

archivers.space

Information gathered during Data Rescue Events through the archivers.space app. See <http://www.ppehlab.org/datarescueworkflow#retiredworkflow> for more information.

Priority *

Number should be 1-10

5

URL *

<https://publications.lbl.gov/>

Agency *

Energy Department

Event

Title

Welcome to The Berkeley Lab Publications System | Welcome to The Berkeley Lab Publications System

Crawled by the Internet Archive? *

Yes

No

Internet Archive URL

https://web.archive.org/web/2019*/https://publications.lbl.gov/

Description

Purpose or significance of data

Berkeley Lab Publications System. This is a collection of papers published 1940-2018. It contains the following:

- Scientific Publications, typically those items published in journals or proceedings.
- Public Reports, for divisions that publish reports that are not otherwise published in journals.

From the FAQ:

Publications.lbl.gov is a public repository of Berkeley Lab Scientific Publications and is the replacement for the existing reports coordination process. It is populated with the existing reports coordination data including historical documents spanning the entire history of the laboratory. It includes a new submission process for reports which allows users to provide additional useful metadata such as the group, project, funding source, and keywords associated with the paper. It also includes a new set of policies and procedures for submission of publications.

What's New?

The Publications System is viewable by the public; an LBNL LDAP is no longer needed to search for and download reports.

Report submitters can now enter publications into the system by DOI and PMID import, as well as RIS/Endnote XML file import.

Report numbers are now assigned immediately after an LBNL report is submitted to the system. To view the report number, go to the report's record in the system, and click the "LBL" tab.

- Do not harvest. All data is small, unstructured, and on a page crawlable by the Internet Archive.
- Page contains dynamic content (e.g., links loaded by JavaScript).
- Page contains interactive visualizations.
- Data is accessible in structured file(s) that can be directly downloaded.
- Data is accessible over FTP.
- Data is accessible using a documented public API.
- Data is only accessible using search queries in a web form.

Recommended approach to harvesting data

All URLs use a format like the following:

<https://publications.lbl.gov/islandora/object/ir%3A1007109>

The number/text after "ir%" varies.

Many of these papers are likely listed in Google Scholar.

File formats

HTML

Estimated size in MB

Related URLs

<https://publications.lbl.gov/>

Where you able to capture all of the data at this URL?

Yes

No

Harvest method used

Manually going through and saving pages to internet archive.

Notes from Harvest

Also included some of the available pdfs which were linked through google scholar on the page. A lot of these pdfs were not downloadable but can be gotten to by clicking on google scholar link.

User certified that to the best of their knowledge this is a well-checked bag that will survive out of context of the site it was harvested from.

Yes

Notes from Bagging

Notes from Describe

Google Forms