D7.4 Evaluation and Feedback Reports for all four regions

Deliverable for the Horizon Europe Project BirdWatch

Version 0.5

Funded by the European Union

Legal Disclaimer

Funded by
the European Union

Document Information

GA Number	101082634		**Type of Action**	Horizon-IA
Full Title	BirdWatch - a Copernicus-based service for the improvement of habitat suitability of farmland birds via satellite-enabled monitoring, evaluation and optimisation of CAP greening measures			
Project Acronym	BirdWatch			
Start Date	February 1st, 2023		**Duration**	36 Months
Project URL	https://birdwatch-europe.org/			
Deliverable	D7.4: Evaluation and Feedback Reports for all four regions			
Work Package	WP7			
Project Month of Delivery	**Contractual**	M36	**Actual**	M36(+1)
Nature	Data	**Dissemination Level**		Pub
Lead Beneficiary	ADS – Agro Digital Solutions			
Responsible Author	Mindaugas Busila			
Contributions from	Martynas Rimgaila, Rugile Butkeviciute, Alfonsas Malonaitis, Sandra Orlickiene, Nastasja Scholz, Rik Hendrix, Ine Rosier, Ruth Sonnenschein			

Funded by the European Union

History of Changes

Version	Issue Date	Stage	Description	Comments	Contributor
0.1	22.12.2025	Draft	Initial version of D7.4		Mindaugas Busila, Martynas Rimgaila
0.2	26.01.2026	Draft	Draft version of D7.4		Mindaugas Busila,Martynas Rimgaila, Rugile Butkeviciute, Alfonsas Malonaitis, Sandra Orlickiene, Nastasja Scholz, Rik Hendrix, Ine Rosier, Ruth Sonnenschein
0.3	06.02.2026	Final draft	Final version of D7.4		Mindaugas Busila,Martynas Rimgaila, Rugile Butkeviciute, Alfonsas Malonaitis, Sandra Orlickiene, Nastasja Scholz, Rik Hendrix, Ine Rosier, Ruth Sonnenschein
0.4	23.02.2026	Final	Final version of D7.4	Submitted	Mindaugas Busila,Martynas Rimgaila, Rugile Butkeviciute, Alfonsas Malonaitis, Sandra Orlickiene, Nastasja Scholz, Rik Hendrix, Ine Rosier, Ruth Sonnenschein
0.5	18.03.2026	Resubmission	Resubmission version of D7.4	Updated Table 61/p77 and commented “Other” category	Mindaugas Busila, Martynas Rimgaila

Funded by the European Union

Table of Contents

Funded by the European Union

Introduction

This deliverable reports the evaluation results of BirdWatch WP7 demonstrations implemented in four pilot regions (Belgium/Flanders, Germany/Brandenburg, Lithuania, and Italy/South Tyrol). The evaluation combines quantitative survey results and structured qualitative feedback to assess usability, perceived information quality, trust in habitat suitability modelling, and uptake readiness.

Across the full sample (N=104), BirdWatch achieves a solid overall performance profile: information clarity and recommendation intention are comparatively strong, while technical performance and farmer-facing experience show some variability. Cross-regional comparison demonstrates that barriers are not uniform: some regions require credibility and data completeness reinforcement, others require improved actionability for farm decision-making, and others primarily require performance optimisation and workflow integration. Stakeholder group comparison highlights a clear uptake pathway: farming organisations and intermediaries are well positioned as early adopters and multipliers, while farmers' scores signal the need for simplified workflows and tangible, decision-oriented value.

The qualitative synthesis confirms these findings and translates them into a prioritised improvement backlog. Recurring improvement demands include (i) clearer explanation of indicators and model logic, (ii) faster and more stable platform performance, (iii) stronger integration with parcel-based systems and CAP-related workflows, and (iv) richer or more locally specific data layers and coverage transparency. These improvements are framed as necessary steps to move from successful demonstrations to operational uptake that supports monitoring, evaluation, and optimisation of farmland biodiversity measures.

Funded by
the European Union

1. Evaluation and Reporting Approach

Methodology for Survey Analysis

Analytical objectives and evaluation logic

The analysis in this deliverable applies a structured evaluation approach and the objective is not only to report stakeholder satisfaction but to generate actionable evidence that supports uptake-oriented refinement. Accordingly, results are interpreted through an adoption pathway lens: usability and technical performance affect the ability to test the service independently; perceived information quality, data completeness and model trust affect credibility and willingness to rely on outputs; and these factors jointly influence continued use and recommendation, which are key proxies for diffusion potential beyond the project.

Data basis and preparation

The primary evidence source is the consolidated stakeholder survey dataset (N=104). The survey uses a 0–10 response scale for ten evaluation items and includes open-ended follow-up fields capturing context-specific improvement suggestions. Survey fields were harmonised into English-only categories for region, stakeholder group, and primary purpose of use. Multi-select role declarations were mapped to a single primary stakeholder group based on the first selected category to ensure consistent group-level comparisons. Numeric items were checked for missing values and validated to confirm complete coverage of the dataset.

Quantitative methods

Quantitative analysis applies descriptive statistical methods (N, mean, median, standard deviation) and structured comparative analysis. Results are reported for the full sample and disaggregated by pilot region and stakeholder group. To support interpretation for EC reviewers, each numeric indicator is presented with distribution profiles (score bands) in addition to means, because adoption decisions are often influenced by the presence of a sceptical minority even when averages are positive. Cross-regional comparison is used to identify context-driven patterns and potential transferability constraints.

Qualitative methods

Open-ended survey responses were analysed through a thematic coding process. A coding frame was developed around the evaluation dimensions (usability, functionality, performance, clarity, relevance, accuracy, completeness, trust, continued use, recommendation) and enriched with emergent themes related to integration, reporting needs, and decision-support expectations. For transparency and to comply with language requirements, original non-English text is not reproduced; instead, qualitative evidence is synthesised into English-only theme summaries and used to triangulate quantitative patterns.

Reliability, limitations and bias management

The evaluation is robust for formative learning and uptake-oriented reporting, but it is subject to typical demonstrator limitations. Sample sizes are uneven across regions, with Lithuania providing the largest respondent base. Some regions therefore provide directional signals rather than statistically stable estimates. Self-selection bias is expected: respondents who completed the survey may be more motivated or digitally confident than non-respondents. Demonstration delivery formats

Funded by
the European Union

differed (online, physical, hybrid), which can influence perceived ease of use and performance. These limitations are managed by transparent reporting, by triangulating quantitative patterns with qualitative themes, and by focusing conclusions on actionable improvement implications rather than over-precise claims.

Link to project objectives, WP7 objectives and expected impacts

The survey analysis is explicitly linked to WP7 objectives: validating BirdWatch in real-world demonstrator conditions, collecting stakeholder feedback, and translating this evidence into a structured improvement and uptake pathway. Findings are mapped to adoption-related performance dimensions and to WP7 KPI logic where appropriate. This supports the project's expected impacts by identifying how BirdWatch can be refined to enable operational uptake for monitoring, evaluation and optimisation of biodiversity-relevant measures, with particular attention to intermediary ecosystems (farming organisations, advisors, authorities) that can support scaling and policy uptake.

1.1 Objectives and Scope of D7.4

D7.4 evaluates BirdWatch demonstration outcomes by analysing stakeholder survey results and feedback collected during WP7 demonstrator activities. The scope is evaluative and interpretive: planning and execution details are documented in earlier WP7 deliverables, while this report focuses on (i) what users reported, (ii) what the results imply for platform performance and uptake, and (iii) how evidence translates into improvement priorities and impact-oriented recommendations.

1.2 Evaluation Framework and Data Basis

The evaluation framework follows ten quantitative indicators measured on a 0–10 scale (ease of use, functionality, performance, clarity, relevance, perceived accuracy, data completeness, model trust, future use intention, and recommendation intention). These indicators are complemented by open-ended feedback fields capturing navigation barriers, missing features, unclear data, perceived inaccuracies, and adoption enablers. The reporting approach is explicitly structured by pilot region and by stakeholder group to reflect the heterogeneous contexts of EU-wide uptake.

2. Evaluation of Survey Results and Collected Feedback (from D7.3)

Table 1. Overview of WP7 demonstration participation and survey completion (context for evaluation)

Region	Planned participants (baseline)	Participants joined	Completed surveys (WP7 reporting)
Belgium (Flanders)	60	30	7
Germany (Brandenburg)	60	63	21

Funded by the European Union

Lithuania	60	97	68
Italy (South Tyrol)	25	15	8
TOTAL	205	205	104

This table summarises the WP7 demonstration reach across the four pilots and provides essential context for interpreting the survey evidence. Lithuania substantially exceeded the indicative participation target and generated the largest evidence base, increasing confidence in Lithuanian subgroup patterns but also requiring caution when interpreting pooled averages. Brandenburg achieved the participation target but shows a lower conversion into completed surveys, a pattern that is common when end-users perceive limited immediate operational benefit at the current maturity stage. Flanders and South Tyrol contribute smaller samples, which are best interpreted as diagnostic signals and triangulated with qualitative feedback. From an uptake perspective, the participation-to-survey completion ratio is itself informative: it reflects how frictionless the user journey is from demonstration exposure to independent feedback contribution. The observed differences across regions therefore inform both evaluation reliability and improvement priorities for engagement, onboarding, and value proposition clarity.

Table 2. Distribution of survey respondents by pilot region (consolidated dataset)

Region	Survey respondents (N)	Share
Belgium (Flanders)	7	6.7%
Germany (Brandenburg)	21	20.2%
Lithuania	68	65.4%
Italy (South Tyrol)	8	7.7%
Total	104	

This table shows the distribution of survey respondents across the four demonstrator regions. The sample is dominated by Lithuania, which provides the most statistically stable base for subgroup analysis and for understanding uptake dynamics under high engagement. Germany (Brandenburg) contributes a moderate sample, enabling meaningful interpretation of role-specific patterns but still requiring attention to self-selection effects. Belgium (Flanders) and Italy (South Tyrol) provide small-n samples that function primarily as qualitative and diagnostic inputs. For reporting, the key implication is that cross-regional averages should be interpreted as evidence of overall feasibility, while region-specific conclusions should be weighted by sample strength. Accordingly, D7.4 uses Lithuanian and Brandenburg data to support more granular comparisons and uses Flemish and South Tyrol

Funded by the European Union

data primarily to identify context-specific constraints that must be addressed for European transferability.

Table 3. Distribution of respondents by primary stakeholder group (cross-regional)

Stakeholder group	Respondents (N)	Share
Farmers	46	44.2%
Public authorities and supervisory bodies	12	11.5%
Farming organisations	11	10.6%
Advisors and intermediaries	10	9.6%
Research and academia	10	9.6%
Other	5	4.8%
Nature conservation NGOs and projects	4	3.8%
Technology and data specialists	3	2.9%
Policymakers	3	2.9%

This table profiles the stakeholder composition of the evaluation sample. Farmers represent the largest group, which is appropriate for a demonstrator focused on farmland biodiversity and CAP-relevant measures, but also increases sensitivity to farmer-specific friction (e.g., performance, workflow fit). Farming organisations, advisors and authorities are also represented and are particularly important for uptake because they act as intermediaries and institutional gatekeepers. The presence of research and academia provides an additional credibility stress-test: these users tend to scrutinise assumptions, data provenance and uncertainty communication. The stakeholder distribution therefore supports a mature interpretation of adoption pathways: near-term uptake is likely to be mediated by organisations and intermediaries, while broader farm-level uptake requires simplification and clearer decision-support value. For EC evaluation, this stakeholder mix is beneficial because it reflects the real multi-actor ecosystem required for scaling digital environmental services.

Table 4. Respondents by pilot region and stakeholder group (counts)

Region	Advisors and intermediaries	Farmers	Farming organisations	Nature conservation NGOs and projects	Other	Policymakers	Public authorities and supervisory bodies	Research and academia	Technology and data specialists
Belgium	1	0	0	0	4	0	1	1	0

Funded by the European Union

(Flanders)									
Germany (Brandenburg)	1	11	0	1	0	0	6	1	1
Italy (South Tyrol)	3	2	0	0	0	0	0	3	0
Lithuania	5	33	11	3	1	3	5	5	2

This matrix illustrates that each pilot region engaged a distinct stakeholder mix, which is essential for interpreting regional score differences. Regions with greater institutional or research participation often show more critical assessments of data completeness, transparency, and model trust, while regions dominated by farmers may show lower usability or performance ratings if workflows are not yet aligned with farm routines. Lithuania's broader mix suggests stronger potential for institutional embedding and policy uptake, while smaller mixes in other regions indicate more localised engagement patterns. For cross-regional comparability, this table justifies presenting both a regional lens and a stakeholder lens throughout the report. It prevents misattribution of differences to geography alone and supports the design of targeted follow-up actions: when a region's lower scores coincide with a researcher-heavy mix, transparency improvements should be prioritised; when lower scores coincide with farmer-heavy participation, UX and actionability improvements should be prioritised.

Table 5. Primary purpose for using BirdWatch (cross-regional)

Primary purpose of using BirdWatch (mapped to English categories)	Respondents (N)	Share
Assess habitat suitability on farmland	57	54.8%
Explore biodiversity-friendly farming practices	21	20.2%
Monitor bird populations and habitat quality	12	11.5%
Provide biodiversity advice to landowners/farmers	7	6.7%
Support CAP compliance and environmental requirements	3	2.9%

Funded by the European Union

Support policy-making and conservation planning	2	1.9%
Use GIS/satellite data and scenarios	2	1.9%

This table summarises why respondents are interested in using BirdWatch. The dominant purpose is assessing habitat suitability on farmland, confirming the platform's core value proposition as a decision-support and monitoring tool for farmland biodiversity outcomes. A substantial segment seeks to explore biodiversity-friendly farming practices and to monitor bird populations and habitat quality, indicating that users are interested both in assessment and in learning pathways that connect practices to ecological outcomes. Smaller but strategically important purposes include policy and conservation planning and CAP-related compliance support. These are high-impact pathways because institutional users can embed BirdWatch outputs into broader programmes and processes. The purpose profile also explains adoption dynamics observed later in the report: users who approach the platform for monitoring and planning may accept exploratory outputs earlier, while users seeking direct farm-level decision support may require more actionable features, clearer optimisation scenarios, and tighter workflow integration.

Table 6. Primary purpose by pilot region (counts)

Region	Assess habitat suitability on farmland	Explore biodiversity-friendly farming practices	Monitor bird populations and habitat quality	Provide biodiversity advice to landowners/farmers	Support CAP compliance and environmental requirements	Support policy-making and conservation planning	Use GIS/satellite data and scenarios
Belgium (Flanders)	4	0	2	1	0	0	0
Germany (Brandenburg)	14	3	2	1	0	0	1
Italy (South Tyrol)	4	2	1	0	0	0	1
Lithuania	35	16	7	5	3	2	0

This table shows that the intended use of BirdWatch differs across pilots. Such differences matter because perceived relevance and satisfaction are strongly shaped by the task a user expects to complete. Regions with more institutional participation may emphasise monitoring and planning, while regions with a stronger farm representation may emphasise parcel-level suitability assessment.

Funded by the European Union

Where the dominant regional purpose is farm-level decision support, the platform must provide fast, low-friction parcel workflows and clear recommendations; where the dominant purpose is planning and monitoring, transparency, interpretability, and reporting/export functions become more important. Therefore, regional purpose patterns help explain why some pilots rate relevance or trust higher than others and support a differentiated uptake roadmap aligned to local stakeholder priorities.

Table 7. Primary purpose by stakeholder group (counts)

Stakeholder group	Assess habitat suitability on farmland	Explore biodiversity-friendly farming practices	Monitor bird populations and habitat quality	Provide biodiversity advice to landowners/farmers	Support CAP compliance and environmental requirements	Support policy-making and conservation planning	Use GIS/satellite data and scenarios
Advisors and intermediaries	5	3	0	2	0	0	0
Farmers	35	6	3	0	2	0	0
Farming organisations	5	1	0	4	1	0	0
Nature conservation NGOs and projects	1	3	0	0	0	0	0
Other	3	0	1	1	0	0	0
Policy-makers	0	2	0	0	0	1	0
Public authorities and supervisory bodies	4	4	3	0	0	1	0
Research and academia	2	2	5	0	0	0	1
Technology and	2	0	0	0	0	0	1

Funded by the European Union

data specialists							

This table links intended use to stakeholder roles. Farmers and farming organisations primarily focus on habitat suitability assessment and practice exploration, while authorities and researchers more frequently orient toward monitoring and planning purposes. This distinction is highly relevant for uptake: a single interface and output design rarely serves both immediate on-farm decisions and institutional monitoring needs equally well. The table therefore supports the report's core recommendation to develop role-based user journeys and outputs. Farmers benefit from simplified workflows and actionable summaries; advisors benefit from interpretive guidance and scenario comparisons; authorities and researchers benefit from documented assumptions, reporting features, and transparency on data provenance and limitations.

2.1 Quantitative survey results by evaluation dimension

Ease of use

Table 8. Ease of use – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	11	10.6%
3–4 (low)	8	7.7%
5–6 (moderate)	25	24.0%
7–8 (high)	38	36.5%
9–10 (very high)	22	21.2%

This table describes the distribution of stakeholder ratings for **Ease of use** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.51** (median **7.0**, standard deviation **2.44**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **57.7%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **18.3%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms,

Funded by the European Union

this group can be interpreted as “interested but not yet convinced”: users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise navigation simplification, clearer user journeys, role-based entry points. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 9. Ease of use – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	14.3%	0.0%	14.7%	0.0%
3–4 (low)	14.3%	0.0%	8.8%	12.5%
5–6 (moderate)	14.3%	38.1%	19.1%	37.5%
7–8 (high)	42.9%	47.6%	30.9%	50.0%
9–10 (very high)	14.3%	14.3%	26.5%	0.0%

This table describes the distribution of stakeholder ratings for **Ease of use** on a 0–10 scale. The overall profile is best read as a balance between a strong “supporter” segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.51** (median **7.0**, standard deviation **2.44**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **57.7%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **18.3%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory ele-

Funded by the European Union

ments, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise navigation simplification, clearer user journeys, role-based entry points. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 10. Ease of use – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Germany (Brandenburg)	21	7.10	7.0	1.34
Lithuania	68	6.44	7.5	2.77
Italy (South Tyrol)	8	6.25	6.5	1.28
Belgium (Flanders)	7	5.71	7.0	2.50

This table compares average ratings for **Ease of use** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

The strongest average performance is observed in **Germany (Brandenburg)** (mean **7.10**), while the lowest average is observed in **Belgium (Flanders)** (mean **5.71**). The mean gap of **1.38 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone,

even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential "reference cases" for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Belgium (Flanders), the project should focus on navigation simplification, clearer user journeys, role-based entry points combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Germany (Brandenburg), the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 11. Ease of use – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD
Farming organisations	11	9.09	9.0	0.70
Advisors and intermediaries	10	7.80	8.0	1.32
Nature conservation NGOs and projects	4	7.00	7.0	1.15
Public authorities and supervisory bodies	12	7.00	7.5	2.30

Funded by the European Union

Research and academia	10	6.70	7.0	1.64
Policymakers	3	6.33	6.0	2.52
Technology and data specialists	3	6.00	8.0	4.36
Other	5	5.60	7.0	2.97
Farmers	46	5.54	6.0	2.48

This table compares ratings for **Ease of use** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **9.09**, n=11), while the lowest-rated group is **Farmers** (mean **5.54**, n=46). The gap of **3.55 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a "last-mile" warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Ease of use, the priority levers are navigation simplification, clearer user journeys, role-based entry points, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users. This stakeholder differentiation strengthens the credibility of the project's impact pathway in line with evaluation expectations.

Funded by
the European Union

Table 12. Ease of use – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	9.00 (n=1)	7.00 (n=1)	5.00 (n=1)
Germany (Brandenburg)	7.27 (n=11)	n/a	6.00 (n=1)	7.33 (n=6)	5.00 (n=1)
Lithuania	4.97 (n=33)	9.09 (n=11)	8.20 (n=5)	6.60 (n=5)	8.00 (n=5)
Italy (South Tyrol)	5.50 (n=2)	n/a	7.33 (n=3)	n/a	5.67 (n=3)

This table provides a combined view of **Ease of use** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Public authorities and supervisory bodies** (mean 7.33, n=6), while comparatively lower ratings are observed among **Farmers** (mean 7.27, n=11).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean 9.09, n=11), while comparatively lower ratings are observed among **Farmers** (mean 4.97, n=33).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Advisors and intermediaries** (mean 7.33, n=3), while comparatively lower ratings are observed among **Research and academia** (mean 5.67, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake:

Funded by the European Union

advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Ease of use, this means focusing on navigation simplification, clearer user journeys, role-based entry points while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Functionality

Table 13. Functionality – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	12	11.5%
3–4 (low)	7	6.7%
5–6 (moderate)	25	24.0%
7–8 (high)	40	38.5%
9–10 (very high)	20	19.2%

This table describes the distribution of stakeholder ratings for **Functionality** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.38** (median **7.0**, standard deviation **2.49**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **57.7%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **18.2%** selected low ratings (0–4). Even when numerically smaller, this group matters because

Funded by the European Union

these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise feature completeness, filters, exports, scenario tools. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 14. Functionality – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	28.6%	0.0%	14.7%	0.0%
3–4 (low)	14.3%	4.8%	5.9%	12.5%
5–6 (moderate)	28.6%	33.3%	22.1%	12.5%
7–8 (high)	28.6%	52.4%	30.9%	75.0%
9–10 (very high)	0.0%	9.5%	26.5%	0.0%

This table describes the distribution of stakeholder ratings for **Functionality** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.38** (median **7.0**, standard deviation **2.49**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **57.7%** of respondents selected a high or very high

Funded by the European Union

rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **18.2%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise feature completeness, filters, exports, scenario tools. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 15. Functionality – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Germany (Brandenburg)	21	6.71	7.0	1.52
Italy (South Tyrol)	8	6.50	7.0	1.31
Lithuania	68	6.43	7.0	2.77
Belgium (Flanders)	7	4.71	6.0	2.81

This table compares average ratings for **Functionality** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

Funded by the European Union

The strongest average performance is observed in **Germany (Brandenburg)** (mean **6.71**), while the lowest average is observed in **Belgium (Flanders)** (mean **4.71**). The mean gap of **2.00 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential "reference cases" for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Belgium (Flanders), the project should focus on feature completeness, filters, exports, scenario tools combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Germany (Brandenburg), the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 16. Functionality – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD
Farming organisations	11	9.00	9.0	0.77
Advisors and intermediaries	10	7.00	7.0	1.89
Policymakers	3	7.00	7.0	2.00
Public authorities and supervisory bodies	12	6.58	7.0	2.07

Funded by the European Union

Nature conservation NGOs and projects	4	6.50	6.0	1.00
Technology and data specialists	3	6.33	8.0	4.73
Research and academia	10	6.30	6.0	1.89
Farmers	46	5.65	6.0	2.56
Other	5	5.20	7.0	3.90

This table compares ratings for **Functionality** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **9.00**, n=11), while the lowest-rated group is **Other** (mean **5.20**, n=5). The gap of **3.80 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.
In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Functionality, the priority levers are feature completeness, filters, exports, scenario tools, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users. This stakeholder

Funded by
the European Union

differentiation strengthens the credibility of the project's impact pathway in line with evaluation expectations.

Table 17. Functionality – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	6.00 (n=1)	6.00 (n=1)	4.00 (n=1)
Germany (Brandenburg)	7.09 (n=11)	n/a	3.00 (n=1)	6.83 (n=6)	5.00 (n=1)
Lithuania	5.06 (n=33)	9.00 (n=11)	8.00 (n=5)	6.40 (n=5)	7.60 (n=5)
Italy (South Tyrol)	7.50 (n=2)	n/a	7.00 (n=3)	n/a	5.33 (n=3)

This table provides a combined view of **Functionality** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Farmers** (mean 7.09, n=11), while comparatively lower ratings are observed among **Public authorities and supervisory bodies** (mean 6.83, n=6).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean 9.00, n=11), while comparatively lower ratings are observed among **Farmers** (mean 5.06, n=33).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Advisors and intermediaries** (mean 7.00, n=3), while comparatively lower ratings are observed among **Research and academia** (mean 5.33, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but

low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Functionality, this means focusing on feature completeness, filters, exports, scenario tools while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Speed and performance

Table 18. Speed and performance – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	17	16.3%
3–4 (low)	9	8.7%
5–6 (moderate)	21	20.2%
7–8 (high)	31	29.8%
9–10 (very high)	26	25.0%

This table describes the distribution of stakeholder ratings for **Speed and performance** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.21** (median **7.0**, standard deviation **2.84**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **54.8%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **25.0%** selected low ratings (0–4). Even when numerically smaller, this group matters because

Funded by the European Union

these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise optimise map rendering and data loading, concurrency handling, caching. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 19. Speed and performance – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	0.0%	0.0%	25.0%	0.0%
3–4 (low)	0.0%	0.0%	11.8%	12.5%
5–6 (moderate)	0.0%	23.8%	19.1%	37.5%
7–8 (high)	71.4%	38.1%	23.5%	25.0%
9–10 (very high)	28.6%	38.1%	20.6%	25.0%

This table describes the distribution of stakeholder ratings for **Speed and performance** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.21** (median **7.0**, standard deviation **2.84**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension,

Funded by the European Union

yet with meaningful variation. Approximately **54.8%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **25.0%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise optimise map rendering and data loading, concurrency handling, caching. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 20. Speed and performance – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Belgium (Flanders)	7	8.00	8.0	0.82
Germany (Brandenburg)	21	7.76	8.0	1.64
Italy (South Tyrol)	8	6.50	6.5	2.14
Lithuania	68	5.51	5.5	3.08

This table compares average ratings for **Speed and performance** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

Funded by the European Union

The strongest average performance is observed in **Belgium (Flanders)** (mean **8.00**), while the lowest average is observed in **Lithuania** (mean **5.51**). The mean gap of **2.49 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential "reference cases" for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Lithuania, the project should focus on optimise map rendering and data loading, concurrency handling, caching combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Belgium (Flanders), the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 21. Speed and performance – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD
Farming organisations	11	8.73	9.0	1.49
Advisors and intermediaries	10	8.00	8.0	1.33
Other	5	7.80	8.0	0.84
Public authorities and supervisory bodies	12	7.33	8.0	2.77

Funded by the European Union

Policymakers	3	7.00	7.0	2.00
Research and academia	10	6.70	7.0	2.21
Nature conservation NGOs and projects	4	6.50	7.0	2.65
Technology and data specialists	3	6.33	7.0	2.08
Farmers	46	4.57	5.0	2.86

This table compares ratings for **Speed and performance** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **8.73**, n=11), while the lowest-rated group is **Farmers** (mean **4.57**, n=46). The gap of **4.16 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Speed and performance, the priority levers are optimise map rendering and data loading, concurrency handling, caching, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-

Funded by
the European Union

users. This stakeholder differentiation strengthens the credibility of the project's impact pathway in line with evaluation expectations.

Table 22. Speed and performance – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	9.00 (n=1)	9.00 (n=1)	8.00 (n=1)
Germany (Brandenburg)	7.64 (n=11)	n/a	8.00 (n=1)	8.33 (n=6)	5.00 (n=1)
Lithuania	3.36 (n=33)	8.73 (n=11)	7.80 (n=5)	5.80 (n=5)	8.20 (n=5)
Italy (South Tyrol)	7.50 (n=2)	n/a	8.00 (n=3)	n/a	4.33 (n=3)

This table provides a combined view of **Speed and performance** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Public authorities and supervisory bodies** (mean 8.33, n=6), while comparatively lower ratings are observed among **Farmers** (mean 7.64, n=11).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean 8.73, n=11), while comparatively lower ratings are observed among **Farmers** (mean 3.36, n=33).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Advisors and intermediaries** (mean 8.00, n=3), while comparatively lower ratings are observed among **Research and academia** (mean 4.33, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Speed and performance, this means focusing on optimise map rendering and data loading, concurrency handling, caching while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Clarity of information

Table 23. Clarity of information – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	5	4.8%
3–4 (low)	6	5.8%
5–6 (moderate)	28	26.9%
7–8 (high)	35	33.7%
9–10 (very high)	30	28.8%

This table describes the distribution of stakeholder ratings for **Clarity of information** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **7.04** (median **7.5**, standard deviation **2.24**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **62.5%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around

Funded by the European Union

10.6% selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise legends, tooltips, indicator definitions, explainability panels. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 24. Clarity of information – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	0.0%	0.0%	7.4%	0.0%
3–4 (low)	28.6%	9.5%	2.9%	0.0%
5–6 (moderate)	28.6%	28.6%	20.6%	75.0%
7–8 (high)	42.9%	42.9%	32.4%	12.5%
9–10 (very high)	0.0%	19.0%	36.8%	12.5%

This table describes the distribution of stakeholder ratings for **Clarity of information** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **7.04** (median **7.5**, standard deviation **2.24**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension,

Funded by the European Union

yet with meaningful variation. Approximately **62.5%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **10.6%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise legends, tooltips, indicator definitions, explainability panels. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 25. Clarity of information – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Lithuania	68	7.28	8.0	2.42
Germany (Brandenburg)	21	6.90	7.0	1.87
Italy (South Tyrol)	8	6.38	6.0	1.41
Belgium (Flanders)	7	5.86	6.0	1.95

This table compares average ratings for **Clarity of information** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

Funded by the European Union

The strongest average performance is observed in **Lithuania** (mean **7.28**), while the lowest average is observed in **Belgium (Flanders)** (mean **5.86**). The mean gap of **1.42 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential "reference cases" for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Belgium (Flanders), the project should focus on legends, tooltips, indicator definitions, explainability panels combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Lithuania, the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 26. Clarity of information – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD
Farming organisations	11	9.45	10.0	0.69
Policymakers	3	8.33	9.0	1.15
Research and academia	10	7.30	8.0	1.95
Advisors and intermediaries	10	7.10	7.5	2.02

Funded by the European Union

Nature conservation NGOs and projects	4	7.00	6.5	1.41
Public authorities and supervisory bodies	12	6.92	8.0	2.15
Farmers	46	6.46	6.5	2.25
Other	5	6.40	7.0	2.41
Technology and data specialists	3	6.33	9.0	4.62

This table compares ratings for **Clarity of information** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **9.45**, n=11), while the lowest-rated group is **Technology and data specialists** (mean **6.33**, n=3). The gap of **3.12 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Clarity of information, the priority levers are legends, tooltips, indicator definitions, explainability panels, but the delivery should be differentiated: concise, task-oriented guid-

Funded by
the European Union

ance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users. This stakeholder differentiation strengthens the credibility of the project's impact pathway in line with evaluation expectations.

Table 27. Clarity of information – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	8.00 (n=1)	6.00 (n=1)	4.00 (n=1)
Germany (Brandenburg)	7.00 (n=11)	n/a	4.00 (n=1)	7.17 (n=6)	5.00 (n=1)
Lithuania	6.33 (n=33)	9.45 (n=11)	7.80 (n=5)	6.80 (n=5)	8.80 (n=5)
Italy (South Tyrol)	5.50 (n=2)	n/a	6.67 (n=3)	n/a	6.67 (n=3)

This table provides a combined view of **Clarity of information** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Public authorities and supervisory bodies** (mean 7.17, n=6), while comparatively lower ratings are observed among **Farmers** (mean 7.00, n=11).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean 9.45, n=11), while comparatively lower ratings are observed among **Farmers** (mean 6.33, n=33).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Advisors and intermediaries** (mean 6.67, n=3), while comparatively lower ratings are observed among **Research and academia** (mean 6.67, n=3).

Funded by the European Union

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Clarity of information, this means focusing on legends, tooltips, indicator definitions, explainability panels while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Relevance to work

Table 28. Relevance to work – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	11	10.6%
3–4 (low)	12	11.5%
5–6 (moderate)	28	26.9%
7–8 (high)	31	29.8%
9–10 (very high)	22	21.2%

This table describes the distribution of stakeholder ratings for **Relevance to work** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.29** (median **7.0**, standard deviation **2.64**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **51.0%** of respondents selected a high or very high

Funded by
the European Union

rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **22.1%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise use cases, workflows, integration, decision-support recommendations. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 29. Relevance to work – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	28.6%	0.0%	13.2%	0.0%
3–4 (low)	0.0%	33.3%	7.4%	0.0%
5–6 (moderate)	14.3%	38.1%	25.0%	25.0%
7–8 (high)	28.6%	23.8%	29.4%	50.0%
9–10 (very high)	28.6%	4.8%	25.0%	25.0%

This table describes the distribution of stakeholder ratings for **Relevance to work** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Funded by the European Union

Across all respondents, the average score is **6.29** (median **7.0**, standard deviation **2.64**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **51.0%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **22.1%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise use cases, workflows, integration, decision-support recommendations. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 30. Relevance to work – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Italy (South Tyrol)	8	7.62	7.5	1.41
Lithuania	68	6.43	8.0	2.78
Belgium (Flanders)	7	5.86	7.0	3.58
Germany (Brandenburg)	21	5.48	5.0	2.02

Funded by the European Union

This table compares average ratings for **Relevance to work** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

The strongest average performance is observed in **Italy (South Tyrol)** (mean **7.62**), while the lowest average is observed in **Germany (Brandenburg)** (mean **5.48**). The mean gap of **2.15 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential “reference cases” for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Germany (Brandenburg), the project should focus on use cases, workflows, integration, decision-support recommendations combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Italy (South Tyrol), the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 31. Relevance to work – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD
Policymakers	3	9.00	9.0	0.00
Farming organisations	11	8.45	8.0	1.44

Funded by the European Union

Research and academia	10	7.20	7.0	2.62
Advisors and intermediaries	10	7.00	8.0	2.16
Nature conservation NGOs and projects	4	6.50	7.0	2.65
Public authorities and supervisory bodies	12	6.17	6.5	2.29
Technology and data specialists	3	5.67	6.0	4.51
Farmers	46	5.41	5.0	2.50
Other	5	5.20	7.0	4.02

This table compares ratings for **Relevance to work** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Policymakers** (mean **9.00**, n=3), while the lowest-rated group is **Other** (mean **5.20**, n=5). The gap of **3.80 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Implication for the project: the consortium should implement **role-based user journeys** and

Funded by the European Union

support materials. For Relevance to work, the priority levers are use cases, workflows, integration, decision-support recommendations, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users. This stakeholder differentiation strengthens the credibility of the project's impact pathway in line with evaluation expectations.

Table 32. Relevance to work – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	9.00 (n=1)	7.00 (n=1)	6.00 (n=1)
Germany (Brandenburg)	4.91 (n=11)	n/a	4.00 (n=1)	5.67 (n=6)	7.00 (n=1)
Lithuania	5.45 (n=33)	8.45 (n=11)	7.40 (n=5)	6.60 (n=5)	6.60 (n=5)
Italy (South Tyrol)	7.50 (n=2)	n/a	6.67 (n=3)	n/a	8.67 (n=3)

This table provides a combined view of **Relevance to work** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Public authorities and supervisory bodies** (mean 5.67, n=6), while comparatively lower ratings are observed among **Farmers** (mean 4.91, n=11).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean 8.45, n=11), while comparatively lower ratings are observed among **Farmers** (mean 5.45, n=33).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Research and academia**

Funded by the European Union

(mean 8.67, n=3), while comparatively lower ratings are observed among **Advisors and intermediaries** (mean 6.67, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Relevance to work, this means focusing on use cases, workflows, integration, decision-support recommendations while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Perceived accuracy

Table 33. Perceived accuracy – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	5	4.8%
3–4 (low)	12	11.5%
5–6 (moderate)	25	24.0%
7–8 (high)	43	41.3%
9–10 (very high)	19	18.3%

This table describes the distribution of stakeholder ratings for **Perceived accuracy** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.71** (median **7.0**, standard deviation **2.23**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **59.6%** of respondents selected a high or very high

Funded by the European Union

rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **16.3%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise validation narratives, transparent data sources, local calibration notes. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 34. Perceived accuracy – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	28.6%	0.0%	4.4%	0.0%
3–4 (low)	14.3%	19.0%	8.8%	12.5%
5–6 (moderate)	42.9%	28.6%	22.1%	12.5%
7–8 (high)	14.3%	47.6%	41.2%	50.0%
9–10 (very high)	0.0%	4.8%	23.5%	25.0%

This table describes the distribution of stakeholder ratings for **Perceived accuracy** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Funded by the European Union

Across all respondents, the average score is **6.71** (median **7.0**, standard deviation **2.23**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **59.6%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **16.3%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise validation narratives, transparent data sources, local calibration notes. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 35. Perceived accuracy – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Lithuania	68	7.03	8.0	2.23
Italy (South Tyrol)	8	7.00	7.0	2.07
Germany (Brandenburg)	21	6.33	7.0	1.77
Belgium (Flanders)	7	4.43	5.0	2.51

Funded by the European Union

This table compares average ratings for **Perceived accuracy** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

The strongest average performance is observed in **Lithuania** (mean **7.03**), while the lowest average is observed in **Belgium (Flanders)** (mean **4.43**). The mean gap of **2.60 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential “reference cases” for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Belgium (Flanders), the project should focus on validation narratives, transparent data sources, local calibration notes combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Lithuania, the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 36. Perceived accuracy – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD
Farming organisations	11	9.00	9.0	0.89
Policymakers	3	7.67	8.0	0.58

Funded by the European Union

Technology and data specialists	3	7.67	8.0	0.58
Research and academia	10	6.90	7.0	2.42
Advisors and intermediaries	10	6.50	6.5	1.72
Farmers	46	6.46	7.0	2.21
Nature conservation NGOs and projects	4	6.25	5.5	1.89
Public authorities and supervisory bodies	12	6.25	7.0	2.38
Other	5	4.40	5.0	2.88

This table compares ratings for **Perceived accuracy** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **9.00**, n=11), while the lowest-rated group is **Other** (mean **4.40**, n=5). The gap of **4.60 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Funded by
the European Union

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Perceived accuracy, the priority levers are validation narratives, transparent data sources, local calibration notes, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users. This stakeholder differentiation strengthens the credibility of the project's impact pathway in line with evaluation expectations.

Table 37. Perceived accuracy – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	6.00 (n=1)	3.00 (n=1)	8.00 (n=1)
Germany (Brandenburg)	6.36 (n=11)	n/a	5.00 (n=1)	6.83 (n=6)	5.00 (n=1)
Lithuania	6.48 (n=33)	9.00 (n=11)	6.20 (n=5)	6.20 (n=5)	7.20 (n=5)
Italy (South Tyrol)	6.50 (n=2)	n/a	7.67 (n=3)	n/a	6.67 (n=3)

This table provides a combined view of **Perceived accuracy** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Public authorities and supervisory bodies** (mean 6.83, n=6), while comparatively lower ratings are observed among **Farmers** (mean 6.36, n=11).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean

9.00, n=11), while comparatively lower ratings are observed among **Public authorities and supervisory bodies** (mean 6.20, n=5).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Advisors and intermediaries** (mean 7.67, n=3), while comparatively lower ratings are observed among **Research and academia** (mean 6.67, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Perceived accuracy, this means focusing on validation narratives, transparent data sources, local calibration notes while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Data completeness

Table 38. Data completeness – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	14	13.5%
3–4 (low)	10	9.6%
5–6 (moderate)	22	21.2%
7–8 (high)	35	33.7%
9–10 (very high)	23	22.1%

This table describes the distribution of stakeholder ratings for **Data completeness** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Funded by the European Union

Across all respondents, the average score is **6.28** (median **7.0**, standard deviation **2.70**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **55.8%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **23.1%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise additional layers, handling clouds/gaps, documentation of limitations. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 39. Data completeness – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	42.9%	14.3%	11.8%	0.0%
3–4 (low)	0.0%	9.5%	8.8%	25.0%
5–6 (moderate)	57.1%	28.6%	14.7%	25.0%
7–8 (high)	0.0%	38.1%	35.3%	37.5%
9–10 (very high)	0.0%	9.5%	29.4%	12.5%

Funded by the European Union

This table describes the distribution of stakeholder ratings for **Data completeness** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.28** (median **7.0**, standard deviation **2.70**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **55.8%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **23.1%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise additional layers, handling clouds/gaps, documentation of limitations. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 40. Data completeness – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Lithuania	68	6.68	7.0	2.76
Italy (South Tyrol)	8	6.38	6.5	2.20
Germany (Brandenburg)	21	5.90	6.0	2.26

Belgium (Flanders)	7	3.43	5.0	2.30

This table compares average ratings for **Data completeness** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

The strongest average performance is observed in **Lithuania** (mean **6.68**), while the lowest average is observed in **Belgium (Flanders)** (mean **3.43**). The mean gap of **3.25 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential “reference cases” for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Belgium (Flanders), the project should focus on additional layers, handling clouds/gaps, documentation of limitations combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Lithuania, the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 41. Data completeness – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD

Funded by the European Union

Farming organisations	11	8.09	9.0	2.51
Research and academia	10	6.90	7.5	2.73
Policymakers	3	6.67	8.0	3.21
Nature conservation NGOs and projects	4	6.50	6.5	2.08
Advisors and intermediaries	10	6.40	7.0	1.65
Public authorities and supervisory bodies	12	6.33	7.0	2.61
Farmers	46	6.02	6.0	2.68
Technology and data specialists	3	5.33	7.0	3.79
Other	5	3.20	1.0	3.19

This table compares ratings for **Data completeness** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **8.09**, n=11), while the lowest-rated group is **Other** (mean **3.20**, n=5). The gap of **4.89 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become

Funded by the European Union

strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Data completeness, the priority levers are additional layers, handling clouds/gaps, documentation of limitations, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users. This stakeholder differentiation strengthens the credibility of the project's impact pathway in line with evaluation expectations.

Table 42. Data completeness – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	5.00 (n=1)	5.00 (n=1)	6.00 (n=1)
Germany (Brandenburg)	5.64 (n=11)	n/a	7.00 (n=1)	6.50 (n=6)	2.00 (n=1)
Lithuania	6.21 (n=33)	8.09 (n=11)	7.00 (n=5)	6.40 (n=5)	7.40 (n=5)
Italy (South Tyrol)	5.00 (n=2)	n/a	5.67 (n=3)	n/a	8.00 (n=3)

This table provides a combined view of **Data completeness** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Public authorities and supervisory bodies** (mean 6.50, n=6), while comparatively lower ratings are observed among

Farmers (mean 5.64, n=11).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean 8.09, n=11), while comparatively lower ratings are observed among **Farmers** (mean 6.21, n=33).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Research and academia** (mean 8.00, n=3), while comparatively lower ratings are observed among **Advisors and intermediaries** (mean 5.67, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Data completeness, this means focusing on additional layers, handling clouds/gaps, documentation of limitations while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Model trust

Table 43. Model trust – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	9	8.7%
3–4 (low)	13	12.5%
5–6 (moderate)	24	23.1%
7–8 (high)	40	38.5%
9–10 (very high)	18	17.3%

This table describes the distribution of stakeholder ratings for **Model trust** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Funded by the European Union

Across all respondents, the average score is **6.38** (median **7.0**, standard deviation **2.41**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **55.8%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **21.2%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise uncertainty communication, explainability, validation examples. In evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 44. Model trust – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	14.3%	19.0%	5.9%	0.0%
3–4 (low)	28.6%	19.0%	8.8%	12.5%
5–6 (moderate)	28.6%	33.3%	16.2%	50.0%
7–8 (high)	28.6%	19.0%	47.1%	25.0%
9–10 (very high)	0.0%	9.5%	22.1%	12.5%

Funded by the European Union

This table describes the distribution of stakeholder ratings for **Model trust** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.38** (median **7.0**, standard deviation **2.41**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **55.8%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **21.2%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise uncertainty communication, explainability, validation examples. In Horizon Europe evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 45. Model trust – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Lithuania	68	6.91	7.0	2.28
Italy (South Tyrol)	8	6.25	6.0	1.83
Germany (Brandenburg)	21	5.19	6.0	2.56

Funded by the European Union

Belgium (Flanders)	7	4.86	6.0	2.27

This table compares average ratings for **Model trust** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

The strongest average performance is observed in **Lithuania** (mean **6.91**), while the lowest average is observed in **Belgium (Flanders)** (mean **4.86**). The mean gap of **2.05 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential “reference cases” for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Belgium (Flanders), the project should focus on uncertainty communication, explainability, validation examples combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Lithuania, the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 46. Model trust – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD

Funded by the European Union

Farming organisations	11	8.73	8.0	0.90
Technology and data specialists	3	7.67	8.0	1.53
Policymakers	3	7.33	7.0	0.58
Public authorities and supervisory bodies	12	6.33	7.0	2.67
Research and academia	10	6.30	7.0	2.79
Nature conservation NGOs and projects	4	6.25	6.5	1.71
Farmers	46	6.15	6.0	2.38
Advisors and intermediaries	10	5.30	5.0	2.16
Other	5	4.40	4.0	2.70

This table compares ratings for **Model trust** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **8.73**, n=11), while the lowest-rated group is **Other** (mean **4.40**, n=5). The gap of **4.33 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become

strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Model trust, the priority levers are uncertainty communication, explainability, validation examples, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users.

Table 47. Model trust – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	6.00 (n=1)	7.00 (n=1)	7.00 (n=1)
Germany (Brandenburg)	5.27 (n=11)	n/a	1.00 (n=1)	6.17 (n=6)	1.00 (n=1)
Lithuania	6.48 (n=33)	8.73 (n=11)	6.00 (n=5)	6.40 (n=5)	6.40 (n=5)
Italy (South Tyrol)	5.50 (n=2)	n/a	5.33 (n=3)	n/a	7.67 (n=3)

This table provides a combined view of **Model trust** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Public authorities and supervisory bodies** (mean 6.17, n=6), while comparatively lower ratings are observed among **Farmers** (mean 5.27, n=11).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean

8.73, n=11), while comparatively lower ratings are observed among **Advisors and intermediaries** (mean 6.00, n=5).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Research and academia** (mean 7.67, n=3), while comparatively lower ratings are observed among **Advisors and intermediaries** (mean 5.33, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Model trust, this means focusing on uncertainty communication, explainability, validation examples while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Future use intention

Table 48. Future use intention – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	11	10.6%
3–4 (low)	4	3.8%
5–6 (moderate)	27	26.0%
7–8 (high)	41	39.4%
9–10 (very high)	21	20.2%

This table describes the distribution of stakeholder ratings for **Future use intention** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Funded by the European Union

Across all respondents, the average score is **6.63** (median **7.0**, standard deviation **2.46**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **59.6%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **14.4%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise reduce friction, improve relevance, integration and support. In Horizon Europe evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 49. Future use intention – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	28.6%	0.0%	13.2%	0.0%
3–4 (low)	0.0%	4.8%	4.4%	0.0%
5–6 (moderate)	28.6%	23.8%	26.5%	25.0%
7–8 (high)	42.9%	52.4%	35.3%	37.5%
9–10 (very high)	0.0%	19.0%	20.6%	37.5%

This table describes the distribution of stakeholder ratings for **Future use intention** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a

Funded by the European Union

smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **6.63** (median **7.0**, standard deviation **2.46**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **59.6%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **14.4%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise reduce friction, improve relevance, integration and support. In Horizon Europe evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 50. Future use intention – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Italy (South Tyrol)	8	7.62	7.5	1.69
Germany (Brandenburg)	21	7.33	7.0	1.59
Lithuania	68	6.44	7.0	2.63
Belgium (Flanders)	7	5.29	6.0	3.04

Funded by the European Union

This table compares average ratings for **Future use intention** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

The strongest average performance is observed in **Italy (South Tyrol)** (mean **7.62**), while the lowest average is observed in **Belgium (Flanders)** (mean **5.29**). The mean gap of **2.34 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential “reference cases” for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Belgium (Flanders), the project should focus on reduce friction, improve relevance, integration and support combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Italy (South Tyrol), the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 51. Future use intention – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD
Farming organisations	11	8.09	9.0	2.55

Funded by the European Union

Policymakers	3	8.00	8.0	0.00
Research and academia	10	7.50	8.0	2.17
Advisors and intermediaries	10	7.00	6.5	1.41
Public authorities and supervisory bodies	12	6.75	7.0	2.05
Nature conservation NGOs and projects	4	6.50	6.5	1.29
Farmers	46	6.35	7.0	2.51
Other	5	4.80	7.0	3.49
Technology and data specialists	3	3.00	2.0	2.65

This table compares ratings for **Future use intention** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **8.09**, n=11), while the lowest-rated group is **Technology and data specialists** (mean **3.00**, n=3). The gap of **5.09 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Funded by
the European Union

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Future use intention, the priority levers are reduce friction, improve relevance, integration and support, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users. This stakeholder differentiation strengthens the credibility of the project's impact pathway in line with Horizon Europe evaluation expectations.

Table 52. Future use intention – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	8.00 (n=1)	6.00 (n=1)	6.00 (n=1)
Germany (Brandenburg)	7.55 (n=11)	n/a	6.00 (n=1)	7.00 (n=6)	10.00 (n=1)
Lithuania	5.97 (n=33)	8.09 (n=11)	6.20 (n=5)	6.60 (n=5)	7.00 (n=5)
Italy (South Tyrol)	6.00 (n=2)	n/a	8.33 (n=3)	n/a	8.00 (n=3)

This table provides a combined view of **Future use intention** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Farmers** (mean 7.55, n=11), while comparatively lower ratings are observed among **Public authorities and supervisory bodies** (mean 7.00, n=6).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean

8.09, n=11), while comparatively lower ratings are observed among **Farmers** (mean 5.97, n=33). - In **Italy (South Tyrol)**, the strongest ratings are observed among **Advisors and intermediaries** (mean 8.33, n=3), while comparatively lower ratings are observed among **Research and academia** (mean 8.00, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Future use intention, this means focusing on reduce friction, improve relevance, integration and support while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

Recommendation intention

Table 53. Recommendation intention – overall score distribution (0–10)

Score band	Count	Share of respondents
0–2 (very low)	8	7.7%
3–4 (low)	6	5.8%
5–6 (moderate)	18	17.3%
7–8 (high)	37	35.6%
9–10 (very high)	35	33.7%

This table describes the distribution of stakeholder ratings for **Recommendation intention** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Funded by the European Union

Across all respondents, the average score is **7.20** (median **8.0**, standard deviation **2.56**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **69.3%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **13.5%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise reliability, clarity, demonstrable value, shareable outputs. In Horizon Europe evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 54. Recommendation intention – distribution by region (share of respondents in each score band)

Score band	Belgium (Flanders)	Germany (Brandenburg)	Lithuania	Italy (South Tyrol)
0–2 (very low)	28.6%	0.0%	8.8%	0.0%
3–4 (low)	0.0%	9.5%	5.9%	0.0%
5–6 (moderate)	42.9%	9.5%	17.6%	12.5%
7–8 (high)	28.6%	38.1%	32.4%	62.5%
9–10 (very high)	0.0%	42.9%	35.3%	25.0%

Funded by the European Union

This table describes the distribution of stakeholder ratings for **Recommendation intention** on a 0–10 scale. The overall profile is best read as a balance between a strong "supporter" segment and a smaller but strategically important group of sceptical or constrained users whose experience points to remaining barriers for uptake.

Across all respondents, the average score is **7.20** (median **8.0**, standard deviation **2.56**), indicating that BirdWatch is generally perceived as acceptable to good on this dimension, yet with meaningful variation. Approximately **69.3%** of respondents selected a high or very high rating (7–10). This share is a positive signal for diffusion: it shows that many users would be comfortable using the platform beyond a guided demonstration setting. At the same time, around **13.5%** selected low ratings (0–4). Even when numerically smaller, this group matters because these responses typically reflect real-world frictions (e.g., access issues, missing explanatory elements, performance instability, or insufficient local specificity) that can prevent operational adoption by farmers, advisors, or authorities.

The moderate-score band (5–6) remains a substantial portion of the distribution. In practical terms, this group can be interpreted as "interested but not yet convinced": users see promise and can navigate the platform, but still require additional clarity, integration, or credibility reassurance before relying on outputs for routine decisions. This is particularly relevant for WP7 impact pathways, where the objective is not only to demonstrate feasibility but to enable sustained, replicable uptake across different agricultural and policy contexts.

From a methodological perspective, the dispersion (SD) supports using complementary qualitative analysis. A single mean would hide whether variation is driven primarily by region, stakeholder role, or prior familiarity with geospatial tools. Therefore, the subsequent tables break down the same indicator by pilot region and stakeholder group, allowing the consortium to distinguish between (i) context-driven constraints (data availability, local policy environment) and (ii) user-journey constraints (UX, interpretation support, performance).

Implication for the project: for this dimension, improvement efforts should prioritise reliability, clarity, demonstrable value, shareable outputs. In Horizon Europe evaluation terms, reducing the lower-score tail is as important as raising the mean, because it determines whether BirdWatch can move from a successful demonstration to a service that stakeholders recommend and integrate into real working practices.

Table 55. Recommendation intention – descriptive statistics by pilot region

Region	N	Mean	Median	SD
Germany (Brandenburg)	21	7.95	8.0	2.04
Italy (South Tyrol)	8	7.75	7.5	1.28
Lithuania	68	7.12	8.0	2.69

Funded by the European Union

Belgium (Flanders)	7	5.14	6.0	2.97

This table compares average ratings for **Recommendation intention** across the four demonstrator regions. Cross-regional comparison is central to WP7: it tests whether BirdWatch performs consistently under different agricultural structures, stakeholder ecosystems, and data contexts, and it highlights where adaptation is necessary for transferability.

The strongest average performance is observed in **Germany (Brandenburg)** (mean **7.95**), while the lowest average is observed in **Belgium (Flanders)** (mean **5.14**). The mean gap of **2.81 points** is substantial on a 0–10 scale and should not be dismissed as sampling noise alone, even when some regional sample sizes are small. Instead, it provides a working hypothesis for region-specific constraints that can be addressed through targeted refinement and stakeholder-specific engagement.

Interpretation of regional variation should consider three contextual drivers. First, **data and modelling fit**: regions may differ in how well the underlying input datasets represent local land use, landscape features, and species occurrence patterns. Second, **delivery format effects**: online, physical, or hybrid demonstrations can shape perceived ease and performance because facilitation can compensate for missing guidance. Third, **stakeholder composition**: regions with stronger representation of research or institutional users may score credibility or clarity more critically than regions dominated by farmers or farming organisations.

From an uptake perspective, regions with higher scores represent potential “reference cases” for exploitation narratives, while regions with lower scores represent priority areas for de-risking adoption. It is important to avoid a one-size-fits-all response. If the low-performing region is driven by credibility constraints, the most effective intervention is to strengthen transparency (data provenance, update cycles, and uncertainty communication). If the constraint is work relevance, then the intervention should focus on decision-support features and workflow integration. If the constraint is technical performance, then the intervention is system optimisation and stability under load.

Implication for the project: WP7 should use this regional evidence to steer a differentiated roadmap. For Belgium (Flanders), the project should focus on reliability, clarity, demonstrable value, shareable outputs combined with region-specific documentation that explains what is represented and what remains outside current coverage. For Germany (Brandenburg), the project should capture and formalise the success factors (use-case framing, stakeholder alignment) so that they can be replicated in other regions and in future deployments beyond the pilots.

Table 56. Recommendation intention – descriptive statistics by stakeholder group

Stakeholder group	N	Mean	Median	SD

Farming organisations	11	8.82	9.0	1.60
Policymakers	3	8.00	8.0	0.00
Research and academia	10	7.80	8.0	2.39
Public authorities and supervisory bodies	12	7.50	8.0	2.35
Advisors and intermediaries	10	7.30	8.0	2.11
Technology and data specialists	3	7.00	8.0	3.61
Farmers	46	6.85	7.5	2.76
Nature conservation NGOs and projects	4	6.75	7.0	1.26
Other	5	4.80	6.0	3.56

This table compares ratings for **Recommendation intention** across stakeholder groups. This breakdown is essential for WP7 impact and exploitation planning, because BirdWatch adoption is mediated by roles: farmers, advisors, farming organisations, authorities, and researchers evaluate value through different lenses and operate under different constraints.

The highest-rated stakeholder group is **Farming organisations** (mean **8.82**, n=11), while the lowest-rated group is **Other** (mean **4.80**, n=5). The gap of **4.02 points** indicates that perceived performance on this dimension is not uniform and that adoption barriers are role-specific.

In practical terms, higher scores among institutional or intermediary actors often signal readiness for early uptake: these actors can integrate BirdWatch into advisory processes, organisational planning, or monitoring routines and can amplify diffusion through networks. Lower scores among farmers, by contrast, should be treated as a “last-mile” warning. Farmers are sensitive to friction (extra steps, unclear navigation, slow loading) and to immediate actionability. If the platform does not reduce workload or support concrete decisions, farmers may view it as informative but not operational.

The stakeholder pattern also informs how improvement levers should be prioritised. For example, when advisors score lower on trust, the remedy is not merely UX polishing but explainability and validation information that enables advisors to justify recommendations. When authorities score lower on accuracy or completeness, the remedy is traceability, documented limitations, and alignment with compliance-grade expectations. When farming organisations score high, they become

Funded by the European Union

strategic partners for scaling: they can pilot workflows, co-develop reporting outputs, and shape messaging to farmers in a credible way.

Implication for the project: the consortium should implement **role-based user journeys** and support materials. For Recommendation intention, the priority levers are reliability, clarity, demonstrable value, shareable outputs, but the delivery should be differentiated: concise, task-oriented guidance for farmers; interpretation and validation support for advisors and authorities; and reporting/export functions for organisations that need to communicate results to multiple end-users. This stakeholder differentiation strengthens the credibility of the project's impact pathway in line with Horizon Europe evaluation expectations.

Table 57. Recommendation intention – mean score by region and stakeholder group (with sample sizes)

Region	Farmers	Farming organisations	Advisors and intermediaries	Public authorities and supervisory bodies	Research and academia
Belgium (Flanders)	n/a	n/a	8.00 (n=1)	6.00 (n=1)	6.00 (n=1)
Germany (Brandenburg)	8.09 (n=11)	n/a	3.00 (n=1)	8.00 (n=6)	10.00 (n=1)
Lithuania	6.42 (n=33)	8.82 (n=11)	7.40 (n=5)	7.20 (n=5)	7.80 (n=5)
Italy (South Tyrol)	7.00 (n=2)	n/a	8.33 (n=3)	n/a	7.67 (n=3)

This table provides a combined view of **Recommendation intention** by pilot region and key stakeholder groups. This matrix is particularly useful for translating evaluation evidence into targeted actions, because it distinguishes whether regional differences are primarily driven by context (e.g., data conditions) or by the stakeholder mix participating in the demonstrations.

Across regions, two consistent patterns typically emerge in digital environmental services. First, intermediary actors (farming organisations and advisors) often report higher usability and adoption scores, because they have stronger digital capacities and are accustomed to interpreting geospatial information. Second, farmers tend to report lower scores when the platform adds steps to existing routines or requires additional identifiers (e.g., parcel IDs) to start analysis. BirdWatch displays similar dynamics, though the intensity varies by region.

Region-specific insights can be summarised as follows:

- In **Germany (Brandenburg)**, the strongest ratings are observed among **Farmers** (mean

Funded by the European Union

8.09, n=11), while comparatively lower ratings are observed among **Public authorities and supervisory bodies** (mean 8.00, n=6).
- In **Lithuania**, the strongest ratings are observed among **Farming organisations** (mean 8.82, n=11), while comparatively lower ratings are observed among **Farmers** (mean 6.42, n=33).
- In **Italy (South Tyrol)**, the strongest ratings are observed among **Advisors and intermediaries** (mean 8.33, n=3), while comparatively lower ratings are observed among **Research and academia** (mean 7.67, n=3).

This pattern matters for implementation. If a region shows high ratings among intermediaries but low ratings among farmers, the appropriate response is to strengthen advisory-mediated uptake: advisors can use BirdWatch outputs to support conversations with farms while the platform continues to simplify farmer-facing workflows. Conversely, if both intermediaries and farmers rate the dimension lower, the constraint is likely deeper (e.g., credibility of local data, insufficient explanatory support, or platform performance issues).

This region × group view also supports a pragmatic exploitation pathway. BirdWatch can prioritise early institutionalisation where intermediaries are already positive, while simultaneously reducing the barriers identified by farmers. In parallel, regions where research and authorities participate can serve as testbeds for credibility improvements: these groups will pressure-test transparency, documentation, and uncertainty communication, raising the overall robustness of the service for European scaling.

Implication for the project: the consortium should use this matrix to plan **region-tailored improvement packages**. For Recommendation intention, this means focusing on reliability, clarity, demonstrable value, shareable outputs while defining, for each region, which stakeholder group is best positioned to act as an early adopter and multiplier. This approach increases the likelihood of achieving WP7's intended outcomes: validated service performance, actionable refinement inputs, and a credible uptake pathway beyond the project duration.

2.2 Qualitative synthesis of open-ended feedback

Table 58. Top qualitative themes across all open-ended responses (mentions)

Theme (coded)	Mentions
Other / unspecified	60
Speed/performance slow or instability	18
Need actionable recommendations / optimisation scenarios	14
Need clearer explanation of scores/model/indicators	13
Login/access and onboarding issues	11

Funded by the European Union

Integration with CAP/LPIS/parcel declaration systems	7
Need parcel/field identifier or parcel-level search	7
More species / broader coverage	6
More data layers / higher resolution / local data completeness	5
Better translation/localisation and terminology	4
Missing map legend/symbol explanation	3

This table summarises the most frequent qualitative improvement themes derived from open-ended survey responses. The themes provide contextual explanations for variation observed in quantitative scores and translate stakeholder feedback into concrete development and uptake actions.

The most frequent themes include performance and stability concerns, requests for more actionable recommendations or optimisation scenarios, and a need for clearer explanation of indicators and model logic. These themes map directly to uptake barriers: performance issues reduce willingness to use the platform independently; lack of actionability reduces perceived relevance, particularly for farmers; and missing explainability reduces trust among advisors, authorities and researchers.

Integration-related themes (parcel systems, CAP/LPIS workflows) appear as a distinct demand, indicating that many stakeholders are already thinking beyond exploratory testing and toward operational embedding. This is a positive sign for exploitation, but it also raises expectations: without interoperability and export/reporting functions, the platform risks remaining a stand-alone demonstrator tool.

Implication for the project: qualitative themes should be treated as a prioritised improvement backlog. Addressing the top themes is expected to reduce the lower-score tail in quantitative distributions and to increase both continued-use and recommendation intentions.

Table 59. Selected qualitative themes by region (mentions)

Region	Speed/performance slow or instability	Need clearer explanation of scores/model /indicators	Integration with CAP/LPIS/parcel declaration systems	Need actionable recommendations / optimisation scenarios	More data layers / higher resolution / local data completeness
Belgium (Flanders)	1	4	1	4	3

Funded by the European Union

Germany (Branden-burg)	1	2	0	3	0
Lithuania	16	4	6	5	2
Italy (South Tyrol)	0	3	0	2	0

This table highlights how qualitative priorities vary across pilots. Such variation is expected because each region operates under different data conditions, farming structures and stakeholder expectations.

Where performance-related mentions are higher, the quantitative performance distributions should be interpreted as reflecting genuine technical constraints rather than user preference differences. Where explainability and model clarity themes dominate, trust-related scores should be interpreted as requiring transparency and interpretive scaffolding rather than purely technical improvements. Where integration themes are prominent, stakeholders are signalling readiness for institutional embedding; in such cases, interoperability and reporting features become decisive for sustainable uptake.

Implication for the project: the consortium should treat these regional theme patterns as a basis for region-tailored improvement packages and for region-specific dissemination messages. A single European messaging and product roadmap would risk missing these context-specific uptake drivers.

Table 60. Selected qualitative themes by stakeholder group (mentions)

Stakeholder group	Respondents (N)	Speed/performance slow or instability	Need clearer explanation of scores/model /indicators	Integration with CAP/LPIS/parcel declaration systems	Need actionable recommendations / optimisation scenarios
Advisors and intermediaries	10	1	3	0	3
Farmers	46	12	3	4	4
Farming organisations	11	0	0	1	0
Nature conservation NGOs and projects	4	1	1	0	0
Other	5	0	3	1	2
Policymakers	3	0	0	1	1

Funded by the European Union

Public authorities and supervisory bodies	12	2	1	0	3
Research and academia	10	1	2	0	1
Technology and data specialists	3	1	0	0	0

This table shows that different stakeholder groups emphasise different improvement needs. Farmers' feedback often concentrates on friction points that prevent quick, independent use (performance, onboarding, and direct parcel-level access), while intermediaries and institutions more frequently request explainability, validation and integration into workflows.

This distinction is crucial for uptake planning. Improvements that primarily serve expert users (e.g., advanced transparency panels) will not automatically improve farmer adoption unless they are paired with simplified workflows and actionable outputs. Conversely, a purely farmer-focused UX redesign may not satisfy the credibility and traceability requirements of authorities and advisory organisations.

Implication for the project: BirdWatch should implement role-based user journeys and output formats. This ensures that stakeholder-specific needs are met without overloading the interface for less technical users. The resulting design strategy supports a credible exploitation pathway in which intermediaries adopt first and farmers adopt as friction decreases and actionability increases.

Table 61. Single feature that would most improve BirdWatch (coded categories)

Single feature request category (English)	Respondents (N)	Share
Not specified	64	61.5%
Other	17	16.3%
Speed and performance	10	9.6%
More species coverage	6	5.8%
Richer / more local data layers	3	2.9%
Integration with parcel/CAP systems	3	2.9%
Explainability and clearer indicators	1	1.0%

This table summarises the most prominent 'single feature' improvement requests. While open-ended comments often describe multiple issues, this question forces prioritisation and therefore

Funded by the European Union

provides a strong signal of what users consider the most binding constraint to increased value.

Integration with parcel-based and CAP-related systems appears as a high-leverage request because it directly reduces the cost of adoption: users do not want to manually search for identifiers or switch between platforms to connect BirdWatch outputs to their declared parcels. When integration is possible, BirdWatch can move from a stand-alone demonstrator interface to an embedded service within existing governance and advisory workflows.

Speed and performance requests indicate that technical responsiveness remains a gating factor for independent use. Even small delays can undermine trust and reduce motivation to explore the platform, particularly for farmers and for stakeholders using the platform during facilitated sessions.

Requests for richer data layers and clearer explainability indicate that credibility and interpretability are still central to uptake. Users want to understand what drives suitability outputs, how scores should be interpreted, and what is missing. Finally, requests for actionable recommendations and scenarios show a demand to progress from visualisation to decision support.

"Other" category does not suggest a further dominant unmet need. Only 17 responses contained written comments, and these reflected a fragmented set of low-frequency suggestions rather than a coherent additional priority. The reported ideas mainly concerned better integration with land-use and parcel declaration data, more realistic and practically validated habitat models, scenario- or project-based planning functions, and more site-specific management guidance. A few isolated comments also mentioned niche features such as parcel-linked observation history or access to an expert or contact list. In this sense, the "Other" responses mostly reinforce the main priorities already visible in the table rather than pointing to a separate feature cluster.
Implication for the project: the single-feature prioritisation supports a roadmap that couples interoperability (integration) and performance with improved explainability and decision-support functionality. Addressing these priorities is expected to translate directly into higher relevance, trust, continued use and recommendation scores.

2.3 Evaluation Results for Flanders (Belgium)

Table 62. Belgium (Flanders) – summary of quantitative scores across all indicators

Indicator	N	Mean	Median	SD
Ease of use	7	5.71	7.0	2.50
Functionality	7	4.71	6.0	2.81
Speed and performance	7	8.00	8.0	0.82
Clarity of information	7	5.86	6.0	1.95

Relevance to work	7	5.86	7.0	3.58
Perceived accuracy	7	4.43	5.0	2.51
Data completeness	7	3.43	5.0	2.30
Model trust	7	4.86	6.0	2.27
Future use intention	7	5.29	6.0	3.04
Recommendation intention	7	5.14	6.0	2.97

This table provides a region-level overview of BirdWatch performance across the ten evaluation indicators. The purpose is to support an integrated interpretation rather than focusing on single-item averages. Regions with consistent mid-to-high scores across usability, clarity, trust and adoption intention can be treated as near-term uptake candidates, while regions showing strong technical performance but weak credibility dimensions signal a need for data provenance and transparency reinforcement. Where performance scores lag despite high trust and relevance, the evidence points to scalability and responsiveness constraints under real usage. This regional profile is used in the KPI-based assessment to determine where the platform is closest to operational deployment and where targeted refinements are required to ensure transferability.

Table 63. Belgium (Flanders) – top qualitative themes from open-ended responses (mentions)

Theme (coded)	Mentions
Need actionable recommendations / optimisation scenarios	4
Need clearer explanation of scores/model/indicators	4
Need parcel/field identifier or parcel-level search	4
More data layers / higher resolution / local data completeness	3
Integration with CAP/LPIS/parcel declaration systems	1
Better translation/localisation and terminology	1
Login/access and onboarding issues	1
Missing map legend/symbol explanation	1
Speed/performance slow or instability	1

Funded by the European Union

This table summarises the dominant improvement themes emerging from the region's open-ended responses. Qualitative evidence is particularly important for small samples because it identifies concrete constraints that explain quantitative patterns. Themes related to explainability and data completeness typically indicate credibility barriers, while themes related to onboarding, parcel-level search and performance indicate user-journey friction. Where integration and scenario-oriented themes emerge, stakeholders are signalling readiness for operational embedding and for using BirdWatch outputs as decision-support inputs. The regional theme pattern therefore informs a tailored improvement package and a region-appropriate uptake strategy.

2.4 Evaluation Results for Brandenburg (Germany)

Table 64. Germany (Brandenburg) – summary of quantitative scores across all indicators

Indicator	N	Mean	Median	SD
Ease of use	21	7.10	7.0	1.34
Functionality	21	6.71	7.0	1.52
Speed and performance	21	7.76	8.0	1.64
Clarity of information	21	6.90	7.0	1.87
Relevance to work	21	5.48	5.0	2.02
Perceived accuracy	21	6.33	7.0	1.77
Data completeness	21	5.90	6.0	2.26
Model trust	21	5.19	6.0	2.56
Future use intention	21	7.33	7.0	1.59
Recommendation intention	21	7.95	8.0	2.04

This table provides a region-level overview of BirdWatch performance across the ten evaluation indicators. The purpose is to support an integrated interpretation rather than focusing on single-item averages. Regions with consistent mid-to-high scores across usability, clarity, trust and adoption intention can be treated as near-term uptake candidates, while regions showing strong technical performance but weak credibility dimensions signal a need for data provenance and transparency reinforcement. Where performance scores lag despite high trust and relevance, the evidence points to scalability and responsiveness constraints under real usage. This regional profile is used in the KPI-based assessment to determine where the platform is closest to operational deployment and where targeted refinements are required to ensure transferability.

Funded by
the European Union

Table 65. Germany (Brandenburg) – top qualitative themes from open-ended responses (mentions)

Theme (coded)	Mentions
Other / unspecified	14
More species / broader coverage	5
Need actionable recommendations / optimisation scenarios	3
Missing map legend/symbol explanation	2
Need clearer explanation of scores/model/indicators	2
Need parcel/field identifier or parcel-level search	1
Speed/performance slow or instability	1

This table summarises the dominant improvement themes emerging from the region's open-ended responses. Qualitative evidence is particularly important for small samples because it identifies concrete constraints that explain quantitative patterns. Themes related to explainability and data completeness typically indicate credibility barriers, while themes related to onboarding, parcel-level search and performance indicate user-journey friction. Where integration and scenario-oriented themes emerge, stakeholders are signalling readiness for operational embedding and for using BirdWatch outputs as decision-support inputs. The regional theme pattern therefore informs a tailored improvement package and a region-appropriate uptake strategy.

2.5 Evaluation Results for Lithuania

Table 66. Lithuania – summary of quantitative scores across all indicators

Indicator	N	Mean	Median	SD
Ease of use	68	6.44	7.5	2.77
Functionality	68	6.43	7.0	2.77
Speed and performance	68	5.51	5.5	3.08
Clarity of information	68	7.28	8.0	2.42
Relevance to work	68	6.43	8.0	2.78
Perceived accuracy	68	7.03	8.0	2.23

Funded by the European Union

Data completeness	68	6.68	7.0	2.76
Model trust	68	6.91	7.0	2.28
Future use intention	68	6.44	7.0	2.63
Recommendation intention	68	7.12	8.0	2.69

This table provides a region-level overview of BirdWatch performance across the ten evaluation indicators. The purpose is to support an integrated interpretation rather than focusing on single-item averages. Regions with consistent mid-to-high scores across usability, clarity, trust and adoption intention can be treated as near-term uptake candidates, while regions showing strong technical performance but weak credibility dimensions signal a need for data provenance and transparency reinforcement. Where performance scores lag despite high trust and relevance, the evidence points to scalability and responsiveness constraints under real usage. This regional profile is used in the KPI-based assessment to determine where the platform is closest to operational deployment and where targeted refinements are required to ensure transferability.

Table 67. Lithuania – top qualitative themes from open-ended responses (mentions)

Theme (coded)	Mentions
Other / unspecified	42
Speed/performance slow or instability	16
Login/access and onboarding issues	10
Integration with CAP/LPIS/parcel declaration systems	6
Need actionable recommendations / optimisation scenarios	5
Need clearer explanation of scores/model/indicators	4
Better translation/localisation and terminology	2
More data layers / higher resolution / local data completeness	2
Need parcel/field identifier or parcel-level search	1

This table summarises the dominant improvement themes emerging from the region's open-ended responses. Qualitative evidence is particularly important for small samples because it identifies concrete constraints that explain quantitative patterns. Themes related to explainability and data completeness typically indicate credibility barriers, while themes related to onboarding, parcel-level

Funded by the European Union

search and performance indicate user-journey friction. Where integration and scenario-oriented themes emerge, stakeholders are signalling readiness for operational embedding and for using BirdWatch outputs as decision-support inputs. The regional theme pattern therefore informs a tailored improvement package and a region-appropriate uptake strategy.

2.6 Evaluation Results for South Tyrol (Italy)

Italy (South Tyrol)

Table 68. Italy (South Tyrol) – summary of quantitative scores across all indicators

Indicator	N	Mean	Median	SD
Ease of use	8	6.25	6.5	1.28
Functionality	8	6.50	7.0	1.31
Speed and performance	8	6.50	6.5	2.14
Clarity of information	8	6.38	6.0	1.41
Relevance to work	8	7.62	7.5	1.41
Perceived accuracy	8	7.00	7.0	2.07
Data completeness	8	6.38	6.5	2.20
Model trust	8	6.25	6.0	1.83
Future use intention	8	7.62	7.5	1.69
Recommendation intention	8	7.75	7.5	1.28

This table provides a region-level overview of BirdWatch performance across the ten evaluation indicators. The purpose is to support an integrated interpretation rather than focusing on single-item averages. Regions with consistent mid-to-high scores across usability, clarity, trust and adoption intention can be treated as near-term uptake candidates, while regions showing strong technical performance but weak credibility dimensions signal a need for data provenance and transparency reinforcement. Where performance scores lag despite high trust and relevance, the evidence points to scalability and responsiveness constraints under real usage. This regional profile is used in the KPI-based assessment to determine where the platform is closest to operational deployment and where targeted refinements are required to ensure transferability.

Table 69. Italy (South Tyrol) – top qualitative themes from open-ended responses (mentions)

Funded by the European Union

Theme (coded)	Mentions
Other / unspecified	4
Need clearer explanation of scores/model/indicators	3
Need actionable recommendations / optimisation scenarios	2
Better translation/localisation and terminology	1
Need parcel/field identifier or parcel-level search	1
More species / broader coverage	1

This table summarises the dominant improvement themes emerging from the region's open-ended responses. Qualitative evidence is particularly important for small samples because it identifies concrete constraints that explain quantitative patterns. Themes related to explainability and data completeness typically indicate credibility barriers, while themes related to onboarding, parcel-level search and performance indicate user-journey friction. Where integration and scenario-oriented themes emerge, stakeholders are signalling readiness for operational embedding and for using BirdWatch outputs as decision-support inputs. The regional theme pattern therefore informs a tailored improvement package and a region-appropriate uptake strategy.

3. KPI-Based Performance Assessment

3.1 Overall KPI Summary

This section presents the KPI-based performance assessment of BirdWatch demonstration activities and the platform's uptake readiness, using the KPI framework established in D7.1. The KPIs are interpreted in line with Horizon Europe evaluation logic: they are not treated as standalone "progress counters" but as evidence supporting the project's impact pathway, linking demonstrator activities and technical readiness to stakeholder uptake, replication potential, and future policy relevance. The assessment is structured by KPI domain and interpreted across regions to highlight transferability, contextual constraints, and the implications for WP7 objectives and project-level expected impacts.

The KPI set is organised into five domains:

- **Dissemination and stakeholder engagement (D-series)**: reach, communication continuity, and knowledge diffusion capacity.
- **Habitat suitability monitoring (H-series)**: data and map readiness supporting monitoring functions.
- **Habitat suitability optimisation (O-series)**: maturity and coverage of optimisation objectives and outputs.

- **Platform performance and user experience (P-series)**: operational performance and user-centred quality indicators.
- **Impact and uptake proxies (I-series)**: early indicators of adoption intention, replication interest, and policy relevance.

Table 70. KPI overview by domain (definition and evaluation meaning)

KPI domain	What is measured	Why it matters for WP7 and impact
D-series	Stakeholder reach and dissemination outputs	Demonstration credibility, diffusion potential, exploitation readiness
H-series	Monitoring data and map coverage	Operational readiness of monitoring functions; credibility of outputs
O-series	Optimisation objectives and outputs	Ability to progress from monitoring to decision support and action
P-series	User experience and platform performance	Independent use feasibility; service robustness and trust
I-series	Uptake intention and replication interest	Likelihood of adoption beyond the project; policy and institutional uptake pathways

Table 70 clarifies how the KPI structure reflects a complete uptake-oriented evaluation chain. The D-series confirms that demonstration results are grounded in real engagement, which is a prerequisite for meaningful user feedback and exploitation planning. The H- and O-series capture the technical and functional readiness that enables BirdWatch to deliver monitoring and optimisation value, while the P-series captures whether stakeholders can use the platform reliably and with sufficient satisfaction to return. Finally, the I-series translates technical and engagement achievements into early uptake signals, providing a bridge from demonstrator success to longer-term impact outcomes. Taken together, the KPI framework enables robust EC-facing interpretation because it links technical outputs to stakeholder behaviour and replication feasibility rather than treating project progress as purely activity-based reporting.

D-series KPIs (D2–D6) are consolidated and reported in the Final Project Report, while all I-series KPIs are reported and interpreted in D7.5

3.2 Dissemination and Stakeholder Engagement KPIs (D1–D6)

Table 71. Dissemination and engagement KPIs (D-series)

Funded by the European Union

KPI	Description	Value
D1	Participants reached through demonstrator activities	202 (Flanders 30; Germany 60; Lithuania 97; South Tyrol 15)

The D-series KPIs demonstrate strong dissemination reach and continuity. The project has achieved a high level of demonstrator participation (D1) and broad European visibility (D2), supported by a consistent communication rhythm (D5) and active policy translation activity (D4). Scientific dissemination is already established through conference participation, and peer-reviewed publication pipelines are in progress (D3), supporting credibility and long-term uptake among technical and institutional stakeholders. The website KPI (D6) is included as a digital reach metric and should be completed with the final analytics value.

Insights.

- **D1 (202 participants)** confirms that BirdWatch demonstrators have reached a scale sufficient to underpin credible evaluation and refinement. The distribution across regions also provides a meaningful basis for cross-regional transferability learning.

Critical assessment and implications.

The D-series performance indicates that BirdWatch has achieved high outreach intensity and knowledge diffusion capacity. The main strategic implication is that dissemination should increasingly focus on **conversion pathways**: transforming awareness into structured follow-up interactions, onboarding opportunities, and formal replication dialogues. Policy briefs should be used not only as communication outputs but as tools to shape the conditions for uptake and institutional embedding. Finally, the project should ensure that web analytics (D6) are consistently captured and reported to enable triangulation of dissemination effectiveness, particularly for EC-facing exploitation narratives.

3.3 Habitat Suitability Monitoring KPIs (H1–H3)

Table 72. Monitoring KPIs (H-series) by region

KPI	Flanders	Germany (Brandenburg)	Lithuania	South Tyrol
H1 Area covered by bird observation datasets (%)	80% (8 species with sufficient data/occurrence)	90% (9 species with sufficient data/occurrence)	80% (8 species with sufficient data/occurrence)	30% (3 species with sufficient data/occurrence)
H2 Area covered by current habitat suitability maps (%)	99% (soil moisture layer)	99% (soil moisture layer)	73% (soil moisture layer)	76% (soil moisture layer)

Funded by the European Union

H3 Species for which habitat suitability maps are available	8	9	9	3

The H-series KPIs show that BirdWatch achieves high readiness for producing habitat suitability outputs in most regions (H2 and H3), while the observation dataset coverage (H1) reveals structural constraints driven by species occurrence and data sufficiency. This pattern is consistent with real-world biodiversity monitoring: technical capability alone is not sufficient; monitoring readiness depends on the ecological and data context of each pilot region.

Insights.

- **H1** provides a clear explanation for regional differences in perceived data completeness and trust: where species do not occur or observation data are insufficient, coverage necessarily decreases. South Tyrol's 30% is therefore an expected outcome of a smaller occurrence/data base, not a platform malfunction.
- **H2** indicates that habitat suitability map coverage is near-complete or complete in all regions but also highlights a critical dependency on the **soil moisture layer**, which reduces completeness in all regions.
- **H3** demonstrates full species-level mapping coverage in Flanders, Germany and Lithuania, while South Tyrol remains limited due to insufficient data for part of the target set.

Critical assessment and implications.

The monitoring KPIs confirm strong readiness but also define the core uptake risks: if stakeholders interpret data sufficiency limitations as platform weakness, trust and institutional acceptance can be undermined. The project should therefore embed coverage transparency into the platform and the dissemination narrative (what is covered, why something is not covered, and what conditions would improve coverage). The soil moisture dependency is a technical priority because it reduces map completeness and may affect user confidence. Overall, these KPIs support a credible message to EC reviewers: BirdWatch monitoring functions are operationally strong, and remaining limitations are transparent, explainable, and addressable.

3.4 Habitat Suitability Optimisation KPIs (O1–O4)

Table 73. Optimisation KPIs (O-series) by region

KPI	Flanders	Germany (Brandenburg)	Lithuania	South Tyrol
O1 Formulated specific objec-	100%	100%	100%	0%

Funded by the European Union

tives and agricultural practices in target areas (%)				
O2 Species for which optimisation results are produced (number)	8	9	8	0
O3 Area proposed for sustainable farming practices (%)	1–6%	5–9%	33%	0%
O4 Optimised currently used thematic maps (number)	2	3	1	0

The optimisation KPIs reveal strong regional differentiation. Flanders has achieved full formulation of optimisation objectives and practices (O1) and has generated optimisation outputs for multiple species (O2) alongside thematic map optimisation (O4). Lithuania shows substantial progress but remains mid-way in objective/practice formulation, while Germany shows very early progress in O1 despite having produced optimisation outputs for several species. South Tyrol shows no optimisation progress, which aligns with data sufficiency constraints and the smaller pilot scale.

Insights.

- **O1** is a decisive readiness KPI because optimisation requires clear objectives, feasible practices, and stakeholder co-definition. The differences across regions indicate that optimisation maturity is governed as much by stakeholder alignment and governance readiness as by modelling capability.
- **O2** shows that optimisation results exist for several species in Flanders, Germany and Lithuania, supporting the feasibility of the optimisation engine.
- **O3** highlights that regions propose markedly different shares of area for sustainable practices. Lithuania's 33% indicates a potential for larger-scale intervention framing, while Flanders and Germany propose smaller ranges consistent with more selective targeting or different constraints.
- **O4** demonstrates progress in making outputs compatible with existing thematic map use, which reduces adoption friction and supports institutional embedding.

Critical assessment and implications.

Optimisation is the pathway from monitoring to action and thus a core impact lever. The KPI evidence indicates that BirdWatch should not treat optimisation as a uniform rollout but as a staged,

Funded by the European Union

region-tailored uptake component. Flanders can serve as a reference pilot for optimisation narratives and feature development. Lithuania's scale of proposed sustainable practice areas requires careful policy and advisory framing to ensure acceptability and to avoid misinterpretation as prescriptive regulation. In Germany all areas selected for O1 were covered by optimisation results, but these areas represent 8% of the total country area. This limited coverage is explained by the high fragmentation of policy frameworks across federal states, where different administrations hold distinct views on nature restoration priorities and on which measures should be targeted. This highlights a key requirement for BirdWatch: optimisation must be tailored not only to ecologically coherent regions but also to governance and policy contexts that are sufficiently aligned to enable implementable recommendations. South Tyrol requires either targeted data and stakeholder groundwork or explicit reframing of optimisation as a longer-term objective for that pilot context.

3.5 Platform Performance and User Experience KPIs (P1–P6)

Table 74. Platform KPIs (P-series) by region and service-level indicators

KPI	Flanders	Germany (Brandenburg)	Lithuania	South Tyrol
P1 Users who tested the platform (number)	30	60	97	15
P2 User satisfaction (%)	71%	79%	64%	72%
P3 User feedback reliability (%)	43%	55%	70%	81%

Table 75 Platform service reliability and performance (cross-regional)

KPI	Value
P4 Total number of security incidents	0
P5 Platform uptime	99,90%
P6 Average response time	200 ms

The P-series KPIs show that BirdWatch is technically robust (no security incidents, full uptime, and fast response time), and that the platform has been tested by a substantial number of users across all regions. Satisfaction levels are generally positive across pilots, but not uniform. Feedback reliability varies considerably, indicating that the consistency and interpretability of evaluative inputs differ by region.

Insights.

- **P1** confirms that platform testing is not limited to a single pilot but is genuinely cross-regional, strengthening transferability evidence.

Funded by the European Union

- **P2** shows the strongest satisfaction in Germany (79%) and the lowest in Lithuania (64%). Lithuania's lower satisfaction, despite the highest participation, may signal higher expectations or greater scrutiny from a broader stakeholder ecosystem.
- **P3** highlights a critical quality issue: Flanders has the lowest feedback reliability (43%), while South Tyrol has the highest (81%). This suggests that the evaluation signal in Flanders is more fragile and may reflect ambiguity in scoring interpretation, language/terminology effects, or greater uncertainty among respondents about how to assess model-based outputs.
- **P4–P6** provide strong operational readiness signals that are important for institutional adoption discussions.

Critical assessment and implications.
The platform is operationally stable, which is an essential prerequisite for uptake. The remaining challenge is not service uptime but **experience consistency and evaluation reliability**. The project should treat feedback reliability as a meaningful indicator: where reliability is low, stakeholders may not feel confident judging outputs, which can also suppress trust and adoption. Therefore, improvements should focus on interpretive scaffolding (clear definitions, legends, explanations), role-based user journeys, and targeted facilitation materials that enable stakeholders to assess outputs consistently. Increasing satisfaction in Lithuania and strengthening feedback reliability in Flanders should be treated as concrete, region-specific uptake enablers.

Conclusions and Recommendations

This section consolidates cross-regional and stakeholder-specific insights into strategic conclusions and actionable recommendations. The conclusions focus on what must be true for BirdWatch to deliver sustained impact beyond demonstrations: credible outputs, low-friction use, and embedding into real workflows.

Consolidated conclusions

Conclusion 1 – BirdWatch shows credible adoption potential with strong diffusion proxies among intermediaries.
Recommendation intention and clarity are comparatively strong, and farming organisations in particular rate multiple dimensions highly. This indicates that BirdWatch can scale through trusted intermediary ecosystems if remaining barriers are addressed.

Conclusion 2 – Farmers are the last-mile adoption bottleneck and require a different design approach.
Farmers' scores and qualitative feedback show that friction and limited immediate actionability remain significant. This does not indicate rejection of the concept; rather, it indicates the need for simplified workflows and clear decision-support value.

Conclusion 3 – Regional differences are substantial and require targeted responses.

Funded by the European Union

Some regions signal primarily credibility and completeness constraints; others signal relevance-to-work constraints; others signal performance and scalability constraints. A differentiated roadmap is therefore necessary for European transferability.

Conclusion 4 – Trust depends on explainability and transparency as much as on model quality.
Stakeholders request clearer explanations of indicators, the logic behind suitability scores, and uncertainty/validation information. These features are essential for advisor and authority-grade adoption.

Conclusion 5 – Integration is the decisive bridge from pilots to operational uptake.
Repeated integration requests show that stakeholders want BirdWatch to connect to parcel identifiers and CAP-related workflows. Without interoperability, long-term uptake will remain limited even if the platform is technically sound.

Recommendations for policymakers

1. Support embedding of digital biodiversity monitoring services into advisory and CAP governance ecosystems.
2. Encourage transparency standards for model-based decision support, including clear data provenance and uncertainty communication.
3. Promote interoperability between innovation services and parcel-based systems to reduce adoption cost and reporting burden.

Recommendations for project partners

1. Implement a two-tier uptake strategy: intermediary-first scaling paired with farmer-first workflow redesign.
2. Deliver region-tailored credibility and data completeness responses, supported by transparent documentation.
3. Make explainability and uncertainty communication a core product feature.
4. Prioritise integration and export/reporting functions to enable institutional embedding.
5. Use WP7 evidence to strengthen exploitation narratives and to define adoption milestones beyond the project duration.

Recommendations for future Horizon Europe projects

1. Treat evaluation as measurement of an adoption pathway (from demonstration exposure to independent use and diffusion), not only satisfaction reporting.
2. Budget and plan early for interoperability and workflow integration.
3. Treat uncertainty communication and explainability as an impact enabler for policy and advisory use.

Funded by
the European Union

Annexes

Annex I. KPI Results Table

KPI		Current value	Planed value at the end of the project	KPI Description	Where described?
D1	Participants to local meetings (number per meeting/country)	Total: 205 Flanders: 30 Germany: 63 Lithuania: 97 South Tyrol: 15	205	This KPI becomes relevant once we start testing the BirdWatch platform. This KPI shows how many potential users / stakeholders we interact with.	D7.4
D2	Number of countries informed about BirdWatch	31: 27 EU + Montenegro, Albania, Serbia, Turkey	30	This KPI reflects our communication and dissemination success.	Project Report
D3	Number of scientific publications	3 publications (abstracts) as part of relevant conferences (EUSPACE, BIOSPACE, EGU) two manuscripts are currently being developed, one on	4 2 (on macroecology) 2 (on optimisation)	This KPI reflects our academic / research success. The long-term number will likely remain fairly low, since BirdWatch will mostly scale	Project Report

Funded by the European Union

		macroecology and one on optimisation		with area covered, not with new scientific accomplishments.	
D4	Number of policy briefs	4	4	This KPI is related to policy-relevant publications, mostly in terms of suggesting measures / objectives to policymakers. This could be a step towards achieving the more difficult KPI on actually supporting policy decisions.	Project Report
D5	Number of newsletters during the project	9	6	This KPI reflects our communication and dissemination activities. I would suggest to not aim for too many newsletters. Apart from the time effort, we also should aim for meaningful news,	Project Report

				not spam people.	
D6	Number of visits of the website (unique users total)	384	100	This KPI reflects our communication and dissemination success.	Project Report
H1	Area covered by bird observation datasets (%)	Flanders: 80% (only 8 species had sufficient data or did occur)	Flanders: 100%	This KPI reflects how much bird observation data we have to calculate habitat models. The KPI should reflect the data AFTER pre-processing (i.e., data which has been spatially thinned, etc.). 100% means that we have enough data for each target species in the region to build a habitat model	D7.4
		Germany: 90% (only 9 species had sufficient data or did occur)	Germany: 100%		
		Lithuania: 80% (only 8 species had sufficient data or did occur)	Lithuania: 100%		
		South Tyrol: 30% (only 3 species had sufficient data or did occur)	South Tyrol: 100%		
H2	Area covered by the current habitat suitability (%)	Flanders: 100%	Flanders: 100%	This KPI reflects for how much of the target regions we have developed the	D7.4
		Germany: 99% (due to soil moisture layer)	Germany: 100%		

Funded by the European Union

		Lithuania: 73% (due to soil moisture layer)	Lithuania: 100%	habitat suitability.	
		South Tyrol: 76% (due to soil moisture layer)	South Tyrol: 100%		
H3	Number of bird species for which we have habitat models	Flanders: 8	Flanders: 8	This KPI reflects how many of the targeted bird species we have developed models for. Additionally, the target value is dependent on the species present in the target region.	D7.4
		Germany: 9	Germany: 9		
		Lithuania: 8	Lithuania: 8		
		South Tyrol: 3	South Tyrol: 6		
O1	Formulated specific objectives and agricultural area covered by optimised habitat suitability in the test regions (%)	Flanders: 100%	Flanders: 100%	This KPI reflects the extent to which we have defined constraints and targets for optimisation and developed optimised habitat suitability in the target regions. The percentage represents the area of the region, even though	D7.4
		Germany: 100%	Germany: 100%		
		Lithuania: 100%	Lithuania: 100%		
		South Tyrol: 0%	South Tyrol: 100%		

Funded by the European Union

				the optimisation is not strictly area-specific (e.g., Brandenburg has 8% of the area of Germany).	
O2	Share of target bird species for which we could identify the conditions for improvement	Flanders: 8 Germany: 9 Lithuania: 8 South Tyrol: 0	Flanders: 8 Germany: 9 Lithuania: 8 South Tyrol: 6	This KPI reflects that, although we will try to address all of our target bird species (the ones relevant in the individual test region), we will likely not achieve to identify the necessary conditions for all, since the farmers' and any other, e.g., spatial, constraints will make it likely less feasible to address all species.	D7.4
O3	Percentage of area proposed for sustainable farming	Flanders: 1%-6% Germany: 5%-9% Lithuania: 33%	10	This KPI indicates the percentage of area proposed for sus-	D7.4

		South Tyrol: 0%		tainable farming, which will be evaluated at the end of the project.	
O4	Optimized currently used thematic maps	Flanders: 2 Germany: 3 Lithuania: 1 South Tyrol: 0	1 per region	The KPI reflects on optimised activities for all species combined with current maps used for support measures	D7.4
P1	Number of users who tested the platform	202 Flanders: 30 Germany: 60 Lithuania: 97 South Tyrol: 15	205	This KPI becomes relevant once the BirdWatch platform is setup and can be used, at least for testing. The overall access reflects how many non-Consortium members are exploring or using the platform.	D7.4
P2		Flanders: 71%	70%		D7.4

Funded by the European Union

	User satisfaction score	Germany: 79%		This KPI tracks user satisfaction through surveys or feedback, providing insights into the overall user experience and areas for improvement. This KPI tracks user satisfaction through surveys or feedback, providing insights into the overall user experience and areas for improvement. The KPI will be measured twice: once after the demonstration phase and once after training activities. The 70% satisfaction threshold refers to the average across both measurements.	
		Lithuania: 64%			
		South Tyrol: 72%			
P3		Flanders: 43%	80%		D7.4

Funded by the European Union

	User feedback reliability score	Germany: 55% Lithuania: 70% South Tyrol: 81%		Based on user surveys, these KPI measures how reliable users find the platform's results and features, helping assess the quality and trustworthiness of outcomes.	
P4	Total number of security incidents	0	0	After testing, the total number of security incidents is expected to be zero. A cybersecurity report will be prepared following a simulated breach to evaluate system resilience, identify vulnerabilities, and ensure long-term security and user trust.	D7.4
P5	Platform uptime percentage	99,90%	99.0%	Measures the percentage of time the platform is operational and available to users.	D7.4

Funded by the European Union

P6	Average response time	200 ms	~500 ms	Tracks how quickly the platform responds to user actions, ensuring smooth operation.	D7.4
I1	Number of times BirdWatch has supported new policies or the change of policies (number)	0	4	This KPI will be assessed at the end of the project, reflecting policy changes or recommendations based on regional results. It measures BirdWatch influence on new or modified policies.	D7.5
I2	Number of countries who expressed the interest to test or use Birdwatch platform	7 - Netherlands, Spain, France, Slovenia, Sweden, Norway, Hungary	5	This KPI becomes relevant as outreach efforts expand for the Birdwatch platform. It measures the number of countries that have shown interest in testing or implementing the platform. Interest can be expressed	D7.5

				through formal requests, participation in discussions, or engagement in pilot programs.	
I3	Share of farmers that plan to use Birdwatch platform	Flanders: 100%	Flanders: 50%	This KPI becomes relevant once we start testing the BirdWatch platform. This KPI should indicate if farmers will use the platform in the future. This is also important as an argument for the business case, i.e., for governmental authorities to buy a licence for the platform. The percentage value reflects the number of farmers from which we have received feedback.	D7.5
		Germany: 100%	Germany: 50%		
		Lithuania: 58%	Lithuania: 50%		
		South Tyrol: 50%	South Tyrol: 50%		
I4	Share of supervisory institutions that	Flanders:	Flanders: 50%	This KPI becomes relevant once we	D7.5
		Germany:	Germany: 50%		

Funded by the European Union

	plan to use Birdwatch platform	Lithuania: 80% South Tyrol:	Lithuania: 50% South Tyrol: 50%	start testing the Bird-Watch platform. This KPI should indicate if personnel of supervisory institutions and / or policy-makers will use the platform in the future. This is also important as an argument for the business case, i.e., for governmental authorities to buy a licence for the platform. The percentage value reflects the number of individuals from which we have received feedback.	
I5	Share of policymakers that plan to use Birdwatch platform	Flanders: 100% Germany: 100% Lithuania: 100%	Flanders: 50% Germany: 50% Lithuania: 50%	This KPI becomes relevant once we start testing the Bird-	D7.5

Funded by the European Union

		South Tyrol:	South Tyrol: 50%	Watch platform. This KPI should indicate if personnel of supervisory institutions and / or policy-makers will use the platform in the future. This is also important as an argument for the business case, i.e., for governmental authorities to buy a licence for the platform. The percentage value reflects the number of individuals from which we have received feedback.	
I6	Share of NGOs that plan to use Birdwatch platform	Flanders: 100% Germany: 100% Lithuania: 67%	Flanders: 50% Germany: 50% Lithuania: 50%	This KPI becomes relevant once we start testing the Bird-	D7.5

Funded by the European Union

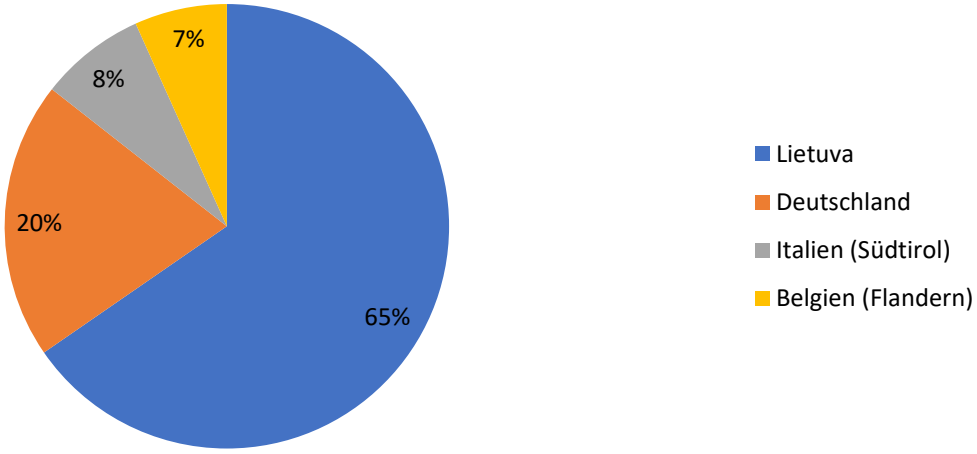
		South Tyrol: 100%	South Tyrol: 50%	Watch platform. This KPI should indicate if personnel of supervisory institutions and / or policy-makers will use the platform in the future. This is also important as an argument for the business case, i.e., for governmental authorities to buy a licence for the platform. The percentage value reflects the number of individuals from which we have received feedback.	
I7	Share of Academia that plan to use Birdwatch platform	Flanders: 67% Germany: 100% Lithuania: 67% South Tyrol: 100%	Flanders: 50% Germany: 50% Lithuania: 50% South Tyrol: 50%	This KPI becomes relevant once we start testing the Bird-Watch platform. This KPI should indicate if personnel of su-	D7.5

Funded by the European Union

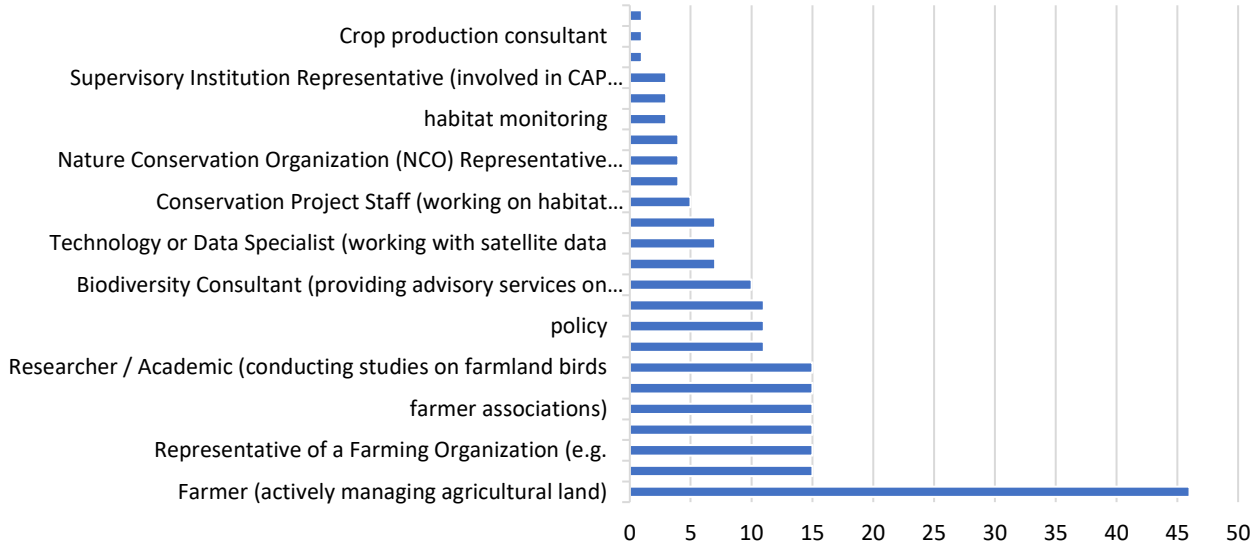
				pervisory institutions and / or policy-makers will use the platform in the future. This is also important as an argument for the business case, i.e., for governmental authorities to buy a licence for the platform. The percentage value reflects the number of individuals from which we have received feedback.	

Annex II. Survey and Qualitative Feedback Summary

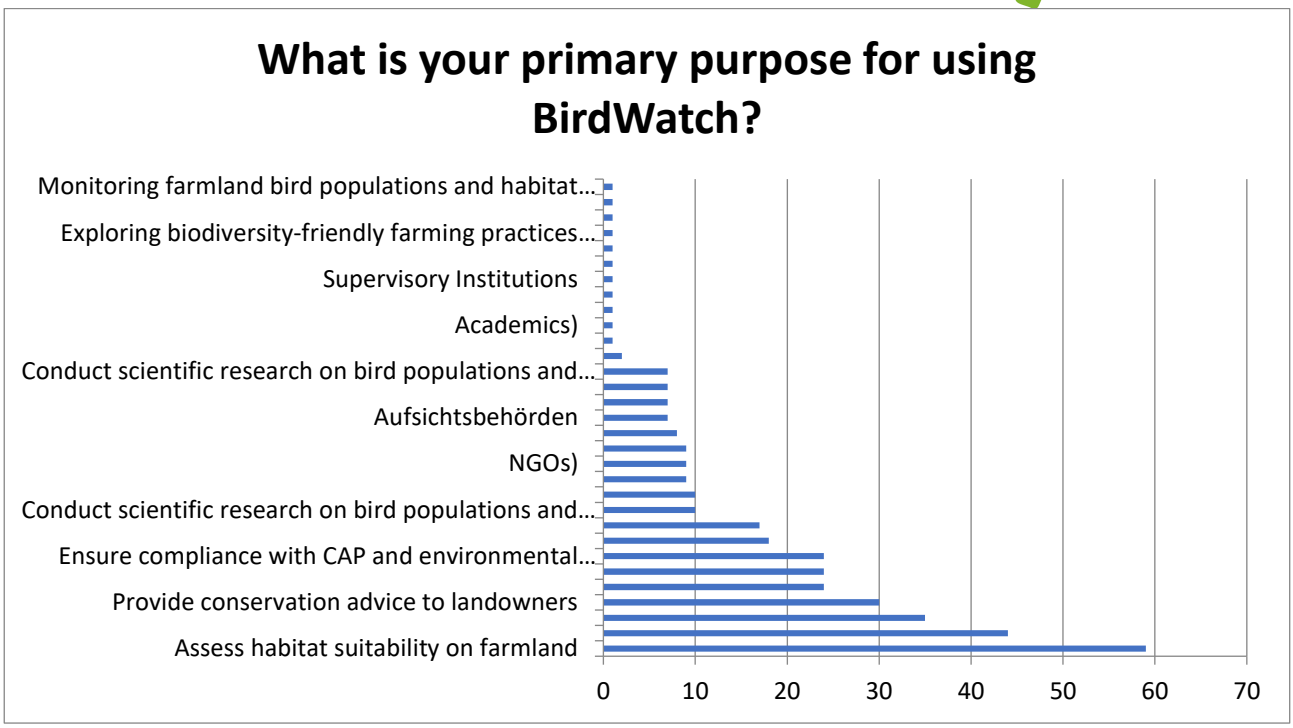
Please select your country/region:



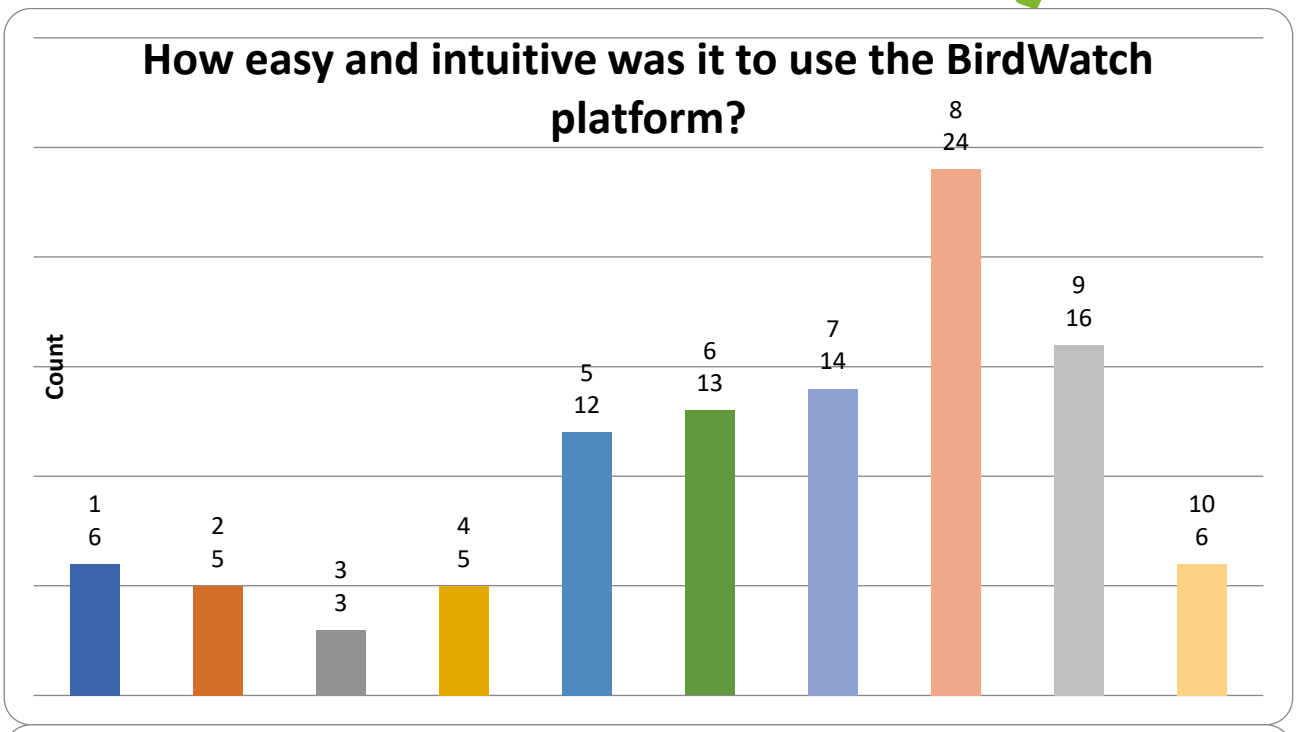
WHICH BEST DESCRIBES YOUR ROLE? (SELECT ALL THAT APPLY) EN VERSION

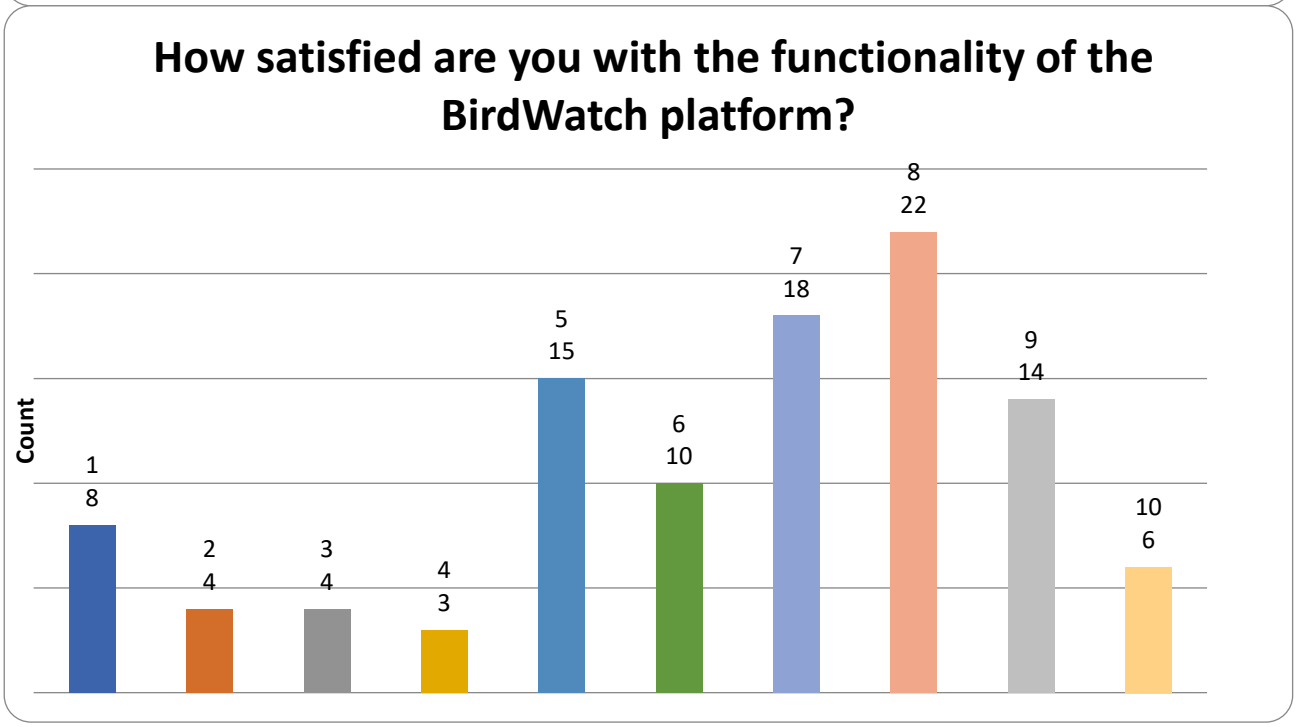


Funded by
the European Union

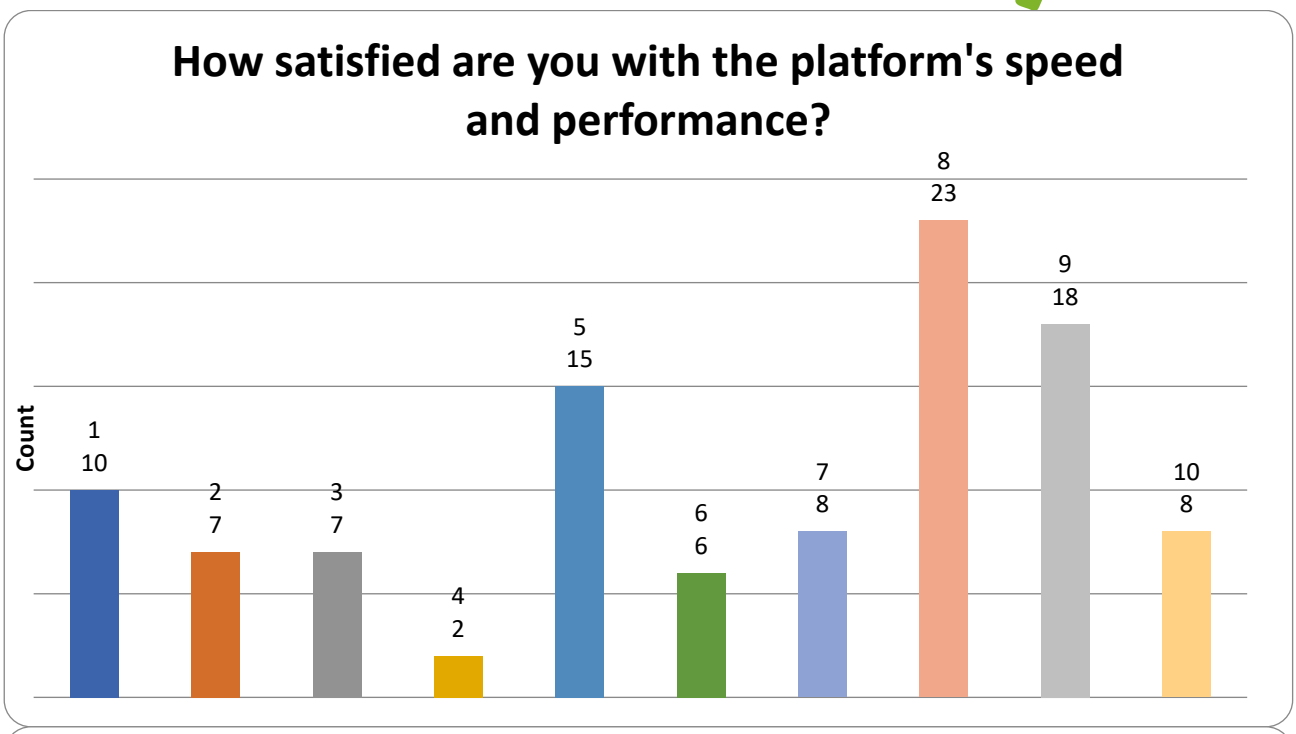


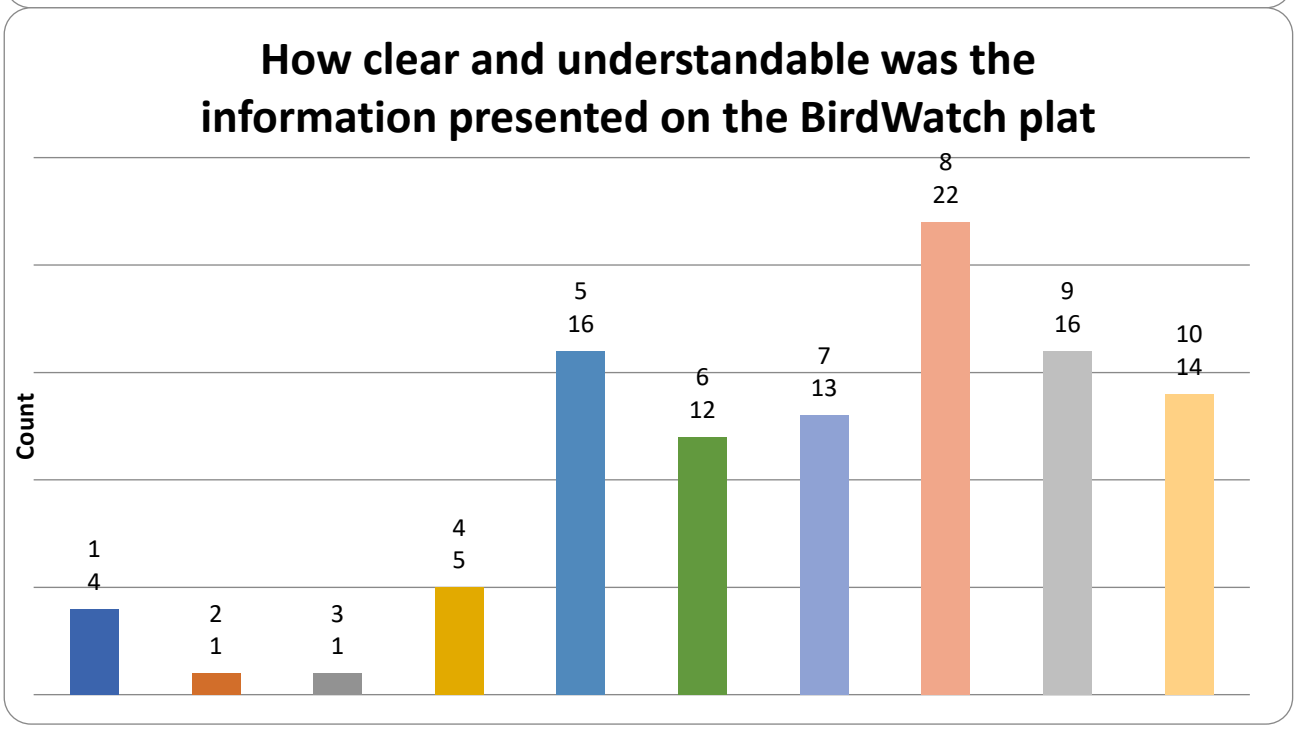
Funded by the European Union



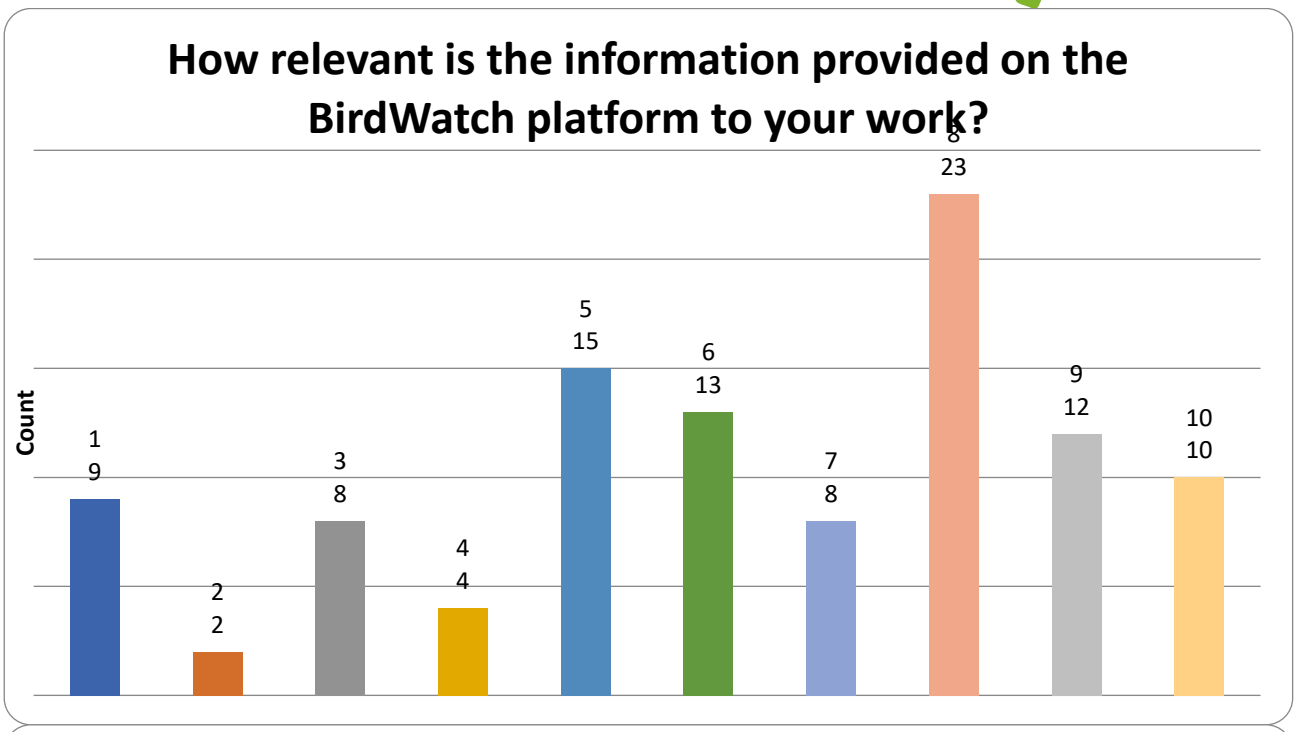


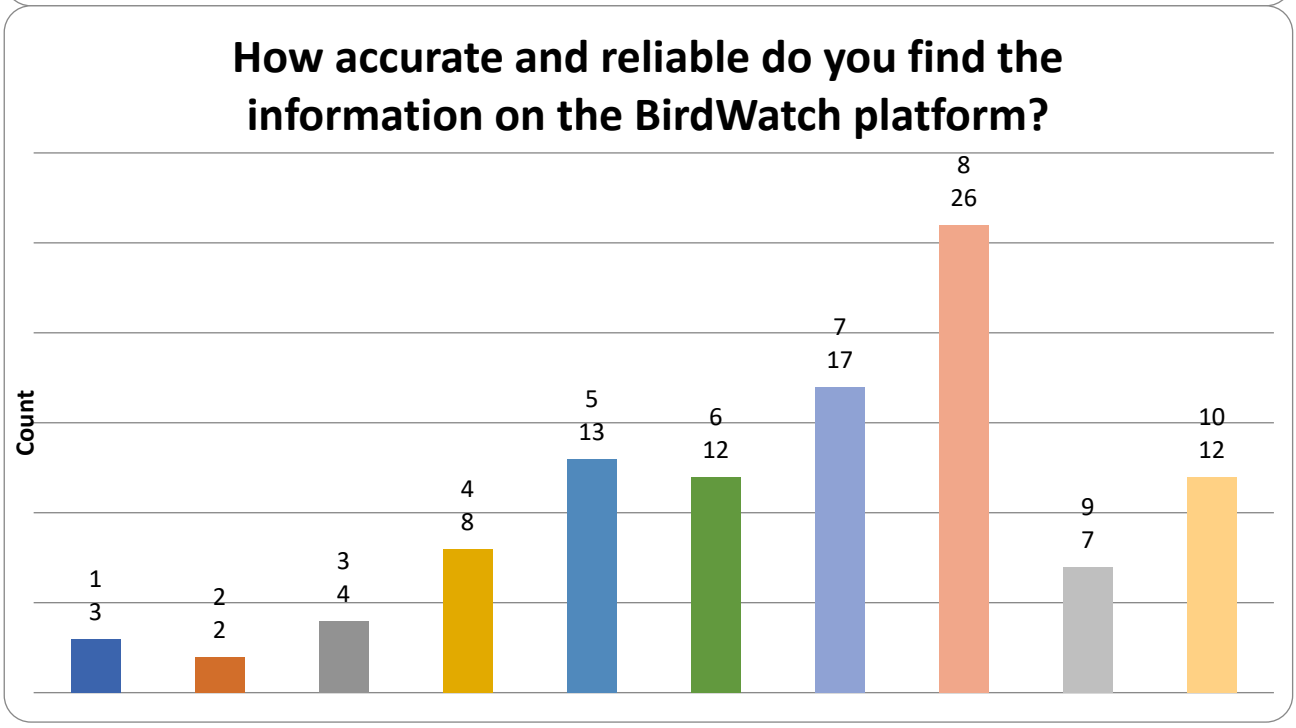
Funded by the European Union



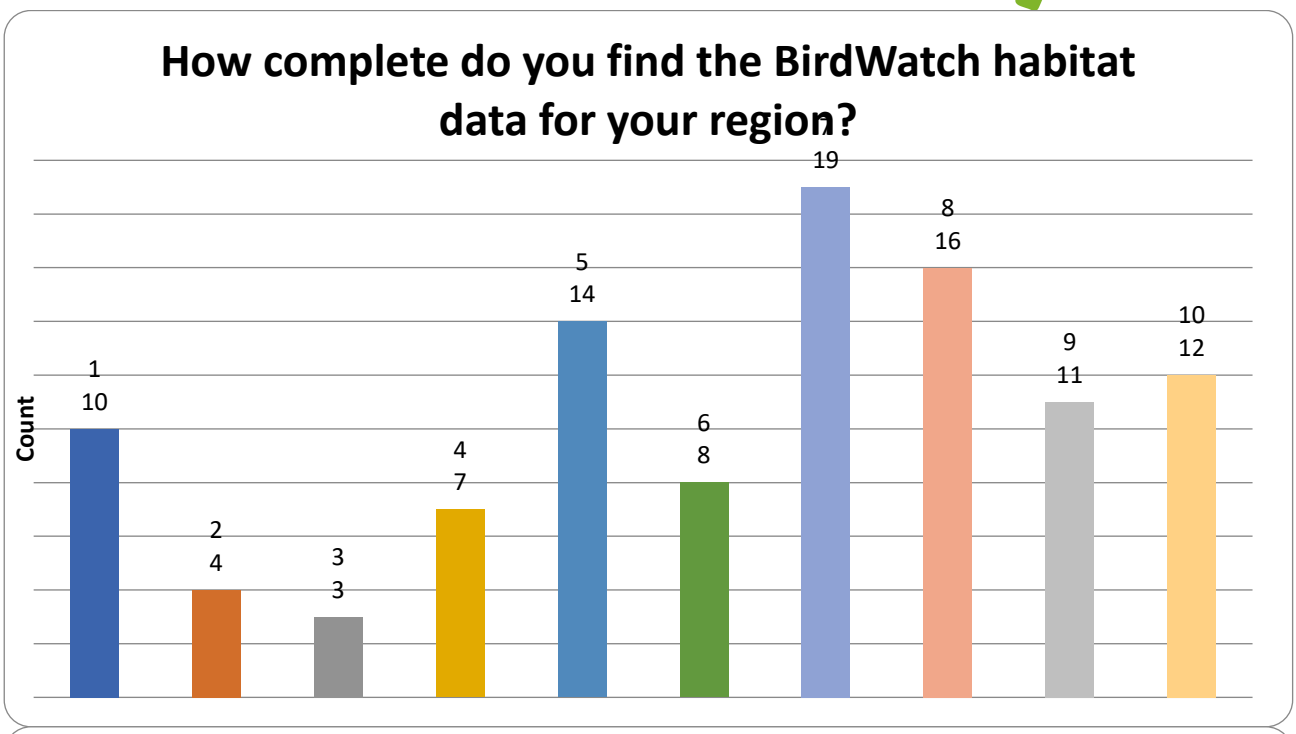


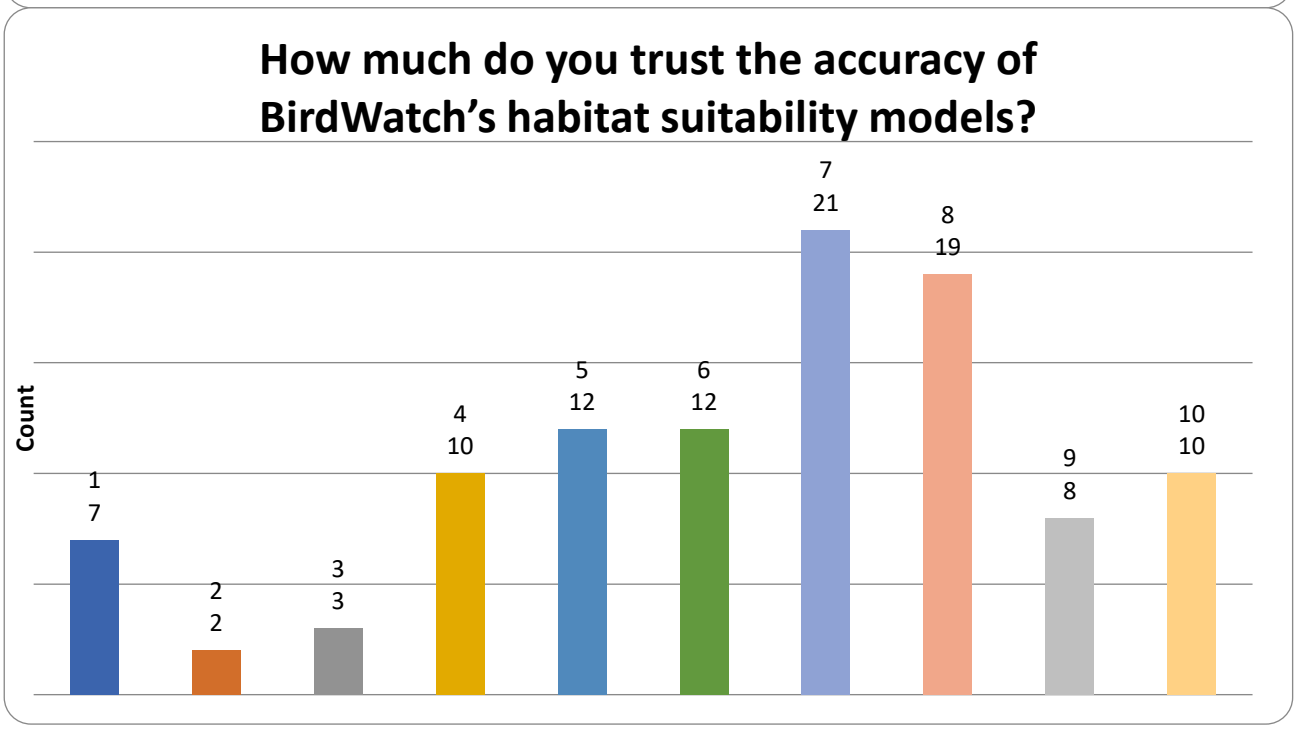
Funded by the European Union



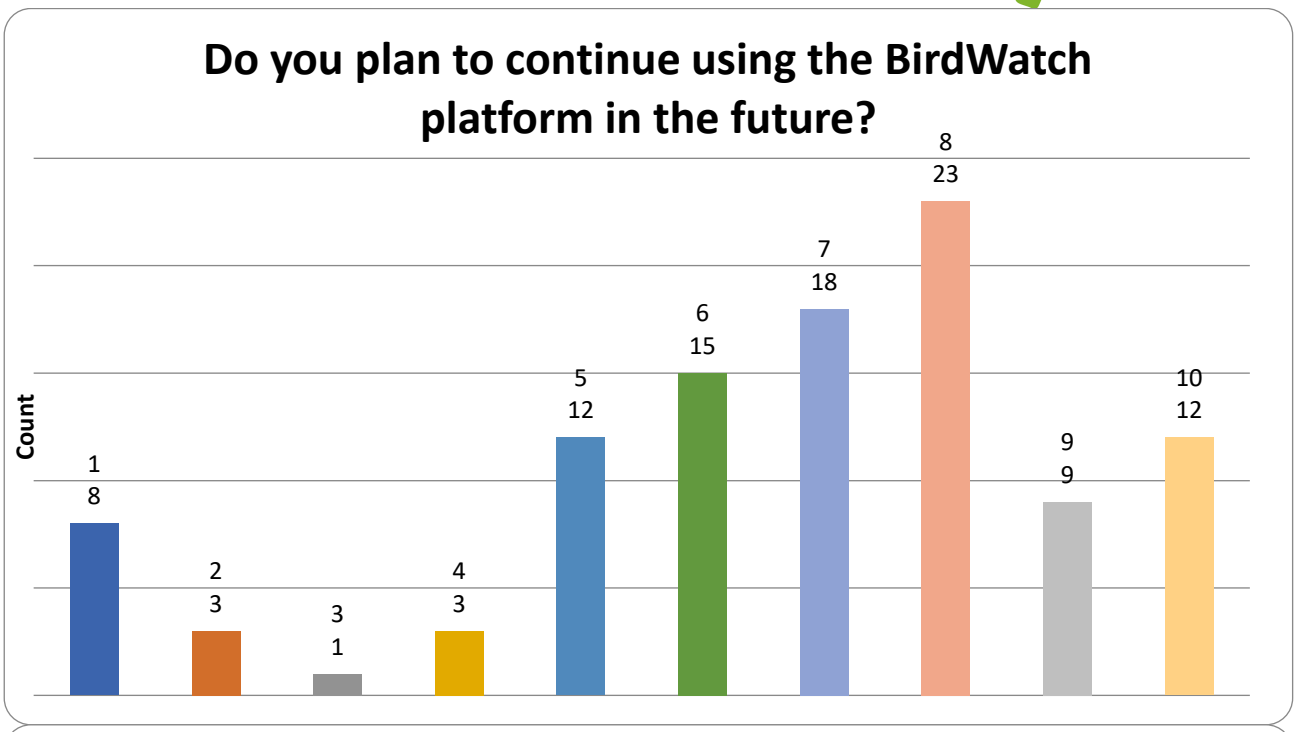


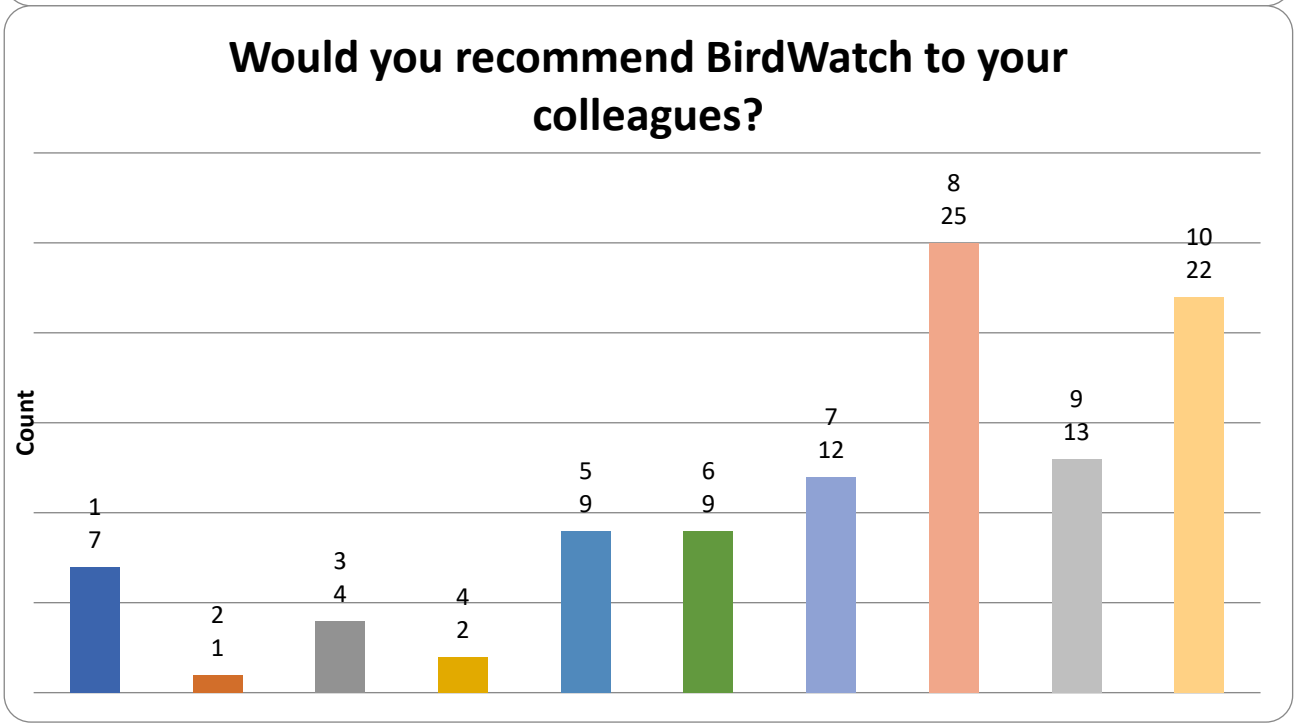
Funded by the European Union





Funded by the European Union





Funded by the European Union