eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC

Enabling FAIR Signposting and RO-Crate for content/metadata discovery and consumption

FAIR-IMPACT Support Offer #2

Stian Soiland-Reyes, The University of Manchester
Herbert Van de Sompel, DANS

Funded by the European Union

Webinar 2023-03-27
https://doi.org/10.5281/zenodo.7773578

Signposting the Scholarly Web

- Consideration:
  1. **Landing pages** support human interaction with scholarly objects
  2. Landing pages are *not* optimized for machines
  3. Programmatic access of data citations currently requires repository-specific knowledge

- Signposting:
  1. Makes scholarly resources on the Web easy to **navigate** for machines
  2. Leverages the **status quo**: scholarly objects are primarily web landing pages
  3. **Enhances** the status quo to empower machines and increase FAIR accessibility
  4. Provides a cross-system API aka **interoperability** for access to resources that are part of a scholarly object

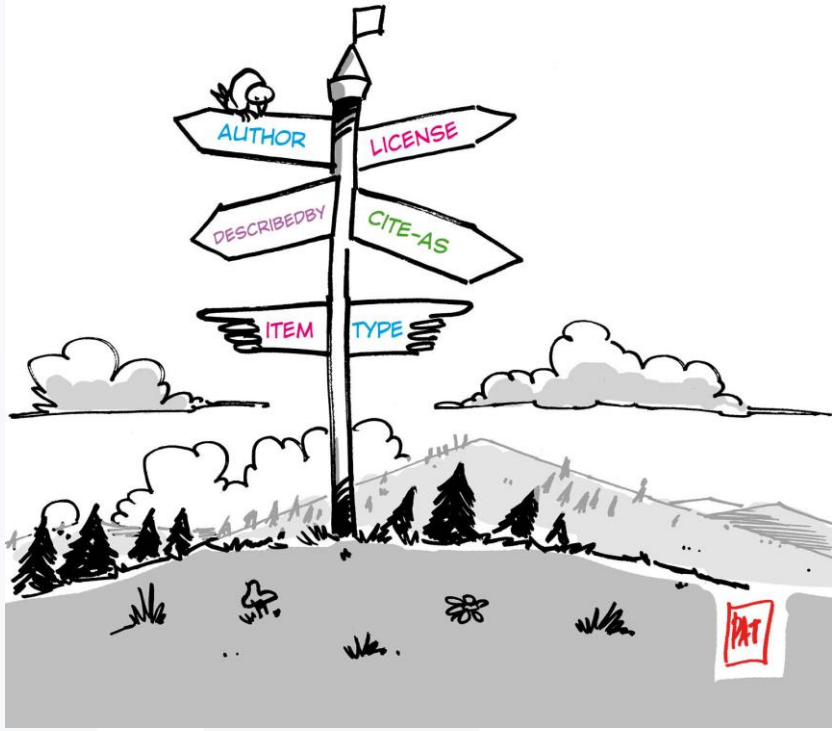


Image courtesy of Patrick Hochstenbach

RO-Crate

- Consideration:
  1. Digital research outputs have increasingly complex **history** and **attributions**
  2. Outputs are stored in disparate repositories with many **identifiers** yet limited **metadata**
  3. FAIR publishing in large repositories is restricted by **pre-determined** capabilities and APIs
- RO-Crate:
  1. Gives **developer-oriented recommendations**, based on existing Linked Data standards
  2. Packages data with external **references**, **contextual** information and rich **metadata**
  3. Propagates **provenance** and **lifted metadata** from custom file formats
  4. Provides long-term **archiving** and **platform-independent** exchange
  5. Extends **generic** capabilities with domain- and application-**specific profiles**

Signposting & RO-Crate: Match Made in Heaven

- The combination of Signposting & RO-Crate:
  - Makes **scholarly resources easy to navigate for machines**
  - Provides an unambiguous way for **machines to access metadata and content in a single request**
  - No need to probe with content-negotiation

Infrastructure independent
Overcome repository silos

Familiar Web technologies
HTTP, JSON

Extensible; One size does not fit all
Expand using link relations, schema.org and Linked Data

Scholarly Objects on the Web

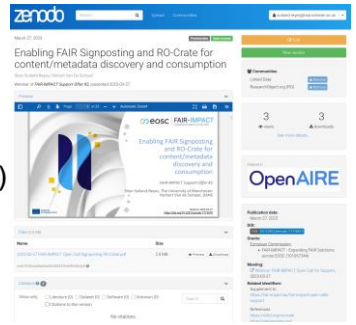
https://orcid.org/0000-0001-9842-9718

https://orcid.org/0000-0002-0715-6126

DataCite

https://zenodo.org/record/7774025

persistent identifier — HTTP redirect(s)

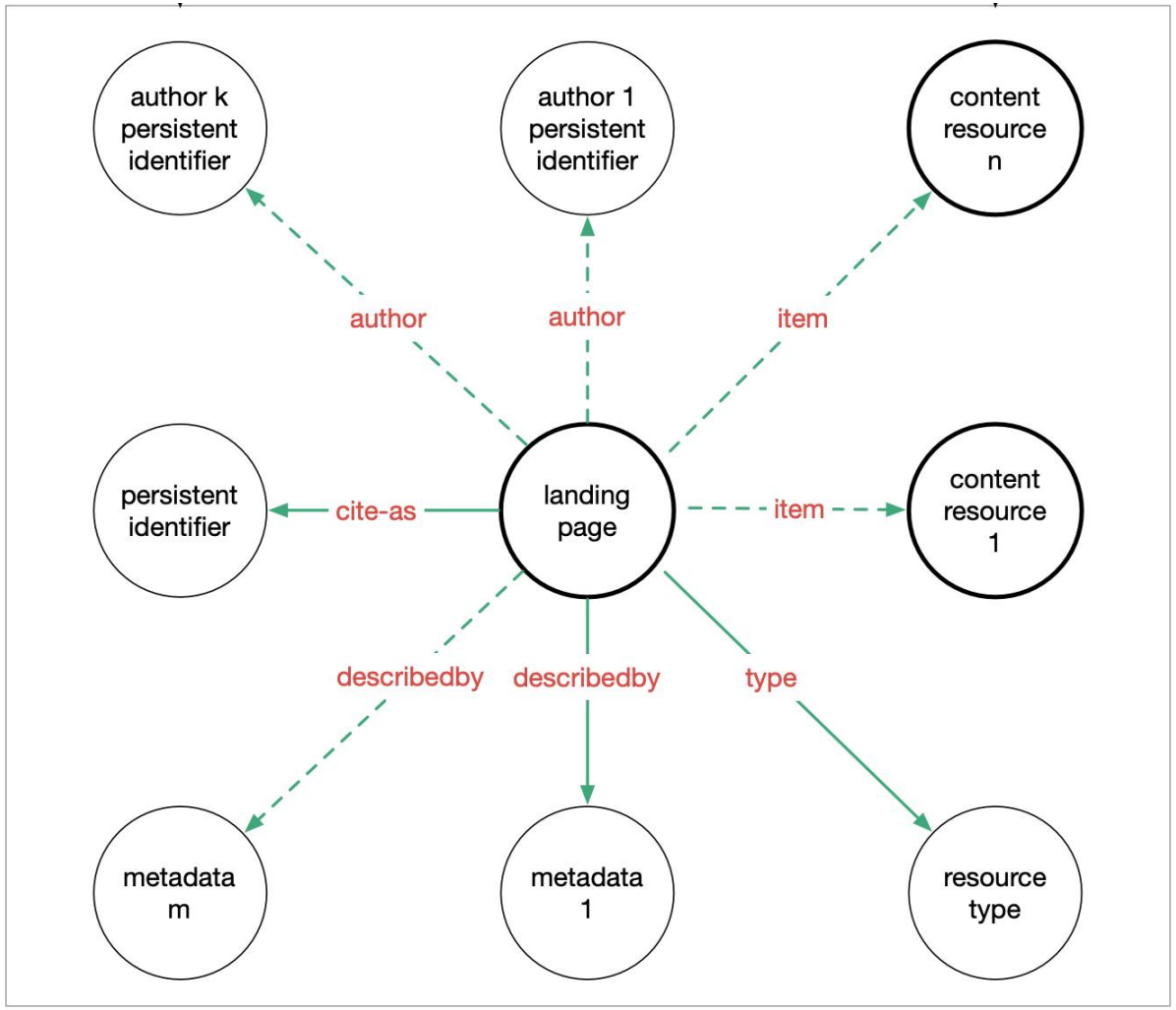


https://doi.org/10.5281/zenodo.7774025

DataCite XML Export

JSON-LD (schema.org) Export

FAIR Signposting: HTTP Links from Landing Page



FAIR Signposting Using HTTP Links

```
$ curl -i "https://example.org/page/7507"

HTTP/1.1 200 OK
Date: Fri, 9 Oct 2020 19:19:22 GMT
Content-Type: text/html
Content-Length: 25414
Link:
 <https://doi.org/10.5061/dryad.5d23f>    ; rel="cite-as"                ,
 <https://orcid.org/0000-0002-1825-0097>  ; rel="author"                 ,
 <https://example.org/meta/7507/bibtex>   ; rel="describedby"
                                          ; type="application/x-bibtex" ,
 <https://schema.org/ScholarlyArticle>    ; rel="type"                   ,
 <https://schema.org/AboutPage>           ; rel="type"                   ,
 <https://example.org/file/7507/1>        ; rel="item"
                                          ; type="application/pdf"

<html lang="en">
 <head>
  <meta charset="utf-8">
  ...
```

Signposting & RO-Crate: Single Request Access

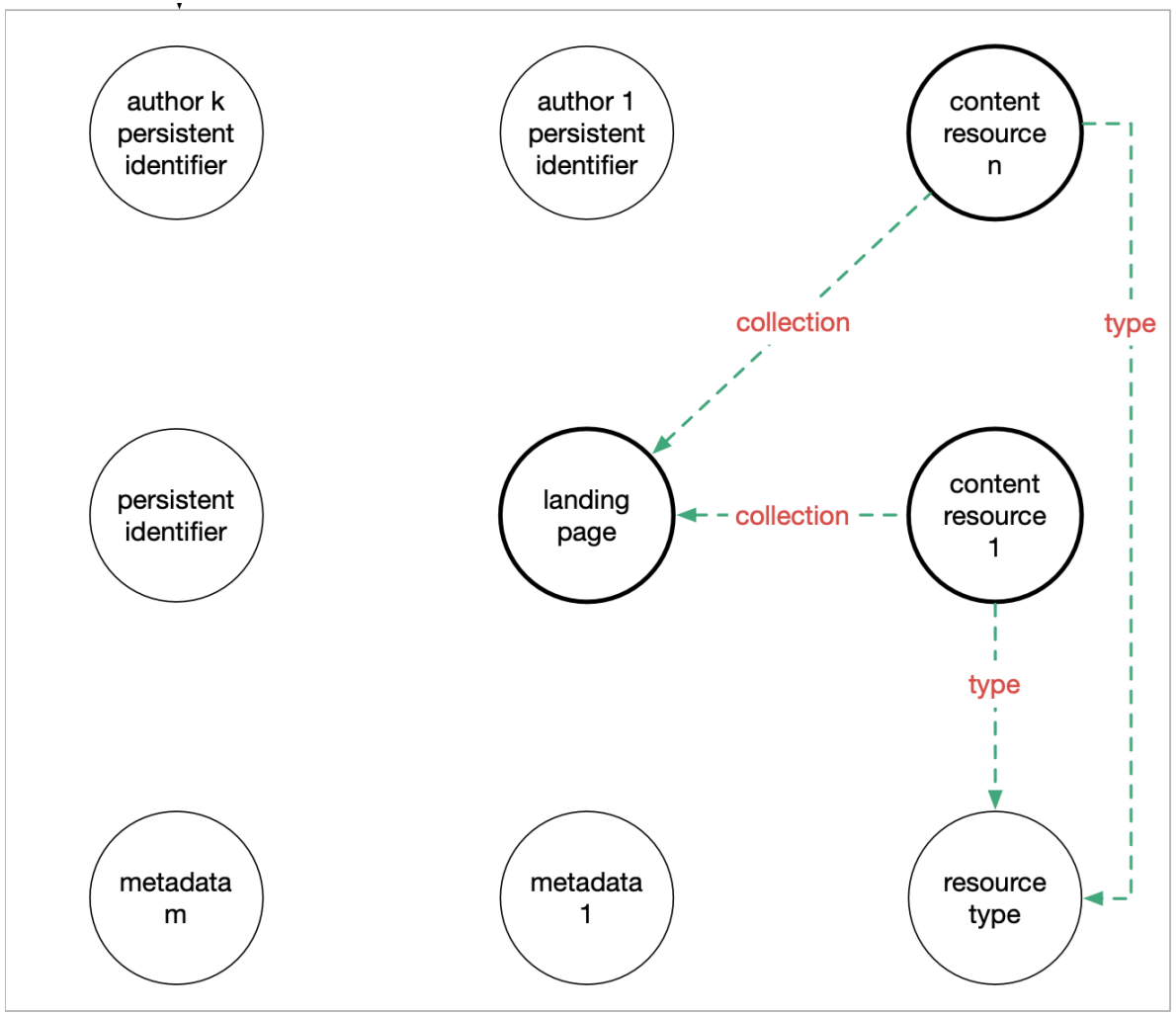
- RO-Crate can be added as additional *side-car* resources – no need to change architecture
- Adds predictable JSON-LD metadata following the RO-Crate profile
- Adds predictable archive of data with embedded metadata

```
<https://example.org/metadata/7507/1/ro-crate-metadata.json>
      ; rel="describedby"
      ; type="application/ld+json"
      ; profile="https://w3id.org/ro/crate",
<https://example.org/file/7507/1>
      ; rel="item"
      ; type="application/zip"
      ; profile="https://w3id.org/ro/crate"
```

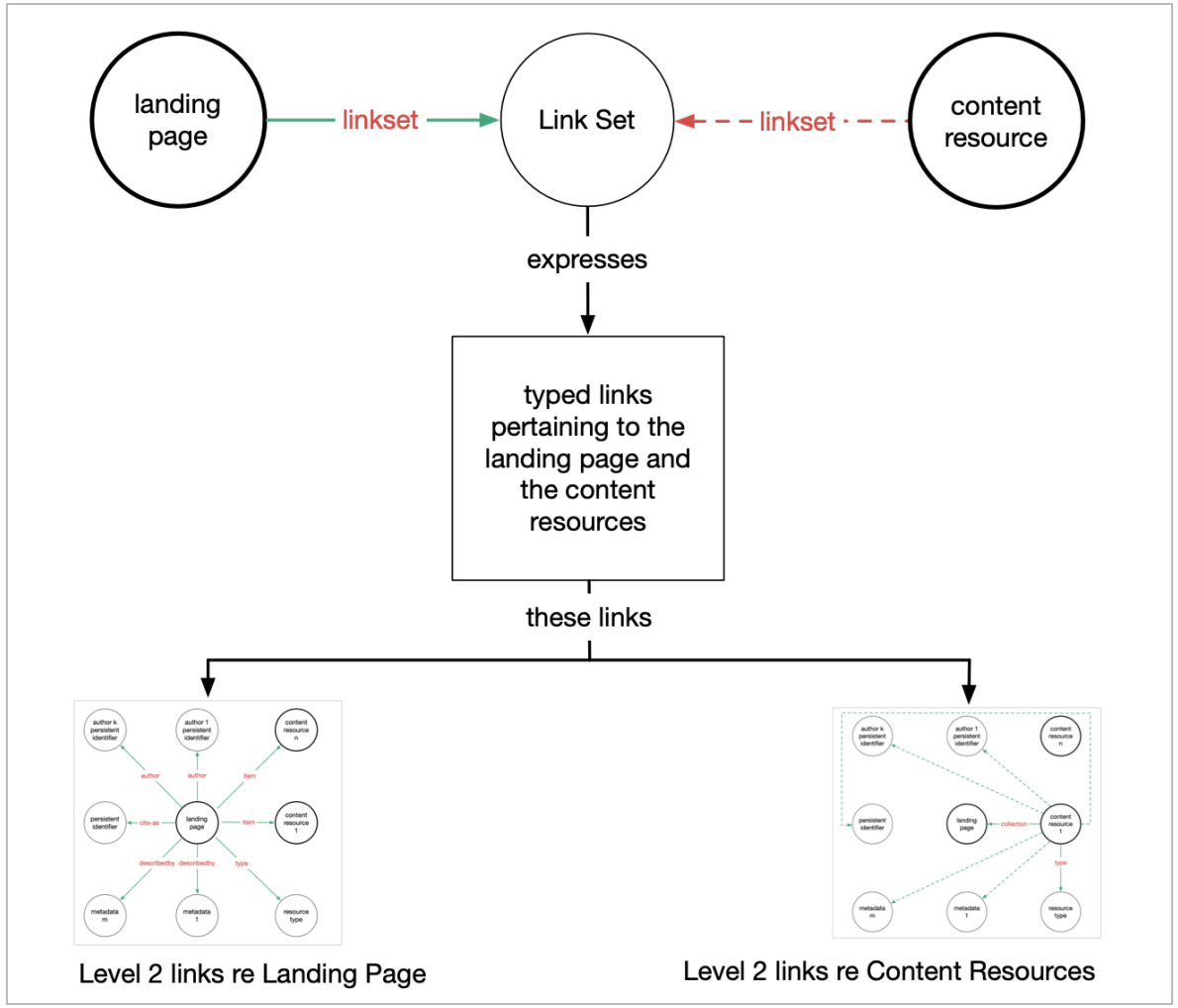
"Detached" RO-Crate Metadata File

ZIP archive including RO-Crate Metadata File

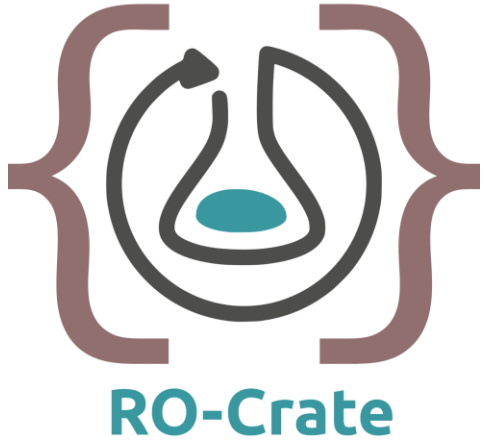
FAIR Signposting: HTTP Links from Other Resources



FAIR Signposting: Links in a Link Set




```
{
  "linkset": [
    {
      "anchor": "https://example.com/repo/d101.html",
      "cite-as": [
        {
          "href": "https://doi.example.org/10.1/d101"
        }
      ],
      "item": [
        {
          "href": "https://…/actual-download.csv",
          "type": "text/csv"
        }
      ],
      "describedby": [
        {
          "href": "https://…/metadata.ttl",
          "type": "text/turtle"
        }
      ]
    }
  ]
}
```



How does RO-Crate use Linked Data?

```
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [
```

JSON-LD preamble

```
    { "@type": "CreativeWork",
      "@id": "ro-crate-metadata.json",
      "conformsTo": {"@id": "https://w3id.org/ro/crate/1.1"},
      "about": { "@id": "./" }
    }
```

RO-Crate **metadata file** descriptor

```
    { "@id": "./",
      "identifier": "https://doi.org/10.5281/zenodo.1009240",
      "@type": "Dataset",

      "hasPart": [
        { "@id": "cp7glop.ai" },
        { "@id": "lots_of_little_files/" },
        { "@id": "communities-2018.csv" },
        { "@id": "https://doi.org/10.4225/59/59672c09f4a4b" },
        { "@id": "SciDataCon-Presentations/AAA_Pilot_Abstract.html"}
       ],

      "author": { "@id": "https://orcid.org/0000-0002-8367-6908" },
      "publisher": { "@id": "https://ror.org/03f0f6041" },
      "citation": { "@id": "https://doi.org/10.1109/TCYB.2014.2386282"},
      "name": "Presentation of user survey 2018"
    },
```

RO-Crate **root** dataset

..aggregates **Data entities**

..described w/ **contextual entities**

```
    { "@id": "cp7glop.ai",
      "@type": "File",
      "name": "Diagram showing trend to increase",
      …
    },
    …
```

Flat list of metadata per entity

Linked Data “by stealth”

Data and **Contextual** entities described *within* RO-Crate Metadata File

Base vocabulary & types: **schema.org**

Cross-references to further contextual entities

RO-Crate **principle**:
Reuse existing PIDs and URLs

..but always **describe entities** which lack a human-readable resolution

“Just enough” use of existing standards

Internet Assigned Numbers Authority

Relation Name	Description	Reference
about	Refers to a resource that is the subject of the link's context.	[RFC6903], section 2
acl	Asserts that the link target provides an access control resource for the link context.	[https://solidproject.org/TR/wac#acl-link-relation]
alternate	Refers to a substitute for this context	[HTML]
amphtml	Used to reference alternative content that uses the AMP profile of the HTML format.	[AMP HTML]
appendix	Refers to an appendix.	[HTML 4.01 Specification]
apple-touch-icon	Refers to an icon for the context. Synonym for icon.	[Configuring Web Applications]
apple-touch-startup-image	Refers to a launch screen for the context.	[Configuring Web Applications]
archives	Refers to a collection of records, documents, or other materials of historical interest.	[HTML5]
author	Refers to the context's author.	[HTML]
blocked-by	Identifies the entity that blocks access to a resource following receipt of a legal demand.	[RFC7725]
bookmark	Gives a permanent link to use for bookmarking purposes.	[HTML]
canonical	Designates the preferred version of a resource (the IRI and its contents).	[RFC6596]
chapter	Refers to a chapter in a collection of resources.	[HTML 4.01 Specification]
cite-as	Indicates that the link target is preferred over the link context for the purpose of permanent citation.	[RFC8574]
collection	The target IRI points to a resource which represents the collection resource for the context IRI.	[RFC6573]
contents	Refers to a table of contents.	[HTML 4.01 Specification]
convertedfrom	The document linked to was later converted to the document that contains this link relation. For example, an RFC can have a link to the Internet-Draft that became the RFC; in that case, the link relation would be "convertedFrom".	[RFC7991]
copyright	Refers to a copyright statement that applies to the link's context.	[HTML 4.01 Specification]
create-form	The target IRI points to a resource where a submission form can be obtained.	[RFC6861]
current	Refers to a resource containing the most recent item(s) in a collection of resources.	[RFC5005]
describedby	Refers to a resource providing information about the link's context.	[Protocol for Web Description Resources

https://www.iana.org/assignments/link-relations/

Linkset
RFC9264

HTTP Web Linking
RFC8288

HTML5
<link>

JSON-LD

schema.org

Support Offer #2

Enabling FAIR Signposting and RO-Crate for content/metadata discovery and consumption

Ideas for repositories and publishers

- **Add FAIR Signposting** headers for machine navigation
  - From landing page to existing downloads, structured metadata, APIs, identifiers, types, authors
  - Signposting back from existing sub-resources

- Support **RO-Crate import**
  - Parse and integrate metadata
- Prototype **RO-Crate export**
  - Embedding existing metadata
  - Add signposting for RO-Crate download

- Support **academic publishers** of scholarly articles and datasets to provide FAIR Signposting

FAIR-IMPACT

Ideas for FAIR tooling implementers

- Improve or integrate **existing Signposting clients** (Python, Ruby)
- Prototype open source **client libraries** for Signposting (e.g. JavaScript, Java, .NET)
- Improve or integrate existing **RO-Crate libraries** (Python, JavaScript, Ruby, Java)
- Make a Resolve-PID-to-RO-Crate tool
  - Following Signposting, falling back to heuristics
- **Publish/archive RO-Crate** in traditional repositories (e.g. Zenodo)
  - Map metadata to repository-specific requirements

Usage

https://pypi.org/project/signposting/

To use **signposting** from Python:

```
import signposting
s = signposting.find_signposting_http(
        "https://w3id.org/a2a-fair-metrics/05-http-describedby-citeas/")
print(s.citeAs)
print(s.citeAs.target)
for d in s.describedBy:
        print(d.target)
        print(d.type)
```

The **signposting.Signposting** object returned points to **signposting.Link** instances which have attributes matching the FAIR signposting profile.

We've started by adding the data entitites. Now we need contextual entities to represent Alice and Bob:

```
from rocrate.model.person import Person

alice_id = "https://orcid.org/0000-0000-0000-0000"
bob_id = "https://orcid.org/0000-0000-0000-0001"
alice = crate.add(Person(crate, alice_id, properties={
    "name": "Alice Doe",
    "affiliation": "University of Flatland"
}))
bob = crate.add(Person(crate, bob_id, properties={
    "name": "Bob Doe",
    "affiliation": "University of Flatland"
}))
```

Next, we express authorship of the various files:

```
paper["author"] = [alice, bob]
table["author"] = alice
diagram["author"] = bob
```

Finally, we serialize the crate to disk:

```
crate.write("exp_crate")
```

https://pypi.org/project/rocrate/

Ideas for FAIR data platforms

- FAIR Signposting or RO-Crate **contributions to open source projects** (e.g. Dataverse, Invenio RDM, RDF libraries and Linked Data platforms)

- **Consumption support** of FAIR Signposting in Linked Data platforms and client software
  - Navigating from PIDs and other resources to their metadata
  - To guide content-negotiation strategies

- Building **aggregated Knowledge Graphs** from RO-Crates and Signposting linksets
  - Including provenance of how the graphs were gathered

Ideas for FAIR outreach practitioners

- **Demonstration** (e.g. Jupyter Notebook) of FAIR Signposting to **navigate and consume** downloadable semantic resources (RO-Crate, RDF data dumps) from PIDs
- Assistance with development and improvement of **training material** and **specifications** for RO-Crate or Signposting
- **Community support** and advocacy of FAIR Signposting and RO-Crate to increase potential uptake of technologies

FAIR-IMPACT Support during the period

- **Slack** channel #signposting-ro-crate
  https://fair-impact-support.slack.com/
  https://join.slack.com/t/fair-impact-support/shared_invite/zt-1s86x15a8-pJdpSns3tdZXgAoruHtuDw
  - Support from the support offer coordinators
  - Collaborate with other participants
- Weekly FAIR-IMPACT "**office hours**"
  - For higher bandwidth support/debugging (Slack + Zoom on demand)
- Good old **email**

Timeline

- 01 June 2023: Deadline for applications
- Beginning July 2023: Successful applicants informed

- September 2023: Joint **introductory** session on support programme
  Virtual **kick-off** workshop inc. networking, lightning talks
  Start **prototyping** and developing ideas (7–10 h)

- October 2023: Virtual **brainstorming** and **scoping** session
  Implement their plans and prepare for demonstrations (15–30h)

- November 2023: Virtual reporting/demo session
- Reporting: **Revise** draft report
- Public dissemination webinar (**demo video**)

- **Exit interview**, draft **Implementation Story** (~30 min)

Scoping

- Nominally 5 days effort per project (incl. ~13h mandated activities) @500 EUR/day -> 2500 EUR
- Hackathon-style: Scope to one feature that can be demonstrated
- Can expand scope by combining multiple challenges (up to 10 days)
- Maximum of 15 teams will be selected

eosc | FAIR-IMPACT

Expanding FAIR solutions across EOSC

@fairimpact_eu /company/fair-impact-eu-project

Funded by
the European Union