

Bug Pictures Beyond The Internet

Jorrit Poelen <https://jhpoelen.nl>

2024-08-19

Guiding Questions

- ▶ How do *you* cite data?
- ▶ How do you look up cited data *now*?
- ▶ How do you look up cited data *40 years from now*?

There's A Bug In Zenodo...



Figure 1: Bug

There's A Bug in Zenodo...

zenodo.org/records/13342373

zenodo

Search records... Communities My dashboard

Published August 19, 2024 | Version v1

Photo Open

Photo of Specimen BMT0009388

Museum für Naturkunde

Uploaded by Plazi for the Museum für Naturkunde Berlin.

Files

BMT121_BMT0009388_stacked_01.tif

The image shows a screenshot of a Zenodo record page. At the top, there is a navigation bar with a bookmark icon, the URL 'zenodo.org/records/13342373', a search bar containing 'Search records...', and links for 'Communities' and 'My dashboard'. Below the navigation bar, the word 'zenodo' is displayed in a large, lowercase, sans-serif font. To the right of 'zenodo' is a search bar with the placeholder 'Search records...' and a magnifying glass icon. Further to the right are links for 'Communities' and 'My dashboard'. The main content area has a blue header bar with the text 'Published August 19, 2024 | Version v1' and two buttons: 'Photo' and 'Open'. Below the header, the title 'Photo of Specimen BMT0009388' is displayed in bold black text. Underneath the title, the text 'Museum für Naturkunde' is shown. A note below that says 'Uploaded by Plazi for the Museum für Naturkunde Berlin.' In the center of the page is a large image of a small, pale, segmented insect larva or nymph, viewed from above. It has long, thin appendages extending from its body. To the right of the main image is a smaller, square thumbnail image of the same specimen, also with a caption 'BMT121_B...'. On the far left, there is a vertical sidebar with a three-line menu icon and a window icon.

So How'd You Cite This Bug?

Museum für Naturkunde. (2024). *Photo of Specimen BMT0009388.* Zenodo. <https://doi.org/10.5072/zenodo.13342373>

The screenshot shows a detailed landing page for a specimen record on Zenodo. At the top, the URL is zenodo.org/records/13342373. The page includes sections for Identifiers, Related works, and Biodiversity, each with specific data points. A central column contains resource type, publisher, rights, and citation information. A bottom section allows for citation style selection (APA) and download.

Identifiers

- URL: [hash://sha256/69df9b5bc5df5619448ade01c92d0a5e20bcec2ab4c8b668d2db56f9e5cce6e7](https://sha256/69df9b5bc5df5619448ade01c92d0a5e20bcec2ab4c8b668d2db56f9e5cce6e7)
- LSID: <urn:lsid:github.com:darktaxon:BMT0009388>

Related works

- Has version
 - Photo: [hash://sha256/69df9b5bc5df5619448ade01c92d0a5e20bcec2ab4c8b668d2db56f9e5cce6e7](https://sha256/69df9b5bc5df5619448ade01c92d0a5e20bcec2ab4c8b668d2db56f9e5cce6e7) (URL)
 - Photo: [hash://md5/f052cbae398bf340a620c421bfefaf16d](https://md5/f052cbae398bf340a620c421bfefaf16d) (URL)
- Is derived from
 - Dataset:
<https://linker.bio/line:hash://sha256/4a4dccbbdc51e11f9ec2c878df263b84b3a995a67520d474e00301a9467abb231> (URL)

Biodiversity

- Catalog number: [BMT0009388](#)
- Institution code: [MN](#)

Resource type: Photo
Publisher: Zenodo

Rights: Creative Commons Attribution 4.0 International

Citation: Museum für Naturkunde. (2024). Photo of Specimen BMT0009388. Zenodo. <https://doi.org/10.5072/zenodo.13342373>

Style: APA

Figure 3: Bug Landing Page

How To Retrieve This Zenodo Bug Now?

Like this?

```
curl -L "https://doi.org/10.5072/zenodo.13342373"\> bug.tiff
```

How To Retrieve This Bug Now?

Wait a minute ...

```
cat bug.tiff  
| cut -z -b1-62
```

```
<!doctype html><html lang="en_US"><head><meta charset="UTF-8"
```

That ain't no bug, it is an HTML page.

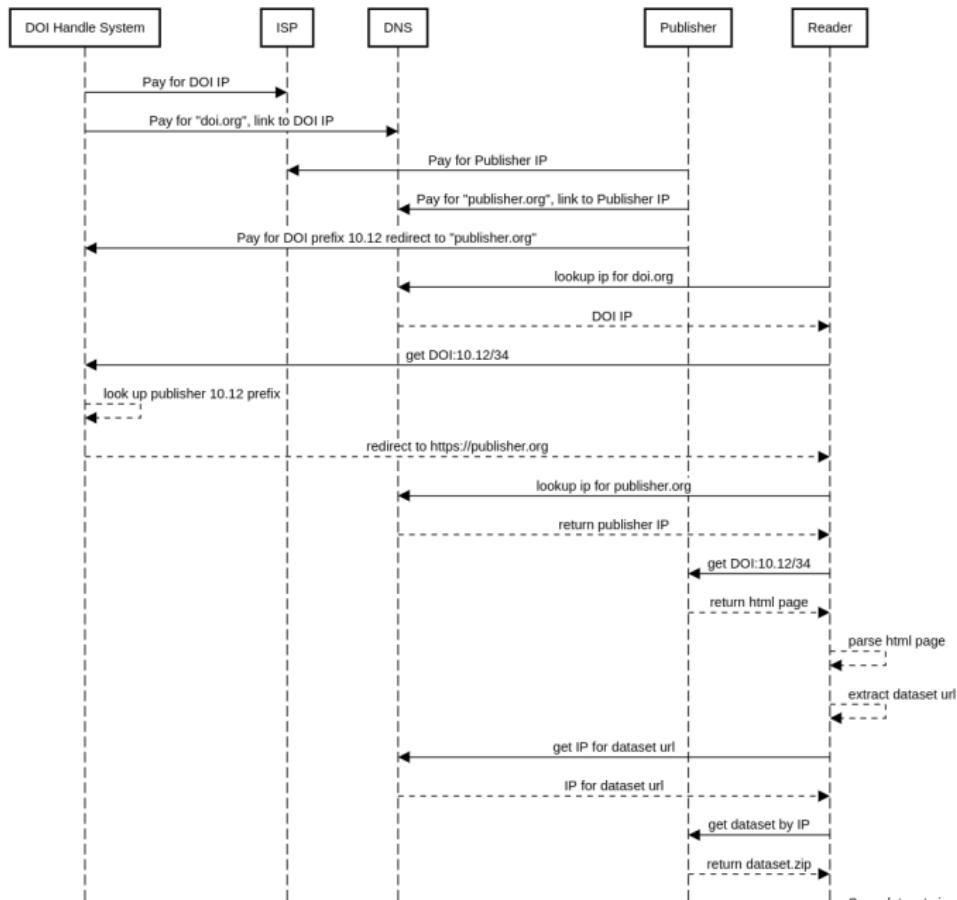
How To Retrieve This Bug Picture Now?

Take 2.

- ▶ turn on the internet (still there?)
- ▶ open a web browser
- ▶ load <https://doi.org/10.5072/zenodo.13342373>
 - ▶ and rely on a delicate and complex socio-technical network
- ▶ inspect page
- ▶ use mouse to click on a link that looks like a bug picture
- ▶ download image and **trust** its authenticity

DOI Economics and Redirection

steps to use DOI to download dataset



How To Retrieve This Cat Picture 40 Years From Now?

Museum für Naturkunde. (2024). *Photo of Specimen BMT0009388.* Zenodo.
<https://doi.org/10.5072/zenodo.13342373>

Likely will not work due to intricate network of dependencies.

How To Retrieve This Bug Picture 40 Years From Now?

Proposal: Sign the citation ¹

... by adding some digital fingerprints of the image.

Museum für Naturkunde. (2024). *Photo of Specimen BMT0009388.* Zenodo.

<https://doi.org/10.5072/zenodo.13342373>

hash://md5/f052cbac398bf340a620c421bfefa16d

¹Elliott M.J., Poelen, J.H. & Fortes, J.A.B. (2023) Signing data citations enables data verification and citation persistence. *Sci Data.*

<https://doi.org/10.1038/s41597-023-02230-y>

hash://sha256/f849c870565f608899f183ca261365dce9c9f1c5441b1c779e0db49df9c2a19d

How To Retrieve This Bug Picture 40 Years From Now?

... by searching the world (in, and beyond, the internet) for the content with the unique fingerprint

hash://md5/f052cbac398bf340a620c421bfefa16d .



<https://linker.bio/hash://md5/f052cbac398bf340a620c421bfefa16d>



https://zenodo.org/records/13342373/files/BMT121_BMT0009388



preston cat --remote <https://zenodo.org>
hash://md5/f052cbac398bf340a620c421bfefa16d



preston cat --remote <https://softwareheritage.org>
hash://sha256/69df9b5bc5df5619448ade01c92d0a5e20bcec2ab



preston cat --remote
<https://linker.bio>,<https://zenodo.org>
hash://md5/f052cbac398bf340a620c421bfefa16d

Let Us Carry Our Bug Pictures Into The Future...

preston cat\

```
--remote "https://zenodo.org"\  
hash://md5/f052cbac398bf340a620c421bfefa16d\  
> bug.tiff
```



Guiding Questions

- ▶ How do *you* cite data?
- ▶ How do you look up cited data *now*?
- ▶ How do you look up cited data *40 years from now*?

Extra

Internet Is Designed For *Exchanging* Information

The internet is a powerful tool for exchanging digital information. But the Internet's contents changes constantly: websites are launched and taken down, webpages change, and content gets archived or lost.²

²Jorrit Poelen. 2024. Unleashing Digital Knowledge Into The Future. Accessed on 2024-02-06 at <https://linker.bio>
line:hash://sha256/8ac18eb75ff20d40d1d60bb6ad5a745eb528093d1ffbe373e3847c131460

Internet Is *Location-based*

*By design, a web address, or Uniform Resource Locator (URL), points to a specific internet location from which a resource, like a webpage, can be retrieved. However, a URL does not provide a way to verify that a retrieved webpage was the one we asked for.*³

³Jorrit Poelen. 2024. Unleashing Digital Knowledge Into The Future.
Accessed on 2024-02-06 at <https://linker.bio>
line:hash://sha256/8ac18eb75ff20d40d1d60bb6ad5a745eb528093d1ffbe373e3847c131460

Finding Content By Their Location Is . . . Tricky

*Imagine using a URL-like reference to find a book at a library: instead of locating a book by what it is (e.g., title, author), you refer to a book by its location (e.g., third shelf on the second row next to the window). With this, a book becomes unfindable if moved to another shelf. And, if you do manage to find a book at the referenced location, how would you know you've found the book you are looking for?*⁴

⁴Jorrit Poelen. 2024. Unleashing Digital Knowledge Into The Future. Accessed on 2024-02-06 at <https://linker.bio>
line:hash://sha256/8ac18eb75ff20d40d1d60bb6ad5a745eb528093d1ffbe373e3847c131460

Finding Content By Their (Summarized) Content Is . . . What Librarians Do

Instead of pointing to where books are located, librarians point to them using a bibliographic reference. For practical reasons, only a few identifying clues are included in such a reference (e.g., author, year of publication, title, and publisher). So, librarians refer to content by what it is, and knowing where it may be located is secondary. ⁵

A bibliographic citation:

Darwin, C. 1859. On the Origin of Species. John Murray.

⁵ Jorrit Poelen. 2024. Unleashing Digital Knowledge Into The Future. Accessed on 2024-02-06 at <https://linker.bio>
line:hash://sha256/8ac18eb75ff20d40d1d60bb6ad5a745eb528093d1ffbe373e3847c131460

