JRC TECHNICAL REPORT

Data-validation solutions for citizen science data on invasive alien species

Tailoring validation tools for the JRC app "Invasive Alien Species in Europe"

Adriaens T, Tricarico E, Reyserhove L, Cardoso AC, Gervasini E, Lopez Canizares C, Mitton I, Schade S, Spinelli FA, Tsiamis K

2021

How to cite this report: Adriaens T, Tricarico E, Reyserhove L, Cardoso AC, Gervasini E, Lopez Canizares C, Mitton I, Schade S, Spinelli FA, Tsiamis K *Data-validation solutions for citizen science data on invasive alien species, Tailoring validation tools for the JRC app "Invasive Alien Species in Europe"*, EUR 30857 EN, Publications Office of the European, Luxembourg, 2021, ISBN 978-92-76-42055-2, doi:10.2760/694386, JRC126140.

Contents

Acknowledgements

Authors

Adriaens T, Tricarico E, Reyserhove L, Cardoso AC, Gervasini E, Lopez Canizares C, Mitton I, Schade S, Spinelli F, Tsiamis K

Abstract

Invasive Alien Species (IAS) constitute one of the main drivers of biodiversity loss globally. Citizen Science (CS) can play a role in influencing the way in which society values biodiversity and can contribute by collecting valuable data. The involvement of the public in IAS recording through CS initiatives using innovative smartphone applications (apps) already complement several existing official surveillance systems in EU Member States. However, the quality of data may hinder the use of CS data for early warning, management and control of IAS. The Joint Research Centre has developed a smartphone app called "Invasive Alien Species in Europe" (JRC IAS app), allowing users to report the presence of IAS of Union concern (EU Regulation 1143/2014) in Europe. The JRC Proof of Concept project "CS data supporting IAS policy in Europe" aims at consolidating the framework for CS IAS data in Europe. Towards this objective, it explores automated solutions for validation of IAS records from the JRC IAS app. The current report includes the description of available tools for IAS validation. Information on currently applied validation strategies and mechanisms was retrieved from existing CS portals. The report describes aspects of data quality in European and global CS portals and how validation of IAS records is carried out. In addition, a refined validation procedure for the JRC IAS app is proposed, taking into account specific taxa and the applicability of methods and tools that are used in similar portals and projects. Further development of a refined validation procedure is expected to increase the quality of the data for science, policy and management, to improve the feedback to and engagement of volunteer recorders, to assist data validators in performing record checks, and eventually to increase the use of the JRC IAS app.

1 Introduction

1.1 Background on alien species

Alien Species are animals and plants introduced accidentally or deliberately into a natural environment where they are not normally found. In some cases, alien species have become invasive, threatening native biota, affecting biodiversity, human health, and the economy. Invasive Alien Species (IAS) constitute one of the main drivers of biodiversity loss globally (IPBES 2019). Despite these negative impacts, reported IAS observations show an increasing trend of introductions of new IAS in recent decades. (Essl et al. 2015; Roques et al. 2016; Seebens et al. 2017).

1.2 The EU Regulation 1143/2014 on invasive alien species

About 14,000 alien species have been reported so far in Europe (Katsanevakis et al. 2015), of which circa 10% are invasive. Recognizing the need for a coordinated set of actions to prevent, control and mitigate the impact of IAS, the European Parliament and the Council adopted the EU Regulation no. 1143/2014 (EU 2014; hereafter referred to as the IAS Regulation) on the prevention and management of the introduction and spread of IAS, which entered into force on 1 January 2015.

The IAS Regulation gives priority at European level to a subset of IAS, named as IAS of Union concern (Art. 4 "the Union list", hereafter IAS of Union concern). Species are included in this list *inter alia* based on risk assessments showing they can cause significant damage to biodiversity and ecosystem services justifying concerted action across Member States and the adoption of dedicated measures at Union level. Currently there are 66 taxa in the Union list, including both animals and plants, but further additions to the list are underway with a further 34 species under consideration to be added in 2021. Under the IAS Regulation, EU Member States must prevent the introduction and spread of listed species through bans on keeping and trading, enforce effective early detection and rapid eradication mechanisms for new introductions, and adopt management measures for species already widely spread. The list of IAS of Union concern has a dynamic character, focusing on species whose inclusion in the list would effectively prevent, minimize or mitigate their adverse impact in a cost effective manner (EU 2014).

The European Alien Species Information Network (EASIN[1]) is the core of the information system supporting EU Member States in the implementation of the IAS Regulation (EU 2014, Art. 25). EASIN facilitates access to data on alien species in Europe, and provides a single repository for accessing all the information necessary to underpin IAS related policy and management decisions (Katsanevakis et al. 2013).

The IAS Regulation prescribes the establishment of surveillance systems which can benefit from the involvement of different sectors and stakeholders, including regional and local communities. This public engagement can contribute to greater understanding and awareness of the IAS problem across society, which is crucial for tackling biological invasions.

1.3 Citizen science for alien species

Citizen Science (CS) can play a significant role in engaging the public in environmental monitoring and management, improved education, and awareness of environmental issues. This awareness is fundamental for alien species policies to be effective (Caffrey et al. 2014). Species occurrence data generated by citizen scientists enable surveillance of emerging and established invaders at larger spatial extents and are essential to scientific activities underpinning IAS policies such as horizon scanning, risk assessment and risk management (Schade et al. 2019). Citizen science can effectively be implemented for passive risk-oriented surveillance, and CS data are used for targeted eradication and control actions (Crall et al. 2012). Moreover, CS has transformative potential influencing the way in which society values biodiversity (Bela et al. 2016) and can contribute by collecting valuable data, including data on IAS. Additional benefits of CS include reduced cost and improved science and technology literacy among participants (Crall et al. 2015). A wide range of projects building on citizen involvement for surveillance, monitoring and management of a range of IAS are currently being implemented in Europe[2]. These range from single-species projects to projects on thematic groups, and span on a range of biomes from the terrestrial, freshwater to the marine realm (Thiel et al. 2014; Schade et al.

---

[1] https://easin.jrc.ec.europa.eu/easin/

[2] https://easin.jrc.ec.europa.eu/CitizenScienceProjects

2019; Johnson et al. 2020). The ambition level of projects and the degree of involvement they want to achieve (from contributory to participatory science *sensu* Haklay 2013) vary, from occurrence reporting or the collection of presence-absence data to more complex measurements and focused research. Increasing the level of participation might also offer interesting opportunities for more detailed studies on invasion dynamics and impact (Cardoso et al. 2017; Groom et al. 2019).

1.4   Smartphone applications on alien species

Citizen science initiatives have flourished thanks to the development of new technologies such as smartphone applications (apps). Many IAS projects use apps (Teacher et al. 2013) or other crowdsourcing tools – such as tools to interpret images from wildlife cameras (Bubnicki et al. 2016; Forrester et al. 2016; Ivan & Newkirk 2016) – to reach a wider audience on the subject of IAS and to involve the public in recording them. Real-time recording and swift validation procedures are key to effective responses for IAS management, and proper data management and swift data mobilization are indispensable to this aim. Additionally, invasion and impact monitoring require more diverse data capture and the generation of innovative (geospatial) information infrastructures (Schade et al. 2013). Projects must address data integration, data openness, data quality, data harmonisation and database interoperability (Crall et al. 2010, 2011; Adriaens et al. 2015; Groom et al. 2017). Awareness on IAS can be raised with different target groups (e.g., anglers, gardeners, beekeepers, hunters, farmers, environmentalists) through dedicated apps for IAS recording. Regardless of the stakeholder groups involved in recording, it is important to implement a quality control on the data by a robust verification procedure.

1.5   The importance of data validation

The design of CS projects is often a compromise between participation and data quality. Mass participation across a larger geographical or temporal space increases the level of engagement and awareness, but projects with less participation but more "expert" citizen scientists can provide more robust repeated measurements (Loiselle et al. 2016). Currently, the majority of CS projects focusing on the recording of IAS have a verification step, whereby data collected by participants (most often images or specimens) are confirmed by experts (Gardiner et al. 2012; Pusceddu et al. 2019; Rowley et al. 2019). A survey carried out by the Alien CSI consortium (unpublished data; Price-Jones 2020) on 103 CS alien species projects in Europe, showed that 89 projects have validation procedures in place. The majority of those are building on expert validation to ensure data quality, others use peer validation, while several well established recording systems have advanced automated flows to aid the validation process (e.g. out of range record checks, automated image recognition). Generally, the most prominent approaches for validation of CS data are peer validation and expert validation, often aided by automatic filtering techniques (e.g. through data mining algorithms, Artificial Intelligence – AI) which can address random variation (e.g. outlier detection) (Balázs et al. 2021). The ease of species identification (and record validation) can vary. Sometimes species identification can be easy and straightforward because species exhibit unique and easily recognizable identification characters. For other species a previous training/experience can be necessary. For example, for plants it can be necessary to take a picture during the flowering period to capture the distinctive traits. Some species look very similar and the identification can be difficult also for an expert without a specimen. Pictures are sometimes not good enough and/or do not report those traits allowing the identification (e.g., juvenile crayfish of different species can be easily confounded). To reduce this problem, spatio-temporal criteria should be included in smartphone apps to help the identification such as the currently known/possible range and seasonal range.

James (2006) describes key features underpinning the quality of biodiversity data:

> ➢ Accurate identification of the thing being recorded (species, habitat, etc.).

> ➢ Precise recording of the geographical locality, depending on survey objectives.

> ➢ Careful documentation of other aspects of the record, such as time or date; the observer who made the record; and the observer who substantiated details of the record subsequently, where relevant.

> ➢ Transparency, robustness and appropriateness of the methods by which collected data are subsequently managed and made available to others.

All these aspects are equally relevant to CS initiatives on IAS. Specifically with regards to regulated IAS, such as the IAS of Union Concern, validation of records (species identification) is of utmost importance, because:

➢ Member States have legal obligations to take appropriate response actions (such as swift eradication under Art 17) in case of new detections of IAS of Union Concern; to avoid delay in notifying new detections, a prioritization approach in validating records could be necessary, especially when the number of submitted observations is high (for example, in the Norwegian Species Observations Service system observations of red-listed and alien species are prioritized).

➢ False positives can lead to false alerts in early warning systems and therefore to unnecessary management interventions and ineffective use of resources.

➢ If data are used for scientific analysis, misidentifications (i.e. a false positive detection) represent a problem for a range of analysis such as site occupancy modeling (e.g. Cruickshank et al. 2019; Altwegg & Nichols 2019) or the calculation of trends in abundance or distribution (Van Strien et al. 2013).

1.6 The "Invasive Alien Species in Europe" smartphone application

The Joint Research Centre (JRC) developed and released in 2017 a smartphone app for iOS and Android devices called "Invasive Alien Species in Europe", allowing users to report the presence of IAS of Union concern in Europe (Tsiamis et al. 2017). The JRC also developed the underlying data management and validation infrastructure. More detailed information on the development of the JRC app can be found in Schade et al. (2019) which outlines the history and the specific policy arena of the app, including the link with the EASIN.

During the last two years, the JRC app has been tailored to cater for specific needs of EU Member States and recently developed CS projects. More concretely, JRC is currently working on establishing collaborations with parties interested in adopting the app, updating its content (adding factsheets of species of regional interest), translating the app's features and factsheets of species of Union concern, improving the data management and validation infrastructure, and developing reporting modules.

The app was specifically designed to aid the implementation of the IAS Regulation through the detection of IAS of Union concern, and some contextual aspects are important for the validation flow:

➢ the taxonomic scope is the EU catalogue of IAS of Union concern (currently 66 species); there is also the possibility to choose "other" species, and species of regional interest when the user is located in specific geographic areas;

➢ registration is mandatory for the user;

➢ species' name, picture, coordinates, coverage/number of individuals (estimated/measured) and environment (marine, terrestrial, freshwater) are mandatory fields (Figure 1);

➢ records are shown on a map (within the app to the user and also on the JRC open science platform) with their validation status: validated, pre-validated, unclear. Validated records are shown on EASIN maps.

Figure 1. Screenshot of IAS in Europe.

1.7 Aim of the report

The Proof of Concept project "CS data supporting IAS policy in Europe" is a JRC project that aims at consolidating the framework for CS IAS data governance in Europe. Towards this objective, this report explores automatized solutions for validation of IAS records for the JRC app "Invasive Alien Species in Europe". This entails the description of available tools, their review and suggestions for a refined validation procedure considering the specificities of the JRC app. The ultimate goal is to improve the current data validation process of the JRC app through automation. This is expected to increase the quality of the data, improve the feedback to data contributors, assist validators in performing record checks and ultimately to increase the functionality and use of the app.

The report highlights:

➢ Information retrieved from existing CS portals and projects on how data quality is ensured within those systems and how the validation of IAS (or other) records is carried out. Contextual descriptions of the existing approaches and tools are provided, together with proposals for their combined use.

➢ A refined validation procedure for the "Invasive Alien Species in Europe" smart app, taking into account specific taxa and the applicability of methods and tools that are used in similar portals and projects. A

graphic representation of the ideal validation workflow for the JRC IAS smart app is also given, including an overview and detailed representations of the individual validation steps.

2 Data validation types in citizen science projects

Data verification and data validation represent critical components of quality assurance. Data **verification** ensures the accuracy of the identification of the things being recorded. Data **validation** involves standardised, often automated checks on the completeness, accuracy of transmission and validity of the content of a record (James 2006). In biological recording, and in this report, both terms are used interchangeably. It is difficult to separate validation procedures from other elements of data quality, which are equally important: survey objectives and design, methods of data acquisition and data management. For instance, the use of a smartphone app for crowdsourcing is already an important mechanism to ensure data quality, as this technology allows for standardized data capture, prevents errors in data transcription and loss of data forms (Crall et al. 2011), prevents errors when manually entering a locality, provides a certainty on the geolocation, ensures quick upload to online databases, and a short time lag between the observation and the reporting. Although the current report touches upon some of these aspects, the main focus is on verification/validation of the content of a record, especially the correct species identification as this is of utmost importance for IAS (see 1.5). The data uploaded in CS portals can be validated in different ways. Here, we make a distinction between four approaches: peer verification, expert verification, automatic quality assessment and model-based quality assessment (Balázs et al. 2021). These approaches consider three specific intervention points: before, during and after the observation took place. Model-based quality assessments, for their complexity, are often performed during the analysis of the data.

2.1 Peer and expert verification

In case of **peer verification**, several experienced, fellow citizen scientists assist in the identification process and validate records. The main goal of this approach is self-regulation by more experienced members within the domain and transfer of this experience to other observers. The result is that data quality is depending on the community of citizen scientists. Examples of European CS projects using a form of peer verification iNaturalist or Vespa-Watch[3] (where validation is performed after sending records to an iNaturalist[4] survey). A large proportion of the European CS projects, however, use **expert verification** (see below) as their main validation approach. In this case, specific contributors or stakeholders are identified as experts within a project, who verify the data generated or collected by volunteers. Continuous expert assistance is required. Examples where expert validation is an important component in the verification process include Mosquito Alert[5] and Waarnemingen.be. Many portals combine expert verification with other validation approaches (e.g. Finland FinBIF and CSMON LIFE app).

2.2 Automatic quality assessment

Automatic quality assessment involves the use of software-based systems to automatically carry out a quality assessment of the data generated or collected by a CS project (Balázs et al. 2021). Artificial intelligence (AI) is currently well-integrated into CS applications as a tool for taxonomic identification (Wäldchen & Mäder 2018; Christin 2019). When a picture is uploaded to the platform, automatic species identification using deep learning recognition models identifies the species (Schermer & Hogeweg 2018) and determines the confidence of this identification. Additionally, statistical analysis is used to determine the plausibility of a species occurrence and the likelihood of a species occurring at a specific time frame in a specific place. When these requirements are met, the observation can be approved and stored. This approach requires training the algorithms on a specific set of species and relevant high quality pictures, to allow the correct identification.

---

[3] https://www.vespawatch.be
[4] https://inaturalist.org
[5] http://www.mosquitoalert.com/en/

2.3   Model-based quality assessment

Model-based quality assessment goes beyond automatic quality assessment (Balázs et al. 2021). Rather than addressing random variation such as naive outlier detection, it focuses on residual errors using an explicit model of how the phenomenon of interest is expected to vary in space and time (Bamford et al. 2009; Kelling et al. 2015). An expert input is required for a concrete understanding of how the phenomena behaves. Model-based quality assessments currently are implemented in Naturgucker and provide additional information as popups at the time an observation is submitted. Often, they are applied in the analysis phase of the data. For example, Kelling et al. (2015) indexed eBird observers' variability using species accumulation curves to account for observer's skill and improve data quality post-hoc. Cruickshank et al. (2019) used a dynamic occupancy model to estimate and adjust for false-negative and false-positive errors, producing accurate estimates of occupancy. The authors validated this model using simulations and applied it to 12 species datasets collected from a 15-year, large-scale volunteer amphibian monitoring program.

2.4   Chatbots and conversational agents

Artificial Intelligence (AI) applications that mimic human conversations, known as "chatbots" or "conversational agents", are increasingly used to interact with users in a conversational format (Klopfenstein et al. 2017). They have the ability to understand spoken language, use speech communication as a user interface and are able to chat with users inside apps. To date, chatbots are mainly used in the private sector to automate communication and manage routine transactions with customers (Poola 2017; Zsarnoczky 2017), although recently, the use of chatbots has been suggested to facilitate communication between citizens and the government (Androutsopoulou et al. 2019). Chatbots can take the form of simple conversational units (Klopfenstein et al. 2017) to more complex systems integrating natural language processing, machine learning and data mining technologies. These complex systems can leverage existing data of various forms, both structured and unstructured and are composed of several services: data management, knowledge processing and application services (Androutsopoulou et al. 2019). These complex systems are highly suitable for approaches including a wide range of citizen's interactions, with high levels of complexity, ambiguity and uncertainty.

However, for relatively simple CS applications that rely on a structured database, clear protocols and a specific CS public, there is no real need to use such complex AI systems. Basically, what is needed is a system that can provide on-the-fly assistance through identification guidelines, notified to the observer once a species name has been entered in the system. This can guide the observer to spot specific species' characteristics needed for a correct identification. Artificial intelligence can also be used to automate feedback to the observer after the expert has validated a report. Human experts are in short supply, and pre-prepared material for feedback is inherently limited in scope. Through the use of a Natural Language Generation (NLG) component, rich information can be provided to the observer (Blake et al. 2012). In the case of incorrect identifications, an NLG component can provide an explanation of this via a feedback to the user.

Results of first experiments with image recognition in the context of the JRC IAS app have been carried out in 2019 and 2020 (Schade et al. 2019; Delipetrev et al. 2020). We explored the possibilities to enhance the image recognition based on the iNaturalist 2017 computer vision model. The experiment included a dataset of 1192 images (of 59 distinct IAS gathered via the mobile app). The developed code is freely available on GitHub[6].

---

[6] https://github.com/deblagoj/iNaturalist-API

3 Validation flows in existing citizen science projects

The COST network AlienCSI (Increasing understanding of alien species through CS, CA17122[7]) performed a large survey on 103 alien species CS projects in Europe probably offering a representative sample of alien species CS projects in Europe (Price-Jones 2020; AlienCSI unpublished data).

Based on this survey, the large majority (91 projects, 86%) of the projects have validation flows in place. Six projects have partially implemented validation systems. Seven projects indicated that the validation system was unknown to them and only two responded they did not have validation in place. Within the subset of projects implementing validation procedures, expert validation was most commonly used, as at least 68 projects indicated to use experts for validation, either solely using experts (57 projects), aided by automated systems (3 projects) or in combination with peer validation (6 projects) and/or automated systems (2 projects). Peer validation and automated validation only represented a minority of projects (4 projects). Eight projects indicated to use expert validation, peer validation and automated systems in combination.

3.1 Qualitative screening of validation systems

Here, we selected several of these projects and recording portals for a more in-depth, qualitative screening of their validation systems, i.e. a description of how the validation is performed and reporting on the experience on quality of the validation flow. We additionally contacted some of the general recording portals, since most have well-established validation workflows. For the selection of these general portals, we focused on well-established systems (active for over 10 years) in Western Europe or global systems used in Europe with well described validation workflows such as iNaturalist, encompassing the current variety of implemented validation procedures (Table 1). As a consequence, there might be other recording portals with alternative validation systems to learn from that are not described here (e.g. https://www.ispotnature.org/, https://www.faune-france.org/). An email was sent to project managers/responsible of app/website asking to provide qualitative information on the validation process:

- ➢ Could you provide us with any information on how the validation is performed in your project/recording portal (web links, figures, descriptions, published information or other)?

- ➢ Do you have any experiences to share about the quality of the validation flow (does it work well, improvements possible)?

On top of this qualitative description of validation systems, scientific evaluation of different systems would be informative. This requires quantitative data on the validation, such as the number of validated records (with or without picture) per taxon. This was outside the scope of the current exercise but could be the subject of a follow-up study. In this context, information about the app users (e.g. age, education) might be relevant. The JRC app "Invasive Alien Species in Europe" does not capture demographic data (e.g. age, education, gender) on its users. However, these could be inferred from or gathered at related activities such as social media communication, training events or bioblitzes. For example, based on JRC training activities and bioblitzes from local partners, we find that the "student profile" (Bachelor/ master students) significantly increased over the past months.

---

[7] https://alien-csi.eu

Table 1. Overview of projects and recording portals contacted to gather qualitative information on their validation systems.

Name	taxonomic scope	validation system	Reference	Publicly available
Waarnemingen.be	general	expert, automated	Kristijn Swinnen pers.comm, Swinnen et al. 2018	no
iNaturalist	general	peer	https://www.inaturalist.org/pages/help#quality	yes
Artsdatabanken	general	peer, expert, automated	https://www.artsobservasjoner.no/Home/About; Toril Moen (pers. comm.)	yes
Naturgucker	general	model based	naturgucker.de	yes
iRecord	general	expert, automated	https://www.brc.ac.uk/irecord/resources-verify	no
KORINA	invasive alien plants in Germany (Saxony Anhalt)	expert	www.korina.info	no
Finvasive (VieKas) LIFE	invasive alien plants in Finland	peer, expert	Markus Seppälä, Mikko Heikkinen (pers. comm.)	yes
Netwerk Ecologische monitoring invasieve vaatplanten	invasive vascular plants in the Netherlands	peer, automated	www.floron.nl/meedoen/exoten-melden	yes
ALPTREES	alien tree species in urban areas and forest in the alpine space region	peer, expert, automated	https://www.alpine-space.eu/projects/alptrees/en/home	no
InvasivApp	invasive plant species in Switzerland and neighboring area	expert, automated	https://www.infoflora.ch/invasivapp	no
CSMON-LIFE	all species (native and alien) in Italy	peer, expert	Martellos & Moro 2017	no
LIFE ASAP	alien species in Italy	expert	Camilla Orlandi (pers. comm.)	no
Plantas Invasoras	invasive alien plants in Portugal	expert	Marchante et al. 2017	no
Project LIFE ARTEMIS	alien species in Slovenia	expert	de Groot et al. 2020	no

3.1.1 The Norwegian Species Observation Service system (Artsobservasjoner)

The Species Observations reporting system Artsobservasjoner[8] is a CS portal where everyone may report species sightings in Norway. Currently, it contains more than 24 million observations. Quality assurance is guaranteed by both expert, peer and automated validation and takes place both before reporting (pre-validation), during reporting and after the observation has been published.

All users must register with a full name as anonymous reporting is not allowed. The main reason is that users must be able to be contacted when validating and performing quality assurance of the observations. Before reporting, the user is encouraged to use the species recognition app Artsorakel[9], identification keys[10] or to ask others with expertise on the relevant species group (e.g. on Facebook or the webpage 'Ask an expert'[11]). All species names and taxonomy are retrieved from the Species Names Database owned and operated by the Norwegian Biodiversity Information Centre (Artsdatabanken).

When an observer starts to register a species in Artsobservasjoner, he/she is notified if the species has not been detected in the county or municipality before. A warning can be sent in case of a rare, red-listed species or an alien species is detected. After publication, the observation may be reviewed by peer validators or experts. Users may comment on each other's pictures and observations in the system, and can report if they discover incorrect or suspicious observations. Additionally, about 140 experts from different biological societies and institutions, trained on validation, are involved in the verification process. They may approve or reject the observation, based on the information provided by the user. A validator cannot process reports to which he/she has participated as an observer or has been involved in confirmation of the species identification.

The validation of observations depends on their importance: observations of red-listed and alien species are prioritized. Observations where there is uncertainty about the species identification, or where species have been reported in unexpected places or unexpected times of the year, are also examined, as well as other species relevant for nature management. The observer can only report on species already found in Norway; for new species, the observer must contact the support system where the likelihood of a correct identification is then examined by the experts.

The validation flow consists of five steps:

Step 1: Review of report:

- ➢ Assessment of phenology (when) and geography (where)
- ➢ Assessment of any photo documentation;
- ➢ Assessment of rarity, possibility of confusion with other species;
- ➢ Assessment of other documentation;
- ➢ Assessment of user;
- ➢ Possible identification / control identification of physical sample (specimen) in a collection.

Step 2: Assign validation category (Table 2). For many observations this is the endpoint of the validation flow.

Step 3: Initiate dialogue with the user.

Step 4: Evaluate new information.

Step 5: Final validation.

---

[8] www.artsdatabanken.no
[9] https://orakel.artsdatabanken.no
[10] https://artsdatabanken.no/artsbestem
[11] https://www.sabima.no/spor-en-ekspert/

Table 2. Validation categories in Artsdatabanken.

Validation category	Based on	Used when/for
Validated	observer's documentation	observations are not photo-documented, but other information is so comprehensive and detailed that they appear credible
	photo or sound	observations have photographs documenting the species identification
	determinator verification	the observation is determined by a third party and this is considered sufficient as documentation
	probability of the finding	observations are not photo- documented or reported in detail, but Other information makes the observation appear credible
Documentation requested		a validator requests documentation of an observation that lacks a picture
Incomplete reporting		used only for birds
Checked, does not need validation		negative observations ("Not found")
Checked, cannot be validated		an observation cannot be verified even if documentation has been received
Not approved		observations that cannot be approved, even after dialogue with the user who reported the observation
Not validated (the observation is not quality assured)		all observations not covered by any of the other categories

When validators log-in, they access lists of observations of the species groups they are validating. These lists are the result of a pre-defined search, showing red-listed species and alien species in the two highest impact categories (SE – severe impact and HI – high impact) that have not been validated in the relevant species group and area in the last month. The results may be presented on a map instead of a list. The validators may also open a map and draw a polygon of the area of interest, resulting in a list of non-validated observations from within the area. Validators may also use the common search to find interesting observations. They have the possibility to search for users, filter by observations they have validated, by red-list- and alien species categories, validation category, etc. (many of these filter options are only available to validators). A third way for validators to find interesting observations is using the 'occurrence catalogue'. This catalogue is a visual comparison of observations from all other data sources compared to Artsobservasjoner, sorted by Norwegian counties, highlighting new observations which should be investigated. Users who show patterns of deviant reporting over time and do not change behavior after attempts at dialogue and repeated messages from the validators and Artsdatabanken, will have their observations hidden by the system. These observations will not be shared nationally and internationally.

Overall, the validation process is perceived by the platform hosts to be fairly good. Most feedback is on small changes to the functionality and not on the overall validation process (pers. comm. Toril Moen).

3.1.2   Waarnemingen.be

The biodiversity portal Waarnemingen.be managed by the nature conservation NGO Natuurpunt was launched in 2008 and yields about 4 million records per year collected by volunteers and the public. To date (2019) it has collected more than 44 million species records from Flanders and the Brussels Capital Region in Belgium. The system belongs to the waarnemingen.be/observations.org/waarneming.nl family with dedicated apps (ObsMapp, iObs, WinObs, ObsIdentify). Since 2019, alien species records are also pushed to GBIF in the highest resolution in dedicated datasets (Vanreusel et al. 2021). The majority of data is collected by non-professional volunteer citizen scientists. The validation procedure is a multi-step process involving AI, expert ("admin") and automated validation; a scheme is available on the website[12].

After submission of an observation, an algorithm checks whether a picture was added as proof of the observation. Observations without supporting material are compared against the autovalidation parameters for that specific species group: a number of species observations of that particular species within a certain radius within a specific time range (Swinnen et al. 2018). If the species' abundance, time and location fall within these autovalidation parameters, the observation is approved. When a picture is included, the automatic species identification using deep learning image recognition models identifies the species (Schermer & Hogeweg 2018), and determines its level of confidence. When the calculated certainty is higher than 90% (the minimum threshold level), the autovalidation parameters are checked. When all requirements are met, the observation is approved. In case one of the requirements is violated, the observation is flagged and reviewed by an expert. Observations supported by a video or sound are currently not processed automatically and are also reviewed by the expert. All rare or very rare species which do not meet the criteria are individually examined by species specialists and are rated based on the probability of correctness. The validation of observations without proof is mostly an interactive procedure in which observers can be asked for additional information by a team of validators (experts per taxonomic group). The remaining observations can be examined by regional species experts or remain non- validated. Validators do have the possibility to also validate records based on user experience.

It is interesting to note that waarnemingen.be organizes "validator days", where validators are discussing verification issues, user friendliness, progress of validation across different taxonomic groups, gaps etc. In this way, the validation system profits from user's feedback.

3.1.3   iRecord

iRecord is a UK recording portal for wildlife records.. Validation is assisted by automated record checks. These checks are run every 30 minutes to highlight records which are outside the known spatial or temporal range of the species, and on records of species which are difficult to identify. If records are highlighted, the observer receives a message on the home page, with the option of editing the record. A full guidance for verifiers   is available on the website[13] as well as a "cheatsheet" of the workflow for verifiers[14].

The verification of records on iRecord applies a two tier approach (Table 3). Verifiers are able to choose from a set of terms that they can apply to a record (Figure 2).

The validation categories are as follows:

Accepted: the record is accepted as meeting the standard required for inclusion

➢ Correct: The verifier is able to confirm that the species has been identified correctly, usually on the basis of photo/s within iRecord (or specimen/s outside iRecord)

➢ Considered correct: The verifier has not seen photo/s or specimen/s but has a high degree of confidence that the record is likely to be correct, based on difficulty of ID, date, location and recorder skills/experience etc.

Not accepted: the record is NOT accepted as meeting the standard required

➢ Unable to verify: The verifier has a high degree of confidence that the record is likely not to be correct based on difficulty of ID, date, location and recorder skills/experience (and where no photo/s or specimen/s are available); or photos are available but do not show enough details to confirm the identification; and/or the record is not sufficiently well documented to confirm (e.g. location is vague)

➢ Incorrect: The verifier is able to confirm that the species has not been identified correctly, or the record is erroneous in other respects, on the basis of photo/s or specimen/s, or on information from the recorder.

Unconfirmed: no decision has been reached as to whether the record meets the standard required for inclusion (i.e. has not yet been looked at, or if looked at no decision yet reached)

➢ Plausible: The record is plausible based on species, date and location, but is from a source of unknown skills/experience

➢ Not reviewed: The record is in the system but has either not been looked at, or has been looked at but a decision not yet taken.

Issues related to validation categories can be discussed between validators through surveys organized with the validators' community. A number of newsletters were issued to inform verifiers about pending issues, technical developments, changes to validation categories, integration of records from other databases such as iNaturalist etc.

Table 3. Two-tier verification terms in iRecords.

Verification status 1	Verification status 2
Accepted	Correct
	Considered correct
Not accepted	Unable to verify
	Incorrect
Unconfirmed	Plausible
	Note reviewed

Figure 2. Screenshot of validation tabs in iRecord[15].
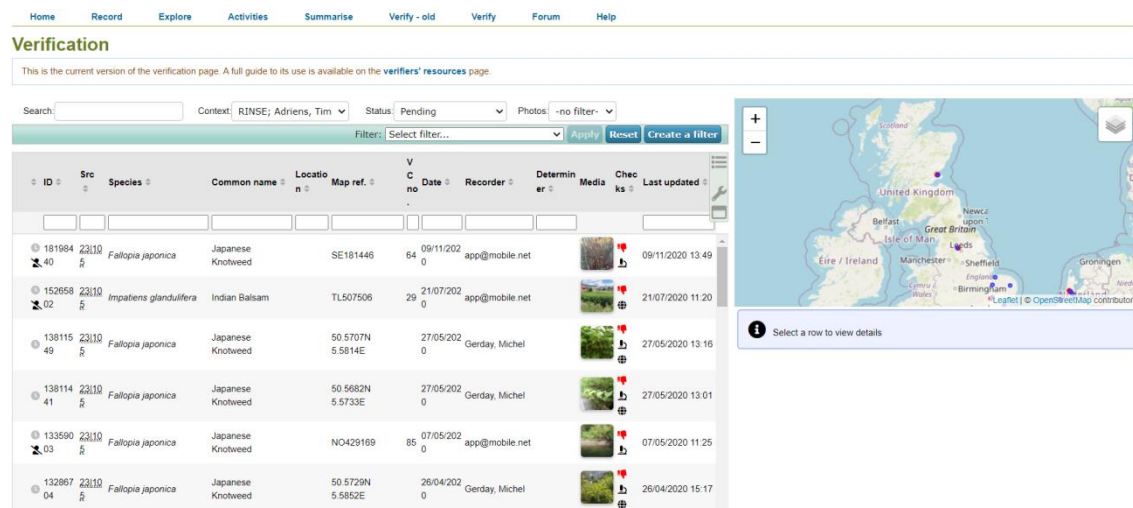
3.1.4 Naturgucker

Naturgucker.de is the biggest recording portal in Germany which also represents a social network for naturalists. Information on how data quality is dealt with is available on the website of the portal[16], supplemented by personal communication with the site managers. The basic principle is to upscale the number of records in such a way that errors in individual records become statistically irrelevant.

Naturgucker.de formally rejects a systematic plausibility check of all observation data on the consideration that such procedures are often potentially manipulative and can corrupt the database. The point of view, is that, albeit some exceptions, individual nature observations cannot be classified as true or false by external people (Munzinger et al. 2017a, b; Ott & Munzinger 2017). Observations do not represent testable, empirical-scientific statements, but are fundamentally based on a subjective experience of an observer who is processing visual and/or acoustic stimuli.

Although there is no general data plausibility check at naturgucker.de, the platform still implements data quality assurance, which is processed exercising the least "manipulative" effect. This data verification runs in parallel at three levels:

(1) **Specialist information** when entering observations: when an observation is very unlikely according to phenological or geographical characteristics, observers receive a corresponding message (described as a "hint") immediately upon recording. Observers can then decide whether they still want to report the observation or not. Detailed information on this procedure is available on the website[17], but there are multiple ways in which an observer is informed about characteristics of his/her observation:

> (a) A popup is displayed as soon as an observer wants to report a "rarely reported" species i.e. a species with a mAI value of ≤ 2 in the database of naturgucker.de. The mAI value is a data characteristic of the system that relates to the frequency of a species being reported (Table 4). It is calculated on a daily basis and consists of:

>> (i) relative frequency of a species (AAI): number of observations of a species in a defined period in relation to the number of observations of the associated species class in that period. For insects (dragonflies, butterflies), the higher taxon is considered the Order level, for other species groups (birds, mammals) the Class level is considered;

>> (ii) the relative distribution (AGI), calculated as the proportion of areas (TK25 quadrants) in Germany with reports of the species in relation to the existing number of areas of the species group within a period of time;

---

[15] https://www.brc.ac.uk/irecord/verification
[16] https://www.naturgucker.info/naturguckerde/datenqualitaet
[17] https://www.naturgucker.info/naturgucker-tipps/hilfe-zu-beobachtungen/fachhinweise-zu-beobachtungen

(iii) a measure for observation intensity (AMI), calculated as the number of observers that detected the species on the total number of observers in a defined period.

(iv) The mAI is then calculated as mAI = (AAI + AGI + AMI) * 100 Using dragonflies, Ott & Munzinger (2017) have shown this index is also useful to draft red lists of threatened species at different geographical scales.

Table 4. Ratings for the mAI value.

Range of values	Frequency in naturgucker.de	Abbreviation
mAI >50	very often	Sh
20<mAI<50	relatively often	Rh
10<mAI<20	moderately frequent	Mh
5<mAI<10	little/isolated	Wv
1<mAI<5	kinda rare	Rs
0<mAI<1	very rare	Ss

(b) Similarly, procedures are implemented to flag/hint records outside normal phenological boundaries or exceptionally high abundance reports of species to the observer at the time of submission.

(c) On top of that, at the level of the higher taxon, there are calculations built into the system, which can be found in most species profiles that ensure further transparency0 (i)a relative statistical measure of the **data resilience (IDB)** of all observations of the species in naturgucker.de. This resilience is based on the number of observations, the number of observers, the number of pictures as well as the geographic distribution of pictures of a species in the system.

(2) **Community control:** since all observations and images are displayed to all visitors to naturgucker.de, everyone can check whether they consider an observation to be probable or an identification to be viable. If not, they can either write to the observer directly by email via the Naturgucker system or leave a publicly visible comment. To do this, it is necessary to be registered and logged into the system. The process of open community "control" is extremely effective and often takes effect within a very short timeframe. In addition, it is highly comprehensible, so that everyone can form their own opinion and is not restricted to any "expert opinions" of individual third parties. Any plausibility check process that may take place becomes comprehensible for everyone. Also, all comments are publicly visible to everyone.

(3) **Advisory board:** The advisory board has been active at Naturgucker.de since spring 2010 and is composed of experienced nature observers and scientists. The members of this board search for unusual reports directly in the database of Naturgucker.de for critical species.

The resulting data quality is deemed very high and on the website is compared to the level achieved by eBird, in which over 500 controllers are active.

3.1.5 iNaturalist

In iNaturalist, the main validation system is through peers;

The building block of iNaturalist is the verifiable observation. A "verifiable" observation is an observation that:

> ➢ has a date;

> ➢ is georeferenced (i.e. has lat/lon coordinates);

> ➢ has photos or sounds;

> ➤ is not of a captive or cultivated organism (this is an important discussion point for alien and domesticated species, cf. Groom et al. 2021).

Verifiable observations are labeled "*Needs ID*" until they either attain *Research Grade* status, or are voted to Casual. Observations become "Research Grade" when the community agrees on species identification, i.e. when more than 2/3 of identifiers agree on a taxon. This is the community ID

If there's a disagreement, a higher level taxon might be chosen that contains of the taxa between which there is disagreement. The algorithm has a slight tendency towards dissent, as iNaturalist have found that dissenters are often right. Observers do have the choice to turn off the whole idea of community ID in their settings.

Observations will revert to "Casual" if:

> ➤ the conditions for Verifiable are not met;
>
> ➤ the community agrees the date does not look accurate;
>
> ➤ the community agrees the location does not look accurate;
>
> ➤ the community agrees the organism is not wild/naturalized;
>
> ➤ the community agrees the observation does not present evidence of an organism, e.g. images of landscapes, water features, rocks, etc.;
>
> ➤ the community agrees the observation does not present recent (~100 years) evidence of the organism (e.g. fossils, but tracks, scat, and dead leaves are ok);
>
> ➤ the community agrees the observation no longer needs an ID and the community ID is at or above family level;
>
> ➤ the observer has opted out of the community ID and the community ID taxon is not an ancestor or descendant of the taxon associated with the observer's ID.

Furthermore, there are caveats and exceptions:

> ➤ "Research Grade" observations will become "Needs ID" if the community ID shifts above the species-level.
>
> ➤ "Research Grade" observations will become "Needs ID" if the community votes that it needs more IDs.
>
> ➤ Observations can be "Research Grade" at genus or any other level below family if the community agrees on an ID at that level and votes that the observation does not need more IDs.

The system will vote that the observation is not wild/naturalized if there are at least 10 other observations of a genus or lower in the smallest county-, state-, or country-equivalent place that contains this observation and 80% or more of those observations have been marked as not wild/naturalized.

3.1.6 Invasoras

INVASORAS[18] is a science communication endeavor that targets the general public (Marchante et al. 2017). Through a WebMapping platform the project intends to engage volunteers to geolocate invasive plants in Portugal. Besides gathering data in support of scientific research and management, the platform raises awareness about invasive plants. It is available since March 2013 and receives data from a smartphone application and a Web application. Validated sightings are shown on an online map. All data is available through a Google Fusion Table and the GBIF database (Marchante et al. 2021).

For INVASORAS, sending pictures with the observation is mandatory. Moreover, besides the pictures, contributors have to fill out a form that includes three compulsory parameters: (a) geolocation (automatic with the smartphone app, manually searchable in Google Maps or by inserting coordinates); (b) density of the plant population; (c) habitat, and three additional, optional, details: (d) phenology; (e) plant development; and (f) degree of certainty in the species identification. This approach has been adopted to meet the two basic requirements: (1) the app is expected to be quick and (2) the app should only request the essential information in order to maintain a low threshold to participation, to be easy in use and to not demotivate users (Roy et al. 2012). If necessary, before validation, the contributor may edit the observation in its personal page to change some features. After the submission, experts verify and validate the observations that are visible on the map

---

[18] https://invasoras.pt/en

on the website. In case of doubts, experts can contact the users to ask for extra pictures or to provide more context. Besides the verification of the species identify, the coherence of geolocation (e.g. aquatic plants should be in water) and phenology (expected for the season) are checked. When users submit a record for a species from a region where experts were not aware of its presence, they may contact the user for more details (Marchante et al. 2017).

Based on the opinion of experts involved in INVASORAS, the system is time consuming, yet experts generally validate observations once a week on a fixed day. In future versions of the app/system, the project envisages to implement "crowd validation". Machine learning/AI recognition are indicated as possible tools of great help for the validation process.

3.1.7 Life projects on invasive alien species

Many LIFE projects developed an app to record alien species.

Finvasive LIFE: the project has developed an app for reporting invasive plants. The app is linked to an online reporting form (https://viekas.laji.fi/). The form is really simplified to facilitate reporting. The form automatically uses the phone's GPS information, or the location can be marked on the map. The form also encourages the user to submit a photo of the species, if available. The project form also asks the user to tell if a reported invasive plant species is in bloom, or growing by a water body. Entries submitted with the form are gathered in the Finnish Biodiversity Info Facility's (FinBIF) laji.fi -database[19]. Species reports submitted to the FinBIF database can be validated by peers or acknowledged professionals registered in the laji.fi -system. Observations have a "quality tag" and a quality comment used to update the observation. These tags allow us to filter the observations. Automatic controls are performed on observations[20]. A large part of the validation flow is crowdsourced.

LIFE ASAP[21]: the project developed a smartphone app (ASAPp) to report alien species of different taxa present in Italy. The app is linked to a platform accessed by experts who were responsible for the validation process. Observers should upload a picture and experts check the picture together with the geographic coordinates, details on abundance and identification suggested by the observer. After validation of the observation, data are published and the observer receives a notification about the outcome. The workflow was simple, although the experts verifying the observations were too few. A larger group with someone supervising the process is recommended to have a better performance and on time validations. Regular notifications should be sent to validators to "encourage" the validation process.

Invazivke (LIFE ARTEMIS[22]): the experts double check and validate the pictures uploaded by the observers with the app Invazivke developed to report invasive alien species in Slovenia, particularly in forest ecosystems. Pictures are mandatory. Once the experts validate the observation, data are published in the online public database and the observer receives a notification about the outcome. Observers can be contacted for further information (de Groot et al. 2020). The system works well, according to the project manager. The "weak point" is the resources necessary to maintain it with an administrator constantly managing the website.

CSMON LIFE[23]: through the portal of the project and the dedicated app, reports on both native and alien species present in Italy and belonging to different taxonomic groups were sent to the experts. Initially, data was manually checked and validated only by a group of a few experts. Observations without a picture were rejected. An automatic feedback was provided to the user in case of a positive or negative identification, while a manual message was sent to the user when the expert needed further information (Figure 3). For certain special activity ("Chiedilo all'esperto" "Ask to the expert"), the expert sent emails to the citizens in case of wrong identification to explain the error, thus providing online training in the identification of different groups of organisms (Martellos & Moro 2017). Validated observations are published online. During the project, to speed up the process, skilled (sometimes trained) volunteers from project partners and stakeholders were involved in quality control with some of them directly using the back-end of the project portal, while others were contacted by project experts only in particular cases. The "consensus" approach was not applied, limiting the quality control to a single expert. Recorded data were aggregated in the National Biodiversity Network. At the end of the project, a mixed system (automated and manual control) was proposed to improve the quality control protocol (Figure

---

[19] https://laji.fi/en
[20] https://laji.fi/about/772
[21] https://www.lifeasap.eu/index.php/en/
[22] https://www.tujerodne-vrste.info/en/project-life-artemis/
[23] http://www.csmon-life.eu/

4). In this proposed protocol, data are classified according to the different levels of quality (Martellos & Moro 2017).

Figure 3. Data quality control protocol adopted during CSMON-LIFE (from Martellos & Moro 2017).
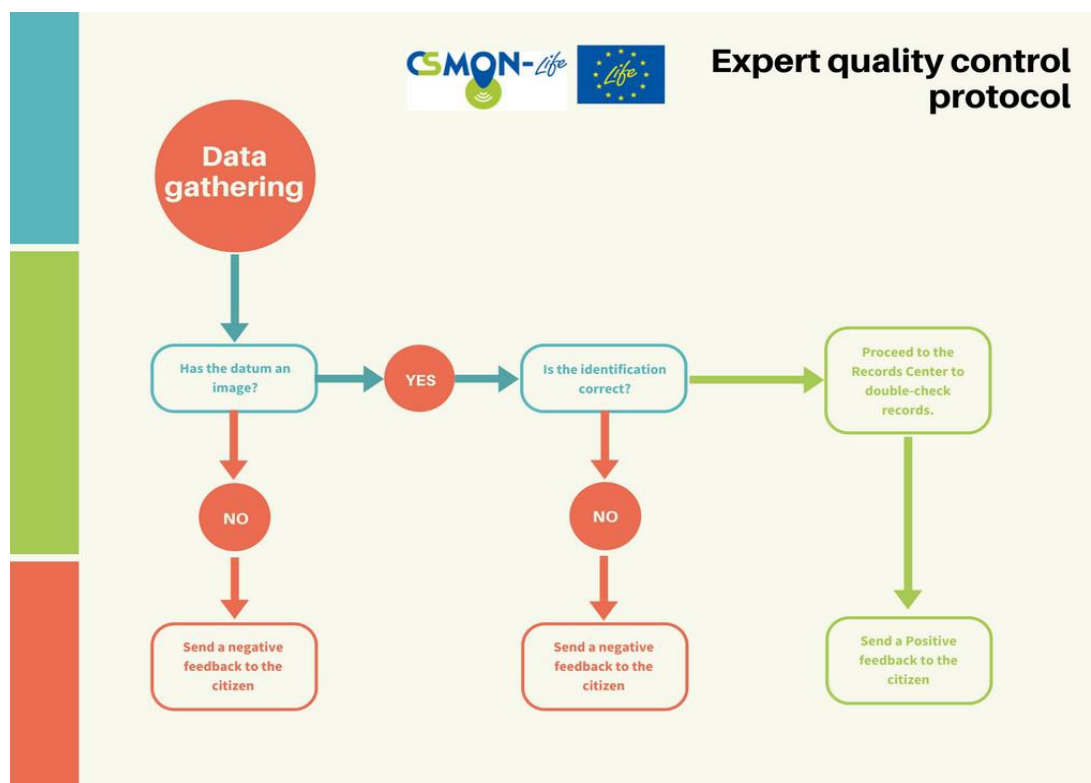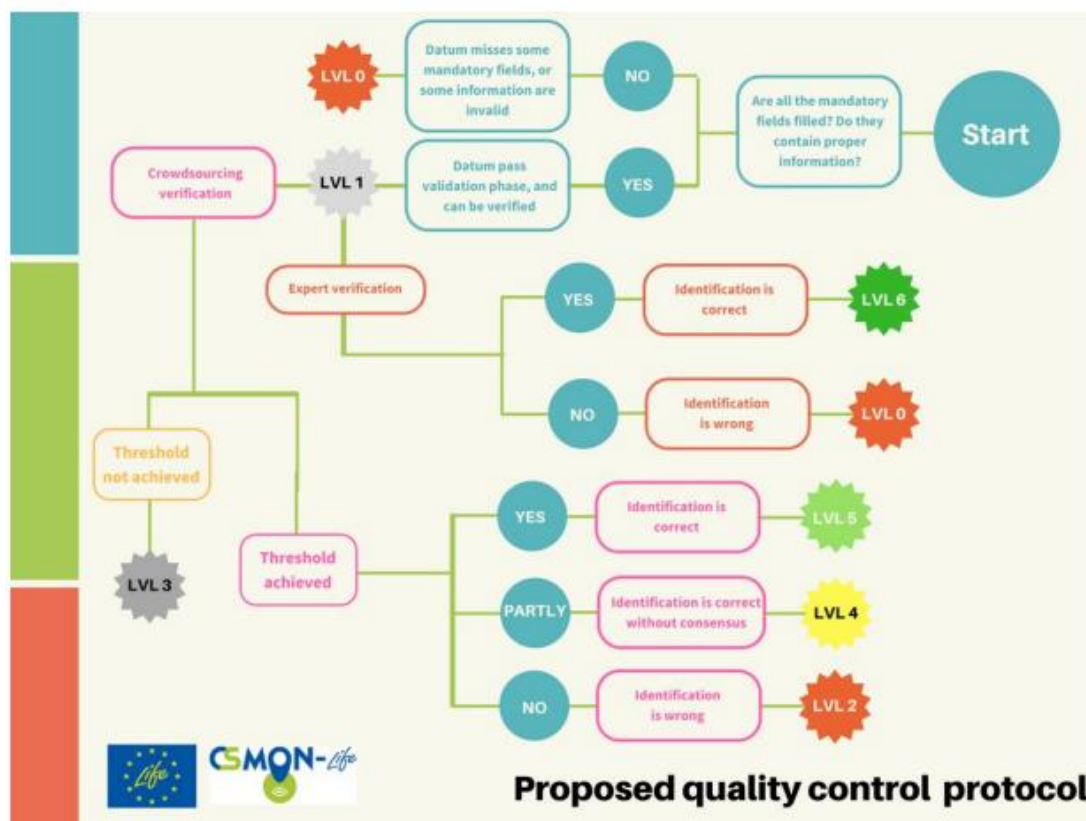

Figure 4. Proposed data quality protocol by CSMON LIFE (Martellos & Moro 2017).

3.1.8 JRC "Invasive Alien Species in Europe" app

The validation flow from the current 'Invasive Alien Species in Europe' integrates both expert and automated validation at three specific stages of the data cycle: before, during and after data gathering (Schade et al. 2019). The application integrates multiple factsheets including information about the invasive species' morphological features, behavior, invasion history and potential confusion with similar species. Complementary information can be found on the EASIN web page. Additionally, technical specifications are provided through the user manual of the app. When reporting a sighting, data quality is guaranteed by providing controlled lists for data entry.

The following information is required: species name, picture, location, coverage and habitat. Once the geospatial data is stored, the picture is analyzed by automated image recognition software to pre-validate the record (scan for inappropriate content and identify images with plants or animals). After this, the record awaits validation by the expert, for which a dedicated tool was developed. An expert can change the attributes, discard images and modify the location. All original information of the observation is stored. Once validated, the observation is stored in the EASIN database.

The validation categories are specified as follows (Schade et al. 2019):

Submitted: the observation at the given geospatial location is visualized on the online map, along with its associated attributes, such as the time of observation. A check is performed on whether the record contains inappropriate content, which is aided by automated analysis of images.

Pre-validated: assigned status once the image and possible free text comments have been assessed and moderated. At this stage, also images and comments are displayed in the viewer.

Validated: after a recognised expert has verified the observation using the information from the picture and additional knowledge about the species and its occurrences. To aid the validator, the back-office includes a guided workflow to filter records. Filters can be used to retrieve only the records of certain species, within a defined time interval, or having a specific status, for example, "pre-validated but still requiring expert evaluation". Records can then be selected and further processed (Schade et al. 2019).

Validated IAS records are then made available as Open Data via a RESTful Application Programming Interface (API).

4 Recommendations for the JRC app "Invasive Alien Species in Europe"

4.1 Quality assurance implementation levels

Quality assurance can be implemented at four different levels:

➢ at the level of defining project settings.

➢ by providing guidance to observers.

➢ through (human) validation of observations.

➢ at the level of the use of the data.

Some of these features are already implemented in the JRC IAS app, while others are proposed in detail in the following subchapters.

4.1.1 Project Settings

At the project setting phase several data quality assurance measures can be defined (see also Figure 5), such as:

1) Define required fields (already implemented). In the JRC IAS app the following information is already required: species name, picture, location, coverage and habitat.

2) User registration (already implemented). Registration is needed to get additional information from and to provide feedback and communicate to observers. Registration is already mandatory in the JRC IAS app.

3) Species list (already implemented). A defined list of species allows the observer to select a specific name instead of typing it. This reduces the risk of erroneous reports. The JRC IAS app already has a defined list of species, i.e., the IAS of Union concern and lists of regional concern. This ensures that the application is fit for reporting on specific species and adds to its policy relevance.

On the other hand, the list of IAS of Union concern includes by 2020 only 66 species. This limited taxonomic scope represents a challenge in terms of fostering mass participation. Citizens are prone to participate and retain their engagement when they can effectively report an IAS and receive feedback on their contributions (Geoghegan 2016), but several species in the list of IAS of Union concern are very rare or do not occur in EU territory. One solution to overcome this barrier to participation is to make the JRC IAS app suitable to report on other species outside the legislative scope of the application. Observations of non-regulated species are of scientific and biodiversity protection relevance and observers should be able to receive feedback on them. These observations can be exchanged with other recording platforms (cf. the iNaturalist integration considered by NBN iRecord). The JRC IAS app should clearly communicate where these records are stored and accessible and how users can see the outcome of their contributions. For example, the Flemish CS project Vespawatch[24] automatically sinks data with a survey on iNaturalist[25] under a generic project profile (Schoonvaere et al. 2020). The functionality to synchronise data with iNaturalist was developed as a stand-alone open source module (Python library[26]), so that other software developers can use it in their applications. This way, data of other species outside the scope of the IAS of Union Concern could also be validated by the community of peers and when validated, uploaded as open data on GBIF.

4) Allow record submission of **"captive" or "domesticated" populations**. Data on the size and geographic distribution of captive populations are important to identify and quantify alien species' propagule pressure. There could however be a requirement to indicate such records from gardens, aquaria or wildfowl collections (cf. a check-box captive/cultivated) by the observer as it is done for instance in other portals (iNaturalist, waarnemingen.be), but for a dedicated IAS application, this option could be opt-in and preferably mandatory. The ethical and juridical aspects have to be considered upon doing this since such records of these regulated IAS could result in juridical action by authorities.

5) Consider the inclusion of **records without pictures**. This would lower the threshold for users to submit an observation, e.g., when the organism is difficult to capture on a picture (e.g. in aquatic environments) or when it is moving too fast (e.g. mammals/birds).

6) Include a specific **training component**. Users should be trained in using the app, understanding the need of quality data and in identifying the focal species. This can be realized by developing online training materials and readings, and by organizing training classes (also in the field). A Massive Open Online Course (Have you seen an alien?[27]), targeting teachers, high school students and the generic public is already available.

4.1.2   Guidance to the user

1) Provide a **user-friendly identification guide** (already implemented). This can guide the observer to spot specific characteristics (e.g., color of the legs in hornets, shape of rostrum in crayfish, flowers or shape of leaves for plants) needed for a correct species identification. This can also stimulate the observer to better document their observation by taking more and better pictures.

2) Put the most **common species on top** in the species selection window. The spatiotemporal components of the observation will be used to filter for specific species names, such as those known to occur within a certain radius and a specified time range. Through a dedicated query calculating the relative frequency of the different species per country, participants are more likely to find species they can report on. This could stimulate the use the JRC IAS app for reporting them.

3) Implement an **image recognition algorithm** in the app to assist the observer in making a correct decision on the species identification. The image recognition software should also notify the observer when the quality of the picture is too low. Several good systems are available (e.g. Tensor Flor is used in iNaturalist, Google Cloud Vision) but these algorithms need to be trained on the specific set of species of the JRC IAS app. To increase the usefulness of the app to border control and nature inspection services, a possibility would be to also include

---

[24] www.vespawatch.be
[25] https://www.inaturalist.org/projects/vespa-watch
[26] https://pyinaturalist.readthedocs.io/
[27] https://academy.europa.eu/courses/have-you-seen-an-alien

training pictures of the species in typical "interception" conditions, which are usually very different from situations of occurrence in the wild (potted plants, containments, controlled holdings, animals in cages or reservoirs...). There is certainly a need with border inspection services to learn about identification of the Union Concern IAS and therefore scope to attract this audience as a target user group of the app. This would however possibly also require some adaptations as to where the data on inspections should go and the degree to which they should be in the public domain.

4) Tailor the record submission for **additional types of media** (e.g. sound recordings, videofiles). Several examples of sound recognition software include Sound Analysis Pro 2011, Avisoft-SASLab Pro, and Song Scope.

5) Tailor the app to **harvest essential data from the uploaded multimedia files**. This could increase user friendliness by automatic filling of location and date (e.g. from the EXIF of a picture file), and allows non-live record submission (e.g. during digital bioblitzes, or after the field work when recorders have more time to upload images).

6) A "**comment" field** is useful to give more information about the observation (already implemented).

4.1.3 Expert validation

After successful submission of the observation to the platform, the post-submission data quality assurance includes a first, pre-validation step of the generated report. The report is then sent to a group of experts for further validation.

1) Apply the **pre-validation** step. This step can remove inappropriate language included in the comments or inappropriate images reported through the app. For this, open source tools such as Microsoft Cognitive Services Content Moderator[28] or WebPurify[29] can be used for the text.

2) Deploy **validation platform** (already implemented). A platform where validators can access the list of observations on the species group they are validating. The validators are also able to view the observations on a map. They also have the possibility to filter by observations they have validated before.

3) Implementation of a **"record cleaner"** functionality, flagging records outside known spatial and temporal boundaries (e.g. records new to a Member State, outside a known grid range, outside phenological boundaries...). The benefit is twofold: this could be flagged to recorders as feedback on their contribution or to ask for more information, and can also assist the validation process by providing a priority on records to be verified.

4) Use **automated validation**. Ideally, all reports are validated by experts. However, the number of generated reports should be manageable. If massive amounts of observations are reported, or if the group of expert validators is too small, automated validation of records will be necessary. Automated validation can be applied to all observations supported by sufficient proof: a predefined number of earlier, validated observations of the species, within a certain range and within a specific time range. This automated validation can reduce the workload of the expert group and can allow for faster processing and up-to-date data flows to support the end user.

5) Extend the validation categories with a **category "unresolved"** (unable to verify). Currently the JRC IAS app hosts 3 validation categories (submitted, pre-validated, validated; see also subchapter 3.8). The "Unresolved" category could include all observations not covered by the previous categories or not verifiable due to the (lack of) feedback received, and generally in case no decision can be taken on the validation status. In addition, the validated category could be split in 2 subcategories:

 ➢ Validated, based on **evidence** (pictures, sound, video).

 ➢ Validated, the finding is **probable**: for observations that are not photo-documented, but where all other information is comprehensive and detailed, where observers have provided feedback that makes a record highly probable and/or where the validator has indications that the observer is very experienced (e.g. based on earlier records submitted by that observer).

6) User feedback. The observer should always be notified of the expert's decision. This feedback is aimed to improve volunteer species identification skills (learning outcome, improved quality of future records) and to

---

[28] https://www.microsoft.com/cognitive-services/en-us/content-moderator
[29] https://www.webpurify.com/

enhance volunteer experience and participant retention. In this context, the use of Natural Language Generation (NLG) technology can be a valuable tool (van der Wal et al. 2016) (already implemented in the IAS Europe app).

7) Gather quantitative data on the validation process, such as the number of validated records (with or without picture) per taxon, the number of initially correct/wrong species IDs and how this evolves in time, the time between observation and the validation of records, etc. Through this process, experts can detect systematic bias in the contributions, and suggest improvements of the training materials or the species fact sheets. Additionally, estimates of the observer's skill or reliability could be calculated. This can be updated based on their history of contributions, used to weigh the value of their submitted data, and set a priority to records for validation.

4.1.4   Quality assurance upon data use

1) Adopt an **open access** approach and advocate for **data interoperability** (already implemented). Reducing the barriers to data sharing will significantly improve our ability to react as quickly as possible to the challenges of biological invasions (Groom et al. 2015). The published data should follow the FAIR principles: Findable, Accessible, Interoperable and Reusable (Wilkinson et al. 2016). Validated records from the JRC IAS app are published in EASIN, which follows international standards and vocabularies allowing for interoperability.

2) **Strengthen the dataflows** of validated records with other apps or CS projects that foster open data. Besides EASIN, the subset of validated or non-sensitive data could be sent to the Global Biodiversity Information Facility (GBIF). This way, data can be harvested for scientific applications and by other recording platforms to improve maps and visualizations. This could also provide a first step to allow for data quality checks upon the use of data for analysis, allowing to capture feedback from end-users after the reporting process. From the perspective of an app for regulated IAS, especially, capturing the use and feedback on data use by public authorities would be valuable.

4.2   Recommendations for long-term engagement and an increased user base of the JRC app

1) Strengthen the user community by developing dedicated communication and training materials, training courses, video materials, webinars, e-learning modules, MOOC, social media etc. The user manual for the app is comprehensive but live training of potential contributors is probably more engaging. Such a program could follow a train-the-trainer approach and could target a group of Invasive Alien Species in Europe *stewards* (e.g. regular contributors) acting as project ambassadors for the app. This creates a user community around the IAS in Europe app and requires investment in communication, app promotion and feedback to users. This is in part already done through EASIN website and social media.

2) Strengthen the peer validation aspect in the app to empower users to also contribute to the quality control by allowing users to take on a validation role (e.g. reacting to other observations, flagging records to expert validators…) on submitted observations. Creating a community of peer validators could create additional engagement for the application.

3) Create a data layer for augmented reality (cf. observations.org). Since smartphones have a compass and accelerometer, they know the direction and angle in which they are travelling. This allows an observation to be projected onto a live image of the surroundings, for example using an app like Layar. If you point your mobile to the landscape, you then get to an icon at the place where a sighting has been reported. This potentially could broaden the group of people that use the data from the app to an audience outside biological recording.

4) Improve the possibilities for communication between experts and observers (on the project website) and between observers (comment moderation). This could be achieved by linking forum possibilities or social media integration but would however require active moderation of these forums/social media.

5) Consider outlining and implementing a quality assurance mechanism to capture feedback from end-users that use the data for scientific purposes ("found any issues with this dataset or its metadata, report them here"). This implies data are freely downloadable and findable (FAIR principles) with appropriate metadata. Additionally, upon download of the data, it would be interesting to collect information on the intended use (management, research, other…).

6) Consider additional functionalities in the sighting map, for instance to visualize protected areas on the map (this is already possible on the EU CS platform geoweb application).

7) Implement a process to liaise with the validation community (cf. validation days) to collectively discuss progress with data verification, technical updates and issues, and potential improvements. This could build on the EASIN Editorial Board expertise (Tsiamis et al. 2017).

5   Conclusions

The IAS in Europe app has had a role in steering discussion on the integration of CS data in the implementation of the IAS Regulation through the EASIN. Importantly, it provides a tool to stimulate citizen reporting of IAS in areas were no such applications are available and appeals through its immediate dataflow to EASIN. It supports data collection in multiple Member States. In the last 2.5 years, the JRC has worked on decentralizing the app with regards to its dataflow and validation flow, to reduce the costs and dependency from the JRC. The assessment of options on the medium term sustainability of the app will be addressed in the last trimester of 2021, and it will be one outputs of the JRC proof of concept "CS data supporting IAS policy in Europe".

In conclusion, based on the review of selected CS projects, a refined validation procedure for the JRC IAS app is proposed, taking into account the specific taxonomic scope (IAS of Union concern) and the applicability of methods and tools that are used in similar portals and projects. This refined validation procedure can be depicted in a schematic workflow given in Figure 5. This suggested procedure is expected to increase the quality of the data of the JRC IAS app, to improve the feedback to volunteer recorders, to assist data validators in performing checks and to increase the use of the smartphone application. In addition, we hope the proposed validation procedure and recommendations can find wider application in the development of similar apps dealing with alien species and biodiversity data.

Figure 5. Schematic overview of the suggested workflow to ensure high quality CS data submitted via the JRC IAS app.
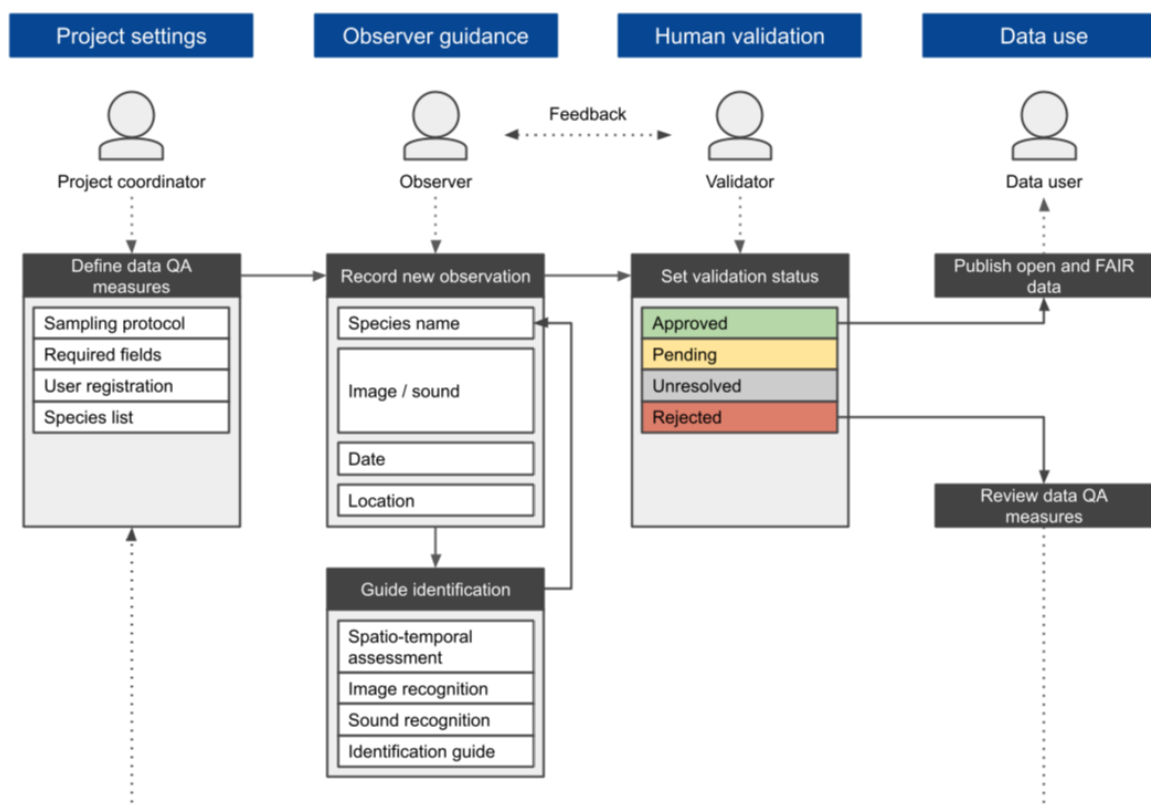
References

Adriaens, T., Sutton-Croft, M., Owen, K., Brosens, D., van Valkenburg J., Kilbey D., Groom, Q., Ehmig, C., Thürkow, F., Van Hende, P., Schneider, K., 'Trying to engage the crowd in recording invasive alien species in Europe: experiences from two smartphone applications in northwest Europe', *Management of Biological Invasions,* Vol. 6, 2015, pp. 215–225.

Altwegg, R., Nichols, J. D., 'Occupancy models for citizen-science data', *Methods in Ecology and Evolution*, Vol. 10, No. 1, 2019, pp. 8–21, https://doi.org/10.1111/2041-210X.13090

Androutsopoulou, A., Karacapilidis, N., Loukis, E., Charalabidis, Y., 'Transforming the communication between citizens and government through AI-guided chatbots', *Government Information Quarterly*, Vol. 36, 2019, pp. 358-367.

Balázs, B., Mooney, P., Nováková, E., Bastin, L., Arsanjani, J.J., 'Chapter 8: Data Quality in Citizen Science', In: K. Vohland et al. (eds.), The Science of Citizen Science, 2021, https://doi.org/10.1007/978-3-030-58278-4.

Bamford, S.P., Nichol, R.C., Baldry, I.K., et al., 'Galaxy Zoo: The dependence of morphology and colour on environment', *Monthly Notices of the Royal Astronomical Society,* Vol. 393, No. 4, 2009, pp. 1324-1352, https://doi.org/10.1111/j.1365-2966.2008.14252.x.

Bela, G, Peltola, T, Young, JC, Balazs, B, Arpin, I, Pataki, G, Hauck, J, Kelemen, E, Kopperoinen, L, Van Herzele, A, Keune, H, Hecker, S, Suskevics, M, Roy, HE, Itkonen, P, Kulvik, M, Laszlo, M, Basnou, C, Pino, J, Bonn, A, 'Learning and the transformative potential of citizen science', *Conservation Biology*, Vol. 30, 2016, pp. 990-999.

Blake, S., Siddharthan, A., Nguyen, H., et al., 'Natural language generation for nature conservation: automating feedback to help volunteers identify bumblebee species', *Proceedings of the 24th International Conference on Computational Linguistics International Committee on Computational Linguistics*, 2012, pp. 311-324.

Bubnicki, J.W., Churski, M., Kuijper, D.P., 'Trapper: an open source web-based application to manage camera trapping projects', *Methods in Ecology and Evolution*, Vol. 7, 2016, pp. 1209-1216.

Cardoso A, Tsiamis K, Gervasini E, Schade S, Taucer F, Adriaens T, Copas K, Flevaris S, Galiay P, Jennings E, Josefsson M, López B, Magan J, Marchante E, Montani E, Roy H, von Schomberg R, See L, Quintas M, 'Citizen Science and Open Data: a model for Invasive Alien Species in Europe', *Research Ideas and Outcomes,* Vol. 3, 2017, e14811, https://doi.org/10.3897/rio.3.e14811

Caffrey, JM, Baars, JR, Barbour, JH, Boets, P, Boon, P, Davenport, K, Dick, JTA, Early, J, Edsman, L, Gallagher, C, Gross, J, Heinimaa, P, Horrill, C, Hudin, S and MacIssac, HJ., 'Tackling invasive alien species in Europe: The top 20 issues', *Management of Biological Invasions*, 5, Vol. 2014, pp. 1–20, https://doi.org/10.3391/mbi.2014.5.1.01

Christin, S., Hervet, É., & Lecomte, N., Applications for deep learning in ecology. Methods in Ecology and Evolution, Vol. 10, 2019, pp. 1632-1644.

Crall, A.W., Newman, G.J, Jarnevich, C.S., Stohlgren, T.J., Waller, D.M., JGraham, J., 'Improving and integrating data on invasive species collected by citizen scientists', *Biological Invasions,* Vol. 12, 2010, pp. 3419-3428.

Crall, AW, Renz, M, Panke, BJ, Newman, GJ, Chapin, C, Graham, J and Bargeron, C., 'Developing cost-effective early detection networks for regional invasions', *Biological Invasions*, Vol. 14, 2012, pp. 2461–2469, https://doi.org/10.1007/s10530-012-0256-3

Crall, A.W., Newman, G.J., Stohlgren, T.J., Holfelder, K.A., Graham, J., Waller, D.M., 'Assessing citizen science data quality: an invasive species case study', *Conservation Letters,* Vol. 4, 2011, pp. 433-442.

Crall, A.W., Jarnevich, C.J., Young, N.E., Panke, B.J., Renz, M., Stohlgren, T.J., 'Citizen science contributes to our knowledge of invasive plant species distributions', *Biological Invasions*, Vol. 17, 2015, pp. 2415.

Cruickshank, S.S., Bühler, C., Schmidt, B.R., 'Quantifying data quality in a citizen science monitoring program: False negatives, false positives and occupancy trends', *Conservation Science and Practice*, Vol. 1, 2019, e54.

De Groot, M., Kus Veenvliet, J., Ogris, N., et al., 'Towards a better early detection and rapid response system against invasive alien species in forests', *Management of Biological Invasions,* Vol. 11, 2020, pp. 633-636, https://doi.org/10.3391/mbi.2020.11.4.06

Delipetrev, B., Schade, S., Mitton, I., Spinelli, F. 'Automated invasive alien species recognition: lessons learned from applying iNaturalist 2017 computer vision model on citizen-science data', Adjunct proceedings of the 34th EnviroInfo conference, Nicosia, Cyprus, September 23-24, 2020.

Essl, F., Bacher, S., Blackburn, T., Booy, O., Brundu, G., Brunel, S., Cardoso, A.C., Eschen, R., Gallardo, B., Galil, B., García-Berthou, E., Genovesi, P., Groom, Q., Harrower, C., Hulme, P.E., Katsanevakis, S., Kenis, M., Kühn, .I, Kumschick, S., Martinou, A.F., Nentwig, W., O'Flynn, C., Pagad, S., Pergl, J., Pyšek, P., Rabitsch, W., Richardson, D.M., Roques, A., Roy, H.E., Scalera, R., Schindler, S., Seebens, H., Vanderhoeven, S., Vilà, M., Wilson, J.R.U., Zenetos, A., Jeschke, J.M., 'Crossing frontiers in tackling pathways of biological invasions', *BioScience*, Vol. 65, No. 8, 2015, pp. 769–782.

EU. "Regulation (EU) No 1143/2014 of the European Parliament and of the Council on the prevention and management of the introduction and spread of invasive alien species". Official Journal of the European Union L315 (2014); 35–55.

Forrester, T., O'Brien, T., Fegraus, E., Jansen P.A., Palmer, J., Kays, R., Ahumada, J., Stern, B., McShea, W., 'An Open Standard for Camera Trap Data', *Biodiversity data journal*, 2016.

Gardiner, M.M., Allee, L.L., Brown, P.M., Losey, J.E., Roy, H.E., Smyth, R.R., 'Lessons from lady beetles: Accuracy of monitoring data from US and UK citizen-science programs', *Frontiers in Ecology and the Environment*, Vol. 10, No. 9, 2012, pp. 471–476, https://doi.org/10.1890/110185

Geoghegan, H., Dyke, A., Pateman, R., et al., *Understanding motivations for citizen science*, Final report on behalf of UKEOF, University of Reading, Stockholm Environment Institute (University of York) and University of the West of England, 2016.

Groom, Q., Strubbe, D., Adriaens, T., Davis, A.J.S., Desmet, P., Oldoni, D., Reyserhove, L., Roy, H.E. and Vanderhoeven, S. Empowering Citizens to Inform Decision-Making as a Way Forward to Support Invasive Alien Species Policy. *Citizen Science: Theory and Practice*, Vol. 4, 2019, pp.1-11, http://doi.org/10.5334/cstp.238

Groom, Q., Desmet P., Vanderhoeven S., Adriaens T., 'The importance of open data for invasive alien species research, policy and management', *Management of Biological Invasions,* Vol. 6, No. 2, 2015, pp. 119-125, https://doi.org/10.3391/mbi.2015.6.2.02

Groom, Q., Weatherdon, L., Geijzendorffer, I.R., 'Is citizen science an open science in the case of biodiversity observations?', *Journal of Applied Ecology,* Vol. 54, 2017, pp. 612-617.

Groom Q., Adriaens T., Bertolino S., Phelps K., Poelen J.H., Reeder D.M., Richardson D.M., Simmons N.B., Upham N. 'Holistic understanding of contemporary ecosystems requires integration of data on domesticated, captive and cultivated organisms', *Biodiversity Data Journal*, Vol. 9, 2021, e65371, https://doi.org/10.3897/BDJ.9.e65371

Haklay, M., *Citizen science and volunteered geographic information: Overview and typology of participation*, In: Crowdsourcing geographic knowledge, Springer, 2013, pp. 105-122.

IPBES. Global assessment report on biodiversity and ecosystem services, summary for policy-makers. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, 2019.

Ivan, J.S., Newkirk, E.S., 'CPW Photo Warehouse: a custom database to facilitate archiving, identifying, summarizing and managing photo data collected from camera traps', *Methods in Ecology and Evolution*, Vol. 7, 2016, pp. 499-504.

James, T., *Improving wildlife data quality: guidance on data verification, validation and their application in biological recording*, Guidance manual, 2006, http://www.nbis.org.uk/sites/default/files/documents/Improving%20wildlife%20data%20quality.pdf

Johnson, B.A., Mader, A.D., Dasgupta, R. and Kumar, P., 'Citizen science and invasive alien species: An analysis of citizen science initiatives using information and communications technology (ICT) to collect invasive alien species observations', *Global Ecology and Conservation*, Vol. 21, 2020, p.e00812.

Katsanevakis, S., Genovesi, P., Gaiji, S., Hvid, H.N., Roy, H., Nunes, A.L., Aguado, F.S., Bogucarskis, K., Debusscher, B., Deriu, I., 'Implementing the European policies for alien species: networking, science, and partnership in a complex environment', *Management of Biological Invasions*, Vol. 4, 2013, pp. 3-6.

Katsanevakis, S., Deriu, I., D'Amico, F., Nunes, A.L., Pelaez Sanchez, S., Crocetta, F., Arianoutsou, M., Bazos, I., Christopoulou, A., Curto, G., et al., 'European alien species information network (EASIN): supporting European policies and scientific research', *Management of Biological Invasions*, Vol. 6, 2015, pp. 147-157.

Kelling, S., Johnston, A., Hochachka, W.M. et al., 'Can observation skills of citizen scientists be estimated using species accumulation curves?', *PLoS One*, Vol. 10, No. 10, 2015, e0139600, https://doi.org/10.1371/journal.pone.0139600.

Klopfenstein, L.C., Delpriori, S., Malatini, S., Bogliolo, A., 'The rise of Bots: A survey of Conversational Interfaces, patterns, and Paradigms', *Proceedings of the 2017 Conference on Designing Interactive Systems*, 2017, pp. 555–565. New York, NY, USA, ACM, https://doi.org/10.1145/3064663.3064672.

Loiselle, S., Thornhill, I., Bailey, N., 'Citizen science: advantages of shallow versus deep participation', *Front. Environ. Sci. Conference Abstract: Austrian Citizen Science Conference*, 2016, doi: 10.3389/conf.FENVS.2016.01.00001

Marchante, H., Morais, M.C., Gamela, A., Marchante, E., 'Using a WebMapping platform to engage volunteers to collect data on invasive plants distribution', *Transactions in GIS*, Vol. 21, 2017, pp. 238-252, https://doi.org/10.1111/tgis.12198

Marchante H, Morais M C, Marchante E, 'Sightings Map of Invasive Plants in Portugal. Version 2.10'. CFE - Centre for Functional Ecology, Department of Life Sciences, University of Coimbra. Occurrence dataset, 2021, https://doi.org/10.15468/ic8tid accessed via GBIF.org on 2021-07-19.

Martellos, S., Moro, A., *A community-driven protocol for data validation*, Action B6, Deliverable n. 1, 2017.

Munzinger, S., Ott, J., Schulemann-Maier, G., 'Citizen-Science-Beobachtungsdaten. Teil 2: Theorie der Plausibilisierung', *Naturschutz und Landschaftsplanung*, Vol. 49, 2017a, pp. 229-235.

Munzinger, S., Ott, V.J., Schulemann-Maier, G., Strub, O., 'Citizen-Science-Beobachtungsdaten, Teil 1: Eigenschaften und Fehlerquellen', *Naturschutz und Landschaftsplanung,* Vol. 49, 2017b, pp. 5-10.

Ott, V.J., Munzinger, S., 'Aussagekraft von Daten Kennwerten aus Citizen-Science-Beobachtungsdaten', *Naturschutz und Landschaftsplanung*, Vol. 49, 2017, pp. 325-333.

Poola, I., 'Making Artificial Intelligence (AI) and Disrupted Business Intelligence (BI) truly Conversational with Humanity Touch, Automated Descriptions and talking Bots', *International Journal of Advance Research, Ideas and Innovations in Technology*, Vol. 3, No. 5, 2017, pp. 573-577.

Price-Jones, *Comparative analysis of European citizen science projects on alien species: Implications for international cooperation and optimized engagement*, Msc Thesis, Anglia Ruskin University, 2020.

Pusceddu, M., Floris, I., Mannu, R., Cocco, A., Satta, A., 'Using verified citizen science as a tool for monitoring the European hornet (*Vespa crabro*) in the island of Sardinia (Italy)', *NeoBiota*, Vol. 50, 2019, pp. 97–108, https://doi.org/10.3897/neobiota.50.37587

Roques, A., Auger-Rozenberg, M.A., Blackburn, T.M., Garnas, J.R., Pyšek, P., Rabitsch, W., Richardson, D.M., Wingfield, M.J., Liebhold, A.M., Duncan, R.P., 'Temporal and interspecific variation in rates of spread for insect species invading Europe during the last 200 years', *Biological Invasions,* Vol. 18, No. 4, 2016, pp. 907-920, doi:10.1007/s10530-016-1080-y.

Rowley, J.J.L., Callaghan, C.T., CutajaR, T., 'FrogID: Citizen scientists provide validated biodiversity data on frogs of Australia', *Herpetological Conservation and Biology*, Vol. 14, No. 1, 2019, p. 16.

Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C., Robinson, L.D., *Understanding Citizen Science & Environmental Monitoring*, Final Report on behalf of UK-EOF, NERC Centre for Ecology & Hydrology and Natural History Museum, 2012.

Schade, S., Díaz, L., Ostermann, F., Spinsanti, L., Luraschi, G., Cox, S., Nuñez, M., De Longueville, B., 'Citizen-based sensing of crisis events: sensor web enablement for volunteered geographic information', *Applied Geomatics*, 2013, pp. 1-16.

Schade, S., Kotsev, A., Cardoso, A.C., Tsiamis, K., Gervasini, E., *et al.,* 'Aliens in Europe. An open approach to involve more people in invasive species detection', *Computers, Environment and Urban Systems,* Vol. 78, 2019, 101384.

Schermer, M., Hogeweg, L., 'Supporting citizen scientists with automatic species identification using deep learning image recognition models', *Biodiversity Information Science and Standards,* Vol. 2, 2018, e25268, https://doi.org/10.3897/biss.2.25268

Schoonvaere, K., Laget, D., Adriaens, T., et al., *Vespa-Watch: Invasiemonitoring van de Aziatische hoornaar met hobbyimkers en het publiek*, Eindrapport in het kader van de oproep Citizen Science van het Departement

Economie, Wetenschap en Innovatie (EWI) van december 2017. Honeybee Valley en Instituut voor Natuur- en Bosonderzoek, Brussels (in Dutch), 2020, https://doi.org/doi.org/10.21436/inbor.19019045

Seebens, H., Blackburn, T.M., Dyer, E.E., Genovesi, P., Hulme, P.E., Jeschke, J.M., Pagad, S., Pyšek, P., Winter, M., Arianoutsou, M., Bacher, S., Blasius, B., Brundu, G., Capinha, C., Celesti-Grapow, L., Dawson, W., Dullinger, S., Fuentes, N., Jäger, H., Kartesz, J., Kenis, M., Kreft, H., Kühn, I., Lenzner, B., Liebhold, A., Mosena, A., Moser, D., Nishino, M., Pearman, D., Pergl, J., Rabitsch, W., Rojas-Sandoval, J., Roques, A., Rorke, S., Rossinelli, S., Roy, H.E., Scalera, R., Schindler, S., Štajerová, K., Tokarska-Guzik, B., van Kleunen, M., Walker, K., Weigelt, P., Yamanaka, T. & Essl, F. 'No saturation in the accumulation of alien species worldwide', *Nature Communications*, Vol. 8, 2017, 14435.

Swinnen, K., Vercayie, D., Vanreusel, W., Barendse, R., Boers, K., Bogaert, J., Dekeukeleire, D., Driessens, G., Dupriez, P., Jooris, R., 'Waarnemingen. be–Non-native plant and animal occurrences in Flanders and the Brussels Capital Region, Belgium', *BioInvasions Records*, Vol. 7, 2018, pp. 335-342.

Teacher, A.G., Griffiths, D.J., Hodgson, D.J., Inger, R., 'Smartphones in ecology and evolution: a guide for the apprehensive', *Ecology and evolution*, Vol. 3, 2013, pp. 5268-5278.

Thiel, M., Penna-Díaz, M.A., Luna-Jorquera, G., Salas, S., Sellanes, J., Stotz, W., 'Citizen scientists and marine research: volunteer participants, their contributions, and projection for the future', *Oceanography and Marine Biology: An Annual Review*, Vol. 52, 2014, pp. 257-314.

Tsiamis, K., Gervasini, E., D'Amico, F., Deriu, I., Katsanevakis, S., Crocetta, F., Zenetos, A., Arianoutsou, M., Backeljau, T., Bariche, M., et al., 'The EASIN editorial board: Quality assurance, exchange and sharing of alien species information in Europe', *Management of Biological Invasions*, Vol. 7, 2016, pp. 321–328.

Tsiamis, K., Gervasini, E., D'Amico, F., Deriu, I., Roglia, E., Shade, S., Craglia, M., Cardoso, A.C. Citizen Science Application, Invasive Alien Species in Europe, EUR 28441 EN, 2017, doi:10.2760/043856

Ueda K, iNaturalist Research-grade Observations. iNaturalist.org. Occurrence dataset, 2021, https://doi.org/10.15468/ab3s5x accessed via GBIF.org on 2021-07-19.

van der Wal, R., Sharma, N., Mellish, C. et al., 'The role of automated feedback in training and retaining biological recorders for citizen science', *Conservation Biology*, Vol. 30, 2016, pp. 550-561, https://doi.org/10.1111/cobi.12705

van Strien, A.J., van Swaay, C.A.M., Termaat, T., 'Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models', *Journal of Applied Ecology*, Vol. 50, 2013, pp. 1450–1458, https://doi.org/10.1111/1365-2664.12158

Vanreusel W, Swinnen K, Gielen K, Vercayie D, Driessens G, Veraghtert W, Desmet P, Herremans M, Waarnemingen.be - Non-native animal occurrences in Flanders and the Brussels Capital Region, Belgium. Version 1.31. Natuurpunt. 2021, Occurrence dataset https://doi.org/10.15468/k2aiak accessed via GBIF.org on 2021-07-19.

Vanreusel W, Barendse R, Steeman R, Gielen K, Swinnen K, Desmet P, Herremans M, Waarnemingen.be - Non-native plant occurrences in Flanders and the Brussels Capital Region, Belgium. Version 1.30. 2021, Natuurpunt. Occurrence dataset https://doi.org/10.15468/smdvdo accessed via GBIF.org on 2021-07-19.

Wäldchen, J., Mäder, P., 'Machine learning for image based species identification', *Methods in Ecology and Evolution*, Vol. 9, 2018, pp. 2216-2225.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, Vol. 3, 2016, 160018, https://doi.org/10.1038/sdata.2016.18

Zsarnoczky, M., 'How does Artificial Intelligence Affect the Tourism Industry?', *Vadyba Journal of Management*, Vol., 31, No. 2, 2017, pp. 85–90.

List of abbreviations and definitions

AI   Artificial Intelligence

Alien CSI   Increasing understanding of alien species through citizen science, CA17122

API   Application Programming Interface

APPS   Smartphone applications

CS   Citizen Science

EASIN   European Alien Species Information Network

EU   European Union

IAS   Invasive Alien Species

IAS Reg.   EU Regulation no. 1143/2014 on Invasive Alien Species

JRC   Joint Research Centre

NLG   Natural Language Generation

List of figures

List of tables

The European Commission's science and knowledge service
Joint Research Centre

JRC Mission
As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub – Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office
of the European Union