

Aug 26, 2025

A Coding Protocol for Labeling Scientific Literature on Carbon Dioxide Removal to Train Machine Learning Models

DOI

<https://dx.doi.org/10.17504/protocols.io.e6nvwqwqwvmk/v1>

Sarah Lueck¹, Christiane Hamann¹, Doménica ichelle Jaramillo Sánchez¹, Ronja Kelch¹, Fariha Mawla¹, Fabian Metz¹, Leon Stephan¹, David Verdugo-Raab¹

¹Potsdam Institute for Climate Impacts



Sarah Lueck

Potsdam Institute for Climate Impacts

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.e6nvwqwqwvmk/v1>

Protocol Citation: Sarah Lueck, Christiane Hamann, Doménica ichelle Jaramillo Sánchez, Ronja Kelch, Fariha Mawla, Fabian Metz, Leon Stephan, David Verdugo-Raab 2025. A Coding Protocol for Labeling Scientific Literature on Carbon Dioxide Removal to Train Machine Learning Models. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.e6nvwqwqwvmk/v1>

License: This is an open access protocol distributed under the terms of the **[Creative Commons Attribution License](#)**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: June 17, 2025

Last Modified: August 26, 2025

Protocol Integer ID: 220414

Keywords: CDR, Carbon Dioxide Removal, Climate Change, machine learning models carbon dioxide removal, coding protocol for labeling scientific literature, labeling scientific literature, global warming, carbon dioxide removal, annotation workflow, relevant scientific literature, assisted evidence mapping, ai, transparent evidence synthesis, machine learning model, dataset, scientific evidence, artificial intelligence, quality training dataset, evidence mapping in the cdr domain, pace with the scientific evidence, facilitating scalable ai, reproducibility

Funders Acknowledgements:

ERC-2020-SyG "GENIE"

Grant ID: 951542

Abstract

Carbon dioxide removal (CDR) plays an important role in any strategy to limit global warming to well below 2°C. As research on CDR rapidly expands, keeping pace with the scientific evidence through rigorous and transparent evidence synthesis is essential for informing sustainable deployment. To support this, we use artificial intelligence (AI) to develop the first comprehensive systematic map of CDR research. A critical first step in this process is the creation of a high-quality training dataset to enable machine learning models to accurately identify relevant scientific literature. This coding protocol provides a structured methodology for generating such a dataset, including the development of a coding scheme, inclusion criteria, and annotation workflow. The protocol ensures consistency and reproducibility, facilitating scalable AI-assisted evidence mapping in the CDR domain.

Attachments



[Coding_Guideline_unl...](#)

757KB

