



HAL
open science

Réutiliser et pérenniser les données de la recherche : l'exemple de grammateus

Elisa Nury

► To cite this version:

Elisa Nury. Réutiliser et pérenniser les données de la recherche : l'exemple de grammateus. *Humanistica* 2023, Association francophone des humanités numériques, Jun 2023, Genève, Suisse. hal-04102863

HAL Id: hal-04102863

<https://hal.science/hal-04102863>

Submitted on 22 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Réutiliser et pérenniser les données de la recherche : l'exemple de grammateus

Elisa Nury

Université de Genève

elisa.nury@unige.ch

Résumé

En l'absence de financement pour la maintenance de projets numériques, comment assurer l'accès et la citabilité des résultats de recherche qui peuvent rester pertinents plusieurs décennies ? Cette présentation se penche sur les solutions adoptées dans le cadre du projet FNS grammateus.

1 Introduction

Cette présentation aborde le thème de la circulation des données à travers grammateus, un projet en Humanités Numériques de l'Université de Genève qui prépare un environnement virtuel de recherche pour étudier la typologie des papyri documentaires grecs de l'Égypte antique. Cet environnement en ligne propose de nouvelles données de recherche sous la forme de métadonnées papyrologiques et descriptions pointues des types de documents, mais repose également sur la réutilisation directe de données au format XML TEI provenant de bases de données telles que *Heidelberger Gesamtverzeichnis der Griechischen Papyrusurkunden Ägyptens* (HGV) et la *Duke Database of Documentary Papyri* (DDbDP) à travers le site *papyri.info*¹.

Les résultats de recherche produits dans le contexte de grammateus par l'équipe, formée de trois papyrologues et une spécialiste Humanités Numériques, doivent rester accessibles et citables pour les futur·e·s chercheur·e·s. Ces résultats sont potentiellement pertinents sur un horizon lointain – plusieurs décennies, comme c'est souvent le cas dans les domaines en Sciences Humaines et Sociales et en particulier en Sciences de l'Antiquité. Dans un contexte académique qui ne finance pas la maintenance de projets numériques tels que grammateus sur le (très) long terme, et en ayant conscience de la fragilité inhérentes aux sites web dépendants non

seulement de données scientifiques externes régulièrement mises à jour, mais aussi des technologies web en constante évolution, la question qui se pose est donc : comment assurer l'accès à la production savante des projets numériques sur le long terme ?

Pour le projet grammateus, nous avons choisi de nous concentrer sur la préparation et l'archivage de données de qualité, en suivant des standards établis. Nous souhaitons présenter et discuter le cas de grammateus sous l'angle du cycle des données, depuis le processus de création et réutilisation de données jusqu'à leur archivage. La typologie des papyri documentaires – les documents de la vie de tous les jours (lettres, contrats, etc.) et non littéraires – comporte une partie descriptive du processus de classification et des différentes catégories établies, ainsi qu'une base de données de papyri qui illustrent chaque catégorie de documents. Les données sont donc divisées en deux catégories : il y a d'une part les données relatives aux papyri et d'autre part les textes des descriptions.

2 Base de données papyrologique

Les données relatives aux papyri sont en partie produites par l'équipe, mais les transcriptions des documents de la bases de données proviennent de la DDbDP via *papyri.info*. Chaque papyrus est illustré par un modèle numérique qui reprend le texte en ajoutant une nouvelle couche d'information : les sections de texte sont soulignées par des couleurs, car elles jouent un rôle dans la typologie des papyri (voir Figure 1).

Ces modèles sont mis à jour à chaque fois que la page est ouverte, les données sont donc dynamiquement intégrées à grammateus. Il faut noter que la communauté des papyrologues a développé un outil collaboratif, l'éditeur papyrologique, qui permet à tous de mettre à jour les données sur *papyri.info* – en ajoutant de nouvelles transcriptions, en mettant à jour celles qui existent déjà ou en cor-

1. <https://aquila.zaw.uni-heidelberg.de/start> (accessed 01.02.2023); <https://papyri.info/> (accessed 15.02.2023).

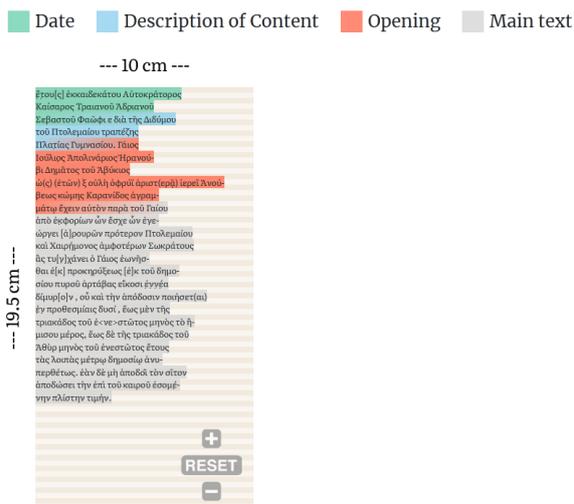


FIGURE 1 – Exemple de texte présenté dans le projet grammateus : Le papyrus P.Mich. 9 572 (TM 12064).

riageant les métadonnées relatives telles que la date ou la provenance d'un papyrus.

Cet aspect d'accès dynamique du projet est une innovation importante, car il démontre qu'il est possible de travailler et présenter les textes les plus à jour possibles, ce qui n'est en général pas le cas. Mais d'autre part, la dépendance à *papyri.info* nécessite une maintenance afin de préserver le fonctionnement de grammateus : outre les changements technologiques web comme le passage au protocole HTTPS, le format d'encodage EpiDoc (Elliott et al., 2007), standard pour les textes épigraphiques ou sur papyrus, ainsi que les stylesheets (Elliott et al., 2008) qui permettent la conversion du texte au format HTML évoluent constamment².

3 Textes savants

Les données textuelles telles que les descriptions sont encodées également en TEI et sont particulièrement sensibles aux problématiques d'accès et de pérennisation. En effet, pour tirer pleinement avantage des ressources papyrologiques en ligne, grammateus ne donnera pas lieu à une publication traditionnelle sous forme d'un livre imprimé et publié par un éditeur. Les textes du projet font donc face aux mêmes problèmes soulevés par Bleier (2021) pour les éditions critiques. En effet, les descriptions typologiques peuvent rester pertinentes sur le très long terme. Pour preuve, des textes savants datant

de plus d'un siècle sont cités par le projet grammateus : par exemple Wilcken (1912) cité par Ferretti et al. (2023, §4). En plus de présenter l'encodage et l'archivage des données sur Yareta, l'entrepôt d'archivage des données de l'Université de Genève, nous discuterons également de la question de l'évaluation par les pairs pour les ressources en ligne : bien que nécessaire pour confirmer la validité de la production savante, il n'existe pas de solution évidente pour grammateus.

Bibliographie

- Roman Bleier. 2021. How to cite this digital edition ? *Digital Humanities Quarterly*, 015(3).
- Tom Elliott, Zaneta Au, Gabriel Bodard, Hugh Cayless, Carmen Lanz, Faith Lawrence, Scot Vanderbilt, and Raffaele Vigiante. 2008. EpiDoc Reference Style-sheets.
- Tom Elliott, Gabriel Bodard, Elli Mylonas, Simona Stoyanova, Charlotte Tupman, and Scot Vanderbilt. 2007. EpiDoc Guidelines : Ancient documents in TEI XML.
- Lavinia Ferretti, Susan Fogarty, Elisa Nury, and Paul Schubert. 2023. *Description of Greek Documentary Papyri: Report of Official Proceedings (Roman)*. In *Grammateus Project*.
- Ulrich Wilcken. 1912. *Grundzüge und Chrestomathie der Papyruskunde. Bd.1 Historischer Teil. 1. Hälfte : Grundzüge*. B.G. Teubner, Leipzig.

2. La dernière mise à jour des guidelines EpiDoc à la version 9.4 date d'octobre 2022 (voir <https://lsv.uky.edu/scripts/wa.exe?A2=MARKUP;1c2d4ed2.2210&S=> accessed 15.02.2023).