Prototype Digital Twin: Recreation and Biodiversity Cultural Ecosystem Services

Simon Rolph, Chris Andrews, Dylan Carbone, Julian Lopez Gordillo, Tomas Martinovic, Nick Oostervink, Dirk Pleiter, Kata Sara-aho, John Watkins, Christoph Wohner, Jan Dick

Recreation and Biodiversity Cultural Ecosystem Services: Prototype Digital Twin

Authors

Simon Rolph, Chris Andrews, Dylan Carbone, Julian Lopez Gordillo, Tomas Martinovic, Nick Oostervink, Dirk Pleiter, Kata Sara-aho, John Watkins, Christoph Wohner, Jan Dick

Abstract

Digital twin approaches have the potential to revolutionise usage, planning and management of cultural ecosystem services i.e. the non-material benefits people obtain from ecosystems, including recreation, tourism, intellectual development, spiritual enrichment, reflection, and aesthetic experiences.

Here we outline our blueprint for a prototype digital twin (pDT) for cultural ecosystem services. The pDT consists of two modelling components; a recreation potential model to quantify the cultural ecosystem services of the physical landscape and species distribution models to quantify the biodiversity component.

It is envisaged that the digital twin will be used primarily by two user types 1. those who wanted to enjoy the area and potentially contribute to citizen science programs and 2. people who want to inform or make evidence-based management decisions (land managers, policy makers, researchers).

Keywords

recreation, tourism, trade-offs, biodiversity, models

Introduction

Cultural ecosystem services refer to non-material benefits people obtain from ecosystems, including recreation, tourism, intellectual development, spiritual enrichment, reflection, and aesthetic experiences. The concept of cultural ecosystem has been less researched by the modelling community (Gould et al. 2019). Biodiversity is recognised as a key ecosystem service in all branches of ecosystem service research but of particular importance in relation to the cultural ecosystem services.

A digital twin of cultural ecosystem services may help understand impacts by providing a dynamic representation of cultural ecosystem services, allowing stakeholders to understand how changes in the environment or management practices affect these services. This may in turn serve as a tool for decision support, by enabling evaluation of potential consequences of various management strategies or policy decisions on the cultural ecosystem services related to recreation and tourism. Additionally, this digital twin could provide a focus for monitoring and assessment of cultural ecosystem services use.

Digital twins do not emerge in isolation but build on existing modelling work. Previous work has focused on modelling recreational cultural ecosystem services within a national park in Scotland. A recreational potential model was developed under a previous EU project (Zulian et al. 2018) and further updated with the additional knowledge gained through stakeholder interaction (Dick et al 2022). In this prior work the recreational model was parametrised for two personas, visitors who prefer high-adrenaline activities that require a high level of fitness (hard recreationists) and those who prefer "calmer" activities that do not require a high fitness level (soft recreationists). However, much of the literature related to the cultural ecosystem service of recreation highlights the diversity of needs of people with a wide range of physical fitness and appreciation of different aspects of nature (Orenstein et al., 2017).

Biodiversity models can be incorporated into the digital twin of recreation potential. For this use case we are motivated by a need to know where and when species are present and observable. Species distribution models (SDMs), also known as environmental niche models, are a versatile and widely used tool to deliver on this need (Zurell et al., 2020). SDMs use statistical or AI techniques to correlate species occurrence data with environmental factors. An SDM can then be used to predict where the species is likely to occur in areas where data is lacking or in the future under different environmental scenarios.

Our prototype digital twin (pDT) focuses on developing a digital twin to support the use, planning and management of cultural ecosystem services, focused on recreation. The digital twin will track changes in services by recording how people use the associated resources. To quantify the cultural ecosystem services of the physical landscape, a recreation potential model is used, while species distribution models are employed to quantify the biotic component. It is envisaged that the digital twin will be used primarily by two user types 1. those who wanted to enjoy the area and contribute to citizen science programs and 2. people who want to make or inform evidence-based management decisions (land managers, policy makers, researchers).

Objectives

The dual purpose of this prototype digital twin is to support personalised knowledge to recreationalists and tourists using a particular area based on their preferences and support the planning and management of cultural ecosystem services by tracking changes in how people

use natural resources. It aims to provide valuable insights into the interactions between biodiversity, human activities, and ecosystem services, allowing for evidence-led conservation policy, adaptive management protocols, and practical decision-making in managing recreational use and biodiversity conservation. Multiple user classes (including recreationalists, wildlife enthusiasts, citizen scientists i.e., amateur/nonprofessional researchers who record biodiversity sightings) are envisaged interacting with the pDT, obtaining knowledge, whilst feeding back biodiversity data. A second set are managing the landscape, either as policy developers and park regulators (including park authority staff or local government staff), or as practical managers (including landowners or scientists aiming to provide in depth knowledge for managers).

Workflow

The Cultural Ecosystem Services pDT comprises of two components: A recreation potential model (Dick et al 2022) and species distribution models implemented using the flexsdm R package (Velazco et al., 2022). Input data is loaded from various sources and processed by the modelling pipeline for each component. The model outputs from each modelling pipeline are transferred to a common repository which can be accessed by the user interface. The user interface overlays the model outputs in a mapping interface to allow users to compare areas of high/low recreation potential against spatial biodiversity trends.
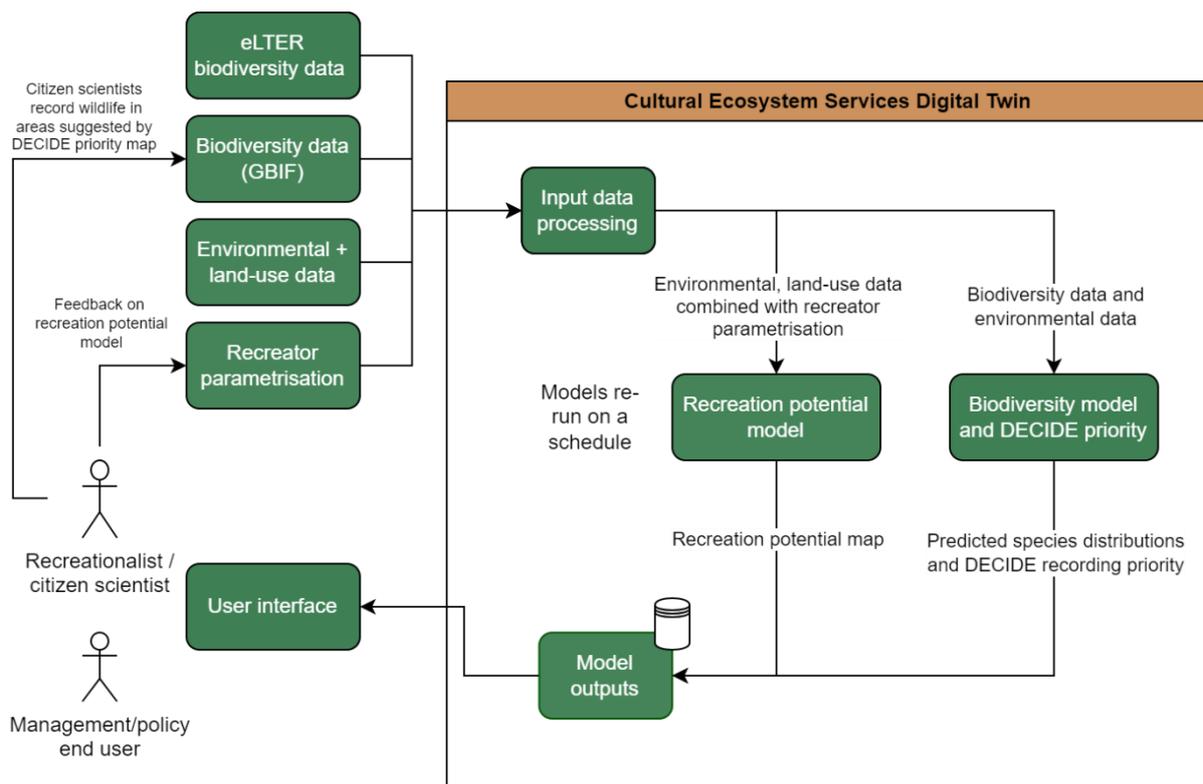

Figure 1: Conceptual schema of the Recreation and Biodiversity Cultural Ecosystem Services, Prototype Digital Twin

Data

Frequent access of data derived from the real system is required to build a digital twin. There are a range of third-party data required to operate this pDT using two research infrastructures (RIs): GBIF[1] and eLTER[2], alongside other external data sources. The recreation potential (RP) component uses external data sources capturing information about the physical (natural and built) environment. These include altitude, slope, land cover, water bodies such as lochs and rivers, footpaths, and roads. The RP model combines these spatial datasets with recreation potential scoring representing persona, initially two personas representing the hard and soft recreationalists. These data are stored as offline data sources, with an aspiration to hold this data online to facilitate recreationalists to personalise their recreation potential maps. The biodiversity component collates data primarily from the GBIF RI and eLTER RI. GBIF provides API access and R package rgbif (Chamberlain et al., 2024) allowing frequent data access and improving the pDT's synchronicity. Relevant eLTER data such as Environmental Change Network (ECN) data is accessed via the eLTER digital asset registry (DAR) and hosted on Environmental Information Data Centre (EIDC). Environmental data for species distribution models are accessed via Google Earth Engine catalogue.

Model

The recreation potential model implementation follows methods described in Dick et al. (2022). The model is a spatial model (as opposed to statistical/mechanistic etc.) implemented in geographic information systems (GIS) software. The recreation potential model analyses a wide range of data, including natural and infrastructure features that influence the potential capacity to provide recreational opportunities e.g. terrain, land cover, proximity to water, and accessibility, to estimate the suitability of various locations for recreational activities. The model creates a recreation potential index that identifies areas with high potential for specific leisure recreator personas, such as 'hard' or 'soft' recreationalists. The model has been ported from QGIS to run in R programming language using the 'raster' spatial processing packages.

The species distribution model (SDM) uses species occurrence data and relevant environmental variables to predict the spatial distribution and habitat suitability of different species using statistical models. The models are implemented in R using the flexsdm package (Velazco et al.) and terra package (Hijmans et al., 2024) for spatial processing. Currently implemented are a gaussian process model, a generalised linear model, a support vector machine and an ensemble model. These model types are chosen to be indicative of SDM performance within a digital twin setting. The pDT incorporates an adaptive sampling approach to continuously improve the biodiversity component of the pDT's representation of the real system using approaches pioneered in the DECIDE project (Pocock et al., 2023) whereby citizen scientists will be directed to areas where the pDT requires biodiversity data to improve its biodiversity models and citizens recreating in nature have the highest probability of viewing a species. It is envisaged that citizen scientists will record the biodiversity they see where they usually record (eg. iRecord, NESBREC, iNaturalist) and will feed into the pDT via existing dataflows.

---

[1] https://www.gbif.org/
[2] https://elter-ri.eu/

FAIRness

The pDT uses various tools and methods to support its FAIRness. A data management plan (DMP) has been completed via the UKCEH FAIR Data Stewardship Wizard (https://fair-wizard.com/), which is a bespoke data questionnaire that prompts consideration of data activities required in a research endeavour and is linked to best practice guidance on many topics. In addition, the pDT is documented in the data management plan of BioDT (Harrison, 2022).

Whenever possible, digital objects will be released to relevant open repositories with assigned persistent identifiers (PIDs) and descriptive metadata. We use the Research Object Crate (RO-Crate) metadata format (Soiland-Reyes et al. 2022), which provides a machine-readable mechanism to communicate the diverse set of digital and real-world resources that contribute to an item of research. Model code is available under the MIT licence on an open source repository within the BioDT organisation on GitHub (https://github.com/BioDT/uc-ces).

We follow the ODMAP (Overview, Data, Model, Assessment and Prediction) protocol (Zurell et al., 2022) with the biodiversity component to describe the model development and application process in human-readable documentation. This ensures transparency and reproducibility, facilitating peer review, evaluations of model quality, and meta-analysis. R code has been developed to automatically generate ODMAP protocol information with each run of the biodiversity component. The workflows are published on the BioDT space in WorkflowHub (https://workflowhub.eu/projects/130) (Goble et al., 2021), a system-agnostic workflow management registry: workflows may remain in their native repositories in their native forms.

Performance

The recreation potential model has currently been run on a laptop but will be run as a single job on a high memory HPC node. The biodiversity models are run as separate jobs per species ensuring straightforward work allocation across nodes. In the pilot study, biodiversity models were run for a target 100 species. Each job (encompassing each model type and ensemble) for each species took between 5 and 20 minutes to complete, depending on the volume of data available for each species, with more data resulting in longer runtime. We found that the model performance was variable, however the ensemble model (mean weighted) of all species achieved an AUC (area under curve, 0-1, larger value = better prediction) >0.7. We found that generalised linear models performed worst of the model types, whereas the ensemble performed the best.

Interface and outputs

Our prototype digital twin requires a means for the target users to access and interact with the pDT. A minimal viable product user interface was developed as a foundation for further development and to enable a stakeholder training workshop. The minimal viable product user interface presents the two pDT components as separate interactive maps however there are aspirations to integrate these components. The biodiversity map shows predicted species distributions for different species groups (Fig. 2 A) and shows the maps for individual species (Fig. 2 B). In the table, species are listed based on their probability of occurrence and the DECIDE recording priority. The RP map (Fig. 2 C) visualises the recreation potential for different personas. The user interface is developed as a module within the BioDT web application ensuring consistent branding and infrastructure to the other pDTs: https://github.com/BioDT/biodt-shiny. The web application was built using the R Shiny framework (Winston et al., 2024). The graphical user interface (GUI) was considered in detail at a workshop

where valuable feedback was received (Cultural ecosystem services - testing pDT with experts | BioDT).
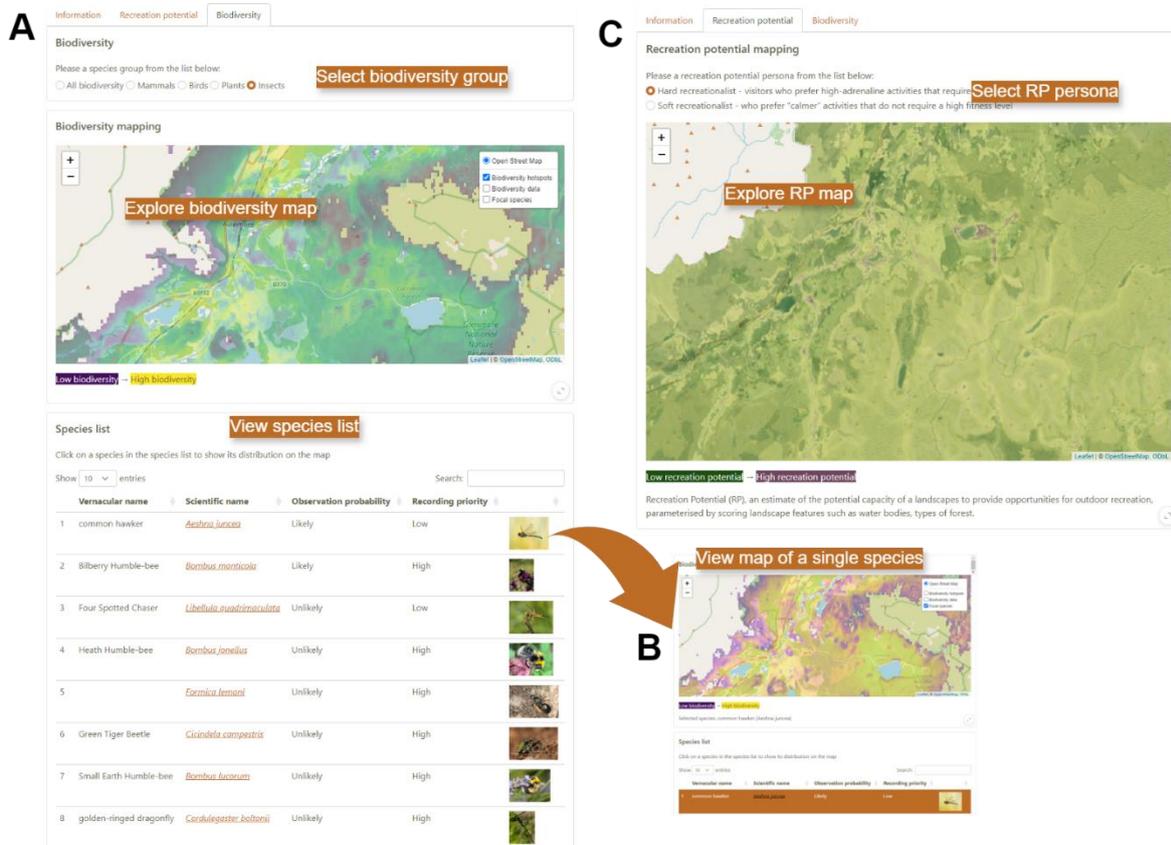

Fig 2: Screenshots from pDT user interface minimal viable product. The page has three tabs, the first (not illustrated) is an information tab with an overview of the pDT. The recreation potential tab offers maps for different RP personas. The biodiversity tab has maps of species distributions available as groups of species or individual species. Further developments of the pDT will provide means to access the data and better integrate between the two components on the pDT.

Integration and sustainability

The sustainability and application of this pDT are in the early stages of development. Initial thoughts have been outlined considering types of users, responsible organisation, computing requirements, and business options. The potential value of the data obtained from the pDT for policy makers is clear but depends on the pDT user allowing their use of the nature areas to be tracked and that depends on the business model adopted. The sustainability of data sources is guaranteed as biodiversity data sources are via European research infrastructures or nationally maintained datasets.

There may also be value to consider linking the pDT to the wellness tourism industry as it is poised for a transformative shift as it begins to embrace wearable healthcare technology. This integration promises to offer travellers a more personalised and proactive approach to maintaining their health while providing business opportunities and challenges to industry stakeholders.

Application and impact

Initial studies scaling up the pDT has highlighted the differences between the modelling approaches adopted for the biodiversity and the recreational potential models. The biodiversity models use

limited data providers (GBIF and eLTER RI) and the data is standardised for all locations which means the same models can be used at any scale. This is not the same for the recreational potential model which uses multiple third-party data sources, some of which were specific to the Cairngorm National Park and therefore not available for scaling up the model to wider geographic areas. Although analogous datasets can be found, it is important to note that model fidelity will decrease during scaling up. In local areas highly specific knowledge can be applied to dataset scoring which improves the useability of the model to the end-user (e.g. scoring a river or lakes suitability for recreation by intimate knowledge of its use and name), however this cannot be applied at wide geographic scales, so more generic scoring is required (e.g. based on river size or type). A bottom-up approach where all datasets are suitable for parameterisation across multiple scales would be complex to implement but should be considered in future iterations of the model.

The implications of our pDT for rare and endangers species was one aspect covered in a workshop held with policy makers and regulators. The conclusion that the recreational potential maps should be parametrised such that sensitive areas are not recommended for any recreational persona was endorsed by most attendees. The dynamic nature of the pDT will enable temporally variable aspects of biodiversity to be accommodated e.g. the breeding behaviour of a ground nesting bird on the IUCN red list, the Capercaillie (*Tetrao urogallus*), is a major tourist attraction in the Cairngorms but uncontrolled access can disturb the birds during the breeding season so restricted access is recommended by ensuring parameterisation of the recreational model never shows the breeding display (lecking) areas i.e. these are given low recreational potential only during the breeding season.

Acknowledgements

References

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2024). rgbif: Interface to the Global Biodiversity Information Facility API [Computer software]. https://CRAN.R-project.org/package=rgbif

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, & J., Dipert (2024). shiny: Web Application Framework for R (1.8.1.1) [Computer software]. https://cran.r-project.org/web/packages/shiny/index.html

Dick, J., Chris Andrews, Daniel E. Orenstein, Yael Teff-Seker, Grazia Zulian, A mixed-methods approach to analyse recreational values and implications for management of protected areas: A case study of

Cairngorms National Park, UK, Ecosystem Services, Volume 56, 2022, 101460, ISSN 2212-0416, https://doi.org/10.1016/j.ecoser.2022.101460 .

Goble, C., Soiland-Reyes, S., Bacall, F., Owen, S., Williams, A., Eguinoa, I., Droesbeke, B., Leo, S., Pireddu, L., Rodríguez-Navas, L., Fernández, J. M., Capella-Gutierrez, S., Ménager, H., Grüning, B., Serrano-Solano, B., Ewels, P., & Coppens, F. (2021). Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory. Zenodo. https://doi.org/10.5281/zenodo.4605654

Gould RK, Morse JW, Adams AB. Cultural ecosystem services and decision-making: How researchers describe the applications of their work. People and Nature. 2019 Dec;1(4):457-75.

Harrison, J., Allas, A.-L., Koivula, H., & Blankett, R. (2022). BioDT D1.2 - Data Management Plan. Zenodo. https://doi.org/10.5281/zenodo.7621030

Hijmans, R. J., Bivand, R., Pebesma, E., & Sumner, M. D. (2024). terra: Spatial Data Analysis (1.7-71) [Computer software]. https://cran.r-project.org/web/packages/terra/index.html

Orenstein, D.E., Katz-Gerro, T., Dick, J., 2017. Environmental tastes as predictors of environmental opinions and behaviors. Landscape Urban Plann. Elsevier B.V. 161, 59–71. https://doi.org/10.1016/j.landurbplan.2017.01.005.

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., 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

Pocock, M., August, T., Rolph, S., Mondain-Monval, T., Jarvis, S., Burkmar, R., Baird, K., Pateman, R., Dyke, A., Wright, E., McInerny, G., & Turkay, C. (n.d.). The DECIDE Recorder Tool: Recording nature where it matters [Web application]. Retrieved January 20, 2024, from https://decide.ceh.ac.uk

Velazco, S. J. E., Rose, M. B., de Andrade, A. F. A., Minoli, I., & Franklin, J. (2022). flexsdm: An r package for supporting a comprehensive and flexible species distribution modelling workflow. Methods in Ecology and Evolution, 13(8), 1661–1669. https://doi.org/10.1111/2041-210X.13874

Zulian, G., Stange, E., Woods, H., Carvalho, L., Dick, J., et al., 2018. Practical application of spatial ecosystem service models to aid decision support. Ecosyst. Serv. 29 (Pt C), 465–480. https://doi.org/10.1016/j.ecoser.2017.11.005.

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., … Merow, C. (2020). A standard protocol for reporting species distribution models. Ecography, 43(9), 1261–1277. https://doi.org/10.1111/ecog.04960