

histat - Zeitreihen zur Historischen Statistik von Deutschland online: Aufbau, Inhalt, Aufbereitung, technische Umsetzung

Rahlf, Thomas; Metz, Rainer; Sensch, Jürgen; Franzmann, Gabriele; Suchi, Stefan; Noack, Andre; Mik, Witali

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Rahlf, T., Metz, R., Sensch, J., Franzmann, G., Suchi, S., Noack, A., Mik, W. (2012). *histat - Zeitreihen zur Historischen Statistik von Deutschland online: Aufbau, Inhalt, Aufbereitung, technische Umsetzung*. (GESIS-Technical Reports, 2012/09). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-325999>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

hizat

Zeitreihen zur Historischen Statistik von Deutschland online

Aufbau, Inhalt, Aufbereitung, technische
Umsetzung

*Thomas Rahlf, Rainer Metz, Jürgen Sensch,
Gabriele Franzmann, Stefan Suchi, Andre Noack,
Witali Mik*

GESIS-Technical Reports 2012|09

hstat

**Zeitreihen zur Historischen Statistik von
Deutschland online**

Aufbau, Inhalt, Aufbereitung, technische Umsetzung

*Thomas Rahlf, Rainer Metz, Jürgen Sensch,
Gabriele Franzmann, Stefan Suchi, Andre Noack, Witali Mik*

GESIS-Technical Reports

GESIS – Leibniz-Institut für Sozialwissenschaften

Unter Sachsenhausen 6-8, 50667 Köln

50667 Köln

Telefon: (0221) 476 94 - 0

Telefax: (0221) 476 94 - 199

E-Mail: thomas.rahlf@gesis.org

ISSN: 1868-9043 (Print)

ISSN: 1868-9051 (Online)

Herausgeber,

Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

Zusammenfassung

Der Datenservice Historische Studien ist Teil des Datenarchivs für Sozialwissenschaften (DAS). Der Studien- und Datenbestand dieses Bereiches zeichnet sich durch eine hohe Qualität aus. Einen bedeutenden Schwerpunkt bilden Studien mit Zeitreihendaten. Sie stellen mittlerweile 70% des gesamten historischen Studienbestandes. Um der interessierten Öffentlichkeit einen komfortablen Zugang zu diesem umfangreichen Zeitreihen-Datenbestand zu bieten, wurde die Online-Datenbank histat von GESIS entwickelt und seit 2004 auf den GESIS-Seiten zugänglich gemacht. Zwischenzeitlich liegt histat in seiner 5. Version vor. Der vorliegende Report soll den Entstehungszusammenhang von histat erläutern, in die Menü-Führung einführen und Einblicke in das Prozedere der Datenaufbereitung für histat -Studien geben. Darüber hinaus wird die technische Struktur der Datenbank vorgestellt.

Summary

The Data Service of Historical Studies is part of the Data Archive for Social Sciences (DAS). Data of historical studies are characterized by a high quality of the collected data. In the meantime the portion of studies with time series is about 70% of all historical studies. Therefore they became an important focus of the historical data collection. In order to offer a comfortable access to this comprehensive inventory of time series for the interested public, the online-database histat was developed by GESIS and published since 2004 via the GESIS-Web. Over the intervening years 5 versions of histat has been published. The report at hand deals with the context in which the database histat was created. Furthermore, an introduction to the menu navigation and a remarkable insight into the procedure of data preparation and processing according to archival standards is given. Additionally we provide information on the technical structure of the data base.

Inhalt

1	Vorwort.....	7
2	histaat auf einen Blick.....	8
3	Bedienung.....	9
3.1	Startseite.....	9
3.2	Menüpunkt Daten.....	9
	3.2.1 Themen.....	10
	3.2.2 Zeiten.....	11
	3.2.3 Namen.....	12
	3.2.4 Neu.....	13
	3.2.5 Top.....	13
	3.2.6 Detailansicht.....	14
	3.2.7 Suche.....	17
3.3	Über.....	19
3.4	Warenkorb.....	19
3.5	Administration.....	20
3.6	Exportfunktion.....	25
3.7	Die Galerie.....	26
4	Vorgeschichte.....	27
4.1	Bedeutung historischer Statistikdaten.....	27
4.2	Daten der Forschung.....	27
4.3	Einbindung der amtlichen Statistik.....	28
4.4	Fortführung der Arbeiten.....	29
5	Zeitreihen in den Wirtschafts- und Sozialwissenschaften.....	30
5.1	Was sind Zeitreihen?.....	30
5.2	Zeitreihen und Zeit.....	30
5.3	Erkenntnisziele der Zeitreihenanalyse.....	32
5.4	Statistische Methoden zur Analyse von Zeitreihen.....	33
5.5	Wer produziert ökonomische und soziale Zeitreihen?.....	36
5.6	Literatur.....	39
6	Inhalt und Nutzung.....	40
6.1	Zeitreihen- und Werteverteilung.....	40
6.2	Überblick über die enthaltenen Studien.....	42
6.3	Nutzungsstatistik.....	53
7	Bestandserweiterung und Akquise.....	55
7.1	Beobachtung der Forschung.....	55
7.2	Anschreiben.....	57

8	Studienbeschreibung und Aufbereitung der Zeitreihen.....	59
8.1	Ziele der Beschreibung und Aufbereitung.....	59
8.2	Die Beschreibung einer Studie.....	59
8.3	Die Aufbereitung der Zeitreihen einer Studie.....	64
8.4	Beispiel für die Aufbereitung der Zeitreihen einer Studie	66
9	Ablauf der Eingabe.....	71
9.1	Formulare.....	71
9.2	Erläuterungen der Formularfelder.....	74
10	Einbettung in das GESIS-Datenkonzept.....	76
10.1	DBK-Integration.....	76
10.2	Versionierung.....	78
10.3	da ra-Registrierung und DOI-Vergabe.....	79
10.4	DDI-Auszeichnung.....	80
11	Technische Umsetzung.....	81
11.1	Logischer Datenaufbau und Datenbankstruktur.....	81
11.2	Web-Oberfläche.....	86
11.3	Das Upload-Modul	95
11.4	Galerie.....	100
12	Anhang	105

1 Vorwort

Der folgende Beitrag stellt mit dem Datenbanksystem **histat** ein Produkt vor, das 2004 in der GESIS entwickelt wurde. Die Online-Plattform hat die Funktion, Zeitreihen aus der historischen, wirtschafts- und sozialwissenschaftlichen Forschung zugänglich zu machen. Der Zugriff auf die Datenbestände sollte auf der einen Seite möglichst einfach erfolgen, auf der anderen Seite wissenschaftlichen Ansprüchen genügen und die angebotenen Zeitreihen in ihrem Entstehungskontext hinreichend dokumentieren.

Aus dieser Idee ist ein erfolgreiches Produkt geworden. Dieses Jahr werden wir über 3.000 registrierte Nutzerinnen und Nutzer aus der ganzen Welt verzeichnen. Die Datenbank umfasst mittlerweile über eine Viertelmillion Zeitreihen aus über 360 Studien. Im letzten Jahr wurden über 70.000 Zeitreihen heruntergeladen. Monatlich werden mehrere Studien ergänzt.

Auch organisatorisch wurde **histat** auf eine solide Grundlage gesetzt. Mit der Auflösung des Zentrums für Historische Sozialforschung und der Etablierung einer Teamstruktur innerhalb von GESIS wird die Datenbank nun dauerhaft vom **Team Datenservice Historische Studien** betreut und weiterentwickelt. Innerhalb des zurzeit zehnköpfigen Teams widmen sich drei Mitarbeiterinnen und Mitarbeiter unterstützt von mehreren Hilfskräften der kontinuierlichen Bestandserweiterung und -pflege.

2011 stellte die GESIS Mittel für einen umfassenden Relaunch bereit. Mit diesen Mitteln wurde die Oberfläche grundlegend überarbeitet und eine einheitliche Benutzerführung über alle Funktionen geschaffen. Hierfür sei GESIS an dieser Stelle ganz herzlich gedankt. Die technische Umsetzung erfolgte durch die Codemanufaktur data-quest, Göttingen, die Gestaltung des Layouts durch Metavor, Hannover. Wir haben die Modernisierung zum Anlass genommen, die bislang für verschiedene Zwecke erstellten Texte zu **histat** zusammenzufassen, zu überarbeiten und in der vorliegenden Form als Technical Report zu publizieren. Für Anmerkungen, Kommentare, Hinweise, Lob und Kritik danken wir insbesondere Jörg Baten, Claude Diebolt, Andreas Kunz, Mark Spoerer, Reinhard Spree, Tobias Vogelsang, Oliver Volckart und Joachim Voth.

Köln, im November 2012

Thomas Rahlf

2 histat auf einen Blick

Studienorientierung	<ul style="list-style-type: none"> • Studien, geordnet nach Sachgebieten. • Innerhalb einer Studie: Untergliederung nach Teilthemen.
Dokumentation der Einzelstudie	<ul style="list-style-type: none"> • Studiendetails: Studientitel, PrimärforscherIn, Studienbeschreibung, verwendete Quellentypen, Quellenverzeichnis, Anmerkungsteil
Ergänzende Texte zur Studie zum Download (PDF-Dokument)	<ul style="list-style-type: none"> • Ergänzende Materialien zur Studie: Publikationen, detaillierte Quellenliste, ausführlicher Anmerkungsteil, Methodenberichte
Metadaten zu einzelnen Zeitreihen	<ul style="list-style-type: none"> • Zu jeder einzelnen Zeitreihe: Definition, verwendete Quellen, Anmerkungen, Tabellenummer und Tabellentitel aus der Publikation; Anmerkungen zu einzelnen Werten in der Zeitreihe.
Datenexport aus einer Einzelstudie	<ul style="list-style-type: none"> • Zitierpflichtige Publikation; • Sämtliche Metadaten zu den ausgewählten Zeitreihen zusammen mit den Daten.
Suchfunktion nach Stichworten: Studienübergreifende Volltextrecherche; thesaurusgestützt	<ul style="list-style-type: none"> • Zeitreihenauswahl aus verschiedenen Einzelstudien mit Hilfe von Suchbegriffen; • Datenexport dieser Auswahl inkl. aller Metadaten zu den einzelnen Zeitreihen.
Administrationsmodul: Nutzerregistrierung; Nutzerstatistik	<ul style="list-style-type: none"> • Anmeldung des Nutzers. Registrierung der Downloads. • Nutzerstatistik: Jahresbericht zur Nachfragefrequenz für die Einzelstudien in histat.

histat ist open source. Sie können den Quellcode unter

<https://github.com/data-quest/histat-web>

und

<https://github.com/data-quest/histat-upload>

frei herunterladen.

3 Bedienung

Ziel der histaat-Anwendung ist es, den Nutzerinnen und Nutzern die Möglichkeit zu geben, in dem Bestand von ca. 250.000 Zeitreihen im Rahmen der ca. 350 Studien zu recherchieren und ihre Rechercheergebnisse bei Bedarf in Tabellenform zu „downloaden“.

3.1 Startseite

Hierzu wird zunächst eine statische Seite Start angeboten:



Der zweite Menüpunkt führt zu Informationen über die angebotenen Daten.

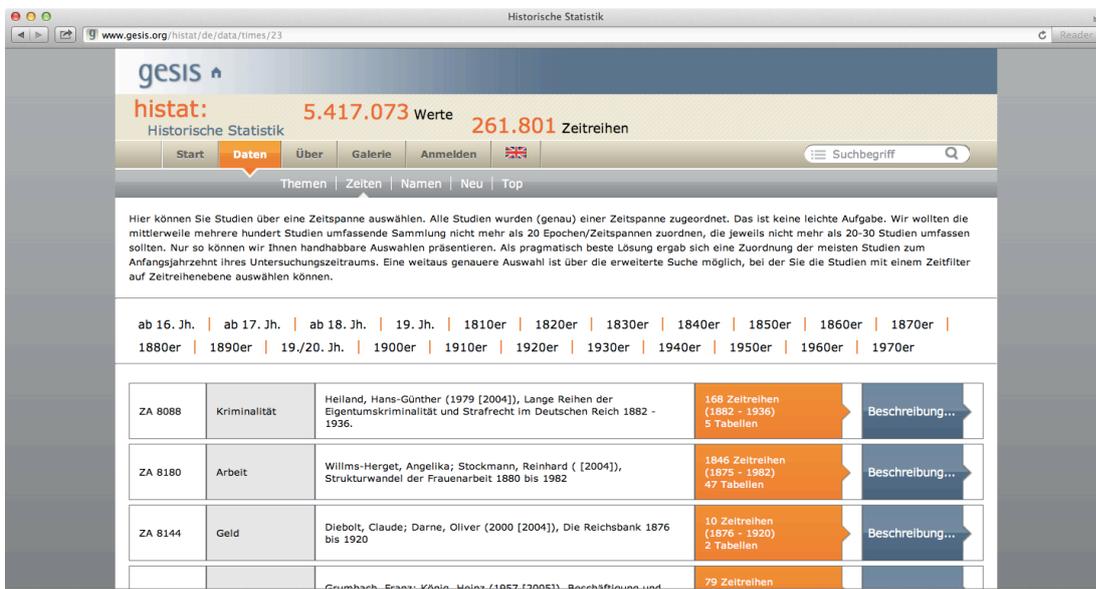
3.2 Menüpunkt Daten

Alle Studien werden in allen Untermenüs in der gleichen Kurzansicht angezeigt:



3.2.2 Zeiten

Die zweite Menüoption ist eine Auswahl von Studien über eine Zeitleiste. Hier können Sie Studien über eine Zeitspanne auswählen. Alle Studien wurden (genau) einer Zeitspanne zugeordnet. Das ist keine leichte Aufgabe. Wir wollten die mittlerweile mehrere hundert Studien umfassende Sammlung nicht mehr als 20 Epochen/Zeitspannen zuordnen, die jeweils nicht mehr als 20-30 Studien umfassen sollten. Nur so können wir Ihnen handhabbare Auswahlen präsentieren. Als pragmatisch beste Lösung ergab sich eine Zuordnung der meisten Studien zum Anfangsjahrzehnt ihres Untersuchungszeitraums. Eine weitaus genauere Auswahl ist über die erweiterte Suche möglich, bei der Sie die Studien mit einem Zeitfilter auf Zeitreihenebene auswählen können (s.u.).



3.2.3 Namen

Der nächste Menüpunkt ist eine Auswahl nach „Namen“. Hier werden alle Autorinnen und Autoren der Studien alphabetisch im Fließtext aufgeführt. Ein Anklicken eines Namens öffnet die Kurzübersicht der betroffenen Studien an der Stelle des Namens:

The screenshot shows the GESIS histat website interface. At the top, the logo 'gesis' is visible, followed by 'histat: Historische Statistik' and statistics: '5.417.073 Werte' and '261.801 Zeitreihen'. Navigation tabs include 'Start', 'Daten', 'Über', 'Galerie', and 'Anmelden'. Below the tabs, there are links for 'Themen', 'Zeiten', 'Namen', 'Neu', and 'Top'. A horizontal menu of letters (A-Z) is present. The main content area is titled 'Liste aller Autorinnen und Autoren' and contains a long list of names, including 'Abel, Wilhelm', 'Albert, Götz', 'Alecke, Björn', etc.

The screenshot shows the GESIS histat website interface with a specific study selected. The page displays a table of studies by Burhop, Carsten. The table has four columns: ID, subject, description, and time period. The selected study is 'ZA 8176' with the subject 'Geld' and the description 'Burhop, Carsten (2002 [2004]), Die Entwicklung der deutschen Aktienkreditbanken von 1848 bis 1913: Quantifizierungsversuche'. The time period is '8 Zeitreihen (1848 - 1883)' and '1 Tabellen'. There are buttons for 'Beschreibung...' next to each row.

ID	Subject	Description	Time Period	Tables	Action
ZA 8176	Geld	Burhop, Carsten (2002 [2004]), Die Entwicklung der deutschen Aktienkreditbanken von 1848 bis 1913: Quantifizierungsversuche	8 Zeitreihen (1848 - 1883)	1 Tabellen	Beschreibung...
ZA 8492	Geld	Burhop, Carsten (2004 [2011]), Die Kreditbanken in der Gründerzeit 1870 - 1879.	201 Zeitreihen (1870 - 1880)	9 Tabellen	Beschreibung...
ZA 8203	Konjunktur	Burhop, Carsten; Wolff, Guntram B. (2005 [2006]), Kompromisschätzung des deutschen Nettosozialprodukts von 1851 bis 1913.	31 Zeitreihen (1846 - 1913)	8 Tabellen	Beschreibung...

3.2.4 Neu

Der nächste Unterpunkt „Neu“ ist eine Auflistung der neuen Studien, also der 20 zuletzt hochgeladenen, in absteigender Reihenfolge ihrer Bearbeitung.

The screenshot shows the 'Neu' page of the hista website. The header includes the 'hista' logo, the text 'Historische Statistik', and statistics: '5.417.073 Werte' and '261.801 Zeitreihen'. Navigation tabs include 'Start', 'Daten', 'Über', 'Galerie', 'Anmelden', and a search bar. The main content area is titled 'Liste der neuen Studien' and contains a table of 20 studies, with the following visible entries:

Studien-ID	Thema	Titel	Zeitreihen	Tabellen	Aktion
ZA 8171	Bevölkerung	Sensch, Jürgen ([2004]), Geschichte der deutschen Bevölkerung seit 1815.	739 Zeitreihen (1500 - 2003)	39 Tabellen	Beschreibung...
ZA 8073	Bevölkerung	Steinberg, Heinz Günter (1985 [005.]), Das Ruhrgebiet im 19. und 20. Jahrhundert: Bevölkerungsentwicklung, Erwerbstätige, Bergbau und Großindustrie.	224 Zeitreihen (1820 - 1980)	10 Tabellen	Beschreibung...
ZA 8270	Bevölkerung	Lösch, August (1936 [2007]), Bevölkerungswellen und Wechsellagen im Deutschen Reich von 1871 bis 1910.	334 Zeitreihen (1803 - 1913)	14 Tabellen	Beschreibung...
ZA 8049	Bevölkerung	Hohorst, Gerd (1978 [2006]), Bevölkerungsentwicklung in Preußen und in dem Kreis Hagen von 1816 bis 1914.	153 Zeitreihen (1748 - 1914)	23 Tabellen	Beschreibung...
ZA 8146	Bevölkerung	Wolf, Christof (1999 [006.]), Religionszugehörigkeit in Westdeutschland von 1939 bis 1987.	332 Zeitreihen (1871 - 1987)	15 Tabellen	Beschreibung...

3.2.5 Top

Der letzte Untermenüpunkt „Top“ ist eine absteigend sortierte Liste von Kurzsichten der meistverwendeten Studien, in absteigender Reihenfolge ihrer Download-Zahlen.

The screenshot shows the 'Top' page of the hista website. The header and navigation are identical to the 'Neu' page. The main content area is titled 'Liste der top Studien' and contains a table of 20 studies, with the following visible entries:

Studien-ID	Thema	Titel	Zeitreihen	Tabellen	Aktion
ZA 8295	Bevölkerung	Besser, Christoph ([2008]), Grunddaten der Bevölkerungsstatistik Deutschlands von 1871 bis 1939.	5205 Zeitreihen (1871 - 1945)	120 Tabellen	Beschreibung...
ZA 8171	Bevölkerung	Sensch, Jürgen ([2004]), Geschichte der deutschen Bevölkerung seit 1815.	739 Zeitreihen (1500 - 2003)	39 Tabellen	Beschreibung...
ZA 8186	Geld	Deutsche Bundesbank (1998 [005.]), 50 Jahre Deutsche Mark. Monetäre Statistiken von 1948 bis 1997.	7004 Zeitreihen (1940 - 1998)	137 Tabellen	Beschreibung...
ZA 8276	Bevölkerung	Kraus, Antje (1980 [2007]), Quellen zur Bevölkerungsstatistik Deutschlands 1815-1875	1289 Zeitreihen (1812 - 1875)	101 Tabellen	Beschreibung...
ZA 8184	Arbeit	Sensch, Jürgen (2001 [2004]), Erwerbstätigkeitsstatistik in der Bundesrepublik Deutschland. Ausgewählte Daten von 1950 bis 2000.	2986 Zeitreihen (1946 - 2000)	79 Tabellen	Beschreibung...

3.2.6 Detailansicht

Studienbeschreibung

Wird von den angegebenen Stellen den Verlinkungen auf die Detailsicht gefolgt, so erscheint eine einzelne Seite mit allen Metadaten zu einer Studie, die bei größerem Umfang einzeln ausklappbar sind ("[mehr]"). Die einzelne Studie wird in histat detailliert dokumentiert, insbesondere wird von den Mitarbeiterinnen und Mitarbeitern des Datenservice Historische Studien eine ausführliche Studienbeschreibung angefertigt. Im Einzelnen sind die folgenden Angaben enthalten: 1. Name der Studie, 2. Leiterin oder Leiter der Studie, 3. die eigentliche Studienbeschreibung, 4. Zeitraum, 5. Anzahl der Zeitreihen, 6. Untersuchungsgebiet, 7. Veröffentlichung, 8. Verwendete Quellentypen (z.B. amtliche Statistiken, Firmenarchive, wissenschaftliche Publikationen), 9. Sachliche Untergliederung der Datentabellen, 10. Tabellenverzeichnis der Publikation (mit direkter Verlinkung auf die Datentabellen), 11. Quellen und Anmerkungen (in der Regel beschränkt auf kurze Angaben, ergänzt um einen Hinweis auf ein PDF-Dokument), 12. Bearbeitungshinweise (Studiennummer, Datum der Archivierung, Datum der Bearbeitung, Bearbeiter oder Bearbeiterin, Bemerkungen, Zugangsklasse).

The screenshot shows a web browser window with the URL `histat.gesis.org/histat/de/project/details/0A84E4690DE01599DEF32838C94C9828`. The page title is 'Historische Statistik'. The main content area is titled 'Studienbeschreibung' and contains the following information:

- ZA 8514 Bevölkerung**: Rothenbacher, Franz, Historische Haushalts- und Familienstatistik von Deutschland 1815 – 1990. (4590 Zeitreihen (1669 - 1991), 145 Tabellen)
- Datensatz (in empfohlener Zitierweise)**: Rothenbacher, Franz, (1997 [2011]) Historische Haushalts- und Familienstatistik von Deutschland 1815 – 1990. GESIS Köln, Deutschland ZA8514 Datenfile Version 1.0.0
- Anmerkungen**: Für jedes Territorium sind für die einzelnen Variablengruppen ausführliche Anmerkungen in dem beigefügten PDF – Dokument dokumentiert. (Button: Mehr)
- Gedruckte Publikation**: Rothenbacher, F., 1997: Historische Haushalts- und Familienstatistik von Deutschland 1815 – 1990. Frankfurt a.M./New York: Campus.
- Studienbeschreibung**: Das vorliegende Datenhandbuch versteht sich als Beitrag zu einer historischen Statistik von Deutschland. Das Datenhandbuch steht dem DFG – Schwerpunktprogramm „Quellen und Forschungen zur Historischen Statistik von Deutschland“ (Förderungszeitraum 1981 - 1991) nahe, da es sich fast ausschließlich auf die amtliche Statistik als Quellengrundlage bezieht. (Button: Mehr)
- Download:** Diese Studienbeschreibung als DDI-XML. (Button: DDI-XML)
- Download weiterer Texte zu dieser Studie im PDF Format (Forschungsberichte, Publikationen, Materialien zur Studie) (Button: 2,51 MB)

Vollständig ausgedruckt ergibt sich hiermit pro Studie ein Umfang von etwa 4-5 Seiten.

In aller Regel wird zusätzlich ein PDF-Dokument zum download angeboten: Dies kann ein eingescannter Ausschnitt aus einer Publikation sein, ein eingescannter Zeitschriftenaufsatz oder andere zusammengestellte Unterlagen.

Tabellenverzeichnis

Historische Statistik

ZA 8514 Bevölkerung Rothenbacher, Franz, Historische Haushalts- und Familienstatistik von Deutschland 1815 - 1990. Tabellenverzeichnis schließen Beschreibung...

A.1.1 Entwicklung der Haushaltsarten, Deutsches Reich, absolut (1871-1939)	10 Zeitreihen
A.1.2 Entwicklung der Haushaltsarten, Deutsches Reich, in Prozent (1871-1939)	10 Zeitreihen
A.1.3 Entwicklung der durchschnittlichen Haushaltsgrößen und von Indexzahlen, Deutsches Reich (1871-1939)	10 Zeitreihen
A.2.1a Haushaltsgrößenkategorien, Haushalte absolut, in Prozent, Deutsches Reich (1900-1939)	22 Zeitreihen
A.2.1b Haushaltsgrößenkategorien, Personen absolut, in Prozent, Deutsches Reich (1900-1939)	22 Zeitreihen
A.3.1. Haushaltstypen, Deutsches Reich (1910-1939)	14 Zeitreihen
A.4.1. Haushaltszusammensetzung, Deutsches Reich (1900-1925)	14 Zeitreihen
A.4.2. Haushaltszusammensetzung, Deutsches Reich (1910-1925)	12 Zeitreihen
A.5.1. Familienzusammensetzung, Deutsches Reich (1910-1925)	24 Zeitreihen
A.6.1. Indikatoren der Bevölkerungsstruktur und Bevölkerungsbewegung, Deutsches Reich (1872-1941)	12 Zeitreihen

zurück 1 2 3 4 5 6 7 8 9 10 11 12 13 14 vor 23 Tabellen

Ein Klick auf den Link zu den Daten führt zunächst zu einer Anmeldemaske:

Historische Statistik

gesis histat: 5.417.073 Werte 261.801 Zeitreihen

Start Daten Über Galerie Anmelden Suchbegriff

Anmelden Registrieren

Anmeldung

Bitte geben Sie Benutzernamen und Passwort ein!

Benutzername:

Passwort:

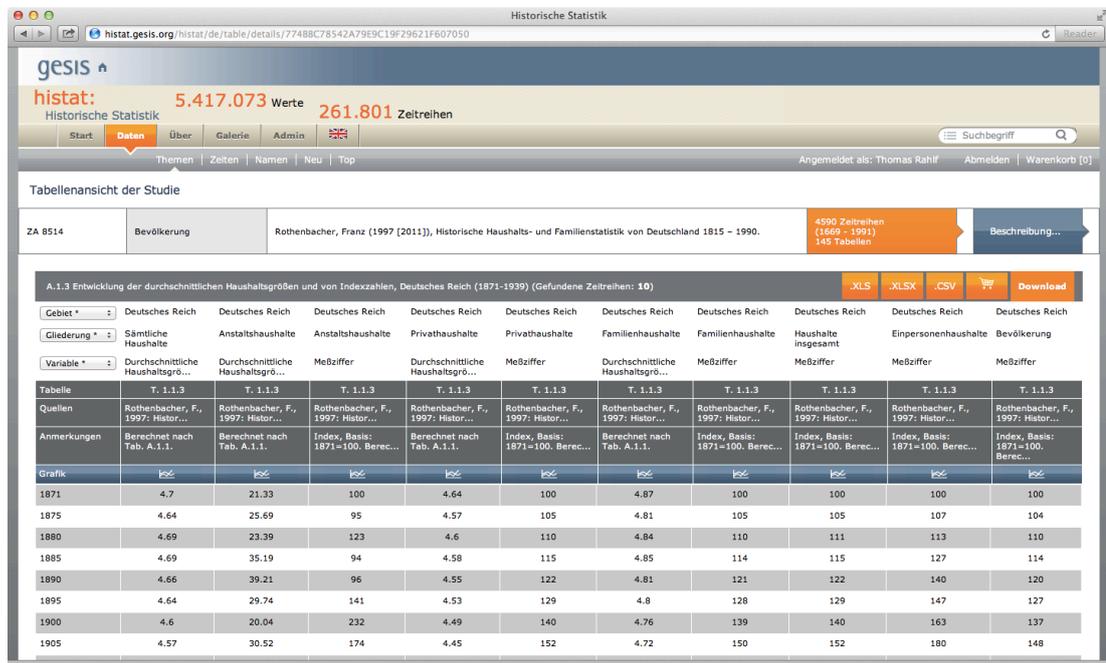
Angemeldet bleiben?

Anmelden

nach oben

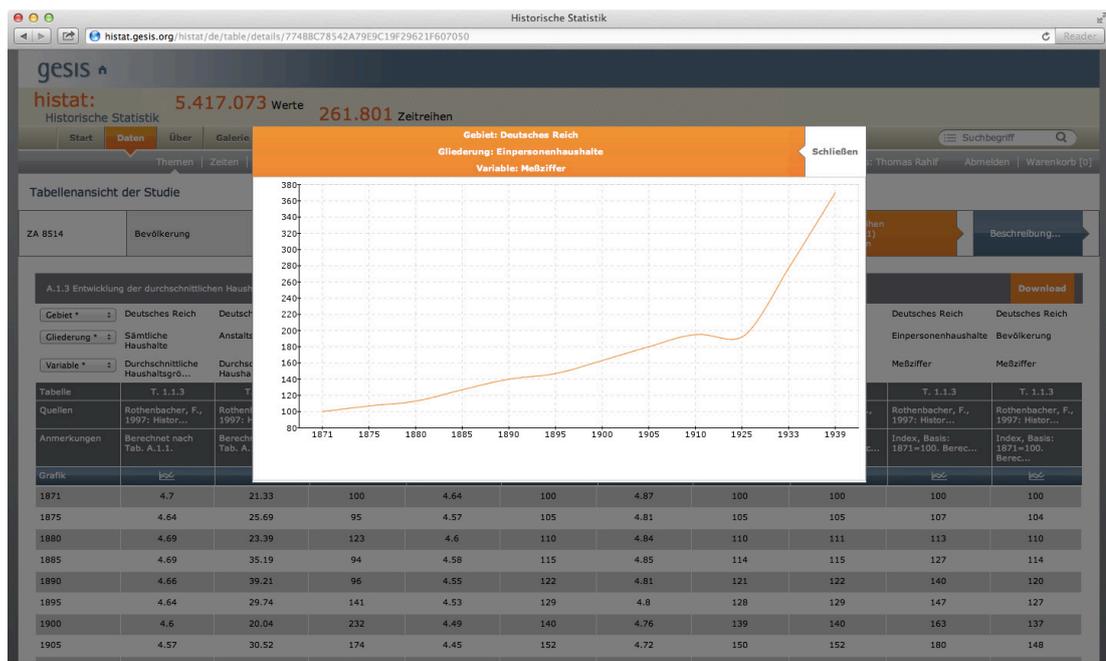
© GESIS Impressum | Sitemap | Last update at 01.11.2012

Anschließend wird die ausgewählte Tabelle angezeigt.



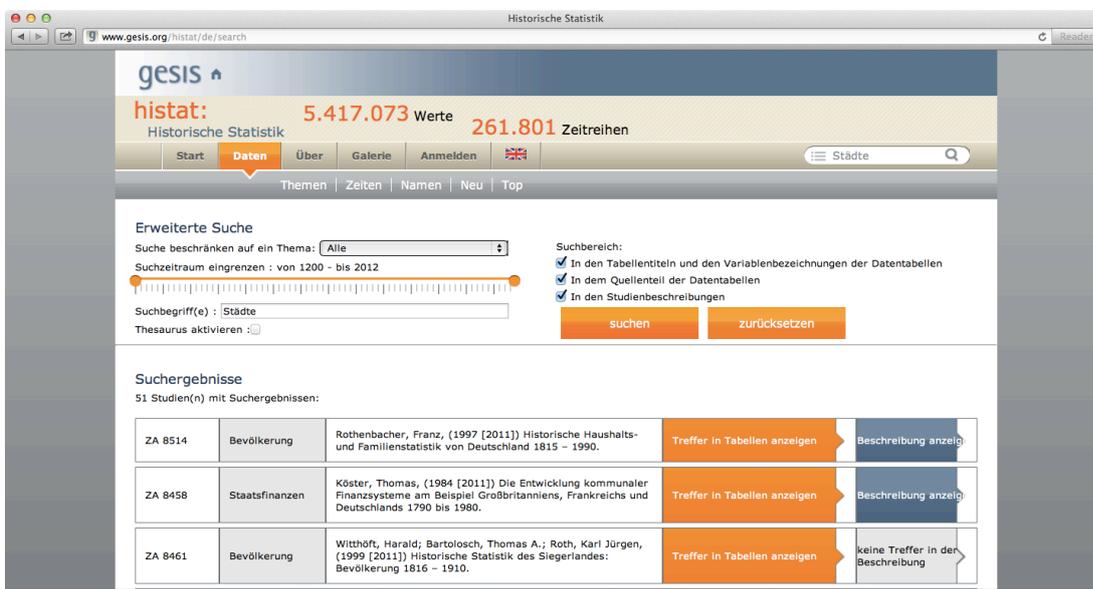
In der ersten Spalte kann in den Kopfzeilen die Tabelle weiter gefiltert werden. Es besteht die Möglichkeit, die Tabelle als CSV-Datei, im XLS- oder XLSX-Format herunterzuladen. Alternativ kann die Tabelle in einen „Warenkorb“ gespeichert werden (s.u.).

Über jeder Spalte befindet sich ein Icon zur Grafik-Schnellansicht, der ein Popup-Fenster mit einer grafischen Abbildung der Zeitreihe öffnet:



3.2.7 Suche

Alternativ zu der Menüauswahl über „Neu“, „Top“, „Zeiten“, „Themen“ und „Namen“ kann auf den gesamte Bestand über eine Volltextsuche zugegriffen werden: mit einer Suchmaske und der Option, entweder „In den Tabellentiteln und den Variablenbezeichnungen der Datentabellen“ oder „In dem Quellenteil der Datentabellen“ oder „In den Studienbeschreibungen“ zu suchen. Alle drei Suchräume können beliebig kombiniert werden. Standardmäßig wird in allen dreien gesucht. Optional kann für die Suche ein Thesaurus verwendet werden. Als Ergebnis wird eine Liste der Studien zurückgeliefert, in deren Tabellentitel und/oder Variablenbezeichnungen die Suche erfolgreich war. Die Ergebnisauflistung erfolgt in der gleichen Ansicht, die Ihnen auch bei der Auswahl über die Menüpunkte „Neu“, „Top“, „Zeiten“ oder „Themen“ präsentiert wird.



Allerdings bekommen Sie hier schon einen Hinweis, ob Ihr Suchbegriff in den Zeitreihen oder in der Studienbeschreibung enthalten ist. Bei einem Klick auf „Treffer in den Tabellen“ anzeigen wird die oben bereits gezeigte Tabellenverzeichnis- und Datenansicht präsentiert, allerdings schon gefiltert auf den Suchbegriff.

Wenn Sie an einem bestimmten Zeitraum interessiert sind, können Sie anhand des Zeitstrahls den Suchzeitraum eingrenzen. Eine Eingrenzung auf den Zeitraum von 1800 bis 1900 liefert im Beispiel mit dem Suchbegriff „Städte“ 27 Studien zurück.

The screenshot shows the GESIS website search results for the term 'Städte'. The page header displays '5.417.073 Werte' and '261.801 Zeitreihen'. The search criteria are set to 'Alle' for the topic and '1800 - bis 1900' for the time period. The search results table lists three studies:

ZA	Thema	Studienbeschreibung	Suchbegriff	Buttons
ZA 8514	Bevölkerung	Rothenbacher, Franz, (1997 [2011]) Historische Haushalts- und Familienstatistik von Deutschland 1815 - 1990.	Städte	Treffer in Tabellen anzeigen, Beschreibung anzeigen
ZA 8420	Bevölkerung	Büch, Henrike, (1998 [2010]) Bevölkerung des Kurfürstentums / Königreichs Hannover 1745 - 1864.	Städte	Treffer in Tabellen anzeigen, Beschreibung anzeigen
ZA 8534	Bauen	Gransche, Elisabeth; Rothenbacher, Franz, ([2012]) Langfristige Entwicklungstendenzen der Wohnverhältnisse in Deutschland 1861 - 1910.	Städte	Treffer in Tabellen anzeigen, Beschreibung anzeigen

Ist der Suchbegriff in der Studienbeschreibung enthalten, wird die Studienbeschreibung auszugswise mit hervorgehobenen Suchbegriffen angezeigt.

This screenshot shows a detailed view of the first study from the previous screenshot. The search criteria remain the same. The study entry for ZA 8514 is expanded to show the following details:

Studienbeschreibung
 Studienbeschreibung [mehr](#)
 sich auf die sieben einwohnerstärksten Bundesstaaten. Daneben werden sieben Großstädte berücksichtigt. Zusätzlich zu diesen Ebenen werden Daten für das Deutsche Reich und die Bundesrepublik Deutschland dokumentiert. Neben der Quellenverzeichnis [mehr](#)
 Amtliche Statistik des Deutschen Reichs; amtliche Statistik ausgewählter Einzelstaaten des Deutschen Reichs; amtliche Statistik ausgewählter Städte.
 Insbesondere wurden die Daten der Volkszählungen berücksichtigt. Ausgewählte Untergliederung [mehr](#)
 Indexzahlen, Preußen, Gesamtstaat (1864-1939) B.1.3. Haushaltsarten, preußische Städte (1867-1910) B.1.4. Entwicklung durchschnittlicher Haushaltsgrößen

3.3 Über

Der Menüpunkt „Über“ ruft eine statische Seite mit Kontaktangaben und weiterführenden Informationen zur Datenbank auf.



3.4 Warenkorb

Alternativ zum direkten Download der Tabellen im XLS-, XLSX- oder CSV-Format können Sie die Tabellen auch in einem Warenkorb zwischenspeichern, um sie mit anderen Rechercheergebnissen gemeinsam und/oder zu einem späteren Zeitpunkt herunterzuladen. Hierzu klicken Sie einfach in der Tabellenansicht auf das Warenkorb-Symbol.

The screenshot shows the GESIS histat website interface. At the top, it displays '5.417.073 Werte' and '261.801 Zeitreihen'. The 'Warenkorb' (shopping cart) section contains two items:

Item ID	Category	Description	Action
ZA 8137	VGR	Ritschl, Albrecht; Spoerer, Mark, (1997 [2011]) Das Bruttozialprodukt in Deutschland nach den amtlichen Volkseinkommens- und Sozialproduktstatistiken 1901-1995.	Tabellen schließen
<input type="checkbox"/>	A.1 Volkseinkommen, Bruttozialprodukt und Bruttoinlandsprodukt in Deutschland, in Mio. M/RM/DM (1901-1950)		10 Zeitreihen
ZA 8514	Bevölkerung	Rothenbacher, Franz, (1997 [2011]) Historische Haushalts- und Familienstatistik von Deutschland 1815 - 1990.	Tabellen schließen
<input type="checkbox"/>	A.1.3 Entwicklung der durchschnittlichen Haushaltsgrößen und von Indexzahlen, Deutsches Reich (1871-1939)		10 Zeitreihen

Buttons at the bottom of the cart include 'Löschen', 'Download', and 'nach oben'. The footer shows '© GESIS Impressum | Sitemap | Last update at 02.11.2012'.

Aus dem Warenkorb können Tabellen einzeln oder zusammen entfernt werden, indem die Tabellen angehakt werden und man anschließend auf den Papierkorb klickt. Der Download funktioniert wie bei einzelnen Tabellen: fährt man mit der Maus über den Button „Download“ werden als mögliche Download-Formate „XLS“, „XLSX“ und „CSV“ eingeblendet. Bei mehreren Tabellen erfolgt der Download in eine ZIP-Datei.

3.5 Administration

Da wir die Daten von unseren Datengeberinnen und Datengebern unter bestimmten Weitergabe-Voraussetzungen erhalten haben, sind wir verpflichtet, die Nutzung von histat kontinuierlich im Auge zu behalten. Zu diesem Zweck und um Missbrauch vorzubeugen können login-Daten von Nutzerinnen und Nutzer von Administratoren auch wieder gelöscht werden.

Benutzerverwaltung

The screenshot shows the 'Benutzerverwaltung' (user management) interface. The top navigation bar includes 'Benutzer Verwaltung', 'Datenänderungen', 'Datenänderung(Tabellenkopf)', and 'Statistiken'. The main content area is titled 'Liste der Benutzer' and displays a table with three rows of user information:

ID	Name	... (blurred)	... (blurred)	... (blurred)	Actions
1	[blurred]	[blurred]	[blurred]	[blurred]	[blurred]
2	[blurred]	[blurred]	[blurred]	[blurred]	[blurred]
3	[blurred]	[blurred]	[blurred]	[blurred]	[blurred]

Weiterhin haben Administratoren die folgenden Einsichtsmöglichkeiten.

Datenänderungen

Neu ist die Möglichkeit, eine Liste der von uns veränderten Daten aufzurufen. Bislang haben wir davon noch keinen Gebrauch gemacht.



Datenänderungen (Tabellenkopf)

Ebenfalls neu ist die Möglichkeit, eine Liste der von uns veränderten Daten im Tabellenkopf aufzurufen. Bislang haben wir auch davon noch keinen Gebrauch gemacht.



Für statistische Zwecke sowie zur Wahrung unserer Rechtsverpflichtungen gegenüber den Datengebern ist es möglich, sich für einen vorgegebenen Zeitraum eine Liste aller Downloads, eine Übersicht über die Anzahl der Downloads pro Studie, eine Übersicht über die Anzahl der Downloads nach Studien und Nutzern, über den Verwendungszwecke der Downloads, die Anzahl der Downloads nach Themen sowie eine Gesamtübersicht anzeigen zu lassen.

Übersicht der registrierten Nutzer

Historische Statistik

www.gesis.org/histat/de/stats/display

gesis
histat: 5.417.073 Werte 261.801 Zeitreihen

Historische Statistik

Start Daten Über Galerie Admin

Benutzer Verwaltung | Datenänderungen | Datenänderung(Tabellenkopf) | Statistiken

Angemeldet als: Thomas Rahf Abmelden Warenkorb [2]

Statistiken

Zeitraum Von: Januar 2012 Bis: Oktober 2012

Übersicht der registrierten Nutzer

Anzeigen

417 Ergebnisse wurden gefunden

Nutzer ID	Nutzername	Name	Anschrift	Kontakt	Registriert am
2798					01.01.2012 11:53:53
2799					04.01.2012 17:19:13
2800					05.01.2012 15:16:55

Übersicht der einzelnen Downloads (Datentabelle)

Historische Statistik

www.gesis.org/histat/de/stats/display

gesis
histat: 5.417.073 Werte 261.801 Zeitreihen

Historische Statistik

Start Daten Über Galerie Admin

Benutzer Verwaltung | Datenänderungen | Datenänderung(Tabellenkopf) | Statistiken

Angemeldet als: Thomas Rahf Abmelden Warenkorb [2]

Statistiken

Zeitraum Von: Januar 2012 Bis: Oktober 2012

Übersicht der einzelnen Downloads (Datentabelle)

Anzeigen

2609 Ergebnisse wurden gefunden

ZA Nr.	Download Nr.	Studientitel	Zugangs-Klasse	Verwendungszweck	Zeitpunkt	Name
ZA 8177	10753	Ausgewählte Zeitreihen aus Studien zur Lohn- und Gehaltsentwicklung und der Einkommensentwicklung in Deutschland von 1850 bis 1985.		Lehre (Hochschulen)	01.01.2012 12:17:01	
ZA 8177	10754	Ausgewählte Zeitreihen aus Studien zur Lohn- und Gehaltsentwicklung und der Einkommensentwicklung in Deutschland von 1850 bis 1985.		Lehre (Hochschulen)	01.01.2012 13:17:32	
ZA 8177	10755	Ausgewählte Zeitreihen aus Studien zur Lohn- und Gehaltsentwicklung und der Einkommensentwicklung in Deutschland von 1850 bis 1985.		Lehre (Hochschulen)	01.01.2012 13:19:35	
ZA 8177	10756	Ausgewählte Zeitreihen aus Studien zur Lohn- und Gehaltsentwicklung und der Einkommensentwicklung in Deutschland von 1850 bis 1985.		Lehre (Hochschulen)	01.01.2012 13:20:25	

Übersicht zur Anzahl der Downloads (Datentabellen) nach Studien

Historische Statistik

gesis
histat: 5.417.073 Werte 261.801 Zeitreihen

Start Daten Über Galerie Admin

Benutzer Verwaltung | Datenänderungen | Datenänderung(Tabellenkopf) | Statistiken

Angemeldet als: Thomas Rahlf | Abmelden | Warenkorb (2)

Statistiken

Zeitraum Von: Bis:

← Januar 2012 → ← Oktober 2012 →

So	Mo	Di	Mi	Do	Fr	Sa	So	Mo	Di	Mi	Do	Fr	Sa
1	2	3	4	5	6	7	30	1	2	3	4	5	6
8	9	10	11	12	13	14	7	8	9	10	11	12	13
15	16	17	18	19	20	21	14	15	16	17	18	19	20
22	23	24	25	26	27	28	21	22	23	24	25	26	27
29	30	31	1	2	3	4	28	29	30	31	1	2	3

Übersicht zur Anzahl der Downloads (Datentabellen) nach Studien

Anzeigen

265 Ergebnisse wurden gefunden

ZA Nr.	Studientitel	Anzahl der Downloads	Anzahl der Zeitreihen	Anzahl Aufrufe mit Downloads
ZA 8186	50 Jahre Deutsche Mark. Monetäre Statistiken von 1948 bis 1997.	162	7522	11
ZA 8450	Industrielle Beziehungen, Bundesrepublik Deutschland 1950 - 1999.	156	2636	3
ZA 8204	Regionale Differenzierung in der ökonomischen Entwicklung Deutschlands im 19. Jahrhundert.	77	3080	5
ZA 8235	Grunddaten der Bevölkerungsstatistik Deutschlands von 1871 bis 1939.	72	4768	28

Anzahl der Downloads (Datentabellen) nach Studien und Nutzern

Historische Statistik

gesis
histat: 5.417.073 Werte 261.801 Zeitreihen

Start Daten Über Galerie Admin

Benutzer Verwaltung | Datenänderungen | Datenänderung(Tabellenkopf) | Statistiken

Angemeldet als: Thomas Rahlf | Abmelden | Warenkorb (2)

Statistiken

Zeitraum Von: Bis:

← Januar 2012 → ← Oktober 2012 →

So	Mo	Di	Mi	Do	Fr	Sa	So	Mo	Di	Mi	Do	Fr	Sa
1	2	3	4	5	6	7	30	1	2	3	4	5	6
8	9	10	11	12	13	14	7	8	9	10	11	12	13
15	16	17	18	19	20	21	14	15	16	17	18	19	20
22	23	24	25	26	27	28	21	22	23	24	25	26	27
29	30	31	1	2	3	4	28	29	30	31	1	2	3

Anzahl der Downloads (Datentabellen) nach Studien und Nutzern

Anzeigen

2603 Ergebnisse wurden gefunden

ZA Nr.	Studientitel	Datum des Downloads	Klasse	Anzahl der Downloads	Nutzer ID	Name, Vorname	Anschrift	E-Mail
ZA 8177	Ausgewählte Zeitreihen aus Studien zur Lohn- und Gehaltsentwicklung und der Einkommensentwicklung in Deutschland von 1850 bis 1985.	01.01.2012		1	2798	[REDACTED]	[REDACTED]	[REDACTED]
ZA 8177	Ausgewählte Zeitreihen aus Studien zur Lohn- und Gehaltsentwicklung und der Einkommensentwicklung in Deutschland von 1850 bis 1985.	01.01.2012		1	2798	[REDACTED]	[REDACTED]	[REDACTED]
ZA 8177	Ausgewählte Zeitreihen aus Studien zur Lohn- und Gehaltsentwicklung und der Einkommensentwicklung in Deutschland von 1850 bis 1985.	01.01.2012		1	2798	[REDACTED]	[REDACTED]	[REDACTED]

Verwendungszweck der Downloads

The screenshot shows the GESIS histat website interface. At the top, it displays '5.417.073 Werte' and '261.801 Zeitreihen'. The main content area is titled 'Statistiken' and shows a calendar for the period from January to October 2012. Below the calendar, there is a dropdown menu for 'Verwendungszweck der Downloads' and an 'Anzeigen' button. The results section shows 33 findings.

Angebener Verwendungszweck	Anzahl der Downloads	Anzahl verschiedener Studien
Lehre (Hochschulen)	1913	232
Abschlussarbeit: erster Studienabschluss, z.B. Master, Bachelor, Diplom, etc	151	52
Forschungsprojekt/-arbeit	128	67
Forschungsprojekt	57	7
ohne Angabe	52	4
Lehre	50	16

Übersicht der Downloads nach Themen

The screenshot shows the GESIS histat website interface, similar to the previous one. The main content area is titled 'Statistiken' and shows a calendar for the period from January to October 2012. Below the calendar, there is a dropdown menu for 'Übersicht der Downloads nach Themen' and an 'Anzeigen' button. The results section shows 26 findings.

Themen	Anzahl Downloads insgesamt	Anzahl der Studien aus denen Downloads erfolgten	Anzahl verschiedener Studien, aus denen Downloads erfolgten
Geld	393	88	20
Arbeit	354	69	21
Bevölkerung	325	127	21
Bildung	203	95	19
Staatsfinanzen	196	88	31
Konjunktur	154	63	23

Gesamtübersicht

Historische Statistik

gesis
histat: Historische Statistik **5.417.073** Werte **261.801** Zeitreihen

Start Daten Über Galerie Admin

Benutzer Verwaltung | Datenänderungen | Datenänderung(Tabellenkopf) | Statistiken

Angemeldet als: Thomas Rahlf | Abmelden | Warenkorb [2]

Suchbegriff

Statistiken

Zeitraum

Von:

Januar	2012					
So	Mo	Di	Mi	Do	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

 Bis:

Oktober	2012					
So	Mo	Di	Mi	Do	Fr	Sa
30	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3

Gesamtübersicht: Registrierungen, Anmeldungen und Downloads

Anzeigen

Nutzer	Erläuterung	Anzahl
Registrierung	Registrierungen insgesamt (Zuteilung eines Passwortes).	417
Nutzer aus Forschungseinrichtungen	Registrierungen mit Angabe einer Institution im Adressfeld.	163
Nutzer mit Angabe einer Privatadresse	Registrierungen ohne Angabe einer Institution im Adressfeld.	254

Aufrufe	Erläuterung	Anzahl
---------	-------------	--------

3.6 Exportfunktion

Neben der Download-Möglichkeit für die Daten im Excel- oder CSV-Format können mit der neuen Version die Studiendetails erstmals in einem strukturierten Format heruntergeladen werden, dem DDI-Format der Data Documentation Initiative. Das Ergebnis ist eine XML-Datei der folgenden Form (Auszug):

```

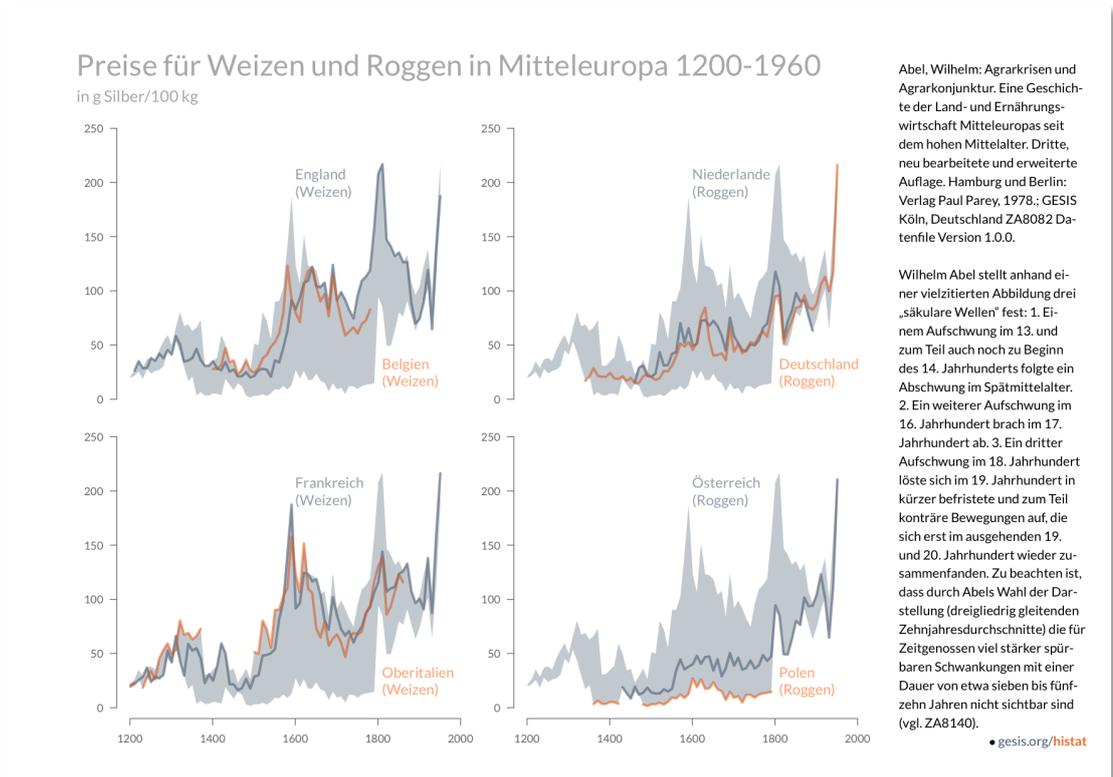
1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <!-- The original DDI 3.1 Schema can be found at http://www.ddialliance.org/DDI/schema/ddi3.1/ -->
3 <!-- from Histat 2.0 on 2012-10-23 -->
4 <!-- DDI3.1 export format -->
5 <!-- Histat 2.0 to DDI 3.1 Export -->
6
7 <!--
8
9
10 <ddi:DDIInstance id="gesis_ZA8142" agency="de.gesis" version="1.0.0" versionDate="2012-08-20" xmlns:ddi="ddi:instance:3_1" xmlns:s="ddi:studyunit:3_1" xmlns:
11 <!--
12 <!--
13 <!--
14 <!--
15 <!--
16 <!--
17 <!--
18 <!--
19 <!--
20 <!--
21 <!--
22 <!--
23 <!--
24 <!--
25 <!--
26 <!--
27 <!--
28 <!--
29 <!--
30 <!--
31 <!--
32 <!--
33 <!--
34 <!--
35 <!--
36 <!--
37 <!--

```

XML | Unicode (UTF-8) | Unix (LF) | 25.629 / 2.464 / 388

3.7 Die Galerie

Aktuell befindet sich eine Galeriefunktion in der Entwicklung. Geplant ist, dass alle Studien durch eine einseitige Abbildung sowie einen kurzen Text charakterisiert werden sollen, um Ihnen einen ersten Eindruck von der Studie zu ermöglichen. Bis zur Fertigstellung präsentieren wir Ihnen an dieser Stelle zur ersten Orientierung ausgewählte Beispiele in statistischer Form.



4 Vorgeschichte

4.1 Bedeutung historischer Statistikdaten

Historische statistische Daten haben weltweit einen hohen medialen und wissenschaftlichen Gebrauchswert. Hierzu wenige Beispiele: GAPMINDER, das inzwischen von Google betriebene, weltweit populäre interaktive Statistikangebot von Hans Rosling ist eine Visualisierung historischer statistischer Daten ab 1800. Mehrere Ökonomie-Nobelpreise wurden an Wissenschaftler für ihre Arbeit mit empirischen (Zeitreihen-) Daten oder unmittelbar für ihre wirtschaftshistorischen Beiträge verliehen. Daron Acemoglu, einer der weltweit führenden Ökonomen, stützt sich bei seinen langfristigen Prognosen auf umfangreiche historische Statistiken. Für Deutschland hat der ehemalige Präsident des Wissenschaftszentrums Berlin für Sozialforschung (WZB), Jürgen Kocka, schon früh auf die Bedeutung historischer Statistikdaten hingewiesen. Auch im Umfeld der Wirtschaftswissenschaften finden sich historisch-statistisch orientierte Wissenschaftler, wie zum Beispiel der am Deutschen Institut für Wirtschaftsforschung (DIW Berlin) tätige Rainer Fremdling oder am CESifo Ludger Wöbmann..

4.2 Daten der Forschung

In Deutschland wurden in den vergangenen Jahrzehnten von verschiedenen Geldgebern umfangreiche Drittmittelprojekte zur Zusammenstellung und Aufbereitung historisch-statistischer Daten sowohl im Bereich der Wirtschafts- und Sozialgeschichte als auch der Sozialwissenschaften finanziert. Aus dem von 1977 bis 1981 erstmalig von der DFG geförderten Projektverbund „*Deutsche Bildungsstatistik 1800-1945*“, das auf breiter empirisch-statistischer Grundlage den langfristigen Strukturwandel des Bildungswesens in Deutschland rekonstruiert und analysiert, gingen zunächst zwei umfangreiche Datenhandbücher hervor. Die DFG hat aber noch in den vergangenen zehn Jahren, also rund dreißig Jahre nach der ersten Förderung, mindestens sechs Einzelanträge zu entsprechenden Datenhandbüchern bewilligt.

Seit 1981 hat ebenfalls die DFG ein Schwerpunktprogramm zum Thema „*Quellen und Forschungen zur Historischen Statistik*“ von Deutschland gefördert. Im insgesamt zehnjährigen Förderungszeitraum (1981-1991) fanden sich 20 Antragsteller mit nahezu 70 wissenschaftlichen Mitarbeitern und Hilfskräften zusammen, um in einer großen Anzahl von Einzelprojekten statistische Daten aufzubereiten. Als Ergebnis sind zwischen 1987 und 2001 insgesamt 26 voluminöse Tabellenbände erschienen.

Auch in den Sozialwissenschaften entstanden umfangreiche historische Datenbestände im Rahmen großer Verbundprojekte. Von der VolkswagenStiftung wurde von 1979 bis 1984 das Projekt „*Vergleichende Analysen der Sozialstruktur mit Massendaten*“ (VASMA) gefördert. Auf der Grundlage der amtlichen Statistik zu den Volks- und Berufszählungen sowie den Betriebs-, Gewerbe- und Arbeitsstättenzählungen des Deutschen Reiches und der BRD seit 1875 befasste sich das Projekt mit der EDV-Erschließung dieser Daten und einer Systematisierung der Berufs- und Arbeitsstättenstatistik.

Ebenfalls von der VolkswagenStiftung wurde bereits in den siebziger Jahren das Projekt „*Historical Indicators of the Western European Democracies*“ (HIWED) von Wolfgang Zapf und Peter Flora gefördert, aus dem das zweibändige Datenhandbuch *State, Economy, and Society in Western Europe 1815–1975*, erschienen 1983 und 1987, hervorging. Die Arbeiten wurden am Mannheimer Zentrum für Europäische Sozialforschung kontinuierlich fortgeführt. Das dort mit drei Wissenschaftlerstellen angesiedelte EURODATA Research Archive bemüht sich dabei systematisch um die Zusammenstellung relevanter Daten. Aus diesem Zusammenhang sind weitere statistische Datenhandbücher zur historischen Entwicklung in Deutschland und Europa hervorgegangen.

Unabhängig von den genannten großen Verbundprojekten sind in den vergangenen Jahren natürlich auch zahlreiche Einzelstudien entstanden, die jeweils spezielle Aspekte der Historischen Statistik zum Gegenstand hatten.

4.3 Einbindung der amtlichen Statistik

Anders als in anderen Ländern hat sich die amtliche Statistik in Deutschland über weite Strecken erstaunlich zurückhaltend zu einer Historischen Statistik positioniert. Bei den vorbereitenden Beratungen und Gesprächen, die 1980/81 zur endgültigen Konzipierung des DFG-Schwerpunkt-programms zur Historischen Statistik führten, waren das Statistische Bundesamt bzw. die Landesämter nicht beteiligt. Auch während der Laufzeit des Projektes beschränkte sich die Zusammenarbeit auf eine informelle Unterstützung. Erst durch das persönliche Interesse des 1983 zum Präsidenten des Statistischen Bundesamtes ernannten Egon Hölder änderte sich das Verhältnis grundlegend. Egon Hölder war von der Bedeutung einer historischen Komponente in der amtlichen Statistik überzeugt und beteiligte sich aktiv an deren Weiterentwicklung. Eine 1989 organisierte Konferenz „Historische Statistik in der Bundesrepublik Deutschland“, die Wissenschaftlerinnen und Wissenschaftler insbesondere aus dem Schwerpunktprogramm und Vertreterinnen und Vertreter der amtlichen Statistik zusammenbrachte, fand unter der Federführung des Statistischen Bundesamts und dem Zentralinstitut für sozialwissenschaftliche Forschung an der Freien Universität Berlin in Wiesbaden statt.

Zwei große Handlungsnotwendigkeiten zeichneten sich ab: Zum einen wurde mit zunehmender Datenproduktion der einzelnen Schwerpunkt-Teilprojekte die Notwendigkeit einer zentralen, übergreifenden Datenbank immer dringender, zum anderen wollte man im Statistischen Bundesamt das lange vergriffene, 1972 anlässlich des hundertjährigen „Geburtstages“ der zentralen amtlichen Statistik erschienene Kompendium „Bevölkerung und Wirtschaft“ überarbeiten und neu auflegen. Es lag nahe, hierzu insbesondere für die früheren Jahre auf die Ergebnisse des Schwerpunktprogramms zurückzugreifen.

Unmittelbar nach der Tagung wurde aus diesen beiden Gründen eine „Arbeitsgemeinschaft Historische Statistik“ mit dem Ziel gegründet, die Forschung und Publikation der Historischen Statistik in Deutschland zu fördern. Über den spontanen Zusammenschluss hinaus wurde auch ein Antrag auf institutionelle Finanzierung an das damalige Bundesministerium für Forschung und Technologie gerichtet. Das Ministerium hatte seine Bereitschaft signalisiert, einen „For-

schungsverband Historische Statistik" mit ca. 8 bis 10 Millionen DM für eine Laufzeit von rund 8 Jahren zu fördern, womit eine Anfangsfinanzierung für eine langfristige Installierung eines Forschungsbereiches Historische Statistik an drei bis vier Instituten geschaffen werden sollte. Eine Neugewichtung der Förderungsprioritäten im Zusammenhang mit der deutschen Einigung hatte dann eine Zurückstellung des Antrags zur Folge. Damit kamen die weiteren Arbeiten zunächst zum Erliegen.

4.4 Fortführung der Arbeiten

Die größte inhaltliche Nähe sowie eine ausreichende institutionelle Basis wies zu dieser Zeit die organisatorische Einheit bei GESIS auf, aus der das Team DHS hervorgegangen ist. Nicht zuletzt aufgrund einer deutlichen personellen Aufstockung in Folge der Auflösung der Akademie der Wissenschaften der DDR konnte hier ein Arbeitsschwerpunkt „Historische Statistik“ etabliert werden, der sich zunächst der ersten Handlungsnotwendigkeit, dem Aufbau einer zentralen Datenbank widmete. Zunächst wurde ein historisch-geografisches Informationssystem in Angriff genommen, in dem nicht nur die historischen Statistikdaten, sondern auch ihre geografischen Bezugsgrößen digital erfasst und zugänglich gemacht werden sollten. Wenig später entwickelte sich ergänzend der Gedanke, eine Datenbank speziell für historische *Zeitreihen* anzubieten. Ausgangspunkt hierfür war das BMBF-Projekt zur Deutschen Innovationsgeschichte, bei dem in GESIS archivierte Daten den im Projekt beteiligten Forschern in einer online Datenbank zur Verfügung gestellt werden sollten. Mit Mitteln des BMBF wurde damit eine erste Version von histat entwickelt, die es nun auch ermöglichte neue, nicht auf die Innovationsstatistik bezogene Studien in die Datenbank aufzunehmen.

Die Arbeiten an dem geografisch-historischen Informationssystem konnten bei GESIS nicht fortgeführt werden. Die Daten wurden daher dem Koordinator des DFG-Schwerpunktprogramms zur Historischen Statistik übergeben. Nach weiteren Vorarbeiten konnte dagegen 2004, also fast fünfzehn Jahre nach der ursprünglichen Anforderung, bei GESIS eine Datenbank zur Historischen Statistik von Deutschland online gestellt werden. Das Ergebnis entsprach dabei weitgehend der ursprünglichen Intention, ging jedoch auch weit darüber hinaus. Der Zugriff auf die Datenbestände sollte auf der einen Seite möglichst einfach erfolgen, auf der anderen Seite wissenschaftlichen Ansprüchen genügen und die angebotenen Zeitreihen in ihrem Entstehungskontext hinreichend dokumentieren. Auf einer vom Vorläufer des Teams DHS gemeinsam mit dem Statistischen Bundesamt 2004 bei der GESIS veranstalteten Tagung konnte eine erste Version präsentiert werden. Aus dieser lange entwickelten Idee ist ein mittlerweile erfolgreiches Produkt geworden.

5 Zeitreihen in den Wirtschafts- und Sozialwissenschaften

5.1 Was sind Zeitreihen?

Als Zeitreihe bezeichnet man die chronologisch geordnete Folge von Messwerten einer Variablen. Die Werte einer Zeitreihe entstehen aus der wiederholten Messung einer ganz bestimmten Größe über die Zeit. Im Idealfall sind die Zeitabstände der Messung konstant (äquidistant). Die Zeitpunkte der Messung sind also zeitlich immer gleich weit voneinander entfernt. Der Abstand kann eine Sekunde, einen Tag, einen Monat oder ein Jahr betragen, um nur einige Beispiele zu nennen. Bei den Werten einer Zeitreihe unterscheidet man zwischen Bestandsgrößen und Stromgrößen. Letztere entstehen durch Aggregation von (Einzel)Werten über einen bestimmten Zeitraum. So ergibt sich der Wochenumsatz eines Ladengeschäfts aus der Summation der Tagesumsätze. Die Werte der Reihe sind also zeitraumbezogen. Im Unterscheid dazu messen Bestandsgrößen die jeweilige Größe zu einem bestimmten Zeitpunkt. Die Einwohnerzahl einer Stadt z.B. ist eine Bestandsgröße, deren Wert sich nur für einen bestimmten Zeitpunkt sinnvoll angeben lässt. Der Wert einer Bestandsgröße zum Zeitpunkt $\{t\}$ ergibt sich aus dem Wert der Bestandsgröße zum Zeitpunkt $\{t-1\}$ plus den Zugängen minus den Abgängen zwischen den beiden Zeitpunkten. Das Besondere an Zeitreihenwerten, im Unterschied zu anderen Messwerten, wie z.B. Querschnittsdaten, wie sie aus Umfragen resultieren, ist demnach, dass sie die Dimension „Zeit“ als stringentes Ordnungskriterium enthalten.

Zeitreihen gibt es in allen Bereichen, in denen Messungen über die Zeit vorgenommen werden können. In den Naturwissenschaften, der Psychologie, Medizin, Umwelt und natürlich auch im sozialen und ökonomischen Bereich. Es kann sich bei einer Zeitreihe z.B. um die Herzfrequenz eines Patienten handeln, es kann sich um den Wasserstand des Rheins bei Köln handeln, der mehrmals täglich zu festgelegten Zeiten vom Wasserwirtschaftsamt gemessen wird, oder, um ein letztes Beispiel zu nennen, es kann sich um die Anzahl der Arbeitslosen in Deutschland entweder zum Monatsende oder im Durchschnitt des Monats handeln. Je nachdem wie lange und wie häufig diese Messung durchgeführt wird, hat man Zeitreihen ganz unterschiedlicher Länge und Dichte. Es gibt Zeitreihen für die Dauer von nur wenigen Minuten mit Messwerten pro Sekunde, z.B. in der Medizin, oder aber für die Dauer von vielen Jahrhunderten mit Messwerten pro Jahr, z.B. in den Geowissenschaften.

5.2 Zeitreihen und Zeit

Die Zeit spielt in den Wirtschafts- und Sozialwissenschaften alleine schon deshalb eine zentrale Rolle, weil menschliches Handeln grundsätzlich in Raum und Zeit stattfindet. Die Zeit ist damit eine Grundkategorie unserer Erfahrung. Sie lässt sich in Vergangenheit, Gegenwart und Zukunft unterteilen. Obwohl damit alle realen Daten, mit denen sich die Wirtschafts- und Sozialwissenschaften und auch die Geschichtswissenschaften befassen, Daten sind, die einem bestimmten historischen Zeitpunkt oder Zeitraum zugeordnet werden können, sind nicht alle diese Daten automatisch auch Zeitreihen. Erst wenn sich die Daten auf ein einheitliches Phä-

nomen beziehen, und für viele aufeinanderfolgende Zeitpunkte (Zeiträume) in gleichen Zeitabständen gemessen werden, lassen sich durch die Anordnung dieser Daten nach der Zeit Zeitreihen bilden.

Zeitreihen, als Messwerte einer Variablen zu historischen und gleichabständigen Zeitpunkten, verleihen also der historischen Zeit eine quantitative Dimension und machen sie damit statistischen Messverfahren und analytischen Theorien zugänglich. Zeitreihen verleihen der historischen Zeit aber nicht nur eine quantitative Dimension, sie ordnen auch jeden Wert der Reihe in den Ablauf dieser historischen Zeit ein. Mit Ausnahme des ersten und letzten Wertes einer Reihe haben Zeitreihenwerte immer ein Vorher und ein Nachher. In diesem Vorher spiegelt sich die Historizität eines bestimmten Wertes wider, welche die Möglichkeit eröffnet, die Frage nach der Kausalität in der Geschichte auch formal quantitativ zu untersuchen. Das Nachher informiert uns darüber, wie sich der Wert, den wir für einen bestimmten Zeitpunkt gemessen haben, weiter entwickelt hat. Ist dieser Zeitpunkt die aktuelle Gegenwart, wissen wir nur etwas über seine Vergangenheit, nicht aber etwas über seine Zukunft.

Betrachtet man lediglich eine Zeitreihe, dann ist die Form der zeitlichen Dynamik, die im Verlauf der Reihe und in den Veränderungen der Werte zum Ausdruck kommt, von besonderem Interesse. Diese Dynamik kann sich ganz unterschiedlich gestalten. Ein dauerhaftes „größer“ oder „kleiner“ der Werte deutet auf Trendentwicklungen hin, die über längere Zeit einer bestimmten Richtung folgen, wobei sich diese im Betrachtungszeitraum natürlich auch wieder ändern kann. Das Ausmaß der Veränderung der Werte signalisiert das Tempo der Entwicklung, das manchmal schneller und manchmal langsamer verläuft. Insgesamt vermitteln diese Trendentwicklungen ein Bild von Kontinuität und Wandel.

Der Verlauf einer Zeitreihe zeigt aber nicht nur bestimmte Entwicklungsrichtungen, er zeigt auch Muster im Auf und Ab der Werte, die sich zu wiederholen scheinen, zwar nicht in exakter Weise, aber doch in mehr oder weniger ähnlicher Form. Geschichte scheint sich hier zu wiederholen, und das Auf und Ab einer generellen Erklärung zugänglich. Zahlreiche Zyklentheorien mit ganz unterschiedlichen Periodendauern und Erklärungen sind aus dieser Beobachtung heraus entstanden. Bei unterjährigen Reihen sorgt die Jahreszeit für solche mehr oder weniger regelmäßigen Schwankungen, weshalb man bei diesen Schwankungen auch von Saison spricht. Interessant ist, dass bei diesen Formen von Entwicklung, die man als Kontinuität und Regelmäßigkeit wahrnimmt, die Vergangenheit notwendigerweise immer einen Einfluss auf die jeweilige Gegenwart hat, da es sonst weder Kontinuität noch Regelmäßigkeit geben würde.

Neben diesen Formen von Kontinuität und Regelmäßigkeit kann es auch extreme und erratische Veränderungen geben, die selten auftreten und entweder dauerhafte oder nur vorübergehende Auswirkungen haben können. Daneben sind auch Entwicklungen vorstellbar, für die wir weder Kontinuität noch Regelmäßigkeit feststellen können. Die Entwicklung erscheint dann als rein zufällig. Sie lässt sich weder systematisch erklären, noch lassen sich aus der Betrachtung der Vergangenheit Aussagen über Gegenwart und Zukunft ableiten.

Die Realität wird immer eine Mischung aus Kontinuität und Regelmäßigkeit auf der einen Seite und Diskontinuität und Irregularität auf der anderen Seite darstellen. Das Interesse bei der Analyse von Zeitreihen gilt aber nicht nur diesen verschiedenen Entwicklungsformen und ihren

Wechselwirkungen, sondern auch den Ursachen dieser Entwicklungen. Dabei geht es nicht nur um die Frage nach der Wirkung einzelner Ereignisse für eine bestimmte Entwicklung, sondern auch und vor allem um den generellen und systematischen Zusammenhang verschiedener Entwicklungen (mehrerer Zeitreihen), m.a.W., es geht um die Frage der Kausalität. Wie sich diese gestaltet, wie sie sich darstellen, modellieren, identifizieren und schätzen lässt, ist Gegenstand einer äußerst umfangreichen Literatur und keinesfalls abschließend geklärt.

Abschließend sei betont, dass alle empirischen Zeitreihen und damit unser Wissen um die historische Zeit endlich sind. Zeitreihen greifen aus dem theoretisch unendlichen Zeitstrahl immer nur einen bestimmten Zeitabschnitt, eine Epoche heraus. Alle Interpretationen und Schlussfolgerungen, die aus der Analyse von Zeitreihen gewonnen werden, unterliegen also genau jener Perspektive, jenem Zeitfenster, die die jeweilige Reihe abdeckt. Da die Zeit davor und die Zeit danach im Dunkeln bleibt, besteht immer die Gefahr der perspektivischen Verzerrung. Das bedeutet, dass sich die aus Zeitreihen abgeleiteten Ergebnisse verändern können und in der Regel auch verändern werden, wenn man den Betrachtungszeitraum z.B. durch eine längere Zeitreihe variiert. Dies ist bei der Betrachtung der Gegenwart sowohl gegenüber ihrer Vergangenheit als auch gegenüber ihrer Zukunft zu berücksichtigen. Einschätzungen vergangener Perioden können sich ändern, wenn man deren Vorgeschichte durch längere Zeitreihen in den Blick nehmen kann und Einschätzungen der Gegenwart können sich ändern, wenn die Zeit in die Zukunft fortschreitet und die bisherige Gegenwart zu Geschichte wird. In kaum einem anderen Bereich wird dies augenfälliger als bei Prognosen der Zukunft mit Hilfe von Zeitreihen.

5.3 Erkenntnisziele der Zeitreihenanalyse

Zeitreihen bieten die Möglichkeit, historische Entwicklungen mit statistischen Verfahren hinsichtlich ihres Verlaufs und ihrer kausalen Ursachen zu untersuchen. Sie sind damit für alle jene Wissenschaften eine wichtige Grundlage, die sich für die Beschreibung und Erklärung von Entwicklungen in der historischen Zeit befassen. In den Wirtschafts- und Sozialwissenschaften geschieht dies:

- aus historischem Interesse. Es geht darum, historische Prozesse zu beschreiben, zu deuten und zu interpretieren. Dabei werden Zustände und daraus resultierende Strukturen sowie deren zeitliche Veränderungen in der Vergangenheit quantitativ erfasst, beschrieben und gedeutet,
- um die Gegenwart in den Gang der historischen Entwicklung vergleichend einzuordnen. Häufig wird dies mit der Frage verbunden, in welcher Weise vergangene Entwicklungen auf die Gegenwart eingewirkt haben und noch einwirken. Damit ist auch die Frage nach Kontinuität und Diskontinuität im geschichtlichen Ablauf verbunden (Die Gegenwart als Ergebnis der Vergangenheit),
- um Aussagen über Form, Ausmaß und Bedeutung kurz-, mittel- und langfristiger Entwicklungen und deren Dynamik in Wirtschaft und Gesellschaft abzuleiten,
- um historische Entwicklungen zu erklären, in dem man versucht, Einflussfaktoren und Interaktionsmuster zwischen verschiedenen Entwicklungsprozessen zu identifizieren (historische Gesetzmäßigkeiten). Bei den Einflussfaktoren kann es sich z.B. auch um po-

litische Entscheidungen handeln, von denen man annimmt, dass sie eine nachhaltige Wirkung auf die weitere Entwicklung gehabt haben (policy analysis),

- um Theorien und Hypothesen zu testen, die generelle / gesetzmäßige Aussagen über ökonomische und soziale Entwicklungen machen (Ökonometrie), und schließlich
- um die Zukunft vorherzusagen (Prognose).

Bei diesen Erkenntniszielen spannt sich ein weiter Bogen von der historischen Deskription, über quantitative Analysen und analytische Theorien bis hin zur Vorhersage der Zukunft. Während sich das historisch deskriptive Interesse der Rekonstruktion und der hermeneutisch deutenden Interpretation der seriellen Vergangenheit widmet, ist das analytische Interesse primär an der Entdeckung von allgemeinen Entwicklungsmustern, an der Identifikation von historischen Gesetzmäßigkeiten sowie am empirischen Gehalt analytischer Theorien interessiert. Welche Zeitperioden und welche und wie viele Zeitreihen dabei mit welchen Methoden analysiert werden, hängt von den jeweiligen Forschungsfragen und vom Analysedesign ab.

5.4 Statistische Methoden zur Analyse von Zeitreihen

Als Methoden der Zeitreihenanalyse bezeichnet man die statistischen Verfahren, die zur Beschreibung, Erklärung und Prognose von Zeitreihen verwendet werden. Die statistischen Methoden der Zeitreihenanalyse sind dabei genauso differenziert und vielschichtig, wie die oben dargestellten Erkenntnisinteressen, die die Ökonomie, die Soziologie oder die Geschichtswissenschaft mit der Analyse von Zeitreihen verbindet. Grundsätzlich lassen sich zwei Ansätze unterscheiden und zwar zum einen der theoriebasierte und zum anderen der datenbasierte Ansatz. Beide Ansätze blicken auf eine lange Entwicklung zurück, die in engem Zusammenhang mit den Problemen und Forschungsinteressen der Volkswirtschaftslehre, der Entwicklung der Statistischen Methodenlehre sowie der Verfügbarkeit von Zeitreihen steht.

Ausgangspunkt der theoriebasierten Ansätze ist eine bestimmte ökonomische oder soziologische Theorie, die nomologische (gesetzmäßige) Aussagen über die zeitliche Entwicklung bestimmter ökonomischer oder sozialer Phänomene macht. Die Aufgabe besteht hier darin, für die entsprechende Theorie ein Modell zu spezifizieren, das aus einer bestimmten Anzahl von Gleichungen besteht, welche einen Funktionalzusammenhang der durch die Theorie festgelegten Variablen spezifizieren. In einem nächsten Schritt gilt es, für die Variablen des Modells die entsprechenden Zeitreihen auszuwählen (Indikatorproblem) sowie den Schätzzeitraum festzulegen, um dann mit Hilfe geeigneter Verfahren die Parameter des Modells zu schätzen und im Hinblick auf ihre statistische Signifikanz zu überprüfen (zu testen). Das Ergebnis wird eine vorläufige Bestätigung oder eine Ablehnung der Theorie sein.

In der Ökonomik sind derartige theoriebasierte Ansätze in den 1920er und 1930er Jahren entstanden. Die Idee war, ökonomische Theorie, quantitative Daten und statistische Methoden systematisch zusammen zu führen und zwar unter dem Primat der Theorie. Damit wurde die Ökonometrie als völlig neues Forschungsparadigma begründet. Verbunden mit der Keynesianischen Lehre, die ja weitgehend makroökonomisch ausgerichtet war, hat dieser Ansatz einen ungeheuren Aufschwung erlebt und die empirische Wirtschaftsforschung bis in die 1960er

Jahre dominiert. Damit verbunden war das Interesse an (langen) Zeitreihen als Indikatoren theoretischer Makrovariablen. Diese Reihen wurden zunehmend im Rahmen der ebenfalls neu entstandenen Volkswirtschaftlichen Gesamtrechnung aufbereitet.

Ausgangspunkt für die datenbasierten Ansätze ist nicht eine bestimmte nomologische Theorie, sondern es sind die Daten/Zeitreihen, bzw. ihre Struktur, die man im Zusammenhang mit einer bestimmten Frage / einem Problem / einer Hypothese untersuchen möchte bzw. als relevant erachtet. Zum datenbasierten Ansatz kann man die Zeitreihenmodellierung, die explorative Datenanalyse, die Chart-Analyse und auch das sog. Data Mining rechnen.

Ihren eigentlichen Ausgangspunkt hatte diese Art der Zeitreihenanalyse in der empirischen Konjunkturforschung, die sich zu Beginn des 20. Jahrhunderts in vielen Industrieländern zu entwickeln begann und mit den daraus entstandenen Barometersystemen (eine Art klassifikatorischer Theorie) ihren Höhepunkt und Abschluss fand. Die dabei verwendeten Methoden dienten der Darstellung der verschiedenen Bewegungskomponenten der Reihen sowie des Zusammenhangs dieser Komponenten bei verschiedenen Reihen. Man hat sich vehement dagegen ausgesprochen, die empirische Konjunkturforschung auf eine bestimmte Theorie zu stützen, da man der Ansicht war, dass dies die zu beobachtende Realität einengen und verfälschen würde. Ein zentrales Problem war der Trend, der die konjunkturellen Schwankungen u.U. verdeckte und deshalb vor der eigentlichen Analyse der Konjunkturzyklen entfernt werden musste. Dazu wurden verschiedene Verfahren entwickelt. Am bedeutsamsten sind die, die den Trend als Funktion der Zeit darstellen (in dieser Zeit wurde auch die logistische Trendfunktion entwickelt) und die „Variate Difference“-Methode. Man kann diese Art der Zeitreihenanalyse als eine Vorstufe der modellbasierten Analyse bezeichnen, da sie Zeitreihen modellhaft aus dem Zusammenwirken verschiedener Einflusskomponenten zu beschreiben und zu erklären versuchte. Im Unterschied zur modernen Zeitreihenanalyse wurden die Komponenten, insbesondere der Trend, jedoch nicht als Realisationen stochastischer Prozesse aufgefasst. Heftige Kritik erfuhr diese Art der Zeitreihenanalyse mit dem Aufkommen der oben erwähnten Makroökonomie und zwar vor allem deshalb, weil sie nicht durch eine ökonomische Theorie fundiert war und daher eine „Measurement without Theory“ sei.

Im Zentrum der modernen datenbasierten Zeitreihenanalyse stehen die Zeitreihenmodellierung und die Vorstellung, dass Zeitreihen Realisationen stochastischer Prozesse sind. Die Theorie stochastischer Prozesse ist deshalb die Grundlage der modernen Zeitreihenanalyse. Das Ziel der Zeitreihenanalyse ist es, ein statistisches Modell zu entwickeln, das der stochastischen Struktur der Zeitreihe möglichst gut entspricht. Das statistische Modell soll die Eigenschaften des datengenerierenden Prozess, der in der Realität prinzipiell unbekannt ist, möglichst gut abbilden. Das statistische Modell ist eine Beschreibung stochastischer Größen und ihres funktionalen Zusammenhangs. Diese statistischen Modelle sind Grundlage für die Beschreibung der kurz- und langfristigen Dynamik und für die Zerlegung von Zeitreihen in spezifische Komponenten, wie Trend, Zyklus und Saison. Auch für die Analyse spezifischer Einflussgrößen, für die Zusammenhangs- bzw. Kausalanalyse und auch für Prognosen werden solche Modelle verwendet. Grundlegende Bedeutung haben in diesem Zusammenhang sog. Autoregressive Integrierte Moving Average (ARIMA-)Modelle, die von Box und Jenkins 1970 in die Diskussion eingeführt wurden.

Die Auswahl eines bestimmten Modells wird als das Grundproblem dieser Art von Zeitreihenanalyse angesehen. Das Problem besteht darin, aus der Vielzahl von verfügbaren Modellen, die i.d.R. ganz unterschiedliche Eigenschaften aufweisen, ein Modell auszuwählen, das sich für die Analyse als „geeignet“ erweist und die Zeitreihen, bzw. den Zusammenhang zwischen ihnen, „angemessen“ beschreibt. „Geeignet“ und „angemessen“ bezieht sich dabei sowohl auf die Struktur der Daten wie auch auf die Fragestellung der Analyse. Im Hinblick auf substanzwissenschaftliche Überlegungen ist eine solche Modellierung zunächst „theorielos“, da das statistische Modell allein aufgrund empirischer Kenngrößen und aufgrund der in den Daten enthaltenen Informationen spezifiziert wird. Damit unterscheidet sich diese Art grundlegend vom oben dargestellten theoriebasierten Ansatz, bei dem man versucht, substanzwissenschaftlich formulierte Modelle empirisch zu testen.

Neben dieser Zeitreihenmodellierung kommt auch deskriptiven Verfahren eine wichtige Funktion bei der datenbasierten Analyse zu. Dabei spielen graphische Visualisierungen der Zeitreihen, verschiedene Transformationsverfahren der Zeitreihenwerte (z.B. Logarithmen) sowie elementare Kennziffern wie Mittelwert, Varianz, Kovarianz und Kreuzkovarianz eine wichtige Rolle. Neben der Darstellung der Reihen im Zeitbereich ist auch deren Darstellung im Frequenzbereich interessant. Im Frequenzbereich sind Periodogramm- und Spektralanalyse wichtige Hilfsmittel insbesondere zum Aufspüren von Periodizitäten in Zeitreihen. Die Spektralanalyse wurde in den 1960er, Anfang der 1970er Jahre als völlig neue Erkenntnismethode in den Methodenkanon der Wirtschafts- und Sozialwissenschaften eingeführt.

Methodisch eng verwandt mit der Spektralanalyse und ebenfalls ein wichtiges Hilfsmittel im Rahmen der datenbasierten Analyse sind sog. Filter. Generell versteht man unter einem Filter einen Algorithmus zur Transformation einer Zeitreihe. Dabei werden die (vermuteten) Komponenten einer Reihe anhand ihrer Schwingungsdauer definiert und dann Filter verwendet, die diese Schwingungen aus der Zeitreihe möglichst exakt herausfiltern. Die Entwicklung solcher Filterverfahren hat ebenfalls eine lange Tradition und in den vergangenen Jahren zu sehr brauchbaren Resultaten geführt.

Nach unseren bisherigen Ausführungen scheinen sich theorie- und datenbasierte Ansätze als zwei unterschiedliche Analysestrategien gegenüber zu stehen. Die neuere Forschung versucht jedoch eine Synthese, in dem sie Modelle konstruiert, die theoretisch fundiert sind, und zwar mikroökonomisch, und die gleichzeitig der Tatsache Rechnung tragen, dass die verwendeten Zeitreihen als Realisationen stochastischer Prozesse aufgefasst werden müssen. Die wohl bedeutsamste Entwicklung in diesem Zusammenhang sind die Dynamischen Stochastischen Allgemeinen Gleichgewichtsmodelle (DSGE = Dynamic Stochastic General Equilibrium), die bei der Politikevaluation und auch bei Prognosen gegenwärtig eine große Rolle spielen.

Abschließend wird man sagen können, dass bei der statistischen Analyse von Zeitreihen sowohl datenbasierte wie auch theoriebasierte Verfahren eine wichtige Rolle spielen und sich gegenseitig ergänzen. Welche Verfahren letztendlich verwendet werden, hängt von der Fragestellung ab, die man mit den jeweiligen Zeitreihen verbindet und natürlich auch von der Länge und Qualität sowie von der Verfügbarkeit von Zeitreihen.

5.5 Wer produziert ökonomische und soziale Zeitreihen?

Zeitreihen resultieren aus der regelmäßigen, systematischen und fortlaufenden Beobachtung bzw. Messung eines bestimmten Phänomens. Sie verdanken demnach ihre Entstehung all jenen Akteuren bzw. Institutionen, die ein Interesse an solchen Daten, bzw. die Aufgabe haben, solche Daten zu erheben und der Öffentlichkeit zur Verfügung zu stellen. Da diese Interessen ganz unterschiedlich sind, gibt es Zeitreihen auch zu ganz unterschiedlichen Phänomenen. Dabei ist zu beachten, dass Zeitreihen in ganz unterschiedlicher Form vorliegen können. Sie können bereits aufbereitet und publiziert sein, sie können aber auch, quasi als Rohmaterial bzw. als historische Quelle in privaten oder öffentlichen Archiven „schlummern“. In der jüngsten Vergangenheit wurden immer mehr Zeitreihen online im Internet angeboten. Dabei kann es sich um Reihen handeln, die auch in gedruckter Form publiziert werden, oder aber auch um Reihen, die nur noch in elektronischer Form veröffentlicht werden, was für die langfristige Archivierung dieser Daten eine große Herausforderung darstellt.

Die amtliche Statistik

Der wichtigste Produzent von ökonomischen Zeitreihen war und ist der Staat, bzw. die Obrigkeit im Rahmen jener Tätigkeiten, die wir heute als amtliche Statistik kennen. Ihre Anfänge hatte diese Statistik in der historischen und empirischen Staatslehre, deren Interesse lange Zeit vornehmlich der Beschreibung von Zuständen galt. Bei der zahlenmäßigen Erfassung der Bevölkerung (Geburten, Sterbefälle, Heiraten) oder der Wirtschaft (Gewerbe und Handel) ging es nicht primär um die zeitliche Entwicklung dieser Variablen, sondern um ihre Größe, ihren Wert zu einem bestimmten Zeitpunkt. Da sich aus diesen Zustandsbeschreibungen, sofern sie fortlaufend und regelmäßig erfolgen, Zeitreihen bilden lassen, ist diese Art von Statistik eine wichtige Quelle für die Rekonstruktion historischer Zeitreihen. Im Unterschied zur heutigen amtlichen Statistik waren diese frühen Tätigkeiten nicht flächendeckend und kaum systematisch. Auch wurden diese Zahlen nicht immer veröffentlicht. Mit der Industrialisierung wurden diese Tätigkeiten zu Beginn des 19. Jahrhunderts zunehmend institutionalisiert und den neu gegründeten statistischen Ämtern übertragen, welche die Daten mehr und mehr regional übergreifend und nach einheitlichen Gesichtspunkten fortlaufend erhoben haben. Heute bestehen in Deutschland das Statistische Bundesamt, 16 Statistische Ämter der Länder und rund 100 selbständige Statistische Ämter in Städten und kommunalen Dienststellen, die ihren umfangreichen Bestand an Zeitreihen auch online anbieten.

Neben den statistischen Ämtern sind die Deutsche Bundesbank sowie die Bundesagentur für Arbeit weitere Anbieter amtlicher Zeitreihendaten. Darüber hinaus sind auch die Sozialversicherungsträger, wie z.B. die Deutsche Rentenversicherung Bund, als Produzenten und Anbieter von Zeitreihen zu nennen. Weiterhin gibt es zahlreiche überstaatliche Einrichtungen, die ebenfalls ein umfangreiches Angebot an Zeitreihen bereitstellen. Auf europäischer Ebene sind hier z.B. Eurostat und die Europäische Zentralbank, auf internationaler Ebene die OECD, die UNO, die Weltbank sowie der IWF zu nennen. Alle diese Institutionen erfüllen mit der Aufbereitung und dem Angebot von Zeitreihen einen gesetzlichen Auftrag.

Die Wissenschaft

Auch zahlreiche durch öffentliche Mittel finanzierte Wirtschaftsforschungsinstitute bieten einen umfangreichen Bestand ökonomischer Zeitreihen an. In Deutschland sind dies:

- Deutsches Institut für Wirtschaftsforschung (DIW Berlin)
- Hamburgisches Weltwirtschaftsinstitut (HWWI)
- ifo Institut für Wirtschaftsforschung, München
- Institut für Makroökonomie und Konjunkturforschung (IMK)
- Institut Arbeit und Qualifikation (IAQ) an der Universität Duisburg-Essen
- Institut für Weltwirtschaft (IfW), Kiel
- Institut für Wirtschaftsforschung Halle (IWH)
- Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Essen
- Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim

An wichtigen Wirtschaftsforschungsinstituten außerhalb Deutschlands seien genannt:

- Arbejderbevægelsens erhvervsråd, Kopenhagen
- Brookings Institution, Washington
- Center for Economic and Policy Research (CEPR), Washington
- DARES, Paris
- Economic Policy Institute (EPI), Washington
- Economic and Social Research Institute (ESRI), Dublin
- Elinkeinoelämän Tutkimuslaitos (ETLA), Helsinki
- Europäisches Gewerkschaftsinstitut, Brüssel
- Institut de Recherches Economiques et Sociales (IRES), Noisy-le-Grand, Frankreich
- Labour Institute for Economic Research, Helsinki
- National Bureau of Economic Research (NBER), Cambridge, Massachusetts
- National Institute of Economic and Social Research (NIESR), London
- Prometeia, Bologna
- Observatoire Francais de Conjonctures Economiques (OFCE), Paris
- Wifo, Wien

In den Sozialwissenschaften spielen Zeitreihen in der Sozialindikatorenforschung eine herausragende Rolle. Diese Forschung hat ihre Anfänge in den 20er und 30er Jahren des 20. Jahrhunderts in den USA. Ihr Ziel war es, Wohlfahrt und sozialen Wandel messbar zu machen. In den 1960er Jahre wurde dann die Sozialindikatorenbewegung initiiert. Grundlage war ein System sozialer Indikatoren, das ähnlich der Volkswirtschaftlichen Gesamtrechnung aufgebaut war und Auskunft geben sollte über den sozialen Wandel und die Wohlfahrt parallel zum wirtschaftlichen Wandel und wirtschaftlichen Wachstum. GESIS bietet mit SIMon (Social Indicators Monitor) einen integrierten und komfortablen Zugang zu Zeitreihendaten von zwei Indikatorensys-

temen, dem System Sozialer Indikatoren für Deutschland und dem European System of Social Indicators. Beide Indikatorensysteme sind Instrumente einer systematischen Gesellschaftsbeobachtung und bieten umfassende und detaillierte Informationen zur Entwicklung der deutschen und europäischen Gesellschaften der letzten Jahrzehnte.

Zeitreihen lassen sich auch aus Umfrage- bzw. Querschnittsdaten konstruieren. Voraussetzung dafür ist, dass die Befragungen regelmäßig durchgeführt werden und identische bzw. vergleichbare Fragen enthalten. Beispiel dafür sind ALLBUS, Politbarometer und Eurobarometer. Werden die Befragungen mit denselben Personen durchgeführt, spricht man von einem Panel. Das prominenteste Beispiel dafür ist das Sozioökonomische Panel (SOEP), das seit 1982 erhoben wird.

Die Revolution des Internet hat dazu geführt, dass heute alle wichtigen staatlichen und öffentlichen Einrichtungen, die mit Statistik befasst sind, ihre Zeitreihen online im Internet anbieten. Die Linksammlung des Zentrums für Europäische Wirtschaftsforschung (ZEW) nennt weitere nationale und internationale Einrichtungen, die ökonomische Zeitreihen online anbieten.

Weiterhin spielen auch jene Zeitreihen eine große Rolle, die im Rahmen von wissenschaftlichen Einzelstudien und Forschungsprojekten entstehen. In diesem Zusammenhang kommt den (sozialwissenschaftlichen) Datenarchiven eine große Bedeutung zu. Deren Aufgabe ist es, die Forschungsdaten der Primärforscher aufzubereiten, zu dokumentieren und langfristig zu archivieren, um sie damit für Sekundäranalysen zur Verfügung zu stellen. Für Deutschland nimmt GESIS diese Aufgabe wahr. Im Archivbestand finden sich zahlreiche Studien, die Zeitreihendaten enthalten. Seit 2004 werden darüber hinaus die Zeitreihen ausgewählter Studien der historischen und wirtschafts- und sozialwissenschaftlichen Forschung in der Online-Datenbank *histat* angeboten. Gegenwärtig umfasst der Bestand etwa 256.000 Zeitreihen und 5,2 Millionen Werte für die Zeit vom späten Mittelalter bis zur Gegenwart. Dieses Angebot ist einzigartig und wichtige Grundlage für ein breit angelegtes Kompendium zur Historischen Statistik Deutschlands.

Die private Wirtschaft

Schließlich sind auch private Unternehmen als Datenanbieter zu nennen. Insbesondere in den Forschungsabteilungen großer Finanzunternehmen besteht ein reicher Fundus an Zeitreihen, die in der Regel jedoch nicht öffentlich angeboten werden. Darüber hinaus gibt es privatwirtschaftliche Unternehmen, die Daten als Dienstleistungen für den Markt aufbereiten und anbieten (Thomson Reuters, Capital IQ, Standard & Poor's). Da es sich hier um privatwirtschaftliche Dienstleistungen handelt, ist die Verwendung dieser Daten mit Kosten verbunden.

5.6 Literatur

Metz, Rainer (2010): Zeitreihenanalyse. In: Wolf, Christof; Best, Henning (Hrsg.): Handbuch der sozialwissenschaftlichen Datenanalyse, Wiesbaden: VS Verlag für Sozialwiss., S. 1053-1090

Metz, Rainer (2008): Auf der Suche nach den Langen Wellen der Konjunktur. Stuttgart: Steiner.

Metz, Rainer (2002): Trend, Zyklus und Zufall. Bestimmungsgründe und Verlaufsformen langfristiger Wachstumsschwankungen (VSWG. Beiheft 165). Stuttgart.

Rahlf, Thomas (1998): Deskription und Inferenz: methodologische Konzepte in der Statistik und Ökonometrie (= Historical Social Research, Supplement No. 9).

<http://nbn-resolving.de/urn:nbn:de:0168-ssoar-286000>.

Metz, Rainer (1993): Probleme der statistischen Analyse langer historischer Zeitreihen, in: Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte (VSWG) 80, S. 457-486.

6 Inhalt und Nutzung

6.1 Zeitreihen- und Werteverteilung

Mit der Online-Freischaltung von histat 2004 waren in der Datenbank 144 Studien integriert. Jährlich kommen mehrere Dutzend hinzu, so dass Ende 2011 der Bestand auf 359 Studien angewachsen ist. Dem entspricht ein jährlicher Zuwachs von mehreren zehntausend Zeitreihen auf mittlerweile rund eine Viertelmillion.

Neue Studien	45	27	42	48	70	28 ⁵
Studien insgesamt ⁴	144	171	213	261	331	359
Neue Zeitreihen	17 926	14 566	16 516	39 462	37 537	20 044
Zeitreihen insgesamt ⁴	121 299	135 865	152 381	191 843	229 380	249 424

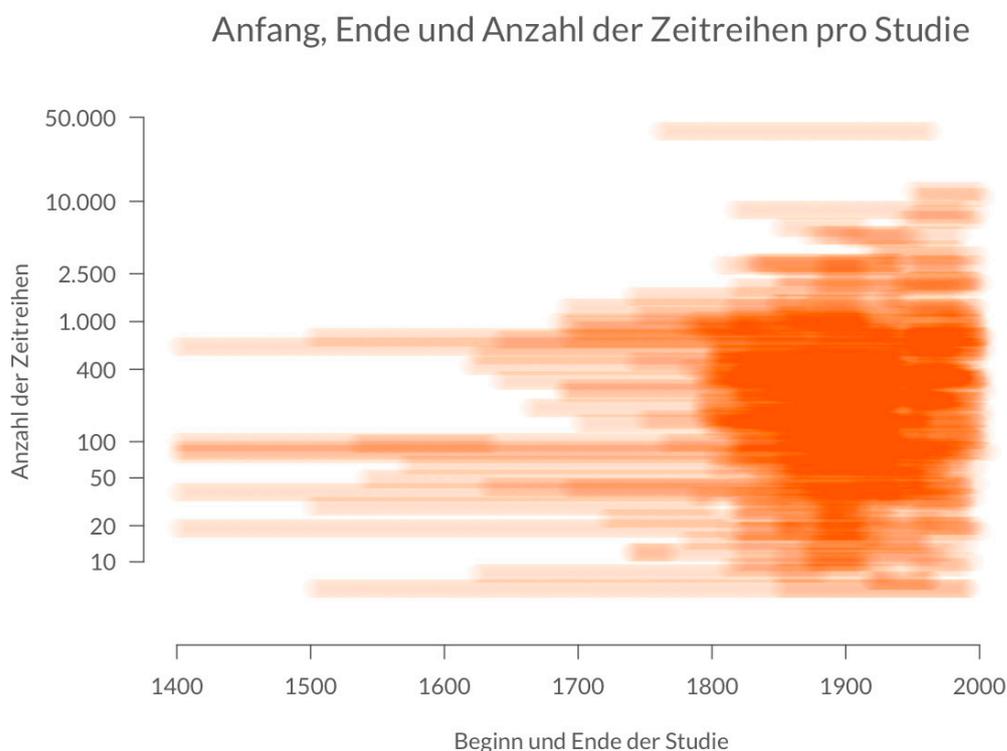
¹Seit 2010 Aufruf mit Ausnahme des downloads ohne Registrierung möglich – ²Angaben aus eTracker, 2010 aus Juli bis Dezember hochgerechnet. – ³Davon 1 982 Downloads durch zwei Nutzer/innen. – ⁴Kumulative Angaben, jeweils Jahresende. – ⁵2011 weniger Neuaufnahmen durch die Vorbereitung des vademecums.

Da histat von Beginn an als System für heterogene Datenbestände konzipiert war,¹ variiert sowohl die Anzahl der Zeitreihen pro Studie als auch die abgedeckte Zeitspanne beträchtlich. Rund zwanzig Prozent der Studien beinhalten bis zu 50 Zeitreihen, weitere 13 Prozent bis zu 100, 40 Prozent zwischen 101 und 500. Jeweils gut zehn Prozent beinhalten 501 bis 1000 und 1001 bis 5000 Zeitreihen. Neun Studien (3 Prozent) umfassen mehr als 5.000 Zeitreihen, die größten drei davon alleine insgesamt 61.274 Zeitreihen.

1-10	11-50	51-100	101-500	501-1.000	1001-5000	5.001-10.000	>10.000
3%	18%	13%	40%	12%	11%	2%	1%

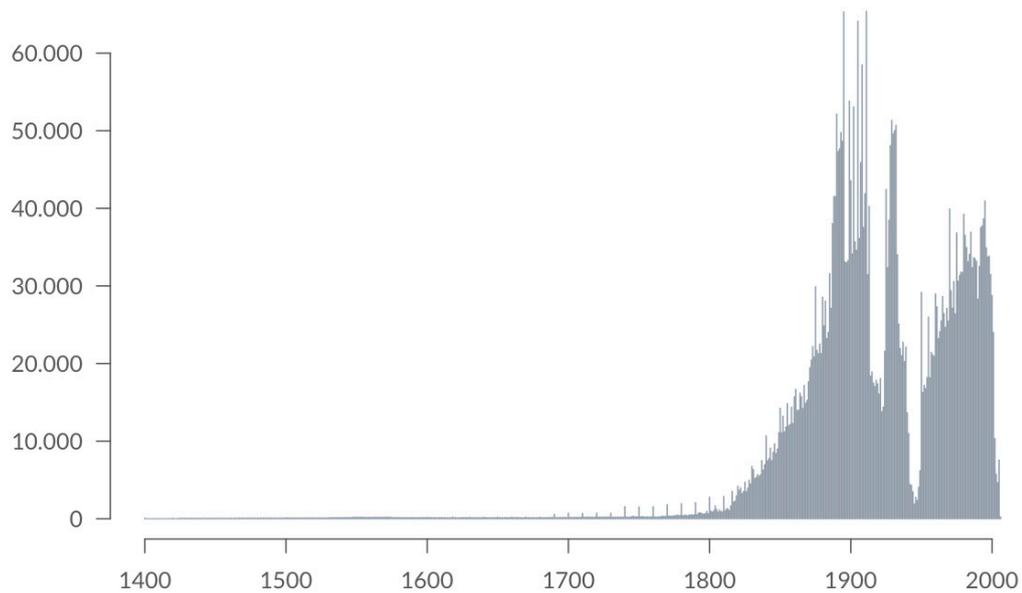
Die folgende Abbildung zeigt die Anzahl der Zeitreihen pro Studie (in logarithmischer Skalierung) sowie ihre Anfangs- und Endzeitpunkte. Man sieht, dass eine ganze Reihe von Studien auch sehr lange Zeiträume von mehreren hundert Jahren umfassen, auch solche mit relativ vielen Zeitreihen. Der Schwerpunkt (dunkle Färbung) liegt jedoch eindeutig in dem Bereich von etwa 1850 bis 1950 und zwischen 100 und 1.000 Zeitreihen. Hierbei ist zu berücksichtigen, dass diese Angaben sich jeweils auf den frühesten und spätesten Wert der in der Studie in mindestens einer Reihe vorhanden ist, beziehen: In einer Studie mit vielen Zeitreihen fangen in aller Regel weder alle Reihen zum gleichen Zeitpunkt an, noch hören sie im selben Jahr auf.

¹ Metz, Rainer / Sensch, Jürgen: HISTAT: Eine Internet-Datenbank der Historischen Statistik im ZHSF. In: Vera Ziegeldorf, Rüdiger Hohls, Daniel Burckhardt (Hrsg.): Geschichte und Neue Medien in Forschung, Archiven, Bibliotheken und Museen: Tagungsband .hist2003. Berlin, S. 211-230.



Betrachten wir daher schließlich den Inhalt der Datenbank auf *Werteebene*. Insgesamt umfassen alle in histat enthaltenen Zeitreihen rund fünf Millionen Werte. Ab 1800 liegen für nahezu jedes Jahr mindestens tausend Werte vor, ab Mitte des 19. Jahrhunderts sind es schon über zehntausend, mit Gründung des Deutschen Reiches über zwanzigtausend. Den Höhepunkt bilden die Jahre um den Beginn des 20. Jahrhunderts: hier stehen für einige Jahre über sechzigtausend Werte bereit. Die beiden Weltkriege führen zu Einbrüchen. Insbesondere für die vierziger Jahre des zwanzigsten Jahrhunderts sinkt die Anzahl der verfügbaren Werte auf unter zweitausend. Für die Nachkriegszeit bis zur Gegenwart liegen zwischen zwanzig- und dreißigtausend Werte vor. Im Vergleich dazu nimmt sich die sogenannte vorstatistische Zeit bescheiden aus. Doch auch hier können wir seit dem 16. Jahrhundert überwiegend mehr als hundert Werte pro Jahr anbieten, seit Mitte des 17. Jahrhunderts zumindest für jedes zehnte Jahr weit über tausend.

Anzahl Zeitreihenwerte in histat pro Jahr



6.2 Überblick über die enthaltenen Studien

Der folgende Überblick über die in histat enthaltenen Studien orientiert sich an der Publikationsform der Primärstudie, aus der Zeitreihen in histat übernommen wurden sowie am Entstehungskontext der Primärstudie. Zusätzlich wird ein Überblick über die in histat enthaltenen Studien vermittelt und abschließend das Projekt „vademecum zur Historischen Statistik von Deutschland“ kurz vorgestellt. Grundsätzlich gilt, dass jeder Studie in histat eine wissenschaftliche Veröffentlichung zugeordnet werden kann. Eine Ausnahme von dieser Regel stellen die sog. DHS-Datenkompilationen dar, da diese auf mehreren Veröffentlichungen basieren. Allerdings sind nicht für alle Studien die in ihr enthaltenen Zeitreihen vollständig in histat aufgenommen worden. Diesbezüglich sind mehrere Varianten zu unterscheiden. histat enthält zahlreiche Studien, die vom Primärforscher zur Archivierung und Langfristsicherung an GESIS übergeben wurden. Für diese Studien ist der gesamte Datenbestand der Studie im GESIS-Datenarchiv vorhanden. In histat wird aber u.U. nur eine Auswahl der Zeitreihen dieser Studie angeboten. Einen anderen Typus stellen jene Studien dar, die nachträglich von GESIS digitalisiert wurden. Auch hier gibt es zwei Varianten. Die eine davon repräsentiert den Gesamtbestand der publizierten Zeitreihen, die andere eine Auswahl.

Publikationsform der Primärstudien

In gedruckter Form werden Zeitreihen in wissenschaftlichen Aufsätzen, entweder in Fachzeitschriften oder Sammelwerken, in Monographien, in wissenschaftlichen Datenhandbüchern oder in Handbüchern/Publicationen der amtlichen Statistik publiziert. Immer häufiger werden Zeitreihendaten nicht mehr gedruckt, sondern in Form elektronischer Datenträger, meist in Verbindung mit einer gedruckten Monographie veröffentlicht.

Die folgende Übersicht zeigt die Anzahl der Publikationsformen der für histat verwendeten Primärstudien:

Aufsätze	84
Monographien/	228
davon Datenhandbücher	21
davon elektronisch (DVD)	2
Amtliche Statistik	15

Entstehungskontext

Wie oben in Kap. 4.5 bereits dargestellt, entstehen Zeitreihen in verschiedenen Kontexten. Wir hatten die amtliche Statistik, die Wissenschaft und die freie Wirtschaft als potentielle Datenproduzenten genannt. Da histat keine Studien enthält, die von kommerziellen Datenproduzenten stammen, sind alle Studien entweder der amtlichen Statistik oder der Wissenschaft zuzuordnen. Bei den in histat enthaltenen Zeitreihen handelt es sich deshalb entweder um amtliche Daten oder um Forschungsdaten. Zur amtlichen Statistik zählen die statistischen Ämter im engeren Sinne und jene öffentlichen Einrichtungen, die durch gesetzlichen Auftrag verpflichtet sind, Zeitreihendaten zu publizieren, z.B. die Deutsche Bundesbank. Diese (amtlichen) Studien dominieren das Angebot in histat zwar nicht in ihrer Anzahl, gleichwohl sind aber einige bedeutende Datensammlungen in histat vertreten, die eine gewisse Repräsentativität beanspruchen und überdies an keiner anderen Stelle online verfügbar sind.

Die folgende Liste nennt 5 Beispiele amtlicher Studien (Anzahl der Downloads von Datentabellen aus den Studien bis Juni 2012 in eckigen Klammern)

- Statistisches Reichsamt (Bearb.) (1925): Zahlen zur Geldentwertung in Deutschland 1914 bis 1923. Sonderhefte zu Wirtschaft und Statistik, 5. Jg., Sonderheft 1, Berlin. (GESIS Datenfile ZA8354) [55]
- Bundesministerium für das Post- und Fernmeldewesen, Hrsg., (1957): Zahlenspiegel der Deutschen Reichspost (1871 bis 1945). 2. erw. u. vervollständigte Auflage. Bonn. (GESIS Datenfile ZA8530) [12]
- Deutsche Bundesbank, Hrsg., (1976): Deutsches Geld- und Bankwesen in Zahlen 1876 - 1975. Frankfurt/M. (GESIS Datenfile ZA8222) [179]
- Deutsche Bundesbank, Hrsg., (1998): 50 Jahre Deutsche Mark. Monetäre Statistiken 1948-1997 auf CD-ROM. München. (GESIS Datenfile ZA8186) [379]

- Statistisches Bundesamt, Hrsg., (2008): Preisentwicklung nach den Volkswirtschaftlichen Gesamtrechnungen, Bundesrepublik Deutschland 1950 bis 2000. Wiesbaden. (GESIS Datenfile ZA8293) [15]

Der größte Teil der in histat enthaltenen Studien entstammt der wirtschafts- und sozialwissenschaftlichen sowie der wirtschafts- und sozialgeschichtlichen Forschung incl. der empirischen Wirtschaftsforschung. Eine genaue Aufteilung der Studien auf einzelne Disziplinen ist nicht ohne weiteres möglich und auch nicht unbedingt sinnvoll. So haben sich viele Vertreter der Volkswirtschaftslehre in Deutschland bis zur Mitte des vorigen Jahrhunderts weitgehend der Geschichte verpflichtet gefühlt. Sie waren primär an wirtschaftshistorischen Fragen und Daten interessiert, weshalb man diese Gelehrten heute eher als Wirtschaftshistoriker oder empirische Wirtschaftsforscher bezeichnen würde.

Die folgende Tabelle nennt die 5 am häufigsten nachgefragten Einzelstudien (Anzahl der Downloads von Datentabellen aus den Studien bis Juni 2012 in eckigen Klammern).

1. Kraus, A. (1980): Quellen zur Bevölkerungsstatistik Deutschlands 1815-1875. (=Quellen zur Bevölkerungs-, Sozial- und Wirtschaftsstatistik Deutschlands 1815-1875, Bd. I), hrsg. von Köllmann, W. Boppard am Rhein (GESIS Datenfile ZA8276). [300]
2. Bundesministerium für Arbeit und Soziales (2006): Statistische Übersichten zur Sozialpolitik in Deutschland seit 1945 (Band SBZ/DDR), Verfasser Steiner, A. unter Mitarbeit von Judt, M. und Reichel, Th., Bonn (GESIS Datenfile ZA8416) [258]
3. Heske, G. (2009): Volkswirtschaftliche Gesamtrechnung für die ehemalige DDR, 1950 bis 1989. HSR, Supplement No. 21/2009, Köln. (GESIS Datenfile ZA8375) [240]
4. Metz, R. (2005): Säkulare Trends der deutschen Wirtschaft. In: North, M. (Hrsg.): Deutsche Wirtschaftsgeschichte – ein Jahrtausend im Überblick. 2. Aufl. München (GESIS Datenfile ZA8179) [194]
5. Heske, G. (2005): Die gesamtwirtschaftliche Entwicklung in Ostdeutschland 1970 bis 2000. Neue Ergebnisse einer volkswirtschaftlichen Gesamtrechnung. HSR, Supplement No. 17/2005, Köln. (GESIS Datenfile ZA8208) [159]

Neben diesen in der Forschung entstandenen Einzelstudien und den Handbüchern der amtlichen Statistik sind hier auch jene Studien zu nennen, die im Rahmen spezieller Forschungsschwerpunkte entstanden sind. Neben den Forschungsprojekten VASMA (Vergleichende Analysen der Sozialstruktur mit Massendaten) und HIWED (Historical Indicators of the Western European Democracies) haben die Schwerpunktprogramme „Quellen und Forschungen zur Historischen Statistik von Deutschland“ sowie „Deutsche Bildungsstatistik 1800-2005“ besondere Bedeutung (siehe oben).

Aus dem von der Stiftung Volkswagenwerk von 1979 bis 1984 geförderten HIWED-Projekt ist ein zweibändiges Datenhandbuch entstanden:

Flora, P.; Alber, J. ; Eichenberg, R.; Kohl, J.; Kraus, F.; Pfenning, W.; Seebohm, K. (Hrsg.): State, Economy, and Society in Western Europe 1815 – 1975. A Data Handbook in two Volumes. Vol. I

(1983): *The Growth of Mass Democracies and Welfare States*; Vol. II (1987): *The Growth of Industrial Societies and Capitalist Economies*. Frankfurt/M.

6 Kapitel des Datenhandbuches mit insgesamt 3280 Zeitreihen wurden unter dem Namen der Bearbeiter in histat aufgenommen:

1. Eichenberg, R.; Singer, D. (1983): *Das Personal im öffentlichen Dienst: Westeuropäische Länder 1850 bis 1975*. Titel des Beitrags im Handbuch: *Governmental Personnel* (GESIS Datenfile ZA8346)
2. Flora, P. (1983): *Öffentliches Schul- und Bildungssystem*. Titel des Beitrags im Handbuch: *Public Education* (GESIS Datenfile ZA8190)
3. Kohl, J. (1983): *Staatsausgaben in Westeuropa 1790 bis 1980*. Titel des Beitrags im Handbuch: *Public Expenditure* (GESIS Datenfile ZA8334)
4. Pfenning, W.; Flora, P. (1983): *Städteentwicklung und Urbanisierung in Westeuropa 1815 – 1975*. Titel des Beitrags im Handbuch: *Cities and Urbanization* (GESIS Datenfile: ZA 8388)
5. Seebohm, K. (1983): *Die Entwicklung der Steuerstruktur in den Ländern Westeuropas 1815 – 1975*. Titel des Beitrags im Handbuch: *Public Revenues* (GESIS Datenfile ZA 8379)
6. Pfenning, W. (1987): *Arbeitskämpfe in Westeuropa von 1865 bis 1975*. Titel des Beitrags im Handbuch: *Labour Disputes* (GESIS Datenfile: ZA8489)

Aus dem ebenfalls von der Stiftung Volkswagenwerk von 1979–1984 geförderten VASMA-Projekt sind 2 Studien mit insgesamt 2462 Zeitreihen in histat aufgenommen worden:

1. Stockmann, R.; Kleber, W.; Willms-Herget, A. (1985): *Die Entwicklung der Betriebs- und Beschäftigungsstrukturen nach Branchen 1875 – 1982*. VASMA-Projekt, Universität Mannheim, Arbeitspapier Nr. 44. (GESIS-Datenfile ZA8181)
2. Willms-Herget, A.; Stockmann, R. (1985): *Frauenarbeit – Zur Integration der Frauen in den Arbeitsmarkt*. Frankfurt. (GESIS Datenfile ZA 8180)

Aus dem von der Deutschen Forschungsgemeinschaft von 1981 bis 1991 geförderten Schwerpunktprogramm „Quellen und Forschungen zur Historischen Statistik von Deutschland“ sind bislang 26 Bände hervorgegangen, von denen die folgenden 12 Bände mit insgesamt 38723 Zeitreihen in histat aufgenommen wurden:

1. Ott, H. Hrsg. (1986): *Statistik der öffentlichen Elektrizitätsversorgung Deutschlands 1890 bis 1913*. Band 1: *Historische Energiestatistik von Deutschland*. Bearbeitet von Herzig, Th., unter Mitarbeit von Ph. Fehrenbach und M. Drummer. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 1), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8409).
2. Borscheid, P.; Drees, A. Hrsg. (1988): *Versicherungsstatistik Deutschland 1750 – 1985*. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 4), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8347)

3. Fischer, W. Hrsg. (1989): Die Statistik der Stahlproduktion im deutschen Zollgebiet 1850-1911. Bearbeitet von J. Krengel unter Mitarbeit von A. Kunz und S. Stöckel. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 7), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8195)
4. Fischer, W. Hrsg. (1989): Statistik der Bergbauproduktion Deutschlands 1850 – 1914. Bearbeitet von Ph. Fehrenbach. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 8), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen (GESIS Datenfile ZA8448)
5. Hohls, R.; Kaelble, H. Hrsg. (1989): Die regionale Erwerbsstruktur im Deutschen Reich und in der Bundesrepublik 1895 – 1970. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 9), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8212)
6. Kaufhold, K. H.; Sachse, W. Hrsg. (1989): Gewerbestatistik Preussens vor 1850, Band 1: Das Berg-, Hütten- und Salinenwesen. Bearbeitet von Albrecht, U.; Kleinecke, D.; Kühl, U.; Laufer, J. und Sachse, W. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 5), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8481)
7. Spode, H.; Volkmann H.; Morsch G.; Hudeman, R. Hrsg. (1992): Statistik der Arbeitskämpfe in Deutschland. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 15), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8407)
8. Fischer, W. Hrsg. (1995): Statistik der Montanproduktion Deutschlands 1915 – 1985. Bearbeitet von Ph. Fehrenbach. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 16), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8400)
9. Fremdling, R.; Federspiel, R.; Kunz, A. Hrsg. (1995): Statistik der Eisenbahnen in Deutschland 1835 – 1989. Bearbeitet von R. Federspiel unter Mitarbeit von Hüchtker, D. und Johansen, A. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 17), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8411)
10. Büch, H. (1998): Bevölkerung des Kurfürstentums / Königreichs Hannover 1745 – 1867, in: Kaufhold, K. H./Denzel, M. A. Hrsg.: Historische Statistik des Kurfürstentums / Königreichs Hannover. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 23), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8420)
11. Kunz, A. Hrsg. (1999): Statistik der Binnenschifffahrt in Deutschland 1835-1989. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 18), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8157)
12. Kaschke, L.; Sniegs, M. (2001): Kommentierte Statistiken zur Sozialversicherung in

Deutschland von ihren Anfängen bis zur Gegenwart. Die Invaliditäts - und Altersversicherung im Kaiserreich, 1891 – 1913. (=Quellen und Forschungen zur Historischen Statistik von Deutschland, Bd. 26), hrsg. von W. Fischer, F. Irsigler, K. H. Kaufhold und H. Ott). St. Katharinen. (GESIS Datenfile ZA8368)

Aus dem ebenfalls von der DFG geförderten Projektverbund „Deutsche Bildungsstatistik 1800-2005“ sind bislang 9 Bände publiziert, von denen die folgenden 7 mit insgesamt 77544 Zeitreihen in histat aufgenommen wurden:

1. Müller, D. K.; Zymek, B. (1987): Datenhandbuch zur deutschen Bildungsgeschichte. Band II: Höhere und mittlere Schulen. 1. Teil: Sozialgeschichte und Statistik des Schulsystems in den Staaten des Deutschen Reichs 1800-1945. Göttingen. (GESIS Datenfile ZA8240)
2. Titze, H. (unter Mitarbeit von H.-G. Herrlitz; V. Müller-Benedict und A. Nath) (1987, 1995): Datenhandbuch zur deutschen Bildungsgeschichte. Band I, Teil 1: Das Hochschulstudium in Preußen und Deutschland 1820 – 1944. Band I, Teil 2: Wachstum und Differenzierung der deutschen Universitäten 1820 – 1944. Göttingen.
3. Hermann, U. G.; Müller, D. K. (2003): Datenhandbuch zur deutschen Bildungsgeschichte. Band II: Höhere und mittlere Schulen. Teil 2: Regionale Differenzierung und gesamtstaatliche Systembildung. Preußen und seine Provinzen, Deutsches Reich und seine Staaten, 1800–1945. Göttingen (GESIS Datenfile ZA8478)
4. Müller-Benedict, V. (unter Mitarbeit von J. Janssen und T. Sander) (2007): Datenhandbuch zur Deutschen Bildungsgeschichte. Band VI: Akademische Karrieren in Preußen und Deutschland 1850-1945. Göttingen. (GESIS Datenfile ZA8274)
5. Lundgreen, P. (unter Mitarbeit von J. Scheunemann und G. Schwibbe) (2008): Datenhandbuch zur deutschen Bildungsgeschichte. Band VIII: Berufliche Schulen und Hochschulen in der Bundesrepublik Deutschland 1949–2001. Teil 1: Berufliche Schulen. Göttingen. (GESIS Datenfile ZA8201)
6. Lundgreen, P. (unter Mitarbeit von J. Scheunemann und G. Schwibbe) (2008): Datenhandbuch zur deutschen Bildungsgeschichte. Band VIII: Berufliche Schulen und Hochschulen in der Bundesrepublik Deutschland 1949–2001. Teil 2: Hochschulen. Göttingen. (GESIS Datenfile ZA8202)
7. Lundgreen, P., (unter Mitarbeit von G. Schwibbe und J. Schallmann) (2009): Datenhandbuch zur deutschen Bildungsgeschichte. Band X: Das Personal an den Hochschulen in der Bundesrepublik Deutschland 1953–2005. Göttingen. (GESIS Datenfile ZA8380)

Eine besondere Art von Studie stellen die DHS-Datenkompilationen dar. Dabei werden aus wissenschaftlichen Einzelstudien und aus Angaben der amtlichen Statistik Zeitreihen zu ganz bestimmten Themen möglichst umfassend zusammengestellt und quellenkritisch dokumentiert. Mit diesen Datenkompilationen verbindet sich der Anspruch, die zu einem bestimmten Thema meist weit verstreut vorliegenden Zeitreihen zu bündeln und damit kompakt verfügbar zu ma-

chen. Die hohe Nachfrage nach diesen Studien zeigt deutlich den großen Bedarf an solchen Datensammlungen seitens der Profession.

Nachstehend eine Liste mit häufig nachgefragten Datenkompilationen (Anzahl der Downloads von Datentabelle aus den Studien bis Juni 2012 in []):

- Besser, Ch. (2008), Grunddaten der Bevölkerungsstatistik Deutschlands von 1871 bis 1939. (GESIS Datenfile ZA 8295) [494]
- Sensch, J. (2004), Geschichte der deutschen Bevölkerung seit 1815. (GESIS Datenfile ZA8171) [431]
- Sensch, J. (2004), Erwerbstätigkeitsstatistik in der Bundesrepublik Deutschland. Ausgewählte Daten von 1950 bis 2000. (GESIS Datenfile ZA8184) [289]
- Sensch, J. (2007), Bevölkerungsstand, Bevölkerungsbewegung, Haushalte und Familien in der Bundesrepublik Deutschland 1947 bis 1999. (GESIS Datenfile ZA8200) [207]
- Franzmann, G. (2006), Bildung in Deutschland: Bildungsstatistische Zeitreihen von 1960 bis 2000 zur Schüler- und Studentenzahl, zum Lehrpersonal und zu den Bildungsausgaben. (GESIS Datenfile ZA8220) [156]

Themen in histat

Alle für histat aufbereiteten Studien werden einem bestimmten Thema zugeordnet. Insgesamt kennt histat die folgenden 26 Themen:

- | | |
|------------------|------------------|
| ▪ Arbeit | ▪ Städte |
| ▪ Bauen | ▪ Umwelt |
| ▪ Bevölkerung | ▪ Unternehmen |
| ▪ Bildung | ▪ Verbrauch |
| ▪ Einkommen | ▪ Verkehr |
| ▪ Energie | ▪ Versicherungen |
| ▪ Geld | ▪ VGR |
| ▪ Gesundheit | ▪ WahlenDie |
| ▪ Handel | |
| ▪ Hanse | |
| ▪ Industrie | |
| ▪ Innovation | |
| ▪ Konjunktur | |
| ▪ Kriminalität | |
| ▪ Landwirtschaft | |
| ▪ Preise | |
| ▪ Sozialstaat | |
| ▪ Staatsfinanzen | |

Anzahl der Studien zu den einzelnen Themen ist recht unterschiedlich, was zum einen Ausdruck des jeweiligen Forschungsstandes sein kann, zum anderen aber auch aus spezifischen Kooperationen mit Primärforschern in der bisherigen Entwicklung der Datenbank zu erklären ist. So gibt es zum Beispiel zum Thema „Energie“ nur eine Studie, wogegen die Datenbank zum Thema „Industrie“ insgesamt 32 Studien enthält.

Neben der Anzahl der Studien pro Thema ist natürlich auch die *Anzahl der Zeitreihen pro Thema* ein wichtiger Indikator für die Qualität der Datenbank. Auch hier variieren die Zahlen pro Thema ganz beträchtlich. Während z.B. das Thema „Unternehmen“ mit nur 548 Zeitreihen vertreten ist, finden sich zum Thema „Bildung“ insgesamt 82087 Zeitreihen.

Neben der Anzahl der Studien und Zeitreihen pro Thema ist auch das Datum der Publikation der Primärstudie bzw. der Erstellung der Datenkompilation von Interesse. Es informiert über die Aktualität der Daten ebenso wie über die Forschungsgeschichte bzw. -intensität einzelner Themen. So datiert die älteste Studie in das Jahr 1905 (Thema „Landwirtschaft“). Die neuesten in die Datenbank aufgenommenen Studien wurden 2009 bzw. 2010 veröffentlicht. Dabei handelt es sich um insgesamt 7 Studien: 1 Studie zum Thema Handel (2010), 1 Studie zum Thema Bildung, 1 Studie zum Thema Landwirtschaft (2009), 2 Studien zum Thema Industrie (2009), 1 Studie zum Thema Staatsfinanzen und eine Studie zum Thema VGR (2009).

Die folgende Übersicht nennt für die Themen, die 10 und mehr Studien umfassen, die bezüglich der Anzahl der Zeitreihen kleinste (KS) und größte Studie (GS) sowie die bezüglich ihres Erscheinungsdatums älteste (ÄS) und jüngste Studie (JS). Die Studiennummer wird in runden Klammern angegeben. Zu jedem genannten Thema wird zusätzlich die Anzahl der Studien und Zeitreihen in eckigen Klammern angegeben.

Thema: Arbeit [29/19.885]

KS (8486): 16	ÄS (8256): 1965
GS (8212): 5234	JS (8181): 2004

Thema: Bauen [16/4097]

KS (8427): 7	ÄS (8430): 1975
GS (8474): 987	JS (8371): 1998

Thema: Bevölkerung [24/25413]

KS (8369): 23	ÄS (8272): 1912
GS (8295): 5205	JS (8162): 2001

pedia of the United States: American History in Numbers". Schließlich ist die didaktisch orientierte Publikation „The first measured century“ zu nennen, die, basierend auf einer Fernsehproduktion, weite Verbreitung gefunden hat. In Großbritannien gilt die *British Historical Statistics* von B. R. Mitchell seit ihrem ersten Erscheinen 1988 als Standardwerk. Vor kurzem ist die aktuelle Auflage als Taschenbuch erschienen (November 2011, Cambridge University Press, 900 Seiten). In Kanada wurde die *Historical statistics of Canada* 1965 in erster und 1983 in zweiter Auflage von der Social Science Federation of Canada and Statistics Canada herausgegeben. Die Publikation ist eine Gemeinschaftsarbeit von Wissenschaftlern, dem Social Science Research Council sowie dem Statistischen Amt. Um der nach wie vor starken Nachfrage zu begegnen, wurde 1999 eine elektronische Fassung erstellt. In Japan enthält ein vom Statistics Bureau und dem Director-General for Policy Planning of Japan erstelltes online-Angebot Zeitreihen ab 1868. Das Angebot basiert auf der Publikation *Historical Statistics of Japan* von 1987/1988, die von der japanischen Statistischen Zentraleinrichtung zeitlich verlängert und erweitert wurde. In Russland wird seit 2010 ein *Electronic Repository of Russian Historical Statistics, 18th - 21st centuries* erstellt. Der Abschluss ist für 2013 geplant. In der Schweiz präsentiert das Bundesamt für Statistik online in der Rubrik „Historische statistische Tabellen der Schweiz“ historische Zeitreihen zu 22 Themen. Diese Arbeit ist noch nicht abgeschlossen, aber für ein Dutzend der Themen gibt es bereits komplette Sammlungen. Zusammengestellt sind etwa 800 Tabellen, basierend auf der 1996 erschienenen großen Kollektion „Historische Statistik der Schweiz“. Für Spanien haben Albert Carreras und Xavier Tafunell 2005 in zweiter Auflage eine nationale historische Statistik in drei Bänden herausgegeben. In Frankreich gibt es im Rahmen der *Annuaire Statistiques de la France* die *annuaires rétrospectifs*: Vol. 58 (1951) und Vol. 72 (1966) sowie die *Séries longues: 1948-1988* (1990), in denen eine Vielzahl von langen Reihen vorhanden ist. In Italien bietet das nationale Statistische Amt ISTAT in der Sparte "Lange Reihen" (serie storiche) unter dem Kapitel Bevölkerung verschiedene Reihen von 1861 bis 2009. 2011 wurde anlässlich des 150. Jahrestages der Einigung Italiens der *L'Atlante Statistico delle Province d'Italia* mit zahlreichen historischen Zeitreihen herausgegeben.

Konzept

Es ist mehr als naheliegend, aus den mittlerweile über 250.000 in *histat* enthaltenen Zeitreihen sowie in Fortführung der überholten Publikation des Statistischen Bundesamtes ein vademecum zu erstellen, das für die Wissenschaft, die Politik sowie die Medien eine kompakte und verlässliche Übersicht mit den „1.000 wichtigsten“ Reihen verfügbar macht. Geplant ist eine „kombinierte Hybrid-Edition“: eine „kombinierte Hybrid-Edition“: Neben einer Taschenbuchausgabe, die jedes zehnte Jahr von 1850 bis 2010 tabellarisch aufführt (25 % des Gesamtumfangs), die wichtigsten Entwicklungen allgemeinverständlich erläutert (25 % des Gesamtumfangs) sowie grafisch abbildet (50 % des Gesamtumfangs), werden die Zeitreihen auf Jahresbasis und wissenschaftlich dokumentiert in der *histat*-Datenbank als „Datenkompilation“ angeboten. Jede Tabelle des Taschenbuches enthält einen kurzen Link-Hinweis auf die vollständig dokumentierte Tabelle in der Datenbank. Sämtliche Grafiken werden darüber hinaus in der Datengalerie des *histat*-Portals angeboten. Eine Vorauswahl der Reihen wird durch die Mitarbeiterinnen und Mitarbeiter unseres Bereiches basierend auf dem *histat*-Angebot vorgenommen. Geplant sind aktuell maximal 1.000 Zeitreihen, eingeteilt in ca. 20 Themengebiete.

Beteiligung der Wissenschaft

In einem zweiten Schritt werden anerkannte Expertinnen und Experten zu den jeweiligen Themen / Tabellen (jedes Thema hat etwa 6-8 Tabellen, damit ca. 50 bis 60 Zeitreihen) mit der Bitte um Bestätigung unserer Auswahl, der Nennung von Literatur- und/oder Quellenhinweisen eingebunden.

Umsetzung

Nach erfolgter Rücksprache mit den jeweiligen Expertinnen und Experten werden die identifizierten Zeitreihen von den Mitarbeiterinnen und Mitarbeitern des Datenservice Historische Studien abschließend zusammengestellt, aufbereitet und dokumentiert. Die Generierung der Druckvorlagen des Taschenbuches erfolgt dynamisch aus einer Datenbank heraus. Die dafür notwendigen Datenbankstrukturen und Satzroutinen wurden in den vergangenen Wochen bereits programmiert. Die Grafiken werden ebenfalls datenbankgestützt nach einheitlichen Regeln mit der neu entwickelten histat-Galeriefunktionalität erstellt.

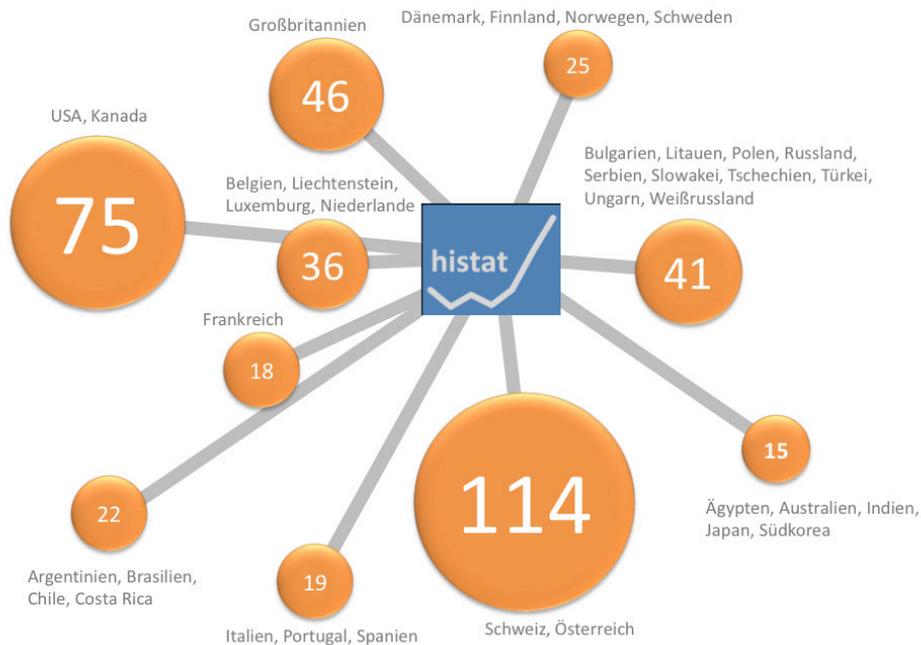
Kooperation

Das Projekt erfolgt in Kooperation mit der Bundeszentrale für Politische Bildung und dem Statistischen Bundesamt.

6.3 Nutzungsstatistik

Seit der Freischaltung verzeichnet histat eine jährlich wachsende Nutzung. Mit der Aufhebung der Registrierungspflicht zur Recherche im Jahre 2010 hat sich die Anzahl der Datenbankaufrufe innerhalb von zwei Jahren mehr als verdoppelt. Dieses Jahr werden wir über 3.000 registrierte Nutzerinnen und Nutzer begrüßen können.

Registrierungen aus dem Ausland, darunter Cambridge, Oxford, Zürich, Berkeley, Princeton, Stanford.



2011 wurde die Datenbank 3.482-mal aufgerufen. Insgesamt wurden 3.809 Tabellen heruntergeladen, die zusammen 71.903 Zeitreihen umfassten. Obwohl einige Studien sehr häufig frequentiert werden, ist die Nutzung des Angebotes insgesamt erstaunlich breit. So wurden von den 2011 in histat enthaltenen 359 Studien 269, also drei Viertel auch tatsächlich benutzt.

	2006	2007	2008	2009	2010	2011
Nutzung						
Datenbankaufrufe ¹	800	1 691	1 607	1 624	≈3 000	3 482 ²
Tabellen-Downloads	326	1 082	995	2 018	2 044	3 809 ³
Zeitreihen-Downloads	5 964	20 057	21 973	37 760	49 660	71 903
verwendete Studien	50	109	133	168	221	269
Bestand						
Neue Registrierungen	181	604	584	554	325	314
Registrierungen insgesamt ⁴	492	1 096	1 680	2 234	2 559	2 873
Neue Studien	45	27	42	48	70	28 ⁵
Studien insgesamt ⁴	144	171	213	261	331	359
Neue Zeitreihen	17 926	14 566	16 516	39 462	37 537	20 044
Zeitreihen insgesamt ⁴	121 299	135 865	152 381	191 843	229 380	249 424

¹Seit 2010 Aufruf mit Ausnahme des downloads ohne Registrierung möglich – ²Angaben aus eTracker. 2010 aus Juli bis Dezember hochgerechnet. – ³Davon 1 982 Downloads durch zwei Nutzer/innen. – ⁴Kumulative Angaben, jeweils Jahresende. – ⁵2011 weniger Neuaufnahmen durch die Vorbereitung des vademecums.

7 Bestandserweiterung und Akquise

7.1 Beobachtung der Forschung

Die bisherigen Ausführungen haben gezeigt, dass die Datenbank einen thematisch vielfältigen Bestand an Studien mit Zeitreihen beinhaltet. Dieser Bestand wird von uns laufend erweitert, und zwar 1) durch die Akquirierung neuer Studien, 2) durch die Aufarbeitung historischer Studien sowie 3) durch die Zusammenstellung von Datenkollektionen zu ausgewählten Themen.

1) Für die Akquirierung neuer Studien ist eine systematische Beobachtung der aktuellen Forschungen notwendig. Hierzu recherchieren wir laufend die Publikationen der Mitglieder des Wirtschaftshistorischen Ausschusses des Vereins für Socialpolitik, da insbesondere diese Fachvertreter quantitativ arbeiten (58 Mitglieder, Stand: September 2012). Zudem recherchieren wir laufend die Publikationen und Forschungsbeiträge ausgewählter Mitglieder der Gesellschaft für Sozial- und Wirtschaftsgeschichte, wobei die ausgewählten Wissenschaftler jene Gruppe repräsentieren, die sich mit seriellen Daten bzw. quantitativen Methoden in der Geschichte befassen bzw. befasst haben. Neben diesen personenbezogenen Recherchen beobachten wir laufend folgende Zeitschriften:

- Cliometrica
- Economic History Review
- Explorations in Economic History
- Jahrbuch für Wirtschaftsgeschichte
- Journal of Economic History
- The Journal of European Economic History
- Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte

Die VSWG hat einen sehr umfangreichen Rezensionsteil sowie eine Rubrik eingegangener Bücher. Beide werden von uns ebenfalls systematisch ausgewertet.

Neben diesen gedruckten Medien werden von uns auch Internetseiten im Hinblick auf Veröffentlichungen, die historische Zeitreihen enthalten bzw. analysieren beobachtet. Darunter befinden sich:

- Cliometric Society
- h-soz-u-kult
- Economic History Association (EH.net)

Eine weitere wichtige Quelle für die Recherche von Forschungsbeiträgen zu historischen Zeitreihen sind die Forschungsdokumentationen von GESIS

Für die für histat als relevant erachteten Studien werden die Forscherinnen und Forscher angeschrieben. Dabei bieten wir ihnen die Archivierung ihrer Daten an. Zusätzlich zur Archivierung offerieren wir die Möglichkeit, die Daten auch in histat online zur Verfügung zu stellen.

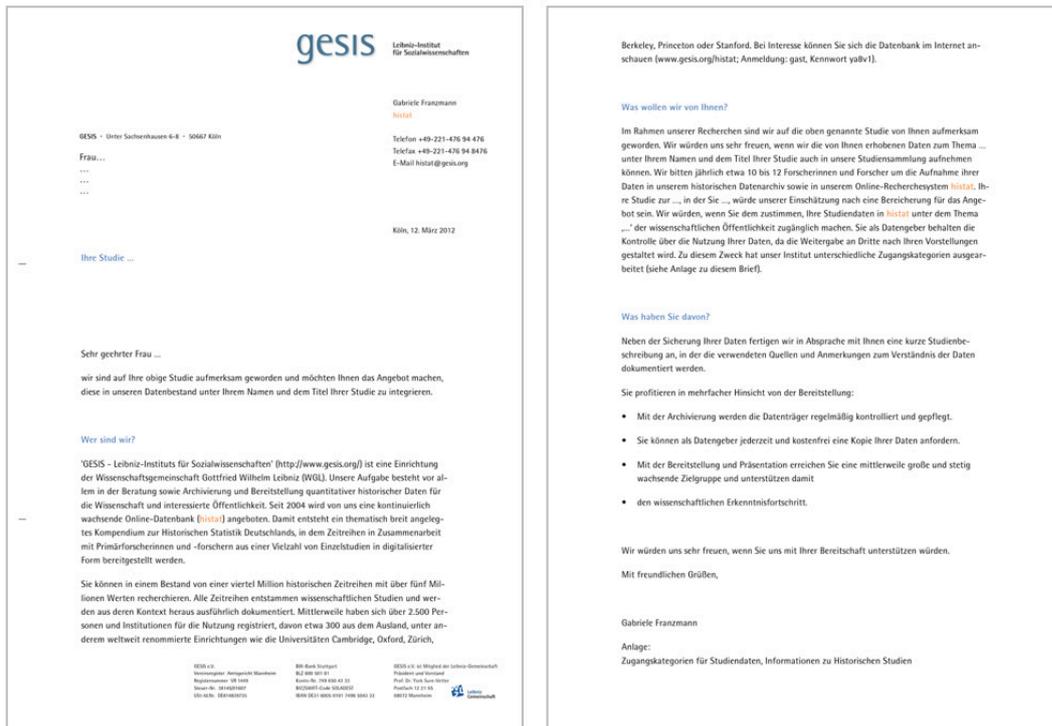
2) Bei der Aufbereitung historischer Studien mit Zeitreihendaten können wir auf einen umfangreichen Bestand an Publikationen aus der Forschung zurückgreifen, die bislang zu jenen Themen entstanden sind, die wir in der Datenbank anbieten, jedoch noch nicht elektronisch vorliegen. Diese Daten sind in ganz unterschiedlichen Kontexten entstanden. Und zwar sowohl im Rahmen wirtschaftshistorischer wie auch im Rahmen wirtschaftswissenschaftlicher Forschungen. Letztere sind besonders deshalb von großer Bedeutung, da sich zahlreiche Vertreter der Volkswirtschaftslehre bis zur Mitte des 20. Jahrhunderts der Empirie und damit auch der Wirtschaftsgeschichte besonders verpflichtet sahen. So sind z.B. im Rahmen der empirischen Konjunkturforschung noch bis in die 1950er Jahre umfangreiche Editionen mit langen Zeitreihen zur deutschen Wirtschaftsgeschichte vorgelegt worden.

Bei der Recherche und Auswahl solcher Studien arbeiten wir themenorientiert. Grundlage ist eine Prioritätenliste der in histat enthaltenen Themen, für die wir ältere Studien recherchieren und auswählen. Auswahlkriterien sind die forschungsgeschichtliche Relevanz der Studie, Anzahl und Länge der in ihr enthaltenen Zeitreihen sowie natürlich die Möglichkeit, die Daten einzuscannen. Selbstverständlich werden vor der Aufarbeitung der Studie auch mögliche Copyright Probleme geklärt. Die in diesen Arbeiten veröffentlichten Zeitreihen werden dann eingescannt, in Tabellen aufbereitet, dokumentiert und zusammen mit der Beschreibung der Studie in histat angeboten. Das Ziel ist es, die Daten zusammen mit ihrem Entstehungs- und Forschungskontext online für Sekundäranalysen anzubieten.

3) Bei der Erstellung von Datenkompilationen geht es ganz allgemein darum, historische Zeitreihen zu ausgewählten Themen zusammen zu stellen. Ausgangspunkt ist ein bestimmtes, spezielles Thema, ein ganz bestimmter Aspekt der Wirtschafts- und Sozialgeschichte, wie z.B. Arbeitszeit, Auswanderung, Löhne, Patente, Gemeindegrößen. Alle diese Spezialthemen lassen sich einem übergreifenden Thema der Datenbank bzw. dem vademecum Historische Statistik zuordnen. Die Themen, für die eine Datenkompilation erstellt wird, ergeben sich aus einer Prioritätenliste, die institutsintern festgelegt wird. Die Arbeit an einer solchen Kompilation beginnt mit der Sichtung des Forschungsstandes: Welche Reihen gibt es zu diesem Thema in der Forschung? Welche Daten liefert die amtliche Statistik? Die Recherche berücksichtigt die Zeit vom 19. Jahrhundert bis zur Gegenwart. Nach Sichtung der Datenlage, die also sowohl Forschungsdaten wie auch Daten der amtlichen Statistik einschließt, geht es darum, festzulegen, welche Variablen / Zeitreihen für dieses Thema bislang relevant waren und welche heute noch relevant sind. Betrachtet man z.B. die Volkswirtschaftliche Gesamtrechnung, dann stellt man fest, dass gegenwärtig das Bruttoinlandsprodukt die zentrale Variable ist. Davor war es, bis in die 1990er Jahre das Bruttosozialprodukt. Im deutschen Kaiserreich und in der Weimarer Zeit war es dagegen das Volkseinkommen, das als Indikator der Volkswohlfahrt betrachtet wurde. In der Datenkompilation werden diese Größen in ihrem historischen und systematischen Kontext rekonstruiert, aufbereitet und dokumentiert. Ergebnis ist eine Studie in histat, die Zeitreihen aus unterschiedlichen wissenschaftlichen Publikationen sowie Veröffentlichungen der amtlichen Statistik für einen historischen Zeitraum zusammenführt und auch in ihrem Entstehungskontext beschreibt.

7.2 Anschreiben

Handelt es sich bei der neu aufzunehmenden Studie nicht um eine gemeinfreie Publikation, etwa mit abgelaufenem Copyright, besteht der erste Schritt darin, die Autorinnen bzw. Autoren anzuschreiben. Das sieht üblicherweise so aus:



Antwortet die angeschriebene Person positiv auf das Akquise-Anschreiben, schließen wir mit ihr den vom GESIS-Datenarchiv ausgearbeiteten Archivierungsvertrag ab, in dem Name der Studie, Studiennummer, die Kriterien für die Datenweitergabe sowie die Serviceleistungen des Archivs hinsichtlich der Langfristsicherung der Daten (Beschreibung auf Studien- und Datenebene und Veröffentlichung der Studienbeschreibung im recherchierbaren Datenbestandskatalog, Pflege und Kontrolle der Datenträger, freier Zugang des Datengeber zu seinen Daten) festgehalten werden.

In den vergangenen Jahren wurden von uns pro Jahr etwa 10 bis 20 Forscherinnen und Forscher um die Aufnahme ihrer Daten in das historische Datenarchiv sowie in das Recherchesystem hista gebeten. Darüber hinaus wurde eine Reihe von Gesprächen mit Verlagen geführt.

Die einzelnen Akquisevorgänge werden im DHS-Team gemeinsam mit einer Sharepoint-Aufgabenliste verwaltet.

The screenshot shows the Microsoft Outlook interface. The title bar reads 'Datenservice Historische Studien - Studienakquisition - SharePoint-Listen - Microsoft Outlook'. The ribbon includes 'Datei', 'Start', 'Senden/Empfangen', 'Ordner', and 'Ansicht'. The left sidebar shows folders like 'Posteingang', 'Entwürfe', 'Aufgaben', etc. The main pane displays a task list with columns for 'Betreff', 'Zugewiesen an', 'Benutzerdefinierte Pri.', 'Fällig am', and 'Geändert'.

Betreff	Zugewiesen an	Benutzerdefinierte Pri.	Fällig am	Geändert	
Hier klicken, um Aufgabe zu erstellen					
Die Herausforderung der Globalisierung, Wirtschaft und Politik in ...	Franzmann, Gabriele	In Bearbeitung	Normal	Ohne	Di 18.10.2011 10:26
Historische Haushalts- und Familienstatistik / Soziale Ungleichheit L...	Sensch, Jürgen	In Bearbeitung	Normal	Ohne	Do 07.07.2011 11:...
Das deutsche Innovationssystem seit der Reichsgründung.	Sensch, Jürgen	In Bearbeitung	Normal	Ohne	Do 07.07.2011 11:...
150 Years of Patent Protection.	Franzmann, Gabriele	In Bearbeitung	Normal	Ohne	Di 28.06.2011 14:45
Marktintegration und Entwicklung der westfälischen Landwirtschaft ...	Sensch, Jürgen	In Bearbeitung	Normal	Ohne	Di 28.06.2011 14:39
Äcker, Wirt, Gaben: Ländlicher Bodenmarkt und liberale Eigentum...	Franzmann, Gabriele	In Bearbeitung	Normal	Ohne	Di 28.06.2011 14:35
Die Kreditbanken in der Gründerzeit.	Franzmann, Gabriele	Erledigt	Normal	Ohne	Di 28.06.2011 14:28
Mit aller Kraft verdrängt, Entrechtung und Verfolgung nicht arisch...	Franzmann, Gabriele	In Bearbeitung	Normal	Ohne	Di 28.06.2011 14:22
Kirchenbücher des Kreises Wittgenstein (MS-Access 2002) - Wittge...	Franzmann, Gabriele	Erledigt	Normal	Ohne	Di 28.06.2011 14:19
Mit hohem Anspruch geschichtet - Die Wohnungspolitik der DDR.	Franzmann, Gabriele	Erledigt	Normal	Ohne	Di 28.06.2011 14:16
Die Industrialisierung im Hagener Raum zwischen 1815 und 1914.	Franzmann, Gabriele	Erledigt	Normal	Ohne	Di 28.06.2011 14:10

8 Studienbeschreibung und Aufbereitung der Zeitreihen

8.1 Ziele der Beschreibung und Aufbereitung

Beschreibung und Aufbereitung einer Studie für histat haben das Ziel, die Zeitreihen der Studie für Sekundäranalysen in Form von Tabellen sowie Kontextinformationen zu den Daten bereit zu stellen. Die **Beschreibung** betrifft alle Tätigkeiten, die die Informationen zur Studie selbst liefern, also die Studie beschreiben. Die **Aufbereitung** dagegen betrifft alle Tätigkeiten, die die notwendigen Informationen für die Beschreibung der Zeitreihen in den Tabellen liefern.

8.2 Die Beschreibung einer Studie

Die Beschreibung einer Studie enthält Studiennummer, das Thema, dem die Studie in histat zugeordnet ist, den Namen des Primärforschers, den Titel der Studie, die Anzahl der Zeitreihen, den Zeitraum, den die Zeitreihen erfassen sowie die Anzahl der Tabellen, in denen die Zeitreihen zusammengefasst sind. Darüber hinaus enthält die Beschreibung folgende Rubriken:

- Datensatz
- Anmerkungen
- Gedruckte (zitierpflichtige) Publikation
- Studienbeschreibung
- Dokumente für den Download
- Verwendete Quellentypen
- Untersuchungsgebiet
- Verzeichnis der verwendeten Quellen

Datensatz: Hier wird die empfohlene Zitierweise der Studie angegeben, wenn die Daten dieser Studie in einer anderen Publikation verwendet werden. In dieser Zitierweise ist auch die Angabe zur Version des Datenfiles enthalten. Die Versionsnummer gibt an, ob die Daten der Studie seit dem ersten upload in histat verändert oder ergänzt wurden. Ist dies der Fall, wird die entsprechende Veränderung im Feld ‚Studienbeschreibung‘ mit Datumsangabe und dem Inhalt der vorgenommenen Änderung dokumentiert. Weiterhin finden sich zwei Jahresangaben zum Studientitel. Das erste Jahr bezieht sich auf die Veröffentlichung der Publikation, das zweite in eckigen Klammern angegebene Jahr nennt die Veröffentlichung der Studie als downloadbaren Datenfile in histat.

Als Beispiel hierzu die Studie ZA 8416: Steiner, André, Statistische Übersichten zur Sozialpolitik in Deutschland seit 1945; Band „Sowjetische Besatzungszone / DDR“.

Steiner, André, (2006 [2010]) Statistische Übersichten zur Sozialpolitik in Deutschland seit 1945; Band „Sowjetische Besatzungszone / DDR“. GESIS Köln, Deutschland ZA8416 Datenfile Version 1.0.0

Anmerkungen: Anmerkungen zur Studie können sich auf zwei Ebenen beziehen, zum einen auf die Ebene der gesamten Studie, zum anderen auf die Ebene einzelner Datentabellen. Methodische Kommentare beziehen sich auf die gesamte Studie. Dazu zählen detaillierte Begriffserläuterungen, das methodische Vorgehen bei der Berechnung oder Schätzung von Werten sowie quellenkritische Erläuterungen. Diese Kommentare werden in dem Anmerkungsteil im Bezug zur gesamten Studie aufgenommen. Beziehen sich Anmerkungen auf die Daten einer Tabelle oder auf einzelne Zeitreihen in einer Tabelle, werden sie datennah im Kennzeichnungsbereich der Datentabelle oder zellenspezifisch notiert. Handelt es sich um umfangreiche Anmerkungen, wird der Anmerkungstext in dem Anmerkungsteil der Studienbeschreibung unter Bezug zu der betreffenden Tabelle aufgenommen. In diesem Fall wird in der Datentabelle auf diesen Ort verwiesen.³

Gedruckte (zitierpflichtige) Publikation: Dieses Textfeld enthält den Titel der in histat aufgenommenen Studie.

Als Beispiel hierzu wieder die Studie ZA 8416:

Bundesministerium für Arbeit und Soziales, Statistische Übersichten zur Sozialpolitik in Deutschland seit 1945 (Band SBZ/DDR), Verfasser André Steiner unter Mitarbeit von Matthias Judt und Thomas Reichel. Bonn 2006.

Studienbeschreibung: Hier wird der Untersuchungsgegenstand der Studie mit der Fragestellung sowie der thematischen Untergliederung erläutert. Die Zusammenfassung der zentralen Inhalte der Studie (Fragestellung, Forschungskontext, Forschungshypothesen) bilden hier den Schwerpunkt, der auch eine kurze Darstellung der Datenerhebung (Quellenauswahl) beinhaltet. Weiterhin finden sich hier im Falle einer Versionierung die genauen Angaben zur Versionierung der Daten. Handelt es sich bei der Studie um ein gefördertes Projekt im Rahmen eines DFG – Schwerpunktprogramms, wird das übergeordnete Ziel des Schwerpunktprogramms, in dem das Teilprojekt eingebettet ist, ergänzend dargestellt. In manchen Studien wurde der Primärforscher mit besonderen Problemen der Methodik sowie der Datenerhebung aufgrund einer schwierigen Quellenlage konfrontiert, die zur Anwendung von Schätz- und Berechnungsverfahren führen. In diesem Fall wird auf diese Probleme sowie auf den vom Autor gewählten Lösungsansatz hingewiesen.

Als Beispiel hierzu wieder die Studie ZA 8416

Im Rahmen der von Bundesministerium für Arbeit und Soziales und Bundesarchiv gemeinsam herausgegebenen „Geschichte der Sozialpolitik in Deutschland seit 1945“ hat André Steiner unter Mitarbeit von Matthias Judt und Thomas Reichel eine vielfach aus Primärquellen gewonnene Sammlung statistischer Daten zur Sozialpolitik der Sowjetischen Besatzungszone Deutschlands und der Deutschen Demokratischen Republik erarbeitet. Sie bietet für zahlreiche statistische Maßgrößen der Sozialpolitik verwendbare Zahlenangaben in Form von Zeitreihen an. Zwar kann den neu gewonnenen Daten kein amtlicher Charakter zuerkannt werden, doch auf zahlreichen Teilgebieten der sozialpolitischen Entwicklung, zu denen die amtliche Statistik schweigt oder keine eindeutigen Angaben liefert, bieten die „Statistischen Übersichten zur Sozialpolitik in Deutschland seit 1945 (Band SBZ/DDR)“ neue, präzisere oder differenziertere Daten, zumindest aber plausible Schätz-

³ Da die Anmerkungen zur Studie 8416 sehr umfangreich sind, wir hier auf ihre Wiedergabe verzichten.

werte. Damit werden neue Grundlagen für die statistische Analyse der sozialen und sozialpolitischen Entwicklung der Deutschen Demokratischen Republik zur Verfügung gestellt.

„Die vorliegende Zusammenstellung statistischer Übersichten zur DDR-Geschichte geht grundsätzlich davon aus, dass die von der DDR-Statistik produzierten Daten für die historische Forschung genutzt werden können und müssen. Dabei sollten aber die Probleme und die Grenzen eines solchen pragmatischen Vorgehens bekannt und bewusst sein. Die Notwendigkeit für Historiker, diese Daten zu nutzen, ergibt sich zum einen daraus, dass eine Um- und Rückrechnung entsprechend westlichen Konzepten sehr aufwendig, zum Teil auch aus inhaltlichen Gründen nicht sinnvoll oder/und (bisher) nicht lösbar erscheint. Zum anderen können historische Prozesse nur dann adäquat beurteilt werden, wenn die zur damaligen Entscheidungsfindung zur Verfügung stehenden Daten berücksichtigt werden“ (Bundesministerium für Arbeit und Soziales, 2006: Statistische Übersichten zur Sozialpolitik in Deutschland seit 1945 (Band SBZ/DDR); Verfasser: Steiner, André unter Mitarbeit von Matthias Judt und Thomas Reichel. Bonn, Einleitung S. XIII).

Zusammenfassend ergeben sich bei der Arbeit mit der DDR-Statistik die folgenden grundlegenden methodischen Probleme: das angewendete Bruttoprinzip, die inkonsistenten Preise als Bewertungsmaßstab, nur beschränkt integrierbare Indikatoren qualitativer Entwicklungen sowie wechselnde Erfassungsdefinitionen und Bezugssysteme.

„Wegen der ... internen Probleme der DDR-Statistik muss auch ihr Vergleich mit der Statistik der Bundesrepublik auf kardinale Schwierigkeiten vielfältiger Natur stoßen, die nur partiell unter Zuhilfenahme verschiedener Vereinfachungen zu lösen sind. Schon die angeführten grundlegenden Probleme der DDR-Statistik führen dazu, dass sowohl die benutzten Primärdaten als auch die aus ihnen errechneten Indizes und Zuwächse nicht per se mit Angaben aus westlichen Statistiken zu vergleichen sind. Außerdem müssen statistische Daten zu Fragestellungen, die als politisch sensibel angesehen wurden, besonders kritisch überprüft werden. ... Gerade im sozialpolitischen Bereich gab es erhebliche Unterschiede in der Systemgestaltung, wenn man beispielsweise an die Finanzierung vieler Leistungen über den Staatshaushalt in der DDR denkt. Das wirft auch für die statistische Darstellung verschiedene Probleme auf. So erhebt sich im Zusammenhang mit der "gesellschaftlichen Konsumtion", also der kostenlosen Bereitstellung von Dienstleistungen sowie der nicht unerheblichen Subventionierung von Gütern und Leistungen durch den Staat, die Frage, wie dies in einer Aufbringungs-, Verwendungs- und Einkommensrechnung nach westlicher Struktur zugerechnet werden könnte und wer als der letzte Finanzier auftritt. Diese Schwierigkeiten spiegeln sich auch in den Umrechnungen auf westliche Konzepte wider, die nur für begrenzte Zeiträume bzw. einzelne Jahre vorliegen und in diese Zusammenstellung soweit im gegebenen Kontext sinnvoll aufgenommen wurden“ (BMAS, 2006: a. a. O., S. XIX, S. XX).

Datentabellen in HISTAT:

Zu den folgernden Themenschwerpunkten liegen jeweils zahlreiche Einzeltabellen vor:

Teil A:

A.1 Bevölkerung und Privathaushalte;

A.2. Staatshaushalt und Sozialsystem;

A.3. Volkswirtschaftliche Gesamtdaten;

A.4. Bilanz der Geldeinnahmen und –ausgaben der Bevölkerung und der Arbeiter- und Angestelltenhaushalte.

Teil B:

B.1. Zur Arbeitsverfassung;

B.2 Arbeitsschutz;

B.3 Arbeitskräftegewinnung und Arbeitskräftelenkung;

B.4 Sicherung bei Krankheit, Unfall sowie Alter, Invalidität und für Hinterbliebene

B.5. Gesundheitswesen;

B.6 Rehabilitation und Hilfe für Behinderte;

B.7 Fürsorge;

B.8. Familien- und Altenpolitik;

B.9. Bildungspolitik;

B.10 Nichtstaatliche soziale Infrastruktur;

B.11 Wohnungspolitik.

Dokumente für den Download: Oft sind der Quellen- und Anmerkungsteil einer Studie sehr umfangreich. Darüber hinaus kann es mitunter erforderlich sein, ein längeres Methodenkapitel für das richtige Verständnis der Daten aus der Publikation dem Nutzer mit den Daten anzubieten. In solchen Fällen werden diese Unterlagen als downloadbares PDF-Dokument dem Nutzer zur Verfügung gestellt.

Als Beispiel hierzu wieder die Studie ZA 8416:

Download weiterer Texte zu dieser Studie im PDF Format (Forschungsberichte, Publikationen, Materialien zur Studie)  [149,04 KB](#)

Verwendete Quellentypen: Hier werden Typen von Quellen genannt, denen die Quellen der Studie zugeordnet werden können, z.B.: Archivunterlagen; Aktenbestände; Kirchenbücher; Zeitungsberichte; offizielle Statistiken; wissenschaftliche Publikationen.

Als Beispiel hierzu wieder die Studie ZA 8416:

Publizierte Amtliche Statistiken:

- Das Gesundheitswesen der DDR, laufende Jahrgänge 1965 bis 1988.
- Statistisches Bundesamt Wiesbaden (Hrsg.): Sonderreihe mit Beiträgen für das Gebiet der ehemaligen DDR: Heft 2, 3, 13, 14, 15, 22.
- Ausgewählte Amtliche Veröffentlichungen der Staatlichen Zentralverwaltung für Statistik der DDR (Hrsg.).
- Statistisches Amt der DDR (Hrsg.), 1990: Jahrbuch Arbeitskräfte und Löhne 1989; Kennziffernsammlung Sozialstatistik 1990.
- Staatliche Versicherung der DDR (Hrsg.): Jahresbericht der Sozialversicherung bei der Staatlichen Versicherung der Deutschen Demokratischen Republik/Staatliche Versicherung der Deutschen Demokratischen Republik (SV StV), laufende Jahrgänge (1980 bis 1988).
- Verwaltung der Sozialversicherung beim Bundesvorstand des FDGB: Statistischer und Finanzieller Jahresbericht der Sozialversicherung der Arbeiter und Angestellten (SV AuA) der DDR, Jahrgänge 1965 – 1989.

Untersuchungsgebiet: Das Untersuchungsgebiet informiert über den historisch geografischen Raum, auf den sich die Studie bezieht. Die sich im Verlauf einer Zeitperiode verändernden Gebietsabgrenzungen können sich auf die Werte von erhobenen Zeitreihen (z.B. die Bevölkerung des Deutschen Reiches mit oder ohne Elsaß-Lothringen) auswirken und bedürfen daher einer genauen Dokumentation. Wenn sich die Studie und ihre Daten auf Gebiete mit sich verändernden Grenzen bezieht, ist eine Beschreibung der Gebietsveränderungen notwendig.

Als Beispiel hierzu wieder die Studie ZA 8416:

Sowjetische Besatzungszone / Deutsche Demokratische Republik, DDR (1945 - 1989). Als Gebietsstand für die vorliegenden statistischen Angaben ist in der Regel das Territorium der DDR einschließlich Berlin (Ost) zugrunde gelegt. In Ausnahmefällen ist Berlin (Ost) nicht mit eingeschlossen, was in Anmerkungen kenntlich gemacht wird.

Verzeichnis der verwendeten Quellen: Hier findet sich ein Verzeichnis der für die Studie verwendeten Quellen, bzw. ein Verweis auf das downloadbare PDF-Dokument, in dem sich das Quellen- und Literaturverzeichnis befindet.

Als Beispiel hierzu wieder die Studie ZA 8416:

Publizierte Amtliche Statistiken:

- Das Gesundheitswesen der DDR, laufende Jahrgänge 1965 bis 1988.
- Statistisches Bundesamt Wiesbaden (Hrsg.): Sonderreihe mit Beiträgen für das Gebiet der ehemaligen DDR: Heft 2, 3, 13, 14, 15, 22.
 - Ausgewählte Amtliche Veröffentlichungen der Staatlichen Zentralverwaltung für Statistik der DDR (Hrsg.).

- Statistisches Amt der DDR (Hrsg.), 1990: Jahrbuch Arbeitskräfte und Löhne 1989; Kennziffernsammlung Sozialstatistik 1990.
- Staatliche Versicherung der DDR (Hrsg.): Jahresbericht der Sozialversicherung bei der Staatlichen Versicherung der Deutschen Demokratischen Republik/Staatliche Versicherung der Deutschen Demokratischen Republik (SV StV), laufende Jahrgänge (1980 bis 1988).
- Verwaltung der Sozialversicherung beim Bundesvorstand des FDGB: Statistischer und Finanzieller Jahresbericht der Sozialversicherung der Arbeiter und Angestellten (SV AuA) der DDR, Jahrgänge 1965 – 1989.

8.3 Die Aufbereitung der Zeitreihen einer Studie

Histat enthält Studien mit ganz unterschiedlichem Umfang an Zeitreihen und Tabellen. Es gibt Studien mit nur wenigen Tabellen bzw. Zeitreihen und Studien mit tausenden von Zeitreihen sowie einer Vielzahl von Tabellen. Eine weitere Unterscheidung ist in diesem Zusammenhang von Bedeutung. So gibt es Studien, die Zeitreihen nur als Graphiken, nicht aber auch in Tabellenform enthalten. In diesem Fall ist man darauf angewiesen, dass der Primärforscher die zugrunde liegenden Zahlenwerte an das Archiv abgibt. Dies ist natürlich bei älteren Studien nicht möglich, da die Daten entweder nicht maschinenlesbar vorliegen und/oder der Primärforscher verstorben ist. In diesem Fall sind die Zeitreihen der Studie nur dann archivierbar, wenn sie als Tabelle in der Publikation abgedruckt sind. In diesem Falle, werden die Zeitreihenwerte eingescannt.

Bei der Aufbereitung werden die Zeitreihen einer Studie in histat-Tabellen übernommen. Die Tabellen sind so aufgebaut, dass den Zeitreihen alle Informationen zugeordnet werden, die für die Sekundäranalyse erforderlich sind. Da Studien in der Regel mehrere Datentabellen umfassen, die sich z.T. in ihrer inhaltlichen Struktur bzw. in ihrer Gliederungstiefe unterscheiden, wurde in Anlehnung an die DIN-Norm 55 301 ein einheitliches Tabellenschema für den Import in histat erarbeitet, das es erlaubt, Tabellen mit unterschiedlichen Gliederungen nach gleichen Kriterien aufzubereiten und darüber hinaus wichtige Metainformationen den Daten hinzuzufügen.

Eine Datentabelle in histat besteht aus zwei Hauptbereichen: Dem eigentlichen **Datenbereich** (Zahlenteil oder Tabellenfeld) mit seinen Tabellenfächern, in welche die Zeitreihenwerte eingefügt werden, und dem **Kennzeichnungsbereich**, bestehend aus Tabellenkopf und Vorspalte, in dem die eingetragenen Daten beschrieben werden. Zum Kennzeichnungsbereich, der in Abbildung X dunkelblau eingefärbt ist, gehören zum einen der Tabellentitel mit Angaben zum Sachverhalt, Raum- und Zeitbezug. Weiterhin werden hier die konkreten Bezeichnungen der Merkmale, nach denen ein Sachverhalt oder ein Merkmalsträger beschrieben wird, genannt (z.B.: Bundesland / Geschlecht). Da durch diese Merkmale die Sachverhalte oder Merkmalsträger geordnet dargestellt sind, werden diese Merkmale auch Gliederungsmerkmale genannt. Zusätzlich enthält der Tabellenkopf die Bezeichnung der Ausprägungen der Gliederungsmerkmale (z.B. für Bundesland: Bayern / und für Geschlecht: Männer, Frauen). Schließlich werden die Quellenhinweise und Anmerkungen für jede einzelne Zeitreihe sowie die Angabe der Tabellen-Nr. in

der Publikation in den Tabellenkopf eingetragen.

Vorspalte		
Tabellentitel	Tabellen-Nr. und Tabellentitel (z.B.: A.01 Bevölkerung in Deutschland nach Bundesländern und Geschlecht (1950-1980))	
Bezeichnung des Gliederungsmerkmals (z.B.: Bundesland)	Ausprägung des Gliederungsmerkmals (z.B.: Bayern)	Ausprägung des Gliederungsmerkmals (z.B.: Bayern)
Ggf. weiteres Gliederungsmerkmal (z.B.: Geschlecht)	Ausprägung des Gliederungsmerkmals (z.B.: Männer)	Ausprägung des Gliederungsmerkmals (z.B.: Frauen)
Maßeinheit	(z.B.: Angabe in 1000)	(z.B.: Angabe in 1000)
Quelle	(z.B.: Stat. Bundesamt (Hrsg.): Fachserie. Ort, Jahr, Seite.)	(z.B.: Stat. Bundesamt (Hrsg.): Fachserie. Ort, Jahr, Seite.)
Anmerkung	(z.B.: zum 31.12. des Jahres)	(z.B.: zum 31.12. des Jahres)
Tabelle	(z.B.: Tab. 03)	(z.B.: Tab. 03)
Zeitpunkt (Jahr, Intervall, etc.)	Wert	Wert
Zeitpunkt (Jahr, Intervall, etc.)	Wert	Wert

Tabellentitel:
Tabellen-Nr. plus sachliche Bestimmung, räumliche Bestimmung, zeitliche Bestimmung, Gliederungsbegriffe.

Tabellenkopf

Datenteil
(Tabellenfeld)

Kennzeichnungsbereich einer Datentabelle in histaat

3 Beispiele zellenspezifischer Anmerkungen:

- Dieser Wert beinhaltet Haushalte mit 8 Personen.
- In diesem Jahr besonders hohe Ernteverluste durch Sturmschäden.
- Wert für dieses Jahr wurde geschätzt.)

Der zweite große Informationsblock einer Studie ist der Datenbereich der Tabelle. Er enthält die Werte der Zeitreihen für die einzelnen Zeitpunkte bzw. Zeiträume. Dabei kann es sich um Monats- und Jahreswerte aber auch um mehrjährige Jahresdurchschnitte handeln. In der Abbildung XX ist der Datenbereich hellblau eingefärbt. Zusätzlich besteht hier die Möglichkeit, für einzelne Werte spezifische Erläuterungen und Anmerkungen hinzuzufügen. Dies wird dann vorgenommen, wenn ein bestimmter Wert sich durch ein spezielles Ereignis oder durch eine Besonderheit in der Quellenlage von den anderen Werten in der Datenreihe unterscheidet. Zellen mit Anmerkungen sind durch eine dickere Umrandung hervor gehoben. Berührt man diese Zellen mit dem Mauszeiger, erscheint ein Textfeld, in dem die Besonderheiten dieses Wertes beschrieben werden.

Tabellentitel	Tabellen-Nr. und Tabellentitel (z.B.: A.01 Bevölkerung in Deutschland nach Bundesländern und Geschlecht (1950-1980))	
Gliederungsmerkmal	Ausprägung (z.B.: Bayern)	Ausprägung
Quelle	(z.B.: Stat. Bundesamt (Hrsg.): Fachserie. Ort, Jahr, Seite.)	(z.B.: Stat. Bundesamt (Hrsg.): Fachserie. Ort, Jahr, Seite.)
Anmerkung	(z.B.: zum 31.12. des Jahres)	(z.B.: zum 31.12. des Jahres)
Tabelle	(z.B.: Tab. 03)	(z.B.: Tab. 03)
Datum (Jahr, Intervall, etc.)	Wert	Wert
	Wert	Wert
	Wert	Wert
	Wert	Wert
Datum (Jahr, Intervall, etc.)	Wert	Wert

8.4 Beispiel für die Aufbereitung der Zeitreihen einer Studie

Als Beispiel für die Aufbereitung von Zeitreihen dient uns die Studie:

Heske, G., 2005: Bruttoinlandsprodukt, Verbrauch und Erwerbstätigkeit in Ostdeutschland 1970-2000. Neue Ergebnisse einer volkswirtschaftlichen Gesamtrechnung. Historical Social Research/Historische Sozialforschung. Supplement/Beiheft No. 17. Köln: Zentrum für Historische Sozialforschung (ZA 8208).

Die Zeitreihen der Studie beziehen sich auf folgende Gebiete:

- ehemalige Deutsche Demokratische Republik 1970 bis 1989
- frühere Bundesrepublik Deutschland 1970 bis 1989
- neue Länder einschließlich Berlin 1970 bis 2000
- neue Länder ohne Berlin 1970 bis 2000
- Deutschland 1970 bis 2000."

Als Beispiel aus den insgesamt 87 Tabellen, die im Anhang der Publikation abgedruckt sind, wählen wir die Tabelle 1.2.0 „Erwerbstätige nach Wirtschaftsbereichen, in 1000 Personen“ aus. Die Abbildung XX zeigt die gesamte Tabelle, in der Form, wie sie in der Publikation abgedruckt ist (S. 224) und die Abb. YY zeigt einen Ausschnitt aus dem Tabellenkopf dieser Tabelle.

Deutsche Demokratische Republik

1.2.0 Erwerbstätige nach Wirtschaftsbereichen, 1000 Personen

Jahr	Insgesamt	Land- und Forstwirtschaft, Fischerei	Produzierendes Gewerbe ohne Baugewerbe		Baugewerbe	Handel, Gastgewerbe und Verkehr				Finanzierung, Vermietung und Unternehmensdienstleistungen			Öffentliche und private Dienstleister	
			zusammen	darunter: Verarbeitendes Gewerbe		zusammen	davon:		zusammen	davon:		zusammen	darunter:	
							Handel, Reparatur von Kfz und Gebrauchsgütern	Gastgewerbe		Verkehr und Nachrichtenübermittlung	Kredit- und Versicherungsgewerbe			Grundsteuerwesen, Vermietung, Unternehmensdienstleistung
1970	8736	950	3576	3266	656	1510	899	103	508	164	61	103	1882	663
1972	8906	925	3604	3285	657	1504	902	95	507	168	61	107	1948	688
1975	8998	851	3645	3306	669	1495	906	83	506	177	60	118	2161	776
1978	9331	865	3736	3386	690	1643	924	85	634	193	61	132	2218	792
1979	9416	852	3769	3415	699	1658	930	86	642	197	61	137	2241	796
1980	9471	845	3789	3431	706	1669	937	87	646	202	61	141	2260	803
1981	9533	849	3810	3449	702	1675	935	93	647	204	60	143	2294	815
1982	9596	862	3825	3461	700	1681	932	99	650	204	60	145	2324	826
1983	9664	880	3832	3467	705	1688	925	108	655	205	59	145	2356	842
1984	9704	882	3848	3479	701	1693	924	113	656	203	59	144	2378	851
1985	9729	884	3843	3472	692	1703	929	117	658	205	58	147	2402	859
1986	9752	884	3842	3472	692	1703	929	117	657	207	58	149	2424	869
1987	9792	885	3839	3464	696	1725	940	121	664	210	59	152	2447	877
1988	9823	884	3837	3460	682	1736	946	123	667	213	59	154	2471	885
1989	9747	874	3704	3421	674	1721	937	122	662	211	58	153	2473	886

Jahr	Insgesamt	Land- und Forstwirtschaft, Fischerei	Produzierendes Gewerbe ohne Baugewerbe		Baugewerbe	Hotel, Gastgewerbe und Verkehr			
			zusammen	darunter:		zusammen	davon:		
				Verarbeitendes Gewerbe			Handel, Reparatur von Kfz und Gebrauchsgütern	Gastgewerbe	Verkehr und Nachrichtenübermittlung
1970	8738	950	3576	3266	656	1510	899	103	508

Ausschnitt aus dem Tabellenkopf der Tabelle 1.2.0

Es handelt sich hierbei um eine sogenannte mehrdimensionale Tabelle. Einige Variablen, wie z.B. das **Produzierende Gewerbe ohne Baugewerbe**, werden noch einmal untergliedert in eine Teilgruppe, die speziell aufgeführt wird, hier das **verarbeitende Gewerbe**. Weiterhin bedeutend für die Darstellung der Tabelle in histat ist der Kontext, in dem die Tabelle in der Publikation wiedergegeben wurde. Die Datentabelle hat in der Publikation eine Überschrift (**Erwerbstätige nach Wirtschaftsbereichen, 1000 Personen**) und eine Tabellenummer (**1.2.0**) erhalten. Sie ist auf der **Seite 224** abgedruckt worden. Die regionale Gruppierung (**Deutsche Demokratische Republik**) steht in der Publikation über der Tabellenüberschrift. Der Name des Autors (**G. Heske**) und der Titel der **Publikation** müssen ebenfalls bei der Umsetzung in das histat-Tabellenformat berücksichtigt werden. Wenn erläuternde Anmerkungen zu dieser Tabelle im Text zu finden sind, müssen auch diese Anmerkungen in den Datenkennzeichnungsbereich der histat-Tabelle aufgenommen werden. Auch die in der linken Spalte stehenden **Jahresangaben** gehören zum Datenkennzeichnungsbereich und müssen übernommen werden.

Die in Excel vorliegende Datentabelle wird nun in eine histat Datentabelle, die, wie bereits erwähnt, einen ausführlichen Datenkennzeichnungsbereich aufweist, transformiert. Diese Datentabelle hat die in Abb. XX dargestellte Struktur. Notwendig ist hierbei, dass jede einzelne Zeitreihe der Datentabelle genau gekennzeichnet ist und sämtliche Informationen zur Herkunft der Zeitreihen-Werte enthält.

Titel der Tabelle (Tabellenummer + sachliche, räumliche und zeitliche Bestim- mung des Tabellenin- haltes)	Z.B. A.1.0 Die Bevölkerung nach Altersgruppen in den deutschen Ländern (1871-1900) oder: 1.2.0 Erwerbstätige nach Wirtschaftsbereichen in der ehemaligen DDR (1978-1989)		
Variable (Sachverhalt)	z.B. Bevölkerung	z.B. Umsatz	z.B. Erwerbstätige
Ggf. Gliederung des Sachverhaltes in Teil- oder Untergruppen	z.B. Geschlecht (oder Altersgruppen, oder in einzelnen Regionen, etc.)	z.B. Wirtschaftsbereiche	z.B. Wirtschaftsbereich
Maßeinheit	Spezifizierung der Men- genangaben, z.B.: pro qm; in 1000; in %	Spezifizierung der Mengenangaben, z.B.: in 1000 DM; oder Index: 1913=100	z.B.: in 1000 Personen
Quelle	Publikation (Datenherkunft)	Publikation (Datenherkunft)	Hier: G. Heske (2005): Bruttoinlandsprodukt, ..., S. 224.
Anmerkung (Hinweise, die für das Verständnis der Daten wichtig sind)	z.B.: Einwohner am 31.12. des jeweiligen Jahres; oder: Jahresdurchschnitte oder: bis 1952 ohne das Saarland.	z.B.: Für 1950 geschätzt; oder: Für 1960 ohne Betriebe des Gastgewerbes	In diesem Fall: Keine Anmerkungen
Tabelle	Tabellenummer der Publikation	Tabellenummer der Publikation	Tab. 1.2.0
Zeitangabe (zeilenweise)	Wert der Zeitreihe (zeilenweise)	Wert der Zeitreihe (zeilenweise)	8738

Nach Überführung in das histat-Tabellenformat hat Tabelle 1.2.0 aus der Arbeit von Heske (2005) folgendes Aussehen. ⁴

⁴ Hier sind zum Zwecke der besseren Lesbarkeit nur die Jahresspalte und die ersten 9 Zeitreihen abgebildet.

9 Ablauf der Eingabe

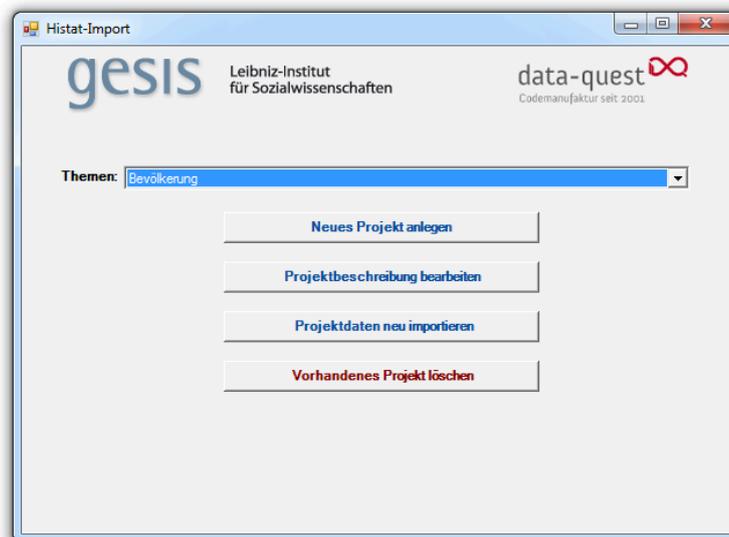
9.1 Formulare

Die aufbereiteten Studien werden unter Verwendung eines speziell für den Datenimport programmierten VB.net Moduls importiert. Die Studiendetails werden in einem Worddokument angefertigt und per copy and paste in ein vb.net-Formular übertragen, die Datenaufbereitung erfolgt in Excel. Die Exceltabellen werden von dem VB.net Modul via ODBC in eine MySQL-Datenbank überführt. In dem Auswahlménü des Import-Programms stehen vier Auswahloptionen zur Überführung einer Einzelstudie in histat zur Verfügung:

Auswahlménü:

Import einer Studie

(innerhalb der DB als
„Projekt“ bezeichnet).



Textfelder der Eingabe- seite 1:

Studententitel, Primärforscher und

Studienbeschreibung.

Histat-Import

Projekt:

Autor:

Beschreibung: Die wesentliche Zielsetzung der Datensammlung zur Bevölkerungsstatistik besteht darin, die Entwicklung der natürlichen und der geographischen Bevölkerungsbewegungen in der modernen deutschen Geschichte zusammenfassend nachzuzeichnen. Bevölkerung ist die nach biologischen Merkmalen (z. B. Geschlecht, Alter) und sozialen Gesichtspunkten (z. B. Familienstand, Beruf, Religionszugehörigkeit usw.) gegliederte Bevölkerungszahl innerhalb eines bestimmten, abgegrenzten Raumes zu einem bestimmten Zeitpunkt (Bevölkerungsstand, z. B. Einwohnerzahl zu einem bestimmten Zeitpunkt bzw. Stichtag oder die mittlere Bevölkerung eines bestimmten Jahres). Die Bevölkerungsbewegung bildet die Veränderung von Zahl und Struktur der Bevölkerung mit den Komponenten „generatives Verhalten“ und „Wanderungen“ ab. Das prozesshafte Geschehen wird wesentlich bestimmt durch die Elemente der Fruchtbarkeit und der Sterblichkeit (Heiratshäufigkeit, Geburtenhäufigkeit, Gestorbene). Neben der natürlichen Bevölkerungsbewegung ist die geographische Bevölkerungsbewegung zu berücksichtigen, d. h. die Zu- und Abwanderungen (Wanderungsbewegungen). Die Datensammlung enthält neben den absoluten Eheschließungs-, Geborenen- und Sterbefallzahlen auch die jeweils berechneten Verhältniszahlen, z. B. die Eheschließungs-, Geborenen- und Sterbeziffer sowie die Wanderungsbilanzen. Diese elementaren Beziehungszahlen besagen, wie viele Eheschließungen, Geborene oder Gestorbene während eines Zeitraumes (meist eines Jahres) auf je 1000 der mittleren Bevölkerung kamen. Die Datensammlung umfasst im einzelnen folgende Themenschwerpunkte:

Zeit:

Letzte Änderung:

Zeitreihen:

Textfelder der Eingabe- seite 2:

Angabe der Publikation,
in der die Daten verwendet werden,

Raumbezug der Daten,
verwendete Quellentypen.

Histat-Import

Veröffentlichung:

Publikationsjahr:

Untersuchungsgebiet: Die geographische Gesamtheit der Datensammlung ist das Deutsche Reich in den Grenzen von 1871 bzw. dessen Vorläufer (Deutscher Bund ohne Österreich, Deutscher Zollverein) und Nachfolgestaaten (Bundesrepublik und ehemalige DDR). Im Mittelpunkt der Datensammlung steht die Bevölkerung Deutschlands, d. h. es geht nicht um die demographische Entwicklung der deutschen Regionen, Staaten, Länder oder Provinzen. Sie werden in dieser Datenbank lediglich in Übersichtstabellen dargestellt, um die Entwicklung der Gesamtbevölkerung zu verdeutlichen. Das ZHSF - Datenarchiv verfügt über ergänzende Daten zur Bevölkerungsstatistik, die regional

Quellen: Die Datengrundlage stellt ausnahmslos die amtliche Statistik dar, d. h. die Angaben stammen zum größten Teil aus den Erhebungen der Statistischen Ämter. Als primäre Quelle dienen für die Zeit bis 1945 die Statistischen Jahrbücher des Deutschen Reiches und für die Zeit ab 1945 die Statistischen Jahrbücher für die Bundesrepublik Deutschland. Grundlegende Statistiken (Ergebnisse von Volkszählungen und statistische Jahrbücher) sind enthalten in: Königlich Statistisches Preussisches Bureau (Hrsg.): Preussische Statistik. Statistisches Amt des deutschen Reichs (Hrsg.): Statistik des deutschen Reichs (Abk.: StDR).

Textfelder der Eingabeseite 3:

DHS – Bearbeitungsvermerke.

The screenshot shows a window titled 'Histat-Import' with the following fields:

- ZA_Studiennummer: 8171
- Datum der Archivierung: Dezember 2002
- Datum der Bearbeitung: Juni 2004
- Bearbeiter im ZA: Birgit Exner/Jürgen Sensch
- Bemerkungen: Version 1.0.0.
- Zugangsklasse: A
- exportierbar:

Buttons at the bottom: << Zurück, Abbrechen, Weiter >>

Textfelder der Eingabeseite 4:

Ausführliches Quellenverzeichnis, Kommentare zur Studie insgesamt und ausführliche Anmerkungen zu einzelnen Zeitreihen.

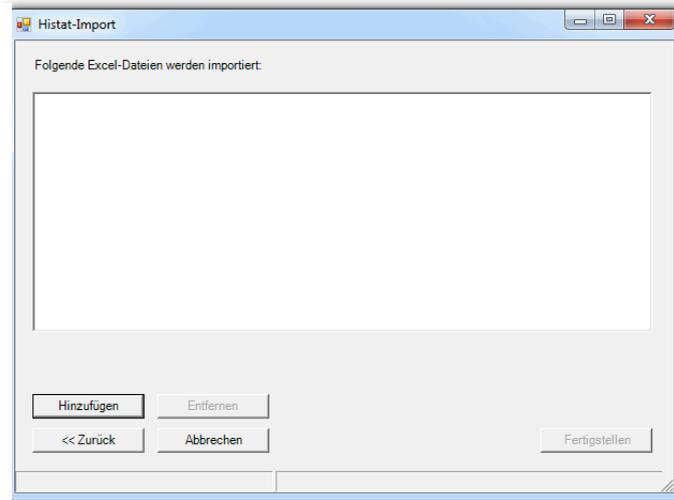
The screenshot shows a window titled 'Histat-Import' with the following content:

- Fundort:** Deutsches Reich: 1840 - 1875: Statistik des Deutschen Reichs, N.F., Bd. 44, Berlin 1892. Jahrbücher für Nationalökonomie und Statistik, hrsg. von Bruno Hildebrand, Jg. 1869, Bd. 12, Jena 1869; Jg. 1870, Bd. 14, Jena 1870. Viehbach, G., Statistik des zollvereinten und nördlichen Deutschland, Bd. 2, Berlin 1862. Vierteljahreshefte zur Statistik des Deutschen Reiches für das Jahr 1873, Jg. 1 = bd. 2 zur Statistik des Deutschen Reiches, Berlin 1874.
- Anmerkungen:** Lange Reihen können nur mit Vorbehalten und Einschränkungen gebildet werden. Bei der Beurteilung und Vergleichbarkeit der hier nachgewiesenen Angaben ist daher insbesondere der regionale Bezug zu beachten. Die in den Tabellen enthaltenen Angaben beziehen sich bis 1945 im Allgemeinen auf das Deutsche Reich und nach 1945 auf die Bundesrepublik Deutschland. Die absoluten Zahlen sind somit wegen der unterschiedlichen regionalen Abgrenzung nicht miteinander vergleichbar. Selbst wenn Angaben vor 1945 für bestimmte Bevölkerungsgrößen auf das Gebiet der heutigen Bundesrepublik umgerechnet wurden, so ist immer noch kein
- Sachliche Untergliederung der Datentabellen:** A. Bevölkerungsstand und Bevölkerungsstruktur
 A.1 Bevölkerungsgröße und Bevölkerungswachstum
 A1.01 Bevölkerungsdichte - Deutsches Reich und Bundesrepublik Deutschland (1816-2001)
 A1.02 Bevölkerung zum Zeitpunkt der Volkszählungen nach Geschlecht, Familienstand und Alter (1871-1939)
 A1.03 Bevölkerungsgröße, Bevölkerungswachstum und Bevölkerungsdichte auf dem Gebiet der
- File selection: Datei auswählen..

Buttons at the bottom: << Zurück, Abbrechen, Fertigstellen

Importmenü für die Auswahl der Excel – Datentabellen:

Pfadangabe und Markierung der Excel – Dateien, die der Studie zugeordnet sind.



9.2 Erläuterungen der Formularfelder

- (1) Neues Projekt anlegen: Studiendetails und Excel-Datentabelle(n) einlesen.

Durch die Wahl dieser Option wird eine neue Studie hinzugefügt.

- (2) Projektbeschreibung bearbeiten:

Sollen lediglich die Textteile in einer korrigierten (oder ergänzten) Version neu eingeladen werden, dann erfolgt durch Auswahl der entsprechenden Eingabeoption kein Datenimport.

- (3) Projektdaten neu importieren:

Diese Option bewirkt ein Überspringen der Textfelder für die Studienbeschreibung. Sind bereits eingeladene Daten lediglich zu ergänzen oder sind korrigierte Daten einzuladen, dann kann der komplette Textteil durch die Wahl der entsprechenden Eingabeoption übersprungen werden. Allerdings muss das gesamte Tabellenset (d.h. sämtliche Untergliederungen einer Studie) neu eingeladen werden.

- (4) Vorhandenes Projekt löschen:

Die Studiendetails und sämtliche Excel-Datentabellen (Untergliederungen der Studie) werden gelöscht.

Ein Projekt (d.h. eine Studie) kann mehrere Excel-Datentabellen umfassen, die jeweils durch den Tabellentitel (Untergliederung; in der Datenbank = Hauptschlüssel) eindeutig definiert sind. Der Projekttitel beinhaltet dagegen die Zuordnung sämtlicher Datentabellen zu einer Studie.

Mit der Option: "Exportierbar" kann die Verwendung (bzw. der Zugang) der jeweils angezeigten Daten eines Projektes gesteuert werden. Ankreuzen bedeutet: die angezeigten Daten können von dem Benutzer direkt auf seinen lokalen PC importiert werden.

Hinzufügen: Als letztes Menü folgt die Auswahl der Excel-Datentabellen unter der Option: "Hinzufügen" mit der Angabe des Pfades und der Auswahl (Markierung der Dateien) der dem Projekt zugeordneten Excel-Datentabellen.

Datei auswählen: Zusätzlich wird in der Datenbank eine von den Studiendetails abgetrennte pdf-Datei mit weiteren Dokumenten (Methodenbericht, evtl. weitere Forschungsberichte, etc.) hinterlegt, die unterhalb des Namens dieser Studie zum Download angeboten wird.

10 Einbettung in das GESIS-Datenkonzept

Der GESIS-Datenarchiv Arbeitsbereich Historische Studien (DHS) erbringt damit „ wesentliche intra- und interdisziplinäre Integrationsleistungen, indem (er) mit seinem differenzierten Serviceangebot den Daten- und Methodentransfer zwischen den historischen und sozialwissenschaftlichen Disziplinen unterstützt...“. Darüber hinaus „werden historische Daten mehr und mehr zu einer zusätzlichen Quelle und einem interessanten Testfeld für die Überprüfung von sozialwissenschaftlichen Theorien“⁵

10.1 DBK-Integration

GESIS bietet auf seiner Homepage den Datenbestandskatalog, kurz DBK, an.⁶ Hier sind sämtliche Studienbeschreibungen zu den Studien der Umfrageforschung sowie den Studien der Sozial- und Wirtschaftsgeschichte des GESIS-Datenarchivs enthalten. Was also im Rahmen der historischen Studien mit Zeitreihen unter dem Oberbegriff ‚Studiendetails‘ erscheint, wird auch in den Datenbestandskatalog von GESIS überführt. Der DBK ermöglicht eine einfache oder erweiterte Volltextsuche in dem gesamten Studienbestand und bietet darüber hinaus für einen großen Teil der Studien Verlinkungen aus der Studienbeschreibung zu ihren Daten an. Im Wesentlichen sind folgende Felder im Datenbestandskatalog enthalten:

Titel: Titel der Studie.

Weitere Titel: Zusätzliche Titel, etwa originalsprachliche Titel der Studie.

Studiennummer (4stellig)

Erhebungsjahr: Jahr der Feldzeit der Untersuchung. Hier ist eine standardisierte Angabe sowie eine freie Angabe möglich. Abweichend von den Umfragestudien wird für historische Studien die Untersuchungsperiode angegeben, also der (historische) Zeitraum, auf die sich die Daten beziehen.

Primärforscher/Wissenschaftlicher Beirat: Namen und Institutionen des/der Primärforscher der Studie.

Inhalt: Inhaltliche Beschreibung der Studie – Abstract. Für Studien, die in histat integriert werden, wird die Studienbeschreibung sowie die sachliche Untergliederung der Studie in das Inhalts-Feld des DBK eingeladen. Die sachliche Untergliederung gibt Auskunft über die erhobenen Daten zu dieser Studie und bietet dem Leser tiefere Informationen zum Datenteil der Studie.

Untersuchungsgebiet: Geografische Einheit, die der Auswahl zugrunde liegt, entweder aufgelistet nach Codes der Norm ISO3166-1/-2 oder als Freitext. Für historische Studien ist die Frei-

⁵ Aus der WGL-Stellungnahme zur GESIS; 2. Sept. 2005, Anlage B: Bewertungsbericht, S. B-6.

⁶ WEB: <http://www.gesis.org/unser-angebot/recherchieren/datenbestandskatalog/>

textangabe sehr wichtig, da aufgrund der vielen deutschen Staaten und der mannigfaltigen Gebietsänderungen in der Geschichte Deutschlands und seiner Staaten eine Standardisierung nicht möglich ist. Es wurde daher eine sehr grobe Einteilung in Anlehnung an die ISO-Norm vorgenommen, die aber immer ergänzt werden muss durch die Angaben im Freitext-Format. Die Grobfassung durch die ISO-Norm umfassen: DQDE: Historische deutsche Gebiete bis 1871, DXDE: Deutsches Reich (1871-1945), DDDE: ehemalige DDR (1945-1990), DE: ehemalige BRD 1945-1990 und Deutschland ab 1990.

Grundgesamtheit und Auswahl: Grundgesamtheit für die Studie und Auswahlverfahren.

Erhebungsverfahren: Typ des Verfahrens. Für historische Studien können Zufallsstichproben aus Aktenbeständen als Erhebungsverfahren genannt werden. Oft muss jedoch der Forscher aus einer Vielzahl von Dokumenten die Daten zusammenstellen. Die Vorgehensweise und die Angabe der verwendeten Quellen, welche der Primärforscher für die Erhebung seiner Daten herangezogen hat, werden genannt. Wenn die Liste der Quellen und Archive zu lang ist, wird sich auf die Angabe der Quellentypen beschränkt.

Analyse-System: (Datensatz) Anzahl der Einheiten; Anzahl der Variablen / Daten-Typ / Analyse-System(e). Für historische Studien mit Zeitreihendaten wird für die Anzahl der Einheiten die Anzahl der gemessenen Zeitpunkte angegeben. Die Anzahl der Variablen entspricht der Anzahl der erhobenen Zeitreihen.

Veröffentlichung: mit dem Datensatz verbundene Veröffentlichungen.

Weitere Hinweise: z.B. weitere Studien.

Inhaltliche Kategorie: Die in GESIS-Datenarchiv archivierten Studien können inhaltlichen Gruppen zugeordnet werden, wie z.B. Staat, Politische Ideologie, Politische Parteien, Arbeit und Beruf, Familie, etc. Eine entsprechende Übersicht der inhaltlichen Kategorien kann auf der Seite des DBKs aufgerufen werden. Im Falle historischer Studien lautet die inhaltliche Kategorie: „Daten historischer Studien“. Recherchiert man nach dieser inhaltlichen Kategorie, erhält man eine Liste mit sämtlichen historischen Studien des GESIS-Datenarchivs.

Zugangskategorie: Die Zugangskategorien regeln, nach welchen Kriterien die Daten für Sekundäranalysen zugänglich sind. Hier werden die Wünsche der Primärforscher und Datengeber berücksichtigt, falls sie einen beschränkten Zugang wünschen oder informiert werden möchten, wer ihre Daten genutzt hat. Studien, welche in histat downloadbar sind, wurden von den Datengebern ohne Beschränkungen für die Nutzung dritter freigegeben. Dies entspricht der Zugangskategorie A oder O.

„Daten & Dokumente, Datenzugang“: Im Fach ‚Datenzugang‘ wird aus dem DBK ein direkter Link zur Studie in histat angeboten. Es wird genau diese Studie in histat geöffnet, so dass der Nutzer die Möglichkeit hat, die Studiendaten downzuloaden.

10.2 Versionierung

Mit Bereitstellung der Studie für die Forschung wird eine Versionsnummer vergeben. Diese Versionsnummer, die im Datenbestandskatalog angegeben wird, informiert den Nutzer darüber, ob, wann und zu welchen Inhalten genau sich die Daten einer Studie verändert haben. Der zur Studie gehörende Datenbestand wird als ‚Analyse-Datensatz‘ gesichert, d.h. die dem Nutzer angebotenen Daten enthalten die notwendigen Metainformationen (Variablendefinitionen und Labels). Werden hier Veränderungen oder Ergänzungen vorgenommen, ermöglicht die Vergabe von Versionsnummern sowie die damit einhergehende genaue Dokumentation der vorgenommenen Veränderungen eine detaillierte Information des Nutzers, die unter der Überschrift ‚Errata & Versionen‘ im DBK eingesehen werden kann. Die Versionsnummer setzt sich aus drei Ziffernblöcken zusammen, über welche die verschiedenen Ebenen einer vorgenommenen Veränderung wiedergegeben werden. Es wird zwischen folgenden Stufen einer Veränderung unterschieden:

Revision: Bei Veränderungen dieser Stufe handelt es sich um nicht bedeutungsrelevante kleine Korrekturen. Dies können z.B. Korrekturen kleiner Rechtschreibfehler im Variablennamen sein, ohne dass sich dabei die inhaltliche Bedeutung der Variable ändert.

Minor: Veränderungen im Minor-Bereich betreffen Korrekturen, die auch die Bedeutung der betreffenden Variable tangieren, wie es z.B. durch eine durchgeführte Recodierung einer Variable der Fall sein kann.

Major: Werden einer Datendatei neue Daten hinzugefügt, wie es z.B. bei Studien mit mehreren Umfragewellen der Fall ist, oder werden Variablen gelöscht, dann handelt es sich um eine tiefgreifende Veränderung, welche der Stufe Major zugeordnet wird.

Überträgt man diese Revision-Minor-Major-Stufen auf Studien mit Zeitreihendaten, so ergibt sich folgendes Schema:

Als Revision werden Veränderungen im Kennzeichnungsbereich der Datentabellen bezeichnet, die keine Auswirkungen auf die Inhaltliche Bedeutung der Zeitreihe hat. Es handelt sich um Korrekturen von Schreibfehlern im Tabellenkopf der Datentabellen (Titel, Variable, Quelle, Anmerkung).

Mit Minor werden Veränderungen in den Datentabellen benannt, die sich auf den Bedeutungsgehalt der betroffenen Zeitreihe auswirken. Es handelt sich dabei um Änderung der Anmerkungen zu einer Zeitreihe sowie zu einzelnen Werten in der Zeitreihe. Auch die Korrektur einzelner Werte in einer Zeitreihe zählt noch hierzu.

Major betreffen Veränderungen, die das Hinzufügen neuer Zeitreihen oder ganzer Datentabellen betreffen.

Die Versionsnummer zählt diese Veränderungen wie folgt: Sie besteht zunächst aus drei Ziffernblöcken, die jeweils durch einen Punkt voneinander getrennt sind: 1.0.0. Der letzte Ziffernblock zählt die einfachen Veränderungen und Korrekturen, die in den Bereich der Revision fallen. Der mittlere Ziffernblock gibt Minor-Veränderungen wieder und der erste Ziffernblock zählt

die tiefgreifenden Veränderungen des Typs Major, so dass ein Datensatz der Version 2.0.0 inhaltlich deutlich erweitert oder verändert wurde.

Diese Versionierung wird in den Studiendetails von histat ebenfalls erfasst. Es wird die Versionsnummer übertragen und die genaue Beschreibung der Versionierung findet sich in den Studiendetails am Ende der Studienbeschreibung unter der Überschrift ‚Hinweise zur neuen Version:‘. Danach folgt die Angabe der neuen Versionsnummer, das Datum der Änderung und die Beschreibung der Veränderung.

Als Beispiel sei hier die Studie ZA8051 von Pfetsch zu der staatlichen Finanzierung der Wissenschaft in Deutschland in der Zeit von 1850 bis 1975 angegeben:

Hinweis zur neuen Version:

Version 2.0.0.

Datum der Änderung: 28.10.2010.

Ergänzung der Tabellen in der Untergliederung B.

Vollständig neu aufgenommene Untergliederungen: A, C.

Zusätzlich wird intern für Archivzwecke eine Liste zu den Studienversionierungen gepflegt, in der, geordnet nach Studiennummer, Datum, Inhalt und Versionsstufe (Revision-Minor-Major) sowie der Name des Mitarbeiters, der die Änderung vorgenommen hat, festgehalten werden.

10.3 da|ra-Registrierung und DOI-Vergabe

da|ra ist die Kurzform für „Registrierungsagentur für sozialwissenschaftliche Forschungsdaten“, die von GESIS seit 2010 betrieben wird.⁷ GESIS will mit dieser Infrastruktur dazu beitragen, dass Forschungsdaten dauerhaft identifiziert, gesichert, lokalisiert und verlässlich zitiert werden können. Zu diesem Zweck werden Forschungsdaten mit einem DOI-Namen versehen. DOI steht für ‚Digital Object Identifier‘ und ist ein dauerhafter persistenter Identifier, mit dem die Verlinkung von elektronischen Forschungsdaten und das Zitieren dieser Quellen sicherer gemacht wird, indem der Link URL-unabhängig auf die Ressource bzw. auf das Objekt verweist. „Darüber hinaus wird ein Objekt durch den DOI-Namen mit aktuellen und strukturierten Metadaten verknüpft. Diese werden langfristig in einer von GESIS zur Verfügung gestellten und dauerhaft gepflegten Datenbank verwaltet und dem Nutzer verfügbar gemacht.“ (WEB: www.gesis.org/dara/home/neue-nutzer/was-ist-ein-doi/)

Der DOI-Name mit dem Link wird im Datenbestandskatalog angezeigt. Jede historische Studie, die in den Datenbestandskatalog eingetragen wird, wird auch bei da|ra registriert und erhält somit einen DOI-Namen.

⁷

WEB: <http://www.gesis.org/dara/>

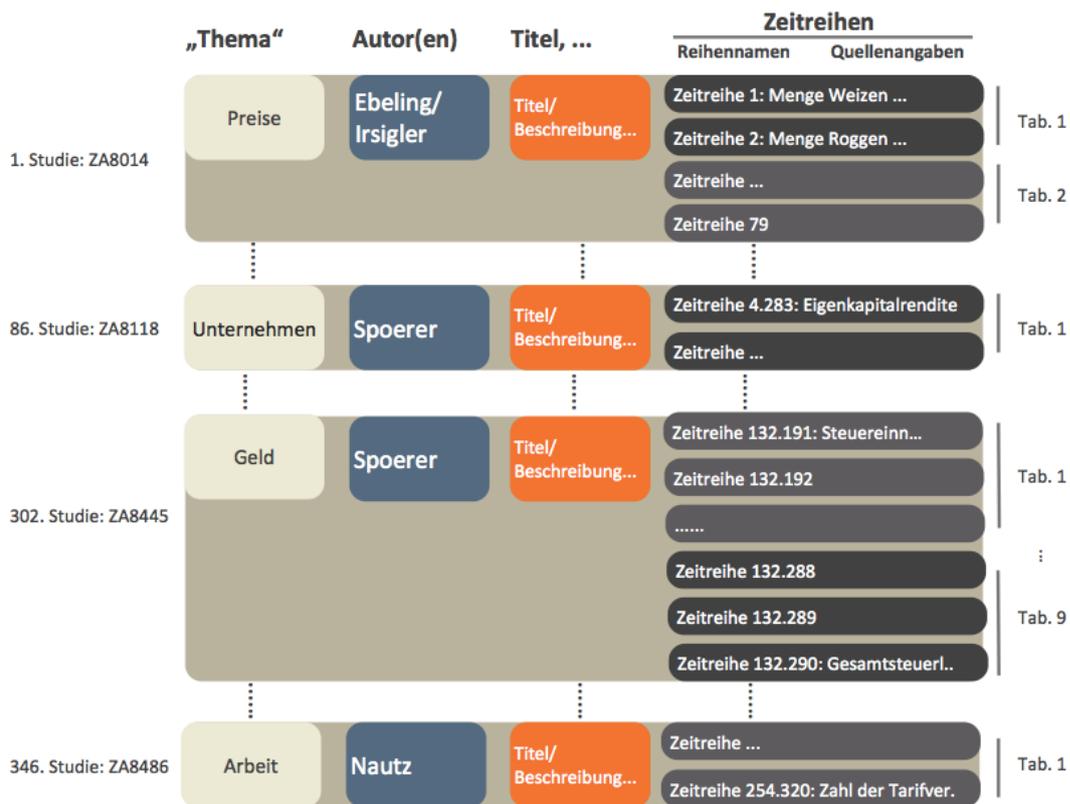
10.4 DDI-Auszeichnung

Mit der aktuellen Überarbeitung von histat wurde auch die Möglichkeit eines direkten DDI-Exports der Studienbeschreibung geschaffen (siehe oben, Abschnitt 3.6).

11 Technische Umsetzung

11.1 Logischer Datenaufbau und Datenbankstruktur

Die Datenbank enthält neben den eigentlichen Zahlenwerten der Zeitreihen die Metadaten, die die o.g. Recherchemöglichkeiten bieten. Dies sind Angaben zum Thema, dem Leiter bzw. der Leiterin der Studie, der Name der Studie, eine Studienbeschreibung, eine Nennung des Zeitraums, die Anzahl der Zeitreihen, Angaben zu Veröffentlichungen, verwendeten Quellentypen, zur sachlichen Untergliederung in einzelne Tabellen sowie Angaben zu Quellen und Anmerkungen (sofern nicht in einem separaten Dokument), schließlich Bearbeitungshinweise. Der Aufbau der Oberfläche (nicht der Daten) wurde konzeptionell so festgelegt, dass sich die Suchmöglichkeiten dabei immer auf eine Studie beziehen. Wird also nach Autoren, Themen oder Angaben zu Zeitreihen etc. gesucht, werden als Ergebnis immer die Studien zurückgegeben, in denen diese Informationen enthalten sind. Die Zeitreihen sind dabei immer einer Tabelle zugeordnet. Das folgende Schema veranschaulicht diesen logischen Aufbau der Metadaten.



Die Meta-Information sind in der Datenbank in den acht Tabellen Aka_Themen, Aka_Projekte, Aka_Schluesselmaske, Aka_Codes, Aka_CodeInhalt, Aka_SchluesselCode, aka_Schluesselindex und Lit_ZR abgelegt.

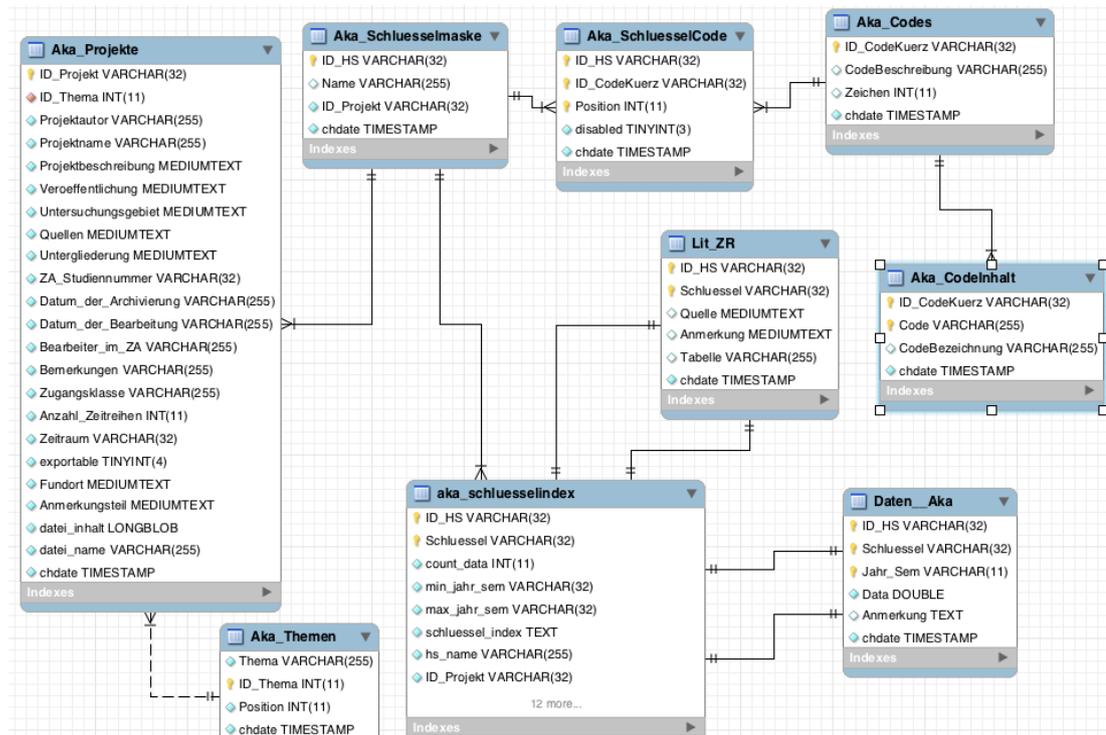


Tabelle	Aka Themen
Inhalt	Thema
Primärschlüssel (technisch)	ID Thema
Sekundärschlüssel (Suche)	Thema
Beispiel	Bildung

Tabelle	Aka Projekte
Inhalt	Beschreibung einer Studie
Primärschlüssel (technisch)	ID Projekt
Sekundärschlüssel (Suche)	Projektname
Beispiel	Berufliche Schulen und Hochschulen in der Bundesrepublik

Tabelle	Aka Schluesselmaske
Inhalt	Hauptschlüssel (Untergliederung, "Tabelle")
Primärschlüssel (technisch)	ID HS
Sekundärschlüssel (Suche)	Name
Beispiel	Das Alter der Studierenden

Tabelle	Aka Codes
Inhalt	Variable
Primärschlüssel (technisch)	ID CodeKuerz
Sekundärschlüssel (Suche)	CodeBeschreibung
Beispiel	Hochschulart

Tabelle	Aka CodeInhalt
Inhalt	Ausprägungen der Variablen
Primärschlüssel (technisch)	ID CodeKuerz + Code
Sekundärschlüssel (Suche)	CodeBezeichnung
Beispiel	Wiss. Hochschulen insgesamt

Tabelle	Aka SchluesselCode
Inhalt	Zusammenstellung und Reihenfolge der Variablen eines Haupt-
Primärschlüssel (technisch)	ID HS + ID CodeKuerz + Position
Sekundärschlüssel (Suche)	-
Beispiel	Das Alter der Studierenden = Hochschulart + Staatsangehörig-

Tabelle	aka_schluesselindex
Inhalt	Metadaten und Suchbegriffe einer Zeitreihe
Primärschlüssel (technisch)	ID HS + Schluessel
Sekundärschlüssel (Suche)	schluessel index
Beispiel	Hochschulart: Wiss. Hochschulen insgesamt, Staatsangehörig-

Tabelle	Lit ZR
Inhalt	Quellen und Anmerkungen zu einer Zeitreihe
Primärschlüssel (technisch)	ID HS + Schluessel
Sekundärschlüssel (Suche)	Quelle + Anmerkung
Beispiel	1954 WS-1959 WS: DHB-VIII

Tabelle	Daten Aka
Inhalt	Datenwerte
Primärschlüssel (technisch)	ID HS + Schluessel + Jahr Sem
Sekundärschlüssel (Suche)	Anmerkung
Beispiel	1954 WS: 1146

Aus den beiden Spalten wird über Schlüssel per Skript on the fly eine Matrix mit den Spalten Jahr, Zeitreihe 1, Zeitreihe 2, usw. erzeugt, die also die Jahre der Studie als Zeilen und die einzelnen Zeitreihen als Spalten enthält. Aus

Studie	Tabelle	Reihe	Jahr	Wert
1	1	1	1851	123
1	1	2	1851	32
1	1	3	1851	3.214
1	1	4	1851	19.872
1	1	1	1852	212
1	1	2	1852	12
1	1	3	1852	4.432
1	1	4	1852	75.643
1	2	1	1851	32
1	2	2	1852	765
1	2	3	1851	4.598
1	2	1	1852	38
1	2	2	1853	764
1	2	3	1852	4576
...				

wird durch das Exportprogramm im Moment des Abrufes folgende Struktur erstellt:

Tabelle 1

Jahr	Reihe 1	Reihe 2	Reihe 3	Reihe 4
1851	123	32	3.214	19.872
1852	212	12	4.432	75.643
1853
...

Tabelle 2

Jahr	Reihe 1	Reihe 2	Reihe 3
1851	32	-	4.598
1852	38	765	4.576
1853	...	764	...
...

11.2 Web-Oberfläche

Allgemeines

Bei der technischen Umsetzung des Webauftritts von HISTAT wurde das PHP Framework „Kohana“ in der Version 3.2 verwendet.

Ergänzend dazu wurden mehrere Plugins mit externen PHP Bibliotheken installiert:

- Bibliothek „Swiftmailer“. Für den flexiblen Mail Versand wurde die PHP Bibliothek „Swiftmailer“ in der Version 4.2.1 verwendet.
- Bibliothek „PHPExcel“: Für einen korrekten Export der Daten in das Excel-Format wurde die PHP-Bibliothek „PHPExcel“ in der Version 1.7.7 verwendet.
- Bibliothek „pChart“: Für die Generierung der Datendiagramme wurde die PHP Bibliothek „pChart“ in der Version 2.1.1 verwendet.
- Javascript jQuery: Bei der Oberfläche wurde das JavaScript Framework „jQuery“ in der Version 1.7 eingesetzt.

Der Einsatz eines vorhandenen PHP-Frameworks bietet etliche Vorteile, wie z.B. den Rückgriff auf bereits vorhandene Ressourcen. Dazu gehören die Dokumentation, verfügbare Erweiterungsmodule, einfache Wartung, einfachere Einarbeitung neuer Entwickler und eine große Community für die Hilfestellung.

Model View Controller

Das PHP Framework „Kohana“ basiert auf dem *Hierarchic Model View Controller* (HMVC) Design Pattern. Es ist eine Erweiterung des *Model View Controller* (MVC) Design Pattern.

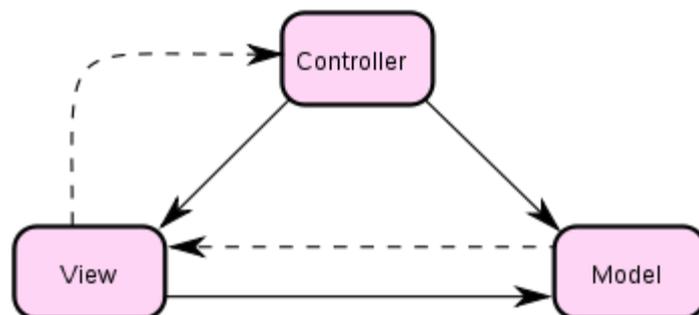


Abb. 1 MVC Design Pattern

Im MVC Design Pattern wird die Software in drei unterschiedliche Einheiten unterteilt. Präsentation (engl. View), Datenmodell (engl. Model) und Programmsteuerung (engl. Controller).

Das Datenmodell liest oder schreibt Daten aus einem Medium. Dies kann eine Datenbank, eine

CSV Datei oder eine Text Datei sein. Die Präsentation formatiert lediglich die Daten und zeigt diese an. Die Programmsteuerung dient als Kommunikationsobjekt der Präsentation und des Datenmodells. Es bestimmt welches Datenmodell und welche Präsentation geladen werden soll. Es kann Daten, die aus der Präsentation (z. B. Eingabefelder) stammen, an das Datenmodell übergeben und auch Daten aus dem Datenmodell laden und an die Präsentation übergeben.

Ziel ist es eine flexible und wiederverwendbare Software zu erstellen. Es ist zum Beispiel möglich die Models der Anwendung wieder zu Verwenden aber die Views komplett um zu gestalten so dass die Applikation auf Mobilen Endgeräten funktionieren würde.

HMVC

Das HMVC Design Pattern ist weniger strikt als das MVC Design Pattern. Es erlaubt Daten direkt vom Model an den View zu übergeben. Dabei wird ein Subcontroller verwendet, der von einem Hauptcontroller aufgerufen wird. Die Programmsteuerung kann über eine URL oder eine Schaltfläche aufgerufen werden, die wiederum weitere Programmsteuerungen mit weiteren Präsentationen und weiteren Datenmodellen aufruft. Somit können Elemente flexibel weiter verwendet werden. Beispiele für wiederverwendbare Elemente sind z.B. modale Dialoge oder Navigationselemente.

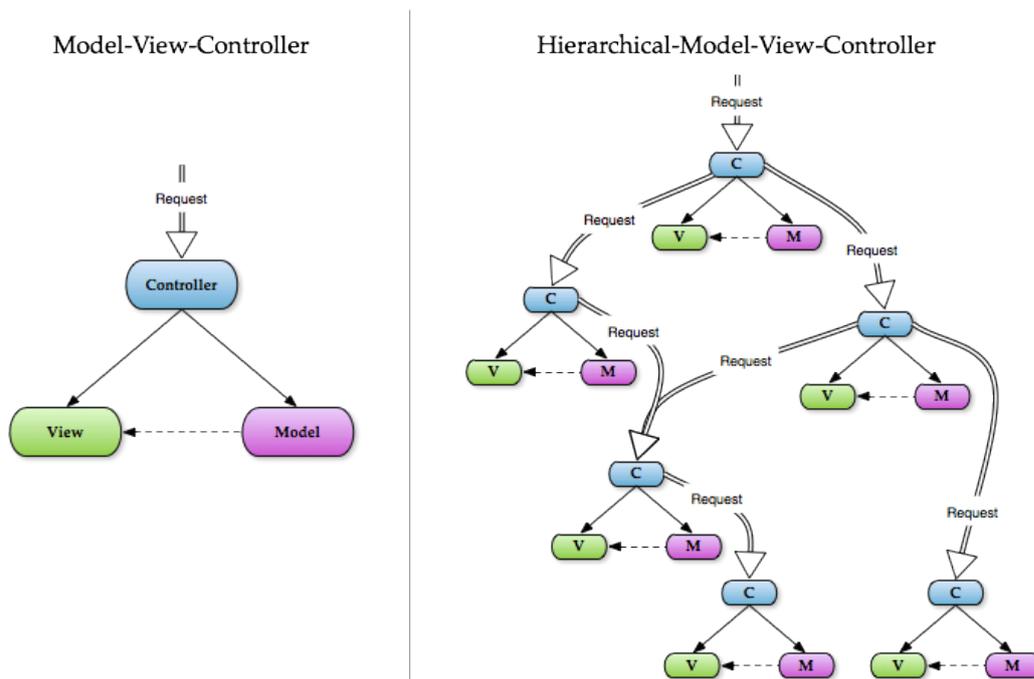


Abb. 2 MVC und HMVC Design Pattern

Auch die Entwicklung selbst wird verbessert, da der Quellcode der Controller durch den Aufruf von Sub-Controllern klein gehalten wird.

Object Relational Mapping

Das Kohana PHP-Framework verwendet Object-Relational-Mapping (ORM) im Datenmodell. Dabei werden einheitliche Methoden für das Lesen und Schreiben der Daten verwendet. Das Ziel ist es, die Models wiederverwenden zu können, unabhängig vom Medium. Man hat die Möglichkeit mittels ORM Daten zu lesen, zu manipulieren und zu schreiben, sei es von einer Datenbank oder Textdatei. Intern wird mit Hilfe des ORM Treibers, Operationen für das eingestellte Medium angepasst. Jedes Medium (MySQL, Oracle SQL, MSSql und NoSQL) hat seine eigenen Spezialitäten. Durch die Verwendung des ORMs ist man in der Lage beliebige Datenbanken auszutauschen ohne den Quellcode der Applikation verändern zu müssen.

Ordnerstruktur

Das Kohana PHP-Framework hat eine feste Ordnerstruktur. Dadurch wird es möglich, dass nur benötigte Klassen bei Bedarf zur Laufzeit nachgeladen werden.

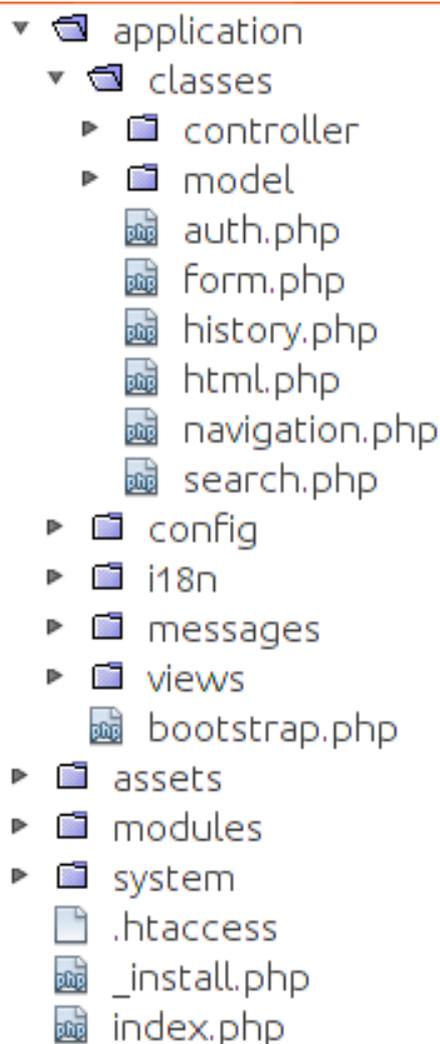


Abb. 3 Ordnerstruktur

Application

in diesem Ordner befinden sich alle Dateien für die zugehörigen Applikationen. Der Ordner hat Unterordner. Die wichtigsten sind:

- **Classes** : Hier sind alle für den Internetauftritt benötigten PHP-Klassen vorhanden. Wenn die Klasse ein Model ist, dann befindet sie sich im Ordner „Models“, wenn die Klasse ein Controller ist, dann befindet sie sich im Ordner „Controller“. Wenn die Klasse weder Model noch Controller ist, dann befindet sie sich im Ordner mit dem Namen „Classes“.
- **Config**: Im Ordner „Config“ sind alle Konfigurationsdateien zur Anwendung hinterlegt. Jede Konfigurationsdatei ist für einen bestimmten Zuständigkeitsbereich vorgesehen der in der Regel bereits vom Namen angedeutet wird. Die Konfigurationsdatei `database.php` ist bspw. zuständig für den Verbindungsaufbau der Datenbank, `email.php` ist zuständig für die Einstellung des Mailversands, etc.
- **l18n**: Dieser Ordner enthält Übersetzungsdateien für die Anwendung. Das Kohana Framework sieht drei Möglichkeiten für die Übersetzung vor: Volltext Übersetzung, Übersetzung von Variablen und Contentübersetzung. Die Volltextübersetzung soll in der Regel für die Übersetzung von Schaltflächen, Navigationselementen oder Beschreibungen eingesetzt werden. Die Variablenübersetzung soll für Meldungen, Warnungen und Informationselemente eingesetzt werden. Die Texte für Variablenübersetzung befinden sich im Ordner „messages“. Die Contentübersetzung dient zur Übersetzung längerer Texte, diese wird durch Views umgesetzt.
- **Messages**: Hier befinden sich Übersetzungstexte für Variablen.
- **Views**: Hier befinden sich die HTML-Views für die Anwendung.

Assets

Im Assets Ordner befinden sich Dateien, welche öffentlich zugänglich sind. Dazu gehören JavaScript Dateien, Stylesheets und Bilder.

Modules

Der Ordner „Modules“ enthält Kohana-Module und externe PHP-Bibliotheken, die zu den Modulen dazugehören.

System

In diesem Ordner ist das Kohana PHP-Framework enthalten. Alle Dateien werden durch das ModRewrite-Modul des Apache Webservers in der `.htaccess`-Datei auf die `index.php` umgeleitet. Aus Sicherheitsgründen kann keine der PHP-Dateien des Entwicklungsframeworks direkt über die URL aufgerufen werden.

Routing

Das Kohana Framework verwendet eine Routingmethode um Controller auszuführen. Die Standardroute ist „<sprache>/<controller>/<action>/<parameter>“. Somit lässt sich anhand der URL direkt erkennen welcher Controller aufgerufen wurde. Wenn bspw. die Adresse „/de/data/top“ aufgerufen wurde um die meistverwendeten Studien anzusehen, kann man daraus schließen, dass die Datei „/application/classes/controller/data.php“ ausgeführt wurde – die „action top“ ist eine Methode die sich in der Datei befindet.

Durch das Routing besteht zusätzlich die Möglichkeit, besondere Routen selbst zu definieren. Dies kann sinnvoll sein, wenn z.B. eine Action in einem Controller mehr als ein Parameter benötigt. Alle Routen werden in der Datei application/bootstrap.php definiert.

Anhand der URL lässt sich auch erkennen, welcher View gerade dargestellt wird. Die Ordnerstruktur ist identisch zur URL. Im Beispiel „/de/data/top“ wird die Datei „application/views/de/data/top.php“ für die Darstellung verwendet. Dadurch ist gewährleistet, dass die Applikation schnell und flexibel gewartet, verändert oder erweitert werden kann.

Authentifizierung

In der neuen Version des Internetauftritts wurde die Authentifizierung komplett überarbeitet und auf RBAC (Role Based Access Control) umgestellt.

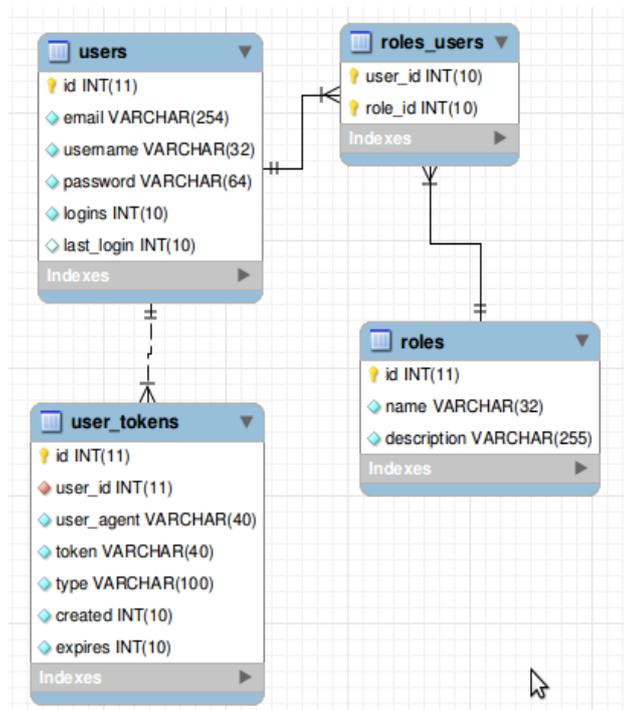


Abb.4 ERD RBAC

Beim RBAC werden verschiedene Rollen definiert. Jeder Benutzer kann mehreren Rollen zugeordnet werden. Rollen sind voneinander unabhängig und bauen nicht auf anderen Rollen auf.

Ein Nutzer mit der Rolle „Administrator“ ist zum Beispiel nicht gleichzeitig ein User, so dass man bspw. einem Benutzer die Rollen „User“ und „Administrator“ zuweisen muss, wenn er Zugriff auf alle Sichten der beiden Rollen haben soll. In der Anwendung werden nacheinander mehrere Rollen abgefragt, um bestimmte Controller aufrufen zu können. Durch die Vererbung der Controller ist man in der Lage eine Hierarchie in den Rollen aufzubauen. Zum Beispiel erbt der Controller_Admin vom Controller_User. Im Controller_User wird überprüft ob der Benutzer die Rolle „User“ besitzt, da der Controller_Admin vom Controller_User vererbt wird, findet die Überprüfung auf die Rolle „User“ automatisch statt.

Ziel dieser Umsetzung ist die einfache Verwaltung von Rollen.

Klassendiagramm

Anhand des Routing wird ein bestimmter Controller mit einer Action aufgerufen. Alle Controller-Klassen beginnen mit der Präfix „Controller_“ und jede Action beginnt mit der Präfix „action_“. Die Controller-Klassen vererben ihre Eigenschaften an die Kind Controller Klassen.

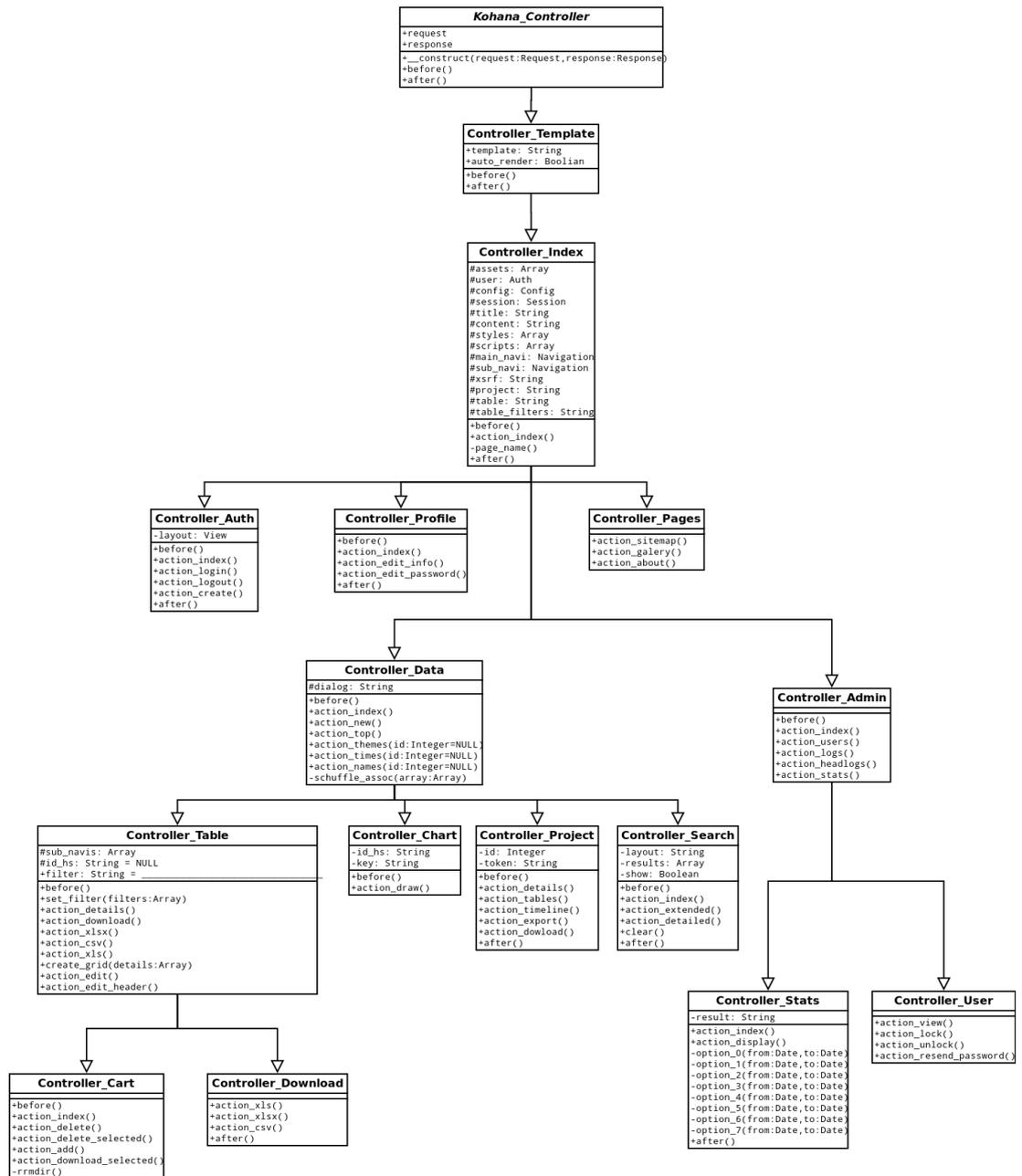


Abb.5 Controller Klassendiagramm

Klassenbeschreibung:

- Kohana_Controller : ist die Basis Controller-Klasse die von Kohana Framework zur Verfügung gestellt wird
- Controller_Template: ist der Basis Front-Controller des Kohana Frameworks. Dieser Controller bestimmt welcher View für das Hauptlayout verwendet wird.
- Controller_Index: ist der Front-Controller der Histat Applikation, in diesem Controller

werden grundlegende Eigenschaften der Histat Anwendung definiert. Alle Controller-Klassen sind von Controller_Index abgeleitet.

- Controller_Auth: ist der Authentication Controller. Der Controller hat Methoden die für den Login bzw Registrierungprozess notwendig sind.
- Controller_Profile: stellt Methoden für den Benutzerprofil zur Verfügung. Zum Beispiel das Ändern des Passworts bzw. Benutzerinformationen
- Controller_Pages: ist für den statischen Inhalt einiger Seiten zuständig. Zum Beispiel Sitemap.
- Controller_Data: ist notwendig um Übersichten und Inhalte der Unterseiten „Daten“ darzustellen.
- Controller_Table: sammelt alle Daten einer Zeitreihe zusammen und stellt diese in einer Tabelle dar. Die Daten können gefiltert werden. Außerdem bietet die Controller-Klasse Methoden um Daten für den Export bereit zu stellen.
- Controller_Chart: dient zur Darstellung der Grafik der Zeitreihe.
- Controller_Project: ist zuständig für die Darstellung der Metadaten einer Studie. Zum Beispiel Anzahl Zeitreihen oder Studien-Details.
- Controller_Search: wird bei der einfachen bzw. erweiterten Suche verwendet.
- Controller_Admin: kann nur von Benutzern mit Administrationsrechten verwendet werden. Die Controller-Klasse stellt Methoden zur Darstellung der Logs und Darstellung der Benutzer zur Verfügung.
- Controller_Stats: zeigt anhand ausgewählter Option und Zeitraum bestimmte Aktivitäten und Statistiken der Anwendung dar, zum Beispiel Anzahl der Downloads.
- Controller_User: bietet Methoden zum Sperren bzw. Entsperren der Benutzer. Gesperrte Benutzer haben keine Möglichkeit mehr sich anzumelden.
- Controller_Cart: wird verwendet für die Verwaltung des Warenkorb eines Benutzers.
- Controller_Download: ermöglicht dem Benutzer ausgewählte Zeitreihen zu exportieren.

Wie im Bereich MVC beschrieben werden die Controller-Klassen nur eingesetzt um fest zu stellen, welches Model bzw welches View aufgerufen werden soll, wenn Parameter über ein Eingabeformular bzw die URL an die Controller-Klasse übergeben werden. So erstellt der Controller eine Instanz eines Models, ruft die Methoden des Models auf, verarbeitet die Daten des Models und übergibt diese an eine Instanz des Views. In der „after()“ Methode des Controllers wird der View ausgegeben.

Models

Die Model-Klassen von Kohana sind abstrakte Abbildungen der Datenbank, durch die Variablen `$_belongs_to`, `$_has_many`, `$_has_one` ist man in der Lage Beziehungen bzw. Verknüpfungen der Tabellen zu Bestimmen.

Jede Model-Klasse hat Eigenschaften `$_table_name`, `$_primary_key`, `$_table_columns`. Die Eigenschaften müssen mit der Struktur der Datenbank übereinstimmen.

Models, die für die Speicherung der Daten zuständig sind, besitzen eine Methode „rules()“ in dieser Methode werden Validierungsregeln für Datenfelder festgelegt. Vor dem Speichern bzw. Ändern der Daten werden die Felder validiert, wenn die Validierung fehlschlägt wird eine `ValidationException` ausgelöst.

Einige Model-Klassen haben (aus Performance-Gründen) Methoden, die mit Hilfe des Kohana Query Builders, Daten aus der Datenbank auslesen und in einem Array zurückliefern.

Quellen:

Abb. 1 MVC Design Pattern: http://de.wikipedia.org/wiki/Model_View_Controller

Abb. 2 MVC und HMVC Design Pattern: <http://techportal.inviqa.com/2010/02/22/scaling-web-applications-with-hmvc>

Abb. 3 Ordnerstruktur : Data-Quest

Abb. 4 ERD RBAC: Data-Quest

Abb. 5 Controller Klassendiagramm: Data-Quest

11.3 Das Upload-Modul

Das Histat-Upload-Modul ist für die Übertragung sowohl der Zeitreihen-Daten aus formatierten Excel-Dateien als auch der Metadaten aus Benutzereingaben in die Datenbank zuständig.

ID_Projekt	ID_Thema	ID_Zeit	Projektautor	Projektname	Projektbeschreibung	Veröffentlichung	Untersuchungsgebiet	Quelle
8B503C80DF968F86DB7B8F2A2DBA570A	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: St
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 Wiesbaden (Hrsg. B...
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 Lander ...
8F59A7DAEB172680D116B4816E6120F2	10	5	Metz, Rainer	Säkulare Trends der deutschen Wirtschaft	Bei der Studie handelt es sich um eine Rekonstrukt...	Metz, Rainer: Säkulare Trends der deutschen Wirtsch...	Deutschland von 1200 bis 2002	Ausgewählte wissenschaftlich Publikationen. A
3849408141F966CF9317FC792820CD95	12	17	Kunz, Andreas; (unter Mitarbeit von Annett Laake u...	Statistik der Binnenschifffahrt in Deutschland 1835...	Die Studie ist die zweite in einer Reihe von Daten...	Kunz, Andreas (Hrsg.), 1999: Statistik der Binnens...	Die geographische Erhebungseinheit ist das Deutsch...	Das Quellenmat historischen Sta
4C210BDE5E0981BD937841C2A0B97433	5	5	Irsigler, Franz; Ebeling, Dietrich	Getreideumsatz, Getreidepreise und Brotpreise in K...	Einer modernen Preisgeschichte werden gewöhnlich z...	Ebeling, Dietrich / Irsigler, Franz: Getreideumsat...	Köln von 1368 bis 1797.	A: Die Fruchtpre von 1531-1674. B: Die Ba...
00AFD9417962E0FB1EEEC3C53C4BA7ED	17	32	Sensch, Jürgen	Erwerbstätigkeitsstatistik in der Bundesrepublik D...	Die Datenkompilation auf der Grundlage der amtlich...	Fritz, W., 2001: Historie der amtlichen Statistike...	Bundesrepublik Deutschland: Früheres Bundesgebiet ...	Über die Erhebungsprogr amtlichen Statis
697A204CD460043B167B45C004AD6FCB	10	20	Blum, Reinhard	Der Kapitalkoeffizient als Instrument der	Die Studie untersucht die Brauchbarkeit des Kapita...	Blum, R., 1971: Der Kapitalkoeffizient als	Deutschland in den jeweiligen Grenzen,	Berechnet nach Originalreihen be

Metadaten: die Tabelle Aka_Projekte in der Datenbank

Systemanforderungen

Die (aktuelle) Version des Upload-Moduls wurde mit dem Microsoft

- Visual Studio 2008, Version 9.0.30729.1 SP

erzeugt und benötigt zur Ausführung das

- .NET Framework, Version 2.0.50727 oder neuer (Version 3.0 oder 3.5).

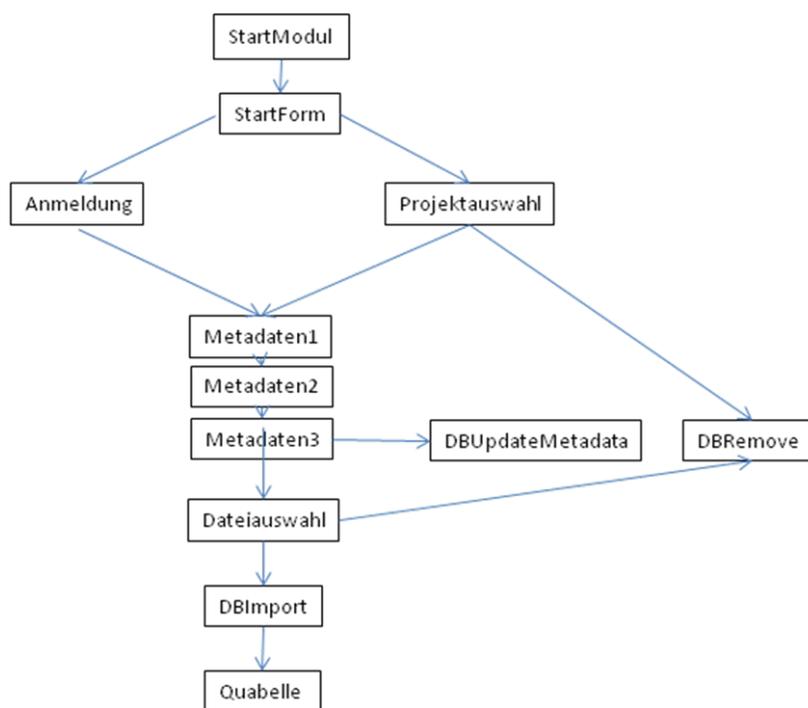
Daher ist ein Betriebssystem wie

- Windows 98, Windows Me
- Windows NT 4.0 mit Service Pack 6a
- Windows 2000
- Windows 2003
- Windows XP mit Service Pack 2
- Windows 7

erforderlich. Außerdem muss der Internet-Explorer 5.01 oder höher vorhanden sein.

Überblick

Die Software für den Histat-Upload enthält neben den Formularen *StartForm*, *Projektauswahl*, *Anmeldung*, *Metadaten1*, *Metadaten2*, *Metadaten3*, *Dateiauswahl* für die Benutzereingaben die Module *StartModul*, *Globals* für anwendungsbezogene Variablen, Konstanten und Prozeduren sowie die Module *DBImport*, *DBUpdateMetadata*, *DBRemove* für die Bearbeitung der Daten in der Datenbank und außerdem das Klassenmodul *Quabelle* zum Überprüfen und Auslesen der Excel-Daten.



Aufrufstruktur des Programms Histat-Upload

In der Datei Histat.log, die sich im Installationsordner befindet, werden erfolgreiche Aktionen und Fehlermeldungen protokolliert.

Formulare

In Abhängigkeit von der getroffenen Button-Auswahl im Formular *Startform* wird die globale Variable *Status* gesetzt und entweder das Formular *Anmeldung* (Anlegen eines neuen Projekts) oder *Projektauswahl* geöffnet. Die Formulare *StartForm*, *Anmeldung* und *Projektauswahl* enthalten Auswahlboxen. Die Zuweisung ihrer Datenquelle erfolgt jeweils in der Load-Prozedur des entsprechenden Formulars mit Hilfe eines System.Data.DataTable - Objekts.

Module

StartModul:

Vorbereiten des Schreibens der Log-Datei *HistatLog*, Anzeigen des Formulars *StartForm*.

Globals:

Dieses Modul enthält global verwendete Konstanten, Variablen, Prozeduren und Funktionen.

Die Variable *status* kann die 4 Werte *onDelete*, *onImport*, *onDataImport* und *onMetadataUpdate* annehmen und dient der Ablaufsteuerung. Sie wird vom Benutzer durch die Betätigung eines der Buttons in dem Formular *StartForm* (siehe oben) festgelegt. In der Konstanten *conStr* wird die ODBC-Datenbankverbindung gespeichert: `Public Const conStr As String = "DSN=histat"`

Die in den globalen Variablen des Moduls *Globals* zwischengespeicherten Benutzereingaben werden von den Modulen *DBImport*, *DBUpdateMetadata*, *DBRemove* als Parameter für die DML - Anweisungen benutzt. Außerdem sind in *Globals* folgende öffentliche Funktionen deklariert: `Public Function checkXITabs(ByRef anz_Zr As Int32, ByRef zeitraum As String) As Boolean`

Aufgabe: Überprüfung der Excel-Dateien und Berechnung der Anzahl der Zeitreihen und des Zeitraums. Output-Parameter: *anz_Zr* - Anzahl der Zeitreihen, *zeitraum* - Zeitraum zwischen dem kleinsten und größten Jahreseintrag in der Form von - bis, Rückgabewert: im Fehlerfall *false*, sonst *true*.

Aufgerufen von: *Dateiauswahl.cmdOK_Click*. Die Funktion verwendet ein *Quabelle*-Objekt; die Kriterien der Überprüfung sind durch dessen Methode *test* gegeben.

`Public Function CreateUniqueKey(ByVal md5 As MD5CryptoServiceProvider) As String`

Aufgabe: Erzeugung von eindeutigen Schlüsseln für die Datensätze der Datenbanktabellen *Aka_Schluesselmaske*, *Aka_Codes*, *Aka_Projekte*. Aufgerufen von: *DBImport.CreateEntries*, *DBImport.Create_Aka_Codes*, *DBImport.Insert_Aka_Projekte*.

`Public Function GetDatei(ByVal filePath As Object) As Byte()`

Aufgabe: liefert zu dem Dateipfad *filePath* den Dateiinhalt. Aufgerufen von: *DBImport.ImportData*, *DBMetadata.Update_Aka_Projekte*.

`Public Sub openLog()`

Aufgabe: erneutes Starten oder Wiederverwenden der Prozessressource für das Schreiben der Log-Datei bei fehlerhafter Excel-Tabelle. Aufgerufen von: Dateiauswahl.cmdOK_Click

```
Public Sub closeForms()
```

Aufgabe: Schließen aller geöffneten Formulare mit Ausnahme des Startformulars.

Quabelle:

Ein Quabelle-Objekt wird von Funktion *checkXITabs* des Moduls *Globals* zur Überprüfung der Excel-Daten gemäß der in 13.2.4 angegebenen Kriterien erzeugt und verwendet.

Weitere Aufrufe erfolgen in den Prozeduren des Moduls *DBImport*, um die Excel-Daten in die Datenbank zu transferieren.

DBImport:

Das Modul *DBImport* sorgt für die Übertragung der Excel-Daten in die beschriebenen Tabellen *Aka_Projekte*, *Aka_Schluesselmaske*, *Aka_Codes*, *Aka_SchlüsselCode*, *Aka_Codelnhalt*, *Lit_ZR und Daten__Aka*. Dazu stellt es folgende zwei Funktionen bereit:

```
Public Function ImportProject(ByVal azr As String, ByVal zeitraum As String) As String
```

Aufgabe: Speichern der Projektdaten eines neuen Projekts. Parameter: *azr* Anzahl Zeitreihen und *zeitraum*, die zuvor von *globals.checkXITab* berechnet wurden. Ruft auf: *Insert_Aka_Projekte* zum Einfügen eines Datensatzes in die Tabelle *Aka_Projekte*. Aufgerufen von: *cmdOK_Click* des Formulars Dateiauswahl.

```
Public Function ImportData(ByVal strXI As String, ByRef exc As Exception, ByVal strID_Project As String, ByVal anzZR As Int32, ByVal zr As String) As Boolean
```

Aufgabe: Speichern der Zeitreihendaten. Parameter: *strXI* Pfad der Excel-Datei, *exc* Verweis auf ein Exception-Objekt für den Fehlerfall, *strID_Project* Schlüssel des Projekts, *azr* Anzahl Zeitreihen, *zr* zeitraum. Aufgerufen von: *importDaten* des Formulars Dateiauswahl.

DBUpdateMetadata

Das Modul *DBUpdateMetadata* besitzt zum Bearbeiten der Metadaten folgende Funktion:

```
Public Function Update_Aka_Projekte() As Boolean
```

Aufgabe: Aktualisieren der Tabelle *Aka_Projekte*. Aufgerufen von: der Prozedur Speichern des Formulars Metadaten3.

DBRemove

Das Modul *DBRemove* bietet zwei Lösch-Prozeduren an:

```
Public Sub del_Projekt(ByVal id_projekt As String)
```

Aufgabe: Löschen eines gesamten Projekts. Aufgerufen von: der Prozedur *Speichern* des Formulars *Metadaten3*

und

Public Sub del_Daten(ByVal id_projekt As String)

Aufgabe: Löschen der Zeitreihendaten, Zurücksetzen der Zeitreihenanzahl in der Tabelle *A-ka_Projekte* auf 0. Aufgerufen von: den Prozeduren *cmdOK_Click* und *importDaten* des Formulars *Dateiauswahl*.

Formatierung der Excel-Dateien

Für einen erfolgreichen Datenimport ist das Beachten einer strikten Formatierung erforderlich. Jede Excel-Datei (Excel-Arbeitsmappe) stellt eine Tabelle mit maximal 65536 Zeilen und 65026 Spalten dar.

Es gelten folgende Bedingungen:

- Tabellen mit höchstens 256 Spalten werden in einem Excel-Arbeitsblatt gespeichert.
- Bei einer höheren Spaltenanzahl können zusätzlich die Spalten aus den folgenden Arbeitsblättern verwendet werden.
- Die erste Spalte ist in allen Arbeitsblättern identisch und hat die Form

Hauptschlüsselname

Codebeschreibung₁

...

Codebeschreibung_n

Quelle

Anmerkung (optional)

Jahrsem₁

...

Jahrsem_m

- Die Spalten müssen fortlaufend belegt sein.

Fehlt in der zweiten Zeile des letzten Arbeitsblattes ein Spalteneintrag, d.h. eine Codebeschreibung oder eine Codebezeichnung, so werden die darauf folgenden Spalten nicht mehr berücksichtigt.

- In den Kopfzeilen ab der 2.Zeile sind keine Leerfelder erlaubt, d.h. keine Angabe zu ei-

ner Codebeschreibung, einer Quelle oder einer Anmerkung, falls diese Zeile vorhanden ist, darf fehlen. Sonst bricht der Import mit einer Fehlermeldung ab.

- Die Codebeschreibungen und ihre möglichen Werte hängen von der betroffenen Tabelle, die durch ihren Hauptschlüsselnamen gekennzeichnet ist, ab. Gleichartige Einträge haben also in verschiedenen Tabellen jeweils eine eigene Bedeutung.
- In den Datenfeldern, d.h. den Zellen, die die Angaben zu $Jahrsem_1$ bis $Jahrsem_m$ enthalten, entspricht ein Leerstring dem Eintrag „nicht vorhanden“. Der numerische Wert 0 wird durch 0 oder 0.000 dargestellt.

	A	B	C	D	E	F	G	H	I
1	Flosserei								
2	Wasserstraße	Neckar	Neckar	Neckar	Neckar	Neckar	Neckar	Neckar	Neckar
3	Stadt	Heilbronn	Heilbronn	Heilbronn	Heilbronn	Heilbronn	Heilbronn	Heilbronn	Heilbronn
4	Wassergebiet (WG)	Rheingebiet	Rheingebiet	Rheingebiet	Rheingebiet	Rheingebiet	Rheingebiet	Rheingebiet	Rheingebiet
5	Güter	Flüsse	Gestöre	Gestöre	Stämme	Stämme	Oblast	Oblast	Oblast
6	Zusatzindikator	insgesamt	Holländer	Gemeine	Holländer	Gemeine	Blöcke (Sagk	Diele und Bö	Rahmschenk
7	Einheit	Anzahl	Anzahl	Anzahl	Anzahl	Anzahl	Anzahl	Anzahl	Anzahl
8	Tabelle	B.1.7.90	B.1.7.90	B.1.7.90	B.1.7.90	B.1.7.90	B.1.7.90	B.1.7.90	B.1.7.90
9	Quelle	s. Beschreibung	s. Beschreibung	s. Beschreibung	s. Beschreibu				
10	1837				5.643	67.287	224	53.730	664
11	1838				9.743	65.545	155	54.803	1.319
12	1839	540	711	6.554	6.176	75.979	293	41.792	1.315
13	1840								
14	1841	183	326	2.243	1.871	22.767	164	18.850	226
15	1842	407	925	4.529	5.173	48.369	44	35.511	656
16	1843	530	865	6.397	4.893	64.531	262	55.471	1.667
17	1844	632	989	7.205	5.230	74.871	123	80.990	931
18	1845	761	1.035	7.024	6.235	74.662	229	76.775	521
19	1846	683	867	7.445	5.127	79.499	37	47.917	135
20	1847	525	889	5.796	4.982	58.561	51	52.763	737
21	1848	222	331	2.732	1.662	26.301	44	31.872	600
22	1849	590	906	6.279	5.436	58.320	212	79.947	794
23	1850	875	1.303	9.425	7.665	98.469	630	90.901	867
24	1851	646	1.357	6.446	8.592	66.558	1.051	89.365	735
25	1852	1.173	1.133	10.974	8.264	113.439	2.724	104.932	2.060
26	1853	1.134	1.531	10.784	10.450	121.870	2.811	96.203	2.360
27	1854	1.254	3.996	10.931	11.075	97.627	3.096	84.975	1.310

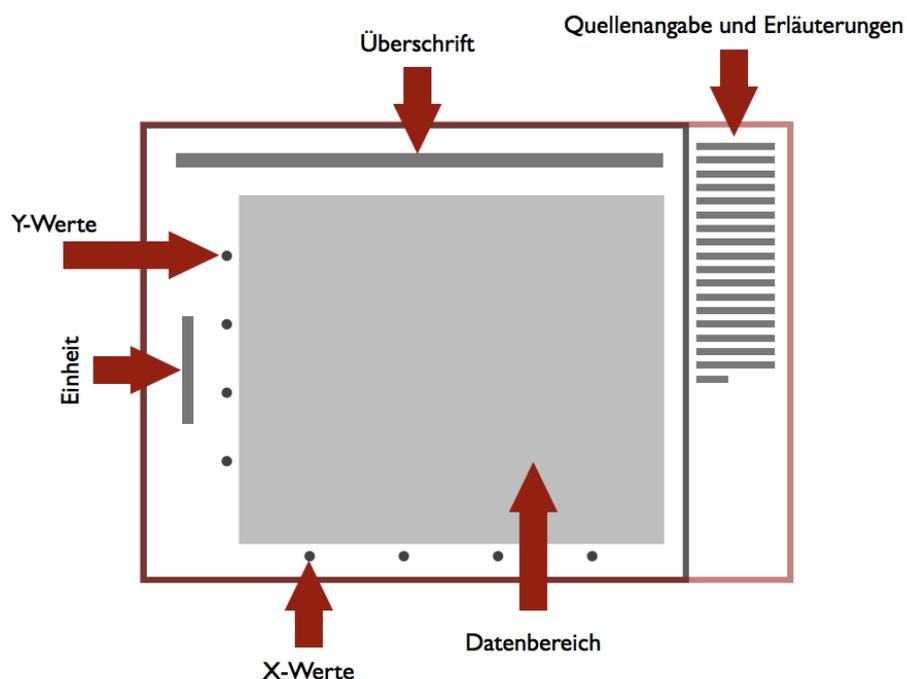
Beispiel aufbereitete Datei aus dem Projekt "Statistik der Binnenschifffahrt"

11.4 Galerie

Langfristiges Ziel ist die Erstellung einer Galerie mit zunächst rund 1.000 Grafiken in etwa 300 Abbildungen, die die 250.000 Zeitreihen von **hista**t möglichst gut repräsentieren sollen. Die Grafiken werden mit Unterstützung durch studentische Hilfskräfte erstellt. Zunächst werden Abbildungen zu Studien erstellt. In einer zweiten Stufe sollen dann Abbildungen folgen, die Zeitreihen aus mehr als einer Studie enthalten.

Die Abbildungen sollen im Internet im Rahmen einer Galerie angeboten werden. In der aktuellen Erstellungsphase werden sie auf der Seite „Galerie“ statisch angezeigt.

Die Abbildungen sind wie folgt aufgebaut:



Eine Grafik kann zusätzlich die folgenden Elemente haben:

- Eine oder mehrere „shape areas“.
- Einen Indexpunkt an einem beliebigen Punkt einer Reihe bzw. an einem Schnittpunkt zweier Reihen.
- Eine oder mehrere senkrechte oder waagerechte rechteckige farbig hinterlegte Bereiche.
- Individuelle Beschriftungen an individuellen Positionen.

Eine Zeitreihe kann mehrere Abschnitte in genau einer anderen Farbe aufweisen. Die Farbe ist abhängig von der Farbe der Reihe, in der diese Abschnitte erscheinen (festgelegt über die Datei `_histat_colors.r`). In der Regel sollte diese Möglichkeit verwendet werden, um eine Verlängerung einer Reihe zu kennzeichnen. (z.B. Aktualisierung durch eine neuere Studie). Sämtliche Daten (= Zeitreihen) werden per SQL aus der `histat`-Datenbank abgefragt. Sämtliche individuellen Werte der Abbildungen / Charts werden in den Datenbanktabellen `charts` und `pictureheadings` vorgehalten (konkrete Datenabfragen, Überschriften, Labels, Positionen der Zeitreihen-Beschriftungen, Spannweiten, siehe Anhang). Die verschiedenen Ausprägungen der unterschiedlichen Stile (grid-Farben, Linienstärken, Farben, Achsentypen, Ränder) werden ebenfalls in einer Datenbanktabelle, `styles`, vorgehalten, siehe Anhang. Sämtliche Verbindungsangaben sowie Layout-Einstellungen werden, ähnlich wie CSS-Dateien, in separaten `r`-Dateien definiert.

<code>_hstat_colors.r</code>	Definition der Farben
<code>_get_data_and_values.r</code>	Daten und Werte einlesen
<code>_hstat_heading.r</code>	Überschrift für die gesamte Abbildung
<code>_hstat_styles.r</code>	Definition des Layouts der Charts

Diese Dateien werden von den konkreten Abbildungs-Dateien unmittelbar oder mittelbar aufgerufen. Darüber hinaus werden optional noch Definitionen von Indexpunkten und shape areas eingebunden:

<code>_hstat_indexpoint.r</code>	Definition eines Indexpunktes
<code>_hstat_shapearea.r</code>	Definition einer shape area

Die r-Abbildungsdatei erzeugt eine PDF-Abbildung mit gleichem Namen. Die PDF-Datei wird in eine LaTeX-Datei gleichen Namens eingebunden, in der sich der Kommentar befindet, und daraus eine neue PDF-Datei erzeugt. Die PDF-Datei wird schließlich in eine PNG-Datei konvertiert. Der Arbeitsablauf ist im Anhang in der Übersicht „workflow“ schematisch wiedergegeben.

Vorgaben

Für jede Abbildung gibt es

- genau eine r-Abbildungsdatei
- genau einen Eintrag in der Tabelle *pictureheadings*
- mindestens einen Eintrag in der Tabelle *charts*

Der Name der Datei (ohne Endung) ist der Primärschlüssel in der Tabelle *pictureheadings*.

Typen von Zeitreihen

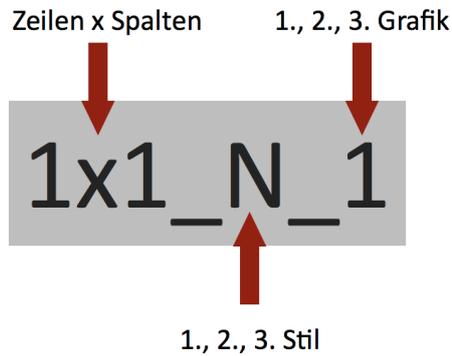
Folgende Typen von Zeitreihen sind möglich:

- Eine Reihe ohne weitere
- Eine Reihe mit Trend
- Eine Reihe, die verlängert wird

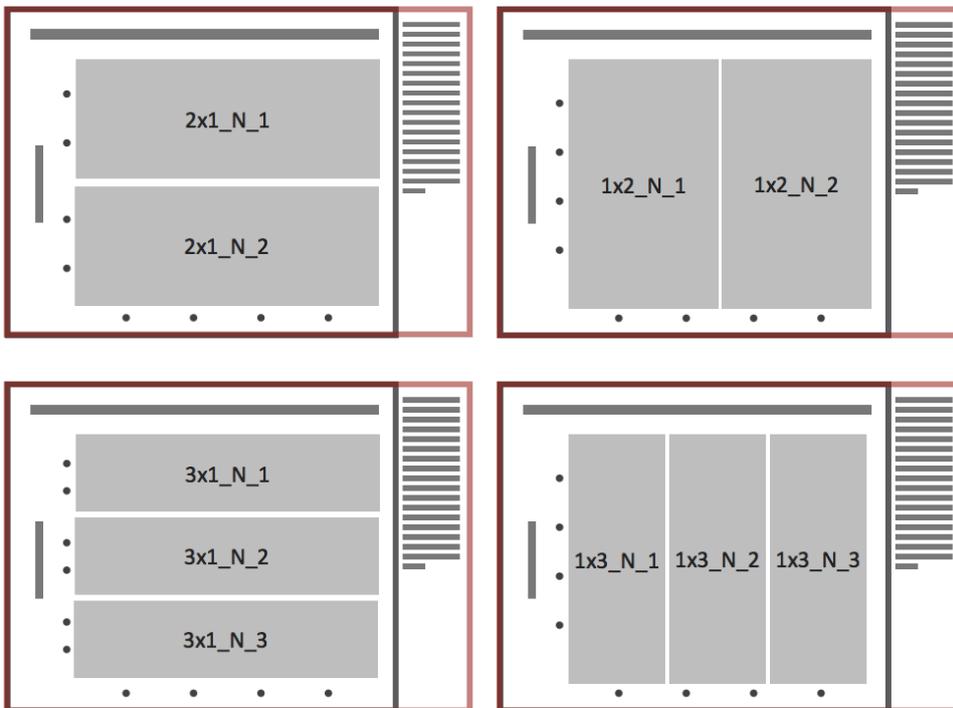
Der Trend wird als dicke, halbtransparente Linie gezeichnet. Die Liniendicke muss abbildungsspezifisch (nicht stylespezifisch) festgelegt werden, da die einzelnen Abbildungen unterschiedliche Skalierungen haben.

Typen von Abbildungen

Derzeit sind pro Abbildung ein, zwei oder drei Grafiken vorgesehen. Die style-Namen einer Abbildung setzen sich wie folgt zusammen:



Daraus ergeben sich die folgenden Grafiken:



Sonderfälle

Zusätzlich können in eine Reihe folgende Gestaltungselemente aufgenommen werden:

- Index: grauer Punkt/Kreis bei 100
- „Shape areas“ (vertikal, unter der Reihe; vertikal, ganze Höhe; horizontal, ganze Breite)

Für Shape areas unter einer Reihe wird die Datei `_hista_shapearea.r` eingebunden, für einen Indexpunkt die Datei `_hista_indexpoint.r`. Rechteckige shape areas sowie weitere evt. Ergänzende Gestaltungselemente werden direkt in die jeweilige `r`-Abbildungsdatei geschrieben. Es kann auch Einträge in `pictureheadings` ohne korrespondierende Einträge in `charts` geben: in diesen Fällen holen die `r`-Dateien die Daten nicht aus der DB. Die Informationen in `picturehea-`

dings werden dann nur für die web-Anzeige und das PDF verwendet, nicht, um daraus eine Abbildung zu erstellen.

Die Farben

Die Farben der verlängerten Reihen sollten aus den Farben der eigentlichen Reihen „abgeleitet“ sein. Alle Farben werden in der Datei `_histat_colors.r` definiert.

12 Anhang

Übersicht über die Eingabe, Datenbank und Ausgabe von histat

