INFRASTRUCTURE FOR OPEN-SCIENCE

A.I. IN DIALOGUE SYSTEMS

Jakub Koperwas, CEO
Łukasz Kobyliński, CSO

sages

Professional IT Education

(since 2008)

IT Events

sages

IT Solutions for Research Management

Data analysis & machine learning 4 business

STACJA.IT

SigDelta

ai & nlp workshop day

kodo/amacz

Postgraduate Studies Programmes

- we are the co-organizers of
  Big Data technical postgraduate studies
  at the Warsaw University of Technology
- the Big Data for Management postgraduate studies
  at the Koźmiński University.
- and the Geostatistics postgraduate studies
  at the Cardinal Stefan Wyszyński University in Warsaw

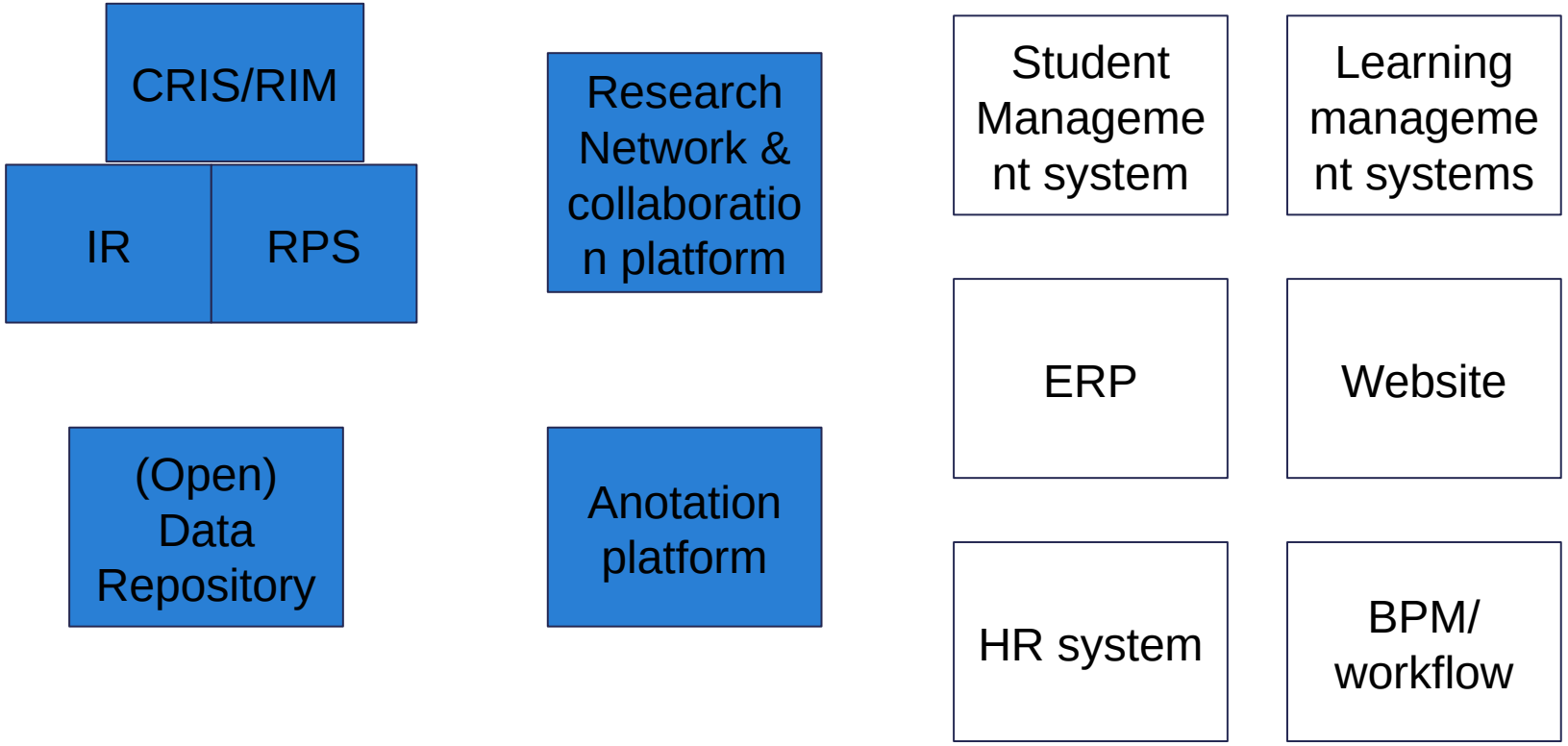
Warsaw University of Technology

UNIVERSITAS CARDINALIS STEPHANI WYSZYNSKI
SOLI DEO
UKSW
VARSOVIAE

KOZMINSKI UNIVERSITY

Infrastructure for Research Management

As a **service provider and contributor of Omega-PSIR CRIS system** we help organizations with optimizing their research infrastructure & Culture.



Products

Warsaw University of Technology

OMEGA-PSIR

IR+CRIS+RPS

PAN
POLISH ACADEMY of SCIENCES

KORPUSOMAT

sages

Labellery.

RIM Goals

Archivsation&Sharing

- Collect research outputs
- Both outputs & Matadata
- Promote Open Access
- Promote Open Data
- Increase productivity of your institution by internal-resource sharing
- Increase your visibility by more citations

Visibility

- Promote research being done
- Build a reputation
- Attract researchers
- Attract PhD students
- Attract funders

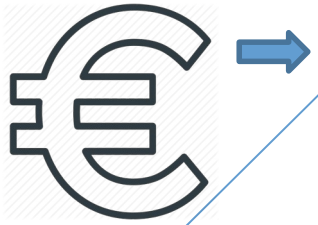
RIM Goals

Reporting

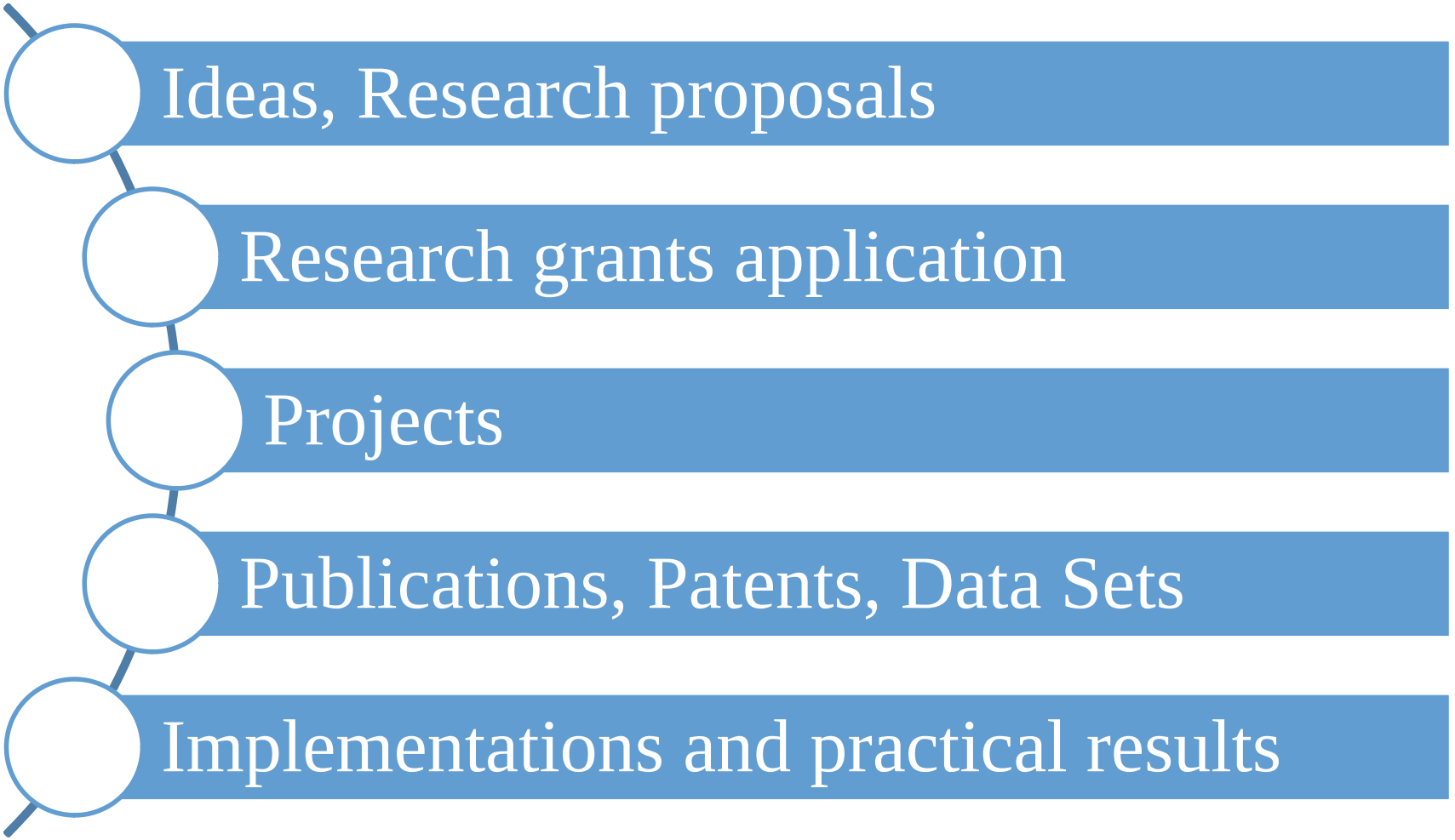
- Report to the Ministry and other Funders
- Distribute Funds into subunits
- Report for promotion
- Annual employee eveluation
- Project scoring

Utilize Research Potential

- Help to discover teams of expertise
- Help to share data internally
- Suggest journals, conferences and research grants
- Encourage and stimulate competing


Publication 1
Researcher 1
TV interview
Is an author
Publication 2
Is performer in
Is an author
Research project
is a result of
PhD Desideration
is a result of
Is a leader
Is a supervisor
Publication 3
Is an author
Researcher 2
Metrics

Complete research lifecycle



Selected types & functionalities

- Books
- Papers & chapters
- Reports
- Translations
- Engineer's / Bachelors theses
- Masters theses
- Postgraduate theses
- PhD theses

- Patents
- Products
- Projects
- Technologies
- Professional activity
- Professional achievements (career)
- Implementations and practical effects (patents, products)
- Architectural & artistic works
- Published journals
- Conferences organized by unit

- Affiliations
- Authors and employees
- Corporate authors and other institutions
- Conferences (events)
- Conferences (series)
- Journals and series
- Languages
- Countries, organizations (for patents)
- Area, domain, discipline, study subject
- Study subjects, specialization

+Infrastructure

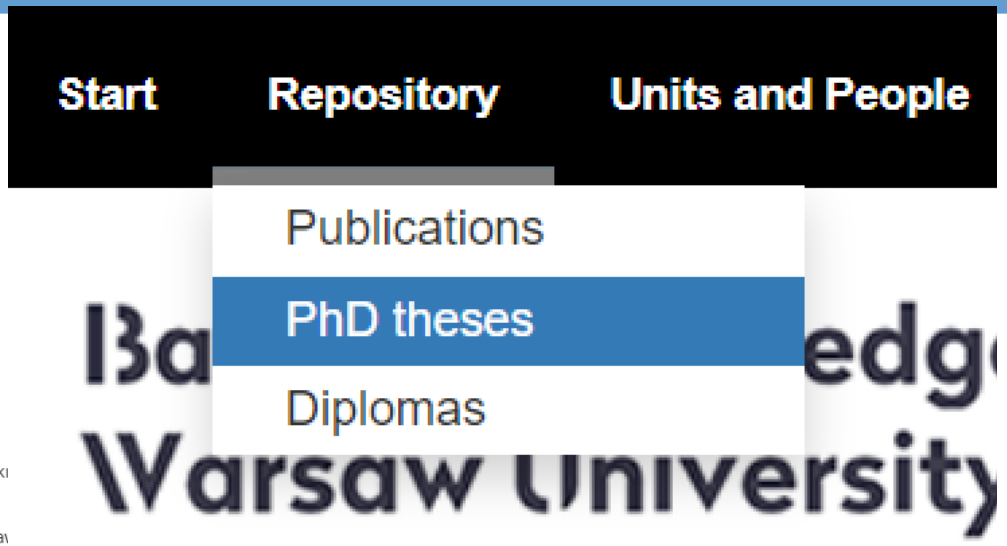
- Full texts
- Metadata

- Searching
- Sharing

- Reporting
- Evaluation

CERIF

Full texts



Rybiński Henryk: Problem optymalizacji reorganizowania zbioru informacyjnego w systemie wyszuki

Rybiński Henryk, Muraszkiewicz Mieczysław: TEST - Baza danych, 2016, Akademicka Oficyna Wyda

Edit

Fronf Matter.pdf	149 KB
PostISMIS_paper__JIIS (66).pdf	242,7 KB
JIIS-D-18-00014 (4).pdf	1,32 MB

blishing,

or Building a Scientific Information Platform, Studies in Computational Intelligence, vol. 390, 2012, Springer, ISBN 978-3-642-24808-5,

Bembenik Robert, Skonieczny Łukasz, Rybiński Henryk, Kryszkiewicz Marzena, Niezgódka Marek (*eds.*): Intelligent Tools for Building a Scientific Information Platform: Advanced Architectures and Solutions, Studies in Computational Intelligence, vol. 467, 2013, ISBN 978-3-642-35646-9, 548 p., DOI:10.1007/978-3-642-35647-6

Kryszkiewicz Marzena, Appice Annalisa, Ślęzak Dominik, Rybiński Henryk, Skowron Andrzej, Raś Zbigniew W (*eds.*): Foundations of Intelligent Systems, Lecture Notes In Computer Science, vol. 10352, 2017, Springer International Publishing, ISBN 978-3-319-60437-4, [978-3-319-60438-1], DOI:10.1007/978-3-319-60438-1

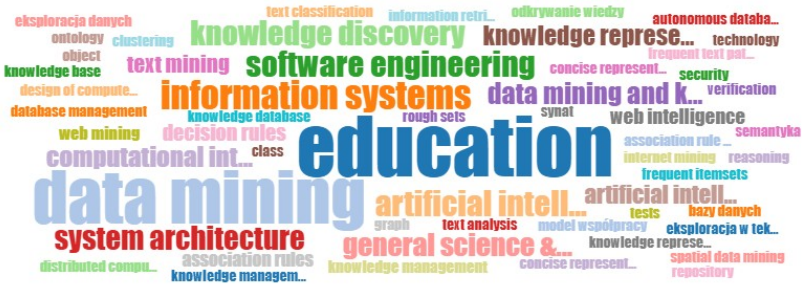
Data Sets

Profile | Publications | PhD | Research data | Projects | BSc and MSc | Activities | Achievements | Products | Citations | Statistics | Cooperation | Edit

Henryk Rybiński, PhD, DSc, Professor

Professor

The Institute of Computer Science
Faculty of Electronics and Information Technology
Phone: +48 22 234 7731
Room no: 304
Consultations: Monday 14.00-16.00



Research data

Number of records: 3.

Export 0 as: Editors pivot table

- Henryk Rybiński (FEIT / IN) Tabela wyników pomiarów rezystancji materiałów alfa The table of measuring results of the resistance of alpha materials [] Wed May 01 00:00:00 CEST 2019 Politechnika Warszawska
- Henryk Rybiński (FEIT / IN) Ontologia OSJ z Polskim językiem OSJ Ontology [] Wed May 01 00:00:00 CEST 2019 POLITECHNIKA WARSZAWSKA
- Henryk Rybiński (FEIT / IN) Dane tabelaryczne z pomiarów rezystancji materiałów \alpha tabular data from measuring resistance of \alpha materials [] Politechnika Warszawska

Gather

Complete, consistent Metadata

Classification of research output

Metrics

Full Texts, Research Data

Financial Data
HR Data

zotero

Reference tools

Export citation

Add to BIBTEX

Crossref

ORCID

automat

OECD

Scopus

THOMSON REUTERS
WEB OF KNOWLEDGE

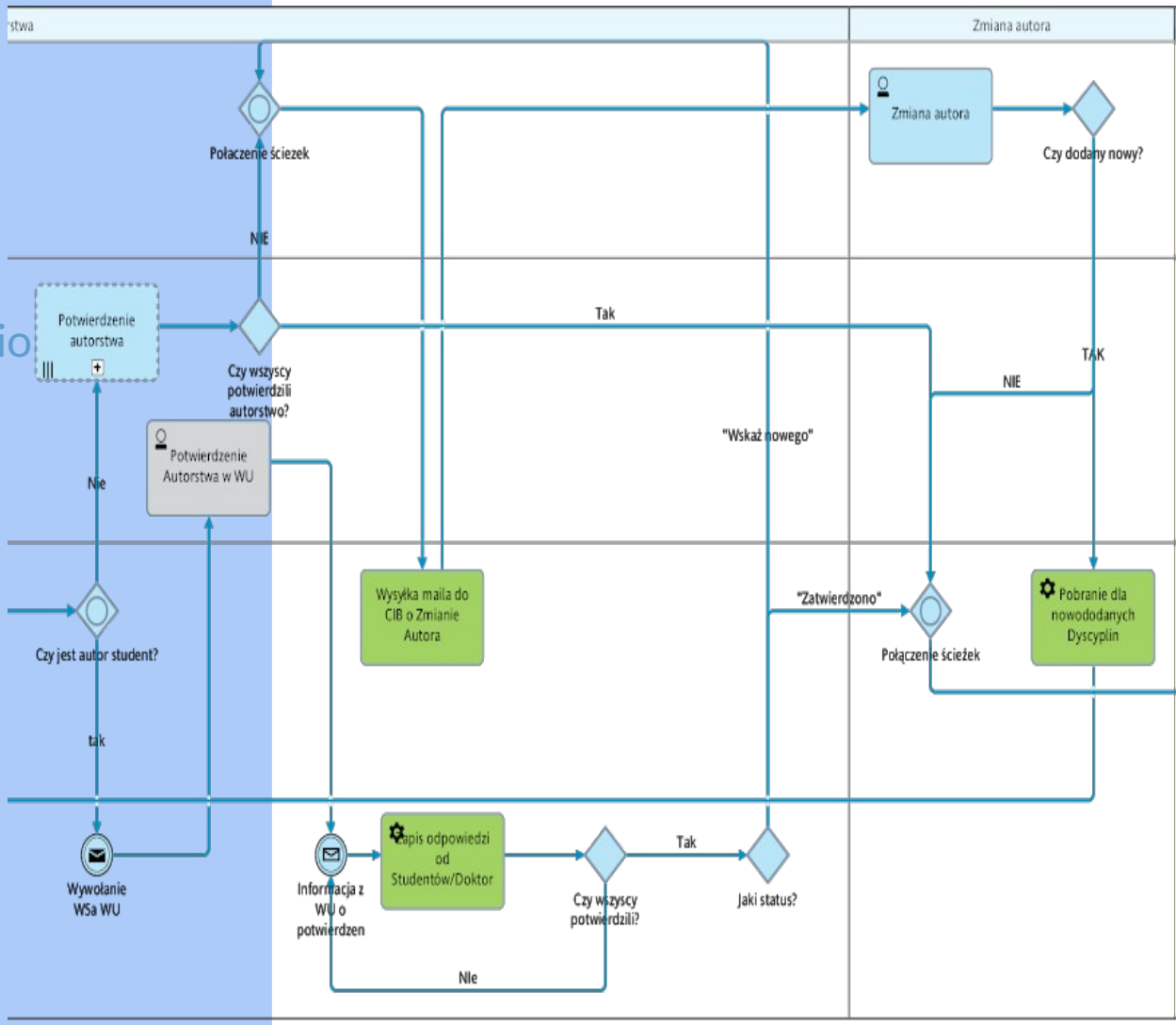
Google scholar

sharing

validation

Authorship validation and adjustments step

1. Every assigned author is required make decisions:
   a. confirm authorship
   b. assign disciplines
   c. confirm affiliation
   d. confirm parametrizatio
   e. confirm description
2. Authors responses are collected in Omega-PSIR during the process.
3. New authors can also be added if needed.



Many sources - > One Profile

Profile | Publications | PhD | Research data | Projects | BSc and MSc | Activities | Achievements | Products | Citations | Statistics | Cooperation | Edit

Henryk Rybiński, PhD, DSc, Professor

Professor

The Institute of Computer Science
Faculty of Electronics and Information Technology
Phone: +48 22 234 7731
Room no: 304
Consultations: Monday 14.00-16.00

Researcher Report

Publications	120
PhD theses	12
Participation in projects	53
Research data	3
Supervised BSc and MSc theses	1
Professional activity	14
Professional achievements	6
Products	1

ORCID profile
Google Scholar profile
Scopus profile
h-index*:15

Get Cited!

Profile | Publications | PhD | Research data | Projects | BSc and MSc | Activities | Achievements | Products | Citations | Statistics | Coope

Henryk Rybiński, PhD, DSc, Professor

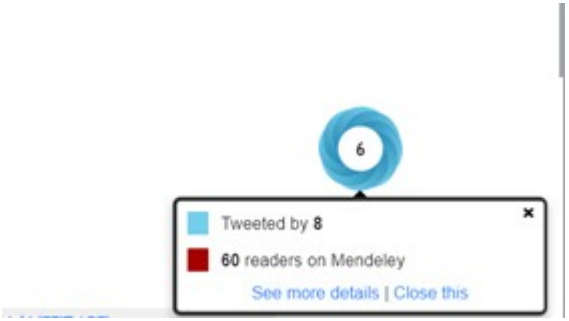
Professor

The Institute of Computer Science
Faculty of Electronics and Information Technology
Phone: +48 22 234 7731
Room no: 304
Consultations: Monday 14.00-16.00

Scholar | Scopus | WoS

hindex = 8, cited by total = 252

no	pub	title	cited	date
1		Kryszkiewicz Marzena, Rybiński Henryk: Computation of Reducts of Composed Information Systems, in: Fundamenta Informaticae, vol. 27, no. 2/3, 1996, pp. 183-195	22	12/05/2019
2		Bembenik Robert, Rybiński Henryk: FARICS a method of mining spatial association rules and collocations using clustering and Delaunay diagrams, in: Journal of Intelligent Information Systems, vol. 33, 2009, pp. 41-64, DOI:10.1007/s10844-008-0076-1	19	12/05/2019
3		Rybiński Henryk: On first order logic databases, in: ACM Transactions on Database Systems, vol. 12, no. 3, 1987, pp. 325-349, DOI:10.1145/27629.27630	17	12/05/2019
4		Kryszkiewicz Marzena, Rybiński Henryk, Gajek Marcin: Dataless transitions between concise representations of frequent patterns, in: Journal of Intelligent Information Systems, vol. 22, no. 1, 2004, pp. 41-70, DOI:10.1023/A:1025828729955	16	12/05/2019
5		Kryszkiewicz Marzena, Rybiński Henryk: Reducing information systems with uncertain attributes, in: Foundations of Intelligent Systems / Raś Zbigniew W, Michalewicz Maciek (*eds.*), Lecture Notes In Computer Science, vol. LNCS 1079, 1996, Springer, ISBN 3-540-61286-6, pp. 285-294, DOI:10.1007/3-540-61286-6_153	13	12/05/2019
6		Podsiadło Mariusz, Rybiński Henryk: Financial Time Series Forecasting using Rough Sets with Time-Weighted Rule Voting, in: Expert Systems With Applications, vol. available on-line, 2016, pp. 1-33, DOI:10.1016/j.eswa.2016.08.066	12	12/05/2019
7		Protaziuk Grzegorz Michał, Kryszkiewicz Marzena, Rybiński Henryk *[et al.]*: Discovering Compound and Proper Nouns, in: Lecture Notes in Artificial Intelligence, vol. LNAI 4585, 2007, pp. 505-515, DOI:10.1007/978-3-540-73451-2_53	11	12/05/2019
8		Gawrysiak Piotr, Protaziuk Grzegorz Michał, Rybiński Henryk *[et al.]*: Text onto miner - a semi automated ontology building system, in: Lecture Notes in Artificial Intelligence, vol. LNAI 4994, 2008, pp. 563-573, DOI:10.1007/978-3-540-68123-6_61	10	12/05/2019
		Rybiński Henryk, Kryszkiewicz Marzena, Protaziuk Grzegorz Michał *[et al.]*: Discovering Synonyms Based on Frequent Termsets, in: Lecture Notes in Artificial Intelligence, vol. LNAI 4585, 2007, pp. 516-525,	8	12/05/2019



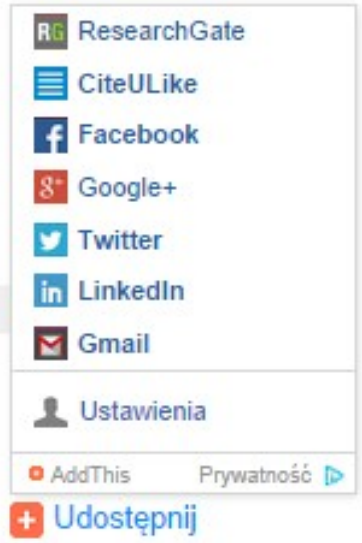
Propagate your achievements everywhere

ORCID

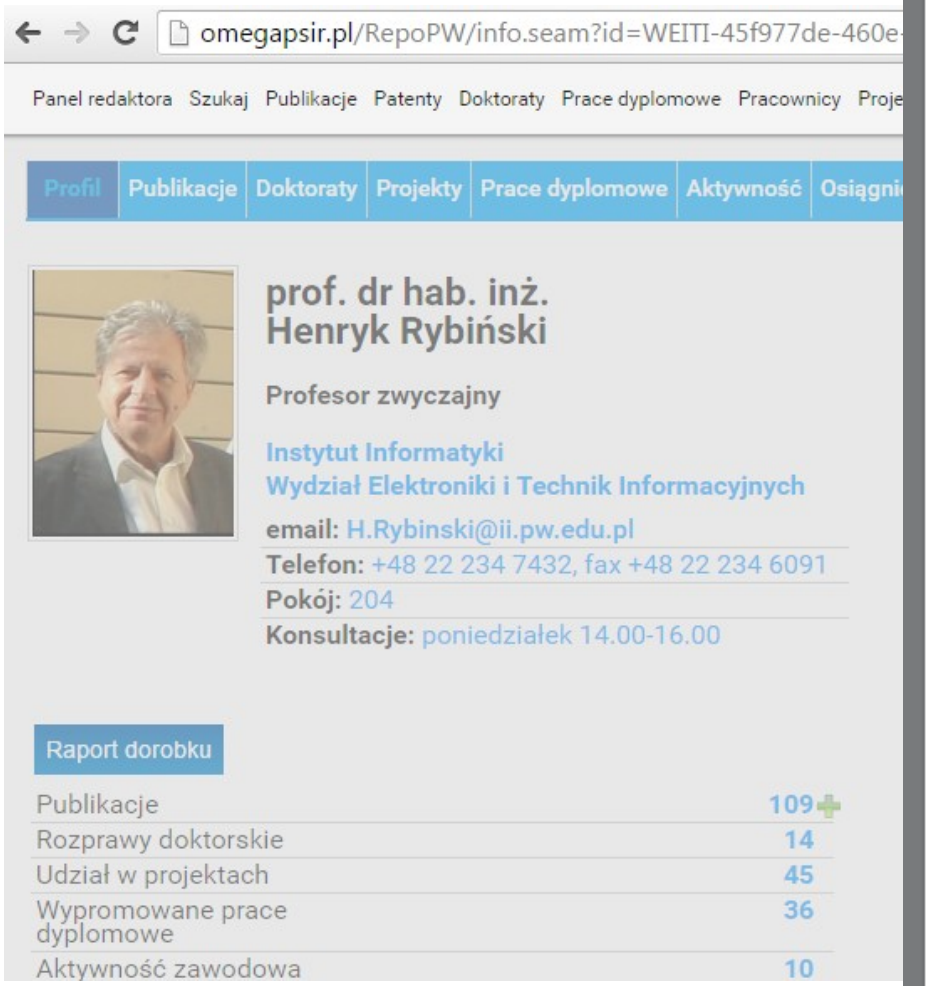
Google scholar

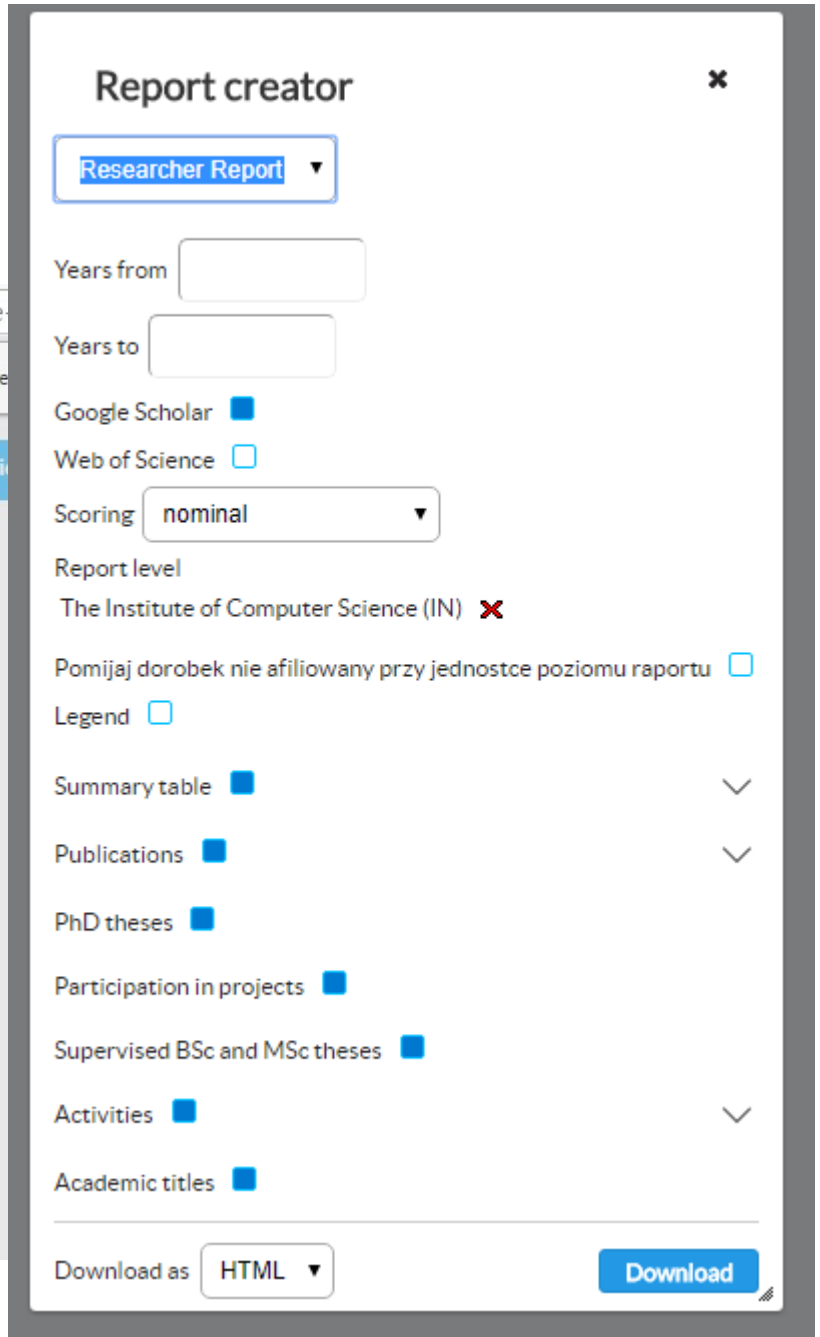
OpenDOAR

OpenAIRE



Reports


omegapsir.pl/RepoPW/info.seam?id=WEITI-45f977de-460e
Panel redaktora Szukaj Publikacje Patenty Doktoraty Prace dyplomowe Pracownicy Proje
Profil Publikacje Doktoraty Projekty Prace dyplomowe Aktywność Osiągni
prof. dr hab. inż.
Henryk Rybiński
Profesor zwyczajny
Instytut Informatyki
Wydział Elektroniki i Technik Informacyjnych
email: H.Rybinski@ii.pw.edu.pl
Telefon: +48 22 234 7432, fax +48 22 234 6091
Pokój: 204
Konsultacje: poniedziałek 14.00-16.00
Raport dorobku
Publikacje 109
Rozprawy doktorskie 14
Udział w projektach 45
Wypromowane prace dyplomowe 36
Aktywność zawodowa 10


Report creator
Researcher Report
Years from
Years to
Google Scholar
Web of Science
Scoring nominal
Report level
The Institute of Computer Science (IN)
Pomijaj dorobek nie afiliowany przy jednostce poziomu raportu
Legend
Summary table
Publications
PhD theses
Participation in projects
Supervised BSc and MSc theses
Activities
Academic titles
Download as HTML
Download

Statistics of a unit

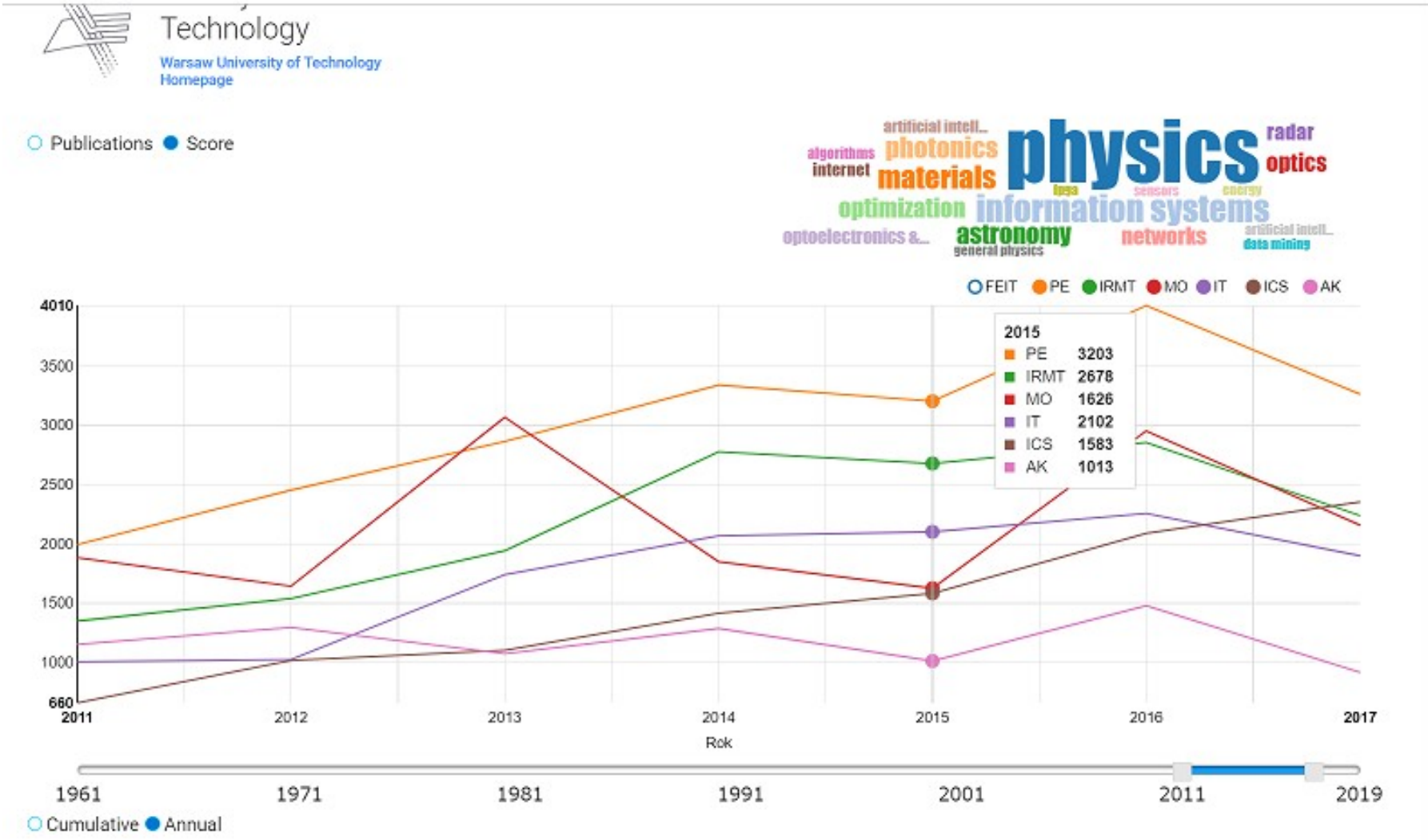
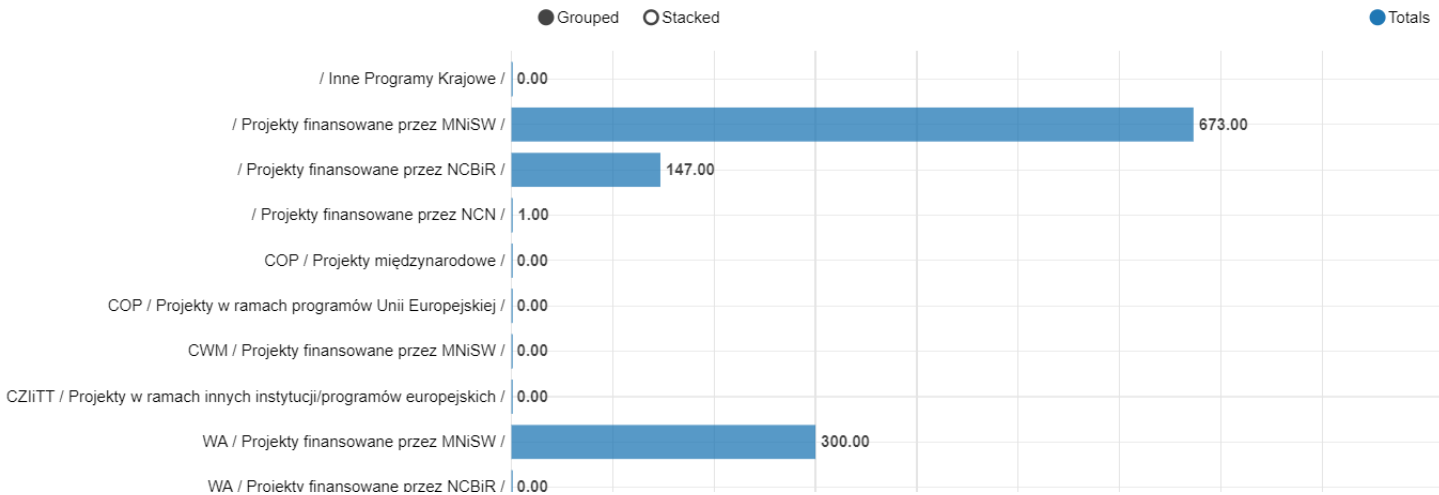


Table | Visualization

Unit (1)	project type (root)	project type (leaf)	Totals
	Inne Programy Krajowe		
	Projekty finansowane przez MNiSW		673 364
	Projekty finansowane przez NCBiR		147 144 894,9
	Projekty finansowane przez NCN		1 698 600
COP	Projekty międzynarodowe		
	Projekty w ramach programów Unii Europejskiej		
CWM	Projekty finansowane przez MNiSW		
CZIiTT	Projekty w ramach innych instytucji/programów europejskich		
WA	Projekty finansowane przez MNiSW		300 000
	Projekty finansowane przez NCBiR		
	Projekty finansowane przez NCN		111 660
WAiNS	Projekty finansowane przez MNiSW		
	Projekty finan		
	Projekty finan		
WBMiP	Projekty finan		
	Projekty finan		
WCh	Projekty finan		
	Projekty finan		

Table | Visualization

Bar chart



STIMULATE FOR GROW

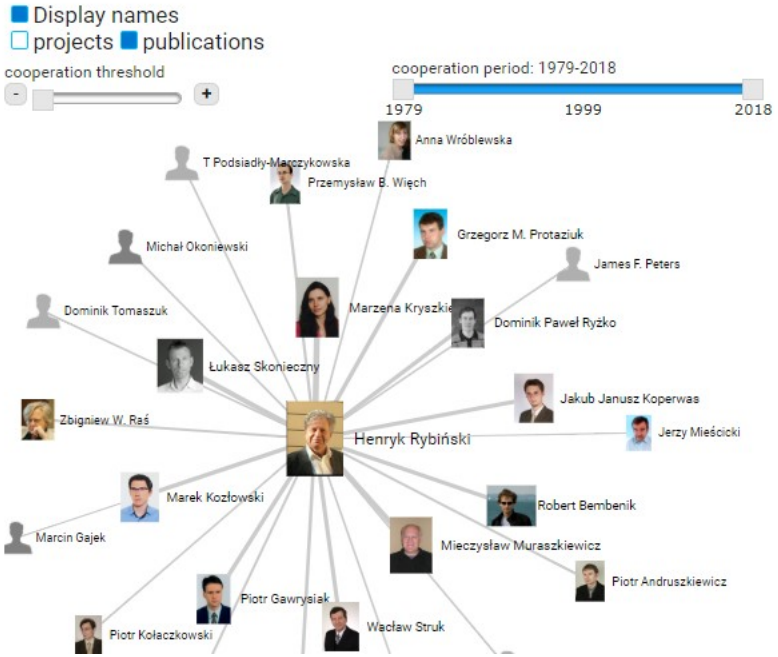
Intstitutional teams

Profile | Publications | PhD | Projects | BSc and MSc | Activities | Achievements | Products | Citations | Statistics | Cooperation | Edit

Henryk Rybiński, PhD, DSc, Professor

Professor

The Institute of Computer Science
Faculty of Electronics and Information Technology
Email: H.Rybinski@ii.pw.edu.pl
Phone: +48 22 234 7731
Room no: 304
Consultations: Monday 14.00-16.00



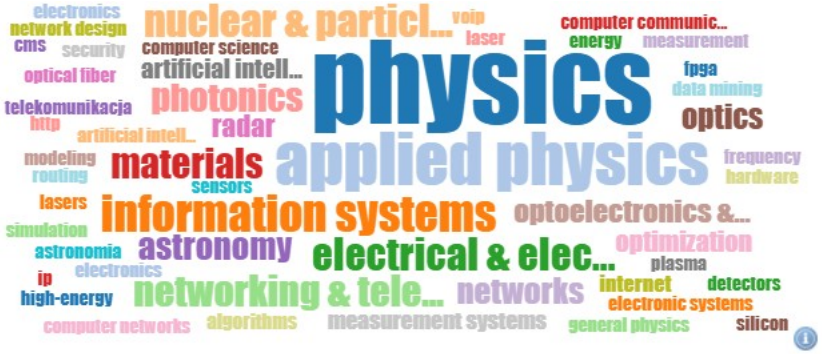
Hints

- Limit the cooperation graph to selected domain by clicking the word cloud
- Double-click a person on the cooperation graph to see his/her profile
- Hover the mouse on the connection between two persons to see the number of publications they did together

External cooperation

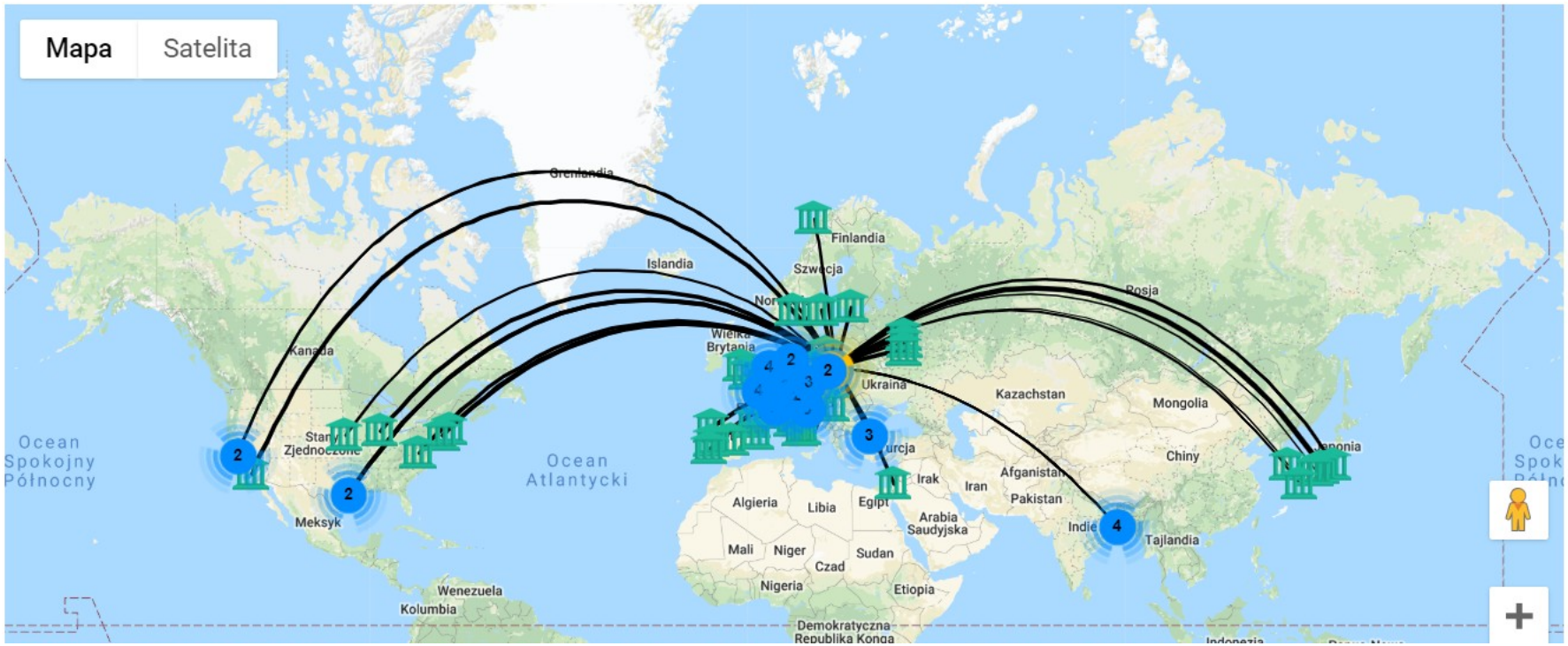
Wydział Elektroniki i Technik Informacyjnych

Politechnika Warszawska
Strona domowa

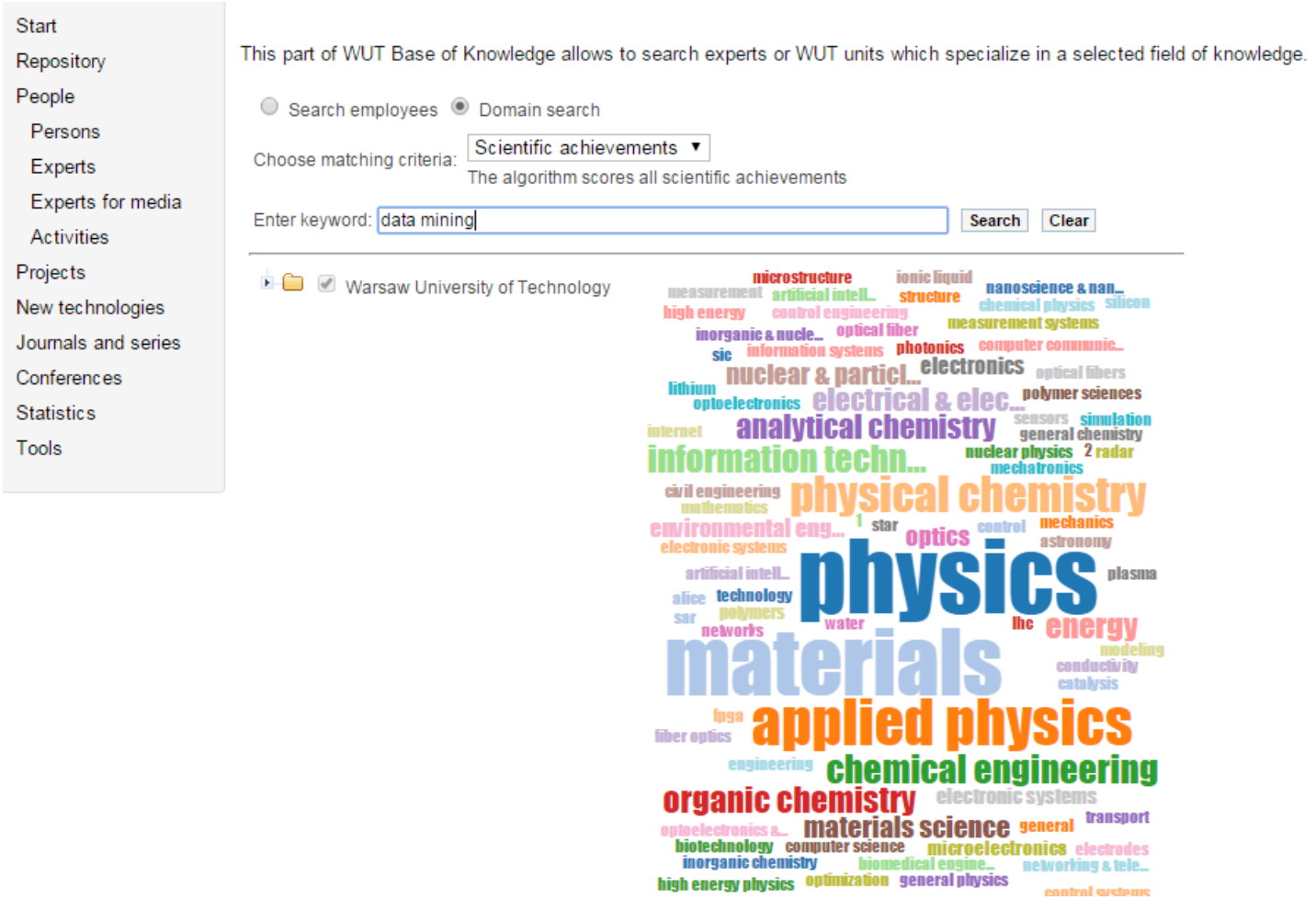


Graf | Mapa

Funkcjonalność w przygotowaniu



Expert search step 1



Expert search step 2

Dziedzina: data mining

Eksperci wg. kryterium Dorobek naukowy ▼ **Algorytm ocenia całość dorobku naukowego**

prof. dr hab. inż. Henryk Rybiński
Profesor zwyczajny
Instytut Informatyki
Wydział Elektroniki i Technik Informacyjnych
email : H.Rybinski@ii.pw.edu.pl
Dopasowanie: [szczegóły]

	w dziedzinie	wszystkie
Publikacje	36	93
Rozprawy doktorskie	7	14
Udział w projektach	7	31
Wypromowane prace dyplomowe		2

prof. dr hab. inż. Marzena Kryszkiewicz
Profesor nadzwyczajny
Instytut Informatyki
Wydział Elektroniki i Technik Informacyjnych
email : M.Kryszkiewicz@ii.pw.edu.pl
Dopasowanie: [szczegóły]

	w dziedzinie	wszystkie
Publikacje	47	83
Rozprawy doktorskie	5	6
Udział w projektach	6	9
Wypromowane prace dyplomowe	2	6

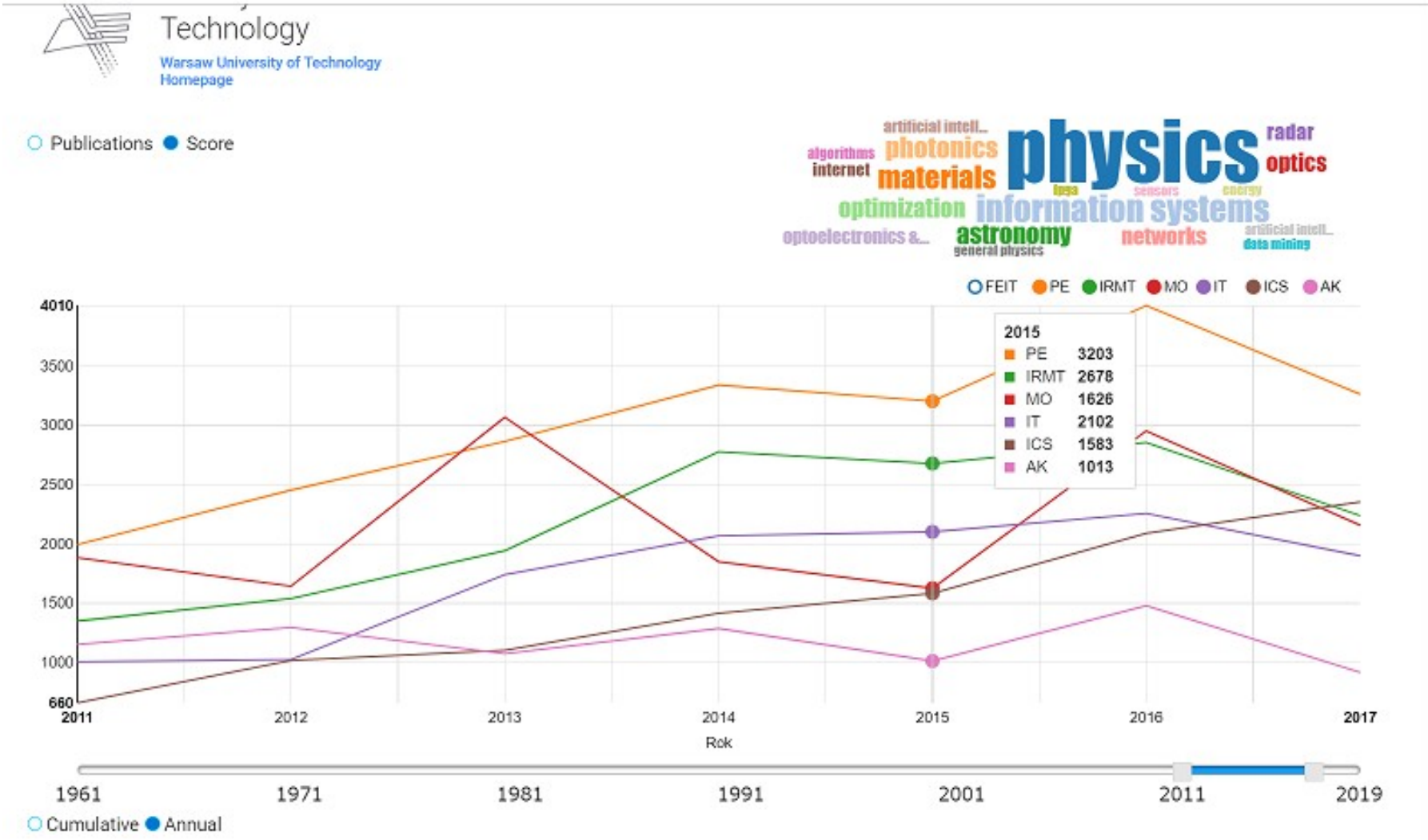
prof. dr hab. inż. Krzysztof Walczak
Profesor nadzwyczajny
Instytut Informatyki
Wydział Elektroniki i Technik Informacyjnych
email : K.Walczak@ii.pw.edu.pl
Dopasowanie: [szczegóły]

	w dziedzinie	wszystkie
Publikacje	17	39
Rozprawy doktorskie	4	4
Udział w projektach	2	2
Wypromowane prace dyplomowe	3	6

dr inż. Tomasz Gambin
Adiunkt
Instytut Informatyki
Wydział Elektroniki i Technik Informacyjnych
email : T.Gambin@ii.pw.edu.pl
Dopasowanie: [szczegóły]

	w dziedzinie	wszystkie
Publikacje	7	23
Rozprawy doktorskie	1	1
Udział w projektach	1	1

Statistics of a unit



Fundings

CONTESTS SEARCH

contest.startDate - from: to:

contest.endDate - from: to:

Open in days - from: to:

Open on

contest.active: ☐ contest.active.true ☐ contest.active.false

contest.regularity: ☐ contest.regularity.regular ☐ contest.regularity.irregular

contest.projecttype:

Tree

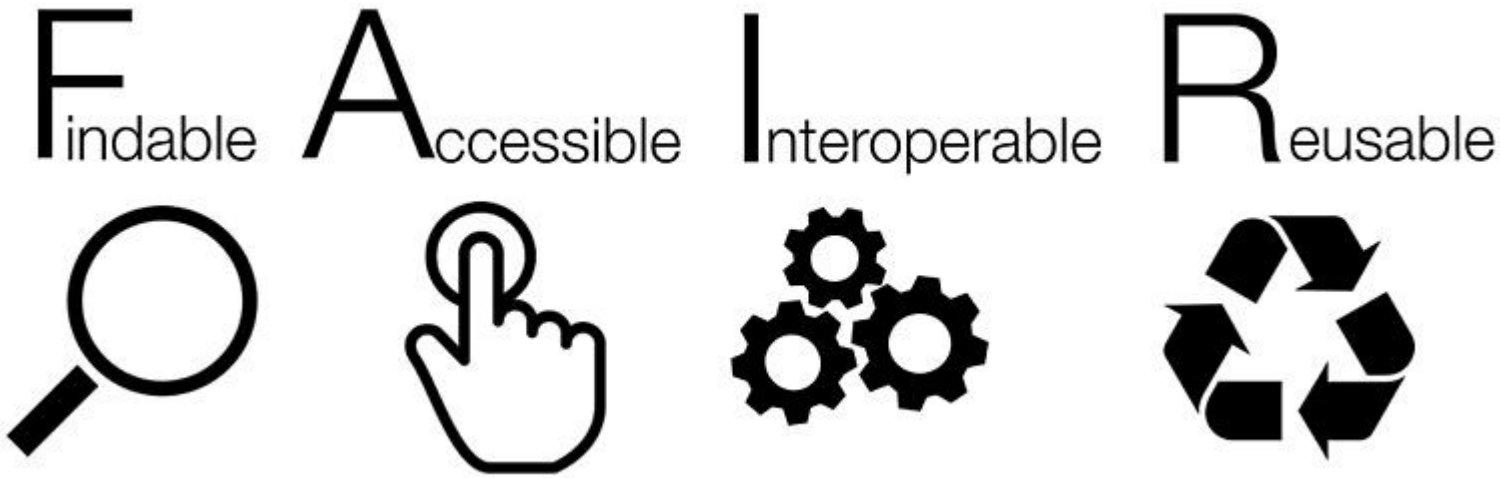
contest.subjectType:
☐ fundamental research ☐ developing research ☐ educational
☐ investment ☐ dissemination of science ☐ norvegian ☐ implementation
☐ construction investment ☐ expertise ☐ opinion ☐ other

Search | Search recommendations | Clear (27)

Note: search results limited to the selected affiliations

OPEN DATA REPOSITORY & ANNOTATION PLATFORM

FAIR Data



Rich metadata+persistent identifier

Metadata and data are understandable to humans and machines

Metadata and data are understandable to humans and machines

Metadata description standarisation, common knowledge representation

Clear License

https://upload.wikimedia.org/wikipedia/commons/a/aa/FAIR_data_principles.jpg

5-star linked open data

☆ Data is available on the Web, in whatever format.
☆☆ Available as machine-readable structured data, (i.e., not a scanned image).
☆☆☆ Available in a non-proprietary format, (i.e, CSV, not Microsoft Excel).
☆☆☆☆ Published using open standards from the W3C (RDF and SPARQL).
☆☆☆☆☆ All of the above and links to other Linked Open Data.

https://www.wikidata.org/wiki/Wikidata:Open_data_publishing

Korpusomat

Korpusomat (aka “Corpus Machine”) is a service used mostly by linguists and translators to **quickly create automatically annotated text corpora**.

Automatic annotation consists of text segmentation into words, recognizing part of speech tags, lemmas and named entities.

The corpus may be queried by the user using a **query language**.

The corpora may be **publicly published** to be used by any user of the service (*this feature is under development*).

Korpusomat

Zapytanie

[base="uczciwy"][base="człowiek"]

KONSTRUKTOR ZAPYTAŃ | METADANE | STATYSTYKI

Liczba wyników na stronę

10

Wyszukaj

Znaleziono 3 wyników.

Lp	Lewy kontekst	Rezultat	Prawy kontekst
1	zgodzę — a ja ani myślę tamtego zrobić".	**Uczciwy** [uczciwy:adj:sg:nom:m1:pos] **człowiek** [człowiek:subst:sg:nom:m1]	! Teraz się wszystko wykryło: „Jestem biedny i
2	i o panu wszystko się wyda; o panu,	**uczciwy** [uczciwy:adj:sg:voc:m1:pos] **człowieku** [człowiek:subst:sg:voc:m1]	, któryś mnie oszukiwał. Nie ma pan pieniędzy,
3	do grona przestępców, świadomych swego złoczynu, aniżeli do	**uczciwych** [uczciwy:adj:pl:gen:m1:pos] **ludzi** [człowiek:subst:pl:gen:m1]	, nagle rzeczą przykrą zakłopotanych. Dwóch lub trzech tylko

Labellery

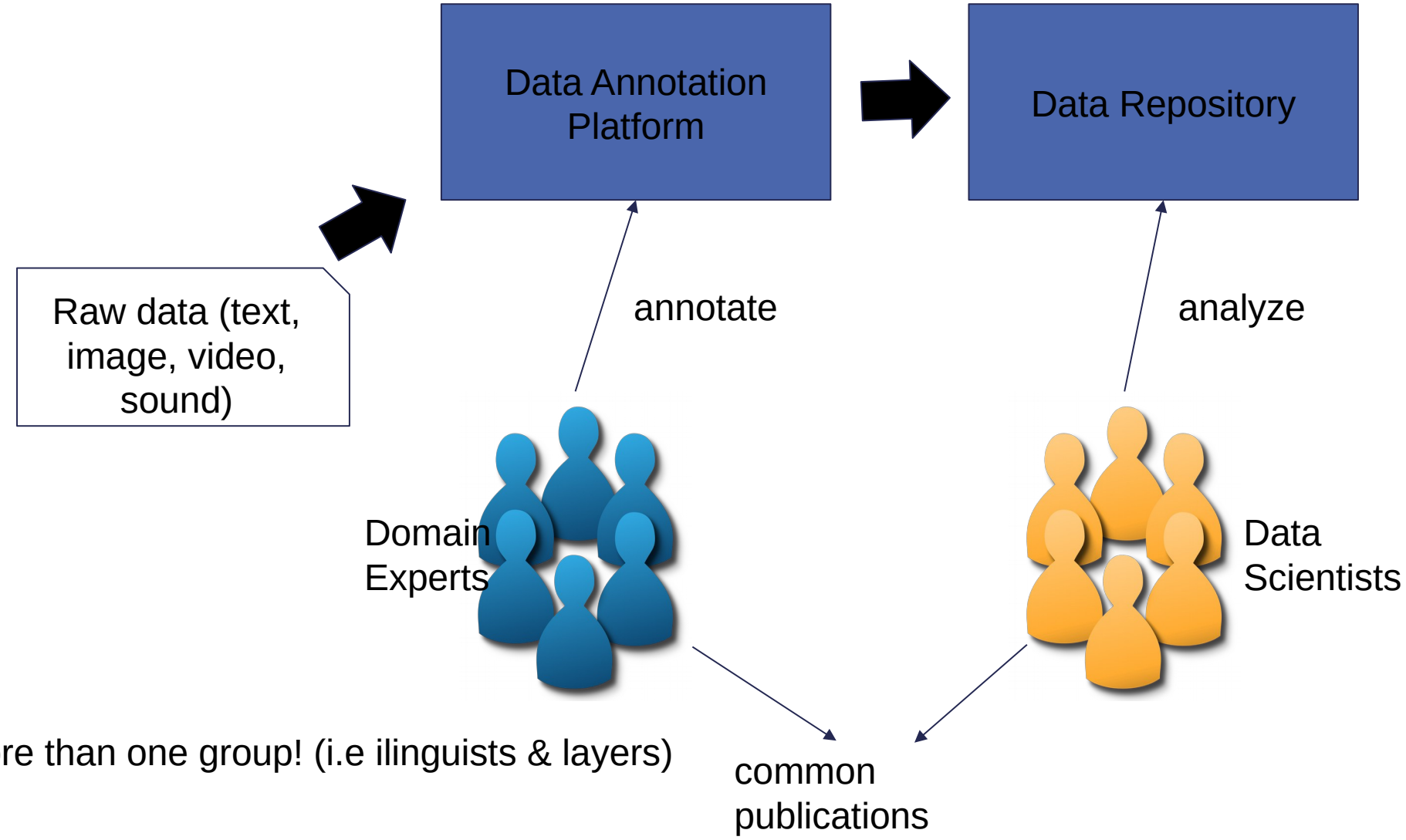
Automatic annotation is great, but in many applications **high quality hand-annotated data** is the crucial element of the process.

For example, **machine learning methods** are currently being used very commonly in many domains. Models used in such approaches need to be trained on pre-annotated data.

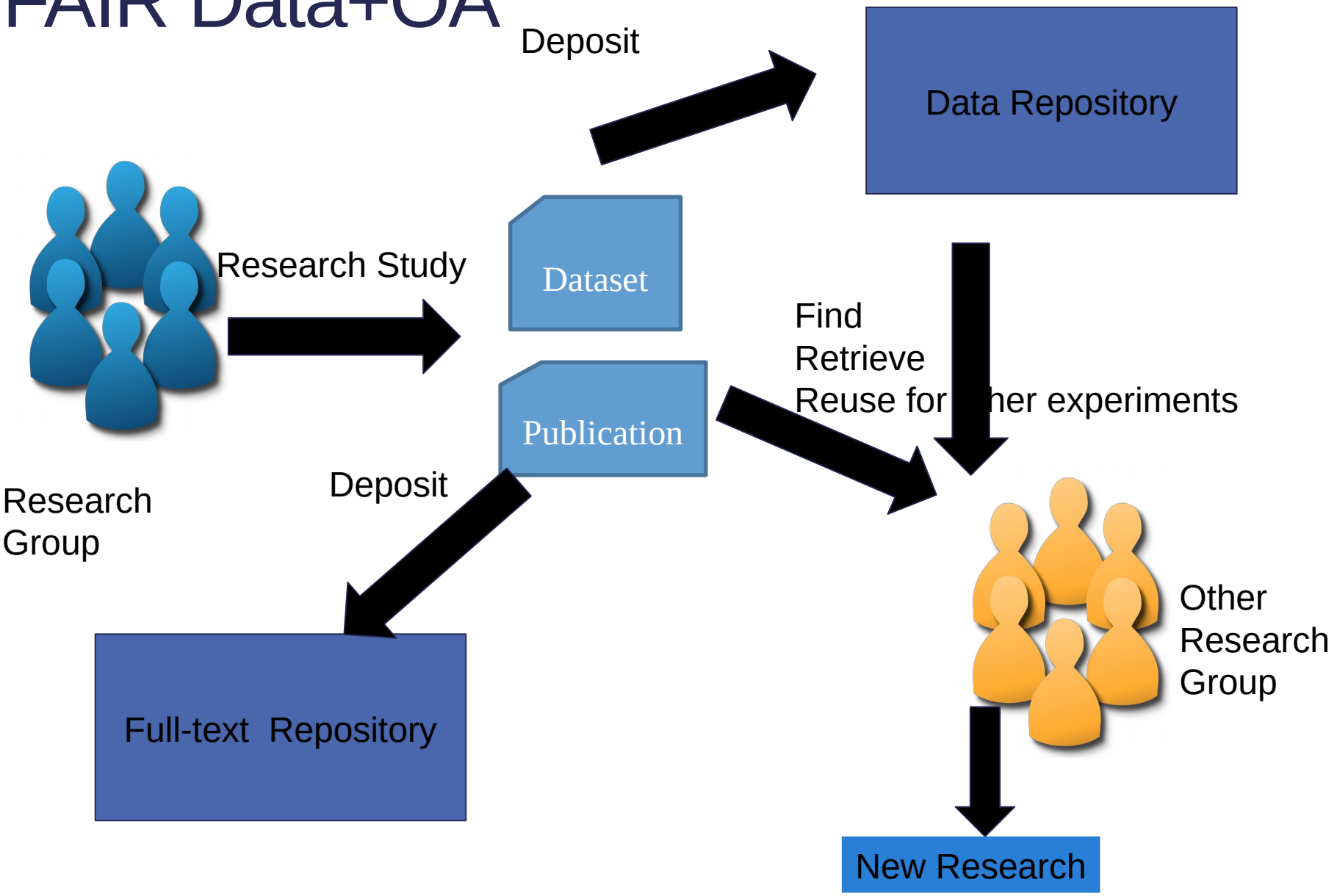
At Labellery we equip annotators with tools to perform their work **quickly and accurately** and match them with organizations needing annotated data.

The data may be **used privately or released openly**, using one of the permissive licenses.

Labellery


Data Annotation Platform
Data Repository
Raw data (text, image, video, sound)
annotate
analyze
Domain Experts
Data Scientists
More than one group! (i.e ilinguists & layers)
common publications

FAIR Data+OA


Deposit
Data Repository
Research Study
Dataset
Find
Retrieve
Reuse for her experiments
Publication
Research
Group
Deposit
Other
Research
Group
Full-text Repository
New Research

AI IN DIALOGUE SYSTEMS

Current issues in dialogue systems

While many NLP-driven dialogue systems exist today, they face numerous problems:

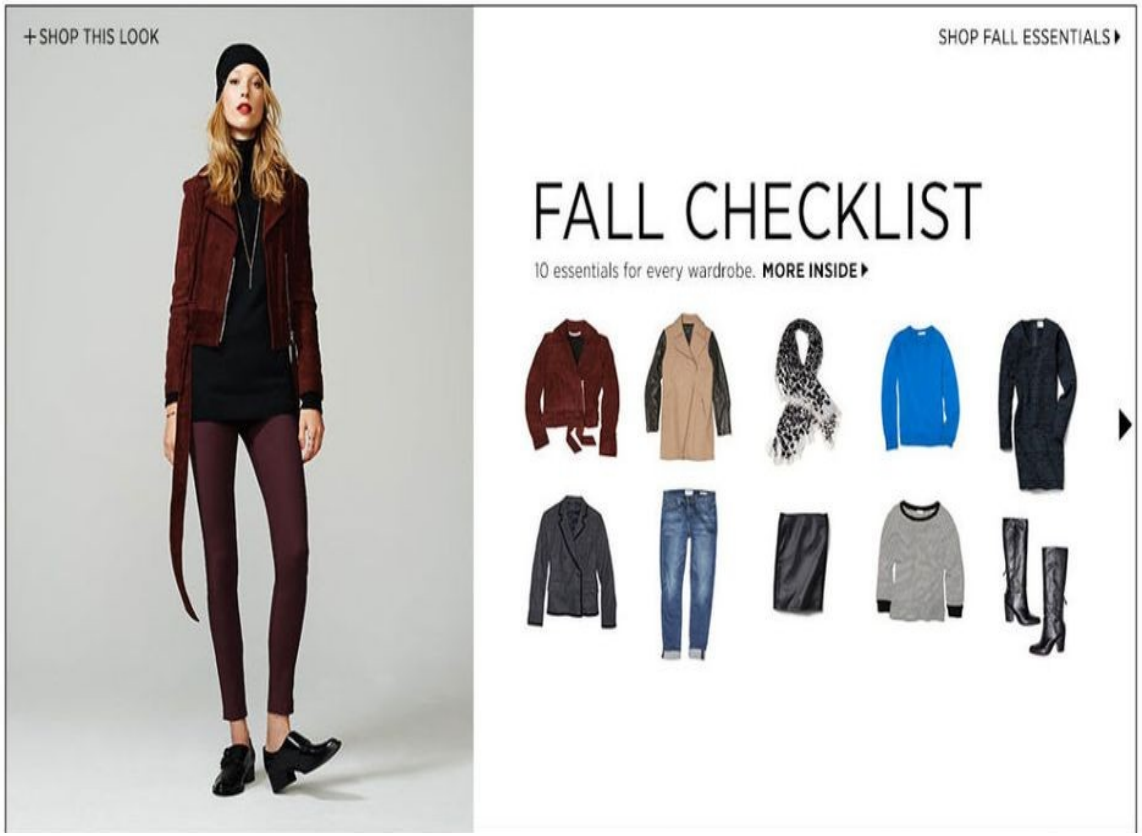
- most of them still rely on pre-determined set of rules (intent-action) and a pre-determined dictionary of entities (e.g. product names),
- the knowledge base, which is used for generating answers needs to be highly structured,
- answers are generated based on manually created templates,
- they are English-focused and systems for other languages are lacking in accuracy compared to English.

An example in the fashion domain

"I need a **dark top** to match **this look.**"

"Show me the **top offers** for today."

- Natural Language Understanding
- Image Recognition
- Context



Natural Language Understanding and context

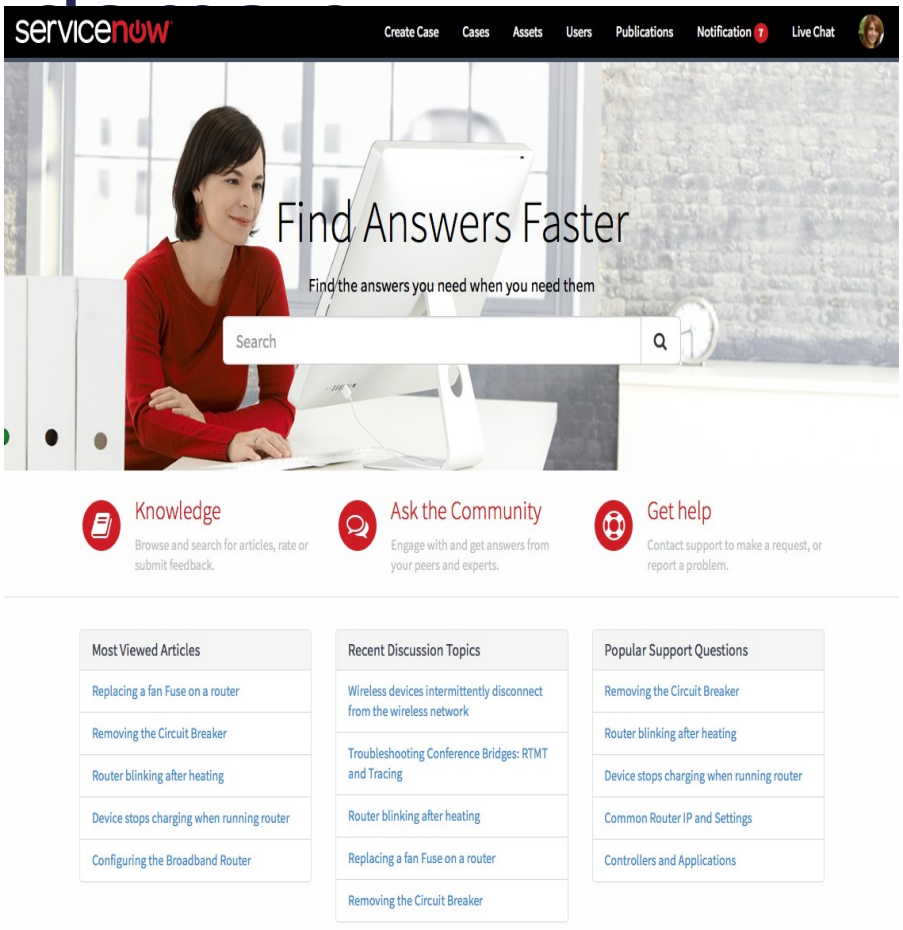
NLU

- named entity recognition
- word sense disambiguation
- word embeddings
- additional resources, e.g. WordNet

Context and Image Recognition

- conversation history, time, date and place of the conversation
- which products are visible or similar to those on the photo?

An example in customer service



"How can I setup Internet access on my new phone?"

"Why is monthly statement higher than usual?"

- Information Extraction
- Combining Knowledge Base with the user profile

Information extraction

Unstructured Knowledge Base

- large, unstructured knowledge bases are usually maintained within companies working a large number of B2C clients
- automated customer service is possible when this unstructured KB is searchable, structured information can be extracted and a natural language answer can be generated

Combining information

- information from the general KB has to be combined with the user profile, her history of communication with the customer service, billing data, etc.

Not only English

An example of Polish language

- Polish is inherently more difficult to automatically process than e.g. English
- A smaller number of researchers are working on Polish
- There is a lesser number of language resources available for Polish

Real world products need to be localized to be successful -> additional work is necessary to maintain localized language resources, train additional models and understand language differences.