An approach to financial information analysis by the Brazilian Federal Police

Renato Kettner Filho[1,*] Douglas D. J. de Macedo[2.]

[1] Universidade Federal de Santa Catarina (UFSC), Departamento de Ciência da Informação, Florianópolis, Santa Catarina, Brasil.

[2] Universidade Federal de Santa Catarina (UFSC), Departamento de Ciência da Informação, Florianópolis, Santa Catarina, Brasil.

Abstract

INTRODUCTION: One of the tasks performed by the Federal Police is the verification and cross-referencing of data contained in Financial Intelligence Reports (FIRs) produced and forwarded by the Financial Activities Control Council (COAF) - an activity in which, among the police involved, the absence of a standard system of execution.

OBJECTIVES: The present work aims to present a study on the form currently in use within the scope of the Federal Police Station regarding the analysis of FIRs from the COAF, then presenting a methodology suggestion, aiming to speed up the process. Regarding open sources, the objective is to carry out a survey of possible complementary repositories not yet used and to expose ways of implementing queries.

METHODS: Pointing out the workflow currently used (displaying it in graphic form) as well as identifying the data sources consulted (open sources and closed sources) during the process,

CONCLUSION: understanding how the FIR analysis system is currently carried out, identifying possible aspects for improvement, and suggesting a methodology to be used, indicating for this the use of files in a specific format (.CSV), exclusion of queries in similar repositories (System "A") and, mainly, the automation of part of the procedure (with the use of the RIBOT prototype software).

Keywords: COAF, Financial Intelligence Reports, FIRs, Polícia Federal, Python, Workflows, Financial Crimes.

1. Introduction

This Currently, we live in a world with a great presence and influence of information and technologies [1], which make it possible for most of the procedures carried out daily to be carried out digitally and using the internet, highlighting among them tasks related to digital commerce and financial transactions involved all of them using principles related to Information Science. Such advances in technological terms have been obtained and studied mainly from to allow adults to live with a better quality of life, have a long productive life and that the state expenses with health are reduced [2]. These are some researches that approached technological aspects from information sciences perspectives [3,4,5,6,7,8]. This increase in the use of digital forms for making payments and commercial receipts, as well as the wide range of possibilities for committing fraud in such procedures, make such activities demand greater attention from police authorities, to identify possible crimes and strange movements, especially money laundry [9]. The importance of using the analysis of Financial Information Reports (FIRs) produced by a federal body called the Financial Activities Control Council (COAF) has grown in Brazil, which is responsible for receiving and processing data on transactions occurring in the Brazilian market, preparation of a report and subsequent submission of such document to the investigative authorities, with a view to

*Corresponding author. Email: uriadm@gmail.com

analysis for the identification of possible crimes. In a simplified way, it is also highlighted that the initial information that gives rise to the FIR prepared by COAF is data forwarded by various bodies and entities present in the Brazilian territory and that are related to commercial and financial transactions, such as Banks, Real Estate Companies, Stock Exchange, among others [10].

Bearing in mind that the author of this article is, in addition to being a student of Information Science, one of the Federal Police officers who carry out activities related to the investigation of financial crimes (including the analysis of FIRs), as well as that one of the investigative bodies receiving the FIRs produced by COAF is the Brazilian Federal Police (PF), a police entity with duties and responsibilities at the national level, responsible for investigating crimes and offenses of national jurisdiction (smuggling, environmental crimes and crimes committed against federal bodies and authorities), an opportunity was identified to carry out a more in-depth study on how the analysis of FIRs is currently processed within the scope of the Federal Police [10]. The main objective of the work is to try to identify more clearly how the process works, as well as to point out ways to improve it, either through the adoption of a better system, or through the creation of tools with the potential to simplify, accelerate its procedures, automate part of the process and, consequently, point out a methodology that can be used to also achieve better dissemination of the FIR analysis procedure among the police officers involved.

Therefore, initially, it was decided to carry out a study of how the activity of analyzing FIRs is currently generally performed by Federal Police officers assigned to the sectors responsible for analyzing such documents issued by COAF (bearing in mind that there is no standard methodology to be used next), to identify possible points that can be improved, aiming to present possible solutions to be adopted.

We sought to understand the format and means by which the FIRs forwarded by COAF arrive at the Federal Police Stations, more specifically at the Federal Police Station in Santo Ângelo/RS, where the work of the present author is carried out.

It should be noted that, given methodological and editorial issues, this article seeks to present a summary compilation of the main points observed in carrying out the work to obtain the title of master's in information science from the Federal University of Santa Catarina (UFSC), the source indicated for those readers who wish to delve further into the subject.

2. Structure of FIR files

As already mentioned above, the body responsible for producing and forwarding the FIRs to the police authorities is the COAF (Financial Intelligence Unit of Brazil - FIU), further highlighting that such remittance can be carried out in two ways: a) automatically (without provocation), where COAF itself produces the FIR and automatically forwards such document to the investigative authorities; or b) provoked, where the police bodies perform a formal procedure to request such a document to COAF (pointing to a list of natural or legal persons), which then sends the result of such consultation to the PF in the form of computerized files. The flow of the procedure reported above is shown below graphically:

Figure 1. Communication flow of the procedure of the production of FIR.

Once the FIR files have been submitted to the investigative bodies, it is important to explain the format and basic structure they have, as they are produced and sent by COAF in two file formats/types: PDF (Portable Document Format) and CSV (Comma Separate Values). The PDF file is a file format developed by Adobe Systems to represent documents independently and in a fast, simple, and light way, being widely used worldwide, because of the facilities already mentioned. Below (Figure 2) is an image of one of the FIR files sent by COAF in PDF format (whose data was anonymized).

The CSV file is, in a very simple way, a text file, where the fields are presented and separated using semicolons, thus facilitating its import and manipulation of data. Below (Figure 3) is an image of a CSV file produced and forwarded by COAF to the PF (whose data were anonymized):

Figure 2. Image of a PDF (Adobe Acrobat Reader) file of an FIR produced by COAF.

That said, the files contained in the information sent by COAF to the PF are listed below: a) 1 (one) PDF file, where the main suspicious and atypical movements identified are found, together with the names of all the people (physical and legal) that carried out transactions (debits and credits) with those involved; and b) 3 (three) CSV files, one of which contains a list of all those involved (individuals and legal entities that transacted with those involved) and another file where there is a list of the main atypical/suspicious movements with more information detailed information about the transacted values, as well as the opinion of the entity that identified that such movement is suspicious.

As seen above, COAF sends the Federal Police two types of files: one PDF (more concise) and 3 (three) CSV files (more analytical). During this study, it was identified that most of the police officers involved in the analysis of FIR files primarily use only the PDF file, because of the greater readability of this format (it is only necessary to have installed Acrobat Reader on your machine), or many times due to lack of knowledge of the content contained in the CSV files (which contain data that is often not present in the body of the PDF file).

In this way, the importance of using the CSV files forwarded by COAF is highlighted to identify details contained only in this file format and which, often, are not used by Federal Police during their analysis.

3. FIRs analysis

Figure 3. Image of a CSV file produced by COAF for an analysis of PF.

Once the FIR files (PDF and CSV) sent by COAF have been received, the Federal Police begin the work of surveying, researching, and crossing data, to add value to the document produced and forwarded by COAF, as well as with the purpose of to identify and confirm points that may characterize a crime or offense, with the purpose of initiate or deepen an investigation. To this end, the Federal Police make use of various internal systems (not accessible to the public) and external systems (some of them open to the public) to obtain additional data and carry out an analysis and identification of persons (individuals and legal entities) who have characteristics of inconsistency regarding financial capacity/size and purpose. Bearing in mind confidentiality issues, the following are just some of the systems used to collect and cross-reference data: Data contained in the Transparency Portal of the Brazilian Federal Government (which has data on the payment of assistance benefits to individuals less favored), Personal and legal data of individuals and legal entities present at the Federal Revenue Service of Brazil (addresses, family members, registration number, etc.), Data present at Interpol (red distribution list, which aims to point out people wanted for crimes), and Internal Systems of the Federal Police (which contain data on indictments and police occurrences of individuals and legal entities), the main ones used for this purpose being the systems herein called "A", "S" and "P" (because of the principles of police secrecy) [11].

4. Proposal

According to the above, it appears that most of the police officers involved with the analysis of FIRs only use the reading and analysis of the PDF file forwarded by COAF, as well as perform part of the procedure for collecting and crossing data using There are 3 (three) main systems, here called System "A", System "S" and System "P", for confidentiality reasons.

Based on this, given the identification that the data entered in System "A" and System "S" have very similar characteristics, it was decided to carry out a comparative study on the result contained in research carried out in both Systems, to seek to identify which one has a more reliable and complete result. Thus, in possession of data present in FIRs

previously analyzed in the database of the Federal Police Station in Santo Ângelo/RS, it was decided to carry out several queries manually to verify the possibility of pointing out a more accurate system. between the two evaluated (System "A" and System "S"), the result of such study is presented below:

Table 1. Results of search on systems available

Nº FIR	Nº Reg.	System "S"	System "A"	System "P"
48557X	107	6	4	15
50357X	54	6	2	5
51728X	8	2	2	0
60269X	75	5	0	0
60479X	154	7	2	4
60885X	61	2	0	4
61060X	40	2	0	4
63936X	58	0	0	5
64416X	499	16	6	30
64596X	116	3	0	4
TOTAL	**1172**	**49**	**16**	**71**
SUM		**4%**	**1%**	**6%**

Based on the data above, it appears that the use of only one of the systems (in this case System "S") and not using the other (System "A"), can generate a representative productivity gain without great loss of quality in the police information produced, since almost all the data are present in both systems and in a more reliable way, especially in the "S" System. Therefore, the first aspect pointed out by the present work is highlighted here: indication of only one system to be used (between System "S" and System "A"), considering that the consultation in two very similar systems presents results they are also very similar, and one of them (System "S") presents more reliable and complete data, thus making the collection of data carried out by the police more agile and quick.

In addition, given the clear definition of the main data repositories used by police officers for the survey and cross-referencing of data on those involved mentioned in the FIR file, the possibility of developing a computerized procedure

that would automate part of the process was identified, through the elaboration of a prototype software for the search of data and its respective crossing.

Therefore, the main idea was, together with the identification and suggestion of a standard methodology to be followed when analyzing FIRs by Federal Police, also the development of a source code that, in an automated way, would cross-reference data contained in the FIR file with other public and private data, to generate a screen of relevant notes identified with the intersection, as well as to generate a draft FIR analysis report containing the main identified suspect points.

In this way, the suggestion of a new methodology [11] for carrying out the analysis of FIR files whose main points are exposed below includes the use of files with the CSV format to the detriment of using only the PDF file (both forwarded by COAF), the elimination of the use of one of the internal PF systems commonly used by Federal Police officers in data crossing (System "A"), and the automation of part of the data collection and crossing procedure, using for this a prototype software – called RIBOT – developed especially for this purpose.

5. RIBOT prototype software

Thus, because of the necessary characteristics, the possibility of developing the prototype software using the Python programming language (one of the most popular and powerful program languages [13]) and the PostgreSQL relational database was identified, both due to the ease of the learning curve and the possibility of using such tools without the need for any financial expenditure, considering that they are free solutions available on the internet.

For the development of the prototype software, open-source tools were used, either for the elaboration of the new methodology identified for the analysis of FIRs or the preparation of the environment and development of the source code of the solution, the main ones being listed below:

Table 2. Applications used to develop the prototype.

Logo	Software	License	Description	Utilization
	Spyder Anaconda	Open Source	Python development environment	RIBOT software development (source codes)
	Flask	Open Source	Microframework	Used to "run" the software developed on the web
	Jupyter Notebook	Open Source	Python web-based interactive platform	Used to test small source codes, in an agile way
	Bizagi Modeler	Open Source	Flow/Business Modeling Software	Modeling of flowcharts

In the period between July 2021 and February 2023, work was carried out to identify the needs of the software and the database and their respective codification, and at the end of the work, approximately 3000 lines of source codes were produced which are responsible for automating the crossing of part of the data contained in FIRs.

The system has the following functionality: selection of an FIR file (CSV format) by the user, selection option of the "checkbox" type to select the systems that the user wants to use for data crossing (Bolsa Família, BPC, Auxílio Emergencial, Interpol and "P" System (internal Federal Police)). After selecting the CSV file and the systems to be used in the data crossing, the user clicks on "Send" and waits for the file to be processed. data (as shown on the screen below), as well as a DOC file (compatible with Microsoft Word) containing a summary of the main points presented on that screen.

In general, after pointing out and loading the FIR file, the RIBOT system checks whether the content of the CSV file is already inserted in the PostgreSQL database (created especially for this purpose) and, if not, performs the insertion. Thus, each time the user carries out the procedure for loading an FIR file in the RIBOT prototype system, the content of a such document is inserted in the PostgreSQL database, and this procedure is of real importance for future queries and crosses. Once this is done, the system crosses the data contained in the FIR file with other data already present in the PostgreSQL database (previously inserted), in addition to performing searches on other platforms through APIs (Application Programming Interface), to obtain a result and standardized consultation with content certification [12]. The number of registers used in such a cross is illustrated below (Table 3):
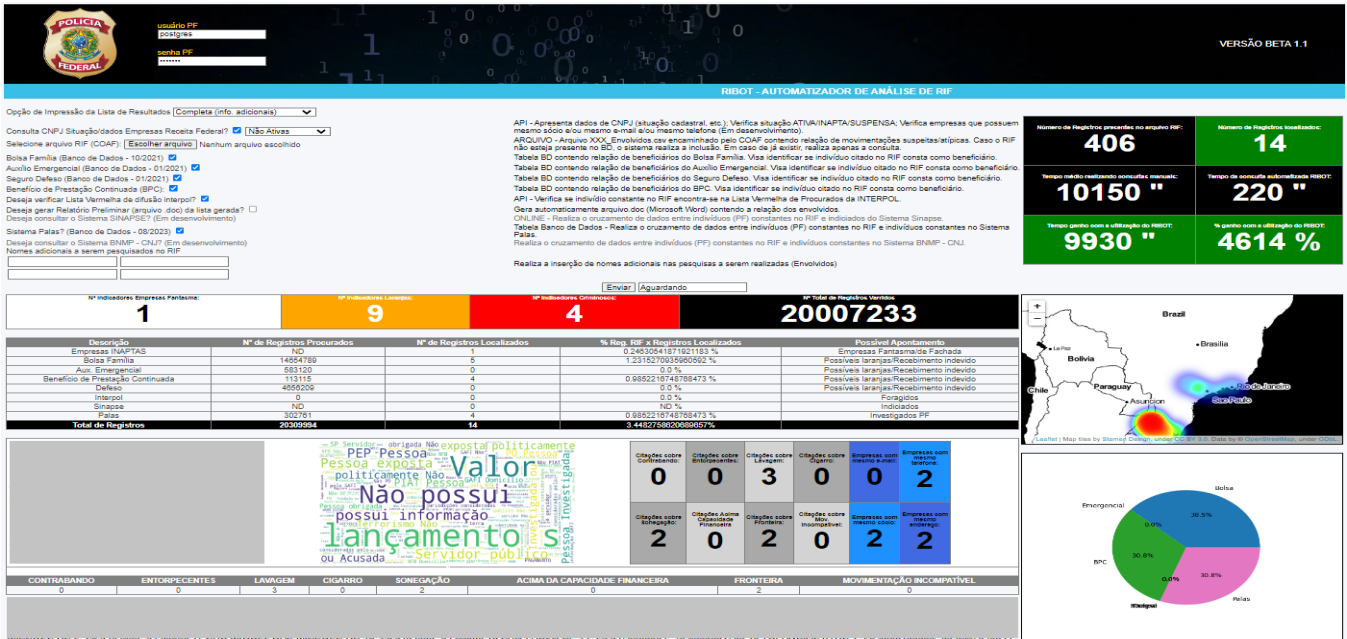
Figure 4. Prototype RIBOT main screen.

Table 3. The number of registers available to search on the database.

Data Base	Nº Registers
Aux. Emergencial (AE)	583120
Bolsa Família (BF)	14654789
Seguro Defeso (SD)	4656209
BPC (BPC)	113115
Palas (PL)	302761
Interpol (IN)	7747
Total	**20317741**

The RIBOT prototype software screen is displayed after processing an FIR file, some of the data being anonymized, because of institutional secrecy issues (Figure 7).

As shown in the image above (Figure 5), you can see some of the notes generated by the system, such as the identification of possible ghost companies, possible "oranges" used in illicit business, and the indication of possible criminals involved in the FIR file.

Figure 5. Prototype RIBOT indicators of suspicious fields (contraband, narcotics, money laundry, cigars, same e-mail, same telephone, money evasion, above financial capacity,

borders, incompatible movement, same partner, same address).

In addition, the system also displays word clouds both related to individuals and to other observations contained in the analytical file sent by COAF (which can expose quotes about possible crimes committed), as well as a heat map that displays the locations where they were carried out. the main suspicious/incompatible financial transactions forwarded by COAF. Based on such indications, the Federal Police who carry out the analysis of the FIR have some data that can help in the preparation of the report and, many times, point out aspects not initially observed.
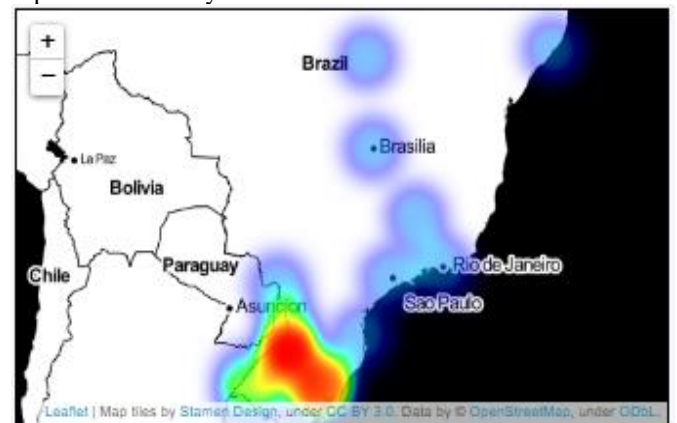

Figure 6. Prototype RIBOT heat map.

The RIBOT prototype software presents fields where comparative performance indices are exposed between the manual used in the analysis of the FIR files (currently used) and with the use of the RIBOT prototype software, where expressive results can be observed regarding the possible productivity gains, with such indicators shown below - example screen (Figure 8).

Figure 7. Prototype RIBOT indicators after the search.

Figure 8. Prototype RIBOT performance indexes.

According to the screen above, it can be seen, as already explained above, that with the use of the RIBOT prototype software to carry out the data crossing, there is a significant productivity gain, in general above 1000% of performance gains, a fact which provides agility to the entire FIR analysis process.

6. Results Presentation

After identifying the format of the files to be used in the data crossing (CSV), the systems to be consulted in the data crossing, as well as after the development of the RIBOT prototype software, it is essential to carry out a study on the results obtained with its use, as well as identification of possible performance gains made possible with the adoption of the RIBOT prototype system in comparison with the consultations and manual crossings currently used. For this, the comparison was performed in two different aspects: a) data crossing results; and b) performance comparison in the survey and crossing of information.

Thus, a data survey was carried out to list those involved who were identified through manual analysis, in comparison with those identified using the RIBOT prototype software, with the numerical results listed below (being consultation times - performance indicators - also featured):

Table 4. Performance indexes search Manual versus RIBOT

FIR	Reg	Productivity			Location involved			
		Manual	RIBOT	% Gain	Manual	RIBOT	Dif.	% RIBOT
*65	153	4590"	100"	4590%	25	10	15	40%
*63	29	870"	62"	1403%	08	04	04	50%
*70	2192	65760"	1549"	4245%	25	18	07	72%
*78	264	6600"	274"	2409%	36	15	21	41,7%
*57	131	3930"	164"	2396%	14	12	02	85,7%
Totals					58	32	26	57,8%

7. Conclusions and Future Works

For the accomplishment of the present work, a study and presentation of how the analysis of FIRs is carried out (in general) currently in the Federal Police was proposed and based on such results, to present a methodology suggestion to better carry out such procedure.

All the data collection procedures were carried out on how the process is carried out (use of repositories, file formats used - PDF and CSV), as well as the steps taken to generate police information based on the data contained in the FIR files.

Based on the results obtained in the data collection phase, it was decided to prepare and present a suggested methodology to be adopted in the FIR analysis procedures, where one of the main differences about the way it is currently carried out is the automation part of the procedure, which is carried out through data processing through a prototype software developed by the author (in Python and PostgreSQL language), especially for this purpose, whose chosen name was RIBOT (acronym of the words FIR and ROBOT).

After subjecting the suggested methodology (with the use of the RIBOT prototype software in the procedure) to efficiency and performance tests, there was great potential for its use by Federal Police in the analysis of FIRs issued by COAF, since both in terms of performance (average speed gains of approximately 3000% compared to the manual procedure), and efficiency (similar results in approximately 58% of the results compared to the manual query), the system has promising rates.

Thus, promising opportunities are identified for improving the way of analyzing RIF within the Federal

Police, especially about the RIBOT prototype system, where it is believed that there is room for the inclusion of new databases, insertion of functionalities linked to artificial intelligence (concerning the identification of crimes according to data contained in RIF files) and, mainly, to broaden the debate among peers of the Brazilian Federal Police on how RIF files are analyzed today, as well as to exchange experiences to build a solution that is increasingly suitable for police work.

Finally, it is reiterated that there is a great potential to be explored and, therefore, further improve the procedures performed by the RIBOT prototype system and, consequently, the RIF analysis methodology within the Federal Police, identifying the opportunity of the present work to serve as a reference/consultation for the development of future studies to be developed, in the related subjects the RIF analysis methodology.

References

[1] Freire, Gustavo Henrique. Ciência da Informação: Temática, Históricas e Fundamentos. Perspect. Ciênc. Inf., Belo Horizonte. 2006 Jan p. 6-19.

[2] Álvarez Diez, R. C., Vega Esparza, R. M., Bañuelos-García, V. H., Villegas Santillán, M. T., Llamas Félix, B. I., Arredondo Luna, V., Hernández Ponce, J. R., García Martínez, F. de M., Alvarado-Peña, L. J., & López-Robles, J. R. (2022). Silver economy and entrepreneurship, an emerging innovative area: An academic, scientific and business framework for building new knowledge. Iberoamerican Journal of Science Measurement and Communication, 2(3). https://doi.org/10.47909/ijsmc.45

[3] Fagundes, P. B.; Freund, G. P.; Macedo, D. D. J. A Produção científica sobre qualidade de dados em big data: um estudo na base de dados Web of Science. Revista Digital de Biblioteconomia e Ciência da Informação, v. 16, p. 194, 2017. http://dx.doi.org/10.20396/rdbci.v16i1.8650412

[4] Macedo, Douglas D.J. de; Wangenheim, Aldo Von ; Dantas, Mario A.R. . A Data Storage Approach for Large-Scale Distributed Medical Systems. In: 2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), 2015, Santa Catarina. 2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems. p. 486. http://dx.doi.org/10.1109/cisis.2015.88

[5] Fagundes, P. B.; Macedo, D. D. J.; Vital, L.; Barros, C. Taxonomias, ontologias e tesauros: possibilidades de contribuição para o processo de Engenharia de Requisitos. Em Questão, v. 26, p. 237-254, 2020. http://dx.doi.org/10.19132/1808-5245261.237-254

[6] Freund, G. P.; Fagundes, P. B.; Macedo, D. D. J.; Dutra, M. L. Mecanismos tecnológicos de segurança da informação no tratamento da veracidade dos dados em ambientes Big Data. Perspectivas em Ciência da Informação (on line), v. 24, p. 124-142, 2019. http://dx.doi.org/10.1590/1981-5344/3348

[7] Freund, G. P.; Sembay, M.; Macedo, D. D. J. Proveniência de Dados e Segurança da Informação: relações interdisciplinares no domínio da Ciência da Informação. Revista Ibero-Americana de Ciência da Informação, v. 12, p. 807-825, 2019. http://dx.doi.org/10.26512/rici.v12.n3.2019.21203

[8] Sembay, Márcio José; Macedo, Douglas Dyllon Jeronimo; Dutra, Moisés Lima. A Proposed Approach for Provenance Data Gathering. Mobile Networks & Applications, v. 26, p. 304-318, 2021. http://dx.doi.org/10.1007/s11036-020-01648-7

[9] CRC/RS, Lavagem de Dinheiro – Um Problema Mundial – Legislação Brasileira, Porto Alegre/RS, 2003.

[10] COAF. Casos e Casos: Coletânea de Casos Brasileiros de Lavagem de Dinheiro/Ministério da Fazenda, Conselho de Controle de Atividades Financeiras – Brasília: COAF, 2016.

[11] Menezes, Romulo Fisch de Berrêdo. Gestão do Conhecimento no Setor Público: o aproveitamento da atividade investigativa da Polícia Federal brasileira. Dissertação. UNB. 2020. Disponível em: https://repositorio.unb.br/bitstream/10482/38735/1/2020_R%C3%B4muloFischdeBerr%C3%AAdoMenezes.pdf. Acesso em: 20/06/2022.

[12] Luiz Pinto, A., de Carvalho Segundo, W. L. R., Dias, T. M. R., Vivian Santos Silva, Gomes, J. C., & Quoniam, L. (2022). Brazil Developing Current Research Information Systems (BrCRIS) as data sources for studies of research . Iberoamerican Journal of Science Measurement and Communication, 2(1). https://doi.org/10.47909/ijsmc.135

[13] Borges, Luiz Eduardo. Python para desenvolvedores. Rio de Janeiro, 2009.