Data integration for paleoenvironmental and archaeological GIS based analysis

Christian Willmes[1], Daniel Becker, Georg Bareth

GIS & RS Group, Institute of Geography, University of Cologne

CAA 2015 - Keep the Revolution going
Siena, Italy - March 31st, 2015

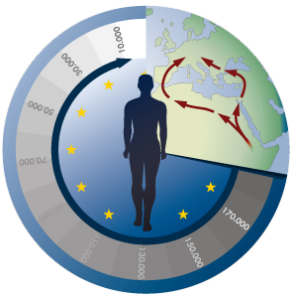
[1] c.willmes@uni-koeln.de

Outline

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Introduction

Collaborative Research Centre 806

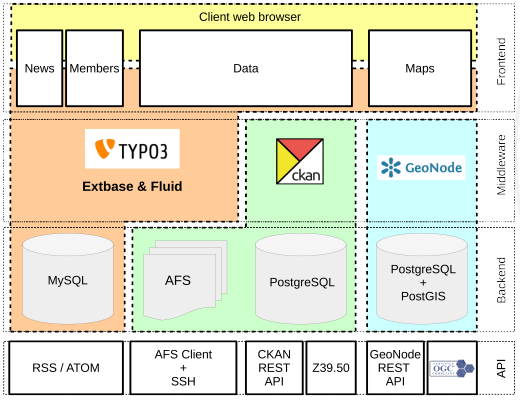


- Interdisciplinary research project
- Concernes the history of mankind, and in particular the migration/expansion process from africa to europe
- Disciplines of archaeology, geosciences and cultural sciences
- Planned for an overall duration of 12 years (3 x 4 year terms), since 7/2013 in second phase
- More information about the project in detail is available at: `http://www.sfb806.de/`

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Introduction

CRC806-Database



http://crc806db.uni-koeln.de/

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Collaborative Research Database

Data Input

- Literature
- Research

Data Import

- Datasets/Databases
- GIS Data
- Internal DBs

Data usage / discovery

- Query, Browse, Filter
- Interactive Maps
- Data Graphs

Data export

- GIS Analysis
- Statistics
- Maps

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

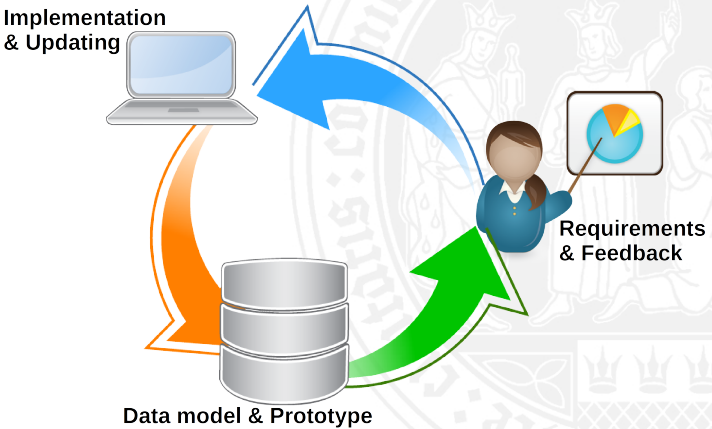
Semantic Mediawiki

- Allows to add **structured data** to a Wiki
- Allows **collaborative editing** of structured Data
- Allows **complex queries**
- Allows **export and display** of query results in many formats

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data model development

Prototyping approach



Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data model development

Mobo Framework

- `https://github.com/Fannon/mobo`
- Allows to develop the data model in **JSON-Schema** notation
- Model adjustments are applied centrally (normally many things to consider in SMW)
- Datamodel can be managed in **VCS like GIT**
- Same model can easily be **deployed to many SMW** instances

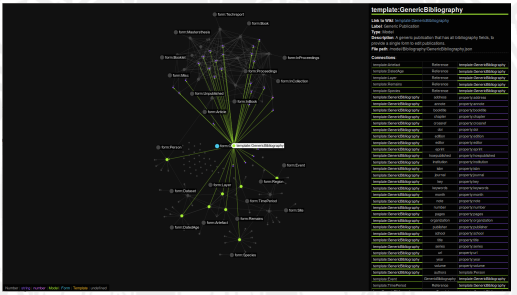
mobo

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data model

The model covers the following domains:

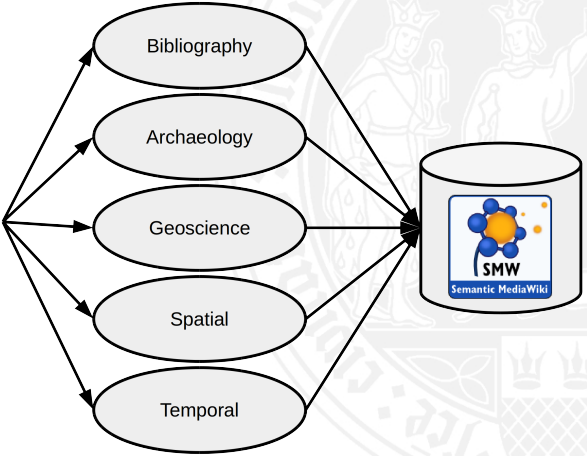
- **Spatial** (Site, Region)
- **Temporal** (Event, TimePeriod, DatedAge)
- **Archaeological** (Artefact, Remains, TimePeriod, Layer)
- **Geoarchaeological** (Core, Sediment, Stalagmite)
- **Paleoenvironment** (Climate, Vegitation, FaunTaxon, Model Simulations)
- **"Recources"** (Bibliography, Datasource, (Geo)Datasources)
- **"Organizational"** (Persons, Cluster, Project)



Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data integration

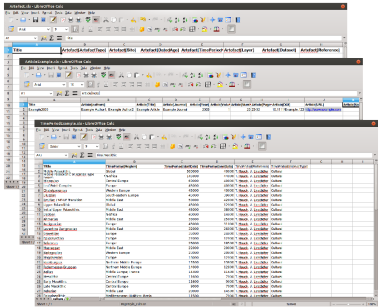
Data from literature workflow

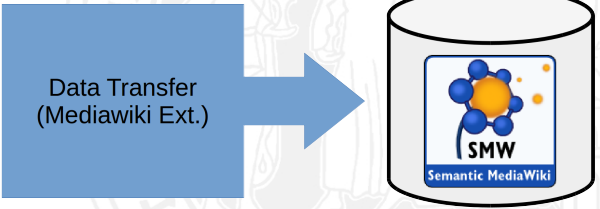


Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data integration

Spreadsheet Template Import



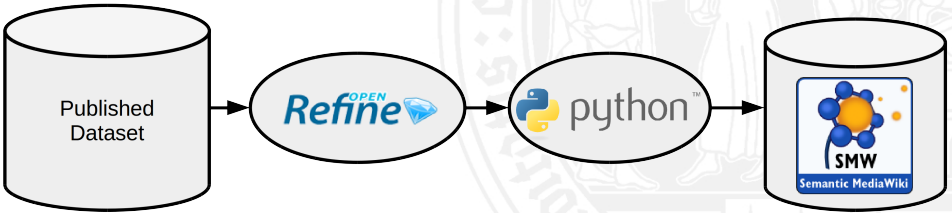


Spreadsheets with fixed schemas (templates) of SMW classes.

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data integration

Integration of datasets workflow



Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data cleaning using OpenRefine

http://openrefine.org/

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data import via Python Script

~/ownCloud/Research/Vortrag/2015_CAA/img/INQUAtoSMW.py • - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

INQUAtoSMW.py

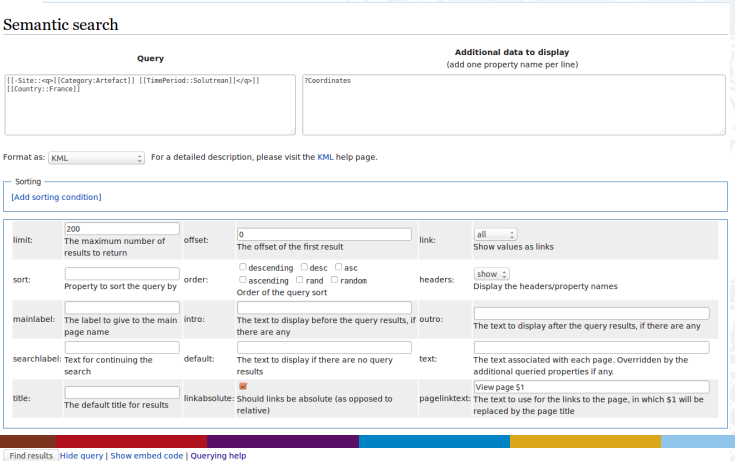
```
#!/usr/bin/python

import xlrd
import xml.etree.cElementTree as ET
import itertools, collections

def consume(iterator, n):
    collections.deque(itertools.islice(iterator, n))

#Excel Definitions
excelFile = xlrd.open_workbook("INQUA_sanitized.xls")

#XML init
root = ET.Element("Pages")

iterator = range(10001, sheet.nrows-1).__iter__()
for i in iterator:
  row = sheet.row_values(i)

  sname = unicode(sanitize_name(row[2]))

  Site = ET.SubElement(root, "Page")
  Site.set("Title", sname)

  [...]

#write output
tree = ET.ElementTree(root)
tree.write("INQUA.xml")

```


Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data discovery

Example:Query Sites by TimePeriod and Region

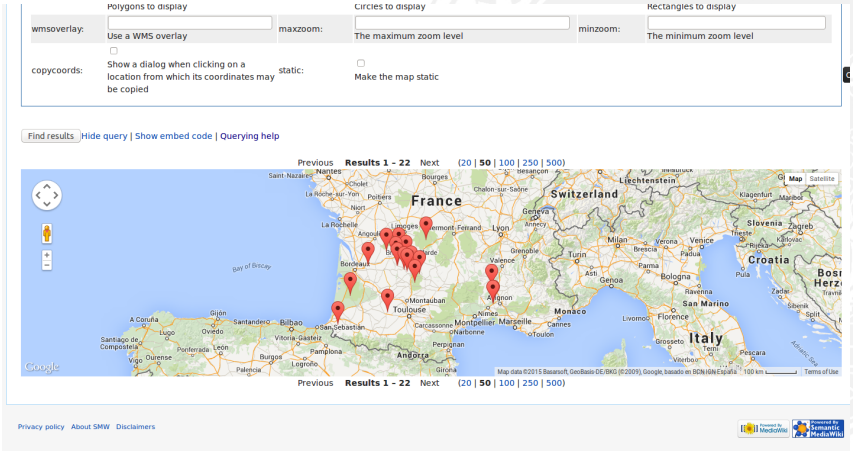
```
[[-Site::<q>[[Category:Artefact]][[TimePeriod::Solutrean]]</q>]][[Country::France]]
```



Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data discovery

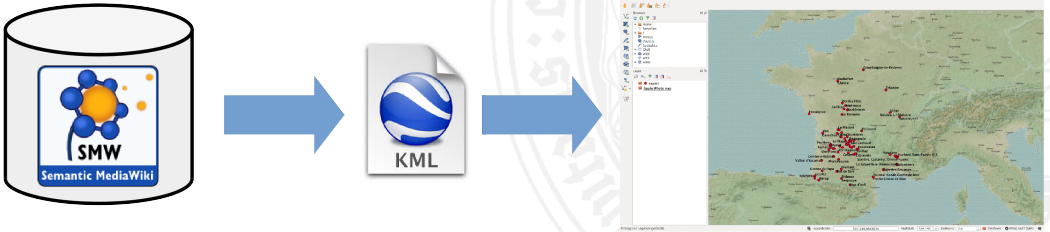
Example:Query Sites by TimePeriod and Region



Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Export data to GIS

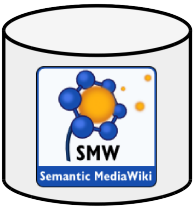
SMW Querysresult in KML result format

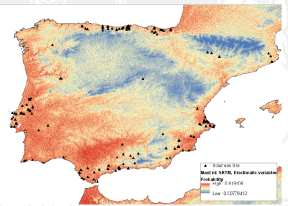


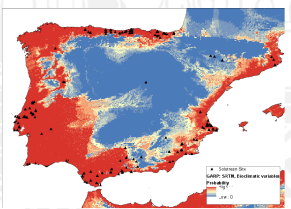
Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

GIS Analysis

Niche- or Species distribution- Modelling



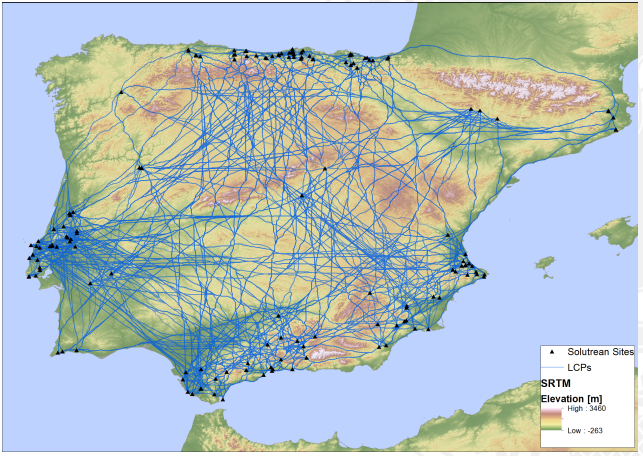




Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Data analysis

Networkanalysis based on LCPs



Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Conclusions and Outlook

- Integrated database of diverse information
- Development of an ”tailor-made” data model
- Mapping the data model to existing Vocabularies
- Publish the database as Linked Data

Christian Willmes, Daniel Becker, Georg Bareth
GIS & RS Group, Institute of Geography
University of Cologne

Thank you very much!

eMail: c.willmes@uni-koeln.de
twitter: @cwillmes
http://crc806db.uni-koeln.de

This research was conducted within the Collaborative Research Centre 806 (www.sfb806.de) funded by the German Research Foundation (www.dfg.de).