COLLABORATIVE ENGAGEMENT ON SOCIETAL ISSUES

WP2 – Pilots implementation and Open call

Report on accessibility to sensitive and privately-owned databases

date 31.03.2022

Deliverable 2.8

Grant Agreement number	:	101006325
Project acronym	:	COESO
Project title	:	Collaborative Engagement on Societal Issues
Funding Scheme	:	[H2020-EU.5. - SCIENCE WITH AND FOR SOCIETY](#)
Topic	:	[SwafS-27-2020 - Hands-on citizen science and frugal innovation](#)
Project's coordinator Organization	:	EHESS-OpenEdition
E-mail address	:	pierre.mounier@openedition.org, alessia.smaniotto@openedition.org
Website	:	https://coeso.hypotheses.org
WP and tasks contributing	:	WP2 – Pilots implementation and Open call
WP leader	:	CRIA
Task leader	:	IRPI
Dissemination level	:	Public
Due date	:	31 March 2022
Delivery date	:	31 March 2022
Authors	:	Lorenzo Bagnoli (IRPI), Luca Rinaldi (IRPI)
Reviewers	:	Arnaud Gingold (AMU-OpenEdition), Alessia Smaniotto (EHESS-OpenEdition)

Contents

What accessibility means

"Accessibility" for the Oxford Dictionary is "the quality of being able to be reached". In journalism, data are accessible when a journalist is able to make them reachable by a targeted audience: in this respect, from a journalistic perspective, it is important to keep together the accessibility for the journalist as a professional, and the accessibility for the readers. From the journalists' perspective, therefore, the data "accessibility" issue is tightly bound to data assessment aimed to provide accurate and understandable data to readers. Fuzzy and inconsistent data are not reliable for journalism.

From the authors' point of view, data accessibility also means making data searchable. "Searchable" means that data can be read and understood by the readers. This process can be a technical challenge with big chunks of data, especially if coming from anonymous sources (i.e. leaks, anonymous tips, spurious datasets, etc.). Data Journalism with data can be seen as a service journalism, because it contributes to the comprehensive and accurate data, rather than concentrating on watchdogging.

Accessibility for authors

- **Quality and data-driven journalism**

  - Data accessibility for authors means obtaining readable and comparable datasets. The readability and the interoperability of different datasets can be considered as key qualities. For this reason, the data-driven approach is not fulfilled with mere data retention. Data-driven journalism is interrogating datasets with the queries that will allow to better assess the value of the data. Larger datasets don't imply more accurate datasets. The more refined the query, the better the outcomes. Collaborations with researchers able to refine big data are crucial to make comprehensive and coherent articles or features/stories.

- **Quality and data accessibility**

  - Data accessibility has an impact on quality journalism. Quality journalism is a definition coming from both academics and professionals. This definition refers to two different environments: the environment of the producers of journalism and the environment of whom is consuming journalism. In terms of data, journalism has quality when: a. is able to provide comprehensive and accurate data (point of view of the professionals); b. is able to provide understandable and clear data (point of view of the readers). The latter can be defined as "data accessibility". For point a. journalists have to rely on people who are able to provide comprehensive and accurate databases such as academics and researchers.

Accessibility for readers

As said above, data accessibility for readers/consumers means reading/consuming data which can be processed and understood. Journalism is always a mediation between "raw information" and the reader/consumer. Data provided to the readers/consumers are always digested first. This digestion is the first phase of the data processing. When data are more accessible, more effective and understandable stories are possible for readers. There are tools in journalism to make data accessible such as infographics and data visualizations. These items are aimed to resume and simplify complex data. The main infographics editing software (i.e. Tableau, Datawrapper et al.) support journalists in the process.

- **Omnichannel communication**

    o   Data visualization is among the ways to disseminate journalism, along with different media such as text, audio, video, pictures, sketches. The ability to combine different media together in specific formats has recently become pivotal. Different formats help to overcome obstacles in both physical (e.g. tools for people with deaf-blindness) and intellectual accessibility (e.g. the context of a piece of information, the clarity of the jargon, the accuracy and the completeness of the content, etc.).

Cooperation between IRPI and Crime&Tech

The goal of the collaboration between IRPI and Crime&tech, the spin-off company of Transcrime – Università Cattolica, is to jointly develop approaches and tools to support the investigation of transnational organised crime (TOC) phenomena. The ultimate aim is to contribute to the understanding of TOC in order to improve research and prevention. The common fieldwork for IRPI and Transcrime is the TOC infiltration into the legal economy. Through the infiltration, dirty money is laundered and reinvested (For details see the Blog Post on Hypotheses.org).

IRPI and Crime&tech believe that, on the basis of the research carried out by Transcrime, it is possible to develop a collaborative space and innovative tools to allow both journalists and public authorities to in-depth investigate the links between criminal actors and legitimate assets, in particular legitimate companies, so as to highlight the risks of corruption, collusion, financial crime and organised crime.

In this light, IRPI and Crime&tech are involved together in a number of projects which try to achieve the above mentioned objective. COESO is one of the steps for reaching this goal, and builds on previous experiences (first of all, project DATACROS) and on current projects (project DATACROS II). The following paragraphs will provide details about these projects (on this, see also COESO Deliverable 2.9).

Background: DATACROS and previous projects

The collaboration between IRPI and Crime&tech started in 2019 with the beginning of DATACROS (www.transcrime.it/datacros). DATACROS aimed at «Developing A Tool to Assess Corruption Risk Factors in firms' Ownership Structure». IRPI was selected as one of the end-users and testers of the tool as representative of journalists and civil society. The other end-users pertained to the domain of law enforcement agencies and anti-corruption authorities (see COESO D.2.9 for details). At that time IRPI was working on one of its investigations about corporate opacities in Europe and offshore jurisdictions. Thanks to the collaboration with Transcrime and the help of DATACROS, IRPI could explore some significant patterns like the connection between opaque corporate structures and blacklisted countries; the anonymisation of the Ultimate Beneficial Owners (UBOs) through the appointment of trusts, fiduciaries and foundations; the hide of certain types of board members or UBOs such as Politically Exposed Persons (PEPs) and local administrators; the link of some companies with entities under sanction or enforcement cases.

Despite this successful experience, it was clear that the cooperation between researchers and journalists, and in particular the development and use of investigative tools, faces some technical and legal challenges, as well as operational challenges, which have to be fully understood and managed. This is the gap which IRPI and Crime&tech are filling with COESO.

Cooperation within COESO

COESO's *Pilot 4 – Tools and databases to increase the impact of investigative journalism* – is aimed at creating a collaborative space between Crime&tech, as representative of the research community, and IRPI, as a representative of civil society/investigative journalists with the aim to (a) improve journalistic investigations on corruption, collusion and financial crime; (b) improve the development of IT tools and (c) setting a sustainable framework for sharing sensitive data. As mentioned, this pilot builds on previous projects which entailed a collaboration between the two entities, in particular on the EU co-funded project DATACROS (see above).

Through COESO, Transcrime and IRPI will use the DATACROS tool in specific high risk economic sectors such as the maritime industry and real estate so as to deep dive into the technical and legal problems which have to be faced when implementing operationally a cooperation between researchers and journalists. Both these economic sectors are formed by companies with usually complex ownership structures in which it is often difficult to identify the real beneficial owner. The DATACROS tool will be used to identify investigative paths that the investigative journalists of IRPI will then try to complete.

Case-studies of cooperation

As mentioned, in COESO and thanks to the use of the DATACROS tool, IRPI will focus on two domains for investigations: Real estate and the Maritime industry. "Cities for rent" has been the first transnational investigation led by Arena for Journalism into the market of corporate landlords. IRPI joined in 2019. DATACROS supported a chunk of the investigation about Milan. The aim of DATACROS is to expand the analysis started with "Cities for rent", adding to the investigation the Transcrime academic approach and the expertise of tracking and identifying complex corporate structures.

The shipping industry is among the main focus of the IRPI transnational investigations. Through COESO, IRPI and Transcrime will share data analysis and insights into this particular business sector. The cooperation between journalists and academics, its challenges, its thresholds and its limits will be explained by posts on the pilot's Hypotheses.org blog, "The back story". Here, we will explore the challenges of the collaboration and the main lessons learnt from the mutual exchange. While on IrpiMedia we will report on the results of the investigations, on Hypotheses.org we will talk about the "behind the scenes" of how the researches are made.

Use of sensitive and privately-owned data: opportunities and challenges

Description of the nature of the data

The data used by IRPI and Transcrime in their collaboration are mainly corporate data and court records. Few data come from confidential sources or leaks. The latter always needs a fact-checking process.

Why it is important for investigative journalists

Corporate data is owned by business registries all over the world and some of them are very expensive. For investigative journalists, corporate data are a key asset to dig into financial crimes, misuse of public money, embezzlement, conflicts of interests, infiltration of TOC into the legal economy. Quantitative data analysis into big datasets is crucial to identify patterns and trends.

Challenges

- **Identification of partner/start-up of the partnership**

  - In order to reach effective storytelling, journalists need to identify academic partners who are able to go beyond the academic jargon and the academic audience. Without these communication skills it is not possible for academics to understand journalism from the professional perspective. A shared methodology on data retrieval and data analysis is the first step needed to set up a fair and healthy work environment. In order to share the same language with academics, journalists are required to improve their technical skills in data analysis, basic computer science and a basic scientific background about the subject matter. Journalists are required to address their subject using a scientific approach rather than the usual journalists' one (e.g. through the 5 W: who, what, where, when, why). A scientific approach implies a methodology based on scientific literature which are tools that journalists often don't know when they approach a topic.

- **Identification of appropriate data sources, consistency and quality**

  - The identification of the appropriate data sources is essential. Data sources can be either official or unofficial. The latter must be independently verified and double-checked by journalists. Comparing data produces storytelling and this is why comparable data have a higher interest for journalists.

- **Coverage of datasets**
  - As mentioned above, the coverage of databases is decisive in terms of details and information provided. In the corporate data analysis, which is the main domain of the IRPI and Transcrime collaboration, the main challenge to face is the huge discrepancies among the worldwide business registries. Complex corporate structure can hide the real owner, who is called the Ultimate beneficial owner (UBO). Hiding the UBO is a red flag for companies which might have illicit business or sanctioned UBOs. Also organised crime went transnational with the aim to exploit the weaknesses of certain jurisdictions. Countries such as Panama or Belize are known for providing with their business registries only the identification code of a company and the formation agent, not the real beneficiaries. In other countries, the details provided are accurate and extensive.

- **Costs and financial challenges**
  - One of the greatest challenges of investigative journalism is being sustainable. Investigations are very expensive and one of the reasons is the costs of the data. Business registries, land registries, court registries are often a major item in the budgets. Fact-checking also relies on these databases. The time devoted to refining and analysing data may not always be sufficient within the newsrooms. Financial resources may not be sufficient in some situations to fund skilled dedicated manpower. The collaboration between researchers and journalists sometimes could be a way to overcome these issues because journalists can get access to expensive databases and they can rely on academics who already know how to refine enquiries.

- **Security challenges**
  - In projects involving the use of large databases, the issue of data security is sensitive. For this reason, at IRPI, employees follow annual training courses to acquire know-how and expertise in data retention and data security. In this way, we can encrypt and secure data in dedicated storage offline and online. Data security is a challenge not only for journalists. Also researchers have to face an increasing risk of data breaching or hacking. Journalists might support researchers in following some protocol to protect data.

Conclusion

In conclusion, we believe that the collaboration between journalists and the academics is not only healthy for both, but necessary. As journalists, having the opportunity to talk and study with those who have an academic background is essential to check the level of comprehension of complex phenomena. It is one of the ways to find the right frame for the investigations. This kind of collaboration gives more strength and consistency to the stories. Sometimes it helps to identify new points of view and to foresee new scenarios. Data is the basis of knowledge, both in journalism and academia. Expensive privately-owned business registries can be inaccessible for journalists either for economic reasons or for lack of expertise about how to search for information. These registries contain sensitive data and they are a goldmine for the companies who own them. Still, the transparency flaws of some of these registries can be exploited by criminals or TOC. Therefore, for the good of the public opinion and the public knowledge, the collaboration between journalism and academia can try to shed light on the complex and opaque structures of some corporates in order to advocate for more transparency.