Research Ideas and Outcomes 10: e125523
doi: 10.3897/rio.10.e125523

Forum Paper

Prototype Biodiversity Digital Twin: Real-time bird monitoring with citizen-science data

Otso Ovaskainen[‡,§], Patrik Lauha[‡], Julian Lopez Gordillo[|], Ossi Nokelainen[¶,§], Anis U. Rahman[§], Allan T. Souza[#], Jussi Talaskivi[¤], Gleb Tikhonov[‡], Aurélie Vancraeyenest[«], Ari Lehtiö[¤]

‡ Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, P.O. Box 65, FI-00014 University of Helsinki, Helsinki, Finland
§ Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Jyväskylä, Finland
| Naturalis Biodiversity Center, P.O. Box 9517 2300 RA, Leiden, Netherlands
¶ Open Science Centre, University of Jyväskylä, P.O. Box 35, 40014, Jyväskylä, Finland
# Institute for Atmospheric and Earth System Research INAR, Forest Sciences, Faculty of Agriculture and Forestry, P.O. Box 27, 00014 University of Helsinki, Helsinki, Finland
¤ Digital Services, University of Jyväskylä, P.O. Box 35, 40014, Jyväskylä, Finland
« CSC – IT Center for Science Ltd., P.O. Box 405, 02101, Espoo, Finland

Corresponding author: Otso Ovaskainen (otso.t.ovaskainen@jyu.fi)

Reviewed v 1

Academic editor: Dmitry Schigel

Received: 16 Apr 2024 | Accepted: 08 Jun 2024 | Published: 20 Jun 2024

Citation: Ovaskainen O, Lauha P, Lopez Gordillo J, Nokelainen O, Rahman AU, Souza AT, Talaskivi J, Tikhonov G, Vancraeyenest A, Lehtiö A (2024) Prototype Biodiversity Digital Twin: Real-time bird monitoring with citizen-science data. Research Ideas and Outcomes 10: e125523. https://doi.org/10.3897/rio.10.e125523

Abstract

Bird populations respond rapidly to environmental change making them excellent ecological indicators. Climate shifts advance migration, causing mismatches in breeding and resources. Understanding these changes is crucial to monitor the state of the environment. Citizen science offers vast potential to collect biodiversity data. We outline a project that combines citizen science with AI-based bird sound classification. The mobile app records bird vocalisations that are classified by AI and stored for re-analysis. Additionally, it shows a shared observation board that visualises collective classifications. By merging long-term monitoring and modern citizen science, this project harnesses the strength of both approaches for comprehensive bird population monitoring.

Keywords

citizen science, bird monitoring, acoustic monitoring, artificial intelligence, species distribution modelling

Introduction

Bird populations are showing rapid and alarming responses to environmental change. One highly conspicuous phenomenon is that of bird migration, in particular the arrival of migratory birds to Europe during spring. Due to climate change, these migratory events are rapidly shifting to earlier, creating ecological mismatches, for example, between the timing of breeding and resource availability. The ongoing rapid changes in bird populations make it increasingly relevant to better understand the mechanisms driving such changes and to continuously monitor the fate of bird populations (Burns et al. 2021).

A great number of people are interested in birds and citizen science has a long history in bird research. While citizen-science projects have provided huge amounts of valuable biodiversity data, a high proportion of the data provided by citizen-science projects suffers from common fundamental limitations. One such limitation is variation in the skills of the observers in species identification, leading to high rates of both false positives (a citizen claims to have observed an species that was not there in reality) and false negatives (an observer failed to report a species that was there in reality). Another such limitation is spatiotemporal bias in observation effort, as citizen-science projects are typically not based on systematic or randomised sampling schemes, but rather on opportunistic sampling. As variation in observer skills and bias in sampling effort can be difficult to quantify and report in the metadata, their effects are often difficult to correct for while using the data to scientific inference, potentially leading to biased inference. Despite these limitations, citizen science has great potential, as it can produce much larger datasets than data acquired by professional researchers (Vohland et al. 2021).

This project aims to combine long-term bird monitoring programmes with citizen science to make the best out of the two worlds. To avoid some of the common pitfalls of citizen-science projects, the data are not based on the identifications made by citizens, but by a new mobile phone application MK (acronym of the Finnish name of the application ”Muuttolintujen kevät”, meaning Spring of Migratory Birds) that we developed for the purpose of this project. The phone application can be downloaded from the Google Play Store (Muuttolintujen kevät 2024b) and the App Store (Muuttolintujen kevät 2024a). The birds’ vocalisations in the audio recorded by the phone app are identified and classified by an AI-based backend (Lauha et al. 2022), removing variation in observer skills in species identification. The phone application returns to the user the most likely species classifications together with their classification probabilities, enabling the user to evaluate their reliability. As the application submits not only the classifications, but also the raw audio files to the server where it is stored, the data can be re-analysed with progressively improving classification models. The size of an audio file depends on the length of the recording, median size being 240.6 kB and 2.5% - 97.5% quantiles 40.0 kB and 2.465 MB,

respectively. The application was launched in spring 2023, attracting 140,000 users who submitted 3 million recordings during 2023. While data from 2023 were acquired with an opportunistic recording scheme, in spring 2024, we published a new version that enables citizen scientists to submit also standardised point counts in preselected locations.

The phone application implements a common observation board where the classifications obtained collectively by all users can be visualised. A key aim of the project, which is still to be implemented, is to use the citizen-science observations to generate continuously updating predictions of bird spatiotemporal distributions and singing activity.

Objective

The objective of this Biodiversity Digital Twin prototype is to investigate if and how citizen science can be employed to real-time bird monitoring, in a way that produces robust data also for scientific analyses. To achieve this, we aim to make the data compatible with existing long-term data on birds by implementing a point count module and generating calibration data by conducting point counts simultaneously by bird experts and by the phone application. We aim to develop an internet portal that shows data and predictions with minimal delay compared to the real-world system, delivering a proof-of-concept of a real-time digital twin of biodiversity. A further important objective of this project is to increase the public awareness of science on nature and the ongoing environmental change.

Workflow

The overall workflow of this prototype digital twin is illustrated in Fig. 1 Citizen scientists record bird vocalisations with a mobile phone application. The audio is sent to a centralised server that runs a CNN model to classify the birds and returns the classifications that are shown in the mobile application. The classifications are compared to prior predictions, based on long-term bird monitoring, the results of which comparison is used to update predictions on bird spatiotemporal distributions and singing activity. The workflow of the overall modelling approach is illustrated in Fig. 2 and the workflow for generating prior distribution is illustrated in Fig. 3.

Data

As illustrated by the yellow boxes in Figs 2, 3, the project combines the following types of data:

- Data recorded by citizen scientists by the mobile phone application MK. The raw audio data consist of .wav files and the metadata contain information about the user (anonymised), date, time, duration, latitude and longitude. The classifications made by AI methods describe for each recording the species classified from the recordings and the reliability of the classifications in units of probability;

- Weather and climatic data were downloaded from Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.f17050d7 (Copernicus Climate Change Service 2019);
- Land-cover data derived from CORINE (European Environment Agency 2019a, European Environment Agency 2019b, European Environment Agency 2019c);
- Transect line counts of birds obtained through collaboration with the Finnish bird monitoring programme (Suomen Lajitietokeskus 2024);
- Earlier citizen-science data on Finnish birds derived from laji.FI (Suomen Lajitietokeskus 2024);
- Systematic audio recordings of birds made by the ERC-synergy project LIFEPLAN (Lifeplan 2024).

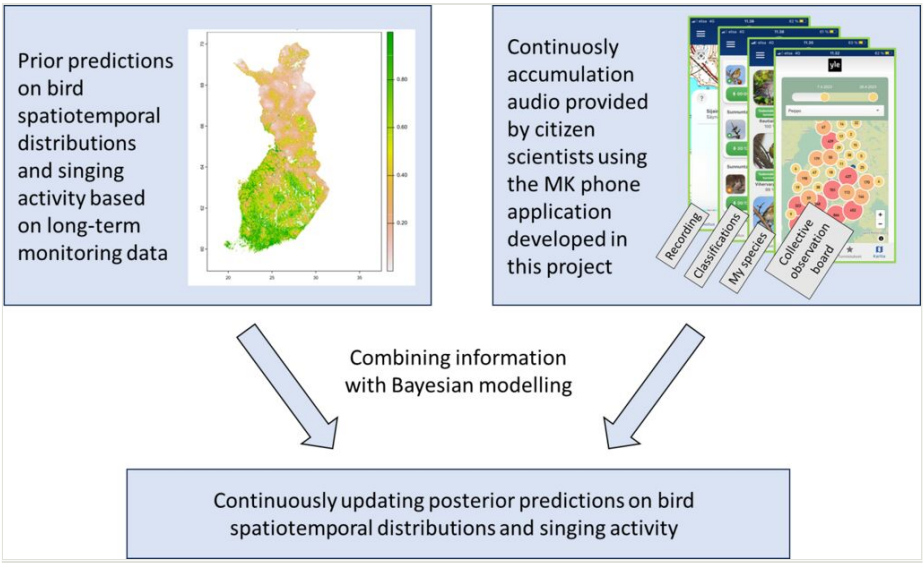

Figure 1. doi

A conceptual diagram of the digital twin prototype. The core aim of this project is to test the feasibility of generating essentially real-time updating predictions on bird spatiotemporal distributions and singing activity by combining prior information, based on long-term monitoring data with continuously accumulating new information provided by citizen scientists.

Model

The overall modelling strategy for combining prior predictions with the MK phone application data and weather predictions is illustrated in Fig. 2. This involves fitting the joint species distribution model HMSC (hierarchical model of species communities) at daily intervals to the continuously accumulating audio data submitted from citizen scientists through the mobile phone application MK. The HMSC model considers the species classifications from the phone application data as the response vector and incorporates prior predictions as an offset, estimating spatial and temporal latent factors. These factors

signify locations and times where bird occurrences deviate from predictions by the prior model. The latent factors are then used to update prior predictions into posterior predictions of current spatiotemporal distributions of birds, which are further multiplied by the vocalisation activity predictions informed by weather forecasts to generate predictions of singing activity. Additionally, the modelling strategy for constructing the prior predictions is illustrated in Fig. 3. It involves a workflow implemented as a combination of R and Python/TensorFlow scripts. The prior predictions are obtained as a product of three probabilities, which model:

1. how common the species is in the spatial location in question during its summer breeding distribution;
2. whether the species is currently in the summer breeding distribution or its overwintering area (relevant only for migratory species);
3. what the vocalisation activity of the species is, given the season, time of the day and weather conditions.

The HMSC model is used in modelling line transect bird count data as a function of environmental (land-use, climate and forest structure) and spatial (latent factors) predictors, used to predict the distribution of Finnish birds at 1 hectare resolution, covering over 30 million grid cells.

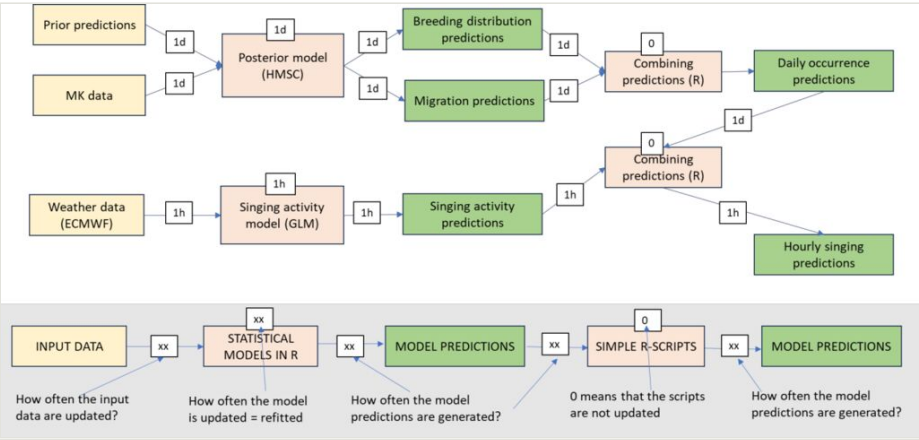

Figure 2. doi

An overview of the modelling strategy for combining prior predictions with the MK phone application data to provide continuous updating predictions of bird distributions and their singing activity. The graphical legend on the bottom of the figure explains the colours and symbols used.

FAIRness

To facilitate the reusability of the data used in this pDT, we will follow the FAIR principles (Wilkinson et al. 2016) by releasing data to relevant open repositories with assigned

persistent identifiers (PIDs) and descriptive metadata. PID systems, such as the widely used Digital Object Identifier (DOI), ensure reliable referencing, boosting discoverability and facilitating proper citation. Metadata provide a comprehensive dataset characterisation and captures contextual information that would otherwise be difficult or impossible to retrieve. Adhering to established community standards and vocabularies enables consistency and interoperability and fosters collaboration. This applies to both the choice of metadata structure, such as Research Object Crate (RO-Crate) format (Soiland-Reyes et al. 2022), as well as the data it describes. While RO-Crate inherently supports the comprehensive Schema.org vocabulary, certain domain-specific aspects cannot be easily captured through it. There are other semantic resources that can be used in conjunction to achieve greater semantic interoperability. For example, for the specific purpose of providing taxonomic information, the *dwc:taxonID* property from the Darwin Core Terms (Darwin Core Maintenance Group 2023) can be incorporated into the context of the RO-Crate metadata file. This extension mechanism can be used with other terms or resources as needed to move towards a richer metadata description.

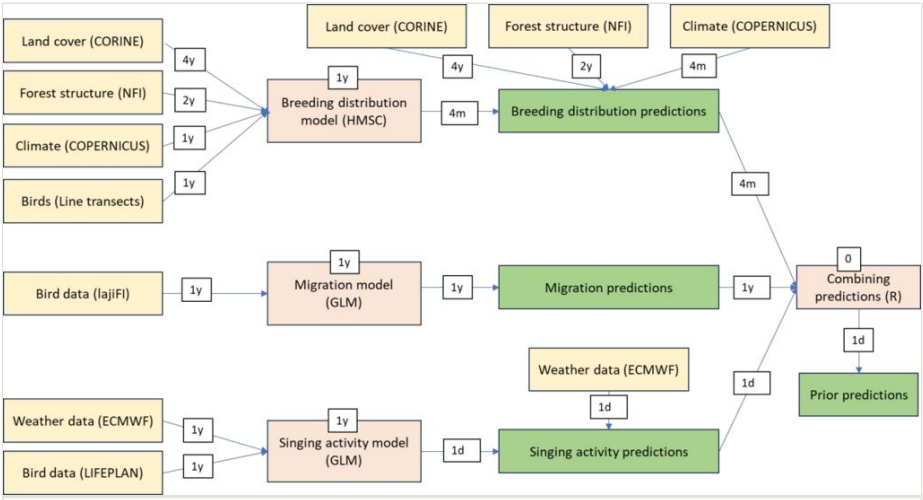

Figure 3. doi

A more detailed description of the modelling strategy used to generate the prior predictions, based on the long-term data. The colours and symbols used follow the graphical legend of Fig. 2.

The project provides open access to non-sensitive data in designated repositories, such as GBIF: The Global Biodiversity Information Facility (2024) and Zenodo (European Organization For Nuclear Research and OpenAIRE (2013). Most datasets will be openly available, while sensitive ones may be restricted. Reasons for restrictions will be stated (e.g. GDPR (2016)) and access requests will be considered ethically and legally. As a concrete example, the raw audio data are not intended to be publicly available (since they may contain sensitive data in the form of personal information and, thus, GDPR applies), but can be provided on request. Furthermore, the associated metadata (which include detailed information about the dataset as well as the derived classification) will be openly

available. Researchers can access data following repository guidelines, ensuring they can locate, retrieve and reuse data. Embargoes may restrict access, with clear terms. Open-access licensing encourages reuse with defined permissions. Quality assurance procedures maintain reliability and include validation, verification and detailed documentation throughout the data lifecycle, ensuring confident reuse.

Whenever possible, adoption of the FAIR principles will extend to other components of the pDT beyond data (e.g. models, workflows), as established by the FAIR Digital Objects (FDO) interoperability framework (De Smedt et al. 2020). The code will be made publicly available via the BioDT GitHub organisation (https://github.com/BioDT) or similar relevant repositories, such as on the space for BioDT on the WorkflowHub registry (https://workflowhub.eu/programmes/22) (Goble et al. 2023). Ultimately, these efforts seek to align with the wider European strategy on the front of FAIRness and open science, as laid out on the EOSC interoperability framework (European Commission. Directorate General for Research and Innovation and EOSC Executive Board 2021).

Performance

The HMSC model is pivotal in our modelling strategy despite its computational intensity. It is used for generating prior predictions and analysing the continuously accumulating audio data from citizen scientists via the MK mobile phone application. The latter consists of model fitting using MCMC approaches and predicting species occurrences at a 1-ha resolution over Finland. Given the daily frequency of these operations, achieving sufficient computational performance is critically important. To address the computational bottlenecks of the R-package Hmsc (Tikhonov et al. 2020), we developed Hmsc-HPC (Rahman et al. 2024), a high-performance computational module implemented in Python/ TensorFlow. Leveraging GPU computation, Hmsc-HPC accelerates model fitting up to 1000 times faster than the R-version (Rahman et al. 2024). Integration of model fitting using Hmsc-HPC on LUMI has been implemented. We have successfully executed the model fitting on LUMI for experimental scenarios, paving the way for making predictions for diverse ecological applications. Furthermore, efforts are underway to leverage the accelerated HPC approach provided by Hmsc-HPC for utilising fitted models in making predictions, thereby enhancing the performance and scalability of our methodology.

Interface and outputs

The RTBM pDT web application is conceived to facilitate interactive engagement, enabling users to interact with the pDT, running simulations and displaying the predictions on a web browser. By selecting specific bird species and spatial and temporal ranges, users can configure the model runs to suit their needs. The interface is currently in the design phase (Fig. 4). The model outputs will be displayed through updated graphs, maps and tables, providing information on the bird breeding distribution, migration and singing activity, offering insights into ecological and behavioural patterns and trends.

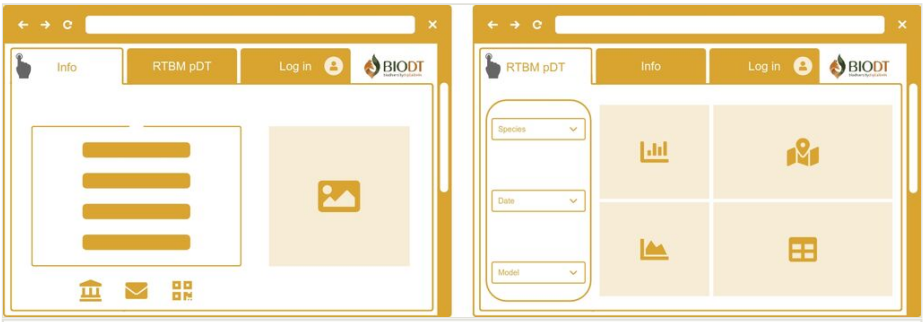

Figure 4. doi

Design of the web application where the users can interact with the RTBM pDT. The figure displays the envisioned features of the web application, including the tabs containing the information on the RTBM, pDT simulation results and user authentication. There will be a selection of inputs on the RTBM pDT tab (on the left-hand side) and a dashboard on the right-hand side of the page displaying the dynamically updated maps, graphs and tables.

Integration and sustainability

The maintenance of the project after the BioDT funding cycle is facilitated by the establishment of the Digital Citizen Science Center that will operate at least until the end of the year 2028 thanks to funding granted by the Jane and Aatos Erkko Foundation (2024) (jaes.fi/en/frontpage/). We aim to integrate the project to Destination Earth (2024) (https://digital-strategy.ec.europa.eu/en/policies/destination-earth) and European Open Science Cloud (2024) (https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud) once doing so is technically feasible, but the details on how to achieve this are still unclear.

Application and impact

Digital twin technologies (DT) have potential to revolutionise biodiversity research, impacting policy frameworks and economic structures and mechanisms related to biodiversity research and conservation. Increasing public awareness of science can inspire masses on environmental initiatives for a common cause: monitoring the state of our environment. Being able to monitor ecological communities in real time through digital technologies can transform biodiversity research. Additionally, it makes possible to scale data from local to global levels, which can facilitate information-based conservation acts faster than before. It is noteworthy that this may include implementing the technology across taxa; a premise, which requires rigorous testing before large-scale reliability could be achieved. Nevertheless, as the information of environmental impact becomes faster and easier through integrating ecological data from various databases, a new era of automated monitoring systems can hasten The European Green Deal (2024) (https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en) and facilitate more sustainable decisions in land use. Examples include ARISE (2024) (https://www.arise-biodiversity.nl/) and BioAcousticsAI (2024) (https://bioacousticai.eu/). Further,

policies promoting innovation fuel technological advancements, shaping future biodiversity research (and market dynamics), because they promote the development of solutions aligned with Biodiversity Strategy priorities. Policy-makers should enact medium and long-term strategies to integrate available DT technology effectively. Collaboration is key to leveraging DT for biodiversity research. Collaboration with stakeholders is vital to maximise benefits, ensure policy alignment and societal impact. Stakeholders include biodiversity research infrastructures, data providers, researchers, policy-makers and industrial actors. In conclusion, policy interventions must align with legislative priorities to drive innovation and achieve sustainable outcomes in the future.

Acknowledgements

We thank executive producer Ville Alijoki (Yle Science, Environment and History) for fruitful collaboration: the phone application fast received a broad user community largely thanks to the cooperation with Yle Nature and its promotion of the application in TV, radio, news articles and social media. The project was funded by the European Union: the HORIZON-INFRA-2021-TECH-01 project 101057437 (Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities, https://doi.org/10.3030/101057437) and the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement No. 856506: ERC-synergy project LIFEPLAN; and grant agreement No. 101123091: ERC-PoC project Breaking the wall between professional science and citizen science by hyperautomation) and the Jane and Aatos Erkko Foundation (grant to establish the Digital Citizen Science Centre for 2024-2028).

Conflicts of interest

The authors have declared that no competing interests exist.

References

- ARISE (2024) The ARISE project. https://www.arise-biodiversity.nl/. Accessed on: 2024-6-03.
- BioAcousticsAI (2024) BioAcousticsAI: Understanding animal sounds with machine learning. https://bioacousticai.eu/. Accessed on: 2024-6-03.
- Burns F, Eaton MA, Burfield IJ, Klvaňová A, Šilarová E, Staneva A, Gregory RD (2021) Abundance decline in the avifauna of the European Union reveals cross-continental similarities in biodiversity change. Ecology and evolution 11 (23): 16647-16660. https://doi.org/10.1002/ece3.8282
- Copernicus Climate Change Service (2019) ERA5 monthly averaged data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Release date: 2024-6-03. URL: https://cds.climate.copernicus.eu/doi/10.24381/cds.f17050d7

- Darwin Core Maintenance Group (2023) Darwin Core List of Terms. Biodiversity Information Standards (TDWG). http://rs.tdwg.org/dwc/doc/list/2023-09-18. Accessed on: 2024-5-28.
- De Smedt K, Koureas D, Wittenburg P (2020) FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications 8 (2). https://doi.org/10.3390/publications8020021
- Destination Earth (2024) Destination Earth. https://digital-strategy.ec.europa.eu/en/policies/destination-earth. Accessed on: 2024-5-28.
- European Commission. Directorate General for Research and Innovation, EOSC Executive Board (2021) EOSC interoperability framework: report from the EOSC Executive Board Working Groups FAIR and Architecture. Publications Office, LU. URL: https://data.europa.eu/doi/10.2777/620649
- European Environment Agency (2019a) CORINE Land Cover 2018 (raster 100 m), Europe, 6-yearly - version 2020_20u1, May 2020. 20.01. European Environment Agency. Release date: 2020-5-13. URL: https://sdi.eea.europa.eu/catalogue/copernicus/api/records/960998c1-1870-4e82-8051-6485205ebbac?language=all
- European Environment Agency (2019b) CORINE Land Cover 2006 (raster 100 m), Europe, 6-yearly - version 2020_20u1, May 2020. 20.01. European Environment Agency. Release date: 2020-5-13. URL: https://sdi.eea.europa.eu/catalogue/copernicus/api/records/08560441-2fd5-4eb9-bf4c-9ef16725726a?language=all
- European Environment Agency (2019c) CORINE Land Cover 2012 (raster 100 m), Europe, 6-yearly - version 2020_20u1, May 2020. 20.1. European Environment Agency. Release date: 2020-5-13. URL: https://sdi.eea.europa.eu/catalogue/copernicus/api/records/a84ae124-c5c5-4577-8e10-511bfe55cc0d?language=all
- European Open Science Cloud (2024) European Open Science Cloud (EOSC). https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud. Accessed on: 2024-5-28.
- European Organization For Nuclear Research, OpenAIRE (2013) Zenodo. CERN. https://doi.org/10.25495/7GXK-RD71
- GBIF: The Global Biodiversity Information Facility (2024) What is GBIF? https://www.gbif.org/what-is-gbif. Accessed on: 2024-5-28.
- GDPR (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). URL: http://data.europa.eu/eli/reg/2016/679/oj
- Goble C, Bacall F, Soiland-Reyes S, Owen S, Eguinoa I, Droesbeke B, Ménager H, Rodriguez-Navas L, Fernández J, Grüning B, Leo S, Pireddu L, Crusoe M, Gustafsson J, Capella-Gutierrez S, Coppens F (2023) EOSC-Life Workflow Collaboratory for the Life Sciences. Proceedings of the Conference on Research Data Infrastructure 1 https://doi.org/10.52825/cordi.v1i.352
- Jane and Aatos Erkko Foundation (2024) Jane and Aatos Erkko Foundation. jaes.fi/en/frontpage/. Accessed on: 2024-6-03.
- Lauha P, Somervuo P, Lehikoinen P, Geres L, Richter T, Seibold S, Ovaskainen O (2022) Domain-specific neural networks improve automated bird sound recognition already with small amount of local data. Methods in Ecology and Evolution 13 (12): 2799-2810. https://doi.org/10.1111/2041-210x.14003

- Lifeplan (2024) Lifeplan | University of Helsinki. www.helsinki.fi/en/projects/lifeplan. Accessed on: 2024-6-03.
- Muuttolintujen kevät (2024a) Muuttolintujen kevät (RFJM) on the App Store (apple.com). https://apps.apple.com/fi/app/muuttolintujen-kev%C3%A4t-rfjm/id1637041247. Accessed on: 2024-6-03.
- Muuttolintujen kevät (2024b) Muuttolintujen kevät – RFJM – Google Play -sovellukset. https://play.google.com/store/apps/details?id=fi.jyu.app.researchforjyumobile&hl=fi. Accessed on: 2024-6-03.
- Rahman AU, Tikhonov G, Oksanen J, Rossi T, Ovaskainen O (2024) Accelerating joint species distribution modeling with Hmsc-HPC: A 1000x faster GPU deployment. bioRxiv https://doi.org/10.1101/2024.02.13.580046
- Soiland-Reyes S, Sefton P, Crosas M, Castro LJ, Coppens F, Fernández J, Garijo D, Grüning B, La Rosa M, Leo S, Ó Carragáin E, Portier M, Trisovic A, RO-Crate Community, Groth P, Goble C (2022) Packaging research artefacts with RO-Crate. Data Science 5 (2): 97-138. https://doi.org/10.3233/ds-210053
- Suomen Lajitietokeskus (2024) Suomen Lajitietokeskus. https://laji.fi/. Accessed on: 2024-6-03.
- The European Green Deal (2024) The European Green Deal. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en. Accessed on: 2024-5-28.
- Tikhonov G, Opedal Ø, Abrego N, Lehikoinen A, de Jonge MJ, Oksanen J, Ovaskainen O (2020) Joint species distribution modelling with the r-package Hmsc. Methods in Ecology and Evolution 11 (3): 442-447. https://doi.org/10.1111/2041-210x.13345
- Vohland K, Land-Zandstra A, Ceccaroni L, Lemmens R, Perelló J, Ponti M, Samson R, Wagenknecht K (2021) Editorial: The Science of Citizen Science Evolves. The Science of Citizen Science1-12. https://doi.org/10.1007/978-3-030-58278-4_1
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1). https://doi.org/10.1038/sdata.2016.18