

Iceland
Liechtenstein Norway
Norway grants grants

Todor Galev

Assets declarations, corruption risk and public procurement

Steps for collecting, integrating and analyzing data

18-20 May 2022

Implemented by:



Overall objective

- Analyse the applicability of asset declarations as corruption prevention and risk assessment instrument, from legal, institutional, and practical (on-the-ground-implementation) perspective.
- Uncover which are the politically exposed companies and/or showcase their presence and impact on the economy.
- Expose cases of misconduct by politically exposed persons, related to the information presented in their asset declarations, including through actions of their immediate relatives, and if possible close business associates, straw persons, etc.

Collection of data

Access to data:

- Online (on a web-site)
- Offline (under request for information)

Format:

- Machine readable (docx, pdf, xlsx, csv, xml, ...)
- Scanned pictures (pdf)
- On paper

Structure of the data:

- Structured data in tables / database (csv, xls, xml)
- Structured data case-by-case (docs, pdf, csv, xml)
- Unstructured data (scanned pictures)

Collection methods

Online sources:

- Manual export of tables / database (if available)
- Manual save /copy of records/tables using third-party software/browser add-ons:
 - Save Page WE (html) : **EXMPL**
 - Table capture / Copytables
- Automatic download (if possible):
 - Bulk download using third-party software/browser add-ons:
 - DownThemAll! (**EXMPL BG AD register in Firefox** - <https://register.caciaf.bg/2020/index.html>) / Simple mass downloader
- OCR (in case of paper / picture based PDF)
 - Adobe Acrobat Pro
 - ABBYY Finereader (Screenshot reader)

Transforming data

Separate XMLs into a single table/sheet:

- XML converters (open source, paid/trial)
- VB scripts in Excel
- Online services /limitations/
- Software tools (e.g. Open Refine, <https://openrefine.org> or Datashare, <https://datashare.icij.org>)

Cleaning and structuring the data

- One “value” per cell or using standard separator (e.g. semicolon)
 - Example: replacing “line breaks in Excel” (CTRL + J)
 - Example: find first number in a text string
 - Example: recode “position”
 - Tip: ask Google

- One record per row (i.e. multiply columns)

Terms in position

(from year – to year)

Prime Minister 2013 - 2022

Mayor of Tirana 2000 - 2011

Minister of Culture, Youth and Sports 1998 - 2000

41	адм. ръководител, председател
42	Адм. ръководител, районен прокурор
43	Адм. ръководител-Окръжен прокурор
44	Адм. ръководител-Председател
45	Адм. ръководител-Районен прокурор
46	адм.р-л-председател
47	Адм.ръководител
48	адм.ръководител - председател
49	Адм.ръководител - Районен прокурор
50	Адм.ръководител- окръжен прокурор
51	Адм.Ръководител -Председател
52	Адм.Ръководител- Председател
53	АДМ.РЪКОВОДИТЕЛ РАЙОНЕН ПРОКУРОР

Matching / merging data:

Asset declarations data:

- PEPs
- Company owned/managed by PEP
- PEP's family members owning/managing company (persons)
- PEP's family members' companies

Challenges in asset declarations' data:

- National specific -> initial and follow-up declarations (disclosing only changes, i.e. need to collect data from previous years)
- De-duplication (if needed)
- Partial disclosure:
 - Not all companies / relatives disclosed
 - Partial data (e.g. company ID)

Corporate data:

- PEP/ PEP's relative / PEC -> check for ownership/management in other companies than disclosed, i.e. companies A & B
- Check companies linked with (related to) A&B (e.g. subsidiaries or mother companies of A & B)
- (optional) Check owners/managers of linked/related companies

Matching / merging data (2):

Challenges in matching asset declarations' and corporate data:

- Duplicate person names (if an unique ID is missing)
- Different spelling of the person /company name ; use of abbreviations / short company name
- (non)existing legal form (e.g. Ltd, JSC, etc)

Possible solutions:

- Find additional data to verify the match (e.g. city or address)
- Manual cleaning (e.g. of abbreviations/short names)
- More sophisticated techniques (with external help), e.g. performing “elastic search”

Adding corporate data:

- Financial data (companies' financial accounts)
- Number of personnel data
- Industry data (e.g. NACE)

Finding / Obtaining corporate data

2 approaches:

- Looking for a single company (ad-hoc case investigation)
- Looking for bulk data (list of companies)

Sources:

- National business registers:
 - Domestic web-site
 - EU e-Justice portal of Business registers, https://e-justice.europa.eu/content_business_registers-104-en.do
- National / international paid corporate database / information system;
 - **EXMPL: Orbis, APIS/CIELA (BG), ...**
- Open sources:
 - Open Ownership Register, <https://register.openownership.org> (16+ mln beneficial owners)
 - OpenCorporates, <https://opencorporates.com>, (206+ mln companies; only publicly available info)
 - Offshore leaks database, <https://offshoreleaks.icij.org>
 - (optional) OCCRP Aleph, <https://aleph.occrp.org>
 - (optional) OCCRP Catalog of databases, <https://id.occrp.org/databases/>

Next steps

Matching AD & corporate and procurement data:

- When we work on a limited scale:
 - Verify how fully all circumstances are declared: matching PEP/PEC with business registers (using both Orbis & national business registers)
 - Compare key economic performance indicators for PECs and peer companies (limitations!)
 - (optional) review media investigations on PEPs & PECs (both for declared and undeclared) using AI-based media monitoring tool (sensika.com)
- When we work on a mass scale:
 - Compare key economic performance indicators for PECs and peer companies
 - Company efficiency
 - Profit margin
 - Compare procurement success of PECs with peer companies

Iceland
Liechtenstein
Norway grants



Norway grants

**Thanks for
watching!**

