PRIME: An Open-Source Data Pipeline for Streamlining Peatland Rewetting

Behzad Mozafari[1], Fiachra O'Loughlin[1], Florence Renou-Wilson[2], Shane Donohue[1], Michael Bruen[1], David Morgan[1]

[1]School of Civil Engineering and UCD Earth Institute, University College Dublin, Ireland
[2]School of Biology and Environmental Science, Science Centre-West, University College Dublin, Dublin 4, Ireland

Abstract

Monitoring systems, data analysis and decision-making tools are becoming increasingly automated and integrated into ecosystem management. Along with new technology, this presents an opportunity to advance peatland management and move beyond traditional labour-intensive and experience-based approaches to create more informed, efficient, and sustainable practices.

In this research, we (i) **introduce a novel data pipeline** (as a proof-of-concept) for Peatland Restoration Integrated with Modelling and Evaluation (named 'PRIME'), which is designed to streamline data processing and analysis in peatland restoration studies; and (ii) **suggest an approach for implementing a standardized and reproducible workflow** for future peatland restoration studies.

The pipeline includes a set of functions that automate the transformation and movement of data between different data sources and types. It also incorporates key data management principles such as versioning, archiving, and citation to ensure the integrity and reliability of data. PRIME is designed to handle large and complex datasets. It automatically performs data cleaning and formatting to ensure data consistency. The pipeline also includes advanced visualization tools that enable users to interact with and explore data in near-real-time.

We argue that such integrated systems with built-in flexibility are crucial to ensure the effective use of both existing and newly collected data for peatland restoration. They can also provide decision support for stakeholders. PRIME is an open-source system that can be customized to address the challenges of interdisciplinary peatland restoration or broader ecosystem management initiatives.

Introduction

Peatlands play a crucial role in carbon storage, clean water provision, and biodiversity conservation. However, human activities, e.g., land use changes and drainage for agricultural, forestry, and peat extraction purposes, have severely degraded many peatlands worldwide, resulting in adverse consequences for the environment [1-2].

Restoration, especially through rewetting, is a vital step towards achieving carbon neutrality and ensuring long-term resilient and sustainable systems.

Digitalization and the fourth industrial revolution (4IR) have opened up exciting opportunities for data-driven natural resource management. Emerging technologies, e.g., digital twins, Internet of Things (IoT), smart sensors, remote sensing, artificial intelligence (AI), and machine learning, offer unprecedented capabilities for monitoring and managing natural resources, enabling precise decision-making and real-time interventions. In peatland ecosystems, these technologies have enormous potential to advance our understanding of peatland processes and support evidence-based management practices for sustainable peatland use [3]. However, realizing this potential requires overcoming significant technical, institutional, and societal challenges, and necessitates collaboration among diverse stakeholders.

The big data generated by these technologies can be a double-edged sword, offering rich insights but also presenting significant data management challenges. To address these challenges, there is a growing need for systematic data collection and management, in accordance with the FAIR (findable, accessible, interoperable, and reusable) principles [4]. Several initiatives are currently providing peat datasets to promote data sharing and enable researchers to extract maximum value from the data. However, the absence of commonly accepted standards and protocols appears to hinder interoperability between datasets and the provision of advanced analytical tools that facilitate data exchange and integration from multiple sources.

Data pipelines are a critical component of data lifecycle management, as they provide an automated and systematic workflow for ingesting, storing, and processing data, allowing near real-time analysis, supporting timely decision-making and action-taking. Additionally, they can help to reduce redundancy and ensure data consistency, which is essential for reliable and accurate analyses. Data pipelines can also provide data privacy and security, and can be customized for different users, such as researchers, policymakers, and practitioners. Therefore, current peatland research appears to significantly benefit from such a pipeline to streamline restoration and rewetting projects.

This poster introduces our novel data pipeline, named PRIME (Peatland Restoration Integrated with Modelling and Evaluation), and highlights its key components.

PRIME Framework

PRIME is an open-source package written in R that contains functions for the data lifecycle management. Using R's built-in functions and packages, the following components are designed as shown in Figure 1.

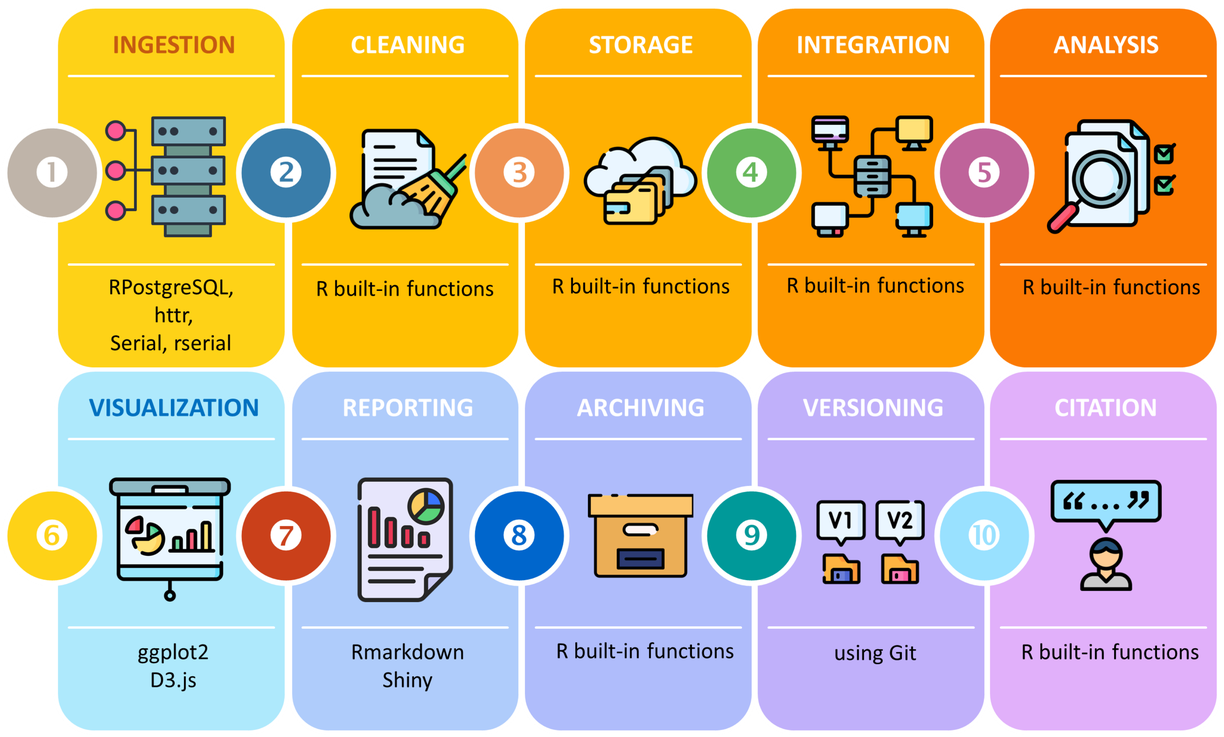

Figure 1. The key components of PRIME data pipeline for each stage of the data lifecycle, which were developed using R's built-in functions or packages. (This image was made using Flaticon.com)

We are examining additional components of the framework, as well as test cases from the Republic of Ireland on peatland rewetting projects. A conceptual hydrological modelling framework is being adopted to the pipeline to help inform timely decision making and be an integral part of the data value chain. Using machine learning tools for data analysis is another possibility.

Conclusion

Advancements in technology have led to an exponential increase in the availability of data, creating a need for effective data management strategies. In response to this, we propose a streamlined and automated data pipeline, named PRIME, to support peatland rewetting projects. We believe that standardized and FAIR analysis workflows are essential for effective data management in peatland ecosystems.

References

[1] F. Renou-Wilson, G. Moser, D. Fallon, C. Farrell, C. Müller, and D.Wilson. Rewetting degraded peatlands for climate and biodiversity benefits: Results from two raised bogs. Ecological Engineering, 127:547–560, Feb. 2019. doi: 10.1016/j.ecoleng.2018.02.014. URL https://doi.org/10.1016/j.ecoleng.2018.02.014.

[2] J. Limpens, F. Berendse, C. Blodau, J. G. Canadell, C. Freeman, J. Holden, N. Roulet, H. Rydin, and G. Schaepman-Strub. Peatlands and the carbon cycle: from local processes to global implications – a synthesis. Biogeosciences, 5(5):1475–1491, Oct. 2008. doi: 10.5194/bg-5-1475-2008. URL https://doi.org/10.5194/bg-5-1475-2008.

[3] N. J. Rowan, N. Murray, Y. Qiao, E. O'Neill, E. Clifford, D. Barceló, and D. M. Power. Digital transformation of peatland eco-innovations ('paludiculture'): Enabling a paradigm shift towards the real-time sustainable production of 'green-friendly' products and services. Science of The Total Environment, 838:156328, Sept. 2022. doi: 10.1016/j. scitotenv.2022.156328. URL https://doi.org/10.1016/j.scitotenv.2022.156328.

[4] M. D.Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A.Waagmeester, P.Wittenburg, K.Wolstencroft, J. Zhao, and B. Mons. The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3(1), Mar. 2016. doi: 10.1038/sdata.2016.18. URL https://doi.org/10.1038/sdata.2016.18.

This research was funded under project 2018-W-LS-18 (SWAMP) of the Environmental Protection Agency (EPA) Research Programme 2014-2020. The EPA Research Programme is a Government of Ireland initiative funded by the Department of Communications, Climate Action, and Environment. It is administered by the Environment Protection Agency, which has the statutory function of coordinating and promoting environmental research.