

Bug Pictures Beyond The Internet

Jorrit Poelen <https://jhpoelen.nl>

2024-08-19

Guiding Questions

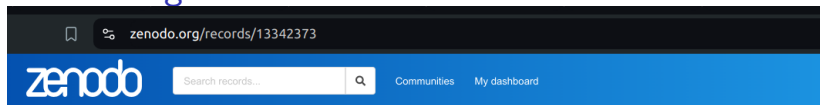
- ▶ How do *you* cite data?
- ▶ How do you look up cited data *now*?
- ▶ How do you look up cited data *40 years from now*?

There's A Bug In Zenodo...



Figure 1: Bug

There's A Bug in Zenodo...



Published August 19, 2024 | Version v1

Photo

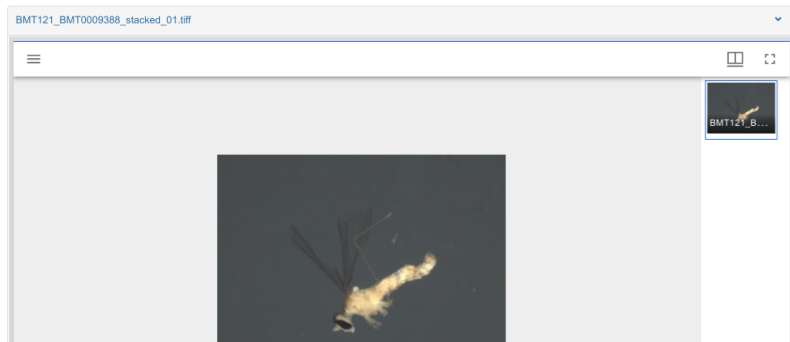
Open

Photo of Specimen BMT0009388

Museum für Naturkunde

Uploaded by Plazi for the Museum für Naturkunde Berlin.

Files



So How'd You Cite This Bug?

Museum für Naturkunde. (2024). Photo of Specimen BMT0009388. Zenodo.
<https://doi.org/10.5281/zenodo.13342373>

The screenshot shows the Zenodo landing page for the record 13342373. The browser address bar displays zenodo.org/records/13342373. The page layout includes a main content area on the left and a sidebar on the right.

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/f052cbac398bf340a620c421bfefa16d](https://md5/f052cbac398bf340a620c421bfefa16d) (URL)
- Is derived from**
Dataset:
<https://linker.bio/line:hash://sha256/4a4dccbbdc51e11f9ec2c878df263b84b3a995a67520d474e00301a9467abb23f/L> (URL)

Biodiversity

- Catalog number** [↗](#)
BMT0009388
- Institution code** [↗](#)
MIN

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/10.5281/zenodo.13342373>

Style APA

Figure 3: Bug Landing Page

How To Retrieve This Zenodo Bug Now?

Like this?

```
curl -L "https://doi.org/10.5281/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-
```

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

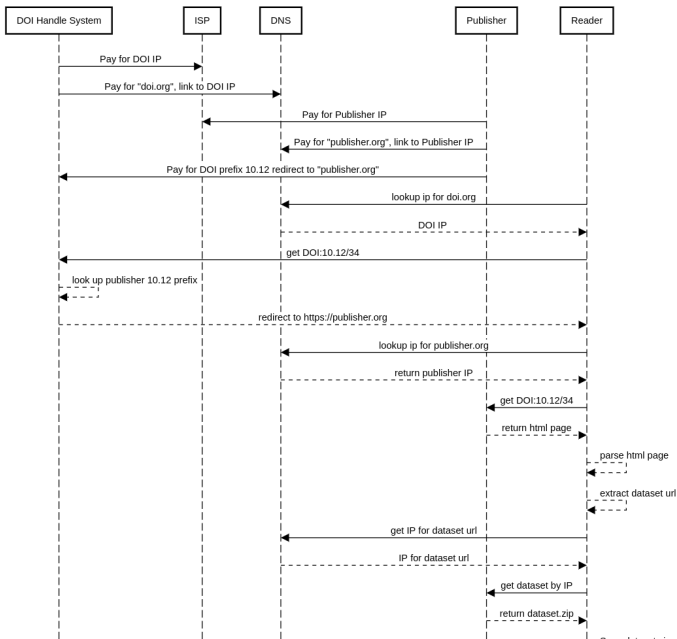
How To Retrieve This Bug Picture Now?

Take 2.

- ▶ turn on the internet (still there?)
- ▶ open a web browser
- ▶ load <https://doi.org/10.5281/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 Bug Picture 40 Years From Now?

*Museum für Naturkunde. (2024). Photo
of Specimen BMT0009388. Zenodo.
<https://doi.org/10.5281/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/10.5281/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

⁴ Jorrit 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

