Sustainability of Research Infrastructures: Some lessons learned from TextGrid and CLARIN-D

Alongside the significant growth of research in Digital Humanities, linguistic and humanistic methods and approaches incline be more and more empirical.

The increase in the amount of digital research data has greatly improved the verifiability of the research results. However, these vital resources are facing serious obstacles not only in keeping research infrastructures technically or organizationally stable, but also in financing their work. Regardless the methodology or technical approaches to research data management – the presentation highlights the experiences gained from the Virtual Research Environment TextGrid and the language infrastructures of the European CLARIN consortium.

A financial framework have to address every single component of a usable, accessible, and secure infrastructure. Particularly, the aspect of reusing data to replicate and approve data-based research requires technical flexibility and development potential in regard to dynamically changing technologies and demands, whereas only a continuous operation and maintenance drives the life-cycle of research data. These priorities raise the issue of financing a long term research data management.

It has been a long time, since German funding organizations included sustainability and stability as criteria in evaluating scientific project applications. The Deutsche Forschungsgemeinschaft (DFG) has established a codex of best practices comprising a retention period of results, tools and data over at least ten years. Due to tight budgets, concepts of data management structures have to meet the financial feasibility and the distribution of costs of a torrential data stream. Considering data preservation as a long-term task of conserving unique knowledge and preventing expensive redundant data assessments, a cursory project-based knock-on funding will not suffice in the future.

TextGrid and CLARIN-D projects set themselves to develop an integral workflow and were able to gather valuable experience in funding evolving digital research data infrastructures in the eHumanities:

- **Building up** research infrastructures usually proceeds on short-term project basis. Third party funding is the traditional source of capital in this initial stadium.
- **Drafting** a faculty profile of research infrastructures should be done with realistic goals. Researchers responsible for the project always have to be aware of the maximum funding period.
- During every stadium of build-up, the development of tools and services from scratch is the primary **cost-pusher**. Drawing on open software and standardized formats are key issues for reasonable technical operations.
- As soon as scientific communities intend to make use of digital research infrastructures, funding organization weighs presumptive costs. Unsound or vague cost schedules on the part of applicants will most certainly be harmful to any application effort. Therefore **types of costs** have to be identified and estimated.
- For **estimating cost,** an inductive approach starting with single individual operations to estimate an overall cost proved to be successful.

- In general, the **long-term life cycle** of research infrastructures will not be funded.
- Research infrastructures need to **dynamically adapt** to evolving demands in order to keep functioning effectively.
- Within a framework of an **open and flexible architecture** vital services are to be identified and institutionalized. Further add-on services will modularly complement core infrastructures and may be funded separately, e.g. with funds from new project basis.
- Assigning capable **key players** ensures the operation of core infrastructures while existent structures and solutions are to be used synergetically.
- Partners may be attracted if the scientific **overvalue** of research infrastructures is visible as well as conclusive (Community Building) and its usage is embedded in an organizationally stable and secure framework.
- Based on experience, **labor costs** usually account for 85% of overall costs of research infrastructures.
- **Overall life cycle costs** are allocated among key infrastructures (10%), long-term archiving (30%), support of tools and services (30%), user liaison, support and training (20%), and administrative management (10%).