

# Projekt ELRC w Polsce

Maciej Ogrodniczuk  
Instytut Podstaw Informatyki PAN





## Przedstawiciele krajowi:

- Maciej Ogrodniczuk (IPI PAN)  
Przedstawiciel krajowy ds. technologii
- Anna Kotarska (NFZ)  
Przedstawiciel krajowy służb publicznych

## Pierwszy warsztat ELRC:

- 9 marca 2016 r.
- 13 prelegentów, 63 uczestników



## Po I warsztacie:

- pozyskano 18 zestawów danych
- łącznie ok. 2 + 178 mln słów
- 60 tys. par zdań
- głównie EN-PL, ale także jeden zasób wielojęzyczny i korpus polskich orzeczeń sądowych

## Dane uzyskane od Państwa:

- MR/ARiMR: kwartalnik Polish Food (3 zestawy danych)
- MSZ: opracowania dot. historii Polski, partnerstwa wschodniego, 10 lat Polski w UE, raport „Młodzi 2011” (3 zestawy)
- UZP: sprawozdania i terminologia wielojęzyczna (3 zestawy)
- ULC: słowniki terminologii lotniczej (2 zestawy)

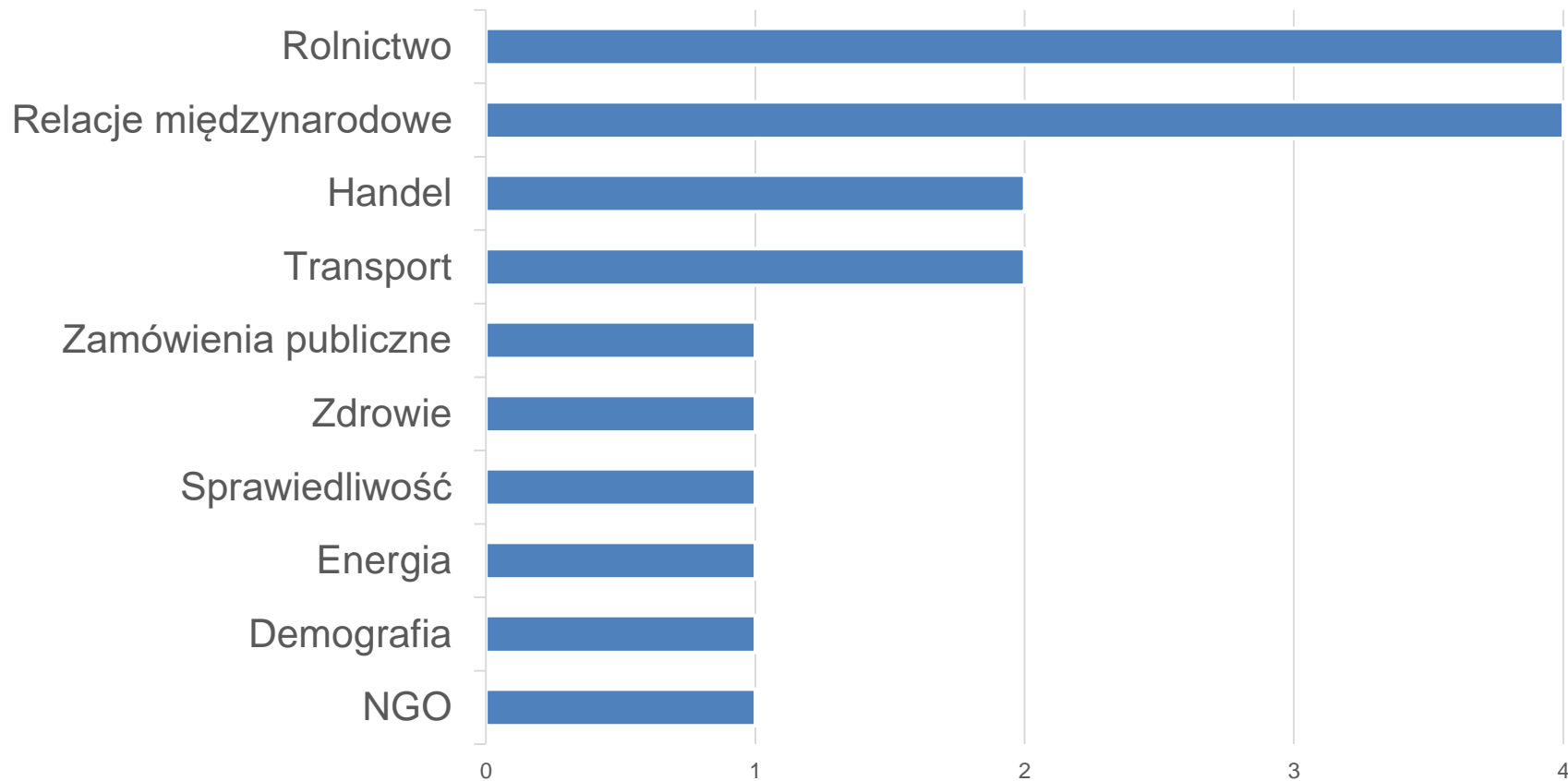
## Dane uzyskane od Państwa:

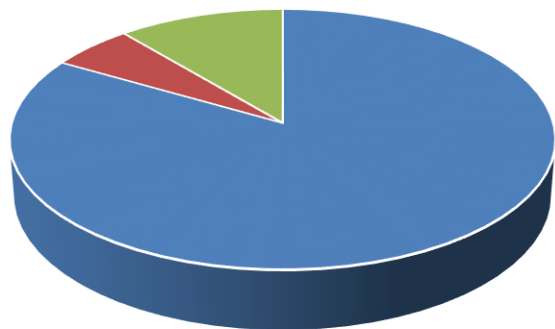
- Centrum Europejskie Natolin: analizy w „Zeszytach Natolińskich”
- GUS: rocznik statystyczny
- MS: orzeczenia sądów powszechnych
- NFZ: publikacja „Jednorodne grupy pacjentów w Europie”
- PAH/Oxfam: seria raportów dot. kryzysów humanitarnych
- PKN Orlen: raporty dot. rynku energii (gaz łupkowy...)



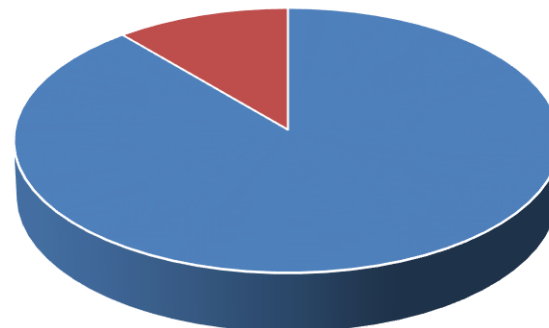
- Centrum Europejskie Natolin
- Główny Urząd Statystyczny
- Ministerstwo Rolnictwa
- Ministerstwo Spraw Zagranicznych
- Narodowy Fundusz Zdrowia
- PAH/Oxfam
- PKN Orlen
- Urząd Lotnictwa Cywilnego
- Urząd Zamówień Publicznych

# Dziedziny tematyczne





- Dwujęzyczne
- Jednojęzyczne
- Wielojęzyczne



- Korpusowy
- Terminologiczny



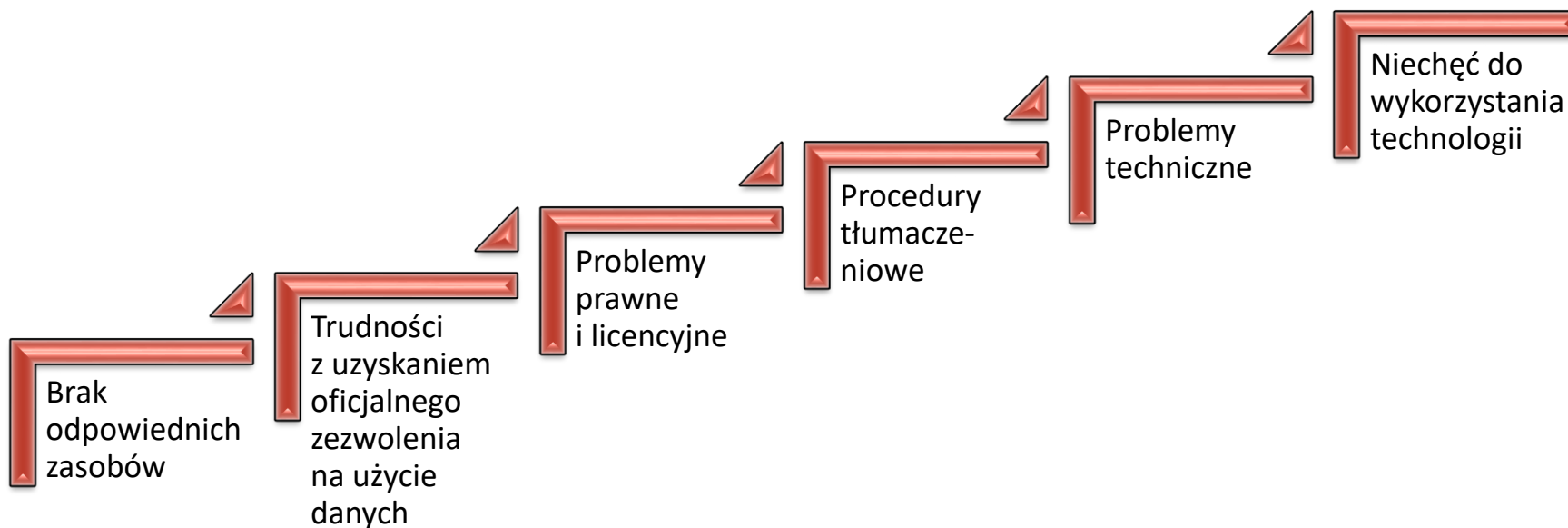
## Przetwarzanie danych:

- ujednoczenie formatu
- zrównoleglenie danych (zdanie do zdania)
- ręczna korekta
- zapis danych w formacie pamięci tłumaczeniowej
- opis metadanymi
- zapis zasobu w repozytorium ELRC Share

<https://elrc-share.eu/>



## Główne problemy:





**Czego  
się  
jeszcze  
nie udało  
zrobić?**

Wprowadzić jednolitych procedur udostępniania danych, tworzenia umów z agencjami itp.

Wprowadzić mechanizmów stałego pobierania danych

Przekonać niektórych instytucji, z którymi się kontaktowaliśmy

Co jeszcze powinniśmy poprawić?



## Co się udało?

Skontaktować się z Państwem!

Dołączyć do europejskiej czołówki  
w pozyskiwaniu danych

Pozyskać dane wielojęzyczne (nie tylko PL-EN)  
i terminologiczne

Pozyskać dane z wielu dziedzin (rolnictwo,  
zdrowie, zamówienia publiczne, prawo, sprawy  
zagraniczne, analizy biznesowe)

Dziękuję za uwagę!

E-mail: [info@lr-coordination.eu](mailto:info@lr-coordination.eu)  
Strona WWW: [www.lr-coordination.eu](http://www.lr-coordination.eu)

