

Decentralized Science – Open, Reproducible and FAIR

Prof. Dr. Philipp Koellinger



DeSci
Foundation



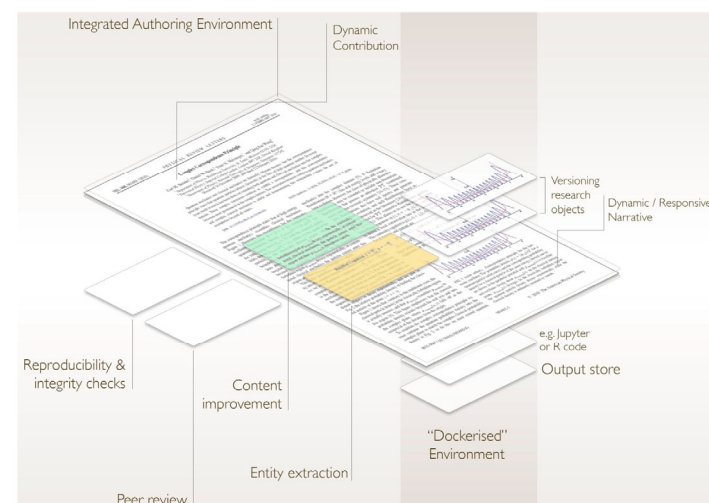
DeSci Labs

The problem

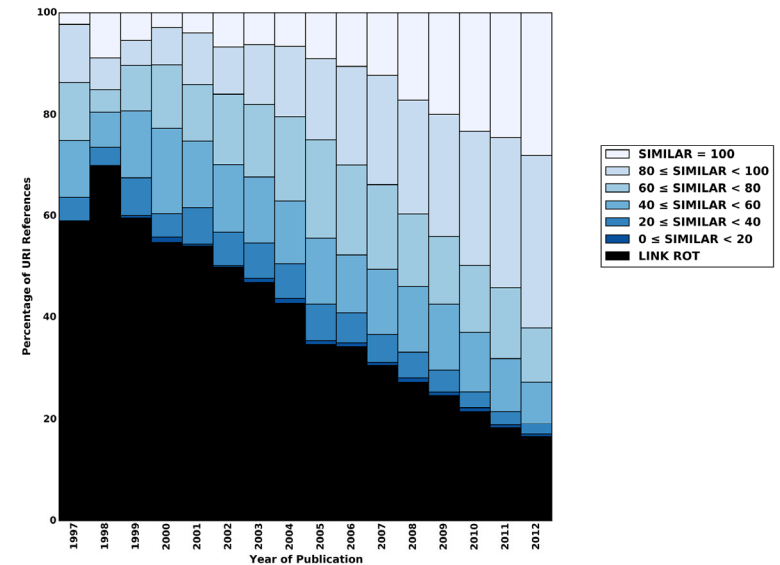
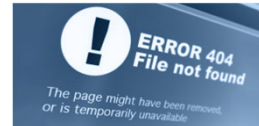
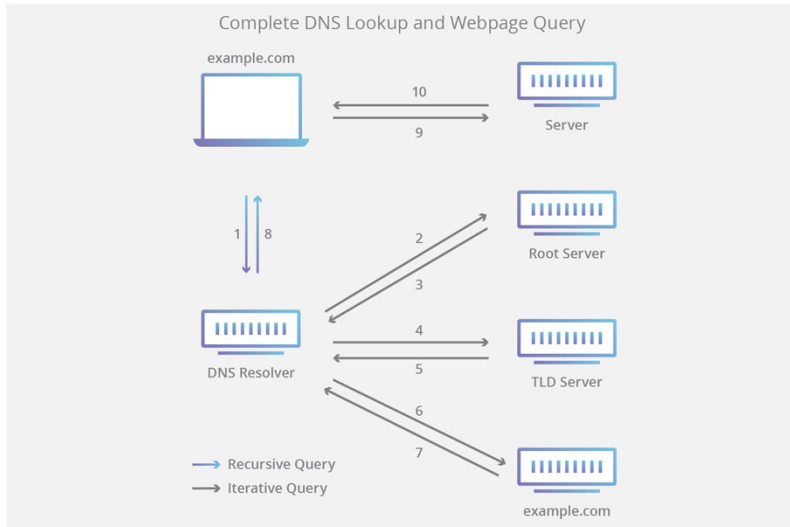
- Outdated, rigid publication infrastructure
- Reproducibility and replicability crisis
 - Decreasing trust in science
- Inaccessible data and code
 - Lost, valuable resources
 - >€10 billion costs due to not having Findable, Accessible, Interoperable & Reusable (FAIR) data
 - European Commission: Cost of not having FAIR data (2018)

The solution

- Interactive and versionable FAIR digital research objects
- Open-state repository
 - No data silos
 - No vendor lock-in
 - Institutional sovereignty
 - Content owned by creators
- Persistent identifiers as irrevocable APIs ([DPID.org](https://dpid.org))
- Pluralistic attestations, e.g.
 - Data & code available
 - Results reproducible
- Affordable data storage
- Fork & merge for composable authorship, e.g.
 - Data stewards
 - Other researchers



Identifiers in the current Internet – URIs



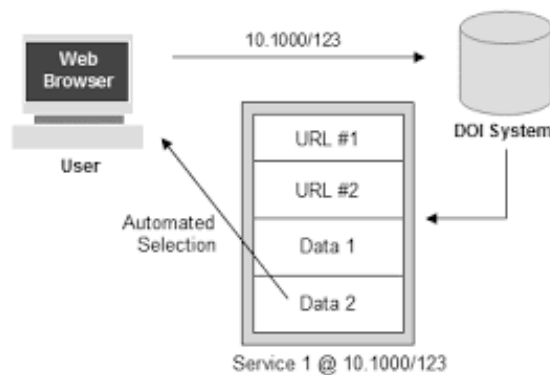
- URLs point to where content is stored, not what the content is
- Link rot (file moved or deleted, 404 error)
- Content drift (content changes over time)
- No version control

- URI citations with link rot or content drift by year of publication, Elsevier corpus (Jones et al. 2016)
- A threat to the integrity and value of the scientific record

Sources:

Jones, S.M., et. al. (2016). Scholarly context adrift: Three out of four URI References Lead to Changed Content. *PLoS ONE* 11(12): e0167475.

DOIs are not persistent identifiers



- DOIs do not correctly resolve to their target resource in ~50% of all cases (Klein & Balakireva 2020)
- Different results for same DOI depending on request method and network environment
- DOIs are matched to URLs in a database
→ Lots of manual updating work for publishers
- Costly, inefficient system for publishers
- DOIs are neither persistent nor unique identifiers

Sources:

Klein, M., Balakireva, L. (2020). On the Persistence of Persistent Identifiers of the Scholarly Web. In: Hall, M., Merčun, T., Risse, T., Duchateau, F. (eds) *Digital Libraries for Open Knowledge*. TPD 2020. Lecture Notes in Computer Science, vol. 12246. Springer.

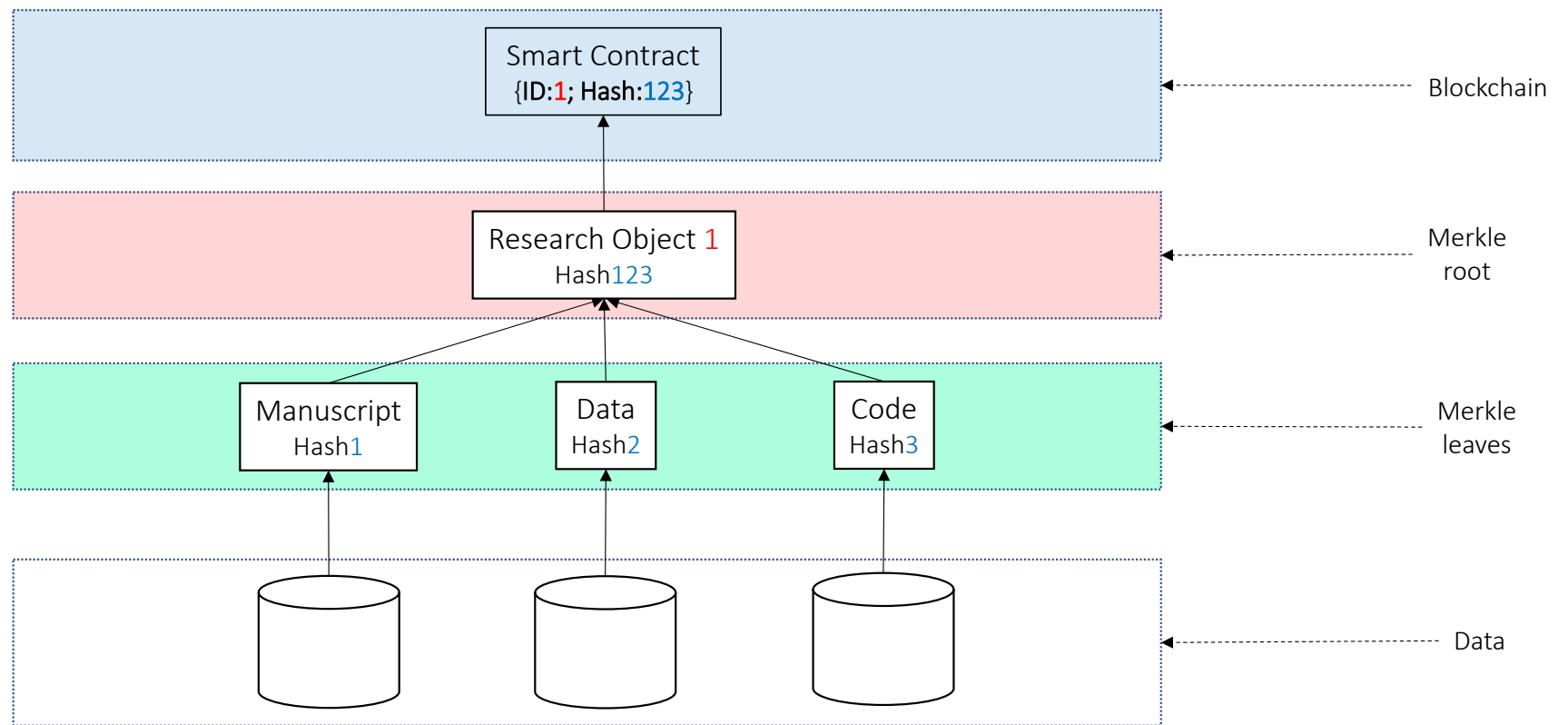
Decentralized content addressing based on hashes

- A cryptographic hash function converts a string of arbitrary length into string of fixed length
 - One-way mathematical function
 - E.g., the SHA-256 algorithm creates a 64 hexadecimal string for any input
 - Changing *anything* in the input (i.e. a word, pixel, comma) will yield a different hash
 - Hashes are *unique*
 - E.g., SHA-256 allows creating 10^{77} different hashes – billions of times more than the number of atoms on Earth
 - For example, SHA-256 hashes:
 - “PSV will win the Champions League 2024” → c554644c23a1ebdcf29f9c7af8492fac3a9964b0fbd0231f66a469580d09cfc9
 - “Feyenoord will win the Champions League 2024” → 24e4996dc8601efcd676e8f94fd312776c3f648cfd0d30a380121e240660ec5e
- Content-addressed storage is immune to content drift and link rot!
 - Interplanetary File System (<https://ipfs.tech/>)

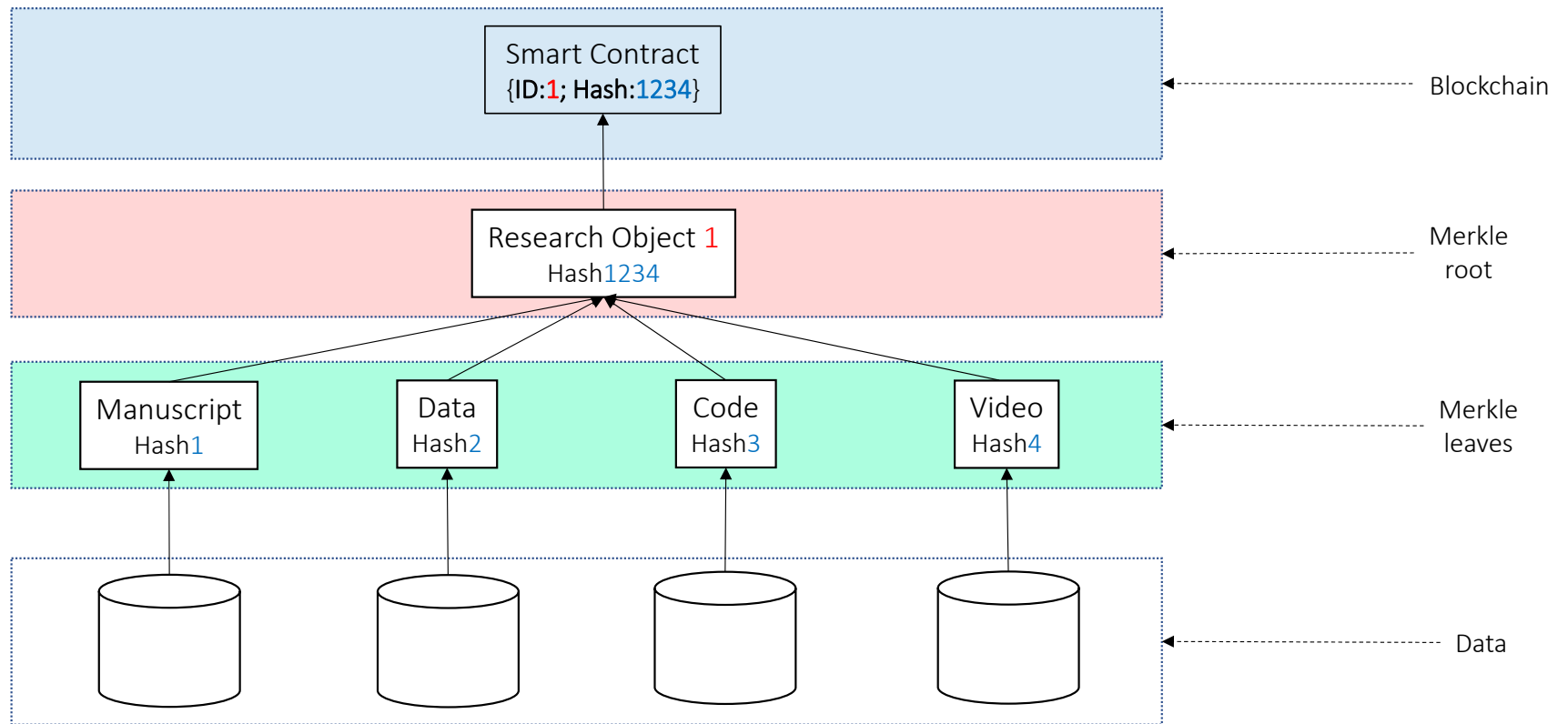
Sources:

<https://xorbin.com/tools/sha256-hash-calculator>

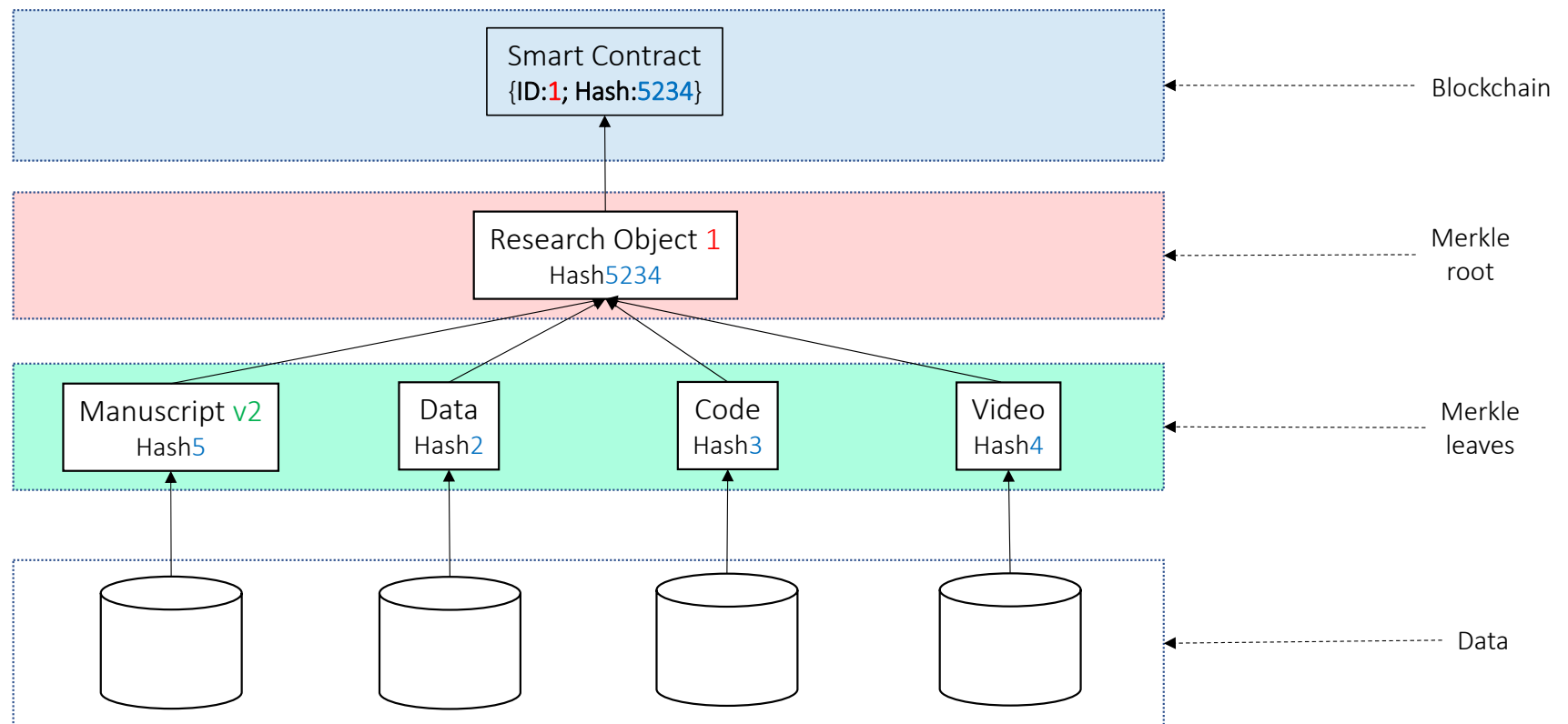
DeSci Nodes: Research objects with hash-PIDs, indexed on a blockchain



Adding a new component to the research object

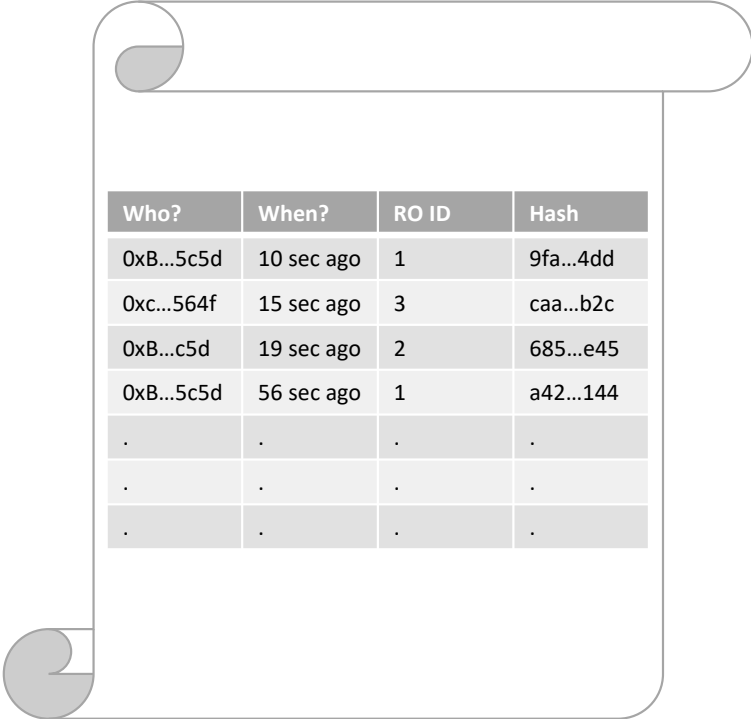


Updating a component of the research object



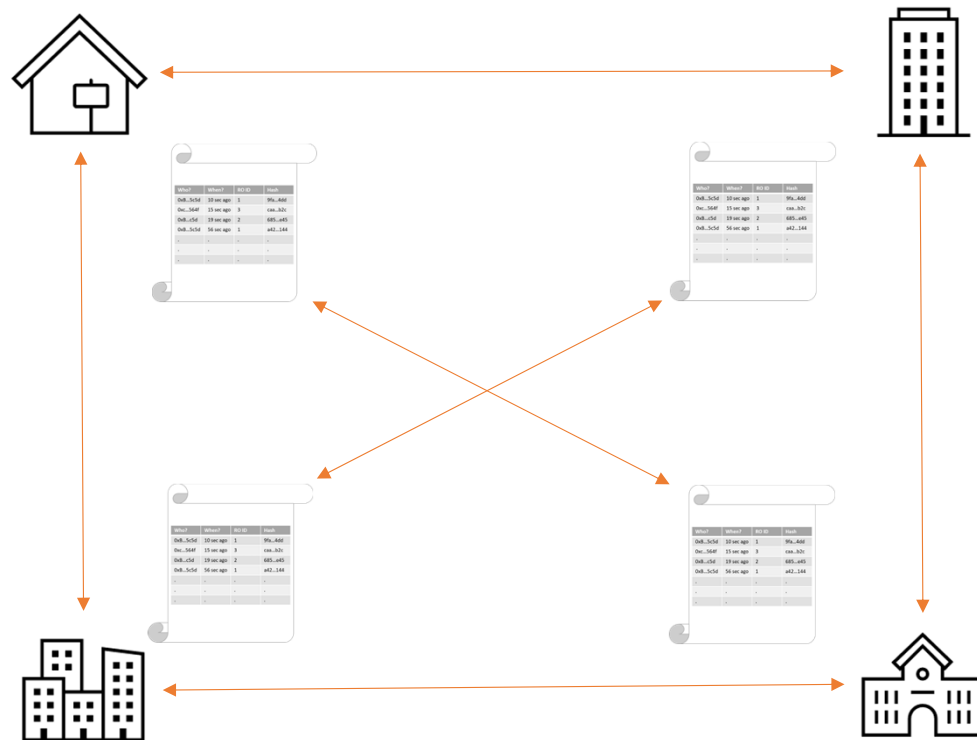
→ The decentralized web allows us to upgrade the scientific record from static manuscripts *without* persistent IDs *or* version control to rich, dynamic, interoperable research objects *with* persistent IDs *and* version control!

Provenance: A public ledger of research objects...



Who?	When?	RO ID	Hash
0xB...5c5d	10 sec ago	1	9fa...4dd
0xc...564f	15 sec ago	3	caa...b2c
0xB...c5d	19 sec ago	2	685...e45
0xB...5c5d	56 sec ago	1	a42...144
.	.	.	.
.	.	.	.
.	.	.	.

...distributed across many servers



DeSci protocol

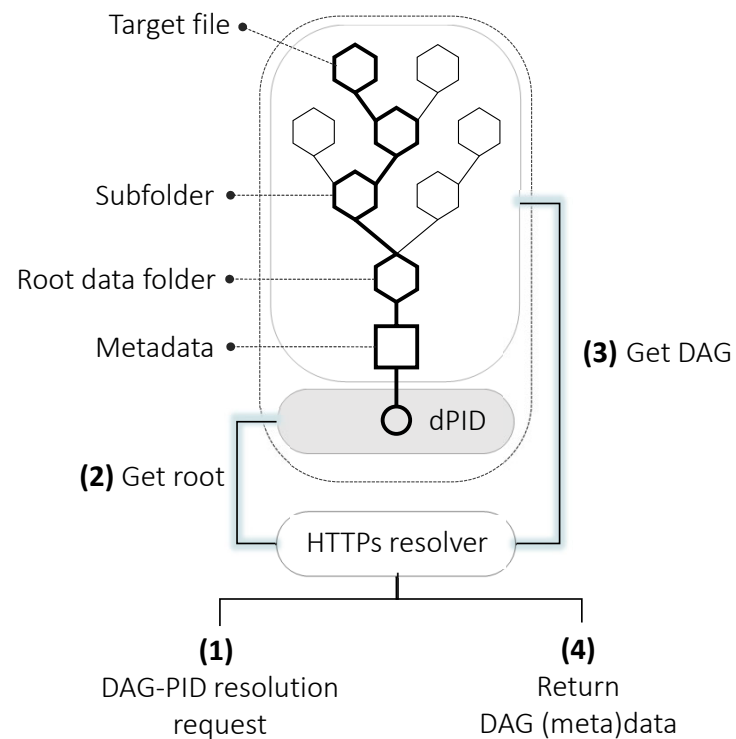
...resolution to (meta)data over HTTP

RESOLVER/[UUID or short PID]/[ROOT_HASH or version]/[Component ID]/[JSON_PATH][optional method suffix]

COMPACT PID TO DATA: dpid.org/42/1/2/1

COMPACT PID TO METADATA: dpid.org/42/1/2/1?jsonld

Open state resolution to (meta)data Over HTTP gateway



Open state read and write
(Irrevocable API)

DeSci protocol

...resolution to (meta)data over HTTP

RESOLVER/[UUID or short PID]/[ROOT_HASH or version]/[Component ID]/[JSON_PATH][optional method suffix]

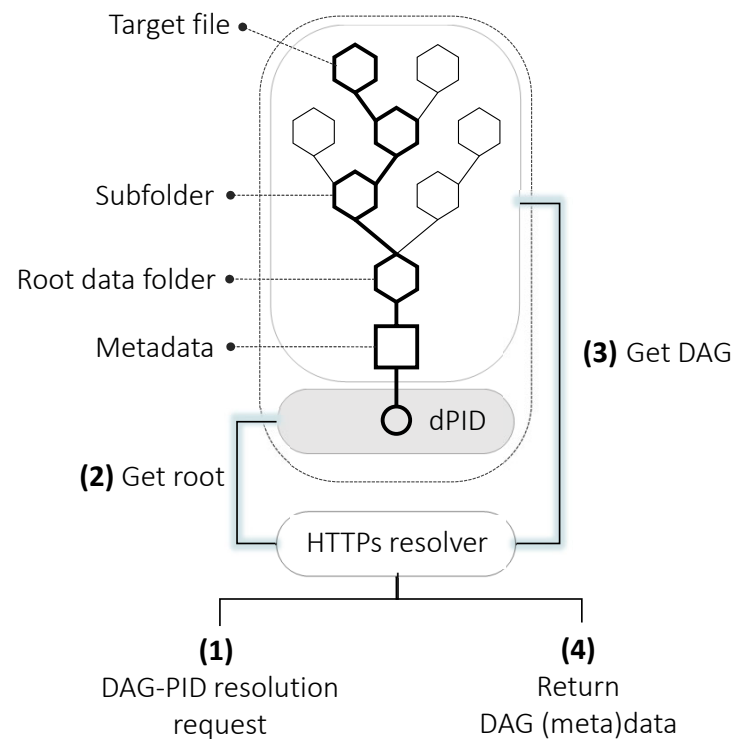
COMPACT PID TO DATA: dpid.org/42/1/2/1

COMPACT PID TO METADATA: dpid.org/42/1/2/1?jsonld

...as addresses for edge compute jobs

```
bacalhau docker run \  
-v PID #PID of the digital object \  
-- magick mogrify -resize 100x100 -quality 100 #program to execute
```

Open state resolution to (meta)data Over HTTP gateway



Open state read and write
(Irrevocable API)

DeSci protocol

...resolution to (meta)data over HTTP

RESOLVER/[UUID or short PID]/[ROOT_HASH or version]/[Component ID]/[JSON_PATH][optional method suffix]

COMPACT PID TO DATA: dpid.org/42/1/2/1

COMPACT PID TO METADATA: dpid.org/42/1/2/1?jsonId

...as addresses for edge compute jobs

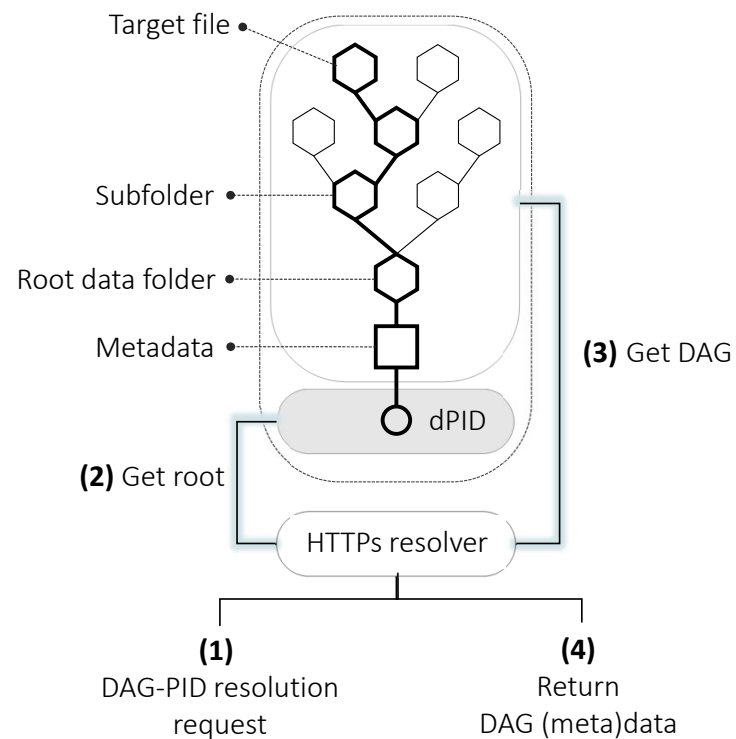
```
bacalhau docker run \  
-v PID #PID of the digital object \  
-- magick mogrify -resize 100x100 -quality 100 #program to execute
```

...as importable resources

```
Import desc as desc
```

```
with desc.fetch \  
import ("42/v1/data/measurement.csv") #import from PID
```

Open state resolution to (meta)data Over HTTP gateway



Open state read and write
(Irrevocable API)

DeSci protocol

...resolution to (meta)data over HTTP

`RESOLVER/[UUID or short PID]/[ROOT_HASH or version]/[Component ID]/[JSON_PATH][optional method suffix]`

COMPACT PID TO DATA: dpid.org/42/1/2/1

COMPACT PID TO METADATA: dpid.org/42/1/2/1?jsonId

...as addresses for edge compute jobs

```
bacalhau docker run \  
-v PID #PID of the digital object \  
-- magick mogrify -resize 100x100 -quality 100 #program to execute
```

...as importable resources

```
Import desc as desc
```

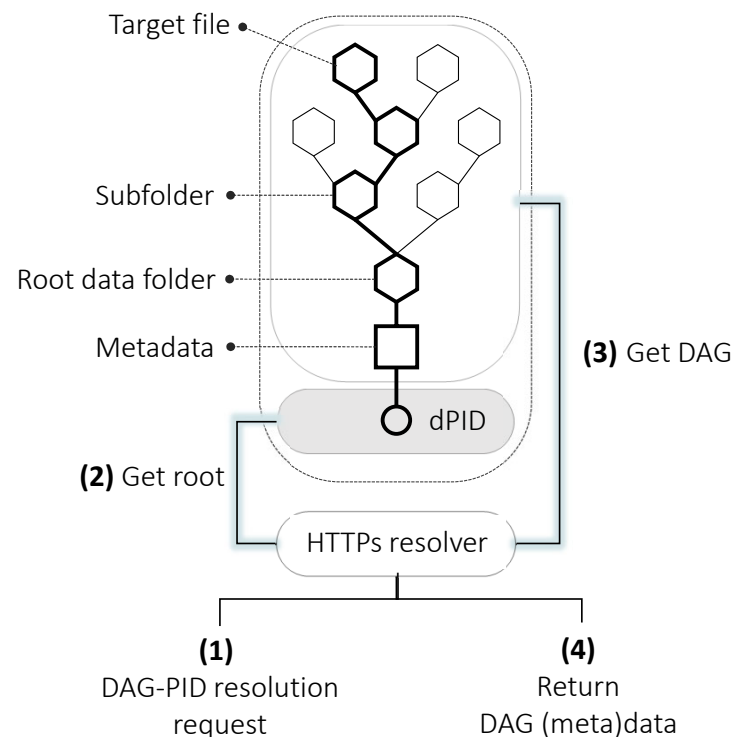
```
with desc.fetch \  
import ("42/v1/data/measurement.csv") #import from PID
```

...as targets for attestations

```
struct Attestation{
```

```
// Generate a PID for the attestation \  
bytes32 uid
```

Open state resolution to (meta)data Over HTTP gateway



Open state read and write
(Irrevocable API)

DeSci Nodes application

- Currently in open beta testing
 - Create and share an interactive, versionable FAIR digital research object
 - Enable reproducibility
 - DPID Fetch (import data and code directly into your local environment)
 - Compute over data with Bacalhau (<https://www.bacalhau.org/>)
 - Free (up to 50GB)
- Major upgrade release in November
 - Improved user experience
 - No wallet needed
 - ORCID integration
 - New website
- Next year and beyond
 - Attestations & commentaries
 - Text-editor integration
 - Fork & merge
 - In-browser compute
 - Automatic meta-data suggestions (e.g. authors and affiliations, keywords for each file)
 - Affordable, flexible data storage
 - Targeted solutions for publishers, academic societies, universities, and funding agencies

DeSci Nodes demo

The screenshot displays the DeSci Nodes web application. The left sidebar contains a file explorer for the project 'EXPLORING LUPUS CLOUDS' (M. ANSDELL ET AL. · DPID 46). It includes tabs for 'Node Drive', 'History', and 'Contributors'. Under the 'Pinned' section, there are links to 'Read The Paper' (exploring-lupus-report.pdf), 'Explore The Code' (exploring-lupus), 'Browse The Data' (data), and 'Go To Publisher Site'. A 'Home' section lists components: 'exploring-lupus-report.pdf' (3 MB), 'exploring-lupus' (873 MB), and 'External Links'. At the bottom of the sidebar, it shows 'ANONYMOUS' and a 'Login or Sign Up' button.

The main area on the right displays the document 'exploring-lupus-report.pdf'. The document features a large, colorful image of a protoplanetary disk (proplyd disk) with a bright blue center and swirling orange and red outer regions. Below the image, the title 'Exploring Lupus Clouds' is prominently displayed. Underneath the title, the subtitle 'Alma Survey of Lupus Protoplanetary Disks I: Dust and Gas Masses' is shown, followed by the year '2016' and the authors 'M. Ansdell et al.'. A 'View Source' button is located at the bottom right of the document content area. At the very bottom of the page, a small text line reads: 'Re-submitted to A&J: 5 May 2016. Preprint format using [B3]X style-emulation v. 12/16/11'.

DPID Fetch

The screenshot shows the GitHub repository page for `descilabs/dpid-fetch`. The repository is public and has 6 commits. The commit history shows that `kadamidev` added the `toCid` function for a `dpid` path. The repository contains files such as `desci`, `.gitignore`, `LICENSE`, `README.md`, `example-run.py`, and `setup.py`. The `README.md` file is displayed, showing the project name `dpid-fetch` and the description: "Resolve data and code from dpid.org into a Python environment." The repository also includes a code block with the following commands:

```
git clone https://github.com/desci-labs/dpid-fetch
pip install ./dpid-fetch
```

Below the code block, there is a link to view `example-run.py` for an example. The right sidebar shows the repository's metadata, including the MIT license, 0 stars, 1 watching, 0 forks, and 0 releases. The contributors list includes `hubsmoke` and `kadamidev`. The languages section shows that the repository is 100% Python.

<https://github.com/desci-labs/dpid-fetch>

Recommendations

- Try to use open-source stats software and data formats (e.g. R, Jupyter, Python)
- One code file for each key result (e.g. figure, table)
 - Use DPIDs from the Nodes application to address data files
 - Annotate your code
- Make use of versionability, e.g.
 - Start with analysis plan
 - Add data and code
 - Add manuscript and connect results to data & code
- Check Dryad recommendations for sharing human subject data
 - <https://datadryad.org/docs/HumanSubjectsData.pdf>
- Check out DeSci Nodes user documentation
 - <https://docs.desci.com/using-nodes/getting-started>

Call to action

- Participate in a meta-science field experiment!
 - Does fully reproducible research have a better chance of getting published?
 - Rewards for participants
 - Reach out to me
- Try out Nodes and make your work reproducible!
 - Any feedback and suggestions are VERY welcome
 - Reach out to me or use the “Feedback” button in the app
- Ask the DeSci Labs team for help to create reproducible research!
 - Reach out to me

Prof. Dr. Philipp Koellinger

p.d.koellinger@vu.nl
philipp@desci.com

DeSci Nodes

nodes.desci.com

Future of Science Seminar & Podcast

<https://descifoundation.org/seminar>



DeSci
Foundation



DeSci Labs



@DesciLabs
@DesciFoundation
@PKoellinger