The Online-Database histat as an Example for Research-Promoting Infrastructures for Studies in Quantitative Historical Research.

La base de données en ligne 'Histat' comme un exemple d'infrastructure pour encourager la recherche

Gabriele Franzmann[1]

Abstract

Usable statistical data belong to the most important source materials, that especially needed by economic and social historians for their research, but also by social scientists, political scientists, and economists interested in historical issues. Since 2004, the team DHS of GESIS-Data Archive archives mainly studies containing historical long time series on Germany's economic and social history. The present article introduces the research- and downloadsystem histat (historical statistics). Histat was especially designed for time series data. The aim was to offer a comfortable data-access for secondary analysis. The structure of histat, the various research- and download facilities as well as the interfaces of histat to other GESIS service offerings are described.

« Des données statistiques utilisables font partie du matériel de source le plus important qui soit, dont (...) particulièrement les historiens en économie et en sciences sociales, mais également des scientifiques sociaux, des politologues et des économes intéressés aux questions historiques ont besoin pour leur travail ».[2] Depuis 2004, l'équipe DHS de l'archive de données GESIS archive des données axées sur les longues séries chronologiques historiques relative à l'histoire économique et sociale de l'Allemagne. Le présent article présente le système de recherche et de téléchargement histat (statistique historique). Histat a été conçu spécialement pour des études avec des données relatives à de longues séries chronologiques, avec comme objectif d'assurer un accès confortable aux données d'études pour des analyses secondaires. La structure d'histat, les différentes fonctions de recherche et de téléchargement ainsi que les interfaces d'histat et d'autres offres de services de GESIS sont représentées.

Key-Words/Descripteurs

Time Series Data; Historical Statistics; Social and Economic History; Data Service; Data Access; Online-Database; histat

séries temporelles; statistique historique; Histoire sociale et économique; service de données; accès aux données; base de données en ligne; histat

---

[1] Gabriele Frazmann, GESIS-Leibniz Institute for the Social Sciences. Team Dataservice Historical Studies (DHS); Unter Sachsenhausen 6-8. D-50667 Cologne. E-mail: gabriele.franzmann@gesis.org.

[2] Wolfram Fischer/ Andreas Kunz 1991: Grundlagen der Historischen Statistik von Deutschland. Quellen, Methoden, Forschungsziele. Opladen: Westdeutscher Verlag, S. IX.
Voir : Wolfram Fischer/ Andreas Kunz 1991: Bases de la statistique historique d'Allemagne. Sources, méthodes, objectifs de recherche. Opladen: Westdeutscher Verlag, S. IX.

0- Indroduction

The team ‚Data Service Historical Studies' (DHS) is part of the GESIS – Leibniz Institute for the Social Sciences in Cologne, Germany. The team is assigned to the GESIS-department 'Data Archive for Social Sciences' (DAS). The DHS's archived study-stock is characterized by a high heterogeneity of the data, coping with aspects of economic and social history of Germany.

The following contribution introduces to the database system histat, which was developed by GESIS-Leibniz Institute for the Social Sciences in 2003 with regard to a special type of research data, the time-series data. This online - platform has the function to allow access to digitally stored research data, stemming primarily from German historical research. This data access should be made possible via a central server. Thus, the time-series data collected in numerous research projects will be made available for purposes of research and teaching, but also for the interested general public. Thus, the team DHS decided to put a much stronger emphasis on the aspect of offering a comfortable online data service in addition to it's task of archiving study data. The access to the data stock should be as simple and transparent to the users as possible, which could be enabled via a portal with easy search and download functions. The histat database system, which shall be presented here as part of the of the Department 'Data Historical Studies' (DHS) and GESIS data service in general, must be regarded as a contribution to the online data services for empirical historical study-data and can be considered an efficient instrument to meet the research and teaching needs of "productive scientists" looking for an independent, central basis of documentation. In all, the system was designed to facilitate the interactive handling of historical data and can be seen as an efficient interface for data administration.

1. GESIS – Leibniz Institute for Social Sciences: Historical Studies in GESIS-Data Archive

GESIS-Leibniz Institute for Social Sciences is financed by the German federal government and the German states. Dedicated to services and research, GESIS provides an extensive range of research-based infrastructural services on a national and international scale, specializing in empirical social research, archiving, documentation and information retrieval. These tasks are, among other things fulfilled by:

a) continuous, interdisciplinary research and development related to the following areas of responsibility (points b to e):

b) description and explanation of social developments in national, international comparative and historical perspective, including data collection, statistical modeling and long-term monitoring;

c) archiving, documentation and long-term-preservation of social science data, including high-quality processing of especially relevant data for secondary analysis;

d) preparation of literature and research information;

e) creating a user-friendly and high-quality access to all relevant empirical research information and data, including the establishment and maintenance of social science portals and communication

networks, development of effective tools for the research, processing, analysis, backup and archiving of all relevant information.

These tasks are described in the GESIS-Constitution of 19th June of 2010:

> *„a) kontinuierliche, interdisziplinäre Forschung und Entwicklung im Zusammenhang mit den in den … (im folgenden unter b) bis e)…) genannten Aufgabenbereichen,*
>
> *b) Beschreibung und Erklärung gesellschaftlicher Entwicklungen in nationaler, international vergleichender und historischer Perspektive einschließlich der Datenerhebung, statistischen Modellierung und Dauerbeobachtung,*
>
> *c) Archivierung, Dokumentation und Langzeitsicherung sozialwissenschaftlicher Daten, einschließlich ihrer Erschließung sowie qualitativ hochwertigen Aufbereitung besonders relevanter Daten für Sekundäranalysen,*
>
> *d) Aufbereitung von Literatur- und Forschungsinformationen,*
>
> *e) Schaffung eines benutzerfreundlichen und hochqualitativen Zugangs zu allen für die empirische Sozialforschung relevanten Informationen und Daten einschließlich des Aufbaus und der Pflege sozialwissenschaftlicher Portale und Kommunikationsnetzwerke, Entwicklung effektiver Instrumente für die Recherche, Aufbereitung, Auswertung, Sicherung und Archivierung der relevanten Informationen, …"*

(GESIS-Constitution of 19.06.2010, http://www.gesis.org/das-institut/der-verein/satzung/)

Thus, among other objectives, GESIS's activities focus on monitoring social developments in historical perspective as well. This concern is met by GESIS on the on hand within the scope of the social indicators research and the resulting online research and download system SIMon (Social Indicators Monitor[3]). With regard to the GESIS-Data Archive, the aim accentuated by GESIS (i.e. monitoring social development processes also from a historical perspective) involves the notion, that in addition to survey data, the acquisition and preservation of data form historical studies, which were collected and processed for clarification of historical research questions, are of importance. Therefore, the archive priorities of the GESIS-Team 'Data Historical Studies' (DHS) are the archiving of collective biographic-data, process produced data, historical political election data as well as cross sectional data of quantitative historical research.

Another important study segment are quantitative studies of the social and economic history as well as research findings on 'Historical Statistics of Germany', reflecting the historical developments in the form of long time series.

The GESIS work area 'Data Historical Studies' (DHS) of the GESIS-Data Archive acquires, archives and documents these historical studies. These studies are – together with the survey studies – recorded with a detailed study-description in the GESIS data catalogue (DBK), which is a central research facility, to find studies and data on the basis of the study descriptions.

---

[3] See WEB-page of GESIS, Social Indicators Research Centre (ZSi): http://www.gesis.org/en/social-indicators/products-of-the-zsi/

2. Collection of Historical Studies in the GESIS - Data Archive for the Social Sciences

The archived quantitative studies on historical and social research are characterized by a high heterogeneity of the collected data. These studies can be grouped into two different dimensions:

1. The first Dimension is the temporal reference of the study.

Historical studies can analyze a question at a particular time in history. In this case, these studies collect cross-sectional data. There are also studies with a design, compiling data on multiple points in time with the aim to compare facts and circumstances on specific time points. Furthermore, there are studies comprising process data or continuous time series, which may serve to describe and analyse individual developments or changes of aggregates.

2. The second dimension is equivalent to the aggregation level of a study.

On the one hand, the archive contains studies which in their turn incorporate various microdata (e.g. personal characteristics collected from different archives such as dates of birth and death gathered from church registers). Other studies stored in the database archive bear macrodata (i.e. aggregate data such as election results or socio-economic and political characteristics referring to territorial units, e.g. constituencies).

General Classification of Data Types in the Historical and Quantitative Studies in GESIS Data Archive (examples are described in separate cells, see below)

		time relation of the studies		
		cross-section data	several points in time	process data / time series
level of aggrega-tion	micro data information on individuals, single companies / enterprises etc.	surveys on key dates, "historical snapshots"	surveys repeated at two or more points in time.	surveys carried out continually, i.e. at equal intervals and over a long period of time examples of process data: collective biographies, e.g. biographies of members of parliament
	macro data aggregation of individual data on a higher level, e.g. communities or the German states (combined in major groups) etc.	survey at a pre-defined point in time, "historical snapshots"	surveys repeated at two or more points in time. examples: data from population and occupational censuses; historical data from electoral research	Surveys performed on a continuous basis at equal intervals / over a long period of time Examples of studies with time series data: *Datenhandbücher zur 'Deutschen Bildungsgeschichte'* (Data Handbooks on the German Educational History); *Datenhandbücher zur 'Historischen Statistik Deutschlands'* (Data Handbooks on German Historical Statistics).

Studies containing time series data are a subset of historical studies of macro data. They account for an essential part (80%) of the quantitative historical studies within the GESIS Archive. Studies with time-series date can be divided into two parts: On the one hand there are studies and data of individual scientific studies, which are thematically highly specialized. These studies and their data cover various topics such as the economic development, historical demography or educational history etc. Furthermore, because of the different research questions these collected data are with significant differences regarding to territory and time period they refer to. Principal investigators, who want to answer their research questions, refer to quantifiable indicators and rely on different statistical, official and non-official sources. In these studies issues from the former German history as well as questions on Germany's recent history are be treated. [4] On the other hand there are data collected and processed in the framework of priority projects funded by the German Research Foundation (Deutsche Forschungsgemeinschaft/ DFG). The collected data are long time series, drawn from official and non-official sources, and they are processed in a source critical way. The aim was, to obtain data collections, which are broad in terms of space and time and detailed in terms of covered subjects. A historical data basis for further historical research should be created and provided. In this context the research priority program "Quellen und Forschungen zur Historischen Statistik von Deutschland" (= Sources and Researches on Historical Statistics of Germany) have to be mentioned. On the basis of this priority program 26 detailed data manuals have been published with critical comments, reaching far back historically and covering a broad thematic range. The project group "Deutsche Bildungsstatistik" (German Educational Statistics) was supported by the DFG, too. Its projects were targeted at describing and analyzing the long-term structural changes of the German educational system on a broad empirical and statistical basis. 10 detailed comprehensive data manuals with long time series are published on the basis of this project group. [5]

As "Added Value" of the data service the team 'Data Historical Studies'(DHS) offers data compilation on selected subjects, compiled, processed and described according to the model of the data manuals. The aim of data compilations is to fill data gaps as well as to provide critical and detailed annotated data material for research. These GESIS-Data compilations focuses on the period of the Federal Republic of Germany, the sources are mainly publications of the official statistics.

Since 2003 the team 'Data Historical Studies' (DHS) in GESIS-Data Archive has addressed himself to the task of acquiring the thematically widely dispersed data from the historical research on

---

[4] Examples are:
- Spoerer, M., 1996: Von Scheingewinnen zum Rüstungsboom. Die Eigenkapitalrendite der deutschen Industrieaktiengesellschaften 1925 –1941. (From Sham Gains to the Armament Boom. The Equity Return of German Stock Corporations from 1925-1941). Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte, Beihefte (123). Stuttgart: Franz Steiner Verlag.
- Thome, Helmut/Birkel, Christoph (2007). Sozialer Wandel und Gewaltkriminalität. Deutschland, England und Schweden im Vergleich, 1950-2000. (Social Change and Violent Crime. A Comparison of Germany, England and Sweden from 1950-2000) Wiesbaden: VS Verlag für Sozialwissenschaften.
– Graf Finck von Finckenstein, Hans Wolfram (1960): Die Entwicklung der Landwirtschaft in Preußen und Deutschland, 1800-1930 (The Development of Agriculture in Prussia and Germany, 1800 to 1930). Würzburg, Holzner.

[5] ‚Historical Statistic of Germany' see:
- Kunz, Andreas (1979): Historische Statistik von Deutschland. Ein Forschungsschwerpunkt der Deutschen Forschungsgemeinschaft 1981 – 1991. In: Historical Social Research 22/2, 1997, S. 236-249.
  Selected publications from the DFG-priority group on the German Educational Statistics (Deutsche Bildungsgeschichte):
- Lundgreen, Peter (unter Mitarbeit von Jana Scheunemann und Gudrun Schwibbe) (2008): Berufliche Schulen und Hochschulen in der Bundesrepublik Deutschland 1949–2001. Datenhandbuch zur deutschen Bildungsgeschichte, Band VIII. Göttingen: Vandenhoeck & Ruprecht.
- Titze, H. unter Mitarbeit von Herrlitz, H.-G., Müller-Benedict, V. und Nath, A., (1987): Das Hochschulstudium in Preußen und Deutschland 1820 - 1944. Datenhandbuch zur deutschen Bildungsgeschichte, Band I: Hochschulen, 1. U. 2. Teil. Göttingen: Vandenhoeck & Ruprecht.

Germany's history. Additionally, results of prior research projects, which were documented in form of printed data tables, should be digitalized. Both, the studies of current historical research and of prior research projects, are documented and processed according to the standards of the GESIS-Data Archive and made available to interested users.

3. Data Processing and Documentation Standards for Historical Studies with Time Series Data

The development of standards serves the purpose to describe and process study data on the basis of consistent and comparable criteria and to ensure a controllable and the highest possible quality. The design of the description should meet the special format of the studies – in our case historical studies – and reflect significant facts, which allow the correct understanding of the study and in particular of the data. The following general objectives are pursued initially with each study description:[6]

- All information about the study and the data have been reported by the primary researcher in his publication has to be reproduced, so that the interested scientific community is able to evaluate the study and it's data properly and to classify it in the scientific context.
- The information to the data that are needed for a secondary analysis must be provided completely.
- The information must be offered in a form that is suitable for standard research tools.

From these general objectives results the following description scheme which is used for survey studies:[7]

- **Title:** Study Title (**Further Titles:** Additional titles, for example original language title)
- **Study Number**
- **Survey Year:** Year of data collection
- **Primary Researcher/Authoring Entity:** Name of Institution or Researcher who carried out the studie
- **Institution:** Institution by which the data were collected, data collector
- **Description, Content:** Study Description
- **Geographic Coverage** Territory of investigation, listed using ISO3166-1/-2 Codes or using free text.
- **Universe and Selection Method:** For example part of population to which a selection method is applied.
- **Mode of Data Collection**
- **Analysis-System**: (data set) Number of Units; Number of Variables / Data-Type / Analysis-System(s)
- **Publications:** Publications on the basis of the data set
- **Further Information:** e.g. further studies

Historical studies differ in the following aspects from survey studies:[8]

- Survey- or investigation period:
In survey research the investigation period is the time period of carrying out the survey and

---

[6] See: Reinke, Herbert (1981): Towards Standards for the Description of Machine-Readable Historical Data. In: HSR, Nr. 18, S. 3.10.

[7] See datacatalogue of GESIS-Data-Archive (DAS), WEB: https://dbk.gesis.org/dbksearch/index.asp?db=e

[8] Studies of 'Oral History' are excluded from this presentation.

therefore it is the time period, in which the data were collected. In the case of historical studies the period of investigation refers to a time period lying in the past, for which retrospectively data are collected from official and / or non-official sources. The investigation period is already part of the research question (for example: Unemployment in Germany during the Weimar Republic from 1918 to 1933. Investigation Period is in this case the period between 1918 and 1933).

- Geographic Coverage:

Similar to survey research, in historical studies the geographic coverage refers to a geographic unit which is the basis of the data selection. It is essential, however, that especially in the case of Germany and the German states borders has often changed. The changing borders and areas over a period of time may affect the values of collected time series (for example the population of the German Empire including or excluding Alsace-Lorraine). Therefore altered boundaries require careful documentation. Attempts to standardize the German historical states and territories with their changing borders have failed so far. Using the ISO-Code, only a very rough classification is possible, which does not meet the requirements for an adequate understanding of the historical situation and therefore for the historical data. In the case of historical studies a detailed description of the territory with its changing borders is necessary, which is done via free-text.

- Universe and Selection Method:

The selection method describes the process of data collection in the survey research. Universe, survey design and used technique of sampling are outlined. For studies of the social and economic history the universe and selection process are defined and explained as well, so that the significance of the data and results can be assessed. Data of historical studies are not collected by interviewing persons - with the exception of oral history -, they were collected from archives, publications of the official statistics and/or other statistical materials. Therefore the sources must be identified, from which the information and data were collected. Source types and individual sources must be listed. Source types are general terms of sources (for example: archival documents, record collections, church registers, official statistics, or scientific publications). Among sources finally the individual sources are accurately named, from which data were collected.

- Critical Comments on the Used Sources:

For Historical Studies a careful critique of sources is very important for assessing the significance of the data. So it may be that information on a population group can only be obtained with restrictions on the basis of the available sources.[9] The source material may be incomplete caused by wars and other influences. Historical sources can differ widely in accuracy and depth, because in time periods before 1871 there existed no uniform guidelines for data collecting processes and additionally these sources and documents were not made for statistical purposes, but for the needs of the contemporary administration of a region. Background information on the sources (what should be originally collected and how was it be done?), the condition of a source (completeness, parts illegible, etc.) and limited access to resources can give indications of possible distortions of the information content of a data set.

---

[9] H. Reinke performs here as an example the entire population of a place as an object of investigation and the control register as source available. In this case, all those who paid no taxes are excluded. (Herbert Reinke (1981): Towards Standards for the Description of Machine-Readable Historical Data. In: Historische Sozialforschung (HSR), Nr. 18, S. 3-10.)

- Notes on Methodology:
In particular for quantitative statistical information there are possibilities to estimate data, or on the basis of new research results to recalculate data or to weight data. These methodological approaches must be documented and discussed or explained on the level of the entire study and additional very closely to the data on time series level and on the level of the single value.

- Analysis-System, data set:
In the case of survey data the number of interviewed persons and the number of variables are specified. Historical studies' data sets are of a higher heterogeneity. In addition to historical data sets, dealing with individual persons, there are data sets dealing with other investigation units (for example grave inscriptions, newspaper articles, etc.). A major part of data, which are of interest for Social- and Economic historical research, are aggregate data in the form of time series. The key information of these data sets are: the number of collected dates (mainly the collected years, but also quarterly time points, monthly dates, term values, or sometimes daily values - for example, the rows of the stock market) and the number of time series (time series are comparable to the variables in survey data).

In summary it can be concluded, that for a correct understanding of quantitative historical data the following additional information are necessary:

• Specification of the historical time reference of the study (= research period).

• Detailed information on borders and changes of boundaries of historical territories (for example: Prussia in borders of 1866; German Empire of 1871 with or without Alsace-Lorraine; German Empire 1933 incl. or excl. Saarland; Germany in borders of 3rd October 1990).

• A detailed list of source types and sources.

• Information on the information value of used sources relating to the research unit.

• Details on the quality of sources.

• Details on estimates and calculations.

• References to used survey systematics of historical aggregate data, especially of historical statistics.

• In order to describe the data file's size precise information on collected units and variables are needed, because in case of historical studies collected units and variables can be very different.

The consequences deriving from the standards for a study description of historical studies described above and from the need for clear assignability of study descriptions and study data to the primary researchers are the following conditions which must be met by a search and download system for time series data:

• The study description must contain the following: a short description of the issues the study deals with; precise and detailed information on the research period, on the territory under investigation, on the used methodology of data collection and if necessary on estimations, detailed information on the used sources and their quality.

• In order to offer a comprehensive information to the user indications on the dataset's size and on the thematical subdivision of the data must to be given.

- The indication of sources and annotations must be possible on the level of the entire study in the study description as well as close to the data for each individual time series and, if necessary, for each single value.

- The list of notes on both levels, in the study description as well as on the data level, must contain detailed information on estimations, recalculations, and breaks in time series. Breaks in time series can be caused by changes in territorial borders, by changes in a measurement unit (for example currency changes), or by changes in data collection systematics.

- The name of the primary researcher or of the data donor, study-title and study description, the publication of the primary researcher which must be quoted as well as the study data must always appear as a cohesive unit.

- There must be given a clear hint on the researcher's publication, which must be quoted, if the data are used for a secondary analysis.

- For the data archive user study descriptions and study data must be accessible in several ways and in a simple and intuitive way. This means that on given thematic blocks, on the names of primary researchers, and by using a key word search studies and data must be searchable.

- There must be a facility to compile time series data on the basis of the complete study data set according to the user's individual defined criteria for secondary analysis. However, the relevant context of principal investigator's name, the researcher's publication and the associated data must always be maintained.

4. The Data Service Project histat

The name of the search and download system histat (Historical Statistics) stems from the priority program "Historical Statistics", which was funded by the DFG from 1981-1991. Ulrike Albrecht and Andreas Kunz created a concept with the name HISTAT with the aim to integrate the collected time series data of the priority program's research projects into a data base.[10] On this concept of Albrecht and Kunz is referred to, but with a clearly different purpose, because the studies which should be integrated into the database do not come from the projects of an individual DFG priority program. The concept of Albrecht and Kunz has to be modified and expanded.

Under the leadership of Professor Dr. Rainer Metz and with significant participation of Dipl.-Soz. Jürgen Sensch the Team 'Data Historical Studies' (DHS) has set itself the target in 2003 to integrate numerous historical studies with time series data into a database and to offer these studies under a uniform user surface. It should be ensured that under histat individual studies on specific topics and issues will be offered. This study compilation in histat is supplemented by studies from the DFG priority programs "Historical Statistics of Germany" and "Data Manuals for Historical Research in Education" as well as by selected data from social and economic historical data manuals.

---

[10] – Albrecht, Ulrike und Kunz, Andreas, 1990: Building a Databank on German Historical Statistics.
In: Metz, Rainer / Van Cauwenberghe, Eddy / van den Voort, Roel (eds): Historical Information Systems. Leuven University Press, 77-86.
- Kunz, Andreas, 1990: Eine Datenbank zur Historischen Statistik von Deutschland.
In: Diederich, N. / Hölder, E. / Kunz, A. u.a., 1990: Historische Statistik in der Bundesrepublik Deutschland.
Band 15 der Schriftenreihe Forum der Bundesstatistik, hrsg. vom Statistischen Bundesamt, Wiesbaden.
Stuttgart: Metzler-Poeschel, S. 160.

As a high-quality data service (added value) the employees of the DHS compile on selected subjects data compilations from official statistics and offer them via the database histat to provide additional long time series.

The studies must be represented by a consistent study description scheme and with a uniform structure of the data tables. Additionally source critical comments on time series and on value level must be available for the user to enable an accurate understanding of the data.
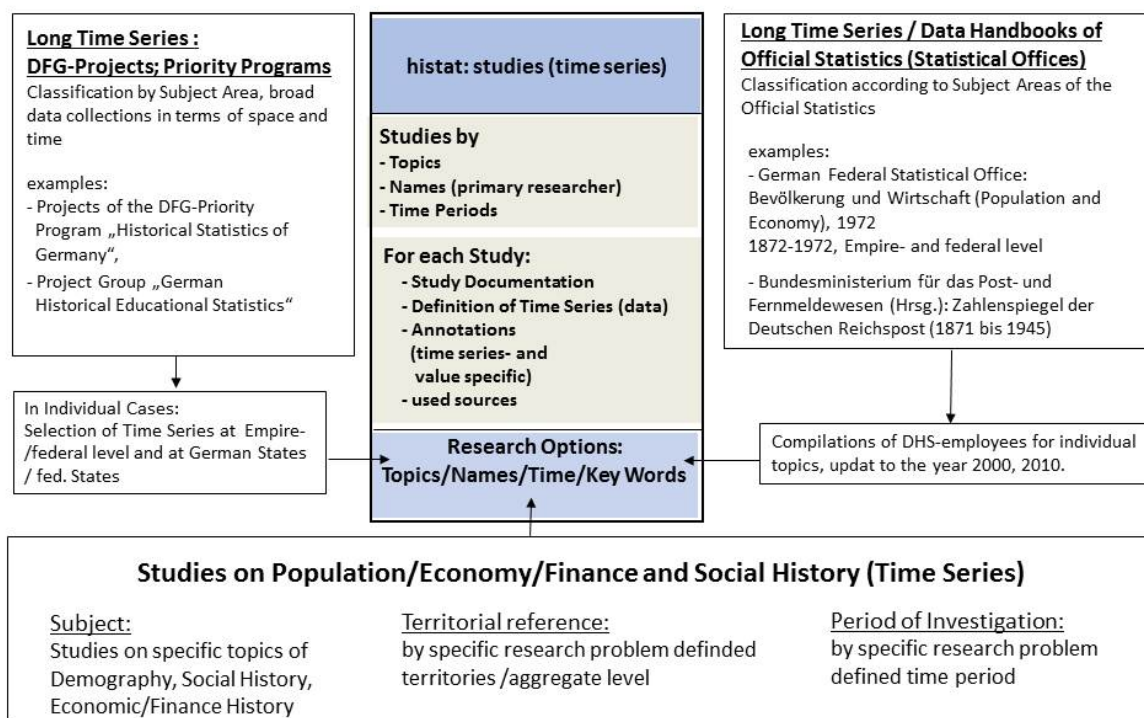

Figure 1: Integration of historical studies with time series into histat

The declared aim is to provide a central and comfortable access to time series data and to enable the user to compile his/her own data set(s) for the purpose of secondary analysis, whereas – this is important to note – the assignment of the respective selected time series data to the corresponding studies and primary researchers must always be preserved.

4.1 Structural Composition of histat

The result of the requirements on a research- and download system for historical studies, which are mentioned above, is the development of a hierarchical structure for histat. According to this concept, the database is subject- and study orientated, i.e. it contains a wide selection of topic titles, e.g. education, demography, etc. The individual studies are allocated to these topics according to their thematical subject. For instance, the data contained in the study *„Titze, H. et. al: Das Hochschulstudium in Preußen und Deutschland, 1820-1944" (Higher Education in Prussia and Germany)* are be found under the data-base topic "Education". The study-data of this study are sub-divided according to their subjects (for example in the case of the mentioned study of Titze there is a data sub-division into subjects and faculties). On the lowest hierarchy level of the system, the user

finds information on individual time series (e.g. the overall number of business students registered at the University of Cologne). In addition to these time series the database comprises a detailed study description, specifications concerning the sources used, comments on individual time series, and, if required, on single values within these series. The following figure 2 illustrates the hierarchical structure and the integration of the studies into the overall structure of histat.
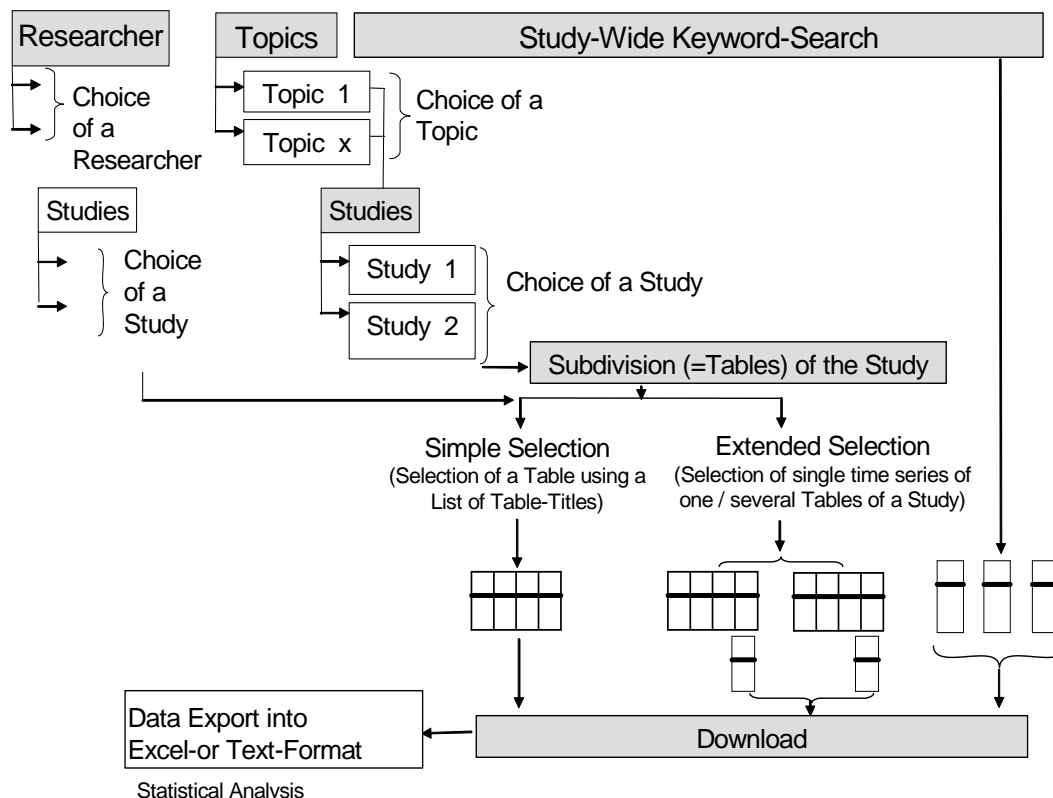

Figure 2: Graphical representation of the hierarchical database structure of histat

The context information (i.e. meta data) are shown at every stage of data-research in histat. Additionally, a PDF-Document contains further details on Data, a methods report, and a detailed list of sources. The data-export contains the complete meta-data to the time series as well as the notion of the publication, which must be quoted, if the downloaded data were used for an own publication. A key-word search offers the user a targeted access to time series of the complete data collection. An administration module offers GESIS-DHS[11], the provider of histat, the possibility to track the study demand and to set up a user-statistics. Figure 3 summarizes the characteristics of the online-database histat.

---

[11] GESIS – DHS = GESIS – Dataservice Historical Studies

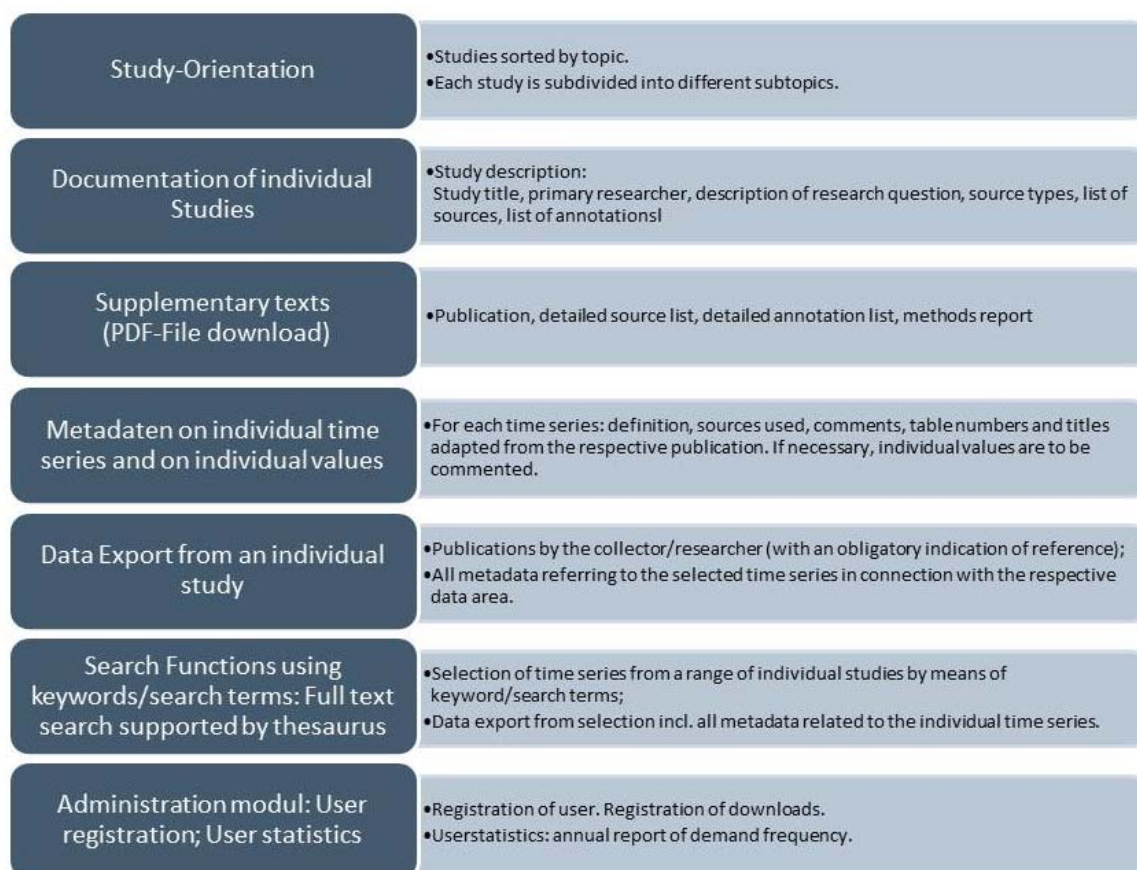|	
Study-Orientation	• Studies sorted by topic. • Each study is subdivided into different subtopics.
Documentation of individual Studies	• Study description: Study title, primary researcher, description of research question, source types, list of sources, list of annotationsl
Supplementary texts (PDF-File download)	• Publication, detailed source list, detailed annotation list, methods report
Metadaten on individual time series and on individual values	• For each time series: definition, sources used, comments, table numbers and titles adapted from the respective publication. If necessary, individual values are to be commented.
Data Export from an individual study	• Publications by the collector/researcher (with an obligatory indication of reference); • All metadata referring to the selected time series in connection with the respective data area.
Search Functions using keywords/search terms: Full text search supported by thesaurus	• Selection of time series from a range of individual studies by means of keyword/search terms; • Data export from selection incl. all metadata related to the individual time series.
Administration modul: User registration; User statistics	• Registration of user. Registration of downloads. • Userstatistics: annual report of demand frequency.

Figure 3: Characteristics of the data base histat

4.2 Documentation of the studies

The documentation of a study is in two parts:

On the one hand the research question of the study is described in the study description. The subdivision of the study data and the number of data tables are shown in the field "content-related subdivision of the data tables". The number of data tables and time series is listed at the download button for the data download. A detailed annotation list is added to the study description with information on the used methodology of data collection or on estimations and recalculations. In addition a detailed list of sources used by the primary researcher and in some cases a methodological report is enclosed. Bibliography, detailed notes and other methodological notes are also offered as a downloadable PDF document for each study (see figure 4, left side „study details of the entire study").

The data part of the studies contains short data descriptions (see figure 4, right part „data part (tables) of the study"). In addition to the description of the individual time series variables the sources used by the author and further remarks on the data of the time series are shown. Since 2007 notes for single values are integrated into histat.

4.3 Data Processing and Data Documentation for histat Studies: Application of DIN-Norm for histat-Data Tables

The study description is – as already mentioned above – into two information parts: the study description contains information on the entire study, the data tables contains the information on the single time series and therefore the data description. In the following the structure of histat will be closer examined.
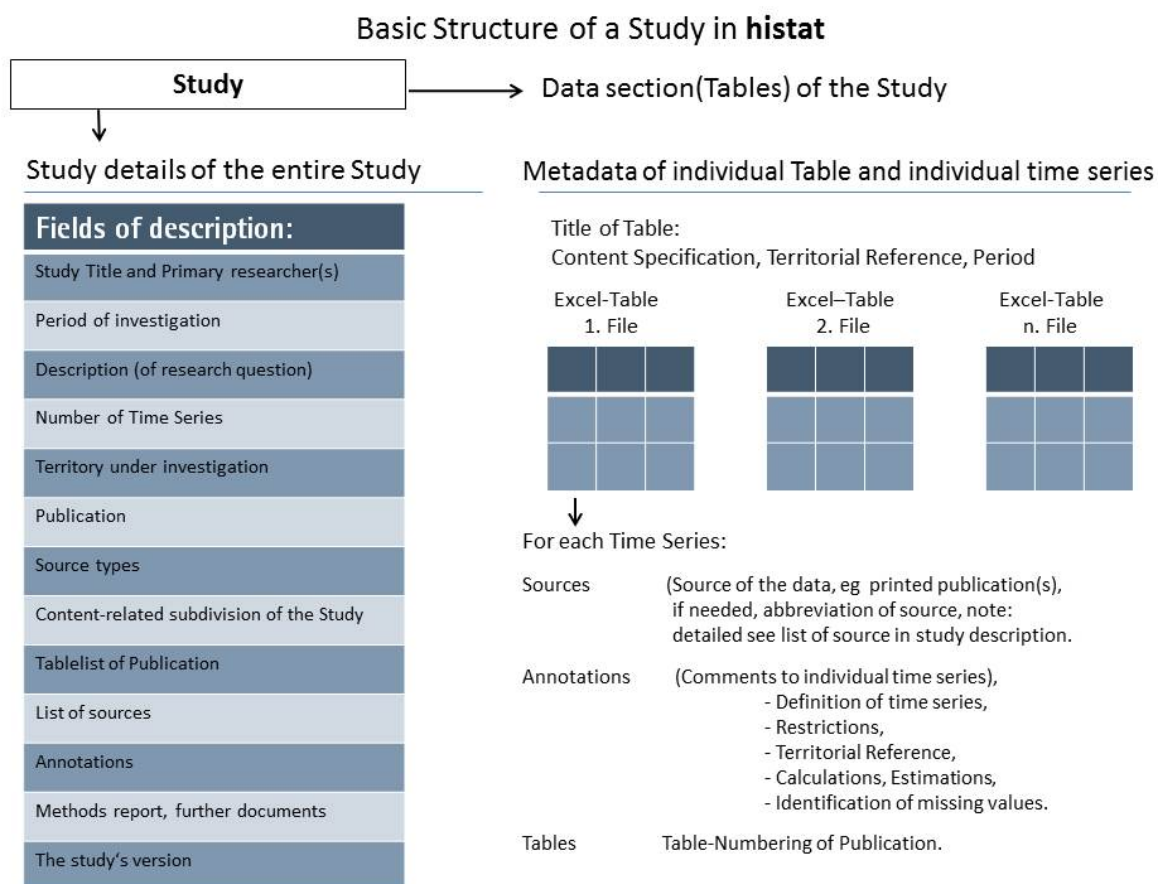
Figure 4: Structur of Studies in histat

Historical studies consist of a set of Data tables, differing in their content structure or rather in their levels of detail. Therefore, on the basis of the DIN-Norm 55 301 a consistently table schema was developed to enable the data import into histat. Tables with different levels of detail and therefore with different content structures, can be processed using the same criteria. Furthermore, important meta information can be added to the data tables.

A histat data table consists of two parts: first there is the data area (number field or table field) in which the time series values are entered. Second there is the identification area, which consists of the table head and the precolumn. The table head is used to describe the entered data according to a clearly defined schema. In figure 5 the identification area is colored in dark blue. The table-title with information on the issue and on time-space-relation, the table-head and the precolumn belonging to the tables' identification area. The specification of the characteristics, according to which an issue, a variable or carriers of characteristics can be described, are entered in the precolumn (for example: federal country / gender). Issues, variables or carriers of characteristics are sorted by these

13

characteristics. Furthermore, the table head contains the different expressions of characteristics (for example: the characteristic ‚federal country' has the expression ‚Bawaria', ‚North-Rhine-Westfalia', etc./ gender= male, female). Finally, information on sources and annotations as well as the table number is listed for each time series in the area of the table head.

Pre-Column ↵ (Definition of characteristics)¶			
Table Title	**Table-Nr. and Table-Title** (eg: A.03 Population in Germany by federal states and sex (1950-1980)		Table-Title¶
Specification of characteristric (eg: federal state)	Expression of characteristic (eg: Bavaria)	Expression of characteristic (eg: Bavaria)	
Specification of characteristic (eg: gender)	Expression of characteristic (eg: male)	Expression of characteristic (eg: female)	
Measurement unit	(eg: in 1000)	(eg: in 1000)	Table-Head↵ (Definition of individual time series)¶
Sources	(eg: Federal Statistical Office (ed.), Title of Publication, Publication details, Page)	(eg: Federal Statistical Office (ed.), Title of Publication, Publication details, Page)	
Annotations	(eg: as at 31.Dezember of the year)	(eg: as at 31.Dezember of the year)	
Table	(eg. Tab. 03)	(eg. Tab. 03)	
Time (eg: Year; Interval)	Value	Value	
Time (eg: Year; Interval)	Value	Value	

Figure 5: Marking Area of a Data table in histat

In summary, the precolumn (first column on the left-hand side of the table) contains the characterstics, according to which issues are subdivided.

The following columns contain the detailed description of the column-contents. First in these columns the expressions or values of the characteristics are entered (for example for the variable Gender the values: male, female). Then the measurement units of the time series data are listed (in 1000 of population; in %; etc.). Finally, primary sources used by the researcher to collect the data of the time series, as well as annotations are named on time series level.

Annotations on time series or single values of a time series could apply to the following circumstances:
(1) Definition of the values in the row; (2) detailed description of the variable; (3) regional reference of the time series, in the case of changing regional borders; (4) regional restrictions; (5) calculation method, (6) identification of missing data under specification of the date (for example the year) of the missing data; (7) identification of values requiring further explanation.

Finally an additional field "table" is created, in which the table number of the researcher's publication is entered.

The following workflow results from the description made above:[12]

1. Data processing

Step 1: Structuring of study data

A thematic subdivision of the study is developed after reviewing the publication, the possible additional information, and the study data. Predominantly the given structure of the primary researcher is adopted, because he has already made the content structure of the data on the basis of his research question. In exceptional cases a structure must be developed with the aim to display the data in histat clearly arranged by content-specific aspects.

In the factual subdivision (thematical order of data tables and the numbering of the tables) context-specific aspects are associated with capital letters. The numbering of the data tables starts with the capital letter of that specific aspect the table is associated with. Then a numerical sequence follows the capital letter (e.g.: A.1; A.2; B.1). Thus, the structure of the study's data part in histat reflects the structure of the researcher's work.

Step 2: Metadata (Data description)

In the following, the identification area of the table is constructed by entering the table-title and by creating the table head according to the clearly defined schema for histat data tables. The table title includes the table number, the issues (for example: destination variable, carriers of characteristics), as well as the territorial and temporal reference of the data. Then, the definition of the characteristics and it's expressions are defined in the head of the single table rows (=naming of variables and it's values). References and annotations are inserted for each time series.

Step 3:

Transferring the data into the prepared table field either by importing, by scanning, or entering the data manually.

2. Writing the study description

The study description is written according to the information of the researcher's publication and possibly additional submitted material of the investigator (content of study description: research question, embedding the study in the research context, period of investigation, territory of investigation, thematically subdivision of the data and a data-table-list, sources and annotations, PDF-document: detailed list of sources, detailed annotations, in some cases further documents and methodological report).

---

[12] The working steps depicted above do refer to the processing of acquired research studies. Data compilations, created by GESIS-DHS-employees, require in advance an additional design phase, to define the scheduled data project as well as to review the sources in detail. Then, a content-oriented division of the data will be made. At the same time, parts of the study description are completed, which are important for sources and the data situation (list of sources, list of annotations, content-related subdivision). Only then the data collection and the creating of the data tables marking area can be done.
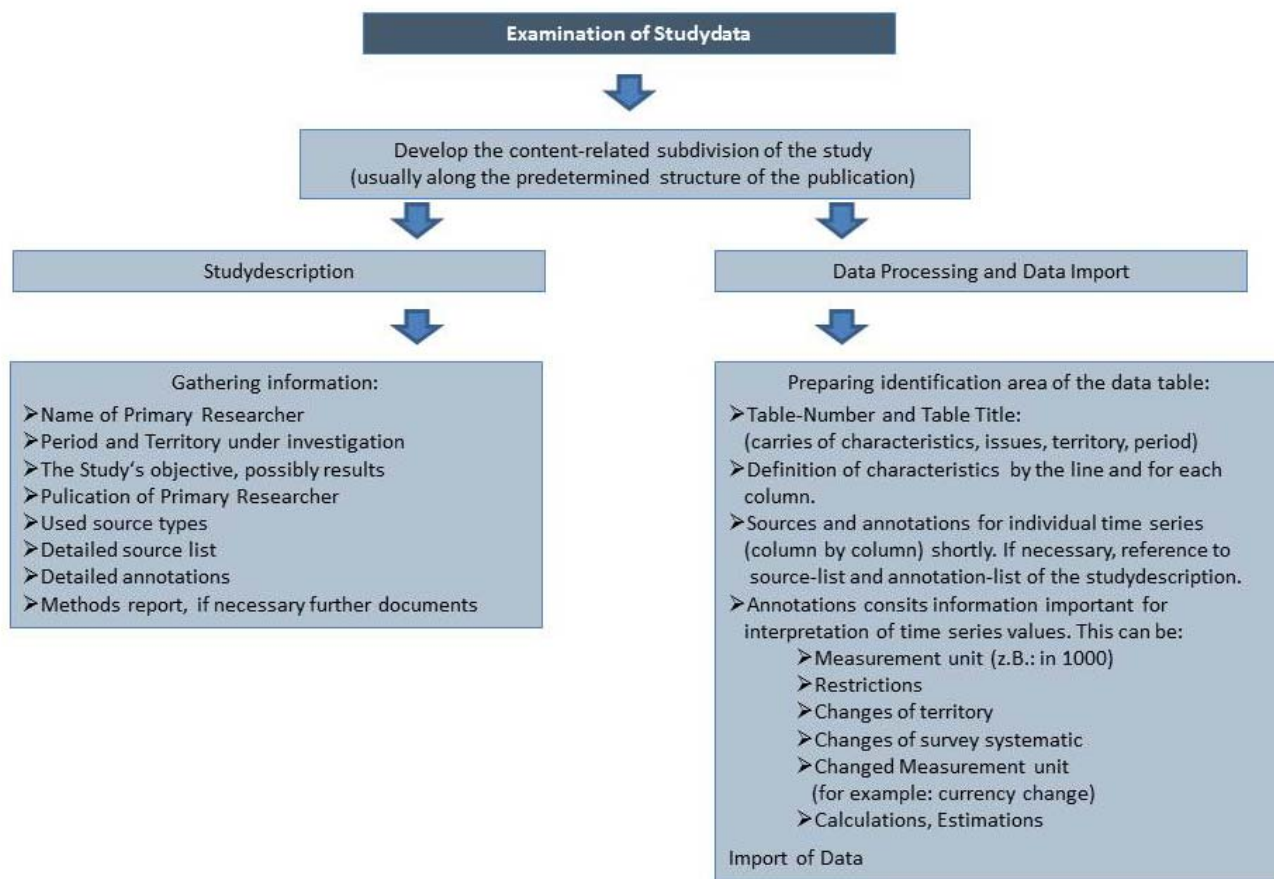
Figure 6: Steps of study and data processing

The processed studies are imported using a special programmed VB.net-import module. Both the study details, which are written in a Word document and the data tables, which are prepared in Excel, are transformed into the format of the online database histat.

5. User Functions in histat

The website http://www.gesis.org/histat/ offers the online-database histat. The use of histat (study- and data research as well as data download) is free of charge. Users are only requested to register when they open the menu indicating the time series, i.e. when they are about to download data. The registration is free of charge. On completion of the registration process via registration form, the user receives an individual user ID and a password via e-mail. This password allows the user a free data download at any time.

5.1 Selection Options
Time series data can be searched in four different ways:

1) **Research by topics and studies allocated to these topics:**

With the start of any data import, each study concerned is allocated to a superordinate topic. The user can call a current topic up by clicking the menu point "Themen" ("topics"). Below any such topic link, the currently available studies are listed (both the title and author of each study is indicated).

16

Each single study within any given topic area offers two display options. On the one hand, the link option "Beschreibung" ("study description", blue button) leads directly to a text view of the desired study description, including a comprehensive list of sources, a commenting part and a downloadable PDF file. The orange button with the information on the number of time series, on the other hand, is linked to the data substructure (i.e. the data tables) of a study. After selecting complete data tables or separate time series from the different data tables contained in a study, this selection is at first displayed in the form of a result. Now the user can opt if he wants to save the previously selected data-file as excel-file or as text-document (CSV). Whenever required, this databased topic catalogue will be supplemented by further studies which might include new topics as well.

Currently 28 themes in histat are offered. The topics are:

- Arbeit (Work)
- Bauen (Housing Construction)
- Bevölkerung (Population)
- Bildung (Education)
- Einkommen (Wages, Income)
- Energie (Energy)
- Geld (Money)
- Gesundheit (Health)
- Handel (Commerce)
- Hanse (Hanseatic League)
- Industrie (Industry)
- Innovation (Innovation)
- Konjunktur (economic situation)
- Kriminalität (Criminality)
- Kultur (Culture)
- Landwirtschaft (Agriculture)
- Preise (Prizes)
- Sozialstaat (Welfare State)
- Staatsfinanzen (Public Finances)
- Städte (Cities)
- Umwelt (Environment)
- Unternehmen (Enterprises, Business)
- Verbrauch (Consumption)
- Verkehr (Traffic)
- Versicherungen (Insurances)
- VGR (National Accounts)
- Wahlen (Elections)
- Übergreifend (Overlapping: studies which have data to several topics)
- Demonstrationsbeispiele (Demonstration-Excamples)

2) Researching names of collectors/researchers and correspondingly allocated studies::

Under the menu option "Namen" ("names"), an alphabetical list with the names of the primary researchers, presented in alphabetical order, can be called up. Like in the "Themen" ("topics") menu, the studies compiled here are marked by their respective title and offer the display options "Beschreibung" ("description") and data-download (the orange button linked with the downloadable dataset).

3) Selection according to starting time periods of time series:

The menu point „Zeiten" (Time Periods) allows a selection of all studies with data dealing with the selected period. The following starting periods of time series are listed in the selection menu:

- ab 16. Jh.
- ab 17. Jh.
- ab 18. Jh.
- 1800er
- 1810er
- 1820er
- 1830er
- 1840er
- 1850er
- 1860er
- 1870er
- 1880er
- 1890er
- 1900er
- 1910er
- 1920er
- 1930er
- 1940er
- 1950er
- 1960er
- 1970er

Many studies with long time series cover several periods. Examples are:

- ZA 8082: *Abel, Wilhelm*: Agrarkrisen und Agrarkonjunktur. Eine Geschichte der Land- und Ernährungswirtschaft Mitteleuropas seit dem hohen Mittelalter. Dritte, neu bearbeitete und erweiterte Auflage. Hamburg und Berlin: Verlag Paul Parey, 1978.
  (Agricultural Crises and Agricultural Markets. A History of Agriculture and Food Sector in Central Europe since the Middle Ages.)
  ▶ Period of investigation: 1201 - 1951
- ZA8270: *Lösch, August*: Bevölkerungswellen und Wechsellagen. Beiträge zur Erforschung der wirtschaftlichen Wechsellagen Aufschwung, Krise, Stockung (Hrsg. von Arthur Spiethoff), Heft 13. Jena: Verlag von Gustav Fischer 1936.
  (Population Waves and Economic Situation. Contributions to the Study of Economic Exchange, Economic Revival, Crisis, and Stagnation.)
  ▶ Period of investigation: 1803 – 1913
- ZA8179: *Metz, Rainer*: Säkulare Trends der deutschen Wirtschaft. In: North, Michael (Hrsg.): Deutsche Wirtschaftsgeschichte – ein Jahrtausend im Überblick. C.H. Beck, München, 2005.
  (Secular Trends in the German Economy.)
  ▶ Period of investigation: 1800 – 2003.
- ZA8523: *Nitsch, Meinolf / Gudermann, Rita*: Agrarstatistik der Provinz Westfalen 1750-1880. Paderborn: Verlag Ferdinand Schöningh, 2009.
  (Agricultural Statistics of the Province of Westphalia 1750-1880.)
  ▶ Period of investigation: 1559 - 1913
- ZA8388: *Pfenning, Winfried; Flora, Peter*: Städteentwicklung und Urbanisierung in Westeuropa 1815 – 1975. In: Flora, P. u.a., 1983: State, Economy, and Society in Western Europe 1815-1975. A Data Handbook in Two Volumes. Volume II: The Growth of Industrial Societies and Capitalist Economies. Chapter 3: Cities and Urbanization. Frankfurt/Main: Campus, S. 247-282.
  (Urban Development and Urbanization in Western Europe, 1815-1975.)
  ▶ Period of investigation: 1815 – 1975.
- ZA8503: *Sensch, Jürgen*: Zur Entwicklung der Arbeitszeit in Deutschland von 1800 bis 2010. histat-Datenkompilation online.
  (To the Development of Working in Germany 1800-2010)
  ▶ Period of investigation: 1800 – 2010

By selecting a period (eg the period 1800), all studies displayed that have time series starting with the selected period (in our example, the starting of the series we search for is between 1800 and 1809):

- ZA8270: *Lösch, August*: Bevölkerungswellen und Wechsellagen. Beiträge zur Erforschung der wirtschaftlichen Wechsellagen Aufschwung, Krise, Stockung (Hrsg. von Arthur Spiethoff), Heft 13. Jena: Verlag von Gustav Fischer 1936.
  (Population Waves and Economic Situation. Contributions to the Study of Economic Exchange, Economic Revival, Crisis, and Stagnation.)
  ▶ Period of investigation: 1803 – 1913

- ZA8179: *Metz, Rainer*: Säkulare Trends der deutschen Wirtschaft. In: North, Michael (Hrsg.): Deutsche Wirtschaftsgeschichte – ein Jahrtausend im Überblick. C.H. Beck, München, 2005.

(Secular Trends in the German Economy.)
  ▶ Period of investigation: 1800 – 2003.

- ZA8503: *Sensch, Jürgen*: Zur Entwicklung der Arbeitszeit in Deutschland von 1800 bis 2010. histat-Datenkompilation online. GESIS Datenarchiv, Köln. histat. Studiennummer ZA8503, Datenfile Version 1.0.0
  (To the Development of Working in Germany 1800-2010)
  ▶ Period of investigation: 1800 – 2010.

4) Research using search words (keywords) covering all studies in the database:

histat enables its users to carry out a cross-study search for time series-related data by means of search words/keywords. In this case, access to the data stock is obtained by means of a full-text research conducted on all studies in the database. Users may optionally launch their individual keyword research in the study details or in the data information (i.e. the table head above the data tables). This search by keywords is supported by the Standard Thesaurus for Economics (STW).[13] Studies and time series, which can be allocated to a certain keyword, are listed to be viewed and downloaded so that the clear attribution of each time series to a certain study and its author is guaranteed. Thereby histat allows the targeted access to specific time series as well as the combination of data gathered from studies of different character and content. In conclusion, the system maintains – and this is worth noting – the allocative efficiency of the database with regard to the combination of time series, studies and their authors. That is to say, the assignment of each selected time series and data table to the relevant study and primary researcher is always maintained.

Of course, studies are imported into histat and thus released for download only with the consent of the primary researcher.

6. Integration of Historical Studies in the Overall Context of the GESIS-Data Archive for the Social Sciences (DAS): Datacatalogue (DBK), Versioning, Digital Object Identifier (DOI)

a) The GESIS Datacatalogue and its direct link to histat

GESIS offers on its website the Data Catalogue (DBK).[14] The DBK contains all the study descriptions of the survey research studies as well as of the studies on social and economic history archived in GESIS data archive. Issues of historical studies treated in their study description, appears also in the GESIS-datacatalogue (DBK). The DBK allows a simple and an advanced full-text search in the entire study stock. Furthermore, the DBK provides for a large part of the studies links from the study description to their data. We will focus on some description fields of the data catalogue (DBK) below with particular reference to the historical studies:

---

[13] The Standard Thesaurus for Economics (Standard-Thesaurus Wirtschaft, STW) provided by the Hamburgisches Welt-Wirtschafts-Archiv (HWWA) and its partner institutes, the Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Kiel (ZBW) and Gesellschaft für Betriebswirtschaftliche Information mbH, München (GBI), is an efficient instrument to make economy and business-related sources accessible. Moreover, it serves as an efficient search tool for terms and expressions within the field of economics.

[14] WEB: https://dbk.gesis.org/dbksearch/index.asp?db=e/

- **Date of Collection:** Year of the study's field time. Information can be given via a standardized specification as well as a free text. Differing from the survey studies for historical studies, the investigation period is specified. So it is the (historical) period to which the study data refer to.

- **Abstract:** Description of the study's content. In case of studies, integrated into the online-database histat (http://www.gesis.org/histat/), the study-description as well as the content-related subdivision of the study-data is entered into the abstract-field of the DBK. The content-related subdivision provides information on the data collected for this study and gives readers deeper information about the data part of the study.

- **Geographic Coverage:** geographical entity, on which the selection is based, listed either by code of the standard ISO3166-1 /-2 or as free text. In case of historical studies, a free text identification is very important, because a standardized indication of territories is not possible as a result of the variety of German states and the diverse territorial changes in German history. Therefore a very rough classification based on the ISO standard has been made, but must always be complemented by the specifications in free text format. The rough version of the ISO standard is: DQDE: Historical German territories before 1871, DXDE: German Empire (1871-1945), DDDE: former Democratic German Republic (GDR) (1945-1990), DE: former Federal Republic of Germany (FRG) 1945-1990 and Germany since 1990.

- **Selection Method:** Method of data collection. For historical studies, random samples taken from file stocks are an example for a data collection method. Often, however, the researcher must compile the data from a variety of printed documents. The approach and the sources used by the researcher for the data collection must be termed. In case of too long source lists the description will be limited to the indication of source types.

- **Analysis-System**: (dataset) number of units; number of variables / Data-Type / Analysis-System(s). For historical studies the number of units corresponds to the number of time points, collected. For the number of variables the number of time series is entered. Time series of historical studies correspond to variables of survey studies.

- **Category:** The archived studies of GESIS-Data Archive can be grouped to different content categories, as for example: State, political ideology, political parties, labor and employment, Family, and so on. For historical studies the content category is: Historical Studies (Daten Historischer Studien).

- **Category of access:** The access categories define the access-possibilities to the data for secundary analysis. Principal investigators and data donors determine the shape of the data access and thus the access category. If they wish restricted access or want to be informed, who used their data, this is taken into account. Studies, which can be downloaded via histat were released for academic research and teaching or without restrictions for everybody by the study donors and primary researchers. Therefore, histat-studies are attached with the access category A or 0.

- **‚Data & Documents, Data access':** Under ‚Download of Data & Documents' the DBK offers a link to the study in the online-database histat. By clicking on the link the study and it's data is activated in histat. The user can now download the study-data from histat.

b) versioning of studies

Having provided the study for the research, a version number is assigned to the study. This version number informs the user by means of a three-digit number according to the scheme "major-minor-revision" whether, when, and which items of the data set have exactly changed. This version number is shown in the DBK under 'Errata & Versions'. Applying this scheme (major-minor-revision) on studies with time-series data, we obtain the following classification:

As a **revision** changes in the labeling field of data tables are named, which have no effect on the content-related meaning of time series (eg correction of spelling errors).

Minor changes are changes in the data tables that have an impact on the content-related meaning of the relevant time series. These are changes or corrections of time series annotations or of individual values in a time series.

Major are changes involving the addition of complete new time series data or entire new tables.

The version number consists of three digits in the form n.n.n, which reflects the changes made in the form major.minor.revision. Study data without changes get the version number 1.0.0. If a simple correction was made in the form of a revision, the study obtains the version number 1.0.1. Changes in the minor-area result into a version number 1.1.0. Profound changes of the type major lead to version 2.0.0, which indicates that the current data content has been significantly expanded or modified.

This versioning is also included in the study details of histat. If there have been made changes at the data, a detailed description of the versioning can be found in the histat study description at the end of the described under the heading 'information about new version'.

c) da|ra registration of historical Studies and assigning of a 'Digital Object Identifier' (=DOI)

Da|ra is the Registration Agency for Social and Economic Data, which is offered by GESIS and ZEW since 2010.[15] "This infrastructure lays the foundation for long-term, persistent identification, storage, localization and reliable citation of research data." In order to facilitate a reliable identification and citation, research datasets and studies are provided with a digital object identifier (DOI). The DOI is a linkage to the research study and data, which is an URL-independent reference to the linked resources. histat studies are also registered in da | ra and thus receive a DOI name.

7. Technical Basis of histat
a) General information

The technical realization of the online database histat as well as the technical maintenance has been undertaken by the company Dataquest in Göttingen (http://www.data-quest.de/unternehmen/). Implemented in 2004 as an entity-relationship model, histat can be technically classified as a MySQL database using the Linux operating system, the Apache webserver and PHP as programming language (= LAMP system). In short, histat relies exclusively on open-source software.

---

[15] http://www.da-ra.de/en/home/

The study description and the processed Excel tables are imported directly into the MySQL database by a VB.net module, i.e. the transformation of the data into the existing database format and their transfer is effected via ODBC (Open DataBase Connectivity) onto the Linux server.

The import of individual studies into the database is mainly based on three components:

1. Import of study details: The study description is written as word-document. The content of the description (ie Study title, name of researcher, description of research question and content-related subdivision of data tables, publication, sources and commenting part) are transferred into the VB.modul via copy and paste.

2. Import of PDF files for download, added by supplementary texts about the respective study. The PDF file is stored with the study details and can be opened by the user.

3. Import of Excel files processed according to a pre-defined structure which constitute the data stock of each individual study.

The study, to be imported into histat, is allocated to an overall topic. At present, histat offers studies for research on a range of 28 topics for download.

b) Technical details[16]

The structure of the database surface is hierarchical, in order to ensure that the offered search facilities and the subsequent presentation of results always focus on the study. Time series are always associated with a table. The meta-information (fields of study description and the contents of the table headers) are stored in the database in the eight tables, named: Aka_Themen (topics), Aka_Projekte (projects), Aka_Schluesselmaske (key mask), Aka_Codes, Aka_CodeInhalt (code content), Aka_SchluesselCode (key code), aka_Schluesselindex (key index), and Lit_ZR (literature time series). The data and the numerical values of the studies are in the table Daten_Aka (data_aka). The values in ‚Daten_Aka' are arranged one below the other as pair of values in the format: time – value.

The hierarchical structure of HISTAT is essentially
- Topics (Aka_Themen)
- Studies (Aka_Projekte)
- Datatables/Subdivision/master key (Aka_Schluesselmaske)
- Time series (aka_schluesselindex)
- Data (Daten_Aka)
In all these five levels, there is a technical primary key and corresponding foreign keys, on which the relationships are fixed. In addition, in each case one or more secondary keys have been created for the purposes of searchable indices.

---

[16] Basis of this presentation is the technical documentation of S. Succhi, A. Noack and N.Witali (DataQuest), published in: histat - Zeitreihen zur Historischen Statistik von Deutschland online. GESIS-WorkingPapers 2012|09, S. 81-100. (downloadbar at: http://www.gesis.org/histat/docs/TechnicalReport_2012-09.pdf)

Overall, the structure of the table structure in histat is as follows:

Table	Aka_Themen (Topics)
Content	Thema (Topic)
Primary Key (technical)	ID_Thema
Foreign Key (Search)	Thema
Example	Bildung (Education)

Table	Aka_Projekte (Projects, this means the study)
Content	Beschreibung einer Studie (Study-description)
Primary Key (technical)	ID_Projekt (Project, ID of the study)
Foreign Key (Search)	Projektname
Example	Berufliche Schulen und Hochschulen in der Bundesrepublik Deutschland 1949-2001 Teil II: Hochschulen (Title of a study, for example: Vocational schools and universities in the Federal Republic of Germany 1949-2001)

Table	Aka_Schluesselmaske
Content	Hauptschlüssel (content-related subdivision, "tables")
Primary Key (technical)	ID_HS
Foreign Key (Search)	Name
Example	Das Alter der Studierenden (title of a data-table, for example: age of students)

Table	Aka_Codes
Content	Variable (carrier of characteristics)
Primary Key (technical)	ID_CodeKuerz
Foreign Key (Search)	CodeBeschreibung
Example	Hochschulart (type of institution of higher education)

Table	Aka_CodeInhalt
Content	Ausprägungen der Variablen (characteristics, variable values)
Primary Key (technical)	ID_CodeKuerz + Code
Foreign Key (Search)	CodeBezeichnung
Example	Wiss. Hochschulen insgesamt (higher education institutions in total)

Table	Aka_SchluesselCode
Content	Zusammenstellung und Reihenfolge der Variablen eines Hauptschlüssels
Primary Key (technical)	ID_HS + ID_CodeKuerz + Position
Foreign Key (Search)	-
Example	Das Alter der Studierenden = Hochschulart + Staatsangehörigkeit + Geschlecht + Alter (Age of students = „type of institution of higher education" + „Gender" + „Age")

Table	aka_schluesselindex
Content	Metadaten und Suchbegriffe einer Zeitreihe
Primary Key (technical)	ID_HS + Schluessel
Foreign Key (Search)	schluessel_index
Example	Hochschulart: Wiss. Hochschulen insgesamt, Staatsangehörigkeit: Deutsche, Geschlecht: insgesamt, Alter: 18 Jahre (type of institution of higher education: scientific university in total, nationality: german, Gener: total age: 18 years)

Table	Lit_ZR
Content	Quellen und Anmerkungen zu einer Zeitreihe (sources and annotations of a time series)
Primary Key (technical)	ID_HS + Schluessel
Foreign Key (Search)	Quelle + Anmerkung (sources + annotations)
Example	1954 WS-1959 WS: DHB-VIII

Table	Daten_Aka
Content	Datenwerte (data, values)
Primary Key (technical)	ID_HS + Schluessel + Jahr_Sem
Foreign Key (Search)	Anmerkung
Example	1954 WS: 1146 Time value 1945 WS 1146

The relationships between the tables of the data base histat:
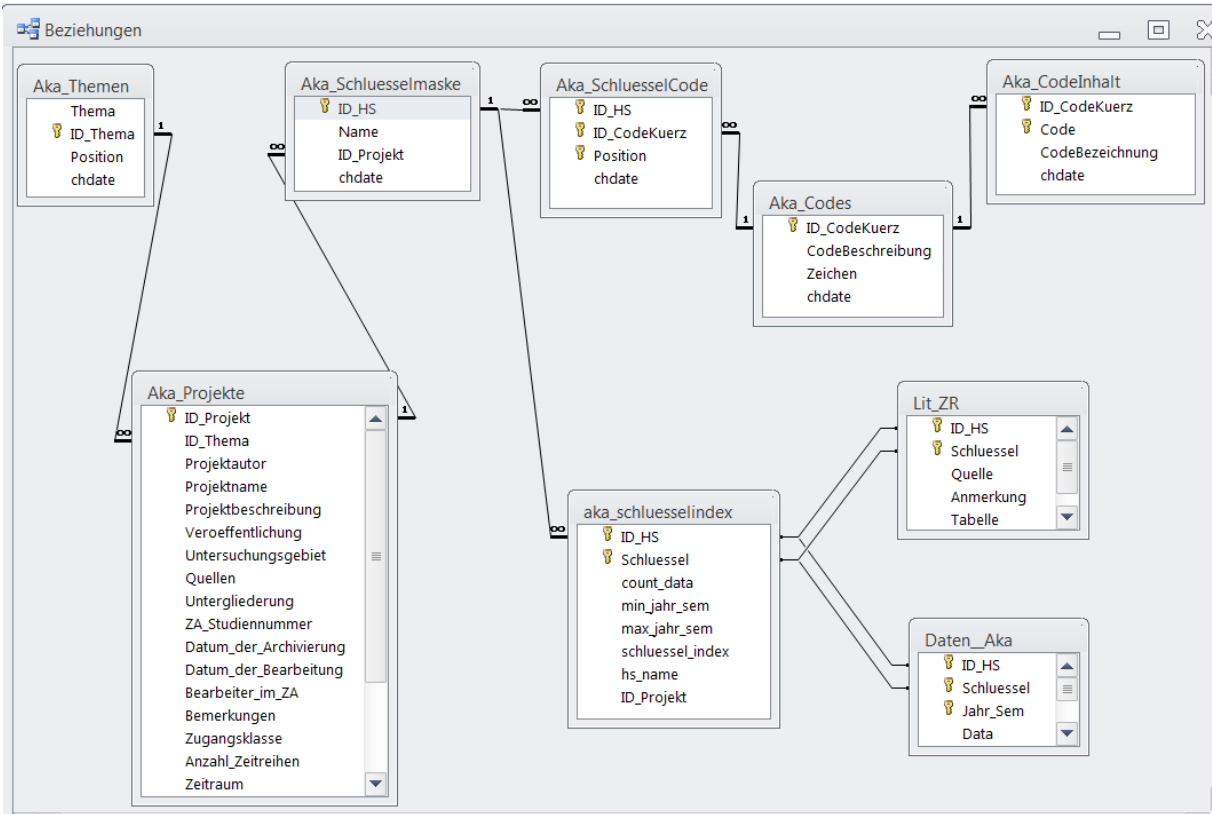

Figure 7: Relationship between the tables of the data base histat

An example of the organization of the data in table "Daten_Aka":

ID_HS	Schluessel	Jahr_Sem	Data	Anmerkung	chdate
118A9BBC44318E8B80B1ECB5DFA1FC7C	00060000000000000000000000000000	1851	396644	NULL	2011-04-05 14:43:10
118A9BBC44318E8B80B1ECB5DFA1FC7C	00061000000000000000000000000000	1851	2.4	NULL	2011-04-05 14:43:10
118A9BBC44318E8B80B1ECB5DFA1FC7C	00070000000000000000000000000000	1851	11859714	NULL	2011-04-05 14:43:10
118A9BBC44318E8B80B1ECB5DFA1FC7C	00071000000000000000000000000000	1851	72.8	NULL	2011-04-05 14:43:10
118A9BBC44318E8B80B1ECB5DFA1FC7C	01080000000000000000000000000000	1851	2080842	NULL	2011-04-05 14:43:10
118A9BBC44318E8B80B1ECB5DFA1FC7C	01081000000000000000000000000000	1851	12.8	NULL	2011-04-05 14:43:10
118A9BBC44318E8B80B1ECB5DFA1FC7C	01090000000000000000000000000000	1851	652115	NULL	2011-04-05 14:43:10
118A9BBC44318E8B80B1ECB5DFA1FC7C	01091000000000000000000000000000	1851	4	NULL	2011-04-05 14:43:10

Time   data, values

Figure 8: Table Daten_Aka in the online-histat database

A data table as it is presented to the user on the screen and as it is available for the download, is made only after a query by using a script, which put together time and value of the database-table 'Daten_Aka'.

For the technical implementation of the histat-database-surface the PHP Framework "Kohana", Version 3.2, was used. Among others, the plugin of the external PHP library "PHPExcel" in version 1.7.7 serves the correct export of data to the Excel format.

The PHP framework "Kohana" is based on the Hierarchic Model View Controller (HMVC) design pattern, consisting of the units 'view', 'model' and 'controller'.

In: Économies et Sociétés, Série «Histoire économique quantitative», AF, n° 50, 6/2015, p. 821-856.

The unit 'model' reads or writes data from the database, the unit 'view' formats the read data and displays it. The unit 'controller' serves as a communication object for the 'model' on the one hand and for the 'view' on the other hand. The 'controller' thus defines which data model and presentation is to be loaded. Thus, a flexible and reusable software was created by DataQuest. Via the histat upload module (or import module) both, the formatted Excel-files and the metadata from the study description are transferred into the database.

The metadata are transferred into the table ‚Aka_Projekte'.

ID_Projekt	ID_Thema	ID_Zeit	Projektautor	Projektname	Projektbeschreibung	Veroeffentlichung	Untersuchungsgebiet	Quelle
8B503C80DF958F86DB7B8F2A2DBA570A	14	14	Sensch, Jürgen	Demonstrationsbeispiel 2: Einfach gegliederte Tabe...	Das zweite Beispiel illustriert einen Datenausschn...	Entfällt.	Gebiet der ehemaligen Deutschen Demokratischen Rep...	Statistisches Bu (Hrsg.), 1994: S B...
C140EA29B5929109D6B14698C9A15E01	14	14	Sensch, Jürgen	Demonstrationsbeispiel 3: Zweidimensionale Verteil...	Dieses Beispiel illustriert eine Datentabelle mit ...	Entfällt.	Deutsches Reich; ab 1950: Bundesgebiet.	Statistisches Bu Wiesbaden (Hrsg B...
B4D82A66376BA34CBBF2E5635212C7F5	14	14	Sensch, Jürgen	Demonstrationsbeispiel 4: Mehrere eindimensionale ...	Das folgende Beispiel illustriert eine Datentabell...	Entfällt.	Gebiet der ehemaligen Deutschen Demokratischen Rep...	Statistisches Bu (Hrsg.), 1990: DI Z...
D401657CB2D314A5A7C2F5A570956DE3	14	14	Sensch, Jürgen	Demonstrationsbeispiel 5: Mehrerer Zeitreihen – Va...	Das folgende Beispiel illustriert eine Datentabell...	Entfällt.	Bundesrepublik Deutschland.	Statistisches Bu Wiesbaden (Hrsg B...
742468FCA13D51B47F5C97BC47112AE1	2	17	Diebolt, Claude	Bildungsausgaben in Deutschland, Frankreich, Großb...	Der thematische Schwerpunkt der vorliegenden Studi...	Diebolt, Claude, 2000: ‚Die Erfassung der Bildungs...	Großbritannien, Frankreich, Deutschland, Spanien u...	Auswahl von nat internationalen Z
A3A29BDE106D2A29B8E2A494CB8AA602	1	17	Steinberg, Heinz Günter	Das Ruhrgebiet im 19. und 20. Jahrhundert: Bevölke...	Die Untersuchung legt ihren Schwerpunkt auf die Be...	Steinberg, Heinz Günter (1985), Das Ruhrgebiet im ...	Das Ruhrgebiet im 19. und 20. Jahrhundert (1845 – ...	Einzelpublikation Statistik der Bur ...
DFC026AFA5E49AF34291BF0627480502	5	5	Abel, Wilhelm	Agrarkrisen und Agrarkonjunktur. Eine Geschichte d...	Das Ziel der Studie von Wilhelm Abel bestand darin...	Abel, Wilhelm: Agrarkrisen und Agrarkonjunktur. Ei...	Mitteleuropa: England, Belgien, Frankreich, Oberit...	Die Preisreihen einzelnen europä Länder ...

Figure 9: Metadata: the table Aka_Projekte of the online-histat database

The current version of the upload module was programmed with Microsoft Visual Studio 2008, Version 9.0.30729.1 SP and requires for the execution .NET Framework, Version 2.0.50727 or a newer Version (Version 3.0 or 3.5).

The current version of the Upload-Module has been programmed using Microsoft Visual Studio 2008, Version 9.0.30729.1 SP. This module requires for execution the .NET Framework, Version 2.0.50727 or a later version (Version 3.0 or 3.5). There is an operating system as
- Windows 98, Windows Me
- Windows NT 4.0 mit Service Pack 6a
- Windows 2000
- Windows 2003
- Windows XP mit Service Pack 2
- Windows 7
required. The Internet Explorer should be available in version 5.01 or higher.

8. Number of Users and Download Statistics

histat was made available for public use in April 2004. The functionality and user-friendliness of this research and download system is constantly tested and checked. At present, the fourth version of histat is in use on our webpages. The fourth revision of histat database was done under leadership of Dr. Thomas Ralph. histat received a new surface design, adapted to the new GESIS web presence. Some new possibilities were added regarding data retrival functionality, it's now quite simple to filter the search results for a specific period. In addition, the DDI standard was introduced for export of the study description. The PHP kohana framework and Javascript jQuery were used as new underlying

technical basis for the presentation layer of the database. The sourcecode of the project is now freely available at GitHub: https://github.com/data-quest/histat-web[17]

The histat database research service, including the download of all data and PDF documents contained herein, is offered free of charge. However, registration is obligatory. The registration is free of charge, too. Before the user starts downloading, he or she is requested to register in the user administration registry by entering his password, he has got via e-mail at his first registration. We request this information for the evaluation of user acceptance concerning our histat offer.

At present (February, the 28th 2015) a total of 435 studies and about 331000 time series are offered by histat, distributed over 28 topics.

Since 2004 until today (December 31, 2014) a total of 4552 individuals and institutions have registered in histat, including about 300 from abroad.

There are among the registrations world-renowned institutions such as the universities Cambridge, Oxford, Zürich, Berkeley, Princton and Stanford.

Since 2007 we have recorded an average of 500 new registrations. In total, since histat is offered, 23218 downloads have been completed, 18326 of which were carried out between 2010 and 2014.

New registrations in histat, 2004-2014

---
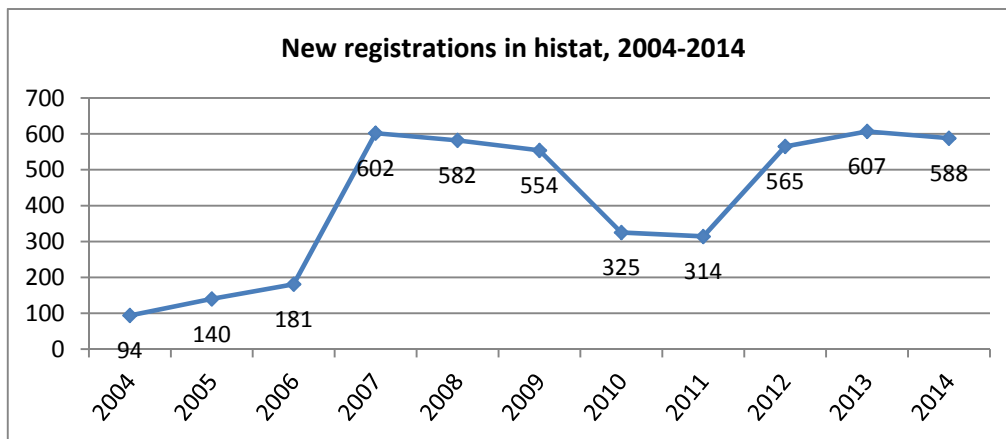
[17] The first three revisions of histat, conducted by Dataquest in collaboration with G. Franzmann, have focused mainly on improving the menu-guided user surface. Furthermore, the histat administration module has been significantly expanded: The user statistics was enhanced in order to differentiate the study-demand according to the topics and to the use of downloaded data (eg: for research, teaching, for a qualifying work or for other purposes). Furthermore, the E-tracker-system, which was acquired and implemented by GESIS on it's WEB-Sites, has been implemented in histat on the study level and data downloads, too.

Download from histat, 2004-2014
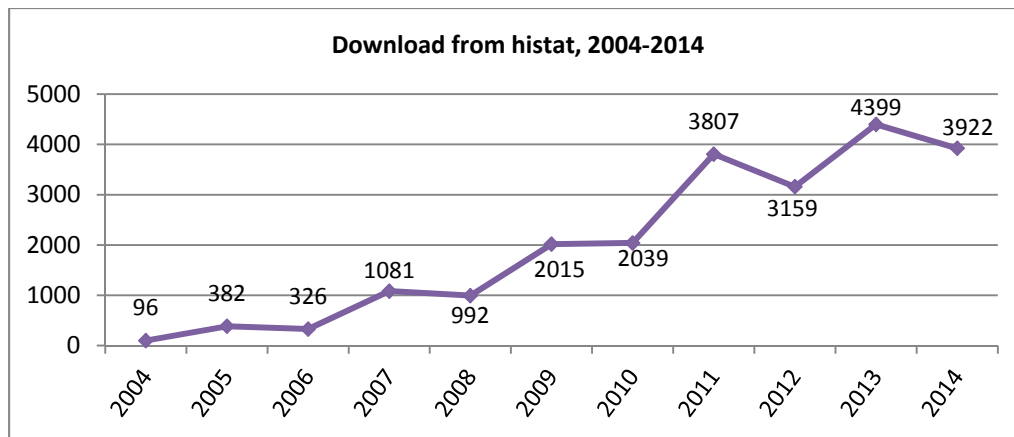
References

Albrecht, Ulrike und Kunz, Andreas, 1990: Building a Databank on German Historical Statistics.   In: Metz, Rainer / Van Cauwenberghe, Eddy / van den Voort, Roel (eds): Historical Information Systems. Leuven University Press, 77-86

Graf Finck von Finckenstein, Hans Wolfram (1960): Die Entwicklung der Landwirtschaft in Preußen und Deutschland, 1800-1930. Würzburg, Holzner.

histat - Zeitreihen zur Historischen Statistik von Deutschland online. GESIS-WorkingPapers 2012|09, S. 81-100. (downloadbar at: http://www.gesis.org/histat/docs/TechnicalReport_2012-09.pdf)

Kunz, Andreas (1979): Historische Statistik von Deutschland. Ein Forschungsschwerpunkt der Deutschen Forschungsgemeinschaft 1981 – 1991. In: Historical Social Research (HSR) 22/2, 1997, S. 236-249.

Kunz, Andreas, 1990: Eine Datenbank zur Historischen Statistik von Deutschland. In: Diederich, N. / Hölder, E. / Kunz, A. u.a., 1990: Historische Statistik in der Bundesrepublik Deutschland. Band 15 der Schriftenreihe Forum der Bundesstatistik, hrsg. vom Statistischen Bundesamt, Wiesbaden. Stuttgart: Metzler-Poeschel, S. 160.

Lundgreen, Peter (unter Mitarbeit von Jana Scheunemann und Gudrun Schwibbe) (2008): Berufliche Schulen und Hochschulen in der Bundesrepublik Deutschland 1949–2001. Datenhandbuch zur deutschen Bildungsgeschichte, Band VIII. Göttingen: Vandenhoeck & Ruprecht.

Reinke, Herbert (1981): Towards Standards for the Description of Machine-Readable Historical Data. In: Historical Social Research (HSR), Nr. 18, S. 3-10.

Spoerer, M., 1996: Von Scheingewinnen zum Rüstungsboom. Die Eigenkapitalrendite der deutschen Industrieaktiengesellschaften 1925 –1941. Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte, Beihefte (123). Stuttgart: Franz Steiner Verlag.

Thome, Helmut/Birkel, Christoph (2007). Sozialer Wandel und Gewaltkriminalität. Deutschland, England und Schweden im Vergleich, 1950-2000. Wiesbaden: VS Verlag für Sozialwissenschaften.

Titze, H. unter Mitarbeit von Herrlitz, H.-G., Müller-Benedict, V. und Nath, A., (1987): Das Hochschulstudium in Preußen und Deutschland 1820 - 1944. Datenhandbuch zur deutschen Bildungsgeschichte, Band I: Hochschulen, 1. U. 2. Teil. Göttingen: Vandenhoeck & Ruprecht.