OLOS.swiss

Pierre-Yves Burgi
Division of Information Technologies
University of Geneva
Geneva, Switzerland
ORCID 0000-0002-4956-9279

Hugues Cazeaux
Division of Information Technologies
University of Geneva
Geneva, Switzerland
ORCID 0000-0002-5618-2670

Andréé Jelicic
Division of Information Technologies
University of Geneva
Geneva, Switzerland
ORCID 0000-0002-6687-1373

Abstract— *During the P-5 program (2019-2020), the DLCM1 services transitioned from pilot to operational status: a data management training program and ad-hoc support were delivered throughout Switzerland, and OLOS (olos.swiss), the long-term solution for research data archiving and publication, became operational. Yareta, powered by the same DLCM technology than OLOS, was launched in June 2019, serving all the Higher Education Institutions of the Geneva Canton. Thanks to OLOS, the next step is to launch in January 2021 an equivalent service at the National level.*

Keywords—research data, archiving, OAIS, architecture, digital preservation.

I. INTRODUCTION

« OLOS » is the name[7] given to the long-term solution for research data archiving and publication, intended to be deployed at Swiss level. OLOS, which is issued from the Swiss DLCM project[8] (Burgi, Blumer, & Makhlouf-Shabou, 2017; Burgi & Blumer, 2018), differentiates itself from other FAIR repositories in that it minimizes the constraints for its customers. Its architecture is highly flexible making it suitable to any kind of research environment. These features are detailed in Section 2. Section 3 presents the OLOS organization with information on how to adhere to it. Final conclusions in Section 4 provide some perspectives on OLOS' future developments.

II. CONCEPTS

A. Architecture

OLOS's architecture is open, modular and scalable. It can integrate to any active research data management (ARDM) solution, adapt to any metadata scheme, and type of storage such as tape, SSD, file systems, and object storage (Burgi, Cazeaux, & Echernier, 2019). The main competitive advantage of OLOS thus comes from its modular and distributed architecture, and its strict compliance to the ISO 14721 (2012) OAIS reference model (Fig. 1). To the three standard packages: submission information package (SIP), archival information package (AIP), and dissemination information package (DIP), we added a pre-ingestion module to facilitate data transition between ARDM and archiving. To further, allow easy interconnections with any research environment, the whole architecture is based on various open and international standards, such as REST for web services, DataCite metadata schema for default metadata, OAI-PMH for exchanging metadata, DOI for sustainable references, and ORCID for unambiguous authorships. The user interface is developed in Angular and has been subjected to user experience design to be more efficient and enhance its, which means entire or whole. intuitive use of the numerous offered functionalities.

[7] OLOS' name takes its origin from the Greek word "Holos", which means entire or whole.
[8] The DLCM project was mandated by swissuniversities.

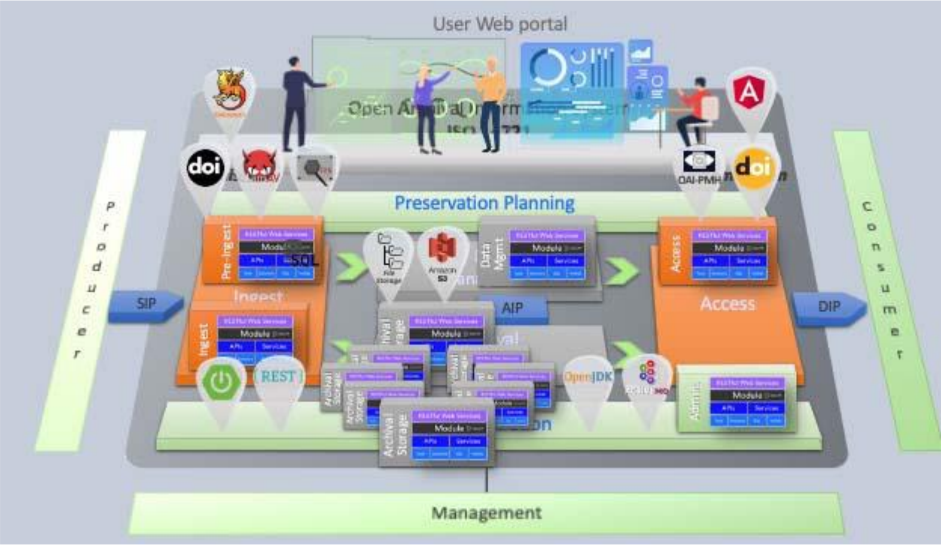

Fig. 1. Architecture of OLOS

The core of OLOS (a.k.a. "backend") is written in JAVA and relies also on several opensource modules such as FITS for automatic format identification, ClamAV for virus checking, S3 for object storage, Elasticsearch as the search engine (based on the Lucene search engine library), and Shibboleth for the authentication, among others (Fig. 1). The whole architecture can be deployed either on premise, fully in the cloud, or a mix of these two modalities; for instance, the SIP and DIP could be deployed in the cloud, while the AIP would be installed on premise to confer more control on long-term preservation to the host institution. Yareta (yareta.unige.ch), powered by the same DLCM technology than OLOS, was launched in June 2019 and successfully operated since then by the University of Geneva to serve all the Higher Education Institutions of the Geneva Canton (Burgi, 2019), which provided a basis for a better understanding of the unique context of each individual research (Bezzi, 2020).

B. *Institutional Benefits*

OLOS is agnostic to the hardware infrastructure. Provided by default on Software as a Service (SaaS) mode as a generic repository suited for research data of any discipline, client institutions can use it without any prior investment. The portal is natively connected to existing storage infrastructures in Switzerland like SWITCHengines, with the aim to create economies of scale at the national level in order to lower the overall research data preservation costs. In some cases (i.e., very large datasets, sensitive information, excess of storage capacity, etc.), a client institution may want to connect its own storage infrastructure to OLOS. Beside the benefits that local data storage may procure, such integration to a compliant preservation service like OLOS would entitle these institutions to recover, through grants, the costs incurred for storing one copy of the archived research datasets for the entire preservation period (usually 15-20 years). Furthermore, OLOS differentiates itself from other FAIR repositories by allowing each research institution to define, implement and monitor its own preservation policy regarding, for instance, the number of copies, the duration of preservation, the copyright licenses, an eventual validation workflow, etc. in accordance to its infrastructure, and its specific institutional characteristics and constraints. A dashboard allows the monitoring of the different phases of the dataset archiving process, and provides key indicators for higher-level data management and research impact assessment.

C. *Key Features For Researchers*

The pre-ingest module provides high flexibility in data management by offering researchers the possibility to manipulate the datasets before final submission. Pre-ingestion thus comes after the ARDM phase, which usually involves intensive data manipulations, but comes before the archiving phase, which prevents any further modifications. Any postarchiving data modification would imply either a new archived dataset, or when permitted, a new version of the original dataset. In OLOS, versioning is not possible, but can be substituted by the use of collections, which could regroup several successive versions of the original dataset. Another key feature akin to the OLOS' modular architecture is the possibility to access all functionalities (deposit, download, search, etc.) from any environments able to activate web services. For instance, Jupyter Notebook connectors allow to search and retrieve data from OLOS, to subsequently process them based on a variety of languages and libraries. For large data volumes (over 100 GB) or high number of files (over hundred), we provide assistance to the researchers so that ingestion is automatized through batches making use of web services. Fig. 2 illustrates such a process involving large volumes.

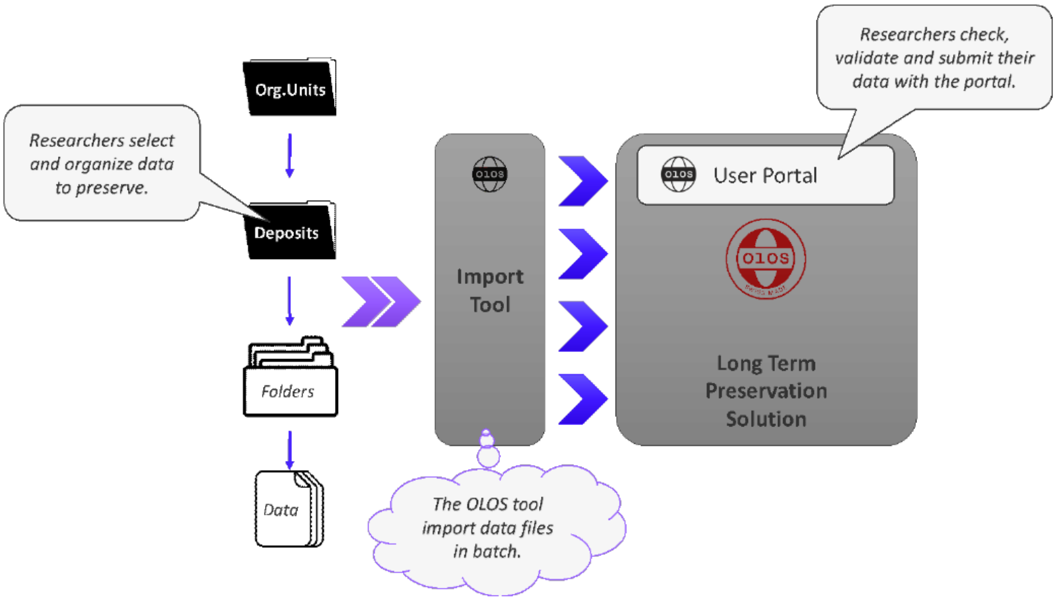
Researchers check, validate and submit their data with the portal.
Researchers select and organize data to preserve.
Org.Units
Deposits
Folders
Data
Import Tool
User Portal
OLOS
Long Term Preservation Solution
The OLOS tool import data files in batch.

Fig. 2. Ingest process for large data volumes

An additional key feature stems from a concept very specific to OLOS: the organizational units. Datasets are organized within units whose granularity can be set at the project, laboratory, department, or institutional level. Such an organization can be a powerful instrument to monitor key indicators (see previous subsection B), and is also convenient to logically structure a lot of datasets. Finally, predefined roles (Fig. 3) provide the possibility to define different user groups, for instance giving the rights to co-authors to edit the dataset while restricting to viewing only for a specific range of users (e.g., visitors). The roles also make possible to setup a quality check, performed either by managers, stewards, or approvers through a workflow. The activation of such a workflow remains optional, and would not make sense if the institution/department/laboratory has no data quality strategy.

III. ORGANIZATION

OLOS is a non-profit association with headquarters in Switzerland, governed by institutional members of the research community. The OLOS association relies on several streams of revenue to avoid a long term dependance on public subsidies. The first one is a preservation fee charged on a per project basis to researchers. Its amount depends on the volume of data associated to preserve for the project, the number of copies (2 is recommended at minimum) and the preservation duration. Default preservation plans are 5, 10, and 15 years. If requested, a special quote can be issued to preserve the dataset for-ever. This is made possible thanks to a close collaboration with industry experts monitoring the developments of storage technologies.

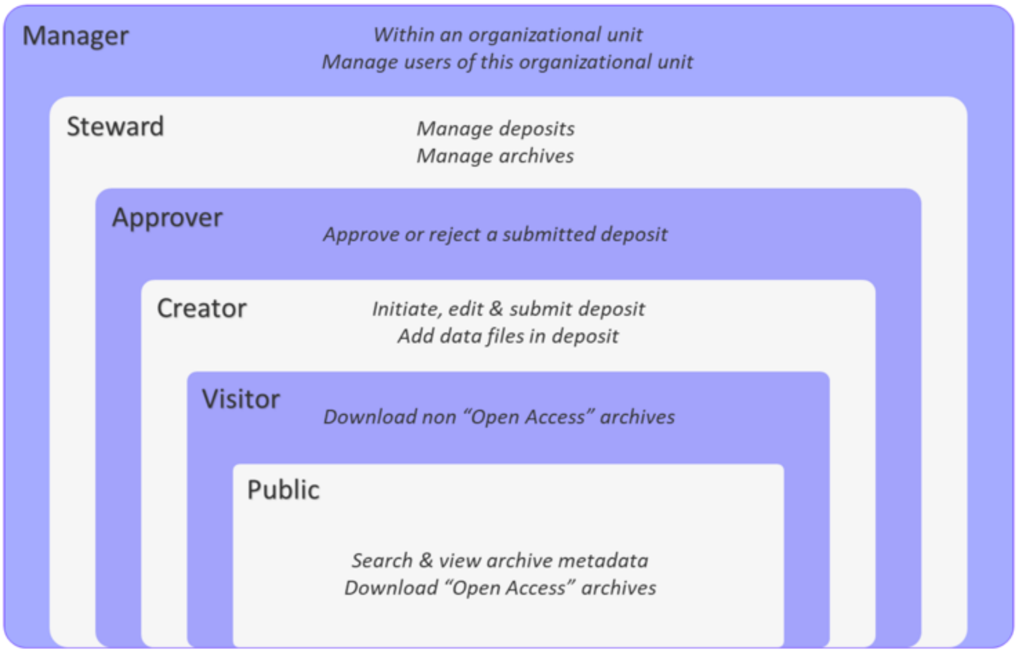
Manager
Within an organizational unit
Manage users of this organizational unit
Steward
Manage deposits
Manage archives
Approver
Approve or reject a submitted deposit
Creator
Initiate, edit & submit deposit
Add data files in deposit
Visitor
Download non "Open Access" archives
Public
Search & view archive metadata
Download "Open Access" archives

Fig. 3. Predefined roles available in OLOS

Annual memberships are another pillar of OLOS financial sustainability. Several membership categories (i.e., bronze, silver or gold) are available to institutions to better suite their needs. Depending on its category, a member institution can, for instance, influence future developments of the portal, suggest further integrations to ARDM solutions or other tools used by its researchers, vote at the general assembly, or ensure a higher level of support for its researchers. To ensure high storage reliability OLOS has partnerships with trusted infrastructures and geographically distant providers, in Switzerland for more security. The integration of storage infrastructures abroad is planned to better suite the needs of international research projects.

IV. CONCLUSION & PERSPECTIVES

OLOS conception represents one of the main outcomes of the DLCM project, which benefited from the expertise of many librarians and IT professionals in the field of data management (Burgi, Blumer, & Makhlouf-Shabou, 2017). Since January 2019 (phase 2 of the DLCM project), we transitioned from a prototype to a functional long-term preservation service, whose technology has already proven itself at cantonal level with the instance called Yareta. OLOS is thus the logical step to extend the offer at National level and is due to operate starting in January 2021.

More than just a tool, OLOS is a service intended to help researchers better manage and organize their data from within their research environments. Next on the roadmap is the Fig. 3. Predefined roles available in OLOS Fig. 2. Ingest process for large data volumes development of the preservation planning module and new dashboard functionalities to provide institutions with fuller control on their assets.

ACKNOWLEDGMENT

OLOS was conceived and developed within the swissuniversities P-5 programme. We are grateful to all DLCM partners who contributed directly or indirectly to this product.

REFERENCES

Bezzi, M. (2020). Préservation des données de recherche : proposer des services de soutien aux chercheurs du site Uni Arve de l'Université de Genève. Mémoire de master : Haute école de gestion de Genève.

Burgi, P.-Y., Blumer, E., & Makhlouf-Shabou, B. (2017) Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs. *IFLA Journal 43*. doi: 10.1177/0340035216678238

Burgi, P.-Y. & Blumer, E. (2018). Le projet DLCM : gestion du cycle de vie des données de recherche en Suisse. In A. Keller & S. Uhl (Eds.), *Bibliotheken der Schweiz: Innovation durch Kooperation. Festschrift für*

Susanna Bliggenstorfer anlässlich ihres Rücktrittes als Direktorin der Zentralbibliothek Zürich (pp. 235-249). Berlin : De Gruyter. doi: 10.1515/9783110553796

Burgi, P.-Y, Cazeaux, H., & Echernier, L. (2019) A versatile solution for long-term preservation of research data: Data Life-Cycle Management: the Swiss Way. In: *iPRES - 16th International Conference on Digital Preservation*. Amsterdam (The Netherlands).

Burgi, P.-Y. (2019) Le Projet de Loi 12146 : Infrastructures et services numériques pour la recherche. *Revue électronique suisse de science de l'information, 20*. http://www.ressi.ch/num20/article_168

ISO 14721:2012 (2012). Space data and information transfer systems - Open archival information system (OAIS) – Reference model