

# Perspektywy rozwoju modeli otwartej nauki w Polsce

Marek Niezgódka  
ICM, Uniwersytet Warszawski  
[marekn@icm.edu.pl](mailto:marekn@icm.edu.pl)

Konferencja BazTech, Kraków, 26.06.2017



UNIVERSITY OF WARSAW  
Interdisciplinary Centre for Mathematical  
and Computational Modelling  
[www.icm.edu.pl](http://www.icm.edu.pl)

# System nauki: obraz dzisiejszy

- Nauka znajduje się w okresie głębokiego kryzysu reprodukowalności:
  - wyniki badań możliwe do niezależnego odtworzenia są bardziej wiarygodne
  - cel: nie tylko weryfikacja ale stymulowanie dalszych badań
  - znaczenie poszerzenia badań eksperymentalnych i obserwacyjnych o umożliwienie szerokiej analizy danych
- Główny bezpośredni powód kryzysu: [systemowa zamkniętość nauki](#)

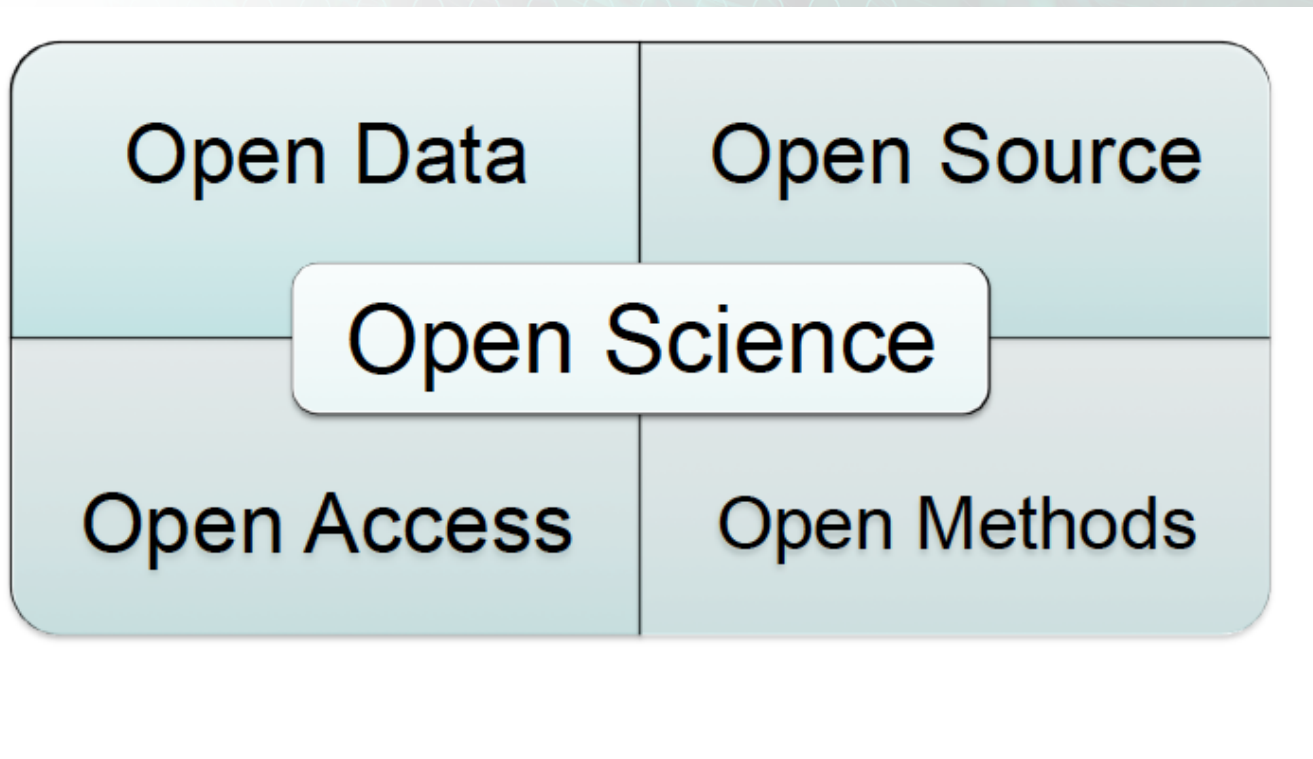
# Kontekst międzynarodowy

- Manchester Manifesto: „*Who own Science?*” (2009)
- Horizon 2020: obowiązek publikowania wyników w modelu otwartym
- Nowy impuls: Prezydencja Holandii w UE: *Amsterdam Call*
- Nowa Dyrektywa Copyright UE, trwająca debata
- Rekomendacje EUA (*European University Association*)
- Komisja Europejska, Horizon 2020: zobowiązanie do otwartości publikacji; rola e-infrastruktury *OpenAIRE*



# 2016: nowe życie

- Priorytet prezydencji NL w EU:



# Istotne novum: problematyka TDM

- TDM: *Text and Data Mining*
- European Commission, **Draft Proposal of the Copyright in the Digital Single Market Directive**, 14/09/2016: *TDM is... ‘any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations’*
- Klauzule wyjątków TDM w Dyrektywie o Copyright
- Kontrowersje co do zakresu wyjątków
- Pozycja wydawnictw: TAK, jako add-on serwis
- Pozycja akademicka, w tym EUA: standard

# Polska: Otwarty Dostęp

- Brak strukturyzacji i modelu finansowania
- Prace podstawowe: MNiSW, Zespół ds. Otwartej Nauki („Nieformalny”)
- **Nowe akcenty**: deklaracje jednoznacznego silnego poparcia przez NCN dla zobowiązania do publikowania w modelu otwartym

# Otwarte dane badawcze

- H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, 26.07.2016
- Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, 21.03.2017
- Concordat on Open Research Data, RC UK, 28.07.2016

# Recommendations on Open Science Publishing, adopted by the Open Science Policy Platform on 25 April 2017

- Five Recommendations:
- 1. Stakeholder communities, member states and the European Commission **should jointly assess and identify how the OA mandate should be achieved by 2020.**
- 2. Progress towards full OA must take into account how fast the publishing system is changing, and how scholarly communications are growing in richness and variety.
- 3-4. **One size does not fit all** -... Issues of compliance, ... should be **... harmonised** in a discipline-sensitive way.
- 5. **From 2020**, The EC must move toward **a broader definition of OA that incorporates the full range of emerging formats and applications of scientific research output.**



# Polska: dylematy, wyzwania

- Brak decyzji o krajowej infrastrukturze danych i publikacji (**NIESPRZECZNE z działaniami oddolnymi!**)
- **Brak spójności decyzyjnej**: powtarzające się wielokrotne finansowanie zbliżonych działań, wciąż silny efekt istnienia silosów sektorowych
- POTRZEBA krajowej koordynacji działań
- **Brak świadomości znaczenia trwałości e-infrastruktur**

# Polska: refleksje

- **Znaczenie integralnej koncepcji nowego krajowego modelu publikowania naukowego i dokumentowania wyników naukowych** (badawczych i edukacyjnych)
- **Kontekst ewaluacji instytucjonalnej i oceny indywidualnej**: pogłębiające się cofnięcie w stosunku do kształtujących się standardów międzynarodowych
- **Znaczenie zdrowych relacji** krajowej e-infrastruktury informacji naukowej (publikacji, danych badawczych i innych form dokumentacyjnych) i dostępu do zasobów zagranicznych **w modelu gwarantującym trwałość**

# Polska: otwarte pola

- **Trwałość zasobów**: modele operacyjne, tryby finansowania
- **Wartość zasobów**: systemowe rozwiązania w zakresie TDM i szerzej, inteligentnej analityki danych
- **Interoperacyjność !!!**

# Polska: perspektywy

- Znaczenie uznania integralnej przynależności dokumentacji badawczej do pełnego procesu naukowego
- Waga upowszechnienia świadomości rzeczywistego miejsca zasady dzielenia się w nauce
- Integracja polskich zasobów w e-infrastrukturach międzynarodowych

# Nowy kształt systemu wydawniczego

- Proces odzyskania podmiotowości nauki i naukowców:
  - Otwarte e-platformy wydawnicze
  - Otwarty proces recenzowania (różnorodność modeli)
  - Model *not for profit*: w skali międzynarodowej radykalne obniżenie łącznych kosztów publikowania i dostępu
- **Rekomendacja:**
  - w perspektywie finansowanie kosztów publikowania a nie kupowanie treści
  - **Efekt:** szacunkowo do 70% obniżenia sumarycznych kosztów

Marek Niezgódka

[marekn@icm.edu.pl](mailto:marekn@icm.edu.pl)



UNIVERSITY OF WARSAW  
Interdisciplinary Centre for Mathematical  
and Computational Modelling  
[www.icm.edu.pl](http://www.icm.edu.pl)

# Załącznik: wybrane aspekty praktyczne dotyczące analizy tekstów i danych w kontekście otwartej nauki

(na podstawie opracowania MN dla EUA)



UNIVERSITY OF WARSAW  
Interdisciplinary Centre for Mathematical  
and Computational Modelling  
[www.icm.edu.pl](http://www.icm.edu.pl)

# TDM: specification

- **Text and Data Mining (TDM):** a machine-based automated or, gradually more and more cognitive, approach to analysis of text and data for any related information
- For analysis, a copy of the content (text and/or data) is usually needed, this potentially infringing Copyright unless by the Copyright owner
- Actually, the notion of TDM is to be meant as encompassing all forms of content and their mining
- TDM may be viewed within text and data analytics paradigm, still its future belongs into the area of cognitive computing:
  - hence references to unstructured content and big data
  - thus new challenging scenarios for deep information and knowledge extraction



# Starting comments

17

- TDM (Text and Data Mining) is a constitutive component of Open Science

## HOWEVER

- TDM does also exist beyond Open Science frames

**TDM workflow**, ref.: Haeussler, M., Molloy, J., Murray-Rust, P. & Oppenheim, C. *Responsible Content Mining*. (2015)

- Mining levels:
  - **Text**
  - **Data**
  - **Content (both text and data)**

# EC, Draft of the Directive “on Copyright in the Digital Single Market”, September 14, 2016

- TDM exception:

... exception for reproductions made by *research organisations* to carry out text and data mining (TDM) for scientific research purposes ...

- Key issues:

- who
- when
- what
- purpose

# Attitude of publishers towards TDM vs. interest of academic community

- A vast majority of STM publishers set rules allowing for limited TDM processing by authorized users, generally confined within non-commercial use, subject to a variety of complementary limitations
- Such policy documents are available for all leading publishers, including Elsevier, Springer Nature and Wiley, among others

# TDM: perspective of publishers and content providers

20

TDM – **publishers' practices observed at accessing content:**

- Technical protection measures
- Crawler traps
- Access restrictions to APIs

TDM: an emerging primary business field from publishers' perspective

Implications of the TDM exception in Copyright Directive:

- Discouraging publishers and content providers from interest in TDM business

# Restrictions, policies

- Typically, publishers introduced some restrictions on the scale of TDM processes and the way textual data can be exploited, as a rule prohibiting any but non-commercial use. E.g., Elsevier's policy has faced strong criticism in a wide range of academic communities [13].
- In general, legal clauses imposed on the allowance of performing TDM are to be carefully examined as they may either show prohibitive and harm effectiveness of processing or prevent exploitation range of its results. Important contributions addressing these problems from the perspective of academic institutions were published by governmental bodies and public organisations in many countries, refer to the UK in the context of research [3] and education [4]. We also recall numerous international initiatives, among them the one on Global Information Commons for Science by Codata [8]
- Most recently, some publishers released statements referring to EC Draft Proposal

# TDM exception: consequences for research sector

- **Necessity** of public (non-commercial) funding for TDM technology, infrastructures and services
- **Need** for developing large(r)-scale, eg. national or supranational platforms:
  - **Content**
  - **Infrastructure**
  - **Competence**

Ex.: PubMed Central, ISTEEX (France), INFONA (Poland), ...