

DEVELOPING AN INFRASTRUCTURE FOR THE EFFICIENT USE OF SAS AT THE FEDERAL STATISTICAL OFFICE GERMANY

Dr. Heinz Stralla
Head of Section in Division
User Service and
Statistical Information Systems

seugi

11 - 13 June
2002

Paris

Agenda

- **Our tasks and the role of SAS at the Federal Statistical Office Germany**
- **SAS Enterprise Guide as the standard user interface**
- **Server configuration and user administration**
- **Tools to support the SAS client/server configuration**
- **Enterprise Application Integration**
 - **Program integration to extend functionality**
 - **Integration of SAS with Genesis, the data warehouse for aggregate data at the Federal Statistical Office Germany**
 - **Interfaces to STATSPEZ, the special software for complex tabulation at the Federal Statistical Office Germany**
- **Realisation of applications with SAS (example: calculation of production indices)**
- **Further application of SAS at the Federal Statistical Office Germany**

Official Statistics

Collection and supply of reliable data following the principles of objectivity, neutrality and scientific independence

Our tasks

- **Methodological and organisational preparation of surveys, coordination of surveys with the Federal Länder**
- **Development of the program of Federal Statistics**
- **Execution of centralized surveys**
- **Aggregation of results of the Federal Länder in decentralized surveys**
- **Data analysis**
- **Publication (Print media, Internet, CD ROM)**

The Federal Statistical Office Germany needs an efficient IT infrastructure to merge and analyze data to create knowledge and publish results.

The role of SAS at the Federal Statistical Office Germany

- Realisation of IT applications for data analysis by subject matter statisticians
 - data management
 - data analysis (e. g. seasonal adjustment)
 - computation of indices, models
 - cross tabulation
 - graphics
- Application of the SAS development environment (AppDev Studio) by the IT division to adjust and optimize the SAS infrastructure at the Federal Statistical Office Germany and to develop individual software for the subject matter divisions.

Reasons for choosing SAS

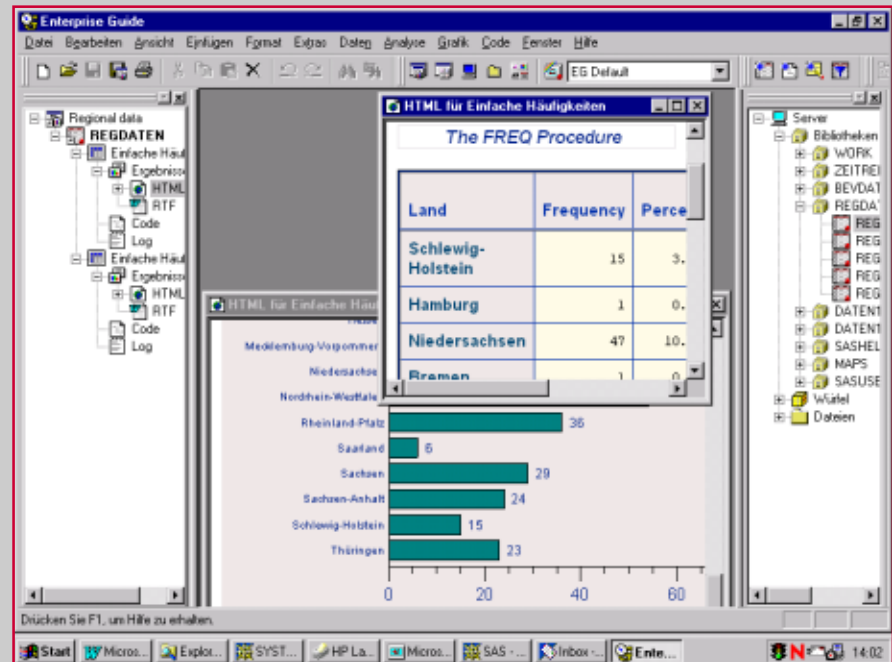
- Efficient building blocks (SAS data step, SAS procedures) to execute data management, statistical data analysis, modelling, tabulation and graphics
- Modern graphical user interface: Enterprise Guide
- Efficient procedure language and macro language to develop complex flows (excellent possibilities for automation)
- Enterprise Application Integration: Possibilities to develop individual tools and applications, individual graphical user interfaces, integration of external programs

**SAS is very well suitable as global analysis tool
for the tasks of the Federal Statistical Office Germany.**

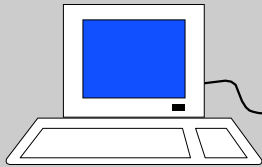
.

SAS Enterprise Guide (EG) as the standard user interface

- Ad hoc data analysis
- Structuring work by use of a EG projekt
- Generating EG process flows for periodically returning tasks
- Generating SAS code by the GUI for modifying it afterwards
- Systematic development of SAS code
- SAS macros and macro variables



SAS configuration at the Federal Statistical Office Germany

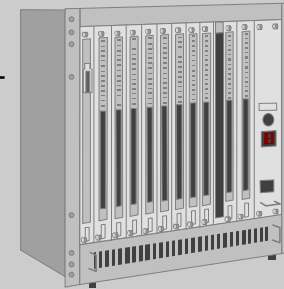


**Client: PC / NT
4.0**

**Enterprise
Guide 1.2**



**File server
(Novell netware)
Application server
Host**



**Server:
SunFire 3800**

**SAS 8.2
(BASE,
STAT, ETS,
IML, GRAPH)**

User administration

user: **UNIX user**

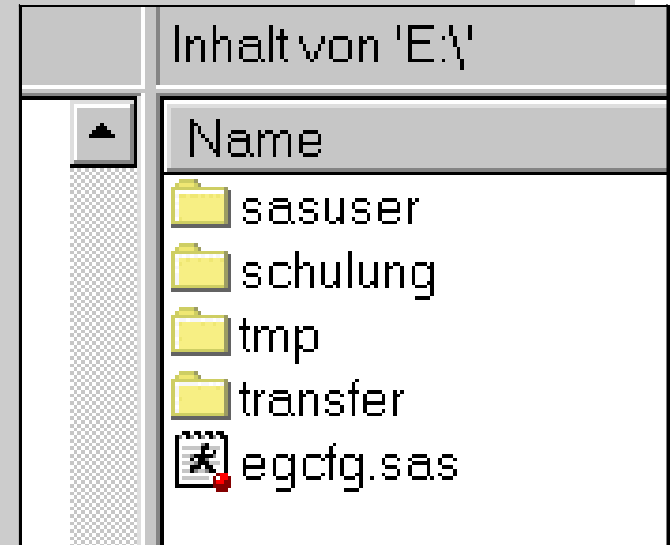
- home folder
- sasuser folder (**SASUSER** library)
- tmp folder (**WORK** library)

group: **UNIX group** (primary and secondary groups)

- **group folder** (read and write access)

for all group members

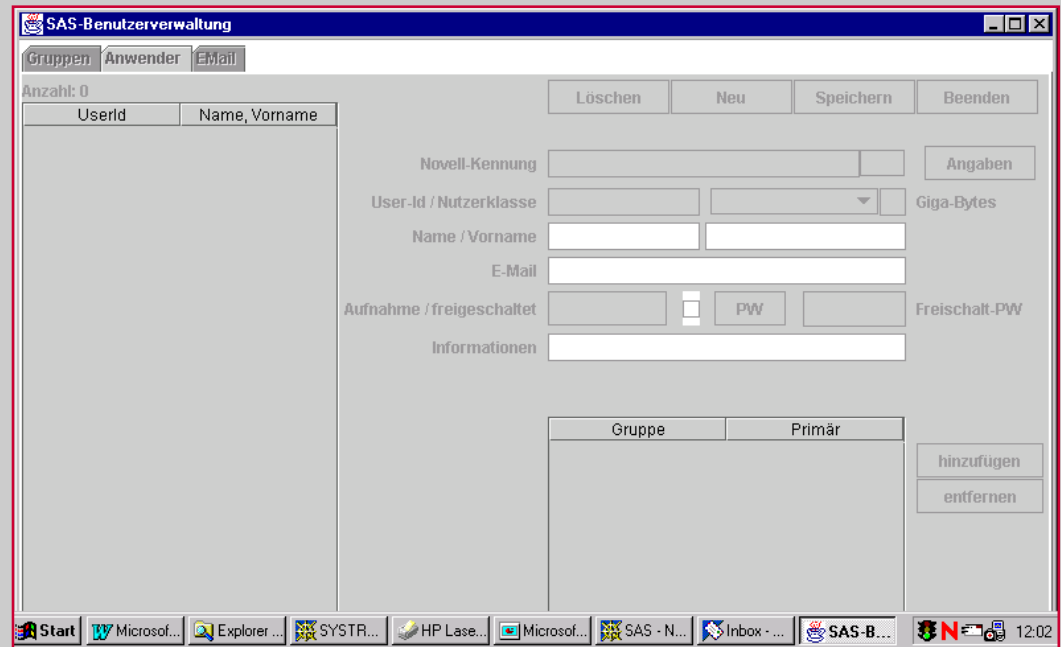
- **transfer folder** (for all users)



Actually we have about 200 users and 90 groups.

Tool for user administration

**Java interface to
the UNIX scripts
we developed
by which the users
on the server are
administrated.**



Tools for the users to support the SAS client / server configuration

General SAS macros

- **UNIX commands, e. g. to assign or delete write protection on data sets, conversion from Windows to UNIX file formats**
- **Information about occupied disk space**
- **Information about server load, possibility to cancel processes on the server started by the Enterprise Guide**
- **Tools to change passwords and adjust them to the Novell password of the user**

Tools for the users to extend functionality

- **Macros to create thematic maps**
- **Macros for the alignment of margins in matrices**
- **Macros to adjust rounding of cells and margins**
- **...**

Enterprise Application Integration

- Program integration to extend functionality

- Integration of original X12-Arima of the US Bureau of the Census for time series adjustment (SAS macro executing the compiled original Windows and UNIX version of X12-Arima)
- Developing a Java interface for X12-Arima including graphics (AppDev Studio)
- Integration of the BV4 method (Berlin method version 4) for time series adjustment (SAS macro executing the compiled Fortran sources available at the Federal Statistical Office Germany, or SAS procedure using SAS/Toolkit)

Enterprise Application Integration

- Integration of SAS and Genesis

Genesis is the developing data warehouse for aggregate data at the Federal Statistical Office Germany

based on a cube data model.

- **Comprehensive metadata for all data objects**
- **Metadata driven access**
- **Retrival system**
- **Tabulation**
- **Export of cubes and tables**

Access to Genesis views by SAS

SAS macro: `%GenesisAbrufTabelle(AUFTRAG,sasuser.auftrag)`



Genesis - Ergebnistabelle Temporär

Abuftabellen | Temporär | Temporär

Bauarten (Indexberechnung)	Deutschland insgesamt									
	Deutschland									
	Monate									
	Januar	Februar	März	April	Mai	Juni	Juli	August	Sept	
	1991									
Hoch- und Tiefbau	82,0	83,2	84,2	85,4	86,6	87,7	88,6	89,4		
Hochbau	78,4	79,4	80,3	81,4	82,4	83,4	84,2	85,0		
Wohnungsbau	61,1	62,1	62,8	63,7	64,5	65,2	65,9	66,7		
Hochbau ohne Wohnungsbau	90,8	91,9	92,9	94,2	95,4	96,6	97,6	98,3		

Table (View) in Genesis

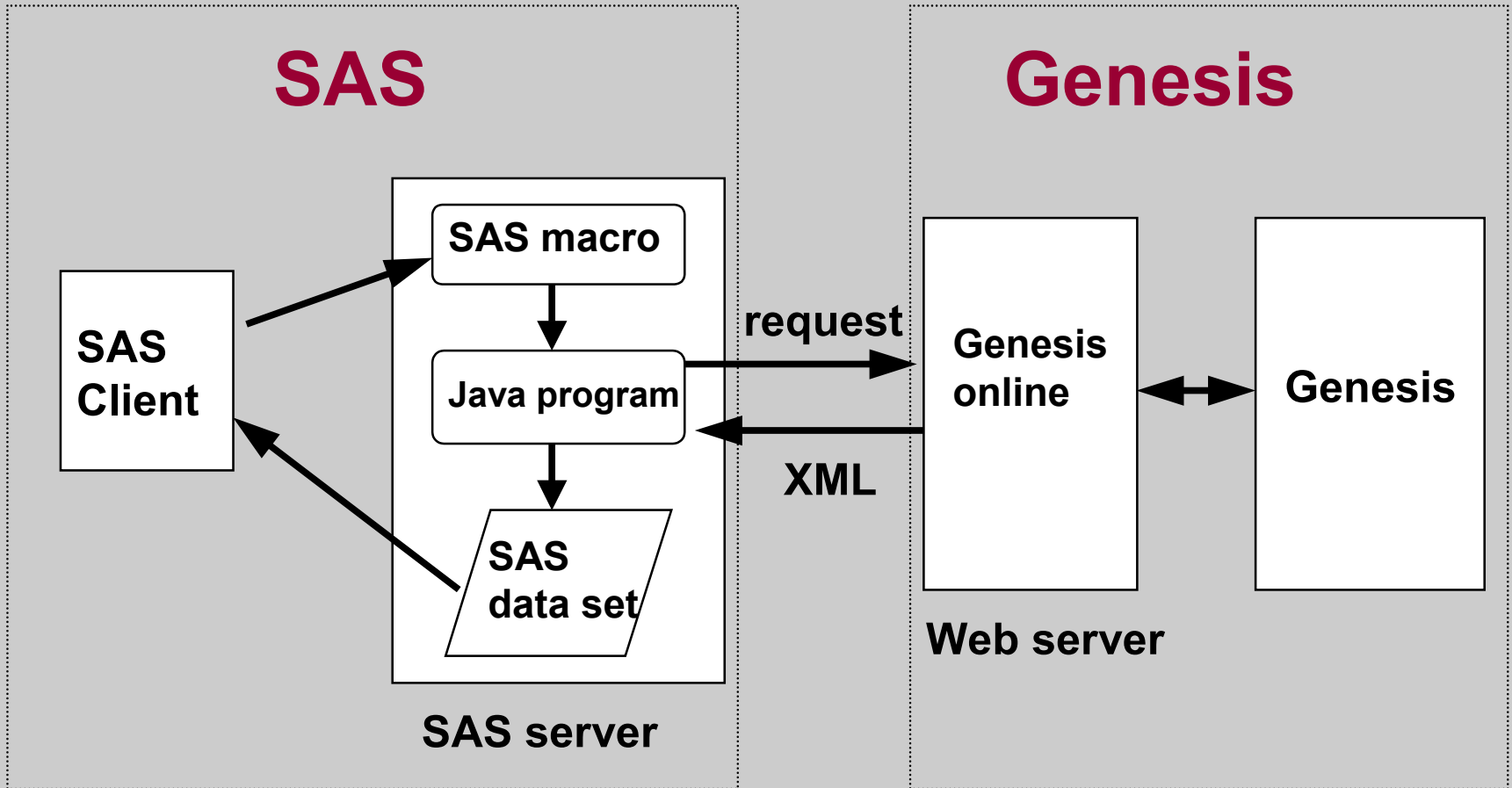
sas\genesis\zeitreihenabruf_seg - [AUFTRAG4 (schreibgeschützt)]

Einfügen | Format | Extras | Daten | Analyse | Grafik | Code | Fenster | Hilfe

	Jahr	Bauarten (Indexberechnung)	Deutschland insgesamt - Monate - Januar	Deutschland insgesamt - Monate - Februar
1	1991	Hoch- und Tiefbau	82	83.2
2	1991	Hochbau	78.4	79.4
3	1991	Wohnungsbau	61.1	62.1
4	1991	Hochbau ohne Woh	90.8	91.9
5	1991	Tiefbau	87.9	89.2
6	1991	Straßenbau	93.4	95.5
7	1991	Tiefbau ohne Stra?	85.5	86.6
8	1992	Hoch- und Tiefbau	91.5	91.7
9	1992	Hochbau	86.9	87.2
10	1992	Wohnungsbau	68.6	69

Corresponding SAS data set

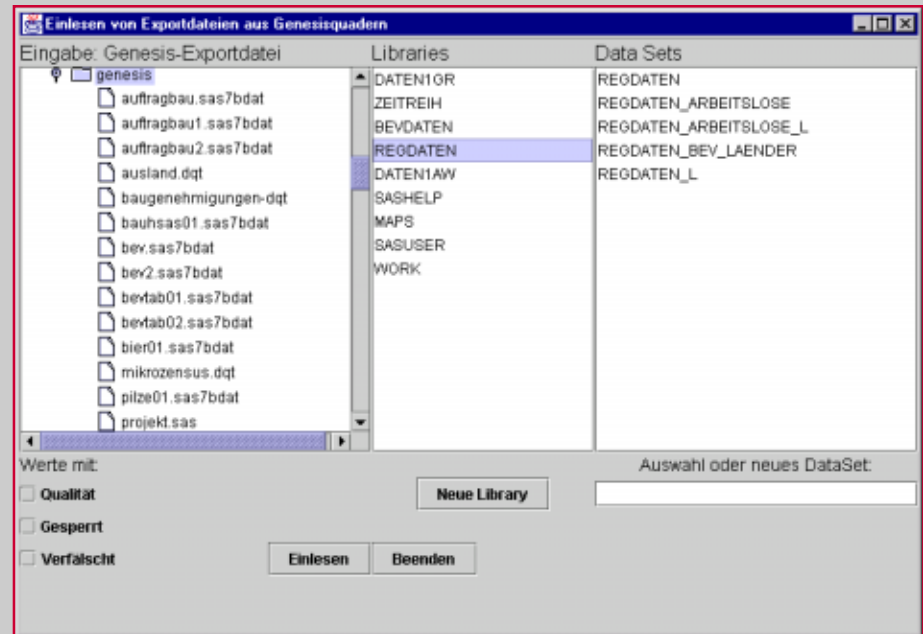
Access to Genesis views



Import of Genesis cubes

A Java program generates a data step to create a SAS data set.

Metadata are used for the labels of the variables.



Integration with STATSPEZ

STATSPEZ is the special software for complex tabulation at the Federal Statistical Office Germany

- **Editor for specifying the structure of data**
- **Editor for specifying thematic objects for tabulation**
- **Tabulation editor**
- **Layout editor**

An XML standard for official statistics is under development:

- **DatML for input data**
- **TabML for tabulated data (aggregates)**

Interfaces to STATSPEZ

Interfaces will be developed

- **to create a data step from DatML and TabML**
- **to produce DatML formats from data sets**
- **to produce TabML from the output of SAS tabulation using ODS XML**

A first prototype to create a data step from TabML was developed. DatML is still in the specification phase.

Implementation of individual applications with SAS

Example:
Computation
of production
indices

System of
SAS macros /
user
interface:
Java client

Monatlicher und vierteljährlicher Produktionsindex

Programm Prozeduren

Eingabe

Periode: Monatlicher Produktionsindex Vierteljährlicher Produktionsindex

Quartal: I. Quartal

Jahr: 1995

Ausgabe (erstellte Datasets)

Auswahl: Eingabe-Daten

Selektieren Feld: Like Wert: Selektieren

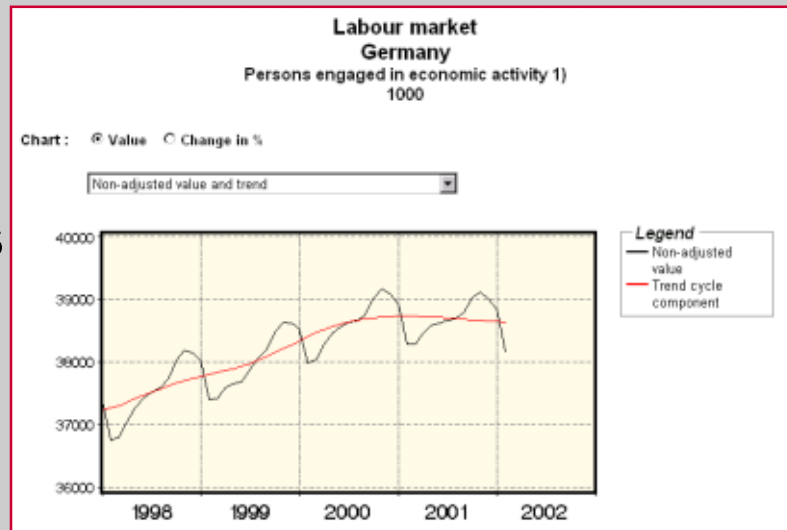
Sortieren Feld: aufsteigend sortieren

Produktionsdaten		Preisindex der Erzeugerpreise (Monat)				Zusatz-Reihen			
EF1	EF2	EF3u1	EF3u2	EF4	EF5	EF5A	EF6	EF9	
E	9	101011300		45	00	1	97	0	
E	9	101012000		45	00	1	97	0	
E	9	102010330		45	00	1	97	0	
E	9	102010350		45	00	1	97	0	
E	9	102010530		45	00	1	97	0	
E	9	103010310		24	00	1	97	5684	
E	9	103010355		24	00	1	97	21306	
E	9	103010357		24	00	1	97	25399	
E	9	103010533		24	00	1	97	2038	
E	9	103010535		24	00	1	97	2839	
E	9	103010553		24	00	1	97	115	
E	9	103010555		24	00	1	97	3721	
E	9	103010557		24	00	1	97	8107	
E	9	103010900		45	00	1	97	3216	
E	9	109900000		98	00	1	97	32648	
E	9	111010000		45	00	1	97	296221	
E	9	111020500		49	00	1	97	956707	
E	9	111020700		49	00	1	97	4237	
E	9	111030000		98	00	1	97	13466	

Bereit

Implementation of individual applications with SAS

Economic indicators on the Internet



activity 1)

Calendar. and seasonally adjusted value (BV4)		Irregular component (BV5)
2000	Change on the previous month in %	1000
2002 Jan	-0,3	-26,0
2001 Dez	-0,3	-12,2
Nov	-0,2	-5,0
Okt	-0,1	2,8
Sep	0,0	18,0
Aug	0,1	20,4
Juli	0,1	-3,2
Jun	0,1	-12,7
Mai	0,2	-11,8
Apr	0,3	-10,5
Mar	0,4	-12,8

Year	Month	Value	Change	Change %	Value	Change	Change %
2002	Jan	38167,8	-0,3	-129,8	38044,9	-0,0	38816,9
2001	Dez	38823,8	-0,3	-98,8	38959,6	-0,0	38838,3
	Nov	38013,8	-0,2	-79,8	38958,5	-0,0	38853,5
	Okt	38120,8	-0,1	-49,8	38999,4	-0,0	38872,2
	Sep	38028,8	0,0	7,8	38990,7	-0,0	38888,7
	Aug	38797,8	0,1	43,8	38992,5	-0,0	38712,9
	Juli	38683,8	0,1	28,8	38702,0	-0,0	38888,8
	Jun	38667,8	0,1	29,8	38708,8	-0,0	38895,9
	Mai	38619,8	0,2	66,8	38717,0	-0,0	38785,4
	Apr	38588,8	0,3	118,8	38724,8	-0,0	38714,3
	Mar	38461,8	0,4	169,8	38730,5	-0,0	38716,7

Example: <http://www.destatis.de/indicators/e/tkarb810.htm>

Further application of SAS at the Federal Statistical Office Germany

- **Use of SAS in many additional subject matter divisions (e.g. implementation of a new statistic on income tax on the basis of SAS), about 100 additional SAS users are expected each year**
- **Support of the cooperation with the scientific community (including work on anonymous remote data access for scientists)**
- **Continuation of the work on supplying tools for the users and perform application integration in order to optimize our SAS infrastructure for knowledge creation.**