



Potencjał otwartych danych badawczych

Katarzyna Patyrak, Bartek Więckowski
Research Intelligence



Plan na dziś

01

Dane badawcze w Polsce

02

Budowanie widoczności

03

W jaki sposób instytucje mogą śledzić swoje dane badawcze?

04

Podsumowanie

Dane badawcze w Polsce

Dane badawcze publikowane przez instytucje Indeksowane przez Data Monitor

41.000+

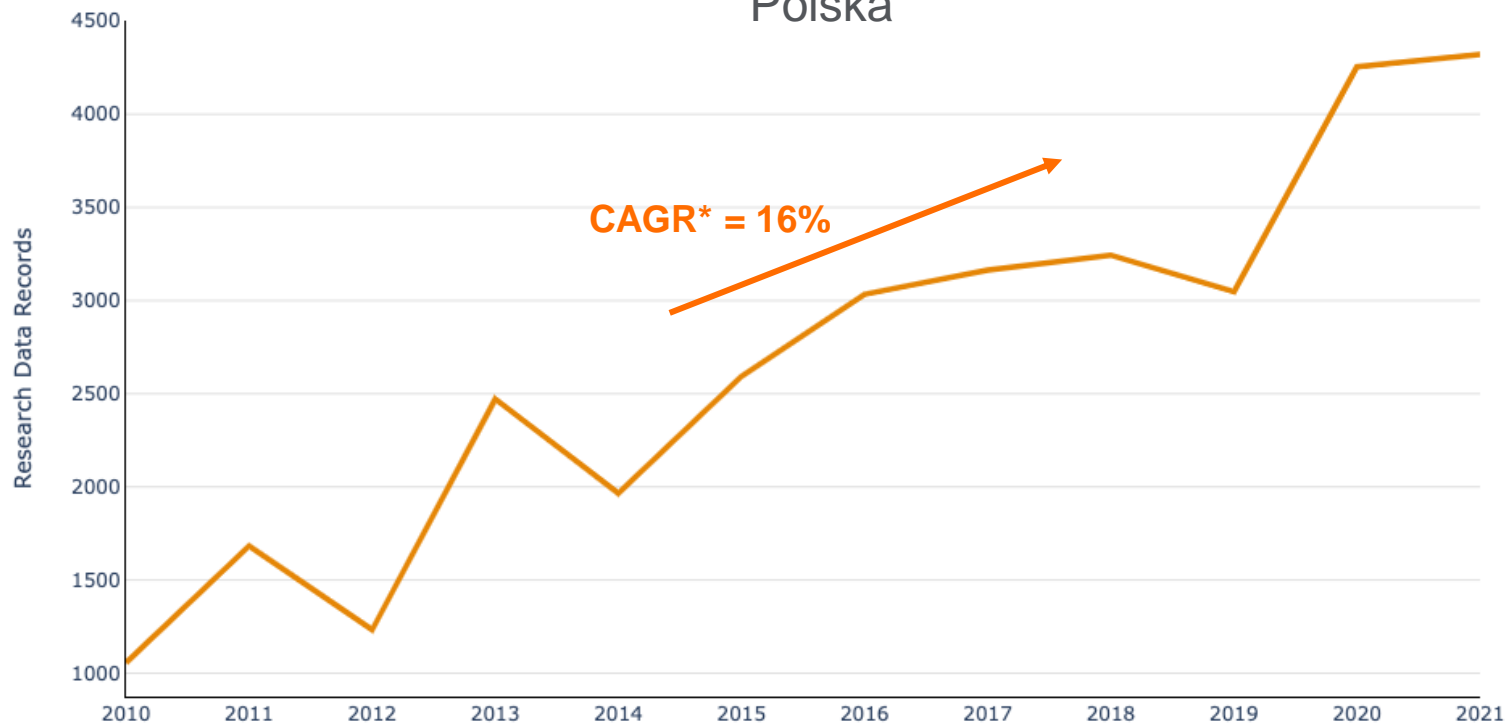
Polska

1.370.000+

EU

Dane dotyczące liczby zestawów danych badawczych według roku

Polska

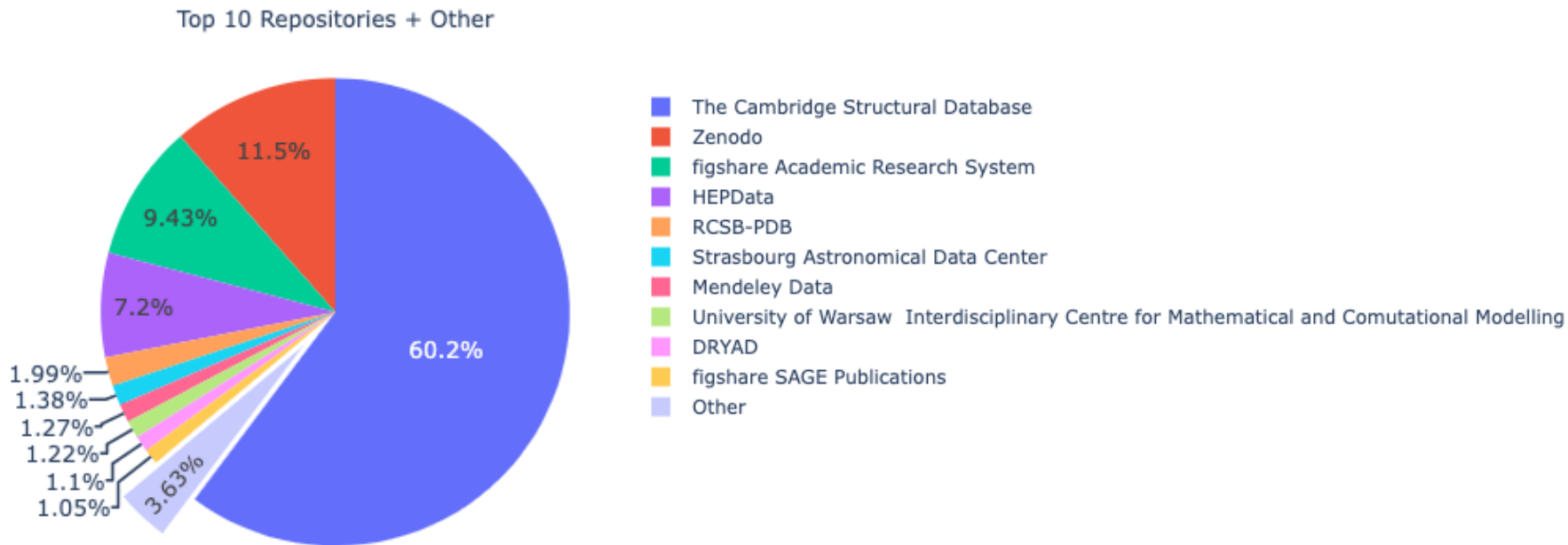


*CAGR (compound annual growth rate) obliczony w okresie 2012 – 2021

Source: **Data Monitor** Corpus
<https://www.elsevier.com/solutions/data-monitor>

% opublikowanych zestawów danych badawczych według repozytorium

Poland

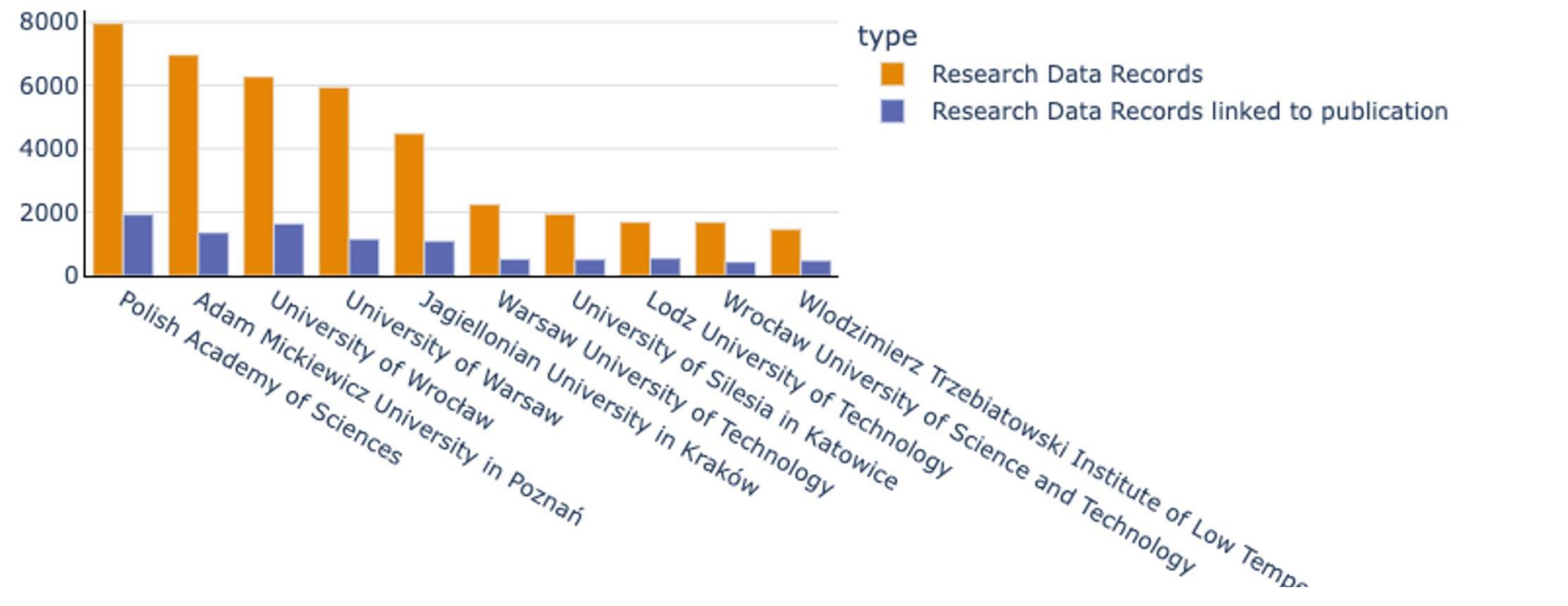


Top 10 Instytucji w Polsce

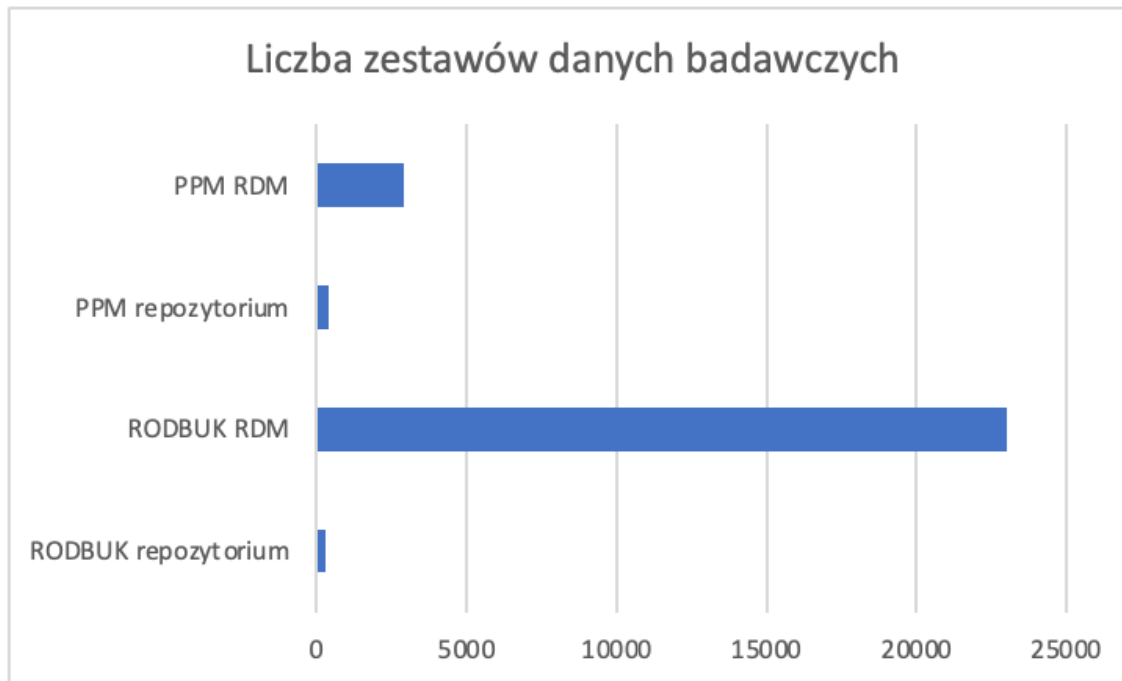
Poland

Source: **Data Monitor** Corpus

<https://www.elsevier.com/solutions/data-monitor>

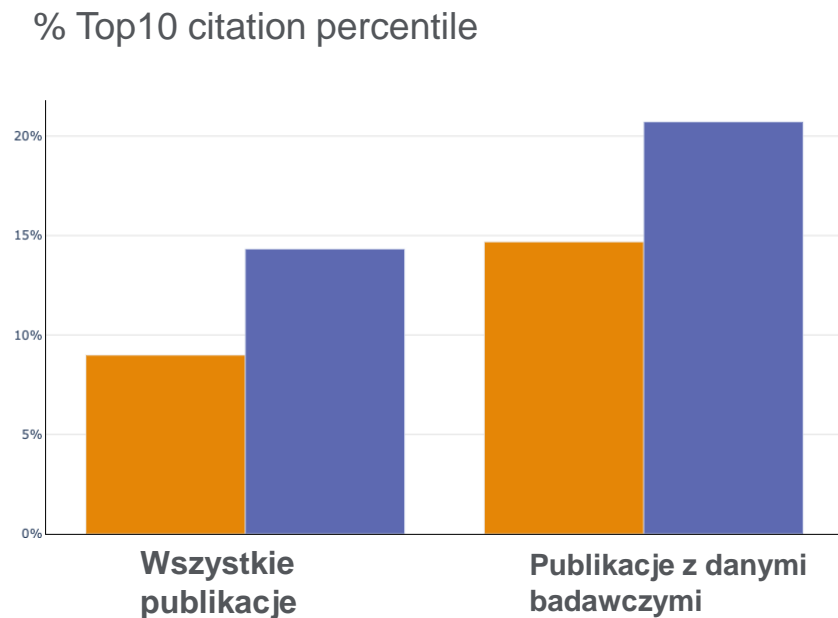
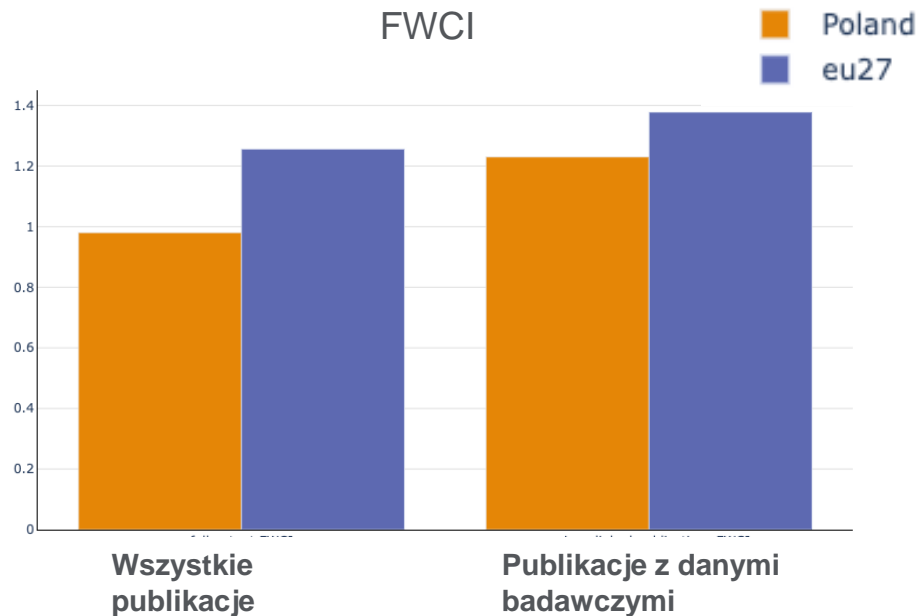


Repozytoria w Polsce



Budowanie widoczności

Wpływ publikacji z danymi badawczymi

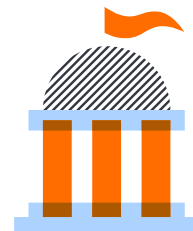


Korzyści

- Wzmocnienie pozycji badacza (mł. naukowiec)
- „Widoczność” z innej perspektywy (nie tylko komunikat naukowy)
- Transparentność procesów badawczych (i weryfikowalność)
- Interoperacyjność
- Oszczędność czasu
- Wzmocnienie procesów dydaktycznych (instytucja → kształcenie)
- Innowacyjność
- Wzmocnienie (dynamiki) rozwoju

Jak instytucje mogą śledzić swoje dane badawcze?

Instytucje zwiększają wysiłki w zakresie RDM

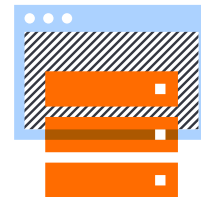


- Zatrudnianie data stewards
- Plany zarządzania danymi
- Repozytorium danych
- Sprawozdawczość (zgodność, ocena itp.) dotycząca danych
- Prezentowanie danych

Bardzo istotne z punktu widzenia strategii

Wyzwanie: Gdzie są dane badawcze uczelni?

Gdzie są Dane Badawcze Twojej Instytucji?



Do **90%**

- jest zdeponowana w jednym z 1000 repozytoriów dziedzinowych / ogólnych
- Jest to zgodne z najlepszymi praktykami społeczności i zaleceniami fundatorów (np. NIH)



Sources:

1: E. Zudilova-Seinstra; A. Zigoni; W. Haak (2020), "Analysis of research data for 11 Institutions – Data Monitor", doi: 10.17632/k5p45z33kb.3

Data Monitor pomaga pokonać 90% wyzwań

2000+
repozytoria
dziedzinowe i
ogólnotematyczne
(w tym DataCite)

Oczyszczanie i
wzbogacanie danych

Data Monitor
(corpus)

Webapp

Znajdź i śledź

API

**CRIS instytucji /
repozytorium**

Jak śledzimy dane badawcze

- Zbieramy metadane z ponad 2000 repozytoriów
- Normalizujemy metadane (schemat OpenAIRE)
- **Czyścimy**, np. usuwamy duplikaty, dane niebadawcze, np. artykuły; **Jak:** Techniki ML, integracja ze Scopusem, eksperci
- **Wzbogacenie metadanych** o publikacje/autora/instytucję **Jak:** Techniki NLP; integracja z Scopus, Scholix, DataCite

Co widzą i otrzymują użytkownicy:

- Webapp
- API
 - Integracja:
 - Pure
 - Other Systems
 - Batch

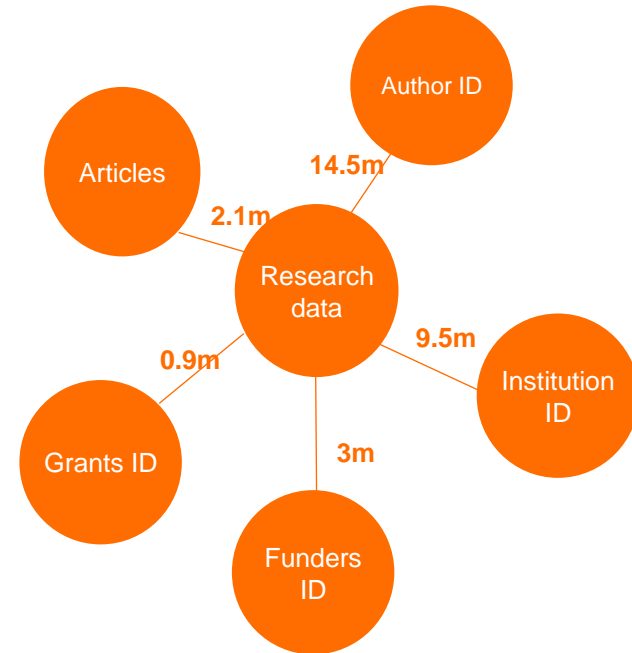
Data Monitor: Ciągłe ulepszanie zakresu, wzbogacanie i jakość

Wzbogacanie Danych

- Afiliacje instytucjonalne
- Dezambiguacja autorów
- Metryki: Cytowania, Wykorzystanie... i inne
- Publikacje, fundatorzy, granty.... i więcej
- Powiązane linki...

Czyszczenie Danych :

- Czyszczenie danych niebadawczych
- Usuwanie spamu i martwych linków
- De-duplikacja
- Zarządzanie wersjami



Source: Data Monitor Q2 2022

Przykłady studiów przypadków: pomoc wiodącym uniwersytetom w śledzeniu danych badawczych

Dane z badań zdeponowane w...

	... repozytorium instytucjonalne	...wszystkie repozytoria (specyficzne dla danej dziedziny, ogólne, instytucjonalne)
UK institution (Figshare)	234	16724
NL institution (Figshare)	534	16372
US institution (Dataverse)	46	3951
Chinese institution (home-grown)	1328	71760
Australian institution (DC Data)	11	785

30x- 70x



Dziękujemy!

Katarzyna Patyrak, Bartek Więckowski
Research Intelligence





Trendy w Zarządzaniu Danymi Badawczymi

Dane badawcze stały się wysokim priorytetem

- **Polityka fundatorów**

Większość fundatorów na całym świecie posiada politykę dotyczącą udostępniania danych badawczych



“No opting out” of RDM”

- **Polityka wydawców**

Wszystkie największe wydawnictwa (w tym Elsevier) prowadzą politykę w zakresie udostępniania danych badawczych.



- **Strategiczne programy rządowe**

Rządy i organy krajowe na całym świecie



Narodowe Centrum Nauki (NCN)

"We have set out to draw up a set of joint European guidelines on research data management and open-access publication methods. A practical guide to uniform European research data management practices was published in January 2019."

National Science Centre

Signed by Zbigniew Błocki, Director



NATIONAL SCIENCE CENTRE
POLAND



Source: https://www.ncn.gov.pl/sites/default/files/pliki/2019_04_03_pismo_dyrektora_NCN_zarzadzanie_danymi_naukowymi_ang.pdf

Rośnie zapotrzebowanie wydawców na RDM

Publikacje z danymi badawczymi



CellPress

THE LANCET

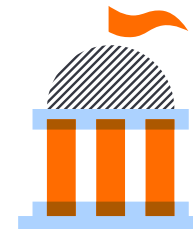


PLOS Open Data policy

Publishing in a PLOS journal carries with it a commitment to make the data underlying the conclusions in your research article publicly available upon publication.



Instytucje zwiększają wysiłki w zakresie RDM



- Zatrudnianie data stewards
- Plany zarządzania danymi
- Repozytorium danych
- Sprawozdawczość (zgodność, ocena itp.) dotycząca danych
- Prezentowanie danych

Bardzo istotne z punktu widzenia strategii