### Deliverable D5.5

**Tools Building and Tailoring Report**

<table>
<thead>
<tr>
<th>Dissemination Level</th>
<th>Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due Date of Deliverable</td>
<td>Month 24, 30.04.2017</td>
</tr>
<tr>
<td>Actual Submission Date</td>
<td>21.04.2017</td>
</tr>
<tr>
<td>Work Package</td>
<td>WP5, Test Beds and Evaluation - Journalism</td>
</tr>
<tr>
<td>Task</td>
<td>T5.3</td>
</tr>
<tr>
<td>Type</td>
<td>Report</td>
</tr>
<tr>
<td>Approval Status</td>
<td>Pending</td>
</tr>
<tr>
<td>Version</td>
<td>2.0</td>
</tr>
<tr>
<td>Number of Pages</td>
<td>22</td>
</tr>
<tr>
<td>Filename</td>
<td>D5.5_Tools Building and Tailoring Report</td>
</tr>
</tbody>
</table>

**Abstract:** This document describes the tools developed under Task 5.5 “Tool building and tailoring” of Work Package 5.

The information in this document reflects only the author’s views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided “as is” without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.
History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Revised by</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>15.04.2017</td>
<td>Nicolas Kayser-Bril</td>
</tr>
<tr>
<td>2.0</td>
<td>20.04.2017</td>
<td>Diana Krebs</td>
</tr>
</tbody>
</table>

Author List

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Name</th>
<th>Contact Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>J++</td>
<td>Nicolas Kayser-Bril</td>
<td><a href="mailto:nkb@jplusplus.org">nkb@jplusplus.org</a></td>
</tr>
</tbody>
</table>

Reviewer

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Name</th>
<th>Contact Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>OKI</td>
<td>Diana Krebs</td>
<td><a href="mailto:diana.krebs@okfn.org">diana.krebs@okfn.org</a></td>
</tr>
</tbody>
</table>
Executive Summary

Task 5.3 of the Work Package 5 developed and adapted tools for journalists to increase their capabilities to deal with budget data and find story leads therein. The tools followed the conclusions of Task 5.1 and moved from techniques to make amounts more visually appealing to data management to, finally, tools that help journalists and user better comprehend budget data.

Overall, the tools were used by dozens of journalists and were made visible to over a million end-users in 15 member-states during the duration of the project.
Table of Contents

Executive Summary 3
Table of Contents 4
Introduction 6
Tools developed 6
The visualizer of large amounts 6
Description of tool 6
Needs 8
How the requirements are fulfilled 10
Simplicity of installation 10
Simplicity of use 10
Integration in the journalistic workflow 11
Outcome 11
The Administrative Layer Clearer 13
Description of tool 13
Needs 13
How the needs are fulfilled 15
Centralizing data easily 15
Identifying payments by address 16
Making sense of payments by entity 16
Outcome 16
The self-learning experience on budgets 17
Description of tool 17


Introduction

Based on the scoping analysis we carried out in Task 5.1, we developed new tools to help journalists make better use of public budget data. A series of feedback loops let us observe users in real-life settings use the tools, which let us improve them further and test them anew during other workshops or meetings. We then sent out questionnaires for workshop participants to complete and advise us on how to continue tool development.

We started from the assumptions that available tools lacked a simple way to contextualize budget data and developed a tool to this end, which took into account the limitations of SpendingStories, a tool developed by Journalism++ and Open Knowledge International previously (see the Visualizer of Large Amounts). Because Task 5.1 showed that this assumption was wrong, this work package changed the focus of the tools under development and moved from a way to better visualize data to a way to better organize it after we showed in Task 5.1 that a large hurdle to working with budget data was the absence of a centralized data set to look at an issue (see the Administrative Layer Clearer). Finally, Task 5.1 and feedbacks gathered as part of Task 5.2 informed us that the vast majority of journalists lacked basic knowledge of budgets and corruption, so that tools to increase the reach of our tutorials were needed (see The Self-Learning Experience on Budgets).

Overall, these tools let us further the scoping analysis as part of a positive feedback loop, informed by user reports during workshops as well. The tools were used by dozens of journalists and reached over a million persons in Europe over the course of the project. All tools are open source and multilingual.

Tools developed

The visualizer of large amounts

Description of tool

The visualizer of large amounts lets journalists visualize amounts in a compelling and easily understandable way. Amounts can be nested with one level of depth. The tool was developed to support the work done by journalists who worked in one of the case studies of Task 5.1, The
Migrants Files, which was an investigation in the budget data related to policies of member states of the European Union and Switzerland designed to restrict the movement of people fleeing wars and poverty.

Quick facts

<table>
<thead>
<tr>
<th>Name</th>
<th>Visualizer of large amounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target group</td>
<td>Journalists</td>
</tr>
<tr>
<td>Reuse</td>
<td>11 cases of reuse in France, Spain, Switzerland, Greece, Austria, Portugal, Czechia, Belgium, Poland, Italy and Sweden.</td>
</tr>
<tr>
<td>Views from end-users</td>
<td>Over 500,000 views in total (incl. reuses)</td>
</tr>
<tr>
<td>Source code</td>
<td><a href="https://github.com/openbudgets/billion-dollar-o-gram">https://github.com/openbudgets/billion-dollar-o-gram</a></td>
</tr>
<tr>
<td>Multilingual</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Figure 1. A static image produced by the Visualizer (truncated)

Needs

One of the assumptions that drove the OpenBudgets.eu consortium at the beginning of the project was that the limitation to the comprehension of budget data by journalists or by their readers lied in the dryness of numerical data and in the inability to compare large amounts
easily. As we wrote in the first deliverable of this work package,\(^1\) we started exploring this field with SpendingStories, a tool developed in 2013 by Journalism++ and Open Knowledge International. The lack of traction gained by SpendingStories was such that we had to go back to our drawing board to reflect on what could be done better.

The mission of SpendingStories was to make budget items more understandable and digestible by making them comparable to other amounts, either from daily life or from the budget.

SpendingStories was a relatively heavy tool. It required specific skills to run and adapt, the technical setup needed server-side computing power and its deployment took, at best, a few hours. It ran on Django, a Python framework, and required a database to run. Besides technological considerations, it required journalists to use a specific interface to add new budget items to the tool. Finally, the quest of SpendingStories to be precise was a double-edged sword because it also made it harder to understand for journalists. The tool would correct all amounts for inflation and adjust amounts in foreign currencies using the exchange rates of the year of the budget item under consideration. That the value of items could differ in the front-end (what users see) and in the back-end (what was input) proved to be a major hurdle to the adoption of the tool.

Finally, SpendingStories did not take into account the need to integrate the tool to the workflow of the journalists. The comparisons created by the tool could be embedded (integrated) in an article, but this had little attraction for journalists. Nice-looking visualization for the web and for the print were not available.

To sum up, SpendingStories was limited in several ways. The needs for a new tool that would have allowed for comparing large amounts were as follow:

- Too complex to install
- Too complex to manage
- Not integrated to the journalists' workflow

\(^1\) Available online: [http://openbudgets.eu/assets/deliverables/D5.1.pdf](http://openbudgets.eu/assets/deliverables/D5.1.pdf)
Figure 2. A screenshot from SpendingStories.

How the requirements are fulfilled

Journalism++ developed a new tool to visualize amounts relatively to one another, which is based on the needs we noted when we collected feedback from users on SpendingStories.

Simplicity of installation

While SpendingStories required many operations to be set up, the new tool requires only one line of code to build and deploy.

Whereas most tools, including SpendingStories, require specific servers to host and run, the Visualizer can be run entirely from free using Google Drive and Github Pages, two services that many data-driven journalists are used to using.

Simplicity of use

Instead of developing a complex back-end, journalists can use the tool directly from Google Spreadsheet. The management of exchange rates and inflation was taken away from the tool. It is up to journalists to input the correct amounts.

Translations are integrated to the same spreadsheet and can be worked on as easily.
The Visualizer is made for integration in an article, whether online or on print. The interactive visualization can be integrated on any website and the integration on mobile has been taken care of, using Pym.js, an open source Javascript library developed by the National Public Radio of the United States.

Specific developments have been made for the print editions. A single command can be run to generate static visualizations in the PNG format, for static graphics, or in SVG, the format used by designers at most newspapers, so that they can rework the visualization and customize it further if needed.

Outcome

The Visualizer was used by a dozen news outlets in Europe and reached hundreds of thousands of end-users (consumers of news). The tool was featured in Libération, Der Standard, Gazeta Wyborcza, El Confidential, the Swiss National Radio, Dnevnik (Bulgaria) and many other news outlets.

While journalists reported satisfaction with the tool, it did not become a standard in the field. This lack of adoption is in line with the findings of Task 5.1, which showed that end-users are not much helped by comparisons of amounts they cannot make sense of.

---

2 Available online at http://blog.apps.npr.org/pym.js/
3 The conclusion was explained in Deliverable 5.3 available online at http://openbudgets.eu/assets/deliverables/D5.3.pdf
Figure 4. The Visualizer used at DerStandard (Austria)

Figure 5. The Visualizer used at El Confidencial (Spain)
The Administrative Layer Clearer

Description of tool

As part of the larger case study of Task 5.1, we developed a tool to centralize money flows between different payers and beneficiaries. The tool lets journalists store money flows from various entities related to the same issue in one place. It lets users browse through the money flows via a variety of filters.

Quick facts

<table>
<thead>
<tr>
<th>Name</th>
<th>Layer Clearer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target group</td>
<td>Investigative journalists</td>
</tr>
<tr>
<td>Reuse</td>
<td>Used in France, Belgium, Switzerland.</td>
</tr>
<tr>
<td>Views from end-users</td>
<td>15,000 page views</td>
</tr>
<tr>
<td>Source code</td>
<td><a href="https://github.com/jplusplus/football-tax">https://github.com/jplusplus/football-tax</a></td>
</tr>
<tr>
<td>Multilingual</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Figure 6. Visualization from the Layer Clearer.

Needs
The assessment of the gaps faced by journalists when working with public budget data in Task 5.1 showed that one of the most difficult tasks was to track down money flows related to a specific piece of infrastructure or to a specific program. In effect, doing journalism around public budget involves doing analytical accounting, that is to do mission-based accounting, as opposed to regular accounting, which does not distinguish between missions (for instance, if two administrations finance the same piece of equipment, each will only account for what it spent and analytical accounting on the equipment itself will be needed to assess its total cost).

Budget data that is useful to journalists never comes in the form of a pre-packaged file prepared by a public institution. On the contrary, data sets must be built by hand from a variety of sources. This has partly to do with the opacity of public administrations by is mostly due to the fact that accounting done by public administrations is of little relevance to the concerns of decision-makers and citizens (and hence, of journalists), for which analytical accounting is what matters. Analytical accounting on specific projects or issues is routinely done by public administrations but rarely published, as the case study on Bonn (related in Deliverables 5.2 and 5.3) showed.

A single project might be funded by many different public entities and involve many different beneficiary entities. A football club, for instance, usually involves both a non-profit entity and a for-profit one, both of which are financed by the city, the region, and the state. To know how much the club received in a given year, there are already 6 transactions to identify and track. If the football club owns a stadium as well and if the investigation runs over several years, the number of transactions to track becomes very large.

Another problems lies in the area covered by each paying authority. In order to answer the question “how much do you, taxpayer, contribute to a given project?” journalists need to know which population is covered by a given administrative body. While it is fairly straightforward to assess the per-capita equivalent of a given expenditure by a city, things get much more complicated when little-known administrations become major funders. Entities such as the new county councils in Ireland or the consortia of cities in France or Germany are little known to the general public, especially when their members are not elected, but manage considerable amounts of money.

Finally, the tools available to manage and visualize flows are not easy to use and require training. Network visualization tools such as Gephi have a steep learning curve and can be
difficult to install. Flow visualizations can be made with tools such as RAW, but they require the data to be reuploaded after each modification.

How the needs are fulfilled

Centralizing data easily

The Layer Clearer lets user input data directly from Google Spreadsheet, a tool data-driven journalists are used to using in their daily routine. Journalists can input all money flows on a given issue in the same spreadsheet, specifying only the payer, beneficiary, amount, and source. Further personalization can be made, such as specifying a category for each item.

Budget items are usually not found in budgetary documents directly. Budgetary documents only show aggregated data, which does not allow for analytical accounting. Instead, journalists must comb through decisions at the governing body of the administration (the city council, for instance) and find the relevant decisions. An alternative for journalists is to go through news archives or press releases, where politicians can be quoted announcing the amount paid by their administration for a specific project. Journalists can also use a reverse mechanism and look at documents from beneficiaries, where subsidies and other payments from public institutions are sometimes listed. Because of this variety of methods and sources, it was important that the back office be as flexible as possible and that it lets journalists aggregate many different sources in the same location.

Figure 7. The data used in an investigation on the Layer Clearer.

4 http://rawgraphs.io
Identifying payments by address

One of the problems local journalists face when collecting budget data from several sources is the difficulty to know what each taxpayer paid. The Layer Clearer lets journalists add files that contain the shape of the territory over which each administration has jurisdiction. The files, in GeoJSON format (a format commonly used in web applications), enable the tool to compute, for a given address, which money flows are relevant.

Making sense of payments by entity

The Layer Clearer automatically displays payments by entity and creates paragraphs of text automatically to tell readers what are the most important points about an entity are. Such pages and automatically-generated blocks of text are available for every paying entity (public administration) in a dataset.

Figure 8. The Layer Clearer’s page on the city of Nice, France.

Outcome

The tool was used at workshops organized under task 5.2 of this work package. Constructive interaction with users let us improve the tool vastly, notably by adding the “search by address” feature, which was requested at the Lille workshop.
The self-learning experience on budgets

Description of tool

The scoping analysis of Task 5.1 informed us that users were not helped by comparisons of large amounts (see Deliverable 5.3, quoted above). Indeed, people can make sense of amounts only in relationship to what they know to be the norm. Task 5.1 also informed us beyond any doubt that budget data itself did not allow journalists to work on a corruption story because the data itself does not allow for it (corruption happens outside of budget data) or because they lack the skills to identify suspicious behavior. The conclusion of Task 5.1 was that there was a general lack of skills among journalists to understand the normality of some practices and, therefore, know how to spot a newsworthy story when they faced one.

The self-learning experience on budgets is a tool that lets journalists teach themselves how corruption works at the local level. It is a bridge to Cooking Budgets, the collection of tutorials created as part of Task 5.2. In an interactive, game-like experience, users play the role of a fictional general manager in a large European city and have to take a series of decisions that show the complexity of corruption and cronyism in the public administration.

Quick facts

<table>
<thead>
<tr>
<th>Name</th>
<th>The Good, The Bad and The Accountant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target group</td>
<td>Journalists, general public</td>
</tr>
<tr>
<td>Reuse</td>
<td>Reused by journalists in Germany, Serbia, Slovakia, Hungary and Spain.</td>
</tr>
<tr>
<td>Views from end-users</td>
<td>Over 1,500,000 page views</td>
</tr>
<tr>
<td>Source code</td>
<td><a href="https://github.com/jplusplus/the-accountant">https://github.com/jplusplus/the-accountant</a></td>
</tr>
<tr>
<td>Multilingual</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Needs

Knowledge of public budgets and, henceforth, knowledge of how they can be abused for personal gain, is largely lacking among European journalists and among the general public, as we showed in Deliverables 5.1, 5.2 and 5.3. We wrote, as part of Task 5.2, a series of tutorials explaining how corruption happens and how it can be researched using open budget data, among other sources. However, traffic to the tutorials remained sparse. We therefore had to
create a tool to generate attention and entice journalists and the general public to read them and increase their knowledge of the issue.

How the requirements are fulfilled

The self-learning experience on budgets lets users experience in a first-person game-like series of interactions how public money can be used for personal gain. The interactivity of the tool ensures a high level of information retention.

Use of public-domain imagery from European cities was made to increase the visual appeal of the tool among journalists and the general public.

Figure 9. A screenshot from the self-learning experience on budgets.

Outcome

The tool was vastly successful. Over 150,000 Europeans used it to learn more about the principles of corruption and how some public officials could employ these principles to their own advantage. The average time spent with the tool was 7 minutes per visit, which, by common standards for web development, is an extremely long time. (According to Google Analytic’s benchmarks, the average session duration for websites in the “news” category in March 2017 was 1:40 minutes).

Some users regretted that the tool was not calibrated to their precise reality (we explained to
them that it was impossible to create a tool targeted at Europe that could mirror the details of every region of every member state) but the feedback was extraordinarily positive, as can be seen by the amount of press coverage received and the comments on Twitter.

Figure 10. A German journalist shares our tool on Twitter.

Below is a select list of the press coverage received by the tool:

- Extralife.fr (France): [Le Bon, La Brute et Le Comptable : un petit serious game sur la corruption](https://www.extralife.fr)
- Libération (France): [Crapule](https://www.liberation.fr)
- Télérama (France): [Jouez à un “serious game” dont vous êtes le héros plus ou moins crapuleux](https://www.telerama.fr)
- Numérama (France): [Le Bon, la brute et le comptable : saurez-vous résister à la corruption en politique ?](https://www.numereama.fr)
- Marsactu (France): [Un jeu pour vivre la corruption de l’intérieur](https://www.marsactu.fr)
- NRC (Netherlands): [Hoe werkt corruptie? Deze game legt het uit](https://www.nrc.nl)
- Le Monde (France): Un « serious game » pour saisir à quel point la corruption est banale
- Le Vif (Belgium): Corruption en politique : testez votre sens de l'éthique
- Capital (France): Le jeu en ligne qui vous met dans la peau d'un élu local face à la corruption
- La Gazette des Communes (France): Un mini-jeu illustre les risques de corruption dans les collectivités
- Ze.tt (Germany): Wie korrupt wärst du, wenn du im Rathaus arbeiten würdest?

We also wrote several articles explaining why and how the game was created, thereby disseminating our know-how with other stakeholders in the field:

- Voici pourquoi nous avons choisi de créer un jeu pour sensibiliser aux dangers de la corruption (HuffingtonPost.fr)
- How we created “The Good, The Bad and The Accountant” (Medium)

The tool was also incorporated into the curriculum in some universities in France, as part of courses in compliance management.

**Conclusion**

By using feedback loops with users and reusing the findings of other tasks in work package 5 and others in the OpenBudgets.eu project, several tools for journalists and for a wider audience were developed. Together, they reached millions of Europeans. By enabling citizens to better understand their administration, to make it more transparent and by making citizens aware of the risks faced by people who staff public administration, we augmented their trust therein.

By making a tool that helps disentangle transactions (the Administrative Layer-Clearer), we simplified comparative analysis for better budget management. One of the main problems when doing comparative analysis is in understanding which administration is responsible and pays for what; our tools are a perfect gateway to do analytical accounting in public budgets.

By listening to the shortcomings of the previous visualization efforts and by adapting state-of-the-art visualizations (nested treemaps, flow charts), we enabled all this project’s stakeholders to visualize public transactions.

Finally, by giving journalists and citizens access to a wide body of knowledge on the intricacies of corruption and budget data analysis, we enhanced accountability within the public sector in
several member states. As a result, we increased the leverage of civil society on the administrative decision process, as is shown by our tools being reused in education, for instance.

Because the tools presented in this report reached a very wide audience, garnered a lot of attention from European journalists and received very positive feedback, we are confident that the work of Work Package 5 made a real difference in the way European journalists approach public budgets and the way some public officials can fiddle with them.