Unleasing Digital Knowledge into the Future

(a working document)

Jorrit H. Poelen

2023-10-11

[edit this page](https://github.com/bio-guoda/preston-service/edit/main/linker.bio/index.md) / [share suggestions](https://github.com/bio-guoda/preston/issues/new)

## Welcome to linker.bio

The internet is a powerful tool to exchange digital information. But content available through the internet changes constantly: websites are launched and taken down, webpages change, and content gets archived or lost.

By design, a web address, or *U*niform *R*esource *L*ocator (URL), points to a specific internet *location* at which a resource, like webpage, can be retrieved. However, a URL does *not* provide a way to verify that a retrieved webpage was the one we asked for. [[1]](#footnote-22)

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 their 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 found the book you are looking for?

Instead of pointing to where a book is located, librarians point to a book using a bibliographic reference. For practical reason, 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.

A bibliographic citation:

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

Thanks to recent advances in mathematics [[2]](#footnote-23), we can add digital fingerprints to bibliographic citations of *digital* content. A digital fingerprint is of fixed size and uniquely describes any digital content (e.g., a webpage, an digital image). A fingerprint is generated by performing a calculation [[3]](#footnote-25) on the content itself. Citations that include a digital fingerprint are also referred to as signed citations [[4]](#footnote-26).

These digital fingerprints open up a way to *automatically* verify, with astronomical certainty, that you got the digital content that you asked for.

## Asking for What You Want

On the internet, we’ve learned to say:

“I’d like to get the latest content from this web address.”

, and *trust* that the retrieved content is what you asked for.

This may work well for a current news website or an internet search engine.

However, for retrieving specific content, like a newspaper article or research paper, we’d like to have a way of saying:

“I’d like to get the content with *this digital fingerprint*.”

, and *verify* that the retrieved content is *exactly* what you asked for.

In using digital fingerprints to refer to content, the location, or even the storage medium, of the content becomes secondary. In other words, *digital fingerprints help preserve references to digital content to a future beyond the internet*.

## The Content-Verse [[5]](#footnote-31)

This is where I’d like to introduce **the content-verse** as the collection of all possible digital content and their associated digital fingerprints. By definition, the content-verse contains all content (or knowledge) ever to be created.

## Using Digital Fingerprints

https://linker.bio builds a bridge [[6]](#footnote-33) from the exciting, dynamic internet to its reliable, boring, cousin — the **content-verse**. In this content-verse, **digital fingerprints are used as links** instead of resource locators (or URLs). Contrary to URLs, these digital fingerprints **do *not* break, nor expire** [[7]](#footnote-40).

Where the internet excels in spreading new information, the content-verse excels at referencing *known* information.

Through digital fingerprints, linker.bio provides a bridge to access billions of openly available biodiversity data records [[8]](#footnote-41), millions of Open Science publications through [Zenodo](https://zenodo.org), over eight hundred thousand datasets via [DataOne](https://dataone.org), billions of open source files [Software Heritage Library](https://softwareheritage.org), and more than ninety seven million freely usable media files from [WikiMedia Commons](https://commons.wikimedia.org/).

The beauty of digital fingerprints is that in fifty years from now, you may use that same fingerprint to find that information, regardless where it may be located, or how it is stored, or transmitted [[9]](#footnote-52).

## How to Request Content

https://linker.bio/ helps to request information, wherever it may be, using a notation like:

https://linker.bio/[fingerprint][.extension]

The extension is optional.

## Examples

For instance, to get a copy of a **scientific paper**, you can ask for:

[https://linker.bio/hash://sha256/f849c870565f608899f183ca261365dce9c9f1c5441b1c779e0db49df9c2a19d.pdf](https://linker.bio/hash://sha256/f849c870565f608899f183ca261365dce9c9f1c5441b1c779e0db49df9c2a19d).

or, to get a copy of a **scientific dataset**, like a historical CO2 Record from the Vostok Ice Core, you can ask for:

<https://linker.bio/hash://md5/e27c99a7f701dab97b7d09c467acf468>

or, perhaps even better, you can also ask for a **picture of a 🐇** (*Oryctolagus cuniculus*) by JM Ligero Loarte -

[https://linker.bio/hash://sha1/86fa30f32d9c557ea5d2a768e9c3595d3abb17a2.jpg](https://linker.bio/hash://sha1/86fa30f32d9c557ea5d2a768e9c3595d3abb17a2).

or, to review an initial draft of **the Hash URI Specification by Ben Trask** -

<https://linker.bio/hash://sha256/3fee21854fb6d81573b166c833db2771b21f0c77daa3095aab542764d89c94c1>

or, if you are ambitious, you can retrieve a digital corpus containing billions of biodiversity records [[10]](#footnote-59) via:

<https://linker.bio/hash://sha256/a755a6ac881e977bc32f11536672bfb347cf1b7657446a8a699abb639de59419>.

For more information and background, see:

Elliott, M.J., Poelen, J.H. & Fortes, J.A.B. Signing data citations enables data verification and citation persistence. Sci Data 10, 419 (2023). <https://doi.org/10.1038/s41597-023-02230-y> [hash://sha256/f849c870565f608899f183ca261365dce9c9f1c5441b1c779e0db49df9c2a19d](https://linker.bio/hash://sha256/f849c870565f608899f183ca261365dce9c9f1c5441b1c779e0db49df9c2a19d.pdf)

## Acknowledgments

Michael Elliott, José Fortes and [Cypress Hansen](https://www.cypresswritesscience.com/) provided comments to help improve a description of todays internet and the benefits of the content-verse.

1. A similar argument can made for a *D*igital *O*bject *I*dentifier (DOI, https://en.wikipedia.org/wiki/Digital\_object\_identifier). [↑](#footnote-ref-22)
2. Sobti, R. & Geetha, G. Cryptographic Hash Functions: A Review. 2012. International Journal of Computer Science Issues (IJCSI) 9, 461–479 <https://www.ijcsi.org/papers/IJCSI-9-2-2-461-479.pdf> accessed at 2023-10-11 with hash://md5/eb8e2fb3e16bd5839443cd40a9a8c3c1 [↑](#footnote-ref-23)
3. Such calculations are referred to as cryptographic hash functions like MD5, or SHA-256 [↑](#footnote-ref-25)
4. Elliott, M.J., Poelen, J.H. & Fortes, J.A.B. Signing data citations enables data verification and citation persistence. Sci Data 10, 419 (2023). <https://doi.org/10.1038/s41597-023-02230-y> [hash://sha256/f849c870565f608899f183ca261365dce9c9f1c5441b1c779e0db49df9c2a19d](https://linker.bio/hash://sha256/f849c870565f608899f183ca261365dce9c9f1c5441b1c779e0db49df9c2a19d.pdf) [↑](#footnote-ref-26)
5. Probably need a better term for this, because of existing uses elsewhere. [↑](#footnote-ref-31)
6. linker.bio is not the only bridge to the content-verse. In fact, linker.bio re-uses existing bridges provided by [Zenodo](https://github.com/bio-guoda/preston/issues/149), [DataOne](https://github.com/bio-guoda/preston/issues/181), [WikiMedia Commons](https://github.com/bio-guoda/preston/issues/239), and [Software Heritage Library](https://github.com/bio-guoda/preston/issues/70) to the massive amount of content they keep. Note also that [Carl Boettiger](https://www.carlboettiger.info/) maintains a bridge from R to the content-verse via the R package [contentid](https://github.com/cboettig/contentid). [↑](#footnote-ref-33)
7. Digital finger prints are cryptographic objects that are mathematically linked to the content they reference. They can be generated with most digital devices. And, by embedding fingerprints into other digital content, you can pretty much reference anything and everything digital using a digital fingerprint that fits on a T-shirt. [↑](#footnote-ref-40)
8. Biodiversity records include snapshot version of digital collections registered with [iDigBio](https://idigbio.org), [GBIF](https://gbif.org), [BioCase](https://biocase.org), [Biodiversity Heritage Library](https://biodiversitylibrary.org), [OBIS](https://obis.org) and [CheckListbank](https://checklistbank.org) [↑](#footnote-ref-41)
9. In other words, digital fingerprints are agnostic of location, technology, and … time. [↑](#footnote-ref-52)
10. To compile this corpus from their referenced parts, you may benefit from using a tool like [Preston](https://github.com/bio-guoda/preston). [↑](#footnote-ref-59)