The FAIR Hourglass: Championing Data beyond FAIR

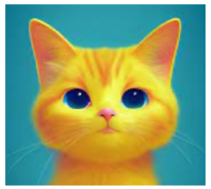
SciDataCon #557
Wednesday 25, Karajan Hall
https://www.scidatacon.org/IDW-2023-Salzburg/sessions/557/#

Organizers
Erik Schultes, GO FAIR Foundation & Leiden Academic Centre for Drug Research, Leiden University
Barbara Magagna, GO FAIR Foundation

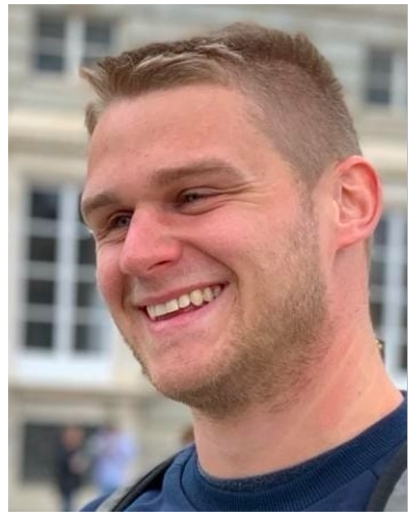
GO FAIR foundation

SciDataCon 2023
organized by
WORLD DATA SYSTEM
CODATA
as part of
23–26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

CEDAR Workbench



DeSci Labs



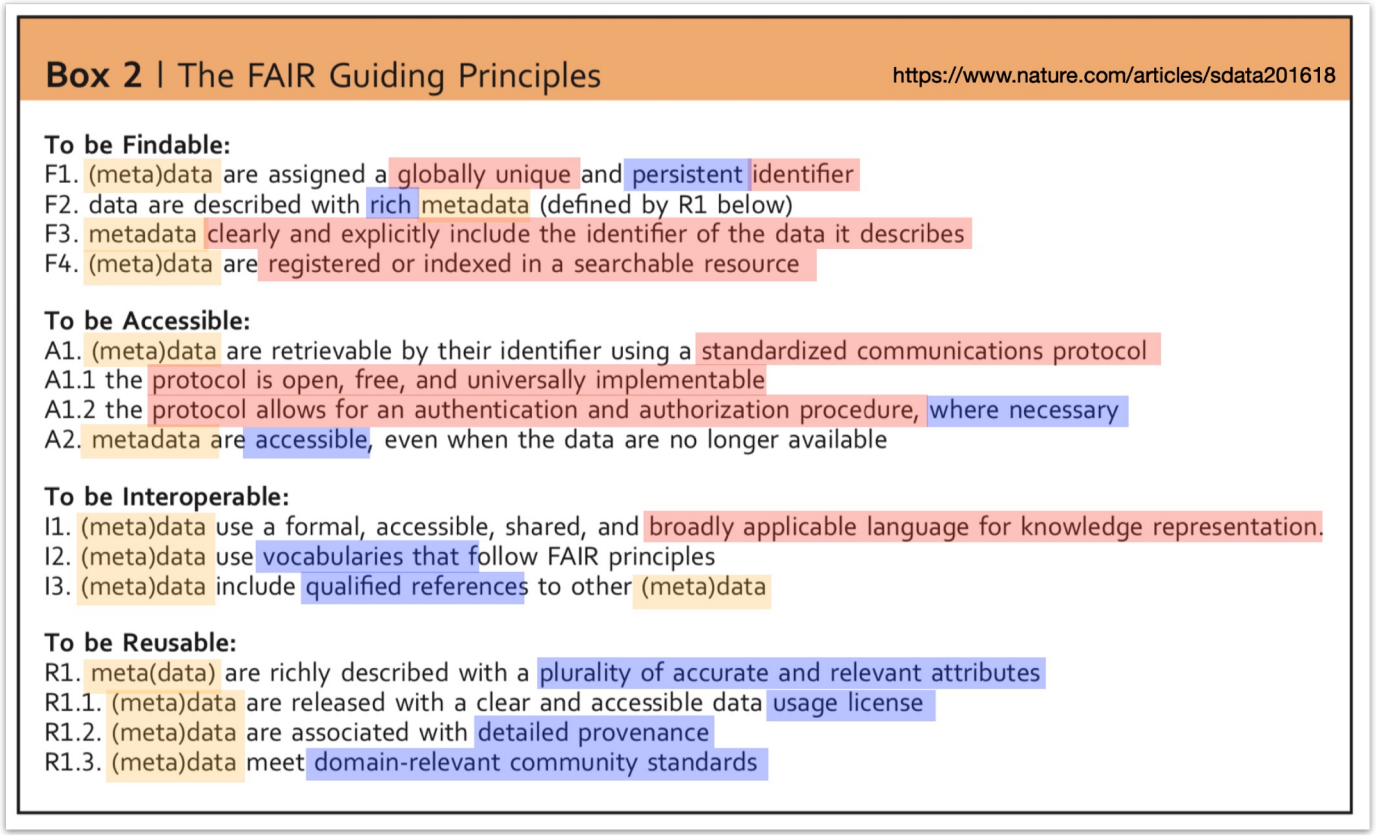
Erik Van Winkle
DeSci Labs

Decentralized Persistent Identifiers Reproducible Research Objects

DeSci Labs

SciDataCon 2023
organized by
CODATA
WORLD DATA SYSTEM
as part of
23-26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

Focus on one thing, do it well



Box 2 | The FAIR Guiding Principles

https://www.nature.com/articles/sdata201618

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

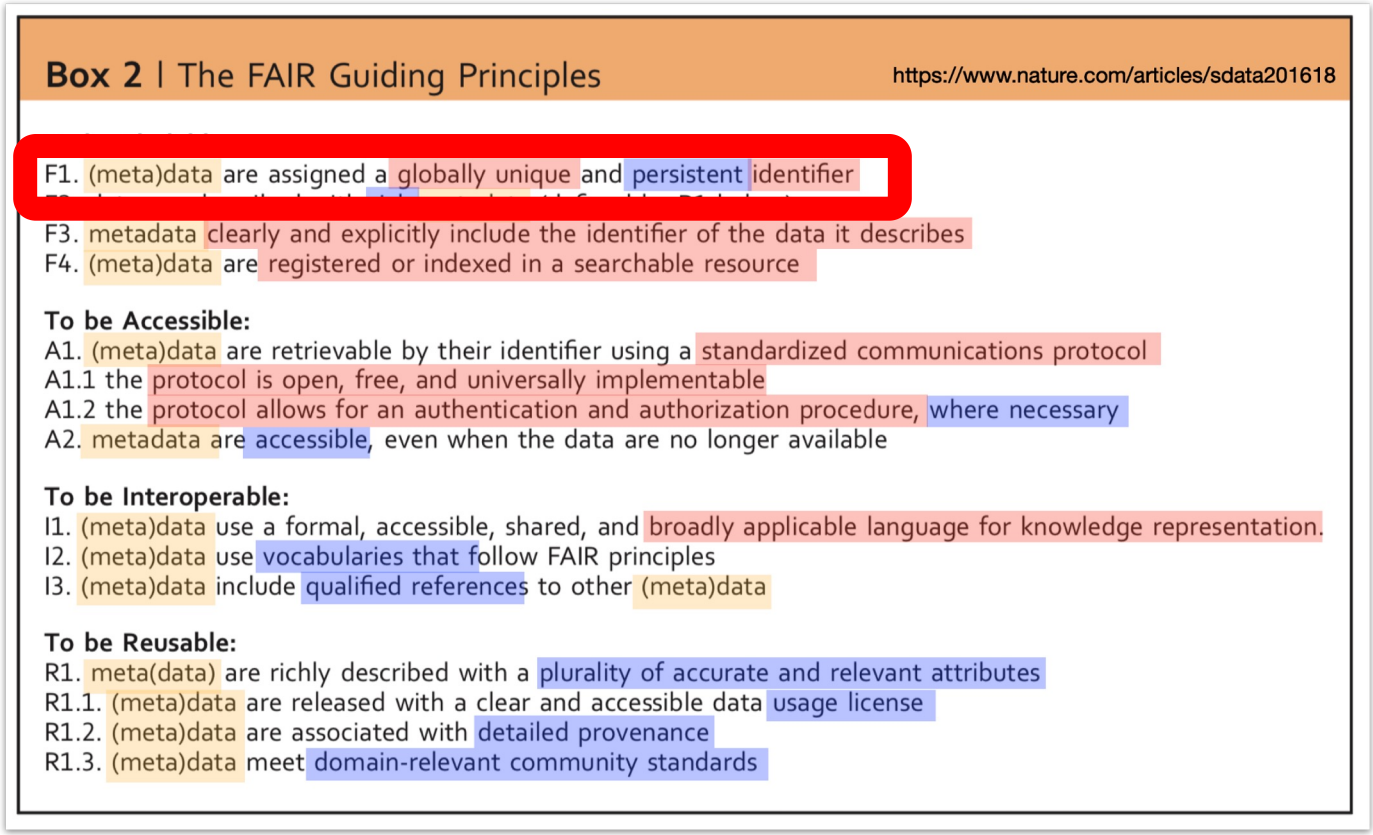
Machine-actionable metadata

Technical infrastructure (accepted generic services)

Social decisions (domain specific agreements)

SciDataCon 2023
organized by
CODATA
WORLD DATA SYSTEM
as part of
23-26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

Focus on one thing, do it well



Box 2 | The FAIR Guiding Principles https://www.nature.com/articles/sdata201618

F1. (meta)data are assigned a globally unique and persistent identifier
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Machine-actionable metadata

Technical infrastructure (accepted generic services)

Social decisions (domain specific agreements)

SciDataCon 2023
organized by
CODATA
WORLD DATA SYSTEM
as part of
23-26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

Principle F1: GUPRIs

“F1: (Meta)data are assigned a globally unique and persistent identifier”

Principle F1: GUPRIs

"F1: (Meta)data are assigned a globally unique and persistent identifier"

FAIR Principles: Interpretations and Implementation Considerations [1]

"F1: (Meta)data are assigned a globally unique, persistent and resolvable identifier"

SciDataCon 2023
organized by
CODATA
WORLD DATA SYSTEM
as part of
23-26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

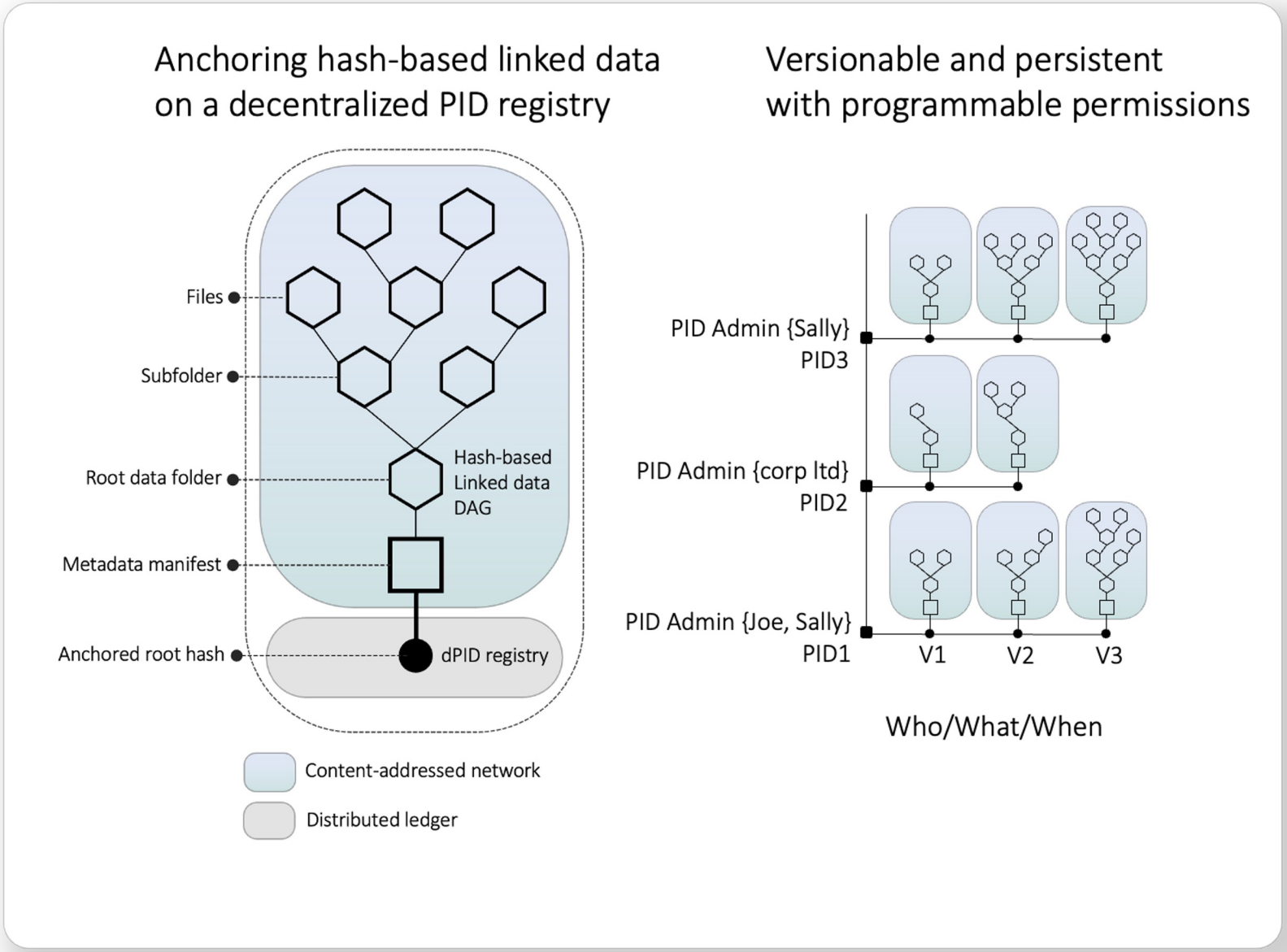
Sources:

Jacobsen, A., et al. (2020). FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence, 2(1-2), 10–29. https://doi.org/10.1162/dint_r_00024

The dPID Protocol

- Is based on persistent folder structures
- Built from the ground up to power the FAIR Principles
- In a fully decentralized, autonomous and user owned fashion
- With the ideals of open access baked into its foundations

Based on Folder Structures for Versionable Data



- Inherit the persistence of Blockchain and Content Addressable Storage. They run without social contracts or human intervention

dPIDs…

dPIDs…

- Inherit the persistence of Blockchain and Content Addressable Storage. They run without social contracts or human intervention
- Act as a PID container of arbitrary size, linking manuscripts, artifacts, sensemaking data, provenance, metadata and more in one place

dPIDs…

- Inherit the persistence of Blockchain and Content Addressable Storage. They run without social contracts or human intervention
- Act as a PID container of arbitrary size, linking manuscripts, artifacts, sensemaking data, provenance, metadata and more in one place
- Are completely immune to content drift and mitigate the effects of link rot

dPIDs…

- Inherit the persistence of Blockchain and Content Addressable Storage. They run without social contracts or human intervention
- Act as a PID container of arbitrary size, linking manuscripts, artifacts, sensemaking data, provenance, metadata and more in one place
- Are completely immune to content drift and mitigate the effects of link rot
- Act as PID APIs, offering standardized resolution pathways for both machines and humans alike.

dPIDs…

- Inherit the persistence of Blockchain and Content Addressable Storage. They run without social contracts or human intervention
- Act as a PID container of arbitrary size, linking manuscripts, artifacts, sensemaking data, provenance, metadata and more in one place
- Are completely immune to content drift and mitigate the effects of link rot
- Act as PID APIs, offering standardized resolution pathways for both machines and humans alike.
- Use decentralized technology to bust data silos and vendor lock-in. Portable across platforms

SciDataCon 2023
organized by
WORLD DATA SYSTEM
CODATA
as part of
23-26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

dPIDs…

- Inherit the persistence of Blockchain and Content Addressable Storage. They run without social contracts or human intervention
- Act as a PID container of arbitrary size, linking manuscripts, artifacts, sensemaking data, provenance, metadata and more in one place
- Are completely immune to content drift and mitigate the effects of link rot
- Act as PID APIs, offering standardized resolution pathways for both machines and humans alike.
- Use decentralized technology to bust data silos and vendor lock-in. Portable across platforms
- And so much more…

SciDataCon 2023
organized by
WORLD DATA SYSTEM CODATA
as part of
23-26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

No Vendor Lock-in - Deploy it yourself

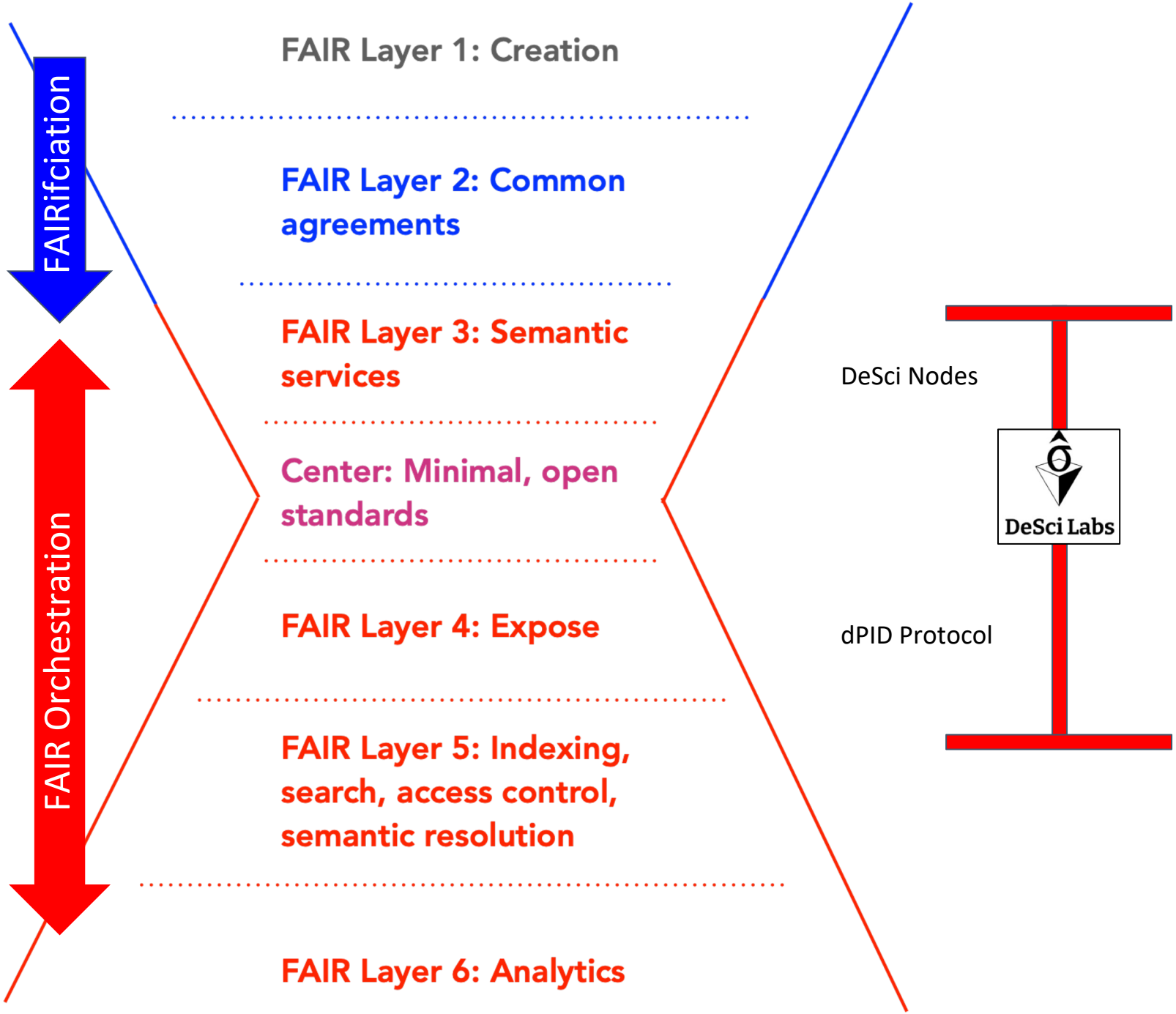
- **Own your PID:** You own the keys to your research. We don’t
- **Own your PID Minting:** You can own your PID minting. Simple on-demand minting through the open source dPID smart contract registry.
- **Own your Data:** Store your own data on your personal or institutional servers. We don’t need to own it to surface it.
- We can help you manage any and all of these functions if requested, but you can do it yourself

The FAIR Hourglass:

an organisational layer sitting on top of the FAIR Principles

Schultes, Erik. (2023) The FAIR Hourglass: A Framework for FAIR Implementation'. FAIR Connect, 1: 13 – 17. https://content.iospress.com/articles/fair-connect/fc221514

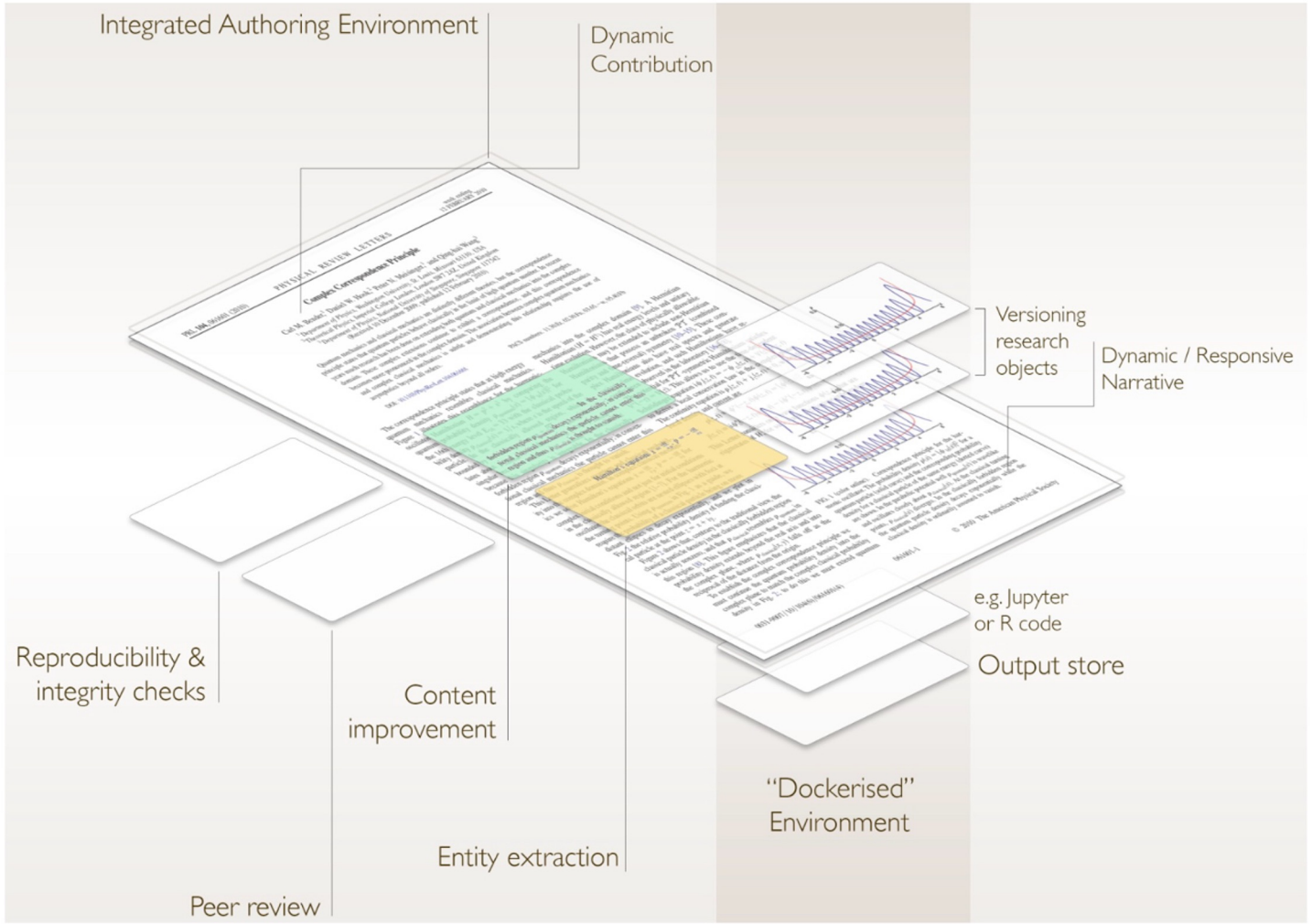
- FAIR
- Freedom to operate
- Open, minimal standards
- Decentralized
- Preempting vendor lock-in



SciDataCon 2023
organized by
CODATA
WORLD DATA SYSTEM
as part of
23-26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

Reproducible Research Objects:

Shifting the Unit of Knowledge



SciDataCon 2023
organized by
WORLD DATA SYSTEM
CODATA
as part of
23–26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

Demo Time!

The DeSci Nodes Application is an example interface that can be built on top of the Open Source dPID protocol

Learn about dPIDs: https://www.dpid.org

See a demo of reproducible research objects: https://beta.dpid.org/46

SciDataCon 2023
organized by
CODATA
WORLD DATA SYSTEM
as part of
23-26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

Call to Action…

- **Researchers**
  - Show us your reproducible research! Have code and data that you're proud of? Publish it!
  - Join our case study on the citation impacts of reproducibility. Prizes included
- **PID Enthusiasts:** Try out dPID. Tell us about your experiments on our open discord community
- **Data Stewards:** Join a user group to talk about mechanisms for appending provenance and metadata in a folder structure based PID
- **Libraries and Repositories:** Spin up a pilot project with dPID. We'll help you set it up, free of charge. Try out versionability of dPIDs.

SciDataCon 2023
organized by
CODATA
WORLD DATA SYSTEM
as part of
23–26 OCT 2023 SALZBURG
International Data Week
A FESTIVAL OF DATA

SciDataCon 2023

organized by

as part of

23-26 OCT
2023
SALZBURG

International
Data Week
A FESTIVAL OF DATA

FAIR well