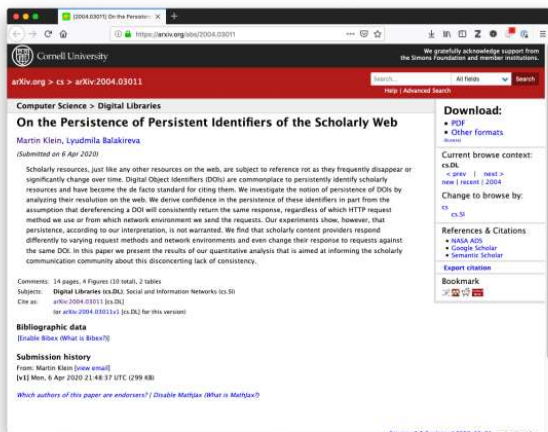
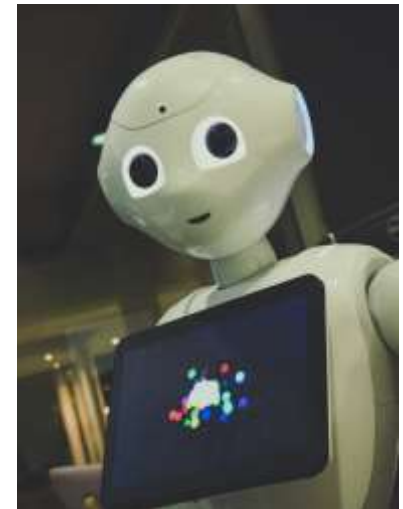
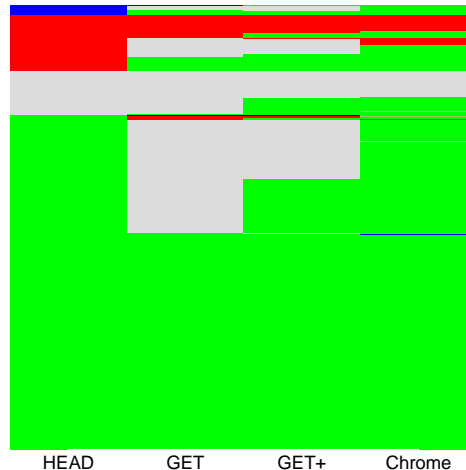


# On the Persistence of Persistent Identifiers of the Scholarly Web

Martin Klein & Lyudmila Balakireva  
Los Alamos National Laboratory  
{mklein, ludab}@lanl.gov



<https://arxiv.org/abs/2004.03011>



On the Persistence of Persistent Identifiers of the Scholarly Web  
@mart1nkle1n  
TPDL, August 2020



# DOIs are very common

The screenshot shows the IEEE Xplore website interface. At the top, there are navigation links for IEEE.org, IEEE Xplore, IEEE-SA, IEEE Spectrum, and More Sites. There are also links for SUBSCRIBE, Cart, Create Account, and Personal Sign In. The main header includes the IEEE Xplore logo, navigation menus for Browse, My Settings, and Help, and an Institutional Sign In button. A search bar is present with the text 'All' and a search icon. Below the search bar, the article title 'Evaluating Memento Service Optimizations' is displayed, along with the publisher 'IEEE' and buttons for 'Cite This' and 'PDF'. The authors are listed as Martin Klein, Lyudmila Balakieva, and Harshar Shankar. The abstract section is expanded, showing the text of the article. A red box highlights the DOI: 10.1109/JCDL.2019.00034. Other sections include 'Document Sections', 'References', and 'Keywords'. On the right side, there is a promotional banner for 'Need Full-Text access to IEEE Xplore for your organization?' and a 'CONTACT IEEE TO SUBSCRIBE' button. Below that, there are sections for 'More Like This' and 'Top Organizations with Patents on Technologies Mentioned in This Article'.

On the Persistence of Persistent Identifiers of the Scholarly Web

@mart1nkle1n

TPDL, August 2020



# DOIs are very common

The screenshot shows a Zenodo page for a preprint titled "Extended data: Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China". The page includes a search bar, navigation links, and a sidebar with document sections. The main content area features the title, authors (Endo, Akira; Abbott, Sam; Kucharski, Adam J.; Funk, Sebastian), a description of supplementary figures and data, and a table of data used from the World Health Organization and a COVID-19 app. A yellow box highlights a revision date of June 28, 2020. The right sidebar displays statistics (567 views, 481 downloads), the OpenAIRE logo, the publication date (April 5, 2020), and the DOI (10.5281/zenodo.3911576), which is highlighted with a red box. The license is Creative Commons Attribution 1.0 Generic.

IEEE.org | IEEE Xplore | IEEE-SA | IEEE Spectrum | More Sites | SUBSCRIBE | Cart | Create Account | Personal Sign in

IEEE Xplore® Browse ▾

zenodo Search [ ] Upload Communities [ ] Log in Sign up

April 5, 2020 [Figure] [Open Access]

## Extended data: Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China

Endo, Akira; Abbott, Sam; Kucharski, Adam J.; Funk, Sebastian

Supplementary figures and underlying data accompanying Endo et al. "Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China".

No primary data collection was involved in this study.

All data used are from:

- World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 38. 2020.
- COVID2019.app - LIVE stats and graphs. 2020 [cited 4 Mar 2020].

28 June 2020: Revised to include additional analysis of differential reproduction numbers

Preview

|             | Total | ImportedChina | ImportedOthers | Local | Unknown | Death |
|-------------|-------|---------------|----------------|-------|---------|-------|
| South Korea | 1766  | 13            | 4              | 605   | 1144    | 13    |
| Japan       | 186   | 28            | 11             | 129   | 18      | 3     |
| Singapore   | 93    | 24            | 0              | 69    | 0       | 0     |
| Australia   | 23    | 12            | 8              | 3     | 0       | 0     |
| Malaysia    | 22    | 18            | 2              | 2     | 0       | 0     |
| Vietnam     | 16    | 8             | 0              | 8     | 0       | 0     |

567 views 481 downloads [See more details...]

Indexed in OpenAIRE

Publication date: April 5, 2020

DOI: DOI 10.5281/zenodo.3911576

License (for files): Creative Commons Attribution 1.0 Generic

Versions: Version 1.1.0 10.5281/zenodo.3911576 Apr 5, 2020

On the Persistence of Persistent Identifiers of the Scholarly Web

@mart1nkle1n

TPDL, August 2020



# DOIs are very common

The image shows two overlapping web pages. The background page is Zenodo, displaying a registration for 'Extended data: Estimating COVID-19 transmission outside China' by Endo, Akira; Abbott, Sam; Kucharski, Adam. The foreground page is OSF Registries, showing a registration for 'A Scoping review protocol to map the evidence on strategies and interventions in neurotrauma and Road Traffic collisions prevention globally'. The OSF page includes a summary, contributors list, and a table of data. A red box highlights the 'Registration DOI' field, which contains the value '10.17605/OSF.IO/TDNP6'.

**OSF REGISTRIES -**

## A Scoping review protocol to map the evidence on strategies and interventions in neurotrauma and Road Traffic collisions prevention globally

Public registration

- Overview
- Files
- Wiki
- Components
- Links
- Analytics
- Comments

### Summary

Provide a narrative summary of what is contained in this registration, or how it differs from prior registrations.

This is a registration for a protocol of a scoping review to map the evidence on strategies and interventions in neurotrauma and Road Traffic collisions prevention globally. There are no prior registrations.

### Contributors

Santhani M Selveindran, Muhammad Mukhtar Khan, Daniel Simadibrata, Peter JA Hutchinson, Carol Brayne, Christine Hill, and Angelos Kollas

### Description

This is a protocol for a scoping review that seeks to identify, quantify and summarize the existing evidence-base on preventative strategies and interventions in neurotrauma, particularly for Road Traffic collisions

### Registration type

Open-Ended Registration

### Date registered

April 2, 2019

### Date created

April 2, 2019

### Registered from

osf.io/g5xkl

### Category

Project

### Registration DOI

10.17605/OSF.IO/TDNP6

CC-BY Attribution 4.0 International

|             | Total | ImportedC |
|-------------|-------|-----------|
| South Korea | 1766  | 13        |
| Japan       | 186   | 28        |
| Singapore   | 93    | 24        |
| Australia   | 23    | 12        |
| Malaysia    | 22    | 18        |
| Vietnam     | 16    | 8         |

On the Persistence of Persistent Identifiers of the Scholarly Web

@mart1nkle1n

TPDL, August 2020



# How does this work via HTTP?

[https://doi.org/10.1007/978-3-540-87599-4\\_38](https://doi.org/10.1007/978-3-540-87599-4_38)

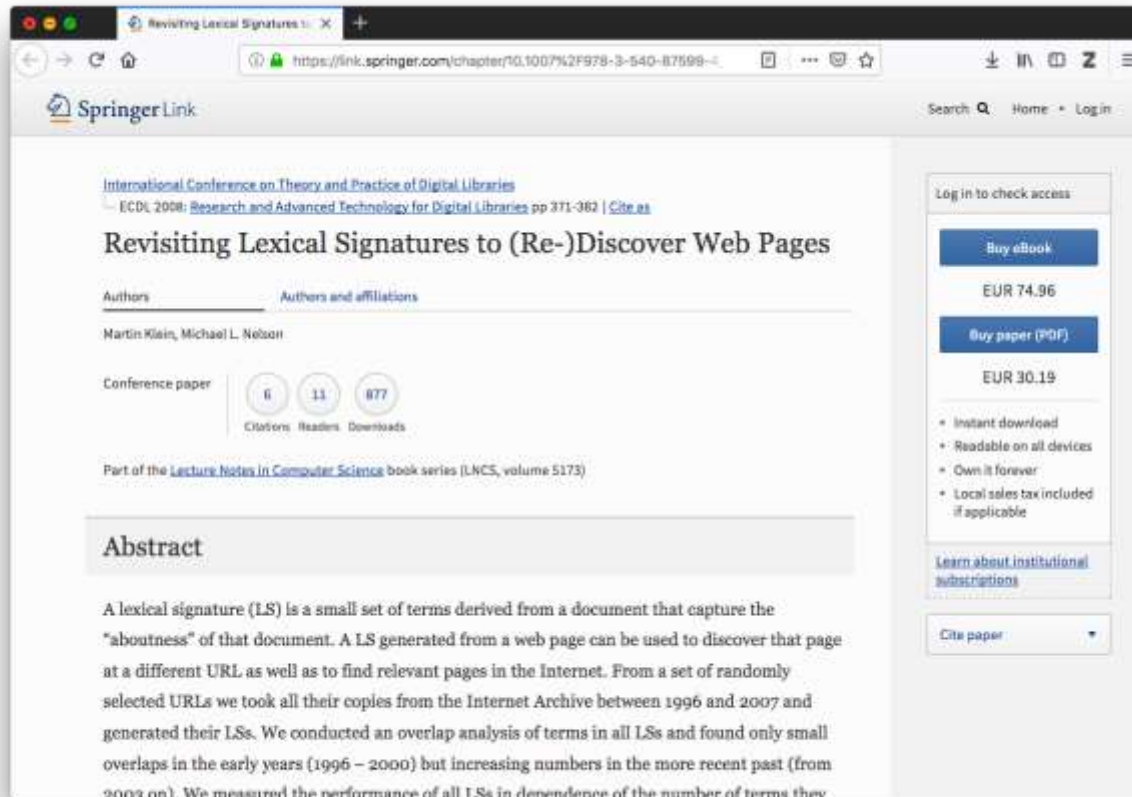


On the Persistence of Persistent Identifiers of the Scholarly Web  
@mart1nkle1n  
TPDL, August 2020



# Arrived at landing page

[https://doi.org/10.1007/978-3-540-87599-4\\_38](https://doi.org/10.1007/978-3-540-87599-4_38)



The screenshot shows a web browser window displaying the SpringerLink landing page for the chapter "Revisiting Lexical Signatures to (Re-)Discover Web Pages". The page includes the SpringerLink logo, a search bar, and a navigation menu. The main content area features the chapter title, authors (Martin Klein, Michael L. Nelson), and a list of statistics: 11 citations, 11 readers, and 877 downloads. The abstract is visible, starting with "A lexical signature (LS) is a small set of terms derived from a document that capture the 'aboutness' of that document." On the right side, there are purchase options: "Buy eBook" for EUR 74.96 and "Buy paper (PDF)" for EUR 30.19. A list of benefits for the eBook is provided, including instant download, readability on all devices, and ownership. A "Cite paper" button is also present.

[https://link.springer.com/chapter/10.1007%2F978-3-540-87599-4\\_38](https://link.springer.com/chapter/10.1007%2F978-3-540-87599-4_38)



On the Persistence of Persistent Identifiers of the Scholarly Web  
@mart1nkle1n  
TPDL, August 2020



# HTTP redirects

[https://doi.org/10.1007/978-3-540-87599-4\\_38](https://doi.org/10.1007/978-3-540-87599-4_38)

→ (HTTP 302 redirect)

[http://link.springer.com/10.1007/978-3-540-87599-4\\_38](http://link.springer.com/10.1007/978-3-540-87599-4_38)

→ (HTTP 301 redirect)

[https://link.springer.com/10.1007/978-3-540-87599-4\\_38](https://link.springer.com/10.1007/978-3-540-87599-4_38)

→ (HTTP 302 redirect)

[https://link.springer.com/chapter/10.1007%2F978-3-540-87599-4\\_38](https://link.springer.com/chapter/10.1007%2F978-3-540-87599-4_38)

→ (HTTP 200)



# Questions...

- How persistent is this DOI resolution?
- Given different clients and network environments:
  - Can we consistently arrive at the same location at the end of the redirect chain?
  - Is the path there (redirect chain) the same?
  - Are there differences between Open Access and non-OA?
  - Subscription vs non-Subscription level content?
  - Do scholarly content providers differ from the popular web?





# Idea...

- Comparative study investigating scholarly publishers' responses
  - To common HTTP requests
  - Against DOIs
- Using different web clients and request methods, resembling
  - Machines "browsing", crawling
  - Humans browsing
- From network environments with different subscriptions/licenses
  - Amazon Web Service EC2 instance
  - LANL internal
- Compare against web servers providing popular web content



# HTTP clients, request methods, dataset, networks

- HTTP HEAD
  - cURL
- HTTP GET
  - cURL
- HTTP GET+
  - cURL + various common parameters e.g., user agent, cookies
- HTTP GET
  - Chrome
- 10,000 DOIs, randomly picked, 100 DOIs from the 100 most frequent publisher domains
- HTTP requests sent from AWS VM and LANL network



# HTTP clients, request methods, dataset, networks



- **HTTP HEAD**

- cURL



- **HTTP GET**

- cURL



- **HTTP GET+**

- cURL + various common parameters e.g., user agent, cookies

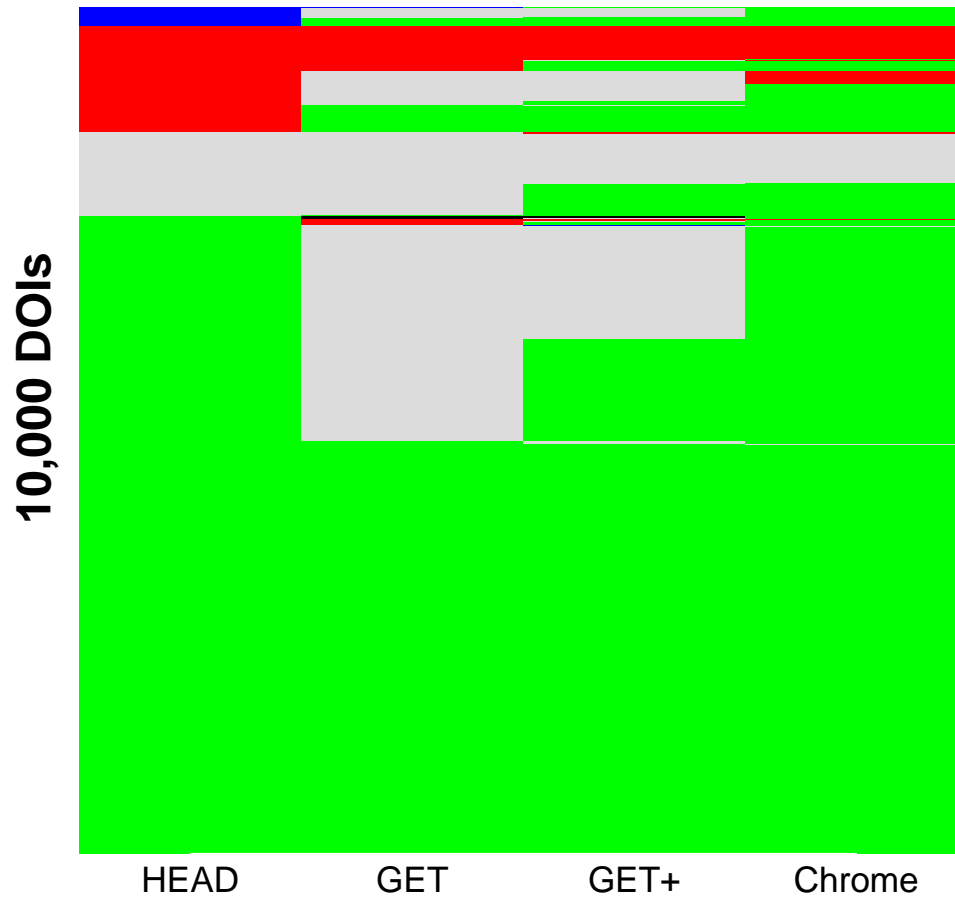


- **HTTP GET**
  - **Chrome**

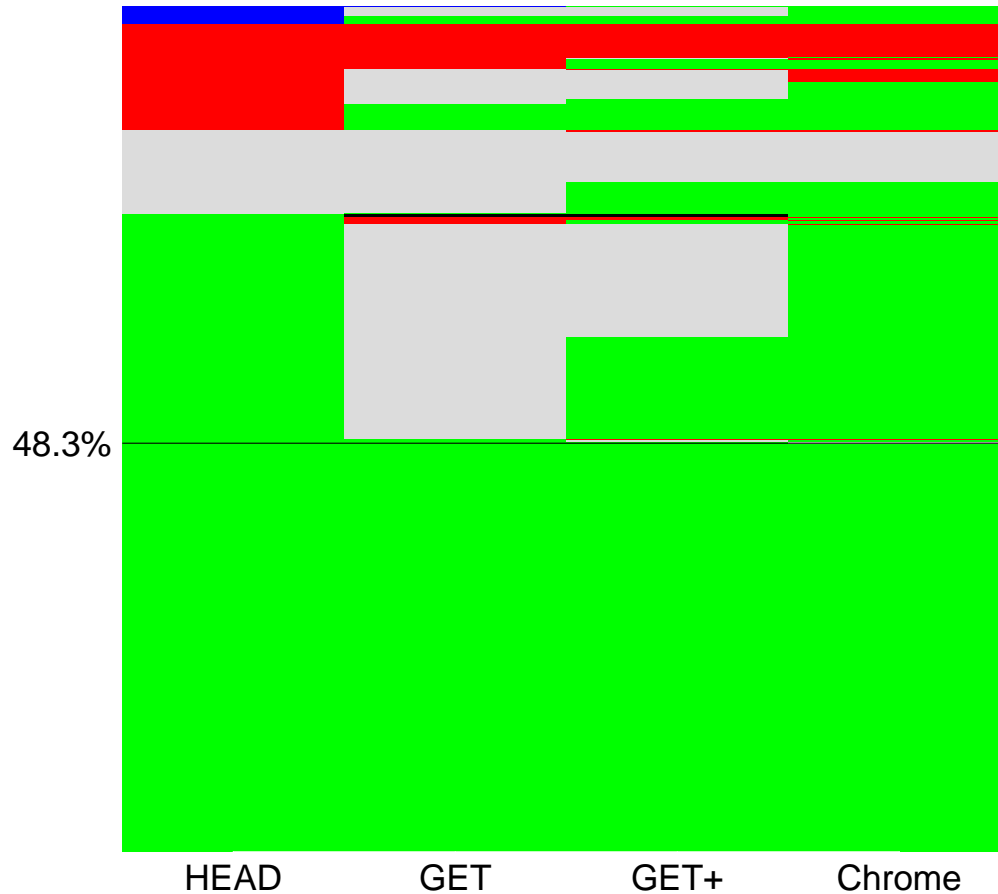
- 10,000 DOIs, randomly picked, 100 DOIs from the 100 most frequent publisher domains
- HTTP requests sent from AWS VM and LANL network



# Response codes of last link in redirect chain by DOI



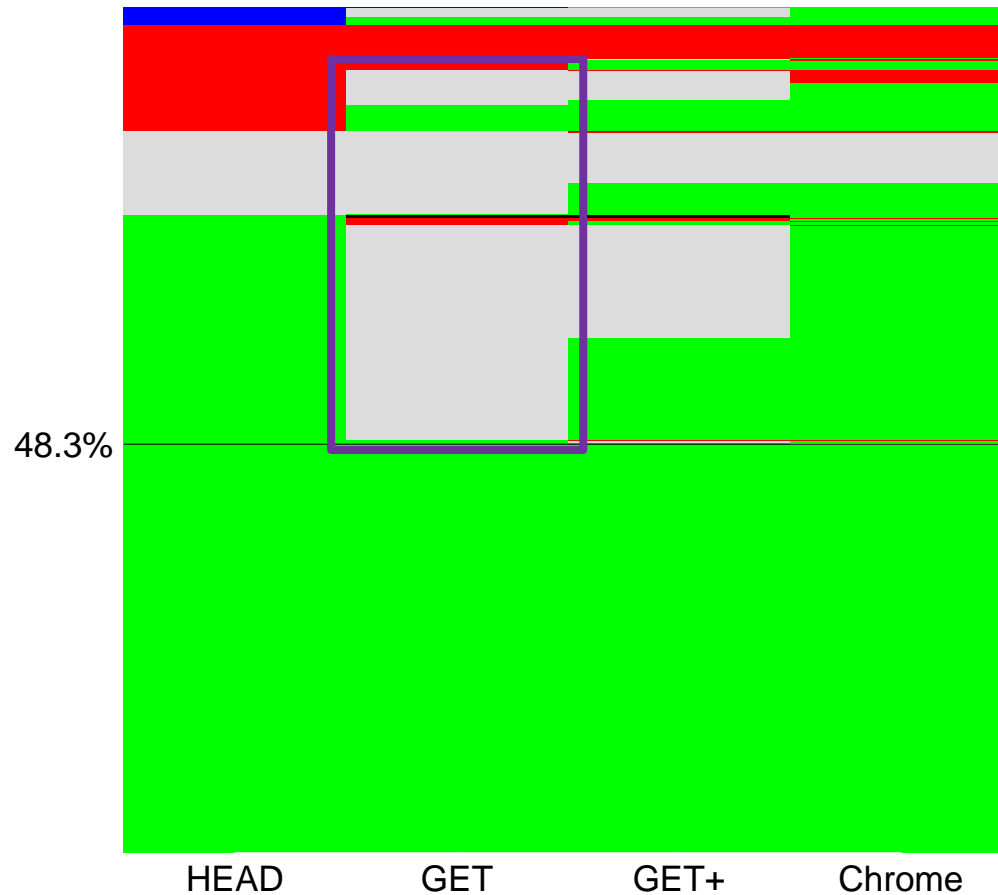
# Response codes of last link in redirect chain by DOI



- < 50% successful requests across all methods



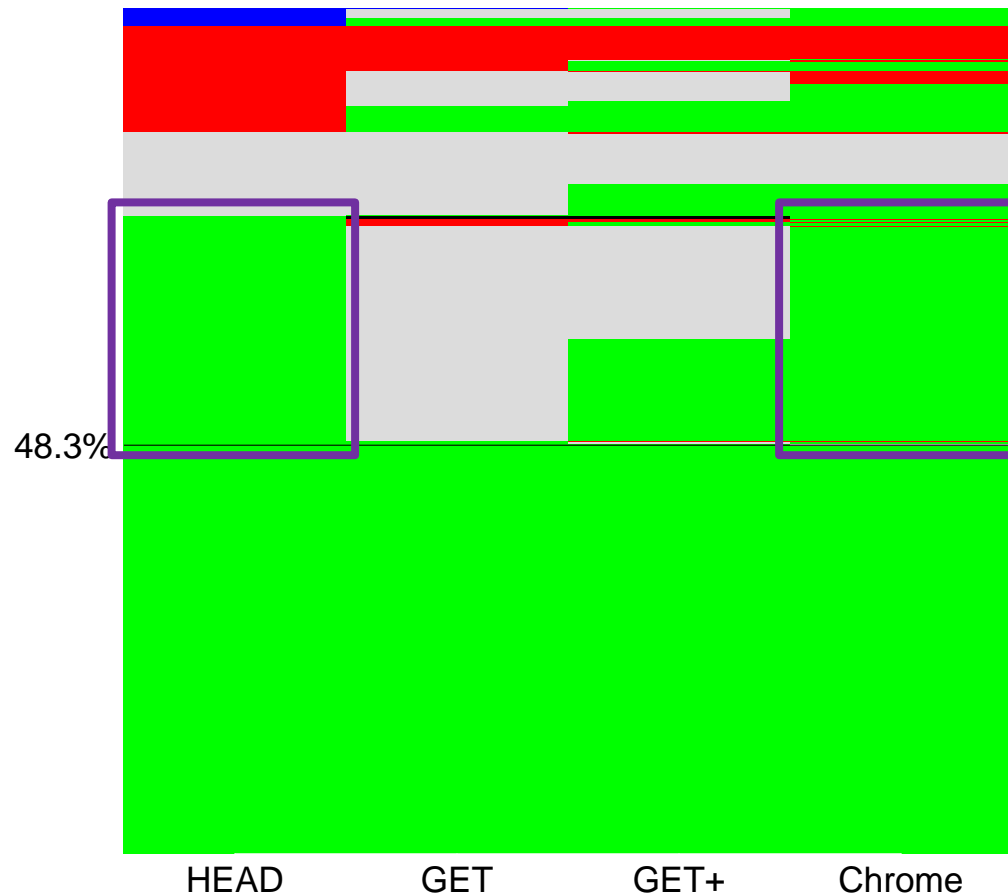
# Response codes of last link in redirect chain by DOI



- < 50% successful requests across all methods
- > 40% 300-level responses w/ GET



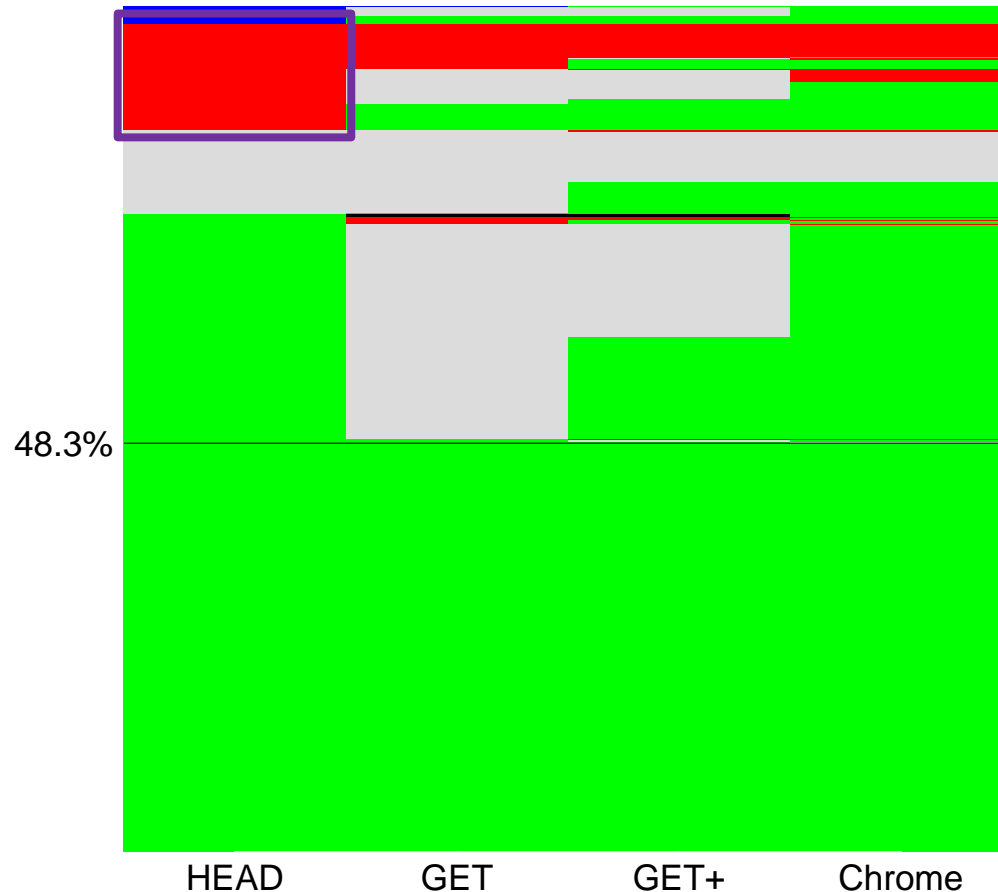
# Response codes of last link in redirect chain by DOI



- < 50% successful requests across all methods
- > 40% 300-level responses w/ GET
  - 25% return 200-level w/ HEAD/Chrome



# Response codes of last link in redirect chain by DOI

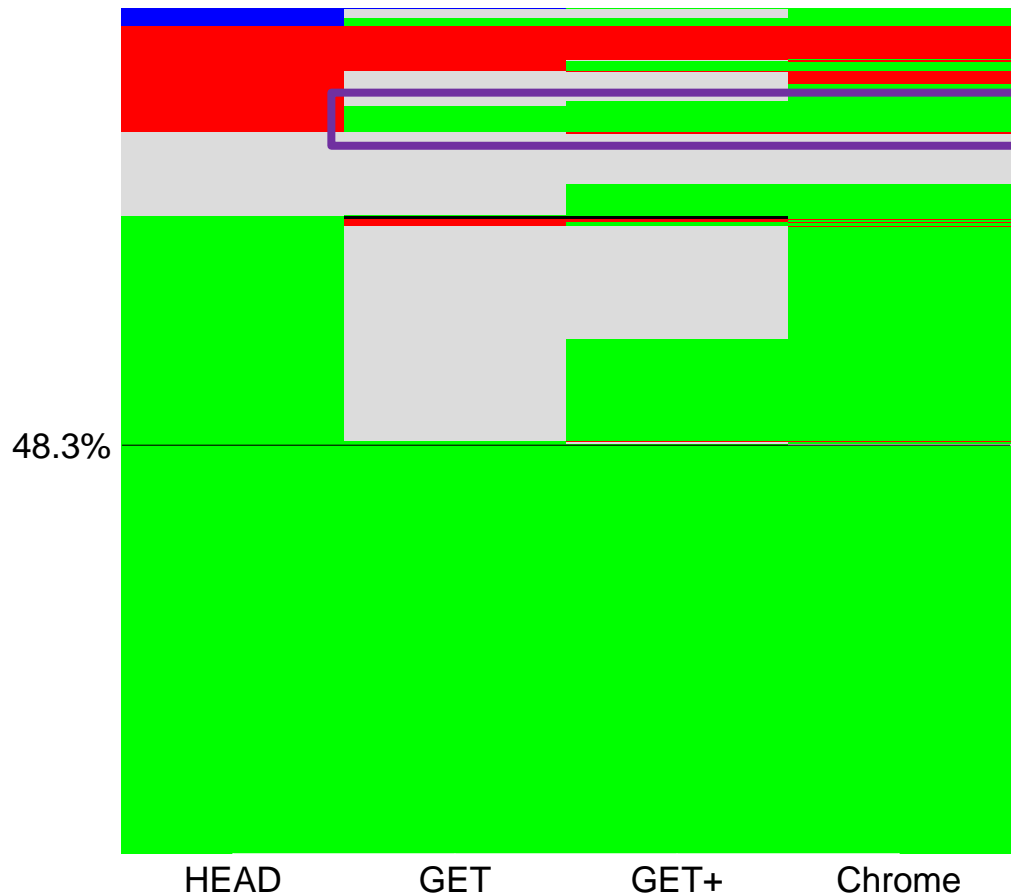


- < 50% successful requests across all methods
- > 40% 300-level responses w/ GET
  - 25% return 200-level w/ HEAD/Chrome
- 13% 400-level responses w/ HEAD





# Response codes of last link in redirect chain by DOI



- < 50% successful requests across all methods
- > 40% 300-level responses w/ GET
  - 25% return 200-level w/ HEAD/Chrome
- 13% 400-level responses w/ HEAD
  - 25% of them w/ 200-level response w/ any other method



# For more background, details, results

The screenshot shows a web browser window displaying the arXiv.org abstract page for the paper "On the Persistence of Persistent Identifiers of the Scholarly Web" by Martin Klein and Lyudmila Balakireva. The page is titled "Computer Science > Digital Libraries" and includes a search bar, navigation links, and a "Download" section with options for PDF and other formats. The abstract text discusses the persistence of Digital Object Identifiers (DOIs) on the web. The page also features sections for "Bibliographic data", "Submission history", and "References & Citations".

<https://arxiv.org/abs/2004.03011>



On the Persistence of Persistent Identifiers of the Scholarly Web  
@mart1nkle1n  
TPDL, August 2020



# On the Persistence of Persistent Identifiers of the Scholarly Web

**Thank you  
&  
stay safe!**

Martin Klein & Lyudmila Balakireva  
Los Alamos National Laboratory  
 [{mklein, ludab}@lanl.gov](mailto:{mklein, ludab}@lanl.gov)

