
</definition>
```



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

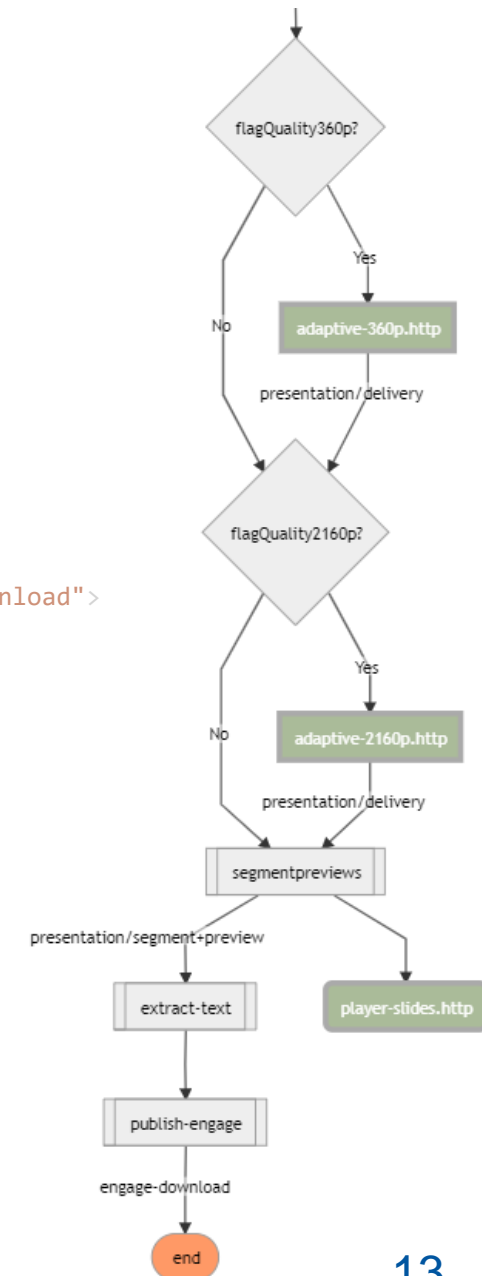
The screenshot shows two directory views. The top view is for /mnt/master_share/master_data/2021/ and lists several subdirectories including 1001692c10. The bottom view is for /mnt/master_share/master_data/2018/648945c29/ and lists two files: presenter.mp4 (761,956 KB) and presentation.mp4 (761,688 KB).



Workflows: cern-partial-publish

- Encoding operations.

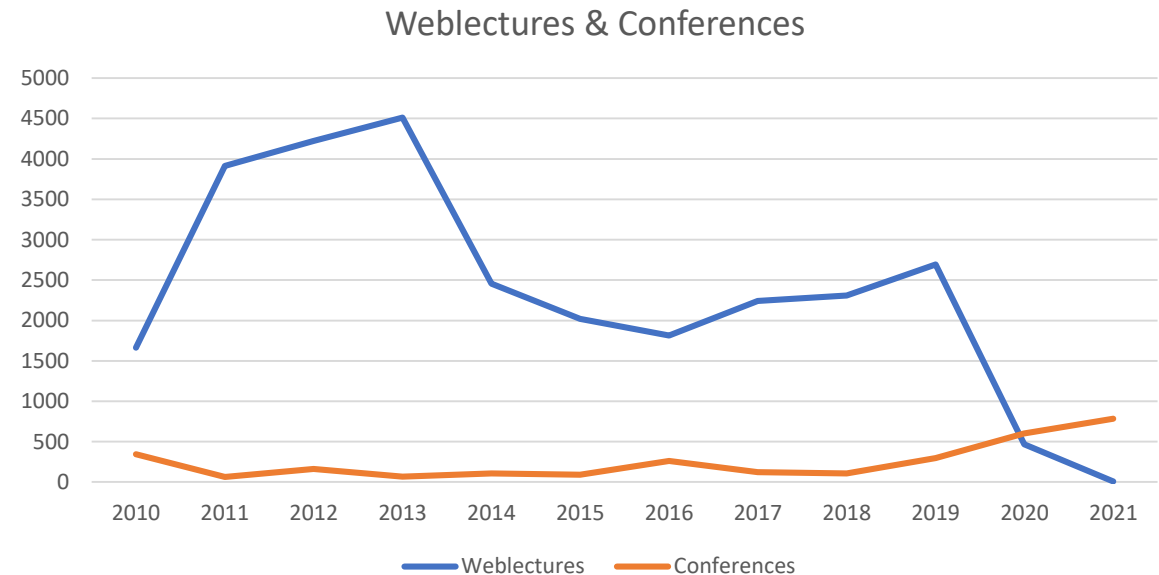
```
<operation
  if="{flagQuality360p} AND NOT {flagQualityMultiple} AND NOT {flagQualityParallel}"
  id="encode"  exception-handler-workflow="cern-encoding-error"  description="Encoding 360p video to MP4 download">
  <configurations>
    <configuration key="source-flavor">*/themed</configuration>
    <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.

Event details - Study of exotic states

Metadata Metadata II Publications Assets Workflows Access policy Comments

ID	Title	Submitter	Submitted	Status	Actions
68762	Transcription: Download and export transcriptions from Server.	mvaleron	May 21, 2021 5:10:12 PM	Succeeded	Details >
31258	Media: Export media files to legacy filesystem.	mvaleron	May 11, 2021 12:06:01 AM	Succeeded	Details >
30983	Indico: Encode and export to CEPH File System	opencast_system_account	May 10, 2021 11:36:32 PM	Succeeded	Details >

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

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

Study of exotic states

Observation of $\Lambda_b^0 \rightarrow X(3872)pK^-$ [JHEP 09 (2019) 028]

- Using LHCb data corresponding to 1.0, 2.0 and 1.9 fb⁻¹ at $\sqrt{s} = 7, 8$ and 13 TeV
- First observation of $\Lambda_b^0 \rightarrow X(3872)pK^-$ with $X(3872) \rightarrow J/\psi\pi^+\pi^-$
 - ✓ Signal yield determined from 2D fit of $M(J/\psi\pi^+\pi^-pK^-)$ and $M(J/\psi\pi^+\pi^-)$
 - ✓ $N_{\text{signal}} = 55 \pm 11$ with significance $> 7\sigma$
 - ✓ pK^- spectrum studied: $(58 \pm 15)\%$ proceed via $\Lambda_b^0 \rightarrow X(3872)\Lambda(1520)$

$$\frac{\mathcal{B}(\Lambda_b^0 \rightarrow X(3872)pK^-)}{\mathcal{B}(\Lambda_b^0 \rightarrow \psi(2S)pK^-)} \times \frac{\mathcal{B}(X(3872) \rightarrow J/\psi\pi^+\pi^-)}{\mathcal{B}(\psi(2S) \rightarrow J/\psi\pi^+\pi^-)} =$$

and its branching fraction is measured relative to that of lambda B to X thirty-eight seventy-two P K times the relative branching fraction of the Jpsi pi pi decays

and using the CMS data corresponding to one hundred and forty inverse femtobarn at thirteen TeV the B S to X thirty-eight seventy-two phi decay was observed

as signal yield is determined from the two D fit of the J psi pi pi mass and K K mass

21/06/2021



Opencast admin web UI.

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

The screenshot shows the Opencast admin web UI. At the top, there's a navigation bar with the Opencast logo and a user profile. Below that, a dashboard shows statistics for Today, Scheduled, Recording, Running, Failed, Todo, and Finished events. The main content area is titled "Events" and shows a table of 19 rows of events. The table has columns for Title, Presenter(s), Series, Date, Start, Stop, Location, Published, Status, and Actions. The events listed are conferences and a congress, all with a status of "Finished".

Title	Presenter(s)	Series	Date	Start	Stop	Location	Published	Status	Actions
Conference 1-0	Miguel Angel Valero Navarro	TEST Category	6/14/21	2:00 PM	2:30 PM	External URI	Yes	Finished	[Icons]
Conference 1-1	Miguel Angel Valero Navarro	TEST Category	6/14/21	2:30 PM	2:50 PM	External URI	Yes	Finished	[Icons]
Conference 1-2	Ruben Domingo Gaspar Aparicio	TEST Category	6/14/21	2:50 PM	3:00 PM	External URI	Yes	Finished	[Icons]
Conference 1-3	Rene Fernandez Sanchez	TEST Category	6/14/21	3:00 PM	3:10 PM	External URI	Yes	Finished	[Icons]
Conference 1-4		TEST Category	6/14/21	3:10 PM	3:20 PM	External URI	Yes	Finished	[Icons]
Conference 1-5	Miguel Angel Valero Navarro	TEST Category	6/14/21	3:20 PM	3:30 PM	External URI	Yes	Finished	[Icons]
Conference 1-6		TEST Category	6/14/21	3:30 PM	3:40 PM	External URI	Yes	Finished	[Icons]
Conference 1-8		TEST Category	6/14/21	3:50 PM	4:00 PM	External URI	Yes	Finished	[Icons]
Conference 1-9		TEST Category	6/14/21	4:00 PM	4:10 PM	External URI	Yes	Finished	[Icons]
Congress 1-1	Miguel Angel Valero Navarro	TEST Category	6/15/21	1:00 PM	1:20 PM	External URI	Yes	Finished	[Icons]

The screenshot shows the Opencast video editing interface. At the top, there's a video player showing a video with a text overlay. Below the player, there's a timeline with a zoom level slider. The timeline shows a video segment with a red bar indicating a segment. Below the timeline, there's a "Segments" section with a table of segments. The table has columns for Start and End times. The segments listed are:

Start	End
00:00:00.000	00:07:52.160
00:07:52.160	01:53:23.767
01:53:23.767	02:01:42.042

- 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.

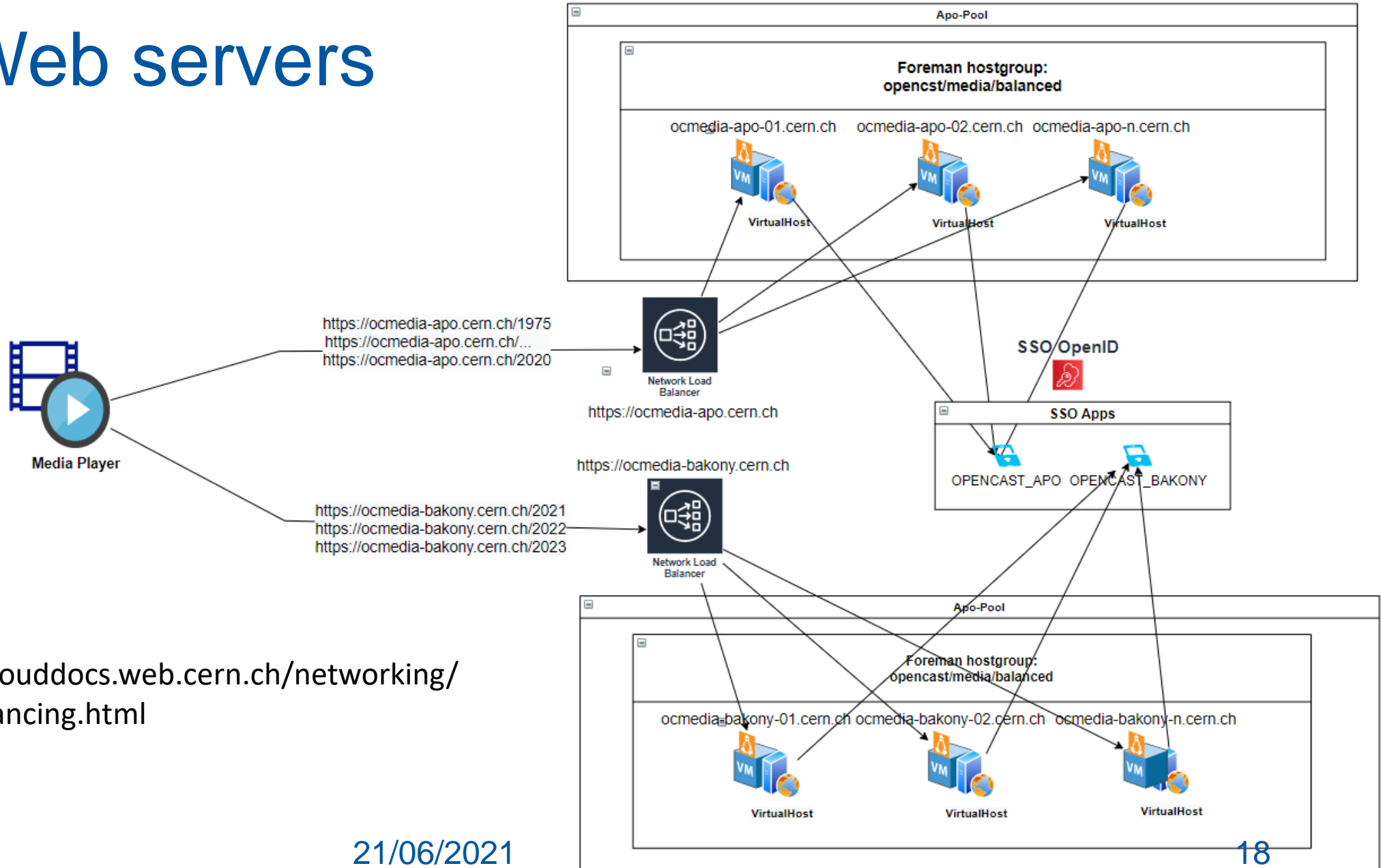


```
<VirtualHost *:443>
  ServerName <%= @balancer %>
  ...
  # Protect any directory not explicitly exposed
  <Directory <%= @web_path %>/ * >
    Require all denied
  </Directory>

  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.

> Applications > My Application

Application: Opencast Media Bakony [2021-2035]

Application details SSO Registration **Roles** Group memberships

Name	Role Identifier	Description	Required?	Multifactor?
Default Allowed Users	default-role	Users must be from CERN or eduGAIN to have access	✓	
admins	opencast-admins	Opencast administrators		
indico-data-managers	indico-data-managers	Role created automatically		
cern-accounts-primary	cern-accounts-primary	Role created automatically		
internal-audit-staff	internal-audit-staff	Role created automatically		
atlas-computing-areas-coordinators	atlas-computing-areas-coordinators	Role created automatically		
indico-atlas-managers	indico-atlas-managers	Role created automatically		
atlas-ecsb	atlas-ecsb	Role created automatically		

```

<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>

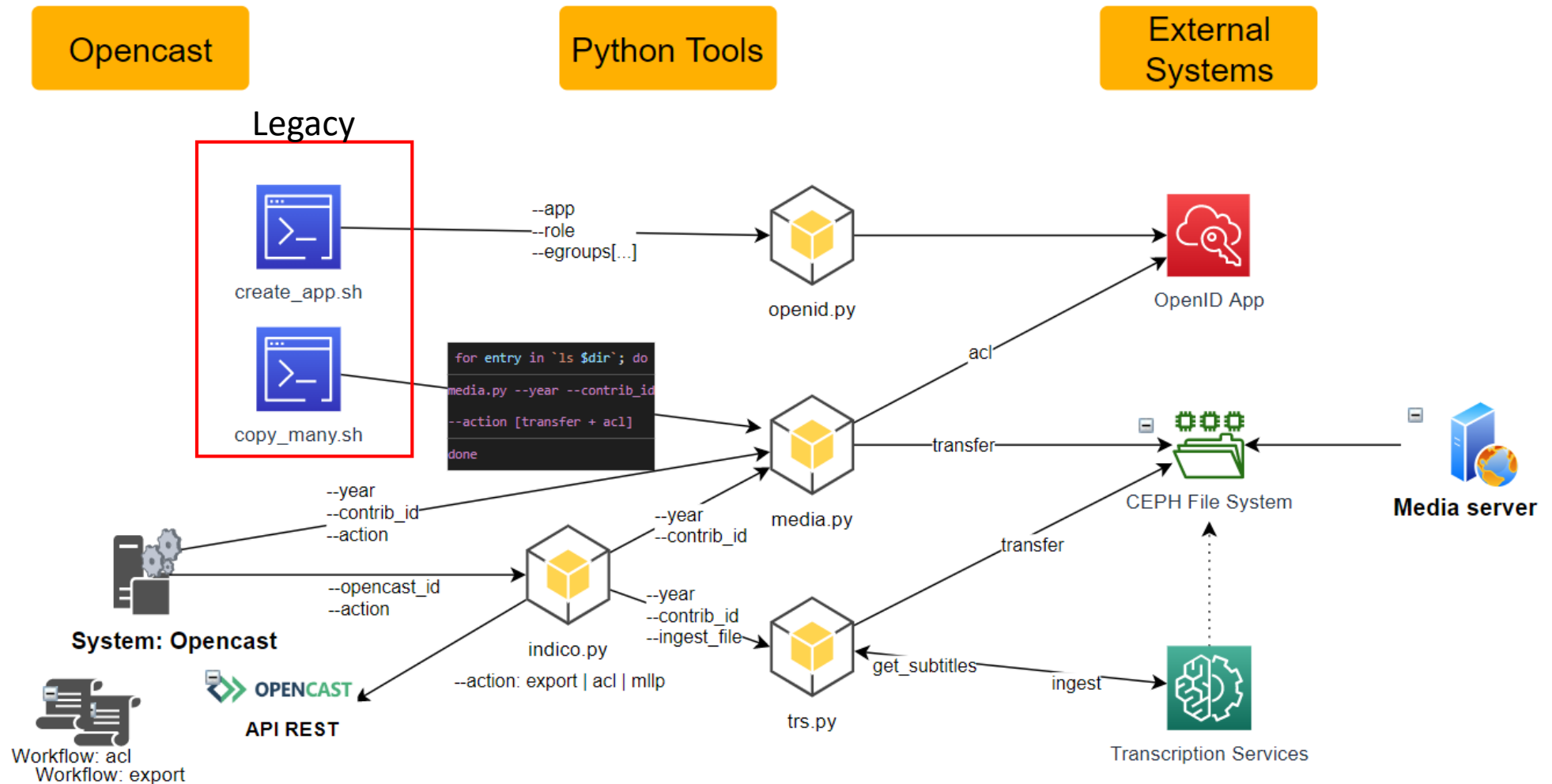
```



21/06/2021



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 Preveessin), ...

The screenshot shows the Rundeck interface for the 'Opencast Media Servers' project. The left sidebar contains navigation options: PROJECTS, DASHBOARD, JOBS, NODES, COMMANDS, ACTIVITY, WEBHOOKS, and PROJECT SETTINGS. The main area displays a list of 16 jobs under 'All Jobs'. A blue arrow points from the 'PyCast Tools: Update Production' job in the list to the 'Edit Job' window. Another blue arrow points from the 'S3 backup - prod - master' job to the script editor. The script editor contains the following code:

```
1 sudo /usr/local/bin/s3_backup.py -l /mnt/media_share/media_data -s -b public -v s3:s3preveessin
2
```

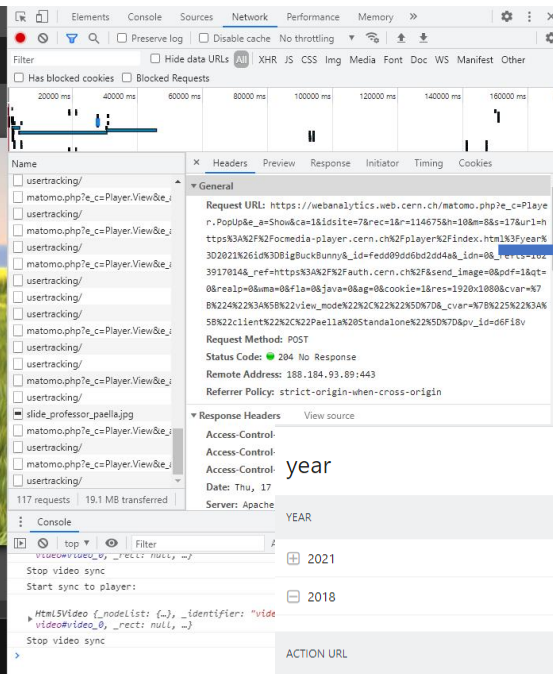
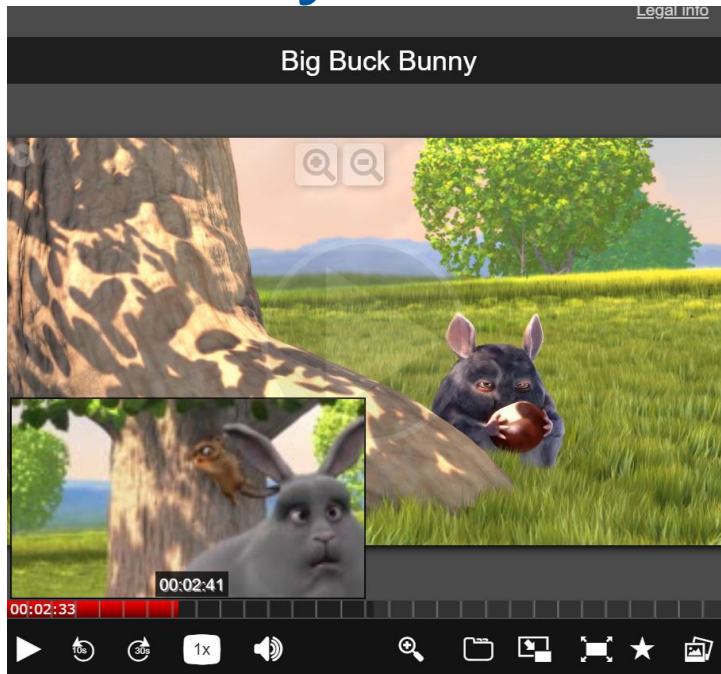
The 'Edit Job' window for 'PyCast Tools: Update Production' is shown. It includes a script editor with the command: `sudo pip3 install git+https://gitlab.cern.ch/webcast/pycast.git@latest`. Below the script editor are tabs for 'Details', 'Workflow', 'Nodes', 'Schedule', 'Notifications', and 'Other'. The 'Nodes' tab is active, showing options for 'Dispatch to Nodes' (selected) and 'Execute locally'. A 'Node Filter' is set to `/((ocworker-(prod)-[0-9]+))ocweb-prod/`. The 'Matched Nodes' section lists 9 nodes: ocweb-prod, ocworker-prod-03, ocworker-prod-06, ocworker-prod-01, ocworker-prod-04, ocworker-prod-07, ocworker-prod-02, ocworker-prod-05, and ocworker-prod-08.

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

https://clouddocs.web.cern.ch/object_store/s3cmd.html
21/06/2021



6. Matomo analytics: Visits and anonymous user actions.



Event Actions	
EVENT ACTION	EVENTS
Seek	50
resize	43
Play	30
EVENT NAME	
EVENT NAME	EVENTS
Introduction and welcome	9
Big Buck Bunny	6
Opencast Lecture 1	5
Capture agent 1-5	2
Conference 1-2	2

YEAR	ACTIONS	UNIQUE ACTIONS	BOUNCE RATE	AVG. TIME ON DIMENSION	EXIT RATE
2021	136	22	27%	00:00:12	86%
2018	4	1	0%	00:00:01	100%

ACTION URL	ACTIONS	UNIQUE ACTIONS	BOUNCE RATE	AVG. TIME ON DIMENSION	EXIT RATE
ocmedia-player.cern...	4	1	0%	00:00:01	100%

```

"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"
},

```

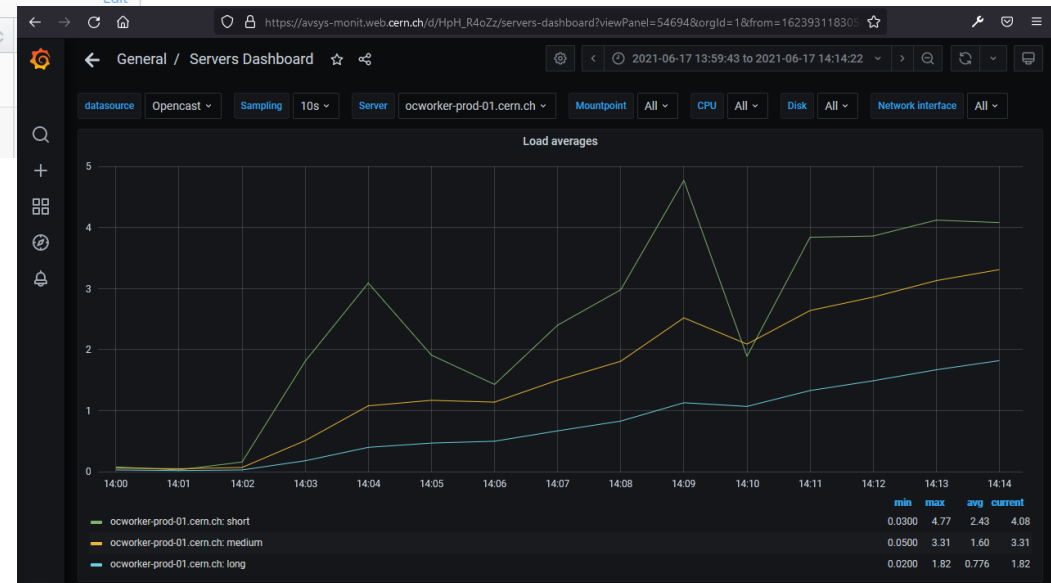
Thanks Pete Jones (IT-CDA-WF) for registering the Sites.



Other tools:Telegraf.

The screenshot shows the OpenCast web interface. At the top, there are statistics for Today, Scheduled, Recording, Running, Failed, Todo, and Finished. Below this is a table of events, with one event 'Conference 1-7' in a 'Running' state. Below the events table is a table of servers, showing two servers: 'ocweb-prod' and 'ocworker-prod-01'. A blue arrow points from the 'Jobs running' column of the servers table to the 'Event details' window.

The screenshot shows the 'Event details - Conference 1-7' window. It displays a table of workflow operations with columns for Status, Title, and Description. The operations include 'defaults', 'include', 'mattermost-notify', 'snapshot', 'encode', 'composite', 'waveform', and 'image'. A small dashboard in the top right corner shows metrics for Processes (253), Threads (780), CPU usage (18.3%), and RAM usage (13.5%).



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

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

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



21/06/2021

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/mediapack

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-prod

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/e473f953-3d83-4!

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

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

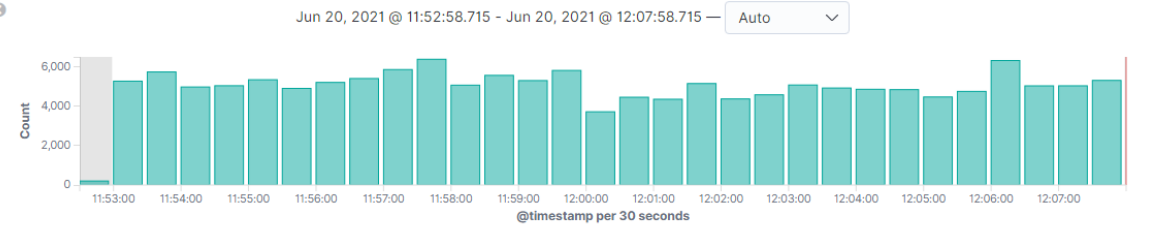
2021-05-08 04:07:12,213 | INFO | (main:2996819) | Indico ID:1001692 Contribution ID:1001692c21 & Series ID: through provided event_id and /api/events/#{id}/metadata endpoint

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



```
Time - _source
> Jun 20, 2021 @ 12:07:57.997 @version: 1 log.file.path: /logs/job.log log.offset: 295,990,213 tags: beats_input_codec_plain_applied
log_source: OPENCAST_LOG_MESSAGE level: DEBUG pathlocation: /opt/app-
root/src/app/services/indico/indico_eagreement_service.py stack: fetch_eagreement_status():87 input.type: log
agent.name: filebeat-1-bfhj6 agent.type: filebeat agent.hostname: filebeat-1-bfhj6 agent.ephemeral_id: d258c967-
50a3-4b68-bdaa-26d8df28aae6 agent.id: 9fec6c89-c819-4b49-b551-388fcb1183d0 agent.version: 7.10.0
> Jun 20, 2021 @ 12:07:57.980 @version: 1 log.file.path: /logs/job.log log.offset: 275,421,254 tags: beats_input_codec_plain_applied
log_source: OPENCAST_LOG_MESSAGE level: DEBUG pathlocation: /opt/app-
root/src/app/services/indico/indico_eagreement_service.py stack: fetch_eagreement_status():87 input.type: log
agent.name: filebeat-1-bfhj6 agent.type: filebeat agent.hostname: filebeat-1-bfhj6 agent.ephemeral_id: d258c967-
50a3-4b68-bdaa-26d8df28aae6 agent.id: 9fec6c89-c819-4b49-b551-388fcb1183d0 agent.version: 7.10.0
```

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

21/06/2021

27



Thank you!

Questions?



www.cern.ch