

Pragmatismus 2.0: Vision trifft Wirklichkeit

The World Atlas of Language Structures Online - eine Fallstudie

Robert Forkel

Max Planck Digital Library

Zusammenfassung

In diesem Beitrag stelle ich das Projekt *The World Atlas of Language Structures Online* (WALS Online)[1] vor, und adressieren damit das Tagungsthema 'Zukunft der Wissenschaft - von elitären Zirkeln zu "hard blogging scientists"'.

Die Entwicklung von Methoden der wissenschaftlichen Publikation lässt sich an diesem Projekt sehr schön ablesen, weil es um die online-Veröffentlichung der zweiten Ausgabe eines gedruckten Buches geht. Dabei will ich insbesondere darauf eingehen, welche Kompromisse in der momentanen Übergangsphase eines umfassenden Medienwechsels nötig sind.

1 Was ist WALS?

WALS ist eine grosse Datenbank struktureller (phonologischer, grammatikalischer, lexikalischer) Eigenschaften von Sprachen, die aus Sprachbeschreibungen (etwa Referenzgrammatiken) gesammelt wurden.

WALS besteht aus 142 Weltkarten mit beschreibenden Texten, die verschiedene Eigenschaften von Sprachen behandeln. Jede dieser 142 Eigenschaften wird für 120 bis 1370 verschiedene Sprachen beschrieben. Insgesamt zeigen die Karten Informationen für mehr als 2600 Sprachen.

Damit macht WALS Informationen zur strukturellen Diversität der Sprachen der Welt auf eine Art zugänglich, die auch für interessierte Nicht-Linguisten geeignet ist, die üblicherweise keine Grammatiken zu exotischen Sprachen lesen.

Zugänglichkeit der Informationen für einen derart erweiterten Kreis ist für die Linguistik besonders interessant, weil Feedback von Sprechern einer Sprache von großem Wert ist.

1.1 Die erste Auflage

Die erste Ausgabe von WALS wurde als Buch mit beigelegter CD-ROM publiziert. Ein “Datenblatt” dieses Buches macht bereits deutlich, daß der Leserkreis dafür tatsächlich ein “elitärer Zirkel” ist [4]:

Price: £450.00 (Hardback)
712 pages, 8 halftones, 4 line illus., 3 color halftones, 161 color maps,
360 X 248 mm

2 Vision: Go Web

Das WALS Online Projekt bestand darin, die zweite Ausgabe im Web zu veröffentlichen [2].

Korrektur müßte der Titel dieses Beitrags lauten: “Visionen treffen sich in der Wirklichkeit”. Denn – im Rückblick eher wenig überraschend – bestehen durchaus unterschiedliche Vorstellungen darüber, was denn “online” bedeutet.

Nicht zuletzt spielte auch die Vision der Max Planck Gesellschaft eine Rolle, die mit der Max Planck Digital Library eine zentrale Einheit geschaffen hat, die den Instituten Web Publishing Expertise und Web Hosting Dienste bietet.

So entstand WALS Online als Kooperation zwischen dem Max Planck Institut für Evolutionäre Anthropologie und der Max Planck Digital Library. Während die Linguisten des Instituts hauptsächlich Fachkenntnis sowie die Erfahrungen aus der ersten Auflage einbrachten, konnte die Digital Editions Gruppe der MPDL ihre Kollaborationsinfrastruktur sowie Web Publishing Erfahrung beitragen.

3 Wirklichkeit: Was bedeutet “online only” heute?

Im Folgenden will ich nun einige Beispiele auflisten, wie und wo die “online only” Vision heute noch mit dem herkömmlichen Wissenschaftsbetrieb kollidiert. Daß das Potential für solche Zusammenstöße in der Linguistik, deren Veröffentlichungen weiterhin von Buchpublikationen geprägt sind, besonders groß ist, soll dabei nicht verschwiegen werden.

3.1 Wie zitiert man WALS Online?

Von Anfang an war die Frage, wie WALS Online zitiert werden soll, ein wichtiges Thema.

In einer idealen “online only” Welt besteht ein Zitat einer anderen Ressource schlicht in einem Hyperlink. Verfolgt man diesen, sollten sich die Metadaten der zitierten Ressource – Autor, Jahr der Veröffentlichung, etc. – leicht erschließen lassen. Wie in dieser idealen Welt aus solchen Verlinkungen eine Rangliste oder Wertigkeit von wissenschaftlichen Veröffentlichungen bestimmt werden kann, ist eine andere – große – Frage. Daß so eine Unternehmung aber prinzipiell möglich ist, hat Google – für das Web allgemein – bewiesen.

In der heutigen Linguistik muß aber weiterhin davon ausgegangen werden, daß der Normalfall darin besteht, WALS Online in einem Buch zu zitieren. In gedruckten Werken gibt es aber keine Hyperlinks, und auch die schlichte Angabe eines URL ist wegen der Fehleranfälligkeit wenig sinnvoll. Bleibt die herkömmliche Referenz via Metadaten – also Autoren, Publikationsjahr, Titel, Verlag.

Für WALS Online bedeutet das, daß nicht nur Aspekte wie die einfache Navigierbarkeit die Aufteilung der Website beeinflusst haben, sondern auch die Frage, welches die einzeln zitierfähigen Bausteine sein sollen. Im Zweifelsfall wurde dabei der – wenn auch einschränkenden – Analogie zur Buchpublikation der Vorzug gegeben. Beispielsweise wurde die Website in “chapter” gegliedert, um übliche Zitierweisen anwendbar zu machen.

3.2 Wie stabil ist WALS Online?

Eine weitere Frage im Umkreis des Themas “Zitierfähigkeit” ist die nach der Stabilität des Zitierten. Wenn der Zitierende nicht mehr eine Kopie des Zitierten in seinem Besitz hat, wie wird dann sichergestellt, was tatsächlich zitiert wurde? Andererseits würden wir die Analogie zum Buch zu weit treiben, wenn wir uns selbst der Möglichkeit berauben würden, Fehlerkorrekturen einfach und schnell online durchzuführen.

Für WALS Online wurde deshalb folgendes Verfahren gewählt:

- Es wird zwischen Kerninhalten (etwa typologischen Daten) und zusätzlichen Inhalten (etwa der Positionierung von Sprachen auf der Karte) unterschieden.
- Zusätzliche Inhalte können jederzeit korrigiert werden, die Korrektur wird aber durch einen Eintrag im Blog nachverfolgbar gemacht [6].

- Kerninhalte können erst bei einer Neuauflage korrigiert werden; bis dahin werden die nötigen Korrekturen in einem Eintrag im Blog in der Kategorie *Errata* [5] bekannt – und referenzierbar – gemacht.

3.3 Wie kann man WALS Karten drucken?

Ein Spezialfall für ein Zitat aus WALS ist die Übernahme einer Karte – also einer Abbildung. Auch hier wurde bei WALS Online einerseits der idealen Welt Rechnung getragen, indem einzelne Karten inklusive Zoom-Stufe und Kartenausschnitt eindeutig per URL identifizierbar sind. Andererseits tauchte recht schnell die Frage auf: “Wie kann ich Karten aus WALS Online drucken?”

Glücklicherweise erlaubt Google Maps die Verwendung von Ausdrucken in wissenschaftlichen Arbeiten (sofern gewisse Bedingungen eingehalten werden). Mit etwas zusätzlichem Programmieraufwand konnte somit auch ein Weg gefunden werden, den Ausdruck von WALS Karten in ausreichender Qualität zur Weiterverwendung in Druckwerken zu unterstützen.

3.4 Offline Modus

Während “online only” in einer Zeit von DSL Flatrates mittlerweile für die meisten Bereiche eine echte Option geworden ist, sieht das in der Linguistik noch anders aus. Hier spielt “field work” immer noch eine große Rolle. Linguistische Ressourcen, die auch bei längeren Aufenthalten im Amazonasurwald verfügbar sein sollen, werden also auch auf absehbare Zeit noch einen “offline” Modus benötigen.

Neue Technologien wie Google Gears werden wohl in der Zukunft eine generelle Lösung dieses Problems möglich machen – also Web Applikationen, die auch offline bedient werden können.

Für WALS Online haben wir uns vorerst für die einfachere Lösung entschieden: Die CD-ROM Version der Daten, die der gedruckten Ausgabe beigelegt war, ist weiterhin frei verfügbar.

3.5 Interoperabilität mit der Vergangenheit

Während Zitierfähigkeit das Thema *Interoperabilität* für herkömmliche Publikationen größtenteils abdeckt, ergeben sich für Publikationen, die Daten enthalten – wie WALS – weitergehende Möglichkeiten.

Und wieder hielte die schöne neue Online Welt passende Lösungen unter dem Stichwort Linked Data [7] bereit. Doch in der echten Welt bedeutet “Datenbank” leider noch nicht “Linked Data”; vielmehr reicht die Bandbreite dessen, was als “Datenbank” firmiert vom Zettelkasten über die Tabelle im Word Dokument bis zu “echten” Datenbankanwendungen.

Damit bedeutet Interoperabilität in den meisten Fällen weiterhin Handarbeit, und ein Austauschformat wie csv kann – trotz fehlender Semantik – meist einfacher verwendet werden.

Gute Zwischenschritte vom Datensilo zu Linked Data stellen wohl momentan XML-basierte Spezialformate wie KML [8] dar, die einerseits die meisten Ansprüche an moderne Formate erfüllen, andererseits aber durch bereits vorhandene Applikationen – im Falle von KML Google Maps – einen sofortigen Mehrwert bieten.

Erschwert wird die Entwicklung von Austauschformaten in der Linguistik übrigens auch dadurch, daß grundlegende Konzepte, wie das einer Sprache, eher selbst Forschungsgegenstand sind, als allgemein akzeptierte Grundlage.

4 Was bedeutet das Web für WALs?

Mit seiner Entstehungsgeschichte zeigt WALs Online einerseits die Weitsicht der WALs Herausgeber, die sich schnell darüber im Klaren waren, daß diese Daten in’s Web gehören; andererseits aber auch eine Möglichkeit, die Hemmschwelle zu überwinden, die immer noch in vielen Bereichen gegenüber Publikationen im Web besteht: Durch die Erstveröffentlichung als Buch wurde einerseits den Autoren der traditionelle Lohn für ihre Arbeit zuteil – in Form eines klassischen Eintrags in ihrer Publikationsliste, andererseits wurde eine Marke WALs etabliert, die der Publikation im Web von Anfang an Autorität verlieh.

Umgekehrt wurde den Ansprüchen moderner Web Publikation Rechnung getragen, indem nicht einfach der “PDF auf dem Server”-Weg gewählt, sondern die Publikation dem neuen Medium angemessen aufbereitet wurde.

Daß offene Verfügbarkeit im Web im Fall von WALs Online eine vielfache Nutzerzahl bedeutet, wird schon deutlich, wenn man den Buchpreis der ersten WALs Ausgabe von 450 Pfund sieht. Die erhöhte Nutzerzahl erklärt sich aber auch damit, dass sich einige der behandelten linguistischen Phänomene durchaus auch dem Laien erschliessen¹.

¹ Etwa das Kapitel “Tea” (<http://wals.info/feature/138>)

Ähnlich wie bei Open Source Software Entwicklung, gilt auch für die Open Access Publikation von Daten: given enough eyeballs, all bugs are shallow ². Zusätzlich wurde mittels eines Blogs für das Projekt ein einfacher Feedback Mechanismus geschaffen. Zusammen mit der Tatsache, daß Web Ressourcen wesentlich einfacher aktualisiert werden können als ihre gedruckten Gegenstücke wird damit tatsächlich eine kontinuierliche Verbesserung der Datenlage möglich.

Eine der erfreulichen Erfahrungen mit WALs bisher ist die rege Benutzung der Feedback Möglichkeiten. Insbesondere das Einrichten von direkten Links von jedem Datenpunkt zum Blog Eintrag, der das Kommentieren ermöglicht, hat dazu beigetragen. Außerdem scheint diese sehr feingranulare Feedbackfunktion auch die Qualität der Kommentare zu erhöhen – so wird kaum ein Datenpunkt in Zweifel gezogen, ohne daß auf entsprechende Literatur hingewiesen wird.

Während es drei Jahre von der Druckausgabe zur zweiten Ausgabe Online dauerte, strebt das Projekt jetzt eine jährliche Neuauflage online an. Daß nicht jedes Erratum sofort korrigiert wird, ist der wissenschaftlichen Natur der Publikation geschuldet. So muss gewährleistet werden, dass die Daten zitierfähig sind, was auch bedeutet, dass sich der zitierte Datenzustand nicht oder nur eindeutig nachvollziehbar ändert.

Als weiterer grosser Vorteil von Web, Standards und Open Access erweist sich für WALs Online die Möglichkeit, die Karten als Google Maps Mashup zu präsentieren. Aber nicht nur WALs kann von anderen Datenquellen profitieren, es bietet sich auch selbst als zusätzliche Datenquelle an, die auch bereits von anderen Applikationen – etwa CALS, dem “Conlang Atlas of Language Structures” [3] – genutzt wird.

Damit wird ein echter Mehrwert durch erweiterte Benutzbarkeit der Forschungsergebnisse geschaffen und gleichzeitig ein Überprüfung der Ergebnisse durch die “Echo Chamber” Internet ermöglicht.

5 Ausblick

Das Potential von Publikationen wie WALs Online für die Wissenschaft wird sich aber sicher erst dann vollständig erschließen, wenn der Netzwerkeffekt greift, also wenn viel mehr ähnliche Datenquellen frei verfügbar und auch miteinander kombinierbar sind.

Nachfolgeprojekte, die in diese Richtung gehen, sind bereits geplant, und

² Zitiert nach http://en.wikipedia.org/wiki/Linus's_Law

werden versuchen, aus der kleinen Dateninsel WALS Online ein Linked Data-Archipel zu entwickeln.

Daß die Ideen von Linked Data für die Publikation von linguistischen Daten ganz besonders geeignet ist, wird schon daran deutlich, dass die Erforschung exotischer Sprachen inhärent dezentral stattfindet. Umso wichtiger ist Interoperabilität, um eine spätere Zusammenführung der Daten zu ermöglichen.

6 Lessons Learned

WALS Online will weniger visionäre Zukunft sein, als vielmehr ein Schritt auf dem Migrationspfad dahin. Der Erfolg von WALS Online bestärkt uns in der Einschätzung, daß solche Zwischenschritte unabdingbar sind.

Die Lehren aus WALS Online sind deshalb:

- Das Problem, wie Online Publikation wissenschaftlich zu werten ist, ist noch nicht gelöst.
- Die (wissenschaftliche) Welt ist noch nicht (annähernd) vollständig digital und online:
 - Mirroring von Websites spielt weiterhin eine Rolle für weniger gut an das Internet angebundene Regionen.
 - Während sich WALS Online gut in das Web einfügt, muss auch klar sein, wie man Teile davon in gedruckten Büchern verwenden kann.
 - “Ausdrucken” ist weiterhin eine gängige Methode auch Online Ressourcen zu nutzen.
- Kompromisse: Es muss ein Ausgleich gefunden werden zwischen missionarischen Ansprüchen (die wissenschaftliche Online Welt besser zu machen) und der Berücksichtigung (alt-)hergebrachter Arbeitsweisen.
- Klassisches IT-Problem: IT muß “Domain Knowledge” bekommen, und den “Kunden” verstehen.
- Datenmigration ist eine Daueraufgabe. Auch WALS Online wird hier nur eine Evolutionsstufe darstellen. Ziel muß es sein, “Daten besser zu hinterlassen, als wir sie vorfinden”.

- Die Zukunft der wissenschaftlichen Kommunikation wird erst beginnen, wenn auch der unterstützende Mikrokosmos migriert wurde.³

Literatur

- [1] Projekt Website.
<https://dev.livingreviews.org/projects/wals/>
- [2] WALs Online.
<http://wals.info/>
- [3] CALS.
<http://cals.conlang.org/>
- [4] WALs 1st edition.
<http://www.oup.com/uk/catalogue/?ci=9780199255917>
- [5] WALs Errata.
<http://blog.wals.info/category/errata/>
- [6] WALs Updates.
<http://blog.wals.info/category/updates/>
- [7] Linked Data site.
<http://linkeddata.org/>
- [8] Wikipedia-Eintrag zu KML.
http://en.wikipedia.org/wiki/Keyhole_Markup_Language

³ Erste Anzeichen, daß das geschieht, gibt es aber bereits: Ist http://en.wikipedia.org/wiki/Martin_Haspelmith bereits “scientific fame 2.0”?