# Processing dialect data using an XML database

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

## The Data

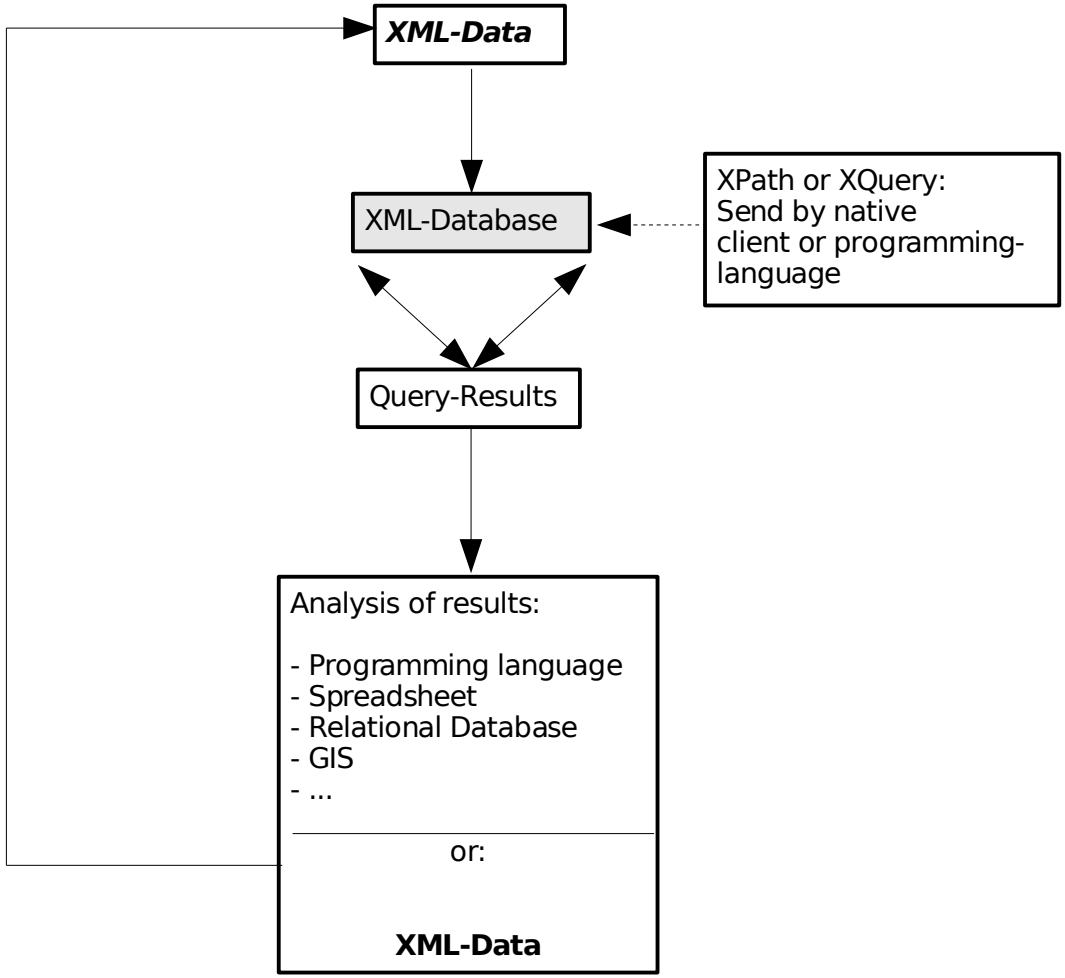
- Ultimately, we will have 200 sites
- At the moment: 66 + 14 without coordinates
- Annotated in XML, using the CLaRK-system
- Ca. 155 words per site

**In Tübingen:**

- Integration into the XML database eXist
- Generation of wordlists in L04-format via XSLT
- Querying the database with the native client and from Java



## The XML Workflow





## The eXist-database

- OpenSource, written in Java
- Runs on Windows, Linux, ...
- A lot of interfaces:  XML-RPC, SOAP, HTTP, ...
- Programing languages: Java, Perl, PHP, ...
- Integrated webserver
- Native client

- XPath and XQuery for querying the database
- XSL transformations included
- XML validation against DTDs

- In Tübingen:
- Most programs are written in Java
- XML-RPC is used as interface from within the Java programs
- Based on Java 5, developed in NetBeans IDE

## XPath/XQuery

- **XPath: Getting all sampa/variant-tags which contain a blank**

```
collection("/db/nbd")//sampa/variant[contains(., " ")]
```

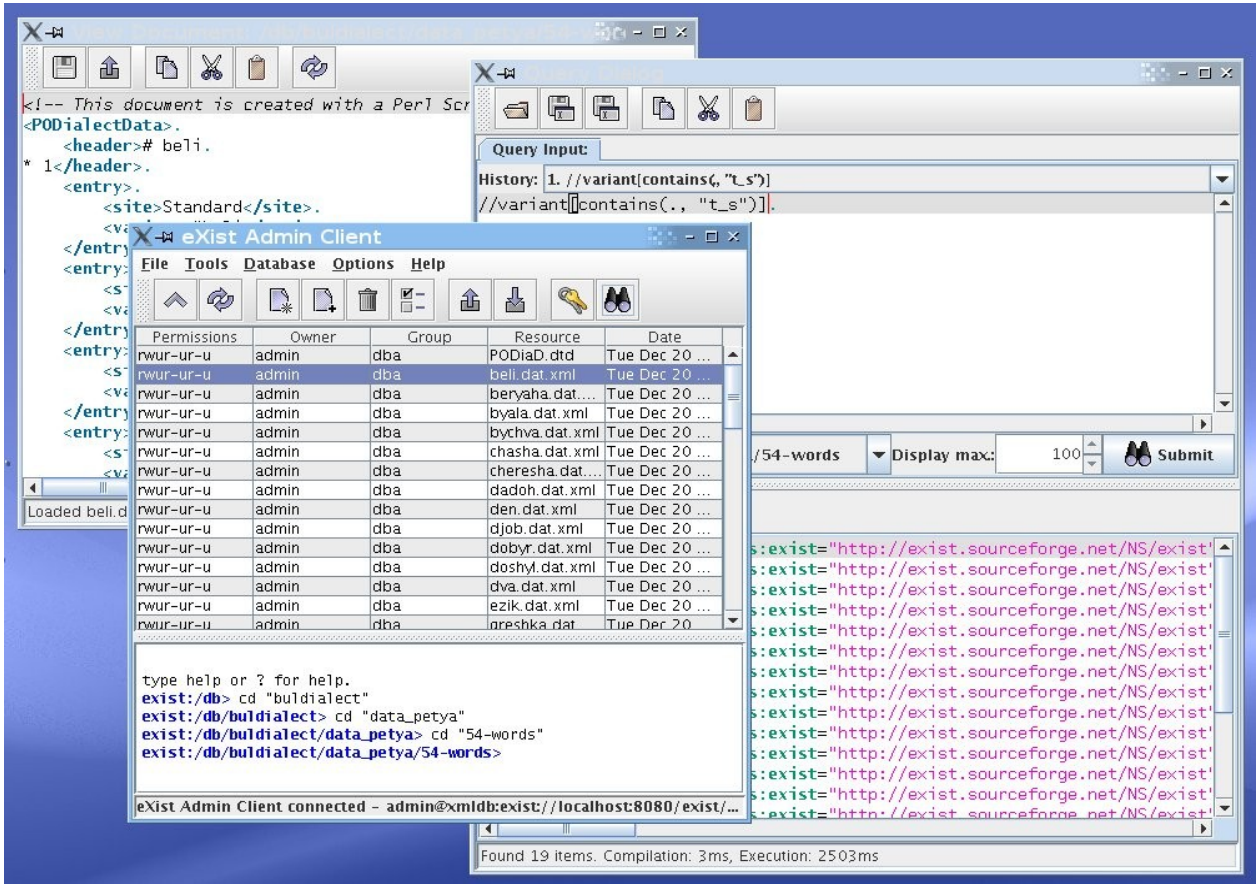
- **XQuery: Building a list of all entries which have an english-tag “lamb”**

```
let $nbd := collection(“/db/nbd”)

for $entry in $nbd//entry where
   string($entry/english)=”lamb”
return <site>{$entry/../name}{$entry/../num}{$entry}</site>
```

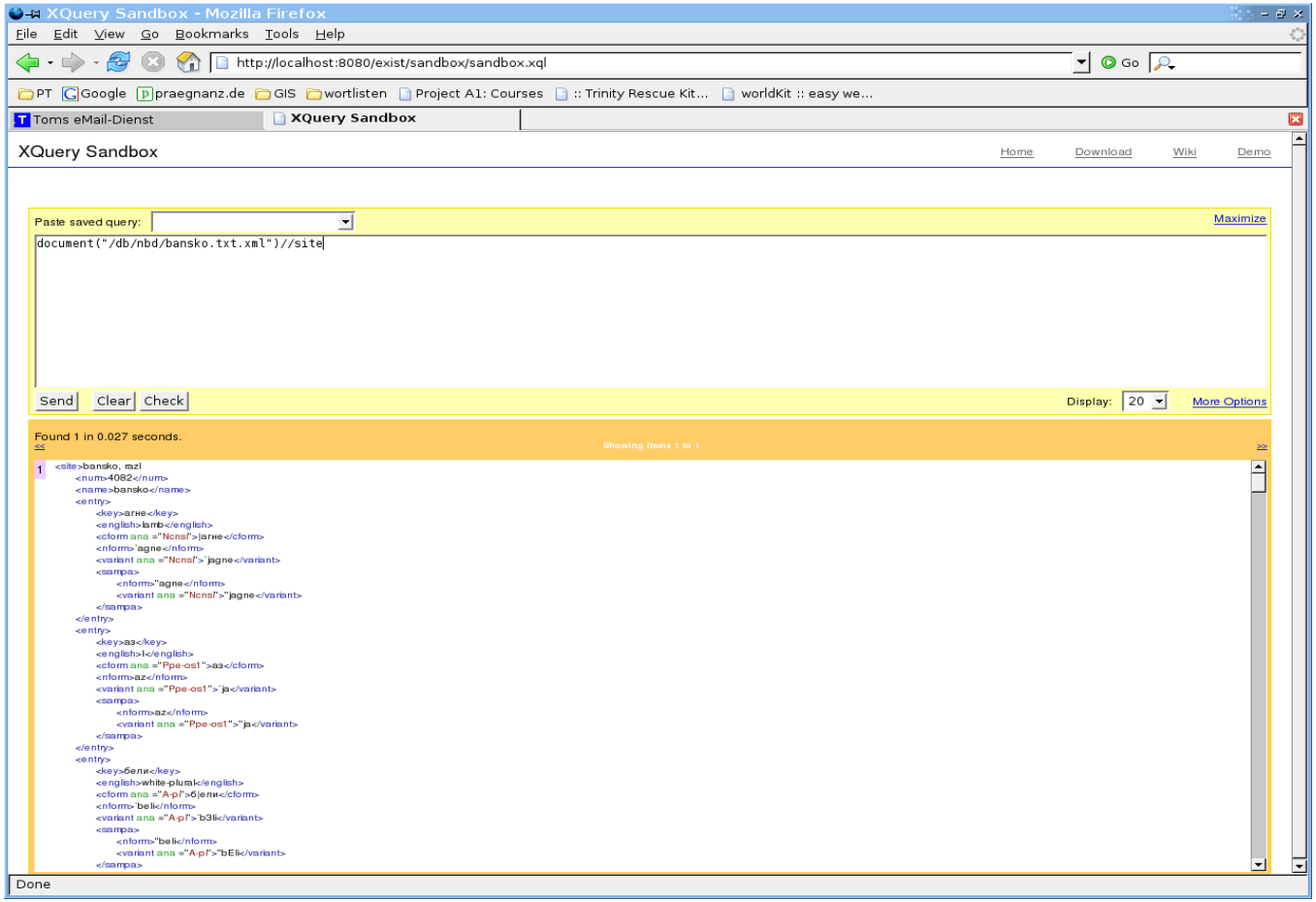
Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

# The eXist database





# The EXISTDATABASE

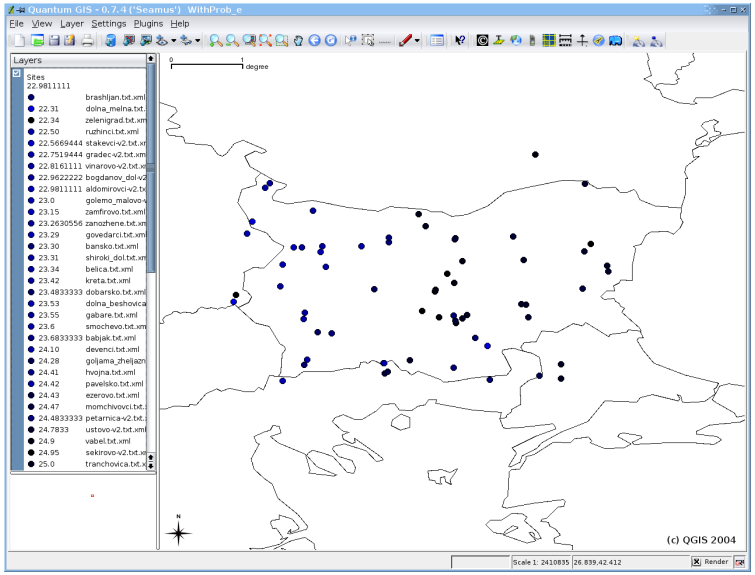


Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

## Visualization with a GIS

- GIS: Geographical Information System
- A geographical coordinate system
- Visual data is arranged in layers
- Layers can contain vector- or raster-data
- Combination of different layers

- QGis, an OpenSource-GIS is used here
- QGis uses XML as data format





## References

- **eXist XML database: http://exist.sourceforge.net**
- **XQuery standard: http://www.w3.org/XML/Query**
- **XQuery tutorial: http://www.w3schools.com/xquery/default.asp**