eTDR@CINES – Business model

Lead Partner:	CINES
Version:	1.0
Status:	Final
Dissemination Level:	Public
Document Link:	https://documents.egi.eu/document/3704

Deliverable Abstract
This document details the business model for the long-term preservation service which has recently been added to the EOSC catalogue as part of the EOSC-hub project.

COPYRIGHT NOTICE

DELIVERY SLIP

Date	Name	Partner/Activity	Date
From:	Olivier Rouchon	CINES	

DOCUMENT LOG

Issue	Date	Comment	Author
v.1	05/11/2020	First version	O.Rouchon

TERMINOLOGY

https://wiki.eosc-hub.eu/display/EOSC/EOSC-hub+Glossary

Terminology/Acronym	Definition
eTDR	European Trusted Digital Repository
FAIR	Findable Accessible Interoperable Reusable
HPC	High Performance Computing

Content

1 Business case

Researchers are required to keep the data they generated as part of their work for at least five years after publication and many funders require a minimum of ten years. Data producers have strong recommendations and incentives to follow the FAIR principles and enable findability, accessibility, interoperability and reusability of the data they produced as part of their research activity.

Institutions can use a significant budget to take care of their cultural heritage by digitizing valuable collections owned by their libraries. The materials produced as part of these programs have to be sustained to justify the investments.

Digital objects can be unstable and difficult to access or (re)use after decades in storage: in addition to a byte stream preservation service, which ensures data integrity from a technical standpoint, there is a need for quality services that preserve information over time. This requires certified (see References section), extra capabilities in the area of metadata/data stewardship, assessment and curation that goes beyond simple data management services. Such a long-term preservation strategy will ensure data can be found, understood, accessed and used in the future.
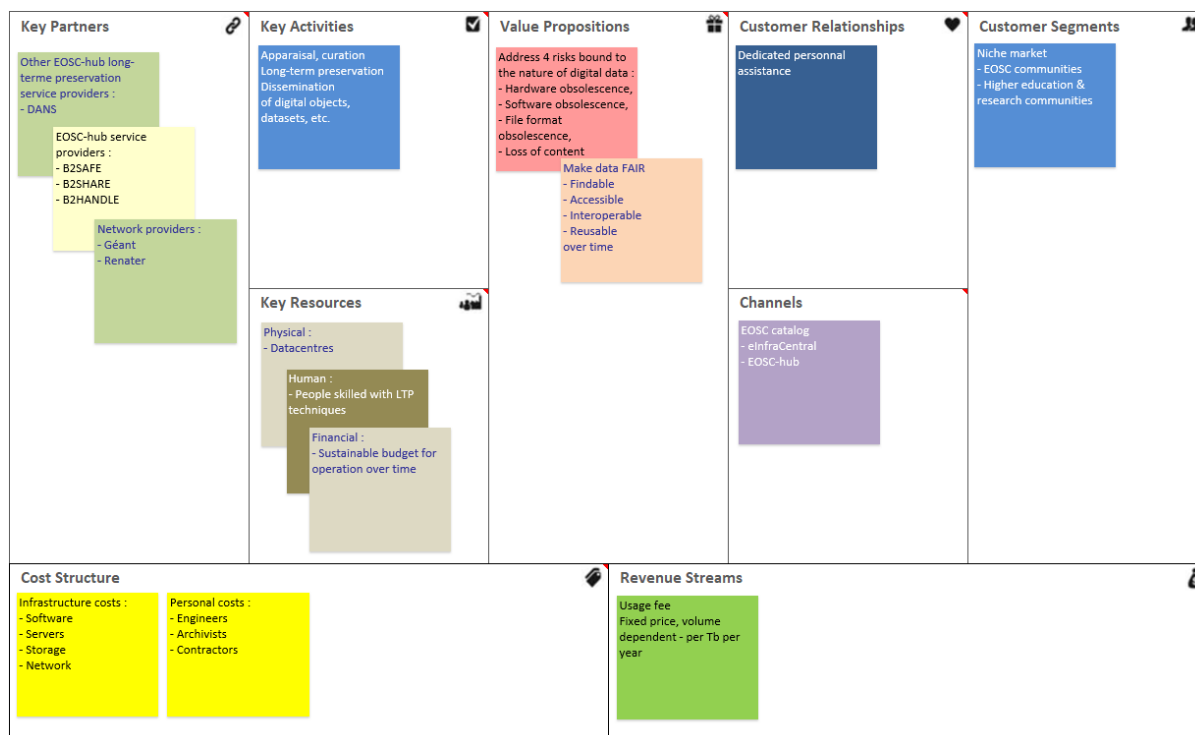
2 Business model

2.1 Overview

Figure 1. Business model canvas

2.2 Key partners

The service relies on mature, "ready-to-use" EOSC-hub/EUDAT services, which have been in production for several years, and provide data transferring functionalities. These include:

- B2SAFE to transfer datasets (digital documents, images, videos, etc.);
- B2HANDLE for PID allocation;
- B2FIND to harvest metadata and publish the catalogue of archived documents.

It also uses the infrastructure (Géant, Renater) which connects the National networks together and enables access, file transfer, etc.

Thus, these organizations can be considered as key partners to the service.

2.3 Key resources

2.3.1 Physical resources

The eTDR is hosted in the CINES premises. Since CINES is a National data centre, it takes advantage of the infrastructure which has been deployed for other services (HPC, etc.) and which is now pooled:

- Machine rooms;
- Power supply/generators;
- Cooling;
- Tape libraries.

2.3.2 Human resources

The Digital Preservation Department is staffed managers, engineers, developers and archivists with the following skills:

- People and project management;
- Network, system and application administration;
- Software development and test;
- Metadata, file formats, storage media expertise;
- Archival expertise;
- User support.

2.3.3 Financial resources

CINES is directly funded by the Ministry of Higher Education, Research and Innovation. The budget is allocated on a yearly basis to allow the provision of the three services that are included in the mandate of the data centre. In addition, other financial resources include EU grants for the contribution to European projects, and revenues from services provided at National level (hosting, long-term preservation, etc.).

2.4 Key activities

eTDR does not provide an off-the-shelf service, it is tailored based on the community needs. There are three phases when adding new collections to the repository:

1. The negotiation phase during which the community needs will be evaluated and technical agreement will be reached;
2. The deployment and operation, during which datasets will be transferred to the eTDR for preservation, and made available for the designated communities;
3. The migration, which will be triggered on an ad hoc basis.

2.4.1 Community on-boarding

- Appraisal of data collections;
- Contractual agreement: roles, responsibilities, service level;
- Technical agreement: file formats, protocols, etc.

2.4.2 Service operation

- Long-term preservation platform operation;
- Transactions monitoring: deposit, access, etc.;
- Community support, follow-up;
- Technology watching.

2.4.3 Migrations

- File format conversion;
- Storage media migration.

2.5 Value proposition

Long-term preservation implies the use of specific, dedicated resources (see Section 2.3) which communities or institutions may see as unaffordable if insourced. eTDR will allow cost-reduction and economies of scale by providing a generic, shared, effective, high-performance, scalable and secure service, which includes:

- Capacity/resource planning to guarantee provision of the service;
- Application of long-term preservation techniques/technologies, quality assurance for metadata, file format and storage media to secure comprehension, integrity, exploitation and valorisation;
- Combination of policies, processes and actions to ensure access to "born-digital" and reformatted data, regardless of the challenges of technological changes or failures.

These will mitigate the impact of risks of loss when they occur. They will focus on scientific data (results of observations, measurements, simulations, etc.) as well as cultural heritage (publications, pedagogics, digitized documents, etc.) and address the following challenges:

Challenge	Service
Loss of content knowledge	Validation of generic, community-specific metadata ; Use of persistent, unique identifiers.
File format obsolescence	Handling of a limited set of durable formats; File format identification, characterization, validation; Management of information representation ; Logical migration (format conversion).
Storage media failure	Monitoring, management of media ageing; Physical migration.
Software or hardware demise	Technology watching, anticipation of technological changes, proactivity.

These procedures are documented for repeatability and improvement purposes; they can be compared/challenged by peers, but most important of all they can be audited and certified in order to strengthen trust from user communities and stakeholders/funding bodies.

2.6 Customer relationships

The use of the long-term preservation service is fully automated: users are assigned credentials to log into the platform and upload the digital objects they wish to archive from their own information system. The ensuing workflow doesn't require any manual/human intervention. However, the on-

boarding phase as well as the user support will be fulfilled with a dedicated, personal assistance to ensure satisfaction.

All commitments will be agreed in a convention to be signed by the two parties at the beginning of the project (see section 2.4.1).

2.7 Customer segments

This is very much a niche market: the service is intended for research communities in Europe that produce digital material they wish to preserve over-time.

2.8 Channels

The service is part of the EUDAT/EOSC-hub portfolio, and as such will be exposed through partner channels: EOSC, EUDAT CDI, eInfraCentral, CatRIS, etc. It will also be promoted through conferences, seminars on the subject matter.

2.9 Cost structure

Three different cost items have been identified: hardware, software and human resources. The fluids (electricity, water, etc.) have been considered marginal and have not been taken into account. They are all reduced to the platform capacity to get an evaluation at the Terabyte level.

2.9.1 Hardware

This includes the purchase costs of application servers, storage servers (spinning disks), and tape libraries (drives, magnetic tapes); it is admitted that these investments are written off after five years for the two formers, and 10 years for the latter – thus, the costs are smoothed over these period to provide an annual amount.

2.9.2 Software

This includes the annual fees for licenses.

2.9.3 Human resources

This includes the annual budget for contractors, as well as third parties to which some activities can be outsourced. Officials/public employees are not taken into account as the service is intended for the Public sector.

2.10 Revenue stream

2.10.1 Model

The model is based on an annual fee derived from the cost structure, and includes three criteria:

- Data volume (storage space used) as this impacts the overall capacity of the system, requires additional attention during migration phases, etc.;

- Level of service (type of storage used : spinning disks – SL1, magnetic tape – SL2) as this impacts the cost of media, performance, etc.;
- Number of objects (archived packages) as it impacts the workload during file format conversion phases, etc.

2.10.2 Pricing strategy

The rate is degressive and the sliding scale includes three levels (1-10TB, 10-100TB, over 100 TB). A weighting factor is used for collections that have a high number of archived packages per Terabyte.

	SL1	*SL2*
1To-10TB	1 043 €	764 €
10To-100TB	680 €	493 €
>100TB	318 €	221 €

Number of package / TB	*Markup*
<250 000	0%
250 – 500 000	10%
500 – 750 000	15%
750 – 1 000 000	20%
>1 000 000	25%

2.10.3 Example

For a collection of 27 TB to be preserved on disks, the fee is calculated as follows:

(10 x 1043) + (17 x 680) = 21 990 € per year.

3 References

No	Description/Link
R1	https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification