

IZ-Arbeitsbericht Nr. 2

**Möglichkeiten der intelligenten Integration
heterogener Datenbestände:
das Projekt GESINE**

Jutta Marx, Peter Mutschke, Marcus Schommler

Dezember 1995

InformationsZentrum Sozialwissenschaften
Lennéstraße 30
D-53113 Bonn
Tel.: 0228/22810
Fax.: 0228/2281-120
email: mutschke@bonn.iz-soz.de
Internet: <http://www.social-science-geis.de>

Herausgeber: Informationszentrum Sozialwissenschaften der Arbeitsgemeinschaft
Sozialwissenschaftlicher Institute e.V. (ASI)
Das IZ ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V.
(GESIS), einer Einrichtung der Wissenschaftsgemeinschaft Blaue Liste (WBL)

Inhalt

1 Vorbemerkungen	3
2 Informationszentrum Sozialwissenschaften	3
2.1 Literatur-Beschaffung und Bearbeitung	4
2.1.1 Literatur-Beschaffung	7
2.1.1.1 Auswertung der Deutschen Bibliographie	7
2.1.1.2 Titeltkarten der Kölner Universitätsbibliothek	8
2.1.1.3 Auswertung von Zeitschriften	9
2.1.2 Lieferung	9
2.1.3 Vorarbeiten	9
2.1.3.1 Bücher und Zeitschriften: Inhaltliche Selektion	9
2.1.3.2 Vorarbeiten bei Datenträger-Zulieferungen	10
2.1.4 SOLIS-Aufnahme und -Redaktion	10
2.1.4.1 Printmedien	11
2.1.4.2 Datenträger-Zulieferungen	12
2.1.5 Freigabe der Daten	12
2.1.6 Problematische Punkte und Vorschläge zur Erfassung von Literatur in SOLIS	12
2.1.6.1 Einscannen von Texten	13
2.1.6.2 Automatische Texterschließung (Deskriptorenvergabe)	13
2.1.6.3 Automatische Rechtschreibkorrektur	14
2.1.6.4 Nutzung der Datenträger-Form der Deutschen Bibliographie	14
2.1.6.5 Elektronische Journalführung	15
2.2 Erhebung über sozialwissenschaftliche Forschungsvorhaben	15
2.2.1 Erhebungsphase	15
2.2.2 Erfassung des Rücklaufs	16
2.2.3 Problematische Punkte und Verbesserungsmöglichkeiten	17
2.3 Auftragsrecherchen	

2.4 SoFid-Erstellung	20
2.4.1 Arbeitsablauf	20
2.4.2 Problematische Punkte im Arbeitsablauf bei der soFid-Erstellung	22
2.4.2.1 Hoher Anteil manueller Tätigkeiten auch bei der inhaltlichen Arbeit	22
2.4.2.2 Organisation der Dateneingabe bei Zuordnungen	22
2.4.2.3 Arbeitsablauf mit vielen Beteiligten bzw. vielen verschiedenen Stationen	23
2.4.2.4 Inhaltliche Erschließung	24
3 GESIS-Außenstelle Berlin	25
4 Zentrum für Umfragen, Methoden und Analysen	26
4.1 Methodenberatung	26
4.2 Fragebogenberatung	27
4.3 Entwicklung und Wartung sozialwissenschaftlicher EDV-Programme	28
5 ALLBUS-Abteilung bei ZUMA und Aufgabebereich des ZA bei der ALLBUS-Erstellung	29
5.1 Kurzbeschreibung des Projekts ALLBUS	29
5.2 Arbeitsgebiete von ZUMA und ZA bei der ALLBUS-Erhebung	30
6 Intra- und interinstitutionelle Problembereiche	31
6.1 Intrainstitutionelle Problembereiche	31
6.2 Interinstitutionelle Problembereiche	32
6.3 Integrierte Recherchekomponente auf der Basis heterogener Datenbestände	32
7 Lösungsansätze	34
7.1 Workflow-Komponenten/Process Reengineering	35
7.2 Behandlung heterogener Datenbestände	37
7.2.1 Virtuelle Integration aller GESIS-Datenbestände	37
7.2.2 Informationsverdichtung und Aggregation	40
7.3 Konsistente Oberflächengestaltung	43
8 Literatur	43

1 Vorbemerkungen

Der erste Teil dieses Arbeitsberichts beinhaltet eine Beschreibung der momentanen Struktur und Arbeitsweise der GESIS-Institute Informationszentrum Sozialwissenschaften (IZ) und Zentrum für Umfragen, Methoden und Analysen (ZUMA) sowie das Zentralarchiv für Empirische Sozialforschung (ZA) soweit es die Aufbereitung und Archivierung der ALLBUS-Erhebung betrifft¹. Hierzu fand eine Analyse der Arbeitsabläufe des jeweiligen Instituts sowie die Sichtung der computerunterstützten Arbeitsschritte statt. Im direkten Gespräch mit erfahrenen Mitarbeitern bzw. durch die Simulation typischer Vorgänge wurde versucht, ein möglichst genaues Bild von der Materialakquisition, -aufbereitung und Archivierung einerseits sowie der Annahme von Benutzeranfragen, deren Verwaltung und Bearbeitung andererseits zu gewinnen. Ein weiteres zentrales Arbeitsfeld der GESIS ist die Produkterstellung (Bücher, Disketten, CD-ROMs) auf der Basis der vorhandenen Datenbestände (cf. Kap. 2.4).

Im zweiten Teil des Berichts werden die bei der Institutsbeschreibung herausgearbeiteten Problempunkte näher beleuchtet und mögliche Lösungsansätze skizziert. Diese lassen sich einteilen in die Gruppen "institutsinterne und institutsübergreifende Prozesse" sowie die "Integration unterschiedlicher Datenbestände" (cf. Kap. 6).

2 Informationszentrum Sozialwissenschaften

Im folgenden sollen einige zentrale und wiederkehrende Arbeitsprozesse des IZ dargestellt und diskutiert werden. Als beispielhaft für Routinearbeiten innerhalb des IZ wurden dabei die folgenden Bereiche herausgegriffen:

- Die Ermittlung relevanter Literatur, ihre Beschaffung und Erfassung in der Datenbank SOLIS
- Die Erhebung über laufende sozialwissenschaftlichen Forschungsvorhaben und die Aufnahme der gewonnenen Daten in der Datenbank FORIS

¹ Analog zu IZ und ZUMA wurde auch vom ZA eine Gesamtanalyse angefertigt und den Gesprächspartnern und Kontaktpersonen weitergeleitet. Diese liegt z. Z. dem Zentralarchiv zur Beurteilung vor und ist daher noch nicht zur Publikation freigegeben.

- Die Durchführung von thematischen Recherchen zu Literatur oder Forschungsprojekten im Auftrag von Kunden
- Die halbjährliche Erstellung der Sozialwissenschaftlichen Fachinformationsdienste (soFid) auf der Basis der Neuzugänge in den Datenbanken SOLIS und FORIS.

Die ersten beiden Punkte betreffen Arbeitsabläufe, bei denen die inhaltliche Fortführung der bestehenden IZ-Datenbanken im Mittelpunkt steht. Die beiden letzten Punkte stellen Arbeitsbereiche dar, in denen auf diesen Datenbanken aufbauend Dienstleistungen in Form von individueller Beratung oder Print-Publikationen erbracht werden. Herausgegriffen wurden diese Punkte, da diese Aufgabenbereiche einen großen Teil der insgesamt verfügbaren Kapazität des IZ binden und es somit als besonders dringlich einzuschätzen ist, daß die mit ihnen verbundenen Arbeitsabläufe optimal gestaltet werden.

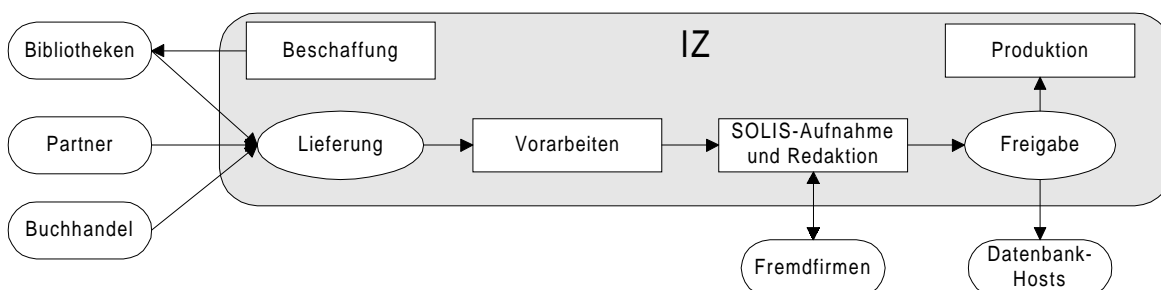
2.1 Literatur-Beschaffung und Bearbeitung

Die Datenbank SOLIS enthält inhaltlich erschlossene und mit Abstracts versehene Nachweise zu wissenschaftlicher Literatur aus dem Arbeitsgebiet des InformationsZentrums Sozialwissenschaften. Jedes Jahr werden in diese Datenbank zwischen 10.000 und 12.000 Literaturnachweise neu aufgenommen, der Gesamtbestand beträgt zur Zeit (Oktober 1995) etwa 200.000 Nachweise. Die Literaturdatenbank ist ein wichtiges Standbein für die Informationsdienstleistungen des IZ und Grundlage vieler zentraler Publikationen.

Bei der Beschaffung und Bearbeitung von Literatur für diese Datenbank sind verschiedene Varianten eines grundlegenden Arbeitsablaufs vorhanden. Die Bearbeitung im Einzelfall ist dabei abhängig von:

- Dem Anfangspunkt der hausinternen Bearbeitung. Liegt die zu erfassende Literatur als Print-Medium vor, ist ein vollständiger Arbeitsablauf durchzuführen. Bei Zulieferung von Daten durch Kooperationspartnern auf Datenträgern sind nur noch einzelne Aufgaben (Fehlerprüfung, Dublettenkontrolle, Übersetzung der Titel) zu erledigen, bevor eine Freigabe der Daten erfolgen kann
- Der Art der Literatur (Monographien, Sammelwerke, Zeitschriftenaufsätze, Graue Literatur)
- Der Vollständigkeit der vorhandenen Angaben (Vorhandensein von Abstracts bzw. Autorenreferaten und englischen Titelübersetzungen)

Unabhängig von diesen Einzelheiten zeigt die folgende Graphik die wichtigsten Stufen der Bearbeitung und beteiligte Einheiten:



Die in der Abbildung dargestellten Arbeitsschritte lassen sich wie folgt kurz charakterisieren:

- *Beschaffung*: Alle IZ-internen Vorgänge, die zur Ausleihe oder zum Kauf von Literatur führen
- *Lieferung*: Jede Anlieferung von in die Datenbanken zu übernehmendes Material, entweder als Printprodukt oder bereits auf Datenträger
- *Vorarbeiten*: Die Arbeitsschritte zwischen der Lieferung und dem erstmaligen „Auftauchen“ der zugehörigen Daten in der Datenbank
- *SOLIS-Aufnahme und -Redaktion*: Die Erfassung der Literaturangaben in SOLIS und ihre Komplettierung bis hin zur Freigabe
- *Freigabe*: Das Entfernen aller Sperrvermerke etc. für die Daten, so daß diese fortan zur Produktion und Recherche genutzt werden können
- *Produktion*: Die anschließende Verwendungsmöglichkeit der Daten für Produkte und Dienstleistungen

Mit diesen Stufen der Bearbeitung im Blick können die Arbeitsschritte für die verschiedenen Arten von Literatur bzw. die unterschiedlichen Arten des gelieferten Materials in ein einheitliches Grundgerüst eingeordnet werden. Die folgende Abbildung ordnet diese einzelnen Arbeitsschritte in diesen Rahmen ein:

Arbeitsschritte für die Erfassung von Literatur in SOLIS

		Printmedien				Datenträger	
		Monographien	Sammelwerke	Graue Literatur	Zeitschriften	Austausch-Format	Token-Format
Beschaffung		Auswertung der Deutschen Bibliographie und der Titeltkarten der Universitätsbibliothek Köln			Sporadische Auswertung nicht abonnierter ZS		
		Lieferung					
Vorarbeiten		Inhaltliche Selektion von Beiträgen/Artikeln				Datenprüfung, Ausdruck u. Korrektur	
						Umsetzung ins Token-Format	
Aufnahme u. Redaktion		Dublettenprüfung				SOLIS-Import und Plausibilitätsprüfung inkl. Dublettenprüfung	
		Bibliographische Erschließung, Eintrag ins Journal				Behandlung der Fehlerprotokolle, Korrekturen	
		Übersetzungen der Titel (Englisch und Deutsch)					
		Verteilung der weiteren Erschließung auf interne Mitarbeiter u. externe Firmen; Journalführung				Inhaltliche Kontrolle	
		Erstellen von Abstracts bzw. Erfassen von Autoren-Referaten					
		Vergabe von Deskriptoren und Klassifikationen					
		Bei Einschaltung externer Firmen: Rücklaufkontrolle und Einlesen erhaltener Daten					
		Schlußkorrektur					
		Freigabe					
		Rückgabe geliehener, Archivierung gekaufter Werke					

Im folgenden werden nun die einzelnen in der Graphik dargestellten Arbeitsschritte und -abläufe ausführlicher beschrieben. Es wird dabei aus Gründen des Platzes und der Übersichtlichkeit nicht auf alle tatsächlich vorkommenden Varianten eingegangen.

2.1.1 Literatur-Beschaffung

Ein großer Teil der eingehenden Lieferungen an Print-Medien und Datenträgern erfolgt aufgrund von Kooperationsvereinbarungen (Zulieferung durch Partner) bzw. Abonnements (Zeitschriften) ohne daß eine vom IZ ausgehende Beschaffung oder Ausleihe von Einzelexemplaren notwendig ist. Beispielsweise stellt die Soziologische Revue (Erlangen) dem IZ die eingehenden Buch-Rezensionsexemplare zur inhaltlichen Erschließung zur Verfügung. Weiterhin liefern Institute wie beispielsweise das BIB (Bundesinstitut für Bevölkerungsforschung) und das IAB (Institut für Arbeitsmarkt- und Berufsforschung) dem IZ Dokumentationseinheiten (DE) bereits auf Datenträgern zu, die nach einigen formalen und inhaltlichen Kontrollen direkt in die Datenbank übernommen werden.

Eine Beschaffung von Literatur durch das IZ erfolgt hauptsächlich zur Komplettierung dieser zugelieferten Bestände. Die ergänzende Beschaffung konzentriert sich auf die Literatur-Arten Zeitschriften, Monographien und Sammelbände.

2.1.1.1 Auswertung der Deutschen Bibliographie

Zur Komplettierung der Datenbank-Bestände an sozialwissenschaftlicher Literatur wird die „Deutsche Bibliographie“ der Deutschen Bibliothek systematisch ausgewertet. In diesem wöchentlich erscheinenden Heft werden alle in Deutschland veröffentlichten Neuerscheinungen nach Sachgruppen aufgeführt. Für die hieraus ausgewählten und nicht bereits anderweitig erfaßten Titel wird eine Beschaffung (Ausleihe) durchgeführt.

Die Bearbeitung der Deutschen Bibliographie erfolgt nach Eintreffen des aktuellen Hefts bewußt mit einigen Monaten Verzögerung. Dadurch ist die präferierte Behandlung von Monographien über Partnerzulieferungen gewährleistet und somit eine überflüssige Bestellung vermieden.

Die Bibliographie wird gelesen und interessant erscheinende Buch-Einträge werden ausgewählt (Markieren durch Ankreuzen und Anbringen von Post-It

Notizzetteln f. d. Seiten). Die Selektion kann hierbei lediglich anhand der gegebenen bibliographischen Angaben (Autor, Titel, Verlag) erfolgen.

Anhand des SOLIS-Bestandes wird überprüft, ob das Werk eventuell schon in der Datenbank vorhanden ist (Dubletten-Kontrolle). Problematisch ist hierbei, daß keine Übersicht darüber besteht, welche Monographien bereits auf dem Bestellweg sind. Eine mehrfache Bestellung kann vorkommen, da auch die Deutsche Bibliographie fehlerbehaftet ist und manchmal Neuerscheinungen mehrfach enthält (in verschiedenen Ausgaben).

Die Bibliographie-Einträge werden ausgeschnitten und auf Leihschein der Bonner Universitätsbibliothek aufgeklebt. Dann werden sie handschriftlich durch Wiederholung der Angaben auf dem rechten Teil des Scheins ergänzt.

In der IZ-Bibliothek wird auf der HBZ-CD-ROM (enthält Literaturbestand der großen Bibliotheken in NRW) zur Feststellung eines Standortes gesucht. Die Ausleihe erfolgt vorzugsweise aus Köln oder Bonn. Ist das Buch nicht in Köln vorhanden, muß unabhängig vom tatsächlichen Standort eine Ausleihe über die zuständige Leitbibliothek Bonn abgewickelt werden, da diese die Anlaufstelle für eine Fernleihe ist.

Die Leihschein werden komplettiert, evtl. wird ein Kölner statt Bonner Exemplar ausgestellt. Anschließend werden die Leihschein an die Kölner Bibliothek (einmal monatlich) bzw. Bonner Bibliothek (einmal wöchentlich) weitergegeben. Ist das Buch in Bonn nicht vorhanden, so werden die Leihschein zurückgegeben, es ist dann eine gesonderte Bestellung per Fernleihe nötig.

2.1.1.2 Titellisten der Kölner Universitätsbibliothek

Eine weitere Ergänzungsmöglichkeit für die Literatur-Datenbank des IZ besteht in der Auswertung der Titellisten für die neu angeschaffte sozialwissenschaftliche Literatur der Universitätsbibliothek Köln. Entsprechende Kopien werden dem IZ seit einiger Zeit zur Verfügung gestellt. Die Titellisten werden im IZ durchgesehen, relevante Titel werden nach einer Dublettenkontrolle bestellt. Hierbei ist eine besonders einfache Ausleihe möglich, da die Titellisten zu den gewünschten Büchern nur noch mit einem IZ-Stempel versehen und zurückgegeben werden müssen.

2.1.1.3 Auswertung von Zeitschriften

Die intern als besonders relevant für die IZ-Datenbanken kategorisierten Zeitschriften (vergebene Kategorien A und B) sind abonniert und werden dem IZ direkt zugeschickt, hier ist somit für die regelmäßige Bearbeitung keine gesonderte Beschaffung nötig. Zeitschriften der Kategorie C (bedingte Auswertung) werden in unregelmäßigen Abständen ausgewertet und bei entsprechendem Bedarf per Ausleihe bestellt.

2.1.2 Lieferung

Die Anlieferung von in die Datenbank zu übernehmendes Material kann entweder in Form von gedrucktem Material oder als Datenträger erfolgen.

Printmedien werden geliefert im Rahmen von Kooperationen (z. B. Rezensionsexemplare der Soziologischen Revue), aufgrund von Abonnements (Zeitschriften) oder als Resultat der Ausleihe bei einer Bibliothek.

Partner-Zulieferungen auf Datenträgern erfolgen in verschiedenen Formen. Beispielsweise erfolgt die Lieferung von Daten entweder auf Magnetbändern oder auf Disketten, wobei auch verschiedene Daten-Austauschformate Verwendung finden. Abhängig von dem Format der Lieferung und der Vollständigkeit der enthaltenen Informationen sind hier jeweils unterschiedliche Prüfungen bei der Aufnahme in SOLIS hinsichtlich Vollständigkeit und Plausibilität der Daten erforderlich.

2.1.3 Vorarbeiten

2.1.3.1 Bücher und Zeitschriften: Inhaltliche Selektion

Nach Erhalt von Büchern und Zeitschriften ist eine Relevanzbeurteilung des Inhalts notwendig. Die aus der Deutschen Bibliographie anhand der bibliographischen Daten „verdachtsweise“ ausgewählten und bestellten Bücher werden ebenso wie die aufgrund von Absprachen eintreffenden Bücher hinsichtlich ihres Inhalts begutachtet.

- Die vom IZ regelmäßig ausgewerteten Zeitschriften (Kategorien A und B) werden durchgesehen und relevante Artikel ausgewählt

- Sammelbände werden Beitrag für Beitrag beurteilt, so daß sie später komplett, gar nicht oder nur in relevanten Teilen aufgenommen werden.

Nach einer zentralen Relevanzbeurteilung der Zeitschriften und Bücher werden die relevanten Werke an zuständige Mitarbeiter weitergegeben, nicht relevante Exemplare werden direkt über die Poststelle zurückgeschickt.

2.1.3.2 Vorarbeiten bei Datenträger-Zulieferungen

Vor der Übernahme der maschinenlesbaren Daten in die SOLIS-Datenbank erfolgen je nach Herkunft bzw. Format der Daten noch Konvertierungen und vorgeschaltete Kontrollvorgänge.

Für Daten, die im sogenannten Austauschformat eintreffen, erfolgt zunächst im Batch eine Plausibilitätsprüfung mittels eines entsprechenden Programms und ein Ausdruck der Daten. Dabei gefundene Fehler werden anschließend nach Möglichkeit korrigiert. Nach Abschluß dieser Arbeiten erfolgt die Umsetzung in das für den SOLIS-Import benötigte, sogenannte Token-Format.

Mit einzelnen Kooperationspartnern wurde eine Lieferung im Token-Format vereinbart. Da sich dieses Format direkt in die Datenbank einlesen läßt, wird beim Eintreffen einer solchen Lieferung ein vereinfachter Prüflauf für die Daten durchgeführt. Die Prüfung beschränkt sich auf formale Aspekte (richtige Formatierung der Daten, gültige Zeilenlängen) und auf eine Kontrolle, ob alle obligatorischen Angaben vorhanden sind.

Als Ergebnis dieser Arbeitsschritte liegen die Daten nun rudimentär auf Vollständigkeit geprüft im Token-Format vor. Diese Daten werden dann in die SOLIS-Datenbank importiert. Während des Imports finden analog zu manuellen Eingaben umfassende Plausibilitätskontrollen hinsichtlich der Vollständigkeit und Konsistenz der Daten statt. Importierte Datensätze, die durch die Kontrollen ausgesondert werden, gelangen in den sogenannten Korrekturbestand und werden von Mitarbeitern korrigiert.

2.1.4 SOLIS-Aufnahme und -Redaktion

Unter Aufnahme und Redaktion in SOLIS werden die Arbeitsschritte zusammengefaßt, die mit der Erstellung eines vollständigen Literatur-Datensatzes verbunden sind.

Unabhängig von der Art der Literatur oder der Lieferung erfolgt vor der eigentlichen Aufnahme eine Dublettenprüfung. Dabei wird auf verschiedene mögliche Weisen überprüft, ob die vorliegende Literatur in der Datenbank bereits erfaßt wurde. Dabei sind beispielsweise Suchverfahren über Autor und Jahresangabe oder Stichwörter des Titels möglich.

2.1.4.1 Printmedien

Nach der Dublettenprüfung für das vorliegende Material wird zunächst die bibliographische Erschließung (Eingabe der Grunddaten wie Autoren, Titel, Standort, etc.) durchgeführt und der Stand der Bearbeitung in einem als Heft geführtem Journal festgehalten. Anschließend werden die DE für Titelübersetzungen an externe Firmen weitergegeben. Der jeweilige Originaltitel (meist in Deutsch, Englisch oder Französisch) wird durch die Übersetzung so komplettiert, daß zu allen DE deutsche und englische Titelangaben vorliegen.

Parallel zu den Übersetzungen erfolgt die Erstellung von Abstracts bzw. die Übernahme vorhandener Autorenreferate. Liegt ein geeignetes Autorenreferat (eigener Abstract bei Zeitschriftenartikeln oder Klappentexte bei Monographien) vor, so wird dieses in der Regel übernommen. Liegen dagegen keine geeigneten Textausschnitte vor, erstellen wissenschaftliche Mitarbeiter des IZ oder externe Firmen die Abstracts.

Im Rahmen dieses Arbeitsschrittes wird die Literatur weiterhin mit Deskriptoren und Methoden-Deskriptoren aus den am IZ erstellten Thesauri verschlagwortet und entsprechend den verwendeten Klassifikationssystemen eingeordnet. Auch diese Aufgaben werden auf IZ-Mitarbeiter und externe Kräfte verteilt.

Wurden Daten zur Bearbeitung außer Haus gegeben, so muß im Arbeitsablauf eine Rücklaufkontrolle vergebener Arbeiten erfolgen. Dies geschieht durch die Eintragungen im Journal und eine regelmäßige Überprüfung der dort noch offenen Posten.

Die von Fremdfirmen gelieferten Daten (Übersetzungen, Abstracts und vergebene Deskriptoren/Klassifikationen) werden nach ihrem Erhalt zu den schon vorhandenen bibliographischen Daten hinzugespielt und stichprobenartig überprüft.

2.1.4.2 Datenträger-Zulieferungen

Bedingt durch die unterschiedliche Datenerfassung bei den einzelnen Kooperationspartnern liefern diese ihre Daten oft so, daß diese zur Verwendung in den IZ-Datenbanken vor einer Freigabe noch ergänzt werden müssen. Beispielsweise werden Übersetzungen des Titels ins Englische für die IZ-Datenbanken gewünscht, diese sind aber in Partnerlieferungen nur selten vorhanden. In solchen Fällen werden die DE in den entsprechenden Teilablauf der Bearbeitung von Printmedien übernommen. Dabei werden die fehlenden Übersetzungen entweder von IZ-Mitarbeitern vorgenommen oder von externen Firmen durchgeführt.

2.1.5 Freigabe der Daten

Sind alle formalen und inhaltlichen Mängel einer DE beseitigt, so erhält diese einen entsprechenden Freigabevermerk und kann somit in den normalen SOLIS-Datenbestand eingehen. Somit ist sie für hausinterne Recherchen, die Produktion und die Weitergabe an die Hostanbieter freigegeben.

2.1.6 Problematische Punkte und Vorschläge zur Erfassung von Literatur in SOLIS

Die Arbeitsabläufe bei der Erfassung der Literatur in SOLIS sind äußerst vielfältig und stark zergliedert. Dies ist durch mehrere Faktoren bedingt. Unter anderem spielen hier die verschiedenen Literaturarten und die unterschiedlichen Vereinbarungen mit verschiedensten Kooperationspartnern über das Format von Lieferungen eine Rolle. Auch durch das nötige Nachfassen für noch nicht vorhandene Literatur, z. B. über die Auswertung der Deutschen Bibliographie und die Titeltkarten der Kölner Universitätsbibliothek werden die Arbeitsabläufe vielfältiger. Es existiert somit eine Vielzahl von Wegen, auf denen das IZ an relevante Literaturangaben gerät.

Verbesserungsvorschläge für effizientere Abläufe in diesem großen Arbeitsbereich beziehen sich zum größten Teil auf Detailbereiche, in denen Verbesserungen möglich und angeraten erscheinen. Durchgreifende Produktivitätsgewinne durch einfache oder zentrale Maßnahmen sind angesichts des komplexen Arbeitsablaufs nicht zu erwarten, die Realisierung von vielen kleineren Einzelmaßnahmen kann jedoch insgesamt auch zu einer wesentlich produktiveren Arbeit führen. Einige dieser möglichen Maßnahmen sollen in den folgenden Unterpunkten kurz angesprochen werden.

2.1.6.1 Einscannen von Texten

Längere Textpassagen wie Autorenreferate könnten sinnvollerweise mit einem Scanner eingelesen werden, um den Erfassungsaufwand zu minimieren. Erste Versuche in diese Richtung haben am IZ bereits stattgefunden. Dabei hat sich gezeigt, daß u. a. die Einbindung in die Applikation (Erfassungs-System) und die Erkennungsleistung der Software von großer Bedeutung für einen möglichen Produktivitätsgewinn sind. Bei den gegebenen Vorlagen ist die Erkennungsleistung aufgrund hoher Graphikanteile der Seiten und oftmals vielen unterschiedlichen Schriftarten z. T. noch unbefriedigend. Die Handhabung ist von Bedeutung, da in dem jetzt installierten System komplizierte Bedienabläufe beim Scannen und der Übernahme in die Erfassungs-Maske die Verwendung erheblich erschweren.

2.1.6.2 Automatische Texterschließung (Deskriptorenvergabe)

Mit der heute eingesetzten Software ist bei der Erfassung von Literatur eine rudimentäre Textanalyse für die automatische Erstellung von Deskriptoren-Vorschlägen möglich. Die Ergebnisse dieser Analyse (Deskriptoren-Vorschläge) sind in der Regel jedoch unbefriedigend, da die Vorschlagsliste sehr lang sein kann und sich unter den Begriffen viele irrelevante befinden. Aus der Durchführung einer entsprechenden Evaluierungs-Studie zu Indexierungs- und Retrievalverfahren könnten Hinweise auf leistungsfähigere Verfahren der Textanalyse hervorgehen, die hier eventuell zu integrieren wären. Ließe sich auf diesem Wege die Güte der Ergebnisse soweit erhöhen, daß mit besserer Treffsicherheit Deskriptoren-Vorschläge erzeugt werden, dann wären im Hinblick auf eine routinemäßige Verwendung die Arbeitsabläufe bei der Erfassung noch so umzugestalten, daß zum Zeitpunkt der Textanalyse die bislang erst später erstellten Abstracts zu den DE bereits in aDIS vorliegen, wenn die wissenschaftlichen Mitarbeiter eine Deskribierung vornehmen.

Momentan bildet der Bereich der inhaltlichen Erschließung mit dem Erstellen von Abstracts und der Deskriptorenvergabe für Inhalte und Methoden ein Arbeitspaket, das für den Großteil der aufgenommenen DE an externe Firmen vergeben wird. Bleibt diese Arbeitsteilung erhalten, so ist bei einer verstärkten IZ-internen Verwendung von Verfahren der automatischen Texterschließung wegen der geringen Zahl komplett im Haus aufgenommener DE keine signifikante Reduktion der Aufnahmekosten pro DE zu erwarten.

2.1.6.3 Automatische Rechtschreibkorrektur

Mit Titeln, Untertiteln und Abstracts enthalten die Datenbanken des IZ Texte in größerem Umfang. Diese haben neben den Deskriptoren für Recherchen eine besondere Bedeutung, da sie grundsätzlich über eine Volltext-Invertierung zu einer erweiterten Suche zur Verfügung stehen würden. Aus diesem Grund ist es erstrebenswert, daß die vorhandenen Texte möglichst fehlerfrei vorliegen. Eine Unterstützung des intellektuellen Korrekturlesens durch eine automatische Rechtschreibkorrektur ist bislang nicht gegeben. Die Gegebenheiten für eine Einführung einer entsprechenden Korrekturkomponente sind durchaus günstig, mit dem am IZ erstellten Thesaurus für die Sozialwissenschaften liegt bereits ein relativ umfangreiches und fehlerfreies Fachvokabular vor, das zur Ergänzung von Standard-Rechtschreibwörterbüchern herangezogen werden kann.

Auch die in den Datenbanken noch unkorrigiert vorliegenden Texte können durch Invertierung und Auflistung nach Worthäufigkeiten relativ leicht zur Ergänzung eines Wörterbuchs herangezogen werden.

Fraglich ist allerdings, wie groß die für eine sinnvolle automatische Rechtschreibprüfung am IZ notwendige Zahl von Wörterbucheinträgen ist und ob diese aufgrund der thematischen Vielfalt der Datenbanken vielleicht sehr große Zahl von Wörtern technisch noch handhabbar ist, ohne daß die Rechtschreibprüfung dadurch extrem langsam wird.

2.1.6.4 Nutzung der Datenträger-Form der Deutschen Bibliographie

Im Bereich Monographien wird die Zulieferung der Soziologische Revue durch die Auswertung der Deutschen Bibliographie komplettiert. Diese ist seit einiger Zeit auch auf Datenträgern (wahlweise Magnetband oder Diskette) erhältlich. Die dort vorhandenen Daten könnten anstelle der Papierform verwendet werden. Eine Bearbeitung auf der Basis der elektronischen Daten verspräche dabei die folgenden Vorteile:

- Weniger Doppelarbeit bei der formalen Erschließung der Literatur durch Übernahme der Daten in aDIS
- Weniger Aufwand beim Ausfüllen von Leihscheinen durch Ausdruck statt handschriftlicher Bearbeitung. Aufbauend darauf ist eine automatische Online-Bestellung ohne Leihzettel anzustreben, sobald eine solche Möglichkeit von einer geeigneten Bibliothek angeboten wird

- Möglichkeiten zur Vermeidung von Mehrfach-Bestellungen durch Realisierung einer Bestellverwaltung

2.1.6.5 Elektronische Journalführung

Bei der Erfassung von Printmedien wird von den einzelnen Bearbeitern handschriftlich ein Journal geführt, das z. B. Auskunft über die fortlaufende Aufnahmeummer, Aufnahmedatum, Weitergabe an externe Firmen etc. gibt. Das Führen dieses Journals ist mit erheblichem Aufwand verbunden, da zu jedem Eintrag Angaben wie Autor und Titel handschriftlich vermerkt werden müssen.

Ein elektronisches Journal würde die Arbeitsbelastung wesentlich reduzieren, da viele der benötigten Daten nach der Erfassung in aDIS bereits zur Verfügung stehen und nur noch wenige weitere Eingaben erforderlich wären. Die im Journal verwalteten Daten könnten in dieser Form auch wesentlich einfacher für statistische Zwecke (Auswertungen) herangezogen werden.

2.2 Erhebung über sozialwissenschaftliche Forschungsvorhaben

Die Datenbank FORIS des IZ enthält laufende, geplante und abgeschlossene Forschungsvorhaben im Bereich der Sozialwissenschaften. Zur Ermittlung dieser Daten führt das IZ einmal jährlich im Herbst (Oktober/November) eine Erhebung bei etwa 4.500 Institutionen dieses Arbeitsbereichs durch. Dabei wurden 1994 über 5.700 Projektmeldungen (Neumeldungen und Aktualisierungen) ausgewertet, der Bestand in der Datenbank für den Zeitraum von 1985-1994 beträgt rund 32.000 Meldungen von Forschungsprojekten.

2.2.1 Erhebungsphase

Die Grundlage für die FORIS-Erhebung bildet das am IZ gesammelte Datenmaterial zu Institutionen der Sozialwissenschaften. Aus dem vorhandenen Datenbestand werden diejenigen Einrichtungen selektiert, die als aktive Forschungseinrichtungen gekennzeichnet sind und die nicht durch Erhebungen von Kooperationspartnern abgedeckt werden. Letzteres trifft z. B. auf alle österreichischen oder Schweizer Einrichtungen zu, da in diesen Ländern eigene Erhebungen stattfinden, deren Resultate anschließend dem IZ zur Verfügung gestellt werden.

Aus dem aktuellen FORIS-Bestand werden zusätzlich diejenigen Forschungsprojekte recherchiert, die in den letzten drei Jahren gemeldet wurden (Neu- oder Wiederholungsmeldungen) und deren Laufzeit noch nicht beendet ist. Zu allen diesen Projekten wird ein aufbereiteter Ausdruck der vorhandenen Daten durchgeführt. Neben den vorhandenen Daten enthält dieser Ausdruck noch einen Fragenkatalog zu den noch fehlenden Angaben, der somit auf einfache Weise von den Projektmeldern komplettiert werden kann. Der gesamte Umfang aller Ausdrücke beträgt dabei etwa 20.000 Seiten.

Für alle ermittelten Institutionen druckt das IZ Etiketten für die Adressen auf den Begleitschreiben und eine dem bisherigen Meldeverhalten entsprechende Anzahl von Etiketten zum Aufkleben auf die Erhebungsbögen. Auf diesen Etiketten finden sich neben der Adresse weitere Daten, die z. B. nach einem Rückversand die Identifikation des Projektmelders vereinfachen (interner Institutionen-Schlüssel).

Die erstellten Etiketten und FORIS-Ausdrücke werden von einer externen Firma übernommen. Diese Firma druckt die Erhebungsbögen, die Begleitschreiben und das Hinweisblatt zum Ausfüllen des Fragebogens. Weiterhin wird von dieser Firma die Vorbereitung des Versands (Zusammenstellen der Sendungen und Aufkleben der Etiketten) übernommen.

Der Versand der Erhebungsunterlagen ist in der Regel bis Ende Oktober abgeschlossen, danach beginnt der Rücklauf der ausgefüllten Bögen.

2.2.2 Erfassung des Rücklaufs

Am Anfang der Rücklaufphase treffen hauptsächlich die vom Projektmelder ergänzten Fragenkataloge zu bereits erfaßten Forschungsprojekten ein, da hier der Ausfüllaufwand relativ gering ist. Bis Ende November trifft dann der Großteil aller Fragebögen ein.

Bei der Ankunft der Unterlagen im IZ wird in der Zentrale nach dem Öffnen eine grobe Statistik über die Zahl der neuen Projektmeldungen (neue Fragebögen), der Fehlanzeigen (z. B. bei abgebrochenen Projekten) und der korrigierten Ausdrücke geführt. Damit ist ein erster Überblick über den aktuellen Stand des Rücklaufs möglich.

Nach dem Aussortieren irrelevanter Fragebögen (thematische Ausreißer) werden diese auf Lesbarkeit der Angaben und Aktualität des Projekts hin durchgesehen. Aktuelle Projekte werden bevorzugt erfaßt, da nur diese für die anschließenden Publikationen relevant sind. Vor dem Erhebungsjahr

abgeschlossene Projekte werden erst später im Sinne einer Bestandspflege in die Datenbank aufgenommen. Von den Dokumentarinnen des IZ werden anschließend die Fragebögen formal erfaßt. Schwierigkeiten machen hierbei vor allem die oft stark verkürzten Angaben zu Institutionen und Geldgebern, die bei der Erfassung in den entsprechenden Datenbanken des IZ nachzuschlagen bzw. neu aufzunehmen sind.

Die wissenschaftlichen Mitarbeiter übernehmen anschließend die inhaltliche Erschließung der Fragebögen. Neben der Vergabe von Deskriptoren und Klassifikationen für das Forschungsprojekt werden dabei auch die eingesetzten Forschungsmethoden deskribiert.

Anfang Dezember wird denjenigen Institutionen in einer Nachfassaktion ein Schreiben mit einem weiteren Exemplar des Fragebogens zugeschickt, die bis zu diesem Datum nicht reagiert haben. Da bis zu diesem Zeitpunkt wegen der großen Zahl eintreffender Erhebungsbögen nicht alle Bögen komplett erfaßt werden können, wird nach dem Eingang der Bögen zunächst einmal in der Datenbank vermerkt, daß überhaupt eine Meldung erfolgte. Somit ist auch bei noch nicht abgeschlossener Erfassung der bereits vorliegenden Bögen eine relativ genaue Auswahl der Institutionen möglich, die für die Nachfassaktion in Frage kommen.

Laufend während der FORIS-Erhebung werden die erfaßten Daten portionsweise in aufbereiteter Form ausgedruckt und dem ZA und dem IAB zur Verfügung gestellt. Beim ZA erfolgt anhand dieses Ausdrucks die Auswahl der relevanten Forschungsprojekte für die Dokumentation „Empirische Sozialforschung 19..“, das IAB verwendet die Angaben ebenfalls zur Herausgabe einer eigenen Forschungsdokumentation und für eine eigene Datenbank.

2.2.3 Problematische Punkte und Verbesserungsmöglichkeiten

Der größte Aufwand bei der FORIS-Erhebung besteht sicherlich in der Erfassung der eingehenden Fragebögen. Problematisch hierbei ist auch, daß in den meisten Institutionen bedingt durch die Verdrängung der Schreibmaschinen durch Computer zur Textverarbeitung die Fragebögen mit Hand ausgefüllt werden müssen. Neben schon vereinzelt bekannt gewordenem Unmut über diese Tatsache bei Ausfüllern sind handschriftlich ausgefüllte Fragebögen oftmals auch schwer lesbar und produzieren somit einen erhöhten Aufwand bei der Erfassung.

In einer Projektskizze zu einem Nationalen Forschungsinformationssystem (NaFoDIS) wird eine elektronische Version eines Fragebogens (online oder als

verteiltes Programm) als wichtiger Punkt für eine effiziente Erhebung angesprochen. Auch die GESIS-Internetgruppe führt in ihrem Projektantrag entsprechende Arbeitspunkte auf. Unabhängig vom Bescheid zu den genannten Projektanträgen könnte diese Idee für die Durchführung der FORIS-Erhebung weiterverfolgt werden. Sollte ein entsprechendes Projekt genehmigt werden, so könnte eine FORIS-Erhebung als Versuchsballon betrachtet werden (Akzeptanz, Rücklauf, Kosten, Probleme ...).

Chancen für eine schnelle Realisierbarkeit und einen positiven Effekt sind aus folgenden Gründen gegeben:

- Es handelt sich um einen relativ isolierten und kleinen (Erfassungsteil!) Arbeitsbereich, der aber bisher einen hohen manuellen Aufwand mit sich zieht (Erfassung der ausgefüllten und eingesandten Fragebögen)
- Die Bereitschaft zum Ausfüllen des Fragebogens am PC ist durchaus vorhanden. Einige Anfragen zur Möglichkeit des Abliefers von Fragebögen per Electronic mail treffen gelegentlich ein. Die Akzeptanz einer solchen Lösung sollte jedoch möglichst in einer (informellen) Vorumfrage geklärt werden
- Je nach Art des Rückversandes können sich auch die Transportkosten reduzieren. Anzustreben wäre entweder eine direkte Bearbeitung via Internet (z. B. mittels eines Dienstes im World Wide Web) oder der Rückversand via E-Mail nach Bearbeitung mit einem lokalen Programm. Das Rücksenden von Disketten ist dagegen relativ aufwendig (Kosten) und fehleranfällig (mögl. Datenfehler durch Transport). Das Versenden des Programms an Kunden via Diskette ist ebenfalls relativ fehleranfällig und teuer (Disketten- und Kopierkosten).

2.3 Auftragsrecherchen

Ein zentrales Dienstleistungsangebot des IZ sind Auftragsrecherchen in eigenen und ergänzend in sozialwissenschaftlichen und thematisch verwandten Datenbanken anderer Anbieter. Es werden etwa 400 Anfragen pro Jahr beantwortet, diese Zahl ist in den letzten Jahren relativ konstant geblieben. Die Nutzer dieses Angebots kommen zum Großteil aus dem universitären und Forschungsbereich.

Anfragen können auf verschiedenen Wegen beim IZ eingehen. Dies reicht vom persönlichen Besuch über Telefonkontakte bis hin zu Briefen und Electronic

mail. Beim Eintreffen der Anfrage müssen zunächst die Daten des Kunden (Name, Anschrift, Telefon, etc.) festgehalten werden. Im Vorfeld einer Recherche ist weiterhin zu klären, in welchen Datenbanken recherchiert werden soll, wie groß der Umfang bzw. die Kosten der Recherche sein dürfen und wie der Versand der Ergebnisse erfolgen soll (per Post, Electronic mail oder Fax). Außerdem sind Informationen zur Rechnungsstellung zu erfragen, beispielsweise werden Studenten und ASI-Mitgliedern (ASI: Arbeitsgemeinschaft Sozialwissenschaftlicher Institute) bei Recherchen in den IZ-eigenen Datenbanken Rabatte gewährt.

Aus dem eingegangenen schriftlichen Material oder im Gespräch wird der konkrete Beratungswunsch des Kunden ermittelt und festgehalten. Dabei kann es beispielsweise auch vorkommen, daß ein Kunde nach einer gewissen Zeit eine durchgeführte Anfrage erneut stellen möchte, um in einem Fachgebiet über neuere Literatur auf dem laufenden zu bleiben.

Die am IZ erstellten Datenbanken werden bislang in ihrer Installation auf dem IZ-eigenen Zentralrechner in der Regel nicht für Kundenrecherchen verwendet, da hier eine leistungsfähige Recherche-Komponente fehlt. Statt dessen wird bei einem Host-Anbieter oder auf den CD-ROM-Editionen von SOLIS und FORIS recherchiert. Hier sollten entsprechende Möglichkeiten geschaffen werden, damit das IZ einen großen Teil der Anfragen auf dem aktuellsten Stand seiner Informationen und ohne externe Kosten beantworten kann. Bei der Verwendung von STN zur Recherche entstehen dem IZ jedoch keine weiteren externen Kosten, da hier ein entsprechender Pauschalvertrag abgeschlossen wurde, der Eigenrecherchen beinhaltet.

Die Ausführung der eigentlichen Recherche ist sehr stark vom gewünschten Abdeckungsgrad und dem vorgegebenen Kostenrahmen abhängig. Soll eine Recherche preisgünstig sein, so wird sie nur auf den IZ-eigenen Datenbanken SOLIS und/oder FORIS ausgeführt. Dazu kann z. B. STN als Host-Anbieter herangezogen werden. Muß sie nicht auf den Punkt aktuell sein, so wird auch hausintern auf der zweimal jährlich erscheinenden CD-ROM-Edition von SOLIS und FORIS recherchiert. Ist größere Aktualität nötig oder wird die Einbeziehung von weiteren Datenbanken gewünscht, wird zur Recherche auf einen Datenbank-Host zugegriffen, der neben SOLIS und FORIS noch weitere Datenbanken zur Verfügung stellt.

Im Laufe der Durchführung einer Recherche kann es zu mehrmaligen Kontakten zwischen dem Bearbeiter und dem Kunden kommen, etwa wenn eine Recherche eine zu kleine oder zu große Ergebnismenge liefert und der Kunde dementsprechend um eine Neuformulierung seiner Frage gebeten werden muß.

Ist die Recherche erfolgreich abgeschlossen, so wird sie dokumentiert und die Ergebnisse werden in der gewünschten Form dem Kunden übermittelt. Anhand der extern entstandenen Kosten und des Umfangs wird der Recherchepreis ermittelt und dem Kunden eine entsprechende Rechnung gestellt.

Zur Vorgangssteuerung bei der Bearbeitung von Anfragen werden am IZ bislang selbst gestaltete Formulare eingesetzt. Für eine zukünftig effizientere Bearbeitung dieses Aufgabenkomplexes wird hausintern bereits eine eigene Software entwickelt.

2.4 SoFid-Erstellung

Die Sozialwissenschaftlichen Fachinformationsdienste (soFid) des IZ dokumentieren halbjährlich neuere relevante Literatur- und Forschungsprojektnachweise in 33 verschiedenen Themengebieten wie beispielsweise Bildungssoziologie, Jugendsoziologie, Sozialpolitik oder Frauenforschung. Grundlage für diese Publikationen sind die Neuzugänge in den IZ-Datenbanken SOLIS und FORIS. Die erste soFid-Produktion eines Jahres beginnt mit dem Abschluß der FORIS-Erhebung für das vorangegangene Jahr jeweils um den 15. März herum, die zweite beginnt etwa Mitte August des jeweiligen Jahres.

2.4.1 Arbeitsablauf

1. Die Grundmenge an relevanten Dokumenten aus FORIS und SOLIS für alle (momentan 33 Themengebiete) zu erstellenden soFids wird bestimmt, indem alle seit der letzten soFid-Produktion neuen DE ermittelt werden.
2. Auf dieser Grundlage erfolgen für die einzelnen soFids durch die jeweiligen zuständigen Redakteure aDIS-Recherchen zur Vorauswahl des relevanten Materials. Die einzelnen Mitarbeiter sind dabei in der Gestaltung der Recherchen frei und verwenden verschiedene Verfahren. Eingesetzt werden Strategien unter Verwendung der in aDIS vergebenen Haupt- und Nebensklassifikationen oder mittels (umfangreicher) Deskriptorenlisten. Die Precision bei der Verwendung von Deskriptorenlisten kann je nach soFid-Themengebiet recht gering sein, so daß der anschließend nötige intellektuelle „Siebeaufwand“ relativ hoch wird.
3. Die Ergebnisse dieser Recherchen werden zur weiteren Bearbeitung von der EDV-Abteilung ausgedruckt. Der Umfang beträgt eine oder mehr Seiten pro DE. Die Vorgabe für den Umfang der einzelnen soFids liegt bei etwa 250 DE pro Band.

-
4. Die Mitarbeiter sehen den Ausdruck des Recherche-Ergebnisses durch und streichen alle nicht relevanten DE. Anschließend oder parallel dazu erfolgt die Zuordnung der relevanten DE zu den einzelnen Sachgebieten des jeweiligen soFids (Hinzuschreiben der Kapitelnummern und/oder Sortierung in „Haufen“ für die Kapitel). Es erfolgt unter Umständen ein intellektueller „Ausgleich“ der DE-Zahlen für die einzelnen Kapitel, wenn sich die Zahlen zu ungleich verteilen. Eventuell werden dann auch noch spezielle Recherchen zu den wenig besetzten Sachgebieten durchgeführt. Der Aufwand für Recherche, Auswahl und Zuordnung der DE liegt etwa zwischen 8 und 25 Stunden pro bearbeitetem soFid.
 5. Eine zentrale Mitarbeiterin erhält die Ausdrücke aller soFid-Redakteure zusammen mit eventuellen Änderungswünschen für die Kapitelgliederung. Es wird für jeden soFid eine Textdatei („Argumentedatei“) erstellt, die die aktuelle Gliederung enthält. Diese Argumentedatei wird um die eindeutigen Schlüsselnummern der DE und den ihnen von Redakteur handschriftlich zugeordneten Kapitelnummern ergänzt. Die Argumentedatei wird ausgedruckt und im Vergleich mit den Angaben im Ausdruck der Redakteure korrekturgelesen.
 6. Auf der Basis der Argumentedatei wird ein Abzug der Daten zu den spezifizierten DE aus aDIS erzeugt. Die Daten werden in Microsoft Word importiert, das entstehende Dokument wird layoutet und das zugehörige Register produziert. Ein Probedruck wird erstellt und auf formale Fehler hin kontrolliert. Die Redakteure kontrollieren anschließend Inhalt und Register auf eventuelle Fehler. Eventuell nötige Änderungen an den DE und am Register werden wiederum von einer zentralen Person vorgenommen.
 7. Es wird ein endgültiger Ausdruck produziert, der in der Druckerei anschließend vervielfältigt wird. Von einer externen Druckerei werden die Bindearbeiten und die Cover-Produktion übernommen. Nach der Lieferung durch die Druckerei wird der Versand an die Abonnenten durchgeführt.

2.4.2 Problematische Punkte im Arbeitsablauf bei der soFid-Erstellung

Im folgenden werden einige Punkte aufgeführt, in denen der heutige Arbeitsablauf extrem aufwendig erscheint. Entsprechende Desiderate für eine Neugestaltung der Arbeitsweise werden formuliert.

2.4.2.1 Hoher Anteil manueller Tätigkeiten auch bei der inhaltlichen Arbeit

Die Zuordnung der Nachweise zu Kapiteln bzw. das Aussortieren nicht relevanten Materials erfolgt auf dem Papier und von Hand, obwohl die zugrundeliegenden Daten in elektronischer Form vorliegen. Der manuelle Aufwand durch Notieren der Kapitelnummern, Unterteilung in Stapel für einzelne Kapitel und Auszählen der jeweiligen Kapitelumfänge ist relativ hoch. Diese Arbeitsweise wird von einigen Redakteuren als archaisch empfunden, andere sind hingegen eher froh, daß sich diese Tätigkeit unabhängig von der Benutzung eines Computers durchführen läßt. Diese Arbeitsweise führt neben dem hohen Aufwand für den Redakteur jedoch auch zu weiterem hohen Aufwand an anderer Stelle, da die handschriftlich ergänzten Angaben des Redakteurs ja irgendwo erfaßt werden müssen, um für die Publikation zur Verfügung zu stehen.

2.4.2.2 Organisation der Dateneingabe bei Zuordnungen

Die zentrale manuelle Eingabe von ISN-Nummer und Kapitelnummer für jede einzelne DE stellt einen hohen Aufwand dar, der eventuell vermieden werden könnte. Bei zweimal jährlich 33 soFids mit mindestens 250 DE müssen nach diesem Verfahren für rund 16.500 DE pro Jahr die nötigen Angaben erfaßt werden. Für jede DE sind dabei zehn bis zwölf Zeichen einzugeben, wobei für die Erfassung erschwerend hinzukommt, daß dabei auch Buchstaben getippt werden müssen („LI“ bzw. „FO“ als Herkunftsangabe Literatur/Forschung der DE).

Grundsätzlich sollte es möglich sein, parallel zum kompletten Ausdruck der soFid-Recherche-Ergebnisse auch eine separate Datei zu erstellen, die zu den ausgedruckten DE lediglich die ISN-Nummern auflistet. Mittels eines kleinen Eingabeprogramms könnte diese Angaben komfortabel um die jeweiligen Kapitelnummern ergänzt werden. Durch die Beschränkung der Eingabe auf die

reine Kapitelnummer reduziert sich der Tippaufwand (Zahl der Anschläge) auf etwa ein Drittel, durch den Wegfall der Notwendigkeit zur Buchstabeneingabe kann effizient mit dem numerischen Block der Tastatur gearbeitet werden.

2.4.2.3 Arbeitsablauf mit vielen Beteiligten bzw. vielen verschiedenen Stationen

Der heutige Arbeitsablauf für die Erstellung der soFids verteilt sich auf viele Beteiligte und verschiedene Stationen. Hier ergeben sich die folgenden, grundlegenden Nachteile:

- An bestimmten Stellen im Ablauf liegt die Arbeitsbelastung bei einzelnen bzw. wenigen Personen. Diese zentralen Personen können extrem in Zeit- bzw. Termindruck geraten, da alle Redakteure ihre Dienste beanspruchen müssen.
- Durch die Verteilung auf viele Stationen entstehen vielfältige gegenseitige Abhängigkeiten in der individuellen Termin- bzw. Arbeitsplanung. Die persönliche Planung einzelner Redakteure muß dabei in starkem Maße auf die Erledigung zentraler Dienstleistungen Rücksicht nehmen.
- Beim Auftreten von Terminüberschreitungen ist es schwierig zu ermitteln, an welcher Stelle im organisatorischen Ablauf ein Engpaß besteht, da viele Stationen involviert sind und somit eine eindeutige Identifizierung der Ursache von Verzögerungen kaum möglich ist.

Im Idealfall würde die Produktion jedes einzelnen soFids weitgehend vom zuständigen Redakteur durchgeführt werden, wobei allerdings im Hinblick auf die termingerechte Publikation aller soFids eine zentrale Koordination und Planung weiterhin notwendig wäre. Der grundlegende Gedanke hierbei ist, daß die im Arbeitsablauf im Ansatz vorhandene Parallelisierung möglichst konsequent ausgedehnt wird, wodurch gegenseitige Abhängigkeiten in der Terminplanung wegfallen und der interpersonelle organisatorische Aufwand reduziert wird. Liegt die Verantwortlichkeit für die soFid-Redaktion weitgehend in einer Hand, so kann der Redakteur seine eigene Zeit frei einteilen, ohne Wartezeiten für die Erledigung von Aufgabenteilen durch andere Mitarbeiter berücksichtigen zu müssen.

Auch wenn sich dieses Idealbild der Produktion nicht vollständig verwirklichen ließe, so könnte doch der Arbeitsablauf an einigen Stellen deutlich vereinfacht werden. Es sollte beispielsweise möglich sein (zumindest mit einer neuen

Groß-EDV), daß ein Redakteur den Druck der Ergebnisse seiner Recherchen selbst anstößt, anstatt hierzu eigens die EDV-Abteilung beauftragen zu müssen.

Einen entsprechenden Ausbildungsstand in der Benutzung von WinWord vorausgesetzt ist es weiterhin nicht einsehbar, warum eine zentrale Person die Korrekturen an der Vorversion der soFids vornehmen soll, wenn die Verantwortung für den Inhalt beim jeweiligen Redakteur liegt. Ebenfalls überflüssig erscheint der zentral vorgenommene Ausdruck der einzelnen soFids, auch diese Arbeit könnte unproblematisch den einzelnen Redakteuren überlassen werden.

Analog sollten Redakteure gewünschte Änderungen an den Gliederungen ihrer Sachgebiete direkt online vornehmen anstatt diese handschriftlich in einem Ausdruck zu spezifizieren und diesen zur Bearbeitung weiterzuleiten.

2.4.2.4 Inhaltliche Erschließung

Der Hauptteil an intellektueller Arbeit bei der soFid-Erstellung liegt in der Durchführung einer möglichst zielgenauen Recherche zur Ermittlung der Grundmenge in Frage kommender Dokumentationseinheiten und in der anschließenden Relevanzbeurteilung und Zuordnung zu den einzelnen Kapiteln.

Die Grundrecherche in aDIS kann über Deskriptorenlisten oder über geeignete Klassifikationen erfolgen, beide Verfahren führen zu einer recht geringen Precision in der Ergebnismenge, das heißt, daß noch viele der recherchierten DE im Laufe der weiteren Bearbeitung als irrelevant verworfen werden. Dementsprechend wird der Aufwand bei der Kapitelzuordnung höher, da das meiste Material nicht zugeordnet, sondern ausgeschlossen werden muß. Dieser Arbeitsschritt der Zuordnung und des Ausschlusses von DE erfordert von den Mitarbeitern eine hohe Konzentration und wird in der Regel für einen soFid auch nicht „am Stück“ erledigt, sondern auf mehrere Tage verteilt und im Tagesablauf mit anderen Arbeiten kombiniert. Dies ist ein Hinweis auf eine hohe kognitive Belastung durch diese Arbeit.

Eine Entlastung durch verbesserte Recherchemöglichkeiten in Richtung auf zielgenauere und somit verkleinerte Ergebnismengen wäre hier anzustreben.

Weiterhin ist in diesem Zusammenhang zu bemängeln, daß die bei der soFid-Erstellung geleistete intellektuelle Arbeit der Zuordnung zu einem entsprechenden Kapiteln momentan nur einmalig für die jeweilige soFid-Publikation

genutzt und nicht in der Datenbank festgehalten wird. Die Abspeicherung dieser vorgenommenen Zuordnungen würde soFid-unabhängig die thematischen Recherchemöglichkeiten in den Datenbanken erheblich erweitern. Bereits gespeicherte Zuordnungen von DE zu soFid-Ausgaben könnten aber auch in Bezug auf die soFid-Erstellung für die Grundrecherchen zu neuerer Ausgaben herangezogen werden. Über die vergebenen Deskriptorenmengen und entsprechende Verfahren ließen sich hier beispielsweise ähnliche neu aufgenommene Nachweise identifizieren und als aussichtsreiche Kandidaten für eine neue Ausgabe festhalten.

3 GESIS-Außenstelle Berlin

Die Außenstelle der GESIS in Berlin besteht seit 1992 und befaßt sich hauptsächlich mit der Bestandssicherung von Forschungsergebnissen der ehemaligen DDR sowie dem Ost-West-Transfer sozialwissenschaftlicher Information. Darüber hinaus leistet sie Beratungsarbeit in Methodenfragen und Möglichkeiten der Informationsgewinnung in den neuen Bundesländern und den osteuropäischen Ländern.

Die Bestandssicherung empirischer Daten der DDR-Sozialforschung wird von der Arbeitsgruppe "Datenakquisition und -aufbereitung Ostdeutschland" wahrgenommen und erfolgt in enger Abstimmung mit dem Zentralarchiv in Köln (cf. Fußnote zu Kap. 1). Bis 1993 konnten so 253 Studien von 1969 bis 1992 archiviert werden. Außerdem akquiriert die GESIS-Außenstelle Berlin Materialien zur Sozialforschung der DDR (Forschungsberichte, graue Literatur, Dissertationen u. ä.) und stellt sie in SOLIS zur Verfügung. Der Zugriff auf den Bonner ADABAS-Server sowie der Austausch von elektronischen Daten und die allgemeine Kommunikation zwischen Berlin und Bonn bringen hierbei einige technischen Probleme mit sich, die im 2. Arbeitspapier des Projekts "Integriertes Sozialwissenschaftliches Informationssystem" (cf. SCHOMMLER 1995b; vgl. auch SCHOMMLER 1995a) näher erläutert werden.

Im Rahmen des Informationstransfers zwischen Ost- und Westeuropa führt die Berliner Abteilung Erhebungen zu osteuropäischen Instituten, Zeitschriften und Projekten durch. Die so gewonnenen Daten werden durch gedruckte Materialien (Literatur, Zeitschriften u. ä.) angereichert, ausgewertet und in Form von Dokumentationen (soFids, Newsletter "Sozialwissenschaften in Osteuropa" etc.) veröffentlicht. Ein Teil dieser Informationen geht in die Bestände von FORIS ein, für andere bestehen lokale MS-Access-Datenbanken oder lediglich eine Datenhaltung in Form von Textfiles. Das grundsätzliche Problem hierbei

besteht darin, daß die Datenstrukturierung in SOLIS und FORIS noch nicht auf die zusätzlichen Bedürfnisse der Osteuropaforschung wie Profile von Zeitschriften (z. B. anhand einer Kurzbeschreibung oder durch Erfassen der Hauptbeiträge, Inhaltsverzeichnisse u. ä.) oder auch Profile von Wissenschaftlern abgestimmt sind. Solche Anforderungen sind darüber hinaus auch für die bereits erfaßten Datenbestände interessant und müssen in das Datenbankdesign eines integrierten Informationssystem einfließen (cf. Kap. 6.3 und 7.2).

4 Zentrum für Umfragen, Methoden und Analysen

ZUMA ist in erster Linie ein Beratungsinstitut, das "Wissenschaftler bei der Durchführung von Forschungsprojekten [unterstützt], die sich Verfahren der Empirischen Sozialforschung bedienen" (ZUMA 1993). Diese Tätigkeiten werden in Kap. 4.1 beschrieben. Ein wesentliches Teilgebiet hiervon stellt die Entwicklung und Überarbeitung von Fragebögen dar (cf. Kap. 4.2). Daneben findet bei ZUMA eigene Methodenforschung statt, die jedoch im Rahmen dieses Arbeitsberichtes (zunächst) unberücksichtigt bleibt, da die hier vorhandene bzw. in Frage kommende EDV-Unterstützung nicht über das hinausgeht, was in Bezug zur Methodenberatung zu erwähnen ist.

Die Abteilung ALLBUS, die mit der Vorbereitung und Durchführung der zweijährig stattfindenden Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) sowie der jährlichen Umfrage des International Social Survey Programme (ISSP) betraut ist, wird gesondert in Kap. 5 beschrieben.

4.1 Methodenberatung

Hauptaufgabe von ZUMA ist die Beratung von Forschern bei der Durchführung von empirischen Studien bzgl. der Erhebungsverfahren und -instrumente, der Festlegung der Grundgesamtheiten und Stichprobenauswahl sowie der Wahl der qualitativen und quantitativen Auswertungsmethoden. Des Weiteren berät ZUMA bei Fragen nach relevanter Literatur zu bestimmten Themen sowie bzgl. Referenzen ähnlicher oder verwandter Projekte. Hierbei werden u. a. auch Recherchen in den Datenbeständen des IZ vorgenommen (per Diskettendienst oder online-Recherche), was allerdings nicht integriert aus der bestehenden Arbeitsumgebung heraus möglich ist. Des Weiteren verfügt ZUMA über eigene Literaturbestände, die nicht in SOLIS enthalten sind und die bei

der Beratung von Kunden konsultiert werden. Hier ist vor allem die sogenannte ALLBUS-Bibliographie zu nennen, die ausschließlich Literatur enthält, die nachweislich ALLBUS-Daten bearbeitet hat.

Darüber hinaus kann bei ZUMA eine Softwareberatung zu Auswertungstools in Anspruch genommen werden.

Neben den täglichen Kurzberatungen, die meist telephonisch ablaufen, betreuen Mitarbeiter von ZUMA Projekte oft über den gesamten Projektverlauf hinweg und führen gelegentlich auch Pretests vor der eigentlichen Testphase durch. Die Projektbeschreibungen werden dabei lokal und individuell als Word-Dateien verwaltet. Eine Anbindung an die Datenbestände des IZ besteht nicht, so daß ein Berater bei der Betreuung eines Neuprojektes lediglich in seinem eigenen Bestand bzw. nach Rücksprache in den Beständen seiner Mitarbeiter nach Referenzen recherchieren kann. Die FORIS-Bestände sind ihm dagegen nicht zugänglich, obwohl hier u. a. Informationen über Forschungsprojekte gehalten werden, die nicht durch ZUMA beraten wurden und somit ein grundsätzlich über die ZUMA-Bestände hinausgehender Korpus verfügbar wäre.

4.2 Fragebogenberatung

Ein wesentliches Betätigungsfeld des ZUMA liegt in der sogenannten Feldabteilung, die u. a. Beratungsarbeit bei der Erstellung von Fragebögen leistet bzw. diese selbst zusammen mit den Kunden entwickelt. Auch bereits vorliegende Fragebögen von Projekten werden evaluiert und gegebenenfalls modifiziert. Hierzu verfügt die Feldabteilung über eine Fragebogensammlung, die die bis dato erstellten Bögen und eventuelle Vorstufen hierzu enthält. So ist ein Rückgriff auf bereits durchgeführte und abgeschlossene Projekte bzw. auf "bewährte" Frageformulierungen und die bei geschlossenen Fragen dem Interviewten angebotenen Antwortsets (sog. "Listenhefte") möglich, so daß die Neuformulierung von Standardfragen überflüssig oder zumindest erleichtert wird. Allerdings liegt die Sammlung nicht maschinenlesbar vor, was die Suche nach relevanten Informationen erheblich erschwert, da der Bestand mittlerweile stark angewachsen ist. Ein EDV-Archiv mit einer entsprechenden Retrievalfunktionalität könnte hier das Auffinden relevanter Fragebögen, aber auch einzelner Frageformulierungen und Antwortsets erheblich erleichtern. Entscheidende Faktoren, die bei der Bewertung von empirischen Studien herangezogen werden und die in ein elektronisches Fragebogenarchiv Eingang finden müßten, sind kontextuelle Informationen wie z. B. das Layout des Fragebogens bei einer schriftlichen Befragung, der Status der jeweiligen

Umfrage (d. i. Interview, schriftliche Befragung, Telefoninterview u. ä.) oder auch das Erfassungssystem, mit dem ein Telefoninterview durchgeführt wurde, da die Funktionalität der einzelnen Tools Einfluß auf die Qualität der resultierenden Daten haben kann.

Im Gegensatz zum Auffinden von Fragebögen können die Mitarbeiter der Feldabteilung bei der Erstellung derselben auf eine EDV-Unterstützung in Form eines Sets von Dokument- und Formatvorlagen sowie Makros in Word für Windows 6.0 (cf. Schneid 1995) zurückgreifen. Diese erlauben eine relativ komfortable Generierung von Standardelementen wie Kategorienlisten und Skalen bei geschlossenen Fragen sowie gleichbleibende Intervieweranweisungen.

Diese Funktionalität ließe sich als ein erstes Anforderungsprofil für eine Art "graphischer Fragebogenbaukasten" nutzen, der in Kombination mit einem elektronischen Fragebogenarchiv und zusätzlicher Quellen wie z. B. diverser Listenhefte (z. B. Liste aller Fernseh-, Rundfunksender u. ä.) sowie bereits vorhandener Standard-Statistiken (z. B. Standard-Statistik Sozialwissenschaften-Bus 1995), die direkte Erstellung neuer Bögen auf der Grundlage recherchierter Formulare ermöglicht (cf. Kap. 7.1).

4.3 Entwicklung und Wartung sozialwissenschaftlicher EDV-Programme

Für einige Teilbereiche der Datenanalyse und -aufbereitung wurden an ZUMA selbst EDV-Lösungen entwickelt bzw. geplant und in Auftrag gegeben. Hier ist an erster Stelle das System TEXTPACK (Mohler/Züll 1995) zu nennen, ein Tool zur computerunterstützten Inhaltsanalyse (MS-DOS und Mainframes), das mit insg. 600 Benutzern bislang bereits erfolgreich vertrieben wurde.

Zur Recherche in den Beständen der Sozialindikatoren-Datenbank, eine Datensammlung aus gesellschaftlichen Dauerbeobachtungen, wurde in Kooperation mit der Berliner Hochschule der Künste das System DISI (Digitales Informationssystem soziale Indikatoren für die Bundesrepublik Deutschland) entwickelt, das auch eine graphische Aufbereitung der Information in Form von Balkendiagrammen ermöglicht.

Im sogenannten Skalenhandbuch, das eine Kollektion psychologischer und soziologischer Umfrageinstrumente (inkl. Anwendungsbeispiele wie z. B. ALLBUS) enthält, kann der Benutzer mit Hilfe des Systems ZIS recherchieren, das neben einem hierarchischen Einstieg über eine Skalenliste eine Volltextrecherche nach Schlagwörtern erlaubt.

5 ALLBUS-Abteilung bei ZUMA und Aufgabenbereich des ZA bei der ALLBUS-Erstellung

5.1 Kurzbeschreibung des Projekts ALLBUS

Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften wird seit Mitte der 70iger Jahre gemeinsam von der ALLBUS-Abteilung des Zentrums für Umfragen, Methoden und Analysen (ZUMA) und einer Arbeitsgruppe des Zentralarchivs für empirische Sozialforschung (ZA) im Rahmen der GESIS-Kooperation durchgeführt.

ALLBUS ist eine systematische gesellschaftliche Dauerbeobachtung in Deutschland anhand einer regelmäßigen (alle 2 Jahre) repräsentativen Bevölkerungsumfrage. Die Hauptziele bestehen dabei in (cf. ALLBUS 1992, 3)

- der wissenschaftlichen "Untersuchung von Sozialstruktur, Einstellungen, Meinungen und Verhaltensweisen in der Bundesrepublik sowie des sozialen Wandels"
- der "Datengenerierung für Studenten und Forscher, die keinen unmittelbaren Zugang zu Primärdaten haben"
- einer "politikrelevanten ... deskriptiven Sozialberichterstattung".

Der Fragenkatalog enthält dabei neben einem konstanten Teil zu Familie, Beruf, Politik und Wirtschaft einen variablen Teil, der Einstellungen und Verhaltensweisen der Bevölkerung zu aktuellen Themen wie Wiedervereinigung oder AIDS erfaßt. Außerdem umfaßt die Erhebung im Anschluß an das mündliche Interview im Rahmen eines sogenannten schriftlichen drop-off Fragen des ISSP-Umfragenkatalogs (International Social Survey Programme), die insbesondere Auskunft geben sollen über den sozialen Wandel in den am ISSP-Programm beteiligten Ländern (cf. ZUMA 1993,46). So können internationale Vergleichsstudien sowie zeitvergleichende Sekundäranalysen auf einem methodisch validierten und qualitativ hochwertig aufbereiteten Datenmaterial vorgenommen werden.

5.2 Arbeitsgebiete von ZUMA und ZA bei der ALLBUS-Erhebung

Die anfallenden Arbeiten zu Planung, Durchführung und Nachbereitung der ALLBUS-Studie teilen sich die GESIS-Institute ZUMA und ZA wie folgt:

- Für das Forschungsprogramm und das Gesamtdesign ist die ALLBUS-Abteilung von ZUMA zuständig. Dies umfaßt vor allem die Erstellung des Fragenkatalogs bzw. des Fragebogens, der aus selbst formulierten Fragen, Fragen des ISSP-Programms, vorheriger ALLBUS-Studien sowie anderer Umfragen verwandter Institute besteht. Dieser wird im Rahmen eines Pretests evaluiert und gegebenenfalls überarbeitet.

Des weiteren ist die Methode zur Stichprobenwahl zu treffen und die Stichprobe zu ziehen. Die Befragung selbst wird von externen Instituten wie z. B. INFRATEST durchgeführt.

Liegen die Daten des externen Zulieferers vor, führt die ZUMA-ALLBUS-Abteilung im Rahmen einer Qualitätssicherung umfangreiche formale und inhaltliche Konsistenzprüfungen durch.

- Die weitere Aufbereitung, Archivierung und Weiterleitung der so gewonnenen Daten fällt in die Zuständigkeit des ZA. Hier erfolgt die Erstellung eines sogenannten Codebuches, das aus einer Auflistung der vercodeten Einzelfragen des Fragebogens sowie der dazugehörigen Antwortkategorien bei geschlossenen Fragen besteht. Weiterhin finden sich hier technische Informationen zu Position einer Variable im Datensatz und zur Feldlänge sowie Instruktionen für den Interviewer. Wesentlicher Bestandteil des Codebuches sind außerdem pro Frage je zwei Tabellen mit den absoluten und relativen Antworthäufigkeiten für Gesamtdeutschland bzw. gesplittet nach Ost und West.

Um die so aufbereitete Studie in den Datenbestandskatalog des ZA aufnehmen zu können, muß eine sogenannte Studienbeschreibung angefertigt werden, die allgemeine Informationen über Art und Umfang einer Studie, Angaben über die Datenlieferer, Literaturhinweise sowie eine Kurzzusammenfassung der in der Studie gestellten Fragen enthält.

Die aus diesem Bearbeitungsschritt resultierenden Daten werden auf Magnetbändern archiviert, die mit einem Inventarisierungssystem verwaltet werden. Dieses erlaubt ZA-Mitarbeitern eine Recherche nach der vergebenen ZA-Nr. und gibt neben dem Standort der Kollektion, d. h. die Bandinventarisierungsnummer auch die jeweilige Studienbeschreibung aus.

Die Weitergabe der ALLBUS-Daten an Interessierte erfolgt ebenfalls durch das ZA, wobei die Nachfrage im Vergleich zu anderen bei ZA gehaltenen Studien als überdurchschnittlich einzustufen ist (GESIS 1993,47).

6 Intra- und interinstitutionelle Problem- bereiche

Aus der Analyse des Ist-Zustandes der drei GESIS-Institute sowie durch die Betrachtung der zu bewältigenden Arbeitsaufgaben haben sich neben dem zentralen Bereich, verschiedene Informationstypen innerhalb eines IR-Systems während eines Rechercheprozesses zu integrieren, zahlreiche Ansatzpunkte für EDV-unterstütztes verteiltes Arbeiten bzw. sog. Workflow-Komponenten ergeben, mit denen effizienteres Arbeiten und deutliche Informationsgewinne hinsichtlich der Erschließung, Vermittlung und Aufbereitung von Information zu erwarten sind. Dies betrifft sowohl Strukturen innerhalb eines jeden Instituts als auch die Kooperation zwischen den Instituten, d. h. eine Integration intra- und interinstitutioneller Prozesse, auf die in den folgenden Kapiteln eingegangen wird.

6.1 Intrainstitutionelle Problembereiche

Auf der intrainstitutionellen Ebene existieren eine Vielzahl von Insellösungen, die hinsichtlich Oberflächengestaltung und Kommunikationsfähigkeit, d. h. Integrität mit anderen Applikationen, nicht mehr dem heutigen Stand der Technik entsprechen. Diese verlangen einerseits, bedingt durch die große Bandbreite der Bedienkonzepte, ein hohes Maß an Benutzerschulung bzw. -vorwissen. Eine weitere Folge dieser voneinander abgekoppelten Arbeitsprozesse und Insellösungen ist die Redundanz von Arbeitsschritten, die aufgrund der nicht gegebenen Kommunikationskanäle bzw. dem fehlenden Informationsaustausch mit anderen inhaltlich eng verwandten Komponenten entsteht. Beispiele hierfür sind die Redundanzen bei der Literaturbeschaffung am IZ oder auch die

unterschiedlichen Adreßbestände verschiedenster Ausprägung an allen Instituten.

An allen GESIS-Instituten gibt es zentrale Domänen, die überhaupt noch keine EDV-Unterstützung erfahren. Ein Beispiel hierfür ist die Feldabteilung des ZUMA, die durch ein Fragebogenarchiv effizient in ihrer täglichen Arbeit bei der Benutzerberatung durch gezielten Rückgriff auf bereits früher erstellte Bögen unterstützt werden könnte.

6.2 Interinstitutionelle Problembereiche

Die interinstitutionellen Prozesse, d. h. die Abwicklung eines Projektes in seiner Ganzheit, angefangen von der Planung und Beratung einer Studie beim ZUMA, über die Akquirierung und Archivierung der Daten beim ZA bis hin zur Aufnahme eines Projekts und der entsprechenden Literatur in FORIS und SOLIS, werden zur Zeit noch nicht optimal unterstützt. Dies läßt sich beispielhaft am ALLBUS-Projekt (cf. Kap. 5) zeigen, da hier eine Vielzahl von Personen unterschiedlichster Aufgabengebiete mit der Bearbeitung gemeinsamer Daten bzw. Dokumente in unterschiedlichsten Entwicklungsstufen involviert sind. Diese Arbeitsteilung kann durch den Einsatz von Groupware-Systemen und darauf aufbauenden Workflow-Komponenten, für die Koordination von Gruppenarbeitsprozessen und die Verbesserung des Informationsflusses unterstützt werden. Zum Informationsfluß zählt hier neben dem Durchreichen der zu bearbeitenden "Primärdaten" (z. B. Literatur, Fragebögen) auch der Austausch von "Metainformationen" wie etwa Terminabsprachen zwischen einzelnen Beteiligten (cf. 7.1).

6.3 Integrierte Recherchekomponente auf der Basis heterogener Datenbestände

Das fehlende informationstechnologische Zusammengehen spiegelt sich bei der Benutzerberatung und Informationsvermittlung entsprechend wider. Es ist für keinen GESIS-Mitarbeiter möglich, in den Datenbeständen der Partner-Institute zu recherchieren, obwohl bei der Kundenberatung ein konkreter Bedarf nach einer integrativen Sicht auf die GESIS-Daten besteht, um z. B. zu einer laufenden Studie inhaltlich ähnliche Projekte oder Literatur zum Thema zu finden. Die Zusammenarbeit beschränkt sich in diesen Fällen auf den Verweis auf die jeweilige Partnerorganisation, was von allen betroffenen Mitarbeitern als Manko empfunden wird.

Ein grundlegendes Problem ist hier die Heterogenität der verschiedenen in den GESIS-Instituten eingesetzten Retrieval- und Informationssysteme (GRIPS, Messenger, ISYS, ZIS, DISI) hinsichtlich der Rechtersprachen und der Datentypen (Fakten, Texte). Effektive Recherchen verlangen vom Benutzer intime Kenntnisse der Datenbank-Schemata und (unterschiedlichen) Retrievalsprachen und sind somit immer noch eine Sache für Insider. Dies sind Voraussetzungen, die von einem sporadischen Informationssuchenden wie dem durchschnittlichen Sozialwissenschaftler nicht erwartet werden können.

Darüber hinaus wird die Suche nach Information aufgrund der Limitierungen der eingesetzten Datenbanksysteme erheblich eingeschränkt: Literatur-, Projekt- oder Studiennachweise sind als separate Informationseinheiten abgelegt, die untereinander nicht in Beziehung stehen. Beispielsweise ist es derzeit nicht möglich, dynamisch von einem GESIS-Bestand in einen anderen zu wechseln, sich über einen komplexeren, die klassischen GESIS-Datentypen (Literatur, Projekte, Skala, Datensatz) übersteigenden Zusammenhang, wie z. B. das Profil einer sozialwissenschaftlichen Institution oder die Entwicklung von Themen, zu informieren oder etwa ergebnisorientiert zu recherchieren. Man denke z. B. an eine zeitvergleichende Fragestellung, bei der der Benutzer einen Schwellenwert hinsichtlich der Anzahl von Singlehaushalten vorgibt (bspw. durch Manipulation der Statistiken im Codebuch) und das System zu einer entsprechenden Wohnstudie des Datenbestandskatalogs verzweigt, in der dieser Wert erreicht wird, sowie gleichzeitig kontextsensitiv adäquate Literatur zum Thema präsentiert.

Auch thematisch zusammengehörige textuelle Daten sind bislang noch nicht miteinander vernetzt. Dies wird an den Studienbeschreibungen des ZA-Datenbestandskatalogs deutlich, die nicht die ursprünglichen Frageformulierungen einer Studie enthalten, sondern lediglich eine Zusammenfassung hiervon. Ein gleichzeitiges Recherchieren in den Studienbeschreibungen und den konkreten Frageformulierungen in den Codebüchern ist allerdings nicht möglich, was eine Rückversicherung über die wahre Intention einer Frage und auch das Suchen nach detaillierten Inhalten verhindert.

Voraussetzung für ein Host-übergreifendes Dokumenten-Management vom individuellen Arbeitsplatz eines Sozialwissenschaftlers aus ist daher zunächst die Homogenisierung des Zugriffs auf die Datenbestände. Der Benutzer soll weder wissen müssen, wie noch wo die Daten physisch gespeichert sind. Um einen einheitlichen Zugriff auf große heterogene Datenbestände (strukturierte und unstrukturierte Daten) und ein integratives virtuelles Informationsmanagement herstellen zu können, müssen daher die unterschiedlichen techni-

schen und formal-sprachlichen Zugangsformen zu den Datenbanken über eine gemeinsame Schnittstelle vereinheitlicht werden.

Für eine inhaltlich erfolgreiche Recherche ist aber nicht nur die Homogenisierung des Zugriffs auf die Datenbestände der GESIS erforderlich, sondern auch eine effektive Retrievalkomponente, die sachlich zusammengehörige Information aus heterogenen Datenbeständen aufeinanderbezieht und dem Benutzer als Informationspakete präsentiert (Idee des *Virtuellen GESIS-Dokuments*). Der Benutzer muß die Vorstellung haben, nur mit *einer* Datenbank zu kommunizieren, in der er dann allerdings nicht nur Informationen über sozialwissenschaftliche Literatur, Projekte und empirische Studien finden kann, sondern auch über komplexere Objekte (Institutionen, Personen, Konferenzen, Zeitschriften etc.), die sich aus den bislang auf separate Datenbanken verteilten Grunddatentypen zusammensetzen. Dazu ist ein intermediäres Metamodell der Basisdaten zu entwickeln, das verteilte Daten logisch aufeinander bezieht und informationelle Mehrwerte durch Informationsverdichtung erzeugt.

Längerfristig ist die Anreicherung mit intelligenten Werkzeugen notwendig, die z. B. eine Relevanzanalyse recherchierter Dokumente im Hinblick auf die Vagheit von Benutzeranfragen durchführen können. Von grundlegender Bedeutung ist auch die Modellierung kognitiver Suchmuster, die z. B. die gezielte Suche nach Überblicks- oder Spezialliteratur, inhaltlich ähnlichen, sich thematisch ergänzenden oder interdisziplinären Publikationen, Projekten oder Umfragen erlauben.

7 Lösungsansätze

Die Skizzierung der bisher sich abzeichnenden Lösungsansätze der oben beschriebenen Problembereiche orientiert sich an der Unterscheidung zwischen Prozessen einerseits (Kommunikation und Datenaustausch innerhalb und zwischen den Instituten) sowie Objekten, d. i. die vorliegenden Datenbestände, andererseits. Während die Prozesseite unter den Stichwörtern "Workflow" bzw. "Process Reengineering" (cf. Kap. 7.1) gefaßt ist, findet sich die Behandlung der zur Disposition stehenden Objekte im Kap. 7.2.

7.1 Workflow-Komponenten/Process Reengineering

Nach erster Sichtung ergeben sich für die ALLBUS-Erstellung folgende Ansatzpunkte für eine Unterstützung durch entsprechende Software-Komponenten:

- Ein gewinnbringender Schritt besteht in der Bereitstellung eines Fragebogen-Editors. Vorarbeiten hierzu liegen in Form eines Sets von Dokument- und Formatvorlagen sowie Makros in Word für Windows 6.0 (Schneid 1995) bereits vor. Die mit diesem Editor erstellten Fragebögen wären mit gemeinsamer Zugriffsmöglichkeit zu halten und stünden so auch den Mitarbeitern beim ZA für die Vercodung und Studienbeschreibungen zur Verfügung. Dies würde den momentan noch betriebenen Aufwand für die wiederholte Erfassung der Fragebögen beim ZA überflüssig machen.
- Weitere Unterstützung bei der Fragebogenerstellung und -verarbeitung könnte ein elektronisches Fragebogenarchiv bieten, das die konkreten Frageformulierungen in Kombination mit dem graphischen Layout enthält. Hier wären adäquate Recherchewerkzeuge zur Suche nach bereits "erprobten" Fragen zur Verfügung zu stellen sowie die Möglichkeit der direkten Übernahme in den Fragebogen-Editor, um ein direktmanipulatives Arbeiten zu gewährleisten.
- Ein weiterer Vorteil zentral gehaltener bzw. allen Beteiligten online zugänglicher Informationen bestünde in einer dann möglichen Unterstützung der Aggregation von ALLBUS-Fragen bei der inhaltlichen Erschließung in der Studienbeschreibung. Dies ließe sich erreichen, indem die Mitarbeiter der ALLBUS-Abteilung bei ZUMA die eventuell bereits bei der Erstellung des Fragebogens entstehenden Vorläuferversionen, Grobklassifikationen oder -gliederungen der Studienbeschreibung zur Verfügung stellen. So ließe sich die Diskrepanz zwischen dem Top-Down-Ansatz bei ZUMA (Überbegriffe/Grobklassifizierung => Feindifferenzierung) und dem Bottom-up-Vorgehen des ZA (Einzelformulierungen => Überbegriffe/Gruppenbildung) minimieren und somit auch der Arbeitsaufwand deutlich reduzieren.
- Die Instrumentarien zur Erfassung der formalen Bestandteile der Studienbeschreibungen wie Erhebungszeitraum, Stichprobenumfang

u. ä. sind ebenfalls online zur Verfügung zu stellen, da diese Informationen von unterschiedlichen Personen in den beteiligten Instituten und eventuell vom externen Erheber (INFRATEST) eingeholt werden müssen bzw. geliefert werden. Indem gemeinsames Bearbeiten zugelassen wird, können die entsprechenden Daten bereits zum Zeitpunkt ihres Vorliegens bzw. Bekanntwerdens eingegeben werden, d. h. im Moment der Genese, was den weiteren Arbeitsaufwand minimieren dürfte (der Aufwand für das Wiederauffinden von Information entfällt). So muß der einzelne Bearbeiter nicht warten, bis das entsprechende Dokument zu ihm durchgedrungen ist. Damit wird der Einzelne in wesentlich stärkerem Maße als bisher von Arbeitsrhythmus und -geschwindigkeit anderer unabhängig, Rückfragen wegen fehlender Einträge aus vorherigen Arbeitsschritten können entfallen oder werden zumindest erheblich reduziert.

- Die bei ZUMA gehaltenen Literaturhinweise zu ALLBUS-Publikationen müssen wie auch die IZ-Datenbanken SOLIS und FORIS online aus der Arbeitsumgebung heraus verfügbar sein, damit sie z. B. für die Verfassung der Studienbeschreibung optimal genutzt werden können. Andererseits sollten ZA-Mitarbeiter neue Literaturhinweise direkt und reibungslos ("papierlos") an die ZUMA-Literaturbestände weitergeben können, um allen Beteiligten ein Optimum an verfügbaren Informationen zur Verfügung stellen zu können.

Weitere Optimierungen des ALLBUS-Erstellungsprozesses durch Einsatz von Groupware- und Workflowelementen sind anhand einer detaillierten Analyse der Anwendungsdomäne zu erwarten und müssen wie auch die oben genannten Punkte auf ihre Realisierbarkeit und Akzeptanz hin geprüft werden.

Die bei diesem Schritt gewonnenen Erkenntnisse über den adäquaten Einsatz von Workflow-Komponenten sind auch auf ihren möglichen Einsatz bei Gruppenarbeitsprozessen im Internet, bspw. bei einer gemeinsamen Erstellung einer Publikation im WWW (cf. GESIS-Arbeitsgruppe "Internet" 1995) zu prüfen.

In punkto intrainstitutionelle Arbeitsgestaltung können die Vorteile eines Process Reengineering besonders wirkungsvoll durch eine Applikation anhand der Produktion von Publikationen auf der Basis vorhandener Datenbasen deutlich gemacht werden. Für dieses Arbeitsgebiet, das GESIS-weit auftritt, bietet sich eine Behandlung am Beispiel der soFid-Erstellung des IZ an. Diese ist bis

dato als eine Abfolge stark separierter Arbeitsschritte ohne dynamischen Informationsfluß realisiert, bei der die Datenbankrecherche getrennt ist vom eigentlichen Publishing-Prozeß, wobei die einzelnen Aufgaben von jeweils unterschiedlichen Mitarbeitern geleistet werden. Dadurch sind sehr viel zusätzlicher Codieraufwand sowie zahlreiche Rückfragen zwischen den einzelnen Personen notwendig. Analog zum Fragebogenarchiv bzw. dem Fragebogenbaukasten sollte auch hier das Ergebnis einer Recherche unmittelbar in eine Aufbereitungskomponente eingehen, in der die Dokumente gesichtet, sortiert, editiert und gesetzt werden können. Durch eine Art Formatvorlage bzw. Formatfilter kann ein homogenes Outfit über alle Themenbereiche einer Publikation gewährleistet werden, die nun von einer Person, angefangen von der Datenauswahl bis hin zu deren Aufbereitung, komplett bearbeitet wird.

Ein erster Prototyp für einen Teilprozeß der soFid-Produktion liegt bereits vor und wird zur Zeit anhand von Benutzergesprächen evaluiert.

Die an diesem Beispiel erprobte Vorgehensweise bzw. die dabei entwickelten Komponenten sind gerade für die im GESIS-Verband geplanten gemeinsamen Themendokumentationen nutzbar, die im Rahmen der mittelfristigen Prioritäten der GESIS definiert wurden (cf. GESIS-Prioritäten 1995).

7.2 Behandlung heterogener Datenbestände

Der Schwerpunkt dieses Projektteils besteht in der Entwicklung von Verfahren, die strukturierte und unstrukturierte Daten (Integration von Text- und Fakteninformationen) in einer GESIS-übergreifenden Arbeitsumgebung gleichermaßen behandeln und dem Benutzer in integrierter Form zur Verfügung stellen. Für die Realisierung eines derartigen Systems bieten sich folgende Arbeitsschritte und Abstufungen an:

7.2.1 Virtuelle Integration aller GESIS-Datenbestände

Die momentan noch nicht mögliche kombinierte Suche in den Datenbeständen der GESIS ist ein grundlegendes Problem und muß daher durch ein Verfahren gelöst werden, daß dem Benutzer erlaubt, in allen GESIS-Datenbeständen gleichzeitig zu recherchieren. In einer ersten Stufe muß das System dabei die aktuelle Recherche automatisch auf jeweils relevante Datenbestände übertragen und damit einen iterativen Rechercheprozeß bzw. eine Ausweitung der Suche ermöglichen. Hierbei sind folgende Problembereiche anzugehen:

- Die Integration von FORIS und SOLIS sollte vordringlich realisiert werden, weil eine flexible Lösung zur Behandlung verteilt gehaltener Textinformation

in strukturierten Dokumenten die Grundlage für die integrierte Suche in den verteilten Fakten- und Textdatenbeständen der GESIS bildet.

- Darüber hinaus sollte der Benutzer gleichzeitig im Datenbestandskatalog des ZA und den Codebüchern recherchieren können, um den Informationsverlust, der durch die von den Fragetexten der Umfragen abstrahierenden Studienbeschreibungen entsteht, durch eine kombinierte Recherche in den Frage- und Antworttexten der Codebücher auszugleichen. Eine Recherche in den Codebüchern ist bislang jedoch nicht möglich. Möchte man z. B. die Beliebtheit einer Person wie Adenauer über die Jahre hinweg beobachten, so muß der Benutzer wissen, daß er in den Studienbeschreibungen nach dem abstrakteren Begriff "Kanzlerpräferenz" suchen muß, während eine Suche nach "Adenauer" auf Frage- und Antwortebene in den Codebüchern vermutlich eher zum Erfolg führen würde.
- Ferner muß eine integrierte Suche in Datenbestandskatalog, Codebüchern, FORIS und SOLIS ermöglicht werden, so daß zu einem bestimmten, vom Benutzer angegebenen Thema bei Bedarf neben empirischen Studien auch gleichzeitig Projekt- und Literaturnachweise vom System gefunden werden. Für die Realisierung eines derartigen Systems hat sich die Möglichkeit einer kombinierten Suche in SOLIS, FORIS und ALLBUS als vordringlich herauskristallisiert.
- Des weiteren bietet sich bei den Codebüchern des ZA ein ergebnisorientiertes Retrieval auf den Randauszählungen als Ausgangsbasis an. Befindet sich der Benutzer z. B. in einer Reihen-Studie über Rollenverhalten in eheähnlichen Gemeinschaften aus dem Jahre 1970, sollte er durch (numerische oder graphische) Manipulation eines Variablenwertes in der Statistik als Ergebnis eine Studie angezeigt bekommen, in der der angegebene Wert erreicht wird. Dadurch kann die langwierige intellektuelle Sichtung einer u. U. großen Studienmenge vermieden werden. Ein derartiges Verfahren ließe sich auch auf die bereits bei ZUMA entwickelten Systeme DISI und ZIS bzw. auf die ihnen zugrundeliegenden historischen Zeitreihen übertragen.

Im Rahmen des an der Universität Regensburg durchgeführten Projekts WING-IIR wurde für die Recherche in Faktendatenbanken das Retrievalwerkzeug WING-GRAPH (cf. Wolff 1995a) entworfen, das in der Lage ist, tabellarisch abgelegte Faktendaten graphisch zu visualisieren (z. B. als Liniendiagramm) und an dieser Ergebnisrepräsentation Manipulationen zuzulassen, die als neue Query-Formulierung an die Datenbank fungieren (d. i. graphisches Ergebnis-Retrieval). Für den erwähnten Teil der GESIS-

Datenbestände hat sich dieses Modul als Ergebnis einer Anwendungsprüfung bereits als geeignet erwiesen, wobei sich eine Reihe interessanter Retrieval-Möglichkeiten ergaben, wie folgende Anwendungsszenarien verdeutlichen (cf. Wolff 1995b:6-7):

Beispiel 1: „Beliebtheit von Politikern, z. B. Adenauer“

Fragen nach der Beliebtheit von Politikern lassen sich in der Regel als Zeitreihe darstellen und sind damit einem Vergleich zugänglich. Konkret könnte das bedeuten, Zeitreihen zu einer bestimmten Fragestellung unabhängig von Zeitpunkt zu vergleichen. Damit ließen sich etwa Beliebtheitsprofile von Politikern vergleichen und kontrastieren.

Gegeben sei bspw. die aktuelle Beliebtheitskurve von Rudolf Scharping; eine vergleichende Fragestellung könnte auf der Basis dieses Profils ähnliche (bessere, schlechtere etc.) Profile recherchieren und anschließend durch die Bestimmung des jeweiligen Zeitkontextes weitergehende Interpretationsmöglichkeiten eröffnen. Natürlich kann bei ausreichend umfangreicher Datenbasis der Zeitbezug auch absolut gesetzt werden, so daß man den Vergleich der Beliebtheitsprofile von Politikern im je identischen Vergleichszeitraum erhält.

Beispiel 2: „Wann lag zum ersten Mal der Anteil der im Haushalt mithelfenden Männer über 50%“

Das Beispiel impliziert nicht unmittelbar eine graphische Retrievalfunktion, da es intuitiv zunächst nahe liegt, die entsprechenden Daten als **eine** Zeitreihe darzustellen bzw. in ein Diagramm einzutragen. Wie schon bei voranstehendem „Politikerbeispiel“ ergibt sich der Übergang zu einer graphischen Recherche erst aus einer vergleichenden Fragestellung, d. h. nicht aus der Frage nach dem Wann an sich, sondern an der möglicherweise nach Beantwortung dieses Informationsbedürfnisses auftretenden Folgefrage, wie z. B. welche vergleichbaren zeitgebundenen Entwicklungen lassen sich hierzu ermitteln?“ „Gibt es dazu radikal unterschiedliche Entwicklungen“. In beiden Folgefragen wäre es angesichts der umfangreichen Datenbestände in der Allbus-Erhebungen unerlässlich, eine Domänenbegrenzung einzuführen, d. h. recherchiert werden dürfte nur nach den Fragestellungen, die mit der

Ausgangsfrage zusammenhängen oder zwischen denen eine interessante Korrelation vermutet wird.²

- Die Integrationsmöglichkeit weiterer GESIS-Datenbestände (Skalen, Soziale Indikatoren, Osteuropa-Datenbank) muß geprüft werden.

Für all diese Datenbestände bzw. für ihre Gesamtheit ist eine Entscheidung über die zu wählende Erschließungs- und Retrievalmethode zu treffen, die wesentlichen Einfluß auf die Güte des Gesamtsystems haben wird.

Gerade in punkto Volltextdatenbanken konkurrieren schon seit langem computerlinguistische und statistische Techniken, ohne daß Erkenntnisse darüber bestehen, welches Verfahren sich für welche Domäne, d. h. Textsorte bzw. für welches Informationsbedürfnis am besten eignet. Bei den klassischen Deskriptorensystemen mit Boolescher Algebra als Abfrageoperatoren sind sehr exakte Anfragen mit exakten Ergebnismengen möglich, wobei allerdings das Problem besteht, daß sie aufgrund der harten Suchlogik alle Dokumente des Rechercheergebnisses gleich behandeln und daher eine Ergebnismenge produzieren, die in sich nicht strukturiert ist. Statistische Verfahren dagegen sind in der Lage, differenziertere Relationen, z. B. Ähnlichkeiten zwischen Dokumenten herzustellen. Allerdings sind diese Verfahren bei einem spezifischen Informationswunsch zu ungenau.

Es stellt sich daher die Frage, ob nicht gerade eine Kombination der beiden Methoden bzw. ihr kontextsensitives Einsetzen je nach Benutzerprofil, Informationsbedürfnis, Textsorte u. ä. einen Qualitätsgewinn der Rechercheergebnisse bringt. Daher müssen diese Ansätze bzw. deren Ausprägungen anhand verschiedener kommerzieller (QSearch, ISYS u. ä.) oder auch wissenschaftlicher Systeme getestet und evaluiert werden, was auf der Basis der SOLIS- und FORIS-Datenbanken erfolgen wird.

7.2.2 Informationsverdichtung und Aggregation

Über diese Grundproblematik einer virtuellen Integration von Text- und Faktendaten hinaus muß das Sonderproblem der Vagheit von Benutzeranforderungen bei der Informationssuche mit Hilfe statistischer oder computerlinguistischer Verfahren des Information Retrieval in Angriff genommen werden. Dieses Problem spiegelt sich insbesondere in folgenden Bereichen wider:

² Ein Problem dürfte hier sicher sein, daß sich gerade interpretativ interessante Korrelationen gerade nicht ex ante bestimmen lassen oder zunächst absurd anmuten mögen. Daher sollte eine Domäneneingrenzung immer durch den Benutzer modifizierbar sein.

Die in Kap. 7.2.1 beschriebene Grundfunktionalität der einfachen Verknüpfung von Datenbeständen, indem das System eine bloße Übertragung der Recherchelogik, d. h. der aktuellen Suchformulierung von einem Datenbestand auf einen anderen vornimmt, birgt das grundsätzliche Problem, das als Rechercheergebnis u. U. große Mengen von Dokumenten aus unterschiedlichen Datenbeständen ungefiltert nebeneinander gestellt und nicht untereinander in Beziehung gebracht werden. Sucht der Benutzer bspw. Informationen zu einer bestimmten Person der aktuellen Forschungslandschaft, so bekommt er als Ergebnis die von dieser Person publizierte Literatur und durchgeführten Forschungsvorhaben oder Studien. Beide Kategorien stünden dann aber in einem wie in Kap. 7.2.1 beschriebenen Grundsystem isoliert nebeneinander, ohne daß die Relationen untereinander präsentiert werden könnten (z. B. Welche Literatur bezieht sich auf welche Studie des Autors? Welche Literatur behandelt das in einer Studie X empirisch untersuchte Grundthema theoretisch? u. ä.). Um eine integrative Sicht auf die GESIS-Daten zu gewährleisten, muß ein integriertes Informationssystem der GESIS daher in der Lage sein, zu ausgewählten Dokumenten eines Ergebnis-Sets inhaltlich ähnliche Dokumente (aus demselben Set oder aus anderen Datenbeständen) zu präsentieren, damit der Benutzer z. B. zu einer laufenden Studie Projekte oder Literatur zum Thema finden kann.

Darüber hinaus kann das Problem entstehen, daß ein und derselbe Suchbegriff aufgrund der möglicherweise unterschiedlichen inhaltlichen Strukturierung bzw. unterschiedlichen Verschlagwortung in den einzelnen Datenbeständen unterschiedlich gut greift. Dies wird deutlich, wenn man sich bspw. die verschiedenen Abstraktionsebenen von Studienbeschreibungen und Codebüchern vor Augen hält, bei denen eine Suche zum gleichen Thema einmal konkret ("Adenauer"), im anderen Fall aber nur abstrakt ("Kanzlerpräferenz") erfolgreich verläuft. Bei der Fragebogen-Vercodung ergibt sich eine ähnliche Problematik hinsichtlich der Abstrahierung der Items von den konkreten Frageformulierungen, wo es i.d.R. keine linguistische Entsprechung zwischen dem Text einer Suggestivfrage und den der Frage zugeordneten Items gibt (s. z. B. 'Todessehnsucht von Jugendlichen'). Ein integriertes Informationssystem muß daher durch eine Komponente angereichert werden, die durch eine kontext-sensitive Analyse von Thesaurus- bzw. Verschlagwortungsstrukturen die Relevanz von Dokumenten (Umfragen, Literatur- oder Projektnachweise) für eine bestimmte Fragestellung bestimmen kann.

Des weiteren bieten die Datenbestände der GESIS über die klassischen, explizit dokumentierten GESIS-Datentypen (Literatur, Projekte, Umfragen, Skalen, Soziale Indikatoren) hinaus viele Ansatzpunkte für Aggregationen dieser Basis-Informationseinheiten zu Meta-Objekten. Typische Beispiele hier-

für sind Beschreibungen sozialwissenschaftlicher Institutionen, Konferenzen, Personen, Zeitschriften, Themen etc., die sich aus in den Basis-Dokumenten enthaltenen und in der Regel auf verschiedene Datenbestände sich verteilenden Einzelinformationen zusammensetzen. Da bisher nur nach den oben genannten Dokumentarten recherchiert werden kann, müssen hier Modellierungen implementiert werden, die solche *virtuellen* Dokumente generieren und dem Benutzer ebenfalls als suchbare Informationsbestandteile eines GESIS-Informationssystems zur Verfügung stellen. So sollte es z. B. möglich sein, daß das System dem Benutzer bei Bedarf (z. B. auf Anforderung oder bei Voreinstellung automatisch) zu einem Thema auch eine (einschlägige) Institution mit einer Beschreibung ihres Forschungsschwerpunktes, ihrer Publikationen und Projekte u. a. m. präsentiert.

Ein weiterer Anwendungsfall, der in den Bereich der Aggregierung von Information fällt, betrifft die Vergleichbarkeit von Skalen in verschiedenen Studien hinsichtlich der Anfrage eines Benutzers. Hier sollte ein Verfahren entwickelt werden, daß unterschiedliche Skalen in den Studien (z. B. 3er-Skala in der einen und 5er-Skala in einer anderen Studie) hinsichtlich der Benutzervorstellung von einer Skala (z. B. 4er-Skala) auf ihre Ähnlichkeit prüft und ggf. durch eine deduktive Aggregierung vergleichbar macht (z. B. 'sehr gut' und 'gut' wird zu 'gut' aggregiert). In diesem Zusammenhang sollte auch das Problem der konsistenten Vergabe von Variablennamen für die Umfragedaten gelöst werden. Es hat sich gezeigt, daß die Einheitlichkeit der Variablen-Label bei gleichartigen Fragestellungen für die Vergleichbarkeit insbesondere von Reihenstudien von großer Bedeutung ist. Hierfür sollte ein Komponente entwickelt werden, die den Inhalt der Fragetexte neu zu archivierender Studien mit den Frageformulierungen bereits vorhandener Studien vergleicht und bei einer semantischen Entsprechung derselben den Variablennamen der älteren Studie vorschlägt.

Über diese Aggregierungsformen hinaus sollte die Recherche durch den Einsatz intelligenter Komponenten, wie z. B. das am IZ entwickelte System AKCESS, verbessert werden, d. h. durch Recherchemodelle, die auf die Besonderheiten der speziellen Datenstrukturen der GESIS-Datenbanken und die Recherchebedürfnisse und -strategien der Informationsvermittler abgestimmt sind und daher über die Aussagekraft herkömmlicher Retrievalmodelle weit hinausgehen. Der in AKCESS realisierte Ansatz interpretiert z. B. komplexe wissenschaftsstrukturelle und thematische Zusammenhänge in Datenbanken unter Vagheitsbedingungen und kommt auf dieser Grundlage nicht nur zu differenzierten Aussagen über die Relevanz von Dokumenten und Wissenschaftlern für die Fragestellung eines Benutzers, sondern auch über die Struktur einer Forschungslandschaft (cf. Mutschke 1994, Mutschke 1995).

7.3 Konsistente Oberflächengestaltung

Für die im Rahmen dieses Projekts entwickelten Komponenten ist eine nach softwareergonomischen Kriterien ausgerichtete konsistente Oberflächengestaltung erforderlich. Die Vielzahl und Unterschiedlichkeit der bislang in den GESIS-Instituten eingesetzten Benutzerschnittstellen (maskenorientierte Systeme wie die FORIS und SOLIS-Oberflächen, Hypertext-Systeme wie DISI, sonstige Windows-Applikationen wie ISYS usw.) sind nicht zuletzt im Rahmen eines integrierten Informationssystems nicht mehr tragbar, da es beim Shift von einer Komponente auf eine andere zu unnötigen Reibungsverlusten bzw. zu Fehlbedienungen aufgrund falscher Analogiebildung kommt.

Im Rahmen des an der Universität Regensburg durchgeführten Projekts WING-IIR konnten aufgrund intensiver Benutzertests umfangreiche Erkenntnisse über eine benutzeradäquate multimodale Oberflächengestaltung für Fakten- und Textdatenbanken gesammelt werden. Das aus dieser Arbeit resultierende WOB-Modell (cf. Krause 1994), eine Theorie zur Entwicklung von Oberflächen, die bereits mehrfach bei kommerziellen Anwendungen die Basis bildete, soll daher zunächst als mögliches Oberflächenkonzept geprüft werden. Es bietet neben seiner empirischen Evaluierung und der modularen Ausrichtung, die eine leichte Erweiterung durch zusätzliche Werkzeuge erlaubt, u. a. den Vorteil, daß es bei gleichbleibender Gestaltung doppelt (d. h. sowohl als Formular- als auch als Werkzeugkastensystem) interpretiert werden kann und somit für Anfänger und für Experten bzw. fortgeschrittene Benutzer gleichermaßen angemessen ist. Damit kann für das breite Spektrum der GESIS-Nutzer, angefangen vom Endverbraucher der CD-ROM-Dienste bis hin zu den hoch spezialisierten Rechercheuren in der GESIS selbst ein einheitliches System zur Verfügung gestellt werden, was Mehrfachentwicklungen überflüssig macht. Auch für eine Präsentation der GESIS-Datenbestände im Internet wird zur Zeit die Anwendbarkeit des WOB-Modells geprüft (cf. GESIS-Arbeitsgruppe "Internet" 1995).

8 Literatur

- GESIS-Arbeitsgruppe "Internet" (1995): Entwicklung eines Daten- und Informationsservice der GESIS für die Sozialwissenschaften im Internet. Beschlußvorlage für die GESIS-Vorstandssitzung am 21. September 1995.
- GESIS-Prioritäten (1995): Mittelfristige Prioritäten der GESIS. Vorstandsvorlage für das Kuratorium und die Mitgliederversammlung der GESIS vom 3.4.1995.

-
- Krause, J. (1994): Das WOB-Modell. Auf der Werkzeugmetapher basierende strikt objektorientierte grafisch-direktmanipulative Benutzungsoberflächen. WING-IIR Arbeitsbericht 53
- Mohler, P.; ZÜll, C. (1995): TEXTPACK PC, Release 5.0, Short Description. ZUMA Mannheim, July 1995.
- Mutschke, P. (1994): Processing Scientific Networks in Bibliographic Databases. In: H.H. Bock u. a. (Hrsg.): Information systems and Data Analysis. Prospects - Foundations - Applications. Springer-Verlag. Heidelber-Berlin. 127-133.
- Mutschke, P. (1995): Uncertainty and Actor-Oriented Information Retrieval in μ -AKCESS. An Approach Based on Fuzzy Set Theory. In: H.H. Bock u.a. (Hrsg.): Data Analysis and Information systems. Statistical and conceptual approaches. Springer-Verlag. Heidelberg-Berlin. 126-138.
- Schneid, M. (1995): Fragebogen schreiben mit Microsoft WORD für Windows Vers. 6.0. ZUMA-Arbeitsbericht Nr. 95/01. Mannheim:ZUMA.
- Schommler, M. (1995a): Strukturierte Zusammenfassung zu den Gesprächen in der GESIS-Außenstelle Berlin am 11./12.9.95. Gesprächsprotokoll.
- Schommler, M. (1995b): EDV-Einsatz am IZ: aktueller Stand und zukünftige Möglichkeiten. Arbeitsbericht Nr. 2 (Dezember 1995).
- Standard-Statistik Sozialwissenschaften-Bus (1995): Standard-Statistik Sozialwissenschaften-Bus. Stand: Januar 1995. Erstellt von ZUMA Mannheim und GFM-GETAS (Gesellschaft für Marketing-, Kommunikations- und Sozialforschung mbH) Hamburg.
- Wolff, C. (1995a). Graphisches Faktenretrieval mit Liniendiagrammen. Schriften zur Informationswissenschaft Bd. XXX. Konstanz.
- Wolff, C. (1995b): Anwendungsmöglichkeiten und -szenarien für eine graphische Retrievalkomponente im Kontext sozial- und wirtschaftswissenschaftlicher Datenbestände. IZ-Arbeitsbericht, September 1995, Bonn.
- ZUMA (ed.) (1993): Jahresbericht 1993. Neustadt/Weinstraße.