

## **Topic: FoLiA: Maintain and continue support and development for the FoLiA format and the rich infrastructure of tools, datasets, and libraries surrounding it.**

**ID:** T108

**Author:** Maarten van Gompel

**What exists:** The FoLiA format has been in existence since 2011 and is widely used in CLARIAH and is strongly coupled to a lot of software we develop and data we deliver.

**What must be adapted / extended / created anew:** The following points are all deemed essential to ensure long-term sustainability for FoLiA:

- FoLiA is extended with new annotation types where necessary, in close contact with users.
- Support is always offered to all users/developers (both within and beyond CLARIAH) with any questions about FoLiA or usage of its various tools and libraries
- The FoLiA programming libraries for Python and C++ are constantly improved, bugfixed, and extended to comply to the latest specification. Both libraries are kept in sync and are an integral part of many of our software (Frog, ucto, FLAT, Gecco, TICCL, etc. . . )
- The FoLiA-Tools (python) and FoLiA-utilities (C++) [largely non-overlapping tools] are expanded when and where new tools are needed.

**Why important for CLARIAH (scientific impact):** FoLiA is being used throughout CLARIAH, by both tools and data sets, and is an important facility to maintain for the future.

**Targeted/Actual users:** tool-developers, data-providers, data scientists

**Actual use (quantify!):** Used by almost all tools provided by Nijmegen, but also beyond, used by many corpora (SoNaR-500, Basilex, Basiscript, DutchSemCor, OpenCGN, VU-DNC, Nederlab), used by lots of other partners.

**Lead:** Maarten van Gompel

**Proposed PM Estimation** (try to justify): 12PM - This should be seen as an ongoing effort stretched over the entire lifetime of CLARIAH-PLUS and has reserved time also for questions or requests coming from the community.

**Proposed Participants + PMs:** Maarten van Gompel (6PM), Ko van der Sloot (6PM)

**Allocated PMs:** 6PM (-6!)

### **Deliverables**

1. [FoLiA Documentation and Reference Guide](#)
  - [folia#43](#) Completely revised FoLiA documentation, turn into more formal specification
  - [folia#46](#) Proper support for data provenance logging
  - [folia#51](#) Increases expressivity for multi-word annotations
2. [FoLiA Schemas and Specification](#)
3. FoLiA Library for Python: [FoLiAPy](#)
4. FoLiA Library for C++: [libfolia](#)
5. [FoLiA Tools](#) (Assorted command-line tools for FoLiA, Python-based)
6. [FoLiA Utilities](#) (Assorted command-line tools for FoLiA, C++)

### **Milestones**

- [FoLiA v2.0](#) - Early 2019 - FoLiA v2.0 is a major update of the FoLiA format, introducing various changes, most notably support for provenance data and completely renewed documentation.