



WP 3.1 Data Management



DESCRIPTION OF WORK

The foreseen **impact** is a significant upgrade of the data management, versioning and workflow automation capabilities that underlie the whole infrastructure. While the infrastructure already **has** multiple distinct processes for data ingestion and versioning, there is a **need** (D:3.1.1) to further integrate and develop these, both to improve efficiency, as well as to cater to the new types of material to be managed in this project. Further, due to the increasing amount of different enrichments being applied to the materials in WP3.3, there is a need (D:3.1.2) to develop better capabilities for workflow automation and version syncing, which will allow these enrichments to be run automatically every time the source data changes, foregoing manual work and the possibility of using stale data versions.

TIMELINE

Quarter	1	2	3	4	5	6	7	8
			D 1					
							D 2	

Status and Plan

A versioning workflow is already in place for NLF Newspaper data. Automatic ingestion for WP3.3 (Enrichment) are not yet in place. Presently there is no monitoring in place to ensure that new versions are created as expected. Also retention times for versions are not agreed (for reproducibility reasons they should be a few years). Versioning needs storage space, if the data is difficult to compress, a lot of storage space. A funding model for continuous support after 2025 needs to be developed.

OUTCOME / Deliverable #1

- Automatic version creation in place in a monitored way.
- Versioning timetable and retention times agreed.
- Funding model after end of project agreed.

OUTCOME / Deliverable #2

- Existing NLF Harvester is extended so it can be used to synchronize data from WP 3.3.