On the way to a ScienceNet ?

Thomas Severiens
severiens@isn-oldenburg.de

Institute for Science Networking Oldenburg

Zuse Institute Berlin
January 14th, 2005

Teile der hier vorgestellten Arbeit wurden gefördert mit Mitteln der
- European Physical Society (EPS) Projekt PhysNet
- Deutschen Forschungsgemeinschaft (DFG) Projekt OAD
- National Science Foundation (NSF) Projekt OAD

Was erwartet Sie?

- Ein Überblick über die letzten Entwicklungen in PhysNet (aus einem Vortrag zusammen mit Christian Thiemann)
- PhysNet und Open Access und Open Archives
- Kooperationsschnittstellen mit Math-Net (sicherlich nur ein Einstieg in die weitere Diskussion)
- Kondensationskerne auf dem Weg zum ScienceNet?

The PhysNet RDF Datamodel

- Vorstellung der aktuellen Entwicklungen im PhysNet
- Folien teilweise auf English, da den Schulungsfolien für das Netz der Administratoren entnommen
- In Kooperation mit Christian Thiemann, Universität Göttingen (thiemann@isn-oldenburg.de)

Former State of the PhysNet Service

What data is stored?

- PhysNet lists links to other websites (homepages, publication lists)
- Link lists reflect geographical and inner-institutional structure

How is the data stored?

- Links are stored in HTML files
- Structure is stored by the use of different HTML tags (like headings and nested definition lists)

Problems of Former Data Storage

Redundant Data

Structure is duplicate in PhysDep and PhysDoc

Some institution (e.g. Institute for Theoretical Physics)
may be listed as a sub-institution of Dept. of Physics
which is a sub-institution of My University in PhysDoc
but listed directly as a sub-inst. of My University in PhysDep.

Inconsistent naming scheme for identic objects

Institutions with sub-institutions in different locations
(e.g. University of California and its campuses)
are not always mentioned the same way at all occurences
(e.g. UC Berkeley, but Univ. of California in Santa Barbara).

Problems of Former Data Storage

Mixture of Data and Visualization

HTML files contain both data and visualization

- Data is hidden in the HTML tags and therefore difficult to maintain
- Simple mistakes (e.g. forgetting to state the end of a definition list) lead to severe data damage (e.g. showing a city as a sub-institution of an university)

Problems of Former Data Storage

Anforderungen an die neue Datenstruktur

- Kompatibilität zum existierenden PhysNet
  (es soll keine vorhandene Information verlohren gehen)
- Erschließung vorhandener Information
  (Aufdeckung vorhandener impliziter Information)
- Innovative Dienste
  (Sortierung und Vernetzung nach Forschungsgebieten,
  Alerting bei Aufkommen neuer Dienste,
  Publikations-Alerting,
  Portale für spezielle Interessengruppen (Dissertationen),
  etc.)
- Verknüpfung von Geo- und Fachinformation
- Schnittstellen für andere Fächer
- Skalierbarkeit

Solution

Separation of data and visualization

- Data is stored in an abstract datamodel
- Scripts generate views on the data (HTML files)

Solution

Advantages of separated abstract data

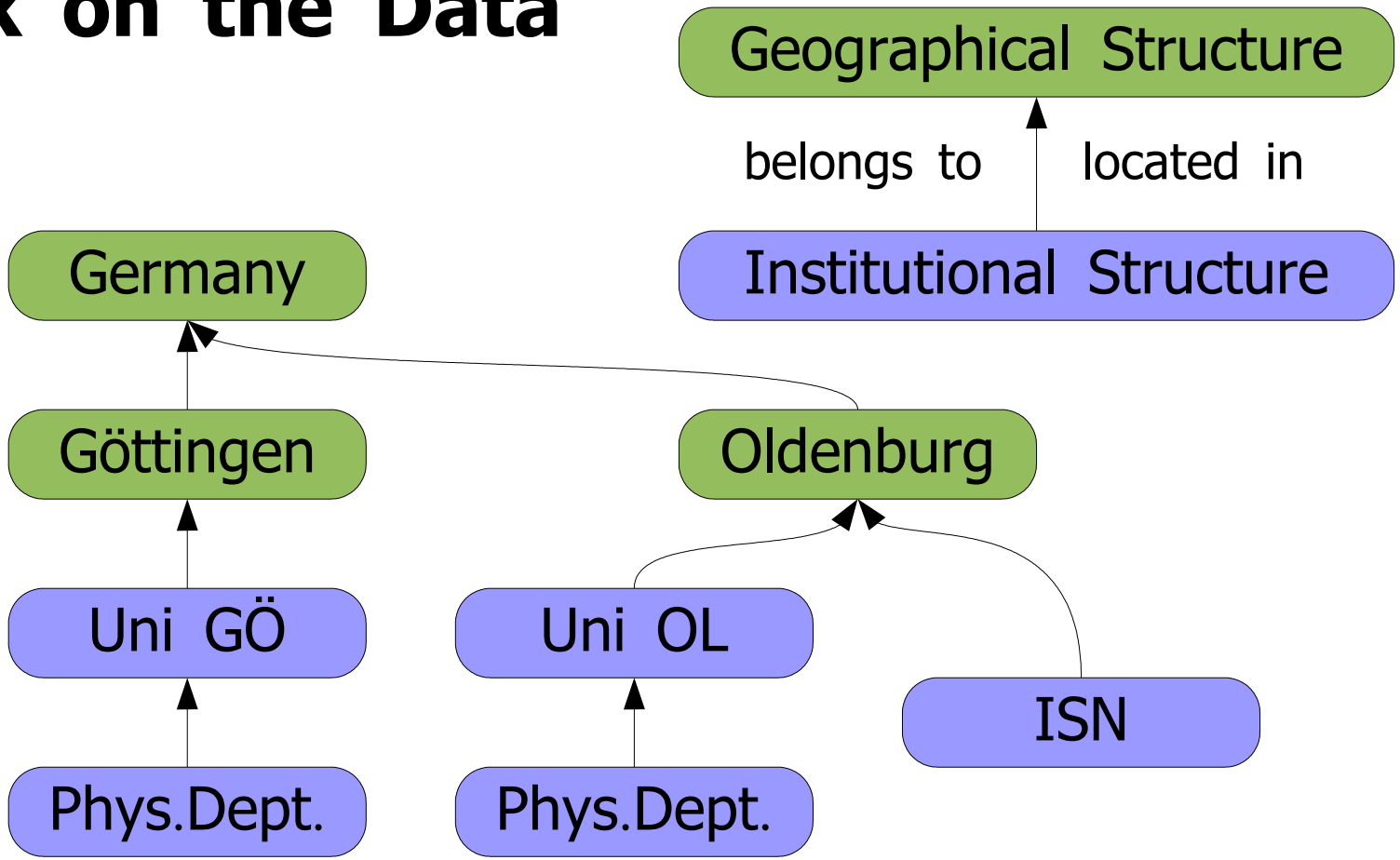
- Storage of pure data without any visualization defilement
- Visualization scripts may produce different views on the same data (PhysDep, PhysDoc)
- Visualization scripts take care of presenting all occurences of the same item (e.g. University of California) in the same style
- Additional information can easily be integrated by
  - telling the datamodel how to store the additional data
  - adding a rule to the scripts how to visualize the data

The PhysNet RDF Datamodel

The Datamodel

Why RDF/OWL?

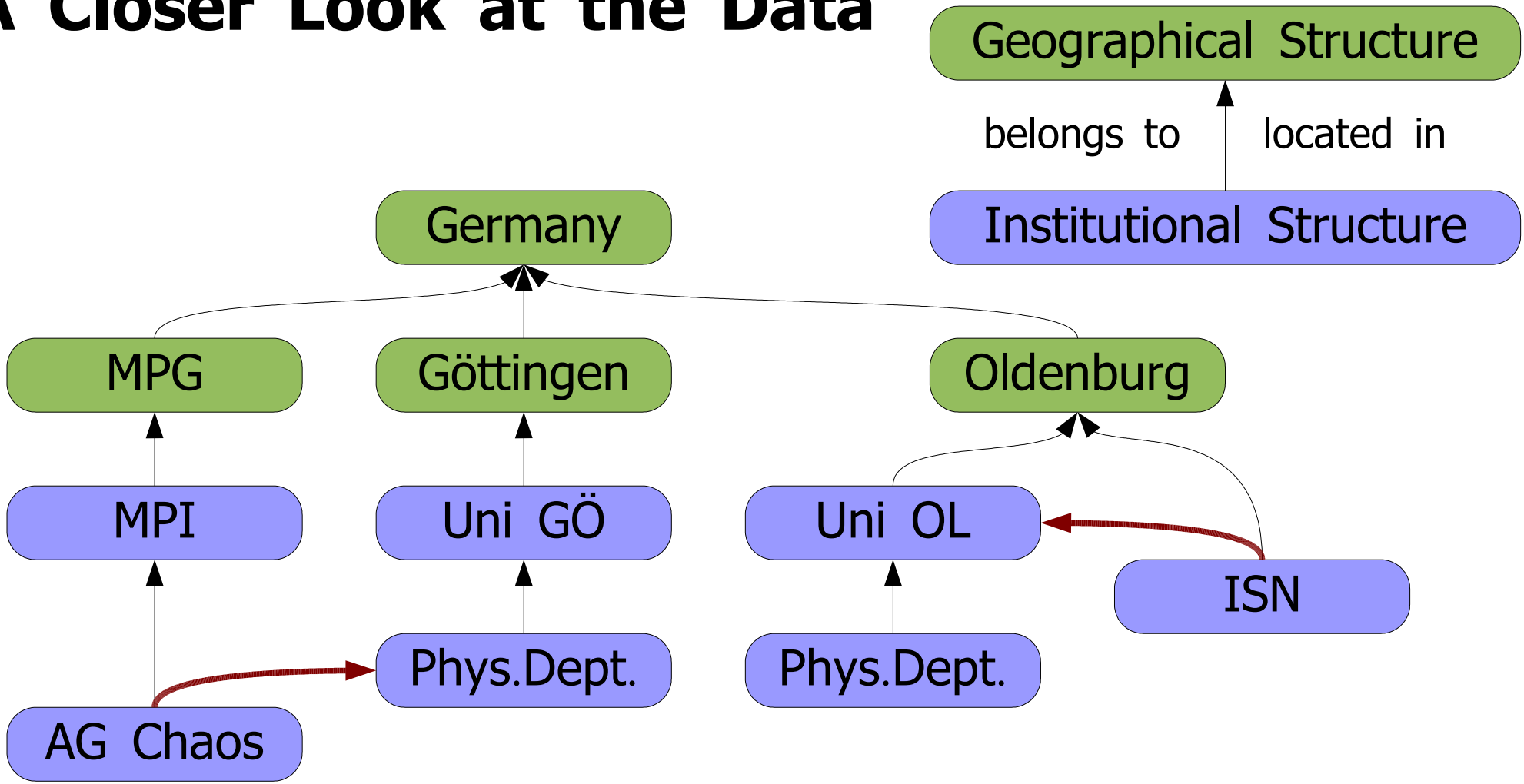
A First Look on the Data



Tree structure <=> XML (EXtensible Markup Language) would do the job, but…

Why RDF/OWL?

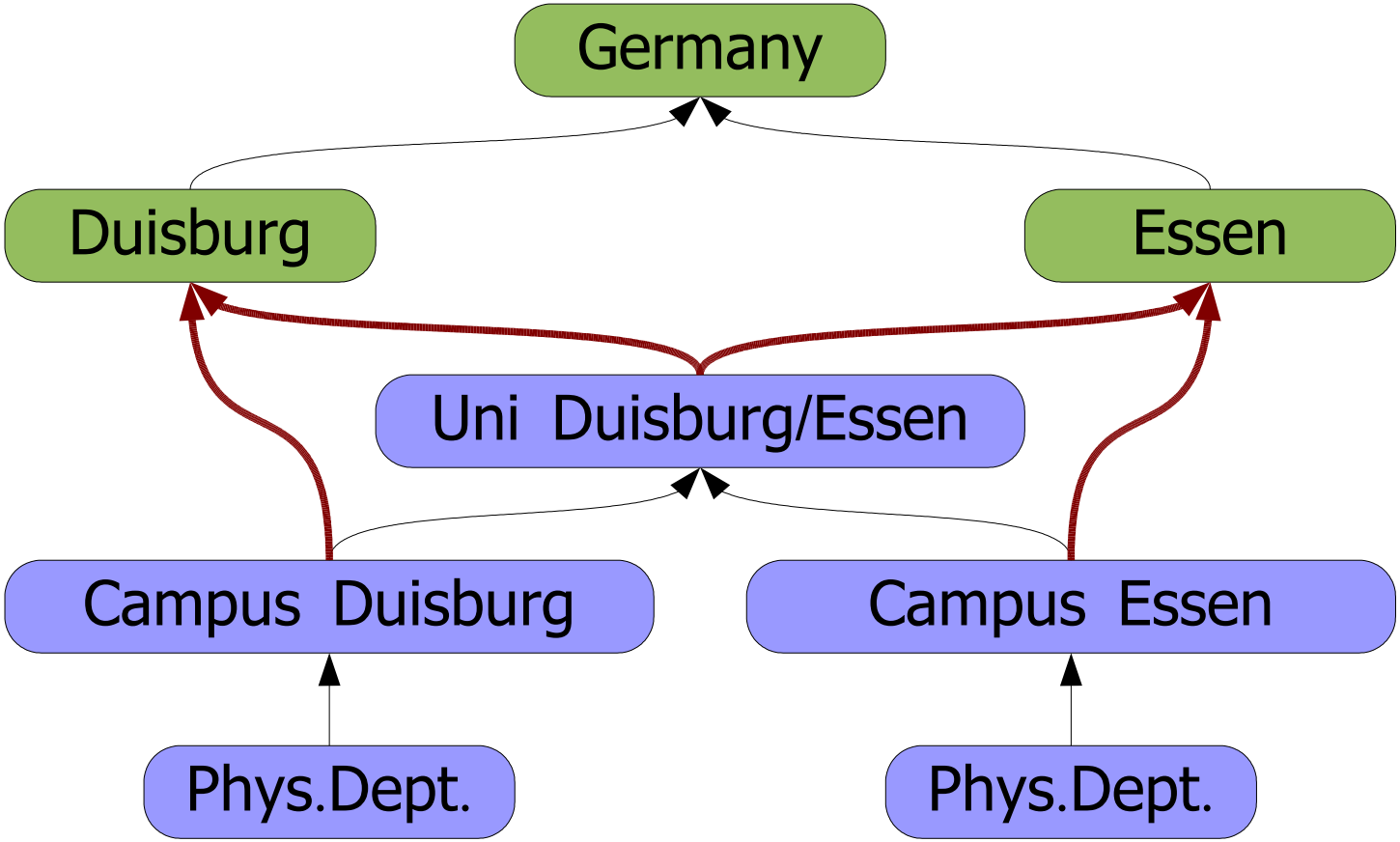
A Closer Look at the Data



Graph structure <=> RDF (Resource Description Framework)

Why RDF/OWL?

A Closer Look at the Data



Graph structure <=> RDF (Resource Description Framework)

Why RDF/OWL?

RDF – Resource Description Framework

- Stores data in a graph
- Offers great flexibility for future data enrichment

RDFS – RDF Schema

- Describes data stored in RDF format
- Offers a way to describe what type of objects are stored in the data and how they relate to each other

Why RDF/OWL?

OWL Lite – Web Ontology Language

- Describes data stored in RDF format
- Extends RDFS by some capabilities we need

Entities

Geographical Objects

- Continents, Countries, States, Cities

Institutional Objects

- Institutions with explicit location information
- Institutions with no special location
  (located in the same location as their parent inst.)

Other Objects

- Physical Societies
- (Physical Journals)

Information on Entities

Web Information

- Content type (homepage / publication list)
- URL, Language, Title

Contact Information

- Postal address
- Electronic contacts (phone, fax, e-mail)

Other Information

- Chairman of a society
- (Publisher of a journal)

Connections between Entities

Geographical Structure

- One GeoObj is part of another (Germany in Europe)
- An Inst is located in GeoObj (Uni OL in Oldenburg)

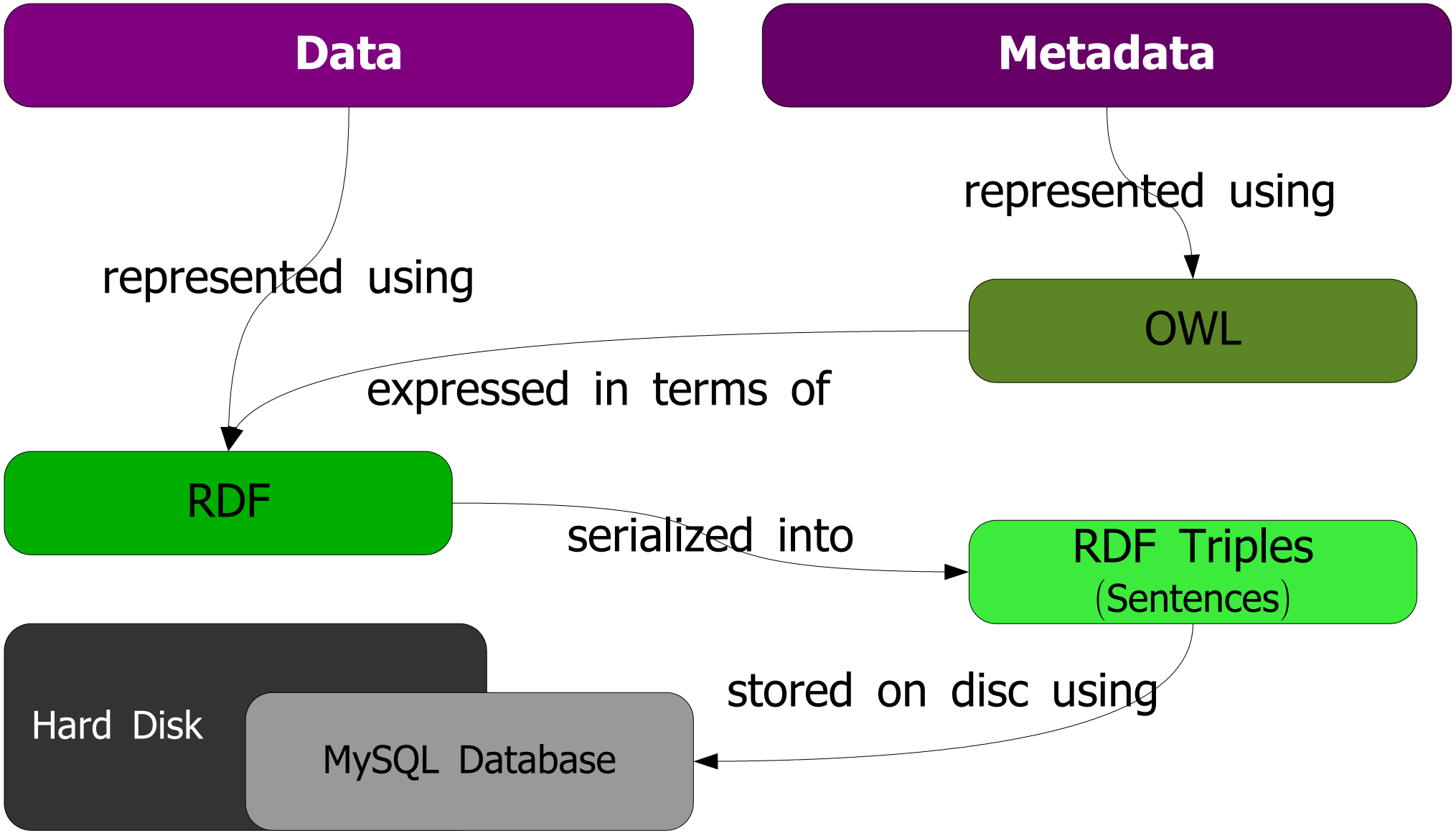
Institutional Structure

- An Inst is sub-inst of another (Phys.Dept. at Uni OL)
- An Inst belongs to a society (MPI to MPG)

The PhysNet RDF Datamodel

Datamodel Implementation

Data Storage

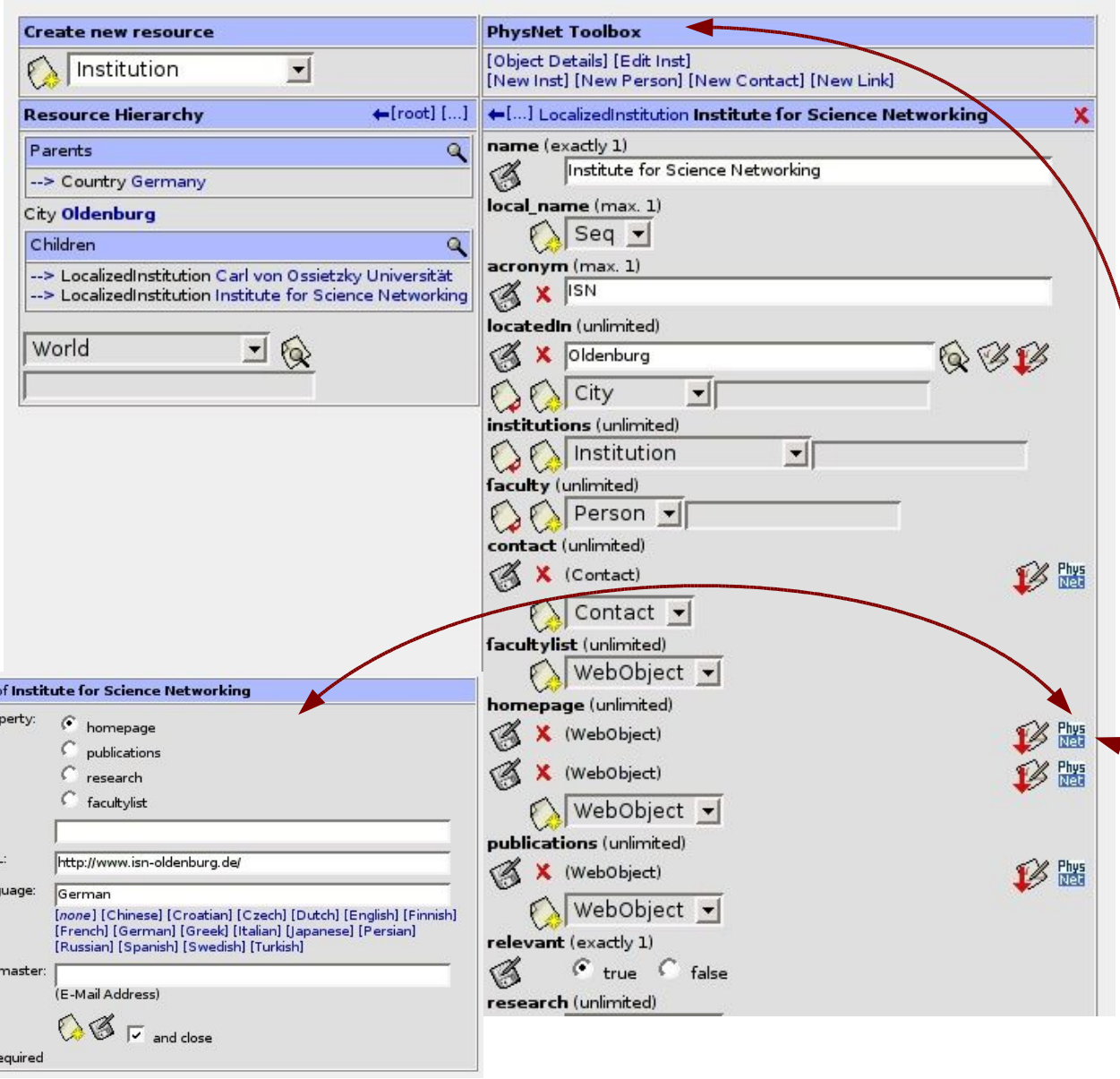


Data Storage

Why RDF/MySQL instead of RDF/XML

- RDF in a MySQL table needs a lot of disc space but XML-representation of RDF is gigantic
- MySQL offers a fast way to access the data while XML needs to be parsed
- MySQL offers better access methods to the data (e.g. searches on the data using SQL queries)

RDFEditor



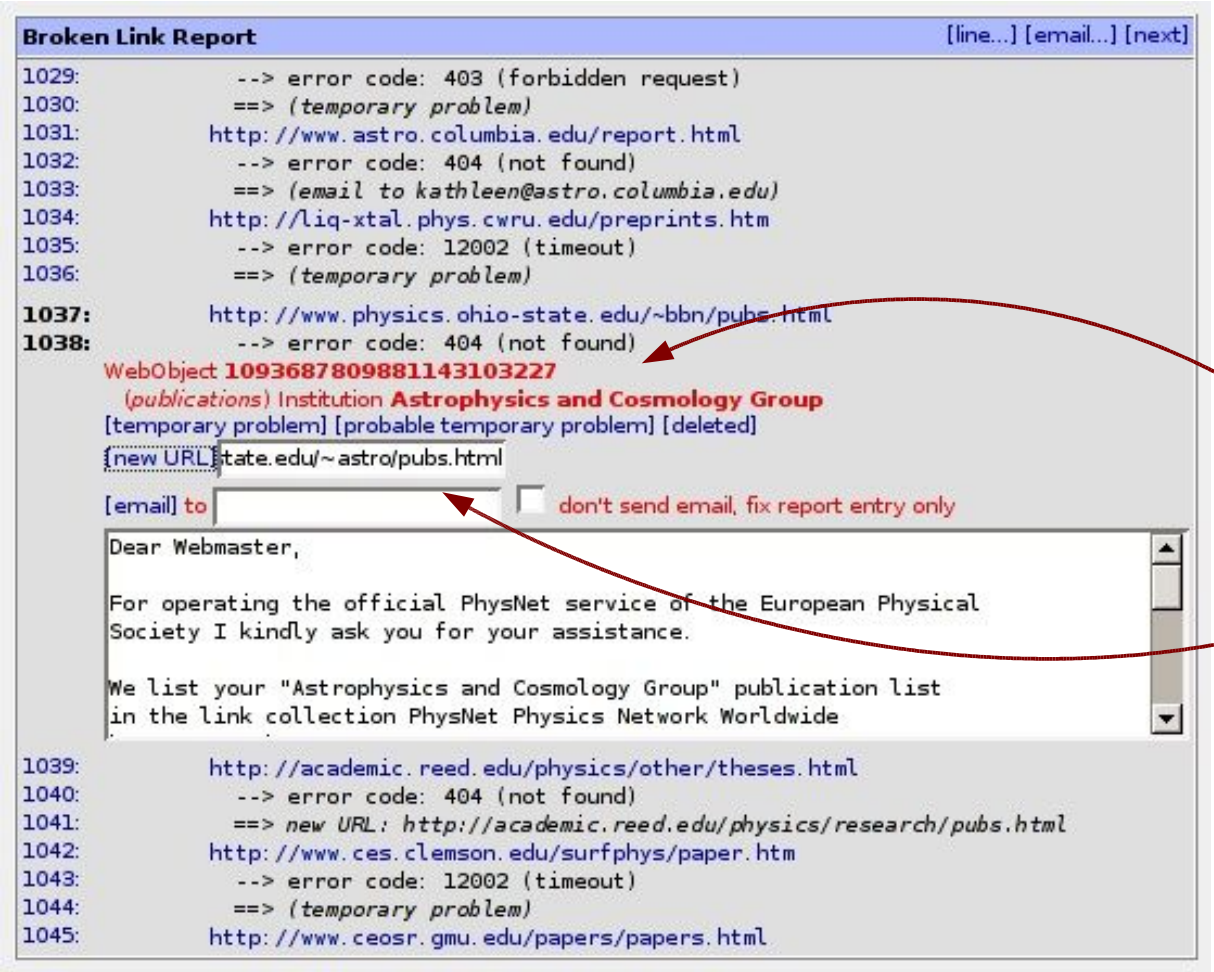
- A web-based tool for editing RDF datamodels
- Special optimizations for RDF stored in MySQL databases (could be extended to other data sources)
- A generic tool for editing any RDF data (RDFEditor adapts to any OWL-Lite-conform metadata)
- Plug-ins simplify recurrent tasks (PhysNet Toolbox)

Visualization Script

RDF2HTML

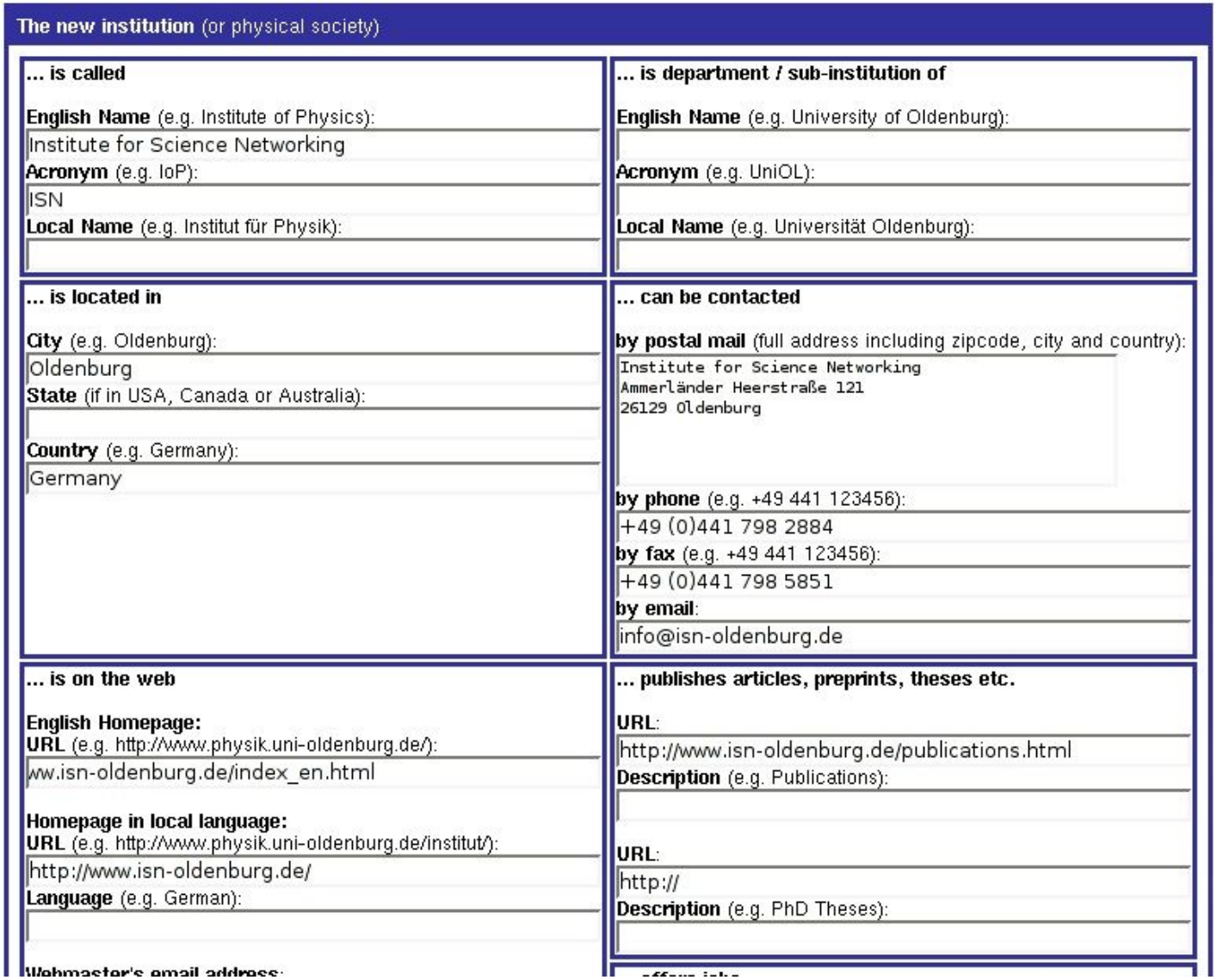
- Is run every night to update the HTML files from the RDF datamodel
- Collects the whole datamodel, analyzes the structure and simultaneously creates four views on the data:
  - PhysDep showing links to homepages of physics-related institutions
  - Societies (as part of PhysDep) showing only links to homepages of physical societies
  - PhysDoc showing links to publications lists of physics-related institutions
  - Details Pages for each institution and society showing the additional information (contact information etc.)

Broken Link Report Plug-In



- RDFEditor plug-in helping to process broken link reports
- Finds the broken link in the RDF data
- Offers several solutions (including ready-to-submit email template)

Example Workflow



The new institution (or physical society)

... is called

English Name (e.g. Institute of Physics):
Institute for Science Networking
Acronym (e.g. IoP):
ISN
Local Name (e.g. Institut für Physik):

... is department / sub-institution of

English Name (e.g. University of Oldenburg):
Acronym (e.g. UniOL):
Local Name (e.g. Universität Oldenburg):

... is located in

City (e.g. Oldenburg):
Oldenburg
State (if in USA, Canada or Australia):
Country (e.g. Germany):
Germany

... can be contacted

by postal mail (full address including zipcode, city and country):
Institute for Science Networking
Ammerländer Heerstraße 121
26129 Oldenburg
by phone (e.g. +49 441 123456):
+49 (0)441 798 2884
by fax (e.g. +49 441 123456):
+49 (0)441 798 5851
by email:
info@isn-oldenburg.de

... is on the web

English Homepage:
URL (e.g. http://www.physik.uni-oldenburg.de/):
ww.isn-oldenburg.de/index_en.html
Homepage in local language:
URL (e.g. http://www.physik.uni-oldenburg.de/institut/):
http://www.isn-oldenburg.de/
Language (e.g. German):
Webmaster's email address:

... publishes articles, preprints, theses etc.

URL:
http://www.isn-oldenburg.de/publications.html
Description (e.g. Publications):
URL:
http://
Description (e.g. PhD Theses):

- Someone submits information using the PhysNet upload formular

http://www.physnet.de/PhysNet/upload.html

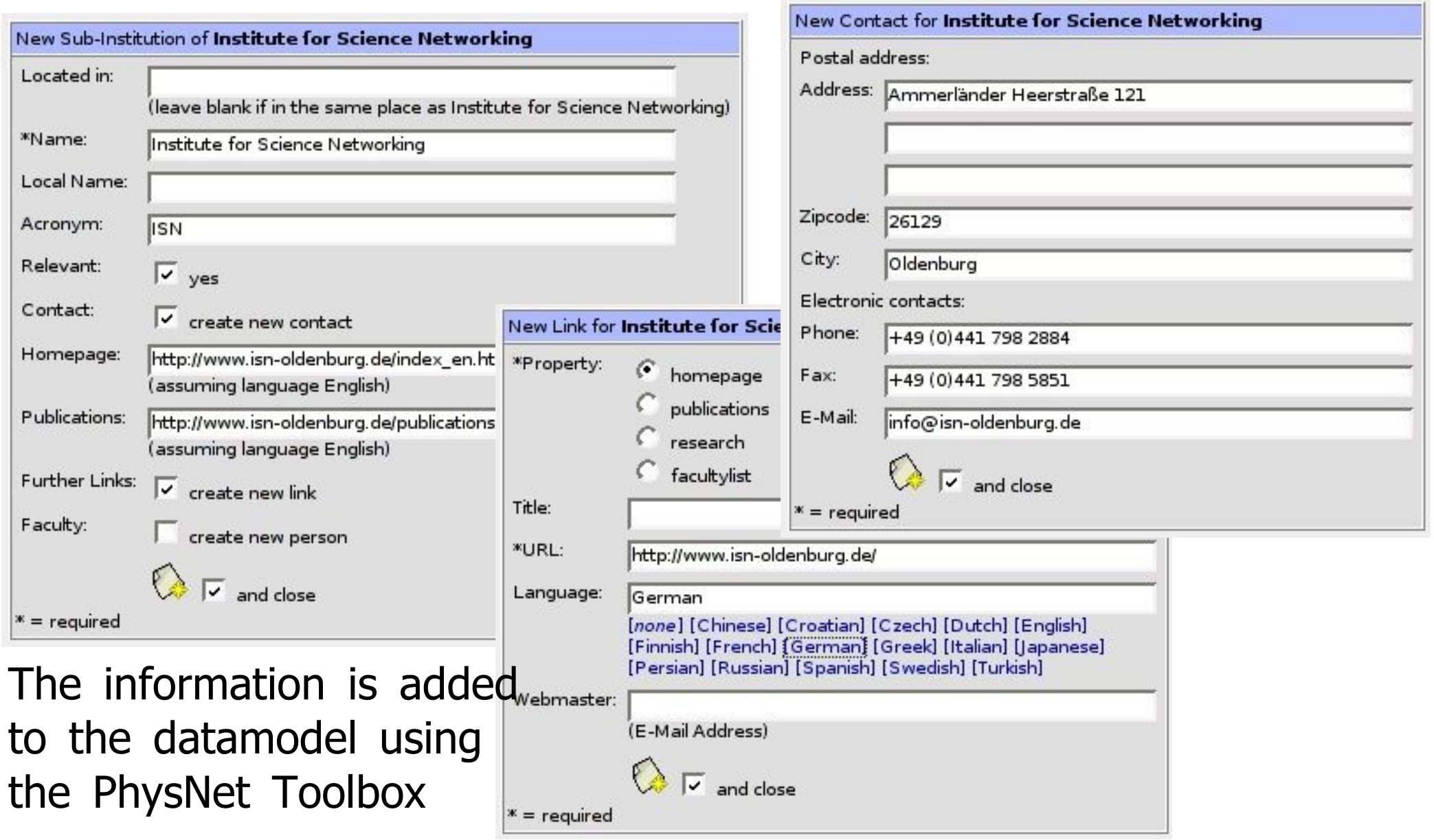
Example Workflow

```
Date: Tue, 11 Jan 2005 23:28:42 +0100
From: WWW daemon apache <wwwrun@physnet.physik.uni-oldenburg.de>
To: physnet@isn-oldenburg.de
Subject: PhysNet

--- 2005-01-11 23:28:42 ---
[form]                  newinst
[inst_name]             Institute for Science Networking
[inst_acro]             ISN
[inst_lname]
[pinst_name]
[pinst_acro]
[pinst_lname]
[inst_city]             Oldenburg
[inst_state]
[inst_country]          Germany
[inst_contact_postaddr] Institute for Science Networking
                        Ammerländer Heerstraße 121
                        26129 Oldenburg
[inst_contact_phone]    +49 (0)441 798 2884
[inst_contact_fax]      +49 (0)441 798 5851
[inst_contact_email]    info@isn-oldenburg.de
[inst_hpeng_url]        http://www.isn-oldenburg.de/index_en.html
[inst_hploc_url]        http://www.isn-oldenburg.de/
[inst_hploc_lang]
[inst_hp_webmaster]
[inst_pub1_url]         http://www.isn-oldenburg.de/publications.html
[inst_pub1_title]
[inst_pub2_url]         http://
[inst_pub2_title]
[inst_jobs_url]         http://
[comments]
[email]
```

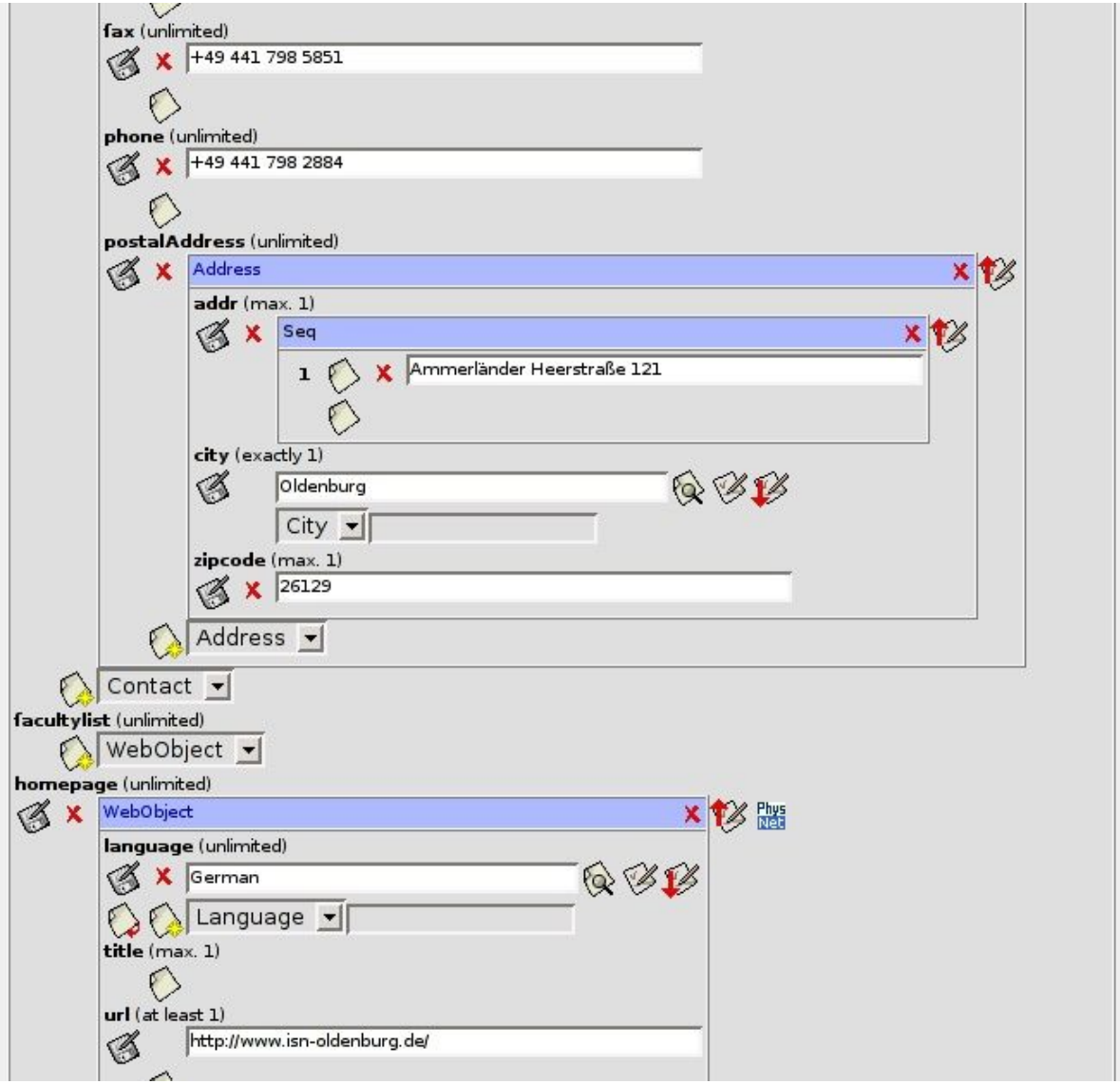
- The PhysNet server mails the information to the PhysNet crew

Example Workflow



The information is added to the datamodel using the PhysNet Toolbox

Example Workflow



- The information is now stored in the datamodel

Example Workflow

Schließen Drucken

Institute for Science Networking (ISN)

Kontakt

Postanschrift

Institute for Science Networking
Ammerländer Heerstraße 121
26129 Oldenburg
GERMANY

Telefon

+49 441 798 2884

Telefax

+49 441 798 5851

E-Mail

info@isn-oldenburg.de

Homepage

English

http://www.isn-oldenburg.de/index_en.html

German

http://www.isn-oldenburg.de/

Publikationen

http://www.isn-oldenburg.de/publications.html

- RDF2HTML updates all views to show the new data
  (only the Details view is shown on the left)

The PhysNet RDF Datamodel

Current work

Current work

Filling current datamodel

- Search for additional information (address, electronic contacts, multi-language websites) and insert into datamodel

Extending current datamodel

- Migrate PhysNet's Journals, Jobs and Conferences sections to the RDF datamodel
- Add new services (like Search in the datamodel)
- Add Geo-Position to Institutions (offer maps etc.)

Improving implementation

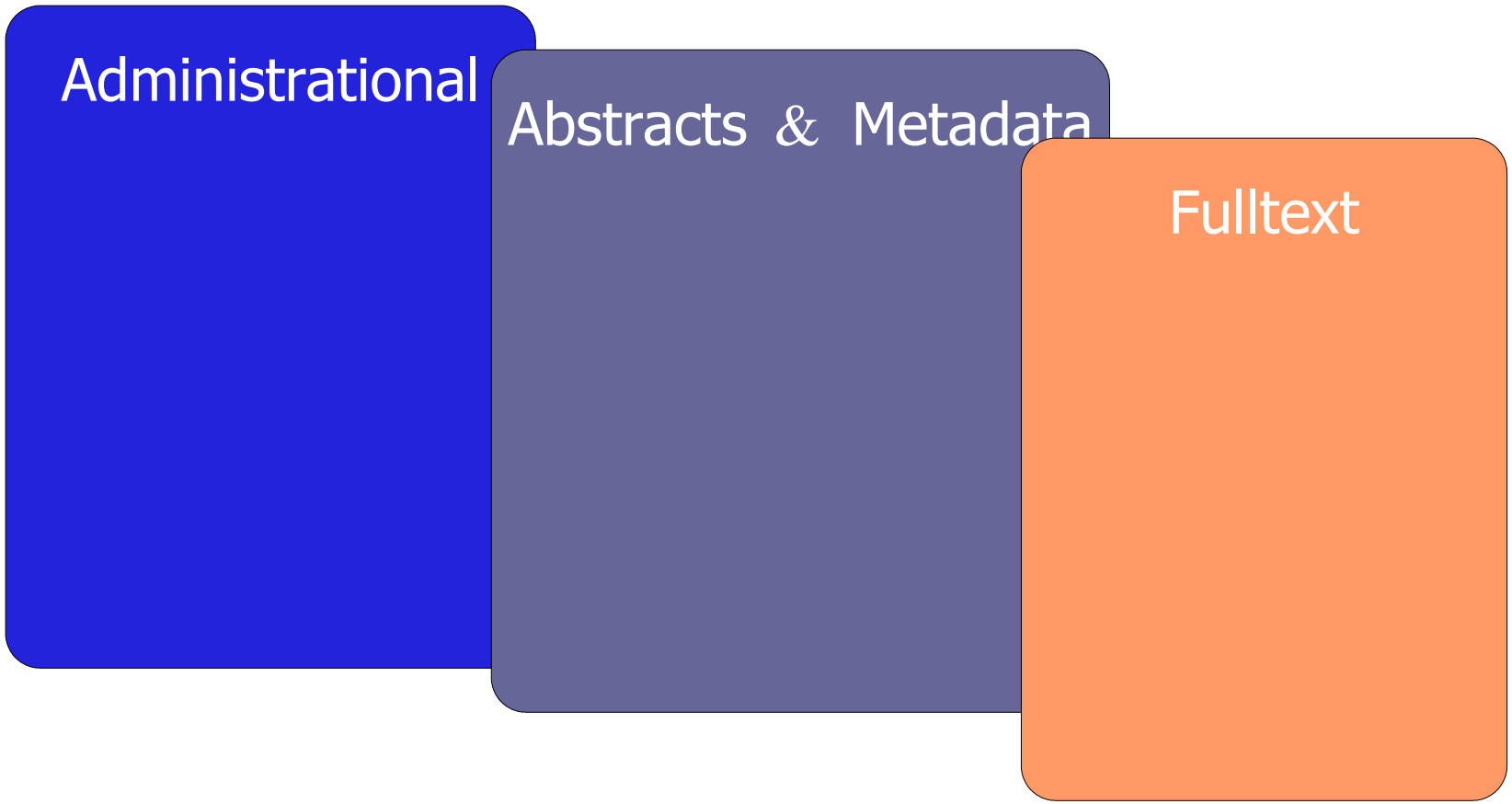
- Enhance RDFEditor and plug-ins

Open Access und OAi in PhysNet

Zugriff auf die Physik-Inhalte

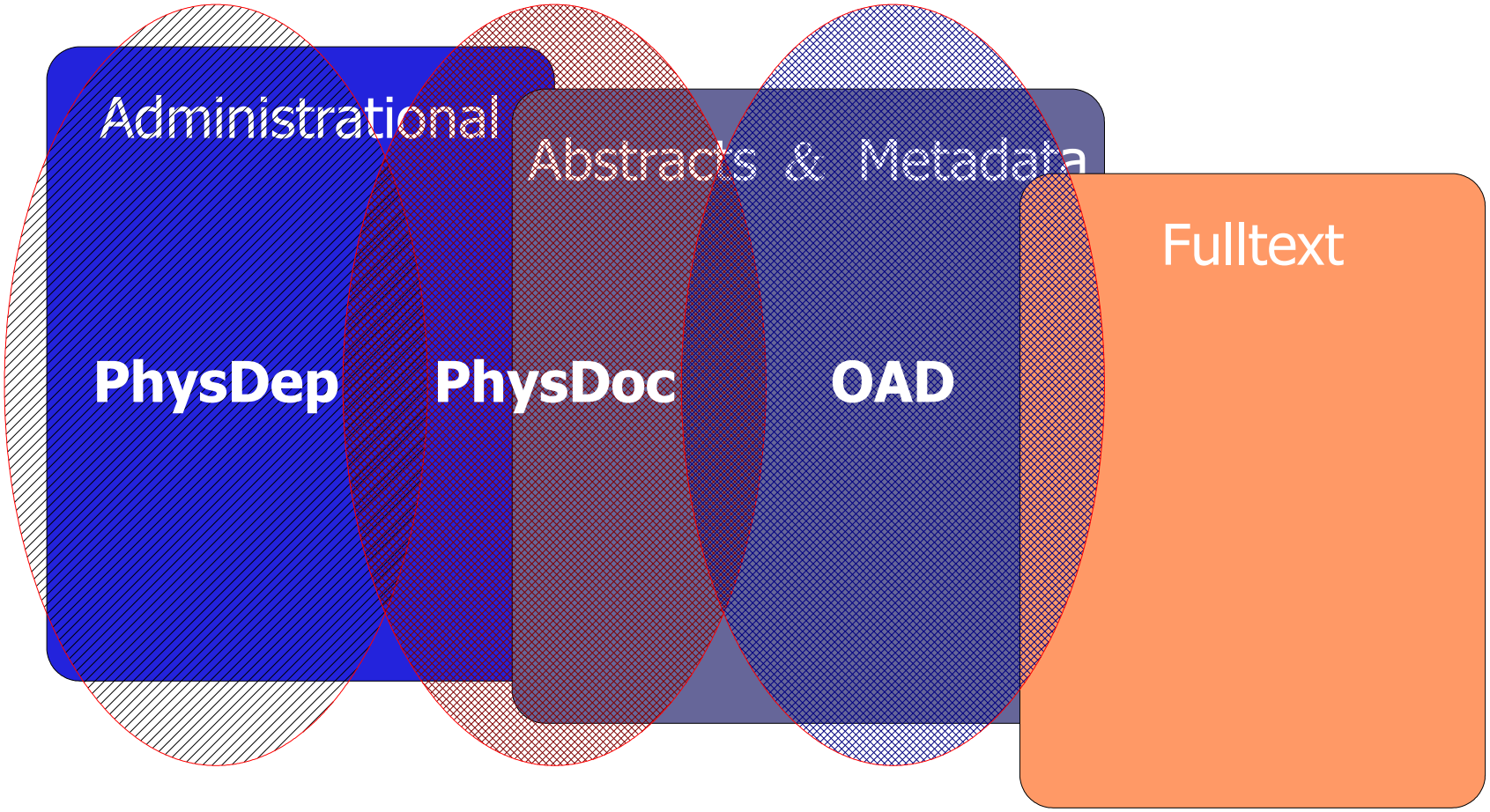
Open Access und OAi in PhysNet

Struktur-Überblick

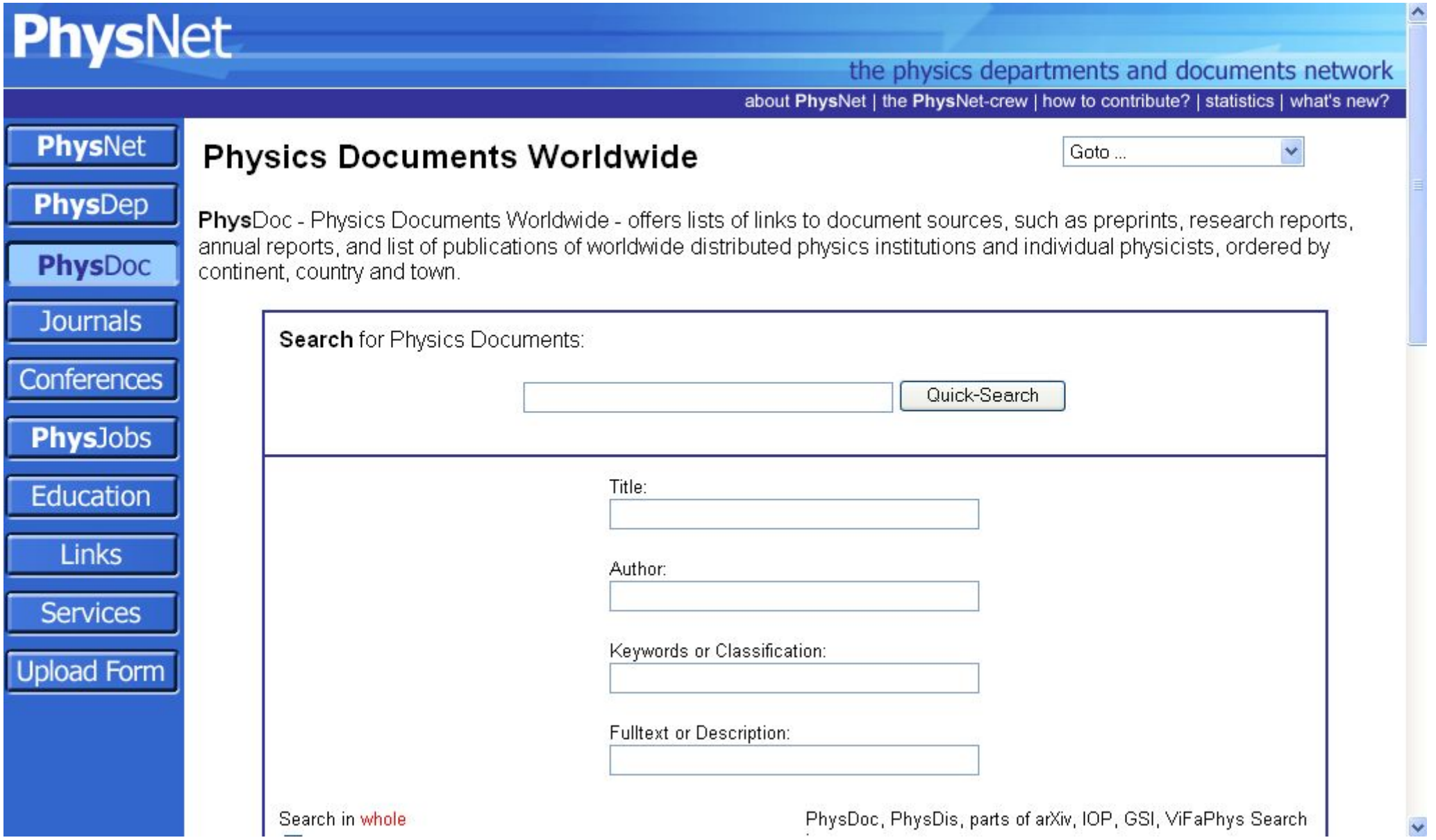


Open Access und OAi in PhysNet

Struktur-Überblick



OAD-Projekt


PhysNet
the physics departments and documents network
about PhysNet | the PhysNet-crew | how to contribute? | statistics | what's new?
PhysNet
PhysDep
PhysDoc
Journals
Conferences
PhysJobs
Education
Links
Services
Upload Form
Physics Documents Worldwide
Goto ...
PhysDoc - Physics Documents Worldwide - offers lists of links to document sources, such as preprints, research reports, annual reports, and list of publications of worldwide distributed physics institutions and individual physicists, ordered by continent, country and town.
Search for Physics Documents:
Quick-Search
Title:
Author:
Keywords or Classification:
Fulltext or Description:
Search in whole
PhysDoc, PhysDis, parts of arXiv, IOP, GSI, ViFaPhys Search

OAD-Projekt

148.393	citebase.eprints.org
146.078	www.hti.umich.edu
121.054	arXiv.org
80.804	memory.loc.gov
54.993	oai.dlib.indiana.edu
43.945	physdoc import
32.816	cdsweb.cern.ch
28.273	ProjectEuclid.org
27.215	hal.ccsd.cnrs.fr
15.292	images.indianahistory.org
14.820	www.mpi.nl
10.543	www.dlese.org
9.843	www.enc.org
9.367	pkp.ubc.ca
9.273	services.nsdl.org

Weitere mit weniger als jeweils 1% Beitrag zum Gesamtvolumen des Dienstes.

Gesamtzahl der Objekte: **918.605**

OAD stellt Software und diesen Dienst bereit, vernetzt die Dissertationssammlungen der Nationalen Initiativen (meist angesiedelt in den Nationalbibliotheken)

OAD-Projekt

Search interface to the PhysDoc OAI Harvester

Querying 983164 Articles

Title: cluster

Author:

Description:

Search | Simple Search

physnet.uni-oldenburg.de/oai/query.php

4379 Records found, 146 pages, showing articles 1 - 30
[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] › »

Rank	Source	Identifier	Date	Format	Author(s)
*****	arXiv	oai:arXiv:astro-ph/0012175 oai:arXiv:astro-ph/0012175	2000-12-07		Ebeling, H. Jones, L. R. Fairley, B. W. Perlman, E. Scharf, C. Horner, D.

Discovery of a very X-ray luminous galaxy cluster at z=0.89 in the WARPS survey

We report the discovery of the galaxy cluster ClJ1226.9+3332 in the Wide Angle ROSAT Pointed Survey (WARPS). At z=0.888 and L_X=1.1e45 erg/s (0.5-2.0 keV, h_0=0.5) ClJ1226.9+3332 is the most distant X-ray luminous cluster currently known. The mere existence of this system represents a huge problem for Omega_0=1 world models. At the modest (off-axis) resolution of the ROSAT PSPC observation in which the system was detected, ClJ1226.9+3332 appears relaxed; an off-axis HRI observation confirms this impression and rules out significant contamination from point sources. However, in moderately deep optical images (R and I band) the cluster exhibits signs of substructure

Kooperations-Schnittstellen

Kooperation PhysNet – XxxNet
(am Beispiel Math-Net)

Kooperations-Schnittstellen

- Organisatorische Schnittstellen
  - Netzwerk von Administratoren
  - Netzwerk von Spiegelservern
- Technische Schnittstellen
  - Gemeinsame Nutzung der geographischen und nicht fachspezifischen Struktur (RDF), Kopplung mittels XQuery oder WebServices.
  - Zugriff auf die Metadaten der verteilten Volltexte mit Oai2.0.
  - Editor und andere Tools

Kooperationsschritte

- Übergreifende Dokumentensuche
  - Gemeinsame Klassifikation notwendig, Crossconcordanz aus PASC bzw. MSC nach z.B. DDC vorhanden
  - Kopplung der Suchmaschinen an Metasuche via WebServices (vascoda-Modell), alternativ XQuery (gemäß SINN, erlaubt Lastverteilung etc.)
  - Alternativ Aufbau einer gemeinsamen Suchmaschine (skaliert aber nur unter Schwierigkeiten)

Kooperationsschritte

- Übergreifende Browsing Strukturen
  - Gemeinsame Klassifikation empfohlen
  - Darstellung der Institutionen entsprechend der Fachauswahl
  - Strukturierung alternativ nach Geographie oder Fachgebiet
  - Kopplung von RDF-Netzen mittels XQuery

Aufbau eines ScienceNet…

Erste Schritte zum Aufbau eines ScienceNet?

Zu einem ScienceNet?

- Einbindung möglichst vieler Fächer
- Aufbau einer Basis-Infrastruktur
  - Übergreifende Klassifikation
  - Gemeinsame Metadaten
  - Sharing von Files und Nutzerverwaltung (Grid?)
  - PlugIn Schnittstelle für Services
  - …