THE BUCHAREST UNIVERSITY OF ECONOMIC STUDIES

DATABASE SYSTEMS JOURNAL

Vol. XV/2024

LISTED IN

RePEc, EBSCO, DOAJ, Open J-Gate,
Cabell's Directories of Publishing Opportunities,
Index Copernicus, Google Scholar,
Directory of Science, Cite Factor,
Electronic Journals Library

BIG DATA
NoSQL
DATA SCIENCE
MACHINE LEARNING
BUSINESS INTELLIGENCE
CLOUD COMPUTING
DATA MINING
DATA WAREHOUSES
DATABASES

ISSN: 2069 – 3230
dbjournal.ro

Contact
CaleaDorobanţilor, no. 15-17, room 2017, Bucharest, Romania
Web: https://dbjournal.ro/
E-mail: editordbjournal@gmail.com

CONTENTS

Automating the Generation of Microservice Architectures in Web Applications

Pavel-Cristian CRACIUN[1]
[1]Bucharest University of Economic Studies
craciunpavel18@stud.ase.ro

The advent of microservice designs, which prioritizes enhancing deployment timelines, scalability, and flexibility, marks an advancement period in software development. This article presents a tool designed to accelerate the construction of microservice architectures. Using an intuitive interface, the solution allows users to create fundamental code and graphically construct structures, which streamlines the typically laborious first coding process. Through the automation of project documentation and scaffolding, the solution reduces resource consumption and speeds up development. The proposed solution's complete approach is demonstrated by a comparison with Spring Initializr, which provides a straight path from conceptual design to deployable code. This highlights the potential of the proposed tool to revolutionize software project development.
Keywords: *Microservices, Software Development Efficiency, Automated Code Generation, Architectural Planning*

1 Introduction

The evolution of web applications demands scalable and flexible architectural solutions to handle the complexities of modern software needs. Microservices are an architectural and organisational approach as much as a technological one, having the goal of accelerating software application deployment cycles, encouraging creativity, ownership and improving a software applications maintainability and scalability. Also, productivity, agility resilience are methods that are the result factor of this shift in software development. This type of architecture represents a foundation of scalable and agile systems which supports flexibility in integrating new features, meeting dynamic demands. A microservice architecture's ability to distinguish services allows developers to concentrate on distinct components, which promotes innovation and makes maintenance easier.

The structured process used for planning, creating, and deploying an information system is SDLC (Software Development Lifecycle) this is compounded from distinct phases such as:

- Requirement analysis: identifying and documenting what is required by stakeholders.
- Design: defining the architecture of the software based on the demanded requirements.
- Implementation: Coding based on the system designed.
- Testing: verifying that the software meets all the requirements.
- Deployment: delivering the software to be used by the users.

The current paper focuses on the Design phase from the SDLC that represents how the architecture will be built.

2 From monolithic to microservice architecture

There are two models that can be approached to have a consistent architecture, these are: Monolith and Microservices. The Monolith represents a standalone architecture that highlights a linear process in which the whole logic of the application and services are dependent one of another [1]. From a management point of view is a less need for coordination between different teams, as everything can be centralized in one place. Everything is in a single codebase thus making the testing process easier to approach. for the Deployment process is straightforward since involves only one build and one release process.

There are certain limitations and other aspects that Monolith architecture highlights:

- Scalability issues, where a notable study by G. Blinowski, et al., underscores the debate surrounding scalability in monolithic vs. microservice architectures. Their investigation into the performance dynamics on a single machine highlighted potential scalability challenges when migrating to microservices, especially for smaller businesses with limited concurrent users, posing implications for scalability strategies [2].
- Complexity in making Changes, which spotlights the revitalization of software design through microservices comes as a response to the rigidity found in monolithic architectures, where the complexity of implementing changes can stifle innovation and prolong development cycles. This is substantiated by the work of Milos Milic, et al., who developed a quality-based model emphasizing the cumbersome nature of adapting monolithic systems to evolving software needs [3].
- Slow Start-Up Times, the agility afforded by microservices directly addresses the sluggish responsiveness endemic to monolithic architectures during initialization phases. The comparison provided by Nahid Nawal Nayim, et al., evaluates performance implications in an e-commerce startup setting, shedding light on the substantial lead time reductions achievable with microservice infrastructures [4].

The addressed limitations of monolithic architectures in the design phase in the SDLC advocate for using microservices instead. Microservices architecture involves designing a system as a collection of smaller independent applications, each performing a specific business function.

Microservice architectures represents an advancement in the design and development of web applications [5].

It is evident that the transition from monolithic to microservices architecture is not merely a technical upgrade but a strategic necessity to overcome profound operational challenges [1]. This type of architectural paradigm named Microservices decomposes applications into small, independently, and deployable services with each of their own having their unique business functionality. Terdal’s work on microservices exemplifies this approach, highlighting the transition of e-commerce web applications from monolithic to microservice architecture, emphasizing the benefits of independent development, testing, and deployment facilitated by modern technologies like Docker containers and Kubernetes [6].

The architectural approach of microservices also offers notable financial advantages, particularly in terms of cost-effectiveness when developing large-scale systems [7]. The perspectives of Nada Salaheddin Elgheriani et. al. [8] highlight even more of the many advantages that microservice designs offer, including improved agility, developer productivity, resilience, scalability, dependability, maintainability, and ease of deployment. All these qualities work together to create a strong structure that can handle future demands in addition to meeting company development's present needs. Considering the above listed academic achievements, software development processes might be greatly streamlined by automating the creation and implementation of microservice architectures. Tools like as "Mono2Micro" from IBM [9], which were put out by A. Kalia et al., show how AI may help ease the shift from monolithic to microservice architectures by providing an automated and effective way forward [10]. Developers may quickly prototype and implement services by including automatic code generation approaches, as covered by Gaetanino Paolone et al. [11]. Although, the solution represents more of an aid in transforming existing codebases and primary focuses on migration rather than the initial creation of microservices from scratch.

The proposed tool offers a novel solution by enabling the rapid setup and deployment of microservices, streamlining the development cycle, and reducing the time and resources traditionally required for architectural planning and documentation. Essentially, the revolutionary potential of automating the creation and implementation of microservice architectures is fully captured by this suggested approach. It holds the potential to redefine the standard for software development, particularly in terms of project conception and execution. As we continue to explore this novel technique, our research delves into the theoretical foundations, real-world applications, and potential impacts of the proposed web application on the effectiveness, scalability, and resilience of web application development frameworks. This investigation aims to illuminate how the proposed solution can fundamentally enhance the architectural landscape of software engineering, promising a transformative shift in how developers approach project development.

3 Efficient Automation of Microservices in Web Applications

In contrast, this paper proposes a solution designed to pioneer the development of microservice architectures from the ground up. As a web application with a user-friendly UI, the solution proposes a solution for creating and interconnecting microservices through the interface. Also, another important objective would be to generate a comprehensive foundational project that includes all necessary dependencies and microservices code. This tool reduces the need for extensive design and planning meetings, allowing developers to directly construct and visualize their architecture through an intuitive UI diagram. Each microservice can be customized based on the goal that it serves and can be documented directly in the interface, where notes and details can be added to enhance clarity and maintainability. This approach not only streamlines the development process but also ensures that all aspects of the system architecture are well-documented from the outset, significantly enhancing the efficiency and effectiveness of project development.

The approach is characterized as empirical research aimed at identifying ways to automate microservices. It was intended to thoroughly analyse the automation of microservices architectures in comparison to with other existing tool, Spring Initialzr. Various strategies were employed to focus on analyse and demonstrate the efficiency of automation in the microservice architecture development process. The research was conducted with the aim of identifying, testing, and documenting the most effective practices to reduce time and resources in the software development lifecycle. This involved a combination of theoretical study and practical application to assess the impact of automation on project efficiency and error minimization.

3.1 Beyond existing solutions

The proposed solution represents a new step in the software development by simplifying the creation of microservice architectures. Through an intuitive and friendly user-interface, developers can visualize and design their system using a diagrammatic approach, which fosters a clear visual understanding of service interconnectivity. The application streamlines the process of setting up microservices by allowing users to add dependencies directly within the UI, thus eliminating the traditional back-and-forth between planning and execution. Going beyond mere diagramming, the proposed tool generates the corresponding microservices code, complete with specified dependencies, thus significantly accelerating the development workflow.

There are multiple solutions that share the same purpose as this paper such as:

- Jhipster [12], a development platform, that generates, develops, and deploys Spring Boot web applications and Spring Microservices.
- Telosys [13], a lightweight code generator capable of creating various types of applications (Web, CLI, etc.) from a simple model defined with text files or database schema.While flexible, Telosys might not offer as much depth in handling complex microservices architectures.
- Jmix [14], an instrument that provides high-level tools for enterprise development with Spring Boot. Specialised in rapid development capabilities but might not offer the same level of microservices-focused functionalities.

Spring Initializr [15] was specifically chosen for comparison due to its widespread recognition and focused utility in bootstrapping Spring-based projects efficiently, making it highly relevant to developers familiar with Spring ecosystems. This context helps highlight the proposed solution's enhancements over traditional methods like those provided by Spring Initializr. A straightforward through the Internet interface is offered by Spring Initializr to create Spring-based projects using pre-made templates. Through the creation of build files and project structures, it provides developers with a rapid start. However, as Table 1 illustrates, its usefulness, is limited to the initial setup and requires human implementation of the comprehensive architectural design and inter-service communication.

Conversely, the proposed solution encompasses a wider range of functions by not just starting projects but also streamlining the complete architecture design process. The suggested approach emphasises the complete perspective of the system architecture by enabling the thorough designing of interconnected services, in contrast to Spring Initializr, which mainly provides project scaffolding. It bridges the gap between initial setup and full-fledged architectural development by providing teams with a more integrated solution.

Table 1. Comparison Solutions Table

Feature	Spring Initialzr	Proposed Solution
Primary Function	Bootstrapping Spring-based projects	Comprehensive microservice architecture creation
User Interface	Simple web-based interface	Intuitive and diagrammatic user interface
Automation Level	Generates basic project structures and build files	Automates microservice code and dependency management

Developing strong software architectures requires meticulous attention to detail, starting with the creation of diagrams and documentation of architectural designs and each microservice's functionality. Research indicates that these preliminary tasks are naturally slow paced and resource-intensive, even though they are essential for comprehending and growing applications [16][17]. Developers expend substantial effort not only in conceptualizing the structure and interaction of services but also in translating these conceptualizations into actionable code which represents difficult, redundant procedure that demands for creativity. While, Spring Initialzr provides a solid foundation for initiating Spring-based projects, the proposed solution, extends far beyond the initial setup. It integrates the entire design and development, saves time and enhances quality which makes it superior.

3.2 Web prototype of the proposed solution

The solution proposed is designed to assist development teams, to avoid the usual obstacles related to the architectural foundation stage. Conventional methods, such as lengthy design and planning meetings with iterative cycles to finalize an architecture, often delay project kick-offs. But these chores would become much simpler when the offered solution would be used. As illustrated in Figure 1, the solution does more than just visualisation of the project architecture: it also converts the visualisation into deployable code. This guarantees that teams leave the design and planning session with more than just conceptual schematics. Instead, they have a workable code foundation that can be expanded upon. The tool is essentially removing the need for fundamental work by automating the creation of microservices infrastructure, freeing up teams to focus on product development and business logic execution. Leaping ahead of the early setup stages results in increased productivity, a shorter time-to-market, and more efficient use of resources.

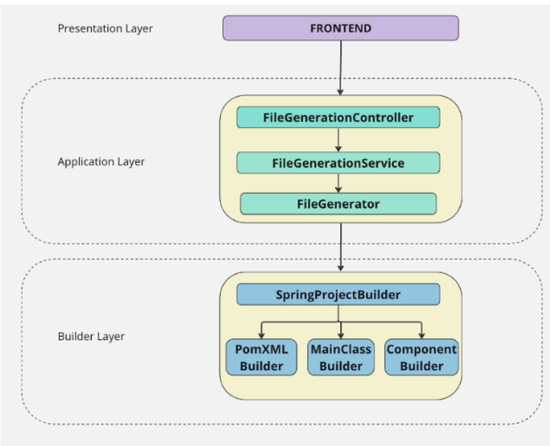


Fig. 1. Solution Architecture

The application is built using Java 17 and employs a microservices architecture. One of the microservices is designed to process JSON inputs, storing templates that represent snippets of code such as classes, records, and interfaces. These templates are then overlaid with additional templates, such as annotations, and transformed into strings. These strings are subsequently modified with parameters derived from the JSON input, utilizing a builder pattern to construct these values and inject them into the codebase. This process allows for dynamic code generation based on the HTTP Request parameters.

Upon accessing the suggested application, developers are prompted to create a new project. Also, the user will have the possibility to edit in the future the project that was created.

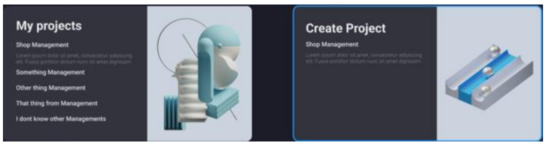


Fig. 2. Main Menu

What can be seen in Figure 2 is an approach of the application to set the stage for a customized microservice architecture. Upon project creation, users will articulate essential metadata, including project's title, the preferred programming language, and an enumeration of global dependencies that will underpin the subsequent architecture. This initial stage is critical as it establishes the project's core framework. Ensuring that a sensible technological and structural baseline is followed by all upcoming microservices.

After creation, the user is redirected to a dashboard that takes the stage for a future placeholder of the microservices that will be connected through diagramming. This represents a centralized core and an orchestrator of the project, which grants the user an oversight over their microservice ecosystem. As shown in Figure 3, the dashboard provides sections for adding new microservices or templates. This is the place where the design, management and configurational utilities come together to provide users with the resources they need so that at the end could have a fully functional project. As new microservices are created, users can attach detailed notes to each one, documenting specific functionalities and architectural decisions. This ensures that each microservice's role and configuration are clearly understood and maintained through the project lifecycle.

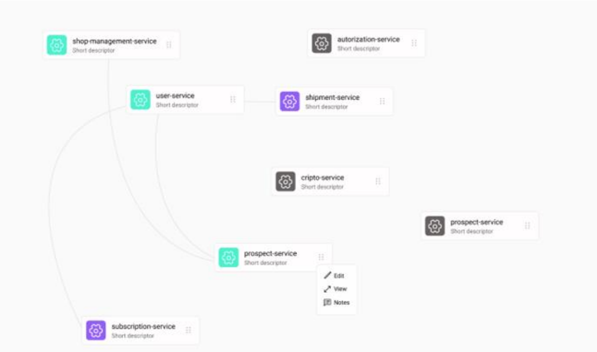


Fig. 3. Dashboard diagram

The "Add Service" option on the dashboard takes users to a **customised design area** where they may access the microservice architecture's granularity. Illustrated in Figure 4, users have the freedom to choose from creating a service from scratch, fully customized by them or to choose from an existing template such as Login Service, User Details service and many more.

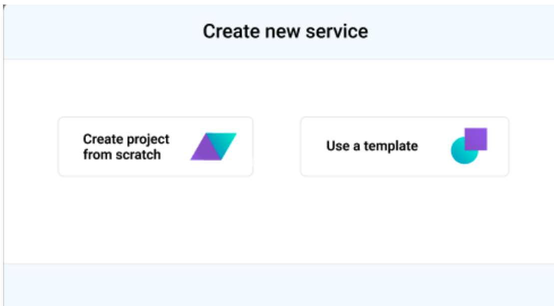


Fig. 4. Create service

These templates are services that are usually present in every application. If the users select to create a new service, the application will grant them freedom to determine the characteristics of the services, such as programming language, nomenclature, and whether to build a custom entity from scratch or instantiate the service from a desired template. Because of its adaptability, every microservice guaranteed to be both precisely functional and to be in harmony with the project's overall architectural philosophy.

One other example of the proposed solution's strength is its ability to specify deep services. As a component of a broader architectural process, every microservice requires careful description of classes, methods, and their interactions. This is made possible by the user-friendly, diagrammatic interface of the suggested solution, displayed in Figure 5, which gives developers the ability to see and control the operational dynamics of any service component.

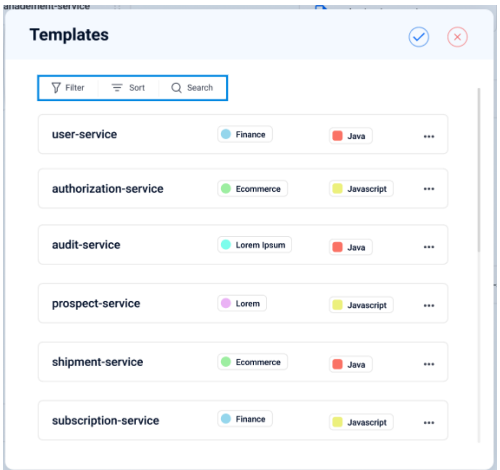


Fig. 5. Create template

One important aspect that the tool provides is the possibility to incorporate dependencies for each microservice, ensuring that each has the necessary tools and libraries for optimal functionality, as seen in Figure 6.

With the proposed solution, developers may examine the complete project layout after a thorough architecture has been built and documented. Users can iterate and revise their design during this comprehensive review phase, which is crucial for assuring alignment with project objectives and functional needs. The suggested tool stands out for being able to provide a codebase that can be used. This essential feature eliminates the need for manual coding by converting the graphically represented architecture straight into an organised, deployable project structure that includes all necessary microservices and dependencies.

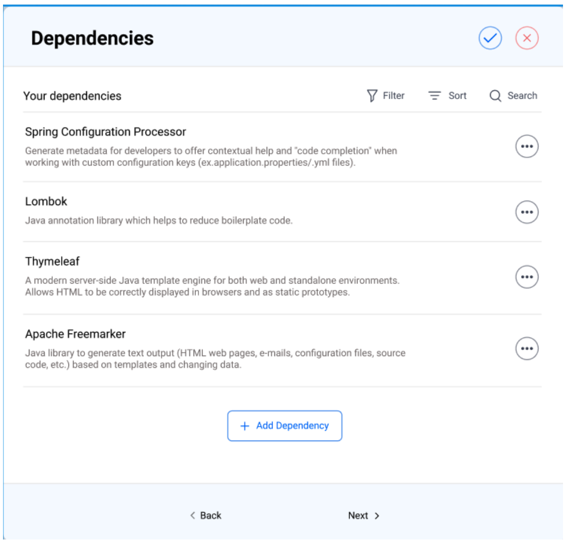


Fig. 6. Add dependencies

At the end, the tool delivers the users a fully deployable code that will only need focus on writing the business logic of the application/product. Through this integration of design, documentation and automated code generation, development teams will help other teams of management and business analysts to understand and have a broader view on the technical landscape which will reversely make the development team be more focused on the product and business side.

4 Conclusion

All things considered, the proposed solution rethinks the architectural foundation of microservice-based apps, transforming a labour-intensive procedure into one that is automated, effective and minimises errors. Looking through the prism of studies that highlights how time- and resources-consuming traditional design methods are, the suggested solution becomes more than simply a tool, and it's a paradigm change. By removing the barrier that separates design from implementation, it enables development teams to jump straight to the process of integrating business logic into precisely created, thoroughly documented code bases. For future enhancements, the proposed solution aims to further boost efficiency and reduce costs by incorporating additional automation across the Software Development Lifecycle, including:

- Infrastructure Automation: Enhancing the solution to automate infrastructure generation could significantly improve system portability through advanced containerization techniques. This would align with modern development practices and address scalability and deployment challenges [18].
- Monitoring Automation: Extending automation to include the monitoring of previously created microservices. This would facilitate proactive management and maintenance, ensuring high availability and performance consistency [19].

In contrast to Spring Initialzr, which mainly facilitates project initialization, the offered solution advances the concepts of agility, speed, and accuracy that form the foundation of contemporary software engineering, therefore capturing the spirit of innovation in software development. The proposed solution facilitated the structured creation of microservices architectures, providing extended functionalities for visualization and code generation. These elements have significantly enhanced the efficiency of the project development process, which confirms the achievement of the established objectives.

References

[1] Kalske, M., Mäkitalo, N., Mikkonen, T. (2018). Challenges When Moving from Monolith to Microservice Architecture. In: Garrigós, I.,

Wimmer, M. (eds) Current Trends in Web Engineering. ICWE 2017. Lecture Notes in Computer Science, vol 10544. Springer, Cham. https://doi.org/10.1007/978-3-319-74433-9_3

[2] G. Blinowski, A. Ojdowska and A. Przybyłek, "Monolithic vs. Microservice Architecture: A Performance and Scalability Evaluation," in IEEE Access, vol. 10, pp. 20357-20374, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3152803.

[3] Milić, Miloš, and Dragana Makajić-Nikolić. 2022. "Development of a Quality-Based Model for Software Architecture Optimization: A Case Study of Monolith and Microservice Architectures" *Symmetry* 14, no. 9: 1824. https://doi.org/10.3390/sym14091824

[4] N. N. Nayim, A. Karmakar, M. R. Ahmed, M. Saifuddin and M. H. Kabir, "Performance Evaluation of Monolithic and Microservice Architecture for an E-commerce Startup," *2023 26th International Conference on Computer and Information Technology (ICCIT)*, Cox's Bazar, Bangladesh, 2023, pp. 1-5, doi: https://doi.org/10.1109/ICCIT60459.2023.10441241.

[5] A. Jan, "What is Microservices Architecture," Sumerge, 09-Nov-2020. [Online]. Available: https://www.sumerge.com/what-is-microservices-architecture/.

[6] Dr. Sujata Terdal, Prasad R G, Vikas Mahajan, Vishal S K, "Microservices Enabled E-Commerce Web Application", International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue VII July 2022, doi: https://doi.org/10.22214/ijraset.2022.45791

[7] I. Papakonstantinou, S. Kalafatidis and L. Mamatas, "A Techno-Economic Assessment of Microservices," 2020 16th International Conference on Network and Service Management (CNSM), Izmir, Turkey, 2020, pp. 1-5, doi: 10.23919/CNSM50824.2020.9269114.

[8] Nada Salaheddin ELGHERIANI, Nuredin D Ali Salem AHME, "MICROSERVICES VS. MONOLITHIC ARCHITECTURES [THE DIFFERENTIAL STRUCTURE BETWEEN TWO ARCHITECTURES]", *MINAR International Journal of Applied Sciences and Technology, 2022, doi:* https://doi.org/10.47832/2717-8234.12.47

[9] https://www.ibm.com/products/cloud-pak-for-applications/mono2micro

[10] Anup K. Kalia, Jin Xiao, Chen Lin, Saurabh Sinha, John Rofrano, Maja Vukovic, and Debasish Banerjee. 2020. Mono2Micro: an AI-based toolchain for evolving monolithic enterprise applications to a microservice architecture. In Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2020). Association for Computing Machinery, New York, NY, USA, 1606–1610. https://doi.org/10.1145/3368089.3417933

[11] Paolone, Gaetanino, Martina Marinelli, Romolo Paesani, and Paolino Di Felice. 2020. "Automatic Code Generation of MVC Web Applications" *Computers* 9, no. 3: 56. https://doi.org/10.3390/computers9030056

[12] https://www.jhipster.tech

[13] https://www.telosys.org

[14] https://www.jmix.io
[15] https://start.spring.io
[16] Przemysław Jatkiewicz and Szymon Okrój. 2023. Differences in performance, scalability, and cost of using microservice and monolithic architecture. In Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing (SAC '23). Association for Computing Machinery, New York, NY, USA, 1038–1041. https://doi.org/10.1145/3555776.3578725
[17] R. Mishra, N. Jaiswal, R. Prakash and P. N. Barwal, "Transition from Monolithic to Microservices Architecture: Need and proposed pipeline," 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 2022, pp. 1-6, doi: 10.1109/INCOFT55651.2022.10094556. keywords:
[18] P.-C. Craciun and S.-C. Necula, "Theoretical and Applied in Automating Kubernetes Resources", Informatica Economică vol. 27, no. 2/2023, 2023, pp. [36-45], doi: https://doi.org/10.24818/issn14531305/27.2.2023.04
[19] D. Boncea, M. Zamfiroiu, and I. Bacivarov, " A scalable architecture for automated monitoring of microservices", Economy Informatics vol. 18, no. 1/2018, 2018, pp. [13-22].

Pavel-Cristian Craciun completed his undergraduate studies at the Faculty of Economic Computation and Economic Cybernetics in 2021 and graduated from the master's program in Informatics Systems for the Management of Economic Resources at the Academy of Economic Studies. He is currently enrolled as a first-year PhD student in Economic Informatics. Professionally, he excels as a DevOps engineer, focusing on developing, maintaining, and monitoring software applications.

Automation in Financial Reporting: A Case Study

Oana-Alexandra DRAGOMIRESCU[1], Adriana-Teodora PARSCHIVOIU[1],
Andreea VINEȘ[1], Andrei NICA[1]
[1]Bucharest University of Economic Studies
dragomirescuoana18@stud.ase.ro, paraschivoiuadriana19@stud.ase.ro,
andreea.vines@csie.ase.ro, nicacatalin19@stud.ase.ro

Financial Reporting is the key in understanding the financial information of an organisation in a clear and organised way. With the advent of Artificial Intelligence (AI), there has been a significant shift towards automation due to its numerous benefits, including improved data quality and integration, cost and time savings, scalability, flexibility, and enhanced operational efficiency. Financial reporting is one area that has particularly embraced these advancements.
This article explores the necessity of automation in financial reporting, focusing on the use of the automation tool Alteryx with the AnaCredit dataset. It examines the outcomes of incorporating automation into daily financial reporting practices, demonstrating the tangible benefits and improvements achieved through this technological integration.
Keywords: *business process automation, financial reporting, Alteryx automation*

1 Introduction

The technological world is evolving rapidly and imposes on companies the need to adapt to these changes quickly, not only to keep up but also to implement and use new technologies to gain a competitive advantage. In the past, computers and software solutions were considered an aid for businesses, but due to digital transformation, they are now at the forefront of innovation, being a major force in the growth of companies. [1]
In the finance industry, numerous areas still remain nowadays without automation. According to an SAP study, some processes with the highest automation opportunities include Financial Reporting, Accounts Receivables, Payroll and Cash management. [2] Also, a study conducted by KPMG states that when discussing about the current state of automation, 59% of the respondents believe that in financial reporting there is very little automation and only 3% states that there is significant automation. [3]
Also, according to a survey by the Institute of Management Accountants (IMA), financial reporting is the area most expected to change significantly or completely over the next five years due to technology evolving. Automation has enabled many companies to streamline the process of producing financial statements. Once set up, the platform can generate monthly financial statements efficiently. This capability contributed to the average financial reporting time for U.S. businesses dropping to just 10 days in 2019, three days less than in 2018, according to Robert Half [4]. Remarkably, 39% of organisations with revenues under $500 million have automated their financial report creation.
Automated financial reporting ensures accuracy, transparency, and credibility, making it useful for various business needs—from bank reports and investor due diligence to earnings reports. This automation frees financial professionals to focus more on analysing results and allows both internal and external stakeholders to receive critical information sooner, aiding in decision-making. [5]
We can mention that automation brings not only operational and financial benefits but also improves employee well-being by reducing repetitive and monotonous tasks. These tasks are much faster and safer when done through such a process. [6]

This article is based on a case study that aims to demonstrate the need and efficiency of process automation in financial reporting. The structure is divided into four main topics: literature review, methodology, the case study, and conclusions. First, the focus is on the need for automation in the financial sector, particularly in reporting activities—why it is necessary, what has been done, and what could be done. The methodology section describes the research method used in this paper, along with the process and main steps involved. The third section presents the automation implemented for financial reporting, and lastly, the conclusions are presented.

2 Financial Reporting nowadays

In today's world, companies have understood the necessity of financial reporting and know how to use it rigorously. According to a KPMG poll on financial reporting and artificial intelligence (AI) [7], businesses are embracing cutting-edge technology more often to improve their financial reporting capacities. Using AI to make financial disclosures more accurate and efficient is one example of its multiple use cases. [8]

A GRI research states that 79% of the top 100 corporations in each of 58 nations (N100) and 96% of the world's largest 250 enterprises (G250) report on environmental, social, and governance (ESG) concerns or sustainability. This demonstrates that big firms throughout the world have a high adoption rate of comprehensive financial reporting methods. [9]

International Financial Reporting Standards (IFRS) are required for listed corporations, and regulations like Directive 2013/34/EU have harmonised financial reporting requirements throughout the European Union. By ensuring that businesses operating inside the EU adhere to a uniform framework for financial reporting, these laws help to make financial statements clear and comparable across national borders. [10]

What is more, a study conducted by KPMG on automation of financial reporting and technical accounting states that 90% of the respondents see value in automation of the group financial reporting and three quarters of respondents reported spending over 10% of their time manipulating data to generate insights. In an average work week, this amounts to more than half a day. Consequently, a third of the respondents dedicate over 10 weeks per year to this task. [11]

Additionally, the integration of advanced technologies such as blockchain and machine learning is poised to further revolutionise financial reporting. Blockchain offers enhanced transparency and security for financial transactions, while machine learning can predict financial trends and detect anomalies with greater accuracy.

The results of McKinsey and Company show that artificial intelligence (AI) greatly increases the efficiency and accuracy of financial reporting by automating repetitive operations and providing predictive analytics for tactical decision-making. However, there are several significant obstacles, like the high cost of AI integration, the requirement for trained staff fluent in AI, and data protection issues. The report also emphasises that one major obstacle to AI adoption in accounting processes is reluctance to change. [12]

Automation reduces the possibility of human mistake in reporting and data input, guaranteeing consistent and accurate data submission. Financial reporting leaders have distinct concerns for AI and GenAI. For AI, the focus is on model transparency and data privacy, while for GenAI, accuracy, data management, and bias are primary concerns. Major barriers to AI adoption include reliance on algorithms, the fast pace of changing regulatory guidelines, and data quality. This is essential to achieving the high standards that regulatory organisations demand. [7] In a survey made by KPMG on the theme “Will AI transform financial reporting and audit?”, the conclusions state that 65% of financial reporting leaders report their functions are using AI, 70% of leaders whose companies are already using

AI report they expect to roll out AI solutions more broadly over just the next two years, 47% of financial reporting leaders agree GenAI will deliver on the hype, while 38% disagree and 72% of financial reporting leaders believe external auditors are ahead of financial reporting functions on using AI and expect AI to enhance audit quality.

Financial reporting leaders believe AI will bring numerous benefits, including increased efficiency (51%) and reduced staff burden, more accurate and reliable data (50%), better ability to identify outliers (48%), solutions to staff shortages (27%) and cost savings (25%). [7]

The amount of time and work needed to gather and report credit data is greatly decreased by automated procedures. This way, institutions are able to concentrate their resources on analysis and strategic decision-making instead of manual data gathering and reporting because of this efficiency benefit. [7]

Automation contributes to regulatory compliance by automatically applying the appropriate standards and reporting criteria. This lowers the possibility of fines for non-compliance and improves the institution's capacity for risk management.

Automated solutions may easily interface with the institution's current data sources and processes, enabling thorough data aggregation and more perceptive reporting. Better strategic planning and decision-making are supported by this integration.

Automation may save financial organisations a lot of money by eliminating the need for labour-intensive manual processing and the related expenses. These savings might be used to fund other important initiatives like customer service and innovation. [9]

Without requiring a substantial increase in expenditure, automated systems may readily expand to accommodate growing data quantities and adjust to changing regulatory requirements. Because of its scalability, institutions may develop and grow without encountering significant operational challenges. [13]

3 Research Methodology

The main research methodology approached for this article was instrumental case study, as it uses the case examined to gain insights into the issue presented. The authors wanted to focus on an area with a high opportunity for automation - financial reporting, as previously demonstrated in the article.

Examining the automation of a financial reporting process using a case study method has numerous important benefits. [14] First of all, it offers a thorough and in-depth analysis that makes it possible to examine each stage in detail and the automation's overall impact. By showing how automation is used in a real-world setting, this approach guarantees practical relevance and increases the applicability and usefulness of the findings for experts in the field. Case studies' distinctiveness enables context-specific outcomes that address special possibilities and difficulties within a certain process or organisation. In addition, compared to abstract theoretical models, the material in a case study is more accessible and easily understood due to its narrative structure. Last but not least, a thoroughly documented case study may become a model for best practices, providing a detailed roadmap that other people can use to get comparable outcomes in their own financial procedures. [15]

In **Fig.1**, a UML activity diagram was created to highlight the six main steps used to conduct the case study.

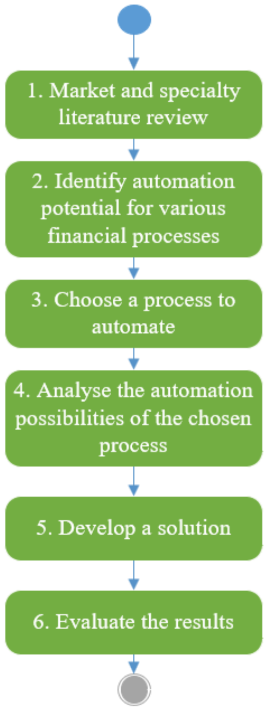


Fig. 1. Activity diagram for case study
Source: Authors' own research

Step 1. Market and specialised literature review
This step involves an in-depth investigation of current trends in financial process automation and the latest publications in the field. It began with a review of the literature by analysing scientific articles, books, and specialised reports to understand the fundamentals and recent developments in financial process automation. Next, a market analysis was conducted to evaluate existing solutions, identify key technology providers, and the available automation tools, such as RPA (Robotic Process Automation), AI (Artificial Intelligence), and ML (Machine Learning). Finally, emerging trends and innovations in the field that can influence or improve the automation of financial processes were identified.
Step 2. Identify automation potential for various financial processes
For the second step, the financial processes within the organisation were evaluated to determine which are most suitable for automation. The main processes within the organisation were identified and documented, such as financial reporting, account reconciliation, invoice management, and others, and the potential benefits of automation for each process were assessed, such as cost reduction, improved accuracy, and efficiency.

Step 3. Choose a process to automate (In this case, a complex reporting process that encounters numerous difficulties was targeted).
The complexity of these processes was then evaluated by analysing the workload, how frequently they are executed, and the degree of repetitiveness, to identify the process that encounters the most difficulties and errors and is the most time and resource-consuming: financial reporting, a process carried out at the end of each month.

Step 4. Analyse the automation possibilities of the chosen process
To identify automation opportunities for the financial reporting process, the current steps of the company were documented, and documentation was carried out by analysing various automation tools and technologies available on the market. Alteryx was used as the automation tool, and the selected reporting type was AnaCredit, a dataset containing detailed information on individual bank loans in the euro area. The reporting process is complex and was chosen because it encounters numerous difficulties observed in daily work. Alteryx is a leading platform for data science and analytics that can assist organisations with data preparation, blending, analysis, and visualisation. Due to its intuitive interface and wide range of tools for advanced reporting, predictive analytics, and data manipulation, users can draw conclusions from their data more effectively. With these elements in place, an automation solution was developed to verify and correct the dataset received from the client, nearly eliminating the need for manual checks.
Step 5. Develop an Automation Solution
The final step in the process involves creating and implementing the automation solution for the selected reporting process. Using the automation tool selected in the

previous step, the solution was developed, and tests were conducted to ensure its correct functioning.

Step 6. Evaluate the results (in this case, a comparison of the proposed solution with the previous method of performing the process)

The purpose of the case study was to observe the effects of automating one of the most frequently used financial processes. To achieve this, three KPIs were established at the end of the study to measure the time saved through automation, the differences in error rates, and the associated costs. The comparison was made between quantitative data from an entity with an average of 750 clients and approximately 3000 financial instruments over a 15-month period (from January 2023 to April 2024) and the output obtained after automation. During this period, the entity conducted the reporting activity both manually and using Alteryx to highlight the differences between the two methods.

4 Automation of AnaCredit Reporting using Alteryx

There are various financial reporting requirements applied in each eurozone country and mandated by regulatory authorities for several reasons, such as: aiding investor decision-making, complying with European-level regulations, assessing creditors, and so forth. However, the primary financial situation analysed is AnaCredit.

AnaCredit is a shared multipurpose dataset containing loan-by-loan information on credit extended by credit institutions to companies and other legal entities. On 18 May 2016 the Governing Council of the ECB adopted Regulation ECB/2016/13 on the collection of granular credit and credit risk data (AnaCredit) establishing Stage 1 of a shared database for the European System of Central Banks (ESCB) as of September 2018. The database contains 88 attributes, updated mostly on a monthly basis, based on harmonised concepts and definitions common to all participating countries.

One of the main purposes of AnaCredit is to provide a detailed and coherent picture of credit exposure and related risks, thereby facilitating financial supervision and macroeconomic stability analysis across the euro area. [16]

To simplify it, AnaCredit is providing a magnifying glass to help analyse credit and credit risk, usually in the EURO area. [17]

In fact, AnaCredit is part of an international model called the "European Reporting Framework" (ERF), which enables all European authorities to have an overview of the banking framework in the European Union.

Since 2018, following this regulation, information regarding credits, corporate financing, and guarantees provided by the financial system has been collected from banks domiciled in the territory, as well as their branches abroad. It is specified that the value of the credit in question for data collection must exceed EUR 25,000. [18]

4.1 Challenges faced in manual AnaCredit reporting processes

Manual AnaCredit reporting processes pose significant challenges for financial institutions. One of the primary hurdles is the labour-intensive nature of data collection and aggregation. Gathering the required data from various sources is a time-consuming task prone to errors. Moreover, manually aggregating data from disparate systems or sources can result in inconsistencies and inaccuracies, undermining the reliability of the reporting process. [19]

Ensuring the quality and integrity of data is another major challenge. [19] Manual data validation and verification processes are susceptible to oversight, especially when dealing with large volumes of data. Without automated checks in place, maintaining data quality becomes increasingly challenging, raising concerns about the accuracy of reported information.

The complexity of AnaCredit reporting requirements further complicates manual processes. Regulatory guidelines are often intricate and subject to frequent updates and revisions. Understanding and interpreting these guidelines manually can be daunting, increasing the risk of non-compliance and regulatory penalties. [20]

Timeliness is also a critical issue in manual reporting. Manual data collection and validation procedures can lead to delays in reporting, jeopardising compliance with regulatory deadlines. Without automated workflows, financial institutions may struggle to meet reporting obligations promptly, potentially facing regulatory sanctions. [21]

The risk of errors and inconsistencies is inherent in manual reporting processes. Human involvement in data entry and manipulation increases the likelihood of errors, such as typos, duplicate entries, and calculation mistakes. Furthermore, inconsistencies may arise in reporting formats and methodologies when different individuals or teams handle the reporting process manually. [22]

Manual reporting processes are resource-intensive, requiring significant human resources in terms of time, effort, and expertise. Allocating staff to manual reporting tasks diverts resources from other strategic activities within the organisation, limiting productivity and efficiency.

Scalability is another challenge faced by manual reporting processes. As data volumes grow or reporting requirements change, manual workflows may struggle to accommodate these shifts effectively. Adding new reporting entities or expanding reporting scope can strain manual processes and increase the risk of errors and delays. [23]

Maintaining a comprehensive audit trail and documentation of manual reporting processes presents additional challenges. Without automated systems to track and document every step of the reporting process, ensuring compliance with regulatory requirements and internal controls becomes more difficult. [21]

In summary, manual AnaCredit reporting processes are fraught with challenges ranging from data collection and validation to compliance and scalability. Addressing these challenges requires the implementation of automation solutions that streamline reporting workflows, improve accuracy, and enhance compliance capabilities. [13]

4.2 Automating example of Financial Reporting

In **Fig. 2**, the workflow of the automation project in Alteryx is depicted. It is structured across multiple levels of data validation in different quadrants. It can be observed that in the first quadrant, CSV Integration Engine-Data INPUT the path where the seven Excel files will be found is specified, and then, based on the file name using the Filter tool, a separate flow is created for each one of them that can be followed in horizontal connections and at every step the before and after status of the files can be checked. It is useful that Alteryx does not overwrite the input documents and keeps each version of them. This way, any changes can be easily observed.

Then the Select Tool focuses on transforming the data into the desired format so that it can be subsequently modelled. For example, initially in the file, the columns containing date information are viewed as String data type, but they need to be converted to Date type to allow comparisons. To process the cells newly transformed into Date type, different formulas are applied to either verify or transform the information according to a set of imposed rules. This is handled by the container called Dates Adjustments Rules Implementation Engine, where we can find the Formula Tool where IF cases are applied for mandatory attributes such as the financial instrument first reporting date (inception date) that obviously should be less or equal to the reporting date. An example of such a condition can be written as: "*IF [inceptionDate]>[reportingDate] THEN [reportingDate] ELSE [inceptionDate] ENDIF*", the structure of the IF clause is simple and the name of the attributes from the condition and output are identified by Alteryx using the square brackets.

The next container, Default Values for Mandatory Attributes Engine, focuses on correcting information and filling in missing information. For these columns, it is mandatory to report various values that are accepted by regulators, so they cannot be missing from the files. If the values are missing then the default values are reported, usually the codification number for these situations means "other" so that it does not affect the quality of the report and it can be changed to a more accurate information if it will be identified in the future, or "Non-Applicable value" can be declared if the element is accepted in the Validation Rules from the regulators.

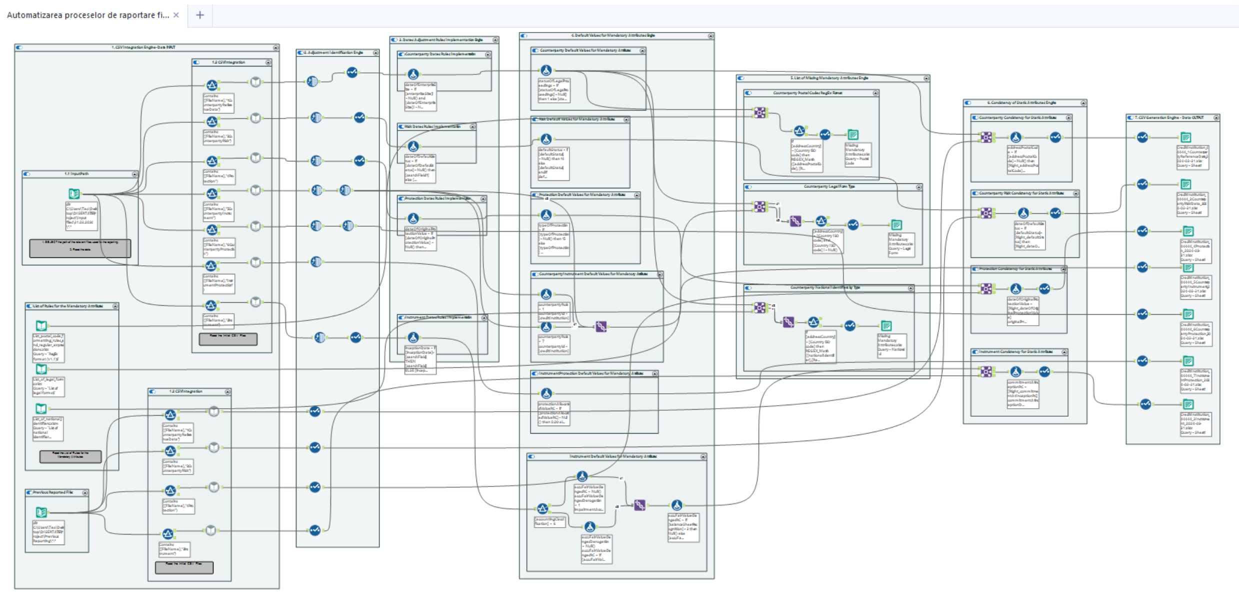


Fig. 2. Alteryx automation workflow

Source: Authors' own research

For the case of correcting the existing data there are several situations based on the category of the declared client, protection, or financial instrument that need to be simultaneously accepted such as for the instruments declared at accounting classification: "IFRS - Financial assets at amortised cost" and balance sheet recognition: "Entirely recognised", for this situation it means the instrument is subject to impairment under the applied accounting standard and in the report it is mandatory for the attributes accumulated Impairment amount, type of impairment and impairment assessment method to be declared and the Non-Applicable values to be used for the accumulated Fair Value Changes Amount. Using Alteryx all these validations and modifications are easily made by using the Filter Tool separating the instruments by their type, the changes are applied in parallel for each case and then all the instruments are united again using the Union Tool.

Then we move on to the fifth container, List of Missing Mandatory Attributes Engine. Within this container, by consulting a list of accepted values and available RegEx formats on central bank websites, the information is verified.

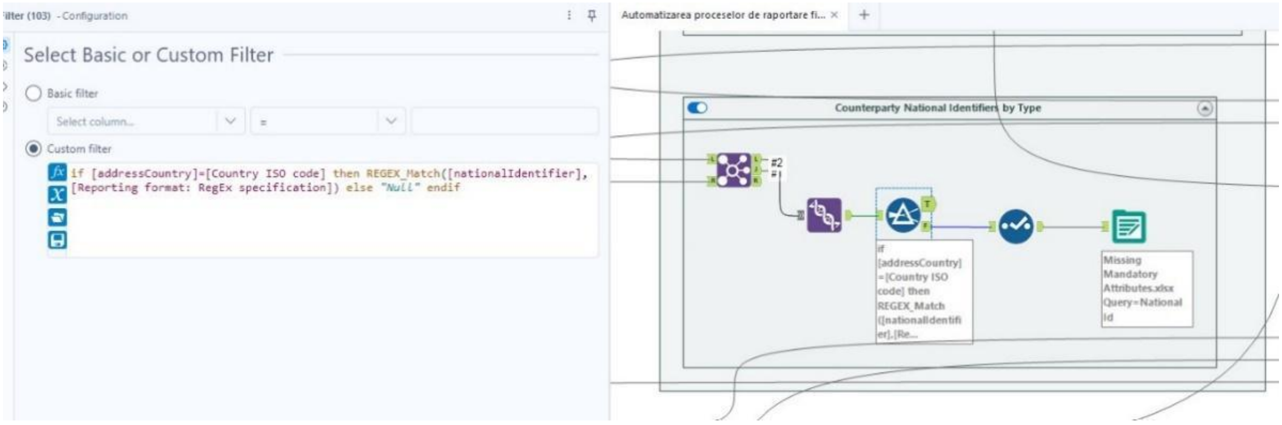


Fig. 3. Counterparty National Identifiers by Type

Source: Authors' own research

This verification is carried out in the previous figure. The latest format of the file, with all the processing done, is taken and then a Join is performed using the country's ISO code (*adressCountry*) and the type of identifier (*typeOfNationalIdentifier*). After this join, there may be unidentified values, so a Union is used to bring in all values, even if there are countries not accepted in the list of National Ids. Then, a new file will be created with only the attributes that do not comply with the format, which will be corrected manually by querying available internal databases or public databases. Although this operation is done manually, the required time is minimal.

In the Excel file, there is a column "Identifier type." Using this, with the "If-clause" structure shown in the figure, the country will be identified and filled in the "Country ISO Code" field, then filtered by type of national identifier, and the formula found in the file for RegEx will be searched, which is filled in the "Reporting format: RegEx specification" column. For example, for Spain, the formula will be *{A-Z0-9}{9}*, and for Japan, it is *\d{13}*. The next column is "Reporting format: description," where the formula from the previous field is described.

Quadrant 6, Consistency of Static Attributes Engine, focuses on static attributes, those that must be reported consistently with the same value, checking their accuracy. For example, two values that always remain the same are the creation date of a financial instrument and the first reported amount. In this step, the comparison is made between the current and previous reports, and if the information is incorrect, then the value from the previous report will be used instead to keep the static attributes from generating errors. The previously reported files are available in the first container, where the path is configured accordingly and using the Join Tool, Formula Tool and Select Tool the current data is compared and corrected.

The final container, CSV Generation Engine – Data OUTPUT, is responsible for saving the resulting files with the Output Data Tool, where the location of the files is configured, also converting the data types back to those accepted by the regulator using the Select Tool. Keep in mind that the corrected files are different from the initial ones and they can be saved to a different location or at the same location with them, but they won't overwrite the initial files.

5 Results and Discussions

At the end, for a better understanding of the benefits obtained using the automation of financial reporting, a comparative analysis was conducted over a period of 15 months (January 2023 – March 2024) between the two reporting methods: manual and automated, within a medium-sized firm. This was done by using a Macro from Alteryx, which compares files to observe differences, and then the data was interpreted in Tableau. [24] The entity for which the statistics were performed consists, on average, of 750 active clients and approximately 3000 financial instruments.

The three situations analysed are as follows:

- The duration of reporting manually versus the duration of reporting using Alteryx;
- The costs involved in manual reporting versus the costs involved in automated reporting;
- The number of errors due to human factors versus the number of errors using Alteryx. Over this period, the entity developed reporting activities both manually and using the Alteryx tool so that the differences of cost, duration and errors could be easily noticeable.

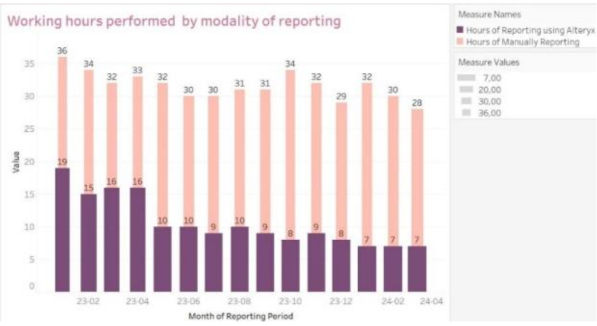


Fig. 4. Working hours performed by modality of reporting from January 2023 until April 2024

Source: Authors' own research

In **Fig. 4**, a comparison is shown between the hours required for manual reporting and those needed for automated reporting. Initially, using Alteryx automation, the duration was longer—reaching up to 19 hours when they began implementing this method in January 2023. However, over time, after employees became familiar with the tool, it decreased to a constant 7 hours for reporting (January, February, March 2024). For manual reporting, a slight decrease in the time required can be observed, from 36 hours in the first month analysed to 28 hours, but comparing the two methods, automation saves significantly more time. Using Alteryx automation for reporting leads to a fourfold increase in speed. Once employees learned how to use this tool, the time required changed significantly from month to month, from 16 hours needed in April 2023 to 10 hours needed in May 2023.

Even though time has passed since the analysis in Fig. 4, we can deduce that time does not significantly improve the manual process. Therefore, this process is much more limited, raising a question mark for companies truly aiming to achieve high performance in their operations.

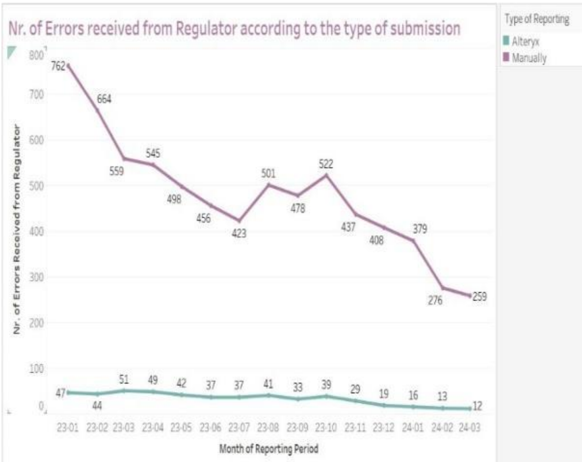


Fig. 5. Number of errors received from regulator according to the type of submission from January 2023 until April 2024

Source: Authors' own research

The figure above highlights the difference between the number of errors in the automated process compared to the number of errors in the manual execution. It can be observed that in both cases the number of errors has decreased, but in manual reporting, although the number has significantly decreased, there is still a large difference compared to the automated method. Automated process errors have consistently been fewer than those made by manual execution, emphasising the reliability and accuracy benefits of automation. By March 2024, there were 259 human errors, compared to 12 errors from automation in March 2023.

It is worth mentioning that most of the errors in automation are due to incomplete data from client files. In some cases, missing, incomplete, or incorrect fields can be filled or corrected, but there are situations where this is not possible.

In such cases, the client is contacted to correct the information.

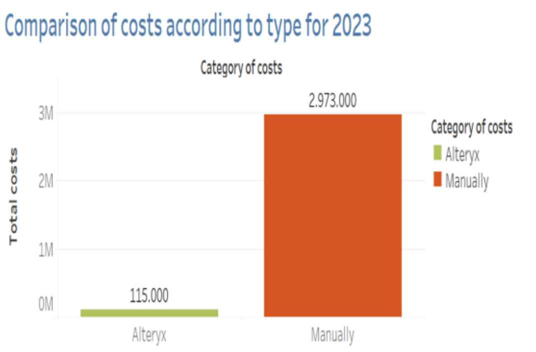


Fig. 6. Comparison of costs according to reporting type for 2023
Source: Authors' own research

A significant difference in costs between the two methods can be observed in **Fig. 6** and **Fig. 7**. In the case of manual reporting, costs are significantly higher due to the greater number of errors and the increased need for resubmissions. Additionally, in the analysed situation, there was a month in which the delivery deadline was missed, resulting in a penalty. When the human factor is involved, there is a higher chance of missing deadlines compared to using a tool that handles data correction.

From **Fig. 6**, we can observe that the difference between the automated procedure with Alteryx and the manual procedure is quite significant. Because of this, any unforeseen costs in the automated process would find it challenging to exceed this difference. If a company chooses to implement the automated method, it can allocate the money saved to other activities within the company.

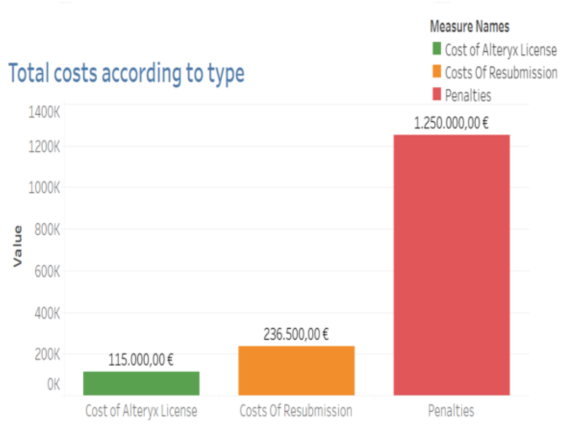


Fig. 7. Total costs according to type for 2023
Source: Authors' own research

If errors occur, it is necessary to resubmit the report to the central banks after correcting them, and each resubmission incurs a fixed cost. The involvement of human intervention significantly increases the likelihood of errors, necessitating the re-submission of reports far more frequently than when utilising automation. In the case of missing the reporting deadline, a fixed penalty is applied, which is shown in the graph as the third column. Of course, this doesn't necessarily have to happen annually or, hopefully, at all. However, even if the expenses for such a fine are eliminated, the costs of automation would still be significantly lower.

In the case of process automation, there is an initial cost involved, specifically the licence for the Alteryx tool, represented by the first green column on the left. Errors can also occur here, necessitating resubmission, but the number of such errors is significantly lower (as discussed in the previous graph). The company may encounter additional costs if it opts to engage a consultant to conduct employee workshops, provide supervision, and facilitate familiarity with the tool. These costs are one-time expenditures incurred at the initial stage of adopting the method. Moreover, the total payment required for these services is not expected to exceed the expenses associated with submitting information to central banks, which are primarily due to the prevalence of human errors.

6 Conclusions

Nowadays, technology is advancing at a rapid pace, and companies should keep up with these changes. With the emergence of artificial intelligence and automation, many traditional daily tasks are becoming obsolete. As this study shows, it is undeniable that automation, where possible, is far superior to manual work.

Of course, there are some processes that cannot be automated—those requiring creativity or human interaction. However, according to the above analysis, it is clear that implementing automation in reporting is beneficial, being faster than human effort, less costly, and leading to significantly fewer errors.

Among the conclusions of the study is that an employee can complete reporting four times faster with automation. This time saving allows the employee to handle reports for four different projects instead of just one. This freed-up time can be redirected to other critical areas such as innovation and customer service, where human input is essential. Finance teams will need to automate their low-value and time-consuming processes if they want to dedicate more time to offering the company value-adding insights and crucial decision-supporting assistance. As part of the changes the finance department must undergo to become a strategic business partner to the organisation, the two are closely intertwined.

Another conclusion is that the number of errors is up to 21 times lower with automation. In most cases, the remaining errors are due to incorrect data received from clients and the impossibility of correcting them either automatically or manually. Here, the human factor can intervene by contacting the client to correct the missing or erroneous information.

Moreover, the costs involved are significantly lower. Although a licence for an automation tool is quite expensive, the overall amount paid is reduced. By reducing errors, the need for resubmitting information—which incurs a fixed cost—is also reduced. Automating the reporting process eliminates the need for employees to manually check the correctness of data field by field, thus reducing human error and work time, and avoiding costly penalties for missing deadlines. Eliminating the costs of resubmission and avoiding fines leads to significant cost savings.

Technology is rapidly transforming business operations across nearly every function. It can significantly speed up processes, improve quality, and enable individuals to focus on areas requiring human involvement. While many finance functions have embraced substantial automation, financial reporting and technical accounting teams risk falling behind. This lag can lead to frustration and potentially harm the business.

Based on all the points presented, it is evident that automation in financial reporting brings substantial benefits and improvements. These tangible benefits underscore the increasing impact of automation and highlight that automation is a significant step towards sustainable success.

Therefore, it is worthwhile for companies to replace manual reporting with the use of a tool to automate the process. AnaCredit reporting process automation has measurable advantages in terms of increased compliance, improved accuracy, and time savings. Financial institutions can maximise their reporting operations and focus on fostering corporate growth and innovation by utilising automation, which streamlines procedures, reduces mistakes, and ensures regulatory compliance.

References

[1] S. Pougkas, “SAP S/4HANA Financial Accounting Certification Guide,” Rheinwerk Publishing, 2017.

[2] SAP, “SAP Intelligent RPA Use Cases By Lines of Business (LoB): oCFO – Record to Report,” https://www.sap.com/documents/2021/02/f6410932-d27d-0010-87a3-c30de2ffd8ff.html, 2021.

[3] KPMG, “Automation of financial reporting and technical accounting,” https://assets.kpmg.com/content/dam/kpmg/uk/pdf/2019/09/automation-of-financial-reporting-and-technical-accounting.pdf, 2019.

[4] R. Half, “How Did Accenture Transform Lead-to-Cash Processes with Ready-toUse AI?,” 2023.

[5] B. I. Dragomirescu O., “Revolutionising Finance: Insights on the impact of Automation,” 2024.

[6] W. Durant, “Using Alteryx automation software to boost the efficiency of creating the net interest income note for company X’s annual report,” 2024.

[7] KPMG, “AI and Financial Reporting Survey – what are companies doing and where do you stand?,” 2023. [Online]. Available: https://kpmg.com/us/en/articles/2023/ai-and-financial-reporting-survey.html.

[8] Deloitte, “Audit & Assurance,” 2023. [Online]. Available: https://www.deloitte.com/uk/en/services/audit.html.

[9] GRI, “Four-in-five largest global companies report with GRI,” 31 October 2022. [Online]. Available: https://www.globalreporting.org/news/news-center/four-in-five-largest-global-companies-report-with-gri/.

[10] E. Comission, “EU rules on financial information disclosed by companies.,” 2023. [Online]. Available: https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/financial-reporting_en.

[11] KPMG, “Automation of financial reporting and technical accounting,” 2019.

[12] M. a. Company, “The impact of AI on accounting practices: A review: Exploring how artificial intelligence is transforming traditional accounting methods and financial reporting,” World Journal of Advanced Research and Reviews, 2024.

[13] M. Ashraf, “Does Automation Improve Financial Reporting? Evidence from Internal Controls,” S&P Global - Market Intelligence, 2024.

[14] J. Gerring, What Is a Case Study and What Is It Good for?, Cambridge University Press, 2004.

[15] T. Gary, How to Do Your Case Study, Sage Publications Ltd., 2021.

[16] O. Hauet, “The AnaCredit regulation is in place—a reflection on data-driven reporting: What has been achieved by the banks and regulators, and what remains to be done?,” Journal of Financial Compliance, 2019.

[17] V. D. R. B. Jean Marc Israel, “The Analytical Credit Dataset. A Magnifying Glass for Analysing Credit in the Euro Area,” ECB Occasional Paper No. 187, 2017.

[18] S. C. Group, 2020.

[19] ECB, AnaCredit Reporting Manual Part I – General Methodology, 2019.

[20] D. C. Thun, “AnaCredit Gives Banks an Opportunity to Improve Data Management, but Challenges Remain,” Analytics, Moody's, 2015.

[21] P. Brown, “International Financial Reporting Standards. What are the Benefits?,” Financial Accounting and Equity Markets, 2013.

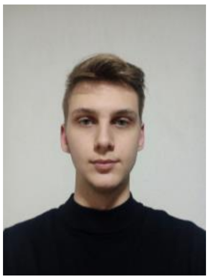
[22] O. Hauet, “The AnaCredit regulation is in place — a reflection on data-driven reporting: What has been achieved by the banks and regulators, and what remains to be done?,” Journal of Financial Compliance, Volume 3, 2019.

[23] A. W. RALF EWERT, “Effects of Increasing Enforcement on Financial Reporting Quality and Audit Quality,” Journal of Accounting Research, 2018.
[24] “Tableau,” 2024. [Online]. Available: https://www.tableau.com.

Oana-Alexandra Dragomirescu is a PhD student at the Academy of Economic Studies from Bucharest, her research topic being “Contributions to the automation of business processes in the financial field”. She graduated from the Master’s degree program Information System for Economic Process and Resources Management (SIMPRE) in 2023. She is currently working as a SAP FI Consultant and her expertise include financial processes optimization, central finance and data integration and migration. Her fields of interest are business processes automation, RPA.

Paraschivoiu Adriana-Teodora is a master's student at the Academy of Economic Studies from Bucharest, her research topic being “Automation of financial reporting processes”. She is a student of the program Information System for Economic Process and Resources Management (SIMPRE), following graduation in 2024. Her fields of interest are: data analysis, agile methodologies and business automation processes.

Andreea VINEȘ is currently a PhD student at the Academy of Economic Studies from Bucharest, where her research subject is ‘Agile solutions for Business Intelligence’. She is currently working as a Data Engineer, and she comes with a wealth of experience in designing, developing and maintaining data solutions, her expertise including building and managing data pipelines, data warehousing and ETL processes, focusing more on the cloud technologies. Her main fields of interest are cloud technologies, big data, data warehousing and BI.



Nica Andrei Cătălin is a master's student at the Academy of Economic Studies from Bucharest. He will graduate from this program in 2024. Currently, he works in the field of marketing and has developed knowledge in working with artificial intelligence. His fields of interest are: artificial intelligence, business process automation, and Business Intelligence.

Mapping Business Process Modeling with the Business Models of Several Energy Community Members

Anca Ioana ANDREESCU[1], Simona-Vasilica OPREA[1], Adela BÂRA[1],
Alin Gabriel VĂDUVA[1], Andreea-Mihaela NICULAE[1]
[1]Department of Economic Informatics and Cybernetics
Bucharest University of Economic Studies
anca.andreescu@ie.ase.ro, simona.oprea@csie.ase.ro, bara.adela@ie.ase.ro,
alin.vaduva@csie.ase.ro, andreea.niculae@csie.ase.ro

Energy communities (ECs) play a major role in energy systems by enabling decentralized production and distribution of renewable energy. This article applies business process modeling to enhance and align the business models of various EC members. Using Business Process Model and Notation (BPMN), it maps the operational workflows of key participants, including prosumers, storage owners, EV charging stations, aggregators, and entities involved in Local Energy Markets (LEM) and Local Flexibility Markets (LFM). Proposed BPMN models provide a structured perspective on essential tasks, decision points, and interactions within the energy market, capturing processes such as energy forecasting, trading, flexibility transactions and daily operations. Through process visualization, the models offer valuable insights for optimizing energy usage, enhancing grid stability and maximizing economic benefits. This approach highlights BPMN's capability to support more efficient, sustainable, and resilient ECs within decentralized systems.

Keywords: *energy communities, business models, business process models, BPMN*

1 Introduction

ECs are groups that aggregate individual and shared resources to participate collectively in energy production, consumption, and management. This model enables distributed energy resources, such as solar panels or shared battery storage, to be pooled among members, who then interact with the grid as a single entity. The cooperative structure enables energy and financial transactions with utility companies, aligning individual incentives with broader community goals and providing members with benefits beyond those available individually [1].

Designing business models for energy markets and ECs is critical because it enables the efficient, sustainable and equitable distribution of electricity, especially as the market adapts to new technologies and renewable energy sources. With the rise of electric vehicles (EVs), renewable integration, and decentralized energy systems, traditional utility models are no longer sufficient to meet dynamic demand and ensure grid stability.

While a business model is strategic and describes what the business does to create, deliver and capture value [2], a business process model is operational in nature and describes how specific workflows and tasks support those strategic goals. Therefore, a business process model provides a functional perspective, outlining how value is generated within the organization by detailing each step involved in achieving desired outcomes [3].

2 Literature review

Existing business process models in the energy sector are designed to streamline, standardize, and optimize various operational workflows within energy production, distribution, and consumption. Scientific research provides a variety of models (both strategical or operational) focused on various aspects of the energy market and ECs. The following highlights some of the most relevant studies in this area.

A comprehensive study explores the growing field of electric vehicle (EV) destination charging, emphasizing critical aspects like charging tariffs, business models, and coordination strategies essential for expanding and sustaining these networks [4]. It identifies destination charging as a crucial complement to residential and public charging, assessing models such as network operator, owner-operator, and integrated approaches for their profitability, user engagement, and feasibility, alongside pricing tactics like time-of-use and real-time options. The findings emphasize the value of user-cantered and flexible pricing models to better synchronize EV charging infrastructure with both demand patterns and operational requirements.

The analysis in [5] focuses on a method for integrating IoT devices into business models using Business Process Model and Notation (BPMN), emphasizing the streamlined, inclusion of real-time data from appliances like sensors and actuators. This approach allows business processes to directly interact with Internet of Things (IoT) devices, such as temperature sensors and automated robots, without increasing BPMN's complexity. The BPMN framework models the overall process flow, while an ontology system manages detailed IoT data, enabling business models to adapt based on real-world conditions. The proposed microservices-based architecture supports these interactions by decoupling IoT devices from the BPMN models, enabling flexible execution that's compatible with varied IoT appliances and technologies.

In [6] the authors are presenting an "energy-as-a-service" business model, targeting aggregators who manage prosumers equipped with distributed energy resources (DERs) such as solar panels and battery storage. Moving beyond traditional volumetric pricing, this model provides prosumers with free electricity through a fixed monthly fee, while aggregators optimize DERs across wholesale markets, including energy and ancillary services. By focusing on predictability and ease, the model offers prosumers a straightforward, risk-free way to engage in energy markets with assured outcomes. Findings show a significant boost in aggregator profitability and a reduction or elimination of energy costs for prosumers.

Another research examines how innovative business models are essential for integrating prosumers into decentralized energy systems [7]. With increasing digitalization and a shift toward sustainability, the energy sector is transitioning from centralized models to flexible systems that empower individuals to locally generate, manage, and trade energy. This shift is supported by technologies like smart grids, IoT, and blockchain, as well as regulatory frameworks promoting prosumer involvement. Business models such as self-consumption, leasing, demand response, and peer-to-peer trading are discussed, each offering distinct benefits and facing challenges shaped by regulations, technology, and market structures. Although these models present economic and environmental benefits, obstacles like costs, infrastructure demands, and regulatory inconsistencies limit widespread adoption.

A different perspective in this area is examined in [8]. The study investigates business models for ECs, including prosumers, storage, EV charging, and aggregators. Six models are proposed to enhance value through local flexibility markets, optimize costs, and drive the energy transition. By shifting energy use to off-peak times, EC members can reduce costs, earn additional revenue, and aid grid operators in load balancing. A case study of 114 apartments in a local flexibility market demonstrates significant financial benefits: consumers earn income from flexibility contributions, while retailers save on energy costs. The potential of these models to support sustainable energy practices, encourage decentralized systems, and provide economic benefits to EC members is emphasized.

Additional relevant research [9] explores emerging business models in LEMs, focusing on peer-to-peer trading, community self-consumption, and transactive energy systems. These decentralized models empower prosumers, consumers, and aggregators to engage more actively in energy trading and management, fostering a collaborative and flexible energy ecosystem. Using the Business Model Canvas framework, the study examines nine key actor categories, shedding light on critical components such as value propositions, customer relationships, revenue streams, and key partnerships. The findings highlight the strengths of these models in increasing flexibility and empowering users, enabling them to contribute to energy sustainability at the local level. However, the review also identifies challenges, particularly in ensuring economic sustainability, regulatory compliance, and technical interoperability.

Our findings indicate that, while there is ample support for describing the roles of ECs and their members, research on business process models for ECs remains limited, suggesting significant potential for further exploration.

3 Methodology

This article applies business process modeling to map the business models of EC members, with a focus on using Business Process Model and Notation (BPMN) as the primary tool.

Business process modeling serves as an effective method for visualizing, understanding, and improving complex processes, particularly in sectors like energy markets where workflows are highly dynamic and require frequent adjustment in response to real-time data. The use of BPMN is specifically selected due to its standardized and intuitive notation, which enhances both communication and operational clarity across stakeholders. It also can be seen as a means to communicate across different languages and cultures [10].

For most BPMN users, graphical representation of models is essential. BPMN provides three main types of diagrams. The process or collaboration diagram is commonly used to depict the flow of a process, including activities, splits, and parallel flows, as well as collaborations between multiple processes with exchanged messages. A single-process version is typically called a process diagram, while a version with interacting processes is referred to as a collaboration diagram. Other types are choreography and conversation diagrams for visualizing complex protocols or an overview of partners and their interactions [11].

Process diagrams are the most widely used and the most intuitive type of BPMN diagram. Three main categories of notations for a process diagram are depicted in Fig. 1.

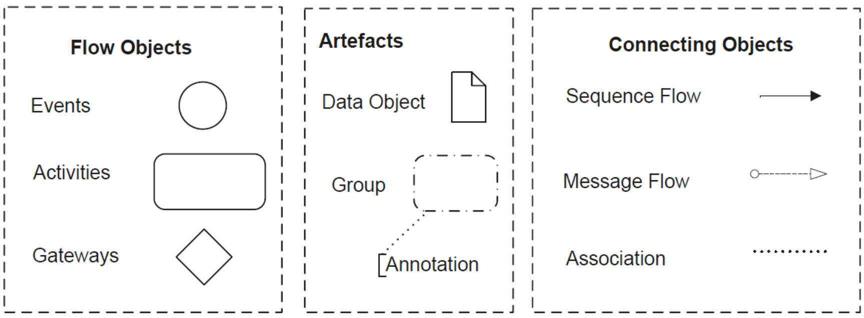


Fig. 1. Detailed sequence diagram for General Settings configurations

In BPMN, flow objects are essential components that include events (signifying process-relevant occurrences), activities (work units within a process), and gateways (controlling path divergence and convergence). Artifacts provide additional details, such as data objects and annotations, adding context without affecting process execution. Connecting objects link these elements: sequence flows outline the order, message flows represent exchanges between participants, and associations connect artifacts to other flow elements [3]. Events are represented by circles, activities by rounded rectangles, gateways by diamond shapes and arcs (called sequence flows in BPMN process diagram) are represented by arrows with a full arrowhead [12].

Business process modeling offers a structured approach for capturing detailed operational steps within an energy market, allowing a clear illustration of each task, decision point, and interaction between entities. This clarity is critical in energy markets where actors, such as energy producers, aggregators, consumers, and market operators, must operate within tightly regulated frameworks while also responding flexibly to fluctuating supply and demand. The visual representation provided by business process models enables both technical and non-technical stakeholders to engage in the workflow, promoting a shared understanding and alignment of objectives [13].

BPMN is chosen as the modeling standard in this research for several reasons. First, BPMN’s standardized notation facilitates interoperability, allowing models to be interpreted consistently across different organizations. This is particularly valuable in energy markets, which involve diverse participants with varying levels of technical expertise. BPMN’s visual syntax, with its distinct elements for tasks, decision points, events, and gateways, allows complex processes to be represented comprehensibly, thereby reducing ambiguity and improving accuracy in process interpretation.

Second, BPMN’s flexibility allows it to capture both simple and highly complex workflows, making it suitable for a range of energy market scenarios, from straightforward trading procedures to intricate decision-making processes that involve real-time adjustments. The ability of BPMN to represent conditional flows and event-driven gateways is especially relevant for energy markets, where processes must adapt dynamically based on market conditions, demand, and resource availability.

The research methodology aims to analyse and design BPMN models tailored to various members in ECs, such as prosumers or aggregators. Each developed BPMN model illustrates specific operational workflows and interactions within the energy market ecosystem. The methodology can be divided into the following steps: (1) Members identification, which involves identifying relevant members of ECs, such as prosumers, EV charging stations, and aggregators, who participate in the energy market; (2) Process analysis, which details the operational processes of each member, including forecasting, energy trading, flexibility transactions, and daily management cycles. Each process flow is analysed based on tasks, events, and decisions required to achieve energy optimization and economic benefits; and (3) BPMN modeling, which creates BPMN process diagrams for each member’s processes, emphasizing workflow, decision points, and interactions with local and flexibility markets.

4 Results

Scientific literature provides a comprehensive overview of several key roles within ECs, with each role contributing uniquely to the production, consumption, storage, and management of energy resources [1], [8], [14], [15], [16]. These members can be classified as follows: (a) active energy participants within the local energy system, such as prosumers, storage owners, or EV (electric vehicle) charging stations; (b) entities and mechanisms that support a decentralized energy ecosystem, including Aggregators, Local Energy

Markets (LEM), and Local Flexibility Markets (LFM); and (c) facilitators and enablers, such as Community Operators, Investors or Sponsors, and Regulatory Bodies or Advisors. In this paper, we focus on the first two categories of EC members, as business models for these roles are generally applicable across global energy markets. While facilitators and enablers, such as Community Operators or Regulatory Bodies, are critical to ensuring compliance and financial stability, their roles often vary significantly based on regional policies and specific community setups. By concentrating on the first two categories, we aim to address the most broadly applicable and impactful aspects of ECs, enabling our findings to be relevant and adaptable across a wider array of markets and regulatory landscapes.
In the remainder of this section, we introduce and provide detailed descriptions of several BPMN business process models to outline the business requirements for the following categories of EC members: prosumer, battery-based storage, EV charging station, aggregator, LEM and LFM.

4.1 Prosumer

Prosumers are consumers who both use and generate energy, distributing any excess electricity to others within the grid [16]. Grids that incorporate prosumers bring notable benefits and opportunities that set them apart from traditional grids. For example, smart prosumer grids improve efficiency by integrating advanced control and communication technologies to optimize the energy use of home appliances.
The BPMN process diagram in Fig. 2 is tailored for a prosumer, depicting a systematic approach for managing appliance control and handling energy surplus within a renewable energy-based structure. The process starts with an event labeled "Each day," marking the beginning of the prosumer's daily operational cycle.

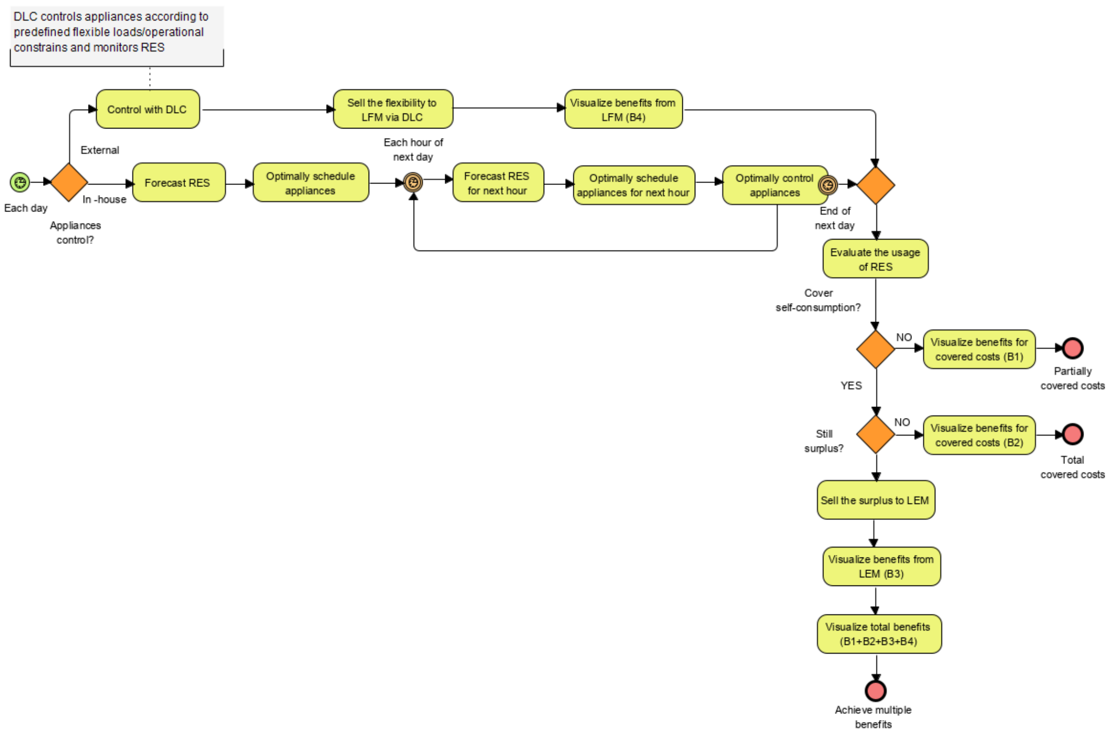


Fig. 2. Business process diagram for a prosumer

The initial decision gateway guides the workflow depending on whether control is managed internally or externally. If the decision is for in-house control, the prosumer manages their own appliances, and the process moves to the task of forecasting reusable energy sources (RES), where the availability of renewable energy for the following day is forecasted. Following this, the "Optimally schedule appliances" task organizes appliance usage based on the forecast to maximize energy efficiency.

An intermediate timer event named “Each hour of next day” initiates a loop where hourly adjustments are made. Inside this loop, three tasks are performed sequentially: “Forecast RES for next hour,” where the forecast is refined for the coming hour; “Optimally schedule appliances for next hour,” which adjusts the schedule to align with the updated forecast; and “Optimally control appliances,” where real-time adjustments are made to optimize appliance performance according to RES availability. This loop continues until the “End of next day” event, which signals the end of the hourly adjustment cycle.

For external appliance control in the initial decision, the workflow bypasses internal scheduling and controlling steps, proceeding directly to the task named "Give control to DLC." Here, appliance control is delegated to the Direct Load Control (DLC) system, which operates appliances based on predefined flexible load and operational constraints while monitoring RES availability, as stated in the note next to the task. Following this, the task “Sell the flexibility to LFM via DLC” comes next, where the prosumer sells flexibility to the LFM using DLC. This path then flows to the “Visualize benefits from LFM (B4)” task, where the benefits obtained from selling flexibility to LFM are summarized and visualized.

Once the hourly adjustments are completed, the workflow advances to the “Evaluate the usage of RES” task, where it is assessed if the renewable energy usage has successfully covered the prosumers’ s self-consumption needs. If the answer is NO the process continues to the “Visualize benefits for covered costs (B1)” task, followed by an end event indicating that some but not all costs were covered.

If renewable energy sources do cover self-consumption, the workflow advances to another decision gateway that checks for remaining surplus energy. If no surplus exists, the prosumer visualizes benefits for covered costs (B2), leading to an end event which confirms that the prosumer’s needs were completely met with no surplus energy left.

If surplus energy remains after self-consumption is covered, the process flows into the task “Sell the surplus to LEM,” where the surplus energy is sold to the LEM, generating additional revenue from the surplus. The benefits gained from this sale are envisioned in the following task, “Visualize benefits from LEM (B3)”. Lastly, a comprehensive task named “Visualize total benefits (B1+B2+B3+B4)” aggregates all benefits obtained across the various stages: covered costs, surplus sales, and flexibility transactions. This overall visualization provides a summary of the financial and operational gains achieved through the day’s energy management activities of the prosumer. The workflow concludes with an end event named “Achieve multiple benefits,” representing the successful realization of multiple benefit categories, including cost reduction, revenue from surplus energy, and income from flexibility sales.

4.2 Battery-based storage

Battery-based storage uses rechargeable batteries to capture and store electrical energy for later use. In an EC, battery-based storage allows for retaining surplus energy from renewable sources like solar or wind, making it available during peak demand periods or when renewable generation is low [17].

Fig. 3 models a scenario for controlling daily trading and flexibility operations in battery-based energy storage. The workflow incorporates forecasting, trading, and active engagement across several energy markets, including the Day-Ahead Market (DAM), Intraday Market (IDM), LEM, and LFM, to enhance the economic value of stored energy. Each step, task, and decision point are crafted to ensure efficient energy dispatch and revenue optimization for battery-based storage.

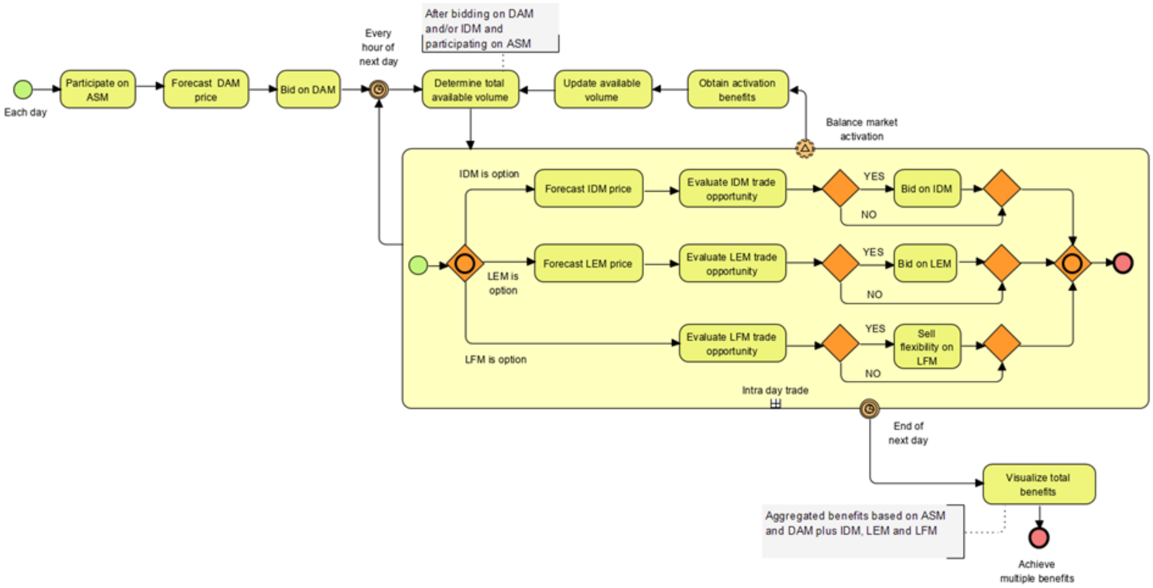


Fig. 3. Business process diagram for a battery-based storage

The process begins with a start time event indicating the start of the daily cycle for managing the battery-based storage's energy resources. The initial task involves participating in Ancillary Services Market (ASM), where the participant provides grid support services, which in turn, generate revenue.
Next, the task "Forecast DAM price" initiates the planning phase for trading in the DAM to determine optimal pricing for energy bids. Following this forecast, bids are submitted to DAM based on anticipated prices and available stored energy.
The process then enters a loop marked by the intermediate timer event "Every hour of next day", to perform hourly assessments and adjustments throughout the day. Within this loop, the first task calculates the amount of energy the battery has available to discharge or store for trading in other markets. Then the flow leads to a subprocess that models the intra-day trade. This subprocess begins with a start event triggered by the need to evaluate an intra-day trade opportunity. The flow then moves to an inclusive gateway, which splits the path based on available options: IDM, LEM and LFM. This gateway has three outgoing paths, each labelled with a condition indicating which option is available, allowing more paths to be taken based on the specific market's availability.
For each of the market options, the subprocess includes two sequential tasks. First, there is a "Forecast Price" task for each option (IDM, LEM, and LFM), where the expected price in the respective market is forecasted. Following the forecasting, the subprocess evaluates the trade opportunity, where the forecasted price is assessed to determine if a trade opportunity on that market is beneficial. After the evaluation, each path encounters a decision point. If the evaluation outcome is positive, the next step involves taking specific trading action in the respective market. For the IDM and LEM markets, this action involves placing a bid, whereas for the LFM market, it involves selling flexibility. If the outcome of the evaluation is negative, no further action is taken on that path, and the flow proceeds towards the end event. After the decision points, the paths converge at a merging

inclusive gateway, that brings together all three paths, ensuring that the subprocess only completes after all available options have been evaluated and, if applicable, trade actions taken. The subprocess will repeat, each hour, until the end of the day.

When a balancing market activation takes place, represented by an intermediate non-interrupting signal event on the boundary of the subprocess, it creates an opportunity for the battery storage system to respond to grid balancing requests. This allows for increased revenue potential through activation benefits while maintaining the regular flow within the subprocess uninterrupted. Following this activation, the available energy volume must be updated to accurately represent the amount of energy that can be supplied or withdrawn in line with market demands, ensuring that responses to balancing requests are based on current energy availability.

An intermediate timer event named "End of next day" marks the closure of the daily trading cycle. After completing trading and flexibility activities, the battery-based storage proceeds to the "Visualize total benefits" step. In this step, the combined benefits from each market (ASM, DAM, IDM, LEM and LFM) are aggregated to give a complete picture of the day's financial performance. The process concludes with an end event, signifying the successful completion of the activities for the day and highlighting the diverse revenue streams achieved through energy trading, flexibility services, and optimal battery use.

4.3 Electric vehicle (EV) charging station

The business process model in Fig. 4 outlines the operations of an electric vehicle (EV) charging station as an active player in the energy market, effectively balancing energy distribution and customer service. It demonstrates how the charging station engages with customers while leveraging opportunities in energy market trading.

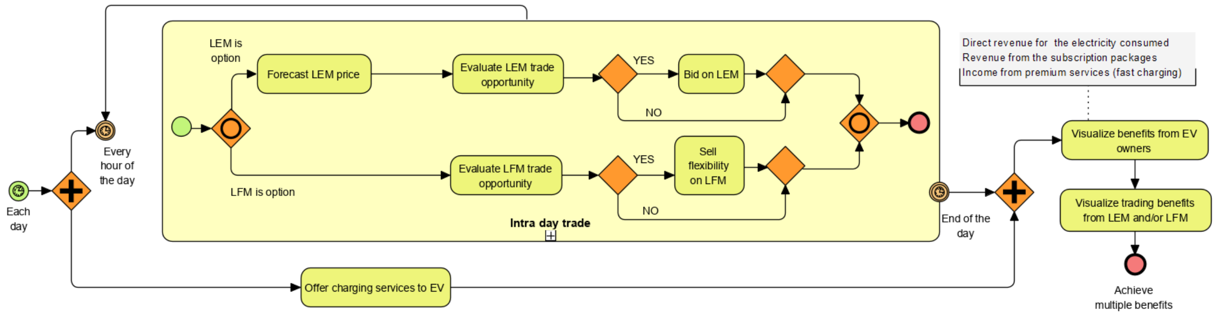


Fig. 4. Business process diagram for an EV charging station

The process starts at the beginning of each day. This daily initiation triggers all subsequent activities and serves as the foundation for the station's activities over a 24-hour period. Following this, a parallel gateway splits the process into two concurrent flows: one for energy market trading and another for EV charging services for EV owners.

In the first flow, the station participates in ongoing energy trading activities. An intermediate time event activates every hour, prompting an assessment of current market conditions. This hourly trigger initiates a trading sub-process, where the station analyses energy market options, forecasts prices, and makes strategic choices to maximize profitability.

Based on trading opportunities, an inclusive gateway divides the process into two paths—one targeting the LEM and the other the LFM—enabling the station to assess and engage with both markets simultaneously. In the LEM path, the station first analyses market data to forecast upcoming energy prices, helping it determine if trading conditions are favourable. A decision gateway then assesses if the forecast justifies placing a bid in the LEM. Similarly, in the LFM path, the station evaluates the potential for selling energy flexibility. If market conditions in either path are favourable, the station proceeds with trading; otherwise, each path concludes for the hour without action.

At the end of each hourly cycle, another inclusive gateway merges both paths, signalling the completion of trading evaluations for that period. The process then resets, preparing to repeat the market assessments in the next hour. An end event marks the conclusion of this hourly trading subprocess.

The second primary flow provides charging services directly to EV customers. Following the initial start event and parallel gateway, the station transitions into the "Offer charging services to EV" task, where it delivers charging based on customer demand, generating revenue from electricity consumption. This charging service operates independently of the market trading activities, allowing the station to meet customer needs while concurrently engaging in energy market activities.

At the end of each day the subprocess stops. Then a second parallel gateway converges both concurrent paths and afterwards, the process initiates the visualization and review of the day’s financial results. First, the station compiles the direct benefits from its charging services. This includes revenue from the electricity consumed, income from subscription packages, and any additional fees collected for premium services, such as fast charging options. This financial overview provides insight into the profitability generated solely from customer interactions.

The station then reviews the financial outcomes from its energy market trading activities, examining the revenue generated from bidding in the LEM and selling flexibility in the LFM. This analysis allows the station to evaluate the effectiveness of its intra-day trading strategy and understand the impact of market participation on overall profitability. The end event signifies the realization of multiple benefits, underscoring the station's ability to maximize revenue by balancing customer charging services with agile market trading strategies.

4.4 Aggregator

Aggregators are essential in enabling small consumers, such as those in residential and service sectors, to participate in electricity markets. While individual energy usage or generation from these consumers may be too limited to impact flexibility significantly, aggregators pool these assets to create a larger, tradable resource. By trading this combined flexibility in different electricity markets, aggregators operate on behalf of consumers, allowing them to indirectly engage in energy markets and contribute to grid stability while supporting innovative business models [18].

Fig. 5 presents another BPMN model, showcasing the aggregator's daily trading and forecasting activities within the energy markets. The model is divided into two main flows: a daily forecasting and bidding sequence, and an hourly intra-day trading cycle. Each path supports the aggregator’s ability to capitalize on different market conditions while meeting the consumption and generation needs of its members.

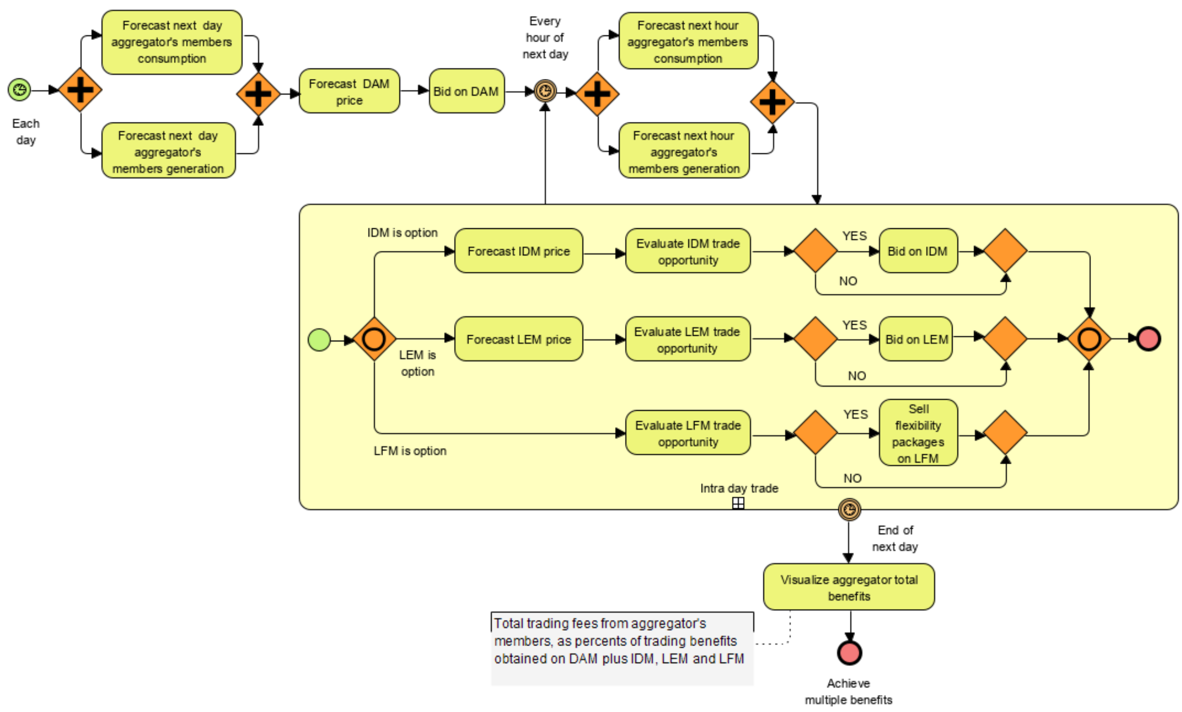


Fig. 5. Business process diagram for an aggregator

At the start of each day, the process initiates with the forecasting of the aggregator's members' daily energy consumption and generation. This step, represented by two parallel tasks, enables the aggregator to anticipate demand and supply for the day ahead. Following this, a forecast for the DAM price is generated, preparing the aggregator to determine if market conditions favour placing a bid. If conditions are favourable, the aggregator places a bid in the DAM.

Next, an hourly loop begins. Every hour, forecasts are generated for the next hour's expected consumption and generation levels of the aggregator's members. This data feeds into an hourly trading sub-process called "Intra-day Trade," which is designed to take advantage of market opportunities in IDM, LEM, and LFM.

Within the intra-day trading sub-process, an inclusive gateway first checks the options available for each market (IDM, LEM, and LFM). If IDM is an option, the process forecasts the IDM price, allowing the aggregator to evaluate potential trades. A decision point then assesses if IDM trading is beneficial; if so, the aggregator places a bid on the IDM. If conditions are not favourable, the IDM path ends for the hour.

Similarly, the aggregator forecasts LEM prices if LEM is an option, evaluating potential trades and deciding whether to bid based on forecasted profitability. If the forecasted conditions are favourable, the aggregator places a bid in the LEM; if no, the step is bypassed.

For the LFM, the aggregator assesses whether selling flexibility packages is viable, adjusting consumption or generation as needed. If the market conditions in the LFM are favourable, the aggregator sells flexibility packages; if not, the LFM path also ends without action.

At the end of each trading cycle, all market paths converge through an inclusive gateway, signaling the completion of the hourly intra-day assessment. This cycle repeats every hour, adapting continuously to real-time market dynamics throughout the day.

At day’s end, the trading results are

compiled. The "Visualize Aggregator Total Benefits" task consolidates the financial outcomes from trading in the DAM, IDM, LEM, and LFM markets, including any fees collected from participants. This visualization offers a summary of the total benefits gained by the aggregator through its trading plan.

4.5 Local Energy Market (LEM)

A LEM provides a decentralized business model enabling participants within a defined geographic area, such as a neighbourhood or community, to trade energy directly with one another. In LEM, prosumers, consumers, and other local stakeholders can engage in buying and selling energy, often sourced from renewables like solar or wind, reducing their reliance on the main electricity grid [8], [15], [16]. Fig. 6 models a business process that describes specific activities for LEM, focusing on integrating, forecasting, optimizing, and trading for local members, such as storage units, prosumers, and EV charging stations.

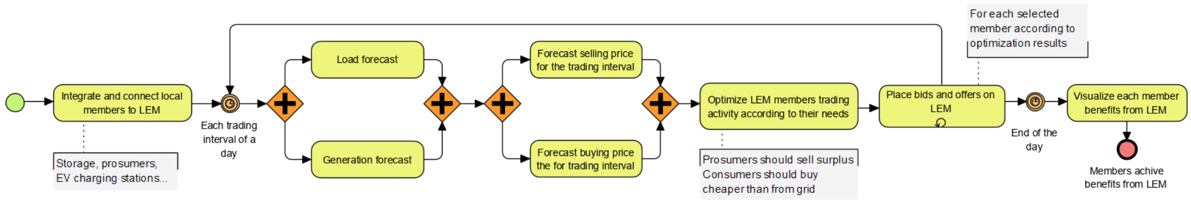


Fig. 6. Business process diagram for LEM

The flow begins with integrating and connecting local members to the LEM. This is followed by a loop that runs for each trading interval within the day, involving separate forecasts for load and generation. These forecasts help in predicting the energy requirements and available generation capacity for that specific interval.
Next, selling and buying prices for the trading interval are forecasted. These forecasts lead to the optimization step, aligning trading activities with the needs of LEM members. Prosumers are encouraged to sell their surplus energy, while consumers are advised to purchase energy at a lower cost than from the grid.
After optimization, bids and offers are placed on the LEM for each selected member, aligned with the optimization outcomes. This is represented as a repetitive task. At the end of the day, the results are visualized, showcasing each member's benefits from the day’s trading. This final step provides insights into the gains achieved by each member through their participation in the LEM.

4.6 Local Flexibility Market (LFM)

In the energy sector, LFM creates opportunities for participants in a defined area to trade flexibility services, helping to balance grid supply and demand. These services involve adapting electricity use, generation, or storage in response to real-time grid needs. Through LFM, decentralized actors such as residential prosumers, commercial buildings, and battery storage owners can provide demand response and other flexibility options, supporting the local grid's stability and ensuring efficiency [1], [8]. Activities specific to energy retailers and consumers on LFM are described in Fig. 7. The model outlines a method to managing LFM and it is repeated on an hourly basis throughout the day. The process begins with connecting local members, including consumers and prosumers, to the LFM platform, establishing a foundation for trading flexibility options aligned with real-time electricity demands. The hourly cycle begins with forecasting the electricity load ($Load_e$) that must be delivered to customers, alongside estimating the hourly electricity price ($price$).

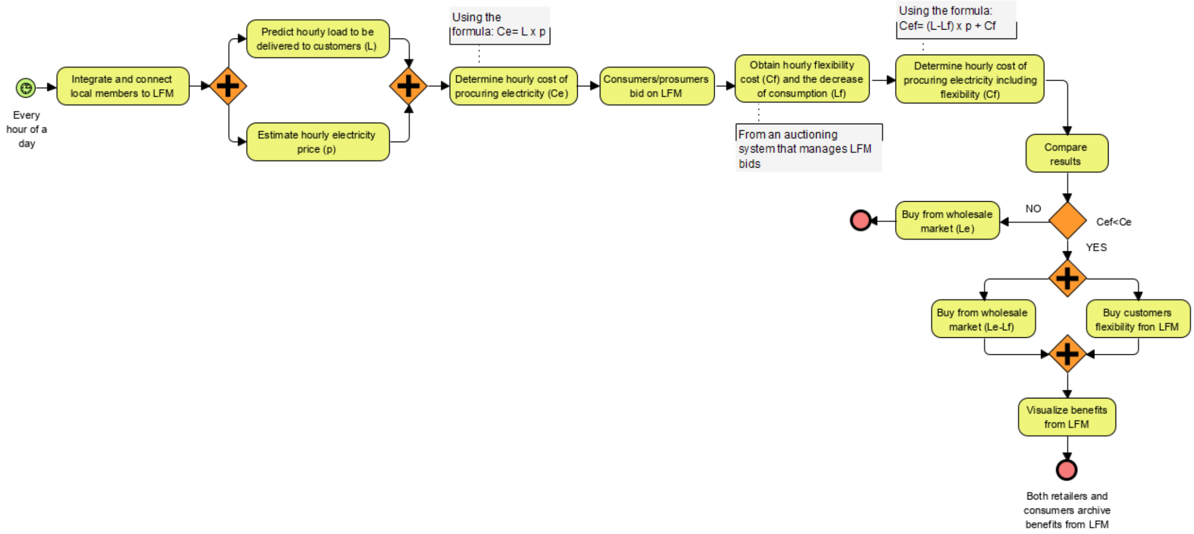


Fig. 7. Business process diagram for LFM

These two forecasts enable the calculation of the hourly cost of procuring electricity ($Cost_e$), determined by multiplying the predicted load with the estimated price, using the following formula:

$$Cost_e = Load_e \; x \; price$$

Once the electricity procurement cost is determined, local consumers and prosumers are encouraged to submit bids on the LFM platform, signaling their willingness to adjust consumption or supply in response to market incentives. The LFM's auction system manages and evaluates these bids, enabling the selection of available flexibility options.
After processing the bids, the workflow moves to determine the hourly flexibility cost ($Cost_{flex}$) and measure the reduction in consumption resulting from flexibility actions ($Load_{flex}$). This step is essential for calculating the adjusted hourly electricity procurement cost when flexibility is factored in. The updated procurement cost, incorporating flexibility ($Cost_{e+flex}$), is calculated using the formula:

$$Cost_{e+flex} = (\,Load_{flex} - Load_e)\; x \; price + Cost_{flex}$$

After determining both $Cost_e$ (the original procurement cost) and $Cost_{e+flex}$ (the procurement cost with flexibility), a comparison is made. If $Cost_{e+flex}$ is lower than $Cost_e$, it indicates that purchasing flexibility on the LFM is beneficial. In this case, the workflow shifts to acquiring flexibility from local members, thereby reducing the amount of energy needed from the wholesale market. The retailer then purchases only the adjusted quantity $Load_{flex}$ - $Load_e$ from the wholesale market, supplementing it with flexibility obtained from the LFM. Alternatively, if $Cost_{e+flex}$ is greater than or equal to $Cost_e$, then it is considered less cost-effective, leading the retailer to fulfill the entire energy requirement directly from the wholesale market without utilizing flexibility options.
In the final step of each hourly cycle, the benefits of participating in the LFM are visualized, offering insights into the advantages for both retailers and consumers. This overview highlights the financial impact of LFM engagement, showing how the flexibility market contributes to optimizing energy procurement costs.

5 Conclusions and future work

Proposed business process models provide a structured framework for analysing and optimizing operational workflow of varied EC members, capturing the intricate interactions and workflows of decentralized energy systems. The models offer a detailed view of tasks and decision points that shape each participant's role, revealing key interdependencies that are essential for effective resource management. By visually mapping critical processes such as energy forecasting, flexibility trading, and daily operations, BPMN helps stakeholders pinpoint bottlenecks and streamline workflows, enhancing both operational efficiency and economic outcomes.

Standardizing EC processes is crucial for creating a consistent framework adaptable to diverse participants. This standardization reduces redundancy, simplifies interactions between market participants, and ensures that each role aligns with the community's objectives. Moreover, the models support the seamless integration of essential market structures like LEM and LFM, which are involved in utilizing and trading community resource flexibility. BPMN's capability to illustrate these mechanisms in detail makes it an invaluable tool for both planning and improving processes in decentralized markets.

A key insight from the study is BPMN's role in enhancing decision-making by visualizing real-time interactions and dependencies. For instance, business process modeling of flexibility trading processes demonstrates how demand response and storage resources can be effectively mobilized during peak periods to ease grid stress and optimize costs. The decision-support function within BPMN aids day-to-day operations and establishes a foundation for strategic planning, enabling activities like optimizing resource allocation during high-demand periods or scheduling storage releases. As a result, proposed models offer guidance to community managers for aligning operations with both market demands and grid stability needs.

Finally, the comprehensive detail of BPMN models enables a clear understanding of the economic value that each EC member can achieve through structured participation in LEM and LFM. By identifying and visualizing value-adding processes, it is revealed how individual contributions—from energy production to storage management—drive collective economic gains. Understanding the interconnected economic contributions within ECs fosters the creation of financially sustainable frameworks, allowing participants to maximize the benefits of their roles.

Future research could explore further customization of business process models to accommodate variations in EC structures, such as those operating in regions with different regulatory requirements or in communities with unique energy needs and goals. Examining BPMN's capacity to model complex, evolving energy technologies could also be valuable, particularly for understanding interactions within communities that incorporate distributed energy resources (DERs) and advanced demand response capabilities. Moreover, future studies could explore more advanced BPMN elements, such as automated decision-making and machine learning integration, to improve real-time adaptability in energy processes.

6 Acknowledgment

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI - UEFISCDI, project number COFUND-CETP-SMART-LEM-1, within PNCDI IV.

References

[1] A. S. Alahmed and L. Tong, "Achieving Social Optimality for Energy Communities via Dynamic NEM Pricing," in IEEE Power and Energy Society General Meeting, 2023. doi: 10.1109/PESGM52003.2023.10253295.

[2] E. A. J. Johnson, "Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers - By Alexander Osterwalder and Yves

Pigneur.," Journal of Product Innovation Management, vol. 29, no. 6, 2012.
[3] M. Weske, Business Process Management: Concepts, Languages, Architectures, 4th ed. Springer, 2024.
[4] J. Y. Yong, W. S. Tan, M. Khorasany, and R. Razzaghi, "Electric vehicles destination charging: An overview of charging tariffs, business models and coordination strategies," 2023. doi: 10.1016/j.rser.2023.113534.
[5] P. Valderas, V. Torres, and E. Serral, "Modelling and executing IoT-enhanced business processes through BPMN and microservices," Journal of Systems and Software, vol. 184, 2022, doi: 10.1016/j.jss.2021.111139.
[6] J. Iria and F. Soares, "An energy-as-a-service business model for aggregators of prosumers," Appl Energy, vol. 347, 2023, doi: 10.1016/j.apenergy.2023.121487.
[7] D. F. Botelho, B. H. Dias, L. W. de Oliveira, T. A. Soares, I. Rezende, and T. Sousa, "Innovative business models as drivers for prosumers integration - Enablers and barriers," 2021. doi: 10.1016/j.rser.2021.111057.
[8] A. Bâra and S. V. Oprea, "A holistic view on business model-oriented energy communities," Kybernetes, 2023, doi: 10.1108/K-07-2023-1235.
[9] J. M. Schwidtal et al., "Emerging business models in local energy markets: A systematic review of peer-to-peer, community self-consumption, and transactive energy models," 2023. doi: 10.1016/j.rser.2023.113273.
[10] J. Fuehrer and W. Almeida, LEARNING BPMN 2.0 A Practical Guide for Today's Adult Learners: An Introduction of Engineering Practices for Software Delivery Teams. 2022.
[11] T. Allweyer, Bpmn 2.0: Introduction to the Standard for Business Process Modeling, 2nd ed. Books on Demand , 2016.
[12] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, Fundamentals of business process management: Second Edition. 2018. doi: 10.1007/978-3-662-56509-4.
[13] E. Taylor, "Business Process Modeling Notation: A Detailed Overview ," https://www.theknowledgeacademy.com/blog/business-process-modeling-notation/.
[14] E. M. Gui, M. Diesendorf, and I. MacGill, "Distributed energy infrastructure paradigm: Community microgrids in a new institutional economics context," 2017. doi: 10.1016/j.rser.2016.10.047.
[15] E. Sorin, L. Bobo, and P. Pinson, "Consensus-Based Approach to Peer-to-Peer Electricity Markets with Product Differentiation," IEEE Transactions on Power Systems, vol. 34, no. 2, 2019, doi: 10.1109/TPWRS.2018.2872880.
[16] Y. Parag and B. K. Sovacool, "Electricity market design for the prosumer era," Nat Energy, vol. 1, no. 4, 2016, doi: 10.1038/NENERGY.2016.32.
[17] A. Saldarini, M. Longo, M. Brenna, and D. Zaninelli, "Battery Electric Storage Systems: Advances, Challenges, and Market Trends," 2023. doi: 10.3390/en16227566.
[18] Ö. Okur, P. Heijnen, and Z. Lukszo, "Aggregator's business models in residential and service sectors: A review of operational and financial aspects," 2021. doi: 10.1016/j.rser.2020.110702.

Anca Ioana ANDREESCU graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 2001. She got the title of doctor in economy in the specialty economic informatics in 2009. At present she is an associate professor in the Department of Economic Informatics and Cybernetics of the Bucharest University of Economic Studies. Her interest domains related to computer

science are requirements engineering, business analytics, modeling languages, business rules approaches and software development methodologies.

Simona-Vasilica OPREA received the MSc degree through the Infrastructure Management Program from Yokohama National University, Japan, in 2007, the first Ph.D. degree in Power System Engineering from the Bucharest Polytechnic University in 2009, and the second Ph.D. degree in Economic Informatics from the Bucharest University of Economic Studies in 2017. She is currently an Associate Professor with the Faculty of Cybernetics, Statistics, and Economic Informatics with the Bucharest Academy of Economic Studies, involved in several research projects.

Adela BÂRA graduated the Faculty of Economic Cybernetics in 2002, holds a PhD diploma in Economics from 2007. She is professor at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from The Bucharest University of Economic Studies and coordinated three R&D projects. Her research interests are focused on data science, analytics, databases, IoT, big data, data mining, power systems, authoring more than 70 papers in international journals and conferences.

Alin-Gabriel VĂDUVA earned his bachelor's degree in Economic Informatics in 2022 and his master's degree in Databases – Support for Business in 2024. He is currently pursuing a Ph.D., focusing on the trustworthiness of artificial intelligence algorithms in business. Professionally, he works as a Data Scientist in the banking sector, contributing to anti-money laundering projects. His research interests include mathematics, machine learning, data mining, deep learning, and generative AI.

Andreea-Mihaela NICULAE – Research Assistant and PhD student at Bucharest University of Economic Studies, Romania; obtained Bachelor's Degree in Economic Cybernetics in 2020, and Master's Degree in Database Support Systems in 2022; former Erasmus+ student at Athens University of Economic and Business, attending Statistics master courses, Athens, Greece; former Erasmus+ student at Hanze University of Applied Sciences, studying IT Management, Groningen, The Netherlands.

The role of Big Data in Climate research

Andreea-Mihaela NICULAE[1], Alin-Gabriel VĂDUVA[1]
[1]Department of Economic Informatics and Cybernetics
andreea.niculae@csie.ase.ro, alin.vaduva@csie.ase.ro

This study explores the growing impact of Big Data in climate change research through a novel approach that combines Big Data analytics with text mining, natural language processing (NLP), and Latent Dirichlet Allocation (LDA). We analysed 7,145 open-access publications from 2011 to 2022 sourced from the Web of Science. Our work highlighted key themes such as urban health, smart technologies, and algorithmic modelling. We observed substantial growth in the use of Big Data in climate research up until 2022, followed by a surprising decline in 2023 that calls for further investigation. Sentiment analysis of the abstracts showed mostly neutral tones, although some exceptions revealed diverse perspectives. This research offers valuable insights into current trends, demonstrating the strength of an integrated analytical approach and the evolving role of Big Data in climate change research. The unexpected downturn in 2023 suggests a shift in research priorities, warranting further exploration.
Keywords: *Big Data, Climate-Change, Bibliometrics, Latent Dirichlet Allocation*

1 Introduction

Climate change is a complex, urgent challenge that calls for new ways to understand and mitigate its effects. Given the vast amounts and variety of climate-related data available today, advanced analytical techniques are essential for drawing useful insights from this information. This study examines the expanding role of Big Data in climate change research, analysing its applications and potential to deepen our understanding of this global issue. Using a dataset of 7,145 open-access publications from the Web of Science (2011-2022), we applied a range of analytical methods, Big Data processing, text mining, natural language processing (NLP), and Latent Dirichlet Allocation (LDA), to identify major research themes. Our analysis points to a substantial increase in climate-focused Big Data research, with clear clusters around urban health, the integration of smart technologies, and advanced modelling techniques. A decrease in publications from 2023, however, raises questions for further investigation. While the application of Big Data in climate change research is rapidly expanding, a comprehensive analysis of its impact and the associated research trends remains limited. This study directly addresses this gap by examining a large dataset of publications, uncovering key thematic areas, and analysing evolving research sentiments. Our findings offer insights into the current state of the field, highlighting areas of strength and weakness, thus contributing to a more informed and strategic approach to climate change research.

The paper is structured as follows: it begins with an introduction, followed by a literature review, a detailed description of the methodology and data sources, and concludes with the results and final insights.

2 Literature Review

The intersection of climate change and Big Data has become a rapidly expanding research area, driven by advancements in data processing and analytics [1]. Big Data's ability to manage vast amounts of diverse, real-time information allows researchers to model complex climate systems [2], predict extreme weather events, and even develop sustainable solutions.

The authors of [1] categorized Big Data applications in climate change research into five key areas through a selective review of over 100 studies in 2019: sustainable urban planning, natural disaster assessment, smart farming, energy efficiency, and other advanced supports. For example, Big Data enables predictive modelling for natural disasters, supports precision agriculture to optimize yields and reduce environmental impacts, and enhances energy efficiency through real-time monitoring. These applications demonstrate Big Data's potential to bolster sustainable practices and increase resilience against climate-related challenges.

Many studies also explore the role of machine learning (ML) in climate change research [3]. While ML and Big Data are distinct domains, they often intersect in climate studies, with ML techniques frequently applied to analyze extensive datasets. The authors of [3] present a bibliometric analysis on ML applications in climate research, examining trends in publication growth, journal prominence, citation patterns, and collaboration networks. Their findings reveal that ML is widely used across various climate topics, including agriculture (e.g., crop predictions), natural disaster modeling (e.g., floods), and weather forecasting (e.g., rainfall predictions).

In addition to bibliometric analyses, the intersection of big data and climate change can be explored through topic modeling, specifically using Latent Dirichlet Allocation (LDA). LDA can analyze various text sources, such as bibliometric abstracts, news articles, social media posts, and other large text collections, to uncover underlying topics and trends. LDA is a useful method in climate change studies as it can capture public sentiment, identify diverse perspectives, and even help categorize broad topics. One research [4] uses LDA on big data obtained from youtube news videos of climate change-related domains in South Korea, where the authors discovered, among others, that many people are scared of climate change, some are relatively aware of environmental issues, few people make serious efforts to suggest alternatives, and that people think about climate change in many ways: social, politics, international, unrealistic, global, lifestyle-centric, and escapist.

3 Methodology

To analyse the large and complex body of research data on climate change and big data, we applied several complementary techniques, including Big Data analytics, Text Mining, Natural Language Processing (NLP), and Latent Dirichlet Allocation (LDA). Together, these methods help us process, categorize, and identify patterns within extensive text datasets, shedding light on recurring research trends and key themes. The following sections describe each part of this approach and its role in analysing the data.

3.1 Big Data

Big Data is a popular term [5] with an increase of usage in recent year due to advancements in computing power that allow for handling vast datasets. This term encompasses a multitude of concepts [6], from the well-known **4 V's**, Volume, Velocity, Variety, and Veracity, to digital techniques and large-scale data that traditional databases cannot accommodate. Big Data generally refers to large volumes of mostly unstructured data from various sources, such as individuals, machines, or sensors. This unstructured data can take many forms, including numerical values, large bodies of text, images, videos, geospatial information, and other types of rapidly generated data.

3.2 Text Mining

To fully leverage the high volume of Big Data, various data analysis techniques have been developed over the years. One notable technique is related to data mining, specifically text mining, which focuses on extracting information and identifying patterns from large volumes of text documents [7]. Text mining serves several purposes, including document classification and clustering, information retrieval and extraction, web mining, concept extraction, and natural language processing (NLP). These techniques capture the essence of textual data and transform information into valuable knowledge.

3.3 Natural Language Processing

Natural Language Processing (NLP) is a field that combines elements of artificial intelligence and linguistics, using machine learning to interpret and generate human language [8]. NLP employs both supervised and unsupervised learning algorithms to handle a range of tasks, including speech recognition, text analysis, and sentiment analysis. By analyzing language structure (using tokenization, parsing, and syntactic analysis) and semantic relationships between words, NLP models can detect patterns, categorize text, and identify themes within large bodies of text. The field also applies linguistic principles to capture nuances in language, such as context, syntax, and sentiment, allowing for applications like language translation, entity recognition, and topic modeling.

3.4 Latent Dirichlet Allocation

A popular method from the NLP family is Latent Dirichlet Allocation (LDA), an unsupervised generative probabilistic model used to discover hidden topics within a significant body of text (or corpus) [9]. In this study, LDA helps identify underlying themes in climate change and big data research by grouping related words into topics. LDA operates by assigning probabilities to words for each topic, under the assumption that each document contains multiple topics. Key preprocessing steps, such as tokenization, removing stop words, and lemmatization, prepared the data for LDA. Model parameters, including the number of topics, were selected to optimize interpretability. A coherence score metric, which measures the semantic similarity among words in each topic, was used to assess topic quality and refine model parameters. Higher coherence scores indicate more meaningful and reliable topics, making the model's results useful for further analysis.

4 Data source

Our research utilizes data retrieved from the Web of Science (WoS) Core Collection database, accessed through an institutional account.

Table 1. Web of Science Core Collection Data Extraction Query

Web of Science filters	**Query**
All Fields	ALL=(clim* chang*) OR ALL=(glob* warm*) OR ALL=(glob* heat*) OR ALL=(clim* warm*) OR ALL=(greenhouse) OR ALL=(environm*) OR ALL=(clim*)
Topic	TS=("big data")
Publication Type	DT==("ARTICLE")
Open Access	OA==("OPEN ACCESS")
Language	LA==("ENGLISH")
Status	NOT (EN==("RETRACTED PUBLICATION"))
Publication Year	NOT(PY==("2025") OR PY==(“2024”))

Using the complex query outlined in Table 1, we obtained bibliographic information for 7,145 publications. Publications from 2024 and 2025 were excluded as they are still in the publication process. We focused exclusively on Open Access articles to ensure easy and free access to the full content for any further assessments. The principal topic of this research is Big Data. To create a comprehensive view of climate change research we selected a variety of terms covering different aspects of the topic.

These include "climate change," "global warming," "global heating," "climatic warming," "greenhouse," "environment," and other derivatives of the terms.

5 Results and interpretation

Initial screening of the dataset revealed 7,145 records and 72 variables. However, 27 variables contained only missing values and were eliminated from the analysis. Among the remaining columns, 17 had more than 50% missing values, making them less useful for further analysis, and these were also removed. The rest of the columns describe various bibliographic aspects, with key variables including Publication Type, Authors, Article Title, Author Keywords and Keywords Plus, Abstract, Affiliations, number of Citations, Web of Science Categories, and Research Areas. For all the key variables, records with missing data were eliminated, ensuring only original information is being analyzed.

Upon further inspection, we found that none of the 7,145 publications were classified as Highly Cited or Hot Paper. This suggests that while the obtained publications contribute to the field of big data in climate change research, they have not achieved a considerable impact or recognition in any domains.

To better understand the obtained dataset, Figure 1 depicts the rising number of publications according to the top 5 most popular research fields found in big data and climate research. Originally, the data spans from 2011 to 2023, with 2011 having only one publication, unfortunately not found in the top 5 research fields. While the query was set to all years except 2024 and 2025, it is important to note that the oldest open-access article focused on the use of big data in climate research is recent, 2011 signaling a rather new research focus. The publication count follows a positive trend, resembling exponential growth, and peaks in 2022, followed by a significant drop in 2023. This decrease may reflect a shift in research interest toward other topics, or perhaps researchers have begun using different terminology to address similar issues. Further analysis is needed to better understand the decline observed in the most recent year.

The dataset contains numerous research areas, with Figure 1 presenting only the top 5 most popular of them. The overall increase in publications, across all research areas, suggests increasing research activity in them and heightened interest to address climate-related issues through diverse research approaches making use of the advantages of using big data. The most popular area is *Computer Science*, followed closely by *Environmental Sciences & Ecology*, signifying their relevance in climate research. *Engineering* and *Science & Technology – Other Topics* show substantial increases along years, but not as notable as the two topics mentioned before. *Telecommunications* shows the lowest number of publications relative to the other fields, but still demonstrates growth.

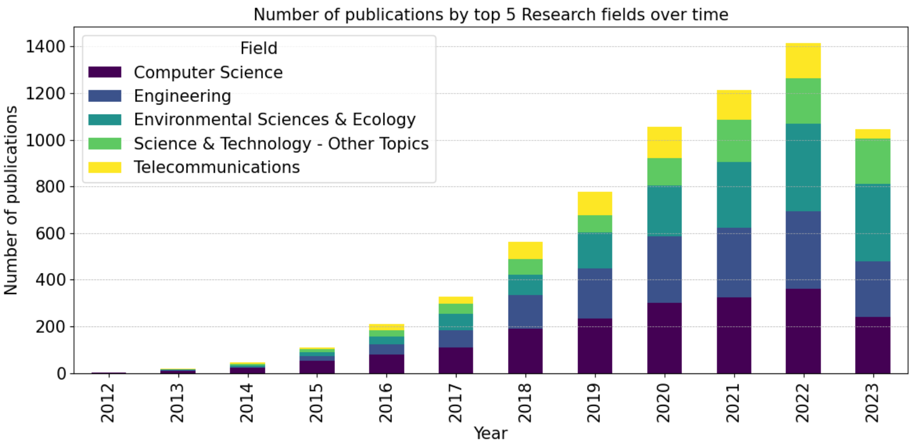


Fig. 1. Number of big data and climate research publications by top 5 research fields

Having examining the growth and distribution of publications over time, we now turn our attention to the content of these publications through text mining techniques, such as generating Word Clouds using Keywords Plus, performing sentiment analysis, or employing Latent Dirichlet Allocation. These techniques allow us to obtain a more in-depth understanding of key themes and patterns found in the big data for climate research publications.
The first analysis we perform is on Keywords Plus, a value created by the WoS database by exploring the articles and extracting common key areas. These values have the same writing style and are easier to use than Author Keywords, which, due to the human's way writing, might include derivations or combinations of terms, making them harder to combine and interpret. Figure 2 presents a Word Cloud obtained from the Keywords Plus of the 7,145 publications. *Big Data* is the dominant theme, seen from its centrality, followed by *Model* and *Management*, indicating a focus on developing *Frameworks* for managing the high volume of climate data effectively. Keywords such as *Prediction*, *Classification*, *Algorithm*, *Performance*, and *Optimization* suggest common analytical techniques used in the field reflecting the multiple methodological approaches being employed in this research. Terms like *Challenges* and *Innovation* suggest that big data presents opportunities in this domain, but it also brings complexities that need to be addressed in climate research. The inclusion of terms such as *Internet, Future, City, Environment* and *Health,* indicates the broader context in which big data and climate change are discussed, linking them to societal impacts. Overall, this word cloud illustrates the variety of elements surrounding big data in the context of climate change, highlighting the role of advanced analytical techniques to uncover insights and address challenges in this critical area of study.

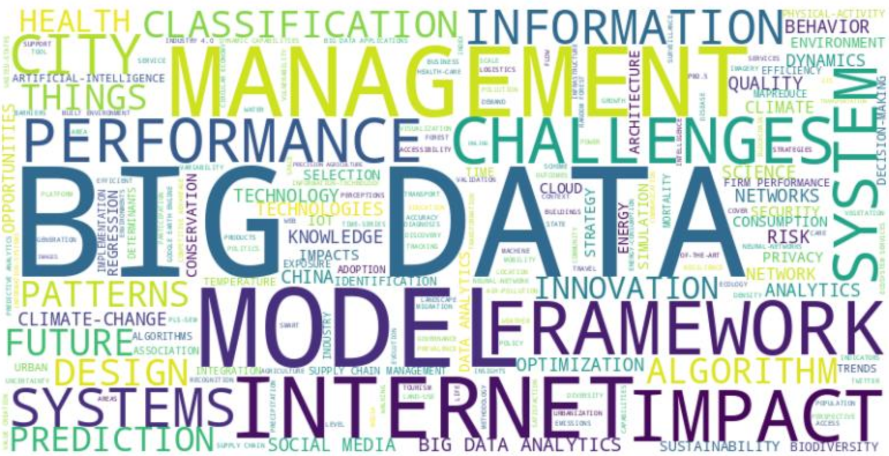


Fig. 2. Word Cloud from Keywords Plus

Using the Natural Language Toolkit (NLTK) package known as SentiWordNet, sentiment analysis was performed on the abstracts, yielding scores for positive and negative sentiments within the texts. In Figure 3, we observe the distribution of both positive and negative scores for each article. Most points are clustered in the lower ranges, suggesting that many abstracts have neutral or low sentiment overall. There appears to be a slight trend where higher positive sentiment scores correlate with increasing negative sentiment scores; however, many abstracts with high positive scores maintain low negative scores. The plot also reveals several outliers, indicating abstracts with exceptionally high sentiment scores. These publications should be further investigated for a better understanding of the implications of big data in climate research. Overall, the results of SentiWordNet analysis indicates that while many abstracts are relatively neutral or slightly positive, there are unique cases that stand out, reflecting varying sentiments in climate change-related discussions.

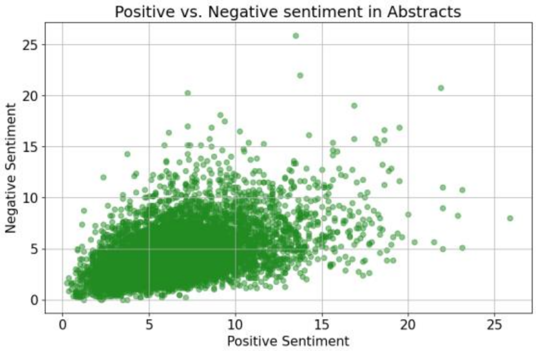


Fig. 3. Analysis of Positive and Negative sentiments from abstract mining

The last text mining approach we use is employing Latent Dirichlet Allocation on all the article's Abstracts to discover common themes. Before showing the results, it is important to mention that LDA's coherence scores for different number of topics were all below 50%, as seen in Table 2. For ease of readability and understanding the hidden topics, we have chosen to firstly represent 3 topics. By increasing the number of topics, while the coherence score increases, the visual representation of topics starts overlapping more and more. With 3 topics, there is no overlap in the plot. We will also present 7 topics to learn more about the implications of big data in climate research, while also showcasing the overlapping themes.

Table 2. LDA coherence scores for different number of topics

Number of topics	Coherence score (%)
3	40
4	39
5	41
6	43
7	48
8	47
9	47

The 3 chosen topics are named after the most important themes identified in them, as it can be seen in Figure 4. Topic 1, *Urban health analysis*, reflects the focus on urban areas and health factors, highlighting spatial analysis and results from studies concentrated on city environments. Topic 2, *Smart technology and Digital research*, emphasises the presence of research and technology, particularly in the context of digital systems and smart solutions, highlighting the advancement and application of modern technologies. Topic 3, *Algorithmic modelling and System development*, focuses on modelling methods, algorithms, and systems in the context of machine learning and research development. These word clouds capture the main themes of each topic, illustrating the varied yet interconnected aspects of climate change research and the use of big data. The prominent terms showcase the multidisciplinary nature of the field, revealing how areas like urban health, digital innovation, and algorithmic methods combine to tackle today's complex environmental challenges.

Fig. 4. LDA Word Clouds for 3 topics

Figure 5 presents LDA's Intertopic Distance Map for the three topics. The significant distance between the topics suggests that they address distinct areas within the larger research context. This separation indicates that each topic focuses on different themes or aspects, highlighting the diversity of research discourse surrounding climate change and big data.

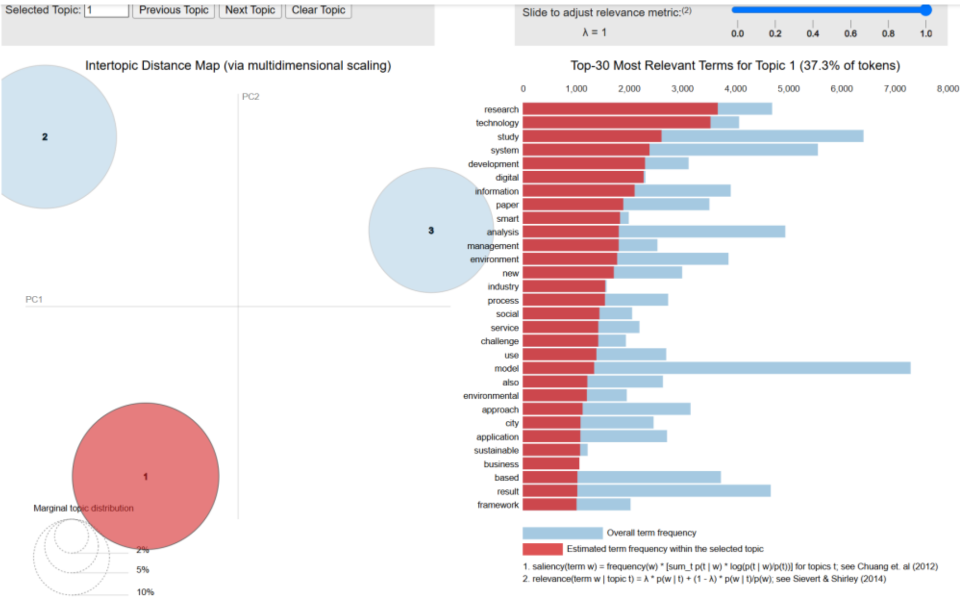


Fig. 5. LDA Intertopic Distance Map for 3 topics

Lastly, we present the obtained 7 topics, the number of topics with the highest coherence score. The new topics are similar to the ones previously obtained, while also containing new information, as seen in Figure 6: Topic 1, Smart technology and Digital management; Topic 2, Water resource management and Climate impact; Topic 3, Algorithmic Systems and Environmental modeling; Topic 4, Urban dynamics and Spatial activity patters; Topic 5, Social development and Digital research; Topic 6, Health prediction models and Risk analysis; Topic 7, Large dataset analysis and Learning methods.

These topics illustrate an interdisciplinary approach, with intersections between technology, social aspects, health, and environmental studies. For example, *Smart technology and Digital management* combines technological advances with management challenges, while *Social development and Digital research* considers societal impacts of technology. Topics focusing on specific application areas, such as *Water resource management and Climate impact* and *Health prediction models and Risk analysis*, suggest that researchers are increasingly applying big data to practical and critical issues, emphasizing the relevance of research in real-world contexts.

Figure 7 shows the Intertopic Distance Map for all seven topics, highlighting how they relate to one another within the research landscape. The topics in the lower half of the plot represent distinct areas of study with minimal thematic overlap, while those in the upper half share common themes, suggesting overlapping concepts or methodologies. This visualization offers useful insights into the connections between various areas of climate change and big data research.

The 7 topics outline a broad and varied landscape of research connecting climate change and big data. They show how different methods and application areas come together to deepen our understanding of complex environmental issues. This creates a useful framework for tackling climate change through interdisciplinary research and innovative data-driven approaches, in contrast to the narrower themes seen in the initial analysis.

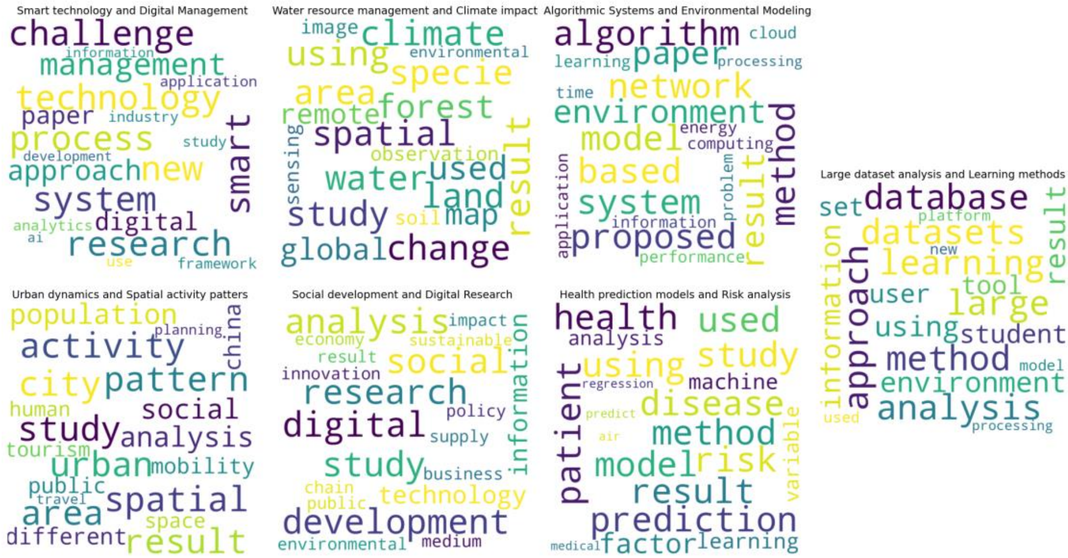


Fig. 6. LDA Word Clouds for 7 topics

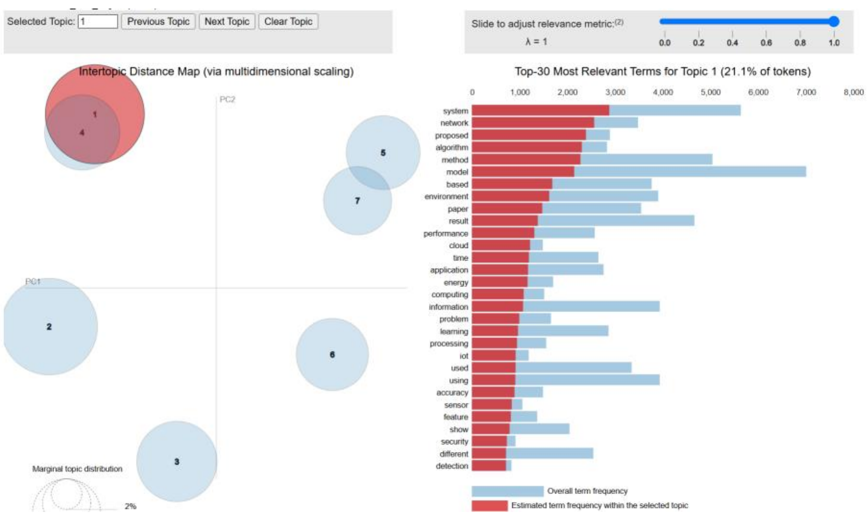


Fig. 7. LDA Intertopic Distance Map for 7 topics

6 Conclusions

This research examined the growing role of Big Data in climate change studies, using a methodology that combined Big Data analytics, text mining, natural language processing (NLP), and Latent Dirichlet Allocation (LDA). We analyzed 7,145 open-access publications from the Web of Science, covering the years 2011 to 2022 (excluding 2024 and 2025 due to ongoing publication). Our findings showed a clear increase in relevant publications from 2011 to 2022, followed by a decline in 2023. This drop raises questions about whether it reflects a shift in research priorities or a data anomaly, and further study is needed to understand this change.

In examining keyword frequencies, 'Big Data,' 'Model,' and 'Management' stood out as dominant themes, pointing to a strong emphasis on building effective frameworks for managing and analyzing large climate datasets. Sentiment analysis of the abstracts revealed mostly neutral sentiment, though with a few notable exceptions suggesting diverse viewpoints on Big Data's role in the field. LDA topic modeling, despite lower coherence scores,

helped uncover key thematic clusters, including urban health analysis, the integration of smart technologies, and advanced algorithmic models.
Overall, these findings highlight the essential role Big Data plays in addressing climate change challenges and underscore the need for interdisciplinary collaboration across scientific domains. Future work should focus on closely investigating the 2023 publication dip, refining sentiment analysis by including full-text data, enhancing the LDA model through parameter optimization, and addressing potential biases due to missing data.

7 Acknowledgment

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI - UEFISCDI, project number COFUND-CETP-SMART-LEM-1, within PNCDI IV.

References

[1] H. Hassani, X. Huang and E. Silva, "Big Data and Climate Change," Big Data Cogn. Comput., 2019.

[2] S. Tatineni, "Climate Change Modeling and Analysis: Leveraging Big Data for Environmental Sustainability," *INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY* , pp. 76-87, 2020.

[3] S.-S. M. Ajibade, A. Zaidi, F. V. Bekun, A. O. Adediran and M. A. Bassey, "A research landscape bibliometric analysis on climate change for last decades: Evidence from applications of machine learning," *Environment,* 2023.

[4] J. Min and H. Lee, "Utilizing Big Data Analysis in Problem Finding of Design Thinking: Deriving Meaning from Climate Change in South Korea Using LDA Topic Modeling," *Archives of Design Research*, pp. 211-221, 2023.

[5] A. D. Mauro, M. Greco and M. Grimaldi, "A formal definition of Big Data based on its essential features," *Library Review*, p. 122 – 135, 2016.

[6] M. Favaretto, E. D. Clercq, C. O. Schneble and B. S. Elger, "What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade," *PLoS One,* 2020.

[7] R. Talib, M. Kashif, S. Ayesha and F. Fatima, "Text Mining: Techniques, Applications and Issues," *(IJACSA) International Journal of Advanced Computer Science and Applications,* pp. 414-418, 2016.

[8] D. Khurana, A. Koli, K. Khatter and S. Singh, "Natural Language Processing: State of The Art, Current Trends and Challenges," *Multimedia Tools and Applications*, 2022.

[9] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li and L. Zhao, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications* , p. 15169–15211, 2018.

Andreea-Mihaela NICULAE – Research Assistant and PhD student at Bucharest University of Economic Studies, Romania; obtained Bachelor's Degree in Economic Cybernetics in 2020, and Master's Degree in Database Support Systems in 2022; former Erasmus+ student at Athens University of Economic and Business, attending Statistics master courses, Athens, Greece; former Erasmus+ student at Hanze University of Applied Sciences, studying IT Management, Groningen, The Netherlands.

Alin-Gabriel VĂDUVA earned his bachelor’s degree in Economic Informatics in 2022 and his master’s degree in Databases – Support for Business in 2024. He is currently pursuing a Ph.D., focusing on the trustworthiness of artificial intelligence algorithms in business. Professionally, he works as a Data Scientist in the banking sector, contributing to anti-money laundering projects. His research interests include mathematics, machine learning, data mining, deep learning, and generative AI.