Central systems for collecting data on science and higher education - role and evolution in the light of the Polish Legislation

Marek Michajłowicz, Małgorzata Paszkowska
National Information Processing Institute, Warsaw, Poland

euroCRIS Strategic Membership Meeting 2018, 27 November, Warsaw, Poland
We are enthusiastic professionals focused on modern Information and Communications Technologies.

We design and develop advanced IT systems to support decision-making processes, conduct research in the fields of machine learning, artificial intelligence, big data, and methods of software production.

- **Big Data & Databases**: We utilise databases as well as big data technologies to persist and process data.
- **Intelligent Algorithms**: We utilise artificial intelligence to serve sophisticated solutions for our clients.
- **Software as a Service**: Our information systems offer services that suit to consumers needs.

Warsaw, Poland
Throughout this presentation …
DIVERSITY
heterogeneous information environment. Most universities and institutes have their own unique information systems in various technologies and architectures (ERP, antiplagiarism systems, repositories, and alike).

CHANGING LAW
The critical issue is the high changeability of law in Poland. During the last five years there were three amendments of the bill about higher education and science in Poland.

DOMINATION OF AD-HOC ANALYZES
difficulties in making long-term comparisons based on consistent data and criteria. Problem with data reliability.

ADMINISTRATIVE BURDEN
distributed reporting obligations in order to obtain similar data by universities and scientific institutes in the form of surveys, reporting forms and similar.
HIGHER EDUCATION AND SCIENCE IN POLAND

Data collected since 2011 in central repositories:

- 1,300,000 students per academic year
- 35,000 patents
- 20,000 projects
- 396 universities
- 21,000 properties
- 164,000 academic teachers and researchers
- 3,100 laboratories
- 17,500 courses
- 1,900,000 theses
- 24,500 research equipment
- 260 publications
- 1,060,000 units
supporting applications for research grants

collecting a wide range of data on science and higher education

collecting detailed data on scientific publications

collecting full texts of thesis

tracking the economic fate of university graduates

supporting universities with antiplagiarism
Integrated Information System on Science and Higher Education in Poland

- SEDN - Evaluation System of Scientific Achievement
- PBN - publications, magazines, articles
- ORPD
- POL-INDEX - Polish Graduate Tracking System

Central system POL-ON

- universities and scientific units
- degrees and titles
- forms and reports
- commercialization
- scientists
- students
- courses
- products
- foundings
- projects
- real estate
- patents and protection rights
- artistic achievements
- investments
- achievements
- infrastructure
- document authorization
- conferences

Over 500 university and institutes systems with different technologies [ERP, CRIS, Student Information Systems, HR Systems, Thesis, Document Repositories etc.]

Use of data

- Ministry of Science and Higher Education
- Ministry of Health
- Ministry of Culture and National Heritage
- National Science Centre
- National Centre for Research and Development
- Central statistical office in Poland (GUS)
- Committee for Evaluation of Scientific Units
- The Degrees and Titles Committee
- Polish Accreditation Committee
- Polish National Agency For Academic Exchange
- Universities
- Polish Academy of Sciences

Data acquisition in one data model
Original plan:

- Unbreakably part of the POL-on
- Independent system with a separate domain
Original plan:
- Personal accounts as a base for repositorium
- Author showing their achievements

Current purpose of the system:
- Gathering official data on publications from scientific institutions
- Source of information for evaluation

From social to social + institutional
POLISH SCIENTIFIC BIBLIOGRAPHY – WHY SEPARATE PUBLICATIONS FORM OTHER DOMAINS? → ACCESS RESTRICTIONS

- Can be accessed strictly by authorized employees of scientific units.
- Data in the system are specified very precisely in legal acts.
POLISH SCIENTIFIC BIBLIOGRAPHY – SOME NUMBERS

- 637,000 monographs
- 71,000 chapters
- 351,000 articles
- 1,060,000 publications
- 247,000 identified authors
- 35,790 journals
- 1,233 reporting institutions

Data for 2013 to date
MAIN GOALS

Funding
- Division of various funds for higher education and science.

Evaluating
- Scientific performance of research units
- Plagiarism of diploma theses
- Quality of higher education
- Diplomas

Controlling
- 75%
- 60%
- 95%

Monitoring
- Careers of graduates
- Financial aid for students

Informing
- Educational offers for candidates
- Financial and statistical reports
- Public data
INFORMING – ANALYTICAL POSSIBILITIES

Authors personal data
Authors employment data
Publications per region
Publication per institution features
MAIN FACTS ABOUT POL-on
THE INTEGRATED SYSTEM OF INFORMATION ON
SCIENCE AND HIGHER EDUCATION

• A centralized databases of highly structured data
• Data model and responsibilities defined on the basis of legal acts
• The system entirely took over the reporting processes of universities and research institutes for the needs of central / government units, including statistics for the Central Statistical Office and financial reporting as well
• Most of the data is calculated automatically
• And we are appreciated ……
EUNIS 2018 Elite Award for Excellence

2018 EUNIS Elite Award Winners

2018 EUNIS Elite Award for excellence in implementing Information Systems for Higher Education winner is the project: POL-on: The Information System of Science and Higher Education in Poland.

The award was handed at the EUNIS 2018 Congress Gala Dinner to Marek Michajłowicz, Marta Niemczyk, Jarosław Protaśewicz and Karolina Mroczkowska from the National Information Processing Institute (OPI) in Poland.

The main purpose of the EUNIS Elite Award is to recognize and promote best practice in some aspect of the use of information systems in higher education in Europe. The competition recognizes excellence, encourages best practice and stimulates wider awareness and use of information systems in research, administration and community service in institutions.

Read more on the Elite Award here.
main features of the current architecture

It has achieved their goal but costs too much effort

Rather silos monolitic architecture
Low interoperability
Wider range of data
Very specific, local data model which was changed to often by the legal acts
Lack of data warehouse – transactional system for all the usage
Unscallable
New law: “The Constitution for Science and Higher education”

Main features in relation to central systems:

- Colloquially referred to as Act 2.0
- Contains paragraphs about the use of ORCID numbers of Polish scientists
- It simplifies and reduces the scope of data needed in the evaluation process of science
- Increases responsibilities regarding the frequency of data updates in central repositories

With new act come opportunities
OPPORTUNITY: INCREASING INTEROPERABILITY

- Deep integration with ORCID
- Managing Web of Science and Scopus data
- Integrating with Polish National Library
- Using data from the Polish Patent Office
OPPORTUNITY: INTERNATIONAL DATA MODELS

Turning to CERIF as a recognized European standard.
OPPORTUNITY: BREAKING THE MONOLITH

Turning endless documents into integrated objects

- JOURNAL
- JOURNAL
- ORGANISATION UNIT
- PUBLICATION
- PERSON
- DISCIPLINE
- PUBLISHER
OPPORTUNITY: BREAKING THE MONOLITH

POL-on 2.0
- Many separated autonomous microservices
- Separated domains (Students, Employees, Financial Data, Scientific Achievements)
- Domain Driven Design approach
- Communication and access mainly through the API REST
- Division of the application structure into a write and readable part (CQRS + Eventsourcing)

Separation of the analytical and integration layer.....
An integrated system of services for science
• Project launched in 2017 by Ministry of Science and Higher Education and National Information Processing Institute
• The main goals:
  • to increase the dissemination of information
  • to produce sophisticated analyses (a warehouse)
  • to increase the interoperability
Thank you for your attention!

Małgorzata Paszkowska – mpaszkowska@opi.org.pl

Marek Michajłowicz – mmichajlowicz@opi.org.pl