

Distr.
GENERAL

Working Paper No.15
11 April 2007

ENGLISH ONLY

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

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

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

Meeting on the Management of Statistical Information Systems (MSIS 2007)
(Geneva, 8-10 May 2007)

Topic (iii): Accessibility and usability of IT applications

**WEB-SERVICES FOR INTEGRATING THE CENTRAL DATABASE IN THE PROCESSES FOR
PRODUCTION, DISSEMINATION AND EXCHANGE OF STATISTICAL DATA**

Invited Paper

Prepared by Dirk Hollmann and Ernst Schrey, Federal Statistical Office, Germany

I. INTRODUCTION

1. The internet as a medium of data exchange between the administration, businesses and citizens has already achieved a lot. The first phase of the eGovernment initiative “BundOnline 2005” in Germany has successfully been completed and several hundred services are online now. Based on the action plan E-Government of the European initiative i2010 and on the experience acquired with BundOnline 2005 and Deutschland-Online, the German Federal Government has adopted the program E-Government 2.0 on 13 September 2006 to concretise part of the overall strategy in the government programme “Future-oriented administration through innovation”.
2. The internet is thus intended to become the preferred communication and distribution channel for administrative services tailored to demand. The largest efficiency potentials can be developed at the interface between businesses and the administration through electronic linkage of their process chains.
3. The Federal Statistical Office is particularly committed to new and progressive IT technologies. Data production and dissemination are part of the core competence of the Office. For some years, data dissemination through the web has mainly been performed through the following applications: Website content, Statistics Shop and GENESIS-Online.
4. The website content provides a wealth of information on statistics and offers access to the other applications. The Shop offers publications as free electronic documents and as printed versions for order. GENESIS-Online allows online access to individually adjusted tables and permits to export data objects, to individually compile and access time series, and to research comprehensive metadata on the objects of official statistics.
5. Web Services are a modern technology that easily and efficiently supports the integration into processes. Making existing online databases available by Web Service is therefore an obvious choice. The

Federal Statistical Office as a modern information provider can thus efficiently and reliably be integrated into process chains of the administration and of businesses.

II. TECHNOLOGY

6. This section presents the two technologies “Web Service” and “GENESIS” before explaining the reason for combining the technologies to form the product “GENESIS-Webservices”.

A. Web Service

7. Since the early years of this decade, there have been specifications and first implementations allowing automated handling of services and business transactions through the internet. They are called Web Services.

8. Even though highly different approaches can be used to handle services via the internet, the term “Web Services” in the narrow sense has become to mean “automated service handling between applications on the internet”. So, what is exchanged between browser and web server is not the HTML documents that are common on the World Wide Web (WWW) and that are mainly used for presentation purposes, but rather documents in the XML format, which can integrate data and their descriptions in a structured manner. What is also used for Web Services is the data exchange protocols, usually http.

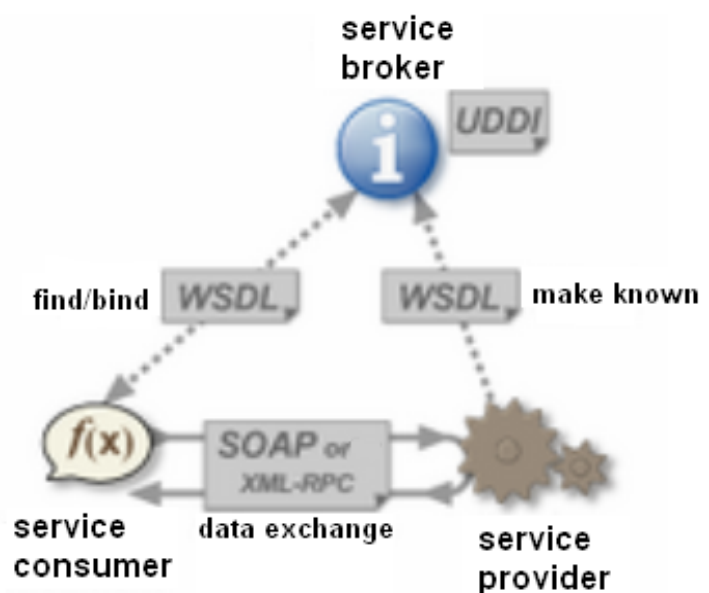
9. For those “Web Services” there are standards such as

- (a) the “business language” SOAP (Service Oriented Application Protocol),
- (b) the “language for services offered” WSDL (Webservice Description Language)
- (c) and “yellow pages” UDDI (Universal Description, Discovery, and Integration) to find services offered

10. Due to the platform-independent standards used, Web Services are able to decode remote method calls of any platform and to transmit them to an application. In that manner, a distributed architecture is created. The communication with Web Services is done through messages that can be transported through several protocols. Those standards have been implemented by several software providers, including open source providers.

11. The development of Web Services is supported, for example, by Java Webservices Development Pack (JWSDP) of the company Sun and AXIS, which is the SOAP 3.0 implementation of Apache. Possibilities to process the results supplied through SOAP are offered in Microsoft Office XP and 2003, in OpenOffice 2.0 and, of course, in individual applications.

Client programs generally send requests to a Web Service, which responds by giving the requested information. Many people therefore say that Web Services are for computers what web pages are for human beings. Even though this describes only part of the possibilities of the Web Services, the statement is true. Web Services are not intended for human users but for software systems exchanging data and/or calling functions on remote computers in an automated manner.



B. GENESIS

12. GENESIS is a statistical information system developed in common by the federal office and the offices of the Länder. GENESIS covers the database with statistical figures and a system of metadata including information on surveys, variables and their items, rules for deriving variables, units of measurement and the existing frames for standard tables.

13. The thesaurus includes all relevant catchwords (terms) required to access the information designed for the general public. Besides, information of the evaluation system like table definitions, result tables and results of retrievals are stored in GENESIS.

14. The logical data model is designed as a data cube, where each dimension (axis) is represented by a statistical variable, each point on an axis represents an item of the appropriate variable.

15. The services are based on a client /server architecture. The software is built in 3 tier architecture using DBMS ADABAS, programming language Natural, Entire X as middleware and Java, XSSL for the frontends. Until the installation of the Web Services, the access to data and metadata needed human interaction using one of the frontends.

16. The software of GENESIS is installed twice:

- (a) on the basic server system (LINUX) for the data and metadata import, internal use of GENESIS by subject matter statisticians, privileged users, data delivery service (data export) and as source system for GENESIS-Online
- (b) GENESIS-Online (UNIX Sun Solaris) for external access via the Internet (guest, customer, premium user) and data delivery service (user and order administration, handling of data packages).

For details of the configuration see figure in section 3.1.

C. Potential

17. A major advantage of Service Oriented Architectures (SOA) is that users (Consumers) can create complete applications by networking individual distributed Services. "Distributed" means that any Services of any operators (Exposers) at any location can be combined to form a System (Composer).

18. On the Composer side, the system looks like a homogeneous closed system. The system user does not realise that loosely coupled components are working in the background. The Consumer/Composer can choose purely by function the Services offered; due to the data exchange standards, it is irrelevant what there is in the background. A disadvantage is the dependence on the Exposer's operational readiness.

19. So, Services can be integrated irrespective of the platform and programming language. Also, the service provider can apply any changes to his infrastructure and can also change implementations; only the interface must be constant. It is very easy to create processes by linking up individual components performing sub-processes. Processes can be adjusted to future changes in a flexible manner.

20. The statistical information system GENESIS already offers centralised keeping of data and metadata as well as research and tabulation functions. GENESIS thus already covers important tasks to be performed by a modern statistical office. Many statistical processes require data and metadata and can only benefit from a centralised system with a standardised interface.

21. By providing "GENESIS-Webservices", both external customers of the Federal Statistical Office and internal processes can benefit from the advantages of service-oriented architectures and from those of GENESIS.

III. THE PRODUCT “GENESIS-WEBSERVICES”

22. With GENESIS-Online we already have an internet interface allowing human beings to create tables, export data and research metadata in a dialogue-based procedure.

23. With “GENESIS-Webservices” an infrastructure has been created allowing to offer, in addition to GENESIS-Online, the functions available there also for automated processing.

24. For customers who have frequently used GENESIS-Online or who have further processed in an automated way the data obtained, faster access to statistical data is thus provided; it involves just a one-time effort for interface programming and it minimises, or even reduces to zero, the regularly recurring work.

25. In addition to improving user comfort for external users of GENESIS-Online, Web Services provide a large potential for integrating the functions and data stocks from GENESIS into internal products and production processes.

26. Altogether 15 methods have currently been implemented, which are subdivided into 3 Services. The ExportService, for example, offers the following 3 methods:

ExportService

- (i) DatenExport
- (ii) TabellenExport
- (iii) Sodi

27. The description of the interface is stored in the freely accessible WSDL file.

```

- <wsdl:definitions targetNamespace="https://www-genesis.destatis.de/genesisWS/services/ExportService">
- <wsdl:types>
- <schema targetNamespace="daten.methods.webservice.genesis">
  <import namespace="https://www-genesis.destatis.de/genesisWS/services/ExportService"/>
  <import namespace="http://schemas.xmlsoap.org/soap/encoding"/>
- <complexType name="ReturnInfo">
  <sequence>
    <element name="code" type="xsd:byte"/>
    <element name="inhalt" nillable="true" type="xsd:string"/>
    <element name="typ" nillable="true" type="xsd:string"/>
  </sequence>
</complexType>
- <complexType name="Quader">
  <sequence>
    <element name="format" nillable="true" type="xsd:string"/>
    <element name="name" nillable="true" type="xsd:string"/>
    <element name="quaderDaten" nillable="true" type="xsd:string"/>
    <element name="returnInfo" nillable="true" type="tns1:ReturnInfo"/>
    <element name="stand" nillable="true" type="xsd:string"/>
    <element name="status" nillable="true" type="xsd:string"/>
  </sequence>
</complexType>
- <complexType name="QuaderAuswahl">
  <sequence>
    <element name="bereich" nillable="true" type="xsd:string"/>
    <element name="namen" nillable="true" type="xsd:string"/>
  </sequence>
</complexType>
- <complexType name="QuaderOptionen">
  <sequence>
    <element name="endjahr" nillable="true" type="xsd:string"/>
    <element name="format" nillable="true" type="xsd:string"/>
    <element name="metadaten" type="xsd:boolean"/>
    <element name="regionalSchluessel" nillable="true" type="xsd:string"/>
    <element name="sachMerkmal" nillable="true" type="xsd:string"/>
    <element name="sachSchluessel" nillable="true" type="xsd:string"/>
    <element name="sprache" nillable="true" type="xsd:string"/>
    <element name="stand" nillable="true" type="xsd:string"/>
    <element name="startjahr" nillable="true" type="xsd:string"/>
    <element name="werte" type="xsd:boolean"/>
    <element name="zeitscheiben" type="xsd:int"/>
    <element name="zusatz" type="xsd:boolean"/>
  </sequence>
</complexType>
- <complexType name="DatenExport">
  <sequence>
    <element name="quader" nillable="true" type="impl:ArrayOf_tns1_Quader"/>
    <element name="quaderAuswahl" nillable="true" type="tns1:QuaderAuswahl"/>
    <element name="quaderOptionen" nillable="true" type="tns1:QuaderOptionen"/>
    <element name="returnInfo" nillable="true" type="tns1:ReturnInfo"/>
  </sequence>
</complexType>
- <complexType name="StrukturElement">
  <sequence>
    <element name="auspraegungen" nillable="true" type="xsd:string"/>
    <element name="fb_art" nillable="true" type="xsd:string"/>
    <element name="fb_vr_ausg_art" nillable="true" type="xsd:string"/>
    <element name="insgesamt" type="xsd:boolean"/>
    <element name="name" nillable="true" type="xsd:string"/>
    <element name="nummer" nillable="true" type="xsd:string"/>
    <element name="regel" type="xsd:boolean"/>
    <element maxOccurs="unbounded" name="strukturElement" nillable="true" type="tns1:StrukturElement"/>
    <element name="strukturElemente" nillable="true" type="impl:ArrayOf_tns1_StrukturElement"/>

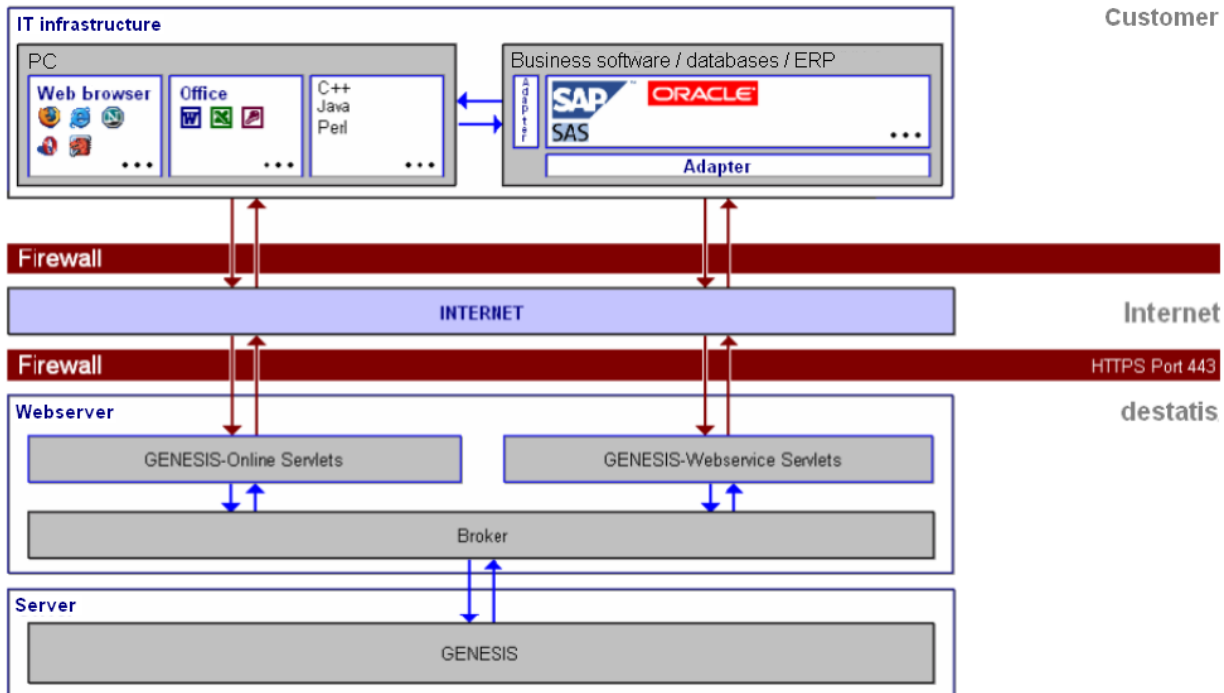
```

Extract from WSDL on the ExportService

A. Architecture

28. The infrastructure needed was already operational through GENESIS-Online. To implement/program the Web Services, two free software libraries (AXIS of Apache-Jakarta and DOM2 of w3c) have been integrated.

Overview of the entire infrastructure



B. Supporting statistical processes through GENESIS-Webservices

29. At the Federal Statistical Office and the statistical offices of the Länder, the standardisation of statistical processes has been defined by the working party SteP ("Standardisation of the production process"). In the following, some processes will be outlined to illustrate the use of "GENESIS-Webservices".

(a) Maintenance and use of metadata

30. GENESIS is a metadata system, thus supporting the storage of metadata and their relations. It is stored, for example, what variables have been covered for a specific statistics and what the legal basis of a statistics is like. Such metadata are crucial for many application cases in the Office and for online customers.

31. Therefore, data can be retrieved through Web Service from GENESIS. Some ministries (e.g. Federal Ministry of Economics) are productive users of that interface already today. The Federal Agency for Cartography and Geodesy is connected through Web Services to GENESIS as part of the geo portal. The German Bundesbank retrieves data from internal production.

(b) Tabulation

32. GENESIS is a tabulation tool. Tables can dynamically be created from the central data stock and be accessed. Pre-defined tables can be provided for use by any users.

33. Web Service users get access to those tables and, where required, can modify them, for example, through selection in terms of time or subject matter. The interface is widely used both within the Office and by big

external customers. In particular, Eurostat is connected as part of the SODI project (“SDMX Open Data Interchange”) on the exchange of indicators in the SDMX-ML format.

(c) Data provision

34. GENESIS allows access to data cubes and metadata also through Web Services. Numerous data objects are available for download. This is widely used both within the Office and by big external customers.

(d) Evaluation

35. GENESIS is a dynamic tabulation tool. During data access, the tables may be modified. For example, it is possible to include totals in the tables and to perform various calculations such as rebasing within a few seconds.

(e) Dissemination

36. GENESIS-Online is available on the internet as a partly chargeable system. Users may do research, compile tables and time series, etc. GENESIS-Webservices are available to a group of GENESIS-Online users as an automated interface.

C. Service Clients

37. It is not the task of the Web Service operator to take care of Clients. Instead, the purpose of Web Services is to allow individual and nearly any applications on the user side to get into direct contact with the Web Service. Individual distributed Services can thus be combined to homogeneous applications and be integrated into the user workflows.

38. For demonstration and to test the GENESIS-Webservice, various Clients have been developed as prototypes. In the following, the “GENESIS-WebClient” (Java), the “Excel-Webservice-Tool” and the access via script languages are explained as examples.

(a) GENESIS-WebClient

39. By means of the tool “wsdl2java” of Apache - AXIS it is possible to create a Java program from the WSDL file in an automated manner; that program will contain all data classes required and the communication. Optionally, the program can be extended to include input and output boxes and templates.

40. In this way we have created the “GENESIS-WebClient” for testing and demonstration purposes. The time required for development is about 1-2 days, with no previous knowledge on Web Services needed. The example shows the result of calling the research Web Service.

41. What is provided are objects (statistics, data cubes, time series, variables, and tables) on a search term, in this case “Hessen AND Schüler” (the German federal state of Hessen and “pupils”). The Service is primarily used by external portals wishing to integrate our data in their offers.

Data are entered in the dialogue boxes and the result is output in specific graphics components

The screenshot shows the GENESIS-WebClient interface. At the top, it says 'Statistisches Bundesamt Deutschland' and 'GENESIS-WebClient V2.231 beta'. There is a login section with 'Kennung:' (username) set to 'hollmann' and 'Passwort:' (password) set to '*****'. Below this is a 'Fehler:' (error) field. On the left, there are icons for 'Export', 'Recherche' (search), and 'Test'. The main area has tabs for 'GetRecords' and 'Recherche'. Under 'Recherche', there is a 'Suchanfrage' (search request) section with 'Suchstring:' (search string) set to 'Hessen and Schüler' and 'Kategorie:' (category) set to 'Tabellen'. A 'Suchen' (search) button is next to the search string. Below this is a 'Trefferliste' (hit list) section with tabs for 'Statistiken', 'Datenquader', 'Zeitreihen', 'Merkmale', and 'Tabellen(20)'. The 'Tabellen(20)' tab is selected, showing a table of results.

Name	Kurztext	€	Gewicht
211110001	Schüler: Bundesländer, Jahr		100.0
11111-0001	Gebietsfläche: Bundesländer, Stichtag	✓	100.0
12411-0009	Bevölkerung: Bundesländer, Stichtag		100.0
12411-0010	Bevölkerung: Bundesländer, Stichtag, Geschlecht		100.0
12411-0011	Bevölkerung: Bundesländer, Stichtag, Altersjahre		100.0
12411-0012	Bevölkerung: Bundesländer, Stichtag, Geschlecht, Altersjahre		100.0
12521-0005	Ausländer: Bundesländer, Stichtag, Geschlecht, Ausgewählte Staaten der Welt		100.0
12631-0004	Ehescheidungen: Bundesländer, Jahr, Gemeinsame minderjährige Kinder, Ehedauer		100.0
14111-0002	Bundestagswahl: Bundesländer, Stichtag, Parteien		100.0
14211-0002	Allgemeine Europawahl: Bundesländer, Stichtag, Parteien		100.0
14221-0002	Repräsentative Europawahl: Bundesländer, Stichtag, Parteien, Geschlecht, Altersgruppen		100.0
21111-0001	Schüler: Bundesländer, Jahr		100.0
21111-0002	Schüler: Bundesländer, Jahr, Geschlecht		100.0

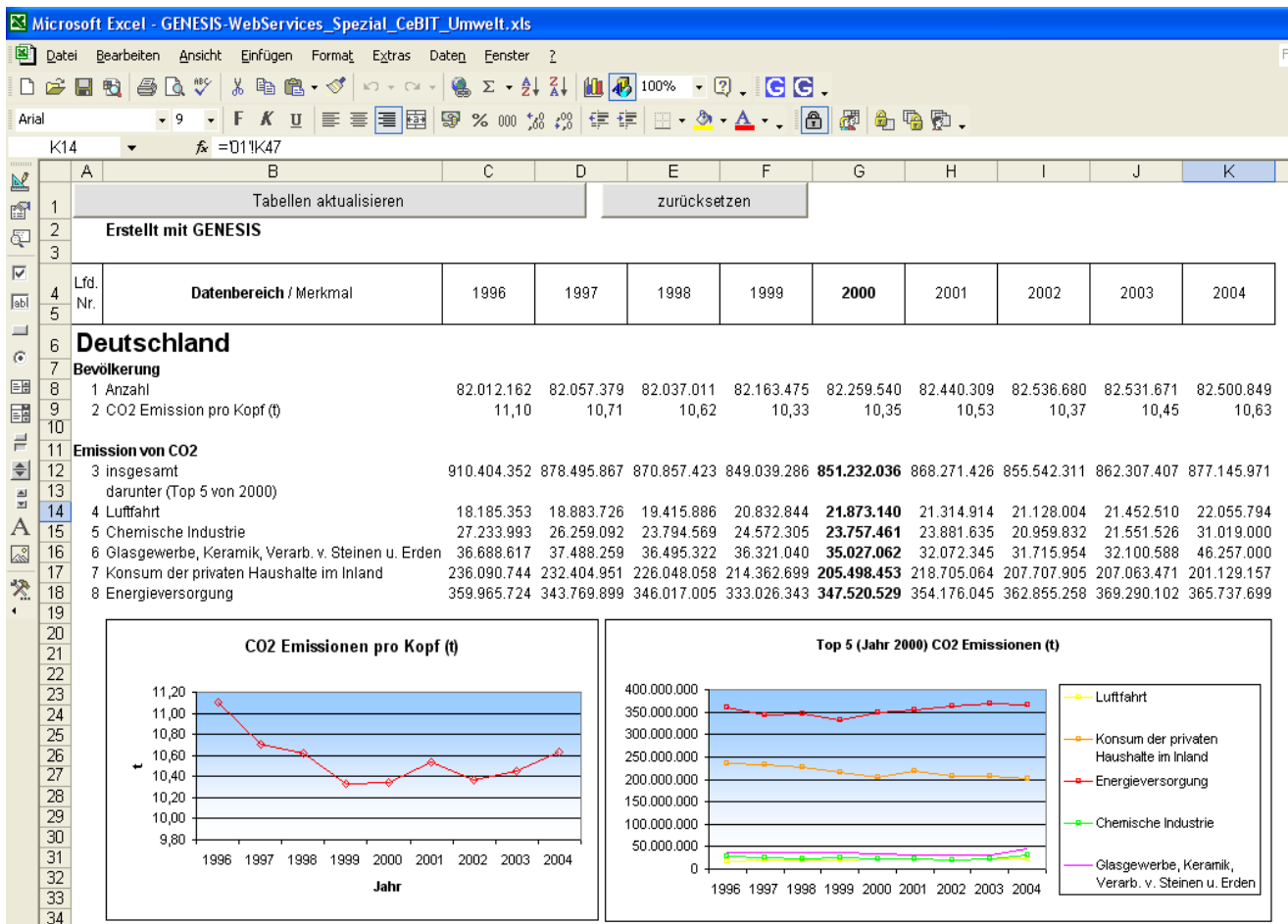
(b) Office products

42. In all reputable office programs (Microsoft, OpenOffice, etc.), Web Services can be used. For this purpose, Microsoft offers the PlugIn “Microsoft Office Webservice Toolkit” for free download. In an automated manner, the Toolkit creates data and communication modules for the call. Only the processing of the data obtained must be programmed by the user himself.

43. In the following, the application in Microsoft Excel will show opportunities available to users. The Excel table in the example contains several data tables which are filled through a Service call. The button “Tabellen aktualisieren” (update tables) on the screenshot below triggers centrally the updating of all tables.

44. The data sheet shown has been compiled through linkages with the data obtained. The VBA Macro in the background, which we programmed ourselves, can be re-used for many application cases. The effort required to create a data sheet like the one shown in the example below is less than one day even for untrained users. An update run will take about 15 seconds.

45. Other examples may show an entirely different run time behaviour because, for instance, extremely large tables with several million value cells can be retrieved, too.



Presentation as table and chart

(d) Script

46. Apart from the above-described Clients it is also possible to start Web Service calls by script direct from an operating system console. As a prototype, we have implemented a script in Perl. It can easily be started time-triggered and can thus be integrated into any processes, e.g. for automated data updating. What is needed is a Perl run time environment and the following very compact script:

Code

```
#!/opt/bin/perl
use strict;
use SOAP::Lite +trace => qw(debug);
my $kennung="TestIt";
my $pw="geheim";
my $server="01.02.03.04";
my $proxy="01.02.03.04";

$ENV{HTTP_proxy} = "http://$proxy"; # unsere Firewall

my $soap = SOAP::Lite
-> uri("http://$server/genesisWS/web/ExportService")
-> proxy("http://$server/genesisWS/web/ExportService");

my @params = (
    SOAP::Data->name(kennung          => $kennung),
```

```

SOAP::Data->name(password      => $pw),
SOAP::Data->name(namen          => "quader"),
SOAP::Data->name(bereich        => "alle"),
SOAP::Data->name(format         => "csv"),
SOAP::Data->name(werte          => "true")
    ->type('boolean'),
SOAP::Data->name(metadata       => "false")
    ->type('boolean'),
SOAP::Data->name(zusatz         => "false")
    ->type('boolean'),
SOAP::Data->name(startjahr      => "2004")
    ->type('string'),
SOAP::Data->name(endjahr        => "")
    ->type('string'),
SOAP::Data->name(zeitscheiben   => "1")
    ->type('string'),
SOAP::Data->name(regionalschlüssel => ""),
SOAP::Data->name(sachmerkmal    => "BILHS1"),
SOAP::Data->name(sachschlüssel   => "HS1720"),
SOAP::Data->name(stand          => "10.10.2001 09:00"),
SOAP::Data->name(sprache        => "de")
);

my $method = SOAP::Data->name('DatenExport')->attr({xmlns =>
"http://$server/genesisWS/services/ExportService"});

my $result = $soap->call($method => @params);
print "RESULT:", $result, "\n";
print "RC:", $result->valueof('//*/*/code'), "\n";

```

IV. SUMMARY

47. Web Services are a modern technology easily and efficiently supporting the integration into processes. Efficiently and reliably incorporating GENESIS-WebServices into process chains of the administration and of businesses is thus an obvious choice.

48. GENESIS-Webservices offer research and tabulation interfaces. It is possible to access metadata; tables can be accessed in an individualised manner and quickly.

49. Some partners in businesses and authorities as well as specialised internal users have been benefiting from the new access channel already since the middle of 2006.

50. Current developments at Eurostat aim at handling by means of Webservices the data supplies from national statistical offices of EU countries to Eurostat. With its central database GENESIS and the GENESIS-Webservices, the Federal Statistical Office is involved in the prototype production process of Eurostat.
