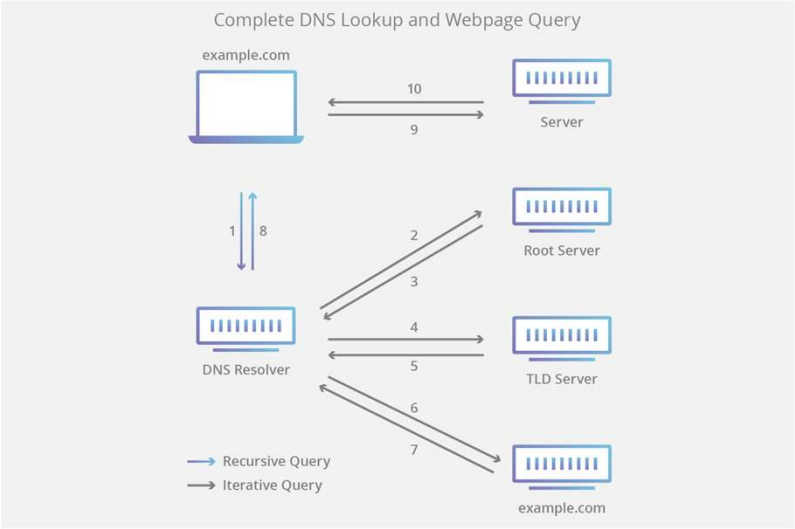
Technological opportunities

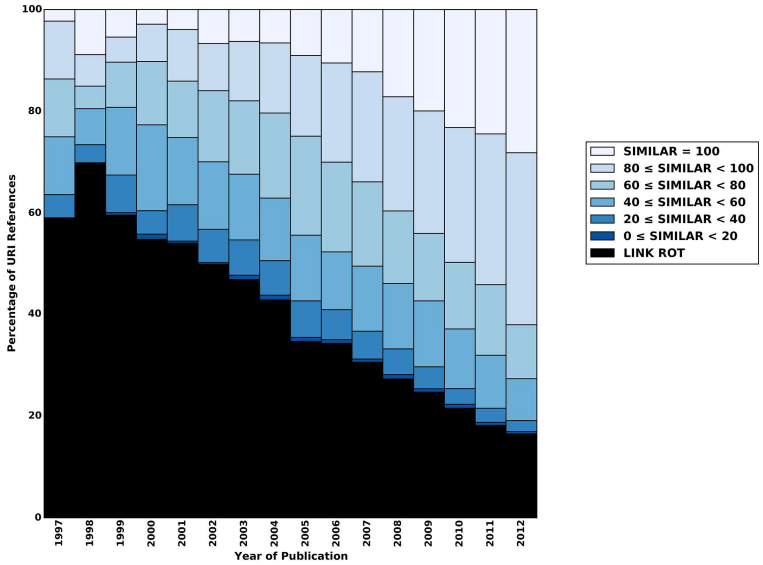
Prof. Dr. Philipp Koellinger

DeSci Labs | DeSci Foundation | Vrije Universiteit Amsterdam (Economics)

URIs

Current Internet





- 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

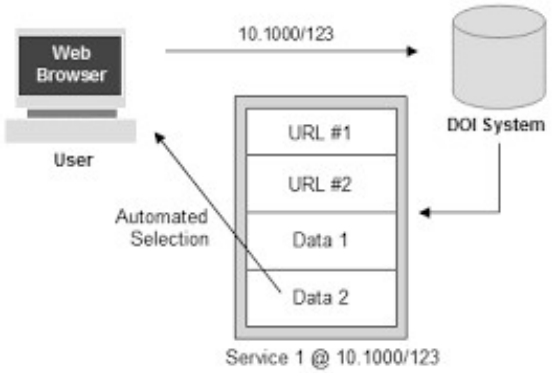
- URI citations with link rot or content drift
  by year of publication, Elsevier corpus (Jones et al. 2016)
- A threat to the integrity and value of the scientific record

Source:

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

Current identifiers



- **DOIs do not correctly resolve to their target resource in ~50% of all cases** (Klein & Balakireva 2020)
- Different results for the same DOI depending on the request method and network environment
- DOIs are matched to URLs in a database
  → Lots of manual updating work
- Costly, inefficient system
- **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*. TPDL 2020. Lecture Notes in Computer Science, vol. 12246. Springer.

Content addressing

Cryptographic hashes

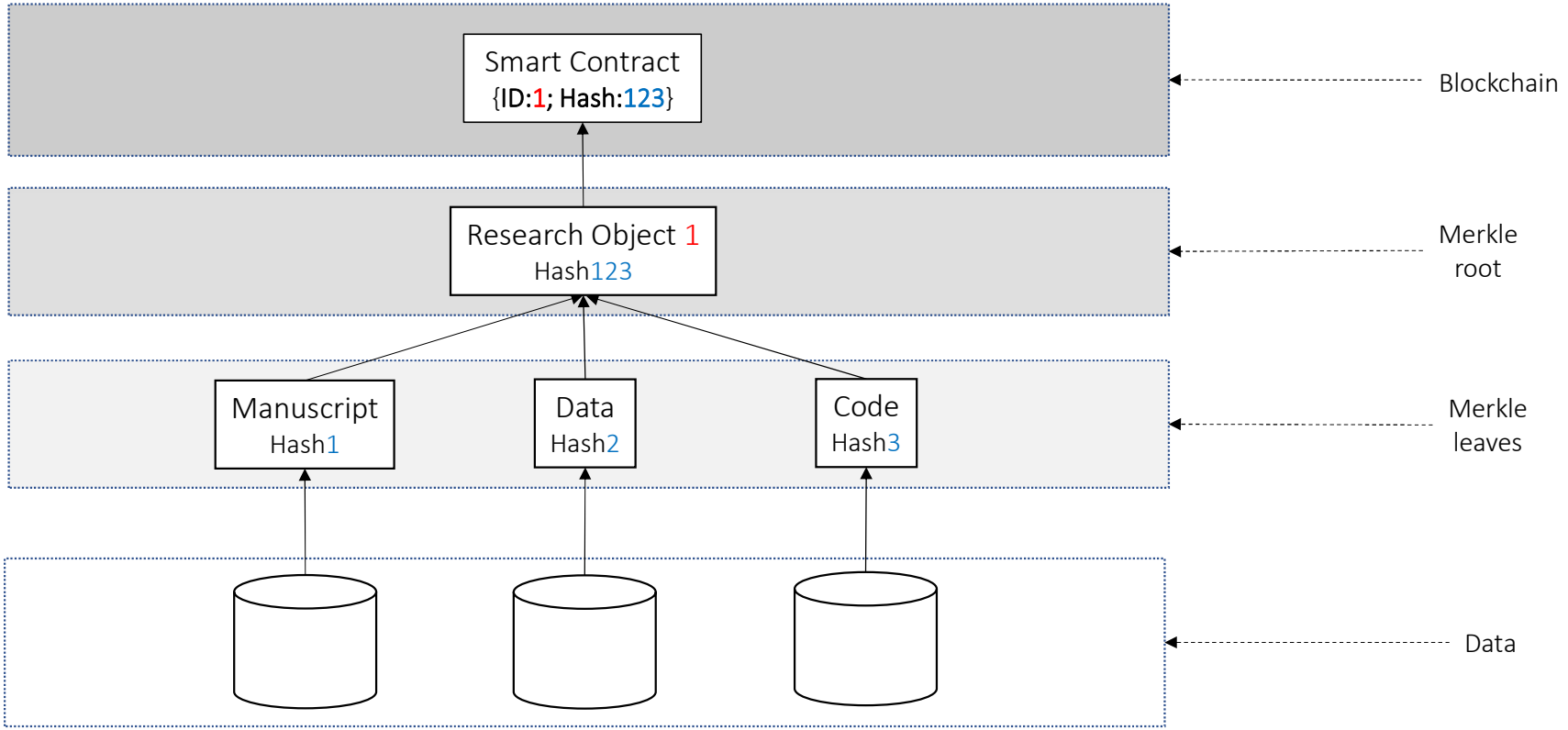
- A cryptographic hash function converts a string of arbitrary length into string of fixed length
  - One-way mathematical function
    - E.g., the SHA-256 algorithm creates a 64 hexadecimal string for any input
  - Changing *anything* in the input (i.e. a word, pixel, comma) will yield a different hash
  - 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:
    - "Chiefs will win the Super Bowl 2024" → 4bb86fab0cb40d173f306ed0570f2cfe083aa02d8e329c35cf57c25d2c431731
    - "Eagles will win the Super Bowl 2024" → 08458fc00e86f9bd2fb8a624e25e51560491b04194878c6a67aaa015f220e6f0

→ Content addressing based on cryptographic hashes is a powerful safeguard against link rot and immune to content drift!

Sources:
https://xorbin.com/tools/sha256-hash-calculator

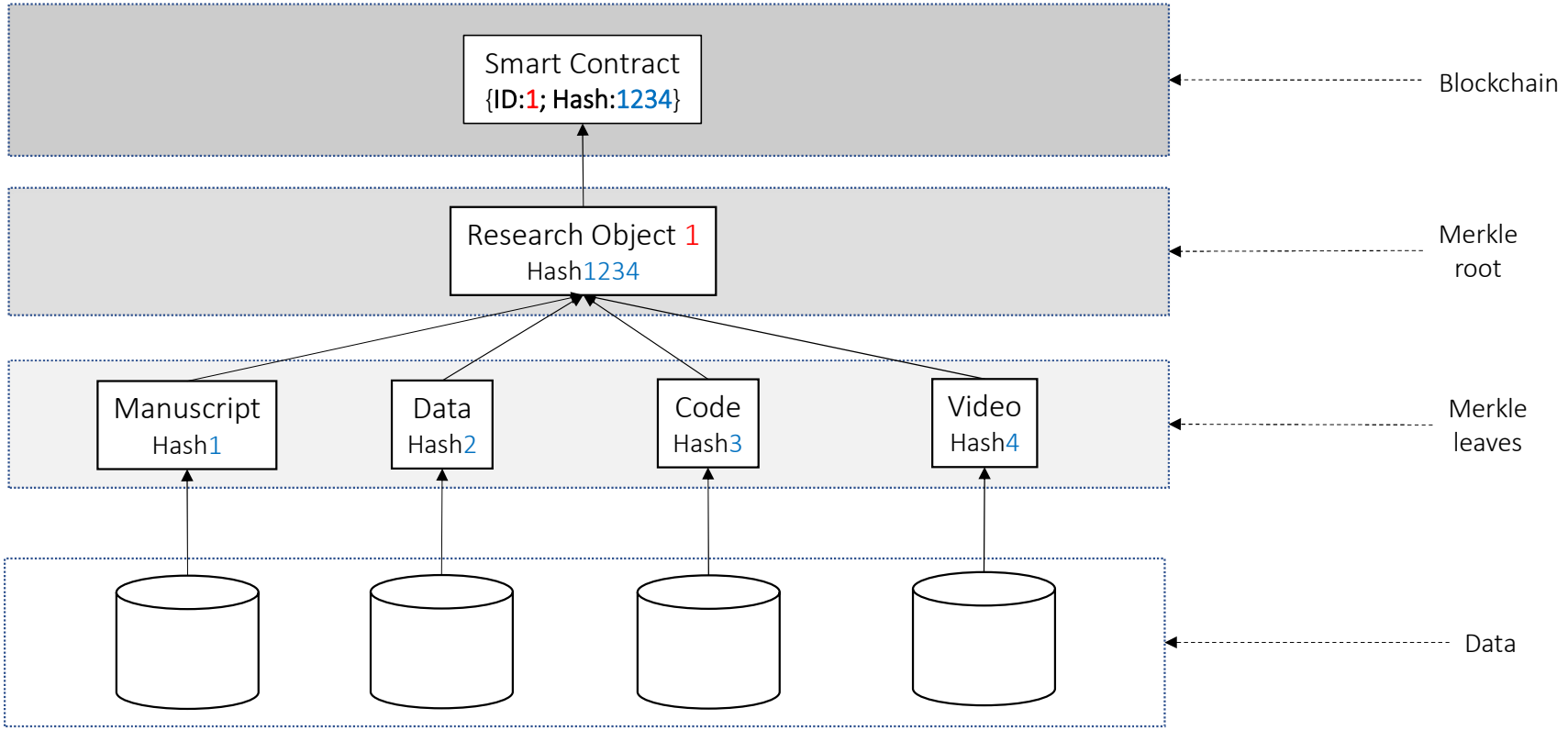
DeSci Nodes

Research Objects



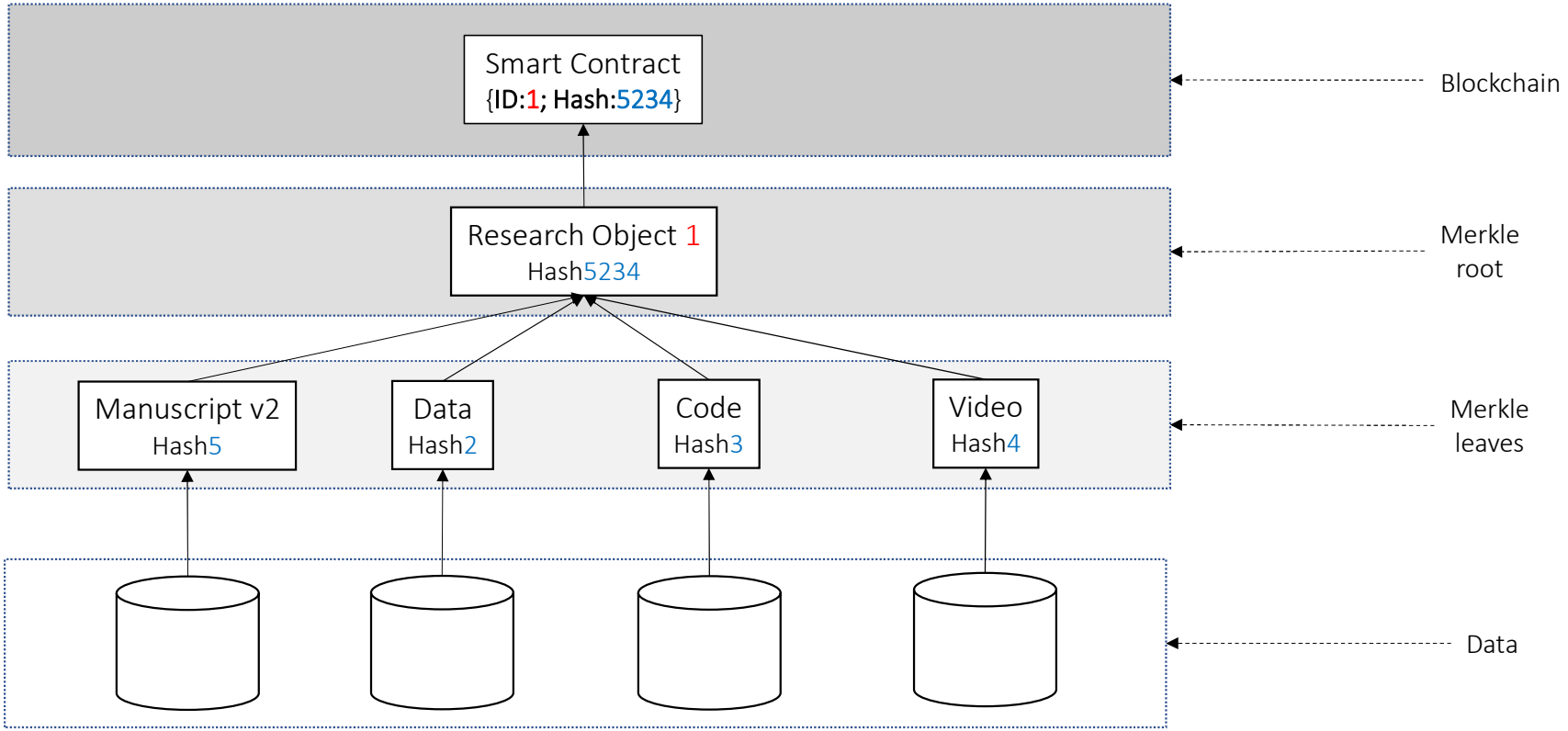
DeSci Nodes

Adding new component



DeSci Nodes

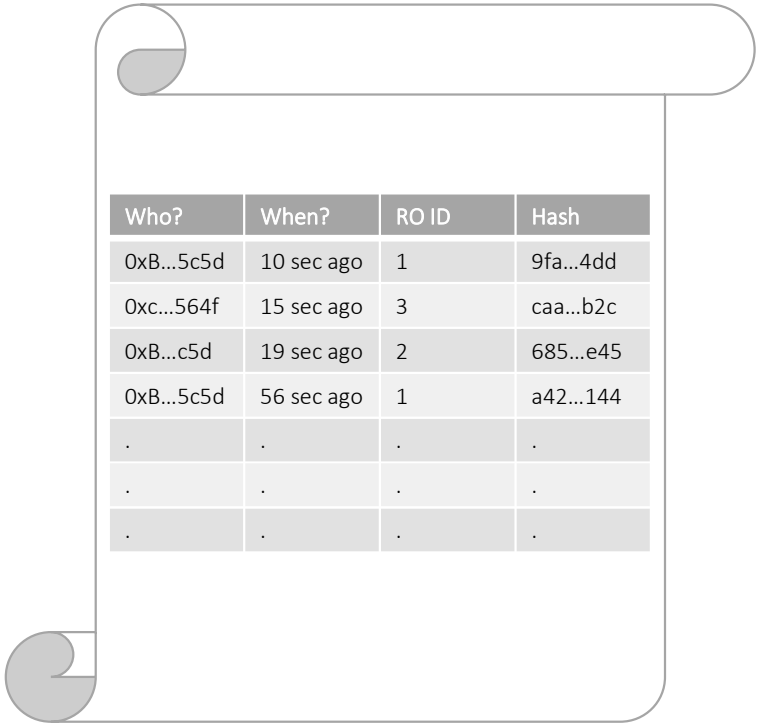
Updating component



→ 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!

DeSci Nodes

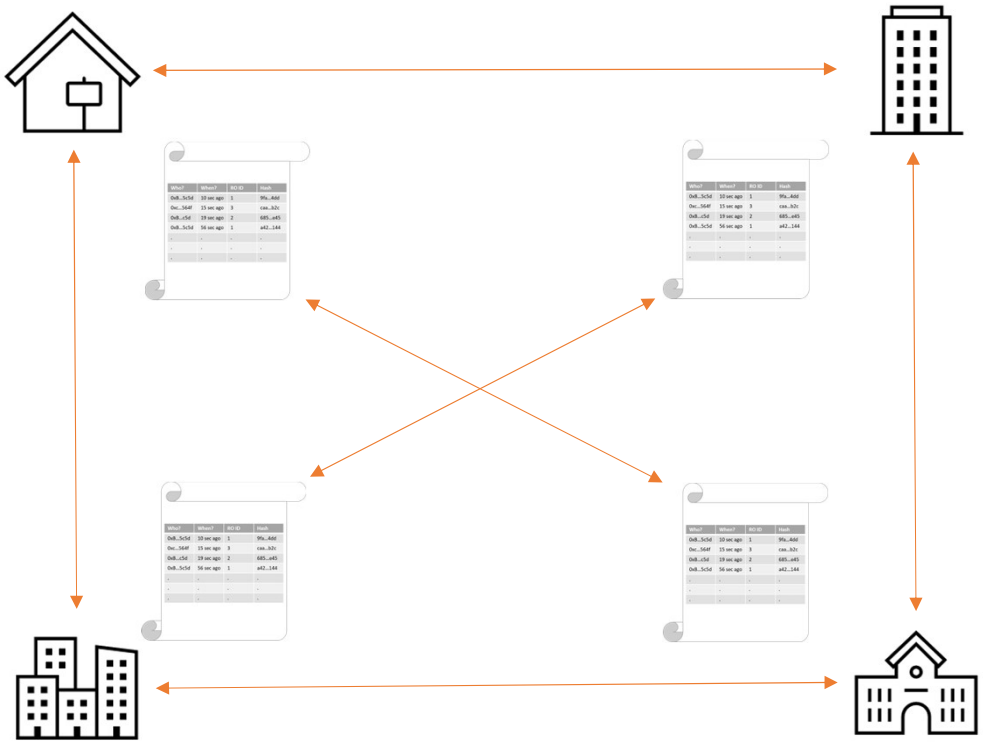
Public Registry of Research



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
.	.	.	.
.	.	.	.
.	.	.	.

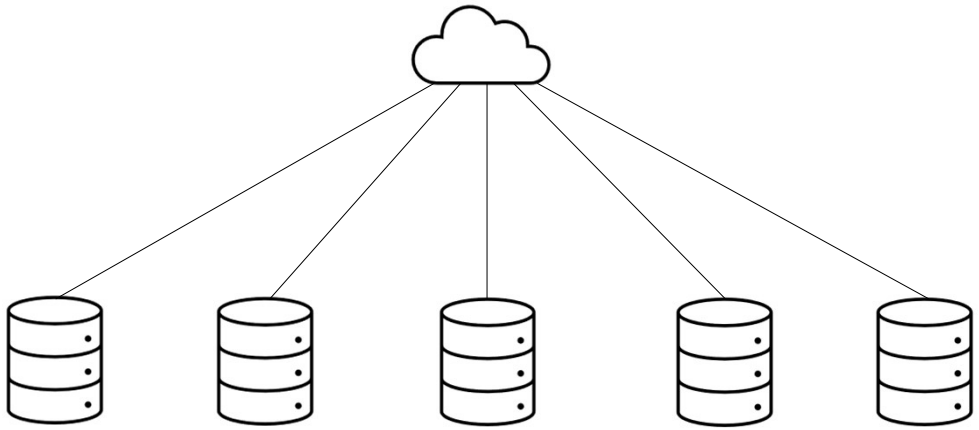
DeSci Nodes

Distributed across many servers



Decentralized data storage

Many copies keep things safe



- IPFS (https://ipfs.tech)
- Many copies keep things safe
- Institutional sovereignty
- Pick the storage providers of your choice or run on own hardware

- PB-scale data storage capability
- Highly competitive prices
  - Competitive marketplace
  - Crypto-economic incentives (Filecoin)

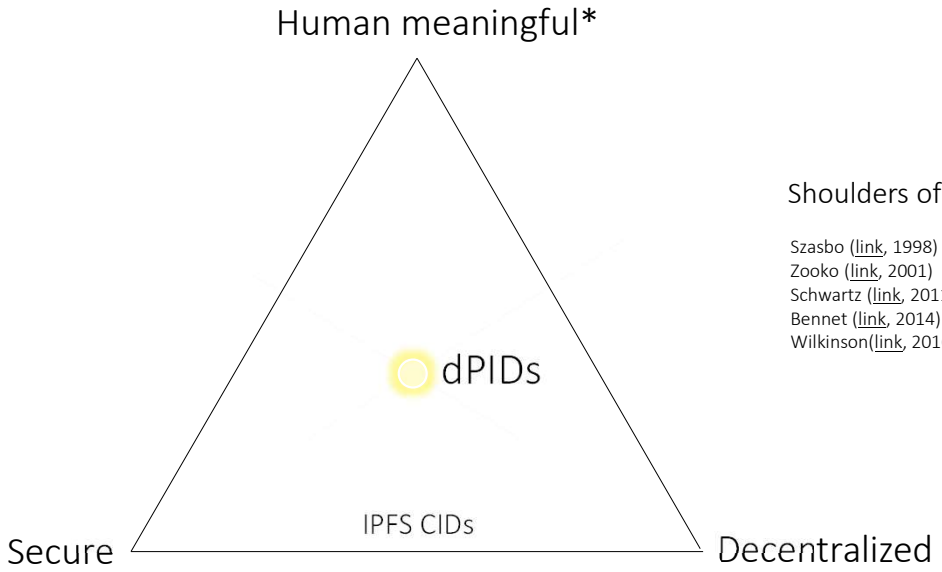
DPID

Decentralized Persistent Identifiers

- PIDs that address the entire linked data structure of a research object directory
- Structure: {Resolver}/{PID}/{version identifier OR CID}/{Component index}/{Component suffix}
- Examples (all the same):
  - Long format:
    - dpid.org/42/bafybeigdyrzt5sfp7udm7hu76uh7y26nf3efuylqabf3oclgtqy55fbzdi/1/measurements.csv
  - Short, human friendly:
    - dpid.org/42/0/1/measurements.csv
    - dpid.org/42/v1/data/measurements.csv

DPID

Zooko's Triangle



Shoulders of giants

Szasbo (link, 1998)
Zooko (link, 2001)
Schwartz (link, 2011)
Bennet (link, 2014)
Wilkinson(link, 2016)

In contrast, DOIs have none of these desirable properties.

www.dpid.org

DPID

Interacting with research

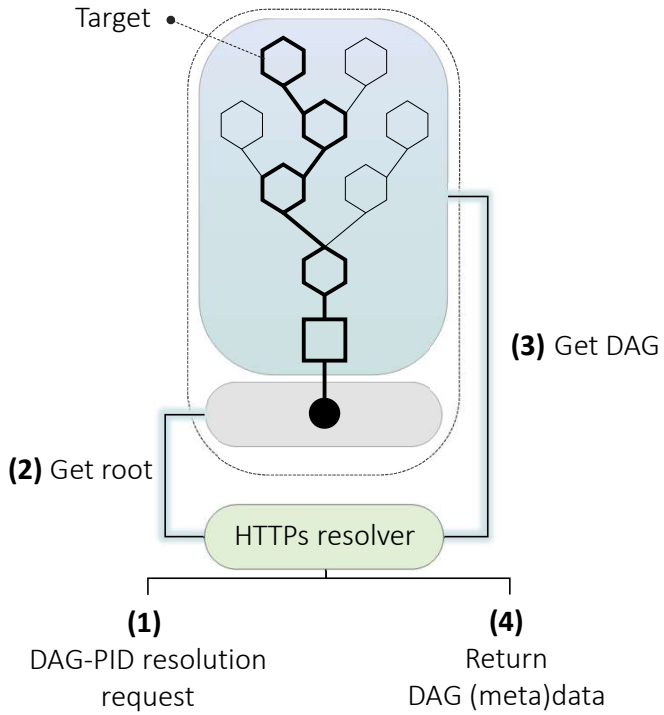
…resolution to (meta)data over HTTP

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)

DPID

Interacting with research

…resolution to (meta)data over HTTP

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

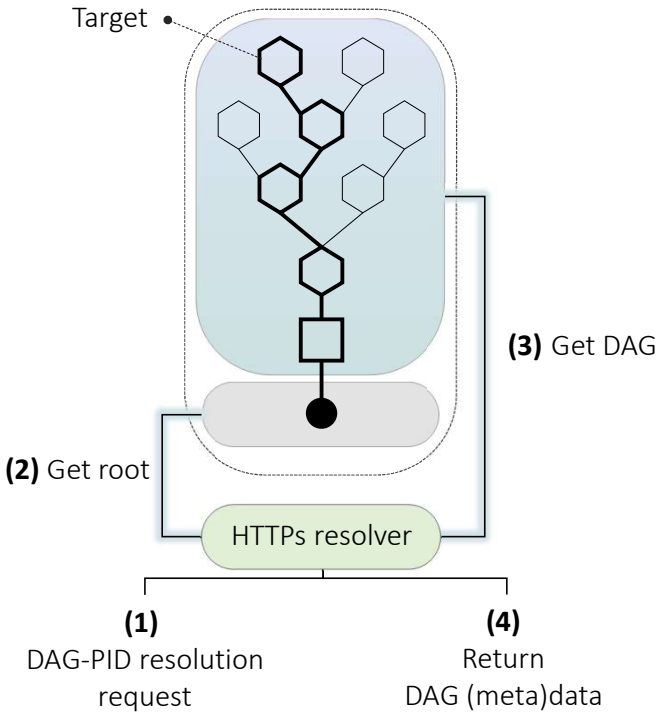
COMPACT PID TO DATA: dpid.org/42/1/2/1

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)

DPID

Interacting with research

…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

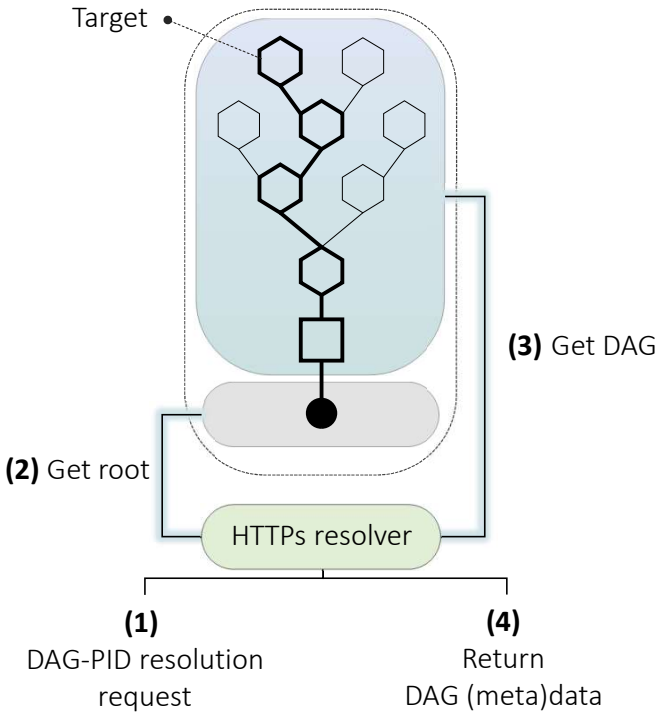
…as addresses for edge compute jobs

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

…as importable resources

```
Import desci as desci
with desci.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)

DPID

Interacting with research

…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
```

…as importable resources

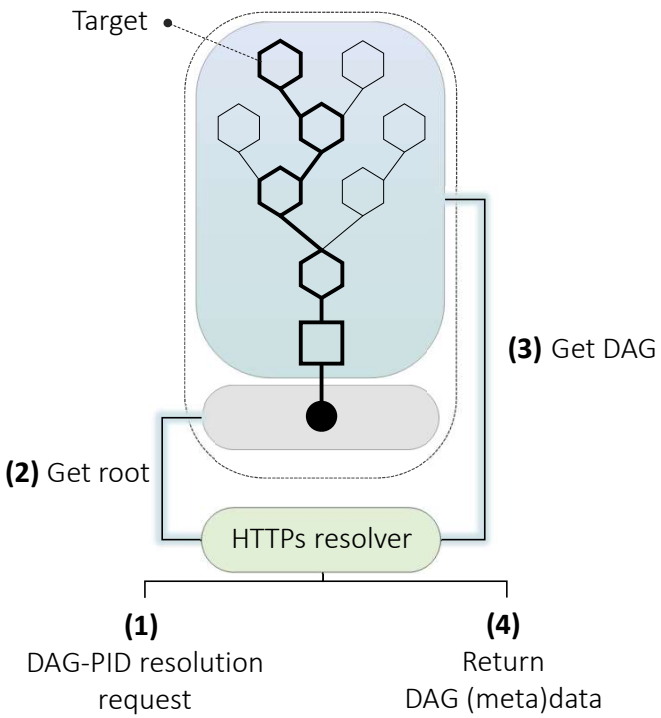
```
Import desci as desci
with desci.fetch
    import ("42/v1/data/measurement.csv") #import from PID
```

…as targets for attestations

```
structAttestation{

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

Open state resolution to (meta)data Over HTTP gateway



Open state read and write (Irrevocable API)

Showing your work

Gateways to the protocol

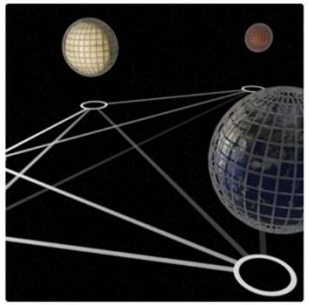
astro automata focuses on automating astrophysics research with AI. It is both a personal research blog and collection of related content.

Miles Cranmer

Incoming Assistant Professor of Data Intensive Science at University of Cambridge. Trying to accelerate astrophysics with AI.

New York City Email Twitter GitHub

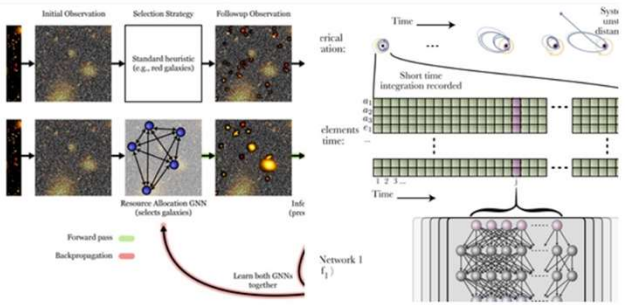
Research (full list)



Rediscovering orbital mechanics with machine learning

Could we discover the law of gravitation without even knowing the masses of planets in the solar system? In this paper we show how.

Read more



Unsupervised Resource Allocation with GNNs

We show how to optimize resource allocation without knowing the true utility. We use this to learn a mock telescope observational survey from scratch.

Read more

Bayesian neural networks for planetary instability

We describe a Bayesian neural network architecture that can accurately learn to predict dynamical (chaotic) instability in compact planetary systems. The network demonstrates surprisingly robust generalization to 5-planet systems.

Read more

https://astroautomata.com/

Summary

- A complete open-science protocol
  - Built on open-source software
  - FAIR by design (Findable Accessible Interoperable Reusable)
  - No link rot or content drift
  - No content paywalls
  - No vendor lock-in
  - No data silos (many copies keep things save and accessible)
  - Copyrights and dPIDs owned by their creators
  - Enabling reproducibility
  - New ways to interact with research
  - New possibilities to create incentive and to validate research
  - Potential for substantial cost savings