Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

MASTER´S FINAL WORK

INTERNSHIP REPORT

DASHBOARD DEVELOPMENT AT THE PORTUGUESE CINEMA INSTITUTE

MICHELE SCOLA

JANUARY - 2026

MASTER
DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK
INTERNSHIP REPORT

DASHBOARD DEVELOPMENT AT THE PORTUGUESE CINEMA INSTITUTE

MICHELE SCOLA

SUPERVISION:
PROF. CARLOS J. COSTA

JANUARY - 2026

GLOSSARY

BI Business Intelligence.

DSS Decision Support System.

DV Data Visualization.

DW Data Warehouse.

ELT Extract-Load-Transform.

ETL Extract-Transform-Load.

EWD Enterprise Data Warehouse.

ICA Istituto do Cinema e do Audiovisual.

IFPS Interactive Financial Planning System.

OLAP Online Analytical Processing.

ABSTRACT

This report presents the development and implementation of dashboards during a four-month internship at the Instituto do Cinema e do Audiovisual (ICA), the Portuguese public institute responsible for supporting and regulating the cinema and audiovisual sector. The work was structured around two distinct projects, both conducted following an adapted CRISP-DM methodology, allowing for a systematic and iterative approach to data analysis, visualization, and knowledge extraction.

The first project focused on internal organizational data, with the objective of designing dashboards to monitor compliance with investment obligations imposed on cinema market operators. This involved cleaning, structuring, and transforming internal datasets, followed by the development of interactive visualizations to facilitate ongoing monitoring and reporting. The second project addressed the need for market intelligence within the cinema industry. External datasets were collected, filtered, and analyzed to provide relevant insights on industry trends, audience behaviors, and strategic opportunities, to support ICA's policy and operational decisions.

This study demonstrates the practical application of business intelligence (BI) methodologies and data visualization principles in a public cultural sector context, highlighting how dashboards can serve as tools for bridging the gap between raw data and strategic decision-making. Furthermore, it explores how these tools can contribute to a data-driven organizational culture, encouraging evidence-based planning and policy implementation. By documenting the methodological choices, challenges, and outcomes, this report provides a valuable case study for the integration of BI approaches in cultural policy administration and offers insights for similar public institutions seeking to enhance their analytical capabilities.

TABLE OF CONTENTS

Glossary **i**

Abstract **ii**

Table of Contents **iii**

List of Figures **v**

1 Introduction **1**

1.1 Context 1
1.2 Institutional Context and Legal Framework 1
1.3 Scope and Motivation 2
1.4 Objectives and Methodology 2
1.5 Report Structure 3

2 Literature Review **4**

2.1 Business Intelligence 4
2.2 Data Visualization 4
2.2.1 Theory and Foundations 4
2.2.2 Application 6
2.3 Pipeline 6
2.4 Data Warehouse 7
2.4.1 Architecture 8
2.4.2 Dimensional Modeling 9

3 Methodology **11**

3.1 CRISP-DM Process Model 11
3.2 Project Methodology 12

4 Project Excecution: from Data to Insights **14**

4.1 Project 1: Investment Obligations 14
4.1.1 Context and Objectives 14
4.1.2 Data and Relational Model 14
4.1.3 Dashboard Design 15
4.2 Project 2: Cinema Industry 17
4.2.1 Context and Objectives 17
4.2.2 Data and Relational Model 18
4.2.3 Dashboard Design 19

5 Conclusion **21**

Bibliography **22**

List of Figures

1 Project 1: Relational Model 15
2 Project 1: Entities Page 16
3 Project 1: Waiting Payments Page 16
4 Project 1: Report Status Page 17
5 Project 2: Overview Page 19
6 Project 2: Market Share Page 20
7 Project 2: Forecast Page 20

ACKNOWLEDGEMENTS

I would like to thank Professor Carlos J. Costa for his guidance and constant availability during the supervision of this work. I am also grateful to Dr. Caterina Correia for her mentorship and for providing me with the opportunity to be part of her team, an experience that contributed to my professional and academic growth.

I would also like to thank the ICA team for the warm welcome, collaboration, and supportive working environment.

1 Introduction

1.1 Context

Business intelligence and data visualization tools are becoming crucial for the transformation of raw data into visuals and insights that can support management decisions and planning. Their implementation in public and cultural sector presents specific challenges like transparency requirements and complexity of institutional processes and regulations.

This report presents the development of data visualization solutions in the context of a four-month internship at the Portuguese public cinema institute, responsible for cinema and audiovisual sector policies. This type of institutions manage public funding schemes, regulatory compliance and sector incentives. Business intelligence tools can support these processes by offering efficient ways to monitor activities, track compliance, and inform strategic decisions.

1.2 Institutional Context and Legal Framework

The internship was carried out at Istituto do Cinema e do Audiovisual I.P. - ICA based in Lisbon. It's a public institute administratively and financially autonomous and indirectly controlled by Portugal State. It operates under the supervision of the Ministry of Culture and counts approximately 50 staff members.

ICA's mission is to support the development of cinematic and audiovisual activities in Portugal managing taxes and public financial supports with the final goal of strengthen and grow the cinema and audiovisual sectors.

ICA is the entity responsible for collaborating with the government in defining public policies for the cinema and audiovisual sectors and for ensuring their execution. Its responsibilities include the management of public support schemes, the promotion of national and international circulation of works, representation in international institutions, collaboration on international agreements, and the collection and dissemination of sector-related statistical information. ICA also plays a strategic role in promoting Portugal as an international filming destination and in leveraging incentive legislation and co-production agreements.

ICA is structured around two main departments: department of management and department of cinema. The internship was carried out within the department of management whose responsibilities include budget planning and control, financial incentives, tax controls, obligation investments compliance, contracts, legal affairs, human resources, and management control. The department of cinema focuses more on the management of production support schemes, competitions, communication, international representation, and sector statistics.

ICA's story started in 1971, with the creation of the Portuguese Institute of Cinema (IPC). Over the decades, the institution evolved into different forms like IPC, IPACA, and ICAM, always keeping legal personality and administrative and financial autonomy.

The current configuration of ICA was established in 2007. Its operational framework follows the Cinema Law (Law no. 55/2012), which defines the principles guiding State intervention in the cinema and audiovisual sectors. This legislation introduced a system based on financial incentives, sector taxes and investment obligations imposed on market operators, including television broadcasters, distributors, exhibitors, and on-demand audiovisual service providers. These constitute the source of funding for public support programs and require structured processes of monitoring, verification, and control.

1.3 Scope and Motivation

The internship includes the integration into the dynamics of the management department, monitoring and participating in the execution of PowerBI dashboards for the various operational areas managed by ICA.

It aims to get into contact with core processes of public administration in the cultural sector, focusing on the areas of management activities, development of monitoring dashboards and data analysis. These activities support internal decision-making and the dissemination of analytical information across different areas of ICA.

Recently, the Institute's budget has increased significantly, reaching 50 million euros, which implies a strengthening of the control and planning of its activity. In this context, the definition and monitoring of activity indicators are essential to support decision-making and ensure more efficient and transparent management.

The scope of the internship is therefore focused on data analytics and business intelligence with the objective of contributing to the development of analytical tools capable of supporting the activities of the management department and other operational units.

1.4 Objectives and Methodology

Goals include the conception and development of monitoring dashboards in Power BI on the ICA platform HAL and the strengthening the integration of data from the different operational groups, ensuring consistency and reliability of information. In addition, the internship aimed to promote an analytical vision that contributes to strategic decision-making in the institute.

In the first part of the internship, an initial analysis was conducted by examining ICA's activities and the reporting context. This phase corresponds to the business understanding stage of traditional Information Systems Development processes and data mining methodologies such as CRISP-DM. Its objective was to identify the activities carried out by the Management Department, together with their information and monitoring needs, in order to define the scope of the analytical work.

Based on this initial understanding, the required data to support operational activities and management decision-making were identified and collected. This phase aligns with the data understanding and analysis stages, during which the available data sources were examined and prepared to support the creation of analytical outputs aimed at addressing the identified business needs.

In the final phase, analytical outputs were developed with the objective of creating dashboards for both operational and management units. This step is consistent with the testing and deployment stages of the development process. The dashboards were made available to the designated users, and feedback was collected through interactions with internal stakeholders in order to refine the solutions and better align them with users' requirements.

Regarding data collection, ICA collaborates with an external technology provider, NTT Data, which developed and maintains a centralized digital platform named HAL that supports ICA's core processes. For the purposes of the internship, selected data were made available in spreadsheet format through structured extractions provided by the external service provider.

In addition to internal operational data, external datasets were also used. These datasets were obtained through third-party providers contracted by ICA. For example, Ampere Analysis provides data related to the global cinema and audiovisual sector, which were used to enrich the analytical context.

1.5 Report Structure

Chapter 2 reviews the literature relevant to the projects executed during the internship, starting with the concepts of business intelligence and data visualization, and then addressing data pipelines and data warehouses. Chapter 3 presents the methodological framework adopted for the projects, based on the CRISP-DM model. Chapter 4 describes the execution and implementation of the projects.

2 LITERATURE REVIEW

2.1 Business Intelligence

Business Intelligence (BI) is a system of processes, methodologies and technical solutions designed to manage and transform data into actionable information to support organizational decision-making. BI combines data gathering, data storage and knowledge management (Negash and Gray, 2008) and can be viewed as data-driven DSSs according to Power's DSS framework (Power, 2002) as they focus the analysis on large structured and unstructured data. The term was first introduced in literature in 1958 by Luhn. Luhn's intelligence system would utilize data-processing machines for auto-abstracting and auto-encoding of documents and for creating interest profiles for each of the "action points" in an organization (Luhn, 1958). In 1989 Howard Dresner used BI as an umbrella term fact-based methods aiming to improve business decisions when he was at Gartner Group. BI has evolved through a series of increasingly sophisticated technological antecedents. These include early decision support systems, fourth-generation financial planning languages like IFPS, executive information systems, the data warehouse, online analytical processing (OLAP), and data mining. Modern software integrates these components into a unified system (Negash and Gray, 2008).

In the organizational context of decision-making, BI supports the interactions between Ackoff pyramid's elements: data, information and knowledge (Ackoff, 1989). Data are raw facts and messages (Choo et al., 2000) and they evolve into information when they are vested with meaning or deliver a message (Davenport and Prusak, 1998). Knowledge is created when the information delivers experience, values, insights, and contextual information (Davenport and Prusak, 1998, Stenmark, 2002). Data Visualization solutions support the transition from information to knowledge by allowing users to dynamically explore patterns, anomalies, and relationships. DV becomes a critical cognitive tool that helps decision-makers transform structured information into actionable organizational knowledge (Chaudhuri et al., 2011).

2.2 Data Visualization

2.2.1 Theory and Foundations

Data visualization serves as a bridge between raw data and human understanding. Islam and Jin (2019) define it as the graphical representation of quantitative information and as the transformation of datasets into visuals that make complexity quickly understandable and interpretable by the human brain. The theoretical foundations of effective data visualization is based on two main concepts: design principles that guide the creation of clear and correct graphics, and the understanding of human visual perception that explains why certain representations succeed while others fail (Islam and Jin, 2019).

Tufte (2001) established fundamental principles for excellence in statistical graphics, arguing that the primary goal is to communicate complex ideas with clarity, precision, and efficiency. His work emphasizes that graphical displays should show the data itself, avoiding distortions and focusing on substance rather than decoration. Tufte's introduced the concept of maximizing the data-ink ratio which is the proportion of a graphic's ink that actually display data. He advocates for the elimination of non-data-ink and redundant elements. Excellent graphics present many numbers in a small space, make large datasets coherent, encourage the eye to compare different pieces of data, reveal information at multiple levels of detail and serve a clear purpose, whether descriptive, exploratory, or analytical.

While Tufte provides prescriptive guidance on what makes visualizations effective, Ware (2004) offers a framework grounded in perceptual and cognitive science. Ware conceptualizes visualization as a four-stage process: data collection and storage, preprocessing to transform data into understandable forms, display hardware and graphics algorithms that render images, and the human perceptual and cognitive system that interprets these images. Visualization effectiveness depends not only on technical implementation but fundamentally on alignment with how humans process visual information. Ware distinguishes between two types of visual representations: sensory representations and arbitrary conventions. Sensory representations are effective because they align with early stages of neural processing and exhibit remarkable stability across individuals, cultures, and time periods. These representations require no training, resist instructional bias, provide sensory immediacy, and maintain cross-cultural validity. Arbitrary conventions derive their meaning from cultural context and are therefore harder to learn, easier to forget, and embedded within specific applications or cultural frameworks. On the other hand, arbitrary conventions possess formal power and can evolve rapidly to meet changing needs. Ware argues that the goal of information visualization should be to create conventions based on sound perceptual principles, leveraging the universal aspects of human vision rather than relying solely on culturally-dependent notation systems.

Costa and Aparício (2019) focus on the relationship between visualization objectives and chart selection in dashboard design. Their work highlights that different analytical purposes require different types of charts and emphasizes the importance of aligning visual representations with decision-making contexts, reinforcing the idea that effective dashboards depend on the appropriate match between data characteristics, user needs, and visualization techniques.

2.2.2 *Application*

In BI data visualization principles are applied to support decision-making and strategy. The main application of data visualization in BI environment are dashboards.

Few (2006) defines a dashboard as "a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance." Dashboards are visual, relying on graphics to communicate information more efficiently. They consolidate different information from multiple sources onto a single screen, ensuring that they can be absorbed quickly and simultaneously by the user.

Shneiderman (1996) articulated the Visual Information Seeking Mantra, which can be apllied to efficient dashboard design: "Overview first, zoom and filter, then details-on-demand." The mantra can be followed working with any type of data and describes the natural progression of visual data exploration: users first gain an overview of the entire dataset, then zoom into areas of interest, filter out irrelevant information, and finally access detailed information about specific items when needed. Additional capabilities such as viewing relationships among items, maintaining a history of actions for undo and refinement, and extracting sub-collections further enhance the exploration process.

Few (2006) identifies common mistakes in dashboard design that can lower their effectiveness in BI contexts. For example, exceeding the boundaries of a single screen, choosing inappropriate display media and cluttering displays with useless decoration that distracts from the data itself. These mistakes are based on a misunderstanding of visual perception principles or from prioritizing aesthetic appeal over functional clarity.

The application of data visualization in BI environments through dashboards demonstrates the practical implementation of the theoretical foundations discussed previously. Tufte's principle of maximizing data-ink translates directly into Few's admonition against decorative clutter, while Ware's insights into visual perception inform both the choice of display media and the organization of information on screen. Similarly, Shneiderman's interaction mantra explain the concept of revealing data at different levels of detail, enabling users to move between strategic overviews and details.

2.3 *Pipeline*

A data pipeline is a structured process that transfer data from its source to a data warehouse or data lake through extraction, processing, and loading steps. It ingests data via APIs or queries, handling issues like redundancy, missing values, and schema changes, before making it available for analytics and BI.

Traditionally, pipelines used the Extract-Transform-Load (ETL) model, transforming data before storage. Modern pipelines often follow ELT, loading raw data into a data lake and transforming it as needed. They support batch and real-time ingestion, ensuring scalable, reliable, and timely data delivery for decision-making (Cottur and Gadad, 2020).

In 2009, ETL systems were introduced to support the initial extraction, transformation, and loading of data into warehouses, although they faced limitations due to closed data sources and lack of standardization. By 2012, research focused on real-time ETL for data warehousing, enabling more timely analytical processing. In 2013, the ELT approach emerged, allowing raw data to be loaded first and transformed only when needed for queries (Cottur and Gadad, 2020).

In 2016, ETL processes were applied in clinical contexts, ensuring data integrity, and Amazon S3 provided cost-effective, high-availability cloud storage, eliminating the need for pre-cleaning or compression. The following year, data lakes further promoted ELT, enabling flexible handling of raw data. In 2018, pipelines were developed for reproducible research using R and SQL, as well as for scalable aggregation of scientific data. By 2019, privacy-enhanced ETL for biomedical data, distributed on-demand ETL systems (DOD-ETL), and financial sector applications using Spark and OLAP were explored, emphasizing real-time processing and low-latency analytics (Cottur and Gadad, 2020).

2.4 Data Warehouse

A data warehouse (DW) is a subject-oriented, integrated, non-volatile, and time-variant collection of corporate granular data in support of management decision-making (Inmon, 2002). It's subject-oriented because classical operation systems are organized around the applications of the company. Integrated because data come from multiple heterogeneous operational sources and are transformed during the loading process. Through activities such as conversion, reformatting, standardization and aggregation, the data is reconciled into a consistent and unified structure. As a result, once stored in the data warehouse, data presents a single, coherent corporate view. DW is non-volatile because data is loaded in snapshot and consequently kept in the DW in a history format. Finally, DW is time-variant: each data element stored in the data warehouse is associated with a precise point in time (Inmon, 2002).

Kimball and Ross (2013) descripes a set of requirements and objective for a DW. First, information must be easly accessible. The DW must present information consistently, meaning they must be credible and carefully assembled implying common labels and definitions must be used across data sources. The DW must adapt to unforeseen changes. It must present information in a timely way. It must protect the information. It must serve as the authoritative and trustworthy foundation for improved decision making. It must be accepted by the business community in order to be succesful (Kimball and Ross, 2013).

A DW stores information with different level of granularity and detail serving different types of needs and depth interrogations. There are four levels: an older level of detail (usually on alternate, bulk storage), a current level of detail, a level of lightly summarized data (the data mart level), and a level of highly summarized data (Inmon, 2002).

Granularity refers to the level of the level of detail of the unit of data. The higher the level of detail, the lower the granularity and vice versa. Some benefits of granularity are: re-usability, because granular data can be used by more users and for different pourposes and corporate functions, analytical flexibility, because the same dataset can be aggregated and examined across various dimensions, time periods, and perspectives, historical completeness, because it allows the retention of a detailed history of corporate activities and events and finally, the adapt to change, thanks to granularity the DW is able to change for future unforeseen business requirements (Inmon, 2002).

2.4.1 Architecture

Following Kimball's architecture (Kimball and Ross, 2013) there are four main components to consider in a DW environment: operational source systems, ETL systems, data presentation area and business intelligence application.

Operational source systems are records of operational business transactions. They can be considered outside of the DW and are the origin of data that will be used in the DW. They are designed for operational support rather then analytical use and they prioritize processing performance and availability. Source systems are queried only through very simple, single-record queries that are essential to daily transaction processing and designed to place minimal load on the operational system. Source systems are not design to keep historical data and enterprise data consistency. A DW can solve these problems by storing history and integrating data, while ERP and MDM systems can help standardize operational data at the source (Kimball and Ross, 2013).

ETL systems sits between the operational source systems and the data presentation area. During the extraction, source data are read and copied into the ETL environment switching from operational to analytical and officially belonging to the DW. The transformation process add value and conform the data thanks to operations like cleansing, standardization, integration, de-duplications. During the final step the system structure and load clean and integrated dimensional data into the presentation area ready for business use (Kimball and Ross, 2013).

The data presentation areas is where data are made available to BI users and applications. Data presentation should store granular data in order to allow ad-hoc and unpredictable queries by users. Dimensional models should reflect business processes events and should be built using conformed dimension in order to ensure integrations (Kimball and Ross, 2013).

BI applications are tools that allow the user to query and interact with the presentation data, which is key practice for data analytics and data-driven decision making (Kimball and Ross, 2013).

There are alternative architectures to the Kimball apporach. one of them is the indipendent data mart architecture, where analytical data is extracted and treated by business department. This aporach can provide very fast queries but can also create redundancy into different storages since different departments can extrat the same data from the same source system (Kimball and Ross, 2013).

Another alternative approach is Inmonn Hub and Spoke Corporate Information Factory. This Architecture include a normalized atomic repository named Enterprise Data Warehouse (EWD) after the ELT. The EWD works as an enterprise data coordinator, the same function of the bus in the Kimball apporach (Kimball and Ross, 2013).

2.4.2 Dimensional Modeling

Dimensional modeling is the preferred technique for presenting analytical data because it delivers understandable data to the business user and provides fast queries. The goal is to make databases simple, making them easier to understand and navigate. Dimensional models are often visualized as a data cube, where dimensions such as product, market, and time are the axes, and measurements like sales or profit are stored at the intersections. This intuitive approach ensures users can easily work with the data (Kimball and Ross, 2013).

In a dimensional model, fact tables store the performance measures deriving from organization's business processes' events. Each row in a fact table represent a measurement event and the data in each row is at a certain level of detail called grain. All the measurements must have the same grain (Kimball and Ross, 2013).

Dimension tables are essential components of a data warehouse, providing descriptive context to business process events. They typically contain multiple attributes, such as product descriptions, categories, and time-related data, and are linked to fact tables through primary keys. These attributes serve as the basis for filtering, grouping, and labeling in queries and reports. The quality and clarity of dimension attributes are key to robust analytics, with clear, business-friendly terminology preferred over cryptic codes. Dimension tables often represent hierarchical relationships and are generally denormalized to simplify access and enhance query performance. This design prioritizes simplicity, ensuring data is easily understandable and queryable (Kimball and Ross, 2013).

The star schema is a dimensional model that organizes data into a central fact table surrounded by dimension tables. This star-like structure, known as a star join, is easy to understand and navigate, making it user-friendly for business users. The schema's simplicity also offers performance advantages, as fewer joins are needed, and database optimizers can efficiently process queries by first narrowing down dimension tables and then joining them to the fact table. One of the strengths of the star schema is its extensibility. It can easily accommodate changes, such as adding new dimensions or facts, without disrupting existing data or BI applications. This structure makes the star schema both intuitive and efficient for analyzing large datasets (Kimball and Ross, 2013).

When the dimension tables are related hierarchically one to another the model is called snowflake. It usually should be avoided cause it can be difficult for business user to understand and it can impact negatively queries (Kimball and Ross, 2013).

3 METHODOLOGY

3.1 CRISP-DM Process Model

Cross-Industry Standard Process for Data Mining (CRISP-DM), by Chapman et al. (2000), is a methodology that provides a structured process for planning and executing data mining and analytics projects. Consisting of six iterative phases that guide analysts from initial business understanding to the deployment and maintenance of analytical solutions. Regarding the projects of my internship, CRISP-DM serves as foundation for the methodology adopted. The 6 phases are:

Phase 1: Business Understanding. In this phase the analyst should analyze in detail the business in order to understand its background and core business processes, describing business objectives and their constraint in order to point out criteria for a successful outcome to the project. Project objectives must be both wide business objectives and technical data mining objectives. Ad the end of this phase, the analyst should have produced a project plan including steps and an initial selection of tools and techniques.

Phase 2: Data Understanding. This phase include the collection and gross description of the data. Following explorations can underline distribution of key attributes, results of simple aggregations, properties of significant sub-populations and simple statistical analyses. This first exploration may address directly the data mining goals. Finally general data quality is addressed.

Phase 3: Data Preparation. This part of the process aim to produce a dataset and its description. The analyst should select and clean the data, run constructive operations, integrate from multiple sources and formatting.

Phase 4: Modeling. Modeling phase include choosing a modeling technique, generate a test that check model's quality and validity, building and assessing the model following evaluation criteria. Assessment happen in two phases: first one is technical and often carried out by a data mining engineer, second one is a business evaluation run by a business analyst.

Phase 5: Evaluation. This step tries to understand if the results meet the business objectives and seeks to determine if there is some business reason why this model is deficient. Eventually process is reviewed and, based on this, successive steps are determined. The analyst should decide whether to finish the project and move on to deployment, or whether to do more work or start a new related project.

Phase 6: Deployment. In this step deployment plan is produced summarizing all the steps and how to perform them. The analyst should design a strategy to monitor data mining results and a maintenance plan in order to avoid incorrect usage. At the end of the project the analyst produce a final report that can summarize the project or present data mining results. Finally, the whole project should be reviewed assessing what was done well and what needs to be improved.

3.2 Project Methodology

For the execution of the projects during the internship, the approach used was based on the 6-phase process proposed by Chapman, with some adjustments to the steps and to the focus of the tasks. These adaptations allow to better align the methodology with the specific characteristics of dashboard development and data visualization projects.

The Business Understanding and Data Understanding phases will follow the CRISP-DM approach (Chapman et al., 2000) and are also consistent with the POST-DS methodology proposed by Costa and Aparicio (2020), which emphasizes the role of structured processes and stakeholder involvement in data-driven projects. In this context, particular attention is given to internal stakeholders and data providers collaboration and communication in order to define KPIs, identify relevant data sources and problems. This will allow alignment between technical capabilities and business needs. In the context of my projects, this involved analyzing ICA's regulatory role and processes, talk with employers, then collaborating with NTT Data to identify and collect the right data from the HAL platform, each representing different aspects of the obligation management process.

In the Data Preparation phase the focus shifts from extensive data cleaning to data modeling. Since data sources are often pre-validated institutional databases, data are already clean and ready to be modeled to enable dashboard designing. As a result, this phase focuses on structuring and modeling data to support dashboard development, designing relational data models based on institutional business logic or conventional dimensional schemas, depending on the analytical objectives and the available datasets.

The Modeling phase will be significantly different from the standard CRISP-DM procedure. This phase will focus on dashboard design and development. Activities include creating visual representations, implementing filters and drill-down functionalities and ensuring visual consistency with already existing dashboards. I designed and structured my dashboards and pages in order to serve different user groups: management and operational users. Visuals followed existing ICA dashboard design and colors to ensure consistency and familiarity.

Evaluation will involve user testing and feedback collection rather than statistical model validation. Dashboards are shared with stakeholders to assess whether visualizations meet their requirements and needs. Feedback was collected from both operational and management users, so that refinements were applied.

The Deployment phase includes making dashboards available to the final users, providing documentation or legends to facilitate adoption.

4 PROJECT EXCECUTION: FROM DATA TO INSIGHTS

4.1 Project 1: Investment Obligations

4.1.1 Context and Objectives

The first dashboard project focused on investments obligations. Portuguese cinema legislation establish a set of investment obligations to all those companies whose revenue exceed a certain amount of euros. ICA is responsible for checking the income declaration, calculating the obligations and monitoring their execution. Companies can decide of not invest directly in the sector but to pay ICA an indirect investment with the same value of the obligation.

Before this project there was no dedicated visualization tool on this matter. Giving this context it is crucial to give ICA a tool to keep track of the companies subjected to the obligations and to monitor investments execution. The dashboard aim to support both management and operational units, providing operational users with monitoring capabilities and offering managers and executives an aggregated overview to support planning and decision making.

4.1.2 Data and Relational Model

The operational users upload all the information and documents regarding investment obligation on HAL and the data are kept by NTT Data. I collected 9 datasets in csv format ready to be upload and modeled on Power BI:

1. Entities: contains information about companies that are registered in HAL that are subject to investment or fees obligations, or that have already submitted communications or reports.

2. Communication: contains information about communication of revenues which determine the amount to invest.

3. Communication by market segment: contains revenues by market segment.

4. Reports: contains information about reports through which companies demonstrates to ICA whether or not it has fulfilled its investments obligations during the relevant period.

5. Reports by modality and sub-modality: contains investment information by modality and sub-modality.

6. Modality list: reference table for investment modalities.

7. Sub-Modality List: reference table for investment sub-modalities.
8. Works: information on works financed within a specific investment report.
9. HAL works: works that are registered in HAL

Since the datasets were extracted directly from the HAL platform, they were already cleaned and validated. The modeling process was performed in Power BI following the logic of institutional business processes, rather than a traditional star schema. Two main branches has been identify starting from the entity table: revenue communication and investment reports. The rest of the tables will follow the branch to which they belong. Primary and foreign keys were used to define relationships between tables, ensuring data consistency and allowing analysis across the different stages of the process (see Figure 1).

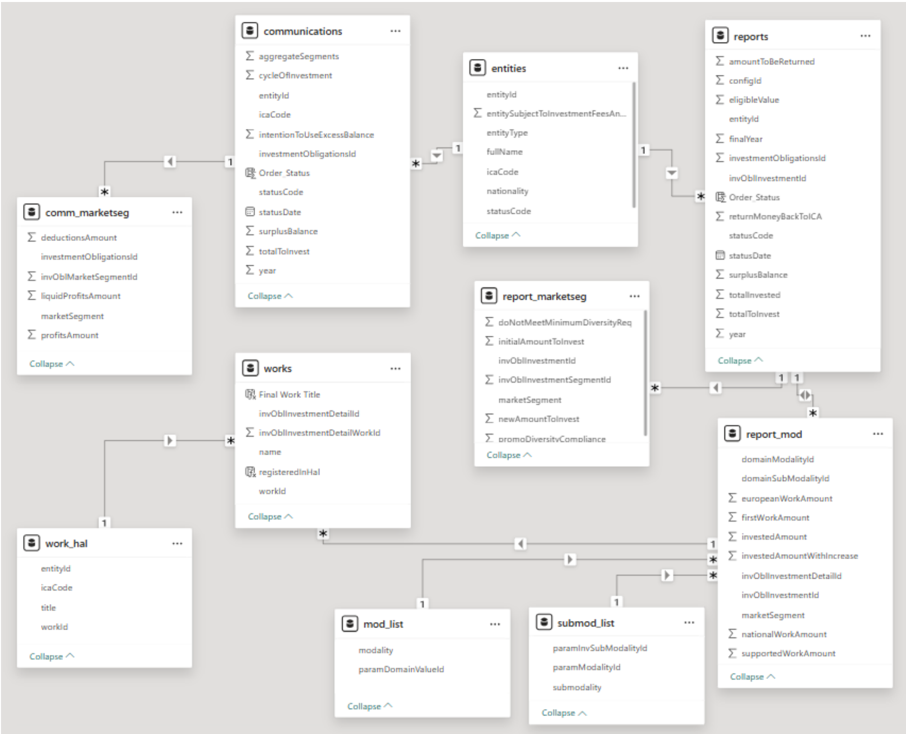

FIGURE 1: Project 1: Relational Model

KPIs were defined with internal stakeholders based on the monitoring and reporting needs of both operational and management units. The dashboards were structured to provide multiple levels of analysis, combining aggregated overviews with detailed exploration capabilities through filters and drill-down functionalities.

4.1.3 Dashboard Design

The dashboards have been organized starting from an overview page (Figure 2) that focuses on entities which main goal is to visualize either if those entities are subject to investments obligations or not. Internal stakeholders (for this page they could be both management and operative users) can visualize all the main measures of investments (communicated, eligible, reported ...) by entity with the possibility of adding filters like nationality, entity type and entity status.

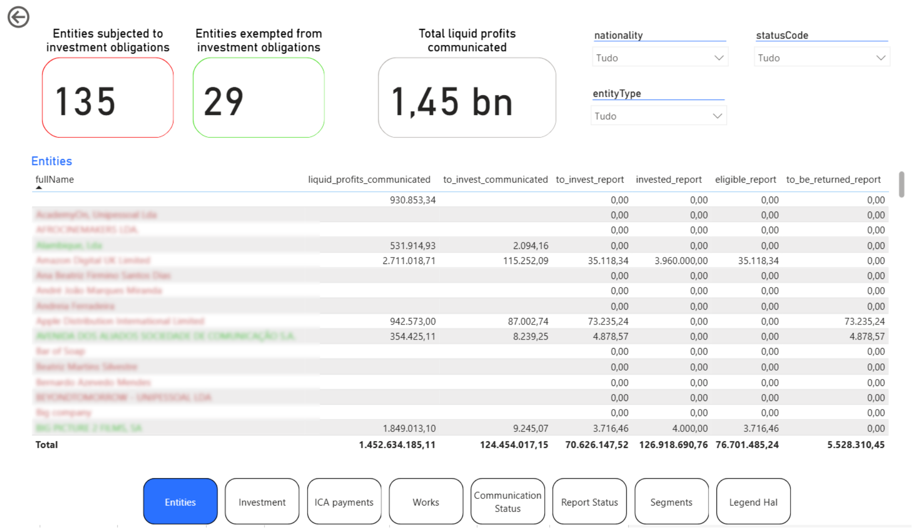

FIGURE 2: Project 1: Entities Page

The following pages are built around the same concepts and designs but focusing on different aspects of the obligations such as investment modalities, eventual pending payments due to ICA following the decision of indirect investments (see Figure 3) and invested works. This pages are designed to be used more by the management, since it can serve to examine financial and core business aspects.

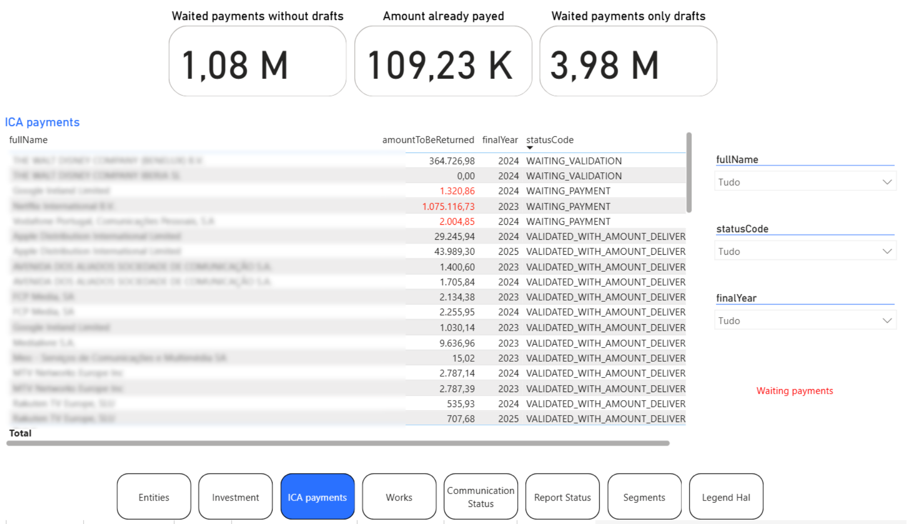

FIGURE 3: Project 1: Waiting Payments Page

The successive 2 pages (one of them in Figure 4) examines the status of communications and reports (draft, waiting validation, validated, waiting payment ...). These dashboards are built for operational users who can keep track on the main workflows of checking and validating the main documents of the obligations.

In order to simplify new users to approach the dashboard a legend has been added that explain where to find the measures of the dashboard on HAL platform.

Regarding visual style, the dashboards were designed in line with existing ICA dashboards made by NTT Data, using same colors and visual elements in order to ensure consistency and familiarity for the users.

Navigation through the pages is enabled through a page indicator at the bottom of the report, allowing users to move easily between sections.

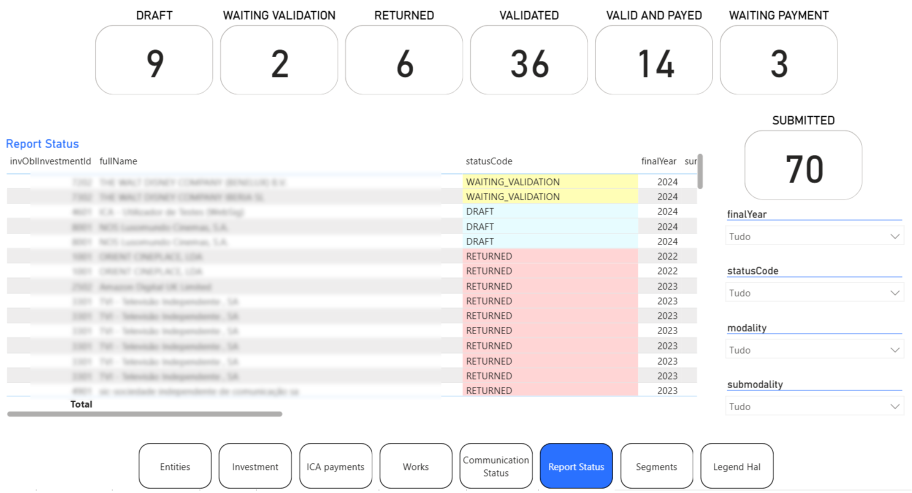

FIGURE 4: Project 1: Report Status Page

4.2 *Project 2: Cinema Industry*

4.2.1 *Context and Objectives*

The second dashboard project regarded market intelligence data provided by Ampere Analysis, a specialized provider of cinema and audiovisual industry insights to which ICA maintains an active subscription. Ampere provides its own visualization interface that include a global datasets on the industry. However, the available information often exceeds ICA's specific needs and interest, making it difficult for users to quickly access relevant insights and navigate through relevant data.

The main objective was to develop focused dashboards that filter and present only the data that fall into ICA's area of interest. A critical aspect of the project involved selecting which data to utilize from Ampere's extensive datasets and determining how to organize and present this information in the dashboards.

The project served two strategic purposes. First, the dashboards were designed to support ICA's evaluation of optimal subscription strategy with Ampere, potentially enabling a transition from continuous annual access to periodic data acquisitions with the final goal of cutting those costs. The project's second goal is to integrate a data-driven culture in the organization by making market intelligence accessible and actionable for both management and operational users.

4.2.2 Data and Relational Model

Data were downloaded from Ampere client area in a singular big spreadsheet. Data included the following measures that have been structured in a fact table: revenue, Revenue Generating Units (RGU), Average Revenue Per User (ARPU), net additions, and market penetration. Dimensions were the following:

1. Time: by quarters. Spanning from 2005 to today, and including a forecast till 2030.

2. Country: 103 countries.

3. Region: world divided in 7 parts

4. Company: 910 companies.

5. Business model: includes pay-as-you-go, subscription, ad-free subscription, ad-tier subscription, free, public, advertising, retail, rental.

6. Business line: includes mobile telephony, fixed telephony, fixed-line broadbend, payTV, subscription OTT, free TV, public TV, VoD and FAST advertising, TV advertising, physical video, theatrical, display advertising, other advertising, other online advertising, other online video advertising and transactional advertising.

7. Technology: this dimension was in the original Ampere study and included different technologies like satellite or terrestrial. The dimension has been completely omitted in this project because not considered relevant for ICA's point of view.

After analyzing the dimensions and ICA's needs I decided to divide the project into 3 dashboards. The first one explores Premium Video Services. It includes payTV, subscription OTT business lines and ad-free subscriptions, ad-tier subscription, subscription business models. The second one focuses on Advertising business model and includes advertising business model and display a display advertising, other advertising, other online advertising, other online video advertising, transactional advertising and tv advertising business lines. The third one is for traditional business and includes free, public, rental an retail business models and free tv, physical video, public tv and theatrical.

In all three cases the tables are been modeled in a classical star schema, with the fact table in the center and all the dimensions around and related to it.

4.2.3 Dashboard Design

All the dashboards have been designed following the same colors and visuals of project 1 and the other existing dashboard from ICA.

The premium video service dashboard starts with an overview page where internal stakeholder can choose the measure through a measure selection navigator (top left in figure 5). It then display the trend from 2005 to 2030, the top 10 companies by the measure value selected and the top 10 companies by YoY growth. The following pages are more detailed and they regard countries and companies performances, business models segments, companies' market share by country (fig 6) and forecast.

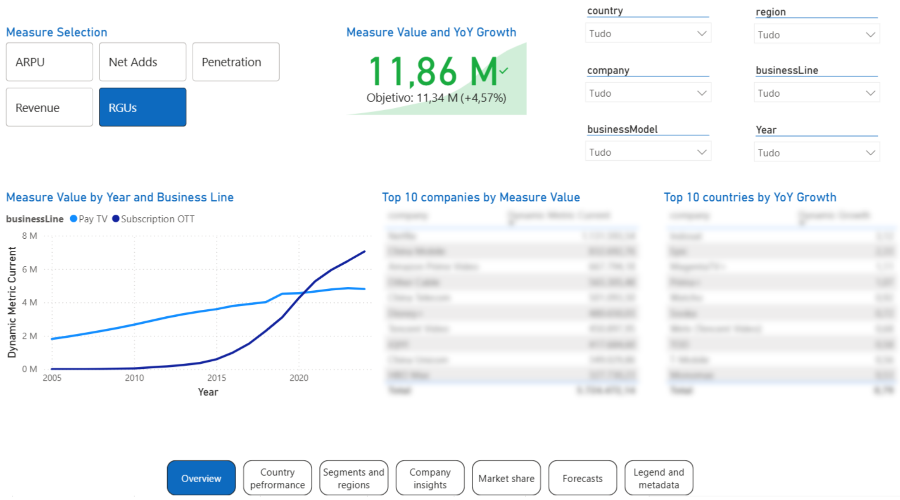

FIGURE 5: Project 2: Overview Page

The second dashboard regard advertising revenues. It follows the same structure and logic of the first one in order to guarantee consistency and usability by alle the stakeholders. After the overview page, showing trend by business line and top 10 companies by revenues and YoY growth, companies and countries breakdowns are displayed. One page is dedicated to the forecast (fig 7).

Finally, the last dashboard is dedicated to the traditional businesses like retail and theaters. The overview page present a revenue trend by business line and some more detailed visuals by countries and companies. The following page present the forecast. The last two pages are dedicated to free tv insights, showing both an overview and a the forecast.

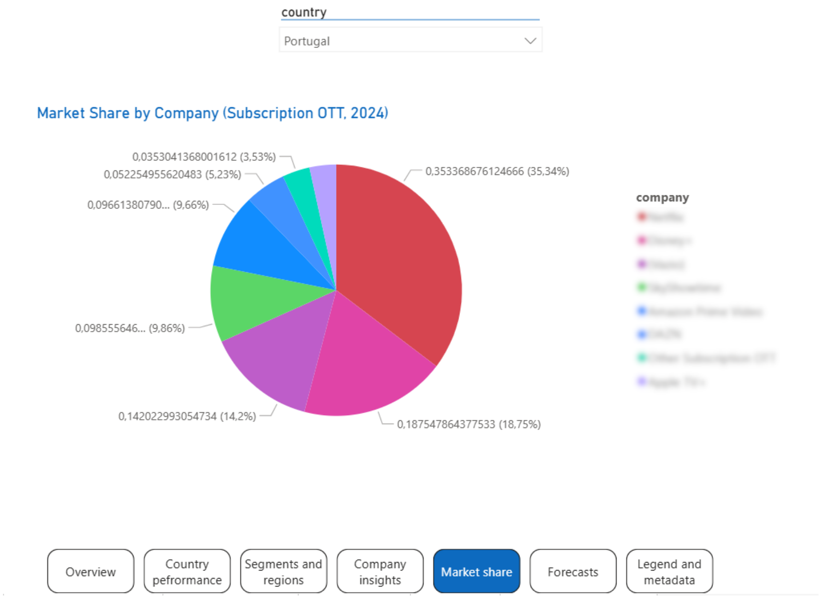

FIGURE 6: Project 2: Market Share Page

All three dashboards have a legend page explaining measurements units, facts and dimensions used. In all pages users have the possibility to apply filters for the right dimensions. Navigation through pages is always possible thank to a navigator on the bottom of all pages.

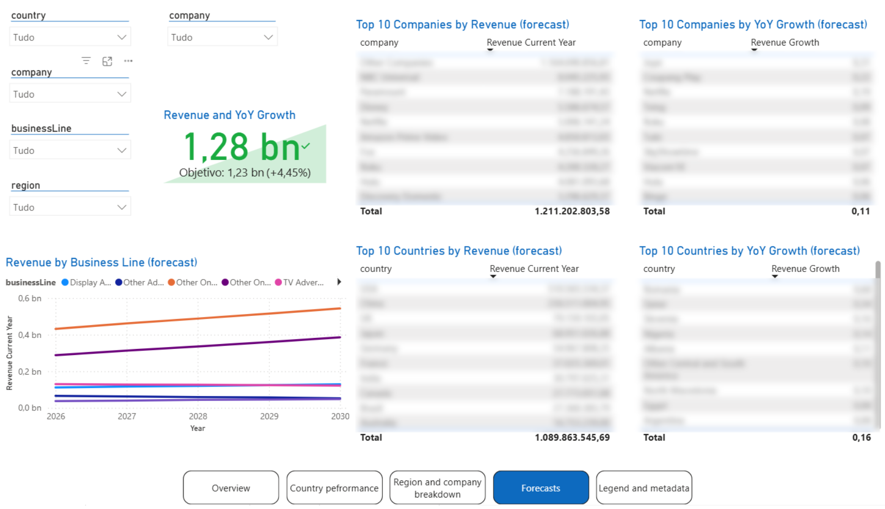

FIGURE 7: Project 2: Forecast Page

5 CONCLUSION

During the internship two projects including four total dashboards have been provided. The first project addressed investment obligations monitoring and the second one provided market intelligence insights from external industry data. The dashboards were designed following user-centered approaches, with KPIs defined collaboratively with stakeholders and visual designs consistent with existing organizational standards.

One of the main challenge regarded the organizational context, characterized by the absence of an internal data specialist team. Decision making regarding data modeling and visualization choices required indipendent judgment and some analytical problems were addressed repeatedly without an internal knowledge base, occasionally slowing down progress.

The two projects also underlined important differences between working with internal and external datasets. The first project was based on ICA's HAL platform data, so it included pre-validated datasets aligned with institutional processes. The cinema industry project required more effort in the data understanding phase to assess data structure, identify relevant subsets, and determine their applicability to ICA's specific context. These differences affected how long each project took and how much work was needed to make the data actionable.

The work performed in the context of this master's final internship allowed for the identification of organizational considerations related to ICA's current approach to data management and analytics. The strong reliance on an external technology provider for analytical development translates into some structural limitations. Any modification to existing dashboards or new analytical requests requires coordination with external actors, resulting in additional time and cost constraints. From an organizational perspective, the development of basic internal business intelligence competencies could reduce these costs. Such an approach would not replace the role of the external provider but would increase autonomy in managing routine analytical needs and support the gradual strengthening of a data-driven culture within the organization.

References

Ackoff, R. L. (1989), 'From data to wisdom', *Journal of Applied Systems Analysis* **16**(1), 3–9.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000), Crisp-dm 1.0: Step-by-step data mining guide, Technical report, SPSS Inc.

Chaudhuri, S., Dayal, U. and Narasayya, V. (2011), 'An overview of business intelligence technology', *Communications of the ACM* **54**(8), 88–98. DOI: `https://doi.org/10.1145/1978542.1978562`.

Choo, C. W., Detlor, B. and Turnbull, D. (2000), *Web Work: Information Seeking and Knowledge Work on the World Wide Web*, Kluwer Academic Publishers. DOI: `https://doi.org/10.1007/978-94-015-9405-9`.

Costa, C. J. and Aparicio, J. T. (2020), Post-ds: A methodology to boost data science, *in* '2020 15th Iberian Conference on Information Systems and Technologies (CISTI)', IEEE, pp. 1–6. DOI: `https://doi.org/10.23919/CISTI49556.2020.9140932`.

Costa, C. J. and Aparício, M. (2019), Supporting the decision on dashboard design charts, *in* 'Proceedings of the 254th The IIER International Conference 2019', pp. 10–15. ISBN: 978-93-89469-21-9.

Cottur, K. and Gadad, V. (2020), 'Design and development of data pipelines', *International Research Journal of Engineering and Technology (IRJET)* **7**(5), 2453–2456.

Davenport, T. H. and Prusak, L. (1998), *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press. ISBN: 978-0-87584-655-2.

Few, S. (2006), *Information Dashboard Design: The Effective Visual Communication of Data*, O'Reilly Media. ISBN: 978-0-596-10016-2.

Inmon, W. H. (2002), *Building the Data Warehouse*, 3 edn, Wiley. ISBN: 978-0-471-08130-2.

Islam, M. and Jin, S. (2019), An overview of data visualization, *in* '2019 International Conference on Information Science and Communications Technologies (ICISCT)', IEEE, pp. 1–7. DOI: `https://doi.org/10.1109/ICISCT47635.2019.9012031`.

Kimball, R. and Ross, M. (2013), *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3 edn, John Wiley & Sons. ISBN: 978-1-118-53080-1.

Luhn, H. P. (1958), ‘A business intelligence system’, *IBM Journal of Research and Development* **2**(4), 314–319. DOI: https://doi.org/10.1147/rd.24.0314.

Negash, S. and Gray, P. (2008), Business intelligence, *in* F. Burstein and C. W. Holsapple, eds, ‘Handbook on Decision Support Systems 2: Variations’, Springer, pp. 175–193. DOI: https://doi.org/10.1007/978-3-540-48716-6_9.

Power, D. J. (2002), *Decision Support Systems: Concepts and Resources for Managers*, Quorum Books. ISBN: 978-1-56720-497-1.

Shneiderman, B. (1996), The eyes have it: A task by data type taxonomy for information visualizations, *in* ‘Proceedings of the 1996 IEEE Symposium on Visual Languages’, IEEE, pp. 336–343. DOI: https://doi.org/10.1109/VL.1996.545307.

Stenmark, D. (2002), The relationship between information and knowledge and the role of it in knowledge management, *in* ‘Proceedings of the 35th Annual Hawaii International Conference on System Sciences’, IEEE, p. 104. DOI: https://doi.org/10.1109/HICSS.2002.994043.

Tufte, E. R. (2001), *The Visual Display of Quantitative Information*, 2 edn, Graphics Press. ISBN: 978-1-930824-13-3.

Ware, C. (2004), *Information Visualization: Perception for Design*, 2 edn, Morgan Kaufmann. ISBN: 978-1-55860-819-1.