

**UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION  
STATISTICAL OFFICE OF THE  
EUROPEAN UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION  
AND DEVELOPMENT (OECD)  
STATISTICS DIRECTORATE**

**Work Session on Statistical Metadata (METIS)**  
(Geneva, Switzerland, 10-12 March 2010)

## **CASE STUDY (UPDATE 2010) – GERMANY**

Edgar Jaensch, Destatis

### **I. INTRODUCTION**

1. Several major changes in the Destatis' work programme for metadata will be taking shape in 2010. First, the GSBPM will become Destatis' new business process model and will at least be partly introduced in the Verbund<sup>1</sup> as well. Second, a new classification server is nearing completion with a prototype already available. Third, the next step – a variable server – is underway as part of the census 2011 metadata system with business specifications finished in 2009 and programming scheduled to begin in 2010.

### **II. ADOPTION OF GSBPM**

2. A major focus in 2009 was the adoption of the Generic Statistical Business Process Model. Over the course of the year, the GSBPM has gained attention and has been accepted for a variety of purposes at Destatis and the Länder offices.

3. After the release of the GSBPM by the METIS steering committee, the metadata unit initiated a working group at Destatis to translate and discuss the model with the aim of issuing a German version. The German version should represent the situation and the terminology used in official statistics in Germany. The working group consists of subject matter statisticians and representatives from the so-called cross-sectional departments (IT, organisation and coordination departments, respectively). The metadata unit is in charge of maintaining the model and editing any changes to it. The working group has kept the structure of the original model and has not altered levels 1 and 2. With the levels 1 and 2 now completed, the focus has shifted to the translation of level 3 and its adoption.

---

<sup>1</sup> Official statistics in Germany is split between the Federal Statistical Office (Destatis) and the statistical offices of the Länder (engl.: states/regions). Together, the statistical offices in Germany are referred to as the "Verbund" (engl.: network/union).

4. At Destatis, there seems to be universal agreement that the GSBPM indeed covers exhaustively all activities in the production of statistical data, being truly an E2E (end to end) process model. The discussion about level 3 will certainly focus on eliminating ambiguities in assigning inputs, outputs and IT systems to each sub-process. However, with the growing acceptance of the model at Destatis and in the Verbund, the discussion about the model is also broadening. With more people taking part in the process, completing a final German version will still take some time.
5. Despite not being entirely finished, several Verbund working groups have started to work with the model. Within Germany's "Standardisation of Production (SteP)" project a sub-working group (called "step 14 – guidelines") was tasked with drafting guidelines for the application of the new standard IT tools designed as part of SteP. The guidelines should encourage the use of new standard IT tools in statistical production. The working group decided to drop an older process model in favour of the new GSBPM and to map the existing standard IT tools to the GSBPM. The result will be an inventory of standard IT tools in compliance with the GSBPM and guidelines how to use them. The decision of the working group has been accepted by the heads of the Länder offices and Destatis, paving the way for the GSBPM to be adopted as the new business process model for SteP.
6. The metadata working group intends to use the GSBPM for modelling production and utilization of metadata along the process chain. As a result, a list of metadata objects produced by or used in every sub process will be established.
7. Another major application for the GSBPM in official statistics in Germany is in the field of data quality. A working group on quality has been tasked with setting up guidelines for the documentation of processes. It has decided to use the GSBPM to develop a framework in that respect.
8. Within Destatis, the heads of department have decided to implement the standard process model (GSBPM) by 2012. As a result, much of what is being discussed in the Verbund working groups will become mandatory within Destatis. There should be standard IT tools and rules for their implementation for all sub-processes. Furthermore, process related documentation should exist for all iterative sub-processes. So-called "Process Improvement Teams" are to be set up in order to assess the work in the subject matter departments, facilitate knowledge sharing and increase quality and efficiency.
9. Given the rapid acceptance of the model, it is quite likely that it will become a cornerstone in Destatis' quality and standardisation efforts. Since its application is largely in its initial stages, it is difficult to assess if any changes will have to be made to the model. So far, the structure of the model has proven to be quite stable. Only the working group on guidelines has mentioned that the collect phase might be somewhat underrepresented on the sub-process level, although this could be made up for by extending the descriptions on level 3. (This is the only area that might have been more comprehensively represented in our older model.) It is quite likely that with more experience, a more thorough assessment will be possible. Up to now, the acceptance of the model in the organisation has itself been an encouraging sign.

### **III. A NEW CLASSIFICATION SERVER**

10. In collaboration with the Bavarian State Office for Statistics and Data Processing and with advice from Statistics Norway, Destatis has started a project for a new classification server. The new classification server – named KlassService – is jointly financed by the Verbund and replaces an older system that was not designed according to any generic model and hence, new classifications could not be added easily. Using the Neuchâtel Terminology as a generic model, new classifications can now be directly imported into the new system. After finishing the first stage in a three staged development

11. KlassService is a universal classification database for the electronic documentation of standard classifications. It is intended to foster the reuse of existing metadata, allowing users to retrieve the information using either web services or download functionalities. KlassService goes beyond pure documentation and metadata reuse by providing support for automatic and semi-automatic coding procedures. As such it is the standard tool in official statistics in Germany for sub-process 5.2 “classify and code” of the GSBPM (v.4.0).

12. In accordance with the Neuchâtel Terminology, multiple versions of one classification are supported. Non-authoritative variants can be added to a specific classification version. Correspondence tables are used to trace version changes over time on the item level. Classifications may be grouped in multiple classification families.

13. KlassService uses an extended CLASET-XML format as its import format. Thus classifications from Eurostat’s RAMON server can be imported. Destatis’ IT department currently designs a converter tool to transform standardised Excel files into CLASET-XML. There is a principal intention to store SDMX standard code lists in the system as well, but no steps have been taken in that direction yet.

14. To support (semi-) automatic coding procedures, KlassService has an elaborated item search function. Classification items can be tagged with keywords by the classification owner. Keywords represent some subclass of the entities classified under a certain item. To amend the keyword list for an item, people working with the classification - coding staff, for example – can suggest new keywords online using the KlassService front end. The suggestions are then checked and can either be approved or rejected by the classification owner.

15. Furthermore, the KlassService employs an elaborated thesaurus. The thesaurus supports stop word analysis, lemmatization, synonym search and generic terms. The functions of the thesaurus are separated from the classification part of the system and can be called via web services by other applications as well. KlassService supports multiple thesauri, thus allowing each classification owner either to use a standard thesaurus or to implement a classification-specific thesaurus. (Semi-) automatic coding procedures often require extensive modifications to the thesaurus in order to improve the precision of the coding procedure<sup>2</sup>.

16. The intention is to connect KlassService to a number of production systems (usually stove pipe systems) and standard tools. For example, it is intended to connect KlassService to our system for electronic questionnaires (a standard tool) with the aim of assisting respondents to classify and code their answers themselves. When paper questionnaires are used, KlassService allows either completely automatic coding<sup>3</sup> or coding by human coders who then have to pick a code from a set of possible codes that are delivered by the web service (or some combination of both).

17. The KlassService data model is fully multilingual, allowing n-language version of one classification to be stored in the system. Currently, an English language version of the KlassService front-end is under construction and the most important classifications will be available in English.

---

<sup>2</sup> cf. Ferillo, A; Machhia, S; Vacari, P (2008): “Different quality tests on the automatic coding procedure for the Economic Activities descriptions”, Proceedings of Q 2008 – European Conference on Quality in Official Statistics, Rome, 8-11 July 2008.

<sup>3</sup> Despite being technically possible, entirely automatic coding is often seen as being to error prone. (cf. Dekker, Arij (2001): “Adapting new technologies to census operations”, Symposium on Global Review of 2000 Round of Population and Housing Censuses: Mid-decade assessment and Future Prospects, New York, 7-10 August 2001.)

18. The system is build around a MySQL database using affordable Java technology and Tomcat. The system makes use of a number of open source frameworks like JSF & MyFaces, Spring and Hibernate.
19. Accessibility has been a major issue in the design of governmental software in Germany for many years. KlassService conforms to strict German accessibility standards (BITV<sup>4</sup>) and allows people with disabilities to access its content using screen readers, for example.

#### **IV. CENSUS METADATA SYSTEM AND OTHER ISSUES**

20. Up to now, a major focus of the metadata unit at Destatis has been the design of a metadata system for the census 2011 for which the business case was finished and approved in 2008. After finishing this project and thereby testing the underlying concepts, the results should be carried over to a generally introduced metadata system.
21. The census system will be built in several modules. The first will be simple document management, which is intended to store methodological documents. After the technical requirements were finished in 2009, we expect to see the implementation in early 2010 based on an amended version of the European Commission's CIRCA system.
22. The second module is a variable server based on a generic model, namely Part II of the Neuchâtel model. The technical requirements were finished in 2009 so that implementation can start in 2010. Further modules are in their initial planning stages. The variable server is to be supplemented by a system for data description built around the "matrix" element in Neuchâtel terminology. A connection to KlassService is also intended.

#### **V. OUTLOOK FOR 2010**

23. This year will see the KlassService going operational. Beyond that a major redesign of our output data base GENESIS – also containing a lot of metadata – is under development. If things go on as planned, the census metadata system will be taking shape with the classification server being supplemented with a variable server.
24. Resource constraints have led Destatis to rethink its current organisation of metadata management. Suggestions include a decoupling of the two most important metadata projects that up to now have been seen together. Therefore the metadata system for the census might be developed under a different organisational structure than the comprehensive official statistics-wide metadata system. (The latter being part of Germany's larger SteP project. It includes the classification server.) The place of a central metadata unit might be taken by a steering committee, consisting of several heads of departments, and a working group, made up of experts involved in quality and metadata management.

---

<sup>4</sup> see: [http://www.einfach-fuer-alle.de/artikel/bitv\\_english/](http://www.einfach-fuer-alle.de/artikel/bitv_english/)

## VI. KLASSSERVICE SCREENSHOTS

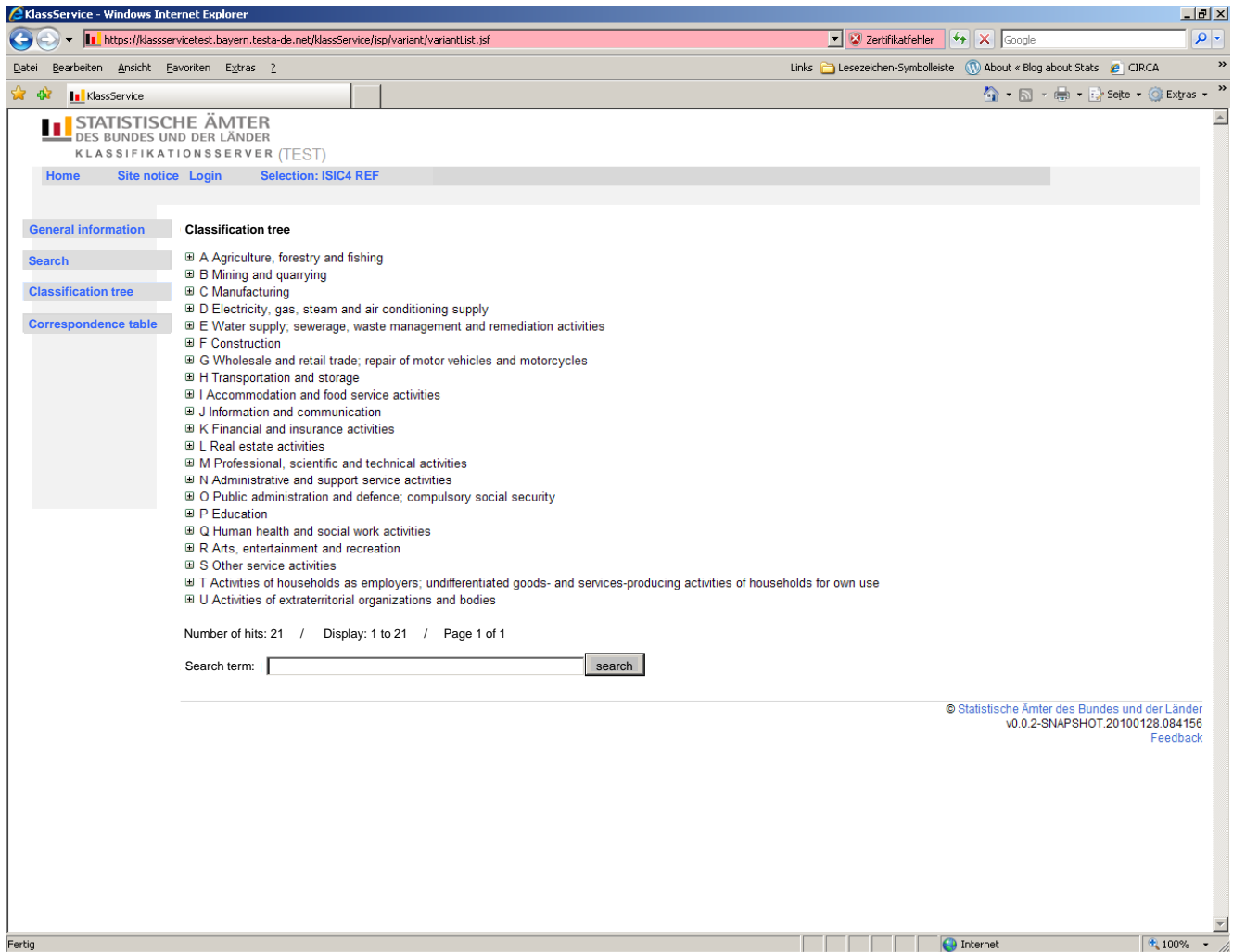


Figure 1: Classification tree

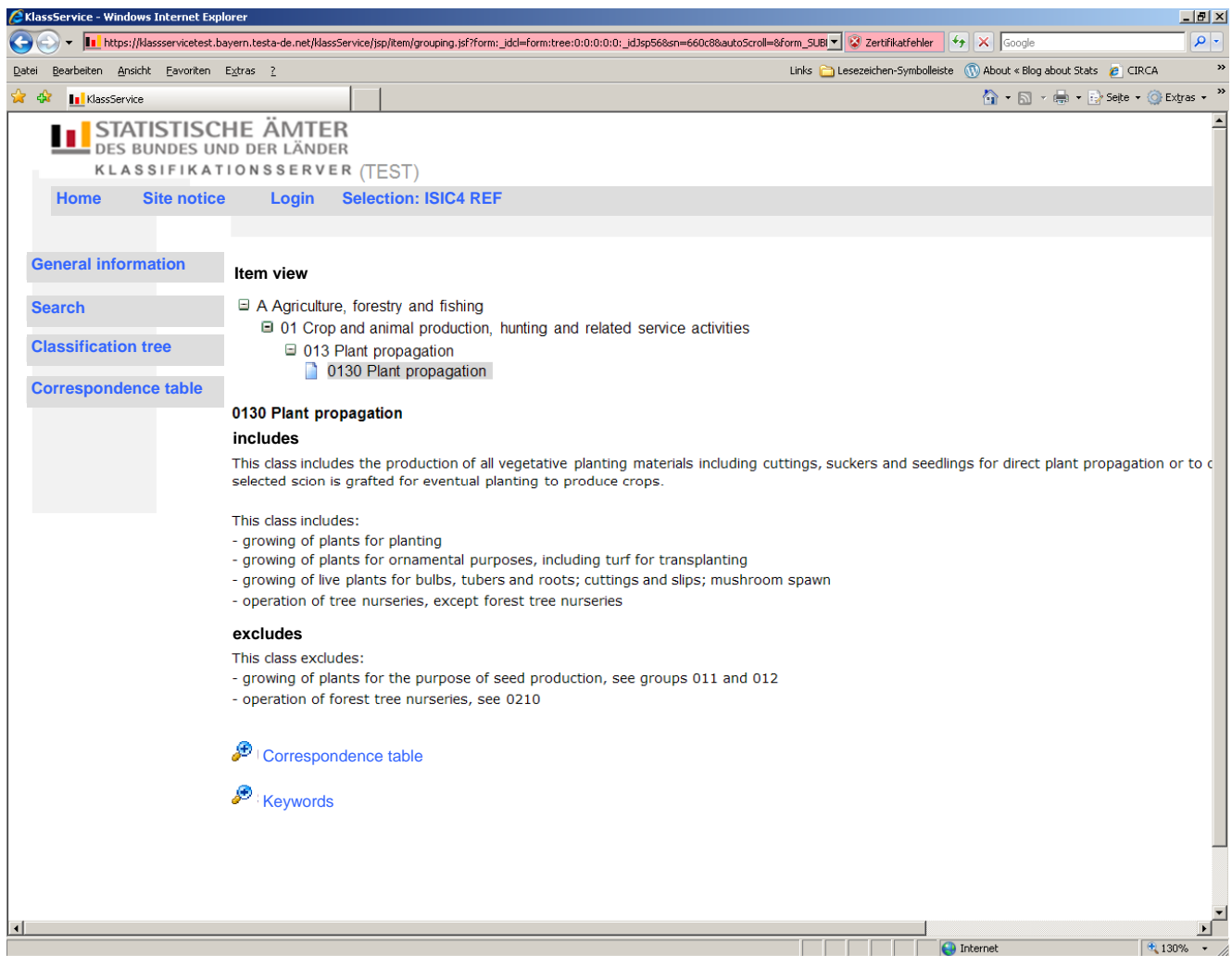


Figure 2: Classification item