



## GT-8 – Informação e Tecnologia

ISSN 2177-3688

### PROPOSIÇÃO DE FLUXO DE INFORMACIONAL PARA OBSERVATÓRIOS GOVERNAMENTAIS

#### *PROPOSAL OF INFORMATION FLOW FOR GOVERNMENT OBSERVATORIES*

**Caio Saraiva Coneglian** - Instituto Brasileiro de Informação em Ciência e Tecnologia

**Diego José Macêdo** - Instituto Brasileiro de Informação em Ciência e Tecnologia

**Ingrid Torres Schiess** - Instituto Brasileiro de Informação em Ciência e Tecnologia

**Milton Shintaku** - Instituto Brasileiro de Informação em Ciência e Tecnologia

#### **Modalidade: Trabalho Completo**

**Resumo:** Os estudos sobre observatórios têm recebido cada vez mais atenção devido às lacunas existentes nesse campo. Embora haja dificuldades em alcançar um consenso teórico-conceitual, observatórios desempenham um papel importante na disseminação de informações, especialmente no âmbito governamental. No entanto, o processo de coleta, armazenamento e tratamento de informações enfrenta desafios na compreensão das ferramentas, técnicas e métodos disponíveis. Assim, o presente estudo tem por objetivo de estudo propor um modelo de fluxo informacional abrangente para coleta, processamento e armazenamento de informações em observatórios governamentais, utilizando uma abordagem exploratória para analisar técnicas e tecnologias de gestão de informação. O resultado consiste em modelo de fluxo informacional, abarcando todas as etapas do processo de coleta, tratamento e armazenamento, considerando as ferramentas, as técnicas e os métodos empregados na literatura e nas práticas atuais. Concluiu-se que a sistematização desse processo contribui para a gestão de informações e tomada de decisão, ao levar em conta tanto os aspectos técnicos do sistema de informação quanto os elementos estratégicos da visão do observatório. Dessa forma, o modelo de fluxo informacional proposto visa garantir a qualidade das informações utilizadas nos observatórios governamentais.

**Palavras-chave:** observatórios governamentais; gestão da informação; sistemas de informação; fluxo informacional.

#### **Abstract:**

The studies on observatories have been receiving increasing attention due to existing gaps in this field. Although there are difficulties in reaching a theoretical and conceptual consensus, observatories play an important role in the dissemination of information, especially in the governmental context. However, the process of collecting, storing, and processing information faces challenges in understanding the available tools, techniques, and methods. Therefore, the present study aims to propose a comprehensive model for the collection, processing, and storage of information in government observatories, using an exploratory approach to analyze information management techniques and technologies. The result consists of an information flow model that encompasses all stages of the collection, processing, and storage process, considering the tools, techniques, and methods employed in the literature and current practices. It was concluded that the systematization of this process contributes to information management and decision-making, taking into account both the technical aspects of the information system and the strategic elements of the observatory's vision. Thus, the proposed information flow model aims to ensure the quality of information used in government observatories.

**Keywords:** government observatories; information management; information systems; information flow.

## 1 INTRODUÇÃO

Observatório, pela formação morfológica do termo em português, significa o local em que se observa. Por isso, o nome dado primeiramente para os lugares de observação astronômica. Entretanto, o que significa observar? Segundo a sua origem latina, *ob* (sobre) mais *servare* (cuidar, manter a salvo), seria sobre manter algo a salvo. Com isso, observar está na mesma família de preservar, conservar, entre outros, alinhando a visão filosófica de observatório, definida por Mora (2001) como ação voltada a vigiar atentamente para determinados fins.

A partir do momento em que os observatórios extrapolaram o seu lugar físico, tradicionalmente conhecido como espaço de observação dos astros, eles passaram a se apresentar como espaços informacionais, nos quais com o crescimento das tecnologias de informação e comunicação conferiu novas funções aos observatórios. Como resultado, surgiram novas formas de atuação e características que nem sempre são claramente compreendidas.

Soares, Ferneda e Prado (2018) evidenciam que estudos sobre observatórios trazem um caráter diverso e plural acerca da temática na atualidade. Embora tragam estudos sobre relatos de implementação de observatórios, constata-se a falta de estudos com análise teórica sobre o tema. Da mesma forma, Vieira, Barbosa, Farias Júnior e Moura (2022) verificam, por meio de uma revisão de literatura, a ausência de um referencial teórico-conceitual homogêneo acerca dos observatórios na atualidade.

Macêdo, Maricato e Shintaku (2021) também relatam as dificuldades de conceituação de observatórios como sistemas de informação, principalmente na ciência e tecnologia, ante aos diversos contextos sociais possíveis de uso. No entanto, para os autores, observatórios podem atuar para ofertar informações de monitoramento, avaliação, análise e previsão, por intermédio de dados coletados. Assim, de forma simplificada, os observatórios precisam coletar, armazenar e tratar as informações para ofertar serviços.

Nesse contexto, evidencia a necessidade de discussão sobre a informatização das atividades voltadas aos observatórios, na medida em que a coleta precisa ser efetuada automaticamente em várias fontes, o armazenamento deve estar de acordo com o tipo de

dados e informações e o tratamento requer os cuidados para cuidar dos dados e informações para a oferta de serviços. Assim, o objetivo deste estudo consiste em propor um modelo de fluxo informacional para a coleta, processamento e armazenamento de dados em observatórios, com ênfase nos observatórios governamentais. Serão consideradas as ferramentas, técnicas e métodos relevantes, a fim de contribuir para a discussão e aprimoramento desses sistemas de informação. Destaca-se a necessidade de diferenciar os observatórios direcionados para a informatização governamental, demandando estudos que respaldam a sua criação.

Neste trabalho, por tratar de um modelo informacional para observatório governamental, entende-se informação como a definida pela Lei nº 12.527 de 18 de novembro de 2011<sup>1</sup>, Lei de Acesso à Informação (LAI). Assim, conforme o artigo 4º, inciso primeiro: “Informação: dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato” e em seu inciso segundo “documento: unidade de registro de informações, qualquer que seja o suporte ou formato” (BRASIL, 2011, p. 1).

## **2 METODOLOGIA**

Este trabalho se encontra vinculado a um projeto mais amplo que visa propor um modelo de observatórios governamentais que irá apoiar as mais diversas organizações no processo da concepção e implementação de seu sistema de informação. Para tanto, estuda, entre outros temas, ferramentas que possam compor um ecossistema de informação com funções de observatório.

Partindo de tal projeto, um dos mais importantes elementos tange à definição do processo de coleta, tratamento e armazenamento de informações, pois serve de base para a oferta de serviços aos usuários. Assim, ao buscar definir um modelo de informações para tais observatórios, este trabalho definiu algumas etapas essenciais para o desenvolvimento do modelo.

Na primeira etapa, foi realizado o levantamento dos observatórios por meio do metabuscador *Google*, com buscas realizadas em quatro modos: pesquisa por observatórios em cada UF, como por exemplo "observatórios de MG" ou "observatórios de Minas Gerais";

---

<sup>1</sup> Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm)

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

pesquisa no portal Gov.br; pesquisa por "*observatory of government*"; e por fim, busca por "*observatory of Science and technology*". Para a seleção dos observatórios participantes, foram escolhidos aqueles que possuíam vínculo governamental.

Foi realizada análise de conteúdo nos observatórios selecionados, seguindo a abordagem proposta por Bardin (2011), a fim de identificar a tipologia documental presente nos respectivos *sites*, no qual deve-se focar no conteúdo das mensagens expressas nos textos apresentados, de forma a possibilitar a identificação dos elementos. Nesse sentido, foi importante identificar a tipologia documental presentes nos observatórios, com o intuito de embasar a definição dos aspectos relacionados ao tratamento das informações fornecidas por eles. Durante o processo de análise dos observatórios, examinou-se cada aspecto descrevendo os diversos elementos presentes e, em particular, destacaram os tipos de informações disponíveis para acesso pela comunidade.

Após a identificação da tipologia documental, o processo exploratório foi iniciado com o objetivo de definir os processos de coleta, armazenamento e tratamento de informações. Essa etapa envolveu o uso de fontes bibliográficas e documentais, que serviram como base para a definição das etapas do modelo de coleta, armazenamento e tratamento de informações.

O processo de identificação das fontes perpassou por buscas no *Google Acadêmico*, com o objetivo de identificar textos que apoiassem a definição das técnicas mais adequadas para o tratamento dos dados previamente definidos na etapa anterior. A escolha do *Google Acadêmico* deu-se firmado nos estudos de Jacsó (2005), Harzing (2017), Zientek, Werner, Campuzano, Nimon (2018) entre outros, que defendem que essa fonte de informação científica é a mais completa, por indexar grande parte da documentação científica publicada no formato digital. Adicionalmente, no processo de identificação das ferramentas, realizou-se buscas, em especial no *Google*, analisando a aderência das ferramentas. Posteriormente, analisou-se as ferramentas de forma individual, buscando extrair mais detalhes para verificar como cada ferramenta poderia ser utilizada.

Por fim, a partir dos elementos de análise dos observatórios, junto com a identificação das técnicas e ferramentas necessárias para o tratamento informacional, definiu-se o modelo de informação que será utilizado para a definição do modelo mais ampliado de observatórios governamentais.

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

Destaca-se que neste trabalho, utilizou-se uma metodologia exploratória, analisando a literatura existente, apoiada com as análises descritas anteriormente. Desta forma, a pesquisa exploratória, junto com o contexto aplicado, levou ao desenvolvimento do modelo proposto.

### **3 RESULTADOS: COLETA, ARMAZENAMENTO E TRATAMENTO DE INFORMAÇÕES EM OBSERVATÓRIOS**

Para compreender os aspectos informacionais dos observatórios governamentais, foram identificados 675 *sites* de observatórios. No entanto, constatou-se que 203 desses *sites* não atendiam aos critérios mínimos necessários para serem considerados observatórios. Um exemplo disso é que muitos deles eram apenas *sites* de notícias. Portanto, foram considerados 472 *sites* para o levantamento da tipologia documental. O resultado consolidado dessa análise (ver Quadro 1) revela que a maioria dos documentos encontrados nos observatórios governamentais é textual e está no formato *Portable Document Format* (PDF), o que pode ser atribuído ao fato de o governo produzir a maior parte dos documentos nesse formato.

**Quadro 1 – Tipologia Documental encontrada nos observatórios analisados**

<b>Tipos</b>	<b>Características</b>
Boletins	Documento em formato de texto ( <i>PDF</i> )
Panorama/Estatística/Indicadores/Painéis/Gráficos/Dashboards (números)	Planilhas e Dados Estruturados ( <i>CSV</i> <sup>1</sup> , <i>XLS</i> <sup>2</sup> )
Relatórios/Resumos/Anuário	Documento em formato de texto ( <i>PDF</i> )
Pesquisas/Projetos	Documento em formato de texto ( <i>PDF</i> )
Publicações	Documento em formato de texto ( <i>PDF</i> )
Dossiê/Legislação	Documento em formato de texto ( <i>PDF</i> )
Artigos	Documento em formato de texto ( <i>PDF</i> )
Notícias	Documento em formato de texto ( <i>PDF</i> )
Mapa/Atlas	Documento em formato de texto ( <i>PDF</i> ) ou Imagens ( <i>PNG</i> <sup>3</sup> , <i>JPG</i> <sup>4</sup> )
Notas: <sup>1</sup> Comma-Separated Values ( <i>CSV</i> ); <sup>2</sup> Abreviação para Planilha do Microsoft Excel. <i>XL</i> em inglês soa como o nome do programa Excel e o <i>S</i> significa planilha ( <i>spreadsheet</i> ); <sup>3</sup> Portable Network Graphic ( <i>PNG</i> ); <sup>4</sup> Abreviação da sigla Joint Photographics Experts Groups ( <i>JPEG</i> )	

Fonte: Elaborado pelos autores (2023)

Com base neste mapeamento, foi possível levantar e analisar técnicas, padrões, métodos e ferramentas que apoiem os processos internos padrão dos observatórios governamentais. Tendo como base um fluxo simplificado de coleta, armazenamento e tratamento, os tipos de informações entradas influenciam nas técnicas e tecnologias a serem utilizadas nestes sistemas de informação.

### 3.1 Coleta de Informação

No âmbito da coleta de informações, no que tange aos observatórios governamentais, destaca-se que tal processo pode acontecer em três contextos principais em que as informações podem estar representadas: coleta de dados estruturados, coleta de informações em ambientes *web* e coleta de informações em banco de dados. No que tange aos dados estruturados, destaca-se diversas técnicas para a sua coleta. Aponta-se que, os dados estruturados, segundo Eberendu (2016, p. 48, tradução nossa)<sup>2</sup> “[...] se referem aos dados que possuem formato definido e comprimento, fácil de armazenar e analisar com alto grau de organização”.

No que tange a este tipo de informação, há técnicas específicas que possibilitam a coleta de dados estruturados, que possui um papel essencial nos observatórios, fornecendo informações relevantes e atualizadas para análises e tomada de decisões.

Complementarmente, no contexto de observatórios é essencial a coleta de informações em ambientes *web*. Tal abordagem é amplamente utilizada para obter informações de fontes *online*, envolvendo a extração sistemática de informações de páginas *web*, incluindo informações como textos, imagens, tabelas e *links*. Adicionalmente, a discussão sobre a utilização de informações passa por dois aspectos, o primeiro que tange a necessidade de coletar informações por *Applications Programming Interfaces* (APIs), o que fornece informações mais estruturadas, mas que, em um segundo aspecto, é necessário também obter as informações que não estão disponíveis para serem acessados via APIs, surgindo assim a necessidade de se coletar tais informações.

Neste contexto, Khder (2021, p. 145, tradução nossa)<sup>3</sup> aponta que:

---

<sup>2</sup> Trecho original: *Structured data refers to data that has definite format and length, easy to store and analyze with high degree of organization.* (EBERENDU, 2016, p. 48).

<sup>3</sup> Trecho original: *APIs can be used across many sites to easily access much of this information, but these are at the discretion of the site owner, and should they choose not to make this information accessible through APIs,*

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

As APIs podem ser usadas em muitos *sites* para acessar facilmente muitas informações, mas isso fica a critério do proprietário do *site* e, se ele optar por não tornar essas informações acessíveis por meio de APIs, é fácil não fazê-lo. A raspagem da *web* ou o rastreamento da *web*, por outro lado, é muito mais rápido e eficaz e pode ser usado para coletar e compilar dados de milhares ou milhões de páginas para processamento e extração de informações.

Por fim, aponta-se a coleta de informações em Bancos de Dados, que podem ser do próprio observatório, de alguma base governamental com acesso ou de alguma organização parceira. Nessa abordagem, as consultas e os comandos são utilizados para recuperar as informações desejadas. As linguagens de consulta mais comuns para coleta de informações em bases de dados são *Structured Query Language (SQL)* para bancos de dados relacionais e consultas específicas para bancos de dados *NoSQL (Not Only SQL)*, como o *MongoDB®*<sup>4</sup>. Essa abordagem é ideal quando as informações estão armazenadas em bases de dados e é necessário realizar consultas complexas para obter as informações desejadas (BEAULIEU, 2019).

A partir destas reflexões, apresenta-se, no Quadro 2, uma sumarização de como os tipos de informações apresentados podem ser coletados, demonstrando os formatos, as técnicas e as ferramentas necessárias para tal.

**Quadro 2 – Tipos de informações e descrição das técnicas e ferramentas para coleta das informações**

<b>Tipo de informações</b>	<b>Formatos</b>	<b>Técnicas</b>	<b>Ferramentas</b>
Dados estruturados	Planilhas no formato <i>CSV</i> , <i>XLS</i> , <i>XLSX</i> <sup>1</sup> , <i>TXT</i> <sup>2</sup>	Coleta de planilhas de forma manual ou por meio de técnicas de <i>RPA</i> <sup>3</sup> para coleta automatizada	Processo de <i>RPA</i> utilizando <i>Selenium</i> <sup>5</sup> ou ferramentas de plataformas como <i>Microsoft</i> ou <i>AWS</i>
Informações em ambientes <i>web</i>	Páginas e planilhas em <i>HTML</i> <sup>4</sup>	Utilização de ferramentas de <i>Web Scraping</i>	<i>Beautiful Soup</i> <sup>6</sup> , <i>Selenium</i> e <i>Scrapy</i> <sup>7</sup>
	Bases públicas ou privadas disponíveis em <i>XML</i> <sup>6</sup> ou <i>JSON</i> <sup>5</sup>	Coleta por meio do uso de <i>APIs</i> para a coleta de informações	Construção de código ou ferramentas para coleta das informações de <i>API</i> de acordo com as suas especificações

---

*it's easy not to. Web scraping or web crawling on the other hand is far faster and more effective, and can be used to gather and compile data from across thousands, or over millions, of pages for processing and drawing information from.* (KHDER, 2021, p. 145).

<sup>4</sup> *MongoDB*. Disponível em: <https://www.mongodb.com/>.

<sup>5</sup> *Selenium*. Disponível em: <https://www.selenium.dev/>.

<sup>6</sup> *Beautiful Soup Documentation*. Disponível em: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.

<sup>7</sup> *Scrapy*. Disponível em: <https://scrapy.org/>.

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

Informações disponíveis em bancos de dados	Banco de Dados Relacionais	Utilização de <i>SQL</i> para coleta	Uso de ferramentas de apoio aos <i>SGBDs</i> , como <i>MYSQL Workbench</i> , que permite a consulta utilizando <i>SQL</i> .
	Banco de Dados <i>NoSQL</i>	Utilização de técnicas de coleta, como <i>JSON</i> ou outros	Uso de ferramentas gerenciadoras de acordo com o banco escolhido que permite a consulta aos dados.
Notas: <sup>1</sup> Abreviação para Planilha do Microsoft Excel. XL em inglês soa como o nome do programa Excel e o S significa planilha (spreadsheet). Depois de 2007, a Microsoft adotou o formato Extensible Markup Language (XML), por isso a letra X. <sup>2</sup> Abreviação em inglês de “Text” <sup>3</sup> Robotic Process Automation (RPA) <sup>4</sup> HyperText Markup Language (HTML) <sup>5</sup> JavaScript Object Notation (JSON) <sup>6</sup> Extensible Markup Language (XML)			

**Fonte:** Elaborado pelos autores (2023)

Além do processo de coleta das informações, é necessária uma análise do processo de armazenamento de tais informações, que será discutido na sequência.

### **3.2 Armazenamento**

As técnicas e ferramentas recomendadas para o tratamento e armazenamento de informações no contexto da modelagem de um sistema para observatório são essenciais para a compreensão de observatórios no contexto atual, que se vincula fortemente à utilização de informações para a tomada de decisões nos mais diversos cenários. Complementarmente, destaca-se que o processo de tratamento e armazenamento se vincula diretamente à qualidade da informação, pois esta é de extrema importância para garantir a confiabilidade e a eficiência das operações do observatório. Portanto, é essencial adotar abordagens adequadas que permitam o correto processamento, organização e armazenamento das informações coletadas.

Ao apresentar e comparar as diferentes tecnologias de armazenamento, Lóscio, Oliveira e Pontes (2011, p. 2) aponta que os sistemas relacionais:

[...] as principais características [...], destacamos: controle de concorrência, segurança, recuperação de falhas, gerenciamento dos mecanismos de armazenamento de dados e controle das restrições de integridade do BD. Outra importante função de um SGBD é o gerenciamento de transações.

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

Já os sistemas *NoSQL* tem o: “[...] objetivo de atender aos requisitos de gerenciamento de grandes volumes de dados, semi-estruturados ou não estruturados, que necessitam de alta disponibilidade e escalabilidade” (LÓSCIO; OLIVEIRA; PONTES, 2011, p. 1).

Neste contexto, apresenta-se a seguir, o Quadro 3 com as características que podem ser utilizadas no processo de armazenamento.

**Quadro 3 – Tipos de armazenamento para as informações coletadas**

<b>Tipo de Armazenamento</b>	<b>Característica</b>	<b>Ferramentas</b>
Banco de Dados Relacionais	Armazenamento de informações estruturadas, oferecendo recursos de consulta e integridade referencial.	<i>MySQL, PostgreSQL</i> ou <i>Oracle</i>
Banco de Dados <i>NoSQL</i>	Armazenamento de informações não estruturadas ou semi-estruturadas, oferecendo flexibilidade no armazenamento de diferentes tipos de dados, como documentos, gráficos e dados em formato de chave-valor.	<i>MongoDB, Apache Cassandra</i> <sup>8</sup> e <i>Redis</i> <sup>9</sup>
<i>Data Warehouse</i>	“[...] uma estrutura de dados otimizada para distribuição, armazenamento em massa e processamento de consultas complexas” (SAHAMA; CROLL, 2007, p. 228, tradução nossa) <sup>10</sup> .	<i>Azure Synapse Analytics, Amazon Redshift, Oracle ADW</i> e <i>Google Big Query</i> .
<i>Data Lake</i>	“Um data lake é uma coleção massiva de conjuntos de dados que: (1) podem ser hospedados em diferentes sistemas de armazenamento; (2) podem variar em seus formatos; (3) podem não ser acompanhados por quaisquer metadados úteis ou podem usar formatos diferentes para descrever seus metadados; e (4) pode mudar autonomamente ao longo do tempo.” (NARGESIAN; ZHU; MILLER; PU; AROCENA, 2019, p. 1986, tradução nossa) <sup>11</sup> .	<i>AWS, Azure, Google Cloud</i> .

**Fonte:** Elaborado pelos autores (2023)

Contando com o processo de coleta e armazenamento definido, é necessário avançar no processo de tratamento das informações, que se mostra essencial para a extração de valor das informações.

### **3.3 Tratamento**

O processo de tratamento das informações contempla diversas atividades, técnicas e ferramentas, como os processos de Normalização, Limpeza, Transformação e

<sup>8</sup> *Apache Cassandra*. Disponível em: [https://cassandra.apache.org/\\_/index.html](https://cassandra.apache.org/_/index.html).

<sup>9</sup> *Redis*. Disponível em: <https://redis.io/>.

<sup>10</sup> Trecho original: [...] *a data structure that is optimized for distribution, mass storage and complex query processing*. (SAHAMA; CROLL, 2007, p. 228).

<sup>11</sup> Trecho original: *A data lake is a massive collection of datasets that: (1) may be hosted in different storage systems; (2) may vary in their formats; (3) may not be accompanied by any useful metadata or may use different formats to describe their metadata; and (4) may change autonomously over time*. (NARGESIAN; ZHU; MILLER; PU; AROCENA, 2019, p. 1986).

Enriquecimento. Essa etapa é importante para preparar as informações, que são oriundas de diversas fontes, para geração de novas informações ou para a oferta de serviços.

A primeira técnica apresentada é a normalização de dados, que trata de um método essencial para garantir a consistência e a uniformidade dos dados armazenados nos mais diversos contextos. Por meio deste processo, é possível evitar redundâncias e inconsistências, facilitando a integração e análise dos dados. Recomenda-se a utilização de padrões de normalização estabelecidos, como o modelo de normalização de banco de dados relacional.

Complementarmente, aponta-se que normalização de dados é definido como o processo “[...] onde os dados são dimensionados para uniformidade. A normalização de dados é necessária para estudar as melhores características dos dados.” (SREE; BINDU, 2018, p. 209, tradução nossa)<sup>12</sup>.

Junto ao processo de normalização, é essencial a realização da limpeza de dados, sendo um processo necessário para identificar e corrigir erros, omissões e inconsistências nos dados coletados. Isso inclui a remoção de valores nulos, a padronização de formatos e a detecção de *outliers*.

A limpeza de dados é definida como:

[...] uma abordagem inicial em que os conjuntos de dados são limpos para identificar quaisquer dados ausentes, remover os dados ruidosos e preparar os dados para análise. A limpeza de dados é necessária para resolver o problema de qualidade de dados. O problema de qualidade de dados é onde a análise pode dar errado em dados confusos. (SREE; BINDU, 2018, p. 209, tradução nossa)<sup>13</sup>.

Após o processo de normalização e limpeza dos dados, a próxima etapa trata da transformação e do enriquecimento de dados, que busca aplicar técnicas que permitem a conversão de dados brutos em formatos mais adequados e a incorporação de informações adicionais para enriquecer o conjunto de dados. Isso pode incluir a agregação de dados, a aplicação de regras de negócio e a incorporação de dados provenientes de outras fontes confiáveis.

---

<sup>12</sup> Trecho original: [...] *where the data is scaled to uniformity. Data normalization is needed to study the best features of the data.* (SREE; BINDU, 2018, p. 209).

<sup>13</sup> Trecho original: [...] *an initial approach where the data sets are cleaned to identify any missing data, remove the noisy data and get the data ready for analyzing. Data cleaning is needed to address the data quality problem.* (SREE; BINDU, 2018, p. 209).

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

Apresenta-se no Quadro 4, a descrição das técnicas e das ferramentas necessárias para a realização das etapas de normalização, limpeza e transformação e enriquecimento.

**Quadro 4 – Tipos de Dados e descrição das técnicas e ferramentas para coleta dos dados**

<b>Etapa</b>	<b>Técnica</b>	<b>Ferramenta</b>
Normalização	Em especial, as ferramentas específicas de gerenciamento e modelagem de dados, também existem bibliotecas e <i>frameworks</i> de programação que oferecem funcionalidades para a normalização de dados. Essas ferramentas permitem a implementação de algoritmos e rotinas personalizadas para realizar a normalização dos dados, levando em consideração as regras de negócio específicas do observatório.	<i>OpenRefine</i> <sup>14</sup> , <i>Alteryx Designer Cloud</i> <sup>15</sup> e <i>Microsoft SQL Server Integration Services (SSIS)</i> <sup>16</sup>
Limpeza	O processo de limpeza de dados exige ferramentas capazes de apoiar a exclusão de adequação da base de dados. Os <i>softwares</i> de limpeza e transformação de dados oferecem recursos avançados para identificar e corrigir erros, inconsistências, valores ausentes e duplicatas nos conjuntos de dados. Eles permitem a aplicação de regras e algoritmos para padronização, normalização e validação dos dados, garantindo sua conformidade com os requisitos estabelecidos.	<i>OpenRefine</i> , <i>Microsoft Power BI</i> <sup>17</sup> e o <i>Tableau</i> <sup>18</sup>
Transformação e Enriquecimento	Essas ferramentas desempenham um papel fundamental na preparação e no aprimoramento dos dados coletados, permitindo que sejam convertidos em formatos mais adequados e enriquecidos com informações adicionais.	<i>OpenRefine</i> , <i>Talend Data Integration</i> , <i>RapidMiner</i> , Bibliotecas <i>Pandas</i> <sup>19</sup> e <i>NumPy</i> <sup>20</sup> (Linguagem <i>Python</i> ):

**Fonte:** Elaborado pelos autores (2023)

Partindo dessa definição das principais etapas de coleta, armazenamento e tratamento dos dados, definiu-se um modelo que será apresentado na próxima seção.

#### **4 PROPOSTA DE MODELO DE INFORMAÇÃO PARA OBSERVATÓRIOS GOVERNAMENTAIS**

Todas as reflexões e discussões acerca do processo de coleta, armazenamento e tratamentos das informações, permite a definição de um modelo para tais processos no

<sup>14</sup> *OpenRefine*. Disponível em: <https://openrefine.org/>.

<sup>15</sup> *Alteryx Designer Cloud*. Disponível em: <https://www.alteryx.com/products/designer-cloud>.

<sup>16</sup> *SQL Server Integration Services*. Disponível em: <https://learn.microsoft.com/en-us/sql/integration-services/sql-server-integration-services>.

<sup>17</sup> *Microsoft Power BI*. Disponível em: <https://powerbi.microsoft.com/pt-br/>.

<sup>18</sup> *Tableau*. Disponível em: <https://www.tableau.com/pt-br/why-tableau/what-is-tableau>.

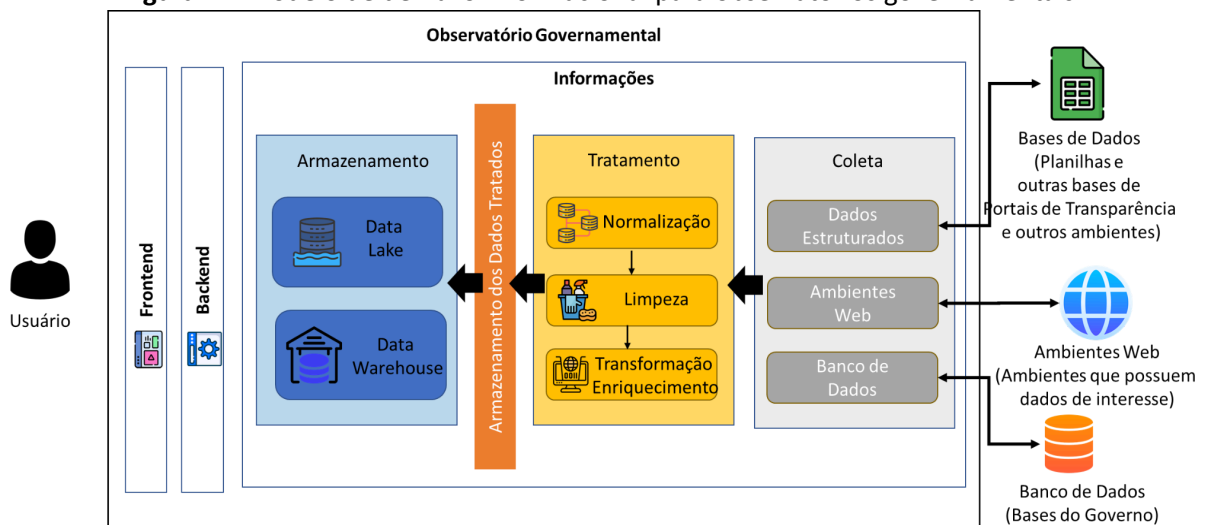
<sup>19</sup> *Pandas*. Disponível em: <https://pandas.pydata.org/>.

<sup>20</sup> *NumPy*. Disponível em: <https://numpy.org/>.

contexto dos observatórios. Tal modelo poderá ser utilizado como um elemento adicional na proposição de um modelo informacional de observatório governamental que poderá apoiar qualquer tipo de organização na montagem de seu observatório.

Ademais, a proposição deste modelo, insere a questão da utilização das novas ferramentas de tratamento das informações no contexto dos observatórios. Assim, apresenta-se, na Figura 1, o modelo de fluxo informacional do observatório governamental.

Figura 1 – Modelo de de fluxo informacional para observatórios governamentais



Fonte: Elaborado pelos autores (2023)

A Figura 1 ilustra a forma como o processo de informações no contexto dos observatórios é abordado, proporcionando uma visão abrangente de um sistema de informação voltado para observatórios governamentais. Destaca-se, que não objetivou-se neste trabalho entrar no contexto nos demais aspectos do sistema, como o *frontend* e o *backend* do sistema, buscando ter como enfoque a camada de informações.

No que tange à coleta, os observatórios obtêm informações, em sua maioria, de fontes governamentais, ou de informações publicadas em ambientes *web*. Assim, pode-se obter planilhas e outras bases de ambientes como portais de transparência, ou ambientes que as informações são compiladas. Já nos ambientes *web*, em processo de *web scraping*, coleta-se informações de ambientes digitais que apresentam fontes valiosas para o conjunto tratado pelo observatório. Por fim, pode-se realizar a integração e obtenção de informações de banco de dados governamentais.

Já no processo de tratamento, realiza-se, após a coleta, os procedimentos de normalização, limpeza e enriquecimento. Durante esse processo, é essencial que se considere o contexto e o escopo do observatório. Como as informações são provenientes de várias fontes, faz-se necessário esse tratamento para armazenamento, de forma a propor serviços consolidados.

Por fim, o Armazenamento, é necessário a criação de ambientes como *Data Lakes* e *Data Warehouses*, para que as informações obtidas, provenientes das mais diversas fontes informacionais possam ser relacionados e tratados como informações que se vinculam, e que possam ser utilizados em conjunto para apoiar o processo de disseminação das informações no futuro.

## **5 CONSIDERAÇÕES FINAIS**

O processo de definição de um modelo de informações para observatórios engloba uma abordagem mais abrangente, com o objetivo de estabelecer um modelo conceitual para o processo informacional em observatórios governamentais. Essa abordagem considera tanto os aspectos técnicos de um sistema de informação quanto os elementos estratégicos inerentes à visão de um observatório.

Dessa forma, ao longo do processo de construção do modelo, foi adotada uma perspectiva voltada para oferecer suporte ao estabelecimento de como um observatório pode coletar, armazenar e processar as informações obtidas. Essa abordagem contribui de forma significativa para a visão e o papel que tais observatórios podem desempenhar no futuro.

Portanto, este trabalho contribui para apoiar os observatórios no intuito de trabalhar com as informações disponíveis, se tornando um agregador das informações disponíveis nos mais diferentes ambientes informacionais digitais. O modelo não busca ser taxativo, considerando os aspectos informacionais das fontes de informação e das ferramentas e métodos para coleta, tratamento, armazenamento de informações, no entanto, a sistematização em uma proposta de modelo pode exemplificar e direcionar as ações para construção de observatórios no âmbito governamental.

A sistematização desse processo pode também contribuir nas estratégias para gestão de informações, especialmente ao considerar as possibilidades de ter informações analíticas e indicadores que podem contribuir para tomada de decisão. Nesse sentido, é importante

ressaltar que um fluxo de trabalho bem definido e estruturado, no qual a coleta, o armazenamento e o tratamento das informações sejam conduzidos de forma eficiente, são fundamentais para garantir a qualidade e a utilidade dos dados resultantes. Além disso, ao adotar abordagens inovadoras e tecnologias adequadas, o processo de gestão de informações pode ser otimizado nos observatórios, permitindo uma análise mais precisa e uma melhor compreensão do cenário em questão. Portanto, os observatórios podem apoiar a gestão de políticas públicas, uma vez que o manejo otimizado das informações nelas trabalhadas contribui para o entendimento da temática pelos gestores e pela comunidade.

O próximo passo, dentro da perspectiva da pesquisa, consiste em analisar e discutir minuciosamente a forma como as informações são estruturadas no âmbito governamental. Essa análise aprofundada permitirá compreender os mecanismos e padrões existentes, viabilizando assim a utilização das técnicas apresentadas como parte essencial para a implementação do modelo em questão.

Os estudos sobre observatórios, no entanto, têm se revelado como um campo fértil para pesquisas interdisciplinares, uma vez que se tornou necessário em face da proliferação desses sistemas de informação. Tal campo de estudo oferece uma oportunidade para aprofundar o conhecimento sobre a natureza, funcionamento e impacto dos observatórios, reunindo perspectivas e abordagens provenientes de diferentes disciplinas acadêmicas.

## **REFERÊNCIAS**

BARDIN, Laurence. **Análise de conteúdo**. Coimbra: Almedina, 2011.

BEAULIEU, Alan. **Aprendendo SQL: Dominando os Fundamentos de SQL**. Novatec, 2019.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Brasília, DF: Presidência da República 2011. Disponível em:  
[https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm). Acesso em: 3 jul. 2023.

EBERENDU, Adanma Cecilia. Unstructured Data: an overview of the data of Big Data. **International Journal of Computer Trends and Technology (IJCTT)**, [s. l.], v. 38, n. 1, p. 46-50, Aug. 2016. DOI: 10.14445/22312803/IJCTT-V38P109. Disponível em:  
<https://ijcttjournal.org/archives/ijctt-v38p109>. Acesso em: 3 jul. 2023.

HARZING, Anne-Wil. **Google Scholar is a serious alternative to Web of Science**. 28 Feb. 2017. Harzing.com [Site]. Disponível em:

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

<https://harzing.com/blog/2017/02/google-scholar-is-a-serious-alternative-to-web-of-scienc>. Acesso em: 1 jul. 2022.

JACSÓ, Péter. Google Scholar: the pros and the cons. **Online Information Review**, [s. l.], v. 29, n. 2, p. 208–214, Apr. 2005. DOI: <https://doi.org/10.1108/14684520510598066>. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/14684520510598066/full/html>. Acesso em: 1 jul. 2022.

KHDER, Moaiad Ahmad. Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. **International Journal of Advances in Soft Computing & Its Applications**, Jordan, v. 13, n. 3, p. 144-168, Nov. 2021. DOI: 10.15849/IJASCA.211128.11. Disponível em: [Web scraping or web crawling: Sta... preview & related info | Mendeley](#). Acesso em: 2 jul. 2023.

LÓSCIO, Bernadette Farias; OLIVEIRA, Hélio Rodrigues de; PONTES, Jonas César de Sousa. NoSQL no desenvolvimento de aplicações Web colaborativas. In: SIMPÓSIO BRASILEIRO DE SISTEMAS COLABORATIVOS, 8., 2011, Paraty. **Mini curso** [...]. Paraty: SBSC, 2011. Material de apoio. Disponível em: [https://www.addlabs.uff.br/sbsc\\_site/download.html](https://www.addlabs.uff.br/sbsc_site/download.html). Acesso em: 2 jul. 2023.

MACÊDO, Diego José; MARICATO, João de Melo; SHINTAKU, Milton. Observatórios: reflexões sobre os conceitos e aplicações em Ciência, Tecnologia e Inovação e relações com a Ciência da Informação. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 17, p. 1-21, 2021.

MORA, José Ferrater. **Dicionário de filosofia**. São Paulo, Ed. Loyola, 2001. t. 3.

NARGESIAN, Fatemeh; ZHU, Erkang; MILLER, Renée J.; PU, Ken Q.; AROCENA, Patricia C. Data lake management: challenges and opportunities. **Proceedings of the VLDB Endowment**, [s. l.], v. 12, n. 12, p. 1986-1989, 2019. DOI: <https://doi.org/10.14778/3352063.3352116>. Disponível em: <https://dl.acm.org/doi/pdf/10.14778/3352063.3352116>. Acesso em: 2 jul. 2023.

SAHAMA, Tony R.; CROLL, Peter R. A data warehouse architecture for clinical data warehousing. In: AUSTRALASIAN WORKSHOP ON HEALTH KNOWLEDGE MANAGEMENT AND DISCOVERY (HKMD 2007), 2007, Ballarat. **Proceedings** [...] Ballarat: Australian Computer Society, 2007. p. 227-232. Disponível em: <https://www.cin.ufpe.br/~aabp/arquivos/14246.pdf>. Acesso em: 3 jul. 2023.

SOARES, Lilian Campos; FERNEDA, Edilson; PRADO, Hércules Antonio do. Observatórios: um levantamento do estado do conhecimento. **Brazilian Journal of Information Science: research trends**, Marília, v. 12, n. 3, p. 86-110, 2018. DOI: 10.36311/1981-1640.2018.v12n3.08.p86. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/7958>. Acesso em: 9 jul. 2023

SREE, K. Dhana; BINDU, C. Shoba. Data analytics: Why data normalization. **International Journal of Engineering and Technology (IAE)**, Amman, v. 7, n. 4.6, p. 209-213, 2018. DOI:

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB  
Aracaju-SE – 06 a 10 de novembro de 2023**

<https://doi.org/10.14419/ijet.v7i4.6.20464>. Disponível em:  
<https://www.sciencepubco.com/index.php/ijet/article/view/20464>. Acesso em: 3 jun. 2023.

VIEIRA, Jeferson Kenedy Morais; BARBOSA, Jessyca Lorena Pereira; FARIAS JÚNIOR, Ivaldir H. de; MOURA, Hermano Perrelli de. Um estudo sobre Observatórios através de um Mapeamento Sistemático da Literatura. **Journal of Information Systems and Technology Management (JISTEM USP)**, São Paulo, v. 19, mar. 2022. ISSN 1807-1775. DOI: <http://dx.doi.org/10.4301/S1807-1775202219003>. Disponível em: <https://www.tecsi.org/jistem/index.php/jistem/article/view/3185>. Acesso em: 2 jul. 2023.

ZIENTEK, Linda R.; WERNER, Jon M.; CAMPUZANO, Mariela V.; NIMON, Kim. The Use of Google Scholar for Research and Research Dissemination. **New Horizons in Adult Education and Human Resource Development**, [s. l.], v. 30, n. 1, p. 39-46, Feb. 2018. DOI: <https://doi.org/10.1002/nha3.20209>. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/nha3.20209#pane-pcw-references>. Acesso em: 3 jul. 2023.