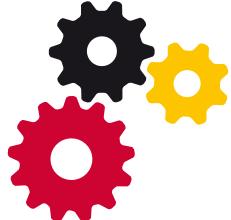


# SteP



2011

Statistisches Bundesamt

## **Your contacts around SteP**

**André Orth**  
E-mail: [andre.orth@destatis.de](mailto:andre.orth@destatis.de)

**Marion Rizzo**  
E-mail: [marion.rizzo@destatis.de](mailto:marion.rizzo@destatis.de)

**Published in May 2011**  
**Order number: 0006028-11900-1**  
**Photo copyrights: © Fancy / F1online / Creative Expressions**  
**© Statistisches Bundesamt**

**© Statistisches Bundesamt, Wiesbaden 2011**  
**Reproduction and distribution, also of parts, are permitted provided that the source is mentioned.**

---

Reasons for standardisation in official statistics	3
Information and contact	4
Glossary	5
New business process models	8
The DatML format family	9
Survey Database	10
Adobe InDesign	11
Basic Applications for Statistical Surveys (.BASE)	12
BASE.pleditor (Data Edit Designer)	13
BASE.sdfeditor (SDF Designer)	14
BASE.formulareditor (Form Designer)	15
BASE.statspez (STATSPEZ)	16
Secure data transmission through online reporting procedures	17
IDEV	18
eSTATISTIK.core (.CORE)	19
Conversion and Distribution Center (KonVertCenter)	20
Collection Network Database	21
PL-Ablaufumgebung (Data Edit Runtime Environment)	22
SAS	23
GENESIS	24

### Why SteP?

Increasing requirements to be met by official statistics and resources that have been shrinking for years require new ways of producing statistics. The statistical offices have therefore set themselves the goal of standardising and optimising sub-processes of statistics production and of creating (IT) tools which can be applied across various sets of statistics. The main purpose here is to optimally support the specialised units of the statistical offices in their work and reduce the burden on them – also regarding co-operation in the network of statistical offices. Consequently, the Standardisation of Processes (SteP) concerns all staff members of the statistical offices.

More concretely, SteP comprises two major tasks:

1. Describing and implementing standardised processes in statistics production
2. Developing relevant standard (IT) tools and interfaces

### The goals of SteP:

By process standardisation, the following goals should be achieved:

- Reducing the variety and complexity of statistics production
- Reducing the burden on respondents through online reporting procedures
- Reducing the burden on the specialised units by ...
  - process automation
  - earlier provision of procedures by using standards
- Providing standard tools and interfaces in the network of statistical offices that can be applied to various sets of statistics
- Improving quality and transparency in statistics production

### The history of SteP:

In many fields of statistics production, binding standards have already been applied in the network of statistical offices. Process standardisation started in 2003 when a relevant decision was taken as part of the Master Plan for the Reform of Official Statistics. To co-ordinate the plan across the network of statistical offices, the SteP Working Group was then set up. Members are representatives of IT and specialised departments of the statistical offices of the Federation and the Länder. Their tasks include testing and releasing SteP tools and developing guidelines for statistics production – the *Geschäftsprozessmodell der Amtlichen Statistik* (GMAS – German Business Process Model of Official Statistics).

### The implementation of SteP in the network of statistical offices:

Since the end of 2009, each statistical office has had a SteP officer. To optimally support the specialised units of the statistical offices of the Federation and the Länder in introducing SteP, the SteP officers provide information and give advice on the opportunities and requirements of standardisation. They make sure that SteP tools and interfaces are introduced where suitable and possible. In addition, they support the specialised units in compiling specifications. Involving the SteP officers of the Federation and the Länder in the development of IT projects is required by the *Vorgehensmodell der Amtlichen Statistik* (VMAS – Procedural Guidelines for Official Statistics) as amended in 2011. In this way, the standard tools presented are successively implemented in the statistical processes.

**About this brochure:**

This brochure addresses staff members involved in the production of statistics. It presents all SteP tools that are ready for production,<sup>1</sup> including their major functions and requirements, and shows how they interact in the SteP process chain.

The tool pages use colour-coded bars to indicate areas of application and targeted end users.

Blue, at the top of a tool page, stands for areas of application:

Not applicable	Applicable	Main area of application
----------------	------------	--------------------------

Green, at the bottom , stands for end users:

Not applicable	Potential user	Targeted user
----------------	----------------	---------------

Further help in reading the brochure is provided by an introductory glossary containing major terms around SteP and a summary of the file formats used in SteP.

**Information on SteP:**

To allow a clear presentation, the brochure does not go into all functions and tasks of the tools. Detailed information is contained in the *Geschäftsprozessmodell der Amtlichen Statistik (GMAS – German Business Process Model of Official Statistics)*, in which sub-processes and responsibilities are described. A detailed presentation of processes and parties involved in the implementation of SteP as part of an IT project is provided by the *Vorgehensmodell der Amtlichen Statistik (VMAS – Procedural Guidelines for Official Statistics)*.

**Your contact:**

The team of the SteP officer of the Federation from Division E 2 – Industry, Construction, Energy – will be pleased to help you by answering general questions on SteP, or on the implementation of SteP, and by indicating competent IT experts:

General questions on SteP: André Orth

Questions on online reporting procedures: Marion Rizzo

SteP officer of the Federation: Birgit Frank-Bosch

E-Mail: step@destatis.de

<sup>1</sup> Tools are considered as ready for production when they have been released by the SteP Working Group after the test and pilot phase.

The standardisation of statistical processes generally involves the development of modern specialised IT applications. This glossary contains major terms around SteP.

**AG-SteP**

SteP Working Group. Working group of the statistical offices of the Federation and the Länder on the “Standardisation of Processes”

**AKIT**

*Arbeitskreis Informationstechnik* (Information Technology Working Group). It gives advice and support to the LA OPTIKO and is responsible for awarding and accepting IT projects concerning the network of statistical offices (e.g. new programming) which do not meet the OPTIKO criteria and, consequently, are not considered as OPTIKO projects.

**Data edit specification**

A data edit specification comprises the full set of metadata managed by the Data Edit Designer, including definitions of variables, check and edit rules, and data edit cycles. It is used for performing data edits and generating web forms and a Java code that executes in the data edit runtime environment.

**Data pool**

A data pool is a central folder containing the common data stock of an organisational unit.

**Data cube**

A GENESIS data cube is a specific data format in which data for the GENESIS platform are produced and stored.

**DatML**

Family of data formats which are specifically designed to meet the needs of official statistics.

**DOI network**

The members of the network of statistical offices are linked up through the communication network of the *Deutschland Online Initiative* (Germany Online Initiative – DOI network). That administrative network is currently replacing the TESTA-D network which has been used up to now. In the DOI network, data are transmitted using line encryption, so that the transmitted data are protected against tapping and falsification up to the point of transfer.

**“Flat” files**

They contain a minimum data structure without metadata. Generally they are described through a data set definition (*Datensatzbeschreibung* – DSB). Examples of “flat” file formats are TXT or CSV.

**GMAS**

*Leitfaden Geschäftsprozessmodell der Amtlichen Statistik* (German Business Process Model of Official Statistics). It describes the existing business processes in German official statistics and defines responsibilities within the network of statistical offices and within the individual offices.

**GSBPM**

Generic Statistical Business Process Model. The GSBPM is an international model defining and describing the business processes of official statistics. It is a practical guide for the production of statistics. The model has been developed jointly by the UNECE, Eurostat and the OECD.

Hierarchical data set	Data set consisting of hierarchically structured statistical units, e.g. household – family – individual.
Interface	Contacts between elements within a software system. They connect individual IT tools with each other and define what data are exchanged between the tools in what way.
LA OPTIKO	<i>Lenkungsausschuss Optimierte Kooperation</i> (Steering Committee on Optimised Co-operation). It consists of the heads of all statistical offices and takes far-reaching decisions on OPTIKO projects (e. g. new programming). This includes awarding projects in a tender procedure, accepting products and abandoning projects.
Mapping	The linkup of variables of different data models.
Metadata	Data containing information on other data (e. g. on variables).
OPTIKO project	An IT project meeting three criteria: 1) it is not a pure maintenance and servicing order; 2) the end of the project is not close (at least 10 calendar months) and 3) larger project volume (estimated programming time of at least 13 person months).
Resources	Documents and files (e. g. survey definition, data set definition, data edit specification, etc.) which are needed for statistics production. The resources created are stored in the survey database under the relevant survey identifier.
Statistical unit	Carrier of statistical information, e. g. an individual, an enterprise or a public institution.
Statistics ID	The statistics identifier is a component of the survey ID. It is a 4-digit serial number for the permanent identification of surveys. When a new survey is created, the central survey administration allocates in a standard way a statistics identifier that is unique all over Germany and cannot be allocated a second time. When the statistics ID is allocated, it is referenced to the EVAS code (Uniform Classification of All Statistics).
SteP interfaces	Refers to the various DatML formats, CORE.connect and interfaces with input control system.
SteP officers	SteP officers inform the specialised units about SteP and give them advice. They make sure that SteP tools and interfaces are implemented where suitable and possible. In addition, they support the specialised units in developing requirement specifications.
SteP tools	Standard tools, applications and infrastructure components that can be applied across various sets of statistics and have been released by the LA OPTIKO.
Supply agreement	The supply agreement contains the subject-related specification regarding statistical reporting via an online procedure.

<a href="#">Survey definition</a>	The survey definition contains formal information on all variables collected from the respondents in a survey. It is created in the SDF Designer and is needed for the compilation of online reports and for the formal checking of incoming data.
<a href="#">Survey ID</a>	The purpose of the survey identifier is to clearly identify a survey and the relevant resources in the survey database. It consists, among other things, of the statistics ID, the time identifier and the Land code.
<a href="#">VMAS</a>	<i>Vorgehensmodell Amtliche Statistik</i> (Procedural Guidelines for Official Statistics). This document presents the bodies involved as well as the major requirements and procedural steps regarding software development and servicing in the network of statistical offices.
<a href="#">XML</a>	Extensible Markup Language. XML is an extensible and hierarchically structured data format that can be adjusted to requirements. XML files can contain any data structures (data and metadata) and can be extended at any time in any way.
<a href="#">ZPD</a>	<i>Zentrale Produktion und Datenhaltung</i> (Centralized Data Production, Management and Storage)

### The GSBPM<sup>1</sup>

The GSBPM is an international model defining and describing the business processes of official statistics. The model has jointly been developed by the UNECE, Eurostat and the OECD. The GSBPM provides practical instruction for the statistical production process. As there are in part considerable differences between various sets of statistics, the GSBPM allows flexible combination of its phases and components. The use of standard terms facilitates communication between statistics producers.

<sup>1</sup> GSBPM stands for Generic Statistical Business Process Model

### The *Leitfaden Geschäftsprozessmodell Amtliche Statistik (GMAS – German Business Process Model of Official Statistics)*

The GMAS is meant as a guide. It transfers the business processes of official statistics in Germany to the internationally recognised Generic Statistical Business Process Model (GSBPM) and describes them in detail, including the *Vorgehensmodell der Amtlichen Statistik (VMAS – Procedural Guidelines for Official Statistics)*. The description, first, defines responsibilities within the network of statistical offices and the individual offices. Second, it allocates standard tools to the various sub-processes. As the GMAS is based on practical work, it is regularly checked and updated.

#### Overview of SteP tools:

Classification to the GSBPM phases and allocation of user groups	Users				GSBPM phase								
	Specialised unit	IT organisers	IT developers	External/other users	1. Specify needs	2. Design	3. Build	4. Collect	5. Process	6. Analyse	7. Disseminate	8. Archive	9. Evaluate
Survey Database													
Adobe InDesign													
BASE.pleditor	■	■											
BASE.sdfeditor	■	■											
BASE.formulareditor			■										
BASE.statspez	■		■							■	■		
IDEV													
.CORE													
KonVertCenter													
Collection Network			■										
Database													
PL-Ablaufumgebung	■												
SAS			■							■	■		
GENESIS		■	■	■						■			



This is where the tool is usually applied  
This is where the tool may be applied

**DatML** refers to a family of XML-based data formats specifically tailored to the needs of official statistics. A major purpose is controlling statistical applications. The manifold requirements and use scenarios have led to the development of specialised data formats, which are briefly explained below.

### DatML/Raw<sup>1</sup>

DatML/Raw is the format used for raw data. This is the format in which the data collected by means of the online reporting procedures of IDEV and .CORE are received at the statistical offices. A DatML/Raw file may contain any number of reports from different respondents, surveys and recipients. The KonVertCenter can automatically check them, split them up into individual reports and pass them on to the relevant offices. Finally they are stored in the Collection Network Database.<sup>2</sup> Since March 2011, DatML/Raw – together with the DatML/RES data format – has been a certified XÖV standard<sup>3</sup> under the name XStatistik.

### DatML/EDT<sup>4</sup>

The DatML/EDT XML format is used for storing data edit specifications. It implements the full data model of the Data Edit Designer, including definitions of variables, checks and edit rules, and data edit cycles. It is used for performing data edits and generating web forms and a Java code that executes in the Data Edit Runtime Environment.

### DatML/SDF<sup>5</sup>

DatML/SDF is the format used to store survey definitions. They are the machine-readable equivalent to supply agreements. A survey definition describes the data model of a specific survey, including all variables to be covered and the relevant formal information (e. g. length, data type of the variable). They are mainly used for the creation of statistical reports in the DatML/Raw format, especially when the .CORE online reporting procedure is used. Also, they are applied to check incoming statistical reports by means of the KonVertCenter. Survey definitions are created through the SDF Designer.

### DatML/ASK (under development)<sup>6</sup>

The DatML/ASK format is needed for the creation of online forms (IDEV, CORE.reporter). It contains information on variables covered as well as formats and design elements. The DatML/ASK files are generated through the two GUI and IDEV Form Designers.

### DatML/RES

The inspection of a DatML/Raw file, performed in the KonVertCenter, is recorded in a DatML/RES file. Such inspection protocols contain not only lists of unique error codes but also lists of the points in the DatML/Raw document where the errors occur. Respondents may view and download the inspection protocols compiled for their data reports.

### DatML/SET (under development)

DatML/SET is an XML format for storing data sets that have passed the data editing process and are considered correct and eligible for production, and information on the data editing process down to data edits performed on single data items. It serves as an interface between data processing and evaluation procedures.

<sup>1</sup> Cf. sections on IDEV, .CORE, KonVertCenter, Collection Network Database

<sup>2</sup> Data from document reading and data entered manually through the specialised applications are also stored in the DatML/Raw format in the Collection Network Database.

<sup>3</sup> XÖV stands for *XML in der öffentlichen Verwaltung* (XML in public administration).

<sup>4</sup> Cf. section on BASE.pleditor

<sup>5</sup> SDF stands for Survey Definition Format.  
Cf. section on BASE.sdfeditor

<sup>6</sup> Cf. section on BASE.formulareditor

## Survey Database

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
Through the <b>survey database</b> , the processes regarding the production of a specific set of statistics are co-ordinated within the entire SteP process chain. It thus makes a major contribution to the automation of collection processes. In the survey database, all resources regarding a set of statistics are stored and managed. An interface allows users to store new or updated resources in the survey database or to retrieve them.								
First of all, a Germany-wide unique and standard statistics identifier (Statistics ID) and a survey identifier are issued for a specific set of statistics. <sup>1</sup> The statistics identifier is part of the survey identifier and will later be used to identify the incoming data. All resources <sup>2</sup> stored for a survey are managed under the survey identifier.								
The survey identifier has 13 digits and is composed as follows:								
Digit 1: version of the identifier <sup>3</sup>								
Digits 2-5: statistics identifier <sup>4</sup>								
Digits 6-11: time identifier <sup>5</sup>								
Digits 12-13: Land code <sup>6</sup>								
A distinction is made between an internal and a public survey database. The public database can be accessed by external users. It shows only contents that have been released for the general public. <sup>7</sup> The internal database contains all resources filed, including those not released. It can be accessed only by the statistical offices.								
Selected users from the specialised units get access rights for their sets of statistics so that they can perform changes and updates. <sup>8</sup>								
<p>The screenshot shows a search interface for survey identifiers. The top navigation bar includes links for 'Startseite' and 'Erhebungen'. The main area displays a table of search results with columns for 'EVAS', 'ErhebungslID', 'Beschreibung', 'Land', 'Periode', 'Gültig ab', and 'Status'. One row is highlighted in blue, showing the identifier '101011000099' and its details: 'Jahreserhebung für Betriebe im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden'. A sidebar on the right titled 'Ansicht' provides navigation options like 'Erhebungen nach EVAS', 'Erhebungen nach StatistikID', 'EVAS-Liste', and 'Filterfunktionen'. Below the main table, there's a section titled 'dazugehörige Ressourcen' with tabs for 'Erhebungsbeschreibung', 'Liefervereinbarung', etc., and a detailed view of the '99 - Erhebungsbeschreibung (DatML/SDF)' entry for identifier '000001'.</p>								
External users can access the survey database via internet at: <a href="https://erhebungsdatenbank.destatis.de/">https://erhebungsdatenbank.destatis.de/</a>								
Specialised units	IT organisers	IT developers	External users					

<sup>1</sup> The request for issue of an identifier is made through an electronic form in the survey database.

<sup>2</sup> Survey definition, data set definition, supply agreement, electronic questionnaire, data edit specification, subject-related information, etc.

<sup>3</sup> The version distinguishes between document identifiers already in use and the new standard survey identifier. Existing document identifiers get the version 0, while the new standard survey identifier gets the version 1.

<sup>4</sup> Non-systematic, serial and unique number allowing permanent identification of surveys. The statistics identifier is issued by the central survey administration. When an identifier is issued by the system, it is linked to the EVAS code.

<sup>5</sup> The time identifier defines the reference period from which the stored resources can be used. It is filed in the form <yyxttt>, with

yy = year (00–99)  
x = 0 → year  
x = 1 → half-year  
x = 2 → semester  
x = 3 → quarter  
x = 4 → month  
x = 5 → week  
x = 6 → day  
ttt depending on x (000–366)

<sup>6</sup> Possible values:  
00 – Federation  
01 to 16 – Länder  
99 – Network of statistical offices

<sup>7</sup> Green checkmark and globe in the internal system.

<sup>8</sup> Requests are made through a form provided in the survey database.

## Adobe InDesign

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
---------------	--------	-------	---------	---------	---------	-------------	---------	----------

**Adobe InDesign** is a professional layout and composition program of Adobe Systems. It is mainly used in communication units, print shops and publishing houses. InDesign is used together with other applications to produce brochures, catalogues, periodicals and other print media.

It allows creating multi-page layouts in any page format, which can then be used as a standard. Many common formats such as PDF can be processed and exported by InDesign.

In official statistics, InDesign is applied as a standard tool in several cross-section units. It is used to produce and design not only official publications, flyers, brochures and logos but also documents for printing paper questionnaires. The paper questionnaires are standardised, i.e. they have all the same layout, irrespective of what survey they are intended for or which statistical office sends them out.

 <b>Ihre Ansprechpartner rund um SteP</b> <p>Armin Bellwied Telefon: +49 (0) 411 / 75 35 27 E-Mail: armin.bellwied@destatis.de</p> <p>Axelle Orth Telefon: +49 (0) 411 / 75 43 77 E-Mail: axelle.orth@destatis.de</p> <p>Mariam Rizzo Telefon: +49 (0) 411 / 75 47 32 E-Mail: mariam.rizzo@destatis.de</p>	 <b>Produktionsreife SteP-Werkzeuge</b> <p>Adobe-InDesign Leistungssystem zur Erstellung von standardisierten Papierfragebögen in Form einer Druckvorlage.</p> <p>BASE Enthält standardisierte Werkzeuge für die Spezifizierung von Datensstrukturen und Platzhalterangaben sowie für die Produktion und Präsentation von statistischen Auswertungen. Die einzelnen BASE-Werkzeuge, wie der</p>	 <b>Erhebungsdatenbank</b> <p>Informationsbestandteile zur zentralen Verwaltung und Verarbeitung von bundesweit einheitlichen Erhebungsdaten sowie zur zentralen Verwaltung und Bereitstellung von Dokumenten und Spezifikationen für interne und externe Nutzer.</p> <p>Konvergenzprotokoll</p> <p>Hier werden einzelne Meldedaten geprägt, in Einzelmeldungen zerlegt und dann in ein „CORE“-Datenschema überführt. Außerdem wird hier ein Konvergenzprotokoll für Melde-Meldungen, die über den zentralen Dateneingang von CORE eingegeben, wenden an die entsprechenden Stellen im Verbund verteilt.</p> <p>SAS</p> <p>Standardwerkzeug zur Datenauswertung mit Komponenten zum Datenmanagement, zur Durchführung von tabellarischen Auswertungen und zur zentralen Datenanalyse. Das System ist modular aufgebaut und ermöglicht die Auswertungen auf der Grundlage von Datenbanken und Metadaten sowie die flexible Zusammenstellung von Beziehungen.</p> <p>GENES</p> <p>Informationssystem für Daten und Metadaten der amtlichen Statistik. Neben der individuellen Tabellierung können die Daten ausreichend visualisiert werden. Zur Integration in andere Systeme stehen vielfältige Schnittstellen zur Verfügung.</p>
 <b>Statistisches Bundesamt</b>		
<b>Kostenstrukturerhebung für das Jahr 2010</b> bei Unternehmen der Energiesversorgung, Wasserversorgung, Abwasser- und Abfallentsorgung, Beseitigung von Umweltverschmutzungen		
Statistisches Bundesamt, Gruppe E2, 65180 Wiesbaden		
 <b>Rücksendung bitte bis</b> <b>Ansprechpartner/-in für Rückfragen (freiwillige Angabe)</b> Name: Telefon oder E-Mail:		
Sie erreichen uns über: Telefon: 0611 75-4209 Telefax: 0611 75-3961 E-Mail: kws-energy@destatis.de		
<small>Falls Anschrift oder Firmierung nicht mehr zutreffen, bitte auf der Rückseite kennzeichnen.</small>		
<b>081</b>		
 Unternehmensnummer <small>(Bei Rückfrage Ihre Angabe)</small>		
 Land		
 Ihre Daten können Sie auch online unter <a href="http://www.idav.destatis.de">www.idav.destatis.de</a> melden.		
 Ihre Zugangsdaten (Kennung und Passwort) stehen auf dem Anschreiben. Weitere Informationen entnehmen Sie unserer Homepage <a href="http://www.destatis.de">www.destatis.de</a>		
<b>online</b>		
<b>Beachten Sie folgende Hinweise:</b> <p>All die Angaben sind für das Gesamtunternehmen ohne Zweig- oder Niederlassungen oder Unternehmesteile im Ausland zu machen.</p> <p>Soweit Ihr Unternehmen verschiedene Bereichen der Energie-, Wasserversorgung, Abwasser- und Abfallentsorgung, Beseitigung von Umweltverschmutzungen oder ganz in anderen Wirtschaftsbereichen tätig ist, bitten wir zusätzlich für einen Teil der Merkmale die Angaben auf dem Berichtsformular für fiktive Unternehmen vorzunehmen.</p> <p>Berichtszeitraum ist das Kalenderjahr 2010. Deckt sich das Geschäftsjahr nicht mit dem Kalenderjahr, so ist das</p>		
<small>Geschäftsjahr zu Grunde zu legen, das im Laufe des Jahres 2010 zu Ende ging. In das Geschäftsjahr sind höchstens 12 Monate einzubeziehen.</small> <p>Sollte der endgültige Jahresabschluss noch nicht vorliegen, geben Sie bitte die Werte aus den entsprechenden Konten oder erörtern Sie dies mit dem Beauftragten.</p> <p>Rechtsgrundlagen und weitere rechtliche Hinweise entnehmen Sie der beigelegten Unterlage, die Bestandteil dieses Fragebogens ist. Bitte beachten Sie bei der Bearbeitung der Fragebögen die Angaben im Kapitel 2 bis 6 in der separaten Unterlage.</p> <p>Vielen Dank für Ihre Mitarbeit.</p>		
<b>A Tätige Personen, Ende September des Geschäftsjahrs</b>		
1 Tätige Inhaber/-innen, tätige Mitinhaber/-innen sowie unterzellt mitarbeitende Familienangehörige 05		
1.1 darunter: weiblich 09		
2 Arbeitnehmer/-innen 06		
2.1 darunter: weiblich 14		

**BASE** stands for “Basic Applications for Statistical Surveys” and currently comprises four standardised tools for designing, building and evaluating statistical surveys. The tools are the following:

- the **BASE.pleditor** (Data Edit Designer) for the specifications of data edit checks and edit cycles,
- the **BASE.sdfeditor** (SDF Designer) for the specification of the data to be supplied through online reporting procedures<sup>1</sup> (survey definitions),
- the **BASE.formulareditor** (Form Designer) for the production of online forms<sup>2</sup>,
- and **BASE.statspez** (STATSPEZ) for the specification, production and presentation of statistical evaluations.



The development of .BASE tools is closely connected with the development of SteP.

In the context of automating statistical processes, the Federal Statistical Office first of all developed specific XML data formats<sup>3</sup> which are used as a standard. The resources required for statistics production, such as the survey definition, the data edit specification and electronic questionnaires, are produced in those formats and are then filed in the survey database.

To produce and maintain those resources, the above tools have been developed and added to the .BASE system.

Through a central server at the Federal Statistical Office, the resources produced are automatically and quickly distributed to the .BASE servers at the statistical offices of the Länder and can directly be used there. For example, existing data edit specifications from the Data Edit Designer can be used across specialised units and in the entire network of statistical offices. The joint development and use of standard resources and of processing and evaluation programs strongly supports co-operation within the network of statistical offices.

One of the components of the .BASE system is an integrated user and object management system which is used to manage different options of handling and using objects. For example, not all users are authorised to create or change objects. But authorised users can make objects available to other users in the network of statistical offices.<sup>4</sup>

On the following pages, the individual .BASE tools are presented in detail.

<sup>1</sup> .CORE, IDEV.

<sup>2</sup> IDEV, CORE.reporter.

<sup>3</sup> DatML, TabML.

<sup>4</sup> For example, in practical work, an office's specialised department responsible for a specific set of statistics could develop and maintain data set definition(s) and relevant specifications, use those items itself for the production of statistical tables and/or make them available to an office doing programming work.

## BASE.pleditor (Data Edit Designer)

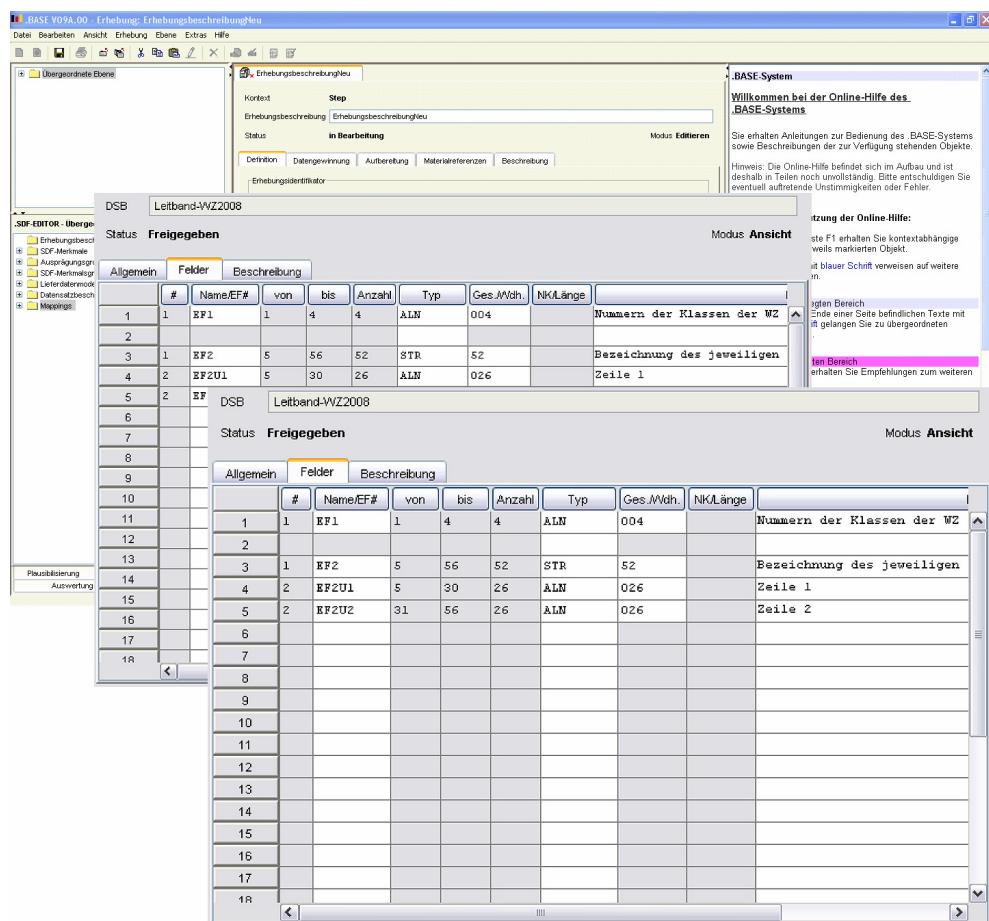
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate				
<p>The <b>Data Edit Designer</b> is an element of the .BASE system. The tool enables the specialised units to create specifications of edit checks by themselves, to maintain them and apply them directly in ongoing processing work.</p> <p>The data edit checks and edit cycles (data edit specification<sup>1</sup>) created in the Data Edit Designer are needed for online data collection<sup>2</sup> and for the processing and evaluation programs<sup>3</sup>.</p> <p>Data edit checks and edit cycles are specified in the Data Edit Designer through a specific user-friendly language. As the .BASE systems of the individual statistical offices are networked, several statistical offices can jointly develop data editing.<sup>4</sup> First of all, the survey variables are defined and grouped into units belonging together in terms of contents (subject fields). Generally, such units correspond to sections of a questionnaire. When a variable is created, checks of codes and value ranges can already be defined and the relevant error texts and correction instructions be described. Combined checks are set up within the subject fields.</p>												
<p>The screenshot shows the 'PL-EDITOR - Haushalterhebungen' window. On the left, there's a tree view of survey variables under 'Haushalterhebungen'. The main panel displays a form for 'm_BausparGuth' with fields like 'Kontext: Haushalterhebungen.EVS-2008', 'Merkmal: m_BausparGuth', 'Status: Freigegeben / Fehlerfrei', and 'Definition' tabs. It also shows 'Ausprägungen' (representations) and 'Wertebereich' (value range) settings. A right-hand sidebar provides help for the online editor, including tips for using F1, blue text for help links, and color-coded regions for specific instructions. At the bottom, there are tabs for 'Auswertung', 'Benutzer', 'Plausibilisierung', 'PL-Struktur', and 'Lieferformat'.</p>												
<p>In a data edit cycle, the specialised unit then describes whether, and in what sequence, the edit checks should be carried out. Different data edit cycles can be generated for different application contexts, e.g. for data collection from respondents, data entry only with value range checks, checking a data set and automated corrections.</p> <p>Once defined, a subject field – including its variables, edit checks and edit cycles – can be reused in the same or in other surveys. This supports the subject-related harmonisation of surveys and reduces the burden on the specialised units.</p> <p>The checks and cycles created in the Data Edit Designer can be tested in the data edit runtime environment before they are released for production.<sup>5</sup></p>												
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #009640; color: white; padding: 5px;">Specialised units</td> <td style="background-color: #009640; color: white; padding: 5px;">IT organisers</td> <td style="background-color: #009640; color: white; padding: 5px;">IT developers</td> <td style="background-color: #009640; color: white; padding: 5px;">External users</td> </tr> </table>								Specialised units	IT organisers	IT developers	External users	
Specialised units	IT organisers	IT developers	External users									

## BASE.sdfeditor (SDF Designer)

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
---------------	--------	-------	---------	---------	---------	-------------	---------	----------

The **SDF Designer** is used by the IT organisers in close co-operation with the specialised units to create electronic supply agreements, also referred to as survey definitions.

It is possible in the SDF Designer to convert variables and subject fields defined in the Data Edit Designer into corresponding SDF variables and SDF variable groups and even to create and post-edit an entire supply data model. If no data edit specification is available yet, the variables and variable groups can be created directly in the SDF Designer.



When all SDF variables and variable groups have been created, they are combined in the supply data model now created.

Finally the survey definition (DatML/SDF) is generated. It is filed in the survey database and is available to both internal and external users<sup>1</sup>. When the survey definition is finished, mapping can be used to define the allocation of variables to the data set rows of the data set definition. On that basis, the INI Mapping can be generated for further handling. It is required by the KonVertCenters where DatML/RAW reports have to be converted into flat data sets<sup>2</sup>.

<sup>1</sup> External users can be either respondents or software companies which use the survey definition to develop modules for automated data collection by means of CORE.connect. See also section on “.CORE”.

<sup>2</sup> E.g. the CSV format. See also the section on “KonVertCenter”.

## Data Edit Runtime Environment

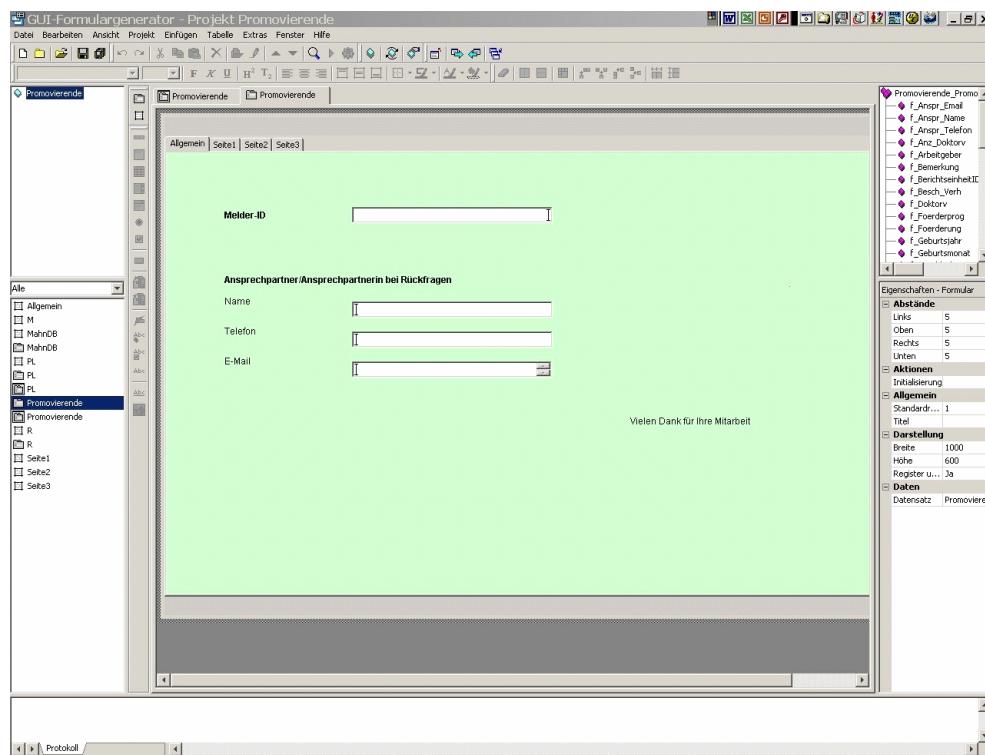
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
---------------	--------	-------	---------	---------	---------	-------------	---------	----------

The **Form Designer** is a tool that is mainly used by IT developers at the statistical offices to create electronic forms for online data collection and for the input interface of the data edit runtime environment.

Working with the Form Designer is based on the data edit specification<sup>1</sup>, which has been created by the specialised units in the Data Edit Designer before. It is indispensable for creating forms.

The Form Designer is currently available in two versions. The GUI Form Designer<sup>2</sup> is needed in particular where special adjustments<sup>3</sup> have to be made because the standard forms for CORE.reporter or for the data edit runtime environment do not meet a survey's requirements. The IDEV Form Designer is needed for creating IDEV questionnaires.

Before compiling a form, the project structure has to be created. Within a project, all relevant data<sup>4</sup> are stored. The graphical interface of the Form Designer allows quick and easy compilation of forms. Those electronic forms may contain text, specific formatting requirements and charts.



To avoid errors, the forms thus created are automatically checked for completeness and consistency. If the forms are free from errors, the final form can be generated in the DatML/ASK format through the form generator and be filed in the survey database. The data produced can be stored in a project file, so that the project can be used again later. Any subsequent changes in the form can then easily be done.

<sup>1</sup> DatML/EDT file.

<sup>2</sup> GUI = Graphical User Interface.

<sup>3</sup> Special adjustments may refer to special formatting, display as registers, charts, explanatory texts, etc.

<sup>4</sup> Including imported data edit specification, created objects and additional information.

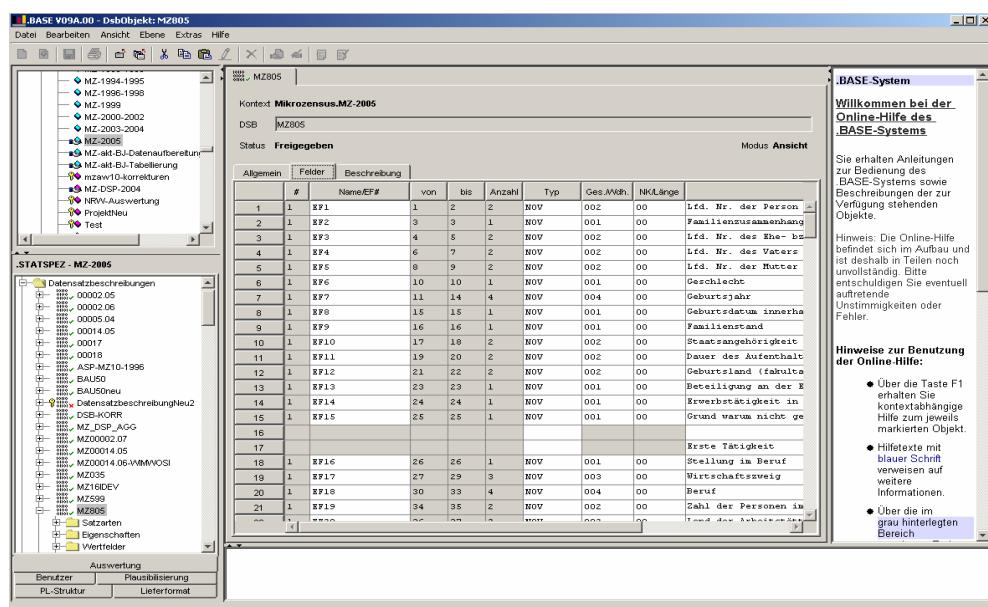
## BASE.statspez (STATSPEZ)

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
---------------	--------	-------	---------	---------	---------	-------------	---------	----------

**STATSPEZ** stands for “statistical table specification” and is a tool of the .BASE system that can be jointly used in the network of statistical offices. It offers manifold functions for the specification, production and presentation of statistical evaluations in the form of tables.

It is closely connected with the other .BASE tools. The dataset definitions and groups of possible values produced in the Data Edit Designer or the SDF Designer can directly be used in STATSPEZ and vice versa.

Although STATSPEZ works with hierarchical data sets<sup>1</sup>, it is also suited to process flat data sets and data sets designed for mainframes. It does not have a database of its own for the data to be evaluated. Instead, it uses data files stored in the file system of the server or computer.



Apart from data processing and specifying the table structure and table layout, STATSPEZ may also be used to calculate the primary confidentiality in table production.

After evaluation, the aggregated data can be further processed<sup>2</sup> as tables with results (ready for printing) or in various other formats<sup>3</sup>. In addition, the GENESIS import format is supported.

The specifications compiled by STATSPEZ, in particular regarding evaluation programs and production jobs, can be used in a standardised way in the entire network of statistical offices. Each organisation has a production system of its own to perform production jobs. Users may choose whether they have production jobs run on the central production server, on a local computer or, as part of ZPD, on a third-party<sup>4</sup> production server.

Production jobs can be done regularly, i.e. at fixed points in time, in an automated manner without users having to trigger them again.

<sup>1</sup> See glossary

<sup>2</sup> For further graphical processing, the finished tables can be exported into other applications, e.g. Excel.

<sup>3</sup> TabML, HTML, CSV, RegioStat, PDF, Excel, Word.

<sup>4</sup> Production server of another organisation.

### Sensitive data require sensitive handling

The activities of the statistical offices are often based on highly sensitive data and many respondents are very reluctant to disclose them. Confidentiality of the data is indispensable, not least of all because of the provisions of the Federal Statistics Law. Also, to protect mutual trust between respondents and statistical offices, it is indispensable to handle the received data with extreme care.

### Risks in data transmission

It must therefore be ensured that the data to be processed are complete and unaltered. Damage arising from a data leak can differ considerably, so that general quantification is not possible. Consequences may include financial losses for respondents, a general loss of trust in, and reputation of official statistics, or an unnecessary burden on colleagues.

There are security risks involved in the transmission channel and in the formats transmitted:

- Special risks are involved in accepting electronic data reports which are not sent through secure transmission channels but e.g. are received on data media by post or in e-mail attachments. Reports may not only be intercepted, read and changed by third parties on their way from the respondents to the statistical offices, but they can also get lost when sent by post. Therefore those transmission channels should be replaced as soon as possible by one of the secure online reporting procedures.
- In addition, data received in an undesired format involve considerable security risks for the entire IT system of the statistical offices. Many formats such as MS Excel files, but also seemingly harmless image files may contain viruses that are able to quickly infect large parts of the system. Such a virus poses a threat not only to the security of the system as such but possibly also to the confidentiality of the data stored in the system. Therefore, the transmission of such risky formats should be replaced as soon as possible by unproblematic formats. What is best suited here is the use of a secure online reporting procedure.

### IDEV and .CORE offer secure transmission channels

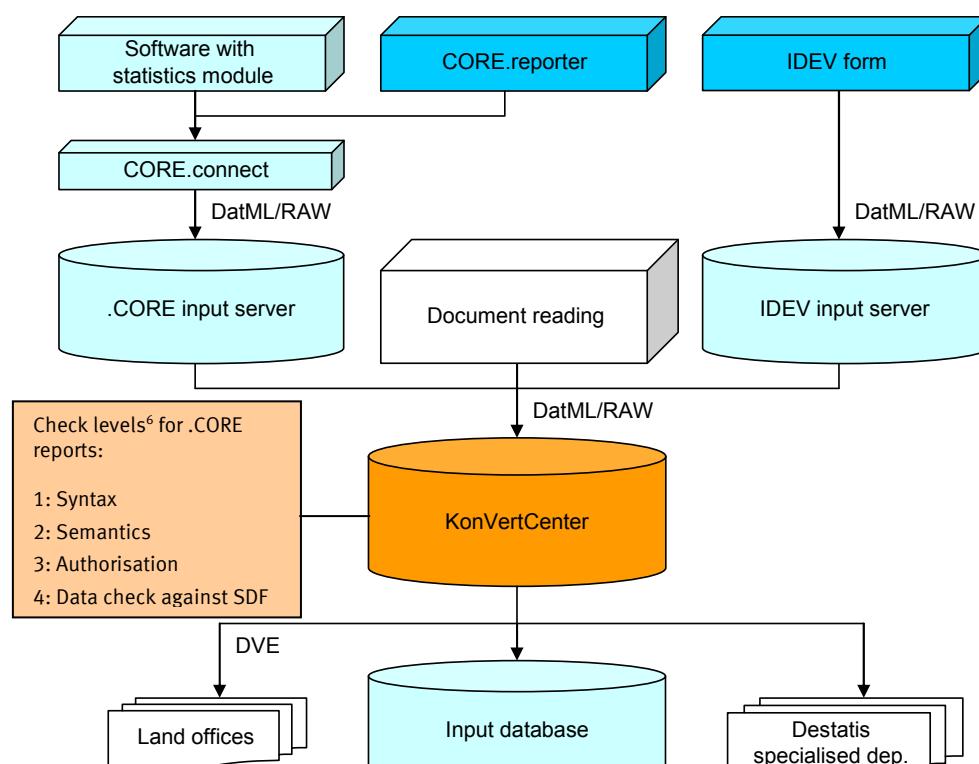
The online reporting procedures IDEV and .CORE allow encrypted transmission of statistical data in the certified standard data format DatML/Raw. They ensure the required data security and optimal support of the SteP process chain. Efforts are being made now to amend the Federal Statistics Law. The goal is that statistical reports should be transmitted in the format required by the statistical offices and through the online reporting procedure offered for the relevant set of statistics. This way of data transmission should be compulsory for public agencies.

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
<p><b>IDEV</b> is an online reporting procedure developed jointly by the statistical offices of the Federation and the Länder and has been used since 2005. It allows form-based manual data input and also data import from a CSV file into the form. IDEV offers many support functions (see box), which considerably reduce the work and time required from the respondents. Data are transmitted through an encrypted connection directly to the IDEV server of the relevant statistical office. There the reports are received in the DatML/Raw format. Every respondent gets individual login details<sup>1</sup>, which provides additional data security.</p>								<p><sup>1</sup> Login details are user ID and password. In contrast to .CORE, they are valid only for a specific set of statistics. They are generally printed on the paper questionnaire or on the letter.</p> <p><sup>2</sup> DatML/EDT file. It is used as an import interface for the development of the forms. The data edit specification is developed in the Data Edit Designer.</p> <p><sup>3</sup> <a href="http://www-idev.destatis.de">www-idev.destatis.de</a></p> <p><sup>4</sup> The scope of the subject-related data edit checks to be carried out in the form by the user is defined by the specialised unit in the Data Edit Designer. The structure and behaviour of the form must be co-ordinated with the IT developers.</p> <p><sup>5</sup> User ID: guest Password: guest For decentralised sets of statistics, the IDEV input servers of the statistical offices of the Länder can be used in the same way.</p> <p><sup>6</sup> Reports transmitted through the guest login are automatically deleted at the central data portal.</p>
<p><b>Internet Datenerhebung im Verbund IDEV</b></p>								
<p>All forms offered through IDEV have the same layout and the same standard user interface. IDEV forms are programmed by IT developers at the Federal Statistical Office or at the Land statistical office responsible for the programming. For the purpose, they need the data edit specification<sup>2</sup> previously created by the specialised unit in the Data Edit Designer. Programming is then done in co-operation with the organisers and the specialised department on the basis of the supply agreement compiled. Finally the AG Design (Design Working Group) checks the IDEV forms with regard to the design standards and releases them.</p>								
<p><b>Benefits of IDEV</b></p> <p>Apart from reporting statistical data, the following services are offered by IDEV:</p> <ul style="list-style-type: none"> <li>• Services for address and contact changes</li> <li>• Information on classifications and legal bases</li> <li>• Pre-filled boxes on the form for recurring data</li> <li>• Pre-checks of completed forms to reduce subsequent queries</li> <li>• Download of user-specific data orders</li> <li>• Archiving and/or printing of reports</li> <li>• E-mail reminder service for fixed reporting deadlines</li> <li>• Lists of reports in the respondent's account</li> <li>• Accessible forms</li> </ul>								
<p>When the finished forms have been released by the AG Design, they are made available on the IDEV web page of the Federal Statistical Office<sup>3</sup> (for centralised surveys) or on the IDEV web pages of the statistical offices of the Länder (for decentralised surveys).</p> <p>In the IDEV form, edit checks can be carried out before data transmission to avoid queries. The specialised unit is responsible for defining the type and scope of edit checks<sup>4</sup>. Before transmission, respondents are informed about boxes filled in incorrectly or incompletely, so that they can correct their data.</p> <p>To test the IDEV system, a guest login<sup>5</sup> is available at <a href="http://www-idev.destatis.de">www-idev.destatis.de</a>, with which all functions can be tested.<sup>6</sup></p> <p>IDEV is now offered for more than 200 centralised and decentralised sets of statistics. With over 400,000 respondents transmitting some 14 million data sets per month through IDEV (March 2011), that reporting channel has achieved wide acceptance.</p>								
Specialised units	IT organisers	IT developers	External users					

## eSTATISTIK.core (.CORE)

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
<p>.CORE stands for “Common Online Rawdata Entry” and has been used as an online reporting procedure since 2005. It supports respondents in the automated retrieval of data requested for statistics from software systems or from other data sets that can electronically be evaluated. It also supports grouping of the data into statistical reports and their encrypted transmission to a statistical office. After successful transmission, the respondent will receive an inspection protocol.<sup>1</sup></p> <p>As in IDEV, the data are transmitted in the standard supply format of official statistics (DatML/Raw). Individual login details<sup>2</sup> (user ID and password) provide additional data security.</p> <div style="background-color: #e6f2ff; padding: 10px;"> <p><b>Benefits of .CORE</b></p> <p>Widespread use of .CORE involves the following benefits:</p> <ul style="list-style-type: none"> <li>• The entire survey process is automated.</li> <li>• The burden on enterprises and public agencies obliged to provide information is reduced.</li> <li>• Earlier data supply and improved data quality of the statistical reports received</li> <li>• Reports can be processed without switches between media in the SteP process chain.</li> <li>• Extended range of services offered by the statistical offices</li> </ul> </div> <p>Respondents may use three channels for data transmission:</p> <p><b>1. CORE.connect</b></p> <p>To transmit .CORE reports, respondents need a statistics module which supports them in compiling and transmitting the requested data. Software companies<sup>3</sup> develop such modules for their products<sup>4</sup>. They control the correct selection of data for the set of statistics chosen and convert the compiled data into the DatML/Raw format. For data transmission, the Federal Statistical Office provides the free communication interface CORE.connect, which is integrated into the modules. Modules are developed on the basis of the survey definition<sup>5</sup>. After transmission, .CORE reports are forwarded to the KonVertCenter of the Federal Statistical Office. There they undergo formal checks<sup>6</sup> and are then forwarded to the relevant statistical office.</p> <p><b>2. Data transmission through the eSTATISTIK.core homepage</b></p> <p>Respondents producing reports in the DatML/Raw format by means of a statistics module which does not have the communication interface CORE.connect can transmit their reports in a secure way to the statistical offices through <a href="http://www.statspez.de/core/uebermitteln.html">http://www.statspez.de/core/uebermitteln.html</a>.</p> <p><b>3. CORE.reporter</b></p> <p>Respondents not using any software with a statistics module can use the CORE.reporter application. Data can either be entered manually into a form or be imported into the form<sup>7</sup> through a CSV file for further handling or be transmitted directly<sup>8</sup>. As in IDEV, edit checks can be carried out in the form before data transmission.<sup>9</sup> Contrary to IDEV, CORE.reporter is also suited for storing, continuously handling and archiving data and reports.</p>	<p><sup>1</sup> The inspection protocols are created automatically in the KonVertCenter. See section on “KonVertCenter”.</p> <p><sup>2</sup> Different login details are needed for sending reports via .CORE and IDEV. Respondents wishing to send their reports via .CORE will receive their login details through online registration. Contrary to IDEV, those login details apply to any set of statistics and in any Land.</p> <p><sup>3</sup> A list of software companies offering statistics modules is available at <a href="http://www.statspez.de/core">www.statspez.de/core</a>.</p> <p><sup>4</sup> E.g. business or administrative software</p> <p><sup>5</sup> The survey definition is created in the SDF Designer. It is then filed in the survey database and made available for external users (green checkmark and globe).</p> <p><sup>6</sup> More information on the definition of “formal checks” and on how they are carried out is contained in the section on the “KonVertCenter”.</p> <p><sup>7</sup> Special requirements to be met by the form can be defined by the IT department of the Federal Statistical Office through the Form Designer in a DatML/ASK file.</p> <p><sup>8</sup> For direct transmission, the data must be available in the CSV format. Before transmission, they will be converted into the DatML/Raw format, just like the form data.</p> <p><sup>9</sup> The scope of the edit checks to be performed in the form by the user is defined by the specialised unit in the Data Edit Designer through a separate data edit cycle.</p>							

## Conversion and Distribution Center (KonVertCenter)

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
<p>The KonVertCenter processes reports received through the two online reporting procedures .CORE and IDEV or through document reading. It checks incoming data reports<sup>1</sup> from formal aspects<sup>2</sup>, divides them into individual reports and forwards them to the responsible statistical office<sup>3</sup>. In addition, where required, it converts incoming data into “flat” data formats. The KonVertCenter operates in a fully automated way in the background and therefore has no user interface.</p> <p>The IDEV reports received at the Federal Statistical Office are intended only for centralised statistics. They are forwarded through the KonVertCenter to the Collection Network Database or to the data pools<sup>4</sup> of the specialised departments. IDEV reports for decentralised surveys are received by the relevant statistical offices of the Länder. They have KonVertCenters and IDEV servers of their own.</p> <p>In contrast, .CORE reports are received at the central data portal of the statistical offices both for centralised and decentralised surveys. They undergo formal checks (see box “check levels”), divided into individual reports and forwarded either to the relevant Land statistical office<sup>5</sup> or to the Collection Network Database or the data pools of the specialised departments of the Federal Statistical Office.</p>								
 <p>Check levels<sup>6</sup> for .CORE reports:</p> <ul style="list-style-type: none"> <li>1: Syntax</li> <li>2: Semantics</li> <li>3: Authorisation</li> <li>4: Data check against SDF</li> </ul> <p>DVE</p> <p>Land offices</p> <p>Destatis specialised dep.</p> <p>Input database</p> <p>KonVertCenter</p> <p>CORE.connect</p> <p>CORE.input server</p> <p>Document reading</p> <p>IDEV input server</p> <p>IDEV form</p> <p>CORE.reporter</p> <p>Software with statistics module</p> <p>DatML/Raw</p> <p>DatML/Raw</p> <p>DatML/Raw</p> <p>DatML/RES</p> <p>DVE</p> <p>1 The reports are received in the DatML/RAW format.</p> <p>2 Formal checks are checks regarding the subject-related descriptions of the variables as defined in the survey definition or the supply agreement.</p> <p>3 Only .CORE reports are forwarded.</p> <p>4 Data pools are the specific data folders of the specialised units.</p> <p>5 Transmission is done through the Data Exchange Procedure in the network of statistical offices (<i>Datenaustauschverfahren im Verbund – DVE</i>).</p> <p>6 Currently all reports involving formal errors of check levels 1-3 are rejected at the data portal. According to a Federation-Länder decision, reports involving formal errors of check level 4 will also be rejected as from 2012.</p> <p>7 Inspection protocols are made available to the respondents either in the respondent's account on the IDEV server or direct in their .CORE statistics module. They are available in the DatML/RES format.</p> <p>8 For the purpose, a suitable INI file (“raw2flat”) has been stored.</p> <p>9 In the Collection Network Database, only DatML/Raw files are stored.</p>								

After checking the .CORE reports, the KonVertCenter automatically produces electronic acknowledgements, the inspection protocols<sup>7</sup>.

The KonVertCenter can convert DatML/Raw files into “flat” files (e.g. into the CSV format)<sup>8</sup> in case the specialised application used for data processing cannot import DatML/Raw files. However, the “flat” files cannot be stored in the Collection Network Database and must be stored in the data pools of the specialised departments.<sup>9</sup>

## Input database

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
---------------	--------	-------	---------	---------	---------	-------------	---------	----------

The **Collection Network Database** is a central infrastructure component in the process of data collection. It manages individual reports<sup>1</sup> across surveys and helps to organise data exchange between the statistical offices in the data collection phase, including full automation.

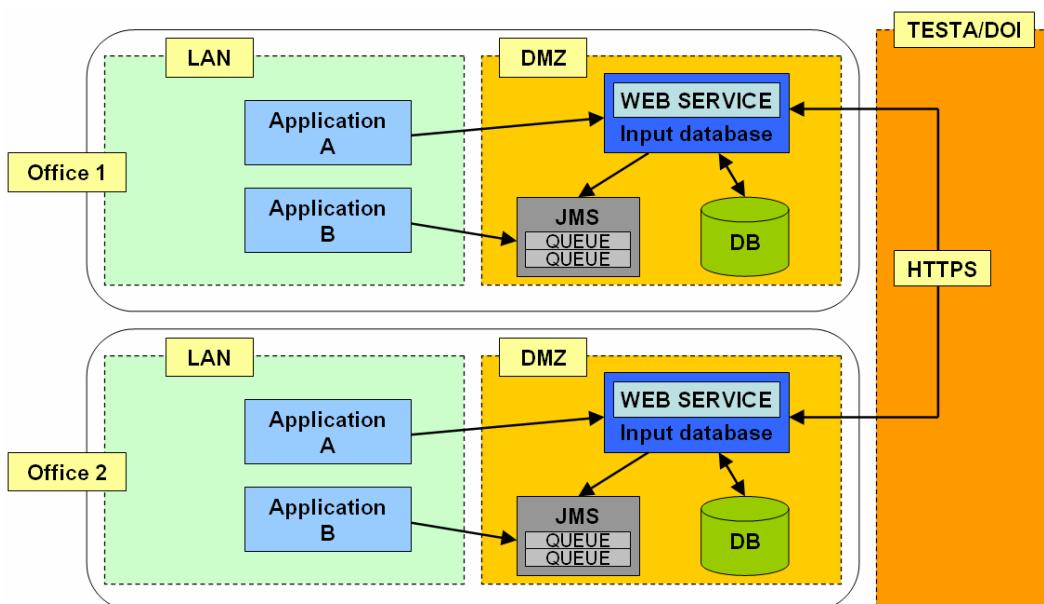
The statistical offices operate a network of Collection Network Database instances that can automatically exchange data with each other. This process is referred to as “routing”. Routing is based on rules of data ownership (collecting office) and the existence of *shared services* for a given statistics. Shared services, such as *Zentrale Produktion und Datenhaltung* (ZPD – Centralized Data Production, Management and Storage) can be associated with a database node in the network, and the Collection Network Database takes care of providing the data they require. Another type of shared service is *Zentrale Erhebung* (Centralized Data Collection).

In the 2011 Census, for example, the reports produced in various document reading centres are automatically routed to the central Collection Network Databases responsible for the relevant surveys.

<sup>1</sup> I.e., data in the DatML/RAW format that have been checked for formal correctness and divided into individual reports in the KonVertCenter and have then been forwarded to the relevant office (see section on “KonVertCenter”)

<sup>2</sup> By means of Land code.

<sup>3</sup> By means of survey identifier.



As an infrastructure application, the Collection Network Database operates in the background and exposes no direct interface to the end user. Instead, users interact with the Collection Network Database through the specialised application only, be it for uploading data, querying the database or retrieving reports for processing.

The Collection Network Database receives data from various sources. Data collected online is automatically uploaded by the KonvertCenter. Other sources are data capture and OCR scans of paper forms.

Via a standardised interface, the Collection Network Database can directly notify input control systems about incoming reports – it is not necessary to make a specific request. When a report has been uploaded to the Collection Network Database, the input control system is notified and can register the receipt of the report. The report can then be imported into the specialised application.

## Data Edit Runtime Environment

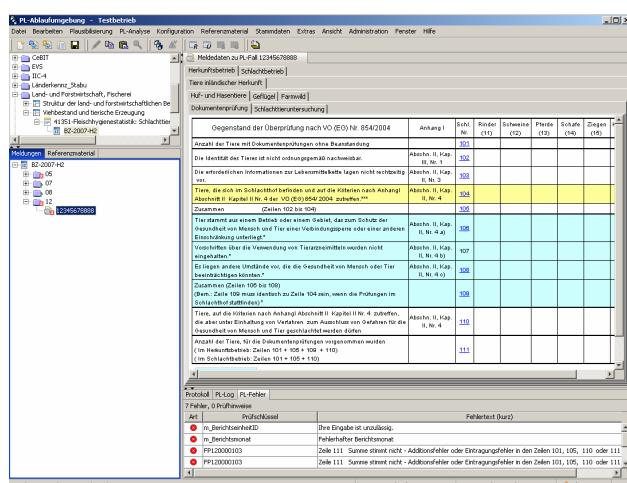
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
---------------	--------	-------	---------	---------	---------	-------------	---------	----------

The **Data Edit Runtime Environment** is a standard application to process reported data. It can be used to manage, check and handle any number of surveys in one application.

Based on the data edit specification<sup>1</sup>, the data edit runtime environment, including the database tables and electronic forms required for data storage and management, is configured for the processing of the relevant survey. Also, the forms can individually be designed by means of the Form Designer. So a specialised application is directly available for the relevant set of statistics. No programming is required.

Already in the setup phase of a set of statistics, the data edit runtime environment plays a major role because it is used as a test environment. The data edit specifications and data edit cycles created in the Data Edit Designer can directly be tested there by the specialised unit at an early stage.

The Data Edit Runtime Environment can read in both reported data in the DatML/Raw format and files in the ASCII CSV format. Users can access the Collection Network Database and import the reported data for processing or view image files of documents that have been read. Manual input of data from a paper questionnaire is also possible.<sup>2</sup>



Modifications applied subsequently to edit checks and data edit cycles in the Data Edit Designer can directly be used by the data edit runtime environment by importing the modified data edit specification.<sup>3</sup>

### Overview of the main features of the data edit runtime environment:

- Any number of surveys can be managed and handled in one application.
- Surveys and reference periods are configured by resources.
- Data edit specification from the Data Edit Designer is used to generate database tables and electronic forms
- Individually designed electronic forms can be integrated.
- Resources in the central survey database can directly be accessed.
- Reports can be imported from the Collection Network Database (access via Webservice).
- Reported data entered manually can be exported into the Collection Network Database.
- Comprehensive range of functions for the data editing.
- After data editing, individual data can be exported.

<sup>1</sup> DatML/EDT file.  
The data edit specification is imported from the survey database or from a local data folder into the Data Edit Runtime Environment.

<sup>2</sup> Data entered manually can subsequently be exported into the Collection Network Database.

<sup>3</sup> Such modifications may be required, for example, if data edit specifications turn out to be unsuitable during processing.

## SAS

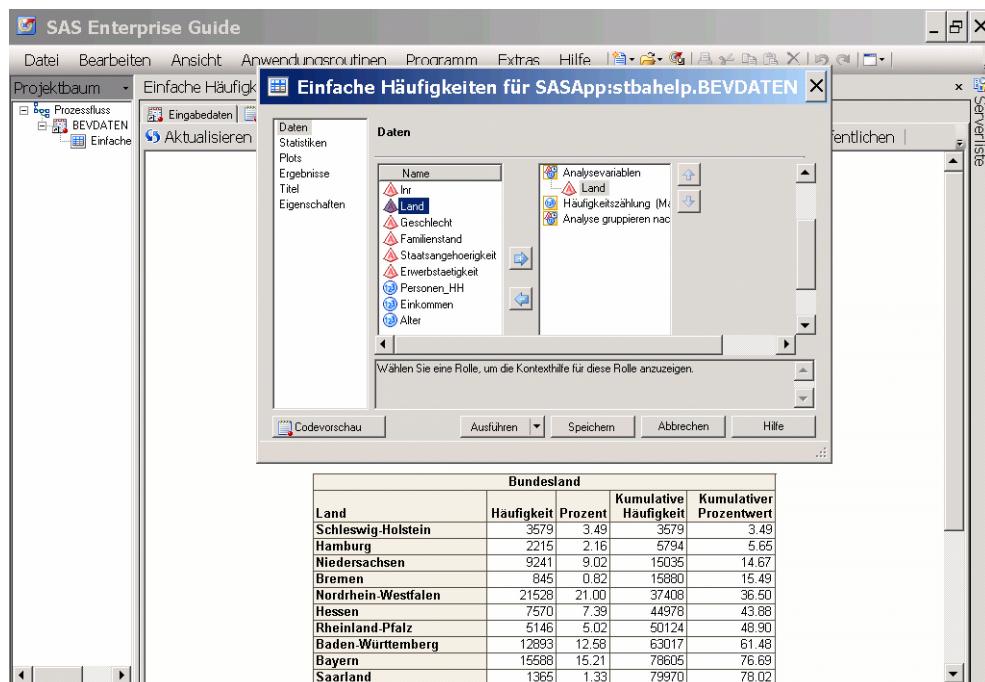
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
---------------	--------	-------	---------	---------	---------	-------------	---------	----------

**SAS** is data evaluation software and includes components for data management, for tabular and graphical evaluation, and for complex data analysis by means of multivariate methods.

Users can apply SAS to calculate descriptive statistics – such as frequencies, totals, mean values, standard deviations, minima and maxima – and produce them as tables in a menu-supported way. It is also possible to do complex calculations, such as index calculation, expansion, error calculation and model calculation. By applying multivariate methods – such as correlation analysis, regression analysis or cluster analysis –, additional information can be obtained from official statistical data. In addition to purely tabular evaluation, it is possible to produce charts for the relevant evaluations, such as histograms, line charts, scatter diagrams and box plots.

Based on the SAS functionalities, data interfaces can be developed for upstream and downstream SteP tools. SAS provides many procedures required in specialised units and – also thanks to its user interface – accelerates the production of specialised applications. The latter can then be used for recurring processes.

“SAS in the network of statistical offices” is a client-server application<sup>1</sup>. It is generally operated through the Enterprise Guide, while the evaluations are done through the server-based SAS software. Many evaluations can be specified through the graphical user interface, which allows flexible ad-hoc evaluation of the data sets. Complex processes can be developed and run in the program development environment of the Enterprise Guide as SAS codes or SAS macros.



Bundesland	Häufigkeit	Prozent	Kumulative Häufigkeit	Kumulativer Prozentwert
Schleswig-Holstein	3579	3.49	3579	3.49
Hamburg	2215	2.16	5794	5.65
Niedersachsen	9241	9.02	15035	14.67
Bremen	845	0.82	15880	15.49
Nordrhein-Westfalen	21528	21.00	37408	36.50
Hessen	7570	7.39	44978	43.88
Rheinland-Pfalz	5146	5.02	50124	48.90
Baden-Württemberg	12893	12.58	63017	61.48
Bayern	15588	15.21	78605	76.69
Saarland	1365	1.33	79970	78.02

Using a server permits efficient processing of large data quantities and integrating them into central production processes.

SAS programs developed individually can easily be exchanged between statistical offices if simple basic requirements are met. Relevant programming guidelines have already been developed, while issues of standardising SAS applications are being settled.

<sup>1</sup> Client-server application:  
A client program (here: Enterprise Guide) communicates with the server software (here: SAS evaluation software). The client program specifies a request, which is carried out by the server program and whose result, once finished, is transmitted from the server to the client. Generally, it is also possible to install both the client and the server software on a PC as a single-user application.

## GENESIS

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Archive	Evaluate
<p><b>GENESIS</b> is the central information system of official statistics. It comprises three major components. First, external users can use GENESIS-Online to access selected current data. Second, internal users can use the GENESIS Client with an extended range of data and functions. With the third component, the GENESIS Webservices, GENESIS can be integrated as a data source into automated processes.</p>								<p><sup>1</sup> For data publication in GENESIS, the cube program is first of all developed in the internal production database.</p> <p><sup>2</sup> SAS, StatSpez, MS Excel or individual programming</p> <p><sup>3</sup> <a href="http://www.destatis.de/genesis">www.destatis.de/genesis</a></p>
<p>For the creation of the GENESIS data cubes<sup>1</sup> and their regular updating, the specialised units have various possibilities<sup>2</sup>.</p>								
<p><b>GENESIS-Online</b></p> <p>For external users, the database available on the internet<sup>3</sup> (GENESIS-Online) provides access to the data provided by the Federal Statistical Office. The data can be retrieved as individually designed tables and the results can be stored in various formats. Suitable tables can be designed graphically as charts or maps. Explanations of the metadata (statistics, variables, values of variables) can be called up through links in the table and ensure comprehensive information of the users.</p> <p>The tables are available free of charge to all users in GENESIS-Online. Additional chargeable functions are also available.</p>								
<p><b>GENESIS Client</b></p> <p>The GENESIS Client is generally available to all staff members of the Federal Statistical Office. With its help, users can very easily create and design tables of their own by just a few mouse clicks. Nearly all tables can be displayed as charts or maps and be transferred into MS Excel. When data sets have been provided in the internal production database, specialist users can do evaluations of their own and allow other specialised units to use the data. For example, the population statistics unit always provides current data in GENESIS, which can be included in a fully automated way in the calculation of other statistics such as "per capita debt". It is thus possible for anyone to compile tables for their own unit and also across several sets of statistics.</p> <p>For all data, GENESIS provides comprehensive information (metadata). It helps users understand the bases of the relevant set of statistics and interpret the figures correctly.</p>								
<p><b>GENESIS-Webservices</b></p> <p>Using the GENESIS-Webservices is a particularly elegant and efficient way of producing tables. For example, in just a few steps, MS Excel can be connected to a table generated in GENESIS. The spreadsheet will then automatically be updated every time the data are revised in GENESIS.</p> <p>Through Webservices, it is also possible to connect databases or portals to GENESIS. For example, an evaluation system of the Federal Ministry of Finance, the geo research at GeoPortal.Bund and some external customers are connected in this way.</p>								
Specialised units	IT organisers	IT developers	External users					