## Thomas Krefeld | Stephan Lücke | Bernd Paeffgen

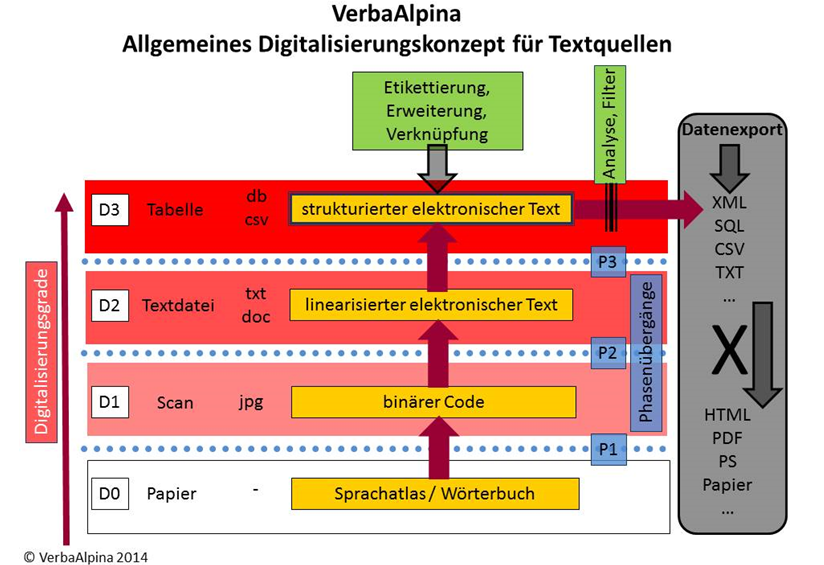
## Römisch-romanische Kontinuität in den nördlichen Alpen und im Alpenvorland: linguistische und archäologische Befunde in der Synopse

Vortrag, gehalten auf dem ZHS-Symposium „Völker und Sprachen“  
20.-21. November 2015

**München**

## Teil 2: Methodische Anmerkungen aus informatischer Perspektive

## (Stephan Lücke)



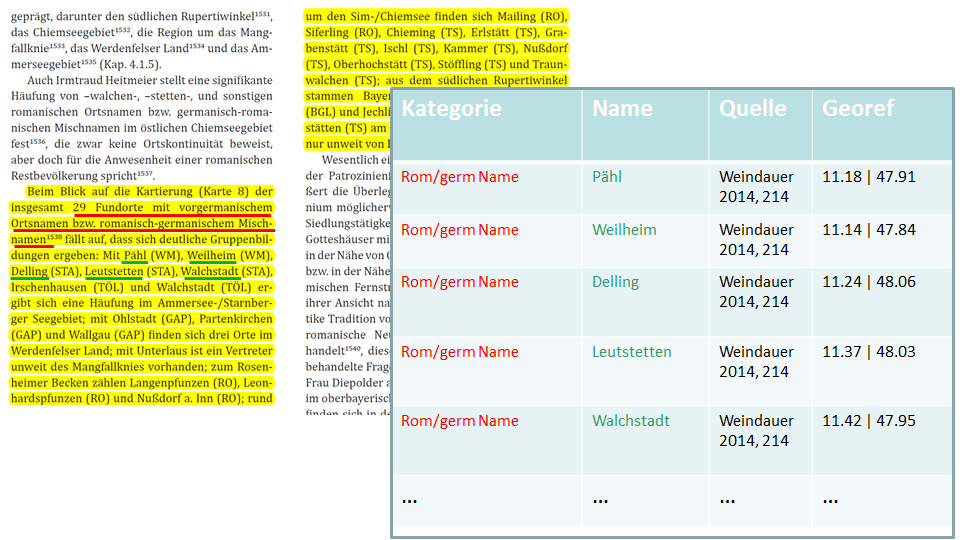
Die hier vorgetragenen methodischen Ansätze zur Kombination linguistischer und archäologischer Befunde vor dem Hintergrund römisch-romanischer Kontinuität basieren zu einem nicht unwesentlichen Teil auf informatischen Konzepten, die an der IT-Gruppe Geisteswissenschaften (ITG) der LMU in den vergangenen Jahren entwickelt wurden und derzeit in einer Vielzahl von Projekten zum Einsatz kommt. Im Zentrum steht dabei die Verwendung des sogenannten relationalen Datenmodells, in dem Informationen im Tabellenformat organisiert werden. Ich möchte im Folgenden ganz kurz exemplifizieren, wie dies im Kontext der Erzeugung einer linguistisch/archäologischen Synopse im Bereich der bayerisch-österreichischen Alpen und ihres Vorlandes zum Einsatz kam.



Eine wichtige Datenbasis für unser Vorhaben ist die von Herrn Päffgen herausgegebene Münchner Dissertation von Franz Weindauer, die sich mit der frühmittelalterlichen Besiedlung unseres Untersuchungsraumes vom 6. bis zum 8. Jahrhundert n. Chr. beschäftigt. Es handelt sich um eine herkömmliche Buchpublikation. Dabei ist es aus informatischer Sicht gar nicht so relevant, ob der Text auf Papier oder in Gestalt einer PDF-Datei - über die wir auch verfügen - vorliegt. Entscheidend ist, dass es sich im informatischen Sinne nicht um strukturierte Daten, sondern eben um Fließtext handelt. Und die dem Wort „Text“ innewohnende Vorstellung vom „Gewebe“ verweist auf den entscheidenden Unterschied zwischen eben „Text“ und strukturierten Daten: Die im Fließtext enthaltenen Informationen sind eng mit dem vom Autor vorgesehenen bzw. schlicht dem gegebenen Kontext verwoben und können an Ort und Stelle, d.h. unmittelbar \*im\* Text, nicht so ohne weiteres isoliert oder ergänzt werden. Aus diesem Grund sind Suchvorgänge mühsam und Sortierungen nach unterschiedlichen inhaltlichen Kriterien nahezu unmöglich. Eine gewisse Abhilfe schaffen traditionell Indizes.



Unsere Vorgehensweise kann man als digitale Tiefenerschließung bezeichnen. Ich präsentiere hier nun zunächst eine Karte aus der Publikation von Franz Weindauer. Die roten Punkte bezeichnen Ortsnamen, die von Weindauer als „vorgermanisch/romanisch“ klassifiziert werden. Der Karte selbst sind dabei weder die Ortsnamen zu entnehmen, noch gibt sie Aufschluss darüber, welche Namen als vorgermanisch und welche als romanisch betrachtet werden - was, nebenbei gesagt, keinesfalls dasselbe ist. Die Informationen, die der vorliegenden Kartierung zugrunde liegen, sind an verschiedenen über das Buch verteilten Orten zu finden. Listen mit den zur Kartierung passenden Ortsnamen finden sich im Anhang der Arbeit.

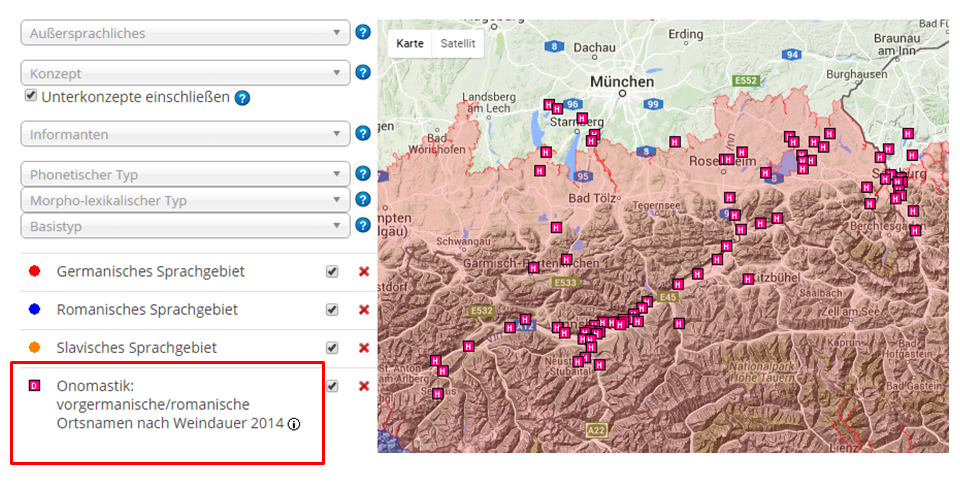


Die entscheidende Arbeit im Rahmen der digitalen Tiefenerschließung besteht in der Trennung der im Text enthaltenen Informationen in Primär- und zugeordnete Metadaten. Im konkreten Fall betrachten wir die Ortsnamen als Primärdaten sowie die Information, dass es sich bei diesen Ortsnamen um vorgermanische bzw. romanische Namen handelt, als Metadaten. Primärdaten können eine unbegrenzte Anzahl und Art von Metadaten besitzen. Neben der Information zum Ursprung der Ortsnamen haben wir Seiten- und Anmerkungszahlen als Metadaten registriert, die den Ursprung der entsprechenden Informationen dokumentieren.

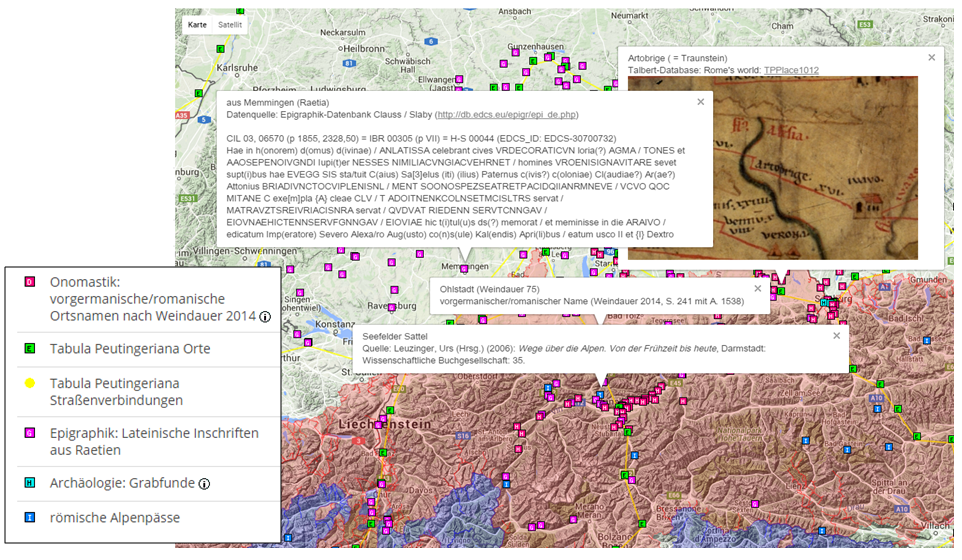
Dem Wesen des relationalen Datenmodells entsprechend, wurden diese sodann in Tabellengestalt überführt, was zu dem hier vorgestellten Ergebnis führt

Daten im Tabellenformat können jederzeit um eine beliebige Anzahl von Spalten oder Zeilen erweitert werden. Handelt es sich um Spalten, spricht man von Attributen bzw. Eigenschaften. Zu unseren Zwecken ist eine Erweiterung um zusätzliche Spalten erforderlich gewesen, die die Daten der von uns durchgeführten Georeferenzierung aufnehmen konnten.

Abschließend erfolgte noch eine Angleichung der Tabellenstruktur an die im Rahmen des Projekts VerbaAlpina für die Verwaltung von georeferenzierten Daten angelegte Tabellenstruktur sowie der Import in eben jene Tabelle.



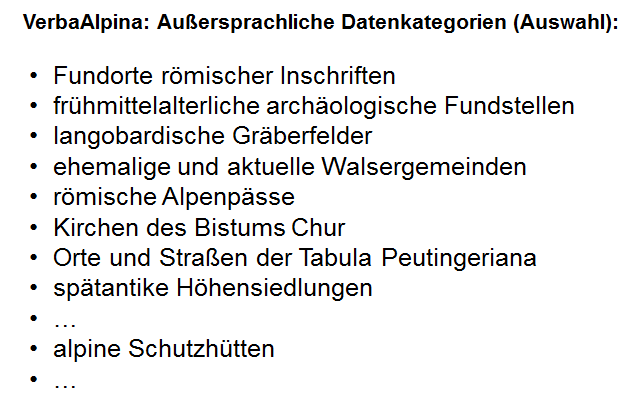
Nunmehr war und ist es möglich, die auf der analogen Karte von Franz Weindauer abgebildeten Symbole auf einer interaktiven, im Internet abrufbaren Online-Karte abzubilden. Der eigentliche Mehrwert dieser Methode besteht aber natürlich nicht in der Erzeugung einer Karte, die in nahezu identischer Gestalt bereits in gedruckter Form vorlag, sondern darin,



dass die präsentierten Daten zum einen um eine beliebige Art und Anzahl von Metadaten ergänzt werden und zum anderen im Sinne der heuristischen Methode mit vermeintlich „sachfremden“ anderen Daten kombiniert werden können. Dies ist eben erst durch die Herauslösung der Daten aus dem Verbund des geschriebenen Textes möglich.

Die VerbaAlpina-Datenbank gliedert sich grosso modo in zwei Teile: Einen, der Sprachdaten enthält, und einen anderen mit außersprachlichen Daten. Allen VerbaAlpina-Daten ist gemein, dass sie georeferenziert sind. Auf der interaktiven Online-Karte sind \*alle\* Daten, also Sprach- sowie außersprachliche Daten, frei miteinander kombinierbar.

Neben der großen Menge an Sprachdaten, die derzeit den Hauptanteil der VerbaAlpina-Datenbank bilden, enthält die Datenbank unter anderem die folgenden außersprachlichen Datenkategorien:



Die Daten werden stetig erweitert und gegebenenfalls korrigiert. In diesem Zusammenhang ist auch beabsichtigt, den Zugang zum Datenmaterial der sog. „Crowd“ zu öffnen, die sodann neue Daten wird beitragen bzw. vorhandene wird kommentieren können.