4. FAIR PRINCIPLES AND GOVERNMENT DATABASE MANAGEMENT: SHARING VITAL RECORDS DATA THROUGH THE GOVDATA DATA LAKE

Cláudio José Silva Ribeiro[113]
Ana Cristina Meirelles Velho[114]

4.1 INTRODUCTION

Information and knowledge use has boosted investigation projects, both in academic environments and business and govern organizations. We are witnessing the emergence of an informational and increasingly globalized (Castells, 1999).

It can be said that we are in the cyberspace era, where collaboration, instantaneous and digital fluency are increasingly present in everyday life (Barreto, 2014). Drive by the notion of "data avalanche", the efforts to gather information to support business management (*Business Intelligence – BI*) have shifted towards *Big Data/Analytics* solutions, causing reflections on the actions currently in development.

Since Castells, it is clear that the informational economy proposes the use of data and information to obtain results, so it appears that we are not facing a novelty in research in the field of information management. The Information Science field has promoted studies that explore this theme, especially in approaches for information assets management on the Web in managerial, tactical and operational levels (Velho, 2007; Ribeiro, 2008) and towards digital curation (Sayão; Sales, 2013).

The identification of standards for metadata, interoperability, sharing, archiving, access, reuse of collection, as well as the process and intelligent discovery of resources through ontologies and taxonomies, have become part of the information manager's concerns (Ribeiro, 2014; Sayão; Sales, 2013). In addition to this, the need to preform predictive analysis in datasets using statistical approach, besides group processing of big collections with *data mining* and simulation, promoted the displacement of the information manager in *BI* to *Analytics* activity (Velho, 2007; Siegel, 2013), materialized in *data lakes* and *Dataponds* structures (Inmon, 2016). Therefore, it is necessary to establish new paradigms of knowledge to deal with those extensive collections of intangible and virtual supplies (Levy, 1996).

The investigation is characterized as a descriptive, exploratory research with a qualitative approach, with bibliographic and documentary research and case study (Gil, 2002). Starting from a literature review on initiatives to

113 PhD in Information Science. Graduated in Engineering. Associate Professor at Unirio where he works in the Department of Technical-Documentary Processes and in the Postgraduate Program in Library Science.

114 Master in Information Science. Graduated in Systems Analysis. Business Intelligence Project Manager at DATAPREV

share government data, this report presents the objective, main components of GovData platform and a clipping for possible uses by different federal entities and public institutions, including Teaching and Research.

In addition to this introduction, this paper presents four more sections: the next one, with the theoretical framework used as reference, followed by the empirical field, the outcomes, and final consideration.

4.2 SHARING GOVERNMENT DATA USING *FAIR* PRINCIPLES

To better understand government data sharing, it is necessary to go back a few years to look for landmarks, for since the launch of the Transparency Portal in 2002, and subsequently the Access to Information Law (LAI – Lei de Acesso à Informação), in 2011, there was the crystallization of processes of dissemination of government information.

Driven by the movement of access to information catalyzed by the "Letter of Service to the Citizen" and by the *Memorandum on Transparency and Open Government* of the American government (Ribeiro; Almeida, 2011), the Brazilian government joined the *Open Government Partnership (OGP)*[115] initiative, improving even more the processes for dissemination, sharing, and reuse of information produced in the governmental sphere (CGU, 2012).

As part of the plans for the implementation of Open Government partnership, the Brazilian Open Data Portal has started, with dataset provision on government actions. As a result of a joint project between the government and the society, represented by standardization bodies, universities and non-governmental organizations, the portal was supported by actions linked to the National Open Data Infrastructure (Infraestrutura Nacional de Dados Abertos - INDA). This last effort presented a set of standards, technologies, procedures, and mechanisms of control necessary to meet the conditions for disseminating and sharing data and public information in the Open Data model, in accordance with the provisions of the Government Interoperability Project (e-Ping project) (Ribeiro; Almeida, 2011).

All these initiatives, directed to the use of technologies aiming at optimizing government internal processes, formed the concept of "electronic government". As of 2015, the government focus becomes "Citizen-centered", a global trend, and it is then called "digital government" (Brasil, [2018?]) .

Within the comprehensive concept of digital govern, the need to implement services in the course of Digital Transformation[116] paved the way for some structuring projects, among which GovData and ConectaGov stand out. They aim , among other things, at facilitating the exchange of information between the public institutions defined by the decrees 8.789 dated in 2016 and 10.046 dated in 2019 (Brasil, 2016; Brasil, 2019), therefore, the interoperability among information under the government institutions. GovData initiative focuses on the intero-

115 International Initiative for governmental transparency and fight against corruption. Its launch was led by Brazil and the Unites States (OGP, 2011; CGU, 2012).

116 Term used in Brazil and in Other countries and that encompasses government projects with the use of digital technologies aiming at increasing the State capacity to offer services to the citizens and data sharing.

perability through database and ConectaGov, through real-time data exchange. Within the scope of this report, the authors' concentration will be given in the evaluation of GovData databases.

4.2.1 GovData Initiative

Decree 8.789, dated June 29, 2016, established the following:

> The bodies and entities of direct and indirect public administration and other entities directly or indirectly controlled by the Union that are holders or responsible for the management of official databases will make available the access to data under their management to the bodies and entities of the direct, autonomous and foundational federal public administration interested in accessing data under their management, under the terms of this Decree (BRASIL, 2016, *online*).

Later, it was revoked and replaced by decree 10.046, dated October 9, 2019, which maintained the same guidelines relevant to this report, namely, the sharing of databases between the institutions and the creation of Central Committee for data Governance[117]. It can be seen that the orientation for sharing databases, in addition to raising awareness on efficient management, has acquired legal force.

The report by the Ministry of Planning, Government Transition 2018-2019, presented 15 structuring themes, including Digital Government. It supports the presentation of the scenario that led to the proposal of GovData as a solution to enable the continuous access to databases by the institutions, without requiring the permission protocols through an agreement that would be necessary for each action.

The history of Brazil in providing technological solutions at the service of society is in line with the world scenario, in which several countries have virtually approached the population via remote channels. Some European countries and the U.S. maintain strategies of Digital Transformation for more than a decade, according to ONU. Furthermore, in this preparatory report for, at the time, the future government that would take over the country in 2018, Denmark was the country of reference, for the amplitude of its initiatives, from fully digital services to data sharing and open government.

It should be noted that in 2020, Brazil ranked first in Latin America as provider of digital services and, in the Americas, it was second Only to the United States (Brasil, 2020).

GovData initiative had as its motivation to meet the need of public managers to define policies based on sufficient information, reducing the empiricism of such decisions (Brasil, [2018?]). Thus, if the State gathers relevant databases that can support analyzes to subsidize actions of its managers, then treating the dispersion of such bases and offer them in a viable way via technologies of access optimize the use of resources and provide more efficiency. This perception of existence of relevant information, not yet adequately available, and the lack of information by managers on the other side, form the core of the justification of this initiative.

117 It is not the objective of this report to discuss possible similarities and differences between the terms data governance and data curation.

The interoperability of information in the federal government was the theme of the 4th Nation Forum of Union Transfer – Sharing, analyze and safety (Brasil, 2019a). At the time, the three components of GovData solution were presented: *data lake*, the access tools and the data science. *Data lake* corresponds to the data lake, a group of databases that, according to legal determination, must be available for access. The access tools (HUE, Qlik, RStudio and MicroStrategy) correspond to resources available so users can work with the databases, enabling from the production of large cross-checking data to dashboards. And, as a methodological support for the use of databases, the data Science techniques, to guarantee the expected results in the data analysis cycle.

In relation to the actors involved in GovData management, the Digital Government Secretariat of the Special Secretariat of Debureaucratization, Management and Digital Government of the Ministry of Economics is responsible for the role of Executive Office of the Data Governance Central Committee, which coordinates the activities of the (Brasil, 2019b). For issues that transcend the scope of data sharing, the Central Committee reports itself to the Interministerial Committee of Governance, established by decree 9.203, dated in 2017 (Brasil, 2017). Among the defined roles are those related to data holding (from data manager, data custodians, interoperability platform manager), and those related to the use (data receiver, data requester) (Brasil, 2019b).

When analyzing GovData features, it is possible to infer that the sharing assumptions pointed out by FAIR principles can be used with the aim of aligning its structures to the reuse of dataset proposed for C&T field.

4.2.2 *FAIR* Principles and the requirements for evaluation

The motivation for proposing principles for sharing and reusing data drove the formulation of *FAIR* principles (*Findable, Accessible, Interoperable and* Reusable). In this report, it is assumed that these principles are already disseminated, as recently, there have been different studies covering the *FAIR* theme in Brazilian universities' context, as observed in Henning *et al*. (2018); in Moreira *et al*. (2019); in Ribeiro (2019) and in Monteiro and Santana (2020).

In essence, *FAIR* principles aim at data interoperability and reuse. This is carried out by meeting the requirements below:

a. data and its metadata using persistent and universal identifiers;

b. data and its metadata represented by a formal, accessible, shared and widely applicable language for the representation of knowledge;

c. data and its metadata must have qualified references for other metadata and data. These elements and their relations need to be described semantically;

d. data and its metadata must have its origin indicated for use and reuse, as well as its transformation process and its history;

e. data and metadata must be clearly licensed;

f. use standards shared by communities.

These requirements were carefully analyzed considering initiatives for evaluation identified in (2019), Monteiro and Santana (2020) and Taco de Bruin *et al*. (2020). The list of questions used for the investigation was structured according to the Organizational, Digital Content and technological view.

The Organizational view covered aspect linked to the management infrastructure and data governance, including the existence of specific politics and profiles suitable for performing management activities

The Digital Content view covered aspects linked to data and metadata, identification strategies, semantic description and indexation.

The technological view covered aspects of standardization, technological infrastructure and communication protocol, in addition to collection and provision services.

In addition, when adopting *FAIR* principles, *GOFAIR* ([20–]) presents *FAIRification Process* as a way to enable institutions to align their *datasets* to those principles. This process is organized in the following stages:

a. gather and analyze datasets;

b. define and represent the semantic model for datasets;

c. link data;

d. verify data licensing;

e. establish metadata for the sets;

f. publish resources as *FAIR* data.

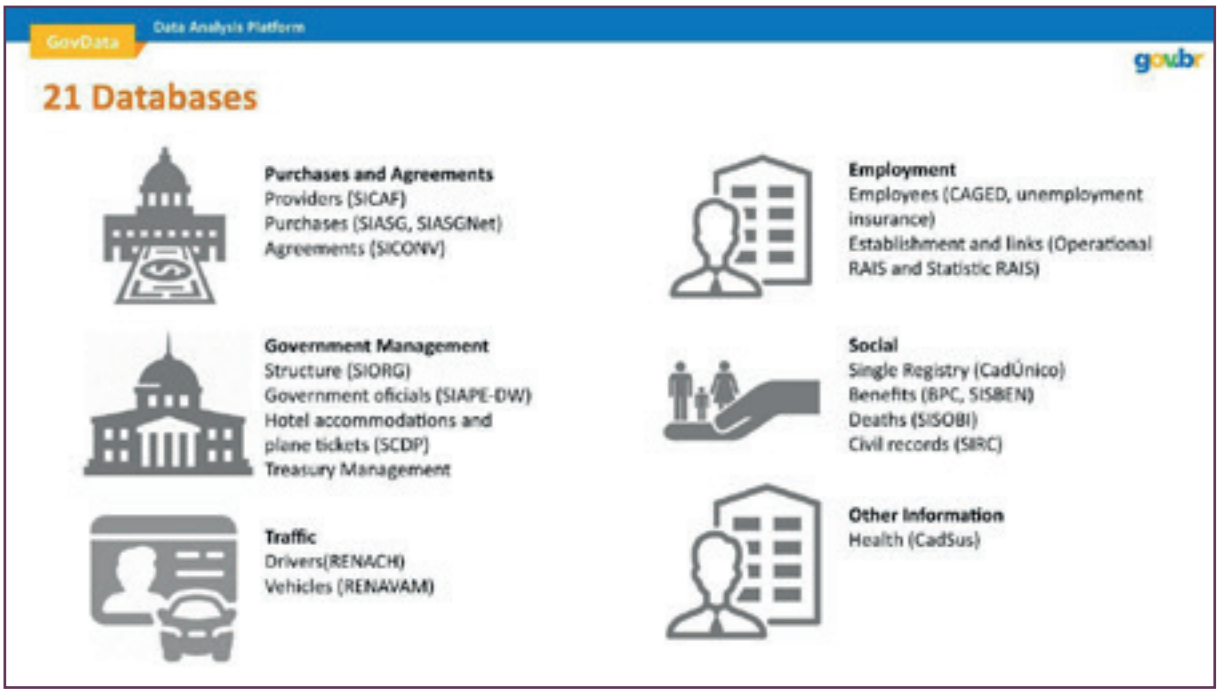
4.3 EMPIRICAL FIELD: SNAPSHOTS OF DATASETS ANALYZED

The Brazilian Social Welfare provides the citizens information about the access to its services, from appointments and orientation to confirmation and granting of benefits. The virtualization of access to these services via the Internet, mobile devices and call centers brings even greater enrichment to this topic. To fulfill its purpose, in addition to the information inherent to its service management, it manages large registers of social information, which are made up of databases of individuals and legal entities in the country, their civil events (birth, marriage, death), their working life and related events. These databases have their origin both in their sources and in external sources provided by other government institutions.

In face of the diversity and correlations between its information, its metadata management is part of its database curation.

At the aforementioned event, 4th National Forum of Union Transfer – Sharing, analysis and safety (Brasil, 2019a), 21 databases available in GovData platform were presented at the time, among which, the Civil Records (SIRC), database who's the analysis snapshot will be used.

Figure 1 – Database presentation



Source: Brasil (2019a).

The National Civil Registry Information System (Sistema Nacional de Informações de Registro Civil – SIRC) is the digital means to obtain civil data on birth, marriage, deaths and stillbirth, which are sent by the registry offices aiming, among others, at eradicating the under-registration in the country, qualify other governmental databases, subsidize public policies and help reduce frauds in granting of benefits and crimes as falsification and human trafficking (Brasil, [2019?]). GovData presents the following data collection:

a. births: identification of the individual, date, and place of birth, date and place of registry, place of residence, gender, identification of family relation;

b. deaths: identification of the individual, nationality, date and place of death, date and place of registry, marital status, gender, identification of names of parents, cause of death, date and place of birth, place of residence, identifier of the social security benefit, occupation;

c. marriage: identification of spouses, nationalities, date and place of the ceremony, date and place of registry, name of spouses' parents, system of marriage, information on the religious marriage, place of residence, information on the dissolution of marriage, occupations;

d. history: previous versions of birth, death, and marriage certificate that have been altered in some way;

e. operational: collection data encoding (for instance, code and description of the system of marriage), services/ registry office register, data on the operations of file upload to feed the database.

Data collection related to birth and death certificates were used for analyzes, due to the need to snapshot this report as well as its importance as acts that represent the demography of the country.

This investigation did not aim at verifying the correlation of information available to the legislation relevant to the Civil Registry in the country. Nevertheless, the importance of future studies with this aim is recognized.

The description of dataset available on birth and death certificate is available with the name of the attribute, its format and extended description, therefore, technical metadata. Information on its update and its custodian are also available. There is no additional information on related data, not being possible to infer if there is no other relations or if it is only its omission (CKAN, [201-]).

In relation to present the dataset, it is possible to verify that the last update occurred more than one year ago, possibly indicating its deactivation. However, it is understood that this fact does not compromise the analysis proposed in this article, since it is an analysis of the implemented initiative.

The available datasets available on the selected themes, births and deaths, have a significant degree of use for understanding these phenomena and its dimensions. These are official administrative records that can support population studies together with other sources from institutions that are recognized in the country, such as IBGE (IBGE, [201-]). Data related to the place of events can show the population movement in relation to the birthplace and place of residence. The occupation, age, and gender according to the place of residence have the statistical potential for historical analysis of deaths. The cause of death, age, gender, and place constitute a group of data that, when combined, can generate indicators of interest to deepen the reality.

4.4 OUTCOMES

Datasets were analyzed considering what was presented in Section 2. The datasets were categorized and gathered for analysis by the themes described. Table 1 gathers the main considerations, resulting from the evaluation process.

Table 1 – Analysis of the compliance to requirements by GovData

View	Requirement	Analysis
Organizational	Organizational and personnel structure defined	Yes. *Datasets* analyzed are protected by Dataprev. The company has specific area for data governance.
	Roles and responsibilities defined	Yes. *Datasets* analyzed are protected by Dataprev. The company has specific area for data governance.
	Current policies include *FAIR* principles	No.
	Teams dedicated to data management, metadata, and data science.	Yes. *Datasets* analyzed are protected by Dataprev. The company has specific area for data governance and for activities related to data science.
Digital Content	Data and metadata with persistent identifier	Partially. Data and metadata identifiers cannot be framed in the concept of persistent ID.
	Metadata enrichment	Partially. The metadata model follows the established by CKAN tool.
	Data and metadata retrievable by the identifier and with standardized, open and free protocol.	Yes. HTTP. Especially for metadata - DCAT (CKAN).
	Metadata available, even after datasets are removed.	There was no evidence.
	Data and metadata include references qualified for Other elements (data and metadata)	There was no evidence.
	Data and metadata are licensed	Metadata licensed with Creative Commons. There is no indication for data.
	Data and metadata with detailed provenance	It does not have.

View	Requirement	Analysis
Technological	The protocol allows procedure of authorization and authentication, when necessary.	Yes.
	Data and metadata with formal, accessible, shared and applicable language for knowledge representation.	DCAT/RDF metadata with Harvesting Java protocol and comparable to OAI-PMH. REST architecture. Data with HUE/Hadoop and RStudio.
	Data and metadata use vocabularies that follow *FAIR* principles	No
	Data and metadata meet domain standards	Partially. They follow local standards to the custodial institutions, but the alignment with the VCGE[118] was not recognized.

Source: elaborated by authors (2024).

The outcome obtained in the context of the organizational view was satisfactory, since the institutions (Dataprev and Serpro) that manage the datasets have high specialization, experience and technological resources for the adequate governance.

There is also a need to prioritize projects that deal with the vision of digital content, since the requirements linked to the metadata representation and the use of semantic models need to be met in its totality to allow the adequate reuse of sets available. Another important requirement is linked to the use of persistent identifiers for accessing data and metadata. The debate about the generation of identifiers for research data can be extended to encompass government datasets as well.

Similar to the organizational view, satisfactory results were obtained for the technological view. Only the lack of alignment with the VCGE was pointed out. According to Ribeiro and Pereira (2015), this vocabulary was created in 2011 and could be in use to improve data semantics.

Finally, after cross-check the requirements in Table 1 with the datasets under analysis, and based on *Fairification Process* presented in section 2.2, it as verifies that it can be possible to search for alignments with *FAIR* principles. The datasets analyzed and described in section 3 demonstrate the wealth of the relations that can be built from the information of the civil registry of births and deaths registered in the country's registry offices.

118 VCGE - Vocabulário Controlado de Governo Eletrônico (Controlled Vocabulary of Electronic Government). Available from: https://www.gov.br/governodigital/pt-br/governanca-de-dados/vocabulario-controlado-do-governo-eletronico. Access on: 15 Oct. 2020.

It is understood that the datasets require some level of anonymization to prevent the identification of individuals, but this transformation is possible in face of the inputs already available and it does not reduce the potential that these databases have to produce new information.

4.5 FINAL CONSIDERATIONS

Data sharing is essential in the development of collaborative actions for research increase. Sharing data is essential in the collaborative development of actions to increase research. Discussing a topic of recent interest, the increase in the speed to obtain results is explicit when analyzing the context of projects linked to the development of coronavirus vaccine (SARS-CoV-2 - COVID-19) and in particular the VODAN project[119].

Reusing data is key element in this increment, therefore, the descriptions in metadata and semantic models are essential items for proper understanding available datasets.

In addition to the context of Science and Technology, it is possible to infer that *FAIR* principles are also applicable in data sharing and government information context. In that matter, the efforts mentioned in this report can serve as a starting point for better data and information dissemination for the organized Society and Teaching and Research institutions.

At the time of writing this report, GovData platform is in the institutional site of the Economy Ministry with its three components already mentioned [see 2.1], however, with no reference to databases available in *data lake*. This fact seems to indicate to the authors that there is priority in offering technological resources that provide access and analysis of data Science than the availability of government datasets.

Finally, this report aims at incorporating to the scientific debate the possibility of engaging Society in the actions of the open government and in the development of open science. It is possible to infer that it is a long path; the first steps taken toward *Citizen Analyst* (Allemang, 2010; Ribeiro; Almeida, 2011) can now be followed searching for *Citizen Data Scientist* (Banker, 2018).

REFERENCES

ALLEMANG, D. **Modalities:** The Magazine of Semantic Web. n. 9. 2010. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.216&rep=rep1&type=pdf. Access on: 7 Oct. 2020.

BANKER, S. The Citizen Data Scientist. **Forbes**, New York, 21 Jan. 2018. Innovation. Available from: https://www.forbes.com/sites/stevebanker/2018/01/19/the-citizen-data-scientist/. Access on: 2 Oct. 2024.

BARRETO, A. A. A aventura de perceber significados. **DataGramaZero**, Rio de Janeiro, v. 15, n. 3, jun. 2014.

119 *Virus Outbreak Data Network* - project based on *FAIR* network for sharing data about COVID-19. Available from: https://www.go-fair.org/implementation-networks/overview/vodan/. Access on: 7 Oct. 2020.

BRASIL. **10 - Governo Digital** [2018?] Available from: https://transicao.planejamento.gov.br/wp-content/uploads/2018/11/10_Governo-Digital_versão_para_publicação.pdf. Access on: 2 Oct. 2020.

BRASIL. **Decreto nº 10.046, de 9 de outubro de 2019.** Dispõe sobre a governança no compartilhamento de dados no âmbito da administração pública federal e institui o Cadastro Base do Cidadão e o Comitê Central de Governança de Dados. Brasília, DF: Presidência da República, 2019b. Available from: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D10046.htm. Access on: 10 Oct. 2020.

BRASIL. **Decreto nº 8.789, de 29 de junho de 2016.** Dispõe sobre o compartilhamento de bases de dados na administração pública federal. Brasília, DF: Presidência da República, 2016. Available from: http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2016/Decreto/D8789.htm. Access on: 2 Oct. 2024.

BRASIL. **Decreto nº 9.203, de 22 de novembro de 2017.** Dispõe sobre a política de governança da administração pública federal direta, autárquica e fundacional. Brasília, DF: Presidência da República, 2017. Available from: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Decreto/D9203.htm. Access on: 2 Oct. 2024.

BRASIL. **IV Fórum Nacional das Transferências da União – Compartilhamento, análise e segurança.** Brasília, DF: Ministério da Economia, 2019a. Slide Share. 43 slides. Available from: https://www.gov.br/plataformamaisbrasil/pt-br/ajuda/Apresentacoes/arquivos-e-imagens/governanca_de_dados_-_sgd_me-1.pdf. Access on: 2 Oct. 2024.

BRASIL. Brasil em primeiro lugar na América Latina. **Govbr**. Brasília, 22 Jul. 2020. Available from: https://www.gov.br/pt-br/noticias/financas-impostos-e-gestao-publica/2020/07/brasil-esta-entre-os-20-paises-com-melhor-oferta-de-servicos-digitais. Access on: 2 Oct. 2024.

BRASIL. **SIRC** – Sistema Nacional de Informações do Registro Civil. [*S. l.: s. n.*], [2019?]. Available from: https://sirc.gov.br. Access on: 2 Oct. 2024.

CASTELLS, Manuel. **A Sociedade em rede**. São Paulo: Paz e Terra, 1999.

CGU. DEFESA – Gestão da informação será o foco da Defesa no "Governo Aberto". **Govbr**, Brasília, 5 Dec. 2012. Available from: https://www.gov.br/defesa/pt-br/assuntos/noticias/ultimas-noticias/05-12-2012-defesa-gestao-da-informacao-sera-o-foco-da-defesa-no-governo-aberto. Access on: 2 Oct. 2024.

CKAN. **CKAN-GOVDATA**. [*S. l.: s. n.*], [201-]. Available from: https://ck.govdata.gov.br/. Access on: 9 Oct. 2020.

GIL, A. C. **Como elaborar projetos de pesquisa.** 4ª ed. São Paulo: Atlas, 2002.

GOFAIR. **Fairification process**. [*S. l.: s. n.*], [20--]. Available from: https://www.go-fair.org/fair-principles/fairification-process/. Access on: 15 Oct. 2020.

HENNING, P. C. *et al*. Desmistificando os princípios fair: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados fair. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Proceedings** [...]. Londrina: UEL, 2018. Available from: http://hdl.handle.net/20.500.11959/brapci/103506. Access on: 2 Oct. 2024.

IBGE. **População**. [*S. l.: s. n.*], [201-]. Available from: https://www.ibge.gov.br/estatisticas/sociais/populacao.html. Access on: 2 Oct. 2024.

INMON, B. **Data Lake architecture:** Designing the Data Lake and avoiding the garbage dump. New Jersey: Technics Publications, 2016.

LEVY, P. **O que é virtual?** São Paulo: Ed. 34, 1996.

MONTEIRO, E. C. S. A.; SANTANA, R. C. G. Repositórios de dados científicos na infraestrutura de pesquisa: adoção dos princípios fair. **Ciência da Informação,** Brasília, v. 48, n. 3, 2019. Available from: http://hdl.handle.net/20.500.11959/brapci/136407. Access on: 2 Oct. 2024.

MOREIRA, J. L. *et al*. Towards findable, accessible, interoperable and reusable (fair) data repositories: improving a data repository to behave as a fair data point. **Liinc em revista,** Rio de Janeiro, v. 15, n. 2, 2019. DOI: 10.18617/liinc.v15i2.4817. Available from: https://revista.ibict.br/liinc/article/view/4817. Access on: 2 Oct. 2024.

OGP. Declaração de governo aberto. **Open Government Partnership**, Sept. 2011. Available from: www.opengovpartnership.org/open-government-declaration. Access on: 2 Oct. 2024.

RIBEIRO, C. J. S.; ALMEIDA, R. F. Dados Abertos Governamentais (Open Government Data): instrumento para exercício de cidadania pela sociedade. *In:* ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília. **Proceedings** [...]. Brasília: UnB, 2011.

RIBEIRO, C. J. S. Diretrizes para o projeto de portais de informação: uma proposta interdisciplinar baseada na Análise de Domínio e Arquitetura da Informação. 2008. 298 f. Tese (Doutorado em Ciência da Informação) - Instituto de Arte e Comunicação Social, Universidade Federal Fluminense / IBICT, 2008.

RIBEIRO, C. J. S. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia (Itec),** João Pessoa/Marília, v. 1, p. 96-105, 2014.

RIBEIRO, C. J. S. Digital repositories maturity model: a way to its adoption in research data management. **Liinc em revista,** Rio de Janeiro, v. 15, n. 2, 2019. DOI: 10.18617/liinc.v15i2.4816. Available from: https://revista.ibict.br/liinc/article/view/4816. Access on: 2 Oct. 2024.

RIBEIRO, C. J. S.; PEREIRA, D. V. A publicação de dados governamentais abertos: proposta de revisão da classe sobre Previdência Social do Vocabulário Controlado do Governo Eletrônico. **Transinformação,** Campinas, v.

27, n. 1, p. 73-82, Apr. 2015. DOI: 10.1590/0103-37862015000100007. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862015000100073&lng=en&nrm=iso. Access on: 2 Oct. 2024.

SAYÃO, L.; SALES, L. DADOS DE PESQUISA: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Pesquisa Brasileira em Ciência da Informação,** João Pessoa, v. 8, n. 2. 2013. Available from: https://pbcib.com/index.php/pbcib/article/view/18634. Access on: 2 Oct. 2024.

SIEGEL, E. **Predictive Analytics.** New Jersey: John Wileu & Sons, 2013.

TACO DE BRUIN *et al.* Do I-PASS for FAIR. A self assessment tool to measure the FAIR-ness of an organization. **Zenodo**, [*s. l.*], 2020. DOI: 10.5281/zenodo.4080867. Available from: https://zenodo.org/records/4080867. Access on: 2 Oct. 2024.

VELHO, A. C. M. **A tomada de decisão na Previdência Social: uma reflexão das ações do produtor de informações da Dataprev.** 2007. 137 f. Dissertação (Mestrado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal Fluminens, Rio de Janeiro, 2007.

How to cite this chapter: RIBEIRO, Claúdio José Silva; VELHO, Ana Cristina Meirelles. FAIR principles ans government databse management: sharing vital records data through the GovData data lake. *In*: SALES, Luana Faria; VEIGA, Viviane Santos de Oliveira; VIDOTTI, Silvana Aparecida Borsetti Gregório; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **FAIR Principles Applied To Research Data Management**: Brazilian Experiences. Brasília, DF: Editora Ibict, 2024. cap. 4, p. 58-70. DOI: 10.22477/9788570131959.cap4.