Marcin Roszkowski

Integration of Polish National Bibliography within the repository platform for science and humanities

The best thing to do to your data will be thought of by somebody else

W3C LLD

Agenda

- Polish National Bibliography (PNB) in a nutshell

- SyNat Project

- Issues of integration

  - e-resources and PNB / indexing online resources

  - metadata interoperability and reusability

The Polish National Bibliography

PNB

- Several bibliographic datasets
- Differentiation : type of resource
- Available free of charge (for non-commercial purposes)
  - PDF files
  - Web application
    - http://mak.bn.org.pl/w10.htm
  - Z39.50 protocol

PNB online

The Bibliographic Guide (Przewodnik Bibliograficzny)

- – monographs,
- – online : 1973 -
- – over 650 000 records

The Index to Periodicals (Bibliografia Zawartości Czasopism)

- – selective bibliography,
- – papers from major polish journals (over 2000 titles),
- – online 1996 –
- – over 750 000 records

PNB online

- Bibliography of New, Ceased and Changed Titles Serials
- Bibliography of Serials
- Bibliography of Electronic Documents
- Bibliography of Sound Recordings
- Bibliography of Cartographic Materials
- Foreign Polonica
  - works published abroad, written in Polish or by Poles or about Poland

Bibliography of Electronic Documents

- Coverage
  - works published since 2001
- Content:
  - monographs, periodicals
  - e-books
    - National Library of Poland E-Book Repository
  - online resources are excluded
- Cataloging
  - ISBD(ER)
  - MARC 21

PNB / cataloging

LDR &b n &c a &d m &e + &f + &g i &h +4500
001 &a b24684776
008 &a 100204s2010++++pl+a+++++++++!001+0+pol++
015 &a PB 13387/2011
020 &a 9788301160937
040 &a WA N &c WA N
041 1. &a pol &h eng
080 &a 159.944.4
080 &a 159.91
100 1. &a Sapolsky, Robert M.
245 10 &a Dlaczego zebry nie mają wrzodów? : &b psychofizjologia stresu / &c Robert M. Sapolsky ; [przeł. Małgorzata Guzowska-Dąbrowska].
246 1. &i Tyt. oryg.: &a Why zebras don't get ulcers? : &b a guide to stress, stress related diseases, and coping, &f 1994
260 &a Warszawa : &b Wydawnictwo Naukowe PWN, &c 2010.
300 &a 448, [3] s. : &b il. ; &c 24 cm.
490 1. &a Biblioteka Psychologii Współczesnej
504 &a Indeksy.
650 .4 &a Stres &x fizjologia
650 .4 &a Stres &x psychologia
700 1. &a Guzowska-Dąbrowska, Małgorzata. &e Tł.
830 .0 &a Biblioteka Psychologii Współczesnej
920 &a 978-83-01-16093-7

PNB / access points

synat
system nauki i techniki

INFINITI
PASSIM

NATIONAL LIBRARY
OF POLAND

SyNaT - System for Science and Technology

- Three-year project (2010-2013) sponsored by National Center for Research and Development
- Goals:
  - to create national system for information science
  - to create a universal, open, repositorial, hosting and communications platform for network stores of knowledge for science, education and open society of knowledge
- Participants: network of 16 institutions
  - universities, libraries, research centers

SyNaT - System for Science and Technology

- Rich-content platform
  - Institutional repositories
  - Digital collections
  - Bibliographic databases
  - Datasets
- Journal hosting services
- Content management systems
  - Adding and managing content / User and institutional profiles
- Business and legal models

SyNaT - System for Science and Technology

- Toolkit:
  - Data processing & analysis
  - Semantic searching
  - NLP applications
  - Image / audio processing
- Tools for social networking

Two complementary parts / two homogeneous applications

Infinity

- Leader: University of Warsaw
- Coverage:
  - project management,
  - platform infrastructure,
  - multimedia indexing,
  - automatic information extraction,
  - ...

Passim

- Leader: Warsaw University of Technology
- Coverage:
  - bibliographic metadata,
  - digital library collections
  - interoperability,
  - business and legal models
  - ...

ArnetMiner

C. Borgman

ALIAS: C. L. Borgman, Christine L. Borgman, Christine Borgman

FOAF Follow

Position:	Professor Presidential Chair
Affiliation:	Department of Information Studies Graduate School of Education and Information Studies
Address:	235 GSE&IS Building Box 951520, UCLA Los Angeles, CA 90095-1520
Phone:	(310) 825-6164
Fax:	(310) 206-4460
Email:	borgman at gseis.ucla.edu

EDIT

[?]

H-index:	18	Uptrend:	0.09	Diversity:	0.64
#Papers:	54	Activity:	0.49	Sociability:	89.34
#Citations:	2099	Longevity:	33	More Statistics…	

Photo by Ed Swinden (07817 363 585)

EDIT

0

No bio information available.

EDIT BIO

Research Interest

Digital Libraries,digital library,Information Science   EDIT

🔴 **Hide Temporal Interests (Do you want to see the change of his/her research interests?)**

Period	Interests
2009-2010	virtual organization,semantic Web,Life Cycles
2007-2008	Digital Libraries,digital library,Data Quality
2005-2006	digital library,Exploratory Study,Scientific Data
2003-2004	digital library,Global Information Infrastructure,digital learning environment
2001-2002	Case Study,Digital Libraries,digital library

See Others:

Anne Gilliland-Swetland   Follow
Coauthor-Count: 6, H-index: 3

Gregory H. Leazer   Follow
Coauthor-Count: 6, H-index: 2

Bruce Croft   Follow
Coauthor-Count: 1, H-index: 50

Jillian C. Wallis   Follow
Coauthor-Count: 7, H-index: 5

Expertise:

Information Retrieval / Probabilistic Indexing (20)
Digital library / Information Access (18)
Software Engineering (3)
Mobile Robot / Hybrid Control (2)
Interactive Computer Systems / Knowledge Acquisition (2)
XML Data (1)

Conference:

JCDL (7),JASIS (7),SIGIR (6),ECDL (4),Inf. Process. Manage. (4),JASIST (3)
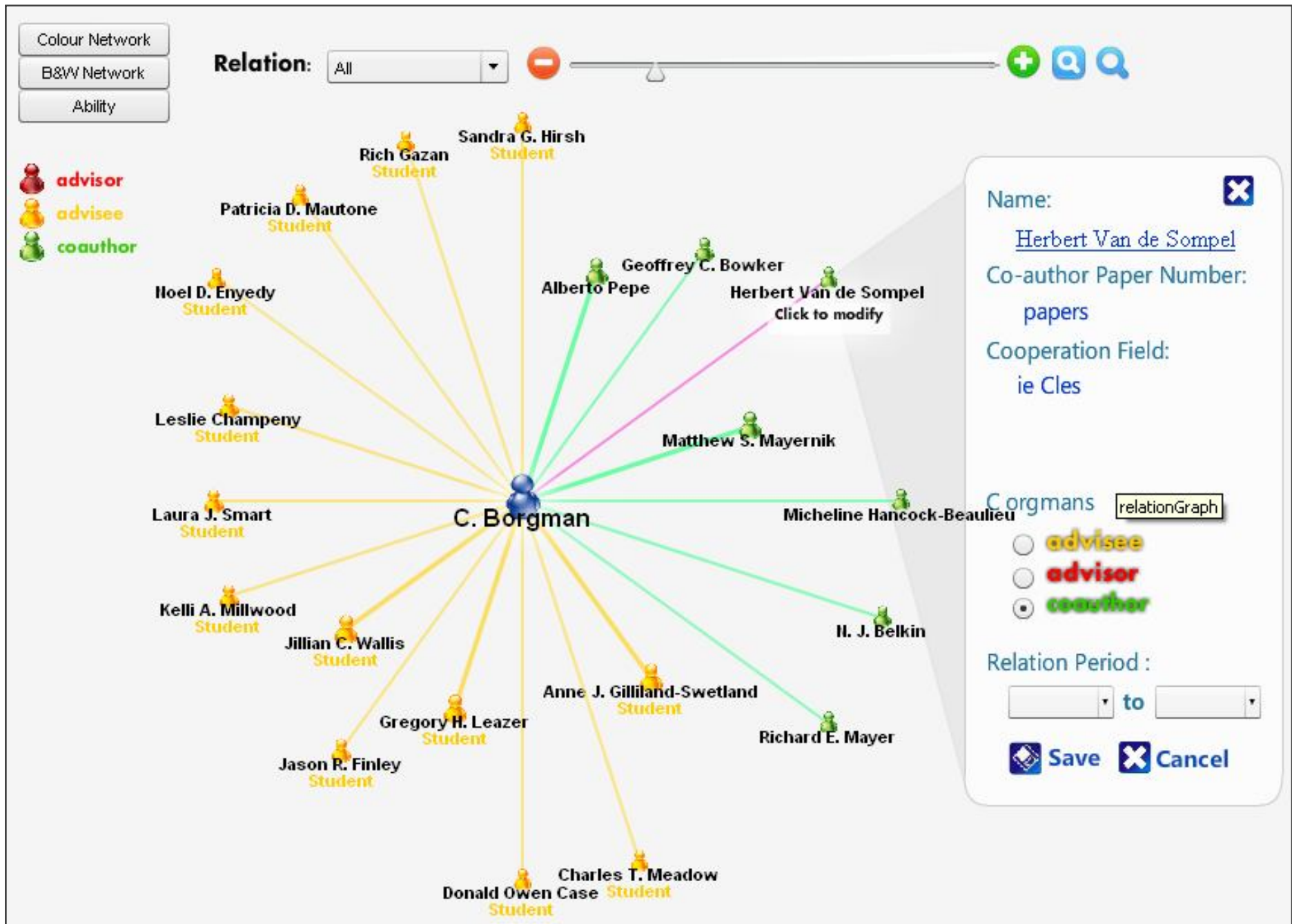
Arnet Miner Social Graph

PASSIM, e-resources and PNB

Web Archiving - As things stand

- Polish law: web documents are copyrighted
  - Exclusion:
    - Document published on public institution websites
    - Official (government, administration) online documents
- National Digital Archives
  - Internet Archive Project (2009-)
  - *.gov.pl / 41 institutional web services / 0,5 TB data

PNB and online resources

- Lack of specific regulation regarding web archiving and e-legal deposit

- PASSIM Deliverable B2-B4: Identification of high-quality web resources as an important type of content for users
  - A chance to develop methodology for indexing online resources for PNB

Indexing Online Resources

- Working Group:
  - Goal : to develop methodology for collecting and indexing web resources for PASSIM platform
    - acquisition, quality assessment, description.
  - Participants:
    - National Library of Poland, Warsaw University of Technology, Jagiellonian University

Indexing Online Resources

- Domain oriented methodology – manual indexing
    - Domains:
        » National Library of Poland:
        - social sciences, arts & humanities, business, economy
        » Warsaw University of Technology:
        - medicine, veterinary, mathematics, physics, astronomy, biology, life sciences, agriculture.
        » Jagiellonian University:
        - chemistry, engineering, informatics.

Indexing Online Resources

- Selection cirteria:
  - Development of a European Service for Information on Research and Education (DESIRE Project)
  - Intute
  - BazTol : Polish Technical Sciences Subject Gateway
  - National Science Digital Library
  - PANDORA : Australia's Web Archive
  - ...

Indexing Online Resources
/ access / internal

Indexing Online Resources
/ access / OPAC

Indexing Online Resources
/ access / Digital Library

PASSIM, e-resources and PNB

- Outcomes for PNB:
  - Functional requirements
  - Methodology for domain oriented indexing of web resources
  - Initial set of records (PNB – over 7000 records)
- Semi-automated approach seems to be inevitable

PASSIM – PNB

metadata interoperability

(ontology vs. MARC21)

PASSIM ontologies

- ## System ontology
  - Based on Semantic Web for Research Communities Ontology (SWRC)
  - Namespaces: BIBO, Dublin Core Terms, Semantic Web Conference Ontology, FOAF, RDF Schema, ...

- ## GIO
  - geographic information objects ontology

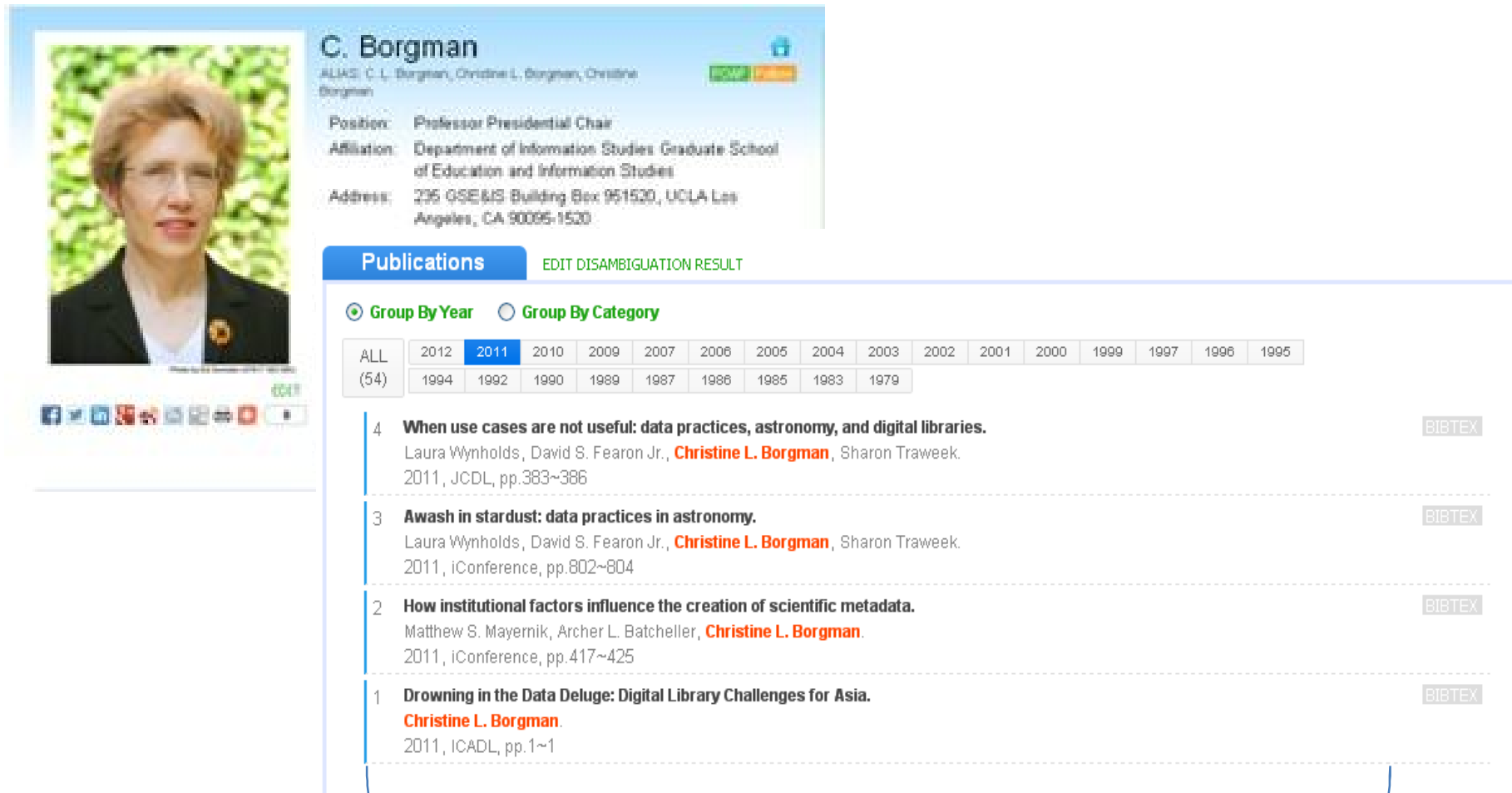- ## Subject areas ontology
  - areas / domains / disciplines

„From tags to triples"

MARC 21					RDF		
Label	Field	Indicator 1	Indicator 2	Subfield	Subject	Predicate	Object
Control number	1				foaf:Document	dc:identifier	xsd:string, xsd:long
Control number identifier	3				foaf:Document	dc:creator	foaf:Organization
ISBN	20			a	bibo:Document, frbr:Manifestation	bibo:isbn, bibo:isbn10, bibo:isbn13	xsd:string
ISSN	22			a	bibo:Periodical, frbr:Manifestation	bibo:issn	xsd:string
Cataloging source	40			a	foaf:Document	dc:creator	foaf:Organization
Language code	41			a	bibo:Document, frbr:Work	dc:language	lvont:Language
Country of publishing/producing entity code	44			a	bibo:Document, frbr:Manifestation		geo:Feature
Main entry – personal name	100			a	bibo:Document, frbr:Work	dc:creator	foaf:Person
Title statement – title	245			a	bibo:Document, frbr:Work	dc:title	xsd:string

http://marc-must-die.info/index.php/MARC_to_RDF_mapping

PASSIM – PNB
metadata reusability

Guidelines: Re-use of records includes downloading or export of batches of records into other databases and/or other computer applications

PASSIM / PNB / person / list of publications

Source : PNB

PASSIM / PNB / affiliation / list of publications
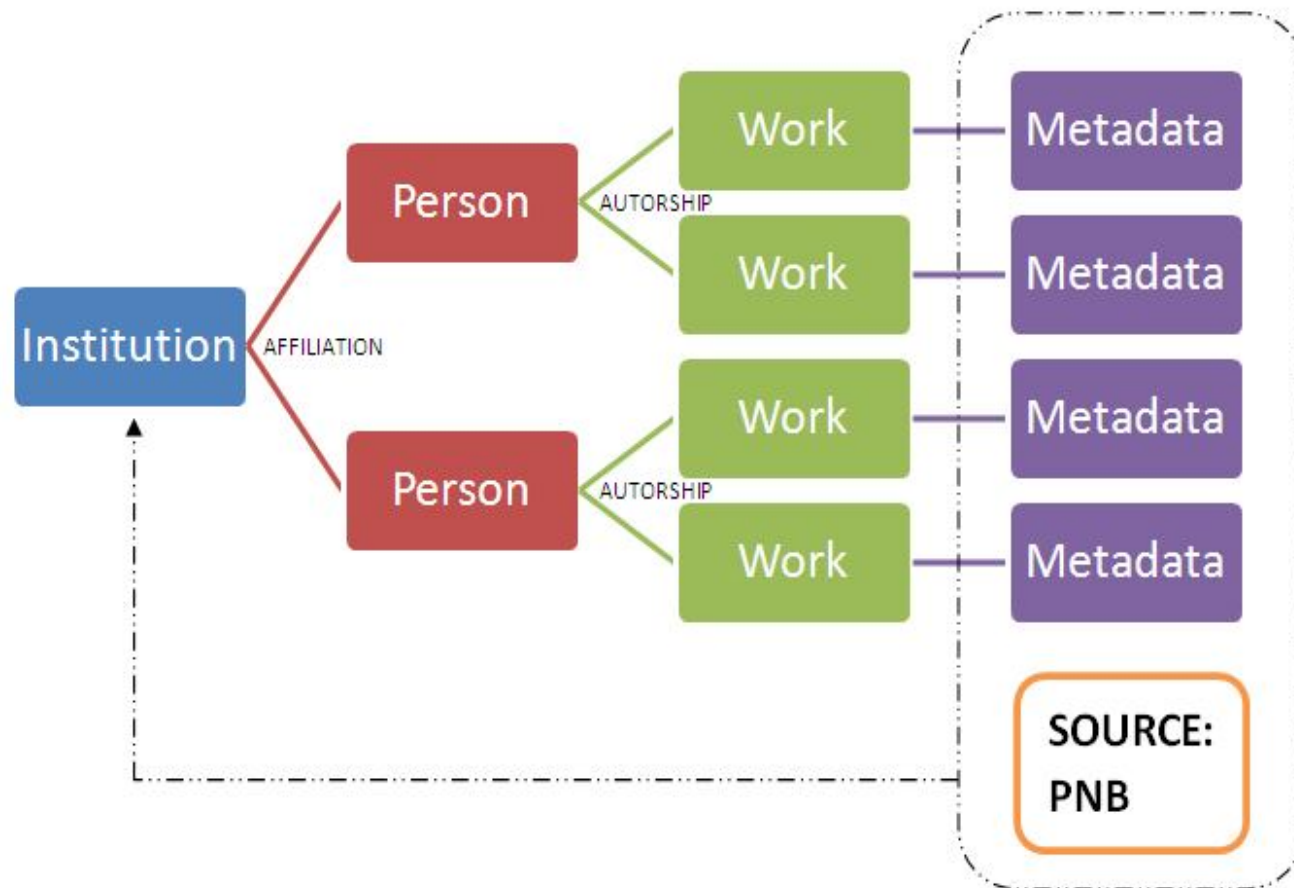
PASSIM / PNB / project / list of publications

Project

Publication 1

Publication 2

Publication 3

Publication 4

- MARC 21
  - 536 ##$a Sponsored by the Advanced Research Projects Agency through the Office of Naval Research
  - $bN00014-68-A-0245-0007
  - $cARPA Order No. 2616

- PASSIM ONTOLOGY
  - Class : Project
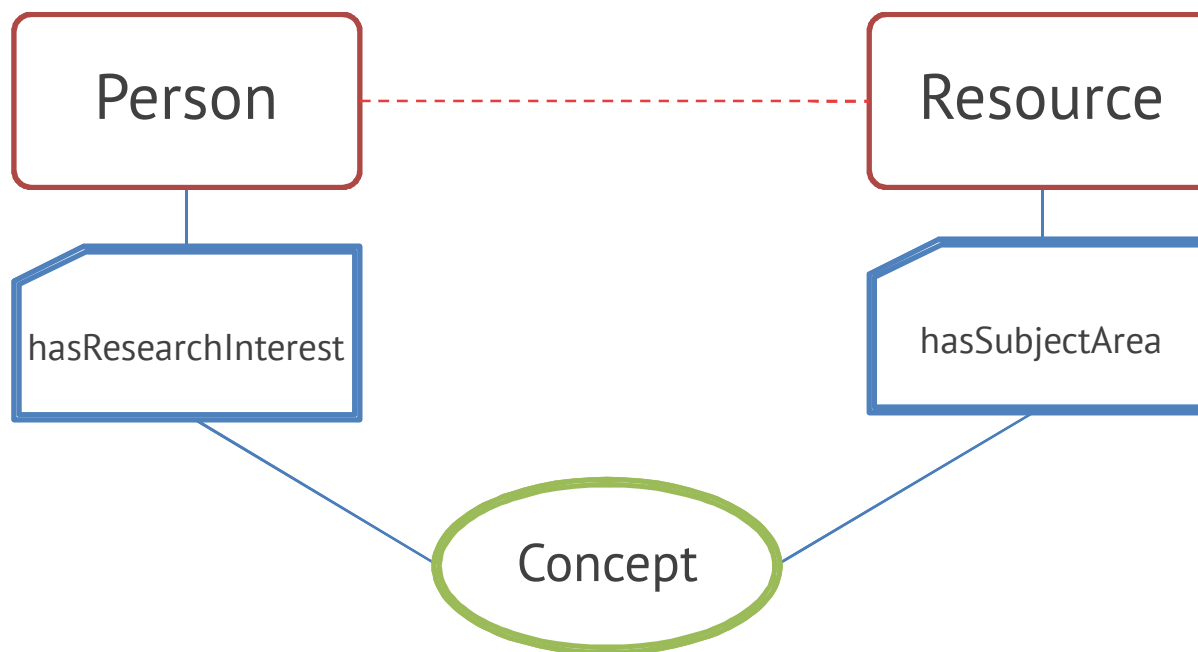 - Attribute: generatedPublication

PASSIM / PNB / personalization

- ❏ Basic Data
- ❏ Position
- ❏ Affiliation
- ❏ Autorship
- ❏ Role in scientific events
  - ❏ Project X participant
- ❏ Research Interest
  - ❏ hasResearchInterest1
  - ❏ hasResearchInterest2
- ❏ ...

PASSIM / PNB / personalization

PASSIM / PNB
/ selective distribution of information

- To select publications from PNB relevant to used search term
- To select the latest publications form PNB according to user research interest
- To import metadata from PNB according to user authorship / other contribution
- To create list of references on topic X...
- ...

Different modes of display / import

PASSIM / PNB / problems to solve

- From MARC21 to PASSIM Ontology
  - Mappings needed
- Implementation of authority files
  - What?
    - name authority file
    - subject headings
  - How?
    - RDF, SKOS, …
- Identifiers
  - PASSIM URI's vs. National Bibliography Number

NATIONAL LIBRARY OF POLAND

Marcin Roszkowski

Thank You

The best thing to do to your data will be thought of by somebody else

W3C LLD

Illustrations

- http://www.howitworksdaily.com/wp-content/uploads/2011/05/Nutshell-small.jpg
- http://www.flickr.com/photos/ivanwalsh/3649492427
- http://dynamicorange.com/uploads/Semantic%20Marcup.pdf

NATIONAL LIBRARY
OF POLAND