

GIVING ACCESS TO BORN-DIGITAL ARCHIVES AT THE ARCHIVES NATIONALES (FRANCE)

The OeDIPus riddle

Levasseur Emeline

*Archives nationales
France*

*emeline.levasseur@culture.
gouv.fr*

Falut André

*Archives nationales
France*

andre.falut@culture.gouv.fr

Fenech Julien

*Archives nationales
France*

julien.fenech@culture.gouv.fr

Ferrera Matias

*Archives nationales
France*

matias.ferrera@culture.gouv.fr

Abstract – Born-digital archives, managed in a repository, are accessible thanks to the delivery of Dissemination Information Packages (DIP). While the DIP is a machine-readable format, it is not easily intelligible for the end user. In practice, making DIP truly accessible appears like a riddle. To make it more human-readable, a DIP must undergo some processing outside the repository and be converted in a new form, which should be discussed. These operations call into question the integrity of the archives, which was ensured until their delivery by the system. How, then, can archivists keep data trustworthy? Following recent requests for access, the *Archives nationales* (France) have provided some answers that could be used as a basis for discussion to solve this OeDIPus riddle. Their experience was one of moving from theory to practice, leading to the creation of an in-house proof-of-concept (POC) and tool: OE-DIP (*Objets et Empreintes de DIP*, Objects and Checksums of DIP).

Keywords – access, DIP, integrity, fixity, checksum
Conference Topics – Digital accessibility, inclusion, and diversity; From theory to practice

I. MAKING DIP TRULY ACCESSIBLE: A NEW RIDDLE FOR OEDIPUS

A. *How Digital Archives Are Accessed At The Archives Nationales: Let Oedipus' Journey Begin!*

At the *Archives nationales*, the digital archiving platform, put into service in 2018, is based on the Vitam software (as a back-end). Vitam is an open-

source software meeting French and international exchange standards and norms (DEPIP, SEDA, NF Z 42-013) to ensure the ingest, management, preservation and long-term access to digital records of administrative as well as archival value [1]. It is interfaced with the *Archives nationales'* archives management software (as a front-end). The platform can deliver digital archives in a DIP, consisting of a zipped container made up of a directory, without any further structure, containing on one hand the retrieved files, which are renamed in a non-meaningful way, and on the other hand their structural, descriptive, technical and management metadata encoded in a XML file (a "manifest"), in accordance with the SEDA (*Standard d'échange de données pour l'archivage*, Data exchange standard for archiving) [2].

The DIP must be delivered in compliance with the *Code du Patrimoine* access rules [3]. Archives are on principle accessible to all. Therefore, they can be consulted on special workstations in the reading room at the *Archives nationales*, copied on hard drive disks or provided through a secure transfer platform. However, there are many exceptions that prevent archives from being immediately accessible, in order to protect the privacy and safety of individuals or national security. In that case, the reader can ask their producer for an exemption. If it is granted, the reader can be allowed either consultation in the

reading room only or full access (consultation and copy).

B. And Then The Sphinx Asks: "How Can Humans Read And Trust The DIP?"

The DIP, meant to be interpreted by a machine, is not easily intelligible for end users unfamiliar with digital archiving concepts and data exchange standards. As a result, a transformation is necessary to make the data and their metadata more human-readable. This requires handling of the DIP contents, which can only be carried out outside the repository, after the delivery. However, such an operation is only acceptable if archives can be proven not to be altered in the process. The chain of integrity must not be disrupted [4].

In order to tackle this double challenge - to meet the concrete needs of users while guaranteeing the integrity of the archives they access - the *Archives nationales* have worked on a proof of concept (POC).

II. OEDIPUS HAS SOLVED THE RIDDLE (AGAIN!)

A. *From DIP To Tree*

The first stage of the POC consists in exporting the archives using ReSIP [5], a tool for processing information packages, usually to prepare SIP (Submission Information Packages), developed by the Vitam program. By interpreting the manifest, ReSIP enables the importation of a DIP and its exportation as a tree of directories and files with their original and meaningful names. A CSV file containing the SEDA descriptive metadata is also exported alongside the tree. This new way of retrieving the archives is easier to understand and seems to meet generic needs, common to all types of users.

B. *OE-DIP: Proving The Archives' Integrity*

The second step of the POC consists in proving that this treatment has not altered the fixity of the accessed archives. To do so, the *Archives nationales* have developed an in-house tool called OE-DIP (*Objets et empreintes de DIP*, Objects and checksums of DIP). OE-DIP performs comparisons between the DIP delivered by the repository on one hand, and the archive tree and its CSV metadata file exported from ReSIP on the other hand. The sequence of instructions first extracts the object hashes from the DIP manifest. It then uses the manifest and the CSV metadata file to locate the corresponding objects in

the archive tree, calculates their checksums and compares them with the hashes it has extracted from the manifest. As a result of this comparison, OE-DIP issues an integrity report informing of the outcome of the operation. A positive report acts as a guarantee that the files' fixity was maintained despite the transformation of the DIP. The tool has been tested on various types of digital archives and on large volumes.

III. SOLVING THE OEDIPUS RIDDLE, LET'S NOT MAKE A COMPLEX OUT OF IT

A. *DIP VS Tree: And The Winner Is...*

In the context of this POC, the *Archives nationales* have deliberately delivered digital archives in this double form to researchers: DIP and tree structure. In spite of this double delivery mode, researchers have so far always chosen to access the files in tree form, demonstrating a clear preference for a mode that differs as little as possible from their usual way of browsing. This feedback clearly demonstrates the need to process the DIP and the relevance of delivering the OE-DIP fixity report to users, as a guarantee of trust.

B. *Expected Aftermath*

As of now, ways of improving the OE-DIP tool are under consideration, especially regarding the related issue of metadata integrity. It should be noted that it remains experimental and can only be used in the technical and functional context of the *Archives nationales*. However, if the process meets a need shared by the archives community in France, it could either lead to the development of a hash-calculating feature in the ReSip tool, or to the integration of tree form exports to the Vitam software. The digital archives could then be rendered both as a DIP and directly in a tree form, in order to avoid handling outside the repository, which would enable greater trust and possibly negate the need for the integrity report. In this manner, giving access to digital archives delivered in DIP will no longer represent an Oedipus riddle... nor a complex!

1. REFERENCES

- [1] Vitam Program presentation, Programme Vitam. https://www.programmevitam.fr/pages/english/pres_english/
- [2] Structuration des Dissemination Information Packages (DIP), Programme Vitam. https://www.programmevitam.fr/ressources/DocCourante/autres/fonctionnel/VITAM_Structuration_des_DIP.pdf

- [3] Code du patrimoine: Chapitre 3: Régime de communication (Articles L213-1 à L213-8), Légifrance. https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT00006074236/LEGISCTA000006159942/#LEGISCTA000006159942
- [4] Digital Preservation Handbook: Preservation issues, Digital Preservation Coalition. <https://www.dpconline.org/handbook/digital-preservation/preservation-issues>
- [5] ReSIP, Programme Vitam. <https://www.programmevitam.fr/pages/ressources/resip/>