Open data and data sharing. The example of public procurement data

Mihály Fazekas
Central European University and Government Transparency Institute
misi.fazekas@gmail.com

Open Research at Cambridge 2021. Open data sharing and reuse. 23/11/2021
Data shared

- Government contracting datasets from around the world

- Each record is a tender/contract/purchase

- Data enriched with corruption risk indicators
Open data: GTI’s database: A unique global scope

- Compiling and standardizing government contracting datasets is the hardest part of our work
- Over 55+ million public contracts from 50+ countries in a standardized format
- Over 6 million government suppliers and 1 million public organisations

GTI national public procurement data country coverage

Interactions

• European data: opentender.eu
  – 77K users in the last 12 months
  – 5K downloads in the last 12 months

• Global data (new portal): procurementintegrity.org/
  – 1-2K users per month
User examples

- Academia: 143 citations to the database and over 1000 citations to papers using the data
- Policy: G7 anticorruption package, World Bank, etc.
- Civil society: Transparency International, Open Contracting Partnership, etc.
- Media: WSJ
- Investigators: European Investment Bank, Swedish Competition Authority
Supporting investigations: Organisational Risk
European Investment Bank (EIB) loan recipients’ risk distribution

- EIB auditors use CRI scores to identify the highest risk entities for audit
- Scope: over EUR 50 billion annually, 1500+ organisations, 500,000+ tenders, 10 tailored red flags
Why publish your data and codes?

• Your funder requires to do so
• You want to increase the impact of your research
• You have way more data than what you can reasonably analyse in a lifetime
• Making your analysis fully reproducible lends your credibility
• You are just annoyed by constant data request😊
Challenges of data sharing

• Difficult to control what people (or bots) do with your data
• Selecting the right license (commercial vs non-commercial)
• Properly sharing is lots of work (but good excercise to yourself too!)
• Replicating your own work from scratch is a very useful excercise
Looking forward to the discussion!
Further resources

http://www.govtransparency.eu/

http://redflags.govtransparency.eu/

https://www.researchgate.net/profile/Mihaly_Fazekas/research