MONITORING ECONOMIC AND INVESTMENT PLAN INFRASTRUCTURE PROJECTS

a Data Scientist's perspective

Cooperation & Development Institute
Instituti për Bashkëpunim dhe Zhvillim

MONITORING ECONOMIC AND INVESTMENT PLAN INFRASTRUCTURE PROJECTS

a Data Scientist's perspective

November 2024

Implemented by:

Supported by:

In partnership with:

Authors:

Gresa Smolica, External Collaborator, Cooperation and Development Institute

With contribution from Ardian Hackaj, Research Director, Cooperation and Development Institute

Publisher:

Cooperation and Development Institute
"Dervish Hima" Street, Building No. 5, Entry 14, 1011, Tirana, Albania
E-mail: info@cdinstitute.eu | Website: www.cdinstitute.eu

© Copyright 2024, Cooperation and Development Institute. All rights reserved

This publication is under Creative Commons Attribution-Non-Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.9).
Should you wish to reproduce parts of this publication, please ensure that you acknowledge the original source and consult with the contributing organization or institution.

Suggested Citation:

Please cite as: Cooperation and Development Institute, Gresa Smolica, Ardian Hackaj, November 2024, "Monitoring Economic and Investment Plan infrastructure projects: a data scientist's perspective."

Keywords:

Data Science, Western Balkans Investment Framework (WBIF), Transparency, Accountability, Data Access

Disclaimer:

This document is produced in the frame of "Connected we Can: Strengthening WB6 CSO-s agency in Economic and Investment Plan" project, implemented by CDI in partnership with regional CSOs and supported by the Norwegian Ministry of Foreign Affairs via SMART Balkans project. The views and opinions expressed therein are the sole responsibility of the authors and do not reflect in any way the positions of the Norwegian Ministry of Foreign Affairs or the SMART Balkans project.

Cooperation & Development Institute
Instituti për Bashkëpunim dhe Zhvillim

TABLE OF CONTENT

INTRODUCTION 1

WBIF WEBSITE AND DATA SCIENCE 2

RUNNING A SIMPLE DATA SCIENCE TASK WITH WBIF WEB DATA 3

COMPARISON WITH DATA ANALYSIS-FRIENDLY INTERNATIONAL FINANCIAL INSTITUTIONS WEBSITES 6

COMPARISON WITH DATA ANALYSIS-FRIENDLY EU WEBSITES 9

FINAL THOUGHTS 10

INTRODUCTION

The ever-evolving technological advancements rarely go hand in hand with the development in the realm of policy-making[1]. The current gaps between the two are reflected in many aspects. This has led different actors to explore methods on how to best use available data and translate it into actions that could help institutions and governments perform better.

As these technological advancements become inevitable, attempts to understand, use and integrate them in governance and policymaking have started to take shape. Enabling researchers, think tanks and Civil Society Organisation (CSO) to use larger institutional datasets, to analyse them at a more advanced level is an important component of accountability and transparency of public policies. This group's adoption of data science methods while still quite advanced when compared with public authorities, is still at an early stage.

This paper discusses the features of data presentation used in an international public institution structure. Its outlays the challenges that researchers encounter when using the Western Balkans Investment Framework (WBIF) data presented in its website for deeper research and analysis. Excluding the region-specific components, this brief focuses on the technical demands and hurdles that hamper a more thorough and quantitative analysis of WBIF data.

The interest around WBIF data has different sources. First, WBIF website provides a rich and unique source of information on EU financing of Western Balkans infrastructure. WBIF states that this data *"…serves as a key tool for improving the visibility of WBIF instrument and all partners*[2]*"*. We believe however that while very useful for general public information, the usefulness of WBIF-presented data for the needs of an infrastructure project cycle management is limited.

However, another important component of WBIF produced information is strongly related to this institution's ability to provide data that allows stakeholders to conduct different assessments at meta-level. In practical term such data presentation and granularity should allow different actors other than infrastructure project partners - think tanks, CSO-s, citizens, researchers, different agencies, etc - to scan this information, track it, monitor the projects, investments, etc., and eventually assess their impact. In short, project data should allow for efficient policy-making with focus on what these investments mean to the end beneficiary - WB6 citizens[3].

---

[1] World Economic Forum. (2019). We must bridge the gap between technology and policymaking. Our future depends on it. Available at: https://www.weforum.org/agenda/2019/11/we-must-bridge-the-gap-between-technology-and-policy-our-future-depends-on-it/

[2] WBIF. (2020). Western Balkans Investment Framework: A guide. Available at: https://bankwatch.org/wp-content/uploads/2024/08/WBIF-Guide_Update_March2020.pdf

[3] CDI. (2022). Building Better Balkans Together: Tirana Connectivity Forum 2021. Available at: https://cdinstitute.eu/wp-content/uploads/2022/05/Building-Better-Balkans-Together-Tirana-Connectivity-Forum-2022.pdf

Think tanks and CSO-s in the WB6 have been monitoring and reporting on WBIF-financed infrastructure projects since 2015. However, as the EU and private investments as well as EU financial instruments grow bigger, more sophisticated and impactful, so does their impact on local economies, on region's growth or on the well-being of its citizen and business content of its enterprises. We notice a similarly significant need to find adapted and optimal solutions that allow project infrastructure stakeholders - think tanks and CSO-s being amongst them - to focus on monitoring governance and policy impact all by facilitating and automatising the day-to-day monitoring of project implementation. *To make this happen, governments and public institutions – such as WBIF - must change the way information is published and rethink the use of published data.*

WBIF WEBSITE AND DATA SCIENCE

The WBIF website presents a rich and powerful resource for researchers, think tanks and CSO-s to better understand the amount of financial suport, the distribution by sector, by SEE6 country, the involvement of international financial institutions, and the overall EU support toward the WB6.

Briefly, the WBIF website's main information is arranged as follows:

- By Priorities - showing the main sectors or areas of support where the EU or IFI financial allocations are / will be spent;
- By Partners - broken down into beneficiaries, international financial institutions, bilateral development banks and EU, and bilateral donors;
- By Investments - showing the type of support (grant, loan, other) and project (service, works)[4]

At first glance, the WBIF appears to be the ideal source for obtaining comprehensive data on project budgets, duration, implementing partners, beneficiaries, and financers. Of particular interest is the overview of investments by country or sector, which provides valuable insights through visually engaging charts that showcase detailed information pertaining to each country, sector, and grant / loan amount.

However, several challenges hinder the effective use of this resource. The first thing a quantitative researcher notices when trying to analyze more thoroughly is the:

- lack of a proper dataset;
- the lack of machine-readable information, and
- data representation is in PDF-s or in embedded HTML files. Those practices are not analysis-friendly, require significant pre-processing and do not allow the usage of this resource for complex research purposes.

[4] WBIF Website can be accessed here: https://wbif.eu

While holding valuable information, the current WBIF web site and the information it provides is not data science friendly.

RUNNING A SIMPLE DATA SCIENCE TASK WITH WBIF WEB DATA

Among the most usual tasks that a data scientist would implement with similar data profile as the ones provided by the WBIF website consist of:

- *Exploratory Data Analysis* to identify patterns and outliers in the data provided – e.g. sectors characterised by the largest delays during project implementation;
- *Forecasting* through statistical modelling or machine learning models to identify trends and recommend solutions depending on the task's goal – e.g. identify red flags during project implementation (delays, budgetary slips, etc.)
- *Data-driven-Alerts* to notify users of significant data changes – e.g. when a project moves into the next phase in its PCM;
- *Data Visualization* to identify patterns, trends, and relationships – e.g. highest / slowest / etc infratsructure project, etc.
- *Etc.*

For EIP stakeholders, national policy-makers, specialised think tanks or CSOs situated in the WB6 that monitor EU-funded initiatives in the region, staying up to date on project updates is a daily priority. The WBIF website is usually the major source of information for such updates. A systematic approach to checking for project updates involves setting aside a specific time during the day to scan the WBIF website for news, checking the project page for any changes or new additions. The alternative is to conduct a Google search and waiting for the search engine's results.

To increase visibility, outreach and accountability, it would be beneficial for WBIF web-masters to incorporate data-driven alerts in their infrastructure project presentation. By providing automatic notifications, these alerts would help users to make informed decisions based on data, all by saving them precious time. Technically, these notifications can be sent through a variety of channels such as email, messaging, or applications. From an impact angle, they are supposed to encourage a change of behavior in the user depending on the notification.

Task:

Develop an automated system to monitor the WBIF website for Project X updates and send a weekly email with a summary of these updates.

Solution:

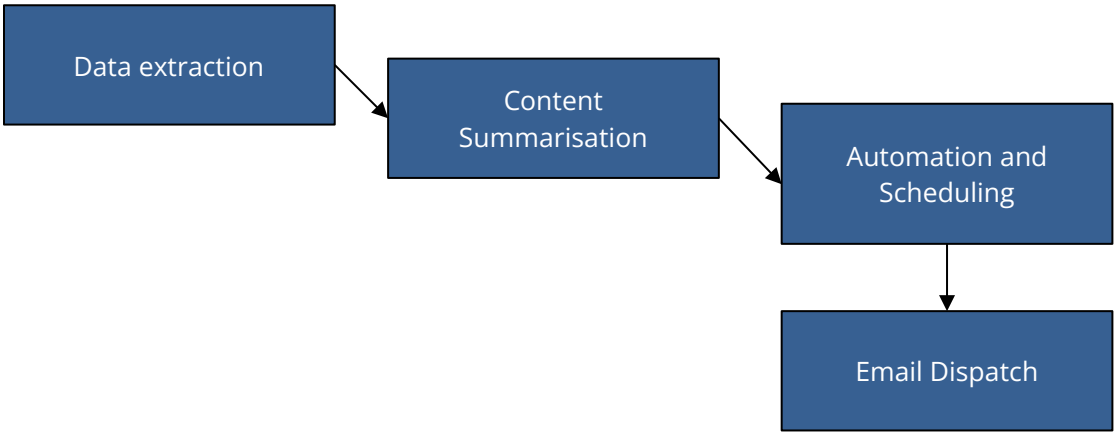


Figure 1. A simple workflow for creating an automatic alert

By employing this approach within the framework of WBIF data presentation, researchers could receive automatic alerts regarding any significant changes in the data they are analyzing respective to the projects that they are following. This would enable them to stay abreast of any development; to modify their approach in time; or put the relevant updates or changes into the larger context looking for correlations and / or explanatory variable. Implementing a simple alert through the WBIF data is shown in Fig. 1. The implementation of a straightforward path as above should not challenging for the WBIF website.

➔ DATA GATHERING

For any simple alert, the first prerequisite is access to clean and real-time data. Those data should have a fixed structure, follow a pattern, and are complete and consistent. The websites and sources providing similar information like the WBIF, usually allow the accessibility to up-to-date datasets, which can be download by users for further research.

Currently, the WBIF website does not provide such a possibility, hence requiring more pre-processing steps to access simple data. The only tabular representation of information on the WBIF website is related to the designation of investments; however, it cannot be downloaded as a whole, and it is only represented as an HTML embedding. HTML embedding is not a friendly practice for research analysis.

Additionally, this information contains hyperlinks to additional information on the project. In principle, this feature would be a very powerful element for a linked database, however in the current setting, this potential remains untapped. As a result, for the moment hyperlinks serve to reach supplementary information, instead of being used to link data points, provide context between entries, support filtering of information, etc. – as would be the case of a linked database.

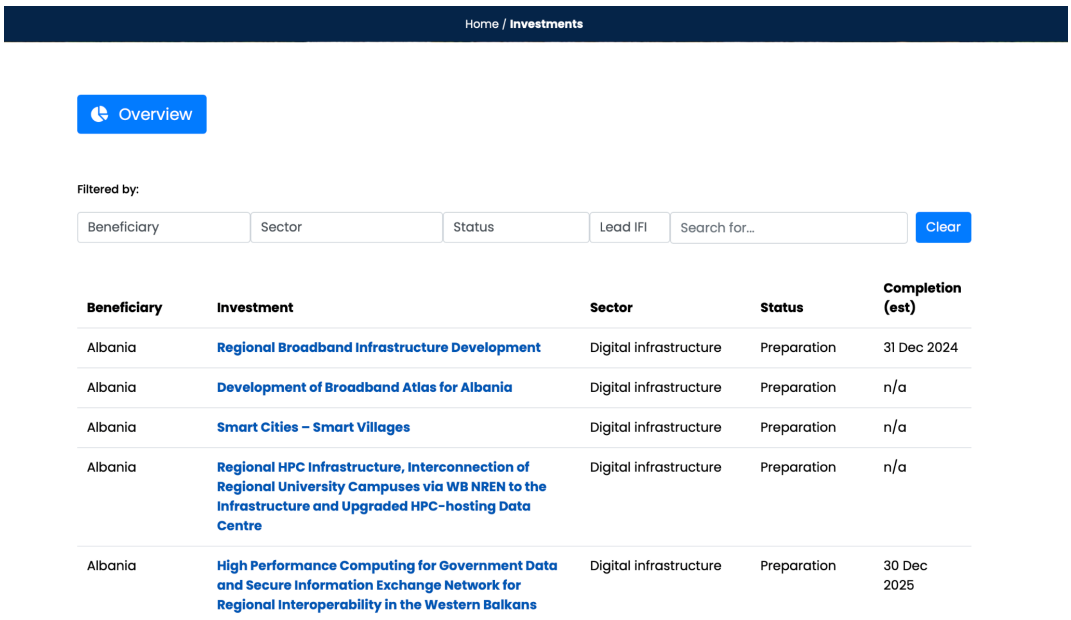


Figure 2. WBIF Investments Overview. Source: https://www.wbif.eu/

➔ ACCESSING WBIF DATA

Accessing or extracting WBIF data from its website presents several difficulties, most of which are connected to the:

- structure of data publication,
- inconsistent updates, and the,
- layout and formatting of the website.

The WBIF website does not provide structured data access or an Application Programming Interface (API) for automated data retrieval, in stark contrast to more structured platforms such as the EU's Funding & Tenders portal. As a result, the scraping procedure must adhere to polite web scraping criteria while using less effective and more error-prone methods. This impacts scrapping output quality and increases the difficulty of obtaining complete and reliable data.

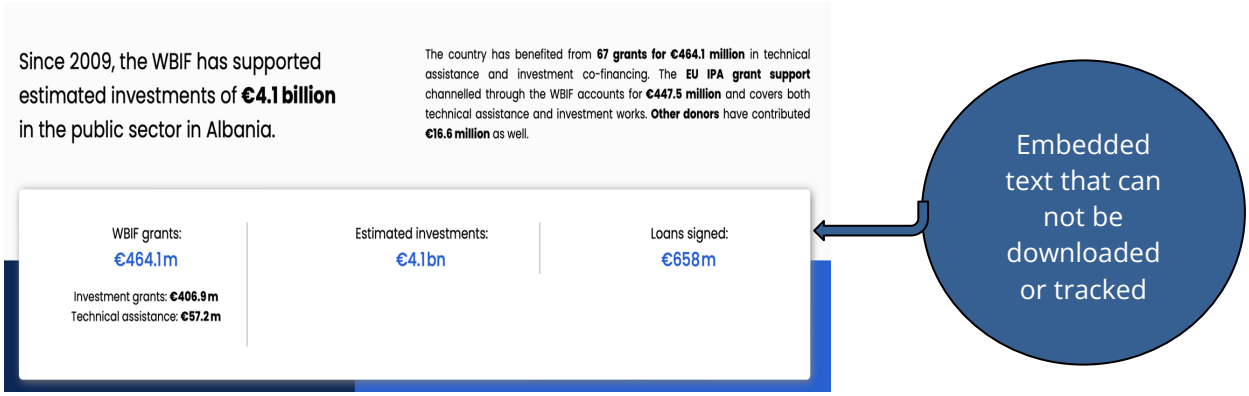


Figure 3. WBIF Beneficiaries page; Albania profile. Source: https://www.wbif.eu/beneficiaries/albania
Annotations: Author

➔ A FRIENDLY ENVIRONMENT FOR ADVANCED RESEARCH

Crucial details about one specific infrastructure project are often scattered over several web pages or tucked away deep into the website, frequently obscured by pagination or intricately nested navigation links. As a result, web crawlers need to be quite clever in their design to be able to navigate these structures and to collect all relevant updates without missing any.

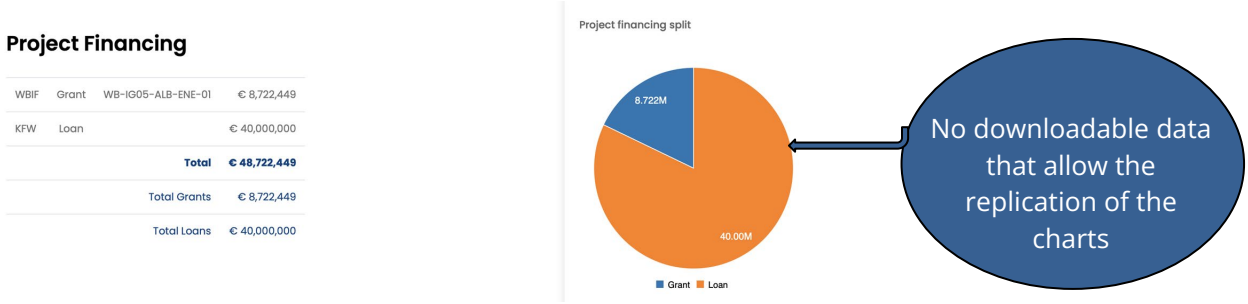


Figure 4. WBIF Project details page; Albania projects. Source: https://wbif.eu/project-detail/PRJ-ALB-ENE-021
Annotations: Author

Furthermore, because project statuses are not consistently updated across all designated parts (some updates are only in the Newsroom, for example, and are not marked as "EIP related", neither referred to the specific project), the alert system may miss significant updates due to the current structure of the website and miss important updates.

COMPARISON WITH DATA ANALYSIS-FRIENDLY INTERNATIONAL FINANCIAL INSTITUTIONS WEBSITES

It apperas from our analysis that the WBIF website is not designed and / or structured for automatic manoeuvres required for a data-science based approach of information

processing. Coupled with irregular data updates and diverse data formats and structures, the WBIF—the primary source of information for EU infrastructure investments in the SEE6 region—further adds difficulties during information gathering and processing – a process that is already difficult for external stakeholders. Adding to structural challenges in obtaining information from WB6 governmental sources, it leaves researchers, think tanks, and CSOs with very limited accessibility and processing options.

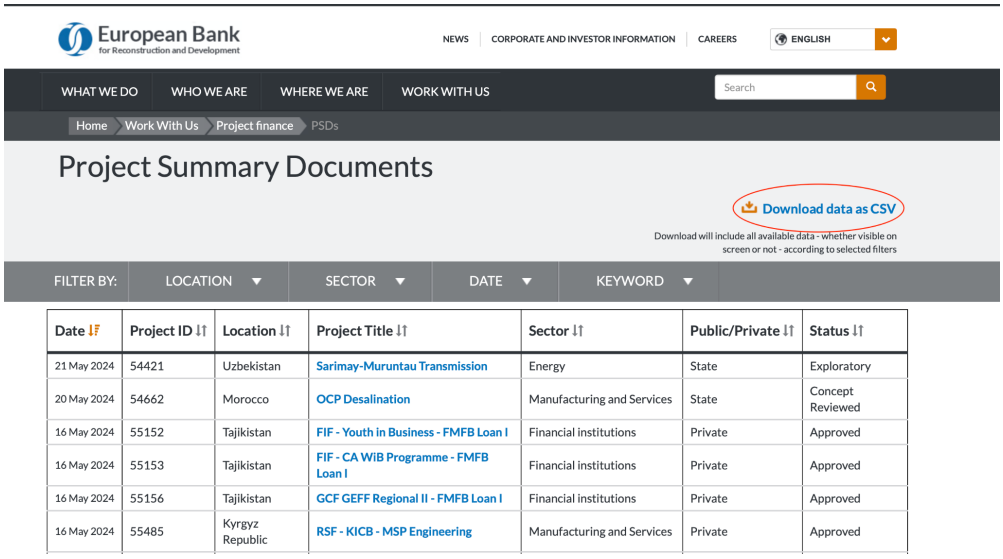
This represents a multilayered challenge that can be overcome by WBIF by evolving the website toward more automated options and enrich it with resources that allow a more advanced level of analysis. Its improvement can be facilitated by integrating data-science related technological advancements from EU, from IFI, from WB6 governments and from specialised think tanks and CSO-s.

Differently to the WBIF, the IFIs provide more possibilities for data sharing and retrieval, making it easier to analyze information and track different EU co-funded projects. The way those IFI present their information can be used as a benchmark. Moreover given the co-financing EU – IFI modality of almost all EU-funded projects in SEE6, a greater coordination between WBIF and IFIs could enhance the way information is served to the public. This should be even easier as those IFI are WBIF members.

The figures below show snapshots of project databases from IFIs such as European Bank for Reconstruction and Development (EBRD), Kreditanstalt für Wiederaufbau (KfW), and Agence Française de Développement (AFD)[5]. Their websites facilitate external research and allow project tracking quite differently from the information provided by the WBIF. Some of the main features – missing from WBIF system - include the downloading data in formats that facilitate data analysis (CSV, JSON, etc.), advanced filtering options, dedicated project databases, API-s, etc.

We have brought below some illustrations from EBRD, from KfW and from AfD.

[5] EBRD Project Summary Documents Database can be accessed here: https://www.ebrd.com/work-with-us/project-finance/project-summary-documents.html |
KfW Projects Database can be accessed here: https://www.kfw-entwicklungsbank.de/Internationale-Finanzierung/KfW-Entwicklungsbank/Projekte/Projektdatenbank/index.jsp?query=*%3A*&page=1&rows=10&sortBy=relevance&sortOrder=desc&facet.filter.language=de&dymFailover=true&groups=1 |
AFD Open Data Portal can be accessed here: https://opendata.afd.fr/page/accueil/



Date	Project ID	Location	Project Title	Sector	Public/Private	Status
21 May 2024	54421	Uzbekistan	Sarimay-Muruntau Transmission	Energy	State	Exploratory
20 May 2024	54662	Morocco	OCP Desalination	Manufacturing and Services	State	Concept Reviewed
16 May 2024	55152	Tajikistan	FIF - Youth in Business - FMFB Loan I	Financial institutions	Private	Approved
16 May 2024	55153	Tajikistan	FIF - CA WiB Programme - FMFB Loan I	Financial institutions	Private	Approved
16 May 2024	55156	Tajikistan	GCF GEFF Regional II - FMFB Loan I	Financial institutions	Private	Approved
16 May 2024	55485	Kyrgyz Republic	RSF - KICB - MSP Engineering	Manufacturing and Services	Private	Approved

Figure 5. EBRD Project Summary Documents. Source: https://t.ly/9VYaj

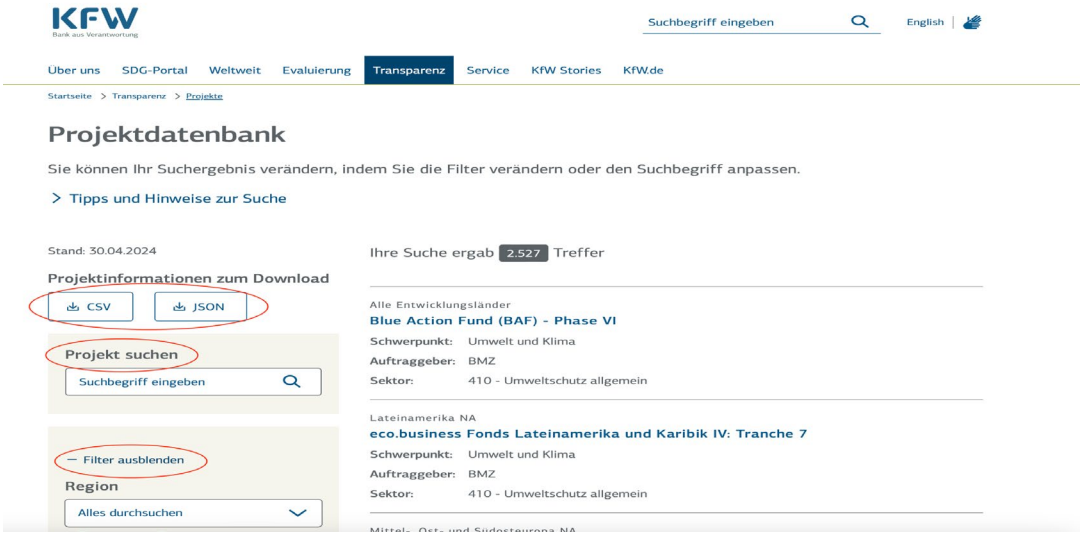


Figure 6. KfW Project Database. Source: https://t.ly/GFONx

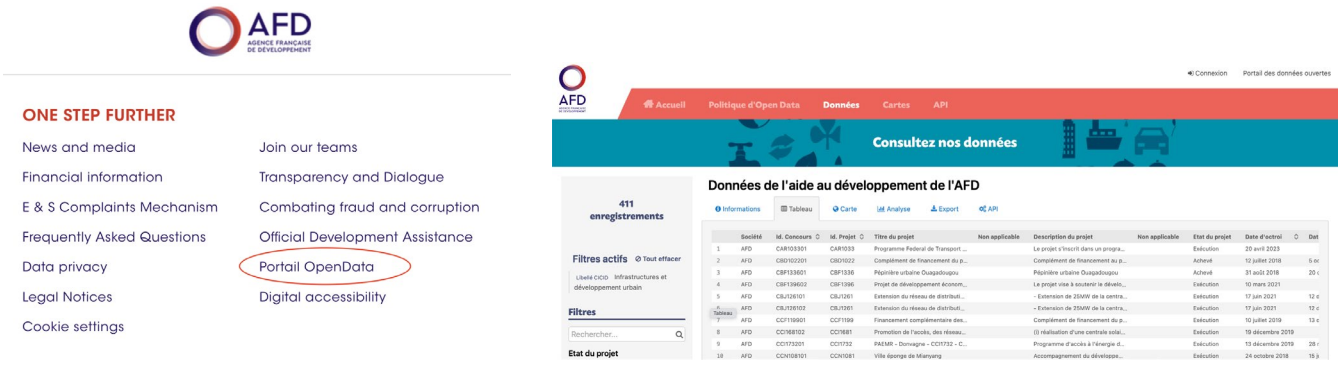


Figure 7. AFD OpenData Portal. Source: https://t.ly/9Cu26

COMPARISON WITH DATA ANALYSIS-FRIENDLY EU WEBSITES

Two similar EU sources that provide similar information to what the users expect from the WBIF website are EUROSTAT and EU Funding & Tenders. Both these sources are distinguished by their high level of accessibility for their users and the large amountof updated information that they provide.

EUROSTAT offers a comprehensive database that is well-organized and easy to access, making data retrieval simple for data scientists.

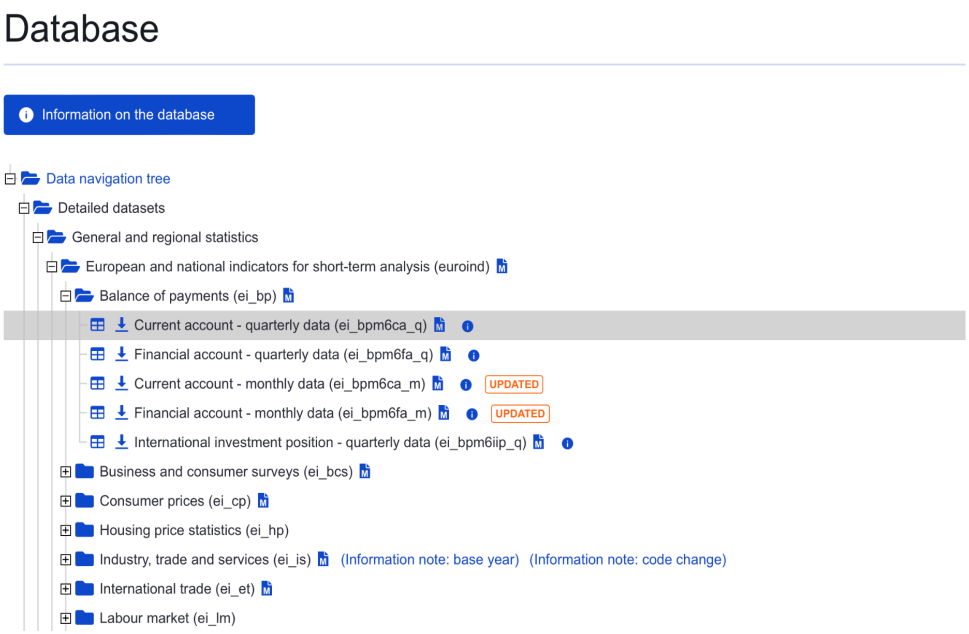


Figure 8. A snapshot of EUROSTAT database representation. Source: https://ec.europa.eu/eurostat

The existence of an API enables automatic data retrieval, making it easier to integrate current information into more advanced research operations without requiring manual intervention. Integrated data visualization methods enable the development of instant insights, which is useful for data scientists who need to conduct quick assessments.

In that way, EUROSTAT's data meets the EU's rigorous standards for reliability, accuracy, and timeliness, guaranteeing the researchers work with the most reliable information available. Regular updates ensure that the data represents the most recent trends, which is essential for timely and appropriate research.

The **EU Funding & Tenders Portal** is the next example that we have chosen. Designed as a one-stop shop for information on all EU funding opportunities and tenders, it carefully categorizes, cross-links and significantly decreases the time required for data collection.

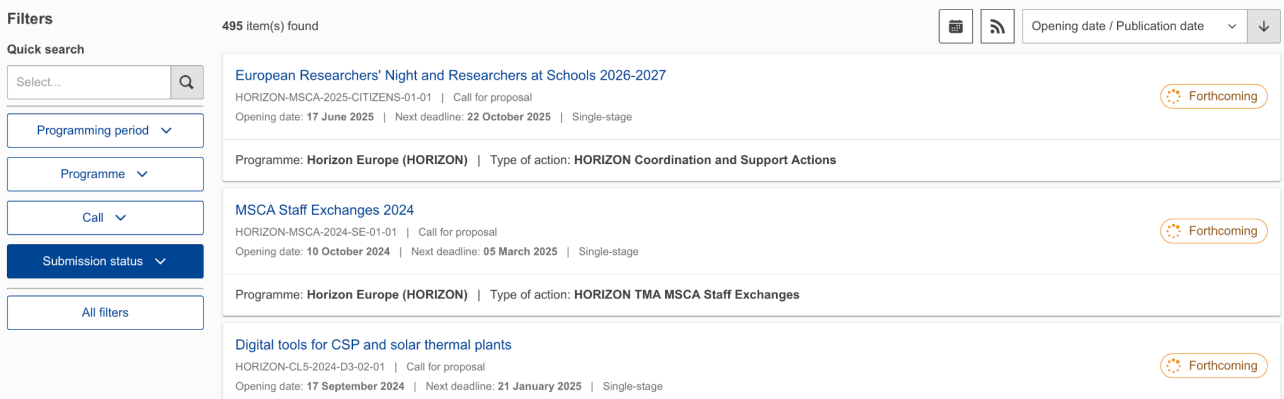


Figure 9. A snapshot of EU Funding & Tenders advanced filtering. Source: https://ec.europa.eu/info/funding-tenders/

Advanced search tools enable users to filter possibilities based on a particular criteria such as the sector, geography, or financing modality, making data collection more targeted and efficient. Continuous updates on new funding opportunities and tender notifications guarantee that data scientists and researchers have access to the most recent data, allowing for timely and effective decision-making.

In addition, the availability of the API allows for different automated actions to be implemented so that users and potential beneficiaries do not miss any updates or new funding opportunities.

FINAL THOUGHTS

The inspiration for developing automated analysis tools comes from a need to make the data retrieval and processing easier and more efficient. Currently, the WBIF website does not meet the demands of WB6 specialised think tanks and CSOs. These stakeholder groups are obliged to spend a significant amount of time actively tracking EU investment news and monitoring updates from institutional representatives in order to gather important project information.

This labour-intensive monitoring process limits the ability of think tanks and CSOs to monitor and engage efficiently in the project cycle maangement steps of Economic and Investment Plan infrastructure projects. As Growth Plan-supported infrastructure projects will be administered by WBIF as well, they will be afected by the same problems.

As a result, rather than maximizing their experience in monitoring, transparency, accountability and policy influence, civil society actors as well as any external EU-suported stakeholder are mired in regular data tracking, confirmation, cross-linking through different pages within the same website, or re-verifying with external sources, so reducing their potential impact on policy creation and project oversight.

This picture contrasts sharply with the more modern, data-driven EU websites such as EUROSTAT or EU Funding & Tenders Portal. Platforms such as EUROSTAT and the EU Funding & Tenders Portal demonstrate how incorporating advanced automated technologies and improved data access may significantly increase the efficiency and effectiveness of data gathering, information processing and ultimately contribute to transparency and accountability. Hence WBIF website improvement is critical not only for ongoing programs like the Economic and Investment Plan but also for newly-adopted financial EU instruments such as the Growth Plan.

Finally, in an era where sophisticated data models can deliver tailored information seamlessly to users, it is imperative for specialised think tanks in the Western Balkans to capitalize on this technological momentum. Leveraging these technological advancements and being more demanding toward WB6 governments and WBIF governance structures, allows think tanks and CSO-s to achieve greater transparency in governance and strengthen their involvement in monitoring and advocating for good governance practices.

Our research has shown that while there is a critical mass of think tanks and CSO-s in the region involved in the EIP project cycle, and that have the critical knowledge to contribute to a more efficient presentation of relevant information. The last take-away from this exercise is to recommend a closer cooperation of WBIF website administrators and their intended audience.

This document is produced in the frame of the project "Connected we Can: Strengthening WB6 CSO-s agency in Economic and Investment Plan".

Implemented by:

Cooperation & Development Institute
Instituti për Bashkëpunim dhe Zhvillim

Supported by:

Norway

SMART Balkans
Civil Society for Shared Society in the Western Balkans

In partnership with:

TRANSPARENCY INTERNATIONAL
Bosnia and Herzegovina

Estima

FOL
LËVIZJA

institut alternativa

European Movement Serbia