Technological tools for transparency and accountability
Landscape of CEE region

By Krzysztof Madejski, ePaństwo, Poland
In cooperation with Anna Kuliberda, Techsoup, Poland

This research is part of the “Technology for Transparency & Accountability Network CEE, Balkans & Eurasia” project funded by the U. S. Department of State.

The data of this research can be found and search through on the website of the project: www.transparencee.org.

In the era of information abundance, IT tools and processes are necessary. We use them to filter and emphasize relevant information, simplify and visualize it so that everyone can understand it.

You can use this publication under the Creative Commons BY-NC-SA 4.0 license.

Initiatives in a Spotlight

Public institutions unwillingness to cooperate, low level of transparency and corruption cases form a part of a superstitious image of the CEE region and there is an element of truth to it. It inspired many CSOs, many of watchdog origin, to independently roll out initiatives that would raise level of transparency and would held officials accountable to their actions. Big part of these initiatives that relies on use of modern technologies is the focus of TransparenCEE Network and this report.

Active CSOs managed to develop tools that challenge public sector or present them best practices in transparency field. There is also some work done on transparency from the public sector side, both national and local. We intentionally leave public initiatives out of
our focus (mostly: data portals and participation-based sites) with exception of those that were done with and by local CSOs or are exceptional feature-wise.

Scope of analyzed tools is quite close to what is understood as civic tech. As the [Knight Foundation report](https://knightfoundation.org/openspending$fields) described #OpenGov: “projects enabling top-down change through the promotion of government transparency, accessibility of government data and services and promotion of civic involvement in democratic process” with exception of the change here being done through bottom-up approach by CSOs, with no or little government collaboration.

I. Solutions by Areas

Most of the existing solutions seem to grow from watchdog-like activities: monitoring parliaments and local councils, courts, procurement, budgets, media and promises made by politicians. There are also some projects developed to strengthen civic engagement and participation: e-petitions, civic reporting, etc. usually run by global rather than local players.

1. Parliament Monitoring

One of the first tech for transparency tools developed dealt with parliament monitoring. It presented how bills were processed: from an idea, through commissions, consultations and hearings all the way to the voting. The tool allowed people to check details about every MP including his/her speeches. Some initiatives extend the scope of statement monitoring and collect data also from social and traditional media.

National level tools can be also implemented locally to monitor representatives and bills at this level.

Solutions in this area include [ParlData](https://parldata.org/) monitoring a dozen of CEE parliaments: Polish [MyCountry](https://mycountry.pl/) and [I have the right to know](https://www.ihr2k.com/), Slovak [Datanest](https://www.datanest.sk/), Slovenian [Legislative Monitor](https://www.monitorpravilnikov.si/), Lithuanian [Mano Seimas](https://www.manoseimas.lt/) are good examples.
2. Court Monitoring

Court monitoring focuses on court rulings with some metrics and aggregations around it. Database of accessible, copiable and structured rulings form its foundations. What is a workload in different courts, how effectively they work, what are the budgets, who is being sued the most? – these among other metrics allow to compare courts, rulings and look for patterns.

Solutions include previously mentioned MyCountry and Legislative Monitor, as well as Lithuanian Open Courts.
3. Budgets

Budgets at city level, not mentioning level of state, are at scale that’s not intuitive to citizens. Budget visualizations and other means of presenting financial data are an attempt of making it less abstract, which is not an easy task. Basic visualizations are using OpenSpending engine or similar designs. Quite recently tax calculators became a thing downsizing budgets at human scale – type in how much you earn and we will show you where your taxes go. You can see it in the Belarusian Cost of the state, Polish Transparent City or MyCountry portals. There are also budget games that allow users to modify existing budgets and propose their own version. When used by a representative number of people – it’s a tool for advocacy on its own.

Interesting examples of budget visualizations come from the cities. On nicely styled infographics they present costs of a city per item: how much does it cost to plant a tree, maintain it, build 100m meters of an asphalt road, bike lane or a pavement? This balances between data aggregated in categories and specific contracts allowing citizens to understand differences in costs.
Solutions include Belarusian Cost of the state, Polish Transparent City and Our Money, and Serbian Vodite Racuna among others.

4. Procurement

How public money is spent has always been the focus of watchdog organizations. Because of the high volume tech tools necessary to process it, solutions vary in features. The base is always to publish procurements with advanced filtering, the variations are navigating between them and simple aggregations (top contractors, top agencies). Some solutions go deeper and link money to specific people rather than companies (making daughter companies more transparent). Some (like Slovak Open Contracts) use crowdsourcing to assess tenders, some use automatic processing using indicators (like Hungarian Red Flags).

Data availability differs, with EU countries obliged to publish every tender above 30k EUR limit. Formats of the data differ too and most of contracts are being published either in HTML or PDFs. Ukraine has a chance to be the first country with standardized data on procurement available through API when Prozorro platform will be officially transferred to the state.
5. Anti-corruption

Anti-corruption activities focus on three areas: procurement monitoring mentioned above, political parties income monitoring, and building corruption cases database. Corruption cases database, not mentioned before, are searchable websites consisting of cases collected either by media monitoring, or anonymous reporting. Reports can be made on the website or through dedicated mobile apps.

6. Truth-o-meter – Fact Checking

Truth-o-meters, known to the US public through the Politifact.com website are quite common in the CEE region. There are at least two major tools being scaled to different countries: Demagog operating in Visegrad region (Poland, Slovakia, Czech Republic and Hungary), and Istinomjer operating in Balkans (Serbia, Bosnia & Herzegovina, Montenegro, Macedonia and recently Croatia).
Both tools are quite simple and offer a similar functionality: they monitor the truthfulness of statements of public figures (politicians, etc.) saying if it’s true, false, a manipulation or it cannot be verified. Apart from checking what has recently been said, a user can also point at a new statement and ask for evaluation. A period of the highest activity of these sites aligns with elections – all the promises made by politicians and political parties are stored and then evaluated throughout the electoral term.

The power of this tool comes from the community of verifiers who are doing the hard work. Search feature makes sure that no registered statement will be forgotten.

7. Smart Voting

Smart voting is a theme for initiatives that describe the candidates we vote for. All of initiatives allow to compare our political views with the ones declared by candidates. Some of them evaluate candidates after elections when they become MPs. One can then compare MPs views with how they actually voted.

Solutions include Bosnian Glasometar, Polish Latarnik Wyborczy and I have the right to know, Lithuanian Mano Seimas, Czech and Slovak Election Calculator.
II. Technical side: Data processing

Data is the key to the most of the aforementioned initiatives and a process of acquiring and transforming data to structured formats is often the most expensive element of a tool development and then its maintenance. A lot depends on the formats of data which is why advocacy for better data quality is inseparable from IT works.

1. Open Data
The most adopted scheme for classifying data quality and its openness is Sir Tim Berners-Lee's 5-star deployment scheme for Open Data. While experts and data receivers should always be consulted on formats, the scale provides a clear guidance on how to enhance quality and gives an overview how much work is there to be done.

**Star One** is publishing data in an open license in any format.

**Star Two** is publishing data in a structured format which is relevant to the type of data. If it’s a text, publish it in a copyable text document. If it’s a table, publish it in an Excel spreadsheet.

**Star Three** is publishing data in non-proprietary formats. While Excel ‘97 spreadsheets are proprietary, new Excel 2007 are non-proprietary. Simplicity is also important: CSV’s are preferred to Excel which can be augmented with styles, formulas, etc.

**Star Four** is using URLs to denote things with unique identifiers, best if they are accessible on the internet at these URLs, ie. state.gov/foia-requests/34.

**Star Five** is linking data between each other using URLs. For example: while publishing an answer to FOIA request, be sure to include a link to the request itself.

Not covered by the 5-star definition, but crucial for interoperability and reusability of IT tools are data standards. Let’s say we’re building procurement monitoring platform. If data is published even at the five star level, but it’s not standardised, one have to process every data feed separately, multiplying the cost. If relevant standards are introduced (Open Procurement in this case) connecting to other data feeds (other institutions, countries) can be technically as simple as inserting a feed URL. Data standards need to match at least Three Star level, and it is the best if they match the five stars.

2. APIs
**Application programming interfaces** are serving data in a standardized manner in a form more advance to simple file download. APIs allow to filter data, retrieve only portions of a content, perform aggregations and also fire actions potentially modifying the data.

3. Scraping

What to do if data is publicly available, but its format doesn’t allow an easy reuse? You are forced to write expensive data scrapers. Scraping, as defined in Wikipedia, focuses [...] on the transformation of unstructured data on the web, typically in HTML format, into structured data that can be stored and analyzed in a central local database or spreadsheet. Inputs for scraping can vary from HTML sites to PDF documents or scanned documents. In every case the cost of creating scrapers is high and its maintenance due to possible changes in underlying data is costly too.

Extreme option for scraping is manually rewriting data into a structured format which is possible if a dataset is not big and necessary if data is really “dirty”.

Dirty data means a lot of mistakes, switched columns, etc. One most popular case are probably address fields: some include postal codes, some not; street names can be written in few variants; it’s hard to distinguish building from apartment number. Enumerated values tend to fall out of reasonable dictionary or use a lot of variants (yes, no, y, y., t, maybe, havent seen, -). There can be IDs that are pointing to nothing, numbers that don’t sum up.

As public sector still doesn’t serve data in standards and through APIs, scraping is in most cases an only way to develop tools based on public information. That’s why the first step in developing independent T&A tools is to go from scraping to publishing structured data, so that others don’t have to repeat this costly step. Polish MyCountry platform is outstanding in this field publishing over 100 processed public datasets through API, as well as ParlData project publishing public data (MPs profiles, debates, votings) on about over 10 CEE parliaments Popolo standard through API.
4. Crowdsourcing

When there is no data or it is too dirty to be fully automatically processed, or there is a need to interpret a lot of information, one can rely on data crowdsourcing. It is an approach that has been successfully implemented in a number of scientific projects like transcribing old ships weather logs or hunting for livable planets. In transparency & accountability field this approach is also being introduced. Warsaw Ninja crowdsource delays of public transportation from its passengers providing often more timely information than official news from transport agencies. Civic data mining engages users to process answers to FOIA requests into machine-readable formats. Slovak’s Open Contracts asks users who feel knowledgeable in the field to fill a survey about a contract being viewed, thus providing rough metrics on them (are they complete, coherent, overpriced, etc.).

Reliability of crowd-sourced information is a crucial issue. Proper statistical methods need to be introduced to filter through gathered input from multiple sources. Quite often crowdsourcing goes in pair with expert review – it is a process of data filtering, where a few most interesting examples are brought up for review by experts.

5. Machine learning

This is a yet to be applied method for analyzing data, with high hopes of using it for procurement solutions. Machine learning allows automatic categorization of data samples (among other things) after a small set of training examples prepared by experts are provided.

One of the recent solutions where machine learning is applied is the Red Flags in Public Procurement platform. It works as an early warning system built on risk indicators which help identify or prevent risky (potentially corrupt) public procurements. Red flagging is done automatically based on a number of indicators. Most flagged cases are sent to experts, investigative journalist, competing companies and other actors for review. After a throughout review of facts, procurements can be further analyzed by the machine
learning techniques and fed back into system to enhance quality of automatic processing, closing the feedback loop.