

# 7. Lessons learned (Statistics Austria)

- It is not a new discovery that the subject of statistical metadata is an extremely complex one. Even now, almost three and a half decades after Bo Sundgren first used the term, different individuals may still mean quite different things or place emphasis on different aspects when speaking of metadata. This phenomenon is even more pronounced when these persons stem from different areas of expertise: senior management, subject matter statisticians, methods specialists, IT experts etc.
- Papers prepared for the METIS work sessions and the Common Metadata Framework (especially the case studies) have proven very useful, as they provide arguments for discussions with statisticians and top management.
- The creation of an integrated system consisting of more than isolated solutions is difficult when there is no organizational unit the main responsibility of which is to deal with the subject of metadata and its usefulness for the NSI - and which is also granted the requisite authority and enjoys the support of top management, so that it can achieve the introduction of integrated and centralized metadata systems even against the possible resistance of subject matter departments.
- That metadata projects are best carried out using an interdisciplinary approach (and not as IT projects) has long been recognized in expert circles. In practice, however, it appears that the qualified subject matter statisticians continually suffer from such a heavy workload that they have no time to spare for complicated conceptual work (e.g., Statistics Austria has reduced the number of personnel by about a third since its separation from the federal civil service in the year 2000).
- Many statisticians associate the concept of "metadata" with the notion of "additional work" (which for instance actually was the case when the standard documentations were introduced). This leads them to resist new metadata systems.
- The idea of developing specialized tools for editing, administrating and (re-)using metadata with an end-to-end approach regarding the statistical life cycle often encounters resistance among statisticians because the introduction of such tools will result in changes to work processes which they have been familiar with for many years.

Statistics, however, is not the only field of activity in which the creation and usage of metadata can be seen as part of the job description. In order to produce software of high quality and in an economic way, the availability of tools - to support the management of "software metadata" (including the source code of the programs) and to provide services to alleviate the software engineers' work - has long been recognized as necessary. Especially when several programmers are cooperating in a software project, the storage and administration of all information items in a central repository seems indispensable.

The production of statistics exhibits a high degree of similarity to the production of software. However, in statistics the advantages offered by specialized tools and a centralized metadata repository are not yet generally accepted.

- As was already said in section 2.3, with regard to the development of systems for the collection and administration of passive metadata, the cost factor presents a particular obstacle. Passive metadata are an integral component of statistical information. Their availability and easy accessibility contribute to the quality of statistical products, but in many cases do not result in cost reductions (they may even increase the work load of subject matter statisticians). Opportunity costs caused by the non-existence of centralized end-to-end metadata systems are rarely found in accounting systems. Thus high investments are accompanied "only" by a gradual gain in quality (which may not even be recognized by all user groups). Under these circumstances it is understandable that in times of economic crisis the willingness to invest in metadata projects is not high.

The concept of "high-quality statistics" is a dynamic one. The needs and requirements of users are changing and will probably increase in the future, e.g. with regard to harmonization of statistics or the linkage of data with *relevant* metadata items (respectively linkage of metadata items with related metadata items), so that they can be accessed at the push of a button. If metadata are stored in the continuous text of bulky documents, these new requirements cannot be met. The management of metadata in an "atomic" and structured form, however, is a challenge with respect to both financial resources and personnel.

The fundamental principles of metadata management, which have been defined by experts during recent years (and which can be found, for example, in part A of the Common Metadata Framework) will become more and more commonly accepted standards and state of the art for the production and dissemination of statistical information.

The task of implementing these standards can certainly not be carried out at short notice. In this respect, it is not easy to answer the question whether to continue building isolated metadata systems whenever the need for one specific system arises, or whether to strive for an integrated system based on a global architecture. The first approach is certainly less expensive in the short run and produces quicker results, but in the long term it will cause quite substantial "repair" costs.

What metadata should actually be collected for and provided to external and internal users, and in what form should they be provided? This is a fundamental question on which opinions within Statistics Austria are divided. The search for an answer should not be postponed just because it is clear from the start that up-to-date solutions will require high investments in time and money. The answer should rather be given as soon as possible in order to ensure from the start that the solutions - which must be planned and implemented step-by-step in accordance with budgetary constraints and on a long-term time scale - will be built to last.