

Control Interface for the Digital Memory Platform

CERN openlab Summer Student Lightning Talks

Marco Donadoni

06/09/2021

Digital Memory Project

Introduction

- Large amount of data generated by CERN that needs to be preserved
 - Research Open Datasets
 - Digital documents (e.g. published papers, reports)
 - Video recordings of lectures and meetings
- Right now, this content is made accessible by multiple information systems
 - CERN Document Server (CDS), CERN Open Data, Zenodo...
- Creation of Digital Memory Platform, an OAIS-compliant digital archive
 - «Open Archival Information System»
 - Single, standardized, long-term archival service
 - Can be used by CERN staff to request the archival of specific resources



Main functions

- Harvesting
 - Fetch data/metadata of a specific record from existing information systems
 - Create a SIP (Submission Information Package)
- Ingestion
 - Generate AIP (Archival Information Package) from SIP
 - Data normalization (e.g. format conversion)
- Long Term Storage
 - Store the ingested archive in a long-term storage facility
 - CERN Tape Archive
- Access to the Archived Content
 - Users must be able to access only content they requested to be archived



Components of the Platform

During the summer programme, we focused on two components

- Platform API
 - Core component of the system
 - Manages all the operations supported by the digital archive
 - Exposes a REST API that can be used to control the Platform
 - Django REST Framework + Celery
- Control Interface
 - Web Interface accessible using a browser
 - Single Page Application
 - React + Bootstrap



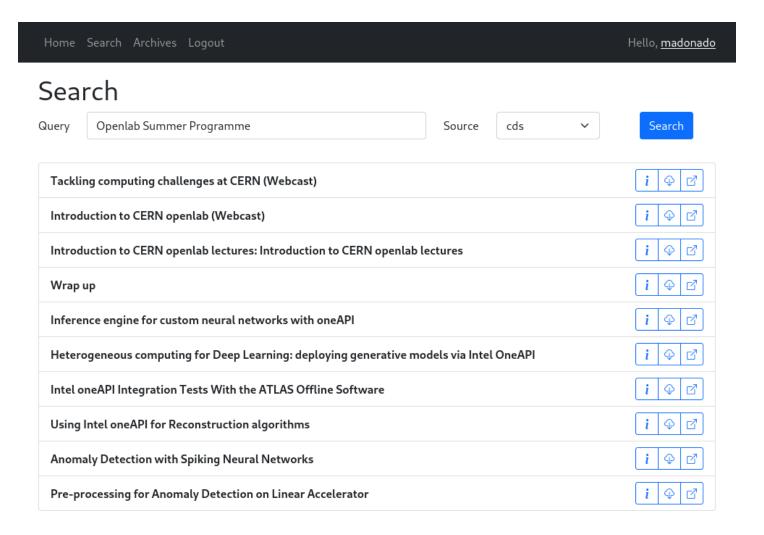
New features

- Login using CERN credentials
- Search for specific records on existing information systems
 - CERN Document Server, CERN Open Data, Zenodo, ...
- Handle harvesting requests
- Approval system for harvesting requests
- Export and ingestion of data and metadata from external systems
 - Using the bagit-create tool developed at CERN
- Permission system to limit the actions available to the user
 - e.g. approval and rejection of harvesting requests



Control Interface

Records Search





Control Interface

Archival Requests

Home Search Archives Logout Hello, admin

Archives

ID	Record	Creator	Creation Date	Status	Actions
6	2737252 (cds)	madonado	2021/08/13 16:42:51	Waiting for Approval	✓ ×
5	2737253 (cds)	madonado	2021/08/13 16:03:44	Pending	
4	2737254 (cds)	madonado	2021/08/13 14:27:37	In progress	
3	2774412 (cds)	madonado	2021/08/13 11:14:23	Completed	
2	2775200 (cds)	madonado	2021/08/13 10:35:12	Failed	
1	2775216 (cds)	madonado	2021/08/13 09:55:56	Rejected	





Deployment

- The Platform is hosted on CERN OpenShift
 - «Platform-as-a-Service» based on Kubernetes
 - Automatic management of resources (VMs, persistent volumes, ...)
- Deployment configuration is described by YAML files
 - Managed by Helm
- Continuous Integration and Deployment
 - Tests are run automatically at each code change
 - Changes are deployed if the tests pass
 - Thanks to GitLab CI/CD pipelines





marco.donadoni@cern.ch

https://www.linkedin.com/in/marco-donadoni

https://gitlab.cern.ch/digitalmemory

