

Digitale Sprachatlanten am Beispiel von VIVALDI und ALD-I. Interoperabilität durch die “Geolinguistic Document Architecture (GDA)”

Marcel Lucas Müller

Dieser Beitrag ist Herrn O. Univ.-Prof. Dr. Hans Goebel zu seinem 65. Geburtstag gewidmet mit den besten Wünschen für Glück und Gesundheit.

1. Einleitung

Ein wichtiges Instrument der auf eine mehr als hundertjährige Tradition zurückblickenden Sprachgeographie ist der Sprachatlas, der die dialektale Realisierungen eines bestimmten hochsprachlichen Stimulus an genau definierten geographischen Punkten darstellt, indem die phonetische Transkription der dialektalen Ergebnisse den entsprechenden Koordinaten auf einer einheitlichen Grundkarte zugeordnet werden.

Die ersten Sprachatlasprojekte, die über die romanische Sprachgeographie hinaus methodologische Maßstäbe setzten, sind der ALF (*Atlas linguistique de la France*) aus den Jahren 1902 bis 1910 und der AIS (*Sprach- und Sachatlas Italiens und der Südschweiz*) aus den Jahren 1928 bis 1940, die in Form von gedruckten, zu mehreren großformatigen Bänden gebundenen Sprachkarten vorliegen. Den ersten Sprachatlas der englischen Sprache verfasste Hans Kurath für den Nordosten der USA (*Linguistic Atlas of New England*, KURATH 1939-1943).

Mit der Entwicklung der Schallplatte stand zwar ab dem Ende der zwanziger Jahre des 20. Jahrhunderts ein Medium zur Verfügung, das eine Konservierung der Original-Tondokumente und deren Verbreitung ermöglichte, aber die Aufnahmepraxis war weiterhin umständlich und kostspielig, so dass sie kaum in der Feldforschung eingesetzt wurde. Es gibt eine Reihe von Schallplatten, z.B. zu italienischen (CORTELAZZO 1974-88) oder zu schweizerdeutschen (HOTZENKÖCHERLE – BRUNNER 1972-76) Dialekten, die jedoch erhebliche Nachteile aufweisen und sich nicht in der Breite durchgesetzt haben – unter anderem deshalb, weil Karten- (Buch) und Tondokument (Schallplatte) nur mühsam zusammenzuführen waren. Wünschenswert für einen Sprachatlas ist hingegen die Integration von Karte, Tondokument und Transkription, die ein multimediales “Erleben” der visuellen (Karte, Transkription) in Verbindung mit den auditiven Elementen (Ton) ermöglicht.

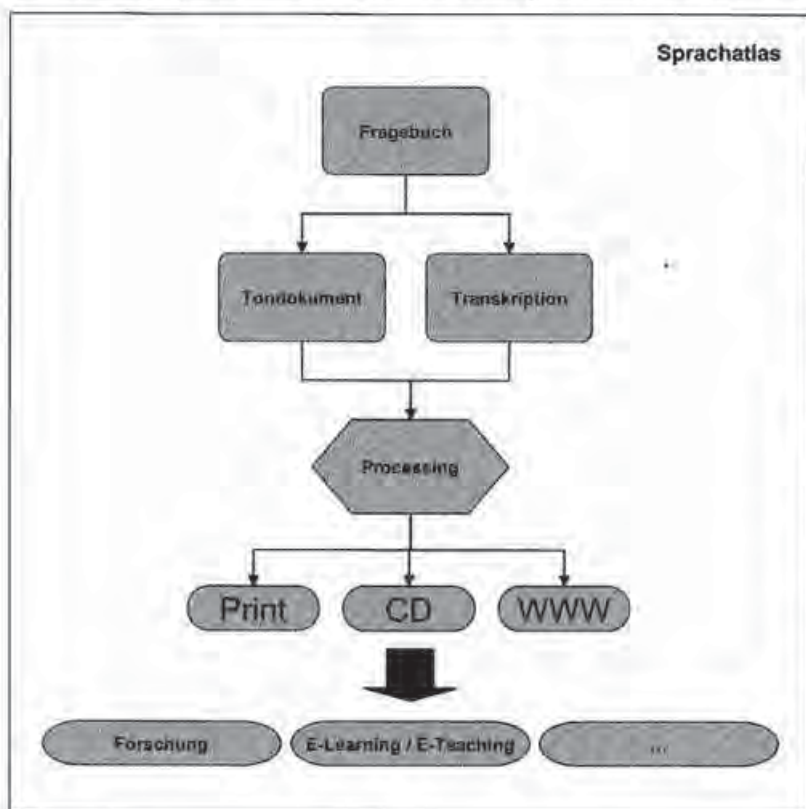
Mit der Einführung des Computers und modernen Verfahren der elektronischen Datenverarbeitung konnte diese multimediale Integration und

Präsentation erstmals realisiert werden. Der Jubilar Hans Goebel hat bereits zu Beginn der neunziger Jahre des vergangenen Jahrtausends diese neuen Optionen erkannt und im Sprachatlas-Projekt ALD-I (*Atlante linguistico del ladino dolomitico e dei dialetti limitrofi*) u.a. wegweisende Methoden der elektronischen Verarbeitung von Sprachatlasdaten und deren multimedialen Präsentation entwickelt.

Mit dem Aufkommen des "World Wide Web" war die elektronische Verbreitung von Informationen jeder Art möglich. Insbesondere die "Hypertext Markup Language" (HTML), eine textbasierte Auszeichnungssprache zur Strukturierung von Inhalten wie Texten, Bildern und Hyperlinks in elektronischen Dokumenten, erlaubte in Verbindung mit grafischen Webbrowsern erstmals, Bild, Text und Ton auf einfache Art und Weise zusammenzuführen – optimal also für die elektronische Verbreitung und ubiquitäre Verfügbarkeit elektronischer Sprachatlanten.

Das vom ALD abgeleitete Projekt "VIVALDI" (*Vivaio Acustico delle Lingue e dei Dialetti d'Italia*) entwickelt einen akustischen Sprachatlas Italiens, welcher die dialektalen Besonderheiten und Minderheitensprachen der 20 Regionen Italiens repräsentativ abbildet. Hierbei wurden die im ALD vorhandenen elektronischen Strukturen konsequent weiterentwickelt und neue Möglichkeiten einer internetbasierten, dynamischen Präsentation geschaffen (MÜLLER – KÖHLER – KATTENBUSCH 2001).

In der Zwischenzeit gibt es eine fast unüberschaubare Vielzahl von Webseiten, die elektronische Sprachatlanten anbieten. Die verschiedenen



Projekte unterscheiden sich nicht nur hinsichtlich der untersuchten Sprachen und geographischen Räume, sondern vor allem auch hinsichtlich der Methoden, wie diese Inhalte angeboten werden. Abbildung 1 zeigt ein stark vereinfachendes Schema eines typischen Sprachatlantes: Aus dem Fragebuch werden in der Exploration für jeden Stimulus Tondokumente erstellt, die – mit oder ohne Transkription – in einer Datenbank pro-

Abbildung 1: Schema eines elektronischen Sprachatlantes

zessiert werden und über ein Offline- (CD, Buch) oder Online-Medium (Internet) verbreitet werden.

Nach einer Beschreibung der beiden Internet-basierenden, "verwandten" Sprachatlanten VIVALDI und ALD-I soll im Folgenden die Problematik der Vielzahl, inhaltlich sich oft überschneidender, in der technischen Umsetzung aber sehr divergierender Sprachatlanten dargestellt, sowie ein Lösungsvorschlag einer standardisierten syntaktischen Basis für Sprachatlanten gegeben werden, welcher die inhaltliche Generierung von Sprachatlasdaten (Konzeption, Exploration, Transkription, Forschung, Lehre) entkoppelt von deren elektronischer Umsetzung.

2. VIVALDI und ALD – Beispiele für digitale Sprachatlanten

2.1 VIVALDI

"VIVALDI" ist das Akronym für "**V**ivaio **A**custico delle **L**ingue e dei **D**ialetti d'Italia", das italienische Wort *vivaio* wird hier verwendet im Sinne von "luogo, ambiente, scuola in cui ci si forma fisicamente o intellettualmente, acquisendo una particolare istruzione, preparazione e sim." (ZINGARELLI 2001). Ziel ist die Sammlung aktueller Dialektdaten in allen 20 Regionen Italiens und deren Bereitstellung unter Nutzung elektronischer Medien, sowie die Weiterentwicklung und Entwicklung neuer Präsentationsformen für den digitalen Sprachatlas. Da keine erschöpfende Erfassung aller Dialekte und Minderheitensprachen Italiens angestrebt wird, sollen pro Region nur etwa 10 bis 12 Orte exploriert werden. Das Fragebuch setzt sich aus verschiedenen Komponenten zusammen, den extra-, sozio- und metalinguistischen Gegebenheiten, dem phonetischen, lexikalischen, morphologischen und syntaktischen Teil sowie, als zusammenhängendem Text, der Übersetzung des "Gleichnis vom verlorenen Sohn". Die Entwicklung des Fragebuches und die Ortsauswahl orientieren sich eng am AIS, um die aktuellen Daten mit den Daten des AIS diachron vergleichen zu können (BAUER 1995; KATTENBUSCH 1995).

Die technologische Entwicklung, die von Carola Köhler und Marcel Müller umgesetzt wurde, erfolgte in verschiedenen Etappen:

- Im Jahre 1998 wurde eine elektronische Eingabehilfe erstellt ("VIVALDI-Transkriptionsassistentin"), die die Erstellung der gesammelten Tondateien und die Eingabe der Transkriptionen unterstützt und beschleunigt. Die Transkriptionsassistentin konnte von Beginn an die vorhandenen Daten in verschiedenen Formaten exportieren und an andere Applikationen zur weiteren Verarbeitung weitergeben.
- Zum Ende des Jahres 1998 erfolgte dann erstmals ein Export im HTML-Format – Ausgangspunkt für die Internet-Präsentation des VIVALDI-

Projektes (<http://www2.hu-berlin.de/Vivaldi>), die ebenfalls im Jahre 1999 frei geschaltet wurde.

- Mit der Fortentwicklung des Internets und dem Aufkommen von komplexen Programmiersprachen zur Erstellung dynamischer Inhalte (z.B. PHP hypertext preprocessor, <http://www.php.net>), Datenbanken (MySQL, <http://www.mysql.com>) und clientseitigen aktiven Prozessen (Java, <http://java.sun.com>) konnte der statische Export von HTML-Dokumenten im Jahre 2005 durch eine datenbankgestützte, dynamische Applikation ersetzt werden, in Übernahme einer für den ALD-I (s.u.) entwickelten Technologie. Insbesondere hat auch die Entwicklung psychoakustischer Kompressionsverfahren (mp3) dazu geführt, dass Tondateien ohne relevanten Qualitätsverlust über das Internet transportiert werden können.
- In der Folge des so genannten “Web 2.0” wurden eine neue Schlüsseltechnologie der Benutzerinteraktion für Webseiten entwickelt: Das Konzept des “Asynchronous JavaScript and XML” (AJAX) bezeichnet eine asynchrone Datenübertragung zwischen einem Webserver und dem Webbrowser, die es ermöglicht, innerhalb einer HTML-Seite eine HTTP-Anfrage durchzuführen, die nur relevante Teile einer Webseite sukzessiv bei Bedarf nachlädt. Zahlreiche Firmen bieten seither meist kostenfrei Komponenten an, welche in eigene Webseiten mit dieser Technologie integriert werden können. Die Firma Google (<http://maps.google.de>) stellt seit dem Jahre 2006 einen Dienst zur Verfügung, der es ermöglicht, Objekte auf einer Karte oder auf einem Satellitenbild der Erdoberfläche darzustellen. Gleichzeitig wurde eine Programmierschnittstelle angeboten, die es ermöglicht, die Kartendarstellung durch eigene Inhalte im Kontext der Karte selber darzustellen – eine reizvolle Möglichkeit, den digitalen Sprachatlas auf eine reale Karten- und Bilddarstellung zu projizieren. Dies wurde für die VIVALDI-Daten im Jahre 2007 realisiert.

Im April 2008 umfasst die Online-Präsentation von VIVALDI sieben vollständig erfasste Regionen (Sizilien, Sardinien, Ligurien, Aostatal, Umbrien, Trentino-Südtirol, Molise) mit zusammen 110 Orten. Das Fragebuch enthält 359 Fragen, insgesamt sind knapp 40.000 Tondokumente mit 20.000 Transkriptionen verfügbar. Die zugrunde liegenden Daten sind in eine dynamische Weboberfläche eingebunden, die sich als zweisprachiges (deutsch und italienisches) Content Management System mit zahlreichen Zusatzinformationen präsentiert.

Mit Hilfe eines in die Webseite eingebundenen Java-Applets (welches zuletzt von Fabio Tosques überarbeitet und optimiert wurde) werden die Karten-, Ton- und Transkriptionsdaten angeboten. Überfährt man einen Ort mit der Maus, wird die Transkription angezeigt und die entsprechende Tondatei abgespielt (Abbildung 3).



Abbildung 2: VIVALDI-Startseite

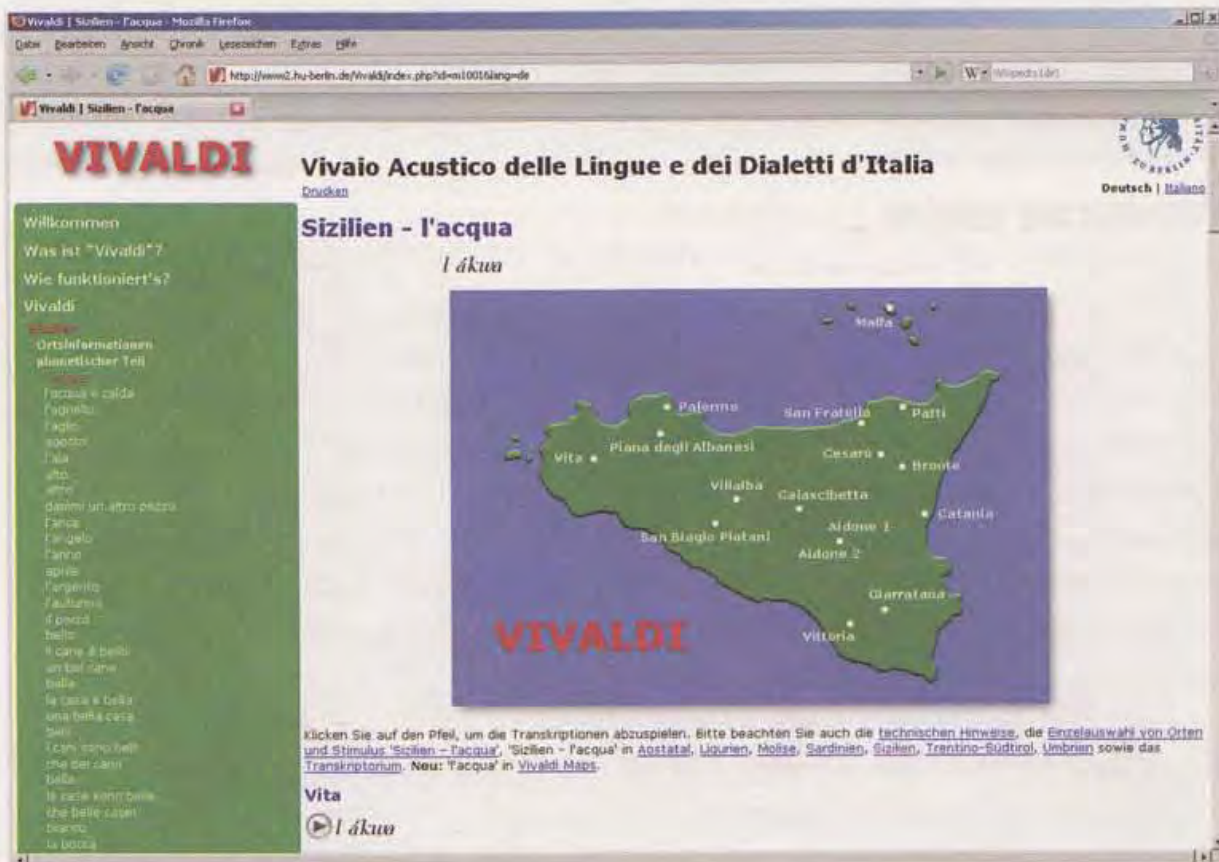


Abbildung 3: VIVALDI – interaktive Kartendarstellung

Zahlreiche Querverweise (z.B. zu anderen Regionen, Metainformationen über Informant, Region und Ort, Informationen auf anderen Webseiten, siehe Abbildung 4) machen die Karte zum Teil eines dichten Netzes an Informationen.

VIVALDI Vivaio Acustico delle Lingue e dei Dialetti d'Italia

Sizilien

Aidone 1

Ort	Provinz	Enna (EN)
	Region	Sizilien
	Sprachatlas/Punkt	AIS: 865, ALI: 1050
	Dialekt/Sprache	Galloitalienisch (Vertreter des alten galloitalienischen Dialekts)
	weitere Informationen	Wikipedia 'Aidone' , Google Maps 'Aidone'
Informant(in)	Geschlecht	m.
	Geburtsjahr	1920
	Beruf	Bergarbeiter
Aufnahme	Datum	12/1993
	Enqueteur(in)	Dieter Kattenbusch

Aidone 2

Ort	Provinz	Enna (EN)
	Region	Sizilien
	Sprachatlas/Punkt	AIS: 865, ALI: 1050
	Dialekt/Sprache	Galloitalienisch (Vertreter des jüngeren sizilianisierten Dialekts)
	weitere Informationen	Wikipedia 'Aidone' , Google Maps 'Aidone'
Informant(in)	Geschlecht	m.
	Geburtsjahr	1954
	Beruf	Lehrer
Aufnahme	Datum	12/1993
	Enqueteur(in)	Dieter Kattenbusch

Bronte

Ort	Provinz	Catania (CT)
	Region	Sizilien
	Sprachatlas/Punkt	AIS: 838, ALI: 1027
	Dialekt/Sprache	Sizilianisch

Abbildung 4: VIVALDI – Metainformationen und Links zu externen Webseiten

VIVALDI MAPS beta

Rhêmes-Saint-Georges (Aostatal)

Map Description: A map of Italy with numerous red location pins indicating dialect collection sites. The map includes a search bar with 'Fagnano' and a language dropdown set to 'Deutsch'. A pop-up window for 'Rhêmes-Saint-Georges (Aostatal)' is visible, showing a play button and the text 'Vivaldi Aostatal'. The map interface includes standard navigation controls like zoom in/out, pan, and map style selection (Karte, Satellit, Hybrid, Gelände).

Abbildung 5: VIVALDI Maps

Mit der Einbindung der Daten in Google Maps (Abbildung 5) können Transkriptionen und Tondokumente in einem optimalen geographischen, regionsübergreifenden Zusammenhang dargestellt werden. Durch Zoomen werden feinste oberflächengeographische Details angezeigt (Abbildung 6).

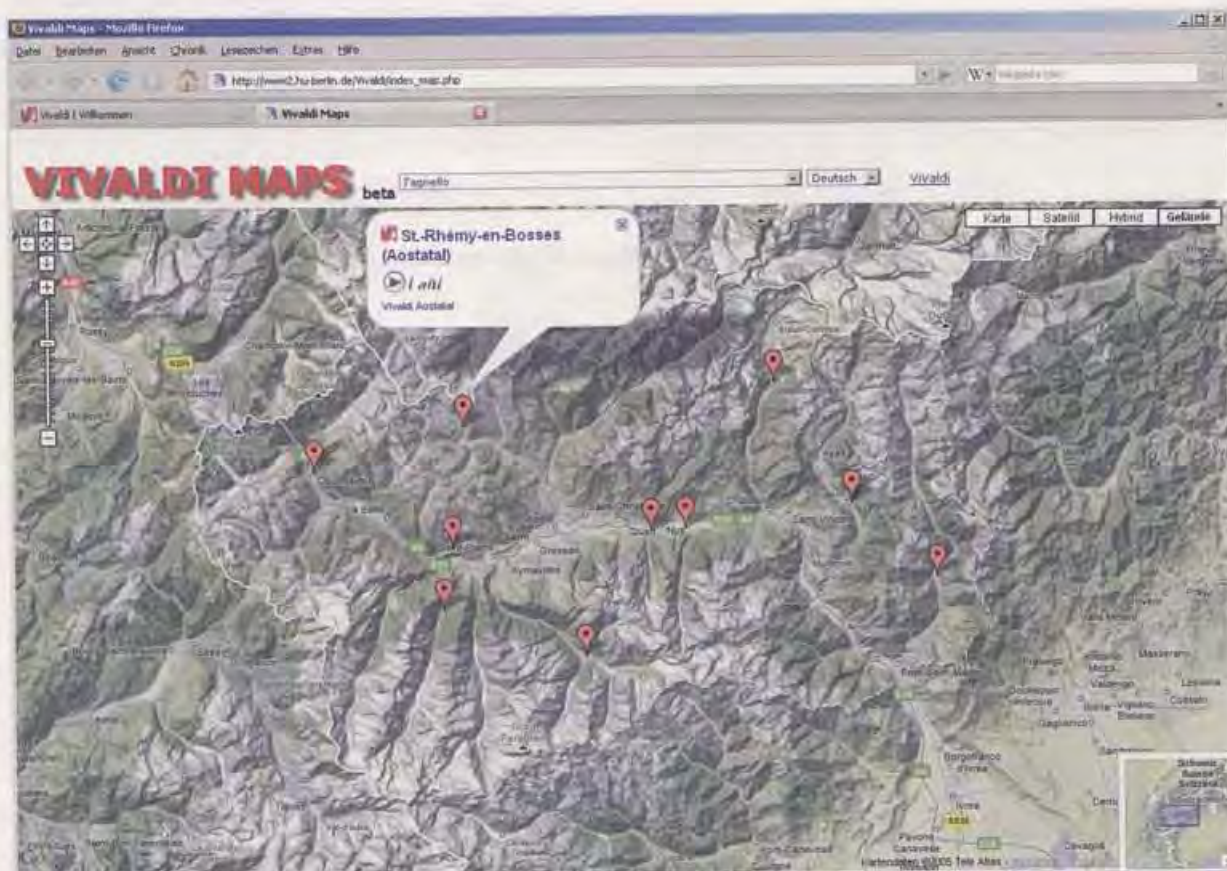


Abbildung 6: VIVALDI Maps, Geländedarstellung und Transkriptionen der Valle d'Aosta

2.2 ALD-I

Der ALD-I stellt den ersten Teil des Gesamtprojektes ALD (*Atlant linguistisch dl ladin dolomitich y di dialec vejins*) dar. Der Publikationsertrag des ALD-I umfaßt den gedruckten ALD-I und den "Sprechenden" des ALD-I. Der gedruckte ALD-I ist 1998 beim Dr. L. Reichert Verlag (Wiesbaden) in sieben Bänden (vier Karten-Bände, drei Index-Bände) erschienen. Er umfasst 217 Messpunkte (in Nordost-Oberitalien und in der Südost-Schweiz) und 884 Atlaskarten (zu vorwiegend phonetischen und morphologischen Themen). Dagegen umfasst der "Sprechende (ALD-I)" nur die 21 Messpunkte der brixnerisch-tirolischen Ladinia (ALD-Meßpunkte 81-101) und dieselbe Anzahl an Karten wie der gedruckte ALD-I. Die Ausarbeitung des ALD-I hat alles in allem (Datensammlung im Feld, EDV-gestützte Datenweiterverarbeitung, Vorbereitung der ebenso EDV-gestützten Publikation, Ausarbeitung aller EDV-Programme etc.) von 1985 bis 1998 gedauert.

Im Jahre 2005 setzte Hans Goebel mit gewohnter Verve die Aufbereitung der in ALD-I vorhandenen Daten für einen sprechenden Sprachatlas im Internet in Gang. Aktuell präsentieren sich die Daten in einem fünfsprachigen (deutsch, ladinisch, italienisch, französisch, englisch) Content Management System, begleitet von einer Menge an Metainformationen zum Projekt.

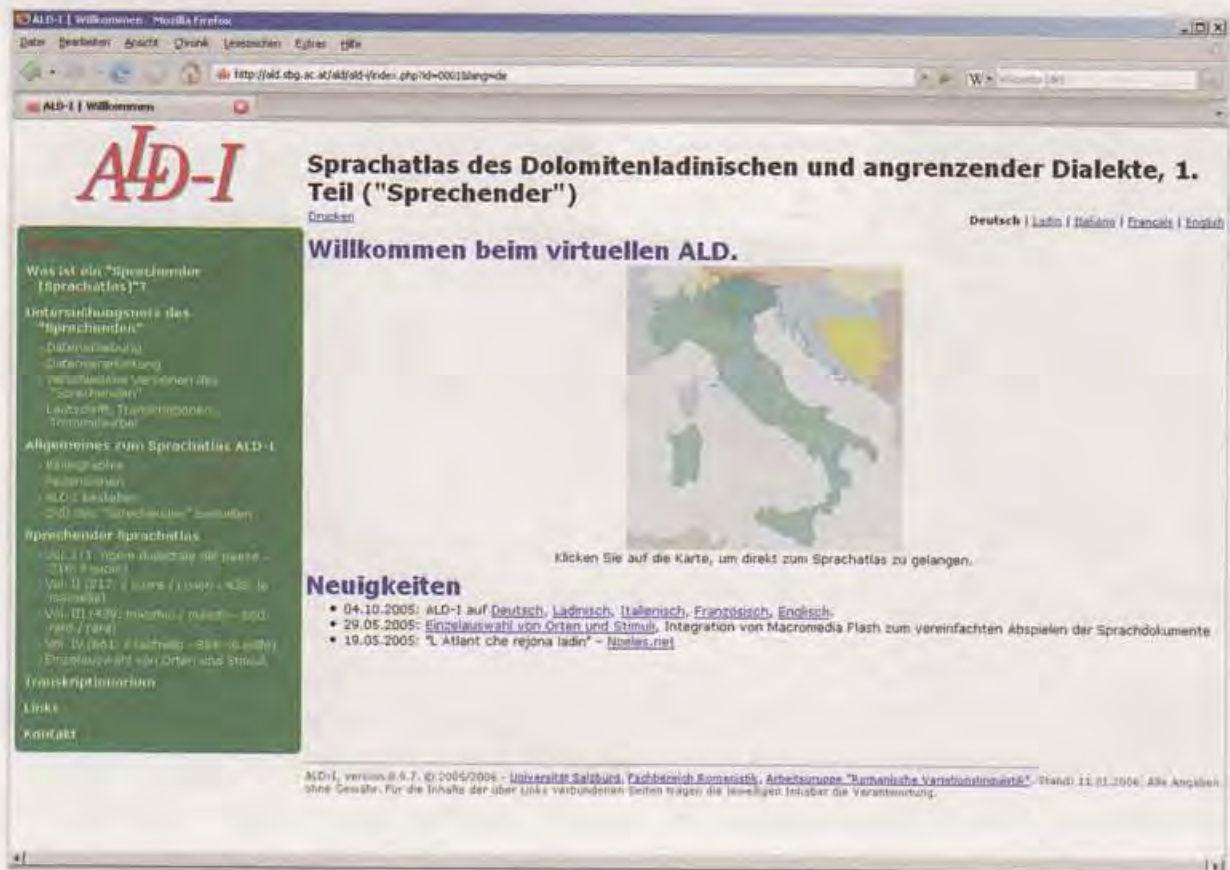


Abbildung 7: ALD-I Startseite

Vor der eigentlichen Programmierung der Web-Applikation mussten mit maschineller Unterstützung die vorhandenen Ton- und Transkriptionsdokumente auf die oft mehrere Einzelstimuli kombinierenden endgültigen Stimuli gemappt werden.

Die vier Bände des ALD-I werden seit 2005 im Internet für die brixnerisch-tirolischen Ladinia in einer entsprechenden Kartendarstellung angeboten (Abbildung 7).

Mit der "Einzeldarstellung" können beliebige Ort- und Stimuluskombinationen angezeigt werden, ein Klick auf den Pfeil spielt die entsprechende Tondatei ab (Abbildung 8), auch hier führen Querverweise komfortabel in die entsprechende Kartendarstellung.

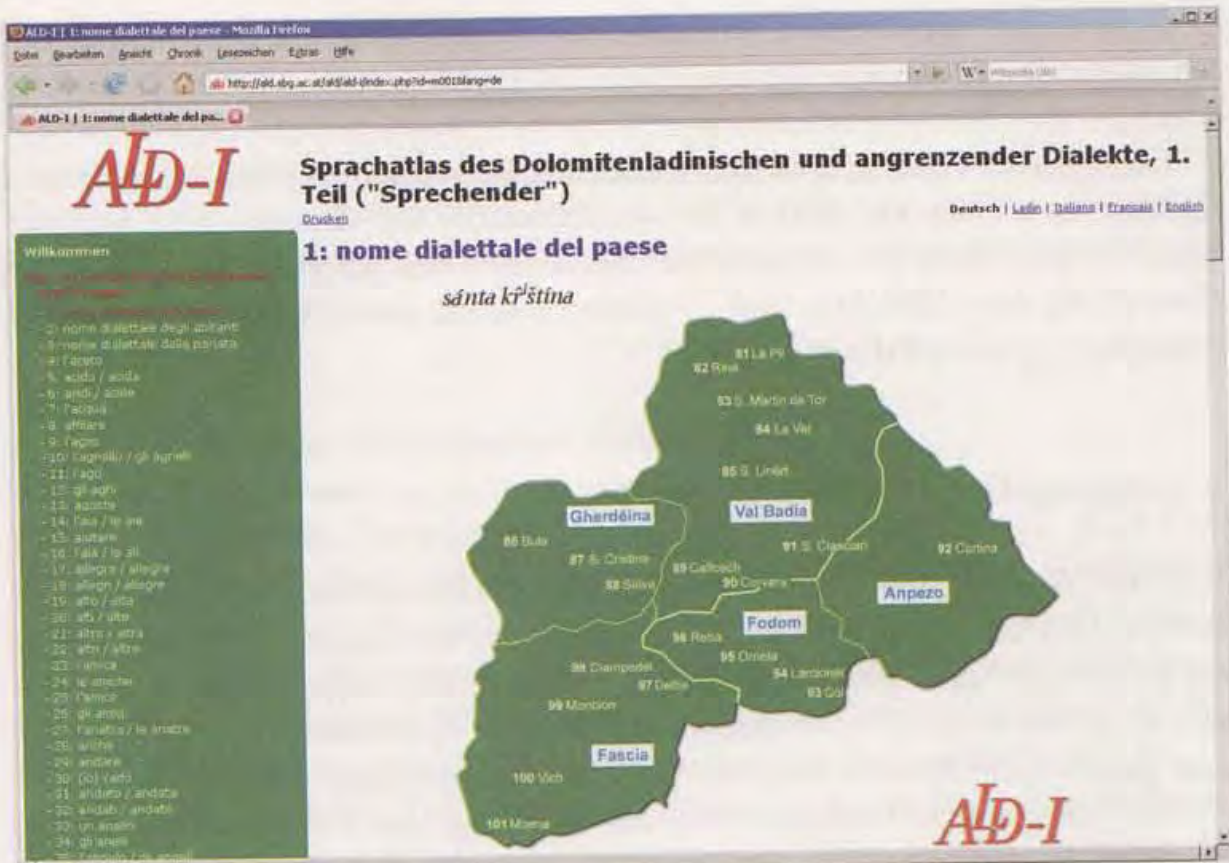


Abbildung 7: ALD-I "Sprechender" – Kartendarstellung

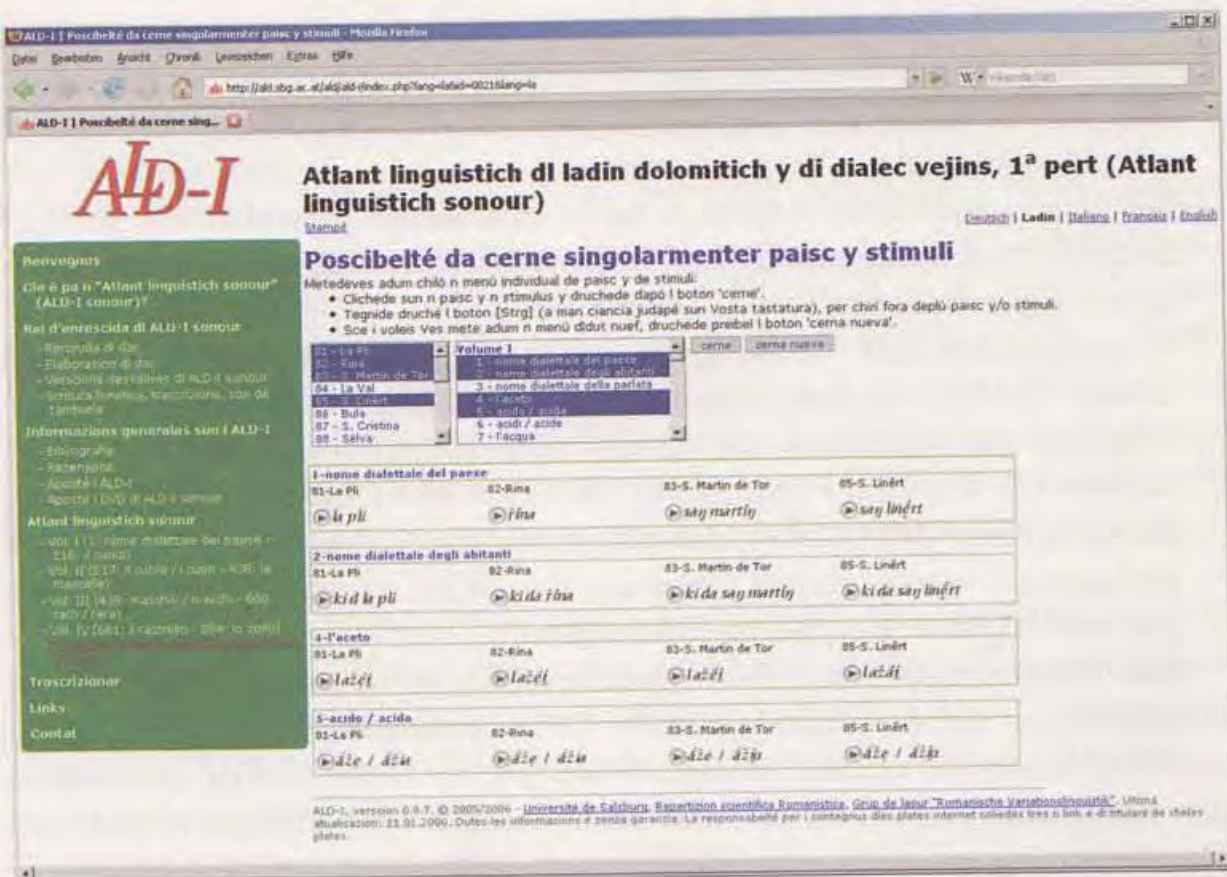


Abbildung 8: ALD-I "Sprechender" – Einzelauswahl von Orten und Stimuli

Die für den "Sprechenden" entwickelten Prinzipien und Mechanismen konnten im Anschluss teilweise für das VIVALDI-Projekt zurück-übernommen werden.

Die Internet-Präsenzen beider Sprachatlanten haben sich gut bewährt und erfreuen sich eines konstanten Besucherzustroms aus dem In- und Ausland. Auch die Bewohner der explorierten Orte freuen sich über eine entsprechende Darstellung ihres Dialektes und Umfeldes, was ein gutes Argument ist für die Rekrutierung neuer Informanten.

3. Interoperabilität digitaler Sprachatlanten

Ein guter Ausgangspunkt für die Recherche nach Sprachatlas-Projekten sind die entsprechenden Wikipedia-Einträge (<<http://www.wikipedia.org>>). Die Suche nach "Sprachatlas", "Atlas linguistique", "atlante linguistico", "linguistic map" etc. zeigt eine fast unüberschaubare Anzahl von Sprachatlas-Projekten. Kaum eine Region oder Sprache wird nicht irgendwo dargestellt.

Hier lassen sich bezüglich der Aufbereitung der Präsentationen einige Auffälligkeiten zusammentragen:

- Die jeweils dargestellten Inhalte unterscheiden sich naturgemäß stark in Abhängigkeit von den geographischen und linguistischen Foci.
- Genauso divergent ist aber auch die Präsentation der Daten, die von eher schlichten bis hin zu extrem komplexen, aufwändigen, dynamischen, interaktiven Webseiten (z.B. "Digitaler Wenker-Atlas", <<http://www.diwa.info>> oder "Linguistic Atlas Projects", <<http://us.english.uga.edu>>) reicht.
- Die eher komplexeren Webseiten bieten eine Menge Informationen an, sind aber mitunter wenig intuitiv und bedienungsfreundlich, kleinere Webseiten bieten nicht selten kaum echte Interaktionen an.
- In geolinguistischen "Ballungsgebieten" (z.B. Deutschland, Norditalien) werden Daten redundant erhoben und aufbereitet. Nicht zuletzt aus historischen und methodologischen Gründen überschneiden sich die untersuchten Gebiete - ALF, AIS, ALI, VIVALDI und ALD haben jeweils variable Schnittmengen gleicher Messpunkte, die aber als solche nicht erkennbar sind, genauso wenig gibt es Querverbindungen, die zu einer Sprachatlas übergreifenden Darstellung beitragen könnten.

Jede Projektgruppe nutzt eigene Mechanismen zur Darstellung ihrer Daten

- es existieren keine Standards, die die Datenspeicherung oder -präsentation vereinheitlichen und erleichtern. Die Darstellung der Daten scheint ein Problem der Ressourcen-Allokation zu sein: Viel Energie wird in Konzeption und Datenerhebung investiert (was unzweifelhaft richtig und wichtiger ist), wenig Energie und informationstechnische Kompetenz wird hingegen oftmals in die

Datenpräsentation und die Nutzbarmachung der Daten für ein breites, auch Nicht-Fach-Publikum angelegt.

Aus didaktischer wie auch aus wissenschaftlicher Sicht sicherlich interessante Querverweise lassen sich durch proprietäre Datenstrukturen nicht herstellen – Methoden der Datenauswertung, die für einen Datenkorpus entwickelt wurden, bleiben “Insellösungen”, was die Vergleichbarkeit und Reproduzierbarkeit der Methodiken erschwert.

4. Vorschlag der “Geolinguistic Document Architecture” (GDA)

Die Entwicklung der Internet-Präsentationen von VIVALDI und ALD-I erfolgte in einem evolutionären Prozess. Beide weisen deshalb eng verwandte informationstechnische Grundlagen auf. Hier zeigten sich viele synergistische Effekte, besonders die sehr ähnliche Datenbankdefinitionen und -strukturen. Diese stellen das Rückgrat für die Präsentation dar und ermöglichen eine rasche Entwicklung und Anpassung an Sprachatlas-individuelle Gegebenheiten.

In der Medizinischen Informatik stand man vor einem ähnlichen Problem, welches gesundheitspolitisch und -ökonomisch sehr brisant ist und die Versorgungssituation von Patienten nachhaltig beeinflusst: Die beispielsweise in einem Krankenhaus erhobenen Daten (Blutdruck, Körpertemperatur, aber auch komplexere Daten wie Operationsprotokolle und Arztbriefe) können mangels geeigneter semantischer und syntaktischer Standards nicht weitergegeben und institutionenübergreifend genutzt werden. Daten werden so mehrfach und redundant erhoben.

Aus diesem Grunde wurde die “Clinical Document Architecture” (CDA), eine syntaktisch und zunehmend auch semantisch standardisierte Dokumentenarchitektur zur Speicherung und Übermittlung klinischer Inhalte, etabliert (DOLIN et al. 2001). CDA basiert auf der “Extensible Markup Language” (XML), einer Auszeichnungssprache zur Darstellung hierarchisch strukturierter Daten in Form von Textdateien. Die Praxis zeigt, dass hiermit eine Institutionen übergreifende Interoperabilität erreicht werden kann (MÜLLER et al. 2005), die CDA gewinnt für die Kommunikation im Gesundheitswesen eine immer wichtigere Bedeutung.

In Analogie zur “Clinical Document Architecture” sei – basierend auf den Erfahrungen der seit 10 Jahren fortschreitenden Entwicklung der Datenstrukturen von VIVALDI und ALD-I – an dieser Stelle die Entwicklung der “Geolinguistic Document Architecture” (GDA) vorgeschlagen, die dazu beitragen könnte,

- redundante Entwicklungen zu minimieren,
- vergleichbare Daten einander zu erschließen,
- einen praktikablen Standard für die Datenpräsentation zu schaffen

- und damit gerade angesichts ohnehin knapper finanzieller und personeller Ressourcen Freiräume für die eigentliche Generierung von Daten zu schaffen.

Eine detaillierte Ausarbeitung eines solchen Standards würde den Rahmen dieses Beitrages sprengen, deshalb seien hier nur einige mögliche Prinzipien dargestellt, die keinen Anspruch auf Vollständigkeit erheben.

Die Daten eines Sprachatlasses könnten in Form einer einzigen, auf XML-basierenden Datei abgebildet werden. Die Datendefinition der ersten GDA-Version (“Level one”) enthält im Wesentlichen zwei Abschnitte:

- Der “Header” (“geolinguistic_document_header”, siehe Abbildung 9) fasst allgemeine Informationen zu Inhalten und Struktur des Atlases zusammen.
- Der “Body” (“body”) beschreibt die Stimuli, deren Transkriptionen, Verweise auf Tondokumente, Metainformationen etc.

Der “Header” umfasst:

- allgemeine Informationen zum Atlas (z.B. geographischer und linguistischer Focus, Mitarbeiter, Erhebungszeitraum etc.),
- Details sowie Verweisnummern (“IDs”) für Regionen, Messpunkte, Informanten und Stimuli sowie die jeweiligen geographischen Koordinaten für die Kartendarstellung in Google Maps.

Im “Body” erfolgen dann – nach einer präzisen Zuordnung zu Messpunkt, Stimulus und Informant – die Darstellung der Transkription sowie der Verweis auf die Tondatei.

Die Vorteile einer solchen Datendefinition liegen auf der Hand:

```

<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE levelone PUBLIC "-//GDA//DTD GDA Level One 1.0//EN" "levelone_1.0.dtd">

<levelone>
  <geolinguistic_document_header>

    <institution>
      <name>Name der den Sprachtlas erstellenden Institution</name>
      <addr>Adresse etc.</addr>
    </institution>

    <linguistic_map>
      <name>Name des Sprachatlasses</name>
      <origination_dtt>Erstellungszeitraum</origination_dtt>
      <covered_area>untersuchter Raum</covered area>
      <phonetic_alphabet>verwendete Lautschrift</phonetic_alphabet>
    </linguistic_map>

    <regions>
      <region id="1" name="Region 1" coordinates="geogr. Koordinaten" />
      <region id="2" name="Region 2" coordinates="geogr. Koordinaten" />
      <region id="3" name="Region 3" coordinates="geogr. Koordinaten" />
      <region id="4" name="Region 4" coordinates="geogr. Koordinaten" />
    </regions>

    <measuring_points>
      <point id="1" region="1" name="Ort 1" coordinates="geogr. Koordinaten" />
      <point id="2" region="1" name="Ort 2" coordinates="geogr. Koordinaten" />
      <point id="3" region="1" name="Ort 3" coordinates="geogr. Koordinaten" />
      <point id="4" region="1" name="Ort 4" coordinates="geogr. Koordinaten" />
    </measuring_points>

    <informants>
      <informant id="1" point="1" name="Name" age="Alter" dialect="Dialekt" />
      <informant id="2" point="2" name="Name" age="Alter" dialect="Dialekt" />
      <informant id="3" point="3" name="Name" age="Alter" dialect="Dialekt" />
      <informant id="4" point="4" name="Name" age="Alter" dialect="Dialekt" />
    </informants>

    <stimuli>
      <stimulus id="1" text="Stimulus" />
      <stimulus id="2" text="Stimulus" />
      <stimulus id="3" text="Stimulus" />
      <stimulus id="4" text="Stimulus" />
    </stimuli>

  </geolinguistic_document_header>

  <body>

```

segue

```

<item id="1" point="1" informant="1" stimulus="1">
  <transcription>l ákwa é ká\6w\1ra</transcription>
  <soundfile>Pfad zur Tondatei 1.mp3</soundfile>
</item>

<item id="2" point="2" informant="2" stimulus="1">
  <transcription>l ákw é ká\6w\1ra</transcription>
  <soundfile>Pfad zur Tondatei 2.mp3</soundfile>
</item>

<item id="1" point="3" informant="3" stimulus="1">
  <transcription>\5n\1y\6e\1(i)t \7i\6S\1t \3E\1 \6N\1gr\7I\6h\1t</transcrip-
tion>
  <soundfile>Pfad zur Tondatei 3.mp3</soundfile>
</item>

<item id="1" point="4" informant="4" stimulus="1">
  <transcription>l ákwa é ká\4QQ\1a</transcription>
  <soundfile>Pfad zur Tondatei 4.mp3</soundfile>
</item>

</body>
</levelone>

```

Abbildung 9: "Geolinguistic Document Architecture" (GDA) – schematisches Datenmodell

- Informationen können durch die Standardisierung besser ausgetauscht werden.
- Redundante Daten werden ohne Mehraufwand zusammengeführt.
- Durch "Zusammenlegung" solcher standardisierter Sprachatlas-Corpora könnten größere Sprachatlanten (mehr Messpunkte und mehr Informanten und Stimuli) generiert werden.
- Eine Software für die Präsentation dieser Daten könnte ohne Mehraufwand andere Sprachatlas-Corpora ebenfalls darstellen (Interoperabilität),
- damit würde die Darstellung der Sprachatlanten vereinheitlicht, beschleunigt
- und wertvolle Ressourcen, die in die elektronische Aufarbeitung investiert werden müssten, könnten gespart werden.

Die Ausformulierung einer "Geolinguistic Document Architecture" wäre – das bestätigen die Erfahrungen mit der Clinical Document Architecture – ein erheblicher Kraftakt, welche einer institutions- und länderübergreifenden Zusammenarbeit bedürfte. Inwieweit der Aufwand hier in einem adäquaten Verhältnis zum Nutzen und zu den dafür allozierbaren Ressourcen sowie dem vorhandenen Interesse steht, ist dabei schwer abzuschätzen.

Bibliographie

- AIS = JABERG, Karl – JUD, Jakob: *Sprach- und Sachatlas Italiens und der Südschweiz*, Zofingen 1928-1940, 8 vol.
- ALD-I = GOEBL, Hans – BAUER, Roland – HAIMERL, Edgar (eds.): *Atlant linguistisch dl ladin dolomitich y di dialec vejins, 1^a part / Atlante linguistico del ladino dolomitico e dei dialetti limitrofi, 1^a parte / Sprachatlas des Dolomitenladinischen und angrenzender Dialekte, 1. Teil*, Wiesbaden 1998, 7 vol.
- ALF = GILLIÉRON, Jules – EDMONT, Edmond: *Atlas linguistique de la France*, Paris 1902-1910, 10 vol.
- BAUER, Roland: *VIVALDI-Sicilia. Documentazione sonora dei dialetti siciliani*, in: RUFFINO, Giovanni (ed.), *Percorsi di geografia linguistica. Idee per un atlante siciliano della cultura dialettale e dell'italiano regionale*, Palermo 1995, 543-550.
- CORTELAZZO, Manlio (ed.): *Profilo dei dialetti italiani*, Pisa 1974-88, 17 vol., einige mit Schallplatten.
- DOLIN, Robert H. et al.: *The HL7 Clinical Document Architecture*, in: "Journal of the American Medical Informatics Association", 8/6, 2001, 552-569.
- HOTZENKÖCHERLE, Rudolf – BRUNNER, Rudolf (eds.): *SDS-Phonogramme. Tonaufnahmen für den Sprachatlas der deutschen Schweiz*, Zürich 1972-76, 16 Langspielplatten.
- KATTENBUSCH, Dieter: *Atlas parlant de l'Italie par régions: VIVALDI*, in: *Estudis de lingüística i filologia oferts a Antoni M. Badia i Margarit*, Barcelona 1995, 443-455.
- KURATH, Hans: *Linguistic Atlas of New England*, New York 1939-43.
- MÜLLER, Marcel L. – KÖHLER, Carola – KATTENBUSCH, Dieter: *VIVALDI - ein sprechender Sprachatlas im Internet als Beispiel für die automatisierte, computergestützte Sprachatlasgenerierung und -präsentation*, in: "Dialectologia et Geolinguistica", 9, 2001, 55-68.
- MÜLLER, Marcel L. et al.: *Cross-institutional data exchange using the Clinical Document Architecture (CDA)*, in: "International Journal of Medical Informatics", 74/2-4, 2005, 245-256.
- ZINGARELLI, Nicola: *Vocabolario della Lingua Italiana*, Bologna 2001.