

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Finding and Accessing Data by Virtual Research Environments

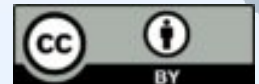
Leonardo Candela

 [0000-0002-7279-2727](https://orcid.org/0000-0002-7279-2727)

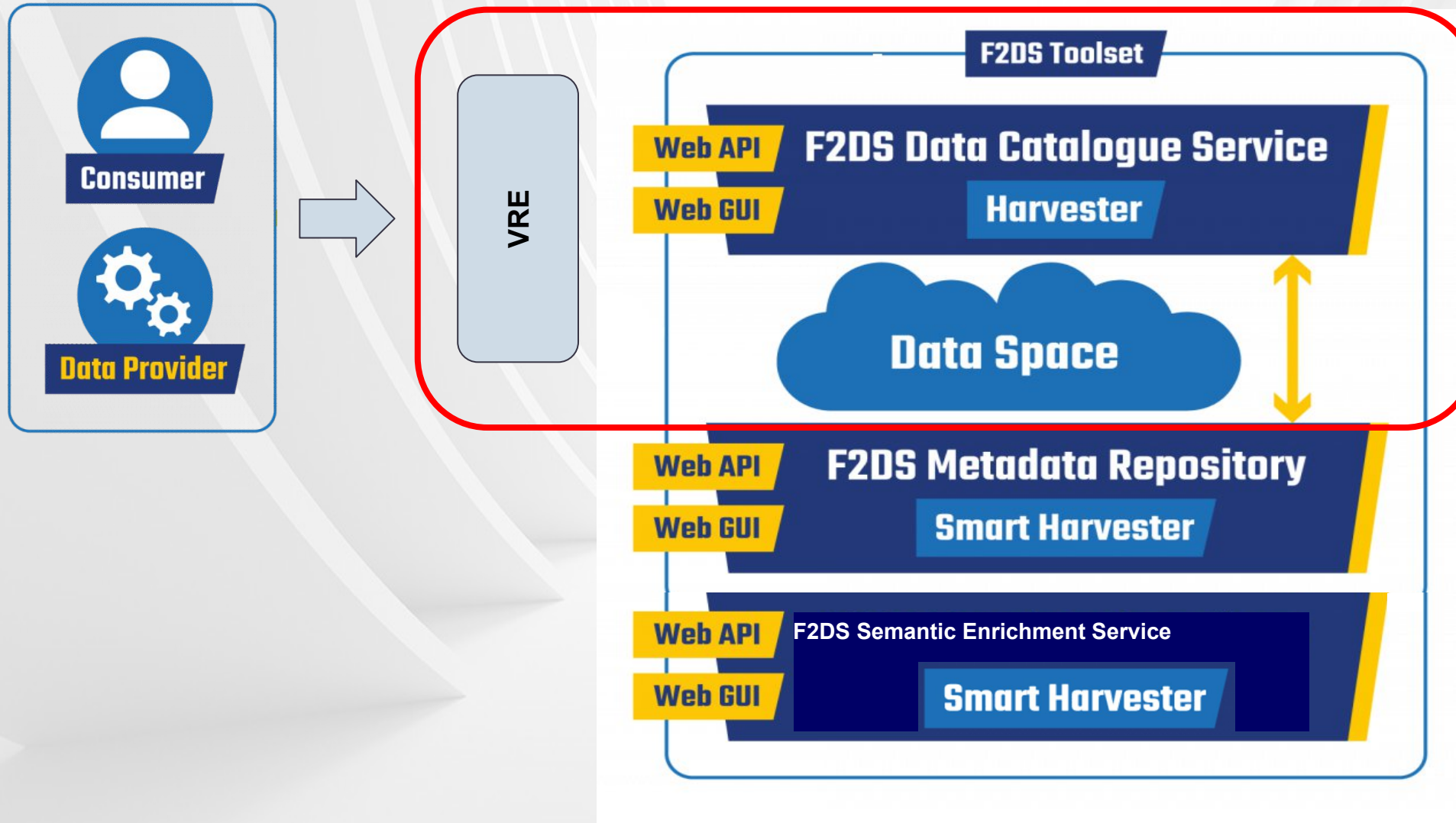
 Follow @leonardocandela



EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.
This material by the EOSC-Pillar is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



The Catalogue side of the story



Why do we need a catalogue?

- * To make data spread across many sources seamlessly **FAIR**

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

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 they describe

F4. (Meta)data are registered or indexed in a searchable resource

- * To easily develop **focused views / collections** for **communities of practices**
 - * by **existing contents** & by facilitating the publishing of **new contents**

Not just a catalogue: A VRE with a Catalogue

A VRE is an innovative, web-based, community-oriented, comprehensive, flexible, and secure working environments conceived to serve the needs of “modern” science delivered as-a-Service and catering for co-creation*

The image displays a screenshot of the EOSC-Pillar Data Catalogue web application and a corresponding system architecture diagram.

Web Application Screenshot:

- Navigation:** Home, Share Link, Publish Item.
- Search:** Search items... (81 items found).
- Filters:** Tags (CMIP6 (20), No Tag (14), air_temper, historical, soil proper, soils (8), geographic, nitrogen_1, pr (7), precipitatic), Formats (Dataset (28), TSV (7), BIN (5), PDF (4), CSV (3), DOCX (2), HTML (2), PNG (2), TXT (2)).
- Item Details:** Somatic variant calling. Description: "These data are used on T6.6 workflow test of Galaxy that demonstrates identification of somatic and germline variants from tumor and normal sample pairs." Includes a list of dataset identifiers and download links.
- Item URL:** <https://data.d4science.org/ctlg/EOSCpillarResDataCtlg/scripts-and-data-of-the-genetic-analysis-of-syrah-x-grenache-progeny>

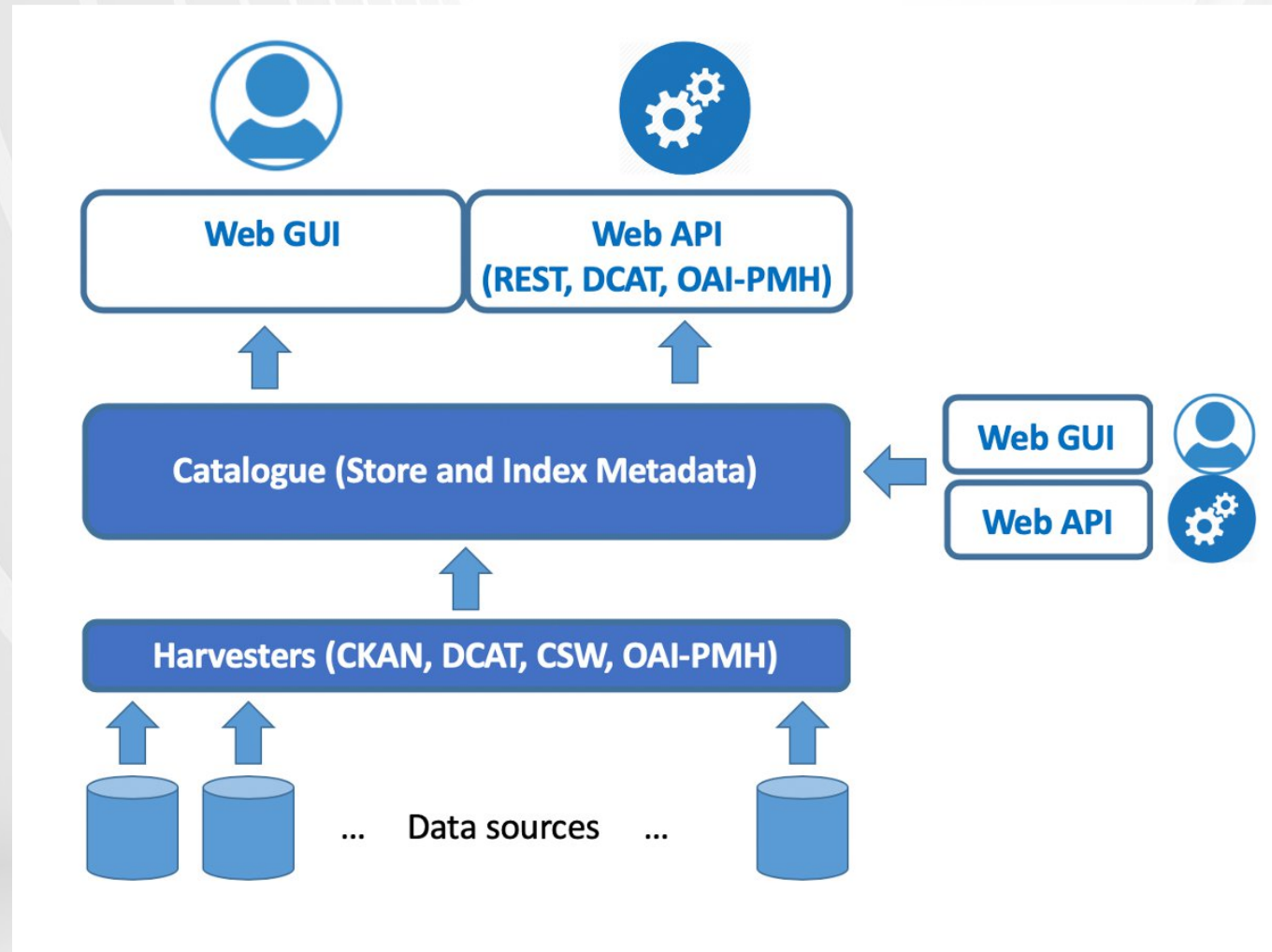
System Architecture Diagram:

- Users:** Represented by a person icon and a gear icon, interacting with the **Gateway**.
- Gateway:** Contains **Web GUI** and **Web API**, which interface with the **VRE**.
- Shared Items:** Flow from the Gateway to the **Workspace**.
- Workspace:** Interacts with **Social Networking** and **Other Services**.
- Social Networking:** Interacts with the **Catalogue** and **Infrastructure**.
- Other Services:** Interacts with the **Catalogue** and **Infrastructure**.
- Catalogue:** Interacts with the **Gateway** and **Infrastructure**.
- Infrastructure:** The foundational layer supporting the entire system.

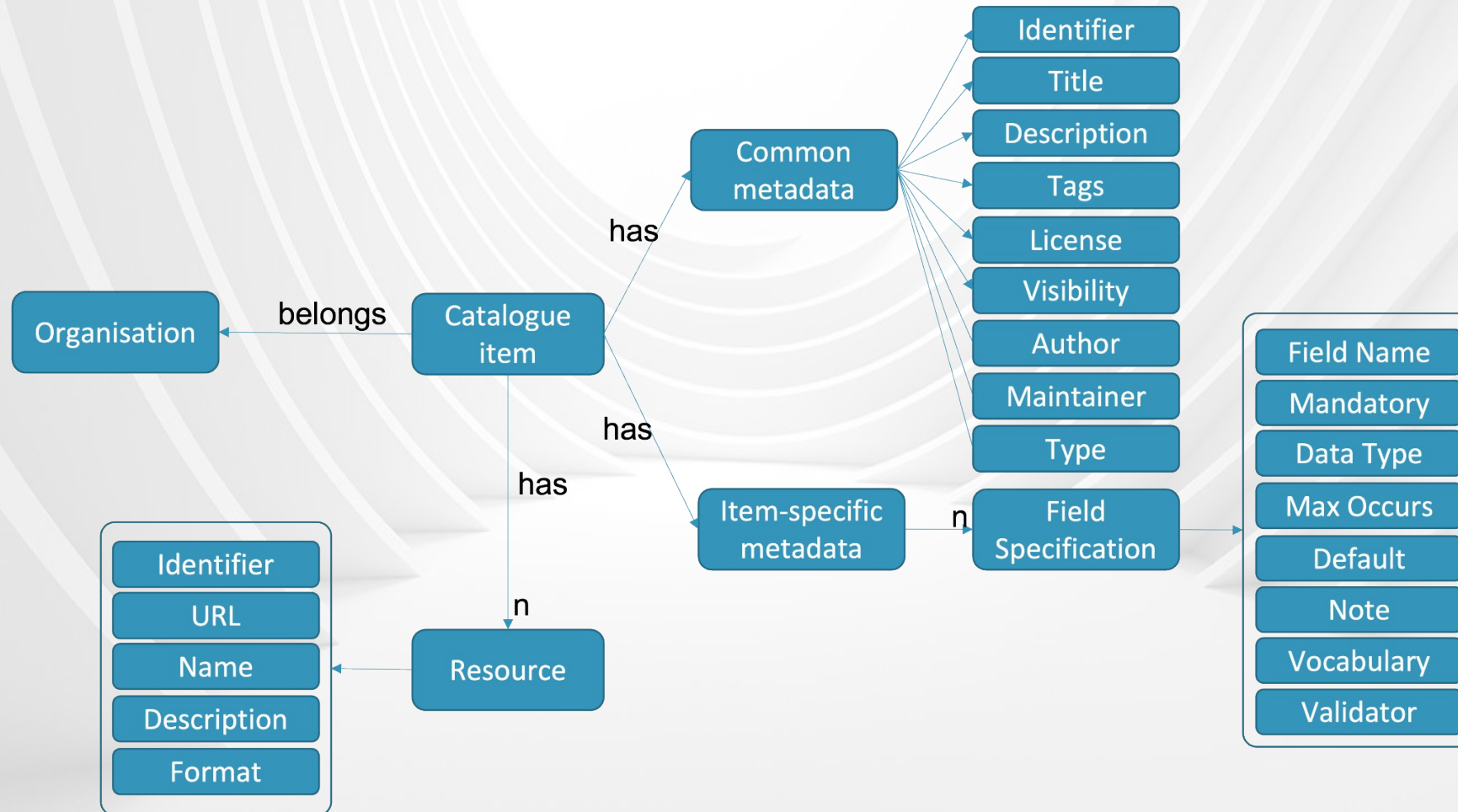
* M. Assante et al (2019) [Enacting Open Science by D4Science](#) Future Generation Computer Systems [10.1016/j.future.2019.05.063](#)

M. Assante et al (2022) [Virtual Research Environments co-creation: the D4Science experience](#) Concurrency Computat Pract Exper [10.1002/cpe.6925](#)

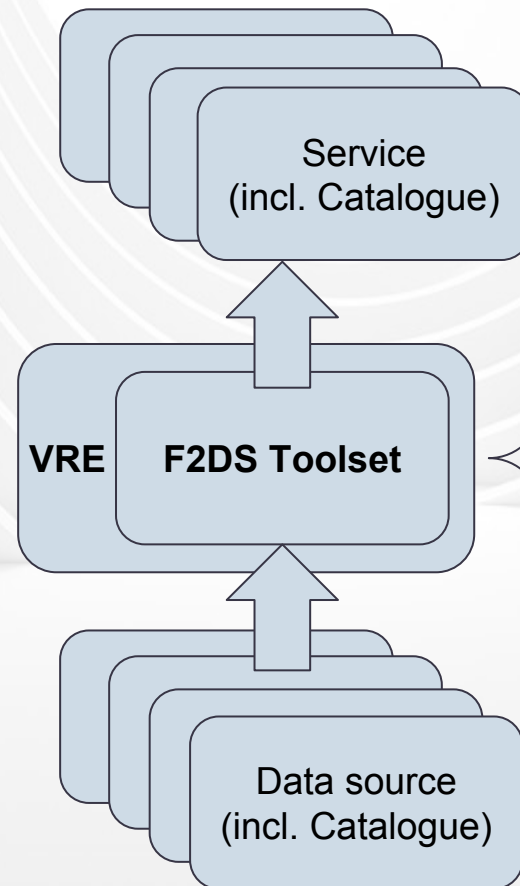
Catalogue Service Architecture



The metadata model of the Catalogue



... enacting an ecosystem of catalogues



Every catalogue has a **“designated community”** and is a **“decision point”** for:

- content to collect
 - standard and ad-hoc harvesters
 - mappings
- “new” content to publish
 - publishing policy
- metadata characterising items
 - profiles and controlled vocabularies
- protocols and standards to use to expose its contents

Data Access / Exploitation Patterns

EOSCPillar Res Data Ctlg
This working environment is mainly conceived to host and operate the overall research data...

Items Activity Stream About

somatic

2 items found for "somatic" Order by: Relevance

Somatic variant calling
These data are used on T6.6 workflow test of Galaxy that demonstrates identification of somatic and germline variants from tumor and normal sample pairs.

GET `https://(base-uri)/items/somatic-variant-calling`

KEY	VALUE	DESCRIPTION
Key	Value	Description

```

{
  "version": null,
  "creator_user_id": "c1fab78-c436-4a44-8979-e004b5458df7",
  "type": "dataset",
  "resources": [
    {
      "cache_last_updated": null,
      "package_id": "288ba5c2-e099-41a6-9b42-7772c29c7f9",
      "datastore_active": false,
      "id": "adc77bef-6ebc-47fb-8f26-deec65ebfe8",
      "size": "582017953",
      "conforms_to": "[\"https://f2ds.eosc-pillar.eu/profile/2f08228e-1789-40f8-84cd-28e328c3604/\", \"https://f2ds.eosc-pillar.eu/profile/c579-43bb-b470-306abdc808c7/\"],",
      "state": "active",
      "hash": "",
      "description": "undefined",
      "format": "GZ",
      "mime_type_inner": null,
      "url_type": null,
      "revision_id": "e9e8269f-0000-0000-0000-000000000000",
      "name": "SLGFSK-N_231335_r1_chr5_12_17.fastq.gz",
      "mime_type": "application/x-gzip",
      "cache_url": null,
      "language": "[\"http://id.loc.gov/vocabulary/iso639-1/en/\"],",
      "license": "http://rdflicense.appspot.com/rdflicense/cc-by-nc-nd3.0/",
      "created": "2022-10-20T23:14:29.084822",
      "uri": "https://doi.org/10.15454/QQM6ZT/MJUDPL",
      "uri": "https://f2ds.eosc-pillar.eu/distribution/2f4aa933-f6c4-4c38-acba-b16c3ae4726",
    }
  ]
}
    
```

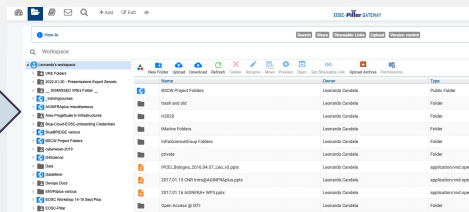
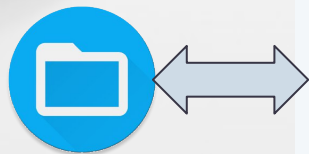
Links to payloads

REST & JSON

dcat:distribution

```

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
  >
  <dcat:Dataset rdf:about="https://f2ds.eosc-pillar.eu/dataset/ae18cd56-3675-4b60-827b-83f660581532">
    <dct:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/en"/>
    <dct:keyword>undefined</dct:keyword>
    <dct:identifier>doi:10.15454/QQM6ZT</dct:identifier>
    <dct:conformsTo>https://f2ds.eosc-pillar.eu/profile/2f08228e-1789-40f8-84cd-28e328c3604</dct:conformsTo>
    <dcat:distribution>
      <dcat:Distribution rdf:about="https://f2ds.eosc-pillar.eu/distribution/d96fe926-169a-49d3-87ea-e6dfed99-43bb-b470-306abdc808c7">
        <dct:title>SLGFSK-N_231335_r1_chr5_12_17.fastq.gz</dct:title>
        <dct:conformsTo>https://f2ds.eosc-pillar.eu/profile/2f08228e-1789-40f8-84cd-28e328c3604</dct:conformsTo>
        <dct:byteSize rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">530427108,0</dct:byteSize>
        <dct:license rdf:resource="http://rdflicense.appspot.com/rdflicense/cc-by-nc-nd3.0/">
        <dct:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/en"/>
        <dcat:mediaType>application/x-gzip</dcat:mediaType>
        <dct:format>GZ</dct:format>
        <dcat:accessURL rdf:resource="https://f2ds.eosc-pillar.eu/dataset/ae18cd56-3675-4b60-827b-83f660581532/2f4aa933-f6c4-4c38-acba-b16c3ae4726"/>
      </dcat:Distribution>
    </dcat:distribution>
    <dcat:contactPoint>
      <vcard:Organization rdf:about="https://f2ds.eosc-pillar.eu/organization/2f4aa933-f6c4-4c38-acba-b16c3ae4726">
        <vcard:hasEmail rdf:resource="mailto:notifications@d4science.org"/>
        <vcard:fn>D4Science Jarvis</vcard:fn>
      </vcard:Organization>
    </dcat:contactPoint>
    <dct:publisher>
      <foaf:Organization rdf:about="https://ckan.eosc-pillar.d4science.org/organization/eab5f866-79c7-4b22-8000-000000000000">
        <foaf:name>etude-entrepot-national</foaf:name>
      </foaf:Organization>
    </dct:publisher>
    <dct:hasVersion>1</dct:hasVersion>
    <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2022-07-06T13:40:54.691158</dct:issued>
  </dcat:Dataset>
    
```



Data Discovery and Consumption Offering Wrap-up

- * data discovery and consumption needs are **very heterogeneous**
 - * global solutions have limitations
 - * cater for **communities of practice**
-
- * **EOSC-Pillar F2DS toolset** fits with this by bringing data close to diverse users and usage needs



EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Thank you!

Get in touch with us!



www.eosc-pillar.eu



[@EoscPillar](https://twitter.com/EoscPillar)



[/company/eosc-pillar](https://www.linkedin.com/company/eosc-pillar)



EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.
This material by the EOSC-Pillar Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

