

# Advanced Exploration of Public Procurement Data in Linked Data Paradigm

Krzysztof Wecel - I2G / Poznan University of Economics  
Vojtech Svatek, Jindrich Mynarz – University of Economics,  
Prague

## Motivation

### Starting point:

- potential benefits for a wide range of players
- mandatory publication of public contract notices

### Weak points:

- restrictions of search interfaces
- GUI in local languages
- no wider analyses available: aggregations, trends, patterns...
- lack of mechanical reasoning
- geographical context not leveraged
- no links to external information

### Opportunity:

- **representation in a form of linked data**



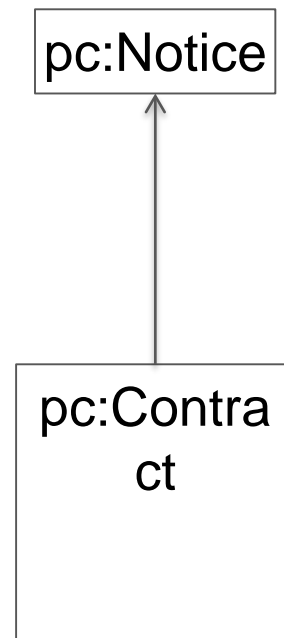
# Public Contracts Ontology



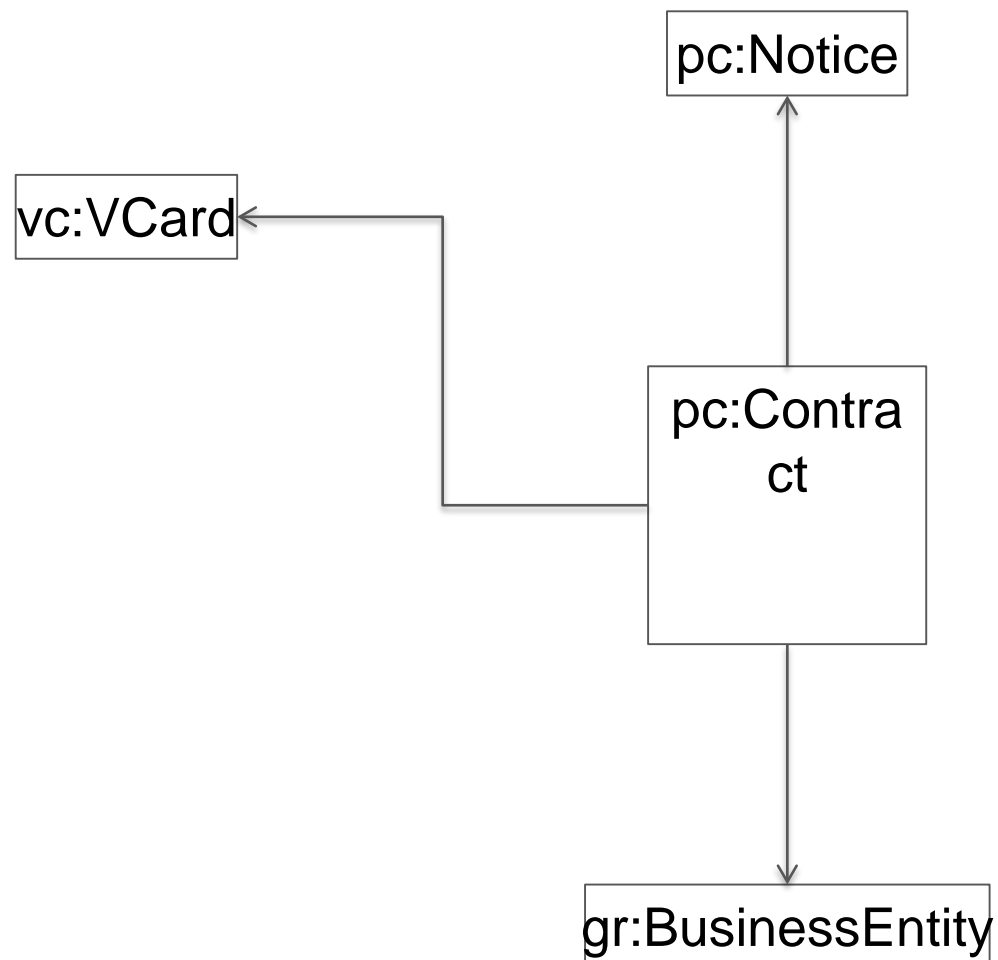
# Public Contracts Ontology

pc:Notice

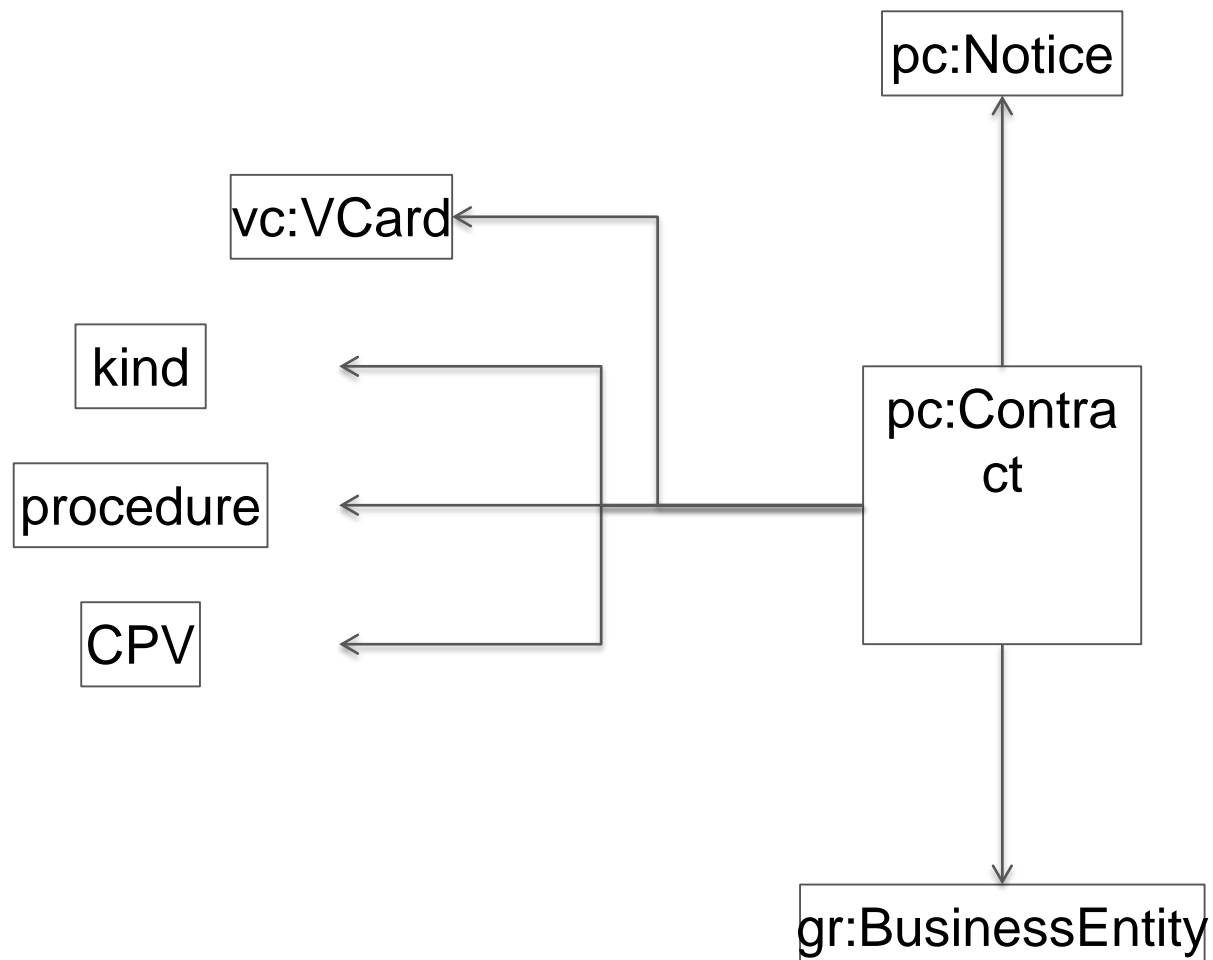
# Public Contracts Ontology



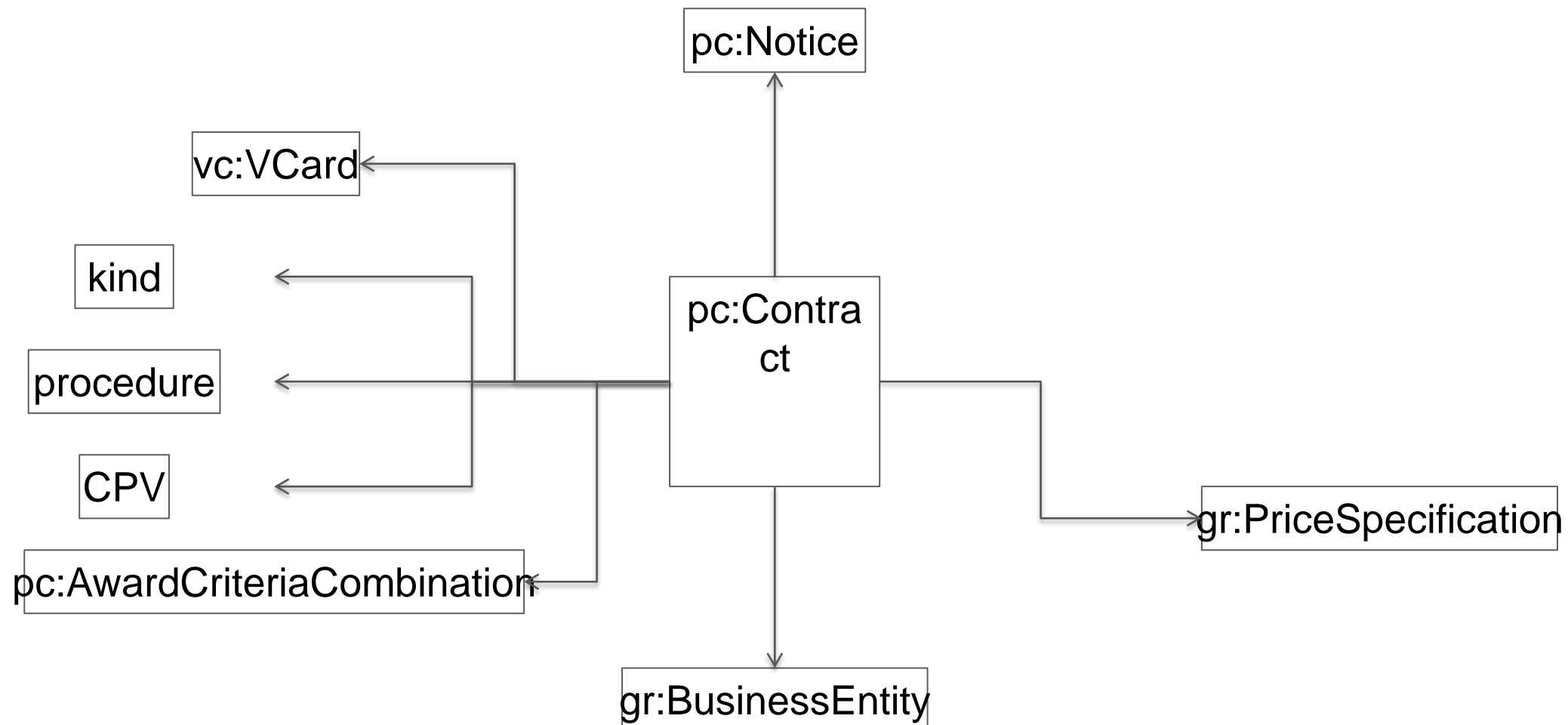
# Public Contracts Ontology



# Public Contracts Ontology

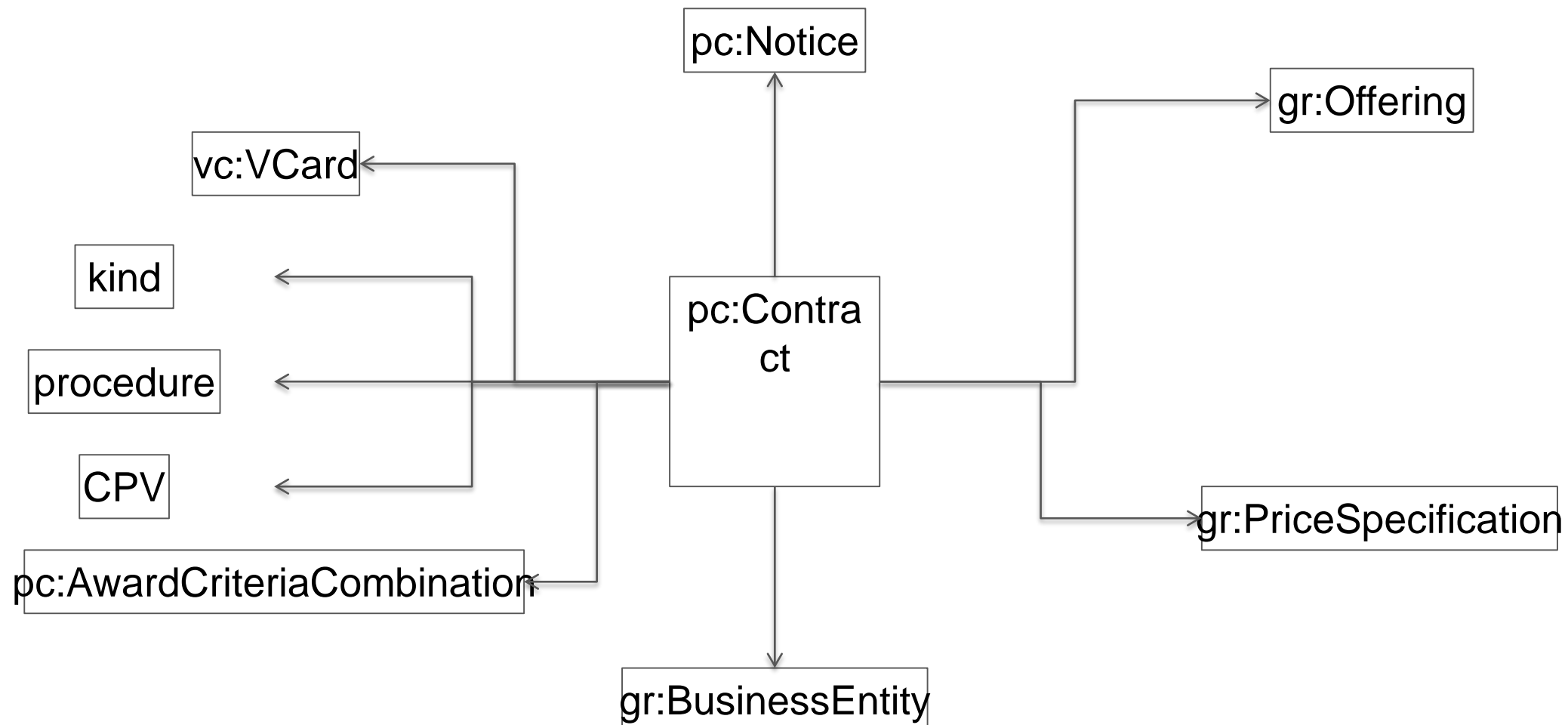


# Public Contracts Ontology

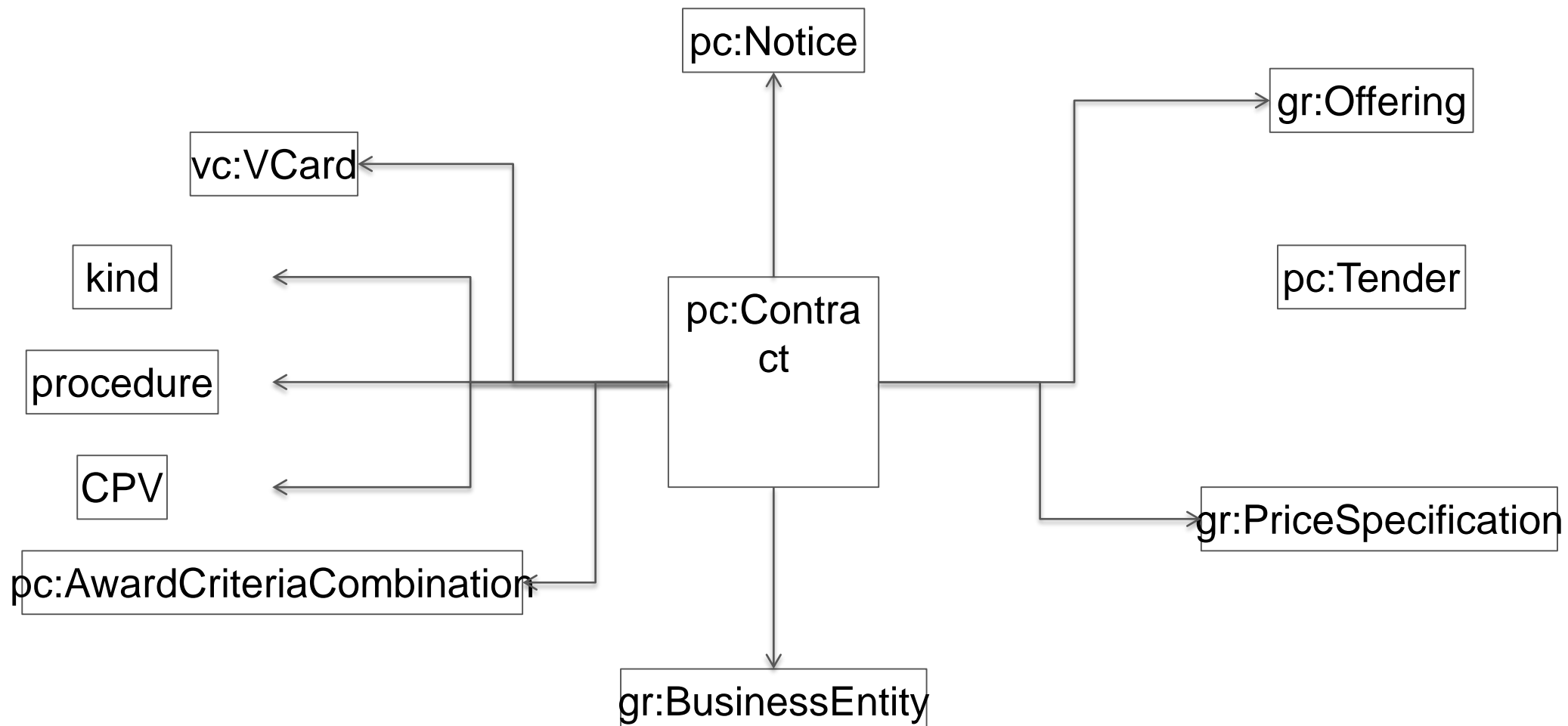




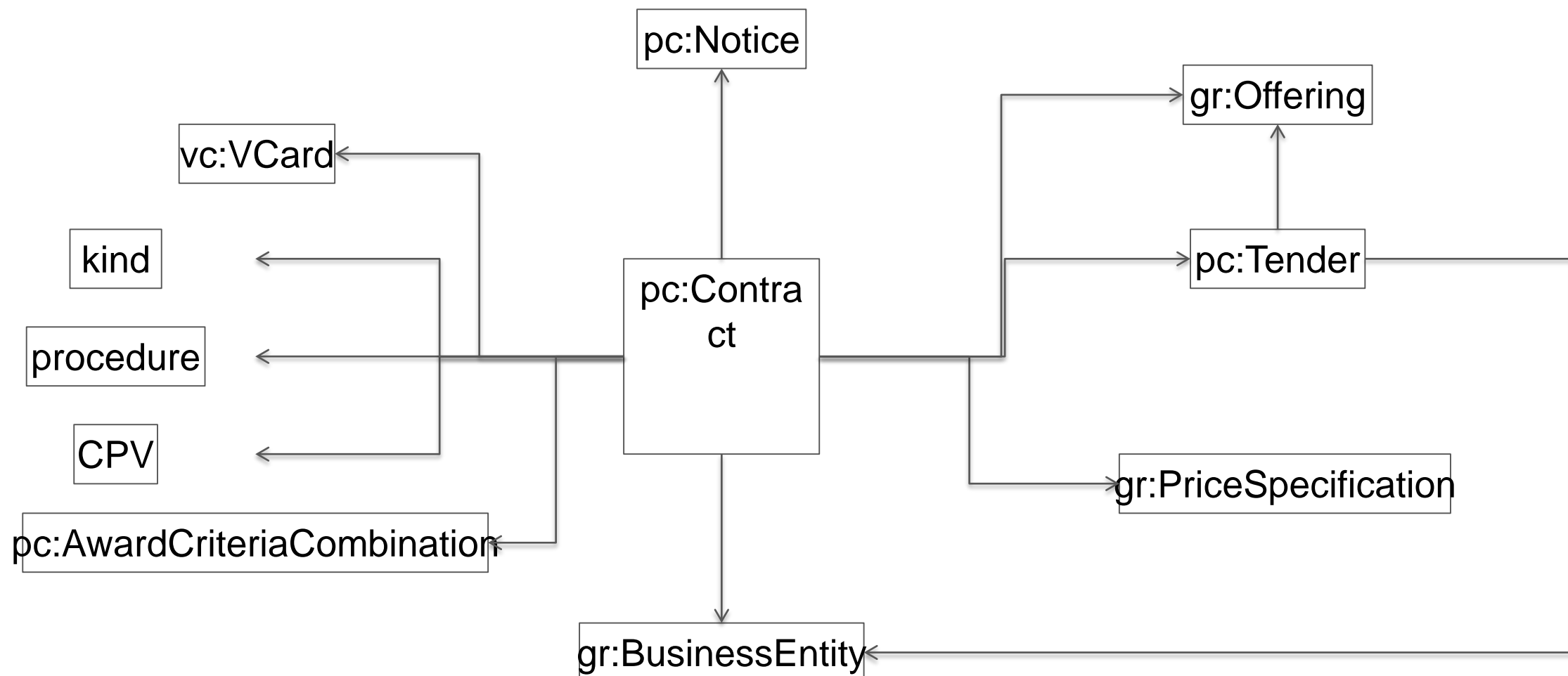
# Public Contracts Ontology



# Public Contracts Ontology



# Public Contracts Ontology



## Data sources and transformation

### Public contract notices:

- HTML – navigation, scrapping
- XML – modeling approach
- mapping issues

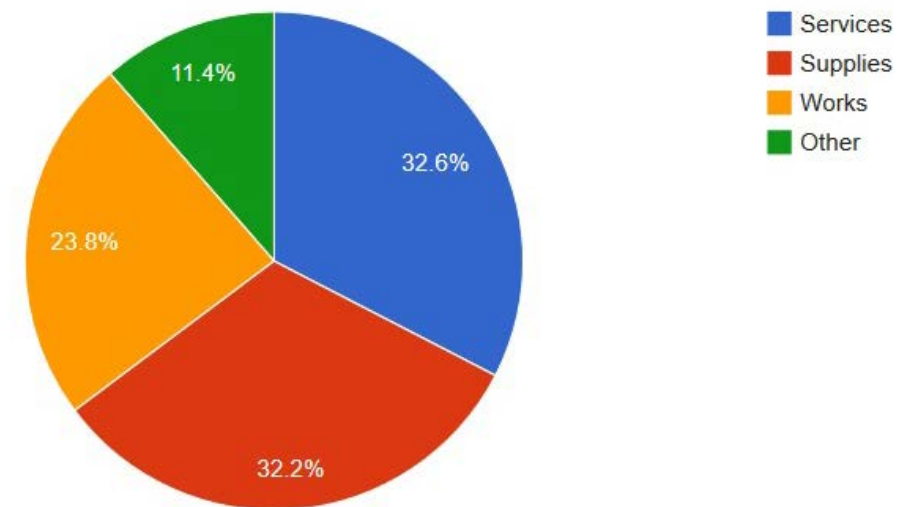
### Additional data – the real value

- business entities (ARES, CEIDG) and their codes (ICO, NIP, REGON)
- geographical codes (NUTS, TERYT)
- geographical coordinates (geocoding)
- CPV and other vocabularies
- optional external information
  - Czech Trade Inspection Authority
  - sentences of Polish National Board of Appeal

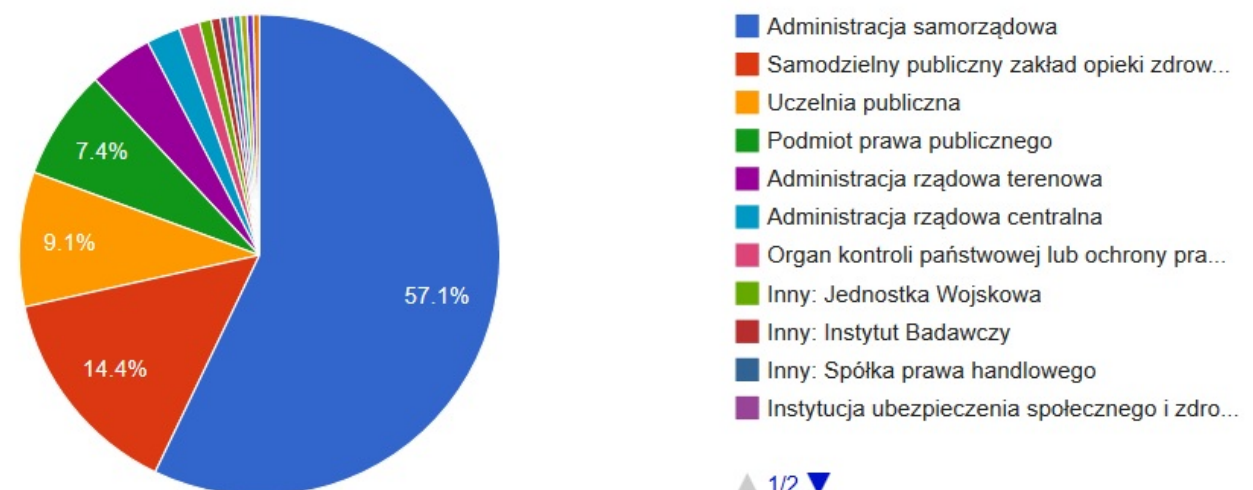
## Polish dataset characteristics (2013)

number of triples: 28,8M  
 notices: 413,382  
 offerings: 922,038  
 contracting authorities: 17,648  
 contractors: 177,136  
 business entities: 194,784  
 unique CPV codes: 11,34<sup>1</sup>

Case: Number of contracts by type (from ZP-400 to ZP-408)



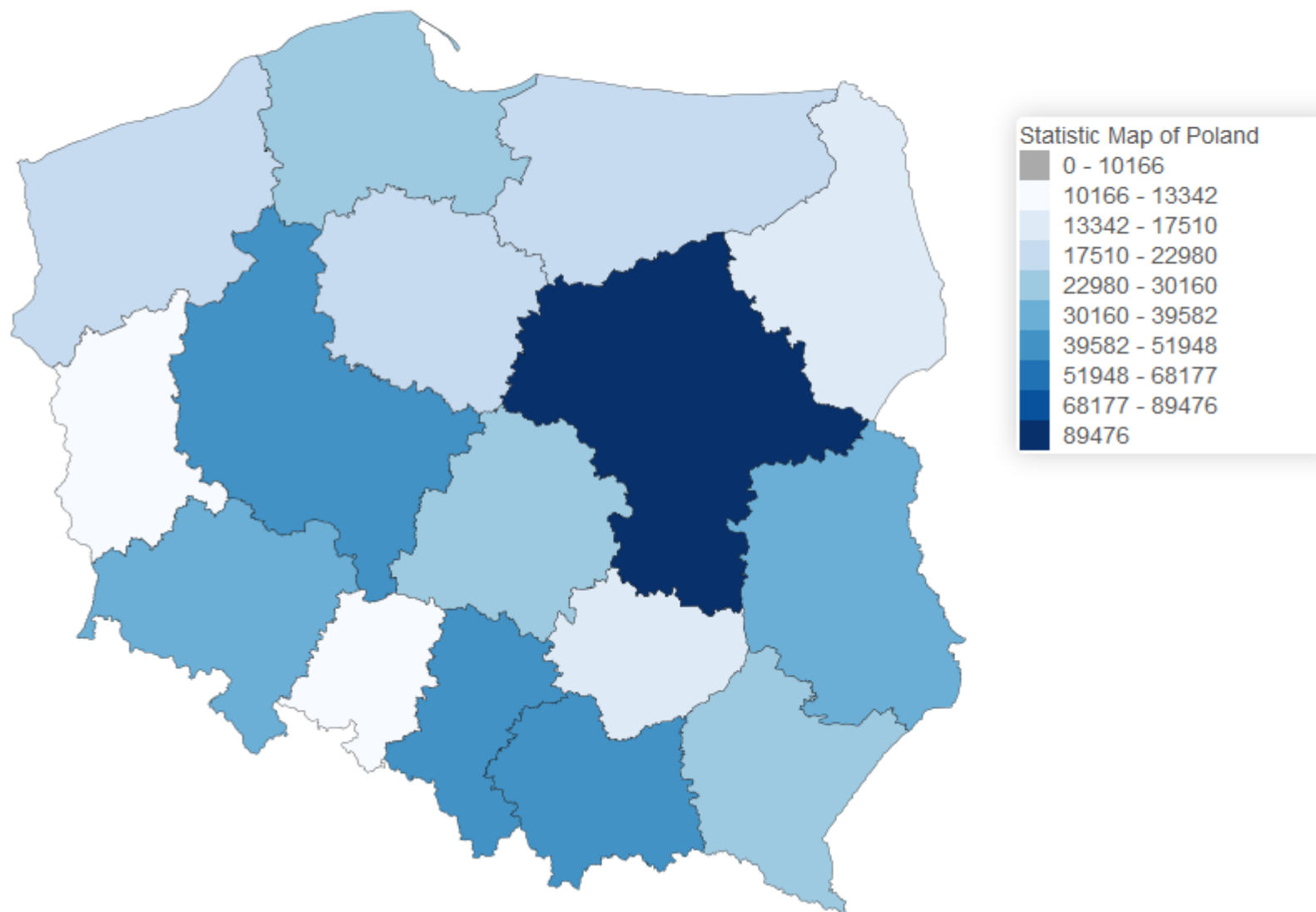
Case: Number of contract notice by type of entity (TOP 15)



## Polish dataset characteristics (2013)

Query:

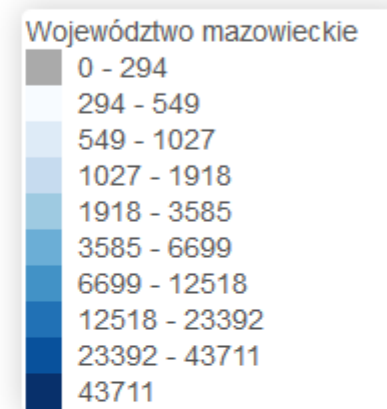
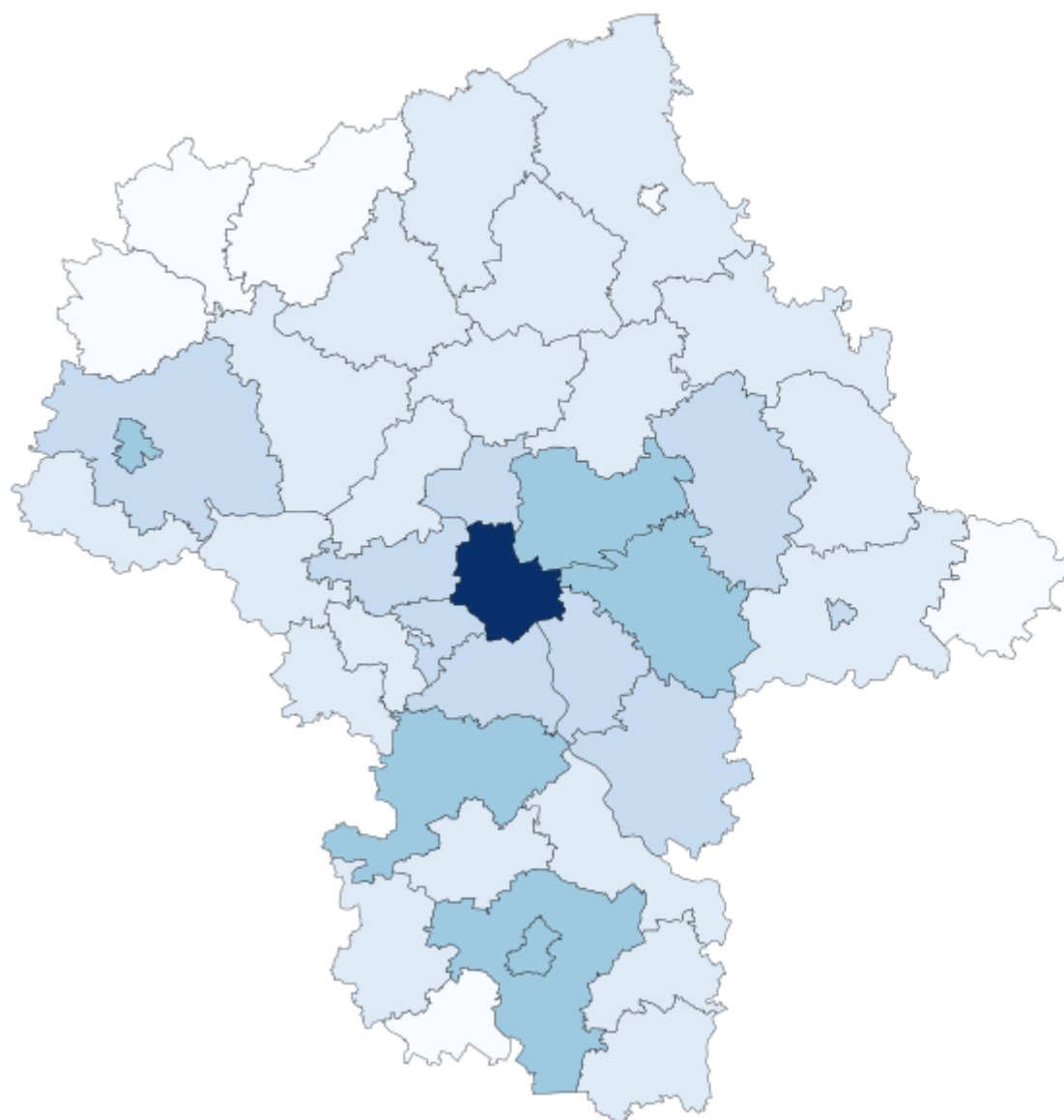
Statistic Map of Poland



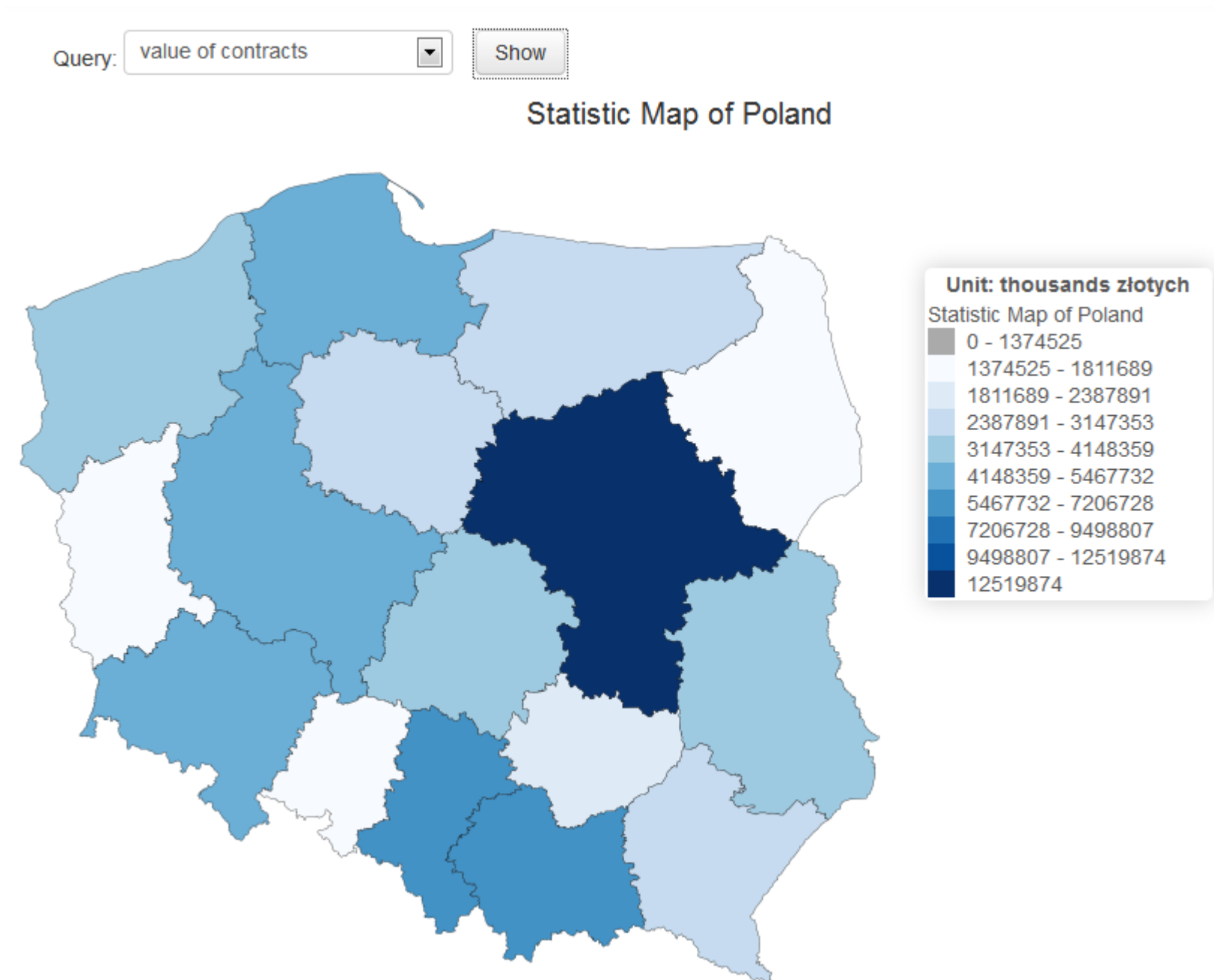
# Polish dataset characteristics (2013)

Query:

Województwo mazowieckie

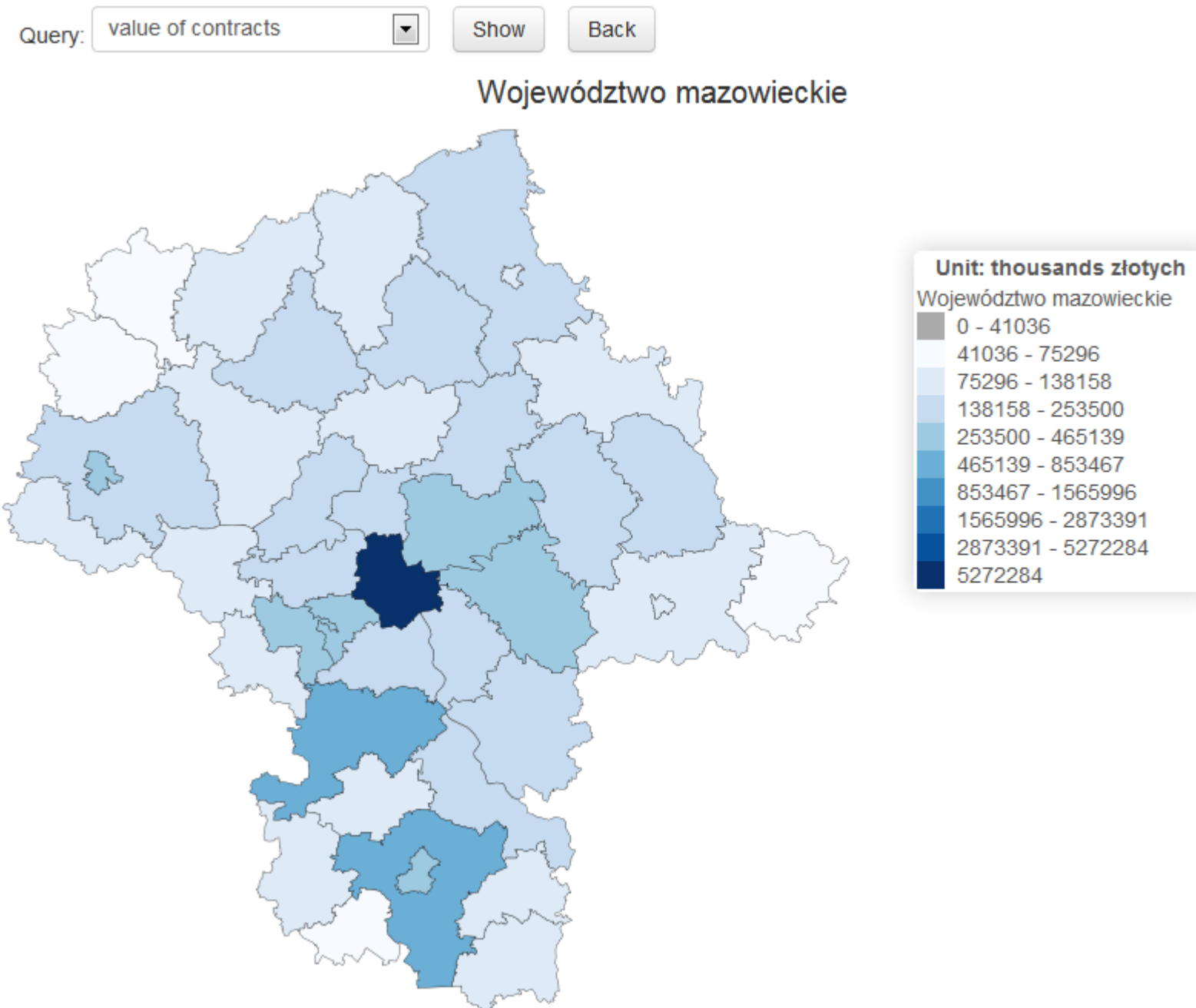


## Polish dataset characteristics (2013)





# Polish dataset characteristics (2013)



## Data mining

leverage analytical and data mining techniques in order to find patterns, trends and anomalies in public contract data

Specific problems of graph data:

- multidimensionality
- big number of potential attributes
- overlap of the classes
- unbalanced learning data (different counts of classes)
- loss of information during transformation from graph to tabular data

## CLUSTERING: looking for similar contracts

beneficiary: bidder

case:

- identify contracts from the past that would be most suitable
- monitor new notices similar the contracts they have already realised
- more expressive than typical search language

beneficiary: contracting authority

case:

- help in preparation of specific contract notice (not only CPV)
- aggregated demand opportunities

## ASSOCIATIONS: ties between various market players

beneficiary: supervisory bodies

case:

- discover anomalies
- contractor-product association: stability of the offer; the tighter the relationship, the more reliable the contractor is
- contractor-authority association: signals the need to check for corruption
- analysis of the depth of the market

## PREDICTIVE MODELS: number of bidders

leverages the link between notices concerning the same procurement process

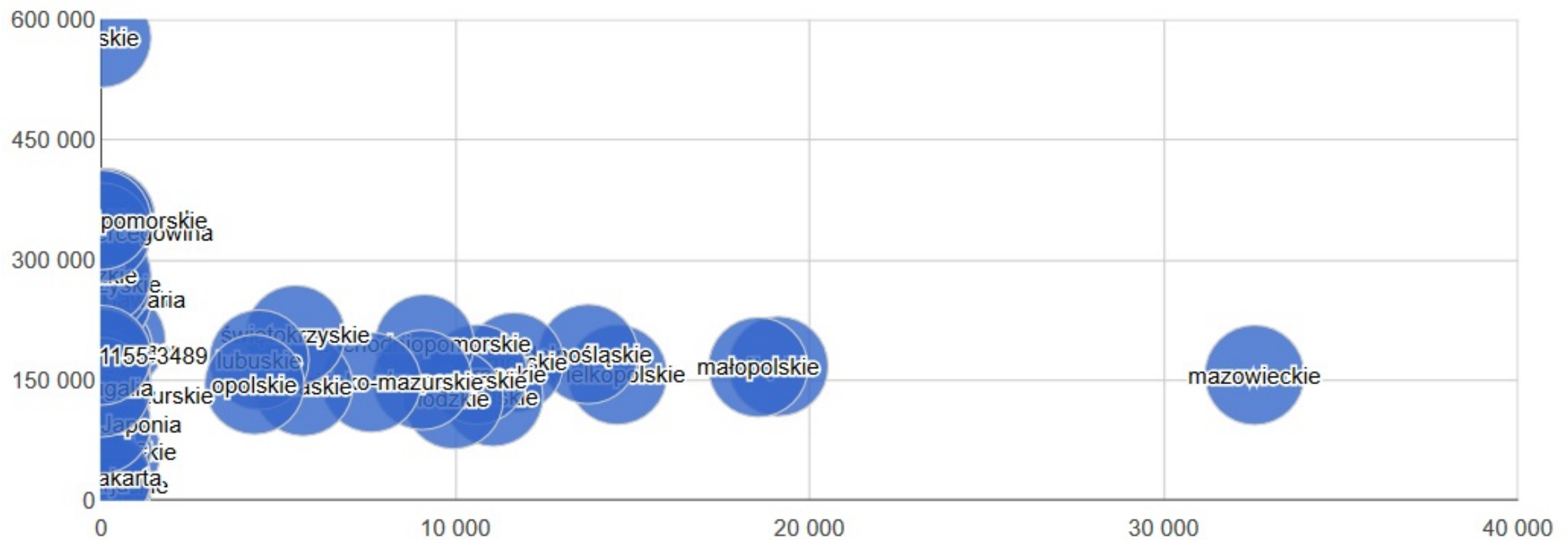
beneficiary: contracting authority

case:

- the bigger number of bidders, the better
- one bidder can mean overspecified contract notice

# Geography vs. value association

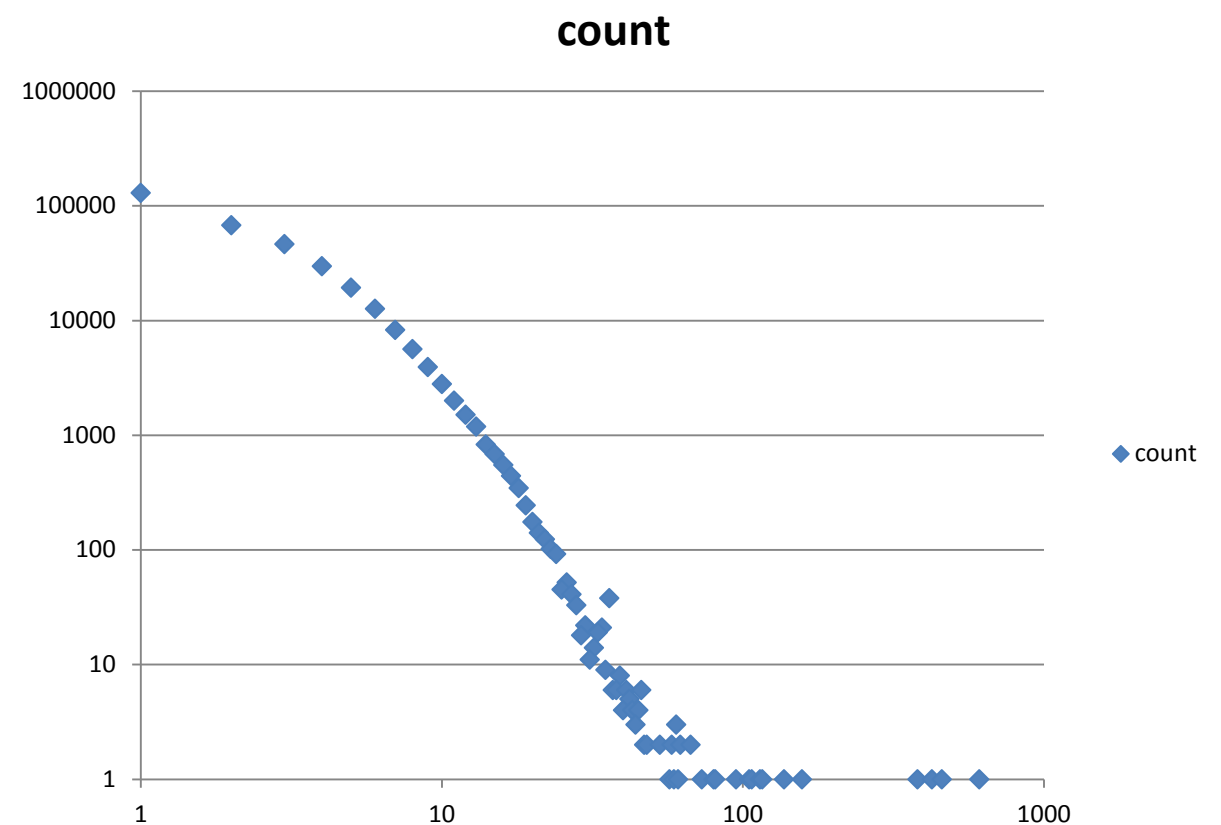
**Case:** Relations of number of contracts average estimated value by province



## Number of tenders - long tail

- 38% of contracts had just one offer
- one bidder = overspecified contract notice?
- one contract had 610 offers
- in some cases large numbers of rejected offers: 298, 245, 111

Number of Tenders	count	Percentage
1	129910	38,2%
2	67929	20,0%
3	46412	13,6%
4	29850	8,8%
5	19367	5,7%
6	12679	3,7%
7	8277	2,4%
8	5639	1,7%
9	3920	1,2%
10	2799	0,8%





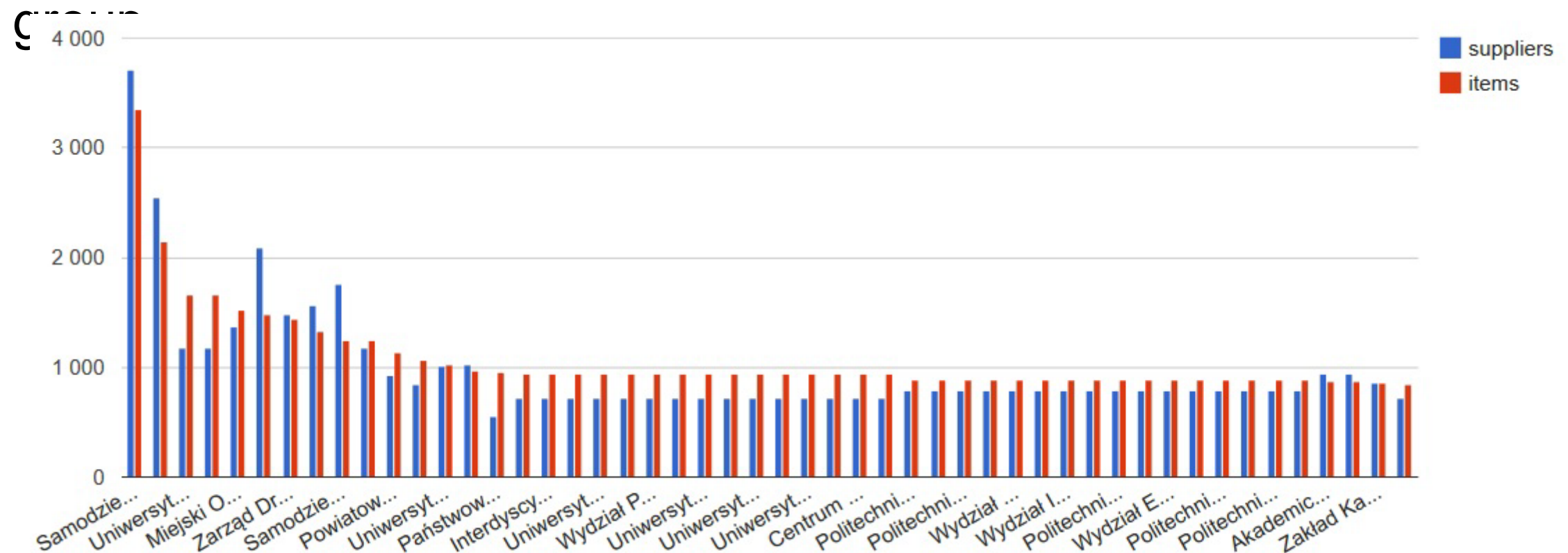
## Some other discoveries

### Type of contract vs. number of bidding entities

- supply contracts are the most popular
- construction work was the least popular
- areas of low competitiveness are more susceptible to abuse

### Number of contractors to number of notices ratio

- should be compared to typical value in similar contracting authorities





## Conclusions

### Public procurement in Czech Republic and Poland

- ontology has been elaborated
- significant amount of data gathered
- we are looking for other interested parties

### Data mining

specific issues of graph data have to be addressed  
old and new tools applied

- similar contracts by clustering
- ties between various market players
- prediction of the number of bidders