Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK

INTERNSHIP REPORT

IMPROVING DATA CONSISTENCY IN OFFICIAL STATISTICS: APPLICATION IN THE CENTRAL BANK OF PORTUGAL

ANA BEATRIZ RODRIGUES CASTOR

FEBRUARY - 2025

Lisbon School of Economics & Management
Universidade de Lisboa

MASTER
DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK
INTERNSHIP REPORT

IMPROVING DATA CONSISTENCY IN OFFICIAL STATISTICS: APPLICATION IN THE CENTRAL BANK OF PORTUGAL

ANA BEATRIZ RODRIGUES CASTOR

SUPERVISION:
SUPERVISOR: PROF. CARLOS J. COSTA
CO-SUPERVISOR: LÍGIA NUNES

JURI MEMBERS:
MÁRIO CALDEIRA
MARTINHA PITEIRA

FEBRUARY - 2025

GLOSSARY

API – Application Programming Interface.

BdP – Banco de Portugal.

DSR – Design Science Research.

ECB – European Central Bank.

ESS – European Statistical System.

GDP – Gross Domestic Product.

JSON – JavaScript Object Notation.

OGD – Open Government Data.

QAF – Quality Assurance Framework.

REST – Representational State Transfer.

RPC – Remote Procedure Call.

SDMX – Statistical Data and Metadata Exchange.

SOAP – Simple Object Access Protocol.

UN – United Nations.

UNECE – United Nations Economic Commission for Europe.

URL – Uniform Resource Locator.

XML – Extensible Markup Language.

ABSTRACT

Central banks play a crucial role in ensuring the reliability and accessibility of official statistics, which are fundamental for economic analysis, policymaking, and public transparency. However, as data volumes expand and statistical dissemination becomes increasingly digitalized, maintaining consistency across multiple sources remains a challenge. This study presents the development of an automated tool designed to validate the consistency of statistical data republished by Banco de Portugal (BdP). Leveraging APIs from primary sources, including the European Central Bank (ECB) and Eurostat, the tool systematically compares datasets to detect inconsistencies and streamline quality control processes. By automating statistical verification, the solution enhances efficiency, reduces reliance on manual checks, and strengthens data reliability.

The study adopts a Design Science Research (DSR) methodology, integrating theoretical foundations with practical implementation. The developed tool successfully identified discrepancies in key statistical series, notably in effective exchange rate indices, where methodological revisions influenced data values. Despite challenges related to system integration and adaptability to structural changes in datasets, the tool demonstrated significant improvements in statistical consistency monitoring.

Future research avenues include incorporating machine learning techniques for anomaly detection, broadening the tool's applicability across various economic indicators, and aligning its framework with evolving international data standards, such as SDMX. This research underscores the transformative impact of automation in statistical quality assurance, reinforcing transparency, accuracy, and trust in official economic data.

KEYWORDS: Data consistency, statistical automation, APIs, Banco de Portugal, data integration, quality control, official statistics, SDMX.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

To begin with, I would like to express my sincere gratitude to the people who made this work possible. I extend my deepest appreciation to Professor Carlos J. Costa for his invaluable guidance throughout the development of this project at ISEG.

I am also immensely grateful to Lígia and Maria for welcoming me into the team and providing unwavering support throughout my internship and the subsequent writing of this report. Their encouragement and assistance made this experience not only enriching but also truly enjoyable.

Furthermore, I would like to thank my university for giving me the opportunity to gain knowledge in a field I am passionate about, but even more so for allowing me to meet two incredible friends - Teresa and Marta. I would never have made it without you, and this journey would not have been as beautiful without your presence.

Lastly, and most importantly, I would like to express my deepest gratitude to my family, especially my mother and younger brother, also known as my greatest supporters. I owe them not only my academic achievements but also their unconditional personal support.

1. INTRODUCTION

1.1 Contextualization

Central banks have a fundamental responsibility in disseminating official information to society, a function that has evolved significantly over time. Nowadays, these institutions aim to be as transparent as possible with the public (Blinder, 2001). Moreover, communication channels have expanded beyond television and newspapers. There is now a strong presence on social media, and since 2010, central banks have been using their websites to share relevant data to achieve a wider audience (Blinder et al., 2022).

As the digital landscape evolves, there is an increasing need for high-quality, reliable statistics to support societal and economic decision-making. Additionally, shifts in user behaviors and expectations have led the European Statistical System (ESS) to systematically reassess and adapt their approaches to collect, produce, and communicate statistical data (Laaboudi et al., 2024).

Currently, user demand and cost-effectiveness are key factors driving the transition toward web publishing. Consequently, most producers of official statistics have now embraced the internet as their primary channel for publishing outputs. (Smith & Rogers, 2008) However, as more data becomes publicly accessible online, finding a balance between preserving confidentiality and ensuring published statistics' quality, timeliness, and integrity is imperative in building trust with the public (SPE, 2024).

A study conducted by UNECE on official statistics highlights the importance of high-quality data as a crucial tool for informed decision-making. These statistics are safeguarded by solid legal frameworks that prioritize independence and objectivity. Beyond guaranteeing impartiality, this data is essential for shaping public policies, socioeconomic analysis, and crisis management (United Nations, 2019).

Given this contextualization and to keep pace with the dynamism of the information world, Banco de Portugal launched, in 2019, the new BP*stat* – a portal to publish official statistics. Besides data, users can access updated news and content aspiring to simplify economic and statistical concepts. This platform serves as a key resource for public access to a wide range of economic and financial data, reflecting the institution’s

commitment to data accessibility. All data is free without the need for registration or subscription (Banco de Portugal, 2019).

To guarantee the accuracy and consistency of its statistics, BdP adheres to international standards and methodologies. This commitment strengthens its credibility and reinforces cooperation with international entities, fostering the exchange of best practices and alignment with global statistical frameworks.

As data volumes continue to expand, Banco de Portugal seeks to optimize its processes further. In this context, automation becomes essential to ensure efficiency, maintain data consistency, and keep statistical information reliably updated, reinforcing the institution's commitment to accuracy and transparency.

1.2 Objectives

This work aims to apply the knowledge acquired during the master's program to propose a solution for data integration that ensures the consistency of data republished by Banco de Portugal.

To achieve this, the project focuses on three key objectives:

- Develop a data consistency verification tool to automate the comparison of statistical data between BP*stat* and its sources, namely the European Central Bank and Eurostat;
- Integrate the tool into the existing processes, ensuring seamless operation and compatibility with current workflows;
- Assess the tool´s performance and continuously refine it, based on real-world application and feedback.

This solution will contribute to Banco de Portugal's commitment to data quality, transparency, and accessibility.

1.3 Structure of the Master's Final Work

This academic work begins with a literature review to establish theoretical foundations relevant to the study. Following this, the chosen methodology is outlined. The empirical work then explores the business context and data sources, leading to the development and implementation of a structured approach. The results section assesses

its integration within existing workflows and evaluates its effectiveness. Finally, the work concludes with insights into future research opportunities and potential limitations.

During this project, AI tools such as ChatGPT and Grammarly were used to enhance the clarity and overall readability of the text.

2. Literature Review

2.1 API – Application Programming Interface

The need for data integration has been a longstanding challenge for organizations (Groth et al., 2014). The rapid advancement of digital technologies and the influx of diverse and complex data sources have exponentially increased the volume of information available (Alma'aitah et al., 2024). However, most of this is stored and processed in various separate systems and databases, often characterized by a lack of uniformity in data definition and how they are described - the metadata (Kellerman & David, 2021).

Application Programming Interfaces (APIs) have emerged as a powerful technology addressing challenges such as data integration, enabling seamless communication between different pieces of software, independently of their location or communication protocol, making them an essential tool for data integration in complex and distributed systems. They first appeared in the 1960s and have been evolving significantly over time. Initially, focused on enabling communication within local systems, APIs have expanded to facilitate over networks (Pedro, 2024).

APIs can be categorized based on their scope of use, which determines how accessible they are and who can use them.

Public APIs, also called open or external APIs, are freely accessible to developers or organizations outside the business. For this project, these are especially relevant, as they allow all the public to access statistical data, being key for improving transparency and enabling the free exchange of data, which is essential for promoting trust in official statistics. Partner APIs, on the other hand, are restricted to specifically selected and authorized outside developers or API consumers. Hence, it can be used to facilitate

business-to-business activities since partners have clear rights and licenses. Therefore, this type generally incorporates stronger authentication, authorization, and security mechanisms, than the public ones. Internal APIs are confined within enterprises, to connect systems and business data. Lastly, the Composite APIs usually combine two or more APIs to craft a sequence of related or interdependent operations (Bigelow, 2024).

The evolution of APIs and the creation of several architectures over the decades show how these technologies have adapted to meet new challenges and business needs (Lamothe et al., 2021).

The set of rules and common practices that define how an API interacts with other components is called API architecture. It provides a framework for designing, developing, and delivering backend services. It dictates how data is transferred between client and server, as well as how services are exposed (Biehl, 2015).

Figure 1 highlights key milestones in this evolution, showcasing the transition from rigid and complex architectures to more developer-friendly solutions that drive innovation in data sharing and system integration.

FIGURE 1 - Evolution of APIs over time from (AltexSoft, 2023)

There are multiple API architecture types, each designed to fulfill different needs and meet specific functionalities or performance requirements (Catchpoint, n.d).

Simple Object Access Protocol (SOAP**)** is a protocol intended for the exchange of structured information in web services. It relies on XML for its message format and provides extensibility through additional standards. It is especially recognized for its strong support in security, transaction compliance, and reliable messaging. On the other

hand, these functionalities increase complexity, making it less flexible than other modern architectures (World Wide Web Consortium, 2007).

Representational State Transfer (REST), introduced by Roy Fielding in his doctoral dissertation, outlines the principles for the operation of distributed hypermedia systems. It is based on stateless client-server interaction, which means that each request must contain all the information and attributes, for the server to understand and process it. This approach emphasizes simplicity, scalability, and performance, making RESTful APIs adopted across various domains (Fielding, 2000).

Remote Procedure Call (RPC**)** is a widely used technique in distributed systems that allows a program to execute a procedure on a remote machine as if it were a local call. RPC relies on clients and server stubs, which act as intermediaries to achieve this transparency (Alma'aitah et al., 2024).

GraphQL is a query language for APIs that provides a complete and understandable description of available data. It empowers the clients to make specific requests, avoiding unnecessary information. This characteristic simplifies API evolution over time and enables powerful developer tools (GraphQL Foundation, n.d).

Communication between the client and the server is defined as a response involving key elements, such as **status codes** that indicate the state of the request; metadata **headers** that provide additional context, and **response bodies** that deliver the requested data. Common response formats such as JSON and XML are widely used due to their flexibility (Chadere, 2024).

Hence, APIs provide an accessible way to extract and share data across organizations, considering that companies are digitally transforming faster than ever to stay competitive and meet growing customer demands. Given that, APIs facilitate digitization, connectivity, and innovation across products and services, serving as a fundamental enabler of these initiatives. In fact, 90% of executives consider APIs to be mission-critical to their businesses. By implementing API-driven strategies, organizations can foster growth and innovation (MuleSoft, n.d).

Additionally, a study conducted by Postman and RapidAPI in 2022, revealed that banks represent a significant share of API users, highlighting the increasing importance of APIs in financial fields, as shown in Figure 2.

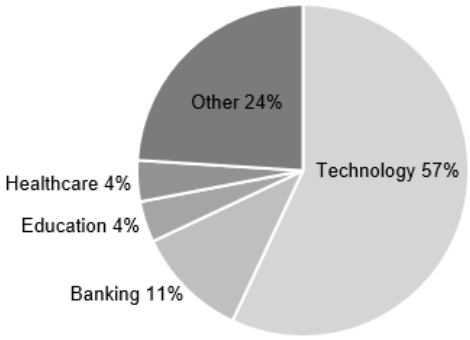

FIGURE 2 - Global API usage breakdown by top industries (Pedro, 2024)

In the context of official statistics, APIs enable the efficient extraction and sharing of data across organizations, enhancing the accessibility and transparency of statistical information.

Among the multiple existent response formats, for this project, it is relevant to refer to the JSON-stat format which is an extension from the traditional JSON projected to represent structured data facilitating data visualization and analysis. This format also includes structured metadata which explains the meaning of each dimension and how the variables connect. JSON-stat format is an ideal companion for the open data initiatives of National Statistical Offices. Designed with simplicity, it is suitable for all kinds of data disseminators (JSON-stat, 2024).

Selecting the appropriate architecture and response format empowers APIs to optimize data exchange, boost efficiency, and enable more cohesive integration across systems. For official statistics, these capabilities are crucial to ensuring improved access to data, and support for transparent, reliable dissemination of statistical information to a wide range of users.

The following table provides a comparative analysis of four major API architectural styles.

TABLE 1 – Main API architectural styles compared

	RPC	SOAP	REST	GRAPHQL
Organization Method	Local procedure calling	Enveloped message structure	Compliance with six architectural constraints	Schema and type system
Response Format	JSON, XML, Protobuf, Thrift, FlatBuffers	XML only	XML, JSON, HTML, plain text	JSON (customizable queries)
Learning curve	Easy	Difficult	Easy	Medium
Community	Large	Small	Large	Growing
Use cases	Command and action-oriented APIs	Payment gateways, identity management CRM solutions	Public APIs, simple resource-driven apps	Mobile APIs, complex systems

Source: AltexSoft, (2023)

2.2 Data Quality in Official Statistics

The need for a set of principles governing official statistics became apparent in the late 1980s, as Central European countries evolved from centrally planned economies to market-oriented democracies (United Nations, 2014).

In addition, the growing demand for statistics, alongside advancements in digital technologies, has posed challenges for statistical institutes in upholding their core values and principles outlined in quality frameworks (Sæbø & Holmberg, 2019).

To ensure reliability and impartiality in the production and dissemination of statistics, the United Nations (UN) established ten fundamental principles (United Nations, 2014).

In Europe, the European Statistics Code of Practice constitutes the cornerstone of the quality framework for the European Statistical System, expanding the UN´s ten principles into sixteen detailed guidelines tailored to the European context. Moreover, it is regularly updated to reflect the evolving statistical landscape (European Commission, 2018).

These sixteen principles are structured into three key dimensions: Institutional Environment, Statistical Processes, and Statistical Outputs, as illustrated in Table 2.

TABLE 2 - The 16 principles of the European Statistics Code of Practice

Principle	Dimension
Professional Independence	Institutional Environment
Coordination and Cooperation	
Mandate for Data Collection and Access to Data	
Adequacy of Resources	
Commitment to Quality	
Statistical Confidentiality and Data Protection	
Impartiality and Objectivity	
Sound Methodology	Statistical Processes
Appropriate Statistical Procedures	
Non-excessive Burden on Respondents	
Cost Effectiveness	
Relevance	Statistical Outputs
Accuracy and Reliability	
Timeliness and Punctuality	
Coherence and Comparability	
Accessibility and Clarity	

Source: European Commission

To ensure a consistent application of the previously mentioned principles, the European Statistics Code of Practice is supported by complementary guidelines, such as the Eurostat Quality Policy, which provides practical tools and methodologies. This policy is structured into four levels of quality assurance: the foundational principles and indicators; the methods and tools to ensure implementation; and standardized quality reporting mechanisms, which assess and monitor statistical outputs. These measures are designed not only for Eurostat but for the entire European Statistical System, ensuring coherence and adherence to high-quality standards across European statistics (Eurostat, n.d).

Additionally, to operationalize the principles outlined in the European Statistics Code of Practice, the Quality Assurance Framework (QAF) provides a detailed roadmap for aligning statistical processes with established quality standards (Eurostat, 2019).

Figure 3 visually represents the four levels of quality assurance to provide a clearer understanding of this multi-level structure and illustrate its relation to the European Statistics Code of Practice.

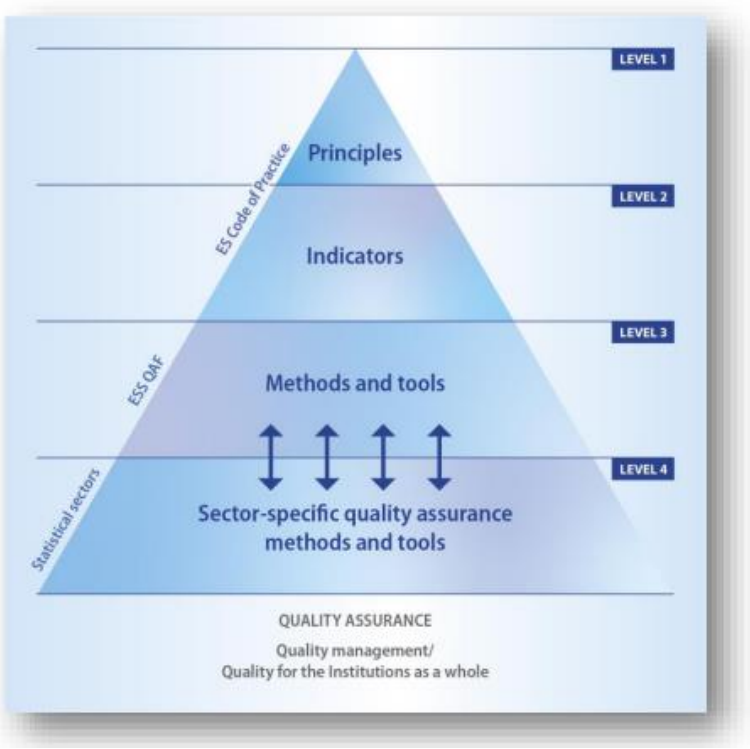

FIGURE 3 - Eurostat Quality Policy

In line with international quality standards, the Banco de Portugal ensures that its compiled and published statistics adhere to rigorous guidelines and best practices. Specifically, these statistics comply with the quality management standards outlined in the Commitment on European System of Central Banks Statistics and the European Statistical System Code of Practice (Banco de Portugal, n.d.).

2.3 SDMX: Statistical Data and Metadata Exchange

The lack of uniformity in data management constitutes a significant challenge to data integration, a fundamental aspect of statistical work. To address these issues, global standards were developed to promote consistency and interoperability (Staab1, 2019).

One of the most significant initiatives in this field is the Statistical Data and Metadata Exchange (SDMX), a global standard designed to harmonize statistical data and metadata exchange to improve efficiency, quality, and interoperability among international organizations and national statistical agencies. Established in 2001 by organizations such as the International Monetary Fund, the European Central Bank, Eurostat, and the United Nations, SDMX addresses both the growing complexity of managing statistical data in the digital age (SDMX, 2022) and the pressing need to harmonize statistical data exchange (Bank for International Settlements et al., 2003).

The goal is to leverage emerging web technologies to standardize and simplify the collection, compilation, and dissemination of statistical information (Bank for International Settlements et al., 2003). To achieve this, SDMX standards are implemented through technologies such as APIs, which facilitate interaction between applications and ensure interoperability between databases. These technologies play a crucial role in distributing, using, and reusing statistical data efficiently, ultimately reducing redundancy (Biao & Assogba, 2024).

SDMX is built on a systematic and modular framework known as the building block approach, which structures statistical data and its accompanying metadata into interoperable formats, enabling seamless integration, interpretation, and dissemination across institutions (Stahl & Staab, 2018). This framework is composed of several key components, including a multidimensional data cube, standardized vocabularies (content-oriented guidelines), a formal schema definition, and various data serialization formats (Biao & Assogba, 2024).

Additionally, central to the SDMX framework is the concept of systematic and self-explanatory keys, which are composed solely of metadata elements. This design ensures that each dataset is uniquely identified and consistently interpreted.

Figure 4 provides a schematic representation of the SDMX structure, summarizing the key characteristics mentioned previously, such as the building block approach, metadata structuring, and the formation of systematic and self-explanatory keys.

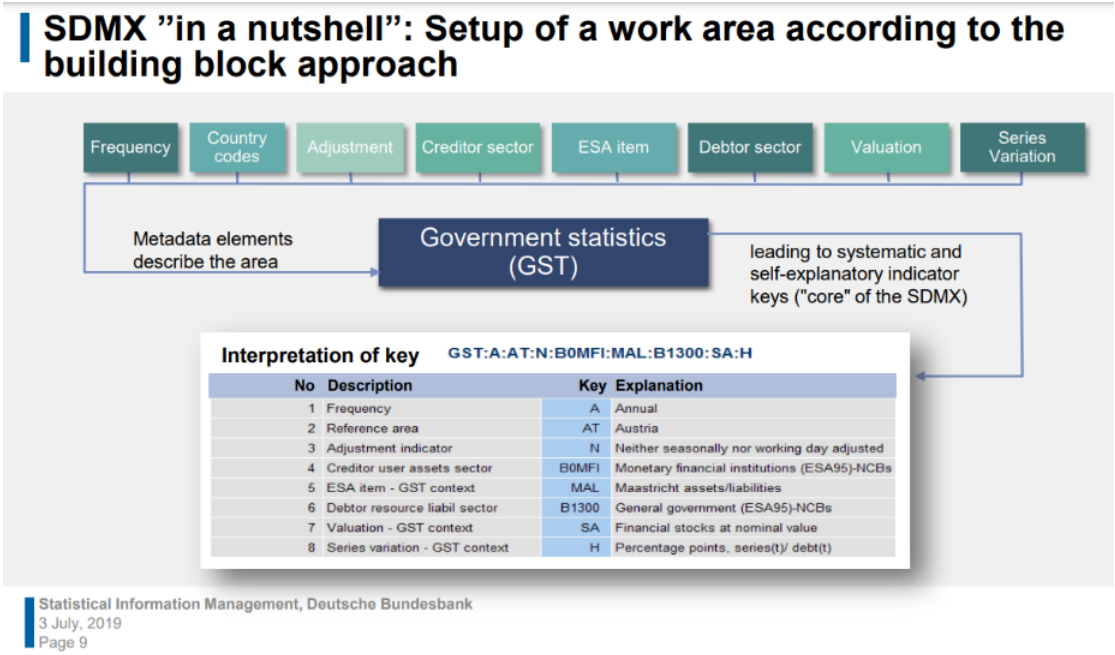

FIGURE 4 - SDMX Structure from Staab1, (2019)

The adoption of this standard brings significant advantages to official statistics, benefiting not only data analysts but also data providers (Stahl & Staab, 2018) by enhancing cooperation and consistency across multiple institutions. Furthermore, SDMX aligns closely with today´s concept of Open Government Data (OGD) for openness, transparency, and ease of access to government-owned data, an aspect that has become increasingly important (Sembiring & Uluwiyah, 2015). As data demands continue to grow, the relevance of SDMX is expected to increase, particularly in areas such as big data integration and real-time analytics (Stahl & Staab, 2018).

For national agencies such as the Banco de Portugal, it serves as a critical tool for aligning with international standards, automating reporting processes, and ensuring data quality. As the global demand for reliable statistics grows, SDMX is poised to remain at the forefront of innovation in statistical systems.

3. METHODOLOGICAL APPROACH

Design science refers to a research field of study that focuses on the research of artifacts, constructs, and other artificial concepts (Aparicio et al., 2023). It is inherently a problem-solving process, and it focuses on creating and evaluating innovative IT artifacts that enable organizations to address important information-related tasks (Hevner et al., 2004).

Given its practical nature, the Design Science Research (DSR) methodology was chosen as the foundation for this work, since the project involves the development of a technical solution for extracting, processing, and comparing data from different sources.

The main stages of DSR include problem identification and motivation, design, and development of the artifact (a solution that addresses the identified problem), evaluation to test and validate the artifact, and communication to share the work's results and contributions.

The following diagram outlines the key phases of the Design Science Research methodology applied in this project.

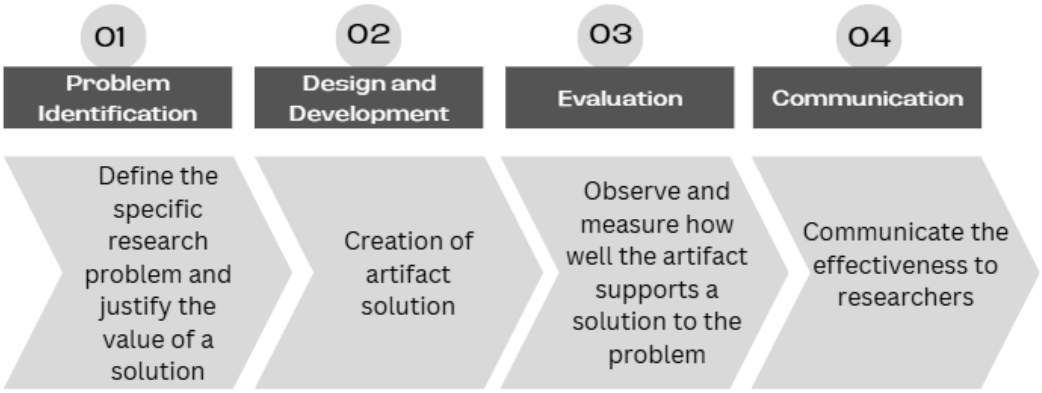

FIGURE 5 - Key phases of DSR methodology (Adapted from Peffers et al., 2006)

3.1 Problem Identification

Following the methodological framework outlined above, the first phase focuses on clearly defining the problem and its broader impact. In this case, the primary challenge motivating this project is the need for consistency and integration among datasets from multiple sources. Since the problem definition will be used to develop the artifact, it is

important to conceptually structure the problem to ensure that the solution captures its complexity (Peffers et al., 2006).

Given that, this challenge aligns directly with the broader responsibilities and objectives of the BdP, specifically with the mission of its Statistics Department, which is responsible for the production and dissemination of official statistics (Banco de Portugal, n.d.-c).

Moreover, this goal also intersects with the Banco de Portugal's Quality and Revision Policy, which underscores the importance of transparency, accuracy, and timeliness in statistical processes. The policy sets standards for consistent data revisions to ensure that statistical information remains reliable and aligned with national and international frameworks. By adhering to these standards, the Bank strengthens public trust and supports effective decision-making (Banco de Portugal, n.d.-b).

Additionally, this topic is crucial to uphold Banco de Portugal´s Strategic Guideline 4, which emphasizes fostering trust and understanding within society. The institution´s core values – integrity, transparency, and responsibility – further reinforce its vision to be a trusted and exemplary entity within the European Union (Banco de Portugal, 2021).

Beyond the institutional context, addressing this problem carries significant implications for society. As highlighted by Feijo and Valente (2005), national statistics services play a vital role in providing policymakers and decision-makers with the essential information needed for effective policy development and planning. Considering that statistics serve an important social and educational function by offering valuable insights into various aspects of socio-economic realities, it is the responsibility of national statistical services to disseminate accurate information, empowering communities to make informed decisions and fostering constructive discussions on national issues.

3.2 Design and Development

To address the identified problem, an automated artifact was developed and implemented to extract, process, and compare datasets from various sources. This solution ensures data consistency over time.

The system was built using a flexible programming framework, leveraging tools for data extraction, manipulation, and interaction with external sources. This approach ensures adaptability, allowing the artifact to integrate seamlessly with the existing infrastructure while remaining adjustable to future technological advancements.

To advance its functionality, the artifact incorporates mechanisms for real-time data extraction and updates, remaining aligned with the latest datasets.

Currently, the tool is operational and actively used by one team of the statistical department of the Bank. Its deployment has streamlined workflows and enhanced the team´s day-to-day efficiency. In addition, the development followed an iterative approach, prioritizing feedback from end-users to ensure that the artifact met the specific needs of the working team.

The evaluation phase will focus on assessing the artifact´s effectiveness in addressing inconsistencies in data related to Portugal and the Euro area.

3.3 Communication

Currently, the tool is utilized by the Statistics Department, yet its adaptability makes it an asset for other teams involved in data dissemination. Furthermore, considering collaboration among institutions, this methodology could serve as a model for other central banks facing similar data integration and consistency challenges. The solution aligns with international standards and best practices and could contribute to regional standardization initiatives.

Looking ahead, several opportunities for further development have been identified. From a technical perspective, the solution could be enhanced by incorporating more advanced data validation techniques, such as machine learning algorithms or artificial intelligence, to detect inconsistencies.

On a broader scale, future work could explore the adaptation of this solution to accommodate evolving international standards, such as updates to the SDMX

framework, ensuring its relevance in a changing regulatory environment. Collaborations with external partners, including other central banks or research institutions, could further expand the potential applications of this work, fostering innovation in data integration and quality assurance across sectors.

4. RESULTS

4.1 Current Workflow and Artifact Integration

As previously mentioned, the developed tool was integrated into the existing internal quality control process, specifically as the final step validation after publication. The evaluation of the process is based on operational effectiveness, ensuring that the tool consistently fulfills its purpose of accurately and efficiently comparing data with the original sources.

Currently, the process is designed to receive data directly from the original producers: European Central Bank and Eurostat. Once the information is processed, files with a specific format are generated and uploaded to BP*stat*. To ensure operational efficiency, routine updates focus solely on the last two years of data, as integrating the entire historical dataset would require excessive processing time. Updated data are republished on BP*stat* during previously scheduled update events and organized into data chunks by data domain.

The developed artifact has become a critical part of this workflow, validating published data in alignment with scheduled update events. After each chunk is updated, the tool is activated to validate the consistency between the entire historical dataset for that specific block of data in BP*stat* with its original sources. This targeted approach minimizes processing time by allowing the selection and validation of specific blocks.

Figure 6 represents the tool's interface upon initialization, allowing users to select the specific data block to be validated. This user-friendly and intuitive interface streamlines the validation process, enabling targeted analysis while optimizing processing efficiency.

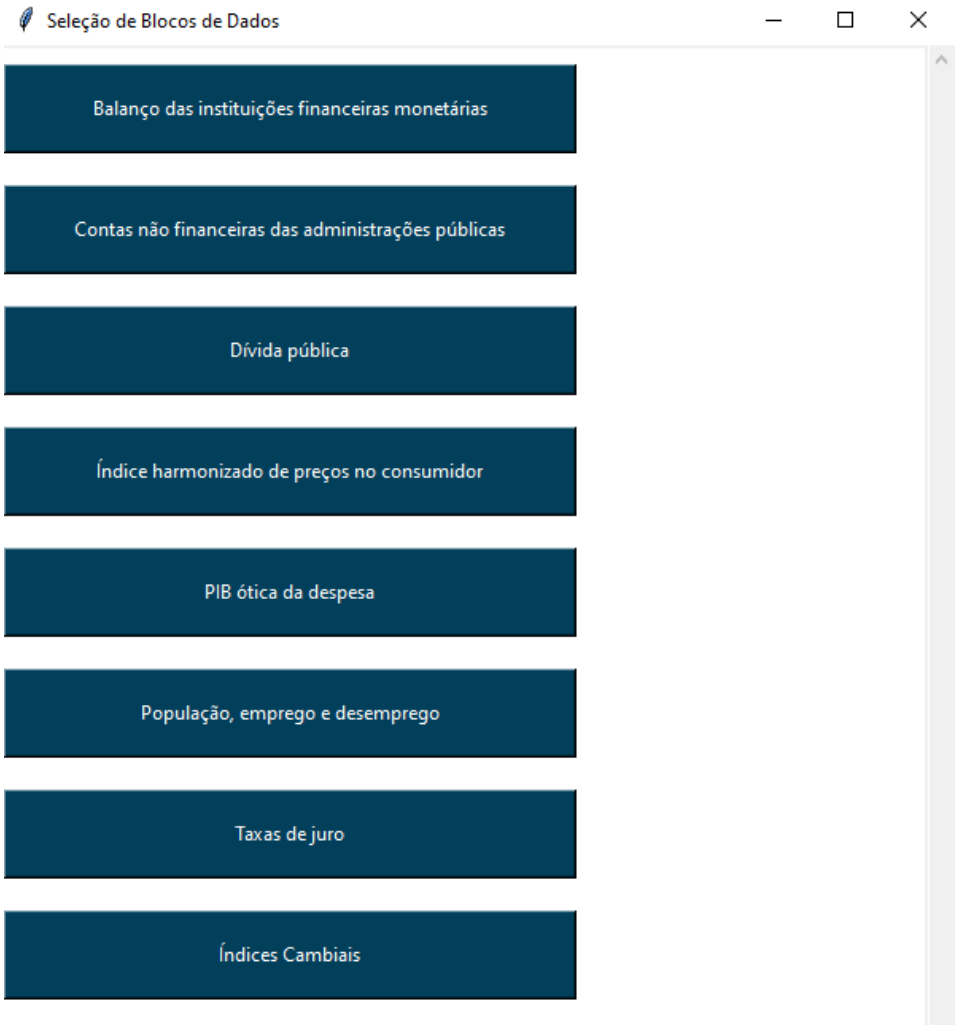

FIGURE 6 - Tool's initial interface

4.2 Project Overview

The following diagram provides an overview of the data consistency validation process developed in this project. It illustrates the main steps of the workflow, while a detailed explanation of each step is provided in the following sections.

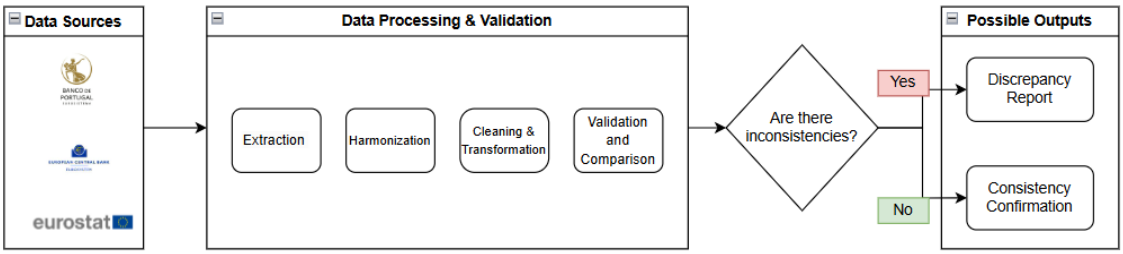

FIGURE 7 - Project Overview Diagram

4.3 Data extraction

Building upon the concepts explored in the literature review, the chosen methodology to extract data from different sources leveraged several public APIs. Given the need for structured data retrieval, the approach required defining appropriate API endpoints and dynamically constructing request URLs.

Firstly, the BP*stat* API provides programmatic open access to a comprehensive repository of statistical data and metadata hosted by BP*stat*. It allows users to make customized requests, facilitating the retrieval of specific sets of data aligned with their needs.

Built on RESTful architecture, it employs the 'GET' method to fetch data, delivering responses in JSON-stat format. The BP*stat* API concedes access to thousands of statistical series organized by statistical domains and datasets, covering several economic and financial areas (Banco de Portugal, n.d.-a).

To retrieve data, the tool first fetched metadata about the requested series via the series endpoint to determine the hierarchical structure that includes the domain and dataset ID associated with it. BP*stat* organizes its data in a hierarchical manner, where domains act as broader categories, encompassing multiple datasets. Once this information is obtained, a second request is made to fetch the historical observations of the series.

The ECB API is based on the SDMX Restful architecture and provides structured access to statistical data. In this project, it was employed to retrieve specific series observations, focusing on data aligned with project´s analytical goals.

Data is organized by datasets and dataflows, where each series is uniquely identified by a combination of dimensions and attributes, serving as a unique identifier (European Central Bank, n.d.).

To build the URL requests, the ECB API requires the specification of both a dataset and a series key to fetch observations. The dataset serves as the overarching structure that organizes dataflows, while the series key acts as a unique identifier for the specific series.

The Eurostat API enables access to statistical data related to the European Union, utilizing a Restful architecture aligned with SDMX standards. Its latest version (3.0) supports operations with large datasets and customized requests and can deliver the responses in multiple formats (Eurostat, n.d.).

In this case, to build the URL request the API requires specifying both the dataset ID and additional filters, in a predefined order. These filters function as attributes that, when combined, form the external key that uniquely identifies a specific series.

The table below presents the base URLs and query structures used for data extraction, illustrating how API requests were dynamically constructed.

TABLE 3 - Dynamic API request structure for data retrieval

API	Base URL	Query Structure
BPStat	https://bpstat.bportugal.pt/data/v1	/domains/{domain_id}/datasets/{dataset_id}/?series_ids={series_id}&lang=EN
ECB	https://data-api.ecb.europa.eu/service/data/	/{dataset_id}/{series_key}?format=csvdata
Eurostat	https://ec.europa.eu/eurostat/api/dissemination/statistics/1.0/data	/{dataset_id}?format=JSON&geo={geo}&unit={unit}&sector={sector}&na_item={na_item}

Based on literature review and the nature of the data, JSON-stat format was selected for BP*stat* and Eurostat. For the European Central Bank API, JSON was chosen to ensure alignment with SDMX standards while maintaining compatibility with the project.

The following table summarizes the main characteristics of the APIs used in this project, complementing the detailed descriptions provided above.

TABLE 4 - Overview of APIs used in the project

API	Architecture	Response Format	Use Case in project
BPStat	REST	JSON-stat	Provides republished data to be validated
ECB	SDMX REST	SDMX-ML, CSV, **JSON**	Provides original data for key financial and macroeconomic indicators (e.g., interest rates, GDP, monetary statistics) used as a reference
EUROSTAT	SDMX REST	SDMX-ML, SDMX-CSV, **JSON-stat**, TSV	Provides original statistical data on socio-economic indicators (e.g., public debt, employment, government finance) used as a reference

These APIs were selected for their reliability and their role as primary sources of the required data. Collectively, they provide a robust and flexible foundation for automated data extraction. In addition, the ability to perform customized requests was essential in achieving the project´s objective, as it enabled the targeted retrieval of information specific to Portugal and the Euro Area. This approach minimized unnecessary processing, optimizing relevance in data handling.

To efficiently structure the automated API requests, a predefined mapping system was implemented using pre-existing working files containing metadata about the series published on BP*stat* and their corresponding external keys were employed. These files served as a reference point, dynamically guiding the construction of URLs for data extraction.

4.4 Data Processing & Validation

For this project, the data preparation step was critical to ensure consistency and comparability across the datasets. Since these APIs use different URL structures and endpoints, a programming language was employed to capacitate integration and future flexibility.

To standardize the retrieved data, specific preprocessing steps were applied. Firstly, observation values were rounded to two decimal places to mitigate precision discrepancies between platforms. Additionally, date formats were standardized to

ensure uniformity across the series, preventing potential conflicts during comparison and facilitating efficient filtering and manipulation.

Thus, to ensure consistency across datasets containing series with varying periodicities, it was necessary to standardize date formats. This step involved retaining only the year for the annual series while preserving full timestamps for higher-frequency data, such as daily or quarterly observations.

```
# Adjust date format depending on the periodicity of the series
if periodicidade == 'A':  #If it´s annual, show only the year
    data['TIME_PERIOD'] = pd.to_datetime(data['TIME_PERIOD'], format='%Y').dt.year
else:
    data['TIME_PERIOD'] = pd.to_datetime(data['TIME_PERIOD'])
```

FIGURE 8 - Code snippet for data preparation.

The datasets were merged using a period-based approach, aligning observations from different sources by their temporal dimension. To achieve this, an outer join methodology was employed, which ensures that all records from both datasets are retained, even if there is no matching entry in the other source. This technique is particularly useful for identifying unmatched values, as it highlights instances where a data point exists in one dataset but is missing in another, facilitating a more comprehensive comparison (IBM, 2025).

In the context of this project, applying an outer join was essential to ensure that no relevant information was excluded during the comparison process. By preserving all available observations, it became possible to detect inconsistencies such as missing periods in one source or values that differed between the datasets.

Finally, data discrepancies, including mismatched values and incomplete periods were possible to identify for further analysis.

Following the harmonization and preparation of the datasets, the tool systematically identifies inconsistencies between data sources by comparing both temporal alignment and recorded values.

To account for minor rounding differences that could otherwise generate false inconsistencies, a tolerance threshold of 0.1 was applied during the comparison process. This ensures that only substantive discrepancies are flagged for further analysis.

4.5 Error handling

Errors in data preparation can negatively impact the analytical process, leading to inaccurate results. To mitigate these risks, automated data validation plays a crucial role in ensuring the integrity and quality of the information for further analysis. Zhang et al. (2023) emphasize that the absence of robust validation mechanisms can result in significant operational setbacks, including increased costs and unreliable datasets. To optimize this process, several error-handling mechanisms were implemented in this project, to prevent data degradation.

- Error Handling in API requests

The process verifies whether requests are successful by analyzing the HTTP status code. If the response is not 200 (OK), an explanatory error message is displayed, interrupting the processing of corrupted data.

- Validation of data returned by APIs

After receiving a response, the tool checks whether the expected data is present. If it is missing, the execution flow is interrupted to prevent errors that could arise from attempting to manipulate an empty object. This validation step helps maintain the integrity of retrieved datasets.

- Specific error handling for the Eurostat API

Due to the added complexity of constructing specific URL requests using the Eurostat API, a dedicated mechanism was implemented to capture errors related to missing attributes that form the key to the request. This ensures that improperly structured requests are identified and handled before causing failures in data retrieval.

- Handling of missing series

If for some reason, the system is unable to retrieve information about a particular series, a clear and informative message is displayed, alerting the user to the issue. This is crucial to the early identification of missing data or changes in working files.

- Retries and Timeouts

To handle temporary failures in API requests, the system implements retries with exponential backoffs and timeouts. These mechanisms prevent the process from getting

stuck indefinitely and help mitigate issues caused by slow or temporarily unavailable services.

A timeout is defined to limit the maximum time the system waits for a response from an API.

If a request fails due to network issues, server unavailability, or rate limits, the system retries the request multiple times before considering it a failure. Instead of retrying immediately, the delay between retries increases exponentially (exponential backoff), reducing the risk of overwhelming the API with repeated requests in a short time.

4.6 Possible Outputs

The tool provides specific outputs based on the results of data validation. If no discrepancies are found during the validation process, a confirmation message is displayed, indicating that the data is consistent with the sources. This approach avoids unnecessary duplication of information in the local memory, ensuring that only essential data is processed and stored and focusing on the main goal which is detecting inconsistencies.

On the other hand, if discrepancies are detected, the tool generates a detailed report in CSV format, saved in the working directory. This report provides a straightforward and efficient way to identify the series with inconsistencies, specifying the periods in question. The goal is to document errors clearly, enabling quick and effective corrections.

Additionally, to simplify further analysis, the discrepancy report includes the external keys associated with the series. This feature allows users to directly reference and consult the original data from the source when necessary, streamlining the resolution process.

The figure below presents an excerpt example from the CSV report generated by the tool, showing the existing data inconsistencies. For presentation purposes, the BP*stat* IDs and external keys shown are fictitious and do not correspond to real datasets.

The *Date_Mismatch* column indicates whether there are observations for the same period in both data sources. A *TRUE* value means that a given period is reported by one

entity, but missing in the other, while *FALSE* indicates both sources contain data for the same period.

Additionally, the difference between the values from the analyzed sources is recorded in the *Difference* column, allowing for the identification of numerical variations between datasets.

To facilitate direct correction and ensure data traceability, the report also includes time series identifiers.

	A	B	C	D	E	F	G	H	I
1	Date	VALUE_BPSTAT	TIME_PERIOD	VALUE_BCE	Difference	Date_Mismatch	ID_BPStat	Chave_Externa	Periodicidade
2	2024-09	98.44	2024-09	99.4	-0.96	FALSE	123456789	M.HHH.PT.NNN.A	M
3			2024-10	99.24		TRUE	123456789	M.HHH.PT.NNN.A	M
4			2024-10	105.47		TRUE	123456789	M.HHH.PT.NNN.A	M
5			2024-10	100.82		TRUE	123456789	M.HHH.PT.NNN.A	M
6	2024-09	204.87	2024-09	205.4	-0.53	FALSE	123456789	M.HHH.PT.NNN.A	M
7			2024-10	205.76		TRUE	123456789	M.HHH.PT.NNN.A	M
8			2024-10	104.59		TRUE	123456789	M.HHH.PT.NNN.A	M
9			2024-10	102.01		TRUE	123456789	M.HHH.PT.NNN.A	M
10			2024-10	105.38		TRUE	123456789	M.HHH.PT.NNN.A	M
11			2024-10	149.31		TRUE	123456789	M.HHH.PT.NNN.A	M
12			2024-10	126.26		TRUE	123456789	M.HHH.PT.NNN.A	M
13			2024-10	95.71		TRUE	123456789	M.HHH.PT.NNN.A	M
14	2024-09	96.99	2024-09	96.31	0.67999999995	FALSE	123456789	M.HHH.PT.NNN.A	M
15			2024-10	95.43		TRUE	123456789	M.HHH.PT.NNN.A	M
16			2024-10	96.52		TRUE	123456789	M.HHH.PT.NNN.A	M
17			2024-10	91.84		TRUE	123456789	M.HHH.PT.NNN.A	M
18	2024-09	91.77	2024-09	91.25	0.52	FALSE	123456789	M.HHH.PT.NNN.A	M
19			2024-10	90.75		TRUE	123456789	M.HHH.PT.NNN.A	M
20			2024-10	91.64		TRUE	123456789	M.HHH.PT.NNN.A	M
21	2024-09	154.57	2024-09	153.88	0.68999999995	FALSE	123456789	M.HHH.PT.NNN.A	M
22			2024-10	153.85		TRUE	123456789	M.HHH.PT.NNN.A	M
23			2024-10	96.06		TRUE	123456789	M.HHH.PT.NNN.A	M
24			2024-10	104.16		TRUE	123456789	M.HHH.PT.NNN.A	M
25			2024-10	101.11		TRUE	123456789	M.HHH.PT.NNN.A	M
26	2024-09	97.85	2024-09	98.83	-0.98	FALSE	123456789	M.HHH.PT.NNN.A	M
27			2024-10	98.78		TRUE	123456789	M.HHH.PT.NNN.A	M
28			2024-10	108.83		TRUE	123456789	M.HHH.PT.NNN.A	M

FIGURE 9 - Example of a discrepancy report generated by the tool

5. PRACTICAL RESULT: APPLICATION OF THE TOOL TO EXCHANGE RATE INDICES

Benchmark revisions are revisions of data sources or methods to estimate national accounts indicators, carried out at least once every 5 years, at European and national level. In 2024, countries, including Portugal, implemented the 2021 edition of the benchmark revisions of national accounts. Although the year is defined, the release timing is not harmonized across countries, i.e., it may occur in different months (Banco de Portugal, 2024).

Banco de Portugal started releasing revised data in September 2024 and the developed tool began gradually integrating into the quality control processes in October to identify inconsistencies across various data blocks, leading to interesting results. One significant case involved data on effective exchange rate indices, where several discrepancies were identified.

These inconsistencies are primarily due to the dynamic nature of the data. Effective exchange rate indices are calculated using deflators (Gross Domestic Product, Consumer Price Indices, and Labour Costs) provided by the ECB, which are frequently revised to reflect updates in national accounts data. These revisions are necessary to ensure accuracy over time but impact all related effective exchange rates and all harmonized competitiveness indicators, as these metrics are interdependent across multiple countries in the Euro Area.

Moreover, since the calculations involve 38 countries, each with its own timeline for updating deflators, revisions occur throughout the year and can impact the full temporal scope of the data. This underscores the challenge of maintaining consistency in such datasets.

The developed tool added significant value to this process by validating the entire historical dataset in a short amount of time, something that would not have been feasible through manual efforts.

By swiftly identifying discrepancies, the tool ensured that inconsistencies were not only flagged and documented in the report but also immediately corrected.

After implementing the tool, its evident strengths and areas for improvement became clear. The following table summarizes these key aspects, highlighting both the benefits and the challenges that remain.

TABLE 5 - Strengths and Improvements of the Tool

Strengths	**Improvements**
Automates data verification, reducing manual effort and errors.	Requires ongoing maintenance to adapt to API updates.
Ensures accurate data comparison and consistency.	Minor rounding differences may cause false positives.
Processes large datasets efficiently.	May need adjustments to support new data sources.
Generates detailed CSV reports for easy error identification.	

6. CONCLUSION

Central banks and statistical institutions play a critical role in ensuring the accuracy and accessibility of official statistics, which are essential for informed decision-making by policymakers, businesses, and citizens. However, the increasing volume of data, the diversity of sources and formats, and the reliance on digital platforms have introduced new challenges in maintaining data integrity and consistency.

This study addresses the critical challenge of data integration by developing an automated tool to ensure the consistency of statistical data republished by Banco de Portugal with original data sources. Given the complexity of managing multiple data sources with varying structures and updating cycles, maintaining coherence across datasets is essential for preserving data reliability. The proposed tool facilitates this process by leveraging APIs and automation to systematically validate statistical information between BP*stat* and its primary sources, the European Central Bank and Eurostat. By streamlining data verification, this solution enhances efficiency, accuracy, and trust in the dissemination of official statistics.

The main contribution of this tool is the automation of statistical data quality control, enabling continuous verification of data consistency between BP*stat* and external sources. So, this approach eliminates the need for manual checks and enhances the reliability of republished statistics. By systematically identifying updates, revisions, and structural changes in datasets, the tool ensures greater accuracy and timeliness in statistical dissemination. Its implementation successfully detected inconsistencies across multiple datasets, particularly in the real effective exchange rate indices, where

methodological adjustments and periodic revisions led to data discrepancies. Furthermore, its integration into Banco de Portugal's internal workflows has strengthened data monitoring processes, enabling real-time validation that is aligned with scheduled updates and reinforcing overall data integrity.

Despite these advancements, the project faced several challenges, particularly in integrating the tool into existing systems and ensuring seamless interoperability between different data sources. While APIs facilitated data retrieval and comparison, variations in data structures, metadata representation, and formatting required additional preprocessing steps, increasing the complexity of integration. Additionally, the tool's reliance on predefined mappings meant that any structural changes in source datasets required frequent manual adjustments, emphasizing the need for greater flexibility and adaptability in automated data validation solutions.

Another key limitation was the restricted timeframe of six months, within which both the conceptualization and implementation of the solution had to be completed. A significant portion of this period was dedicated to understanding the business context, internal workflows, and the specific data structures used by Banco de Portugal. While this initial learning curve was crucial to ensuring the tool's alignment with institutional requirements, it also limited the available time for further development, testing, and refinement.

The initial deployment of the tool within Banco de Portugal has demonstrated its potential to enhance statistical consistency and streamline validation processes. Building upon these successes, future research presents several opportunities to further develop and expand its capabilities.

One promising avenue for advancement is the integration of machine learning techniques or artificial intelligence languages for anomaly detection and immediate discrepancy resolution. While the current rule-based approach effectively identifies inconsistencies, incorporating AI-driven models could enable the tool not only to detect complex discrepancies but also to suggest corrections in real-time. This would enhance the system's ability to proactively address inconsistencies, further strengthening statistical reliability.

Beyond technical enhancements, expanding the tool's applicability to a broader range of datasets and statistical domains represents another valuable direction. By adapting the tool to accommodate a wider variety of economic and financial indicators, it could serve as a comprehensive validation framework for multiple statistical outputs, reinforcing data quality and consistency.

Additionally, there is potential for scaling the tool for broader institutional use, particularly among other central banks and statistical organizations facing similar data integration challenges. As international data harmonization efforts continue to evolve, aligning the tool with emerging statistical frameworks such as updates to the SDMX standard, could ensure long-term relevance and interoperability. Collaborative efforts with external institutions could foster knowledge exchange and cross-institutional standardization, ultimately contributing to a more cohesive and transparent global statistical system.

Overall, this research highlights the transformative potential of automation in modern statistical dissemination. By addressing data consistency challenges through technological innovation, this project enhances transparency, accuracy, and public trust. As data-driven decision-making continues to shape economies and societies, ensuring the integrity of statistical information is not only a necessity but a cornerstone of transparency and accountability in the modern world.

REFERENCES

Alma'aitah, W. Z., Quraan, A., AL-Aswadi, F. N., Alkhawaldeh, R. S., Alazab, M., & Awajan, A. (2024). Integration Approaches for Heterogeneous Big Data: A Survey. *Cybernetics and Information Technologies*, *24*(1), 3–20. https://doi.org/10.2478/cait-2024-0001

AltexSoft. (2023, January 17). *Comparing SOAP vs REST vs GraphQL vs RPC API*. https://www.altexsoft.com/blog/soap-vs-rest-vs-graphql-vs-rpc/

Aparicio, J. T., Aparicio, M., & Costa, C. J. (2023). Design Science in Information Systems and Computing. In S. Anwar, A. Ullah, Á. Rocha, & M. J. Sousa (Eds.), *Proceedings of International Conference on Information Technology and Applications* (Vol. 614, pp. 409–419). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-9331-2_35

Banco de Portugal. (n.d.-a). *BPstat Data API*. Retrieved January 25, 2025, from https://bpstat.bportugal.pt/data/docs

Banco de Portugal. (n.d.-b). *Função estatística | Banco de Portugal*. Retrieved January 23, 2025, from https://www.bportugal.pt/page/funcao-estatistica

Banco de Portugal. (n.d.-c). *Organograma do Banco de Portugal | Banco de Portugal*. Retrieved January 23, 2025, from https://www.bportugal.pt/page/organograma-do-banco-de-portugal

Banco de Portugal. (2019). *Sobre o BPstat*. BPstat. https://bpstat.bportugal.pt/sobre-o-bpstat

Banco de Portugal. (2021). *Plano Estratégico do Banco de Portugal 2021-2025*.

Banco de Portugal. (2024, July 9). *A partir de setembro, algumas estatísticas publicadas no BPstat, vão ser revistas. Saiba mais aqui.* BPstat. https://bpstat.bportugal.pt/conteudos/noticias/2213

Bank for International Settlements, European Central Bank, International Monetary Fund, Organisation for Economic Co-operation and Development, Statistical Office of the European Communities, & United Nations Statistical Division. (2003). *Report on Common Open Standards for the Exchange and Sharing of Socio-economic Data and Metadata: The SDMX Initiative*. https://sdmx.org/wp-content/uploads/StatisticalCommission2003.pdf

Biao, H.-D. K., & Assogba, E. (2024). *Statistical data governance based on the SDMX*. *CEUR Workshop Proceedings, ISSN 1613-0073*.

Biehl, M. (2015). *API Architecture*. API-University Press.

Bigelow, S. J. (2024, December 23). *What Are the Types of APIs and Their Differences? | TechTarget*. Search App Architecture. https://www.techtarget.com/searchapparchitecture/tip/What-are-the-types-of-APIs-and-their-differences

Blinder, A. S. (2001). *How Do Central Banks Talk?* Centre for Economic Policy Research.

Blinder, A. S., Ehrmann, M., de Haan, J., & Jansen, D.-J. (2022). *Central Bank Communication with the General Public: Promise or False Hope?* (Working Paper No. 30277). National Bureau of Economic Research. https://doi.org/10.3386/w30277

Catchpoint. (n.d). *API Architecture Patterns and Best Practices*. https://www.catchpoint.com/api-monitoring-tools/api-architecture

Chadere, I. (2024, March 27). *Understanding API Response Types and Formats: A Comprehensive Guide*. https://apidog.com/blog/understanding-api-response-types-and-formats-a-comprehensive-guide/

European Central Bank. (n.d.). *Overview | ECB Data Portal*. Retrieved January 25, 2025, from https://data.ecb.europa.eu/help/api/overview

European Commission (Ed.). (2018). *European statistics code of practice: For the national statistical authorities and Eurostat (EU statistical authority)*. Publications Office. https://doi.org/10.2785/914491

Eurostat. (n.d.). *Data Browser—User guides—Eurostat*. Retrieved January 25, 2025, from https://ec.europa.eu/eurostat/web/user-guides/data-browser

Eurostat. (2019). *Quality Assurance Framework—Quality—Eurostat*. https://ec.europa.eu/eurostat/web/quality/european-quality-standards/quality-assurance-framework

Eurostat. (n.d). *Eurostat Quality Policy*. https://ec.europa.eu/eurostat/documents/64157/4392716/Eurostat-Quality-Policy-EN.pdf

Feijo, C., & Valente, E. (2005). *AS ESTATÍSTICAS OFICIAIS E O INTERESSE PÚBLICO*. *15*.

Fielding, R. T. (2000). *Fielding Dissertation: CHAPTER 5: Representational State Transfer (REST)*.

https://web.archive.org/web/20210512181107/https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm

GraphQL Foundation. (n.d). *GraphQL | A query language for your API*. https://graphql.org/

Groth, P., Loizou, A., Gray, A. J. G., Goble, C., Harland, L., & Pettifer, S. (2014). API-centric Linked Data integration: The Open PHACTS Discovery Platform case study. *Journal of Web Semantics*, *29*, 12–18. https://doi.org/10.1016/j.websem.2014.03.003

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

IBM. (2025, January 24). *IBM Data Virtualization Manager for z/OS 1.1.0*. https://www.ibm.com/docs/en/dvm/1.1.0?topic=functions-outer-join

Kellerman, M., & David, E. (2021). Implementation of the SDMX International Standard in Statistical Data Specification. *Bank of Israel*.

Laaboudi, C., Karlberg, M., & Islam, M. (2024). Open data dissemination at Eurostat: State of the art. *Statistical Journal of the IAOS*, *40*(1), 81–90. https://doi.org/10.3233/SJI-230115

Lamothe, M., Guéhéneuc, Y.-G., & Shang, W. (2021). A Systematic Review of API Evolution Literature. *ACM Comput. Surv.*, *54*(8), 171:1-171:36. https://doi.org/10.1145/3470133

MuleSoft. (n.d). *What is an API? (Application Programming Interface)*. https://www.mulesoft.com/api/what-is-an-api

Pedro, B. (2024). *Building an API Product: Design, implement, release, and maintain API products that meet user needs*. Packt Publishing Ltd.

Peffers, K., Tuunanen, T., Gengler, C., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research. *Proceedings of First International Conference on Design Science Research in Information Systems and Technology DESRIST*.

Sæbø, H. V., & Holmberg, A. (2019). Beyond code of practice: New quality challenges in official statistics. *Statistical Journal of the IAOS*, *35*(2), 171–178. https://doi.org/10.3233/SJI-180463

SDMX. (2022, January 4). *History of the SDMX Initiative | SDMX – Statistical Data and Metadata eXchange*. https://sdmx.org/about-sdmx/history/

Sembiring, J., & Uluwiyah, A. (2015). Data and Metadata Exchange Design with SDMX Format using Web Service for Interoperability Statistical Data. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, *14*(2), Article 2.

Smith, A., & Rogers, S. (2008). Web 2.0 and official statistics: The case for a multi-disciplinary approach. *Statistical Journal of the IAOS*, *25*(3–4), 117–123. https://doi.org/10.3233/SJI-2008-00696

SPE. (2024, October). *Replicabilidade e Controlo de Confidencialidade*. SPE Sociedade Portuguesa de Estatística. https://www.spestatistica.pt/publicacoes/publicacao/boletim-spe-outono-de-2024

Stahl, R., & Staab, P. (2018a). History of SDMX. In R. Stahl & P. Staab (Eds.), *Measuring the Data Universe: Data Integration Using Statistical Data and*

Metadata Exchange (pp. 73–83). Springer International Publishing. https://doi.org/10.1007/978-3-319-76989-9_11

Stahl, R., & Staab, P. (2018b). *Measuring the Data Universe: Data Integration Using Statistical Data and Metadata Exchange*. Springer.

United Nations. (2014). *Fundamental Principles of Official Statistics*. https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx

United Nations. (2019). *Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics*. UN. https://doi.org/10.18356/418997cc-en

World Wide Web Consortium. (2007, April 27). *SOAP Version 1.2 Part 0: Primer (Second Edition)*. https://www.w3.org/TR/2007/REC-soap12-part0-20070427/#L3631

Zhang, L., Howard, S., Montpool, T., Moore, J., Mahajan, K., & Miranskyy, A. (2023). Automated data validation: An industrial experience report. *Journal of Systems and Software*, *197*, 111573. https://doi.org/10.1016/j.jss.2022.111573