



DUVA 4 CS

Grundlagen

Stand April 2010



Inhaltsverzeichnis:	Seite
Vorbemerkungen	2
1. Der DUVA-Metadatenansatz	6
1.1 Grundsätzlicher Prozessablauf statistischer Arbeit	6
1.2 Granulierung - Grundlage für die Metadatenmodellbildung	10
1.3 Das DUVA- Metadatenmodell	16
1.3.1 Verschlüsselungen	16
1.3.2 Merkmale	18
1.3.3 Satzaufbauten	18
1.3.4 Dateibeschreibungen	21
1.3.5 Dateischablonen	25
1.3.6 Regeln	26
1.3.7 Produktion	30
1.3.8 Thesaurus, Suchfunktione	32
1.3.9 Import/Export	34
1.3.10 Benutzerverwaltung	35
2. DUVA-Module	36
3. DUVA im Einsatz	40

Handbuch: Tilman Häusser

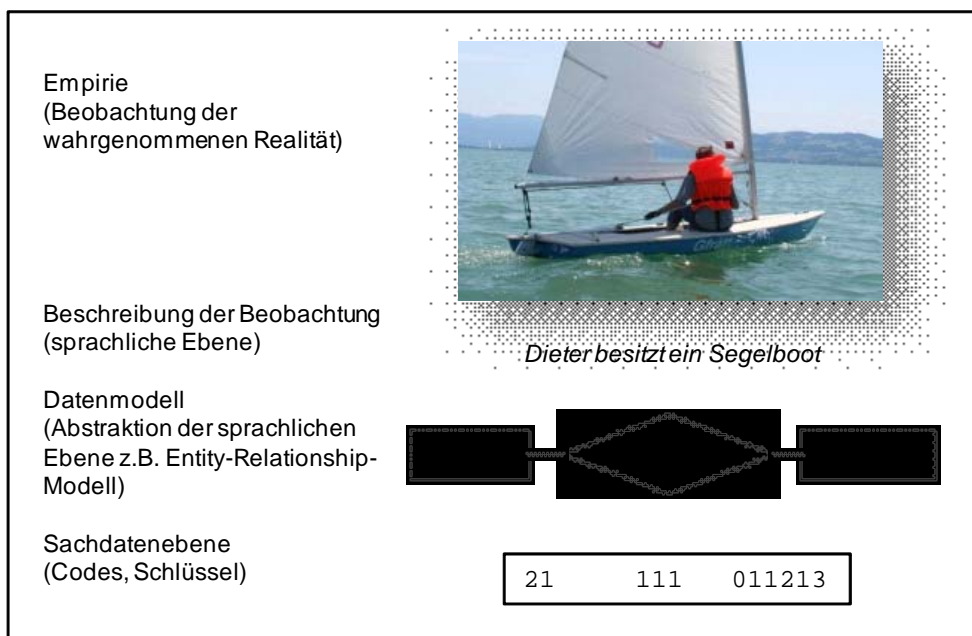
Betreuende Stelle: KOSIS-Gemeinschaftsprojekt DUVA
c/o Amt für Bürgerservice und Informationsverarbeitung
Stadt Freiburg Postfach
79095 Freiburg

Internet: <http://www.duva.de>
info@duva.de

Vorbemerkungen

In öffentlichen Verwaltungen wie auch in Unternehmen der Privatwirtschaft ist eine wachsende Vielfalt von Verfahren zur Datenerfassung, Datenhaltung und -Verwaltung sowie der Datenaufbereitung, Datenanalyse und Datenauskunft festzustellen. Selbst bei prinzipiell einheitlichen Verfahren, wie etwa dem Einwohnerwesen in der Kommunalverwaltung, gibt es länderspezifische Verfahren oder auch unterschiedliche Erfassungs- und Verarbeitungssysteme innerhalb eines Bundeslands. Daneben werden in den einzelnen Fachbereichen wie etwa dem Bauwesen, dem Sozialbereich, dem Verkehrswesen usw. unterschiedlichste Systeme und Architekturen eingesetzt. Diese heterogene Struktur von Verfahren, Systemen und fachlich basierter Definitionen führt dazu, dass die Kommunen ohne ein übergeordnetes Informationsmanagement grundsätzlich nicht mehr in der Lage sind, auf ihr Informationspotential vollständig zuzugreifen und es nach gegebenen Fragestellungen auszuschöpfen. Dies gilt in gleichem Maße für die Privatwirtschaft. Wissensbasierte Datenanalysen und flexible, fachübergreifende Informationsnachfragen als Grundlage für Entscheidungsprozesse sind bei rein sektoralen Systemlösungen und Betrachtungsweisen kaum möglich. Auch hier geht es darum, alle möglichen oder erforderlichen Daten- und Informationsquellen über eine Informationsmanagementschicht zu integrieren und unabhängig von ihrem operativen Einsatzzweck in sinnvollem Kontext zu nutzen.

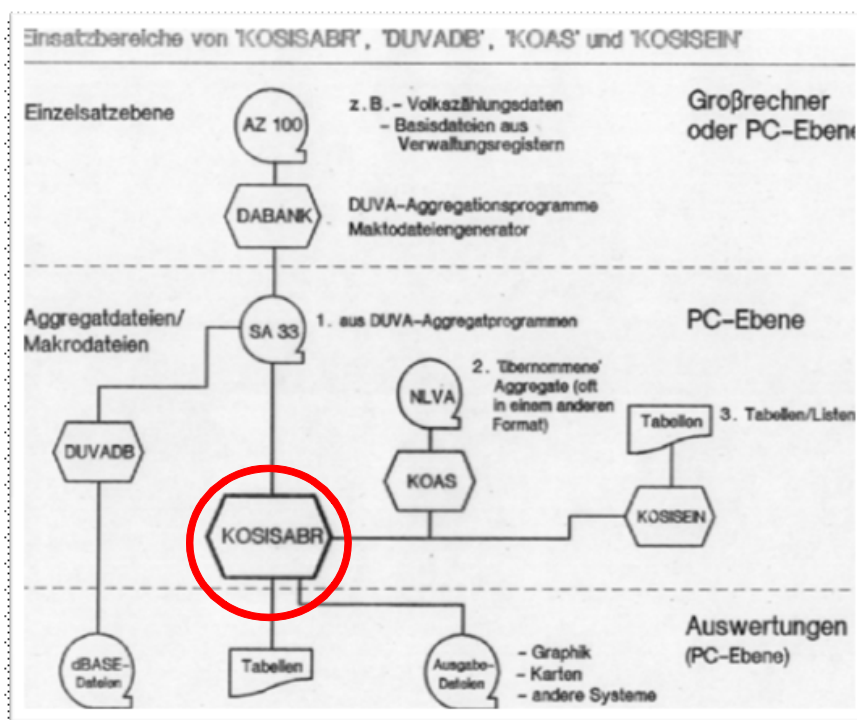
Data-Warehouse-Konzepte bzw. –Systeme bieten seit einigen Jahren branchenspezifische Lösungen oder proprietäre Implementationen auf Anwenderseite an. Branchenübergreifende Lösungen, die auch Anpassungsmöglichkeiten an verschiedene IT-Umgebungen vorsehen, sind dagegen eher selten. Der Grund dafür liegt in der Konzentration auf branchenspezifische Sachdaten. Gefordert sind aber fachübergreifende Metadaten. Dies sei an einem kleinen Beispiel dargestellt:



Im obigen Beispiel sind die Abstraktionsstufen von einer Beobachtung in der Realität über die sprachliche Fixierung zum Datenmodell bis hin zur Sachdatenebene, die in Form von Zeichen (Schlüssel) oder Werten (numerische Variablen) Grundlage der Datenspeicherung ist, dargestellt. Reduziert man das Beispiel auf ‚branchenspezifische‘ Metadaten, so lassen sich diese anhand des Datenmodells und der Codes oder Schlüssel auf der Sachdatenebene ableiten. Das bedeutet, die Metadaten beschreiben zunächst eine Relation zwischen zwei Objekten (Mensch, Fahrzeug), die

durch eine bestimmte Operation (besitzen) spezifiziert ist. Desweiteren sind die Begriffe ‚Mensch‘, ‚besitzt‘ und ‚Fahrzeug‘ selbst beschreibende Daten und somit Metadaten, denn sie erklären die Schlüssel der Sachdatenebene (z.B. 21=Mensch, 111=besitzt, 011213=Fahrzeug). Anhand dieses kleinen Beispiels wird bereits ersichtlich, dass es unterschiedliche Klassen oder Kategorien von Metadaten gibt. Hier sind es Metadaten, die den Zusammenhang zwischen Objekten untereinander beschreiben und Metadaten, die Abstraktionen bzw. Abstraktionsebenen erklären (z.B. Erklärung der Sachdaten). Branchenspezifisch ist nun die Anwendungsbreite unseres Beispieldatenmodells: es handelt sich um zwei Objekte und einen Operator. Statt ‚Mensch‘ könnte auch ‚Gebäude‘ eingesetzt werden, statt ‚besitzt‘ ‚ist ausgestattet mit‘ und statt ‚Fahrzeug‘ beispielsweise ‚Fahrstuhl‘. Ein fachübergreifender Ansatz sollte jedoch nicht eine begrenzte Anzahl von Objekten und Operatoren zum Inhalt haben, sondern für unterschiedlichste Arten von Objekten, Operatoren und den daraus resultierenden Daten und deren Organisation geeignet sein.

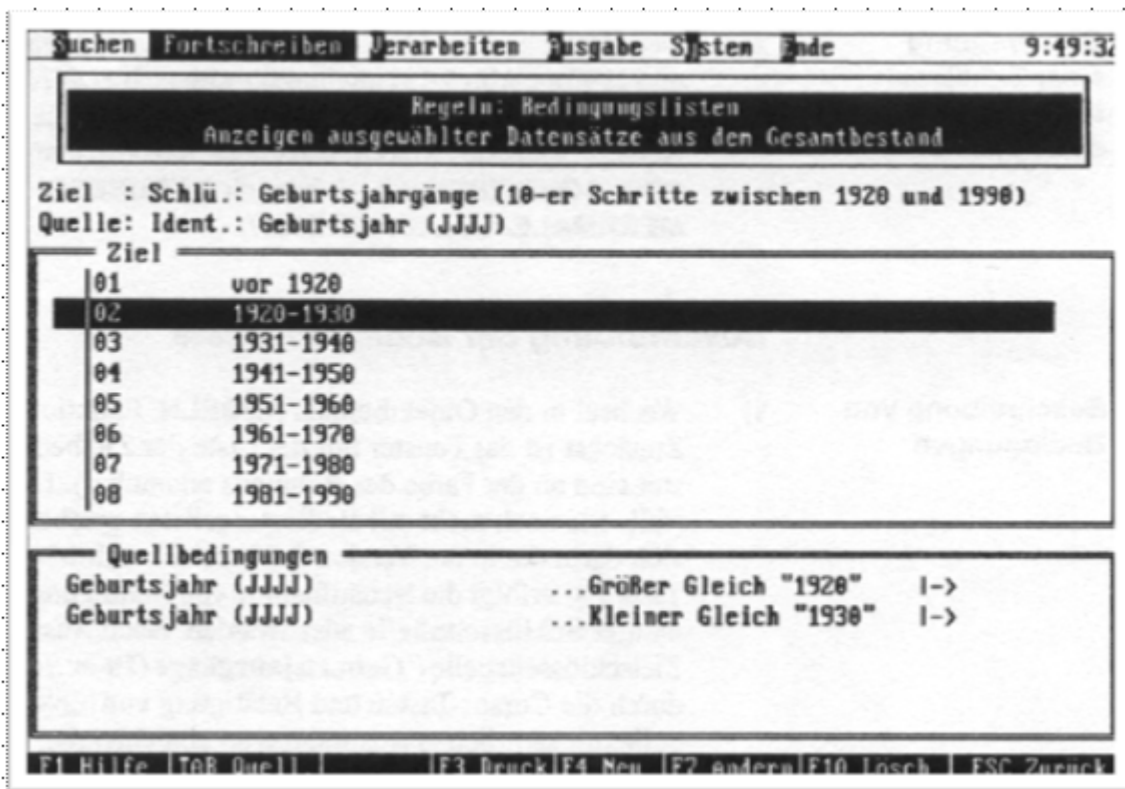
Dieser Grundgedanke begleitet das metadatenbasierte Informationssystem DUVA seit seiner Entstehung zu Beginn der neunziger Jahre. Der Name DUVA steht für ein Gemeinschaftsprojekt des KOSIS-Verbands unter der Trägerschaft des Verbands Deutscher Städtestatistiker (VDSt). Der KOSIS-Verband hat sich das Ziel gesetzt, durch kooperative Verfahrensentwicklungen von Statistischen Informationssystemen sowie deren Pflege und Nutzerunterstützung Verbesserungen der Planungs- und Entscheidungsfähigkeit von Kommunen zu erreichen. Die Aufgabe des DUVA-Projekts war, historisch gesehen, die Bereitstellung von Datenverarbeitungsinstrumenten zur Unterstützung von Volkszählungsauswertungen der Volkszählung 1987 in Westdeutschland. Zunächst handelte es sich dabei um Programme, die in den Statistikstellen der Kommunen eine PC-seitige Weiterverarbeitung der Volkszählungsdaten zu DIN-gerechten Tabellen ermöglichten.



Um die Tabellierungsprogramme an die jeweilige Informationsnachfrage anzupassen, müssten beispielsweise die Beschreibungen von Merkmalen und ihren Ausprägungen, also die Metadaten, aus umfassenden Katalogen in die Programmsyntax übernommen werden. Diese mühevollen und nicht medienbruchfreie Arbeit, führte zu der Forderung nach einem System, das sowohl die Beschreibungsdaten verwaltet, als auch die Sachdaten zu den gewünschten Produkten wie Tabellen und Diagrammen verarbeitet. Eine weitere Forderung war die, dass die Kommunikation zwischen

System und Anwender nicht auf syntaktischer Ebene sondern in natürlicher Sprache („Klartext“) erfolgt. Der Anwendungsbereich sollte sich außerdem nicht nur auf die Auswertung von Volkszählungsdaten beschränken, sondern für alle Arten von Verwaltungsregisterdaten geeignet sein.


Die Basisüberlegungen für die Ableitung des dafür erforderlichen Metadatenmodells gingen deshalb auf die elementaren Grundlagen von Statistik bzw. des statistischen Produktionsprozesses zurück. Statistik hat ja nicht nur deskriptive und schließende Methoden zum Inhalt, sondern umfasst das gesamte Spektrum von grundlegenden Fragen einer Messung oder Erhebung (z.B. Stichprobenmethode, Erhebungsart) über das Datenmanagement, Einsatz von Auswertungsmethoden bis hin zu Ergebnistabellen, Grafiken, Karten und textlicher Erläuterungen und Interpretationen. Mit anderen Worten: es handelt sich um den grundlegenden Formkreis, der bei der Datenentstehung beginnt und im interaktiven Wissensaustausch als Endstufe einen neuen Zyklus öffnet. Und dies unabhängig von der Art einer materiellen Statistik. Das heißt, es könnte sich um die einfachste Art einer Statistik handeln, wie etwa ein Adressbuch, oder um den komplexen Prozess eines Web-basierten Buchhaltungsverfahrens. Die erste Realisierung des metadatenbasierten Informationssystems DUVA erfolgte auf DOS-Ebene unter der Desktop-Datenbank Paradox. Sieht man von einigen Erweiterungen auf Objektebene ab, die sich durch Anwendererfahrungen und technologischem Fortschritt ergaben, entspricht die Metadatenbank der aktuellen Version immer noch dem Kernmodell aus den neunziger Jahren. Den wissenschaftlichen Hintergrund für das Metadatenmodell lieferten maßgeblich Bo Sundgren (Statistics Sweden, Stockholm University) und Günther Appel (bis 2000 Präsident des Statistischen Landesamts Berlin, TU Berlin).



Die weiteren Entwicklungsstufen auf Basis der Borland Database Engine und der kostenfreien Paradox-Runtime-Umgebung unter Windows 95 und nachfolgenden Windows-Versionen schafften für die Anwender der DUVA-Module eine weitgehende Systemunabhängigkeit. Grund für diese Unabhängigkeit war die getrennte Datenhaltung von Sachdaten (z.B. als Textfiles, CSV-Dateien, DB-Dateien etc.) und den Metadaten in der Paradox-Datenbank als zentrales Repository.


Die aktuelle Version DUVA4 CS steht als Client-Server Anwendung unter Oracle, MS-SQL-Server und PostgreSQL zur Verfügung. Der Datenbankzugriff erfolgt über ADO. Nach wie vor gilt das Prinzip der Trennung von Sach- und Metadaten, d.h., es lassen sich unterschiedlichste Sachdateiformate (Textfiles in festem Format, variable Textformate, Datenbanktabellen) laden und verarbeiten und über die Metadatenbank des DUVA-Nachweissystems als Kernmodul steuern. Als Betriebssystem kann Windows XP oder höher eingesetzt werden, die Hardwareanforderungen stehen in Abhängigkeit zur eingesetzten Datenbank bzw. der entsprechenden Datenhaltung (minimal 2,8 GHz-Prozessor, 2GB RAM und entsprechende Festplattenkapazität insbesondere für die Sachdaten).

Die nachfolgenden Darstellungen und Erläuterungen geben einen Überblick zum Thema DUVA, angefangen vom Metadatenansatz bis hin zu den Einsatzbereichen und Funktionalitäten der einzelnen Module. Aufbauend auf die Grundlageninformationen stehen weitere Dokumente wie Installationsanweisungen, Benutzerhandbücher und technische Erläuterungen für den speziellen Anwendungsfall bzw. Detailfragen zur Verfügung.



Informationsmanagement


Das Projekt



Organisation und Aufbau

Das KOSIS-Gemeinschaftsprojekt DUVA setzt sich aus derzeit 60 Einrichtungen und Institutionen zusammen. Darunter befinden sich Kommunen, kommunale Rechenzentren, Bundes- und Landesbehörden sowie Bildungseinrichtungen.

Die Mitglieder der Gemeinschaft stellen das beschließende Gremium dar, die betreuende Stelle (derzeit die Stadt Freiburg im Breisgau) vertritt die GbR nach außen. Eine von der jährlich stattfindenden Mitgliederversammlung gewählte Lenkungsgruppe ist für die inhaltliche und technologische Entwicklung des Projekts zuständig. Die betreuende Stelle hat den Vorsitz in der Lenkungsgruppe und wird durch eine hauptamtliche Geschäftsstelle unterstützt.



1. Der DUVA-Metadatenansatz

Die Frage ‚was sind denn eigentlich Metadaten‘ wird häufig mit dem Satz beantwortet ‚Daten über Daten‘. Wie jedoch bereits in den Vorbemerkungen angedeutet, ist der Zusammenhang komplexer und erfordert Kategorisierungen, um das Thema Metadaten vollständig zu erfassen. Eine wesentliche Eigenschaft von Metadaten ist der Bezug zu Objekt- oder Sachdaten. Metadaten beschreiben niemals direkt Objekte der Realwelt sondern nur Daten (z.B. Schlüssel) oder Strukturen (z.B. Satzaufbauten) oder Abhängigkeiten in einer Prozesskette (z.B. Anwendung von statistischen Methoden oder Ableitungsregeln). Ein knapp abgefasster Exkurs über den statistischen Produktionsprozess soll beispielhaft für einen allgemeinen Ansatz stehen.

1.1 Grundsätzlicher Prozessablauf statistischer Arbeit

Das Wesen der Statistik besteht ganz allgemein darin, Daten zu erheben oder zu sammeln, sie zu strukturieren und zu ordnen, sie mit Hilfe deskriptiver und schließender Methoden zu analysieren sowie sie in Form von tabellarischen, grafischen und textlichen Darstellungen zu präsentieren. Obwohl die Daten unterschiedlichsten Ursprungs sind und je nach Systematisierungsansatz verschiedensten Fach- oder Sachgebieten zugeordnet werden können, haben sie doch eine Gemeinsamkeit: sie bilden mit mehr oder weniger hohem Genauigkeits- und Abstraktionsgrad Ausschnitte der Realität ab.



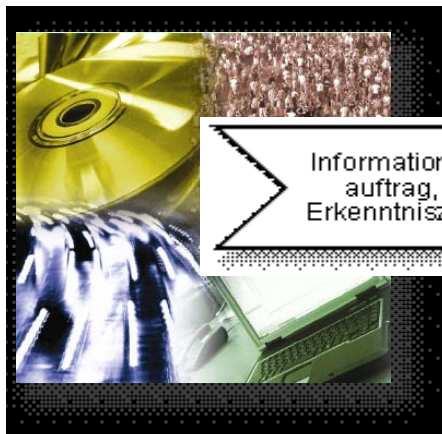
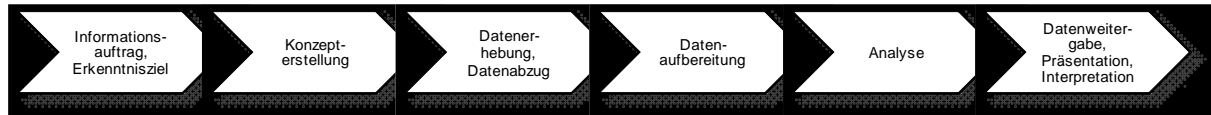
Dies trifft für Kundendaten eines Betriebs genauso zu wie für Messdaten der Grundwassergüte oder Daten aus dem amtlichen Einwohnerwesen, Daten aus Forschungslaboren oder die persönliche Adresskartei. Dabei bestimmt die jeweilige Messvorschrift oder Erfassungsmethode das Maß an Genauigkeit und Objektivität der Informationsübertragung aus der Realität. So ist die physikalisch exakt definierte Messung von Strecken oder Gewichten mit einer anderen Objektivität verbunden als z.B. Datenerhebungen im sozialwissenschaftlichen Bereich wie etwa eine Erfassung des Wohlbefindens von erwachsenen Personen mit Hilfe freier Interviews.



Im ersten Fall bestimmt eine allgemein festgelegte Skala (z.B. 1 m = 100 cm) und ein geeichtes Messinstrument (z.B. Meterstab) die Qualität der Messung. Messwiederholungen führen jeweils bei gleicher Anordnung und gleichen Grundbedingungen zum gleichen Messergebnis. Anders gestaltet sich die Situation, z.B. bei einer Erhebung des Wohlbefindens. Hier fehlen allgemein festgelegte Messskalen und Messinstrumente. Diese sind im Einzelfall auf Grundlage von Theorie und Hypothesen in einem Datenerhebungskonzept zu definieren. Bei jeder Art von Gewinnung statistischer Daten - als allgemein nachvollziehbare und verlässliche Übertragung von Ausschnitten der Realität - sind Messvorschriften oder Erhebungskonzepte unbedingte Voraussetzung.

Versucht man statistische Arbeit als Prozessablauf darzustellen, so ergeben sich sechs Stufen. Diese einzelnen Arbeitsstufen lassen sich in der Praxis nicht immer klar voneinander trennen und erfahren aufgrund spezieller Fragestellungen oder gesetzlicher Vorgaben - z.B. im Bereich der amtlichen Statistik - mehr oder weniger starke Modifikationen. Die folgende, oberflächliche Darstellung der

Grundzüge dieser Arbeitsstufen soll nicht darüber hinwegtäuschen, dass bereits Teile dieses Themas (z.B. Datenerhebungstechniken, statistische Analysemethoden) 'lehrbuchfüllend' sind.



Statistische Arbeit dient nicht dem Selbstzweck, sondern basiert auf einem Informationsauftrag oder einem Erkenntnisziel. Im Zusammenhang mit der öffentlichen Statistik sind es vielfach Normen (z.B. Meldegesetze), die operative Verfahren und Statistiken fordern. Hinzu kommen Erkenntnisziele, die von Politik, Wirtschaft oder Wissenschaft als Entscheidungsgrundlagen benötigt werden (z.B. das Statistische Amt einer Stadt erhält den Auftrag eine Studie über den Zusammenhang zwischen Zu- und Wegzügen und dem lokalen Immobilienmarkt anzufertigen). Nicht zu vergessen der demokratisch begründete Informationsanspruch der Öffentlichkeit.

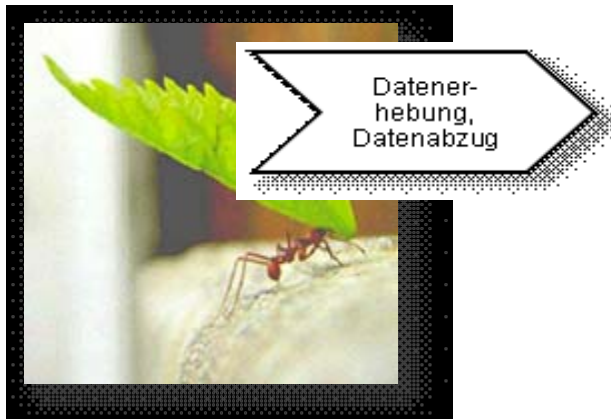
Liegt ein Erkenntnisziel fest, so folgt die theoretische Beschäftigung mit dem Gegenstandsbereich (z.B. Literaturanalyse, Methodenrecherche, Klärung, ob für den zu untersuchenden Gegenstandsbereich bereits Studien, Theorien vorliegen, Abklärung rechtlicher Rahmenbedingungen etc.).



Die Konzepterstellung beginnt mit der Konzeptspezifikation (Begriffsklärungen, Festlegung von Definitionsmerkmalen) und der Operationalisierung (Umsetzung der einzelnen Erkenntnisziele z.B. in konkrete Fragen und Antwortmöglichkeiten, Festlegung von Messskalen). Leitfragen sind: 'wie liegt ein Sachverhalt in der Realität vor?' und 'durch welche Erhebungsmerkmale kann ich ihn möglichst umfassend, und vor allen Dingen strukturgetreu, abbilden?'. Dann folgen Festlegungen zur Stichprobenart und -höhe sowie zur Befragungs- oder Messmethodik.

Als wichtiger Bestandteil eines Erhebungskonzepts ist außerdem die Vorbereitung des Datenmanagements (z.B. Fragen der EDV-gestützten Datenaufbereitung, Formate, Schnittstellen, Datenweitergabe), des Einsatzes von Verfahren zur Gewährleistung der Datenqualität (Plausibilisierung) und der statistischen Geheimhaltung zu nennen. Aus dem Erhebungskonzept resultiert - wenn z.B. eine Erhebung als Befragung konzipiert ist - der Fragebogen und die Gesamtstrategie für das weitere Verfahren. Die Erarbeitung von Erhebungskonzepten ist nicht nur für Primärstatistiken (erstmalige Erhebungen) grundlegend, sondern auch für Sekundärstatistiken. Besteht beispielsweise bei einem Statistikabzug (Datenabzug aus einem operativen Verfahren) kein Zugriff auf das Erhebungskonzept, so fehlen entscheidende Informationen für die Nutzung und Interpretation der zur Verfügung stehenden Daten. Erhebungskonzepte sind für die statistische Arbeit von höchster Relevanz, da sie Inhalt, Möglichkeiten und Methoden der statistischen Bearbeitung und folglich die Qualität der weiterzugehenden Informationen bzw. des abzuleitenden Wissens determinieren.

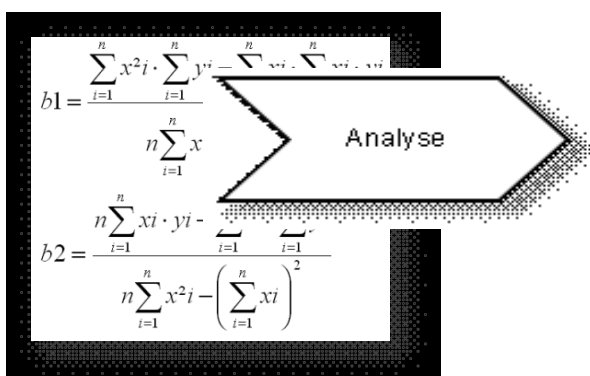
Einen Zwischenschritt oder eine Controllingmöglichkeit zwischen der Konzepterstellung und der Datenerhebung bilden Voruntersuchungen (Pretests). Sie ermöglichen die Überprüfung des Erhebungskonzepts hinsichtlich Validität (sind die Ergebnisse gültig?) und Reliabilität (sind die Ergebnisse zuverlässig?).



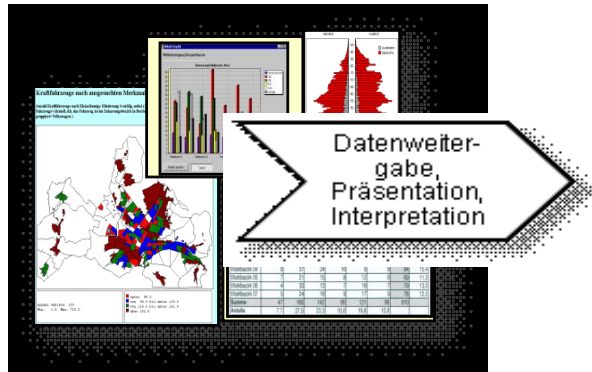
Unter Datenerhebung ist die Durchführung z.B. einer Befragung, Zählung etc. mit der im Erhebungskonzept festgelegten Methode und der vorgesehenen Stichprobenhöhe zu verstehen. Handelt es sich bei den zu erhebenden Daten nicht um Primärdaten sondern um Sekundärdaten, so geht es darum, aus bereits erhobenen Daten Teile im erforderlichen Umfang und in der erforderlichen Qualität als Statistikabzug zu erhalten. Statistikabzug bedeutet, dass aus einem operativen Verfahren (z.B. Einwohnerdatei) eine separate Statistikdatei erzeugt wird (z.B. Einwohnerbestand zum 31.12 eines Jahres).



Je nach Art des Datenträgers, auf dem sich Primär- oder Sekundärdaten befinden, ergeben sich mehr oder weniger zeitintensive Arbeiten zur Datenaufbereitung (Überführung der erhobenen Daten in eine strukturierte, auswertbare Form). Dies bedeutet bei Primärdaten: Einlesen oder Aufnehmen z.B. von Fragebogeninhalten (z.B. über Erfassungsprogramme, Scanner). Bei Sekundärdaten, die bereits als Dateien gehalten werden, sind vielfach Umsetzungen oder Umformatierungen erforderlich. Datenaufbereitung beinhaltet auch Plausibilitätskontrollen und, falls erforderlich, Anonymisierungsverfahren.

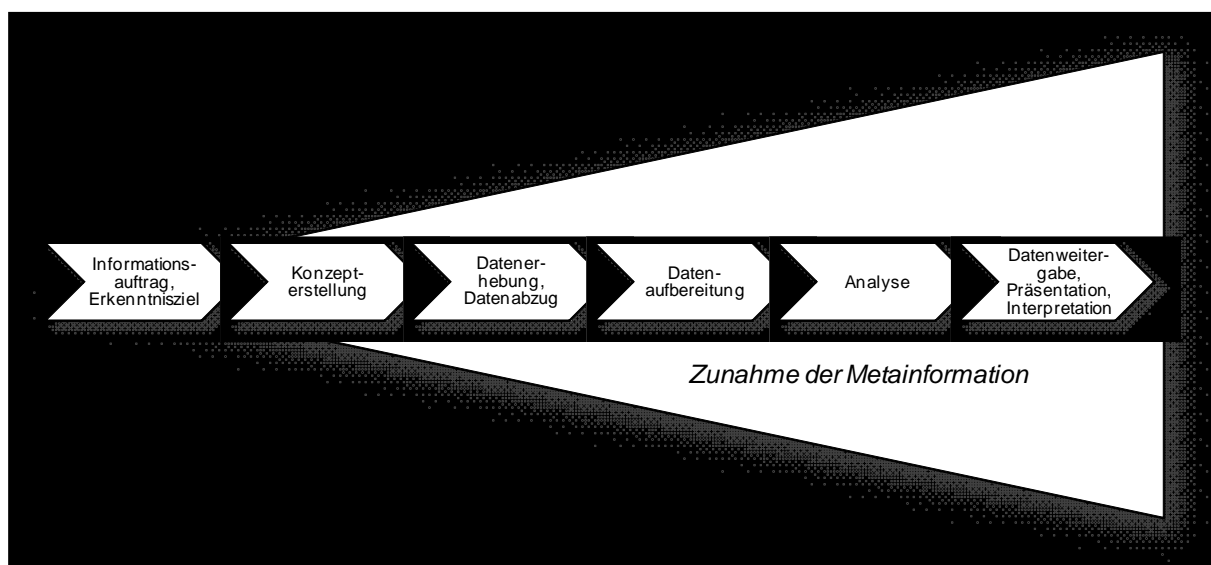


Die Datenanalyse hat den je nach Erkenntnisziel festgelegten Einsatz deskriptiver oder/und schließender Methoden zum Inhalt. Hier lassen sich unterschiedliche Softwareprodukte einsetzen.

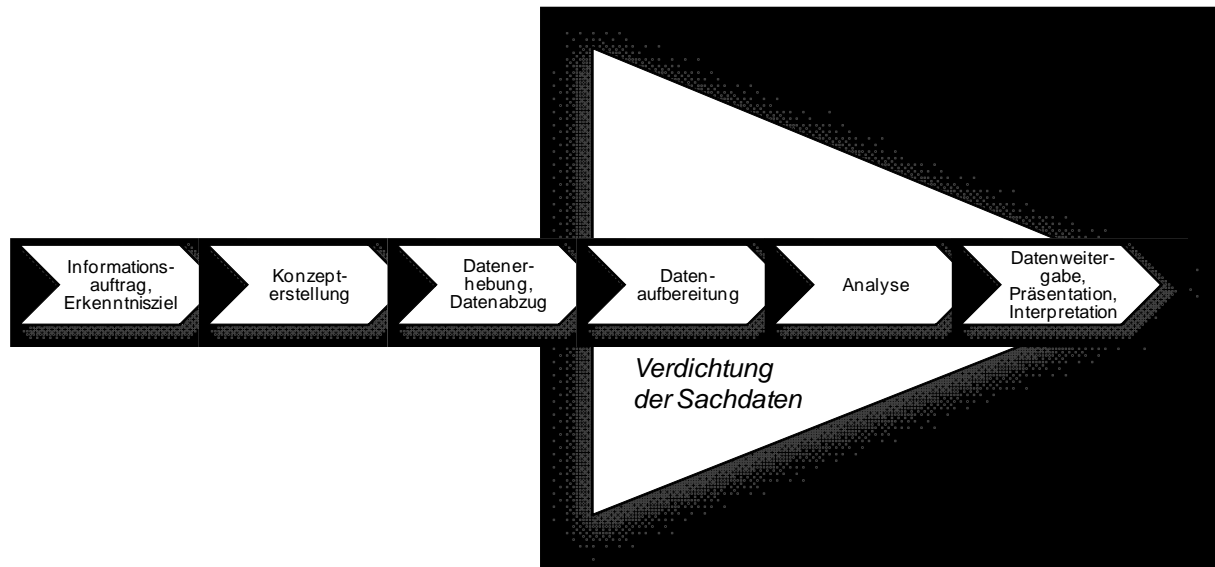


Die Datenweitergabe der Analyseergebnisse erfolgt im Regelfall in Form von Tabellen, Diagrammen, Kennzahlen und textlichen Ausführungen (Erklärungen, Interpretationen), wobei das Weitergabemedium besondere Bedeutung hat (dynamische Web-Tabellen und Grafiken oder kartografische Darstellungen).

Entscheidend für sämtliche Formen der Datenweitergabe ist, dass von jeder Bearbeitungsstufe aus eine Bezugnahme auf den zugrundeliegenden Realitätsausschnitt sowie die im statistischen Produktionsprozess wachsende Menge an Beschreibungs- oder Metainformation möglich sein muss. Nur auf diese Art und Weise können Analyseergebnisse und statistische Führungsinformationen fair und sachlich richtig interpretiert werden.



Sachinformationen als Ergebnisse von Messungen oder Erhebungen entstehen auf der Stufe Datenerhebung (bei Datenabzügen sind es die originären Verfahren). Die Informationsmenge umfasst zunächst sämtliche Merkmale und ihre Ausprägungen, die für jede einzelne Erhebungseinheit oder jeden Merkmalsträger ermittelt wurden. Da sich diese Informationsmenge auf Einzeldatenebene schlecht zur Wissensvermittlung eignet, findet eine Informationsverdichtung über mehrere Abstraktionsebenen statt. Darunter ist z.B. die Verdichtung von Einzeldaten zu Aggregatdaten (z.B. Häufigkeitstabellen) oder die weitere Verdichtung zu statistischen Maßzahlen (z.B. Mittelwerte, Streuungsmaße, Kontingenzmaße etc.) auf Ebene der Analyse zu verstehen.



Als wesentliche Eigenschaft des statistischen Produktionsprozesses bleibt somit zum einen der Verdichtungs- oder Konzentrationsvorgang auf Ebene der Sachdaten und zum anderen die zunehmende Menge an Metainformation festzuhalten. Daraus leitet sich die Forderung ab, dass ein metadatenbasiertes statistisches Informationssystem die Prozesskette und die Beziehung unter den einzelnen Gliedern über Metadaten abbilden muss. Diese Forderung lässt sich auch auf einen allgemeinen Ansatz übertragen. Um die relevanten Metadatenobjekte aus einem Prozess abzuleiten, bedarf es zunächst der Granulierung der einzelnen Prozesskomponenten und ihrer Abhängigkeit zueinander und dann der Übertragung in ein Metadatenmodell.

1.2 Granulierung - Grundlage für die Metadatenmodellbildung

Granulierung von Prozesskomponenten oder Zusammenhängen bedeutet Auflösung in kleine oder kleinste Einheiten. Bereits mit dem Informationsauftrag oder dem Erkenntnisziel erfolgt eine Fixierung von Metadaten. Zu nennen sind hier z.B. Rechtsgrundlagen, Definitionen, Erhebungsart/Datenquelle, der Zeit- und Raumbezug und der Merkmalsträger sowie das Sachgebiet, das aus systematischen Gründen eine Rolle spielt. Diese Metadatenobjekte gehören zu der Objektgruppe Definitionsmerkmale, die grundsätzliche Fragen ‚Wer‘, ‚Was‘, ‚Wann‘, ‚Wo‘ und ‚Wie‘ erhoben oder gemessen werden soll, beantwortet.



Auf Ebene der Konzepterstellung folgt eine Fixierung der Definitionsmerkmale und die Festlegung der Objektgruppe Erhebungsmerkmale. Erhebungsmerkmale setzen sich aus einem Merkmalsnamen und dem Zuordnungsobjekt Verschlüsselung zusammen. Die Verschlüsselung beinhaltet die eigentliche Qualität eines Merkmals, da über sie die Verbindung zur Sachdatenebene hergestellt wird. Die Sachdaten sind die eigentlichen Messergebnisse, die entweder als Zeichen (Schlüssel) oder als numerische Größe (Wertebereich) vorliegen. Sie sind als konsensuale, symbolische Darstellung eines Befundes aus der Realität zu verstehen und werden durch die Metainformation in den Ausprägungstexten erklärt.

Schlüsseltableneinträge

	lfd. Nr.	Schlüssel	Ausprägung
	1	1	Männlich
	2	2	Weiblich

Beispiel einer Verschlüsselung:

Objekt Schlüsseltabelle

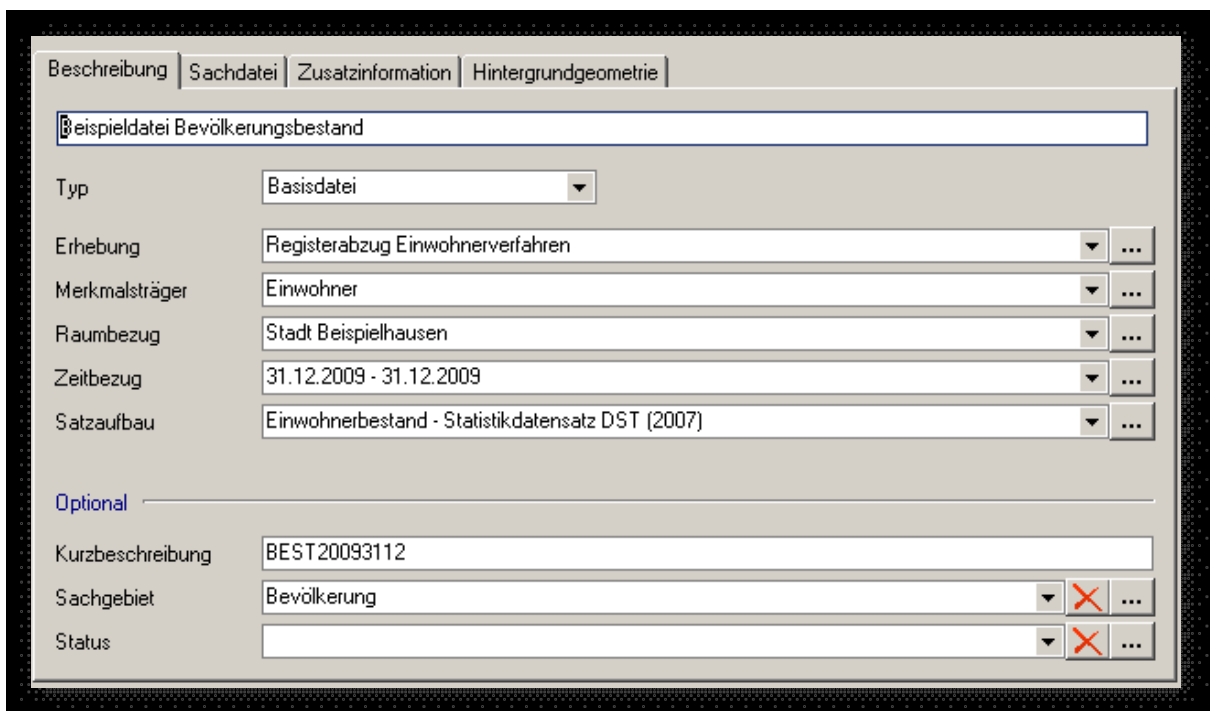
Die Schlüssel 1 und 2 werden durch das Metadatenobjekt 'Ausprägung' (hier die Ausprägungen ‚Männlich‘ und ‚Weiblich‘) beschrieben.

Sämtliche Erhebungsmaße und die entsprechend zugeordneten Verschlüsselungen erhalten in einem Satzaufbau ihre strukturelle Ordnung. Das Metadatenobjekt Satzaufbau setzt sich aber nicht nur aus Merkmalsnamen und den zugeordneten Verschlüsselungen zusammen, sondern enthält durch die Metadatenobjekte Merkmalstyp und Alias weitere Attribute. Der Merkmalstyp verweist auf die Datenstruktur einer Sachdatei, also dem Objekt, das z.B. als Textdatei oder Datenbanktabelle die Ergebnisse einer Erhebung oder eines Datenabzugs enthält. So kann die Datenstruktur entweder Einzeldatensätze enthalten (sequenzielle Struktur) oder Datensätze mit Häufigkeiten (aggregierte Struktur). Aliase sind dann von Relevanz, wenn die Datenhaltung in Datenbanktabellen vorgesehen ist, bzw. wenn Datenabzüge als Datenbanktabellen geliefert werden.



lfd. Nr.	Position	Merkmal	Verschlüsselun	Typ	Länge	Feldtyp	Feldalias
1	1	Gemeindegemeinschaft der Basisgemeinde	String 8-stellig	I	8	Ordnungsmerk	GEMEINDESCHLUESSEL_D
2	9	Straßenschlüssel der Basisadresse	String 5-stellig	I	5	Ordnungsmerk	STRASSENSCHLUESSEL_D
3	14	Hausnummer der Basisadresse	String 4-stellig	I	4	Ordnungsmerk	HAUSNUMMER_DER_BASIS
4	18	Hausnummernzusatz der Basisadresse	String 2-stellig	I	2	Ordnungsmerk	HAUSNUMMERNZUSATZ_DE
5	20	Laufende Nummer des Datensatzes an der Basis	String 4-stellig	I	4	Ordnungsmerk	LAUFENDE_NUMMER_DES_
6	24	Kleinräumige Gliederung	String 7-stellig	I	7	Übriges Merkmal	KLEINRAEUMIGE_GLIEDE
7	31	Wohnungsstatus der Person an der Basis	Wohnungsstatus c	S	1	Übriges Merkmal	WOHNUNGSSTATUS_DER_P

Die Verknüpfung der Metadatenobjekte aus denen sich der Satzaufbau zusammensetzt (Erhebungsmerkmale) mit den Metadatenobjekten, die den Rahmen einer Datenerhebung oder eines -Abzugs darstellen (Definitionsmerkmale), erfolgt in der Konzeptbeschreibung (DUVA-Terminologie: Dateibeschreibung). Eine Konzeptbeschreibung dient als Grundlage für Entwürfe von Fragebogenlayouts, Online-Befragungsmasken oder als Grundlage für die Definition von Datenabzügen z.B. aus operativen Verfahren. Obwohl die Konzeptbeschreibung eine explizite Datensatzbeschreibung enthält und z.B. auch Pfadangaben, muss eine Datei mit Sachdaten physikalisch noch gar nicht vorhanden sein. Liegt eine Sachdatei bereits vor, ist es unproblematisch, den in der Prozesskette (theoretisch) vorgelagerten Schritt an Beschreibungsinformation nachzuordnen.



Beschreibung | Sachdatei | Zusatzinformation | Hintergrundgeometrie

Beispieldatei Bevölkerungsbestand

Typ: Basisdatei

Erhebung: Registerabzug Einwohnerverfahren

Merkmalsträger: Einwohner

Raumbezug: Stadt Beispielhausen

Zeitbezug: 31.12.2009 - 31.12.2009

Satzaufbau: Einwohnerbestand - Statistikdatensatz DST (2007)

Optional

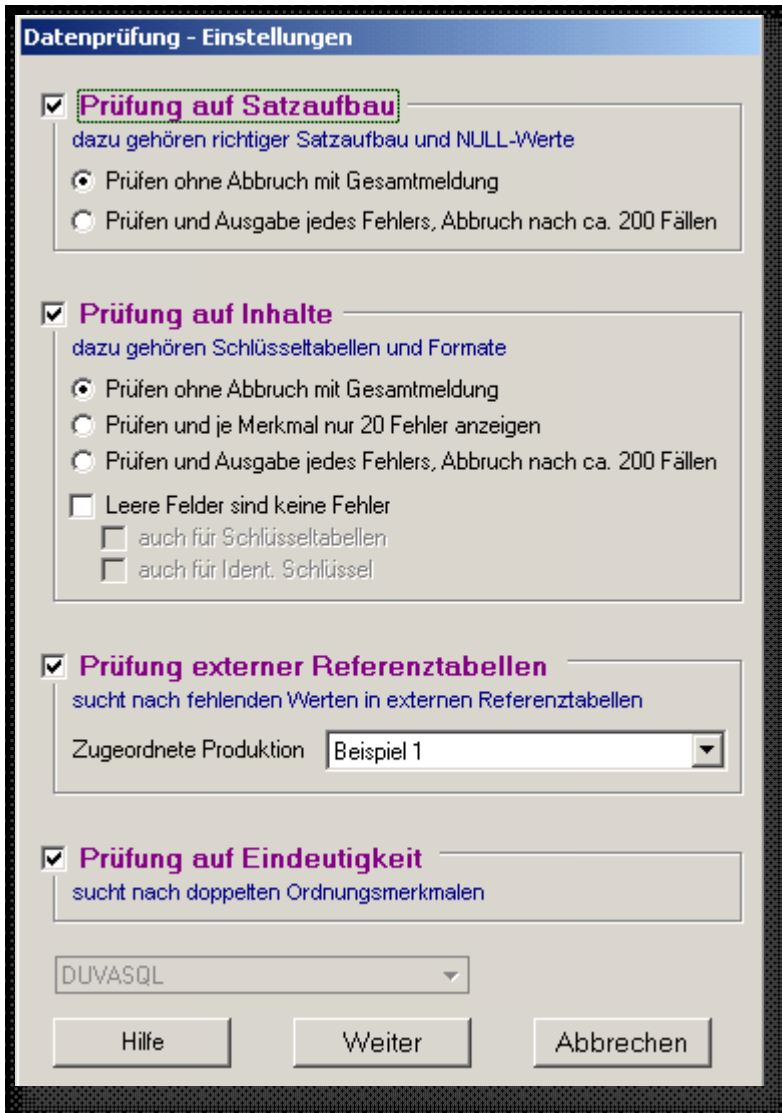
Kurzbeschreibung: BEST20093112

Sachgebiet: Bevölkerung

Status:

Weitere Inhalte einer Konzeptbeschreibung können auch unstrukturierte Metainformationen (Metatexte) sein. So zum Beispiel die Gesamtheit an methodischen, organisatorischen und personellen Anweisungen, die unter Umständen gefordert sind oder benötigt werden. Zur Aufnahme dieser Informationen stehen mehrere Metadatenobjekte zur Verfügung (z.B. Definitionen, Zusatzinformationen, Qualität). Die Festlegung von Metadatenobjekten aus denen sich Konzept- oder Dateibesreibungen zusammensetzen, führt zu einer strukturellen Standardisierung, die Grundlage für Mehrfachnutzungen von Konzeptbeschreibungen ist. Hat beispielsweise ein Anwender eine Konzeptbeschreibung erstellt, kann sie ein weiterer Anwender nach geringfügigen Veränderungen (z.B. Zeit-/ Raumbezug etc.) verwenden. Im Zusammenhang mit der Datenerhebung oder dem Datenabzug entstehen weitere Metadaten (z.B. Rücklaufquote), die sich jedoch nachträglich in die Dateibesreibung aufnehmen lassen bzw. Ergänzungen darstellen.

Die Prozessstufen Datenaufbereitung und –Analyse lassen sich nicht immer klar trennen, denn die Aufbereitung einer Datei mit Einzeldatensätzen zu einer Aggregatdatei stellt bereits aufgrund von Merkmalsselektionen einen Analyseschritt (deskriptive Methode) dar. Kern einer Datenaufbereitung sind zunächst Qualitätsüberprüfungen oder Plausibilisierungen und – falls erforderlich – der Einsatz von Anonymisierungsverfahren.



Datenprüfung - Einstellungen

Prüfung auf Satzaufbau
dazu gehören richtiger Satzaufbau und NULL-Werte

Prüfen ohne Abbruch mit Gesamtmeldung
 Prüfen und Ausgabe jedes Fehlers, Abbruch nach ca. 200 Fällen

Prüfung auf Inhalte
dazu gehören Schlüsselstabellen und Formate

Prüfen ohne Abbruch mit Gesamtmeldung
 Prüfen und je Merkmal nur 20 Fehler anzeigen
 Prüfen und Ausgabe jedes Fehlers, Abbruch nach ca. 200 Fällen

Leere Felder sind keine Fehler
 auch für Schlüsselstabellen
 auch für Ident. Schlüssel

Prüfung externer Referenztabellen
sucht nach fehlenden Werten in externen Referenztabellen

Zugeordnete Produktion:

Prüfung auf Eindeutigkeit
sucht nach doppelten Ordnungsmerkmalen

DUVASQL

Hilfe Weiter Abbrechen

Plausibilisierungen dienen der Überprüfung, ob die auf der Metadatenebene beschriebenen Schlüssel mit den Schlüsseln auf Ebene der Sachdaten übereinstimmen. Weitere Überprüfungsschritte sind Strukturüberprüfungen (z.B. Satzlänge) oder auch die Eindeutigkeit von Primärschlüsseln (Ordnungsmerkmale), die jeden einzelnen Datensatz eindeutig machen. Bestehen Referenzierungen (Umschlüsselungen), die eine Ableitungsregel zwischen zwei voneinander abhängigen Dateien beschreiben, so hat hier die Plausibilisierung die Aufgabe der Überprüfung von Zuordnungseindeutigkeit und -vollständigkeit.

Der Einsatz von Anonymisierungsverfahren ist dann erforderlich, wenn Dateien oder Tabellen aus Gründen der statistischen Geheimhaltung keine Fälle mit Häufigkeiten von 1 oder 2 enthalten sollen. Anonymisierungsverfahren lassen sich sowohl auf Ebene von Dateien mit Einzeldatensätzen als auch bei Aggregatdatensätzen anwenden (Beispiel Programm zur Geheimhaltung SAFE).



Weitere Teilprozesse der Datenaufbereitung sind Merkmalsselektionen und Merkmalsableitungen wie Gruppierungen, Filter oder der Einsatz von Algorithmen. Im Allgemeinen gibt es hier zwei unterschiedliche Vorgehensweisen: entweder man führt die Selektionen und/oder Ableitungen auf Ebene der originären Statistikdatei durch oder man produziert ein neues Dateiojekt (Zieldatei) unter Beibehaltung der ursprünglichen Datei (Quelldatei). Der stringentere Weg, der vor allem auch reversibel ist, setzt sich aus einer Produktion zusammen. Produktionen und Ableitungsregeln stellen somit weitere Metadatenobjektgruppen dar. Das heißt, Metadaten dienen zur Prozesssteuerung einer Produktion.

The screenshot shows the 'Editor für Produktionsbeschreibung - (Beispiel 2)' window. The main area displays a table with the following data:

		Merkmal	Verschlüsselung	Typ	Länge	Feldtyp	Ableitungsregel
1	1	Wohnort	Orte im	S	8	K	BAS1[2];
2	9	Eingekaufte Artikel	Artikel	S	1	K	BAS1[4];
3	10	Einkaufsbetrag (absolut)	Anzahl 4-stellig	N	4	G	
4	14	Einkaufsbetrag (durchschnittlich)	Zahl (3,2)	N	3,2	G	
5	20	Kunden	Anzahl 4-stellig	N	4	G	

A context menu is open over the 'Einkaufsbetrag (absolut)' row, listing the following options:

- Direkte Merkmalsübernahme
- Modifizierte Übernahme
- Interne Referenztafel
- Externe Referenztafel
- Bedingungsliste
- Algorithmus
- Zwischendateischlüsselreferenz
- Benutzerdefinierte Regel
- Aggregatfunktion (Summenfeld) ▶
- Regel lokalisieren...

Red boxes in the image highlight 'Quelldatei' (pointing to the source file path), 'Zieldatei' (pointing to the target file path), and 'Ableitungsregeln' (the context menu).

SPSS/SAS

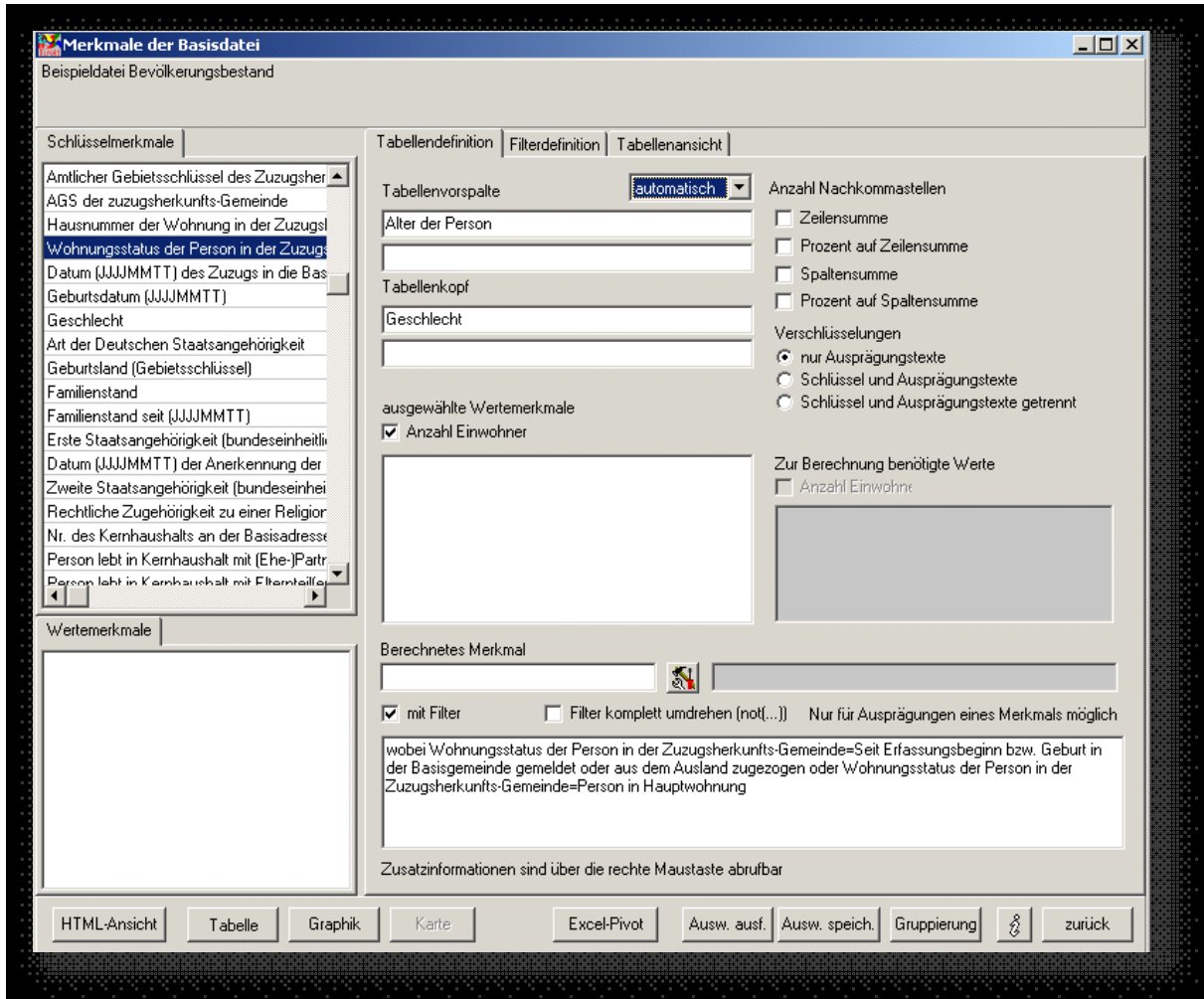
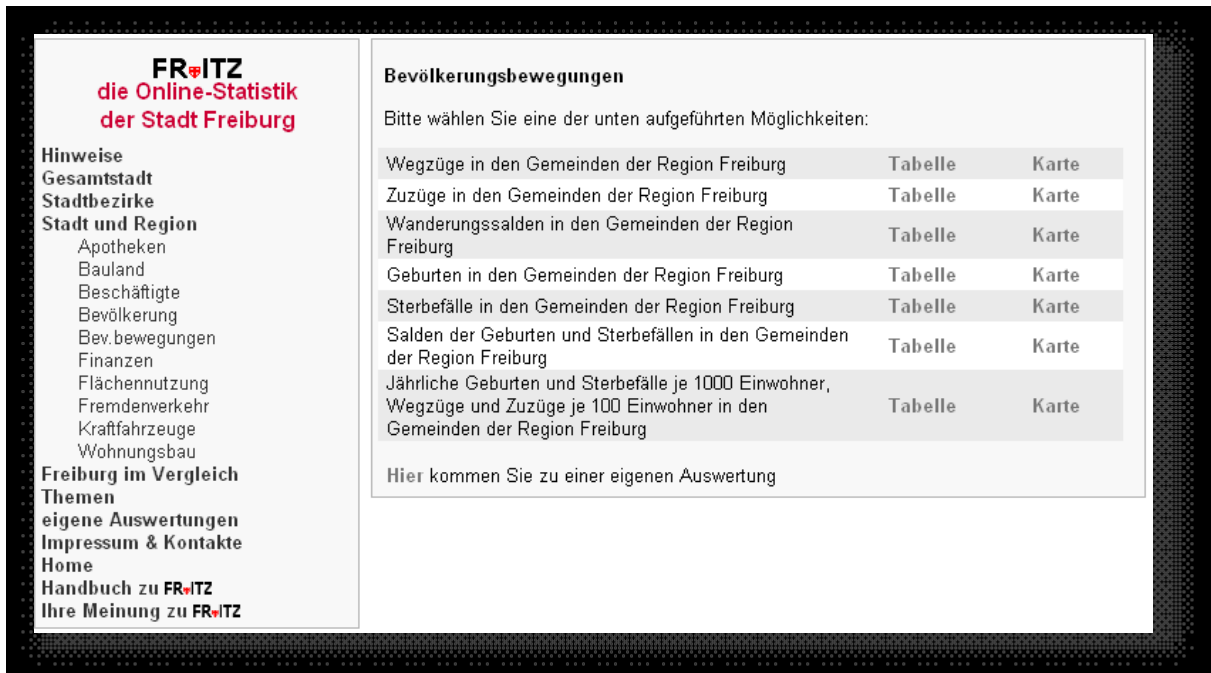
Aufbau | SPSS Code | SAS Code

```

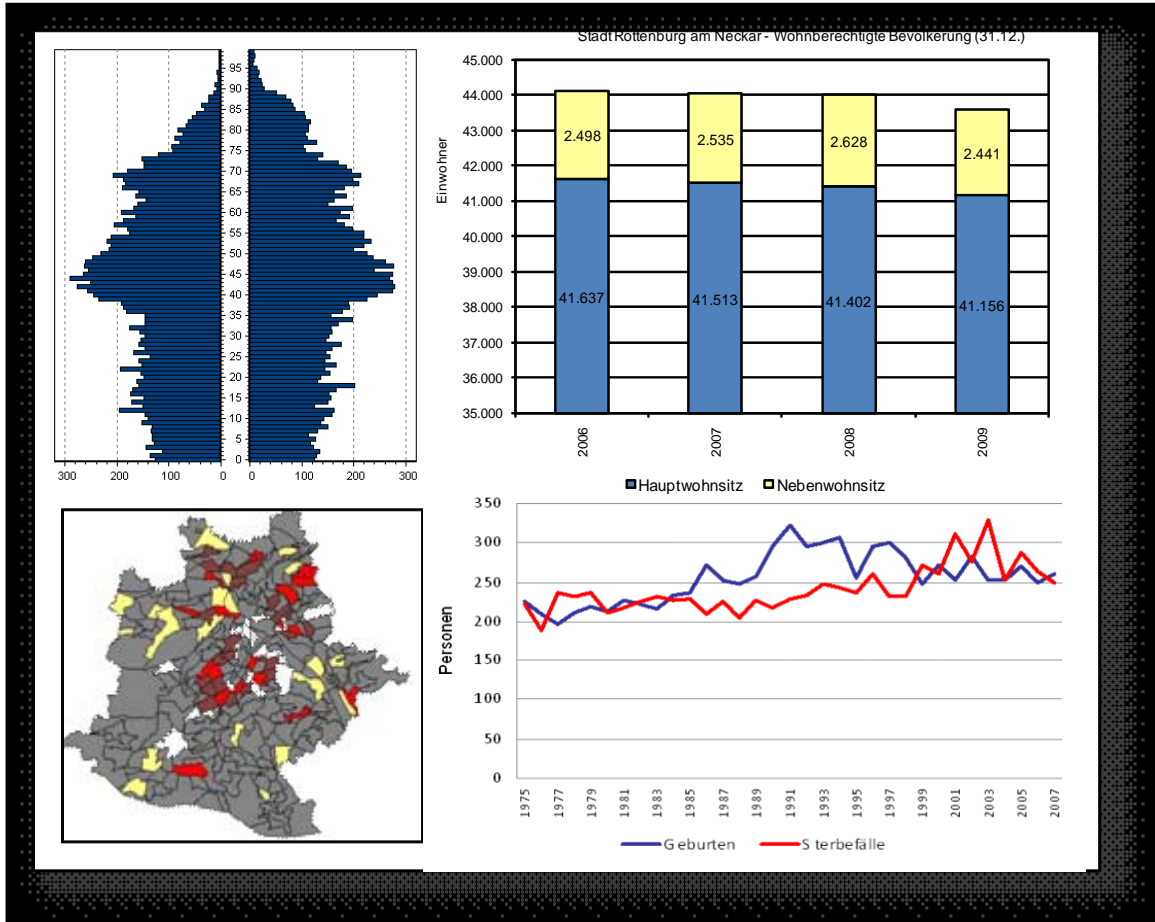
Title 'Beispieldatei Bevölkerungsbes
data list file = "DUVA4S::t0000"/
GEMEINDESCHLUESSEL_D 1 - 8 (a)
STRASSENSCHLUESSEL_D 9 - 13 (a)
HAUSNUMMER_DER_BASIS 14 - 17 (a)
HAUSNUMMERNZUSATZ_DE 18 - 19 (a)
LAUFENDE_NUMMER_DES_ 20 - 23 (a)
KLEINRAEUMIGE_GLIEDE 24 - 30 (a)
WOHNUNGSSTATUS_DER_P 31 (a)
DATUM_JJJMMTT_DES_E 32 - 39 (a)
    
```

The screenshot shows the 'Faktorenanalyse' dialog box. The 'Variablen:' list is empty. The 'Auswahlvariable:' field is also empty. The 'einfluss' variable is selected in the list on the left.

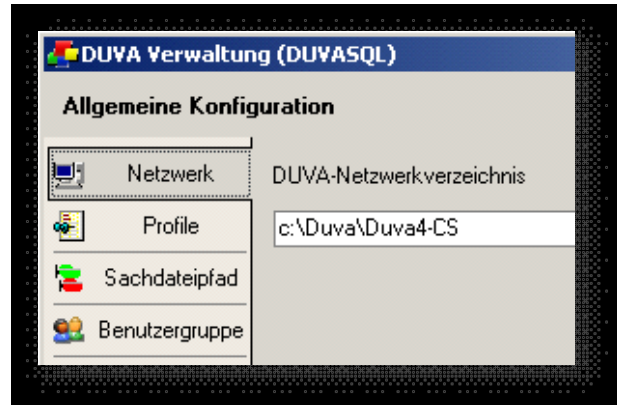
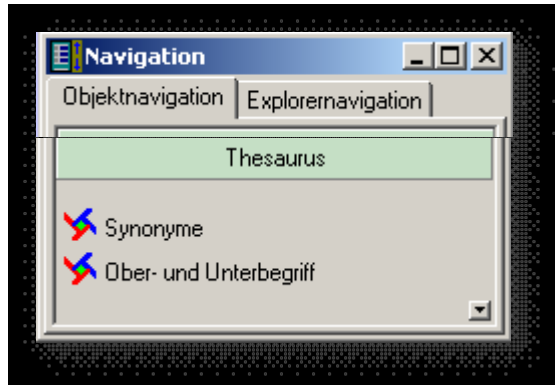
Die Datenanalyse im engeren Sinne beinhaltet statistische Verfahren (deskriptive und schließende Methoden), die Verdichtungen auf Ebene der Erhebungsmerkmale durchführen. Hier ist entscheidend, dass nicht nur die Schlüssel oder Werte (Sachdaten) vom entsprechenden Analyseinstrument bearbeitet werden, sondern dass auch die Beschreibungsinformation (Metainformation) mitgeführt wird. Dafür bedarf es entsprechender Metadatenchnittstellen oder eines Direktzugriffs von Modulen auf die Metadatenbank (Beispiel DUVA-Modul ‚Mdirekt‘, unten DUVA-Internet-Assistent).

Die Datenweitergabe in Form von Tabellen, Grafiken, Karten sowie textlichen Erläuterungen und Interpretationen bilden die letzte Stufe im Statistischen Produktionsprozess. Metadatenbasiertes Informationsmanagement hat den Vorteil, dass die entsprechenden Darstellungen (mit Ausnahme der textlichen Erläuterungen und Interpretationen) sowohl Sachinformationen als auch Metainformationen inklusive der Definitionsmerkmale enthalten.



Um die sprachliche Interaktion zwischen metadatenbasierte Informationsverarbeitung und den Benutzern eines Systems zu ermöglichen bedarf es einer Organisation der Beschreibungstexte inklusive deren möglicher semantischer Beziehungen wie Über-/ Unterordnung oder Synonyme. Dies kann durch die Einbindung eines Thesaurus erfolgen, der die Schnittstelle zwischen Benutzer und System herstellt. Metadatenbasierte Informationsverarbeitung umfasst aber nicht die Kommunikation eines Benutzers mit einer Metadatenbank, sondern mehrerer oder theoretisch unendlich vielen Benutzern. Die logische Konsequenz ist, dass auch die Systemadministrierung durch Metadatenobjekte organisiert und gesteuert werden muss.

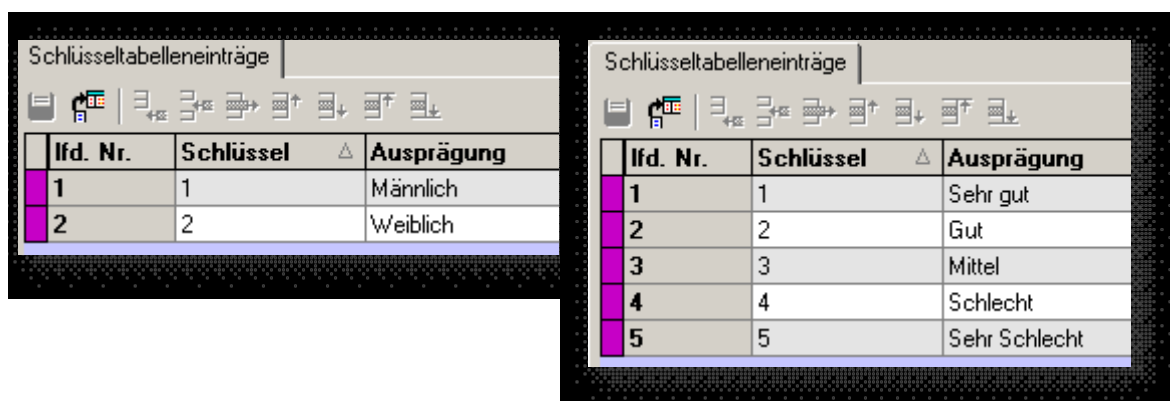


1.3 Das DUVA- Metadatenmodell

Das vorangegangene Kapitel gab einen Überblick über den Ablauf des statistischen Produktionsprozesses, wobei die Aufzählung der Metadatenobjekte nicht abschließend war, sondern lediglich das Thema Granulierung von Prozesskomponenten anreißen konnte. Ausgehend vom Granulierungsansatz (und nicht aus Sicht des Prozessablaufs) sollen folgend die Metadatenobjekte, wie sie in DUVA organisiert sind, in ihrer Hierarchie dargestellt werden. Die hierarchische Darstellung dient allerdings nur der Gliederung, in der praktischen Arbeit mit DUVA ist einer hierarchische Vorgehensweise nicht zwingend.

1.3.1 Verschlüsselungen

An der hierarchisch untersten Stufe stehen Merkmalsausprägungen, die je nach Messung einen Schlüssel oder einen Wert darstellen. Bei den Schlüsseln ist zu untergliedern, ob sie gleichrangig nebeneinander stehen (Nominalskala) oder eine natürliche Reihenfolge abbilden (Ordinalskala). Die Beschreibung der Schlüssel erfolgt durch Ausprägungstexte. Zuordnungen von Schlüsseln (Sachdaten) zu Ausprägungstexten (Metadaten) werden als Schlüsselstabellen bezeichnet.

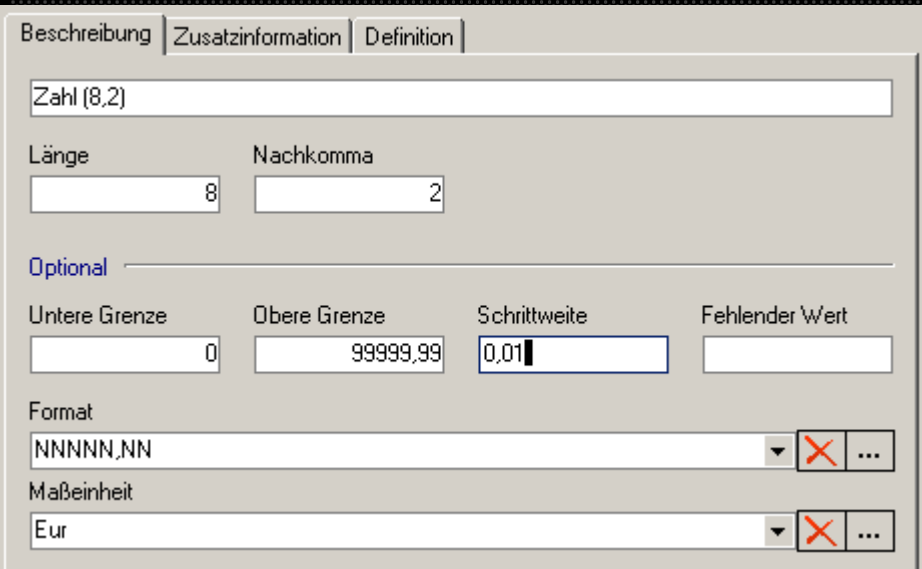


Im Beispiel links ist einmal eine Schlüsselstabelle Geschlecht mit den Schlüsseln 1 und 2 und den Ausprägungstexten männlich und weiblich (Nominalskala) dargestellt. Bei der Schlüsselstabelle rechts handelt es sich um eine Befindlichkeitsskala nach v. Zerssen (Ordinalskala). Schlüsselstabellen setzen

sich aus einem Beschreibungstext (s. unten), der Längenangabe (Stellen), den optionalen Angaben von fehlendem Wert, Gültigkeit, Sachgebiet und den Schlüsseltableneinträgen zusammen.

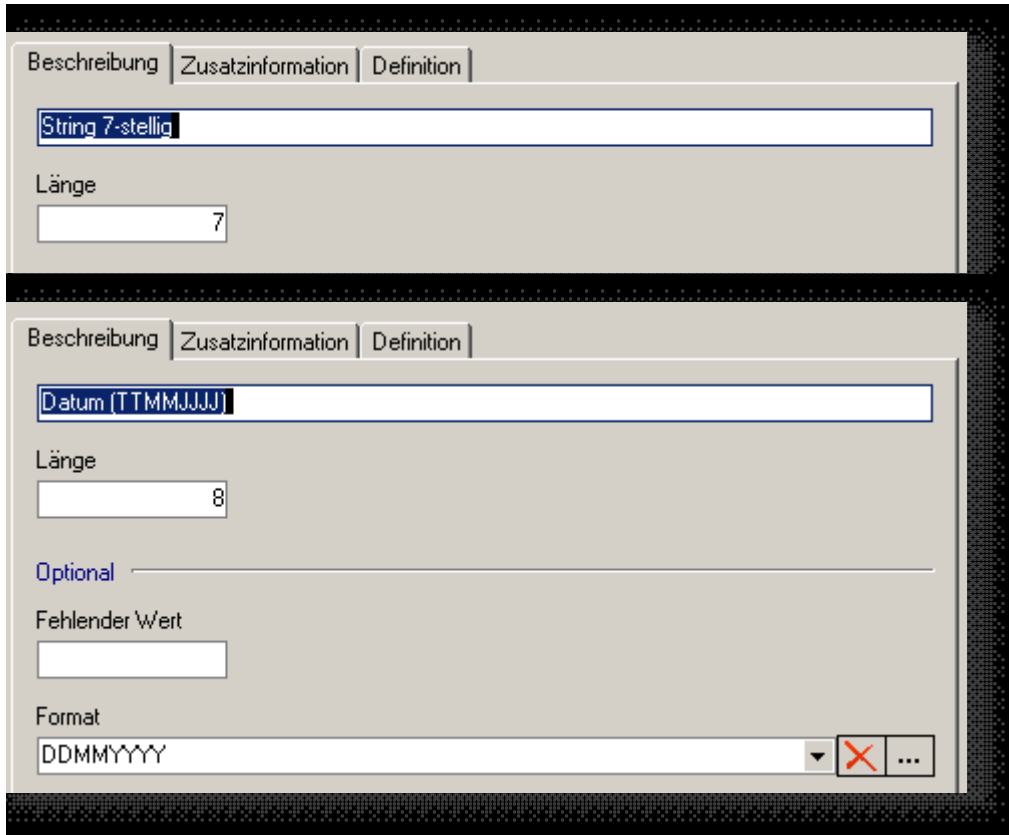


Handelt es sich bei Merkmalsausprägungen um Werte, so sind die Beschreibungsinformationen keine Ausprägungstexte, sondern Länge (Anzahl an Stellen insgesamt) und Formatangaben (in welchem Zahlenformat liegen die Daten in der Sachdatei vor) und eine Maßeinheit. Optionale Angaben sind Angaben von Minimum und Maximum sowie Schrittweite und fehlender Wert.



In der angewandten Statistik kommt es häufig vor, dass Schlüssel nicht explizit durch Ausprägungstexte beschrieben werden. Beispielsweise ist dies bei einer laufenden Nummer oder bei

einigen räumlichen Ordnungssystemen der Fall (z.B. Kleinräumige Gliederung). Diese Verschlüsselungsart trägt bei DUVA den Namen Identifizierender Schlüssel (Schlüssel erklärt sich selbst ohne zusätzliche Textinformation).



The image shows two screenshots of the DUVA software interface for defining keys. Both screenshots have three tabs: 'Beschreibung', 'Zusatzinformation', and 'Definition'.
 The top screenshot shows a key definition for a 'String 7-stellig'. The 'Beschreibung' tab is active, and the text 'String 7-stellig' is entered in the main field. Below it, the 'Länge' (length) is set to 7.
 The bottom screenshot shows a key definition for a 'Datum (TTMMJJJJ)'. The 'Beschreibung' tab is active, and the text 'Datum (TTMMJJJJ)' is entered in the main field. Below it, the 'Länge' (length) is set to 8. There is an 'Optional' checkbox, a 'Fehlender Wert' (missing value) field, and a 'Format' dropdown menu set to 'DDMMYYYY'.

Im Beispiel oben ist eine Zeichenkette mit 7 Stellen eingetragen. Diese Verschlüsselung könnte beispielsweise für das Merkmal Kleinräumige Gliederung mit ihren 7 Stellen verwendet werden. Identifizierende Schlüssel eignen sich auch zur Verschlüsselung von Dummy-Feldern oder auch von Texteinträgen (Textfelder mit maximal 255 Stellen). Ein weiterer wichtiger Einsatzbereich von Identifizierenden Schlüsseln sind Datumseinträge. Dann ist allerdings die Angabe eines entsprechenden Formats erforderlich.

1.3.2 Merkmale

Im Unterschied zu vielen Statistikprogrammen, bei denen die Verschlüsselung unmittelbar mit einem Merkmal verknüpft ist, sieht DUVA diese beiden Metadatenobjekte jeweils eigenständig. Somit reduziert sich die Beschreibung eines Merkmals lediglich auf den Merkmalsnamen. Die Verknüpfung von Merkmalsnamen und der dazugehörigen Verschlüsselung erfolgt im Satzaufbau einer Dateibeschreibung. Die Trennung von Verschlüsselung und Merkmal hat auf der einen Seite systematische Gründe (Stringenz), auf der anderen Seite lässt sich so eine Verschlüsselung mehreren Merkmalen zuordnen. Setzt sich beispielsweise ein Fragebogen aus 20 Erhebungsmerkmalen, die mit ja und nein beantwortet werden können zusammen, so sind für die Metadatenbeschreibung zwar 20 Merkmalsnamen einzutragen aber nur eine Verschlüsselung (nämlich die Antwortkategorie ja/nein).

lfd. Nr.	Position	Merkmal	Verschlüsselung	Typ
1	1	Laufende Nummer des Fragebogens	String 4-stellig	I
2	5	Orte im Landkreis Beispielhausen	Orte Landkreis Beispielhau	S
3	13	Verkehrsmittel	Verkehrsmittel	S
4	14	Lebensmittel	Antwortkategorie ja/nein	S
5	15	Drogeriewaren	Antwortkategorie ja/nein	S
6	16	Schuhe	Antwortkategorie ja/nein	S
7	17	Kleidung	Antwortkategorie ja/nein	S
8	18	Spielwaren	Antwortkategorie ja/nein	S
9	19	Sonstige	Antwortkategorie ja/nein	S
10	20	Gesamtbetrag für alle Einkäufe	Ganzzahl (4-stellig)	W

1.3.3 Satzaufbauten

Satzaufbauten beinhalten die Erhebungsmerkmale (Merkmalsnamen und ihre Verschlüsselungen) und die Strukturinformation (welches Merkmal steht an welcher Stelle) einer Dateibeschreibung. Je nach Vorgehensweise im Statistischen Produktionsprozess hat ein Satzaufbau die Qualität einer Strukturvorschrift (ein Datenabzug soll exakt nach dieser Vorgabe erzeugt werden) oder einer nachgeordneten Beschreibung (Beschreibung einer bereits vorliegenden Sachdatei). In Abhängigkeit vom Aufbau einer Sachdatei unterscheiden sich Satzaufbauten nach Basisdateisatzaufbauten und Makrodateisatzaufbauten. Ein Basisdateisatzaufbau setzt sich aus Einzeldatensätzen zusammen, ein Makrodateisatzaufbau aus Aggregatdatensätzen.

Der Satzaufbau von Basisdateien lässt sich am einfachsten anhand einer Sachdatei (Textdatei) mit sequenziellem Aufbau erklären. Bei dieser Form ist für jedes Merkmal exakt ein Feld im Satzaufbau vorgesehen (andere Strukturen sind auch bearbeitbar). Die Feldlänge je Merkmal ist durch die Länge eines Schlüssels (z.B. 1-stelliger Schlüssel zur Verschlüsselung des Verkehrsmittels) festgelegt. Jedem Merkmal ist jeweils nur eine Verschlüsselung zugeordnet. Die Anordnung der einzelnen Merkmale im Satzaufbau ist je Datensatz identisch. Jeder einzelne Datensatz entspricht einer Erhebungseinheit. Sind beispielsweise Personen Merkmalsträger, so ist für jede Person jeweils ein Datensatz abgebildet. Eine Basisdatei bzw. ihr Satzaufbau bezieht sich jeweils nur auf einen eindeutigen Merkmalsträger, der über einen Ordnungsbegriff definiert ist.

Im Beispiel auf S. 19 ist das erste Merkmal im Satzaufbau (Laufende Nummer des Fragebogens) mit der Verschlüsselung String 4-Stellig (einem 4-stelligen Identifizierenden Schlüssel) als Ordnungsmerkmal festgelegt. Dann folgen die weiteren Merkmale und ihre Verschlüsselungen. Unabhängig von ihrer inhaltlichen Qualität sind sie vom Feldtyp Übriges Merkmal. An den Feldtyp schließt sich in der Beschreibung noch der Aliasname eines Merkmals an (ist hier ausgeblendet).

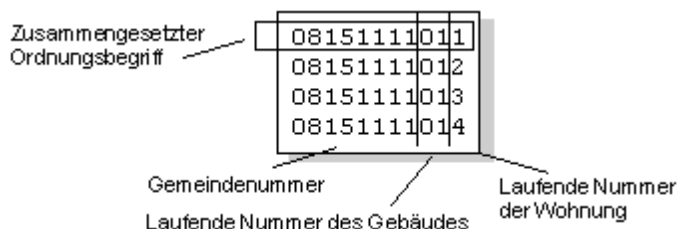
Ild. Nr.	Position	Merkmal	Verschlüsselung	Typ	Länge	Feldtyp
1	1	Laufende Nummer des Fragebogens	String 4-stellig	I	4	Ordnungsmerkmal
2	5	Orte im Landkreis Beispielhausen	Orte Landkreis Beispielhau	S	8	Übriges Merkmal
3	13	Verkehrsmittel	Verkehrsmittel	S	1	Übriges Merkmal
4	14	Lebensmittel	Antwortkategorie ja/nein	S	1	Übriges Merkmal
5	15	Drogeriewaren	Antwortkategorie ja/nein	S	1	Übriges Merkmal
6	16	Schuhe	Antwortkategorie ja/nein	S	1	Übriges Merkmal
7	17	Kleidung	Antwortkategorie ja/nein	S	1	Übriges Merkmal
8	18	Spielwaren	Antwortkategorie ja/nein	S	1	Übriges Merkmal
9	19	Sonstige	Antwortkategorie ja/nein	S	1	Übriges Merkmal
10	20	Gesamtbetrag für alle Einkäufe	Ganzzahl (4-stellig)	W	4	Übriges Merkmal
11	24	Geburtsdatum (JJJJMMTT)	String 8-stellig	I	8	Übriges Merkmal

0001	08151222	5	0	1	1	1	0	1	0050	05051942
0002	08151333	2	0	1	1	0	0	1	0034	18011969
0003	08151333	4	0	0	1	0	0	1	0021	02121981
0004	08151111	1	1	0	0	1	1	1	0044	19111951
0005	08151111	1	0	1	0	0	1	0	0022	07031953
0006	08151111	4	0	1	0	1	1	0	0039	18071972
0007	08151666	2	0	1	1	0	1	0	0031	01081962
0008	08151666	6	1	0	0	0	1	1	0032	25101945
0009	08151777	1	1	0	0	0	1	1	0033	09071939
0010	08151444	1	1	1	1	0	0	1	0045	06021942
0011	08151777	1	1	0	0	0	1	0	0022	24011952

Beschreibung Satzaufbau

Sachdatei

Ein Ordnungsbegriff kann sich aus einem oder mehreren hierarchisch aufgebauten Ordnungsmerkmalen zusammensetzen. Im Beispielfall oben ist die 4-stellige Nummer (z.B. eines Fragebogens) das Ordnungsmerkmal. Das Ordnungsmerkmal/ die Ordnungsmerkmale in einem Satzaufbau sind für sämtliche Verarbeitungsprozesse (Stichwort Produktion) grundlegend. Kommt ein als Ordnungsbegriff definierter Schlüssel in einer Sachdatei mehrfach vor, kann eine Produktion nicht erfolgreich sein, da die Eindeutigkeit des Zählobjekts nicht mehr gegeben ist.



Beispiel: Ordnungsbegriff, der sich aus den Ordnungsmerkmalen 'Gemeindenummer', 'Laufende Nummer des Gebäudes' und 'Laufende Nummer der Wohnung im Gebäude' zusammensetzt. Geht man z.B. davon aus, dass die 'Laufende Nummer der Wohnung im Gebäude' für jedes Gebäude immer bei 1 beginnt, dann wäre dieses Merkmal für sich genommen nicht als Satzidentifikator oder Primärschlüssel geeignet. Erst in Kombination mit der 'Gemeindenummer' und der 'Laufenden Nummer des Gebäudes' ergibt sich ein zusammengesetzter Ordnungsbegriff, über den jede Wohnung als Zählobjekt eindeutig ist

Satzaufbauten von Makrodateien haben aufgrund ihrer Struktur (Aggregatdatensätze) einen anderen Aufbau. Sie setzen sich aus Schlüsselfeldern und Summenfeldern sowie einem Zählfeld, das die entsprechenden Häufigkeiten der möglichen Ausprägungskombinationen enthält, zusammen. Die Häufigkeiten beziehen sich auf den zugrundeliegenden Merkmalsträger. Makrodateien lassen sich entweder über den Weg einer Produktion erzeugen oder liegen als Datenabzug bereits in dieser Struktur vor (z.B. Abzüge aus der Beschäftigungsstatistik der Bundesagentur für Arbeit). Bei Makrodateien ist die Bestimmung eines Ordnungsbegriffs nicht erforderlich, da die Zeichenfolgen der Schlüsselfelder definitionsgemäß eindeutig sind.

lfd. Nr.	Position	Merkmal	Verschlüsselung	Typ	Länge	Feldtyp	Feldalias
1	1	Orte im Landkreis Beispielhausen	Orte Landkreis Beispielhau	S	8	Schlüsselfeld	ORTE_IM_LANDKR
2	9	Verkehrsmittel	Verkehrsmittel	S	1	Schlüsselfeld	VERKEHRSMITTEL
3	10	Kunden	Ganzzahl (4-stellig)	W	4	Zählfeld	KUNDEN

08151111	1	0013
08151111	2	0001
08151111	3	0002
08151111	4	0002
08151111	6	0001
08151222	1	0007
08151222	3	0003
08151222	4	0003
08151222	5	0001
08151333	1	0010
08151333	2	0003
08151333	3	0002
08151333	4	0001
08151333	5	0001
08151444	1	0005
08151444	4	0001

Der Schlüssel im ersten Feld steht für einen bestimmten Ort, im zweiten Feld stehen die Schlüssel für die (hier insgesamt 6 vorkommenden) Verkehrsmittel. Im dritten Feld stehen die Häufigkeiten. Somit kommt im Beispiel hier die Kombination Ort= 08151111 und Verkehrsmittel=1 insgesamt 13mal vor.

Beschreibung Satzaufbau

Sachdaten

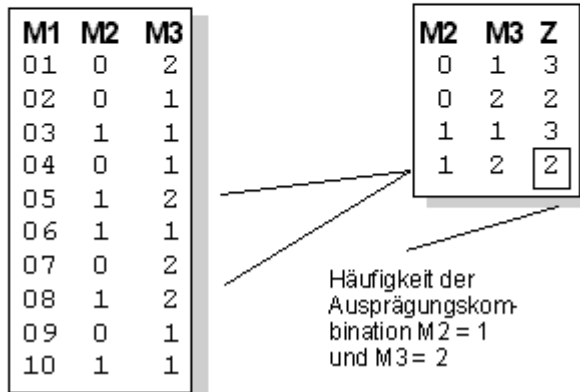
Um die Beziehung zwischen der Satzstruktur einer Basisdatei und einer Makrodatei auf Ebene der Sachdaten zu skizzieren, ein kleines Beispiel: die unten stehende Abbildung zeigt eine Basisdatei mit sequenziellem Aufbau (Merkmale 'Laufende Nummer' (M1), 'Neinsager/Jasager' (M2) und 'Geschlecht' (M3)). Die daraus generierte Makrodatei enthält sämtliche möglichen Ausprägungskombinationen der Merkmale 'Neinsager/Jasager' und 'Geschlecht' sowie ein Zählfeld mit der Angabe der Häufigkeiten (die Kombination 0 = Neinsager und 1= männlich hat eine Häufigkeit von 3 etc.).

Basisdatei

- M1** = Laufende Nummer
- M2** = Schlüsseltablette
'Neinsager/Jasager'
(0 = Neinsager, 1 = Jasager)
- M3** = Schlüsseltablette 'Geschlecht'
(1 = männl., 2 = weibl.)

Makrodatei

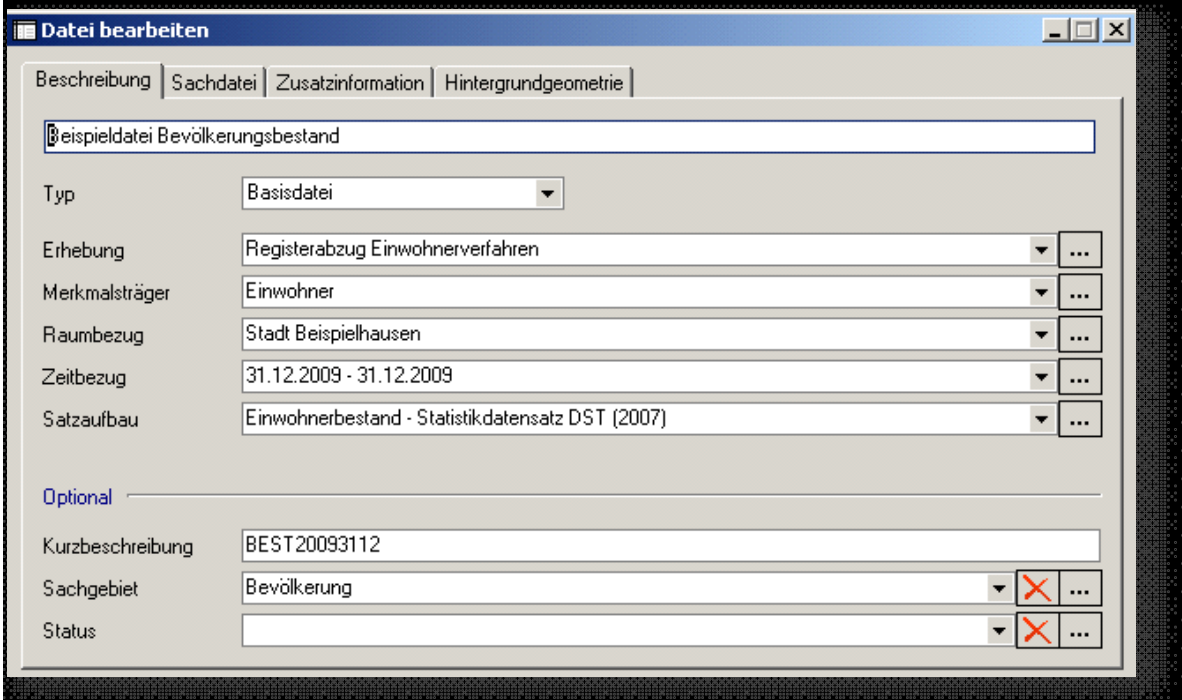
- M2, M3** = Schlüsselfelder
aus der Basisdatei
- Z** = Zählfeld



Aufgrund der Struktur des Satzaufbaus von Makrodateien (Kombination von Schlüsseln) lässt sich die Anzahl an Datensätzen aus dem Produkt der Ausprägungen je Merkmal berechnen. So hat beispielsweise eine Sachdatei (Makrodateiformat) mit den Merkmalen 100 Altersgruppen und Geschlecht 200 Datensätze (100 Ausprägungen bei den Altersgruppen und 2 Ausprägungen des Merkmals Geschlecht), und dies unabhängig davon, ob sich die zugrundeliegende Basisdatei aus 1.000 oder 100.000 Datensätzen zusammensetzt.

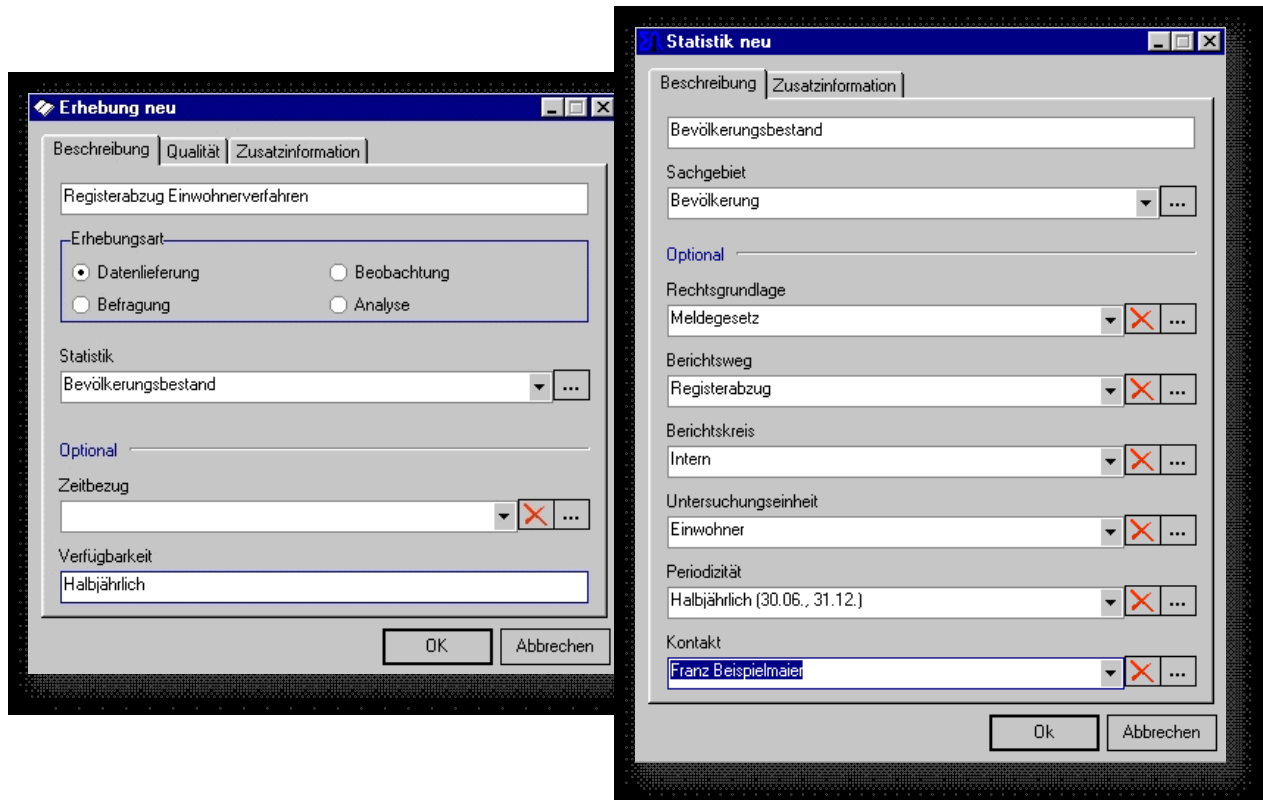
1.3.4 Dateibeschreibungen

Im allgemeinen Ansatz (Kapitel 1.2) wurden Dateibeschreibungen auch als Konzeptionen bezeichnet, da sie nicht nur die im Satzaufbau enthaltenen Erhebungsmerkmale enthalten, sondern sämtliche Definitionsmerkmale und zusätzliche Metadaten. In Abhängigkeit vom zugeordneten Satzaufbau kann es sich um Basisdateibeschreibungen oder Makrodateibeschreibungen handeln. Bezüglich der zugeordneten Definitionsmerkmale gibt es keinen grundlegenden Unterschied zwischen Basis- und Makrodateibeschreibungen. Am Beispiel einer Basisdateibeschreibung sollen im folgenden die einzelnen Zuordnungsobjekte vorgestellt werden:

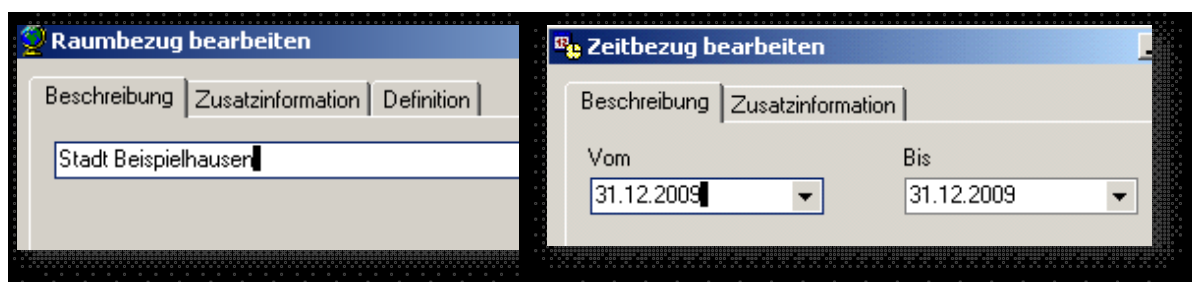


Dateibesreibungen werden im System über einen Beschreibungstext (hier: Beispieldatei Bevölkerungsbestand) abgelegt, recherchiert und für weitere Arbeiten aufgerufen. Das Auswahlkriterium Typ legt die Kategorie (Basis- oder Makrodatei) fest.

Das Objekt Erhebung (s.S. 22) gibt über die Art der Datenquelle oder die Art der Erhebung Auskunft, über die Qualität der Daten (Rücklaufquote, Systematische Fehler), optional über den Zeitbezug und die Verfügbarkeit. Ein Zuordnungsobjekt zur Erhebung stellt das Objekt Statistik dar. Statistiken fassen mehrere Dateibesreibungen innerhalb eines Sachgebiets zusammen. So kann beispielsweise das Sachgebiet Bevölkerung die Statistiken Bevölkerungsbestand, Bevölkerungsbewegungen, Haushalte, Bevölkerungsprognose, Mikrozensus Bevölkerung und Haushalte enthalten. Optionale Zuordnungsobjekte zur Statistik sind Rechtsgrundlage, Berichtsweg, Berichtskreis, Untersuchungseinheit, Periodizität und Kontaktdaten.

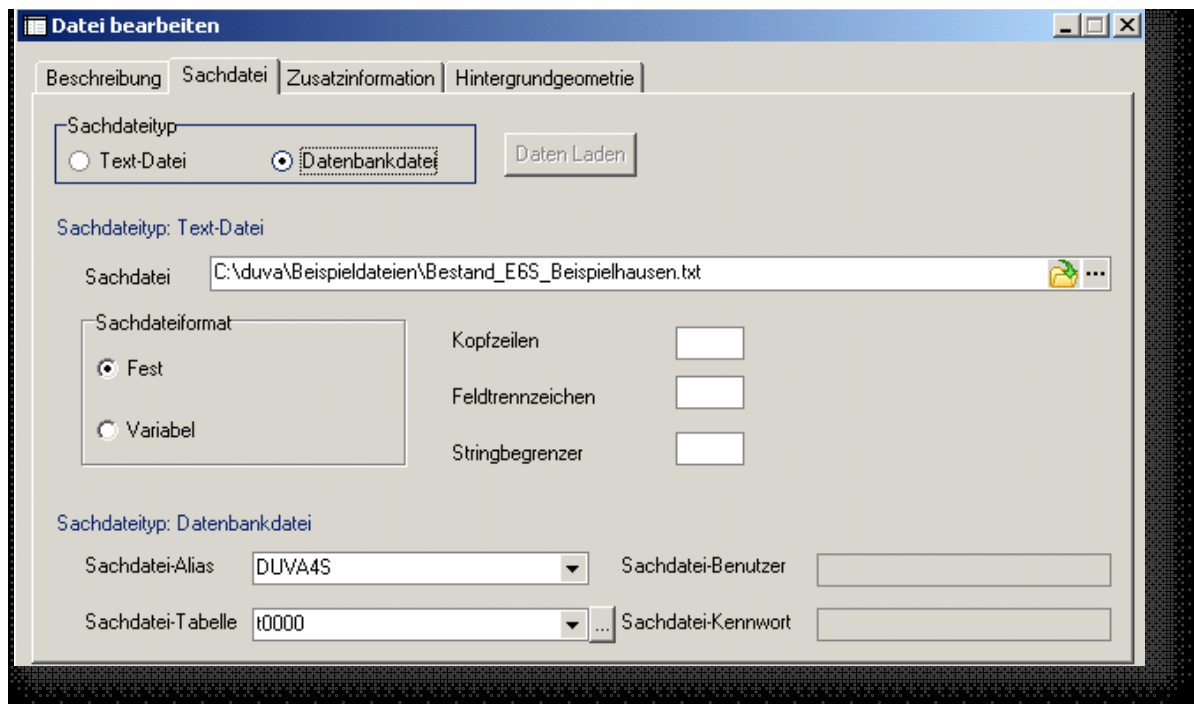
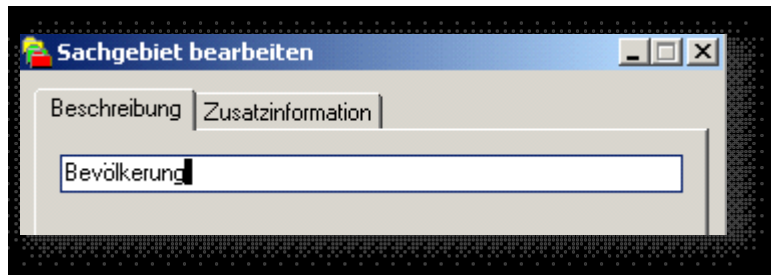


Zurück zur Dateibeschreibung: die Zuordnungsobjekte Raum- und Zeitbezug dienen der Eindeutigkeit einer Dateibeschreibung. Beim Zeitbezug sind die Angaben Vom...Bis entweder identisch (Bestandsmasse) oder geben ein Zeitintervall an (Bewegungsmasse).



Die Zuordnung eines Satzaufbaus zu einer Dateibeschreibung ist zwingend, wohingegen die Angaben Kurzbeschreibung (Kürzel, das bei Systematisierungen oder Katalogen verwendet werden kann), Status (Spezifikation, ob die Sachdaten z.B. durch ein Anonymisierungsverfahren verändert wurden) und Sachgebiet optional sind.

Unter Systematisierungserfordernissen spielt das Sachgebiet allerdings eine wichtige Rolle, da es eine fachliche Einheit umfasst. Im Unterschied zu Statistiken, die sich aus mehreren Sachgebieten zusammensetzen können, ist das Sachgebiet durch eine Gruppe von Merkmalsträgern abgegrenzt. So z.B. das Sachgebiet Einwohnerwesen durch die Merkmalsträger Personen und Personenzustände. Das Sachgebiet bildet eine Zuordnungseinheit zu Dateien, Satzaufbauten, Statistiken, Verschlüsselungen, Regeln und Produktionen.



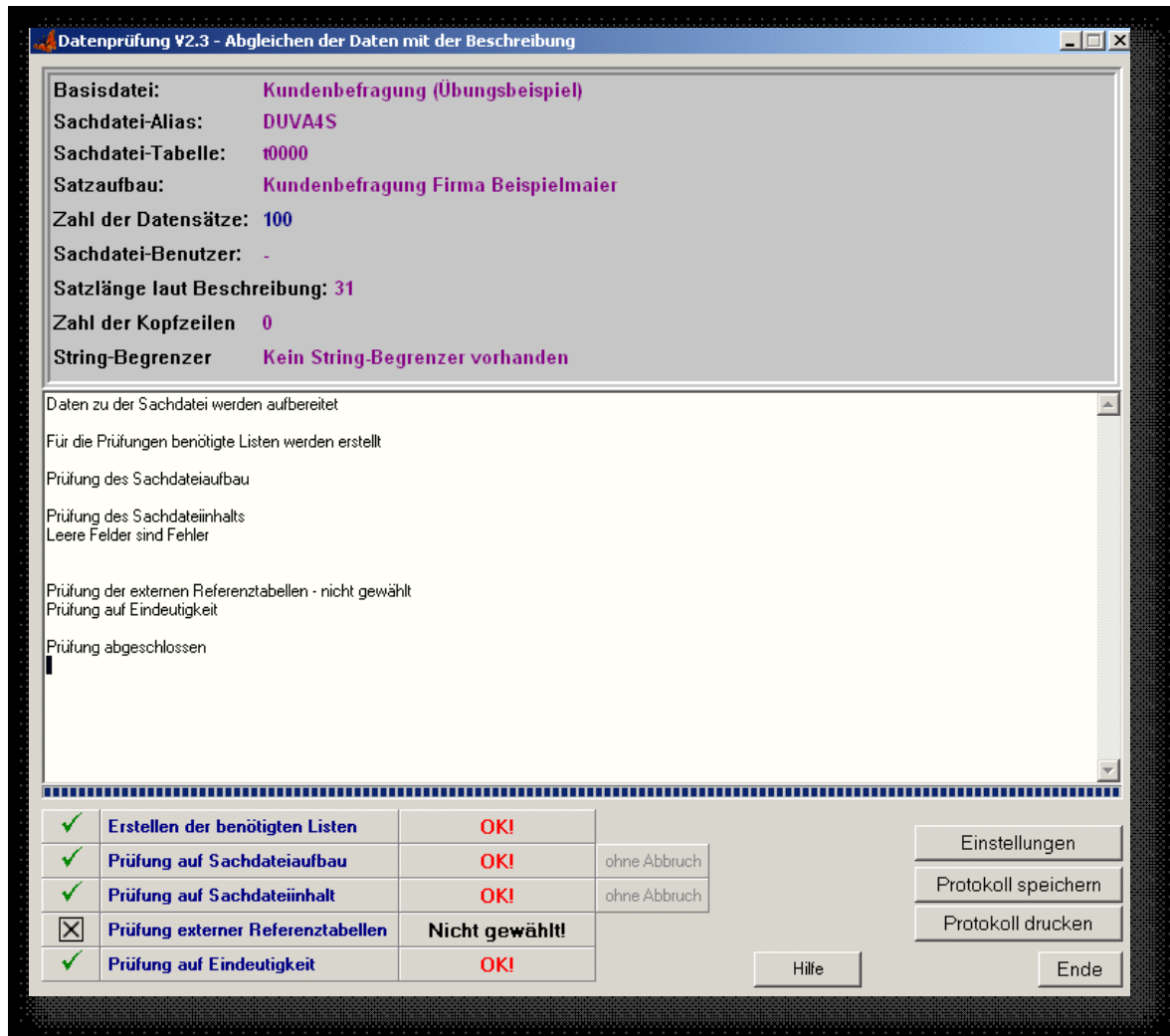
Über die Karte Sachdatei erfolgt die Verknüpfung einer Beschreibung zu den Sachdaten, die entweder als Textdateien (festes oder freies Format) oder als Datenbanktabellen vorliegen. In der aktuellen Version wird zu Verarbeitungszwecken jede Textdatei in eine Datenbanktabelle konvertiert (Daten Laden). Je nach Wahl einer Datenbank bei der Installation von DUVA ergibt sich das Format der Tabellen.

Die Karte Zusatzinformation ermöglicht die Eingabe von Metatexten oder Verweisen. Die Karte Hintergrundgeometrie bietet die Auswahl aus der Liste aller Geometrien an. Das Objekt Geometrie verweist auf eine Datei, die Geometriedaten als Kartengrundlagen enthält (z.B. Shape-Dateien).

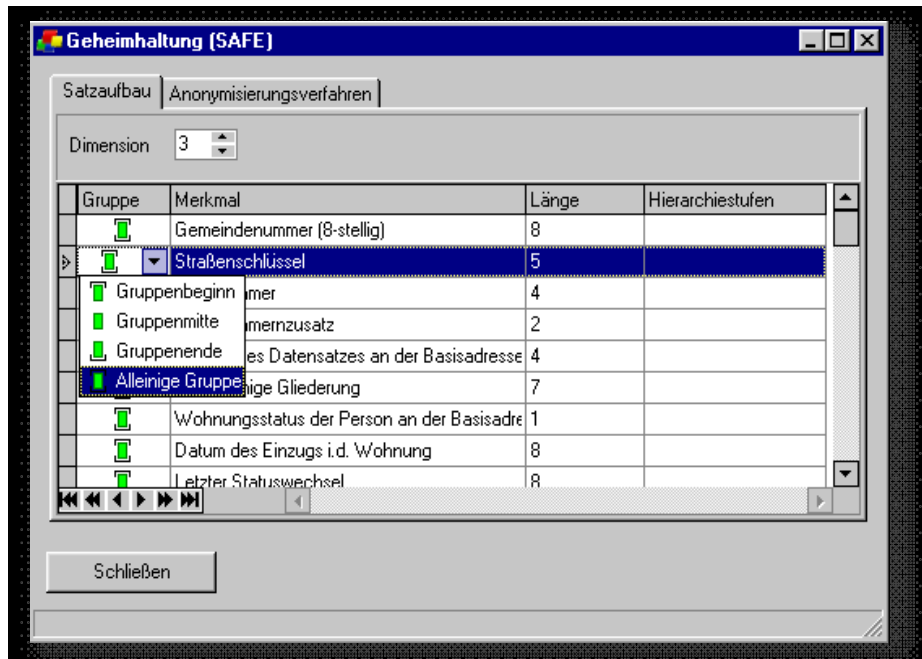


Über ein separates Objekt Raumbezugsebene findet die Verbindung einer Geometrie zu der Merkmalsebene (bzw. der dahinter stehenden Verschlüsselung) statt.

Fertigt man Dateibeschreibungen als Konzepte oder Anforderungskatalog für einen Datenabzug an, so ist es nach Erhalt der Sachdaten empfehlenswert, eine Überprüfung auf Plausibilität durchzuführen. Das integrierte Prüfprogramm liest die entsprechenden Metadaten aus der Dateibeschreibung ein und führt eine formale und inhaltliche Kontrolle der Sachdaten durch (s. Prüfprotokoll, S. 24).

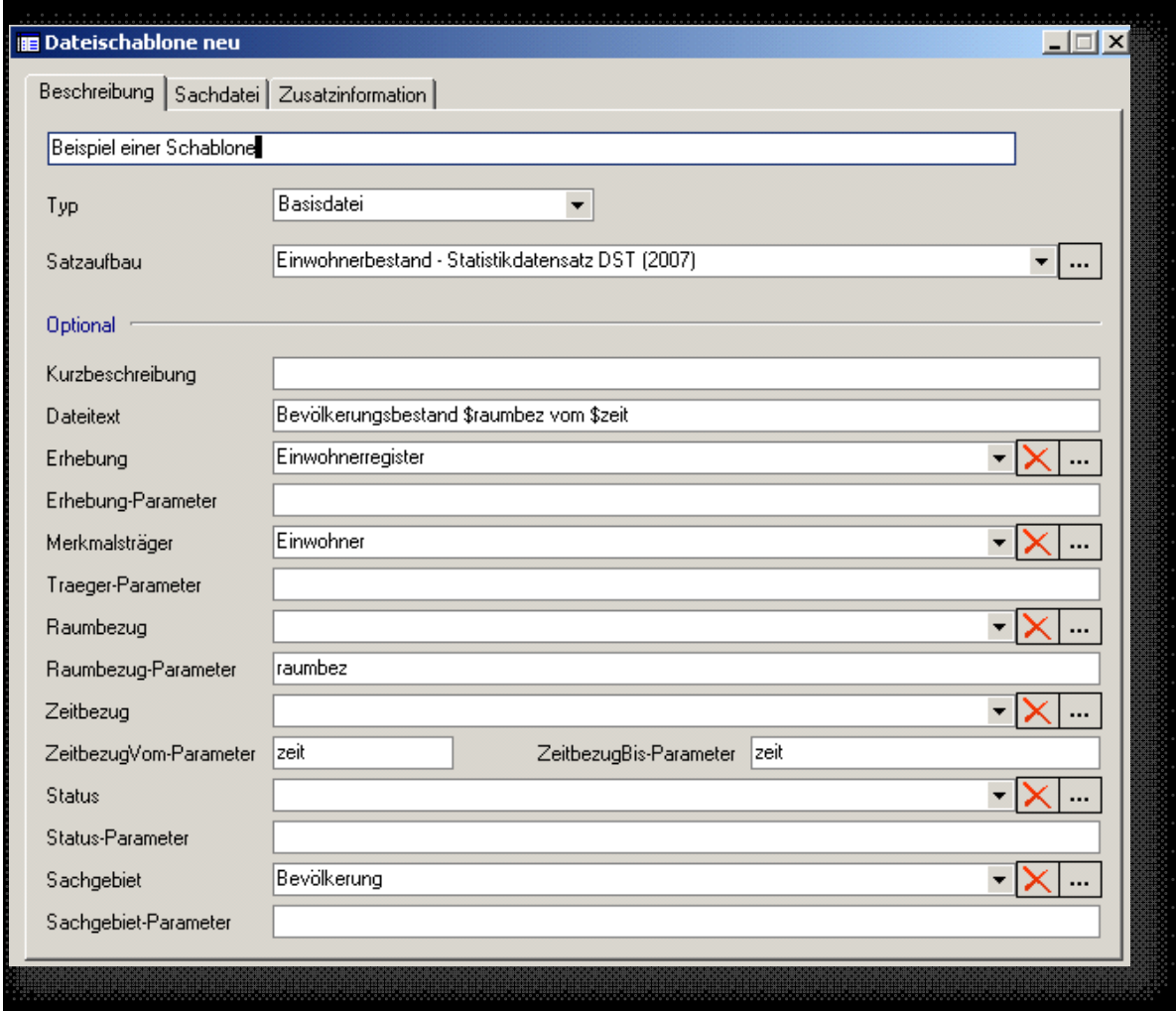


Um den Erfordernissen der statistischen Geheimhaltung und des Datenschutzes zu genügen, lässt sich auf Grundlage einer Dateibeschreibung und einer vorliegenden Sachdatei das Geheimhaltungsprogramm SAFE zur Anonymisierung von Einzeldaten einsetzen. In der Konsequenz führt das Programm zu einer Umbuchung der Sachdaten, so dass im Ergebnis Ausprägungskombinationen mit Häufigkeiten von 1 oder 2 entweder die Häufigkeit 0 oder 3 erhalten.



1.3.5 Dateischablonen

Aus den praktischen Erfahrungen in der Arbeit mit DUVA und dem Wunsch von Prozessautomatisierungen entstand das Objekt Dateischablone. Analog zu Dateibeschreibungen lassen sich Basis- und Makrodateischablonen beschreiben. Die Beschreibung einer Datei enthält eindeutige, festgelegte Zuordnungen von Objekten wie Dateiname, Erhebung, Merkmalsträger, Raumbezug, Zeitbezug, Status, Sachgebiet, Sachdateityp und Sachdateiname. Bei der Beschreibung von Dateischablonen können diese Zuordnungen durch Vergabe eines oder mehrerer Parameter variabel gehalten werden (im Beispiel unten raumbez für Raumbezug und zeit für den Zeitbezug. Die Parameter lassen sich auch als Teile von Beschreibungstexten verwenden (dann aber mit vorangestelltem \$-Zeichen). Zulässig ist die Vergabe von maximal 64 Parametern. Der Einsatz von Dateischablonen ist immer dann sinnvoll, wenn periodische Verarbeitungen von Dateien anfallen, die sich z.B. nur im Dateinamen, Zeitbezug und dem Namen der Sachdatei unterscheiden (wie etwa Einwohnerbestände zum 31.12. eines Jahres).



Auf Ebene der Produktion (s. Kapitel 1.3.7, in einem Automatisierungsprojekt) erhalten dann die Parameter konkrete Zuweisungen oder Werte und es besteht die Möglichkeit aus einer Schablone eine oder mehrere Dateibesreibungen in der Metadatenbank abzulegen. Dies ist allerdings nicht zwingend. Wenn aber eine Dateibesreibung übernommen werden soll, dann müssen neben Dateityp und Satzaufbau die Einträge von Dateitext, Erhebung, Merkmalsträger, Raum- und Zeitbezug sowie Sachdateiverweis entweder als Beschreibung oder als Parameter vorliegen.

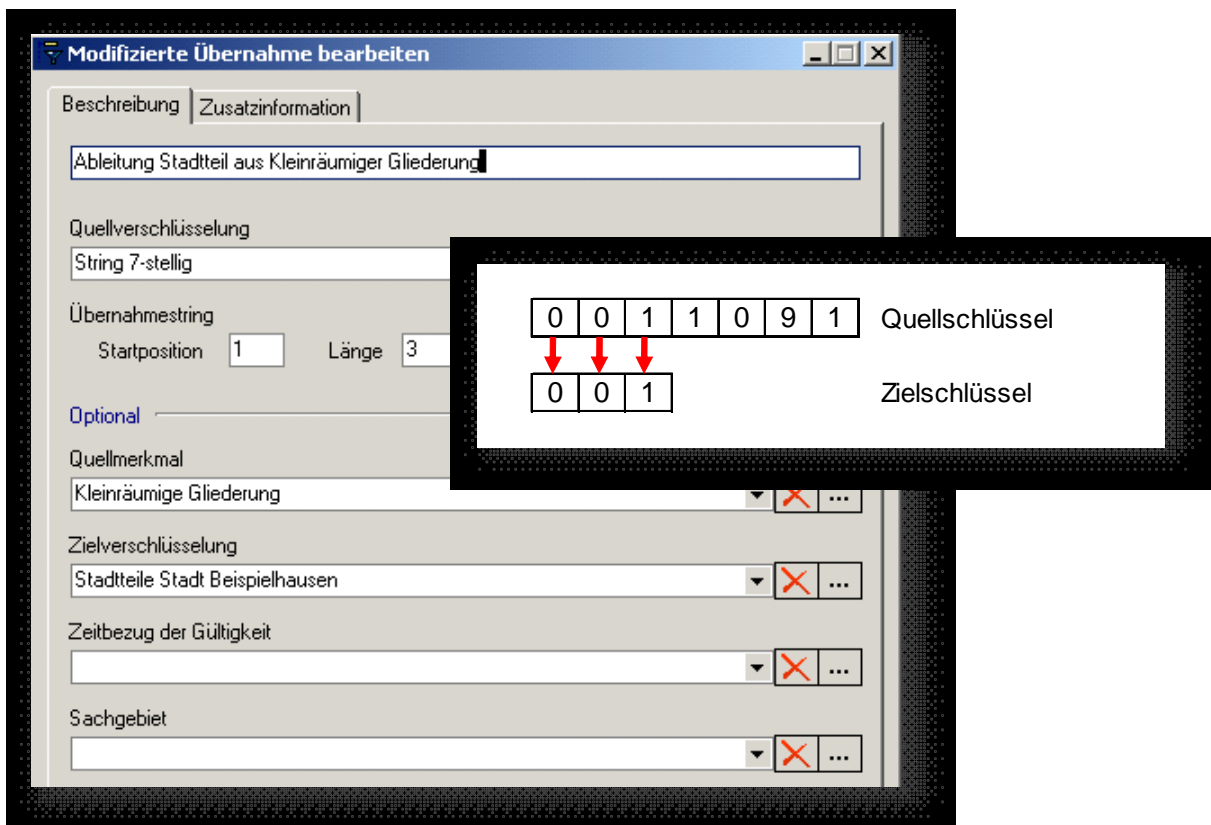
1.3.6 Regeln

Regeln oder auch Ableitungsregeln bilden eine Objektgruppe, die bei Produktionen zum Einsatz kommt. Obwohl von Merkmalsableitungen gesprochen wird, handelt es sich um die Beziehung von zwei oder mehreren Verschlüsselungen bzw. deren Ausprägungen. Die Standardregel bei einer Produktion wäre eine 1:1-Übernahme eines Merkmals und seiner Verschlüsselung, d.h. ein Merkmal wird bei einer Datenaufbereitung oder in einem Analyseschritt nicht verändert. Als Beispiel für einfache Ableitungsregeln wären Gruppierungs- oder Klassifizierungsverfahren anzuführen.

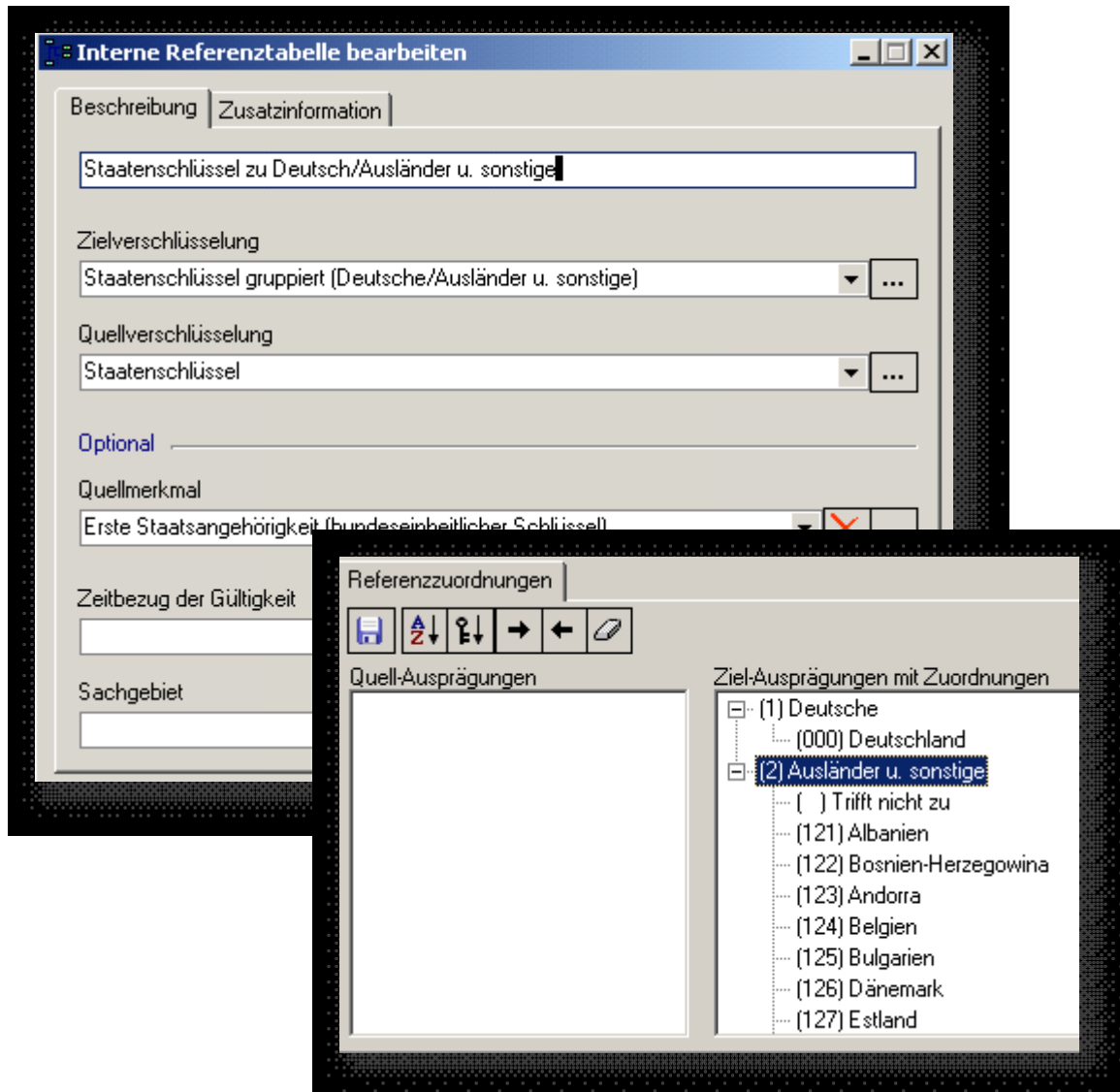
Komplexere Regeln wie etwa Bedingungslisten und Algorithmen (z.B. zur Berechnung von statistischen Maß- und Kennzahlen) können auch merkmalsübergreifend sein (Mittelwerte, Streuungsmaße, Messziffern, Zusammenhangsmaße etc.). Für Merkmalsableitungen stehen folgende Regeln zur Auswahl:



Modifizierte Übernahme bedeutet, dass aus einem Schlüssel ein Teilschlüssel übernommen oder ausgeschnitten wird (Substring), der für sich einen Sinn ergibt. So zum Beispiel die Ableitung des Stadtteils als räumliche Einheit aus der Kleinräumigen Gliederung.



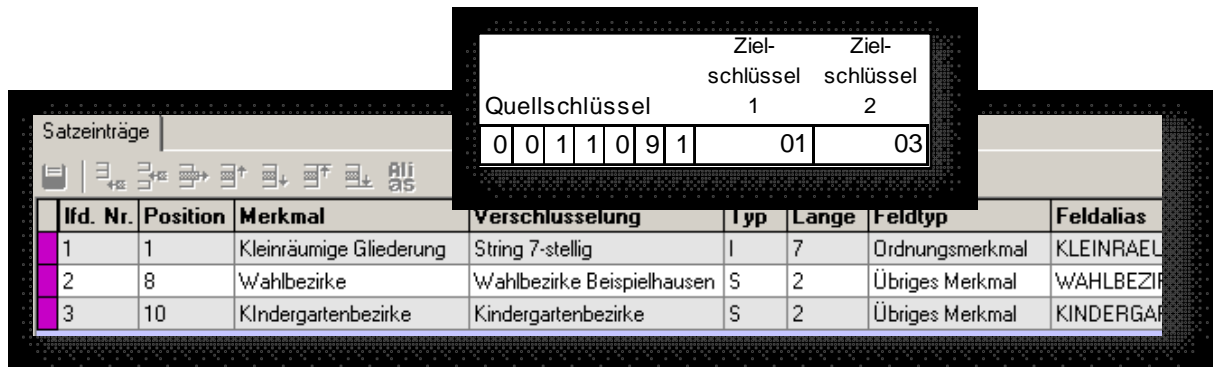
Interne Referenzierung steht für eine Beziehung zwischen zwei Schlüssel Tabellen und hat eine Gruppierung oder unter Umständen eine Klassifizierung zum Inhalt (Beispiel: Gruppierung aller Staatenschlüssel zur Zielverschlüsselung Deutsche/ Ausländer und sonstige).



Externe Referenztablelle bedeutet eine Zuweisung oder Gruppierung von Quell- zu Zielausprägungen über eine vorhandene Tabelle (z.B. Textdatei mit festem Format oder Datenbanktablelle). Eine Externe Tablelle hat vom Prinzip her den selben Aufbau wie eine Basisdatei, allerdings mit nur einem Ordnungsmerkmal (als erstes Merkmal im Satzaufbau). Das jeweilige Ordnungsmerkmal repräsentiert die Quellverschlüsselung bzw. das Quellmerkmal, die übrigen maximal möglichen 254 übrigen Merkmale die Zielverschlüsselungen bzw. Zielmerkmale. Als Quellschlüssel kommt jede der drei möglichen Verschlüsselungsarten (Schlüsseltabelle, Identifizierender Schlüssel oder Wertebereich) in Frage. Bezüglich der Ausprägungen oder der einzelnen Werte gilt die gleiche Erfordernis wie bei Ordnungsbegriffen: im Quellschlüssel darf eine Ausprägung oder ein Schlüsselwert aus Gründen der Eindeutigkeit nur ein einziges Mal vorkommen.

Als Zielschlüssel können ebenfalls alle drei Verschlüsselungsarten eingesetzt werden. Im Satzaufbau erhalten sie jeweils die Typzuweisung 'Übriges Merkmal'. Die Externe Tablelle kann in ihrem Satzaufbau mehrere Zielverschlüsselungen bzw. -Merkmale aufweisen. In diesen Fällen erfolgt die Festlegung des Zielschlüssels, auf den referenziert werden soll, auf Ebene der Produktion. Externe Tablellen werden in der Praxis häufig für Raumreferenzierungen eingesetzt. Steht beispielsweise eine Externe Tablelle zur Verfügung, die ausgehend von einem geeigneten Quellschlüssel (z.B. 'Kleinräumige Gliederung') Bezüge zu Raumkategorien (z.B. Wahlbezirke, Schulbezirke, Kindergartenbezirke etc.) enthält, so lässt sich z.B. eine Einwohnerdatei nach allen diesen

Raumkategorien auswerten. Voraussetzung ist, dass das Referenzmerkmal (hier 'Kleinräumige Gliederung') auch im Satzaufbau der Quelldatei (z.B. Einwohnerdatei) vorkommt.



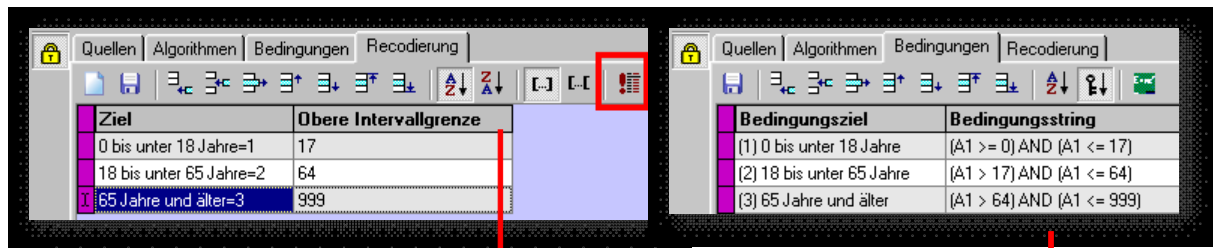
The screenshot shows a software interface with a table of mappings and a source key value. The mapping table is as follows:

Quellschlüssel	Ziel-schlüssel 1	Ziel-schlüssel 2
0 0 1 1 0 9 1	01	03

Below this, a table lists field mappings:

lfd. Nr.	Position	Merkmal	Verschlüsselung	Typ	Länge	Feldtyp	Feldalias
1	1	Kleinräumige Gliederung	String 7-stellig	I	7	Ordnungsmerkmal	KLEINRAEL
2	8	Wahlbezirke	Wahlbezirke Beispielhausen	S	2	Übriges Merkmal	WAHLBEZIF
3	10	Kindergartenbezirke	Kindergartenbezirke	S	2	Übriges Merkmal	KINDERGAR

Bedingungsliste – Recodierung: die Recodierung lässt sich zur Gruppierung oder Klassifizierung von Wertebereichen oder Identifizierenden Schlüsseln einsetzen. Als Zielschlüssel kommt hier ausschließlich der Typ Schlüsseltabelle in Frage. Die Zuweisung von Quellausprägungen oder -Werten zu den Ausprägungen der Zielschlüsseltabelle muss vollständig sein. D.h., die gesamte Spannweite eines Wertebereichs muss lückenlos referenziert werden. Die Zuordnung erfolgt über Operatoren ('größer', 'kleiner', 'gleich', 'kleinergleich', 'größergleich', 'ungleich', 'und', 'oder' etc.). Im Beispiel ist die Gruppierung von Altersangaben aus einem Quellschlüssel 'Alter in vollen Jahren' (Wertebereich, 3-stellig ohne führende 0) zu den Ausprägungen 'unter 18 Jahre', '18 bis unter 65 Jahre' und '65 Jahre und älter' einer Ziel-Schlüsseltabelle 'Altersgruppen' dargestellt. Durch Angabe der entsprechenden Intervallgrenzen erzeugt die Funktion Recodierung einen Bedingungsstring.



The screenshot shows two side-by-side windows of the 'Recodierung' tool. The left window shows the mapping of source key ranges to target keys:

Ziel	Obere Intervallgrenze
0 bis unter 18 Jahre=1	17
18 bis unter 65 Jahre=2	64
65 Jahre und älter=3	999

The right window shows the resulting condition strings for each target key:

Bedingungsziel	Bedingungsstring
(1) 0 bis unter 18 Jahre	(A1 >= 0) AND (A1 <= 17)
(2) 18 bis unter 65 Jahre	(A1 > 17) AND (A1 <= 64)
(3) 65 Jahre und älter	(A1 > 64) AND (A1 <= 999)

Bedingungsliste ohne Recodierung: im Unterschied zur Recodierung sind bei Bedingungslisten, sämtliche Verschlüsselungsarten als Quell- bzw. Zielverschlüsselungen einsetzbar. Es lassen sich auch mehrere Quellverschlüsselungen einer Zielverschlüsselung zuordnen (merkmalsübergreifende Regel). Außerdem braucht die Zuordnung der einzelnen Quellausprägungen zu den Zielschlüsseln nicht vollständig zu sein. Dafür müssen jedoch die einzelnen Bedingungen über die Tastatur bzw. Mausklicks eingetragen werden.

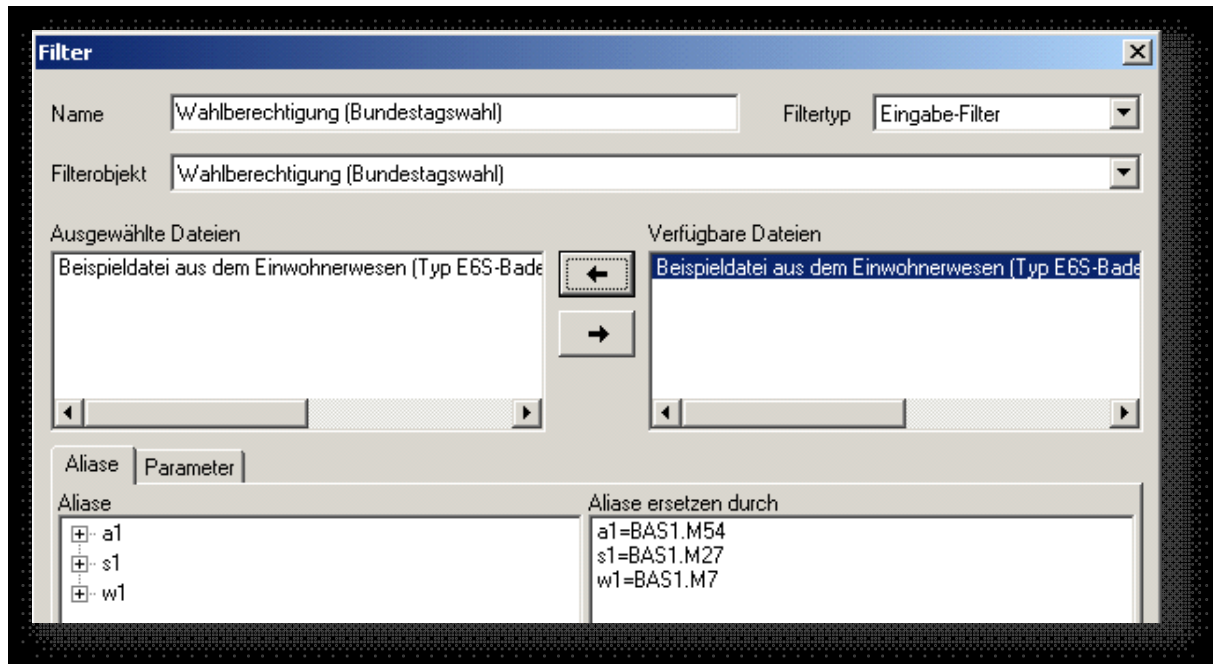
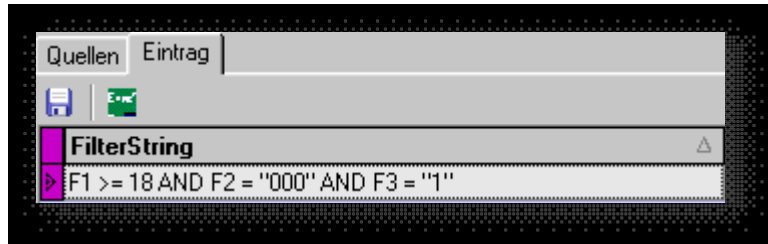
Bedingungsziel	Bedingungsstring
(1) Jugendliche ausländische Mädchen (14 bis 18 Jahre)	a1 >= 14 AND a1 < 18 AND g1 = "2" AND s1 <> "000"
(2) Junge deutsche Erwachsene (18 bis 21 Jahre)	a1 >= 18 AND a1 < 21 AND g1 = "1" AND s1 = "000"
(3) Restliche Bevölkerung	ELSE

Algorithmen: mit Hilfe der Regel Algorithmen lassen sich ganz allgemein Ableitungen zwischen Quell- und Zielverschlüsselungen (auch merkmalsübergreifend) frei formulieren. Zur Verfügung stehen arithmetische Operatoren (+, -, *, /), relationale Operatoren (<, >, =, >=, <=, <>) und die bool'schen Operatoren (AND, OR, NOT). Ein einfaches Beispiel für einen Algorithmus 'Berechnung von DM-Beträgen in EURO-Beträge': das Quellmerkmal ist ein Wertebereich 'Betrag in DM', das Zielmerkmal ist ebenfalls ein Wertebereich und lautet 'Betrag in EURO'.

Der Algorithmus zur Merkmalsableitung lautet: 'Betrag in EURO' = ('Betrag in DM' / 1,9558).

AlgorithmString
K1 / 1.9558

Filter führen zu einer Selektion der Datensätze nach den formulierten Bedingungen. Die Bedingungen lassen sich durch einfache Ausdrücke oder auch als komplexe Ausdrücke formulieren. Filter können sich auf eine oder mehrere Dateien beziehen. Bei der Verwendung von Filtern auf Ebene der Produktion besteht die Auswahlmöglichkeit zwischen Eingabe- und Ausgabefiltern. Eingabefilter führen bei einer Produktion zu einer Reduktion der Datensätze einer Quelldatei (die Filterbedingungen werden bereits beim Einlesevorgang einer Quelldatei berücksichtigt), was eine schnellere Verarbeitung nach sich zieht. Ausgabefilter beziehen sich nur auf Zieldateien. D.h. bei der Verarbeitung werden zunächst sämtliche Datensätze einer Quelldatei eingelesen, anschließend aber nur die Datensätze ausgegeben, die den Filterkriterien entsprechen. Die Festlegung, ob ein Filter Ein- oder Ausgabefilter sein soll, erfolgt auf Ebene der Produktion (Filterzuordnung).

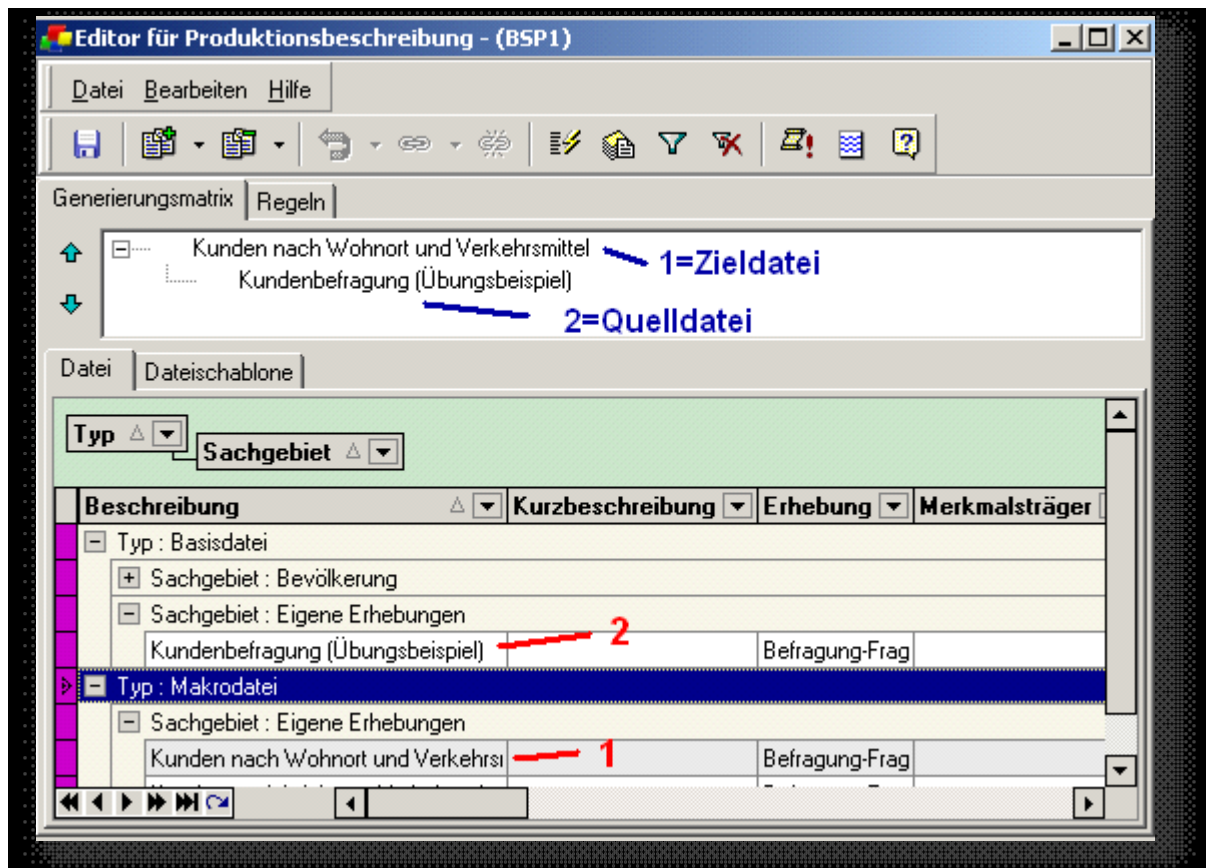


1.3.7 Produktion

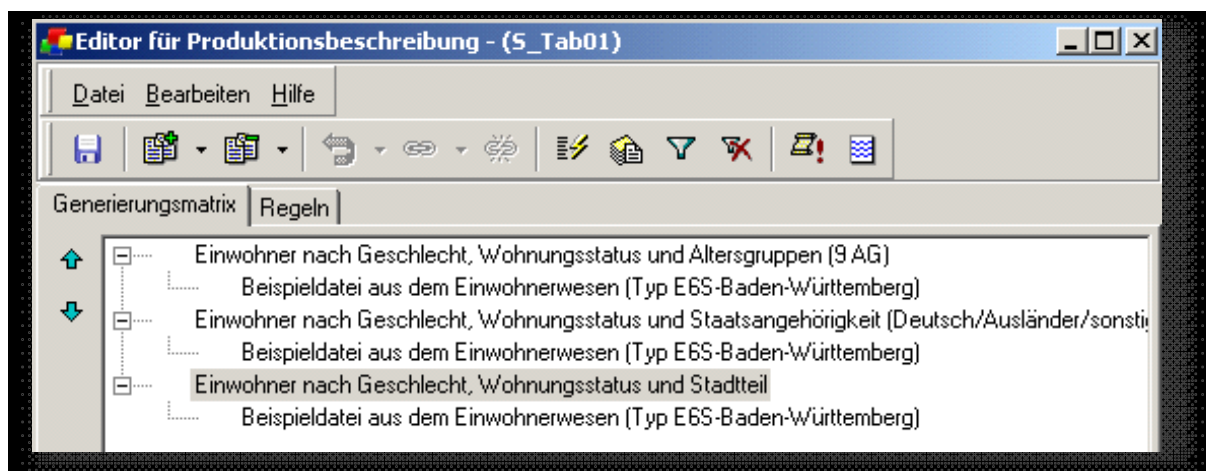
Unter Produktion ist ganz grundsätzlich die metadatenbasierte Verarbeitung von Quelldateien (Basis- oder Makrodateien) zu Zieldateien (Basis- oder Makrodateien) zu verstehen. Bei der Produktion wird von einem temporär aktiven Modul (Zieldateigenerator) eine oder mehrere Sachdateien im entsprechend gewählten Format (Textdatei oder Datenbanktabelle) erzeugt. Die Formatspezifikation erfolgt auf Ebene der Zieldateibeschreibung. Grundsätzlich sind folgende Kombinationen von Dateibeschreibungen oder Dateischablonen möglich:

Quelldatei	Zieldatei
Basisdatei (en)	Basisdatei (en)
Basisdatei (en)	Makrodatei (en)
Makrodatei (en)	Makrodatei (en)

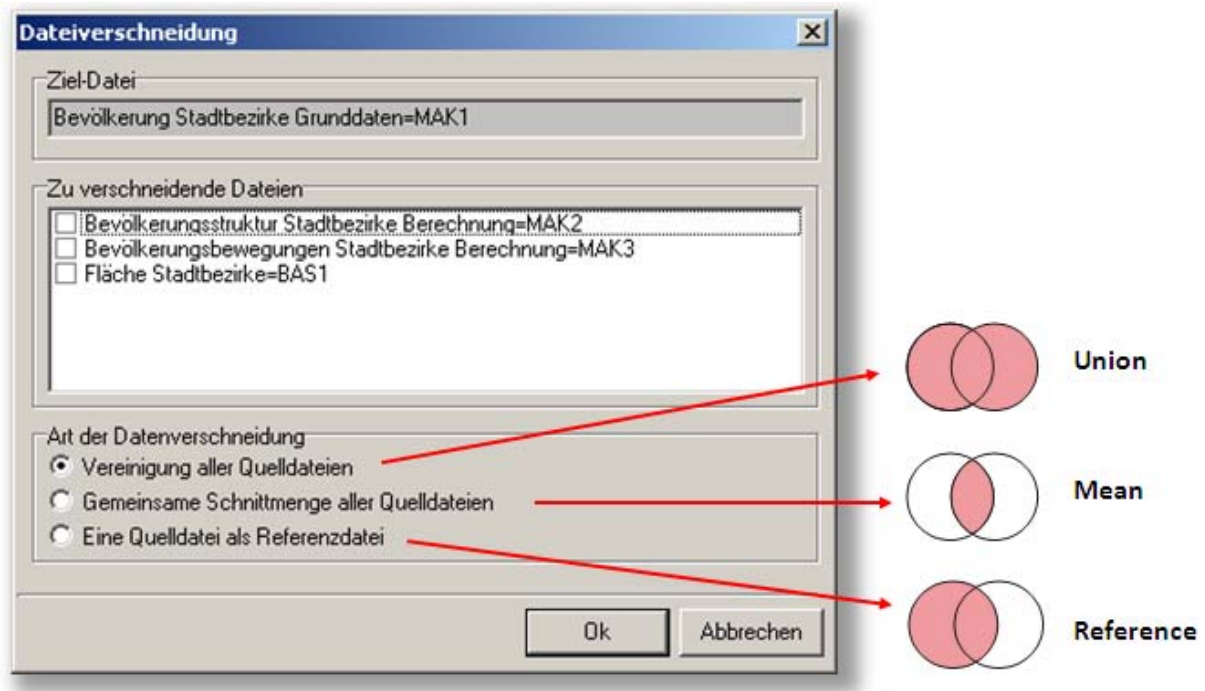
Produktionen stellen die Kernarbeit der Datenaufbereitung dar und lassen sich nur durchführen, wenn die Beschreibung einer Zieldatei (oder mehrerer Zieldateien), die Beschreibung einer Quelldatei (oder mehrerer Quelldateien) sowie die entsprechende(n) Quell-Sachdatei(en) vorliegen.



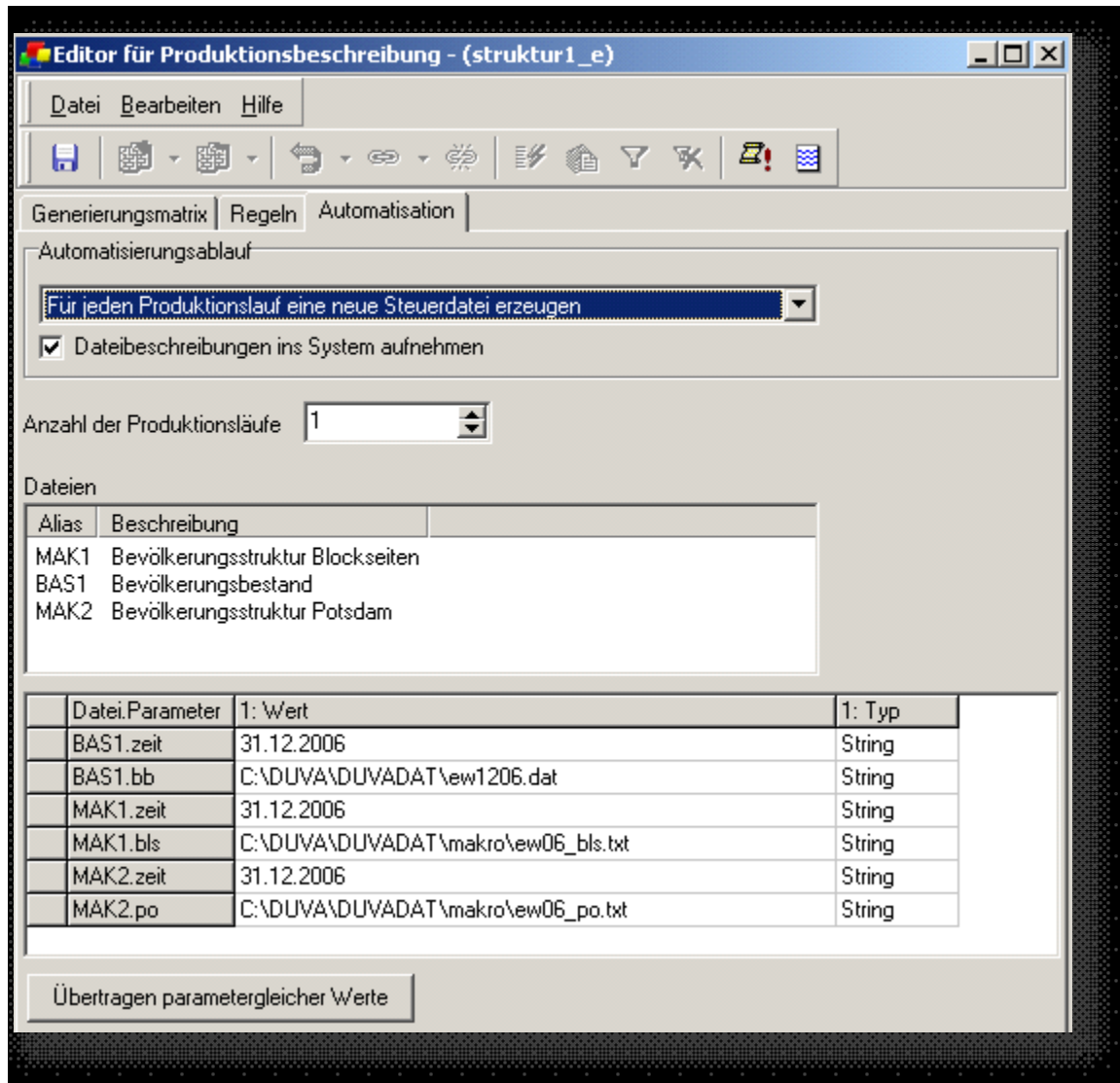
Einfache Produktionen setzen sich aus einer Zieldatei und einer Quelldatei zusammen. Dabei können die Merkmale aus der Quelldatei 1:1 übernommen werden (direkte Merkmalsübernahme) oder durch den Einsatz von Regeln abgeleitet werden. Es besteht aber auch die Möglichkeit, in einem Produktionslauf aus mehreren unterschiedlichen Quelldateien mehrere unterschiedliche Zieldateien zu erzeugen oder in einem Produktionsablauf zuvor beschriebene Produktionen hintereinander laufen zu lassen. Jede Produktionsbeschreibung stellt ein Metadatenobjekt dar und kann somit beliebig oft eingesetzt werden und zwar inklusive aller definierten Regeln.



Komplexe Produktionen können aber nicht nur aus mehreren Quell- und Zieldateien bestehen, sondern auch aus Dateiverschneidungen der Quelldateien. Das bedeutet, dass unterschiedliche Dateien mit unterschiedlichen Merkmalsträgern zur Bildung von Kenngrößen oder Indikatoren herangezogen werden können.



Produktionen unter Verwendung von Schablonen, auch Automatisierungen genannt, eignen sich insbesondere für wiederkehrende Auswertungen und Analysen. Kennzeichnend für wiederkehrende Auswertungen ist, dass sich die Sachdaten und wenige Beschreibungsobjekte verändern (z.B. Zeitbezug, Name der Sachdatei, Dateibeschreibung), der Rest bleibt jedoch unverändert (z.B. Satzaufbau, Datenquelle, Merkmalsträger, Raumbezug). Die grundsätzliche Vorgehensweise für die Beschreibung einer Produktionsdatei mit Schablonen ist analog zur normalen Dateibeschreibungen. So lassen sich einfache oder komplexe Produktionen durchführen, sämtliche Regel-, Filter- und Levelinstellungen usw. zuordnen. Desweiteren ist auch eine Mischung von Schablonen und vollständigen Dateibeschreibungen in einem Automatisierungsprojekt möglich. Der einzige Unterschied zur Produktion mit vollständigen Dateibeschreibungen besteht darin, dass statt des Zieldateigenerator-Aufrufs die Automatisierung gestartet wird. An dieser Stelle sind die variabel gehaltenen Beschreibungsobjekte (Parameter) zu ersetzen und es ist festzulegen, ob aus den Schablonen vollständige Beschreibungen in die Metadatenbank aufgenommen werden sollen oder nicht.

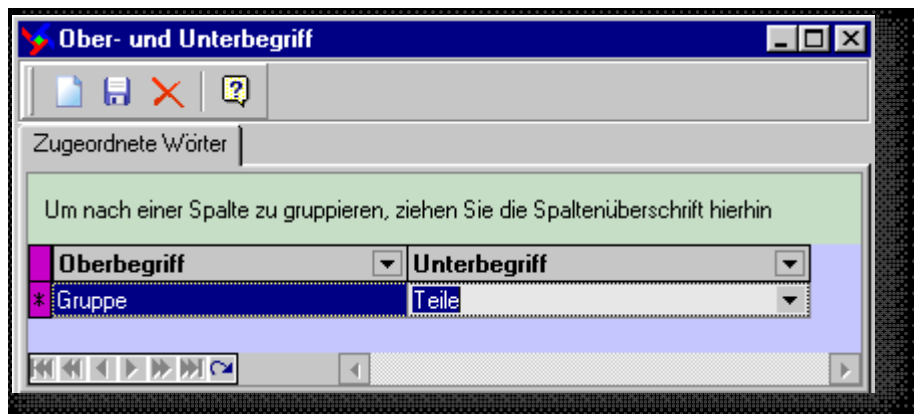
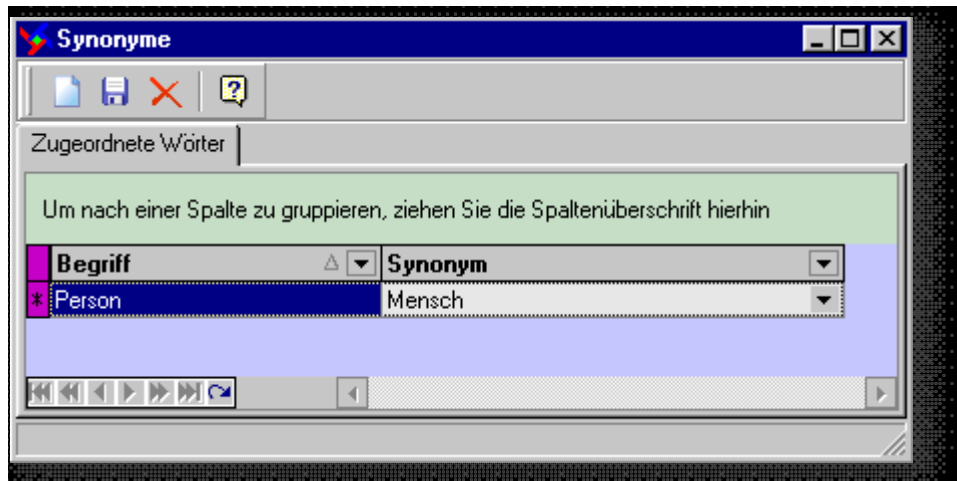


Liegt eine Produktionsbeschreibung als Automatisierungsprojekt vor, so reduziert sich der Beschreibungsaufwand auf die wenigen Arbeitsschritte: Aufruf der Produktionsdatei, Klick auf Automatisierung, Parameter eintragen und Start der Automatisierung. Dabei ist es völlig unerheblich, ob es sich um eine sehr komplexe Produktion mit vielen Quell- und Zieldateien, Filter- und Regelzuweisungen usw. oder eine ganz einfache Produktion handelt.

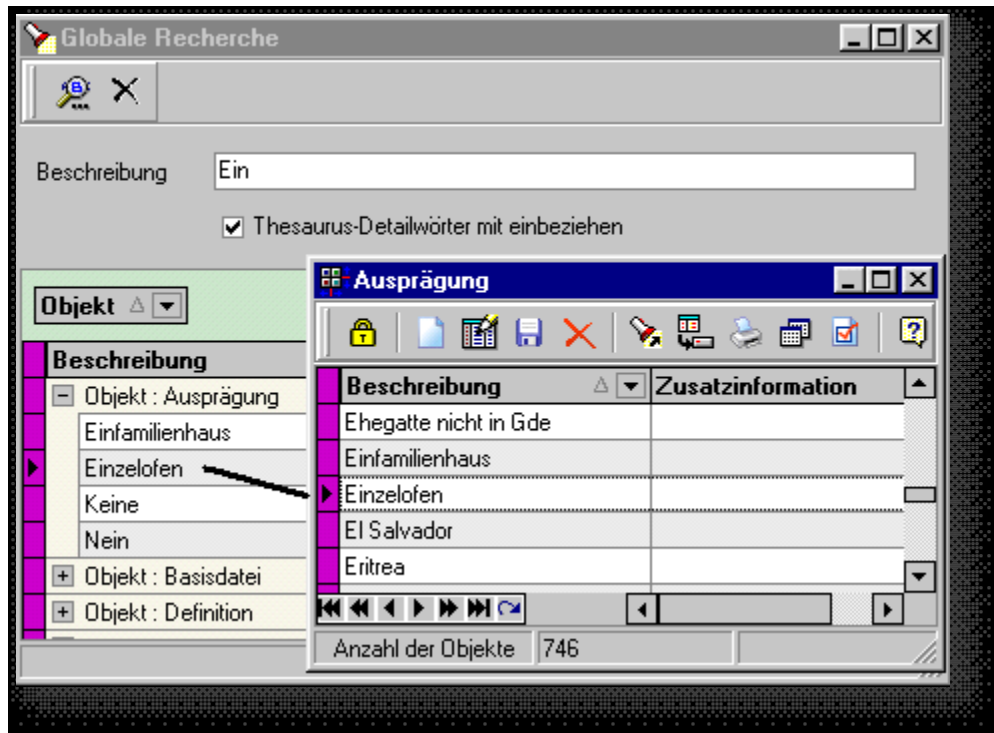
1.3.8 Thesaurus, Suchfunktionen

Der Thesaurus organisiert die semantische Verknüpfung von Beschreibungstexten unabhängig von den Beziehungen zwischen den Metadatenobjekten. Somit kommt ein Begriff nur einmal im System vor, weist aber theoretisch n Verknüpfungen zu anderen Begriffen auf. Die Relationen zwischen den Metadatenobjekten sind durch die Struktur des Datenmodells festgelegt. Das bedeutet, es handelt sich um zwei Ebenen, die für die stringente Organisation der Metadaten und die Eindeutigkeit der Bezeichnungen unerlässlich sind.

Mit Hilfe des Thesaurus ist es möglich, Wortbeziehungen wie Synonyme oder auch Über-/Unterordnungen zu beschreiben. Damit ergeben sich bei Such- und Recherchevorgängen komfortable Möglichkeiten, da sich der Informationszusammenhang sowohl über die originären Begriffe als auch über die semantisch zugeordneten Begriffe erschließen lässt. Bei der Schlagwortsuche, die auf Zeichenketten basiert, sind sowohl phonetische Suchfunktionen als auch unscharfe Suchfunktionen (Fuzzy-Logik) einstellbar.

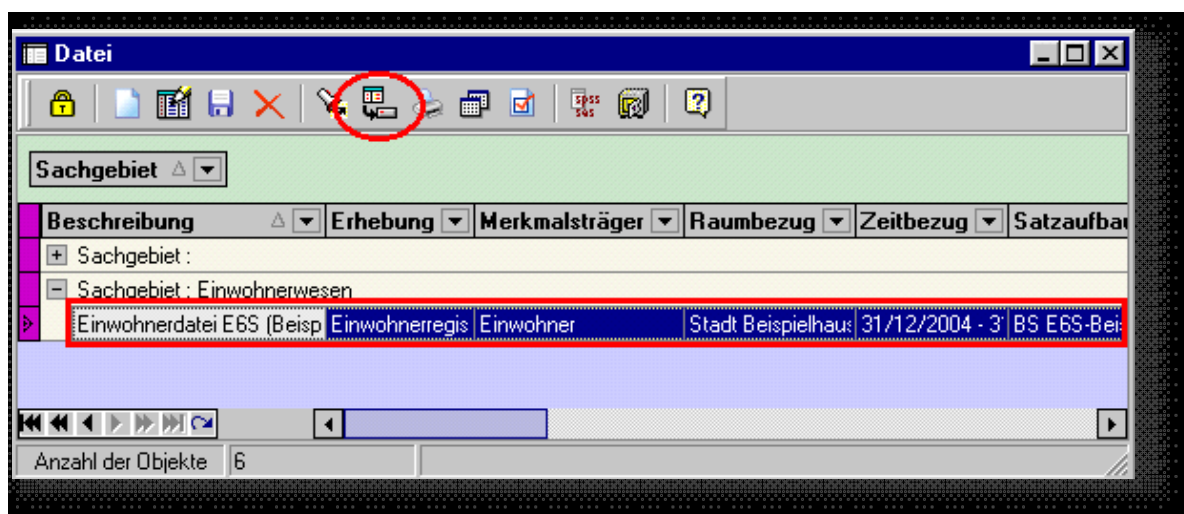


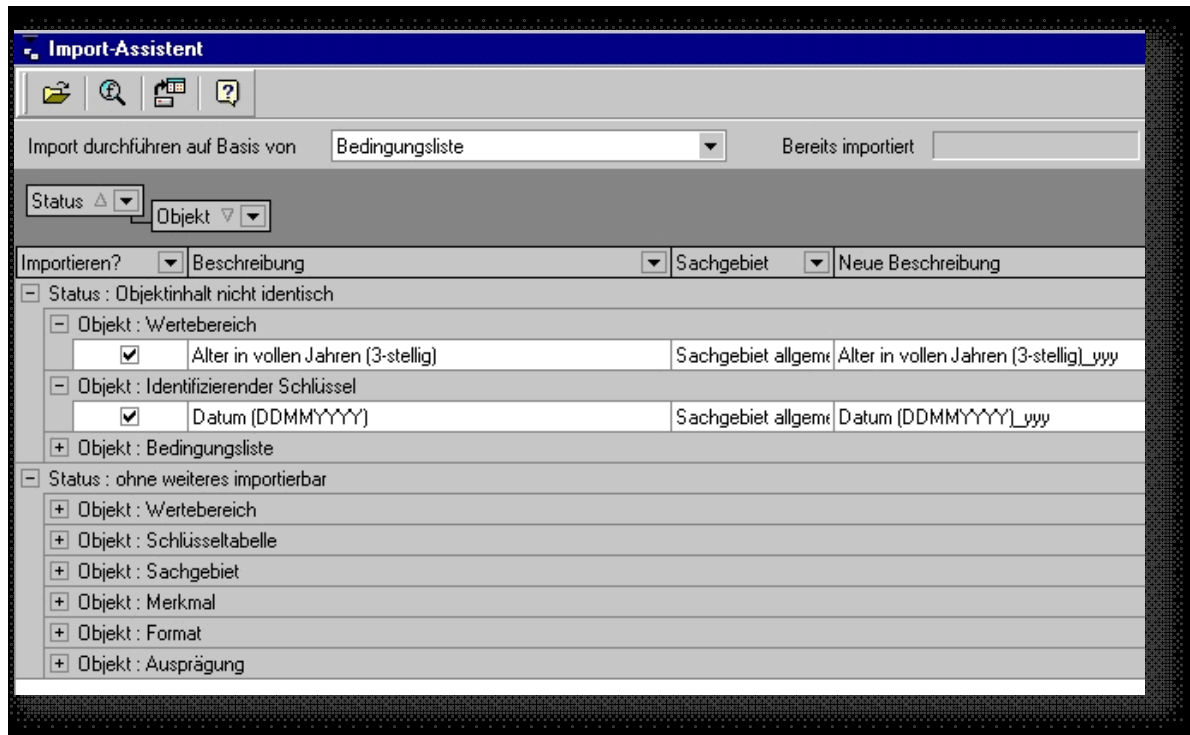
Die Informationssuche beginnt auf der begrifflichen Ebene durch Eingabe eines Wortes oder auch nur einer Silbe des eigentlichen Suchbegriffs oder eines semantisch abhängigen Begriffs (Synonym, Ober-/Unterbegriff). Über die Anzeige der entsprechenden Metadatenobjekte lässt sich eine Ebenenanalyse durchführen, die sämtliche Verknüpfungen aufzeigt (beispielsweise: welcher Schlüsseltable ist eine Ausprägung zugeordnet, welchem Satzaufbau ist diese Schlüsseltable zugeordnet, in welcher Produktion kommt dieser Satzaufbau vor etc.).



1.3.9 Import/Export

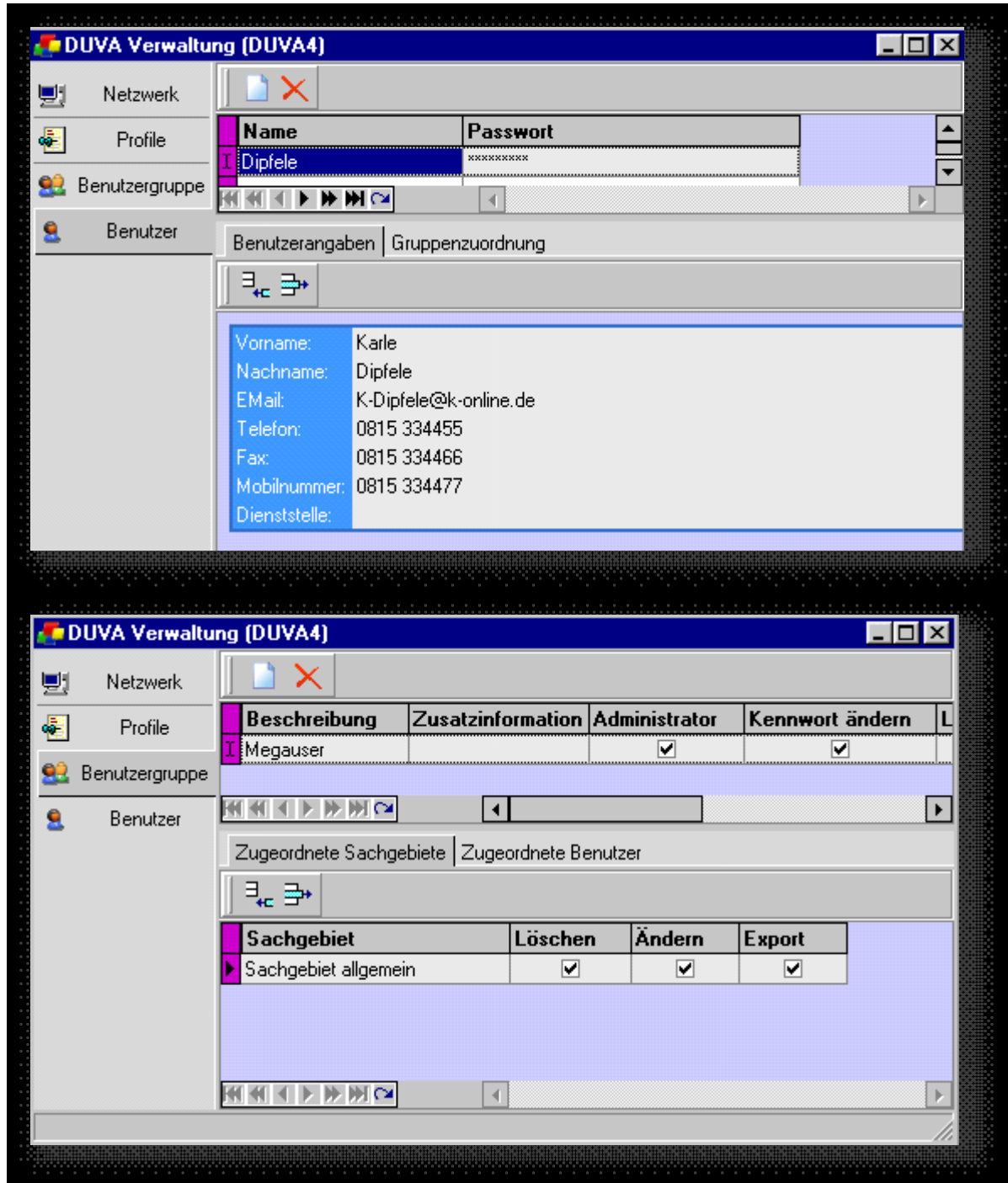
Die DUVA-Metadatenbank ist inhaltlich zunächst auf eine organisatorische Einheit beschränkt, die im Client-Server-Betrieb von unterschiedlichen Fachbereichen genutzt wird. Über die Möglichkeit des Im- und Exports von Objekten oder auch Objektgruppen (Produktionen inklusive Dateibeschreibungen und Regeln) können Beschreibungen unter den DUVA-Anwendern ausgetauscht werden, bzw. schaffen einen Pool an Beschreibungen. In der Konsequenz führt dies zu erheblichen Reduktionen des Zeitbedarfs für Erhebungs- und Erschließungskonzepte und durch die Quasi-Standardisierung zu einer verbesserten Vergleichbarkeit von Ergebnissen (z.B. bei interkommunalen Vergleichen und Raubeobachtungen).





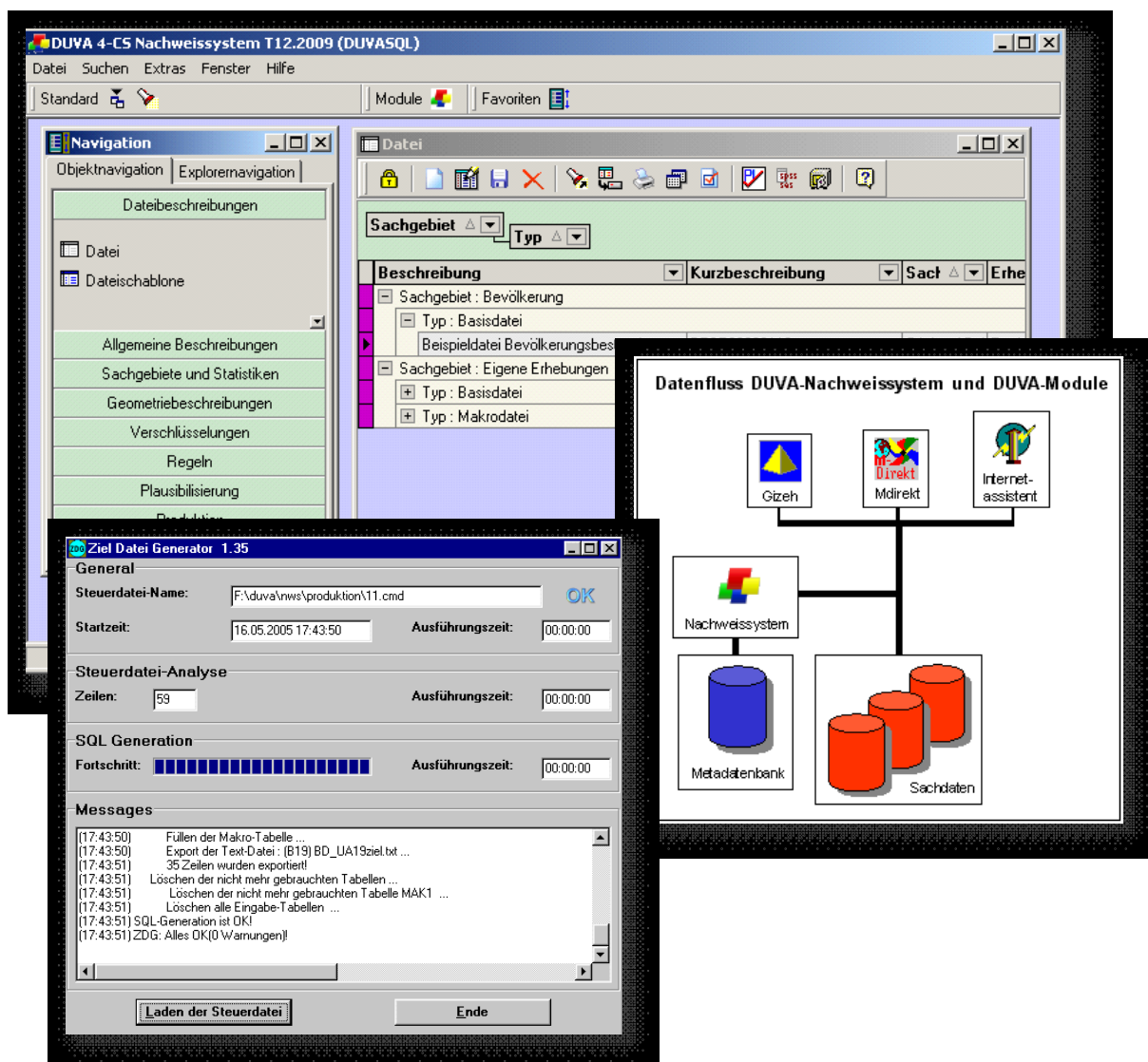
1.3.10 Benutzerverwaltung

Das DUVA-Metadatenmodell umfasst nicht nur die fachlichen Metadaten, die sich aus dem statistischen Produktionsprozess ergeben, sondern auch die Metadaten im Zusammenhang mit der System- und Benutzeradministrierung. Dies ist eine Voraussetzung zur für die Beschreibung von selektiven Zugriffs- und Benutzerprofilen.



2. DUVA-Module

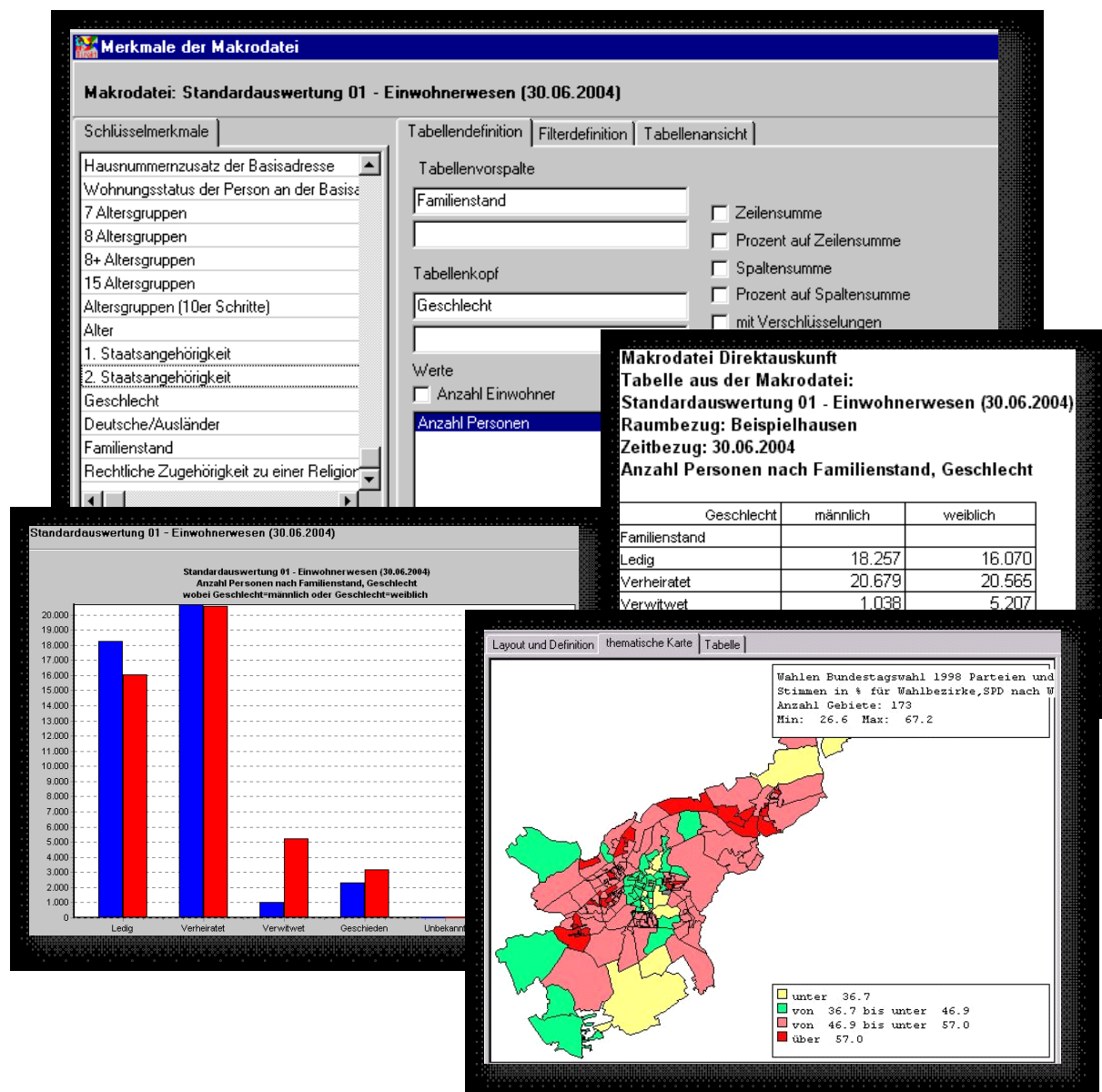
Bei der Darstellung des DUVA-Metadatenansatzes wurde bereits auf einzelne DUVA-Module hingewiesen, allerdings nicht im Systemzusammenhang. Den Kern der gesamten Metadatenverarbeitung stellt das Modul Nachweissystem dar. Im Nachweissystem werden Metadaten erfasst, verwaltet und zu Produktionen eingesetzt. Dazu greift das Nachweissystem auf die Metadatenbank zu und veranlasst bei Produktionen über das Modul Zieldateigenerator den Zugriff zur Sachdatenbank. Die bereits erwähnten Module zur Datenüberprüfung (s.S. 23f) und zur Anonymisierung werden vom Nachweissystem aufgerufen und greifen dann ebenfalls auf die Metadatenbank bzw. die Sachdaten zu.



Über die Weiterverarbeitungsmodulare Mdirekt, Internet-Assistent und Gizeh, die auch unabhängig vom Nachweissystem eingesetzt werden können, lassen sich Auswertungen in Form von Tabellen, Diagrammen und einfachen thematischen Karten (Mdirekt und Internet-Assistent), bzw. Bevölkerungsdiagrammen (Gizeh) erzeugen. Diese Module werden über die DUVA-Administrierung

verwaltet und ermöglichen in Abhängigkeit zu den Benutzerrechten einen Zugriff auf die Metadatenbank bzw. die Sachdatenbank.

Das Modul Mdirekt (Auskunftssystem: Erzeugung von Tabellen, Diagrammen und einfachen thematischen Karten):



Merkmale der Makrodatei
Makrodatei: Standardauswertung 01 - Einwohnerwesen (30.06.2004)

Schlüsselmerkmale | Tabellendefinition | Filterdefinition | Tabellenansicht

Tabellenvorspalte
 Familienstand Zeilensumme
 Prozent auf Zeilensumme

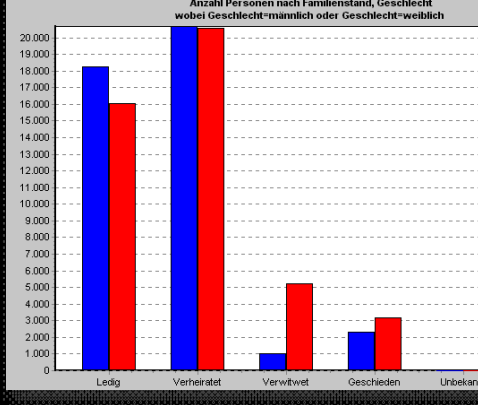
Tabellenkopf
 Geschlecht Spaltensumme
 Prozent auf Spaltensumme
 mit Verschlüsselungen

Werte
 Anzahl Einwohner
 Anzahl Personen

Makrodatei Direktauskunft
Tabelle aus der Makrodatei:
Standardauswertung 01 - Einwohnerwesen (30.06.2004)
Raumbezug: Beispielhausen
Zeitbezug: 30.06.2004
Anzahl Personen nach Familienstand, Geschlecht

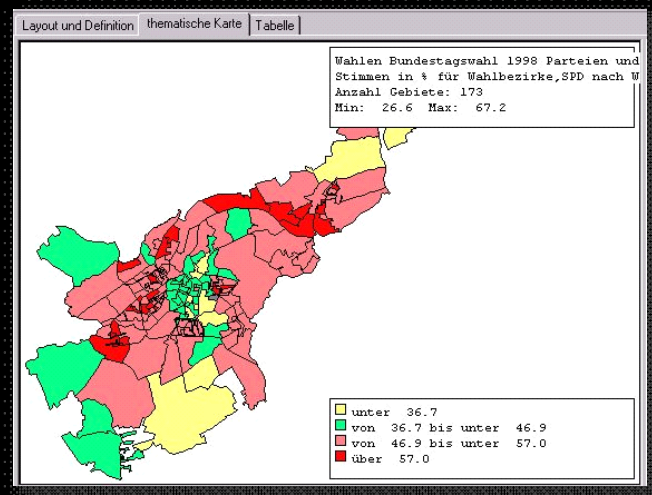
	Geschlecht	männlich	weiblich
Familienstand			
Ledig		18.257	16.070
Verheiratet		20.679	20.565
Verwitwet		1.038	5.207

Standardauswertung 01 - Einwohnerwesen (30.06.2004)
 Standardauswertung 01 - Einwohnerwesen (30.06.2004)
 Anzahl Personen nach Familienstand, Geschlecht
 wobei Geschlecht=männlich oder Geschlecht=weiblich



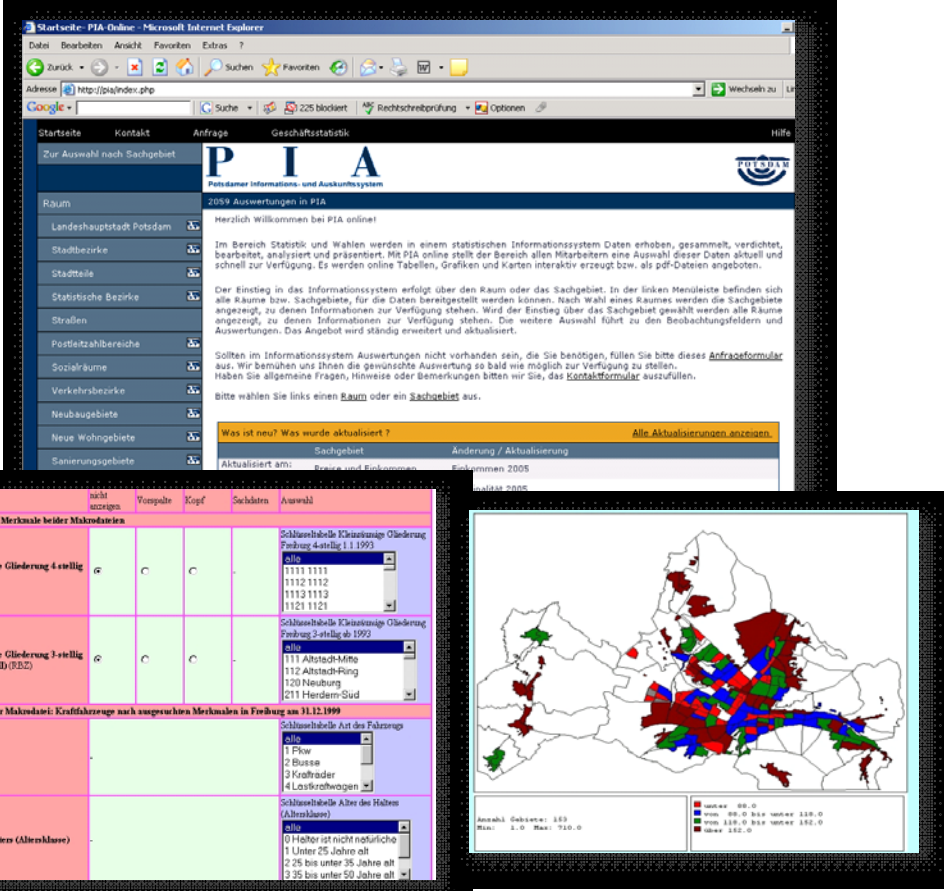
Layout und Definition | thematische Karte | Tabelle

Wahlen Bundestagswahl 1998 Parteien und Stimmen in % für Wahlbezirke, SPD nach W
 Anzahl Gebiete: 173
 Min: 26.6 Max: 67.2

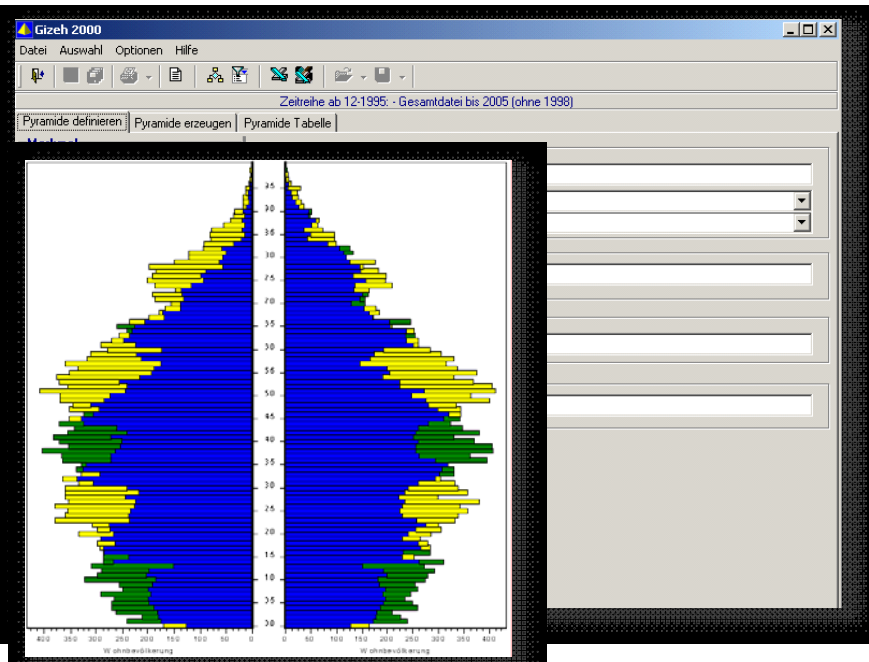


unter 36.7
 von 36.7 bis unter 46.9
 von 46.9 bis unter 57.0
 über 57.0

Das Modul Internet-Assistent (Auskunftssystem, Darstellungen von Informationen in Form von dynamischen Tabellen, Grafiken und Karten im Internet oder Intranet):

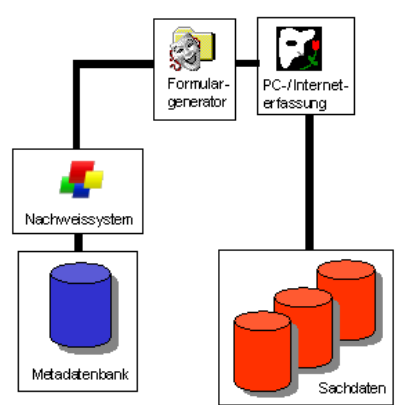


Das Modul Gizeh (Erzeugung von Pyramidengrafiken, Bevölkerungsdiagrammen)



Das Modul Formulargenerator (erzeugt aus Dateibeschreibungen Erfassungsmasken zur Datenaufnahme im Client-Server-Betrieb oder auf Web-Seiten (Internet, Intranet)).


Datenfluss DUVA-Nachweissystem und DUVA-Module



```

graph TD
    FG[Formular-generator] --- NS[Nachweissystem]
    FG --- PIE[PC-/Internet-erfassung]
    NS --- MD[Metadatenbank]
    PIE --- SD[Sachdaten]
            
```

Anmeldung bei Datenbank



DUVA Internet Erfassung

Erfassungsdatei: Museumserfassung

Bitte Daten erfassen und speichern:

Nr.	Bezeichnung	Eingabe
1	Laufende Nummer des Fragebogens	<input type="text"/>
2	Letzter Besuch in einem Museum	<input type="text" value="0 keine Angabe"/>
3	Geschlecht	<input type="text" value="0 keine Angabe"/>
4	Alter	<input type="text" value="0"/>
5	Besuchsentschluß	<input type="text" value="0 keine Angabe"/>
6	Besuchszeitpunkt	<input type="text" value="0 keine Angabe"/>
7	Benutztes Verkehrsmittel zur Anreise	<input type="text" value="0 keine Angabe"/>
8	Geschätzte Aufenthaltsdauer im Museum (min)	<input type="text" value="0"/>

Geserricht am 07.04.2005 um 10.10.36 Uhr
Internet-Erfassung V2.0.B01

Laufende Nummer des Fragebogens

Verkehrsmittel

Drogeriewaren

Kleidung

Sonstige

Geburtsdatum (JJJJMMTT)

Orte im Landkreis Beispielhausen

Lebensmittel

Schuhe

Spielwaren

Gesamtbetrag für alle Einkäufe

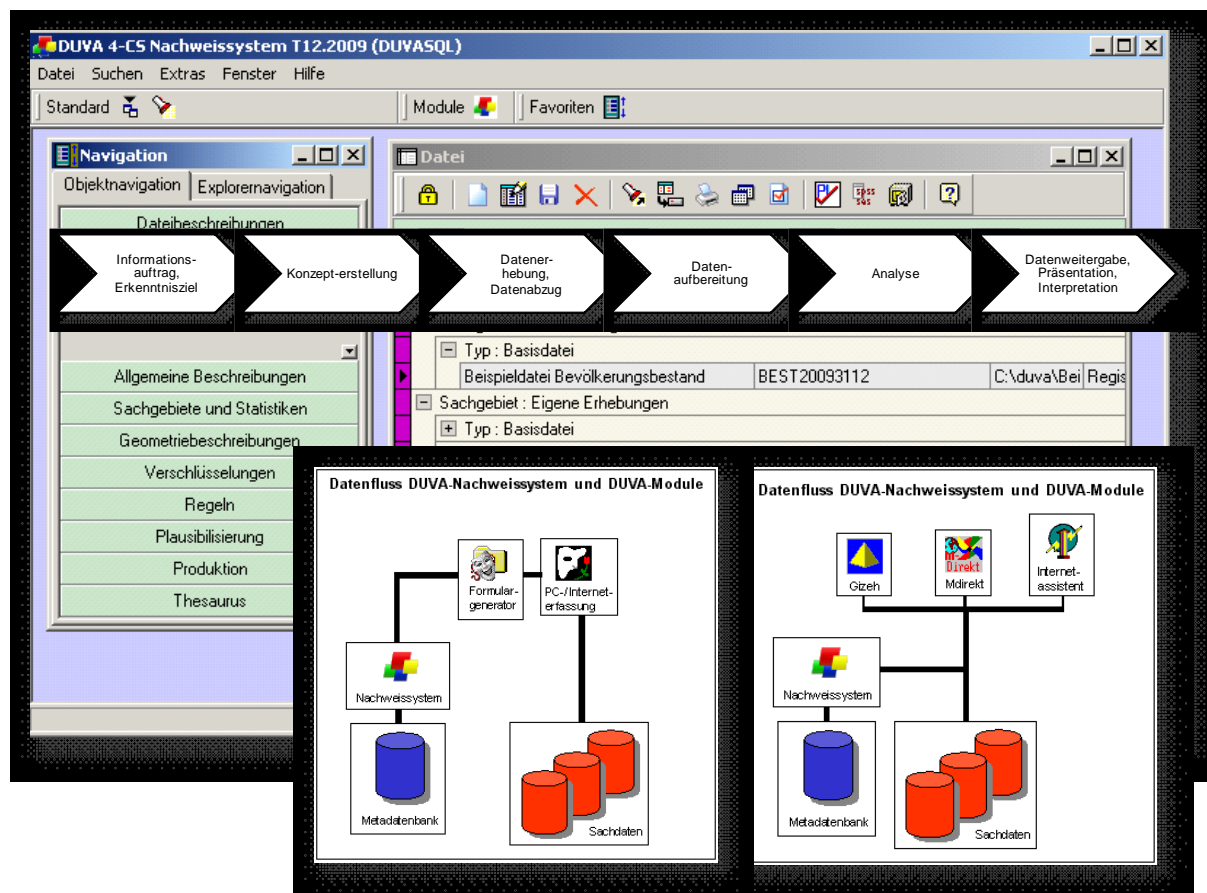
DUVA 2010 Informationsmanagement

© KOSIS-Gemeinschaft DUVA

Seite 47

3. DUVA im Einsatz

Das DUVA-Nachweissystem und die einzelnen Module ermöglichen ein metadatenbasiertes Informationsmanagement, indem sie die während des statistischen Produktionsprozesses entstehenden Metadaten bei ihrem ersten Auftreten erfassen, sie verwalten, zur Prozesssteuerung einsetzen und zu Auswertungszwecken vollständig nutzen.



Die Einbindung von DUVA in den organisatorischen Rahmen, beispielsweise einer Kommunalverwaltung, beinhaltet informationstechnologische Aspekte (Einsatz von bestimmten Softwareprodukten, Datenbanken, Konfigurierung des Client-Server-Betriebs etc.), fachliche Aspekte (welche Fachbereiche werden involviert), personelle Aspekte (welcher Kreis übernimmt konzeptionelle Arbeiten, wer führt das Auskunftssystem, wer betreut die vom Internet-Assistenten versorgten Webseiten). Die langjährigen DUVA-Anwender verfügen hier jedoch über einen großen Erfahrungsschatz, der sich über die eingerichteten Anwenderforen mobilisieren lässt (www.duva.de). Die Informationsplattform der DUVA-Foren dient aber nicht nur zum Wissensaustausch bezüglich organisatorischer Fragestellungen, sondern wird im täglichen Einsatz zum Austausch von Konzepten oder zur ad-hoc-Problemlösung frequentiert.

Ausführliche Dokumentationen zu Fragen der Installation und der informationstechnologischen Voraussetzungen sowie zu der Funktionsweise der einzelnen Module liegen in Form von Online-Dokumenten und Handbüchern vor.