Ref. Ares(2024)9207323 - 20/12/2024

D4.8 – Synthesis of Observatory Findings – Draft 1

Work Package 4 - Observatory Data Collection and Analysis

Authors: Joep Tummers (WR), Bobby Tsvetkov (WR), Sinne van der Veer (WR), Lan van Wassenaer (WR), Aikaterini Kasimati (AUA), Cristina Virto (INTIA), Nasser Seyni (CTIFL), Marine Louargant (CTIFL), Charlotte Niyomizero (ILVO), Jürgen Vangeyte (ILVO), Sari Vainikainen (VTT), Sajad Ashouri (VTT), Arash Hajikhani (VTT), Rasa Gofman (LITH), Kristina Šermukšnytė-Alešiūnienė (LITH), Ioannis Gitas (AUTH), Eleni Gkounti (AUTH), Azadeh Abdollahnejad (AUTH), Christos Avdellas (RFF)

Date: 20.12.2024

Funded by the European Union

Full Title	**4Growth** - Synthesis of Observatory Findings – Draft 1			
Project number	101082130		**Acronym**	4Growth
Start date	01.01.2024		**Duration**	36 months
Granting authority	European Research Executive Agency (REA)			
Project Coordinator	STICHTING WAGENINGEN RESEARCH (WR)			
Date of delivery	**Contractual**	M12	**Actual**	M12
Type	R - Document, report		**Dissemination level**	PU - Public
Lead beneficiary	Wageningen Research (WR)			
Lead author	Joep Tummers (WR), Bobby Tsvetkov (WR), Sinne van der Veer (WR), Lan van Wassenaer (WR)		**Email**	joep.tummers@wur.nl
Other authors	Cristina Virto (INTIA), Nasser Seyni (CTIFL), Marine Louargant (CTIFL), Charlotte Niyomizero (ILVO), Jürgen Vangeyte (ILVO), Sari Vainikainen (VTT), Sajad Ashouri (VTT), Arash Hajikhani (VTT), Rasa Gofman (LITH), Kristina Šermukšnytė-Alešiūnienė (LITH), Ioannis Gitas, Eleni Gkounti (AUTH), Azadeh Abdollahnejad (AUTH), Christos Avdellas (RFF)			
Reviewer(s)	Aikaterini Kasimati (AUA), Eirini Chlouveraki (AUA) Sjaak Wolfert (WR)			
Keywords	Digital Agriculture; Digital Forestry; Survey study; Adoption;			

Document Revision History				
Version	**Issue date**	**Stage**	**Changes**	**Contributor**
1.0	20.12.2024	Draft	First draft for input from observatories	Wageningen Research (WR)
1.1	22.11.2024	Draft	Added results from observatories experiences + Automated data collected	WR + VTT
1.2	03.12.2024	Draft	Filled in section 2.3	WR
1.3	06.12.2024	Final	Ready for internal review	WR
1.4	18.12.2024	Final	Incorporated internal review comments	WR
1.5	20.12.2024	Final	Final edition	WR

4Growth Consortium

	Participant organisation name	Short name	Country
1	STICHTING WAGENINGEN RESEARCH	**WR**	**NL**
2	EVENFLOW	**EVF**	**BE**
3	GEOPONIKO PANEPISTIMION ATHINON	**AUA**	**EL**
4	REFRAME FOOD ASTIKI MI KERDOSKOPIKI ETAIREIA	**RFF**	**EL**
5	LE EUROPE LIMITED	**LEE**	**IE**
6	DAHEIM CORNELIA	**FI**	**DE**
7	SIMBIOTICA SL	**VIZ**	**ES**
8	EIGEN VERMOGEN VAN HET INSTITUUT VOOR LANDBOUW- EN VISSERIJONDERZOEK	**EV ILVO**	**BE**
9	INSTITUTO NAVARRO DE TECNOLOGIAS E INFRAESTRUCTURAS AGROALIMENTARIAS SA	**INTIA**	**ES**
10	CENTRE TECHNIQUE INTERPROFESSIONNEL DES FRUITS ET LEGUMES	**CTIFL**	**FR**
11	TEKNOLOGIAN TUTKIMUSKESKUS VTT OY	**VTT**	**FI**
12	AGRIFOOD LITHUANIA DIH	**LITH**	**LT**
13	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	**AUTH**	**EL**

Glossary of terms and abbreviations

List of Abbreviations and Acronyms	
AI	Artificial Intelligence
CDP	Consumer Data platform
DTDDs	Digital Technologies and Data-driven solutions
DIHs	Digital Innovation Hub
EU	European Union
MMFT	Market Monitoring & Forecasting Tool
NACE	Nomenclature of Economic Activities
QR	Quick Response (code)
RAG	Retrieval Augmented Generation
WP	Work Package

Contents

Table of Tables

Table of Figures

Executive Summary

This document represents the first deliverable (D4.8) of Task T4.2 in Work Package 4 (WP4) within the 4Growth project. It synthesizes findings of the 4Growth observatory ecosystem, presents the initial portfolio of stakeholders and delineates the results of the first wave of data collection. This sets the groundwork for subsequent data collection and analysis.

The current observatory ecosystem comprises eight observatories, each dedicated to monitoring specific business ecosystem, digital innovation ecosystem, and data ecosystems. These observatories function as central hubs for planned surveys and other data collection methods to generate data and information, which is then utilized in forecast and foresight modules. Together, these interconnected components form a dynamic ecosystem essential for understanding and managing digital agriculture and forestry. Observatories play a pivotal role in gathering comprehensive data on digital technology adoption and its impact, informing strategic decisions and policy formulation in the agriculture and forestry sectors.

This first wave of data collection provided valuable insights into the adoption of digital agriculture and forestry technologies, though the number of responses fell short of project expectations. The first wave of survey responses highlighted significant gaps in data completeness, with 271 unique responses but only 200 providing sufficient data for analysis. The average response time was 14.3 minutes, though completion times varied, with engaged respondents taking approximately 20 minutes. Response rates varied across countries, with Spain (68 responses) and Greece (65 responses) contributing the most, while no responses were received from Hungary or Poland. Stakeholder distribution included 97 agriculture, 83 forestry, and 19 agriculture and forestry stakeholders, with low representation from upstream actors like data providers and platform operators. Enhancing response rates in future waves will require targeted strategies to engage underrepresented groups and leverage the 4Growth consortium's networks effectively.

An extensive analysis of the observatories' experiences with the first wave of data collection revealed key insights for improvement. Survey design issues, such as excessive length and overly technical language, significantly contributed to participant drop-out and engagement challenges. In-person engagement emerged as the most effective data collection method, yielding higher response rates, but it is resource-intensive and not scalable across all contexts. To address these challenges, a mixed-method approach will be adopted moving forward, balancing the effectiveness of direct interaction with the scalability of digital and automated methods to ensure broader and more inclusive participation in future waves.

Besides traditional methods for measuring the adoption of digital technologies, such as the survey outlined above, emerging approaches like automated data collection offer promising alternatives. We leveraged web scraping and AI-driven analysis to, efficiently extracted insights from company websites, identifying trends in digital transformation, technology integration, and data-sharing practices. By streamlining data collection, these techniques can significantly reduce manual workload and provide scalable, real-time insights, addressing gaps in traditional methods. While limitations remain in capturing nuanced or sensitive data, this approach demonstrates strong potential to complement surveys, enhancing the efficiency and depth of data collection efforts in the second wave.

Adjustments for the second wave include shortening the survey, simplifying language, Increased stakeholder engagement strategies, and exploring innovative techniques such as AI-driven web scraping to streamline data collection. These findings will inform updates to the data visualization platform and support modelling tasks in future project phases.

This document is the first deliverable of the task T4.2 “Data collection via observatories”. The organization and planning activities of this task will then be reported in two subsequent versions of D4.8 in M21 (D4.9) and M30 (D4.10), outlining waves 2 and 3 respectively.

1.Introduction

This section provides an overview of the 4Growth project, focusing on its background and objectives. The aim of this section is to highlight the project's key points for the reader to understand how the findings of the 4Growth observatory fit within the overall project's objectives.

Background

In the contemporary landscape of agriculture and forestry, the promise of digital technologies and data-driven solutions (DTDDs) stands as a beacon of potential transformation. Digital innovations offer the promise of enhanced sustainability, economic performance, and working conditions within these critical sectors (Kamilaris et al., 2017; Wolfert et al., 2017). Although the potential and promises of DTDDs are widely acknowledged, policymakers and other stakeholders frequently lack comprehensive and timely insights into their adoption and impact in agriculture and forestry. While initiatives like the FAO's AgriTech Observatory and the Digital Agri Hub (Digital agri hub, 2024; FAO, 2024) provide valuable platforms for advancing digital agriculture and fostering transparency in digital ecosystems, their focus primarily lies in regional development and aggregating innovations in low- and middle-income countries, respectively, which differ from the broader, real-time monitoring and adoption insights envisioned for 4Growth .

Objectives

The 4Growth project aims to contribute to the uptake of digital solutions by (i) documenting the current state-of-play and projecting the future evolution (forecasting and foresight) of the sector; (ii) making insights available to the wider community of decision makers and value chain actors – through the 4Growth Visualisation Platform; (iii) collecting a wide range of ground truth data and identifying key factors or constraints for uptake; and (iv) producing sets of key policy recommendations and best practices to encourage/facilitate further uptake.

WP4 of 4Growth is focused on (iii) and (iv) in close collaboration with other WPs that work on (i) and (ii) (WP2).

The objectives of WP4 are to:

- organize an array of distributed observatories across technologies, subsectors, and regions
- analyse the framework conditions, technical aspects, governance models, socio-economic benefits and data sharing practices associated with the adoption of digital agriculture & forestry
- build synergies with other key European initiatives through which further data can be analysed and outcomes of the project can be sustained
- produce policy recommendations for governance actors and best practice guides for value chain actors to encourage the adoption of digital technologies.

WP4 consists of 5 tasks. Task 4.1 aims to develop and maintain a portfolio/catalogue of various agricultural and forestry stakeholders who will be contacted to gather data on the adoption and use of digital technologies. The establishment of observatories is fundamental to initiating Task 4.2, which encompasses all outreach and data gathering activities conducted by 4Growth observatory partners within their respective ecosystems. The overview of the observatories can be found in Table 1. Data collection will occur over three dedicated waves, during which observatory partners will engage with users and stakeholders to understand their needs and preferences. These interactions will involve multiple touchpoints, including surveys.

Table 2: Organisation of observatories. Adopted form D4.1: Organization of observatories

Observatory node	Node type	Location and region coverage
1	Agriculture	Netherlands (Western Europe)
2	Agriculture	Greece (AUA) (Eastern Europe)
3	Forestry	Greece (AUTH) (Eastern Europe)
4	Agriculture	France (Western Europe)
5	Agriculture	Belgium (Western Europe)
6	Agriculture	Spain (Southern Europe)
7	Agriculture	Lithuania (Eastern Europe)
8	Forestry	Finland (Northern Europe)

Insights gathered during these waves will inform the content presented on the 4Growth Visualisation Platform, such as uptake figures. They will also shape policy recommendations and best practices for value chain actors. Data collection activities rely on the Digital Agriculture and Forestry Uptake Grid, developed under Task 2.2, to ensure consistency and objectivity in addressing all relevant topics. Each wave concludes with a report synthesizing the outreach conducted and the data collected, documented in deliverables D4.8, D4.9, and D4.10, corresponding to M12, M21, and M30.

This deliverable is structured as follows: Section 2 outlines the technical features of the survey and details its distribution methods and data quality. Section 3 discusses the experiences of data collection, including best practices, challenges, and suggestions for future improvements. Section 4 presents the planning for Wave 2, including strategies for targeted surveys, interviews, and the potential integration of automated data collection methods. Finally, the report concludes with a synthesis of findings.

2.Outreach of data collection

This chapter delves into the outreach efforts for collecting data through the execution of a carefully designed survey. The survey aimed to gather insights into the adoption and impact of digital technologies in agriculture and forestry across Europe. By targeting various stakeholders through a combination of digital and in-person methods, the survey sought to capture a wide range of perspectives and data. In this chapter, we discuss the technical features of the survey, the methods employed for its distribution, and an evaluation of the responses received, including their quality and relevance to the 4Growth project's objectives.

2.1 Technical features of the survey

The survey was based on the *Digital Agriculture and Forestry Uptake Assessment Grid* developed in *WP2: Uptake of Digital Agriculture & Forestry Technologies*, described in *Deliverable D2.2: Development of the Digital Agriculture and Forestry Uptake Assessment Grid.* This grid served as a standardized framework for capturing key data points from stakeholders across various sectors.

Survey Implementation

The survey was created using the *Consumer Data Platform (CDP) (Wageningen Economic Research, 2024b) ,* a tool developed by Wageningen Economic Research to facilitate consumer science reports in various national and EU projects. While originally designed for consumer research purposes, the CDP tool proved versatile and was easily adapted to meet the specific needs of the 4Growth project. It enabled the construction of reusable question components aligned with the thematic blocks of the grid (e.g., governance model, technology adoption, economic impact) (see Figure 1).

Although the CDP supports survey creation, it requires an external platform, such as Qualtrics, for distribution to respondents. Its multilingual support ensured accessibility, offering surveys in a basic English version and translations into Greek, Spanish, French, Dutch, Finnish, Lithuanian, Swedish, Hungarian, and Polish. The integration of the CDP with the project's needs allowed for efficient and consistent data modeling, which was essential for collecting and analyzing stakeholder insights effectively. ,By using the CDP tool, a data model could be exported to Wageningen Economic Research's Adagio platform (Wageningen Economic Research, 2024a), which supports the OData protocol. OData (Open Data Protocol) is a standardized data access protocol designed to enable seamless integration with external systems (Chappell, 2011). This export capability facilitated efficient data sharing with WP2's dashboard visualization from *T2.3 Visualisation Platform of Digital Agriculture & Forestry Uptake* by providing structured datasets for direct integration into the project's visualization tools.

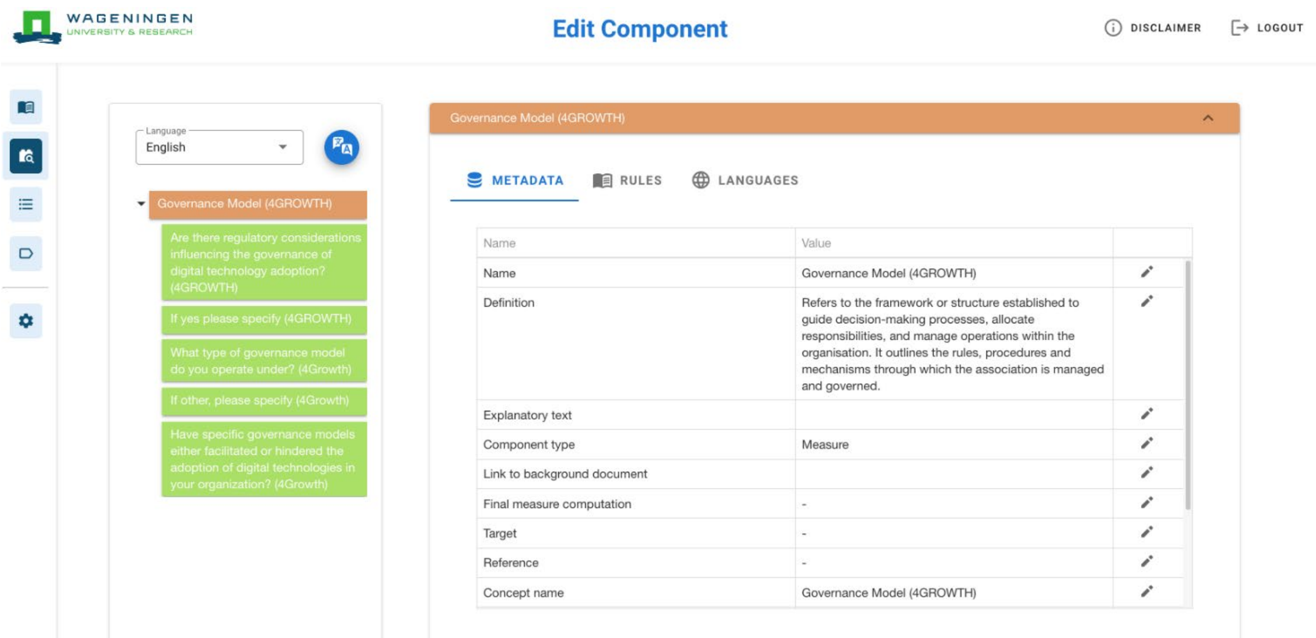

Figure 1: Screenshot of the Governance model survey component in the CDP tool

Survey Interface

The actual survey interface was implemented in Qualtrics (Qualtrics, 2020), which allowed the creation of language-specific surveys. Each survey began with a consent form in the respondent's native language. The consent form can be found in Appendix A. These consent forms were developed in collaboration with WP1: Coordination and management.

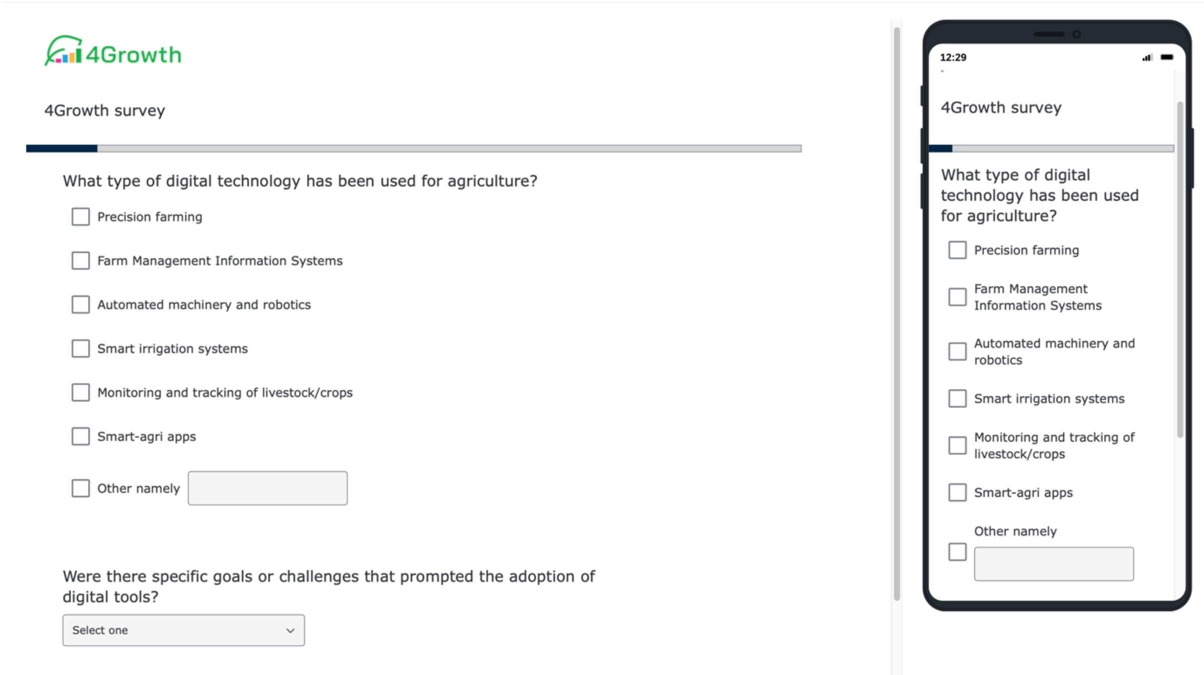

Figure 2: Preview of the 4Growth Survey

Following consent, respondents were guided through a tailored set of questions based on the sector that are active in Sector: Agriculture, forestry, or both, and their role, which were grouped to: Tech operators, tech providers, or data sharers. An example of a survey page can be found in Figure 2.

This routing logic minimized respondent's burden while maximizing data relevance. The survey routing (see Figure 3: Survey Routing) ensured that only the applicable questions were displayed, improving response accuracy and engagement. The survey is available in Appendix B.

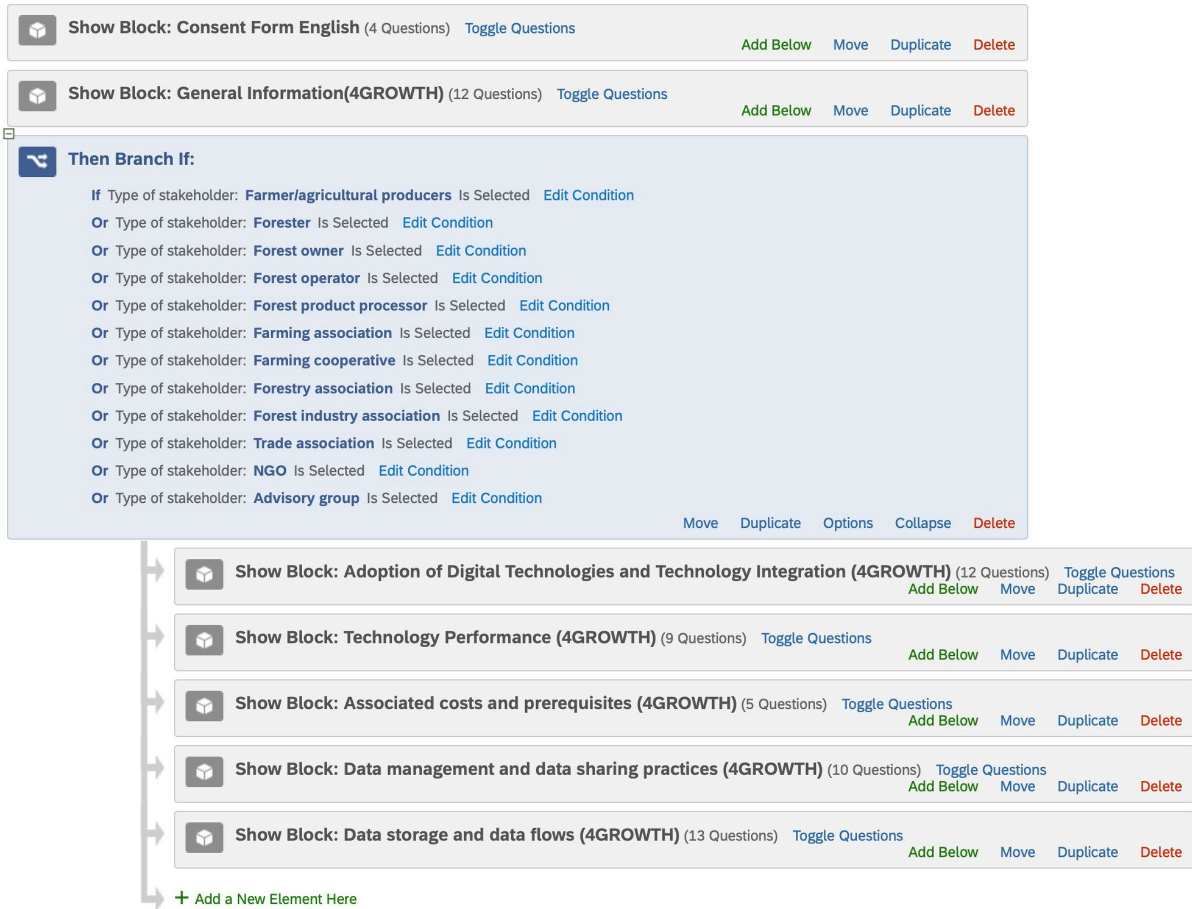

Figure 3: Part of survey routing for the English survey

The combined capabilities of the CDP and Qualtrics platforms ensured the survey was both robust and adaptable, meeting the multilingual, multi-stakeholder demands of the 4Growth project.

2.2 Survey distribution

Each observatory was responsible for data collection within specific geographical regions. Wageningen Research focused on agriculture in the Netherlands, ILVO on agriculture in Belgium, the Agricultural University of Athens on agriculture in Greece, and the Aristotle University of Thessaloniki on forestry in Greece. INTIA covered agriculture in Spain, CTIFL focused on agriculture in France, VTT handled forestry in Finland, and Agri-Food Lithuania DIH was responsible for agriculture in Lithuania, Hungary, and Poland. This allocation ensured representation from diverse regions in the European Union, capturing insights into digital technology adoption.

The surveys were distributed through a variety of channels. Most observatories used email marketing, social media, direct calls or personal outreach, and in-person events. QR codes were frequently generated for each survey, allowing stakeholders to easily access the questionnaire via smartphones. Additionally, the English-language survey contained links to all translated versions, ensuring consistency and accessibility across languages. The surveys were also distributed via general 4Growth channels, including LinkedIn posts and the 4Growth website, to reach a broader audience.

The survey promotion was further extended at generic events, such as the Synergy Days held in Barcelona. The Synergy Days, known for fostering collaboration across digital agriculture and forestry initiatives, provided an excellent platform for engaging diverse stakeholders. During these events, 4Growth representatives actively promoted the survey through dedicated headers and materials, distributing it further among attendees.

The distribution methods used for survey outreach varied across the seven observatories. The most commonly employed method was email marketing, utilized by all observatories. Social media was the second most frequently used, adopted by five observatories. Four observatories organized in-person events to promote the survey, and three observatories relied on direct calls or personal outreach. Two observatories mentioned other methods, specifically newsletters focused on the sector, and leveraging direct connections. Notably, none of the observatories used advertisements or sponsored posts for survey distribution. The effectiveness and challenges of these methods, including what worked and what did not, are explored further in Chapter 3.

2.3 Responses and data quality

Figure 4 below represents a summary of the survey responses from the first wave. There is overall a large gap in the number of unique responses who have started the survey and the amount that have answered enough questions to provide enough data, that is useful for the 4Growth project. The surveys marked as not providing enough data usually only provided their name and the farm type or stakeholder type.

The distribution of stakeholder types overall is relatively even, but the responses are skewed towards each observatory which is expected as they focus on either the forestry or agricultural sector. There are notable disparities between collection rates of the different country surveys which is explained in the next section on experiences on data collection. The average response time is not overly long, but the figures are skewed by the large amount of people who opened the survey and answered one or two questions and exited. The response time is closer to **20** minutes for respondent who completed the survey.

Table 3: Statistics of survey respondent measures

Survey	Average response time - seconds	Unique responses	Completed responses	Agriculture stakeholder	Forestry stakeholder	Agriculture & Forestry Stakeholder
English	61469	44	33	6	24	3
Dutch	21168	26	15	15	0	0
French	14305	28	12	11	0	1
Spanish	52284	68	55	53	0	2
Finnish	20378	23	21	0	17	4
Swedish	2104	1	1	0	1	0
Hungarian	0	0	0	0	0	0
Lithuanian	16419	16	16	12	0	3
Polish	0	0	0	0	0	0
Greek	33822	65	47	0	41	6
Total	**14.3 minutes**	**271**	**200**	**97**	**83**	**19**

In figure 5 we can also see that there is a large representation of responses from Spain and Greece and a more even response rate from Belgium, Finland, France and Lithuania. While at least there is representation from Western, Southern, Eastern and Northern Europe, more can be done to ensure a more equal distribution of country responses and involving multiple countries from each region. It is important to note that we could not directly match the selected countries of the survey respondents with the observatories. This is due to several factors, such as having two observatories in Greece. Additionally, the survey did not explicitly ask respondents which observatory had invited them, and we were unable to trace responses back to a specific observatory via the survey link or other tracking mechanisms. For future iterations, it is recommended to establish a system to capture this information, enabling better evaluation of observatory performance and outreach efforts.

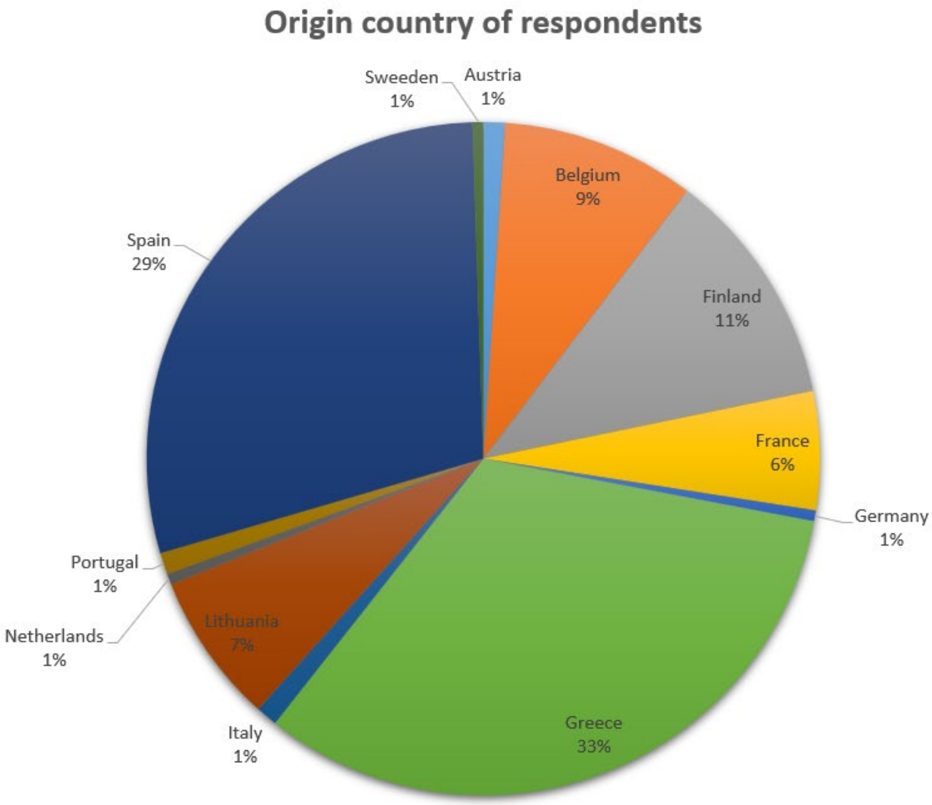

Figure 5: Selected country of survey respondents as a % of all responses.

In figure 6 we can see the breakdown of responses by stakeholder type which gives further insight into how representative wave 1 has been. We can see observatories have been successful in getting responses from primary stakeholder in the agriculture and forestry value chains and advisory groups. However, we see low response rates from data providers, infrastructure providers, platform providers and other stakeholders who are data brokers and operate in the upstream section of value chains. Prioritising these types of stakeholders is crucial in the next data collection waves so we can get a true insight into the uptake and market for digital technologies in agriculture and forestry sectors. One reason for the low response rates is that observatories are well connected to primary producers but have less reliable network for other stakeholders who are only possible to reach by email or third parties. This will require the 4Growth project to find new innovative ways of collecting data on these upstream actors. Additionally, the 4Growth consortium will need to mobilise its network to better reach and communicate with these data brokers and technology providers.

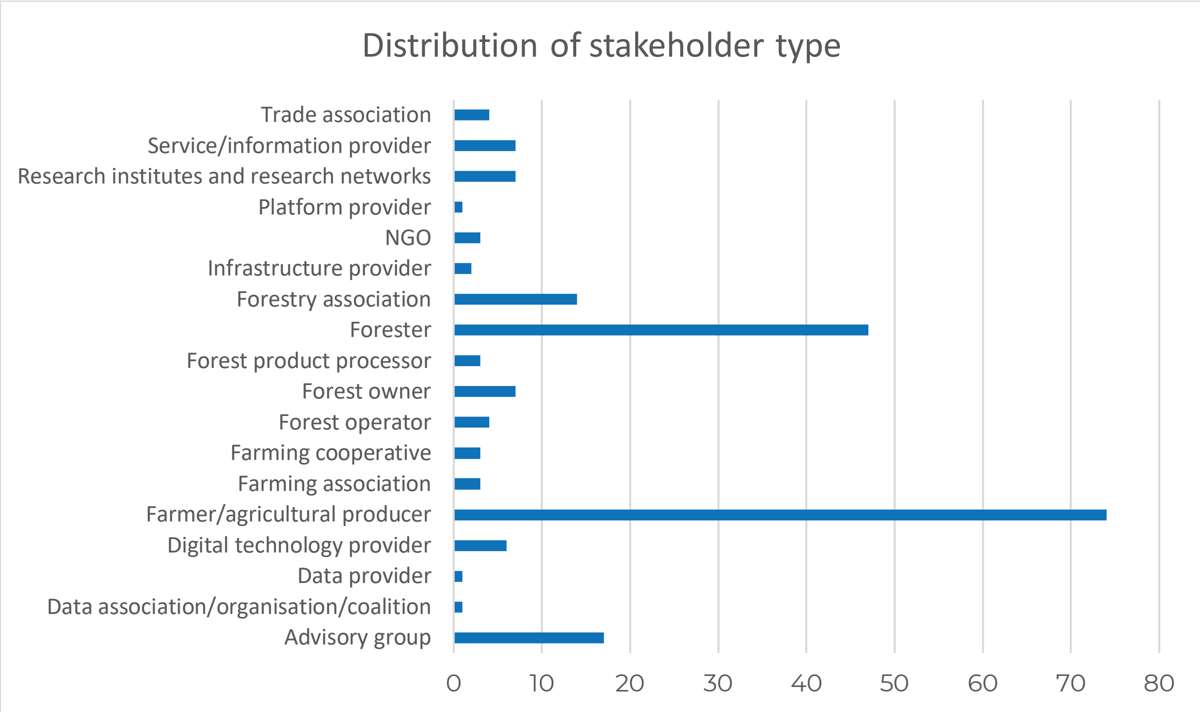

Figure 6: Distribution of stakeholder type of survey respondents.

3.Experiences with data collection

To gain insights into the experiences of data collection, we decided to survey the surveyors themselves. After all, who better to understand the challenges of collecting responses than those who do it regularly? All seven observatories participated in this meta-survey, providing valuable reflections on what worked, what did not, and what could be improved in future waves.

3.1 Lessons learned

Effectiveness of Data Collection Methods

Observatories were asked to identify the most and least effective data collection methods. In-person events emerged as the most effective approach, with four observatories selecting this option. Direct calls or personal outreach followed, with two observatories identifying it as their preferred method. Email marketing and other methods, such as utilizing direct connections, were less frequently mentioned. These results highlight the importance of direct and personal engagement in achieving successful data collection efforts.

In contrast, when asked about the least effective methods, email marketing was identified as the least effective by five observatories, followed closely by social media, cited by four observatories. One observatory also mentioned newsletters as a less effective option. These findings suggest that less personalized approaches, particularly those relying on broad, impersonal outreach, may yield lower engagement levels compared to direct and interpersonal strategies.

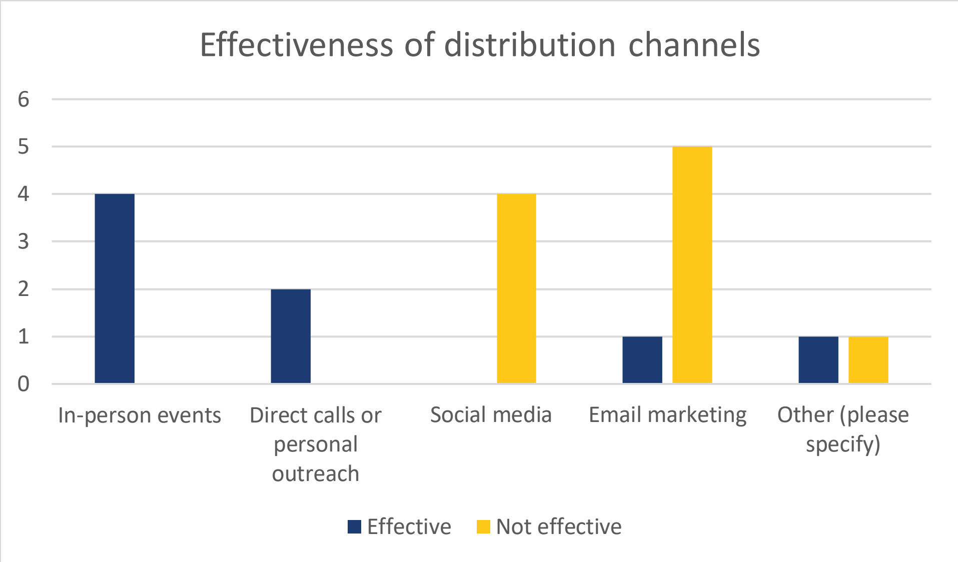

Figure 7: Effectiveness of survey distribution according to observatories (n=7)

Strategies to Prevent Response Bias

Observatories employed various strategies to minimize response bias in their surveys. Conducting random sampling and providing explanations to clarify complex questions were the most commonly used approaches, with three observatories implementing each. These methods aimed to enhance the reliability of responses by reducing potential misunderstandings and ensuring diverse representation.

Other strategies included targeting diverse demographics, mentioned by two observatories, and reaching out to less digitally engaged stakeholders, used by one observatory. Additionally, observatories explored customized approaches such as targeting individuals with varying levels of digitalization, compiling comprehensive lists of companies and organizations across different stakeholder groups, and leveraging professional networks to pass along survey invitations. Some observatories also provided opportunities for stakeholders to respond during in-person events, further diversifying the respondent pool.

Challenges in Reaching Certain Groups

Several (5) observatories reported encountering challenges in reaching specific groups, while two others indicated no significant difficulties. Among those facing challenges, low response rates across stakeholder groups and difficulty engaging farmers in particular regions, such as Poland and Hungary, were common issues. Other frequently identified issues were:

1. **Survey Saturation and Complexity:** The abundance of surveys in the sector and the use of technical vocabulary in lengthy questionnaires were cited as barriers. Farmers, often with limited time and education levels, found these factors overwhelming.
2. **Timing Constraints:** Contacting professionals, especially farmers, during their busiest seasons (e.g., harvest or sowing periods) led to low response rates. The ideal time for engagement was identified as the winter months.

3. **Remote Engagement Barriers:** Observatories noted difficulties in engaging stakeholders through email or newsletters, with limited visibility on which channels were effective.
4. **Regional Challenges:** In Poland and Hungary, reliance on local partners was critical due to limited capacity for in-person engagement.

Challenges During Survey Distribution

Survey distribution faced several challenges, particularly in ensuring broad participation and effective communication with stakeholders. One observatory had an issue with outdated contact information when relying on social media and websites to gather emails and phone numbers, leading to a high number of undelivered email notifications. This posed barriers to reaching intended respondents and required additional efforts to update contact lists and explore alternative channels.

The complexity and design of the survey also created obstacles. Farmers, in particular, struggled with accessing and completing the survey due to unfamiliarity with digital tools and terminology. The length and repetitive nature of the questionnaire further discouraged engagement, with respondents in some regions, such as Belgium, expressing fatigue with similar surveys. These challenges underscored the need for more accessible language, concise question design, and targeted outreach to increase response rates and improve the overall survey experience.

Willingness and Barriers for Stakeholder Participation

Observatories reported various barriers that impacted stakeholder willingness to participate in the survey (see Figure 8). Time constraints were the most frequently mentioned issue, highlighted by six observatories, as stakeholders often struggled to find time to complete the survey amidst competing priorities. Five observatories noted that the survey length posed a significant challenge, with lengthy questionnaires frequently resulting in unfinished responses. Four observatories identified issues with survey quality, such as overly technical or unclear questions, which discouraged participation, especially among stakeholders with less familiarity with digital topics.

A lack of interest in digitalization and related survey topics was cited by three observatories, reflecting limited stakeholder motivation to engage. Two observatories mentioned other barriers, including general disinterest in survey participation or fear of clicking on unfamiliar links. One observatory reported concerns about data privacy and usage, which deterred some stakeholders, while another observed that variations in digital literacy affected stakeholders' ability to complete the survey. These findings emphasize the need for shorter, simpler surveys, tailored outreach strategies, and reassurance about data security to improve participation in future waves.

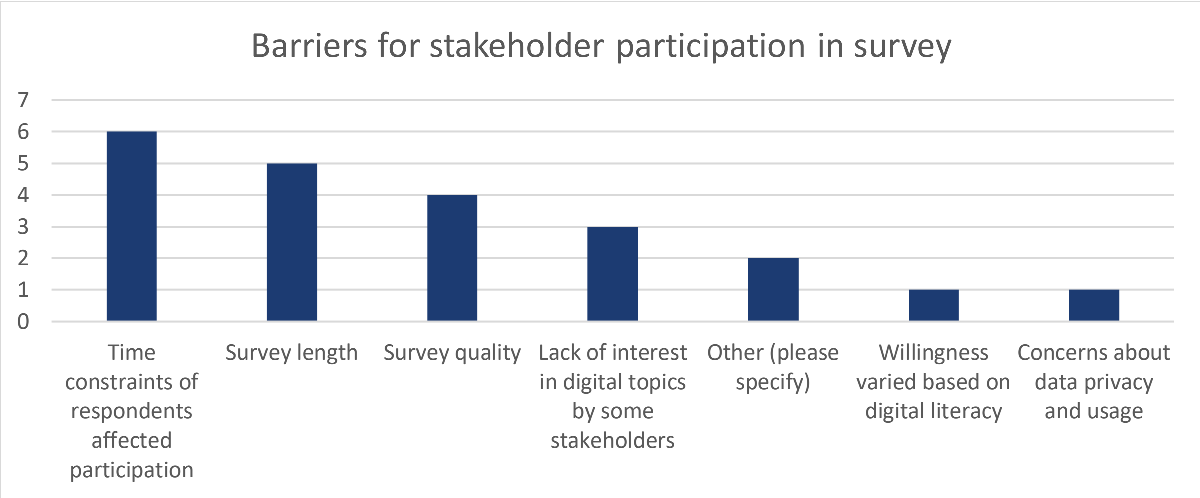

Figure 8: Barriers for participation according to the observatories (n=7)

3.2 Good practices

One of the key practices highlighted by observatories is the importance of tailoring survey distribution methods to the target audience. While in-person approaches are highly effective for fostering engagement and clarifying questions, they are often time-consuming and not scalable for reaching larger groups. Email emerged as a practical and efficient method, enabling stakeholders to respond at their convenience and reaching a broader audience. Leveraging personal networks and contacts of experts also proved successful, particularly in sourcing relevant stakeholders through trade publications and forestry-specific media.

Timing and context were also emphasized as crucial factors for successful survey distribution. Scheduling outreach during less busy periods for stakeholders, such as the off-season for farmers, increased the likelihood of participation. Direct outreach, such as in-person questionnaires or interviews, was highlighted as particularly beneficial for addressing complex questions and ensuring respondents felt supported throughout the process. This combination of strategic timing, diverse channels, and personalized interaction was seen as instrumental in maximizing participation and response quality.

3.3 Suggestions for future iterations

To improve survey effectiveness and stakeholder engagement in future iterations, observatories provided several suggestions focused on simplifying the design and enhancing the user experience. A key recommendation is to reduce the length of the survey, with some observatories suggesting that the duration should not exceed five minutes. Keeping the survey concise and focused on essential questions will minimize respondent fatigue and increase completion rates. Informing stakeholders in advance about the short duration can also encourage participation.

Simplifying the language of the survey emerged as another critical priority, particularly for audiences like farmers who may not be familiar with technical terminology. Observatories emphasized the importance of creating a user-friendly environment by avoiding complex phrasing and tailoring questions to the audience’s level of understanding.

Other recommendations included removing redundant options, such as “I do not know,” to focus responses and improve the clarity of the results. Clearer objectives and a well-structured questionnaire aligned with specific survey goals were highlighted as essential elements to enhance the overall quality and utility of the survey data. Collectively, these adjustments aim to create a streamlined, accessible, and effective survey experience for stakeholders in future waves.

4.Planning for data collection in Wave 2

This section outlines the planned adjustments for Wave 2 of data collection, incorporating the findings and suggestions from Wave 1. It highlights the steps to address the challenges identified in Chapter 3, such as survey length and language complexity, and explores new methods, including automated data collection, to streamline the process and reduce respondent's burden. The results of these adjustments aim to enhance data quality and increase participation in the next wave.

Chapter 4.3 introduces an additional method to complement traditional data collection approaches, which was not part of the original task plan. Automated data collection and analysis leverages web scraping and AI-driven tools to gather insights directly from public sources, such as company websites. This approach aims to address gaps identified in Wave 1 by broadening the dataset and reducing the reliance on resource-intensive survey methods.

4.1 Distributing targeted surveys

In Q4 2024 and Q1 2025, the suggestions made in Chapter 3 will be implemented to refine the survey for Wave 2. These improvements focus on shortening the survey, simplifying language, and ensuring questions align with the primary objectives. Adjustments will be developed through discussions within the consortium and finalized during the annual meeting in February 2025. This collaborative approach ensures that the survey design benefits from diverse perspectives and addresses the challenges identified in Wave 1.

The updated survey will aim to engage a broader range of stakeholders while maintaining the consistency and focus provided by the Digital Agriculture and Forestry Uptake Assessment Grid. Distribution strategies will also be reviewed to incorporate a mix of methods, including digital outreach, in-person engagement, and automated approaches where feasible, as detailed in subsequent sections.

4.2 Interviews & Events

For the entire project, the goal is to utilize the Assessment Grid in over 5000 cases, encompassing various actors, entities, and sectors such as agriculture, horticulture, and livestock. Regardless of the method of data collection, whether it is through interviews or surveys, the final dataset will be standardized for each observatory. Using the WR meta-data platform “Adagio”, the consortium will be able to collect and aggregate questions from different formats into a standardized form, which can then be used for further data analysis purposes.

4.3 Automated data collection and analysis

Besides traditional methods for measuring the adoption of digital technologies, such as the survey outlined above, emerging approaches like automated data collection offer promising alternatives. This subsection explores how automated techniques, including web scraping and AI-driven analysis, can streamline data collection efforts, potentially reducing the workload for the second wave while providing valuable insights into digital technology adoption.

Automated data collection and analysis: Introduction

The aim of this automated data collection was to get more information about the use of digital technologies in the context of the forestry sector. The aim is to complement the information collected by the survey. In this study we tested how well we could automatically find answers to the selected questions in the survey by using information on companies' websites.

Company websites are valuable sources for understanding companies' use of digital technologies and other activities because they often highlight their latest innovations, tools, and strategies to attract customers and stakeholders. They typically provide insights into their digital transformation efforts, product offerings, and industry focus through sections like blogs, case studies, and news updates. Websites are also an authoritative and direct source, ensuring the information aligns with the company's official messaging and branding.

The automated data collection methodology leverages Large Language Models (LLMs) to efficiently gather and analyze information from companies' websites. This approach uses advanced natural language processing and retrieval augmented generation (RAG) to scan and interpret website content, identifying relevant information based on pre-defined criteria according to the survey questions. By collecting and processing the content of the selected websites, the methodology can spot the relevant information, streamlining the data extraction process and ensuring high relevance in the information collected. This automated method reduces manual effort in identifying companies' activities related to forestry, which provides scalable solutions for comprehensive data collection and processing. As a proof of concept, this approach was applied exclusively to the forestry sector and did not include agriculture. Future iterations may explore expanding this method to include agricultural data collection if the concept proves successful.

Automated data collection and analysis: methods

Selection of companies

We used the Orbis database, which provides detailed information on companies and entities worldwide, to select forestry and forestry sector-related companies for the analysis and to identify their website addresses for inclusion in the study. Orbis, developed by Moody's, is a comprehensive resource for company data, offering insights into financials, ownership structures, and industry classifications(Moody's, 2024) .

Search steps in Orbis included these definitions:

- Status: Active companies
- NACE code definitions
- World region/Country/Region in country: European Union (27)
- All companies with a web site address

- Number of employees: min=5, Last available year, exclusion of companies with no recent financial data and Public authorities/States/Governments

For more general companies such as technology companies additional “Activity text search” step was included to get companies that work in the field of forestry.:

- Activity text search:
  Must have 'forestry OR "forest management" OR "forest monitoring" OR "forest conservation" OR "forest ecosystem" OR "forest health" OR "forest inventory" OR deforestation OR reforestation' (in Brand names, Description and history, Size estimate, Full overview, History, Main activity, Main customers, Main distribution site, Main domestic country, Main foreign countries or regions, Main production site, Main sales representation sites, Membership of network, Primary business line, Primary national activity, Product and services, Secondary activity, Secondary business line, Secondary national activity, Strategic alliances, Strategy organization and policy, Trade description)

Different set of companies were defined by Nomenclature of Economic Activities (NACE) codes (See Table 3) as potential source for collecting information. Challenges in NACE classifications are ambiguities in accurately categorizing businesses with diverse or overlapping activities, which can lead to misclassification. Companies working in forestry can be found for example under “Computer programming activities”, “Manufacture of air and spacecraft and related machinery”, “Freight rail transport” or “Other research and experimental development on natural sciences and engineering”.

In addition of using definitions of NACE codes we separately picked technology companies relating for example to forestry and earth observation by using reports such as EUSPA market report 2024 (EUSPA Programme, 2024) and VTT internal report about Drone Market in Europe and Finland (Jutila & Sinkkilä L., 2023).

Table 3 NACE code definitions by different groups of companies and total count of companies when all the search steps were defined.

Group name	NACE code definitions	Count
Forestry group	0210 - Silviculture and other forestry activities, 0220 - Logging, 0230 - Gathering of wild growing non-wood products, 0240 - Support services to forestry, 1610 - Sawmilling and planing of wood, 1711 - Manufacture of pulp, 1712 - Manufacture of paper and paperboard	8176
Forestry machinery	2830 - Manufacture of agricultural and forestry machinery, 4661 - Wholesale of agricultural machinery, equipment and supplies	6583
Selected tech company list	List of defined companies, linked to Orbis data	44
Technology companies	3030 - Manufacture of air and spacecraft and related machinery, 6201 - Computer programming activities, 6202 - Computer consultancy activities, 6203 - Computer facilities management activities, 6209 - Other information technology and computer service activities, 6820 - Renting and operating of own or leased real estate, 7112 - Engineering activities and related technical consultancy, 7219 - Other research and experimental development on natural sciences and engineering + Activity text search	223
Other miscellaneous	1920 - Manufacture of refined petroleum products, 2611 - Manufacture of electronic components, 2620 - Manufacture of computers and peripheral equipment, 2630 - Manufacture of communication equipment, 2651 - Manufacture of instruments and appliances for measuring, testing and navigation, 2670 - Manufacture of optical instruments and photographic equipment, 4613 - Agents involved in the sale of timber and building materials, 4651 - Wholesale of computers, computer peripheral equipment and software, 4652 - Wholesale of electronic and telecommunications equipment and parts, 4920 - Freight rail transport, 4941 - Freight transport by road, 5210 - Warehousing and storage, 5221 - Service activities incidental to land transportation, 5222 - Service activities incidental to water transportation, 5224 - Cargo handling, 5229 - Other transportation support activities + Activity text search	247
Forest Industry – other manufacture of paper products	1721 - Manufacture of corrugated paper and paperboard and of containers of paper and paperboard, 1722 - Manufacture of household and sanitary goods and of toilet requisites, 1723 - Manufacture of paper stationery, 1724 - Manufacture of wallpaper, 1729 - Manufacture of other articles of paper and paperboard	5474
Forest Industry – other manufacture of wood	1621 - Manufacture of veneer sheets and wood-based panels, 1622 - Manufacture of assembled parquet floors, 1623 - Manufacture of other builders' carpentry and joinery, 1624 - Manufacture of wooden containers, 1629 - Manufacture of other products of wood; manufacture of articles of cork, straw and plaiting materials	14 245

Selection of survey questions and prompt design

The survey includes highly detailed questions targeting various stakeholders; however, it is unrealistic to expect that all of these can be answered through information available on company websites. Therefore, we selected a subset of questions from the survey, prioritizing those that companies are more likely to address or discuss on their websites. We selected 46 questions from total of 88 questions in the survey (See Table 4).

Table 4 Selected survey questions for the automatic analysis

Survey Code	Survey question
type	Type of stakeholder:
prim.forest	Primary Area of Operation in forestry
GM02	What type of governance model do you operate under?
GM04	Are there regulatory considerations influencing the governance of digital technology adoption?
itegrated.digi.tech	Has your organisation integrated digital technologies into its workflows?
digitaltech.forestry	What type of digital technology has been used for forestry? - Selected Choice
goals.to.adopt	Were there specific goals or challenges that prompted the adoption of digital tools?
lvl.digitalisation	How would you rate the level of digitalization in your farming/forestry practices on a scale of 1 to 5 (1 being low, 5 being high)
prim.function.tech	What are the primary functions of these technologies in the agriculture or forestry value chain?
adopt.level.tech	What is the adoption level of these technologies?
challenges.tech.adopt	Have you encountered challenges in the adoption of digital technologies?
further.integration	Are there specific barriers hindering further integration?
digitech.userneeds	To what extent do digital technologies meet evolving user needs within your organization?
adv.tech	What are the advantages of the used technologies? - Selected Choice
limitations.tech	Have you encountered any perceived limitations or challenges in utilising these technologies?
Dmdsp7.1	Is data collected from your farming/forestry activities?
Dmdsp7.2	What type of data sharing practices related to digital technology does your organisation use?
Dmdsp7.3	What type of data do you collect?
Dmdsp7.4	Do you pay for this data?
Dmdsp7.5	What type of tools or platforms do you use to collect data?
Dmdsp7.6	Do you share this data?
Dmdsp7.7	Do challenges exist in sharing and interoperability of agricultural and forestry data?

Dmdsp7.9	How do these practices contribute to or impede the overall effectiveness of technology adoption?
DSDF8.2	Do you use cloud services/data centers?
DSDF8.4	Are there economic implications associated with data flows in these sectors?
DSDF8.5	Do data flows enhance productivity and efficiency in agriculture and forestry?
DSDF8.6	Do you use data analytics for decision-making?
DSDF8.7	Where do you receive data from and how much?
DSDF8.8	What type of data do you receive or provide?
DSDF8.10	Where and how do you store this data?
DSDF8.11	What do you do with this data?
DSDF8.12	To whom and where do you send derived information or data?
social.benefits	Have you experienced social benefits through the use of digital technologies?
job.creation	How have digital technologies impacted job creation?
social.impact	What is the overall social impact of adopting digital technologies?
digitech.costsavings	Have digital technologies resulted in cost savings or increased efficiency?
digitech.savinginputs	Have you seen savings in inputs due to digital technologies?
impact.digitech	What is the overall economic impact of implementing digital technologies?
digitech.sustainability	Have digital technologies contributed to sustainability and environmental practices?
digitech.impacts.footprint	Have you observed positive impacts on resource conservation or environmental footprint?
digitech.energy.efficiency	Have digital technologies contributed to energy efficiency?
digitech.biodiversity	Have you observed any positive or negative effects on biodiversity in agricultural and forestry areas due to digital technology adoption?
digitech.track.sustainability	Do you use digital technologies to track and ensure adherence to sustainable farming practices and forestry activities?
plan.upgrade.digitech	Are there plans to expand or upgrade your current digital infrastructure?
facilitate.expansion.upgrade	What would help facilitate the expansion/upgrade of digital infrastructure in the future?
type.developments	What type of developments do you anticipate in the near future?

These questions were design as prompts for the AI analysis. An example of such prompt is presented in the Table 5.

Table 5 An example prompt design for the survey question relating to Primary Area of Operation in forestry

Survey Code	Query

prim.forest	Identify the organization's primary area of operation in forestry. Classify the organization's primary area as one of the following categories: "Reforestation", "Forest conservation - thinning, pruning, weed & pest control," "Felling," "Transportation of logs", "Non-Timber Forest Products (NTFPs)", "Forest Fire Management," "Forestry inventory and mapping", "Wildlife management" or "Other." Provide both the selected primary area and a detailed explanation of why this area was chosen. If the primary area is "Other," provide a detailed description of what this organization's primary area of operation in forestry entails. If no relevant information is available regarding the query, respond with "No information".
	
	Return the information in the following format:
	{
	"prim.forest": "The selected primary area",
	"prim.forest.description": "Detailed explanation of why this area was chosen"
	}
	If there is no relevant information, respond with:
	{
	"prim.forest": "No information"
	}

Web scraping company websites

Web scraping is a powerful technique for automatically gathering information from websites. The methodology of web scraping in this experiment is a structured approach to web scraping that utilizes Python and several libraries to create a sophisticated data collection tool. The scraper begins with an initialization phase, where parameters are set to control the scope and depth of the data collection. Two key settings include the maximum depth and the token count threshold. Maximum depth determines how many layers of linked pages the scraper can visit from the original webpage, while the token count threshold limits the volume of text gathered. Since the volume of the content might be different in different websites, the token limit breaks the massive number of contents into smaller chunks, which facilitates the processing of the collected information. Once the parameters are initialized, the scraper accesses a webpage by sending a request to retrieve its content. This request returns the raw HTML of the page, which contains everything from text to images, scripts, and styles. To make the information usable, the scraper uses BeautifulSoup, a tool for parsing HTML. BeautifulSoup "cleans" the HTML content to leaves behind only the meaningful text, which is then stored in a structured format.

A unique feature of this scraper is its ability to handle URLs carefully. The scraper verifies each URL's format and adjusts it if needed, ensuring smooth access to the page. This feature helps maintain efficiency in data collection by eliminating issues caused by improperly formatted URLs. After retrieving and cleaning the main page's content, the scraper has the ability to follow links within the page to gather further information. This process is known as "recursion" and is controlled by the maximum depth setting. At the first level, the scraper gathers content from the initial page and stores it. Then, if the maximum depth has not been reached, the scraper identifies all the links within the current page and checks whether each link leads to a new page on the same website. If a link meets these criteria and hasn't already been visited, the scraper will access it, retrieve the content, clean it, and add it to the data collection. This depth-controlled recursion enables the scraper to expand its scope gradually without becoming overwhelmed by a seemingly infinite chain of linked pages. Finally, the scraper compiles all the gathered content into a structured format, organized by depth.

The scraper also includes error handling for common web-related issues such as timeouts, SSL certificate errors, and connectivity problems. If any of these issues occur, the scraper displays an error message and stops the data collection for that page, allowing the rest of the process to continue uninterrupted.

It should be mentioned that the web scraping process in this experiment is designed to ensure privacy and adhere to ethical guidelines. The scraping procedure does not collect or store any personal information, focusing solely on publicly available textual content from websites. Furthermore, the scraped data is processed in real-time to extract and process relevant insights, after which it is discarded and not stored in any format. This approach complies the privacy policy of website by limiting data collection to non-sensitive, publicly accessible information and avoiding the retention of web-scraped data. The methodology emphasizes transparency, ethical practices, and compliance with privacy standards throughout the scraping process. To ensure the security of data processing, all textual information was analysed exclusively on VTT's internal infrastructure. Moreover, we only report aggregated results, which prevent the identification of individual companies based on the findings.

AI analysis process

In this project, we employed the Retrieval-Augmented Generation (RAG) model for information retrieval from extensive textual datasets. RAG is a methodology that combines the strengths of traditional retrieval-based approaches and generative capabilities of language models. This hybrid approach allowed us to address the challenges inherent in analysing large volumes of unstructured text while ensuring the scalability and adaptability of the retrieval process. Below, we detail the methodological steps undertaken to implement, optimize, and evaluate the RAG framework. The Figure 9 also shows the overall methodology has been implemented in this task.

ك

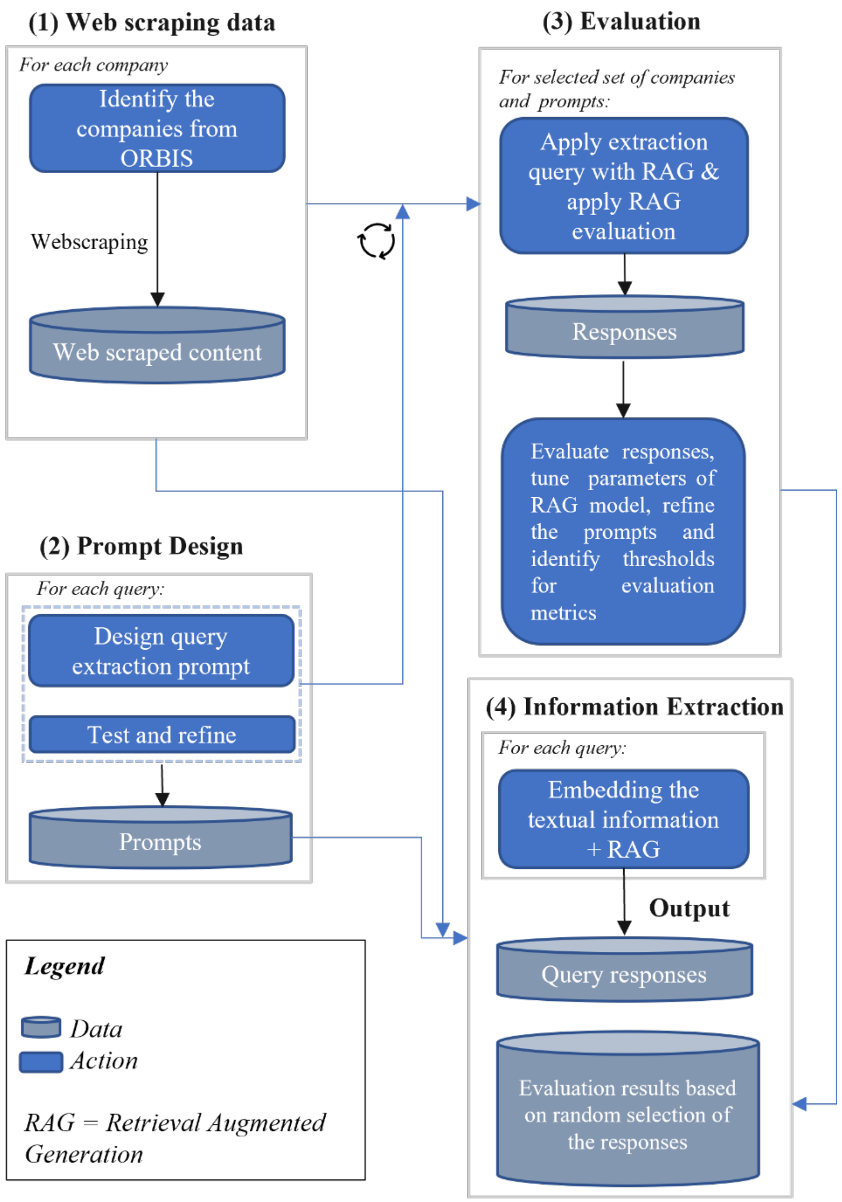

Figure 9: Methodology of the data collection and analysis

Utilizing RAG for Information Retrieval

The first step in our methodology involved deploying the RAG framework to analyze massive amounts of textual data. One of the primary advantages of using RAG lies in its ability to retrieve relevant information from extensive corpora by dynamically combining retrieved knowledge with the generative capabilities of language models (Lewis et al. (2020), Gao Xiong Y. Gao X. et al. (2023)). Unlike traditional information retrieval systems that rely solely on keyword matching or fixed retrieval pipelines, RAG leverages pretrained language models to enhance the relevance and depth of retrieved information. This made it particularly effective in handling the complexity, diversity, and volume of textual data in our dataset. The RAG process involves transforming user prompts into semantic representations to retrieve relevant information from an indexed external database. The retrieved data is then combined with the LLM's pre-existing knowledge, producing responses that are both more accurate and reliable.

Auto-merging retrieval

To further enhance the performance of the retrieval process, we implemented auto-merging retrieval. This method proved superior in collecting scattered data points distributed across various sections of the textual dataset. Auto-merging retrieval offers a sophisticated solution by establishing a hierarchical structure that organizes smaller chunks of text under larger parent chunks. Each smaller chunk is linked to a corresponding parent chunk, which can encompass multiple child chunks. During the retrieval process, a heuristic is applied to merge smaller chunks into their parent chunk if they surpass a predefined percentage threshold. This ensures that, instead of returning disjointed snippets, the system retrieves a cohesive and comprehensive parent chunk, resulting in a smoother flow of information.

However, to enhance the performance of the retriever on the same dataset, certain parameters can be adjusted to improve the model's efficiency. We evaluate the retriever's performance using RAG triads, as explained in the following subsection.

Evaluation process and metrics

Considering the different parameters for the retriever including the chunk size, as well as number of chunks fed into the large language models, can affect the performance of the RAG model. We tuned the RAG model parameters to optimize its performance. The optimization process utilized RAG triad evaluation metrics, which focus on RAG triad metrics. This framework assesses the performance of the Retrieval-Augmented Generation (RAG) implementation and its outputs across three critical metrics: Answer Relevance, Context Relevance, and Groundedness.

Answer Relevance evaluates the alignment of the LLM’s response with the given prompt, ensuring that the response remains on topic and effectively addresses the question posed. This metric serves to confirm that the LLM provides meaningful and pertinent answers rather than generating unrelated or tangential content.

Context Relevance focuses on the relationship between the user’s prompt and the information retrieved from the underlying documents. This metric assesses whether the retrieved data contains content that is relevant to the user’s query, thereby measuring the effectiveness of the retrieval process in supporting the LLM’s response generation.

Groundedness, sometimes referred to as Faithfulness, examines the degree to which the LLM’s response is substantiated by the information retrieved from the documents. This ensures that the answers are evidence-based and accurately reflect the source material, mitigating the risk of hallucinations or unverified claims in the generated content.

Through iterative adjustments to key model parameters, such as the number of retrieved documents, or chunk sizes, we fine-tuned the RAG model. This iterative tuning process also assisted in refining the prompts as well as identifying the threshold for the evaluation metrics.

In this experiment, we tuned the parameters of the RAG model based on the evaluation of 15 websites randomly selected from the sample. We also selected 8 prompts from the query set to run the information extraction process. Therefore, the final sample resulted in 120 instances. Subsequently, we executed the RAG pipeline across the samples, which included evaluating the RAG triad.

The experiments were conducted over multiple sets of retriever parameters. Specifically, we tested two configurations for chunk sizes: a two-level hierarchy with chunk sizes [512, 128] and a three-level hierarchy with chunk sizes [2048, 512, 128]. Additionally, we evaluated four different values for the number of top similar chunks retrieved by the retriever: 5, 10, 15, and 20.

The results of the RAG triad evaluation were reviewed in detail. This evaluation not only provided insights into the effectiveness of the parameter configurations but also guided minor modifications to the prompt design. The final scores for the RAG triad indicated that the optimal parameter configuration was the two-level hierarchy of chunk sizes [512, 128] combined with selecting the top 15 similar chunks. The RAG model also used OpenAI gpt4o-mini for the large language model, and text-embedding-3-large for the embedding model.

In addition to reviewing the results, analyzing the RAG triad evaluation metrics assisted in finding the thresholds for determining the validity of responses. After reviewing the results of the evaluation, the thresholds were established as follows: 0.56 for groundedness, 0.60 for answer relevance, and 0.33 for context relevance. These thresholds represent the minimum acceptable scores for a response to be considered valid.

In the initial evaluation dataset, the average scores for these metrics were 0.76 for groundedness, 0.96 for answer relevance, and 0.56 for context relevance. However, in instances where the queried information was not found on the website, the scores reflected this lack of information: answer relevance and context relevance were reported as 0, while groundedness was consistently reported as 1, indicating that the model grounded its response in the provided content despite the absence of relevant information.

The evaluation further revealed that 80% of the responses exceeded the threshold for groundedness, while 74% exceeded the threshold for context relevance. These results highlight the robustness of the RAG pipeline in generating valid and contextually grounded responses, even when dealing with diverse and challenging queries.

Running the RAG Model Across the Main Dataset

Following parameter optimization, the tuned RAG model was deployed across the main dataset. This step involved running each prompt over the entire corpus of textual data to extract relevant information.

Final Evaluation of Model Performance

To assess the RAG model’s performance comprehensively, we conducted a final evaluation on a randomly selected subset of the data points. This evaluation allowed us to validate the model's relevance, efficiency, and cost metrics on unseen data. The random selection ensures that the evaluation results were representative of the model’s overall performance across the dataset.

Lessons learned

During the experiment, we also attempted other approaches to address the automatic data collection and analysis. Below is a summary of the implemented methodologies and experimentations.

We employed a Large Language Model (LLM) with extended context windows, specifically the GPT-4o-mini model, to implement a structured prompt for information extraction. Instead of utilizing multiple queries in the RAG (Retrieval-Augmented Generation) framework, this method integrates all the queries into a single, comprehensive prompt, which is then provided to the language model. While this approach simplifies the process and reduces the need for multiple retrieval steps, it increases the risk of hallucinations due to the complexity and density of the prompt. In contrast, the advantage of the RAG approach lies in its ability to identify and retrieve context relevant to each query. This selective retrieval not only reduces the likelihood of hallucinations but also facilitates the evaluation of the generated answers by grounding them in specific, relevant source material.

Another challenge was identified with the use of long prompts. The current evaluation metrics, such as those used in TrueLens-type frameworks, are not well-suited for assessing the outputs of lengthy and complex prompts. These metrics require prompts to be more focused and concise to provide accurate and meaningful evaluations. Consequently, while structured prompts offer certain efficiencies, they demand careful design and evaluation strategies to mitigate risks and ensure reliable results. This also introduces challenges in evaluating the cost-efficiency of the RAG model, as increasing the number of prompts directly escalates the computational expenses of the experiment. The trade-off between the granularity of prompts and the associated costs necessitates careful consideration to optimize both performance and resource utilization.

Automated data collection and analysis: results

We implemented the methodology on the first group of companies selected based on their NACE codes. In this phase, the web scraping and text processing of websites faced several challenges and considerations, which ultimately reduced the sample to 1,549 companies.

One significant consideration was the processing of large websites. As the procedure involved processing website information down to the third layer of the domain, very large-scale websites resulted in an excessive number of tokens. To manage this, we decided to defer processing these websites to later stages. Among the 8,176 companies in the initial sample, 1,279 companies had websites containing more than 120,000 tokens within the first three layers, which were excluded for this phase.

Additionally, web scraping was not feasible for all websites, either due to problematic URLs listed on ORBIS or a lack of permission to scrape their domain. Therefore, a total of 3,479 websites were dropped out due to deduplication issues or being unsuitable for web scraping. Furthermore, 1,875 websites were excluded as they did not contain any information relevant to the survey questions. This left a final sample of 1,549 websites for the first phase of analysis. It is important to note that the processing of large-scale websites will be carried out in subsequent phases.

To provide an overview of the sectors and countries of origin of the companies, Figure 10 and Figure 11 present a summary of the 1,549 companies analysed in the first phase, categorised by country and NACE codes.

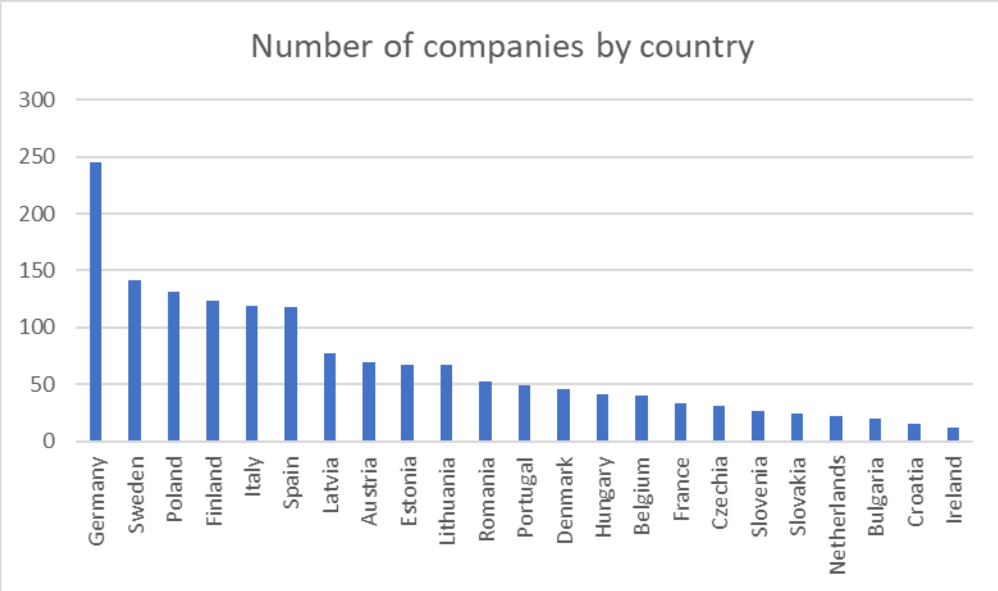

Figure 10: Number of companies by countries in the test set.

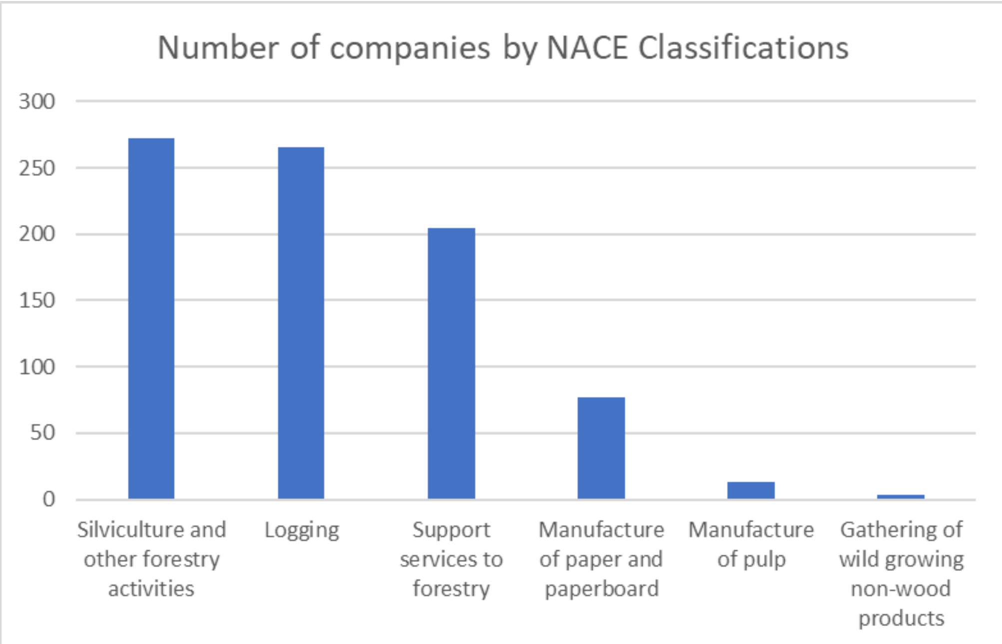

Figure 11: Number of companies by NACE classification in the test set.

We also analysed the coverage of responses to the 46 queries within this sample to assess how effectively the analysis was able to provide answers to various queries. The top five queries with the highest number of answers identified by the analysis were:

- Type of stakeholder:
- Primary Area of Operation in forestry
- What type of governance model do you operate under?
- What type of data do you receive or provide?
- What type of data do you collect?

In contrary, the queries for which the analysis found the fewest answers were:

- Have you encountered challenges in the adoption of digital technologies?
- Are there specific barriers hindering further integration?
- What is the overall economic impact of implementing digital technologies?
- Do you pay for this data?
- Do challenges exist in sharing and interoperability of agricultural and forestry data?

See Annex C for more detailed analysis of answer retrieval effectiveness for various queries.

The Table 6 provides an overview of how well companies were able to answer various queries on their websites. The "Count_queries" indicates the number of queries for which the analysis was able to find an answer on a company's website. In other words, 94% of companies provided answers to 10% of all queries, while 14% of companies answered 50% of the queries. There were 53 web sites (3%) that did not have any information relating to digital technologies in forestry.

Table 6: Coverage of answers to queries by companies. Count_queries shows for how many queries analysis was able to find an answer on a company web site.

Metric	Count	%
Not_any_information	53	3
Count_queries >=10%	1468	94
Count_queries >=20%	1121	72
Count_queries >=30%	797	51
Count_queries >=50%	231	14
Count_queries >=60%	70	4
Count_queries >=70%	26	1
Count_queries >=80%	4	0

Automated data collection and analysis: Discussion and Conclusion

Company websites serve as a resource for identifying the integration and application of digital technologies within organizations. By analyzing the content of these websites, we can gain insights into how companies leverage technological tools to provide services to customers and integrate them into their workflow. The analysis of company websites reveals priorities in digital transformation efforts, as they often showcase technological adoption and innovation through case studies or product announcements. Through the websites, companies like to highlight their adoption of emerging digital technologies and give information about technological trends within the industry. Company websites are regularly updated to reflect the latest innovations and achievements, providing researchers with up-to-date data to track evolving trends in digital technology utilisation.

This approach offers numerous opportunities, but it also has limitations. Compared to survey questions, which delve deeply into challenges related to technological adoption, including detailed insights into costs and barriers, such information is typically absent from company websites. However, company websites do provide more visible insights into the adoption of digital technologies and the integration of technology into business operations.

Furthermore, analyzing company websites allows for a broader perspective on the forestry sector across Europe, including insights into how stakeholders throughout the value chain collaborate using digital technologies. This approach supports a more holistic understanding of the industry's digital transformation. However, very small companies often lack websites, or if they do have them, the content is often limited to brief descriptions and contact details.

To define and identify relevant companies for analysis, we used NACE code classifications. While this serves as a good starting point, it presents challenges in distinguishing companies outside the forestry sector that still provide services to it. Another important consideration is determining how far along the value chain the analysis should extend.

In the current analysis, we focused on forestry-related companies and expanded the scope to include forestry processing companies, such as sawmills, pulp, and paper manufacturers. Additionally, we included technology companies and forest machinery manufacturers. We compiled different lists of forest industry-related companies, such as other manufacturers of paper and wood, as well as companies in related sectors, including transportation, warehousing, and the manufacture of instruments or electronic components. However, we did not include all the categories in the analysis at this stage.

In this study, we employed a survey structure to the responses, facilitating their integration into a visualization platform if the results prove relevant for inclusion. However, the survey relies on predefined response categories, which may not align seamlessly with the diverse and dynamic information available on websites or the outputs of LLM-based analysis. The implemented automated analysis not only classifies data into categories but also provides concise explanations for the category selections. To enhance flexibility and gain deeper insights, we included an “Other” category in some queries, allowing us to capture and explore additional types of categories that might emerge.

These extracted descriptions serve as a possible data source for further analysis. They enable the creation of summaries, offering an overview of topics related to the adoption of digital technologies in forestry. Such insights can inform other project tasks, such as horizon scanning in Foresight (T3.2). Additionally, this study contributes to T2.4 by exploring innovative approaches to market analyses, potentially supporting development of alternative ways of collecting data in monitoring of digital technology uptake in agriculture and forestry.

The analysis pipeline can be refined to adapt to evolving needs. For future research within the project, we could identify key tasks, revisit prompt designs, and focus on obtaining answers to more targeted questions. Furthermore, with the rapid advancements in generative AI and LLM technologies, experimenting with emerging features and models may provide enhanced tools to advance our work as the project progresses.

5.Conclusion

The first wave of data collection provided valuable insights into the adoption of digital agriculture and forestry technologies, despite the number of responses falling short of the project's expectations and needs. The main takeaway was that the survey design played a significant role in limited participation. Its length often led to participant drop-out, and the complexity of the language presented challenges for stakeholders, particularly those less familiar with digitalization concepts. These issues will be addressed in the next iteration of the survey to improve accessibility and completion rates.

The next version of the survey, planned for Q1 2025, will focus on reducing the survey length and simplifying the language. This adjustment aims to enhance respondent engagement and ensure that data collection efforts yield a greater volume of high-quality responses. The findings from this wave have established a strong foundation for subsequent data collection, but there is a clear need to significantly increase response rates in the future.

In-person engagement proved to be the most effective method for data collection, producing the highest response rates. However, it is also the most labor-intensive approach, making it impractical for large-scale implementation. Therefore, a mixed-methods approach will remain essential, balancing the effectiveness of in-person outreach with the scalability of digital methods such as email and social media. To effectively capture insights from upstream actors such as data brokers and technology providers, the 4Growth project will need to explore innovative data collection methods and leverage its consortium's extensive network for targeted outreach and communication.

Looking ahead, there is a potential to integrate AI-driven web-scraping technologies to identify certain indicators. As explored in Section 4, this approach could reduce the need for some survey questions, streamlining the data collection process. These innovations, alongside adjustments to survey design, will shape the next wave of data collection. Looking ahead the insights gained from this data collection will not only inform updates to visualization tools and modelling tasks but also provide a robust evidence base to shape future policy directions, supporting more targeted and effective strategies for the adoption of digital technologies in agriculture and forestry.

The results from the first wave will directly contribute to updates in the data visualization platform of Work Package 4 and provide input for modelling tasks supporting growth conservation. The project team will continue to refine its strategies to maximize stakeholder engagement and data quality in future waves.

References

Chappell, D. (2011). Introducing odata. *Data Access for the Web, The Cloud, Mobile Devices, and More*, 1–24.

Digital agri hub. (2024, December). *Digital agri hub*. Https://Digitalagrihub.Org.

FAO. (2024, December). *AgriTech Observatory: A virtual hub to monitor digital agriculture developments in Europe and Central Asia*. Https://Www.Fao.Org/Agroinformatics/News/News-Detail/Agritech-Observatory--a-Virtual-Hub-to-Monitor-Digital-Agriculture-Developments-in-Europe-and-Central-Asia/?Utm_source=chatgpt.Com.

Gao Xiong Y. Gao X. Jia K. Pan J. Bi Y. Dai Y. Sun J. Guo Q. Wang M. & Wang H., Y. (2023). Retrieval-Augmented Generation for Large Language Models: A Survey. In *ArXiv Preprint: Vol. nan*. https://arxiv.org/abs/2312.10997v4

Jutila & Sinkkilä L., E. (2023). Drone market in Europe and Finland. In *VTT Analyst Report: Vol. nan*.

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*, 23–37.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., & Rocktäschel, T. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, *33*, 9459–9474.

Moody's. (2024, December). *Orbis*. Https://Www.Moodys.Com/Web/En/Us/Capabilities/Company-Reference-Data/Orbis.Html.

Programme, E. U. A. for the S. (2024). *EO & GNSS Market Report 2024: Vol. nan*. https://www.euspa.europa.eu/publications-multimedia/publications/eo-gnss-market-report

Qualtrics. (2020). *Qualtrics survey software*. https://Qualtrics.com

Wageningen Economic Research. (2024a, September 4). *Adagio Data Warehouse*. Adagio.Wecr.Wur.Nl.

Wageningen Economic Research. (2024b, December). *Consumer Data Platform*. Https://Cdp.Wecr.Wur.Nl/Auth/Login/.

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big data in smart farming–a review. *Agricultural Systems*, *153*, 69–80.

ANNEX

Annex A: Survey informed consent form

Consent Form for the Survey

Project details

Project name	Project number	Start and end date
4Growth	2282300621	01-02-2024 – 01-02-2028

Please take some time to read this information and ask questions if anything is unclear. Contact details can be found in this document.

Brief project information

Wageningen Economic Research is coordinating the 4Growth Horizon Europe project and the research for this study is being undertaken by consortium partners. This survey is part of the activity *Observatory Data Collection and Analysis,* undertaken by *all partners except EVF,* This project is funded through the Horizon Europe funding program of the EU Commission (grant agreement No. 101016807). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

The 4Growth Horizon Europe project, aiming to advance digital solutions in agriculture and forestry, involves the collection and processing of certain data from stakeholders in the agriculture and forestry sectors. To ensure compliance with the General Data Protection Regulation (GDPR), we seek your explicit and informed consent before proceeding.

Purpose of Data Collection

The collected data will be used for the development and implementation of the Digital Agriculture & Forestry Uptake Assessment Grid. This tool aims to document various aspects of technology adoption, integration, costs, prerequisites, governance models, socio-economic benefits, and more. The insights gathered will contribute to advancing digital solutions in the agriculture and forestry sector. Aggregated data will be used for reporting and visualizations.

Deliverables

■ Paper ■Report ■ International research ■Other: Visualization dashboard.

Data Handling and Security Measures

We assure you that your data will be handled with the utmost care and confidentiality. Security measures, including encryption and access controls, will be implemented to safeguard the information collected. Personal data will be anonymized in any reports, infographics or publications.

The information will be retained by *Wageningen Research* and will only be used for the purpose of research. The security of personal data and by the processors takes place on the basis of generally accepted standards and best practices. Please refer to *the organization information security policy: e.g.:* *WUR information security policy* and the *organization data policy: e.g.:* *WUR data policy* for more details.

Which data is collected, where is it processed and for how long

Data category (you can add and delete)	Location/ country of processing	Storage period
Company Name	Netherlands	10yrs
Country of residence	Netherlands	10yrs
Survey/interview responses	Netherlands	10yrs

Organizations, institutions, and countries with which the data is shared

After conducting the activity personal data are pseudonymized as soon as possible. Access to the 'key' file is restricted to researchers analyzing the data. Analyses are conducted only on the basis of pseudonymized data. The pseudonymized data are only accessible to the researchers analyzing these data within the framework of the 4Growth Horizon Europe project, composed of the following organisations:

Organization/institutions which are part of the 4Growth Consortium	Countries within the EU (incl. NL)
EVENFLOW	BE
GEOPONIKO PANEPISTIMION ATHINON	EL
FOODSCALE HUB GREECE ASSOCIATION FOR ENTREPREUNERSHIP AND INNOVATION ASTIKI MI KERDOSKOPIKI ETAIREIA	EL
LE EUROPE LIMITED	IE
DAHEIM CORNELIA	DE
SIMBIOTICA SL	ES
EIGEN VERMOGEN VAN HET INSTITUUT VOOR LANDBOUW- EN VISSERIJONDERZOEK	BE
INSTITUTO NAVARRO DE TECNOLOGIAS E INFRAESTRUCTURAS AGROALIMENTARIAS SA	ES
CENTRE TECHNIQUE INTERPROFESSIONNEL DES FRUITS ET LEGUMES	FR
TEKNOLOGIAN TUTKIMUSKESKUS VTT OY	FI
AgriFood Lithuania DIH	LT
ARISTOTELIO PANEPISTIMIO THESSALONIKIS	EL

Contact details project leader, researcher(s) and data privacy officer

Name	Function	WUR mail address	Phone number
George Beers	Project Leader	George.beers@wur.nl	070 3358 330
Lan van Wassenaar	Senior Researcher	Lan.vanwassenaar@wur.nl	070 3358 330
Joep Tummers	Researcher	Joep.tummers@wur.nl	070 3358 330
Sinne van de Veer	Researcher	Sinne.vandeveer@wur.nl	070 3358 330

Research results

You may request a summary of the research findings by contacting the task leader, Lan van Wassenaar of the Wageningen Economic Research Lan.vanwassenaar@wur.nl At any time, you are free to withdraw consent by contacting the task leader.

Concerns or complaints

If you have any concerns about the project, your involvement in it or this consent form, please discuss this with the researcher undertaking the *survey* to find out how your concern will be addressed. If your concern is not addressed you can contact the Data Protection Officer -DPO- who supervises the application of and compliance with the GDPR via: *privacy@wur.nl*

Annex B: Survey

4Growth English Wave 1 Survey Flow

Standard: Consent Form English (4 Questions)
Block: General Information(4GROWTH) (12 Questions)

Branch: New Branch
If
If Type of stakeholder: Farmer/agricultural producers Is Selected
Or Type of stakeholder: Forester Is Selected
Or Type of stakeholder: Forest owner Is Selected
Or Type of stakeholder: Forest operator Is Selected
Or Type of stakeholder: Forest product processor Is Selected
Or Type of stakeholder: Farming association Is Selected
Or Type of stakeholder: Farming cooperative Is Selected
Or Type of stakeholder: Forestry association Is Selected
Or Type of stakeholder: Forest industry association Is Selected
Or Type of stakeholder: Trade association Is Selected
Or Type of stakeholder: NGO Is Selected
Or Type of stakeholder: Advisory group Is Selected

Block: Adoption of Digital Technologies and Technology Integration (4GROWTH) (12 Questions)
Block: Technology Performance (4GROWTH) (9 Questions)
Block: Associated costs and prerequisites (4GROWTH) (5 Questions)
Block: Data management and data sharing practices (4GROWTH) (10 Questions)
Block: Data storage and data flows (4GROWTH) (13 Questions)

Branch: New Branch
If
If Type of stakeholder: Data provider Is Selected
Or Type of stakeholder: Research institutes and research networks Is Selected
Or Type of stakeholder: National and European networks Is Selected

Block: Governance Model (4GROWTH) (5 Questions)
Block: Data management and data sharing practices (4GROWTH) (10 Questions)
Block: Data storage and data flows (4GROWTH) (13 Questions)

Branch: New Branch
If
If Type of stakeholder: Infrastructure provider Is Selected

Block: Adoption of Digital Technologies and Technology Integration (4GROWTH) (12 Questions)
Block: Technology Performance (4GROWTH) (9 Questions)
Block: Associated costs and prerequisites (4GROWTH) (5 Questions)

Branch: New Branch

If

If Type of stakeholder: Data association/organisation/coalition Is Selected
Or Type of stakeholder: Platform provider Is Selected
Or Type of stakeholder: Service/information provider Is Selected
Or Type of stakeholder: Digital technology provider Is Selected

Block: Governance Model (4GROWTH) (5 Questions)
Block: Adoption of Digital Technologies and Technology Integration (4GROWTH) (12 Questions)
Block: Technology Performance (4GROWTH) (9 Questions)
Block: Associated costs and prerequisites (4GROWTH) (5 Questions)
Block: Data management and data sharing practices (4GROWTH) (10 Questions)
Block: Data storage and data flows (4GROWTH) (13 Questions)

Block: Social benefits and impact (4GROWTH) (3 Questions)
Block: Economic benefits and impact (4GROWTH) (3 Questions)
Block: Environmental and sustainability impact (4GROWTH) (5 Questions)
Block: Future outlook (4GROWTH) (3 Questions)
Block: Additional comments (4GROWTH) (1 Question)

Page Break

Start of Block: Consent Form English

Start of Survey

Q1
4Growth- Consent Form
Enquête in het Nederlands Kysely suomeksi Enquête en français   Undersökning på svenska Encuesta en español Ankieta w języku polskim   Felmérés magyar nyelven Έρευνα στα ελληνικά Apklausa lietuvių kalba

The 4Growth Horizon Europe project, aiming to advance digital solutions in agriculture and forestry, involves the collection and processing of certain data from stakeholders in the agriculture and forestry sectors. To ensure compliance with the General Data Protection Regulation (GDPR), we seek your explicit and informed consent before proceeding.   By giving your consent, you declare that: You have provided the data voluntarily   The data you provide will only be used for the purpose for which you provided it   You have the right to inspect, delete, correct, or limit the processing of personal data, as well as the right to object and the right to data portability   Any use of the information beyond the scope or duration of this project will require the researchers to contact you for (renewed) consent.   There are no known risks in taking part in this study.   Read more about the consent form: Consent form English

---

*

Q2 If you agree, please confirm the following statements: I have read the information presented in this consent form. I have had the opportunity to ask any questions related to this research and received satisfactory answers to my questions.   I understand that relevant sections of the data collected during the research may be accessed by individuals from the 4Growth project.   With full knowledge of all the foregoing, I agree that my answers will be processed as part of the 4Growth project.   I understand that relevant sections of the data collected during the research may be looked at by individuals from the 4Growth project. I give permission for these individuals to have access to my responses.

○ Yes (1)

○ No (2)

Q3 I agree to be contacted again by the researchers for clarification or elaboration on my input in the discussion (Optional)

- ◯ Yes, e-mail: (1) ____
- ◯ No (2)

Q4 Name:

____

End of Block: Consent Form English

Start of Block: General Information(4GROWTH)

Q1.1 Organisation Name:

____

Q1.2 Sector (Agri/Forestry/Both):

▼ Forestry (1) ... Both (3)

Q1.3 Type of stakeholder:

▼ Farmer/agricultural producers (1) ... Advisory group (20)

Q1.4 Location (Country/Region)

▼ Albania (28) ... Vatican City (city-state) (32)

Display This Question:

If Sector (Agri/Forestry/Both): = Agriculture

Or Sector (Agri/Forestry/Both): = Both

Q1.5 Primary Area of Operation in Agriculture

- ☐ Crop cultivation - grains (1)
- ☐ Crop cultivation - vegetables (2)
- ☐ Crop cultivation - legumes (3)
- ☐ Crop cultivation - fruits (4)
- ☐ Plant propagation (5)
- ☐ Livestock farming - meat (6)
- ☐ Livestock farming - dairy (7)
- ☐ Livestock farming - other (8)
- ☐ Mixed farming (crops and animal) (9)
- ☐ Agricultural machinery and equipment services (10)
- ☐ Crop services (monitoring) (11)
- ☐ Farm management services (12)
- ☐ Post-harvest handling services (13)
- ☐ Other namely (14)

Display This Question:

If Primary Area of Operation in Agriculture = Other namely

Q1.6 Other namely ...

____

Display This Question:

If Sector (Agri/Forestry/Both): = Forestry

Or Sector (Agri/Forestry/Both): = Both

Q1.7 Primary Area of Operation in forestry

- ☐ Reforestation (1)
- ☐ Forest conservation - thinning, pruning, weed & pest control (2)
- ☐ Felling (3)
- ☐ Non-Timber Forest Products (NTFPs) (4)
- ☐ Transportation of logs (5)
- ☐ Forest Fire Management (6)
- ☐ Forestry inventory and mapping (7)
- ☐ Wildlife management (8)
- ☐ Other namely (9)

Display This Question:

If Primary Area of Operation in forestry = Other namely

Q1.8 Other namely ...

____

Q1.9 Organic farming operation

▼ Yes (1) ... Don't know (3)

Q1.10 Agriculture/Forestry organisation size

▼ Small-scale/Local (1) ... Large-scale/National-International (3)

Q1.11 Specific regional or subsector considerations to take into account

▼ Yes (1) ... Don't know (3)

Display This Question:

If Specific regional or subsector considerations to take into account = Yes

Q1.12 Considerations to be taken into account

____

End of Block: General Information(4GROWTH)

Start of Block: Adoption of Digital Technologies and Technology Integration (4GROWTH)

Q3.1 Has your organisation integrated digital technologies into its workflows?

▼ Yes (1) ... Don't know (3)

Skip To: End of Block If Has your organisation integrated digital technologies into its workflows? = No

Display This Question:

If Sector (Agri/Forestry/Both): = Agriculture

Or Sector (Agri/Forestry/Both): = Both

Q3.2 What type of digital technology has been used for agriculture?

- ☐ Precision farming (1)
- ☐ Farm Management Information Systems (2)
- ☐ Automated machinery and robotics (3)
- ☐ Smart irrigation systems (4)
- ☐ Monitoring and tracking of livestock/crops (5)
- ☐ Smart-agri apps (6)
- ☐ Other namely (7) ____

Display This Question:

If Sector (Agri/Forestry/Both): = Forestry

Or Sector (Agri/Forestry/Both): = Both digitaltech.forestry What type of digital technology has been used for forestry?

- ☐ Forest Fire Prediction and Monitoring systems (1)
- ☐ Automated machinery and robotics (2)
- ☐ Drones for Forest Monitoring (3)
- ☐ Forest Inventory Management Software (4)
- ☐ Other namely (5) ____

---

goals.to.adopt Were there specific goals or challenges that prompted the adoption of digital tools?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Were there specific goals or challenges that prompted the adoption of digital tools? = Yes

specify.challenges.tech.adopt If yes, please specify

____

---

lvl.digitalisation How would you rate the level of digitalization in your farming/forestry practices on a scale of 1 to 5 (1 being low, 5 being high)

▼ 5 (1) ... 1 (5)

---

prim.function.tech What are the primary functions of these technologies in the agriculture or forestry value chain?

- ☐ Data management (1)
- ☐ Harvesting and distribution (2)
- ☐ Crop health and disease detection (3)
- ☐ Planning and Management (4)
- ☐ Decision-making (5)
- ☐ Supply chain optimisation (6)
- ☐ Monitoring (7)
- ☐ Production phase (8)
- ☐ On-farm activities (9)

---

adopt.level.tech What is the adoption level of these technologies?

▼ Preliminary (1) ... Fully integrated (3)

---

challenges.tech.adopt Have you encountered challenges in the adoption of digital technologies?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Have you encountered challenges in the adoption of digital technologies? = Yes

specify.challenges If yes, please specify

____

---

further.integration Are there specific barriers hindering further integration?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Are there specific barriers hindering further integration? = Yes

specify.barriers If yes, please specify

____

End of Block: Adoption of Digital Technologies and Technology Integration (4GROWTH)

Start of Block: Technology Performance (4GROWTH)

digitech.userneeds To what extent do digital technologies meet evolving user needs within your organization?

▼ Not at all (1) ... Completely (3)

---

adv.tech What are the advantages of the used technologies?

- ☐ Enhanced safety and monitoring (1)
- ☐ Improved management (2)
- ☐ Smart irrigation and water conservation (3)
- ☐ Economic benefits (4)
- ☐ Early detection of issues (5)
- ☐ Traceability and transparency (6)
- ☐ Efficient resource allocation (7)
- ☐ Improved decision-making (8)
- ☐ Increased efficiency and productivity (9)
- ☐ Other namely (10) ____

---

limitations.tech Have you encountered any perceived limitations or challenges in utilising these technologies?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Have you encountered any perceived limitations or challenges in utilising these technologies? = Yes

specify.limitations If yes, please specify

____

network.connect Do you have network connectivity?

▼ Yes (1) ... Don't know (3)

Skip To: End of Block If Do you have network connectivity? = No

network.connectivity What network connectivity do you use?

- ☐ Low-power Wide-area network (1)
- ☐ Private networks (2)
- ☐ Fiber optic networks (3)
- ☐ IoT networks (4)
- ☐ Satellite internet (5)
- ☐ Cellular networks (6)
- ☐ Wireless internet (7)
- ☐ Wired internet (8)

reliability.network How reliable is the current network connectivity? (1 being not reliable, 5 being very reliable)

▼ 5 (1) ... 1 (5)

barriers.connectivity Are there any specific barriers to accessing connectivity?

____

devices.network What type of devices are commonly used to access the network?

- ☐ Agricultural machinery equipped with IoT (Internet of Things) sensors (1)
- ☐ GPS devices (2)
- ☐ Smartphones (3)
- ☐ Tablets (4)
- ☐ Laptop computers (5)
- ☐ Desktop computers (6)

End of Block: Technology Performance (4GROWTH)

Start of Block: Associated costs and prerequisites (4GROWTH)

ACP6.1 What are the most significant costs associated with the adoption of digital technologies in your organisation

- ☐ Initial investment (1)
- ☐ Connectivity infrastructure (2)
- ☐ Maintenance and upgrades (3)
- ☐ Energy (4)
- ☐ Integration with existing systems (5)
- ☐ Training and skill development (6)
- ☐ Data security and privacy measures (7)
- ☐ Other namely (8) ____

Acp6.2 What is the level of direct costs?

▼ High (1) ... Low (3)

Acp6.3 Unexpected or hidden costs?

▼ Yes (1) ... Don't know (3)

Acp6.4 Have you identified organisational prerequisites (skills, workforce, education) necessary for successful technology integration?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Have you identified organisational prerequisites (skills, workforce, education) necessary for successful technology integration? = Yes

Acp6.4.1 If yes, please specify identified organisational prerequisites

____

End of Block: Associated costs and prerequisites (4GROWTH)

Start of Block: Data management and data sharing practices (4GROWTH)

Dmdsp7.1 Is data collected from your farming/forestry activities?

▼ Yes (1) ... Don't know (3)

Dmdsp7.2 What type of data sharing practices related to digital technology does your organisation use?

▼ Open sharing (1) ... No sharing (3)

Dmdsp7.3 What type of data do you collect?

- ☐ Crop and yield data (1)
- ☐ Soil data (2)
- ☐ Weather and environmental data (3)
- ☐ Pest and disease data (4)
- ☐ Inventory and equipment data (5)
- ☐ Market and economic data (6)
- ☐ Remote sensing and geospatial data (7)
- ☐ Livestock data (8)
- ☐ Financial and operational data (9)

Dmdsp7.4 Do you pay for this data?

▼ Yes (1) ... Don't know (3)

Dmdsp7.5 What type of tools or platforms do you use to collect data?

- ☐ Field Data Collection Apps (1)
- ☐ Precision Agriculture Technology (2)
- ☐ IoT Devices (3)
- ☐ Remote sensing platforms (4)
- ☐ Farm Management Software (5)
- ☐ Forest Management Software (6)
- ☐ Forest Inventory Tools (7)
- ☐ Traceability systems (8)
- ☐ Research Databases (9)

Dmdsp7.6 Do you share this data?

▼ Yes (1) ... Don't know (3)

Dmdsp7.7 Do challenges exist in sharing and interoperability of agricultural and forestry data?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Do challenges exist in sharing and interoperability of agricultural and forestry data? = Yes

Dmdsp7.8 If yes, please name the challenges associated with sharing and interoperability

____

Dmdsp7.9 How do these practices contribute to or impede the overall effectiveness of technology adoption?

▼ Contribute (1) ... No impact (3)

Dmdsp7.10 Approximately what percentage of overall decisions made are based on data analytics in your organisation?

▼ 25% (1) ... 90% (4)

End of Block: Data management and data sharing practices (4GROWTH)

Start of Block: Data storage and data flows (4GROWTH)

DSDF8.2 Do you use cloud services/data centres?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Do you use cloud services/data centres? = Yes

DSDF8.3 If yes, please name which cloud services/data centres

____

DSDF8.4 Are there economic implications associated with data flows in these sectors?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Are there economic implications associated with data flows in these sectors? = Yes

DSDF8.4.1 If yes, please name the main implications

____

DSDF8.5 Do data flows enhance productivity and efficiency in agriculture and forestry?

▼ Yes (1) ... Don't know (3)

DSDF8.6 Do you use data analytics for decision-making?

▼ Yes (1) ... Don't know (3)

DSDF8.7 Where do you receive data from and how much?

____

DSDF8.8 What type of data do you receive or provide?

- ☐ Farm-level data (1)
- ☐ Earth Observation (EO) data (2)
- ☐ Environmental data (3)
- ☐ Socio-economic data (4)
- ☐ Supply chain data (5)
- ☐ Research and Development data (6)

DSDF8.9 Do you pay for this data?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Do you pay for this data? = Yes

DSDF8.9.1 If yes please specify (type/amount)

____

DSDF8.10 Where and how do you store this data?

- ☐ On-premises servers/local storage facilities (1)
- ☐ Cloud-based platforms (2)
- ☐ Data warehouses (3)
- ☐ Agricultural information management systems (4)
- ☐ Geographic Information Systems (GIS) (5)
- ☐ Hybrid storage solutions (on-premises and cloud) (6)
- ☐ Secure data centres (advanced security measures) (7)

DSDF8.11 What do you do with this data?

____

DSDF8.12 To who and where do you send derived information or data?

____

End of Block: Data storage and data flows (4GROWTH)

Start of Block: Governance Model (4GROWTH)

Q2.1 Are there regulatory considerations influencing the governance of digital technology adoption?

▼ Yes (1) ... Don't know (3)

Display This Question:

If Are there regulatory considerations influencing the governance of digital technology adoption? = Yes

Q2.2 If yes please specify

____

Q2.3 What type of governance model do you operate under?

▼ Traditional/Subsistence Agriculture or Forestry (1) ... Other namely (7)

Display This Question:

If What type of governance model do you operate under? = Other namely

Q2.4 Other namely ...

____

Q2.5 Have specific governance models either facilitated or hindered the adoption of digital technologies in your organization?

▼ Facilitated (1) ... No impact (3)

End of Block: Governance Model (4GROWTH)

Start of Block: Social benefits and impact (4GROWTH)

social.benefits Have you experienced social benefits through the use of digital technologies?

▼ Yes (1) ... Don't know (3)

job.creation How have digital technologies impacted job creation?

▼ Substantial impact (1) ... Negligible impact (3)

social.impact What is the overall social impact of adopting digital technologies?

▼ Negative (1) ... Positive (3)

End of Block: Social benefits and impact (4GROWTH)

Start of Block: Economic benefits and impact (4GROWTH)

digitech.costsavings Have digital technologies resulted in cost savings or increased efficiency?

▼ Yes (1) ... Don't know (3)

digitech.savinginputs Have you seen savings in inputs due to digital technologies?

▼ Yes (1) ... Don't know (3)

impact.digitech What is the overall economic impact of implementing digital technologies?

▼ Negative (1) ... Positive (3)

End of Block: Economic benefits and impact (4GROWTH)

Start of Block: Environmental and sustainability impact (4GROWTH) digitech sustainability Have digital technologies contributed to sustainability and environmental practices?

▼ Yes (1) ... Don't know (3)

---

digitech impacts footprint Have you observed positive impacts on resource conservation or environmental footprint?

▼ Yes (1) ... Don't know (3)

---

didgitech energy efficiency Have digital technologies contributed to energy efficiency?

▼ Yes (1) ... Don't know (3)

---

digitech biodiversity Have you observed any positive or negative effects on biodiversity in agricultural and forestry areas due to digital technology adoption?

▼ No impact (1) ... Positive (3)

---

digitech track sustainability Do you use digital technologies to track and ensure adherence to sustainable farming practices and forestry activities?

▼ Yes (1) ... Don't know (3)

End of Block: Environmental and sustainability impact (4GROWTH)

Start of Block: Future outlook (4GROWTH) plan upgrade digitech Are there plans to expand or upgrade your current digital infrastructure?

▼ Yes (1) ... Don't know (3)

---

facilitate expansion/upgrade What would help facilitate the expansion/upgrade of digital infrastructure in the future?

- ☐ Better connectivity/Infrastructure (1)
- ☐ More income/Access to funding (2)
- ☐ Standardisation efforts/Regulatory support (3)
- ☐ Better training and education (4)

---

type of developments What type of developments do you anticipate in the near future?

▼ Emergence of new technologies (1) ... No significant changes anticipated (3)

End of Block: Future outlook (4GROWTH)

Start of Block: Additional comments (4GROWTH)

add.info Please share any other input that could be relevant to the questionnaire

____

End of Block: Additional comments (4GROWTH)

Annex C: Analysis of answer retrieval effectiveness for various queries

code	query	Has information %	No information %
type	Type of stakeholder:	96	4
prim.forest	Primary Area of Operation in forestry	91	9
GM02	What type of governance model do you operate under?	89	11
DSDF8.8	What type of data do you receive or provide?	84	16
Dmdsp7.3	What type of data do you collect?	72	28
type.developments	What type of developments do you anticipate in the near future?	71	29
facilitate.expansion.upgrade	What would help facilitate the expansion/upgrade of digital infrastructure in the future?	65	35
digitech.userneeds	To what extent do digital technologies meet evolving user needs within your organization?	57	43
DSDF8.11	What do you do with this data?	57	43
Dmdsp7.9	How do these practices contribute to or impede the overall effectiveness of technology adoption?	54	46
Dmdsp7.2	What type of data sharing practices related to digital technology does your organisation use?	52	48
social.impact	What is the overall social impact of adopting digital technologies?	51	49
adopt.level.tech	What is the adoption level of these technologies?	47	53
adv.tech	What are the advantages of the used technologies? - Selected Choice	41	59
DSDF8.7	Where do you receive data from and how much?	41	59
GM04	Are there regulatory considerations influencing the governance of digital technology adoption?	38	62
prim.function.tech	What are the primary functions of these technologies in the agriculture or forestry value chain?	38	62
DSDF8.12	To who and where do you send derived information or data?	37	63
itegrated.digi.tech	Has your organisation integrated digital technologies into its workflows?	32	68
DSDF8.6	Do you use data analytics for decision-making?	30	70
Dmdsp7.6	Do you share this data?	28	72
DSDF8.10	Where and how do you store this data?	28	72
lvl.digitalisation	How would you rate the level of digitalization in your farming/forestry practices on a scale of 1 to 5 (1 being low, 5 being high)	24	76
job.creation	How have digital technologies impacted job creation?	20	80

digitaltech.forestry	What type of digital technology has been used for forestry? - Selected Choice	19	81
goals.to.adopt	Were there specific goals or challenges that prompted the adoption of digital tools?	19	81
digitech.costsavings	Have digital technologies resulted in cost savings or increased efficiency?	18	82
digitech.sustainability	Have digital technologies contributed to sustainability and environmental practices?	16	84
Dmdsp7.5	What type of tools or platforms do you use to collect data?	14	86
DSDF8.4	Are there economic implications associated with data flows in these sectors?	14	86
DSDF8.2	Do you use cloud services/data centres?	12	88
Dmdsp7.1	Is data collected from your farming/forestry activities?	11	89
plan.upgrade.digitech	Are there plans to expand or upgrade your current digital infrastructure?	8	92
limitations.tech	Have you encountered any perceived limitations or challenges in utilising these technologies?	7	93
digitech.savinginputs	Have you seen savings in inputs due to digital technologies?	7	93
digitech.track.sustainability	Do you use digital technologies to track and ensure adherence to sustainable farming practices and forestry activities?	7	93
digitech.impacts.footprint	Have you observed positive impacts on resource conservation or environmental footprint?	6	94
digitech.energy.efficiency	Have digital technologies contributed to energy efficiency?	6	94
DSDF8.5	Do data flows enhance productivity and efficiency in agriculture and forestry?	4	96
digitech.biodiversity	Have you observed any positive or negative effects on biodiversity in agricultural and forestry areas due to digital technology adoption?	4	96
social.benefits	Have you experienced social benefits through the use of digital technologies?	3	97
challenges.tech.adopt	Have you encountered challenges in the adoption of digital technologies?	2	98
further.integration	Are there specific barriers hindering further integration?	1	99
impact.digitech	What is the overall economic impact of implementing digital technologies?	1	99
Dmdsp7.4	Do you pay for this data?	0	100
Dmdsp7.7	Do challenges exist in sharing and interoperability of agricultural and forestry data?	0	100