The WissKI Project
A scholarly communication infrastructure

Georg Hohmann
Germanisches Nationalmuseum Nürnberg
Department of Museum Informatics

CIDOC 2009
Santiago de Chile, 30. September 2009

Overview

- Project funded by the DFG (German Research Foundation) 2009 - 2011
- Partners
  - Friedrich-Alexander-University of Erlangen-Nuremberg (FAU), Chair of Computer Science 8 Artificial Intelligence (Prof. Günther Görz)
  - Germanisches Nationalmuseum (GNM) Nuremberg, Head of Museum Informatics (Dr. Siegfried Krause)
  - Zoologisches Forschungsmuseum Alexander Koenig (ZFMK) Bonn, Head of Biodiversity Informatics (Dr. Karl-Heinz Lampe)
- Staff
  - Dipl.-Inf. Martin Scholz (FAU)
  - Georg Hohmann M.A. (GNM)
  - Dipl.-Inf. Mark Fichtner (ZFMK)
  - 2 Student Assistants

Goal

- To create a wiki-based software system, that ...
  - supports scientific communication.
  - supports scientific documentation in memory institutions.
  - provides long-term availability of research results.
  - assures the identity of authorship.
  - assures the authenticity of information.
  - enables the persistence of citations.
  - supports quality management.
  - supports the preparation of scientific publications.
  - supports semantic content analysis based on ontologies.
  - serves as a platform for curated knowledge.

Use Cases

- ZFMK
  - Biodat specimen database
  - Database of the research diaries of Wilhelm Aerts (1885-1964)
- GNM
  - Database of Goldsmith's Art in Nuremberg 16th – 19th century
  - Database of the Early Duerer Research Project

Unique Features

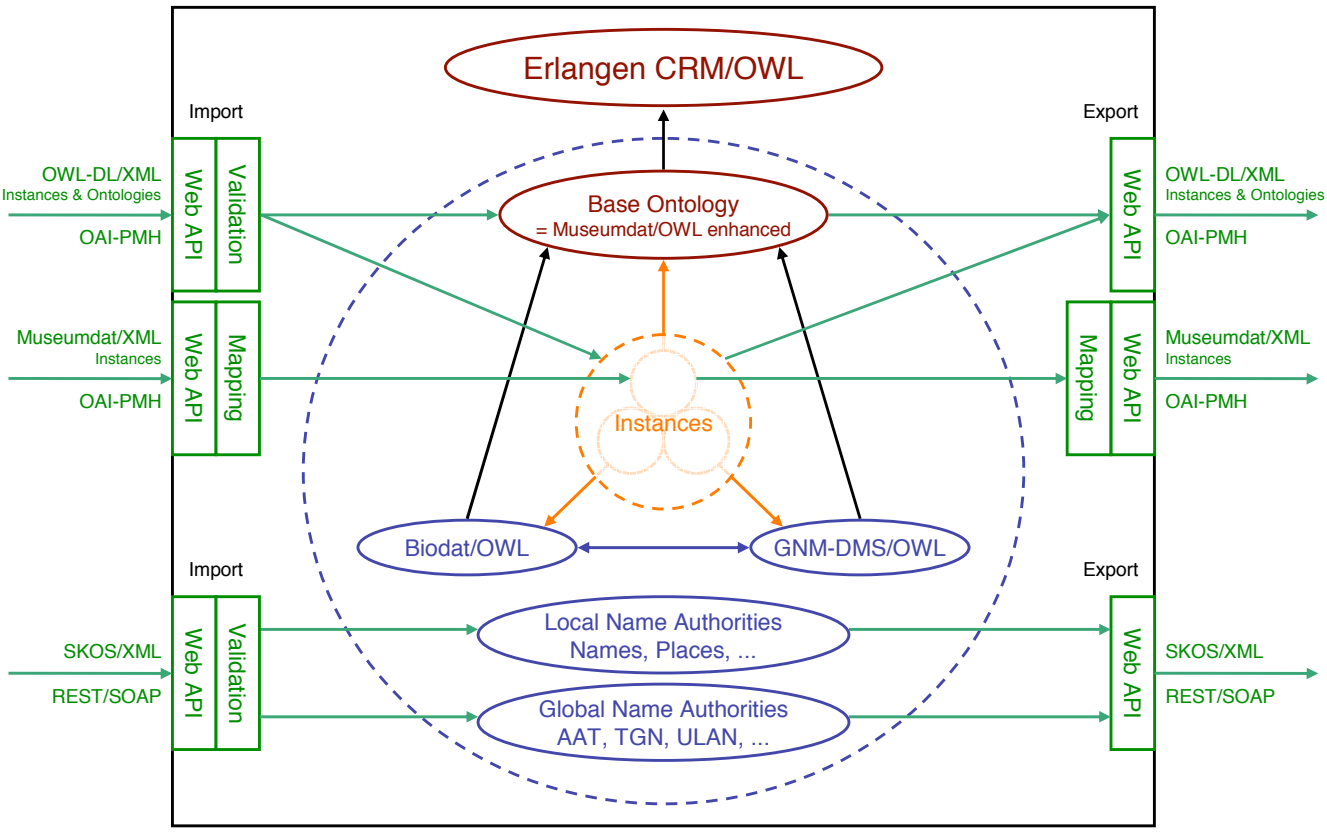
- Other projects dealing with wiki technology and knowledge management, e.g.
  - KiWi - "Knowledge in a Wiki" - project (http://www.kiwi-project.eu)
  - OntoWiki (http://ontowiki.net)
  - WIKINGER – WIKI Next Generation Enhanced Repository (http://www.wikinger-escience.de)

- Unique Features of WissKI
  - Pure OWL-DL architecture
  - Accentuation of communicational aspects
  - The only project with focus on humanities data
  - Tools for scientific documentation
  - Automatic text analysis for semantic annotation

Workpackages 2009

- Creating an integrated overall concept
- Selecting a software framework to build on
- Preprocessing of data from the use cases
- Creating suitable ontologies
- Incorporating of common authority files
- Creating local authority files
- Developing a first software prototype

Overall Concept



= Base Components = Local Data Models = Data = Interfaces
= Hierarchical Ontology Connection (is-a) = Equivalent Ontology Connection (same-as) = Instance Of = Data Flow

Software Framework

Drupal

Core

Modules

Third-Party

- CCK
- Views
- WikiTools
- WysiwygAPI
- ImageAPI
- ...

WissKI

- OWL/RDF System
- Discussion System
- Automatic Text Annotator
- Authority Files Management
- Import/Export API
- ...

Database

- Triple Store

All software used is available under free software licences.

Ontologies

- Available as first drafts
  - Erlangen CRM / OWL
    OWL-DL 1.0 implementation of the CIDOC CRM 5.0.1
  - WissKI Base Ontology
    Enhanced OWL-DL 1.0 implementation of Museumdat
  - Biodat OWL
    OWL-DL 1.0 Ontology of the ZFMK Biodat Database & Aerts Diaries
  - Duerer OWL
    OWL-DL 1.0 Ontology of the Early Duerer Research Project Database
- In development
  - Gold OWL
    OWL-DL 1.0 Ontology of the GNM Goldsmith's Art Database
  - GNM-DMS OWL
    OWL-DL 1.0 Ontology of the GNM Documentation Database

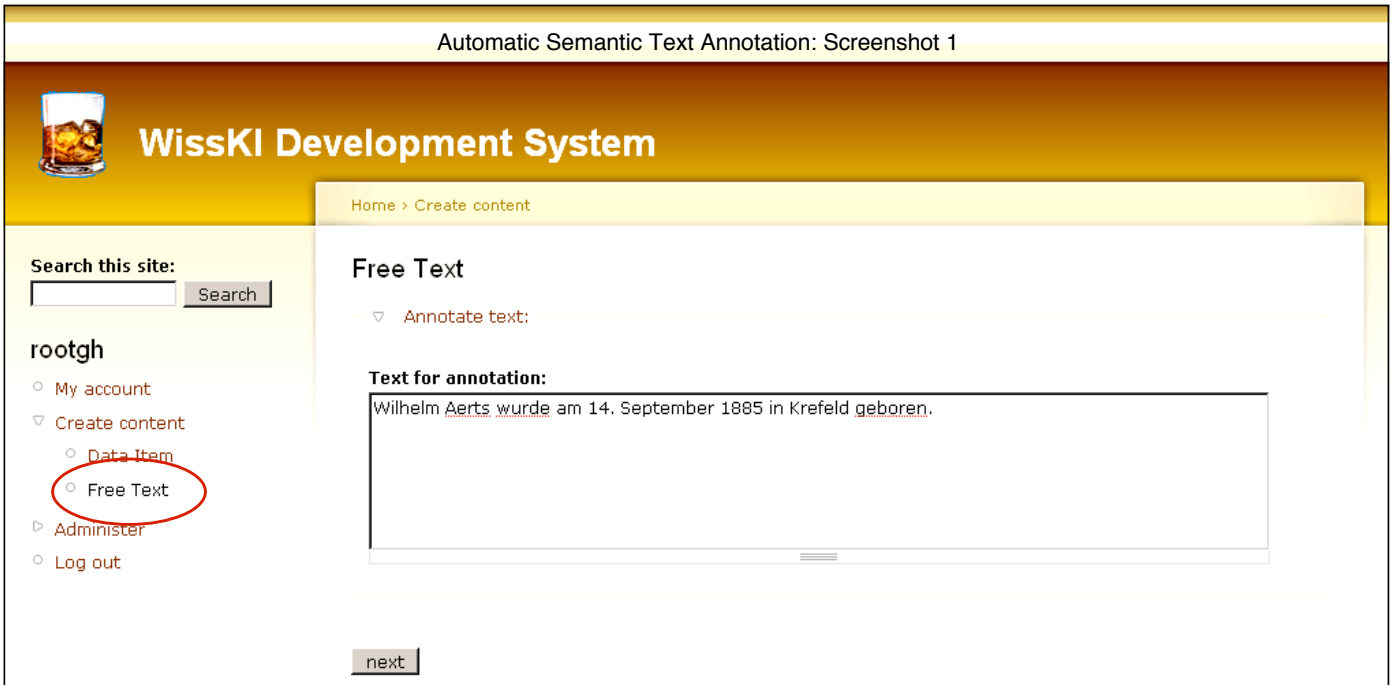
Content Creation

- Each item is an instance of a class in a given ontology.
- Each instance is represented by RDF-triples and stored inside a triple store.
- Instances can be created in three ways:
  - Via submitting a free text.
  - Via traditional webforms.
  - Via uploading of an image or other multimedia files.
- Every manual or automatic data entry is assisted by authority files.

Automatic Semantic Text Annotation

- Assumption
  - Scientists are used to communicate and record data by writing texts (papers, articles, books, etc.).
- Approach
  - WYSIWYG-Editor enhanced with capabilities of automatic semantic markup.
  - Automatic recognition of named entities (place names, person names, calendar date etc.) based on name authorities.
  - User should be motivated to revise annotations.
  - The text itself is treated as an instance of E31.Document.

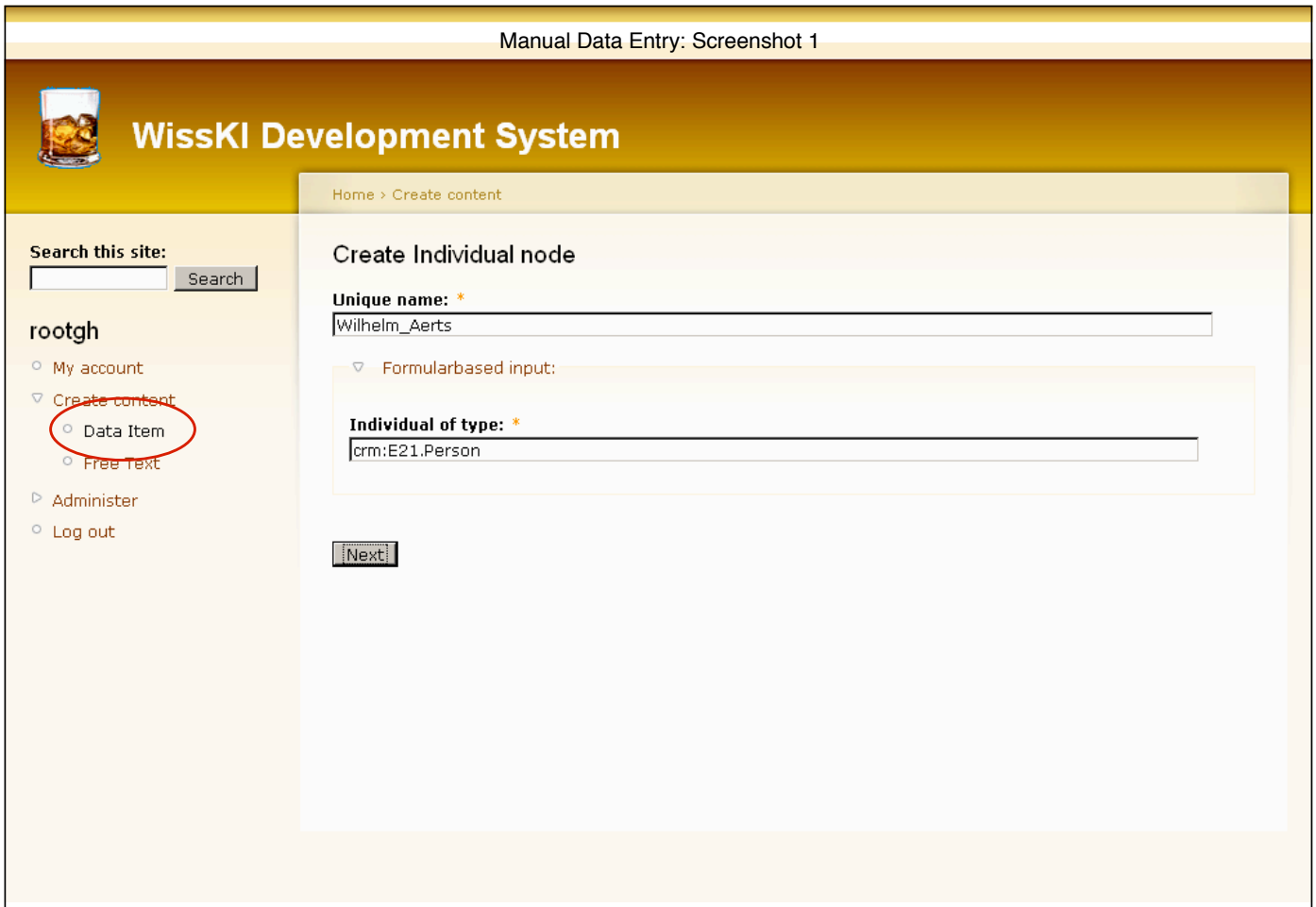
Automatic Semantic Text Annotation: Screenshot 1



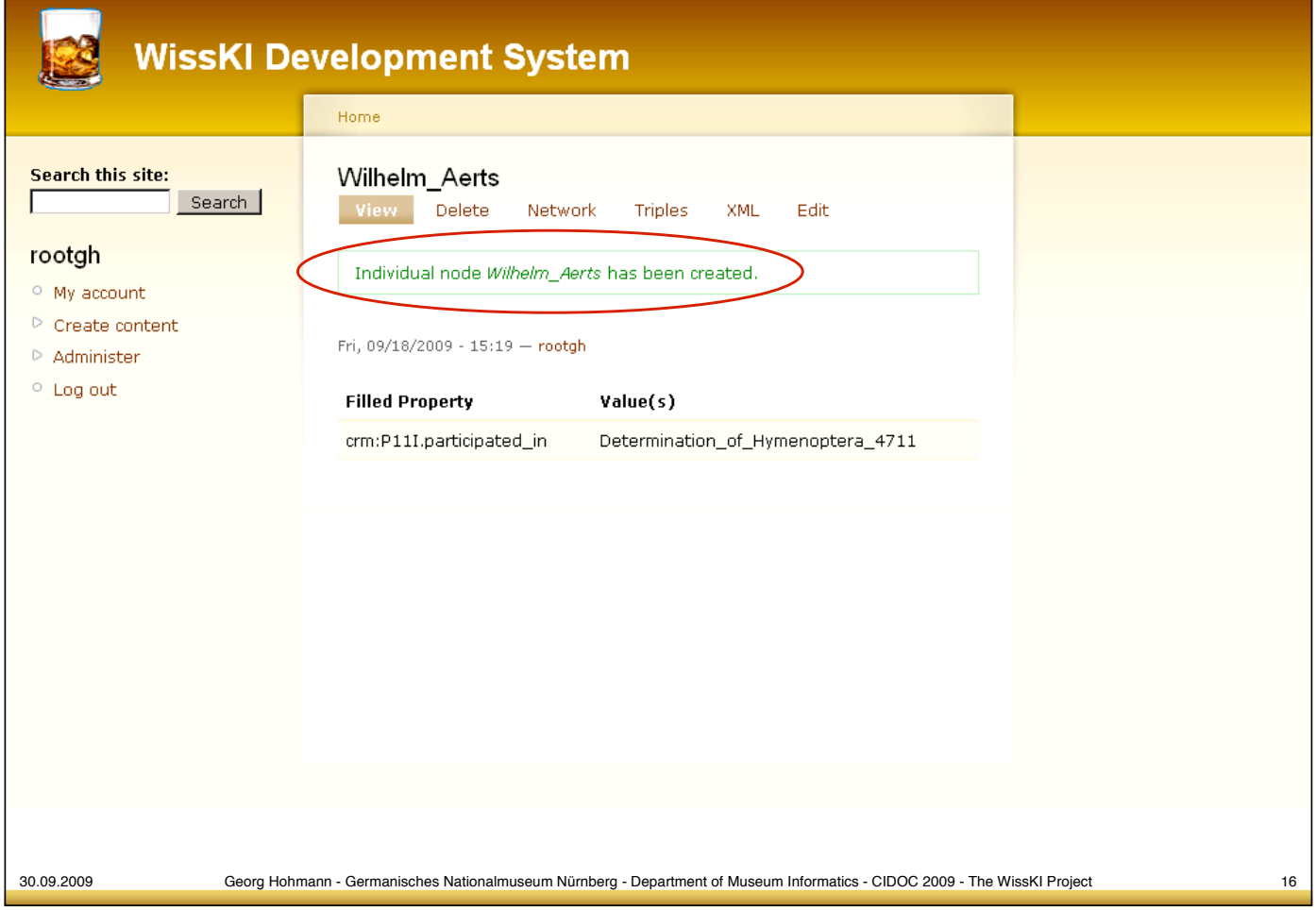
Manual Data Entry

- Assumption
  - Some scientists are used (and like) to record data by entering data manually in a database with a forms-based interface (MS Access etc.).
- Approach
  - Forms are dynamically built from the concept description inside the used ontologies.
  - Available properties are shown depending on their definition.
  - Data is standarized with regard to the loaded ontology.
  - Input of values is supplemented by the data from authority files.

Manual Data Entry: Screenshot 1



Manual Data Entry: Screenshot 2



Communication

- The choosen software framework itself provides several community functionalities like mailing(lists), forums, blogs, support for multimedia files, user profiles etc.
- Instances can be discussed.
  - Data generated by these discussions can replace or modify the previous data.
- A curator decides which statements represent the official statements of a project.
- Facilities for cooperative preparation of scientific publications
- Automatic data exchange between different installations of WissKI.
- Import/Export of data via widely adopted standards and based on standarized vocabulary.

Discussion: Screenshot 1

Achievements so far

- A prototype is already in development.
- Process in most working packages is just in time.
  - Automatic semantic text annotation (proof-of-concept).
  - Automatic form creation based on ontology concepts (proof-of-concept).
  - Integration of authority files. Byproduct: A tool to convert Getty authority xml to SKOS.
- 6 ontologies are available in different stages of development.
- Data of some use cases is already prepared to test the system.
  - E.g. Aerts Diaries ~ 177 places, 12829 determination events, 1069 examinations events ~ approx. 1.4 Million triples.
- Project website online (http://wiss-ki.eu).
- Papers published and talks given related to the project.