

From: "Jan Hajic" <hajic@ufal.mff.cuni.cz>
To: "Ufal-I mail conference, local DL" <ufal-l@ufal.mff.cuni.cz>
Sent: Monday, 30 December, 2024 19:29:41
Subject: P.F. 2025 with thanks for your work in 2024

Dear all,

As has become tradition, I am writing to you after you have already seen Bára's overview of 2024. This letter is primarily addressed to my ÚFAL colleagues involved in the LINDAT/CLARIAH-CZ Research Infrastructure. However, it is also for those who have contributed to it in any way during 2024, whether directly or indirectly. Even if you haven't been involved yet, you may contribute in the future; in any case, I think it's always helpful for everyone to stay informed about each other's activities.

2024 marked the second year in the ongoing four-year cycle of LINDAT's funding, and it has been the quietest of the four. Having moved past the more turbulent initial year, we are now entering a year in which we will start the preparation for the next cycle, as I'll elaborate on below. However, there have still been significant developments.

In addition to the new projects that Bára has already mentioned - such as EOSC CZ OS I NRP, EVERSE, and ATRIUM, in which LINDAT directly participates - we have submitted several additional projects with a direct or supporting role of LINDAT. These include CEDMO 2, OP JAK MeS for industrial cooperation, two Digital Europe projects, one Horizon Europe project, and OP JAK Excellent Science. While the latter two projects are still under evaluation, the others are either already underway or expected to start early in 2025, pending the usual formalities.

Several ongoing projects have also made significant progress. Notably, the OP JAK funding for equipment, as well as our involvement in CEDMO 2 and EOSC CZ OS I NRP projects, have allowed us to significantly expand our computing cluster. This would not have been possible without the dedicated work of the ÚFAL/LINDAT IT team, who managed everything from specifications through the public tenders to installation. Data preparation, funded by LINDAT and various other sources (such as GAČR, MŠMT, NAKI, TAČR, etc.), has continued as well. While I may forget some of the smaller ones, the largest datasets include: the four-million-token PDT-C 2.0 with new, full manual annotation of the a-layer, new discourse annotations from the discourse team, thousands of corrections, and updates to the now again fully compatible morphological and valency lexicons; SynSemClass 5.1, which has just been released; and, of course, the regular updates to Universal Dependencies (now at version 2.15), along with all the upgraded UDPipe models available through LINDAT repository and services. The U.S.-Czech UMR project is also set to release UMR version 2.0 on LINDAT very soon, now including (automatically converted) Czech and (manually annotated) Latin treebanks from our teams. LINDAT services, whether updated or not, continue to handle tens of millions of requests yearly.

In addition to these, many other projects supported by LINDAT - both within and outside of ÚFAL - have continued. This includes the CVHM project, which is already integrated into LINDAT but, for historical and significance reasons, we still report as a separate initiative, as well as the HPLT data- and LLM-building project we are leading. Several other supported projects, both national and international, are also ongoing, all led or co-led by our ÚFAL colleagues.

We are also focused on securing future funding for both LINDAT's development and for combined efforts with ÚFAL colleagues, often involving national and international partners. For 2025, we are in a relatively strong position, assuming that the OP JAK "industrial" and Digital Europe OpenEuroLLM

projects are approved. EOSC CZ OS I NRP, which is supporting CLARIN DSpace development, will run until 2028, and the humanities- and arts-focused EOSC CZ OS II is expected to launch in late 2025 or early 2026, running until the end of 2028 as well. LINDAT itself is funded through the end of 2026, and HPLT will be extended until the end of 2025. Some other projects, including international ones, extend beyond 2025 or even 2026. However, we must continue seeking new funding opportunities, as the reduction in funding expected for 2027, and especially the drop after 2028, will be significant. It is unlikely that any continuation of OP JAK after 2027 will exist, or it will have extremely limited funding allocation.

The primary focus for 2025, however, will be securing LINDAT's funding for 2027-2030. This funding is the core support for many of our activities, and we must put in our best efforts to secure it once again. The usual "Evaluation 2025" process will take place, which will include an international panel review and a thorough collection of scientometric data on LINDAT and its users. The evaluation report must be prepared in the spring of 2025. I will rely not only on the LINDAT team but will likely need to call on colleagues from ÚFAL to help craft a compelling report that will impress both the panel and the Ministry. As many of you know, this is a very serious matter: Czech Research Infrastructures are evaluated on a 1-5 scale, and only those receiving a grade of 4 or 5 are allowed to continue, with significant funding cuts if a grade 4 is awarded.

Regarding other funding opportunities, the LINDAT team is, of course, happy to assist with negotiations, writing, and submission of any projects that can help secure further funding for ÚFAL, even if LINDAT is only playing a supporting role. As always, we are ready to provide data and services support for everyone. Please don't hesitate to bring us your ideas and apply for grants - LINDAT is here to help and support your projects.

I haven't mentioned teaching yet; while LINDAT's main purpose is not teaching per se, we are always happy to support students at all levels, openly and globally, the more so if they are "ours." If any of you who teach our courses need additional support from LINDAT, please let us know, and we will do our best to assist you with new data, tools, or services.

To conclude, I would like to express my sincere thanks to everyone for their contributions to the LINDAT infrastructure, data, projects, and for keeping our collective efforts strong and thriving. This includes all my colleagues directly involved in the core LINDAT development, the teams working on data preparation, processing and packaging, and developers of our tools and services, both within and outside of LINDAT, our excellent teachers for educating students who use or later contribute to LINDAT, and the IT team for not only securing new equipment but also for maintaining our hardware, software, and networking systems. I am also grateful to the increasingly important project management and administration team, whose excellent work is essential to the success of all our projects - without them, none of this would be possible.

I would like to extend my special gratitude to Bára, as head of ÚFAL, for her leadership and coordination across all of ÚFAL and her personal support for LINDAT. I also wish to thank Vice-Dean Prof. J. Sgall and Dean Prof. M. Rokyta, both of whom have been incredibly helpful in addressing our often unconventional requests and challenges, even if initially they are often not particularly happy. And, of course, I cannot forget the various departments at the Dean's Office: while our cooperation isn't always seamless, their work is indispensable to our continued success.

I wish you all a prosperous 2025, filled with good health, positive spirits, minimal stress, and plenty of joy in both your professional and personal endeavors. May your travels be safe, and your scientific contributions be significant and long-lasting, used and cited for many years to come.

I look forward to seeing you all again in 2025!

Jan Hajič
PI and coordinator, LINDAT/CLARIAH-CZ
ÚFAL MFF UK