

Received: 23 December 2020 | Revised: 18 February 2022 | Accepted: 22 February 2022

DOI: 10.1002/cpe.7261

## **RESEARCH ARTICLE**

WILEY

# Evaluation of a scientific data search infrastructure

Alexandru Iulian Orhean<sup>1</sup>  | Anna Giannakou<sup>2</sup> | Katie Antypas<sup>3</sup> | Ioan Raicu<sup>1</sup> |  
Lavanya Ramakrishnan<sup>2</sup>

<sup>1</sup>College of Computing, Illinois Institute of Technology, Chicago, Illinois, USA

<sup>2</sup>Scientific Data Division, Lawrence Berkeley National Lab, Berkeley, California, USA

<sup>3</sup>NERSC, Lawrence Berkeley National Lab, Berkeley, California, USA

### Correspondence

Alexandru Iulian Orhean, College of Computation, Illinois Institute of Technology, Chicago, IL, USA.  
Email: [iorhean@cs.cornell.edu](mailto:iorhean@cs.cornell.edu)

### Summary

The ability to search over large scientific datasets has become crucial to next-generation scientific discoveries as data generated from scientific facilities grow dramatically. In previous work, we developed and deployed ScienceSearch, a search infrastructure for scientific data which uses machine learning to automate metadata creation. Our current deployment is deployed atop a container based platform at a HPC center. In this article, we present an evaluation and discuss our experiences with the ScienceSearch infrastructure. Specifically, we present a performance evaluation

