

AD-II: 4. Arbeitsbericht (2006)

Hans Goebel, Edgar Haimerl,
Andreas Wagner

1. Vorbemerkung

Der hiermit präsentierte vierte Arbeitsbericht zum *AD-II* bezieht sich auf den etwa zwischen Pfingsten 2006 und Pfingsten 2007 gelegenen Zeitraum und betrifft damit grosso modo das dritte (und somit letzte) Jahr der derzeit laufenden Förderungstranche des FWF,¹ die für die Jahre 2004–2007 ausgesprochen wurde. Wenn also im Titel dieses Beitrags das Jahr 2006 aufscheint, so dient das vor allem zur Markierung des Platzes dieses Arbeitsberichtes im Rahmen der vorhergehenden und der ihm nachfolgenden Darstellungen.

Zwischenzeitlich wurde vom österreichischen Forschungsförderungsfonds FWF die Unterstützung des Projektes *AD-II* für den Nachfolgezeitraum 2007–2010 genehmigt.²

¹ FWF: “Fonds zur Förderung der wissenschaftlichen Forschung in Österreich”, Wien. Die Nummer der derzeit noch laufenden Förderungstranche des *AD-II* lautet: 17.326.

² Die entsprechende Projekt-Nummer lautet: 20.047. Die Bewilligung erfolgte in der FWF-Sitzung vom 25.6.2007.

2. Bericht des Projektleiters (Hans GOEBL)

2.1 Personalstand, Mitarbeiter, Arbeitsabläufe

Im Jahr 2006 waren für den *AD-II* die folgenden Damen und Herren in den unten näher bezeichneten Funktionen tätig:

als vollbeschäftigte Mitarbeiter (beiderlei Geschlechts):

für die Datenerhebung im Feld:

- Walter Strauß: Seine Tätigkeit beim *AD-II* begann am 1.10.2001 und wird – nach Ablauf der beim FWF üblichen Sechs-Jahres-Frist – am 30.9.2007 enden.

für EDV-Arbeiten:

- Slawomir Sobota: Seine vom FWF finanzierte Mitarbeit dauerte vom 15.10.1999 bis zum 14.10.2005. Seither wird sein Verbleib im Projekt *AD-II* durch den Einsatz FWF-externer Mittel³ und durch die Mithilfe der Universität Salzburg sichergestellt.
- Xavier Casassas: Er hat seine Tätigkeit beim *AD-II* am 1.12.2005 begonnen und damit innerhalb des FWF-Projektes 17.326 offiziell die Nachfolge von S. Sobota angetreten. X. Casassas stammt, wie im vorhergehenden *AD-II*-Bericht bereits erwähnt worden war, aus Katalonien, hat in Spanien Linguistische Informationsverarbeitung und Arabistik studiert und lebt seit geraumer Zeit in Salzburg. X. Casassas teilt sich derzeit die Erledigung der EDV-Agenden des *AD-II* mit F. Tosques, wobei er sich ganz besonders um Belange der EDV-Erfassung und -Archivierung der im Feld gesammelten Tondokumente (siehe dazu den Abschnitt 3.2) kümmerte und den laufenden Kontakt mit unserem EDV-Supervisor E. HAIMERL und A. WAGNER, dem Vollender des Kartenerstellungs-Moduls im Rahmen des Programms DGM (siehe dazu seinen Bericht in Abschnitt 3.1), sicherstellte.
- Fabio Tosques: Er hat am 1.4.2007 die Nachfolge der am 30.9.2006 aus dem Projekt *AD-II* ausgeschiedenen Trentiner Linguistin Ilaria Adami angetreten. F. Tosques hat in Berlin an der Humboldt-Universität bei Prof. Dieter Kattenbusch, einem der Pioniere des *AD-I*, Romanistik (Italianistik) studiert und diese Ausbildung mit einem regulären Fachstudium aus Informatik kombiniert. Er ist somit für die Arbeit in der Endphase des *AD-II* ein geradezu ideal qualifizierter Mitarbeiter.

³ Siehe dazu die in Abschnitt 2.5 genannten Institutionen.

als Feldexplorator/inn/en auf Werkvertragsbasis:

- Ilaria Adami (Trient),
- Brigitte Rühlringer (Pisa) und
- Tino Szekeley (Graz).

als Programmierer für die AD-II-Software DMG (“Dialect Map Generator”):

- E. HAIMERL (Blaustein, derzeit San Francisco),⁴
- A. WAGNER (Hallein) und
- F. Tosques (früher Berlin, jetzt Salzburg).

Wie in den früheren Jahren auch arbeiten die genannten EDV-Experten (A. WAGNER, X. Casassas, F. Tosques, S. Sobota) eng mit E. HAIMERL zusammen, wobei sie sich des öfteren der heutzutage weit verbreiteten Internet-Telefonie mittels *Skype*⁵ bedienen. Insofern haben die konkreten Arbeitsfortschritte auf dem EDV-Sektor unter der im wahrsten Wortsinn trans-kontinentalen Arbeitsorganisation nicht gelitten.

Die EDV-Erfassung der im Feld gesammelten Daten lag in den Händen der folgenden Personen, die diesbezüglich zum Teil bereits in den vorhergehenden Jahren tätig geworden waren:

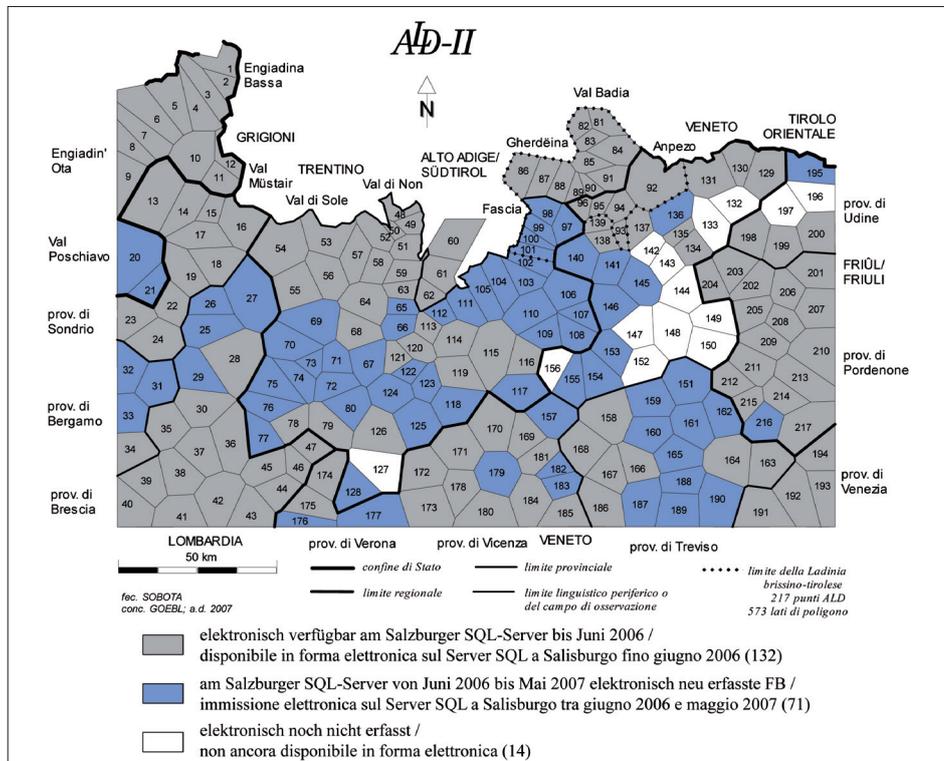
- I. Adami (Trient),
- Susanne Heißmann (Hyères, Frankreich),
- Liza Klinger (Salzburg),
- Gertraud Klingler (Salzburg),
- B. Rühlringer (Pisa),
- Zanita Sobota (Salzburg).

Wie aus den Angaben der verschiedenen Heimatorte ersichtlich ist, erfolgte die EDV-Erfassung der Fragebücher zum Teil dezentral, wobei die externen Mitarbeiter/innen den Ertrag ihrer Erfassungsarbeit über Internet an den Salzburger Server schickten. Traditionsgemäß besorgte daran anschließend auch im Jahr 2006 Pavel Smečka den Ausdruck der Korrekturausdrucke und die Eingabe der darauf von den Korrektoren festgestellten Eingabe-Irrtümer in den AD-II-Server.

Bei der Korrektur der EDV-erfassten Felddaten wurden einerseits die Dateneingebenden selber und zusätzlich über Werkvertrag in das AD-II-Projekt eingebun-

⁴ E. HAIMERL hatte bereits im Jahr 2006 einige Monate in Seattle (USA) verbracht und im Jahr 2007 seinen Wohnsitz mit seiner ganzen Familie in die USA verlegt.

⁵ Dabei kann neben dem Ton auch das Bild der Gesprächspartner (weitgehend kostenlos) übertragen werden.



Karte 1: Derzeit (1. Juni 2007) im Salzburger *ADM-II*-Server abgelegte Daten: 203 voll erfasste und korrigierte *ADM-II*-Fragebücher

Grau: Datenbestand bis Juni 2006 (man vergleiche dazu die Figur 1 in GOEBL/HAIMERL 2006, 206)
Eingefärbt: Neuzuwachs im Berichtszeitraum (71 Stück)

dene Kooperanten tätig. Die für den Postversand der betreffenden Korrekturausdrucke nötigen Agenden wurden vom Projektleiter erledigt. Insgesamt wurden auf diese Weise innerhalb des Berichtszeitraums (siehe Karte 1) die Daten von 71 *ADM-II*-Fragebüchern erfasst bzw. behandelt.

Gertraud Klingler sind darüber hinaus zahlreiche koordinierende Eingriffe zu verdanken, die für die Aufrechterhaltung der projektinneren Ordnung und Übersichtlichkeit sehr nützlich waren.

2.2 Feldarbeiten

Wie Karte 2 deutlich zeigt, hat die Menge der im Berichtszeitraum erledigten Feldenquäten (glücklicherweise) abgenommen: so dokumentiert diese Karte die Durchführung von 23 Feldenquäten, während die analoge Übersicht beim vorhergehenden (dritten) *ADM-II*-Arbeitsbericht auf 49 durchgeführte Feldexplorationen

2.3 EDV-Entwicklung und -Anwendung

Für den Großteil des Berichtszeitraums bestand unsere EDV-Mannschaft aus E. HAIMERL und seinen drei Kooperanten S. Sobota, X. Cassasas und A. WAGNER, wobei die Hauptarbeit der EDV-Entwicklung das Kartenproduktionsmodul von DGM betraf und somit bei A. WAGNER lag. Insofern bedeutete der Eintritt des philologisch und informatisch bestens ausgewiesenen F. Tosques eine sehr spürbare Erleichterung bzw. Bereicherung im Rahmen der ganzen EDV-Arbeit.

Daneben wurden – stets unter Anleitung von E. HAIMERL – die Arbeiten für die Sounddaten-Bank (SDB) vorangetrieben und im Mai 2007 mit der Schaffung einer Eingabe-Maske abgeschlossen, mit deren Hilfe es nunmehr möglich ist, fortan alle seit Beginn der *AD-II*-Explorationen im Feld erstellten Tonmitschnitte im Wege eines regulären Transfers der Tondaten von den Mini-Disks auf die Festplatte nicht nur zu digitalisieren, sondern auch in eine projektspezifisch erstellte Datenbank einzuspeisen, um sie von dort bei der Redaktion der *AD-II*-Kartenblätter bei Bedarf akustisch abrufen zu können.

2.4 Archivarbeit

Auch hier sind gegenüber dem dritten *AD-II*-Arbeitsbericht keine wesentlichen Änderungen zu vermerken. L. Klinger hat sich erneut um die Etikettierung, Einsortierung und Archivierung der im Feld gesammelten Diapositive gekümmert, während die bibliothekarische Erfassung der Neuzugänge für die *AD*-Bibliothek weiterhin in den Händen von P. Smečka lag.

In Erwartung des Beginns der routinemäßigen Produktion von *AD-II*-Karten im Herbst 2007 wurden die folgenden Vorarbeiten getätigt:

- Kompaktierung der in früheren Jahren erstellten Kartenausdrucke des *AD-I* auf insgesamt zwei der sieben am *AD*-Archiv vorhandenen Metallkartei-Schränke (im Hängeformat A 3), so dass für die erste Zeit der Kartenproduktion beim *AD-II* fünf leere Metallkartei-Schränke (samt der nötigen Anzahl an Hängeordnern im Format A 3) zur Verfügung stehen. Zur leichteren Bewältigung der damit verbundenen Transfer-Arbeiten wurde vom Projektleiter ein entsprechend dimensionierter Roll-Wagen gefertigt.
- Zur Einrichtung des für F. Tosques vorgesehenen Arbeitsplatzes mussten in den beiden *AD*-Räumlichkeiten Umadaptierungen vorgenommen werden.

Dies betraf sowohl die Möblierung der beiden Räumlichkeiten als auch den Ankauf neuen EDV-Geräts.⁷

- Zusätzlich musste auch der Aufstellungsort des für die Produktion der AD-II-Karten vorgesehenen Ink-Jet-Plotters (für die Formate A 2 und 1) in das neue Möblierungs-Konzept eingepasst werden. Dazu wurde dieses schon vor längerer Zeit angeschaffte (und gar nicht so kleine) Gerät aus dem Server-Raum des AD in den “EDV-Raum” des AD-Archivs transferiert und dort in ergonomisch günstiger Weise aufgestellt. Damit haben die EDV-Mitarbeiter des AD-II diesen Plotter unter ihrer direkten Kontrolle.

Schließlich wurden von jenen AD-Grundkarten,⁸ auf denen die Probeausdrucke der AD-II-Karten erfolgen sollen, jeweils rund 2.000 Exemplare bei einer Salzburger Druckerei bestellt und von dieser auch in der gewünschten guten Qualität geliefert. Zur Aufnahme einer größeren Anzahl dieser Grundkarten und zu deren komfortabler Bereithaltung in unmittelbarer Nähe des oben erwähnten großen Ink-Jet-Plotters wurde vom Projektleiter ein weiterer Roll-Wagen verfertigt, der seinen normalen Aufstellungsort unter jenem Tisch haben wird, worauf der Ink-Jet-Plotter steht.

2.5 Finanzierung

Die Finanzierung der AD-II-Arbeiten wurde so wie in den vorhergehenden Berichtszeiträumen aus den folgenden Quellen bestritten:

- aus Mitteln des FWF-Projektes 17.326,
- aus Förderungsmitteln des *Istitut ladin Micurà de Rü*, S. Martin de Tor (BZ),

⁷ So wurden für F. Tosques ein neuer, leistungsfähiger Rechner und für fast alle AD-Mitarbeiter neue (augenschonende) Flachbildschirme angeschafft. Dazu kamen – in Ergänzung eines bereits im Jahr 2006 angeschafften externen Speichers über 200 GByte – zwei extrem große externe Speicher (mit jeweils 750 GByte) für die Zwecke der Sounddaten-Bank und ein zusätzliches Abspielgerät der Firma Sony (Walkman Hi-MD MZ-RH1) zur EDV-Einspeisung von Mini-Disks alter und neuer Machart.

⁸ Es handelt sich dabei um zwei Sorten von Grundkarten: 1) mit dem auch beim Druck des AD-I verwendeten hellblauen Grundnetz, 2) ohne dieses Grundnetz, jedoch mit sieben in verschiedenen Farben gehaltenen Prüfpfaden. Bei der Produktion der AD-II-Karten erfolgt der erste Probe-Ausdruck immer auf einer Prüfpfad-Karte, deren Inhalt dann unter Zuhilfenahme entsprechender Listen genau auf Korrektheit und Stimmigkeit geprüft wird. Erst beim nachfolgenden zweiten Probe-Ausdruck kommt die Grundkarte mit dem blauen Kartengrund zur Anwendung.

- aus Förderungsmitteln des (früheren) Bundesministeriums für Bildung, Wissenschaft und Kunst,⁹ Wien,
- aus Förderungsmitteln des Istitut cultural ladin “Majon de Fascegn”, Vich (TN),
- aus Förderungsmitteln des Landes Tirol, Innsbruck,
- durch eine Förderung von Seiten der Universität Salzburg: Diese betraf nicht nur die fortdauernde Bereitstellung der Infrastruktur des *AD*-Archivs (Räume, Mobiliar, Beleuchtung, Heizung, EDV-Anschluß etc.) sondern auch die Durchführung der rechnerischen Verwaltung der vom FWF gewährten Personal- und Sach-Mittel sowie die Gewährung eines größeren Betrages für die Weiterbeschäftigung von S. Sobota.

2.6 Danksagungen

Diese richte ich als Projektleiter nicht nur an die im vorliegenden Abschnitt genannten Institutionen, sondern mit besonderer Herzlichkeit auch an alle inner- und außerhalb Salzburgs tätig gewordenen Mitarbeiter (semper utriusque sexus), unter denen ich ganz besonders jene erwähnen möchte, die an der Weiterbeförderung des Projektes *AD-II* den allergrößten Anteil haben: dies sind neben dem hauptamtlich tätigen Explorator W. Strauß die zu denselben Bedingungen am *AD-II* mitarbeitenden EDV-Mitstreiter X. Casassas, S. Sobota und (neuerdings) F. Tosques sowie E. HAIMERL und A. WAGNER, die mit dem *AD-II* zwar nicht über einen Dienstvertrag, aber doch “mit Herz und Seele” verbunden sind.

Den institutionellen Förderern seien abschließend zwei Dinge versichert:

- dass das Projekt *AD-II* gegenüber früher publizierten Zeitplänen aufgeholt¹⁰ hat und somit der für 2011 vorgesehene Projektabschluss an Wahrscheinlichkeit deutlich gewonnen hat,
- dass meinerseits alles daran gesetzt werden wird, dass dieser Termin “hält”.¹¹

⁹ Seit 1. März 2007 trägt dieses Ministerium einen neuen Namen: “Bundesministerium für Unterricht, Kunst und Kultur” (bm:ukk).

¹⁰ Siehe dazu die Figur 2 im 2. *AD-II*-Bericht (GOEBL/HAIMERL 2005, 113).

¹¹ Dies hat auch mit dem jetzt schon feststehenden Zeitpunkt meiner Emeritierung am 30.9.2012 zu tun.

3. Berichte aus der EDV-Werkstatt

Dieser Berichtsteil wird – anders als früher – nicht von E. HAIMERL allein, sondern auch von A. WAGNER bestritten werden, der im Berichtszeitraum nicht nur das Kartengenerierungsmodul von DGM zur vollen Anwendungsreife entwickelt, sondern auch zu diesem Entwicklungsprozess eine im Frühjahr des Jahres 2007 an der Naturwissenschaftlichen Fakultät der Universität Salzburg approbierte Diplomarbeit verfasst und damit den Titel eines Diplom-Ingenieurs erworben hat. Dazu sei ihm an dieser Stelle herzlich gratuliert.

3.1 Kartengenerierung mit den Programmen *MapGenerator* und *GeoTools* (Andreas WAGNER)

3.1.1 Was ist *MapGenerator*?

MapGenerator ist eine eigenständige Programm-Komponente, die als Verbindung zwischen den Programm-Komponenten *DMG Client* und *GIS¹² Frameworks Geotools* agiert. Der Fokus von *MapGenerator* liegt in der Erweiterung von in GIS-Systemen geläufigen Funktionen. Das Ziel ist, solche Projekte zu unterstützen, die Karten als eine alternative Form der Präsentation von Daten verwenden wollen. Dies soll möglichst wenig Aufwand erforderlich machen, jedoch unabhängig von der Art der Speicherung der projekteigenen Daten möglich sein. Weiters soll dem Anwender von *MapGenerator* (hier: dem Kartenautor beim AD-II) bei der Einzeichnung seiner Daten in Karten möglichst viel Bewegungsfreiheit geboten werden. In Abschnitt 3.1.3 werden einige von uns dafür entwickelte Konzepte kurz vorgestellt.

Im Laufe der Ausarbeitung der für die Kartengenerierung beim AD-II notwendigen Funktionen stellte sich heraus, dass viele der dabei entworfenen Konzepte sich in einen allgemein gültigen Teil und einen konkret auf die AD-II-Karten zugeschnittenen Teil einordnen lassen. Der allgemeingültige Teil wurde auf Klassenebene getrennt und kann somit auch für andere Projekte weiter verwendet werden: dies wäre demnach ein Prozess der Generalisierung.

Der wesentliche Vorteil für das AD-II-Projekt besteht neben der Wiederverwendbarkeit in der einfachen, iterativen Generierung der Atlaskarten. Anpassungen in

¹² Heute allgemein übliche Abkürzung für "Geographic Information System".

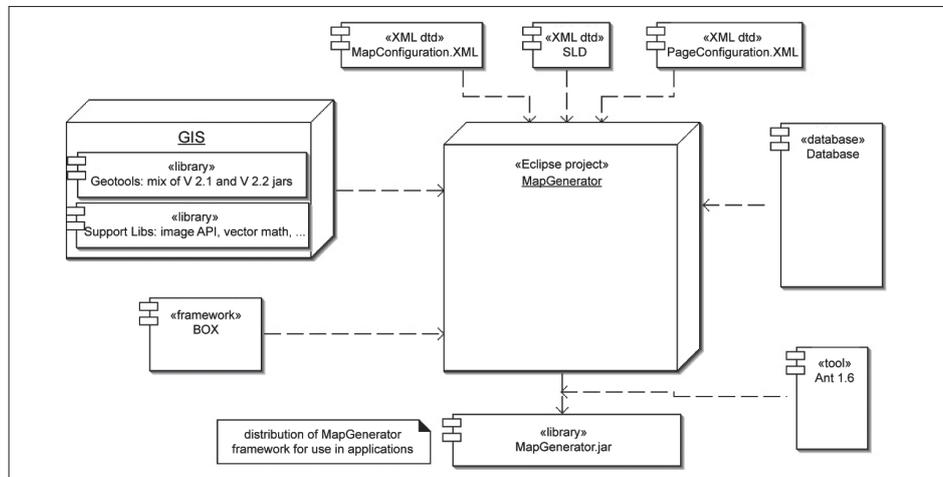


Fig. 1: Komponenten von *MapGenerator*

Dieses Diagramm zeigt die am Kartenerzeugungsprozess beteiligten Komponenten und deren Interaktion.

größerem Umfang mussten nur bei jenen spezifischen Algorithmen vorgenommen werden, die für die Erstellung der einzelnen Schichten der Karten (Erstellung der lautschriftlichen Karteneinträge (Transkripte) und der verschiedenen Legenden) zuständig sind. Dabei werden das Einlesen der benötigten Daten, die Erstellung und Verarbeitung der Definition jener Schichten, aus welchen eine Karte besteht, die komplette Verwaltung und auch das *Rendering* (d. h. die Erzeugung der Karte auf Bildschirm oder über den Drucker) von *MapGenerator* und teilweise von Geotools durchgeführt.

Die Figur 1 zeigt die wichtigsten Komponenten, die in *MapGenerator* und damit in den Prozess der Erzeugung von *AD-II*-Karten involviert sind:

- *GIS*:
Als GIS-Komponenten werden aktuelle *Geotools 2.1* und einige Funktionen aus *Geotools 2.2* verwendet. Viele der Funktionalitäten, die *MapGenerator* anbietet, werden von dieser GIS-Komponente ausgeführt (z.B. das *Rendering* der Karten).
- *MapConfiguration.XML*:
Dieses XML-File definiert die verschiedenen, beim *AD-II* zur Verwendung kommenden Karten, ferner ihren Aufbau aus diversen Schichten sowie die für die eigentliche Kartenerzeugung benötigten Daten und deren Quellen.

- *PageConfiguration.XML*:
Bei der Definition der Karten ist es auch möglich, spezielle Papierformate in der Form einer XML-Sprache zu definieren.
- *SLD (StyledLayerDescriptor)*:
Standardformat¹³ zur Festlegung von Darstellungsinformationen von Kartenschichten. Damit können den jeweiligen Schichten einer Karte die erforderlichen Darstellungsdetails zugeordnet werden (z.B. Linienstärken, Fonts etc.).
- Datenbank:
Wichtig ist die persistente Speicherung der dargestellten Daten. Dabei muss man zwischen der thematischen Datenbank und den geographisch relevanten Daten unterscheiden.
Die Struktur und die Zugriffsform der thematischen Datenbank sind für jene Applikation, die *MapGenerator* anwendet, spezifisch. Dagegen ist das Format der geographischen Daten vorgegeben. Von *MapGenerator* werden die folgenden geographischen Formate unterstützt: *MapInfo*, *ESRI-Shapefile*, *Post-GIS* sowie *GML (Geographic Markup Language)*.
- *MapGenerator*:
Ergänzt die Funktionalität von *Geotools*, wovon einige Funktionen im Abschnitt 3.1.3 näher beschrieben werden. Neben den Erweiterungen fungiert *MapGenerator* als Bindeglied zwischen den anderen Komponenten und versucht, für die anwendende Applikation den teilweise recht komplizierten Zugriff auf die *Geotools*-Komponenten zu vereinfachen.

Neben den beschriebenen Komponenten wird von E. HAIMERL noch das *BOX Framework* für *GUI*-Applikationen verwendet. Hier sind nützliche und allgemein verwendbare *Java*-Klassen zu finden.

Die Komponenten werden dann mit dem *ANT*-Werkzeug der Apache Group¹⁴ zu einem *Java*-Archiv (*jar*) zusammengefasst. Dieses *jar*-File wird in DMG verwendet, um damit die Karten zu erzeugen. So muss nicht mehr direkt auf die Files der *Geotools*-Bibliothek zugegriffen werden.

¹³ Die Standardisierung stammt von OGC (*Open Gis Consortium*).

¹⁴ Siehe dazu <<http://ant.apache.org/>>.

3.1.2 Erweiterung um Funktionalitäten, die in GIS-Systemen nicht vorgesehen sind bzw. die von *Geotools* nicht unterstützt werden

Die hier angeführten Erweiterungen stellen keine vollständige Auflistung dar; es werden nur wenige neue Funktionen herausgegriffen, die für die Verwendung im *AD-II*-Projekt als wesentlich angesehen werden. Für all jene neuen Funktionen, die *MapGenerator* über *GeoTools* hinaus bereitstellt, musste zudem gewährleistet werden, dass sie die bereits in *Geotools* vorhandenen Möglichkeiten optimal nutzen.

3.1.2.1 Zur Kombination von vorhandenen linguistischen Daten mit geographischen Formaten

Um genuine Dialekt-Daten in eine Karte einzeichnen zu können, muss man deren Verortung im geographischen Raum¹⁵ kennen. Diese erfolgt üblicherweise über die Lage jener geometrischen Objekte (wie Punkte, Linien oder Polygone), die in geographischen Datenformaten (z.B. in *MapInfo*) definiert sind. Wenn nun thematische Daten wie laut- oder normalschriftliche Einträge zu diesen geometrischen Objekten hinzugezeichnet werden sollen, dann müssen sie den letzteren zugeordnet werden.

Geographische Formate bieten zwar Regeln, wie thematische Daten zugeordnet werden, aber diese Regeln sind in den meisten Fällen nicht angemessen. Normalweise haben gewisse Applikationen jene Daten, die sie in Karten präsentieren wollen, bereits in Datenbanken gespeichert. Es wäre nunmehr ein enormer Aufwand, diese Daten nur für den Prozess der Kartenerzeugung zu duplizieren und dem Regelwerk der geographischen Formate anzupassen. Außerdem würden die betreffenden Daten unter dem Problem der Redundanz leiden. Änderungen bei den Daten müssten sich in der Datenbank und in der Repräsentation für die Kartenproduktion niederschlagen.

Eine der wesentlichen Funktionen von *MapGenerator* löst dieses knifflige Problem. Es ist möglich, eine Vielzahl bekannter geographischer Formate einzulesen. Für die aufrufende Applikation besteht kein Unterschied, welches Format verwendet wird. Nach dem Einlese-Prozess wird das *geographic linking*-System

¹⁵ Die geographische Position ist durch jene *x-/y*-Koordinaten definiert, die der Verortung des betreffenden Datums in der realen Welt entsprechen.

von *MapGenerator* verwendet, um die geographischen mit den thematischen Daten zu kombinieren, die wiederum in einer Datenbank vorliegen. Die wesentliche Arbeit wird dabei von *MapGenerator* erledigt. Die aufrufende Applikation muss nur spezifizieren, in welcher Form die Daten vorliegen. Nach Ausführung des *linking*-Prozesses sind die gewünschten Daten mit geographischen Positionen verknüpft und können in den für die Generierung der einzelnen Kartenschichten verantwortlichen Zeichenalgorithmen verwendet werden.

3.1.2.2 Definition der Schichten einer Karte in einer einfachen XML-basierten Sprache

Wie die Figur 2 zeigt, werden die einzelnen Karten aus mehreren Schichten zusammengesetzt. Dabei entsteht durch die Kombination mehrerer Schichten die komplette Karte. Die in Figur 2 gezeigte Karte besteht nur aus zwei Schichten. Daher wird eine Schicht mit linguistischen Transkripten über eine Schicht aus Streckenzügen gelegt, die einen Lesepfad darstellen.

MapGenerator sieht eine auf XML beruhende Sprache vor, die die einzelnen Schichten der Karte definiert. Pro Schicht werden die jeweils notwendigen geographischen Files sowie weitere benötigte Daten angegeben. *MapGenerator* kümmert sich auch um das Einlesen dieser Definitionen sowie um deren Umwandlung in weiter verarbeitbare und auf den Bildschirm oder über den Drucker reproduzierbare Objekte.

3.1.2.3 Kompletter Zugriff auf die Zeichenmethoden mit minimalem Aufwand (*Custom Layer*-Konzept)

Geotools bietet die Möglichkeit, eine Vielzahl verschiedener Arten von Schichten zu zeichnen. Wie die Daten einer Schicht gezeichnet werden sollen, kann in *SLD* definiert werden. Der in Figur 2 sichtbare Lesepfad kann mittels dieser Sprache unter Verwendung von *Geotools* realisiert werden. Es gibt aber Fälle, in denen die Möglichkeiten von *SLD* und auch von *Geotools* nicht mehr ausreichen. Dazu gehören beispielsweise das exakte Positionieren, das Erstellen genauer Umbruchregeln sowie die Darstellung von Turm-Zeichen in den AD-II-Transkripten.

Aus diesen Gründen wurde in *MapGenerator* das so genannte *custom layer*-Konzept (*CL-K*) entwickelt. Das *CL-K* bietet den Programmieren die Möglichkeit, in den Zeichenalgorithmen die kompletten Funktionen der sehr mächtigen *Java2D API* zu verwenden. Dazu ist es notwendig, nur eine einzelne Methode zu programmieren, die den

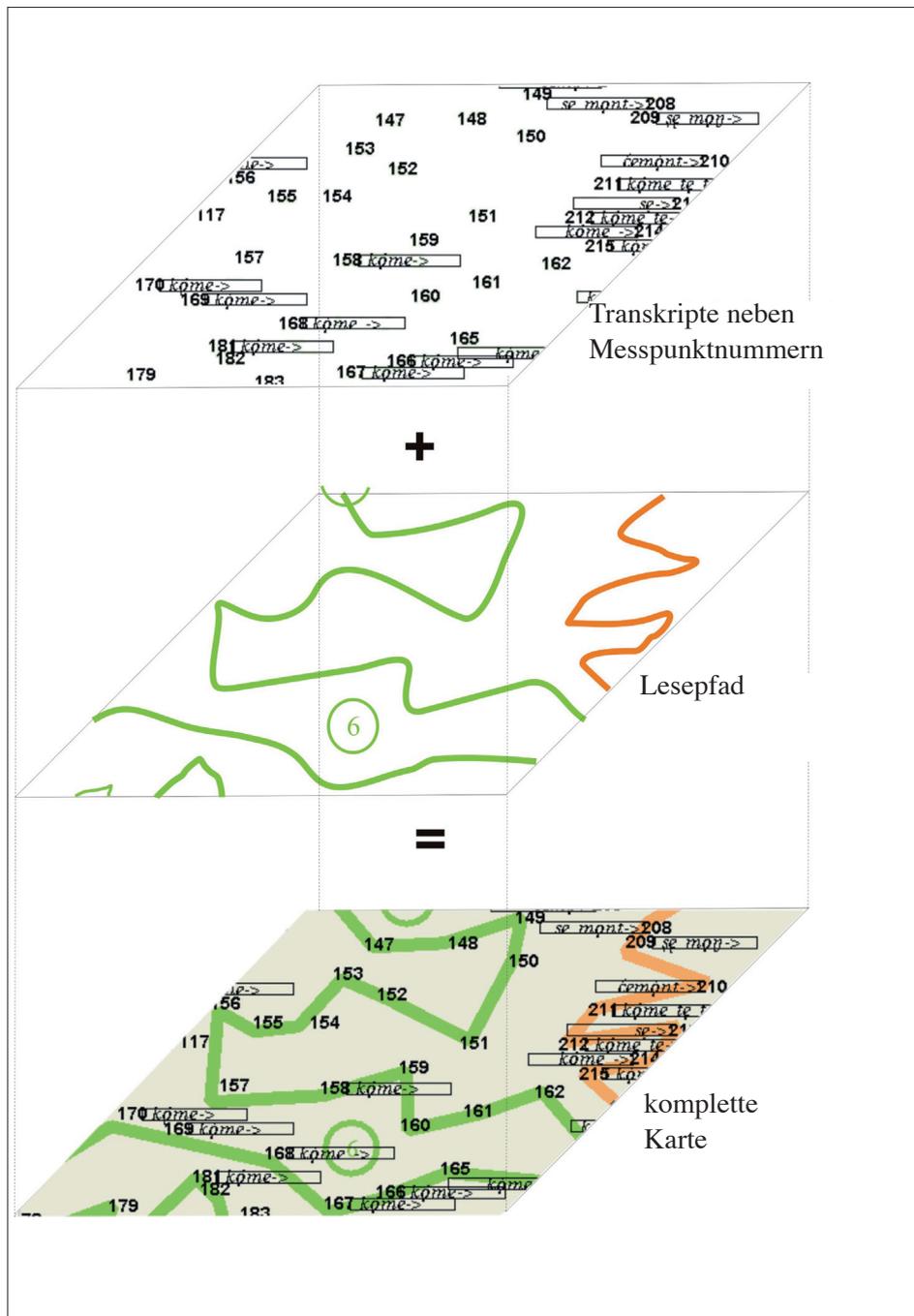


Fig. 2: Zusammensetzung einer *AD-II*-(Probe)Karte aus mehreren Schichten.
Diese Figur zeigt die Positionierung einiger Transkripte entlang eines Lese- oder Prüfpfads. Die zur Aufnahme der Transkripte dienenden Bounding-Boxes sind dabei nicht ausgeblendet worden.

tatsächlichen Zeichenalgorithmus beinhaltet, der dann aber an jedem Punkt ausgeführt wird. *MapGenerator* übernimmt dabei die folgenden Funktionen: die Umrechnung der geographischen Koordinaten auf äquivalente Positionen am verwendeten Ausgabegerät (Bildschirm oder Drucker), die Positionierung an diesen Punkten sowie das Auslösen des eigentlichen Zeichenprozesses. Weiters ist für jene Applikation, die *MapGenerator* verwendet, keine Unterscheidung zwischen den verschiedenen Schichten des *CL-K* und des *Geotools Frameworks* erforderlich.

Im Abschnitt 3.1.3.1 werden die einzelnen Schichten der *AD-II*-Karten vorgestellt. Außerdem besprechen wir, warum dieser Ansatz notwendig ist bzw. welche konkreten Vorteile er bringt.

3.1.3 Konkrete Anwendung von MapGenerator beim *AD-II*-Projekt

Nach der eher allgemein gehaltenen Beschreibung von *MapGenerator* beschäftigt sich dieser Abschnitt mit dem konkreten Prozess der Erzeugung von *AD-II*-Karten. Dabei werden jene Schichten, die in den *AD-II*-Karten enthalten sind, sowie der Fluss jener Informationen beschrieben, der die Verteilung der Daten auf die verschiedenen Schichten betrifft. Dann wird noch das automatisierte Layouting-System erwähnt, das vor allem auf die selbsttätige Erkennung von Kollisionen von Lautschriftsymbolen ausgerichtet ist. Dieses automatische Redaktions-System stellt einen wesentlichen Fortschritt gegenüber der manuellen Redaktionsarbeit dar, die noch beim *AD-I* notwendig war. Durch dieses neuartige System ist eine enorme Zeitersparnis bei der Produktion der Karten des *AD-II* zu erwarten.

3.1.3.1 Die einzelnen Schichten einer *AD-II*-Karte

Grundsätzlich lassen sich die Schichten einer *AD-II*-Karte in zwei Gruppen unterteilen:

- in *thematisch* oder *geographisch* relevante Kartenschichten:
Dieser Gruppe zeichnet sich dadurch aus, dass für jedes geographische Objekt derselbe Zeichenalgorithmus ausgeführt wird. Beim *AD-II* betrifft das die geographische Lage der 217 Messpunkte oder aber auch die einzelnen Lineamente eines Lesepfades.
- in Legendens:
Die Legendens werden einmal auf der Karte positioniert und stellen – so wie die Kommentare – eine zusätzliche Information dar.

Beim *AD-II* werden drei verschiedene, thematisch relevante Kartenschichten verwendet:

1. eine Schicht mit Polygonen, die den Status der Datenerfassung in den Regionen widerspiegeln.
Aufgrund der Statusfelder der *AD-II*-Datenbank werden jene Polygone entsprechend eingefärbt, deren Daten einen bestimmten Erfassungs-Status zeigen. So wird ein schneller optischer Überblick über den jeweiligen Zustand der Datenerfassung in den einzelnen Abschnitten des *AD*-Netzes gewährleistet.
2. eine Schicht mit Lesepfaden, entlang derer das Auge den Lesers geleitet wird. Mittels speziell eingefärbter (und nummerierter) Linienzüge wird das Auge des Lesers in geordneter Weise durch die Transkripte einer Atlaskarte geleitet.
3. eine Schicht mit der Darstellung der linguistischen Information zu den exakten Positionen der Messpunkte.

Mittels ausgefeilter Methoden, die *Java2D* bietet, werden die in der *AD-II*-Datenbank gespeicherten Transkriptionsdaten in speziell dafür entwickelte Fonts umgewandelt. Weitere Funktionen sind in diesem Zusammenhang: Erstellung und Ausrichtung von Turm-Transkripten sowie die Positionierung von hochgestellten Lautzeichen, Ausführung von Umbrüchen und Prüfung der Möglichkeit, ob die zu verschiedenen *AD*-Messpunkten gehörenden Transkripte überschneidungsfrei auf die Karte gezeichnet werden konnten. Dies ist ein sehr komplexer Prozess, der in den nächsten Abschnitten näher beschrieben werden soll.

Während den Anforderungen der beiden ersten Schichten durchaus allein mit *Geotools* Genüge getan werden kann, sprengen die Anforderungen der dritten Schicht den Rahmen der in *GIS* vorgesehenen Möglichkeiten. Dafür musste nun das schon erwähnte *custom layer*-Konzept (*CL-K*) herangezogen werden.

Die nachfolgenden drei Schichten betreffen drei verschiedene Legenden:

4. die Header-Legende:
Dies ist eine allgemein informierende Legende, die Grundinformationen der Karte (wie etwa die betreffende Frage des Fragebuchs) enthält.
5. die Kommentar-Legende:
Diese kann optional verwendet werden, um bestimmte Besonderheiten einer Karte hervorzuheben.
6. die Overflow-Legende:
Diese bietet zusätzlichen Platz für jene Transkripte, die aufgrund der Umbruchregeln nicht mehr in die primäre Schicht der Transkripte aufgenommen werden konnten.

3.1.3.2 Automatischer, schichtenübergreifender Datenfluss

Ein Blick auf die Figur 3 zeigt den bei der Produktion einer Karte ablaufenden Datenfluss. Nach der Ausführung des Schritts Nr. 1 – bestehend aus der Kombination von AD-II-Daten mit den betreffenden geographischen Positionen – beginnt der eigentliche Kartenerzeugungsprozess. Im Rahmen dieses Prozesses werden einige Regeln angewandt. Dabei kann es als Folge der Anwendung der Regeln 2, 3 und 5 notwendig werden, dass jene Daten, die ursprünglich für die Präsentation auf der Transkriptionsschicht der Karte vorgesehen waren, in eine andere Schicht bzw. sogar in die Supplement-Bände¹⁶ übertragen werden müssen.

Zunächst wird versucht, alle gewünschten Daten auf der Karte zu präsentieren. Wenn aber die Gesamtheit der lautschriftlich vorliegenden Daten aufgrund unserer Layouting-Regeln nicht in die für die Transkripte vorgesehene Schicht passen sollten, dann werden sie an den Stellen 4 und 5b in die nächste Kartenschicht, d. h. in die Overflow-Legende, transferiert.

Im Rahmen der Figur 3 werden nur jene Daten zwischen zwei verschiedenen Kartenschichten transferiert, die “zur Laufzeit bekannt” sind. Dieses Prädikat (“nur zur Laufzeit bekannt”) weist darauf hin, dass vor dem Beginn der Produktion der betreffenden AD-II-Karte nicht vorhergesehen werden kann, welche Daten in die Transkriptionsschicht passen werden und welche durch die Layouting-Regeln ausgeschieden bzw. einer Spezialbehandlung unterzogen werden müssen. Je nach Inhalt und Definition der betreffenden AD-II-Karte variieren daher die Menge und das lautschriftliche Aussehen jener Daten, die zu den einzelnen Messpunkten dazugezeichnet werden müssen. Diese große Variabilität betrifft nicht nur die Länge der Transkripte (definiert durch die Menge von Lautzeichen), sondern auch das Vorkommen bestimmter Lautzeichen oder die Kombinationen von diesen. Ganz besonders problematisch sind in diesem Zusammenhang die Turm-Notationen, bei denen die (sehr variable) Höhe und die optisch gefällige Ausrichtung der beiden Komponenten ganz besondere Herausforderungen an das automatische Layouting stellen.

Zur Bemeisterung all dieser Probleme war beim AD-I ein langwieriger Layout-Prozess nötig, bei dem die einzelnen Bestandteile der Transkripte händisch auf die Transkriptionsschicht und die Overflow-Legende verteilt (“balanciert”) wer-

¹⁶ Die Supplement-Bände des AD-II werden gegenüber dem AD-I ein Novum darstellen. Sie werden all jene lautschriftlichen Daten aufnehmen, die weder in den 217 Bounding-Boxes noch in der Overflow-Legende einer AD-II-Karte Platz finden.

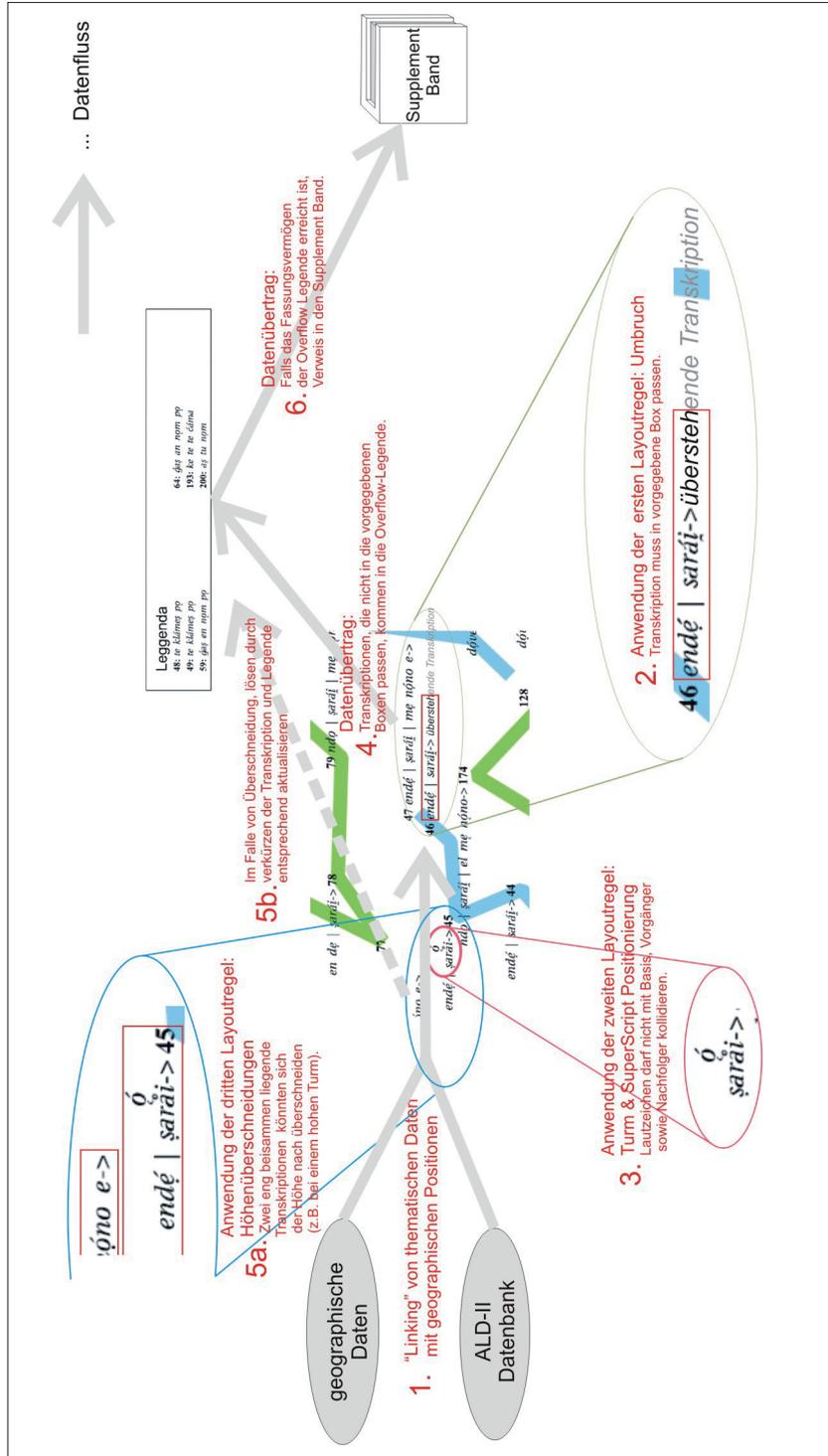


Fig. 3: Einzelstapen der automatischen Layouterstellung mit Hinweisen zum Fluss der Daten. Diese Figur zeigt den gestaffelten Einsatz mehrerer Layouting-Regeln bei der Erstellung einer *AD-II*-Karte. Die gezeigten Transkripte haben einen rein demonstrativen Charakter und sind daher fiktiv. Die Bounding-Boxes wurden zur besseren Verdeutlichung des Gesamtvorgangs nicht ausgeblendet.

den mussten. Obwohl der automatisierte Ansatz de facto auf weniger Regeln zurückgreift, als es bei einem exklusiv händischen Layouting der Fall wäre, sind die Balancierungen, die dieses System erzeugt, bereits sehr überzeugend und kommen dem als definitiv zu betrachtenden Layout bereits sehr nahe. Freilich sind abschließende Feinjustierungen von Hand nicht völlig auszuschließen. Doch können auf diese Weise viele Probe- und Zwischenausdrucke und die damit zusammenhängenden Korrekturarbeiten eingespart werden.

Die zuletzt erwähnte Regel Nr. 6 verweist auf den Fall, dass der gesamte verfügbare Platz der Overflow-Legende sich als nicht hinreichend erweist. Sofern dies zutrifft, müssen die nicht auf der Karte platzierbaren Transkriptions-Daten für einen der Supplement-Bände vorgesehen werden.

3.1.3.3 Die automatisierte Positionierung von Turm- und Hochstellungs-Transkripten

Die beim AD-II verwendeten Transkripte können auch Türme (“Turmnotationen”) und hochgestellte Lautsymbole beinhalten. Die Gestalt und Gesamtdimension dieser Zeichen können hinsichtlich ihrer Höhe und/oder des Vorhandensein von Akzenten (etc.) stark variieren. Wenn man bedenkt, dass nicht nur das Basiszeichen solcher Türme, sondern auch die vorhergehenden und die nachfolgenden Lautsymbole nach Größe und Kombination sehr verschieden ausgeprägt sein können, dann versteht man, dass sich daraus eine große Anzahl an graphischen Konfigurationen ergibt, die allerdings in für das Auge möglichst angenehmer Form dargestellt werden müssen.

Es gelang, auch die damit verbundenen Positionierungsarbeiten zu automatisieren. Bevor ein Turm oder ein Superskript¹⁷ tatsächlich zu Papier (oder Bildschirm) gebracht wird, prüft ein Algorithmus die relationale Position dieses Zeichens in seinem Umfeld. Dazu werden die Konturen des Basiszeichens sowie jene des Vorgänger- und Nachfolger-Zeichens auf Überschneidungen mit dem zu erstellenden Turm oder Superskript geprüft. Erst wenn die volle Überschneidungsfreiheit nach allen Seiten bzw. allen Richtungen hergestellt worden ist, wird das betreffende Zeichen tatsächlich platziert.

¹⁷ Unter “Superskript” ist ein hochgestelltes (Laut)Zeichen zu verstehen.

3.1.3.4 Automatisierung der Umbruchsregeln

Wie bereits in den letzten Abschnitten ausgeführt, werden die Transkripte auf Grund eines Regelwerks automatisch getrennt bzw. portioniert. Welche Regeln sind nun notwendig, um eine Trennung zu vollziehen und welche Randbedingungen sind bei der Spezifikation eines solchen Regelwerks zu bedenken?

Grundlegend lassen sich für die Umbruchsregeln zwei Ebenen identifizieren:

1. das Fassungsvermögen der Bounding-Boxen:
Jeder Karte des *AD-II* liegt dasselbe Ortsnetz zu Grunde. Darum ist es möglich, mittels genau definierter Bounding-Boxen eine nahezu optimale Platzverteilung zu definieren. Diese Regel ist auf Figur 3 unter Punkt 2 zu finden.
2. die Frage der Höhenkollisionen:
Das Eintreten von störenden Höhenüberschneidungen kann bereits im Layout der Bounding-Boxen berücksichtigt werden. Allerdings kann es je nach Konstellation der Daten zu Situationen kommen, bei denen sich vertikal (d.h. entlang der y-Achse) eng beisammen liegende Transkripte überschneiden können. Eine solche (vertikale) Kollision könnte beispielsweise durch das Auftreten eines sehr hohen Turm-Transkripts in der darunter liegenden Bounding-Box erzeugt werden. Würde man dieses Problem bereits beim Layout der Bounding-Boxen berücksichtigen, so würde man ab ovo viel Platz verlieren. Da aber dieses Problem wahrscheinlich nur selten auftreten wird, sollte es gesondert geprüft werden (siehe dazu Punkt 5a in Figur 3).

In Befolgung der beiden Umbruchsregeln muss also das in der Datenbank des *AD-II* gespeicherte Transkript verkürzt werden. Leider kann dies nur ganz selten an jener Stelle passieren, an der die fragliche Kollision eintreten würde. Der Umbruch muss also an einer Stelle vollzogen werden, an der durch eine Trennung der thematische Zusammenhang der Transkription möglichst wenig gestört wird. Daher muss ein Umbruch innerhalb eines Wortes oder zwischen ganz eng zusammengehörigen Wörtern ausgeschlossen werden. Dennoch sollte sichergestellt sein, dass möglichst viele Transkripte auf einer *AD-II*-Karte Platz finden.

Auf Figur 3 wird dieses Problem beim Punkt 2 kurz dargestellt. Die vorhandene Transkriptionsmenge würde die Bounding-Box nicht ausfüllen. Nur der in Grau gedruckte Text müsste in die Legende abwandern. Da aber innerhalb eines Wortes nicht getrennt werden darf, muss das ganze Wort wegbewegt werden, obwohl noch ein Großteil davon in der Bounding-Box Platz hätte.

3.1.4 Dokumentation von *MapGenerator*

Zum Projekt *MapGenerator* liegt eine geschlossene Dokumentation in Form einer Diplomarbeit vor, die am Fachbereich “Angewandte Informatik” der Naturwissenschaftlichen Fakultät der Universität Salzburg (WAGNER 2007) erstellt wurde. Diese Arbeit enthält die folgenden Hauptteile:

1. Beschreibung des beim Projekt *AD-II* angewandten Development-Prozesses für Software:
Damit der komplexe Prozess der Software-Entwicklung (Erfassung der notwendigen Funktionalität, Design der Programm-Architektur, Implementierung und Testen) in einer geordneten Form abläuft, werden im *AD-II*-Projekt Ansätze aus der *Software Engineering*-Methodologie, aus dem *Rational Unified Process (RUP)* sowie aus dem *Extreme Programming (XP)* verwendet. Ein weiterer entscheidender Faktor ist die effiziente Zusammenarbeit der Programmierer (*Collaborative Software Development*), die über einen relativ langen Zeitraum sogar in verschiedenen Kontinenten sesshaft waren.¹⁸ Diese konnte durch E-Mail und die Internet-Telefonie mittels *Skype* hergestellt werden.
2. Einführung in die verwendeten informatischen Techniken:
XML-Sprachen, Entwurfsmuster für Softwarekomponenten sowie für die automatisierte Erzeugung der Projekte aus Programmcodes spielen beim *MapGenerator*-Projekt eine wesentliche Rolle. All diese Techniken werden kurz vorgestellt.
3. Beschreibung der allgemeingültigen Komponenten von *MapGenerator*:
Unsere Diplomarbeit liefert ferner eine detaillierte Beschreibung der implementierten Funktionalitäten, wobei Begründungen dafür angegeben werden, warum diese Funktionalitäten in der beschriebenen Wege umgesetzt wurden und welche Alternativen denkbar wären.
4. Konkrete Anwendung und notwendige Erweiterungen von *MapGenerator*, um die *AD-II*-Karten zu erzeugen:
Man findet in diesem Abschnitt eine exakte Beschreibung, wie auf die *AD-II*-Datenbank sowie auf jene (Karten-)Schichten zugegriffen wird, die notwendig sind, um druckreife Versionen von *AD-II*-Karten zu produzieren.

Die Abschnitte 3 und 4 stellen eine detaillierte Beschreibung der Implementierung auf Klassenebene dar und liefern damit eine geschlossene Dokumentation

¹⁸ Hier soll nochmals daran erinnert werden, dass unser EDV-Supervisor E. HAIMERL im Jahr 2006 größtenteils in den USA gelebt hat.

des gesamten Projekts. Dies soll jenen Programmierern, die dereinst Änderungen und Anpassungen an der erarbeiteten Programmstruktur vornehmen müssen, das Verständnis des benützten Programmcodes wesentlich erleichtern.

Jedoch ist der Nutzen der in den einführenden Abschnitten 1 und 2 gebotenen, eher allgemein gehaltenen Informationen auch nicht unbeträchtlich. Darin erhalten jene Mitarbeiter, die neu in die Entwicklung von Software für das *AD-II*-Projekt einsteigen, eine zentrale Plattform, um sich grundlegendes Wissen über den *Software Development Process* beim *AD-II*-Projekt anzueignen. Die Dokumentation kann somit als eine Art allgemeiner Schulungsunterlage betrachtet werden. In erster Linie ermöglicht sie den raschen Erwerb eines Überblicks über die verwendeten Techniken und eröffnet darüber hinaus die Möglichkeit, durch die Nachverfolgung zahlreicher bibliographischer Hinweise all diese Techniken vertieft zu studieren.

3.2 Zur Sound-Datenbank¹⁹ (SDB) des *AD-II* (Edgar HAIMERL)

3.2.1 Die Anforderungen von Seiten der Benutzer

Die Sound-Datenbank des *AD-II* soll all jene Funktionen übernehmen, die beim *AD-I* durch C-90 Kassetten erfüllt wurden bzw. noch werden:

1. Während der Kartenredaktion müssen fallweise Abhörungen von Sound-Datenträgern zur Beseitigung von Zweifeln vorgenommen werden.
2. Die Mitschnitte der Feldenquëten stellen eine wertvolle Datenquelle des Gesamtprojektes *AD-II* dar, die in sehr willkommener Weise die gedruckten Daten ergänzt.

Beim *AD-II* repräsentieren – anders als dies noch beim *AD-I* der Fall war – die im Feld erstellten Tondatenträger in materieller Hinsicht nicht mehr mehrere Dutzend mit C 90-Kassetten prall gefüllte Kartons, sondern nehmen – in rein elektronischer Form – auf einer Festplatte ca. 100–120 GByte in Anspruch. Daher kann auch daran gedacht werden, über die SDB des *AD-II* Tonsequenzen aus den Aufnahmegesprächen via Internet direkt verfügbar zu machen.

¹⁹ Wenn hier geläufig der Anglizismus *Sound* anstelle des deutschen Worts *Ton* verwendet wird, so ist das eine Emanation des am *AD*-Archiv seit vielen Jahren etablierten, betriebsinternen Jargons. *Sit venia verbo!*

Beim *AD-I* wurden die im Feld aufgenommenen C-90 Kassetten mehrfach kopiert, um den Originalsound den ladinischen (und trentinischen) Kulturinstituten zur Verfügung zu stellen. Beim *AD-II* kann durch die Existenz einer netz-basierten SDB diese mühsame Kopierarbeit entfallen. Dies hindert aber nicht, dass bei Bedarf immer noch die Soundfiles aller Interviews relativ einfach von der Festplatte des *AD-II*-Servers auf eine oder mehrere DVD kopiert und an Interessenten verschickt werden können.

3.2.2 Das Arbeiten mit der SDB

Wie soll mit der *AD-II*-SDB bei der Erstellung und Redigierung der Karten gearbeitet werden? Der Kartenredaktor wird an einem großen Schreibtisch die Korrekturausdrucke der *AD-II*-Karten bearbeiten. Die bisherigen Korrekturen und Datenprüfungen erfolgten immer aus Sicht eines Fragebuchs, d.h. alle Antworten eines Interviews waren relativ leicht zu übersehen. Bei der Kartenredaktion ergibt sich jedoch eine neue Optik: erstmals erscheinen alle 217 landschaftlich variierenden Antworten auf ein und dieselbe Frage in geographischer Synopse. Durch dieses "Kippen" der Datenmatrix von der Dimension der Fragebücher in jene der Atlaskarten werden neue, bislang völlig unbekannte Zusammenhänge deutlich. Es können dabei naturgemäß Fragen auftreten, die durch einen punktgenauen, direkten Rückgriff auf die Tondokumentationen geklärt werden müssen. Dazu ist ein schnelles Auffinden der betreffenden Antwort quer durch 217 Ortschaften und rund 1.100 Teilfragen notwendig.

Beim *AD-I* wurde dieses Problem in der folgenden Weise gelöst: manuelles Heraussuchen der betreffenden C 90-Kassette aus einer größeren Menge wohl sortierter Archiv-Schachteln, Einlegen der betreffenden C 90-Kassette in einen Kassetten-Recorder, Betätigen des raschen Vor- oder Rücklaufs bis zur Auffindung der gesuchten Stelle. Es war das ein sehr zeitaufwändiges Verfahren. Zudem konnte der Zeitaufwand beim wiederholten Aufsuchen sogar derselben Passage nicht verkürzt werden. Für die Redaktion des *AD-II* soll nunmehr dieser Aufwand deutlich reduziert werden:

1. Statt auf C 90-Kassetten befinden sich die Aufnahmen aller Interviews in Form von Soundfiles auf einem Server und sind über eine Internet-Seite abrufbar (Gesamtvolumen: ca. 120 GByte; die Größe der Soundfiles variiert zwischen 1 und 15 MByte).
2. Unter Entfall des mechanischen Rück- und Vorspulens auf dem Kassettenrecorder können die einzelnen Soundfiles in einem *Playerapplet* auf der Inter-

net-Seite abgespielt werden. Der Player kann die gesuchte Stelle automatisch durch Interpolation aus vorher bereits gespeicherten Fragepositionen (*Marker*) im File annähernd positionieren. Diese Marker wurden von den Exploratoren während der Aufnahme in das Fragebuch eingetragen.

3. Bei der SDB-Benutzung abgehörte Antworten können im *Playerapplet* zusätzlich als ergänzende Marker gespeichert werden. So lernt das System bei der Benutzung und die Suche nach denselben (oder benachbarten) Fragepositionen wird durch dieses “Training” immer exakter und besser.

Nicht gefordert (und auf Grund des hohen Arbeitsaufwands auch nicht durchführbar) ist eine schnittgenaue Segmentierung der Antwortreflexe dergestalt, dass jedes Soundfile nur eine Antwort enthält. Diese Anforderung ist nur für einen “Sprechenden Sprachatlas” wesentlich. Zudem würde die Segmentierung aller Antworten für die Erstellung der *AD-II*-SDB eine erhebliche Zeitverzögerung darstellen und außerdem keine signifikante Unterstützung der Redaktionsarbeit darstellen. Deshalb steht hier nicht ein “Sprechender Sprachatlas”, sondern eine reine SDB zur Diskussion.

Bei dieser werden die Offsets²⁰ in den Soundfiles gespeichert und nicht Ausschnitte aus einem Soundfile in ein neues File kopiert. Im Laufe der Arbeit mit dem *Playerapplet* wird die SDB kontinuierlich vervollständigt, indem die Offsets von immer mehr Antworten in ihr gespeichert werden. Die technische Lösung dieser Anforderung sieht folgendermaßen aus: die während der Enquêtes aufgenommenen Soundfiles werden auf den Server kopiert, wobei eventuell Konvertierungen nötig sind. Die in den Fragebüchern bereits notierten Positionen bestimmter Fragen innerhalb der Soundfiles (*Tags*) werden in eine Datenbank übernommen. Damit ist eine grobe Positionierung der gesuchten Frage im betreffenden Soundfile möglich, die während der konkreten Arbeit mit dem *Playerapplet* noch weiter präzisiert werden kann.

3.2.3 Zu den Details des *Playerapplets*

Ein Applet ist eine Applikation, die in eine Internet-Seite integriert werden kann. Sie kann aber auch leicht als eigenständige Applikation gestartet werden. Durch diese Java-Technologie können Anwendungen unabhängig von verschiedenen

²⁰ Unter einem “Offset” versteht man den Beginn einer Tonsequenz.

Betriebssystemen und Internet-Browsern und ganz ohne eigenen Installationsaufwand betrieben werden.

Figur 4 zeigt das *Playerapplet*. In der ersten Zeile werden die Abfrage-Parameter *OrtNr* (Orts-Nummer) und *FrageNr* (Frage-Nummer) eingegeben. Durch die Betätigung des Buttons "Suchen" wird die Abfrage an den Server geschickt. Dieser lokalisiert dann das entsprechende Soundfile und

schickt ein File-Segment vordefinierter Länge an das *Applet* zurück. Die Länge dieses Segments in kByte kann in "Länge eines Segments" festgelegt und gespeichert werden. Das *Applet* arbeitet mit Segmenten von Soundfiles, da die Übertragung des gesamten Soundfiles erhebliche Wartezeiten erzeugen und den Server unnötig belasten würde. Mit Hilfe des Buttons ">" in der zweiten Reihe kann das Sound-Segment einmal oder mehrfach abgespielt werden. Die exakte Position innerhalb des Segments ist durch die Stellung des Schiebereglers in Zeile 3 ersichtlich und kann durch dessen Verschiebung abgeändert werden.

Wenn nur die im Fragebuch notierten Positionen bzw. Marker verfügbar oder erst wenige Fragepositionen gespeichert worden sind, kann bzw. wird es vorkommen, dass sich die gesuchte Antwort nicht in dem vom Server geschickten Segment befindet. In diesem Fall ist die auf Interpolation basierende Positionierung noch recht ungenau. Dann kann der Benutzer durch die Betätigung der Rück- und Vorsprungtasten in der Zeile 2 (<<, >>) eine beliebige Anzahl von Ton-Segmenten überspringen²¹ und damit vom Server ein anderes Sound-Segment anfordern.

In der Praxis empfiehlt es sich, die einmal erkannte Fragenummer eines abgehörten Sound-Segments in das Feld "aktuelle FrageNr" einzutragen und mit dem Button "Positionsangabe speichern" im Server zu speichern. Damit werden die Genauigkeit der Positionierungen bzw. die Zugriffspräzision der ganzen SDB für spätere Anfragen erhöht.

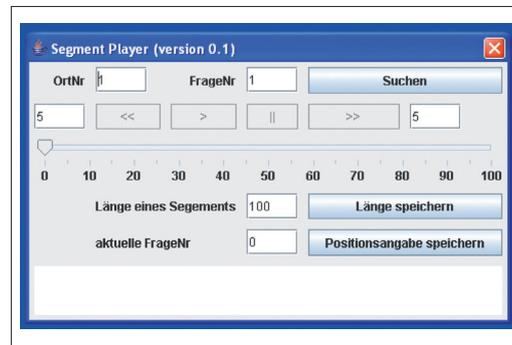


Fig. 4: Ansicht des *Playerapplets*.

Mit diesem Dialog kann das Abspielen von ausgewählten Tonsequenzen aus der *AD-II*-Sounddaten-Bank (SDB) interaktiv gesteuert werden.

²¹ Der *default*-Wert ist 5, im Textfeld ganz links oder rechts.

3.2.4 Überlegungen zur Performance der SDB

Die Leistungsfähigkeit (Performance) der SDB hängt wesentlich von der Länge der Sound-Segmente ab, da dank extrem optimierter relationaler Datenbanken das Aufsuchen der Daten heute keine Probleme mehr aufwirft. Zu große Sound-Segmente (1 MByte und größer) belasten den Server, die Server-Client-Verbindung sowie das *Applet* unnötig. Deshalb wird die Segmentierung der Soundfiles vom Server vorgenommen und nur das angeforderte Segment (siehe dazu das Eingabefeld "Länge des Segments") übertragen. Ein einmal übertragenes Segment ist im *Applet* gespeichert, so dass beim erneuten Abhören desselben Sound-Segments keine weiteren Anfragen mehr an den Server erfolgen.

In der Praxis hat sich gezeigt, dass die Übertragung von Soundfiles über *WebService* (*AXIS*) im Vergleich mit direktem *file download* nicht mit merkbarer Performance-Einbußen verbunden ist. Zudem ist die Implementierung recht einfach, denn das *AXIS framework* übernimmt die Umwandlung in ein übertragbares Format und die Rückwandlung im *Client*.

3.2.5 Architektur

3.2.5.1 SOA – Service Oriented Architecture

Das in Figur 5 gezeigte *deployment*-Diagramm gibt einen Überblick über die Architektur des gesamten Systems, dessen wesentlichste Komponenten hier kurz erwähnt seien:

- *Client*: Er besteht im Wesentlichen aus dem *Playerapplet* (siehe oben), der die Eingaben des Benutzers in Anfragen an den *WebService* umbaut und die Soundfiles vom *WebService* holt.
- *SDB_Server*: Dieser *WebService* erfüllt Anfragen wie z.B die folgende: "Gib mir das Segment der Länge X für FrageNr Y und OrtNr Z"
- *DMG_Service*: Für die Bereitstellung von Daten aus der DMG-Datenbank, z.B. von Ortsnamen und von Transkriptionen, wird über diesen *WebService* auf die *AD-II*-Datenbank zugegriffen.
- *Daten-Speicher*: Auf dem einem Rechner werden die Soundfiles gespeichert (*file storage*), auf einem anderen die *AD-II*-Datenbank. Durch die Verteilung auf verschiedene Rechner erreicht man eine Steigerung der Performance sowie eine Erhöhung der Sicherheit durch mehrere *firewalls*.

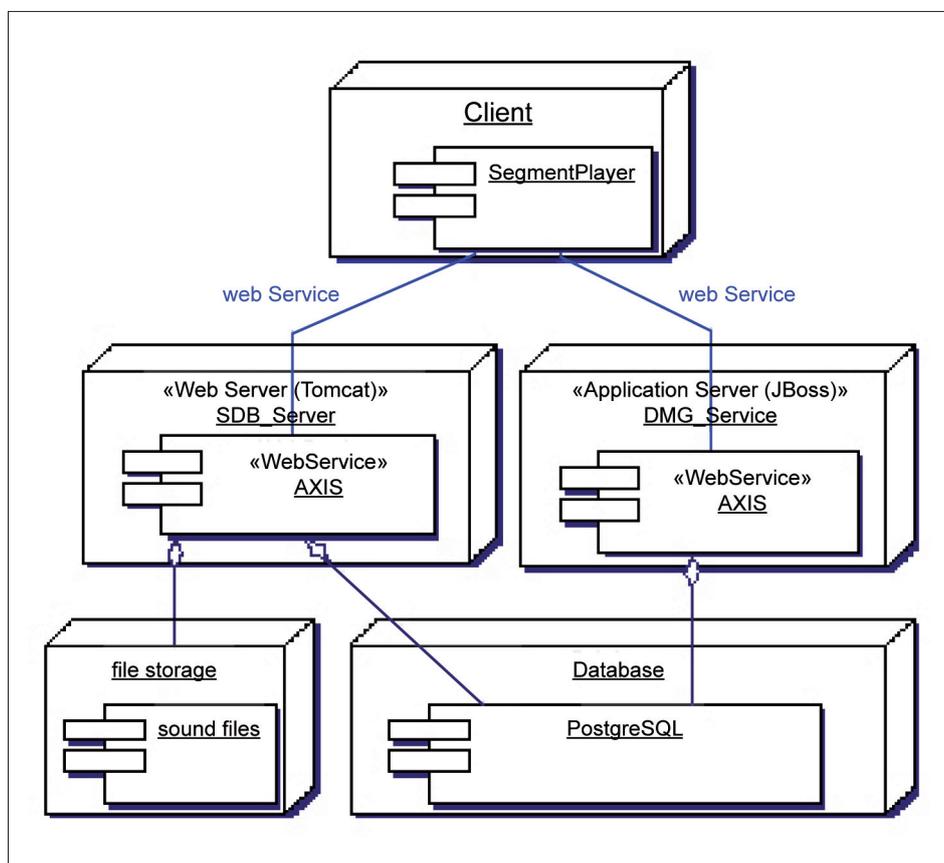


Fig. 5: *Deployment*-Diagramm des Servers der Sounddaten-Bank (SDB). Dieses Diagramm zeigt die Architektur der verschiedenen SDB-Komponenten und deren Interaktionen.

Der Vorteil der *SOA* besteht vor allem in der Entkoppelung der Komponenten. Während der *Server* ganz auf den schnellen Zugriff auf die Soundfiles ausgelegt ist, bleibt der *Client* schlank und hält immer nur die gerade nötigen Daten im Speicher. Zudem können Komponenten auf verschiedenen Servern genutzt werden: so können z.B. die Transkripte vom DMG-Server abgefragt werden, während die anderen Daten vom SDB-Server bezogen werden. Die verteilte Architektur erleichtert und reduziert darüber hinaus den Implementierungsaufwand, indem bestehende Komponenten in ein neues Gewand gehüllt und problemlos weiter verwendet werden können.

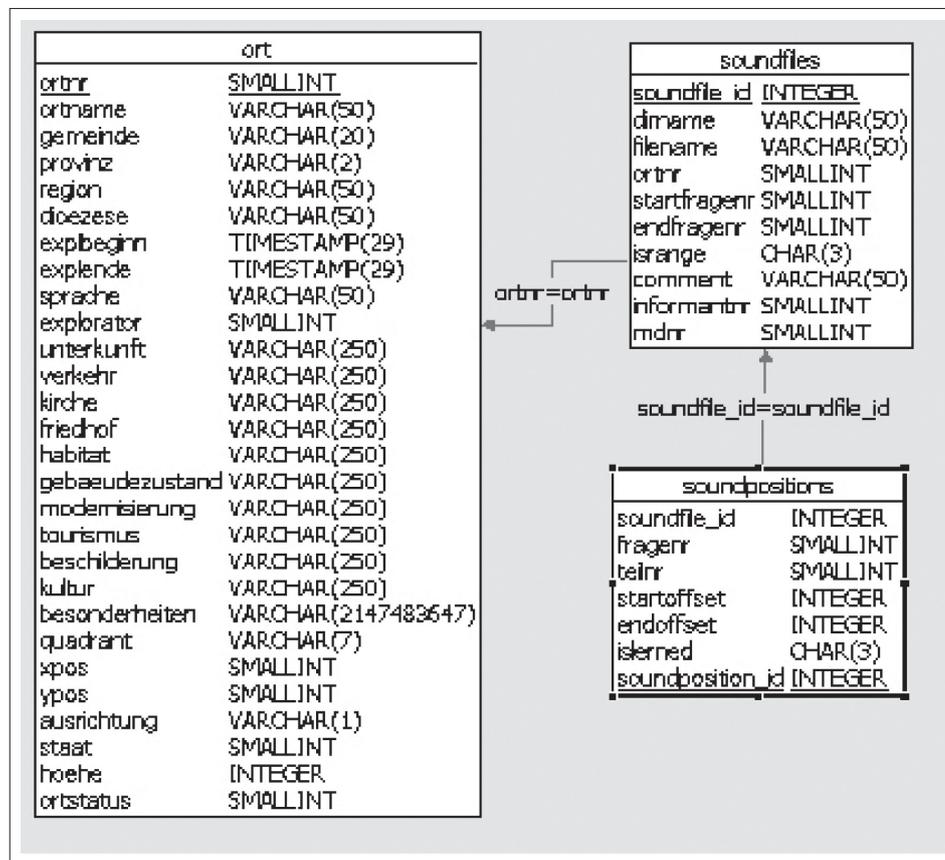


Fig. 6: Entity Relationship-Diagramm:

Dieses Diagramm verweist über einen Ausschnitt aus der *AD-II*-Datenbank auf jene Tabellen und deren Spalten, die in der Sounddaten-Bank (SDB) verwendet werden.

3.2.5.2 Hibernate für O-R mapping

Wesentlich für den Server ist der problemlose und schnelle Zugriff auf die Positionsdaten in der Datenbank. Deshalb wird zum Zweck des schnellen Zugriffs (auf relationale Datenbanken) auf etablierte Tools zurückgegriffen. So haben wir mit dem Open-source-Tool *Hibernate* (Version 3.2) sehr gute Erfahrungen gemacht. Als *O-R (object relational) mapping Tool* übernimmt es die Konvertierung von Datenbank-Werten in Java-Klassen und umgekehrt und kann durch wenige Konfigurationsparameter für jede relationale Datenbank konfiguriert werden. Wir testeten dies mit *MySQL*, *PostgreSQL* und *MS SQL Server 2000*.

Das ERD (*entity relationship diagram*: siehe dazu die Figur 6) zeigt jene Tabellen und deren Spalten, die für die Speicherung der Daten der SDB nötig sind:

- Die Tabelle *ort* ist die Tabelle der 217 Erhebungspunkte, die als Referenz für die Sounddaten verwendet wird.
- Die Tabelle *soundfiles* enthält die Information zu den Soundfiles, d.h. Informationen zu den einzelnen Enquêtes wie Dateiname, Verzeichnis und FrageNr-Bereich, wie er während der Aufnahmen im Fragebuch notiert wurde.
- In der Tabelle *soundpositions* werden die Positionen von im Verlauf der SDB-Benützung im File (= *offset*) aufgefundenen Antworten (*fragenr – teilnr*) gespeichert. Dies geschieht in den Spalten *startoffset* und *endoffset*.

4. Bibliographie

- GOEBL, Hans et al.: *AD-II: 1. Arbeitsbericht/1a relazione di lavoro (1999–2003)*, in: “Ladinia”, XXVIII, 2004, 115–199.
- GOEBL, Hans/HAIMERL, Edgar: *AD-II: 2. Arbeitsbericht (2004)*, in: “Ladinia”, XXIX, 2005, 107–124.
- GOEBL, Hans/HAIMERL, Edgar: *AD-II: 3. Arbeitsbericht (2005)*, in: “Ladinia”, XXX, 2006, 203–221.
- WAGNER, Andreas: *The Map Generator Framework. Map generation and interaction using Geotools open source GIS library and the Java2D API. Applied in the Dialect Map Generator (DMG) application*, Salzburg 2007, [Magisterarbeit zur Erlangung des akademischen Grades Diplom-Ingenieur an der Naturwissenschaftlichen Fakultät der Universität Salzburg].

Resumé

Chesta 4. relazion sun i progress de laour fac en cont de *AD-II* reverda i laours realisés entant l'ann 2006 con prospetives enchina l' meis de mei dl 2007. La relazion vegn partida su – en gros – te does pertes: te na pruma pert, che se referesc al numer y a les ativités di colaboradours, al finanziament dl projet, al azertament di dac dialetai sun l' post y a la elaborazion informatica de chisc, la seconda pert trata l' svilup de n software general de *AD-II* che à inom DMG (*Dialect MapGenerator*).

Tla pruma pert végnel danter l' auter conté che tl temp tout en conscidrazion per la relazion él gnù fat endut 23 enquêtes sun l' ciamp y che al vegn vedù dantfora de scluje ju dutes les esplorazions per l' auton dl ann 2007. Emplù végnel conté che al é gnù tout su 71 registrasions sun l' ciamp te na maniera digitela y che an pò fé i conc che per la fin dl auton dl 2007 sarà a desposizion tl server de Salzburg dutes les 217 registrasions *AD-II* te na forma che la mascinn é bona de lieje.

Tla seconda pert él i doi colaboradours *AD-II* A. WAGNER y E. HAIMERL che conta dla elaborazion de n program de layout dret efizient per la creazion automatica de chertes *AD-II* enjignedes per gnù stampedes (A. WAGNER) y dla programazion de na banca de dac nueva per la memorisazion y l'aministrazion de duc i tais acustics dles 217 *AD-II* esplorazions (E. HAIMERL).

La relazion vegn acompagneda da does chertes y da sis figures.