FAIRification STEP 2 – FAIR principle F3

Wednesday 3 Feb 2021, 10:00 → 12:10 Europe/Copenhagen

Relevance and Importance of splitting META-DATA from DATA

Erik Schultes, PhD
International Science Coordinator
GO FAIR Foundation
Leiden Center for Data Science

erik.schultes@go-fair.org
https://www.go-fair.org
http://orcid.org/0000-0001-8888-635X

FAIRification STEP 2 – FAIR principle F3

Wednesday 3 Feb 2021, 10:00 → 12:10 Europe/Copenhagen

10:00 → 10:10	**Welcome and Introduction** **Speaker**: Mr Bert Meerman (GFF)	10m
10:10 → 10:25	**FAIR assessment results** **Speaker**: Dr Andreas Jaunsen (NeIC)	15m
10:25 → 10:55	**Relevance and Importance of splitting META-DATA from DATA** **Speaker**: Dr Erik Schultes (GO-FAIR)	30m
10:55 → 11:05	**BREAK**	10m
11:05 → 11:20	**Some recommendations for the practical, machine-friendly implementation of the FAIR F3 principle** **Speaker**: Dr Robert Huber (Univ of Bremen)	15m
11:20 → 11:35	**File level identification support in Dataverse/DataverseNO** **Speaker**: Mr Philipp Conzett (UiT / The Artic University of Norway)	15m
11:35 → 11:55	**Q & A** **Speaker**: Mrs Josefine Nordling (CSC)	20m
11:55 → 12:00	**Call to Action and Close** **Speaker**: Mr Bert Meerman (GFF)	5m

Acknowledgments

Nikola Vasilijevic
Technical University of Denmark

Mark Wilkinson
Universidad Politécnica de Madrid

2016

nature > scientific data > comment > article

SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

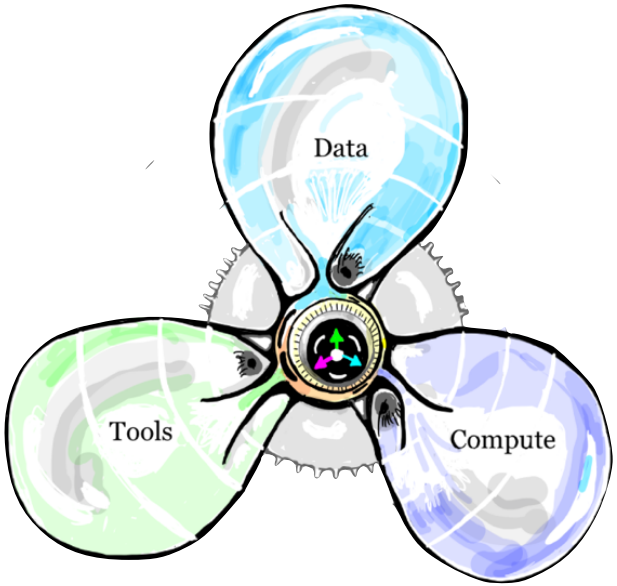
The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Download Citation

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

Data and services that are findable, accessible, interoperable, re-usable both for machines and for people.



Relevance and Importance of splitting META-DATA from DATA

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

Relevance and Importance of splitting META-DATA from DATA

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

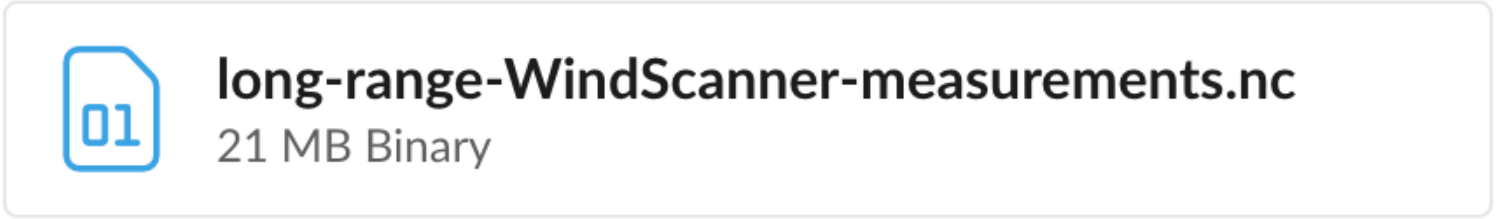
To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

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

self-documenting data

https://www.unidata.ucar.edu/software/netcdf/documentation/NUG/_best_practices.html

NetCDF files, example:



- 1 hour of 5 Hz measurements
- 1.5 milions of samples per variable

The embedded metadata…

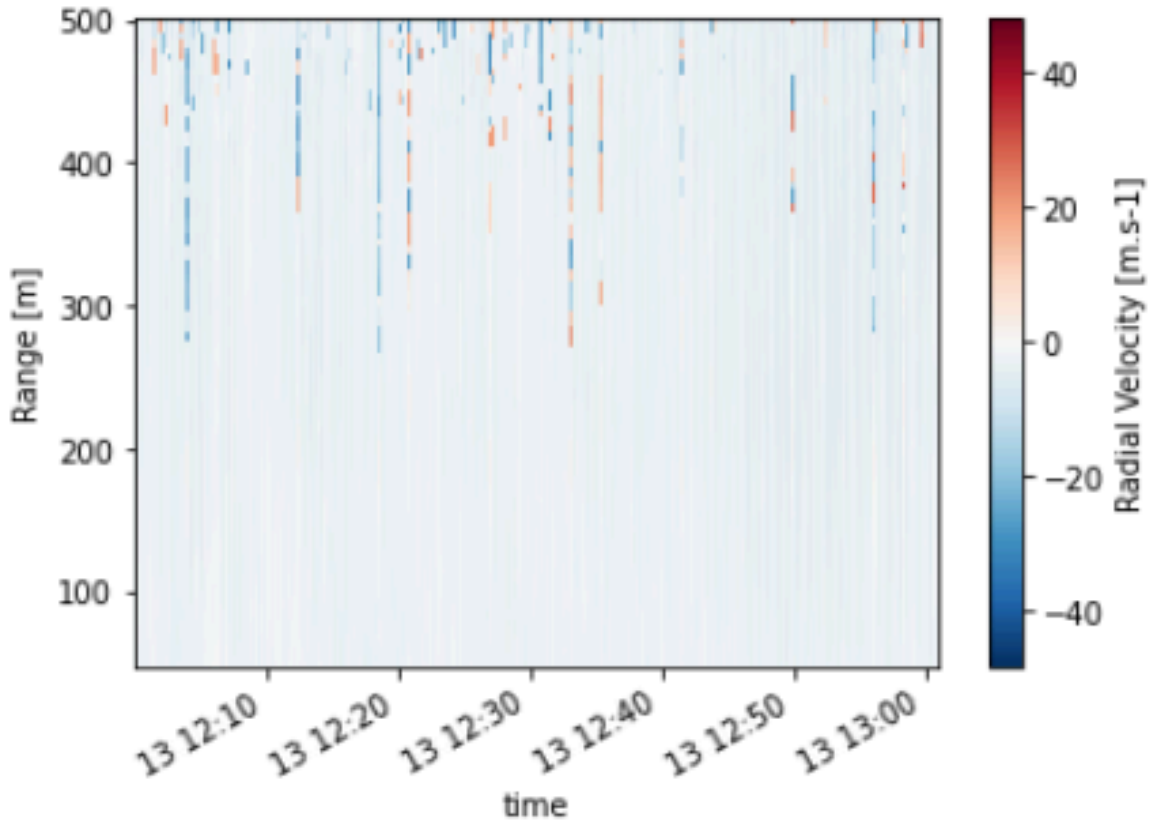
```
<xarray.Dataset>
Dimensions:                     (los_no: 180, range: 91, time: 14317)
Coordinates:
  * range                       (range) int32 50 55 60 65 70 ... 485 490 495 500
  * time                        (time) datetime64[ns] 2018-08-13T12:00:01.005000114 ... 2018-08-13T13:00:59.855000019
  * los_no                      (los_no) int32 1 2 3 4 5 ... 176 177 178 179 180
    mask                        (range, time) int64 1 1 1 1 1 1 ... 1 0 0 1 1 1
Data variables:
    radial_velocity             (range, time) float32 ...
    CNR                         (range, time) float32 -17.398 ... -23.877
    spectral_broadening         (range, time) float32 ...
    los_number                  (time) int32 142 143 144 145 146 ... 55 56 57 58
    azimuth_angle               (time) float32 117.541 121.105 ... 121.13 117.57
    elevation_angle             (time) float32 26.86 26.548 ... 10.418 10.557
    azimuth_angle_calculated    (los_no) float64 160.0 158.6 ... 115.8 112.1
    elevation_angle_calculated  (los_no) float64 2.047 2.113 ... 29.82 30.01
Attributes:
    title:                Long-range WindScanner multi-rotor wake measurements
    authors:              Nikola Vasiljević
    summary:              Data set containing 5 Hz radial velocity measuremen...
    data_owner:           DTU Wind Energy
    DOI:                  10.11583/DTU.9896459
    licence:              CC BY 4.0
    logbook:
    averaging_period:     200 ms
    temporal_resolution:  200 ms
    site:                 Risoe campus
    instrument:           Long-range WindScanner system
    instrument_id:        https://doi.org/10.3390/rs8110896
    latitude:             55.686014
    longitude:            12.097596
    height:               0
```

```
<xarray.DataArray 'radial_velocity' (range: 91, time: 14317)>
[1302847 values with dtype=float32]
Coordinates:
  * range    (range) int32 50 55 60 65 70 75 80 ... 470 475 480 485 490 495 500
  * time     (time) datetime64[ns] 2018-08-13T12:00:01.005000114 ... 2018-08-13T13:00:59.855000019
    mask     (range, time) int64 1 1 1 1 1 1 1 1 1 1 1 ... 1 1 1 1 1 1 0 0 1 1 1
Attributes:
    standard_name:  radial_velocity_of_scatterers_toward_instrument
    concept_id:     http://data.windenergy.dtu.dk/controlled-terminology/wind...
    short_name:     Vrad
    long_name:      Radial Velocity
    units:          m.s-1
```

… makes for easy processing

```
In [73]:  1  one_hour_data.radial_velocity.plot()
```

```
Out[73]: <matplotlib.collections.QuadMesh at 0x12a01ba60>
```



Why decouple data from metadata?

- Metadata can often be lightweight compared to data
  Kb instead of Gb or Tb

- Cost of maintaining metadata lower than data
  Metadata can be more persistant than data **A2**

- Convergence on FAIR Implementations

  Can use same tech stack for building/serving metadata

  Higher chance of converging to a fewer metadata standards (templates) vs jungle of “standards”

  We can use the same metadata format **I1**
  RDF is a way to go in any of its serializations, such as JSON-LD, TTL, etc.

  Leads to machine-actionablity even though data will always be depended on the underlaying data standards (not all data is SPARQL friendly)

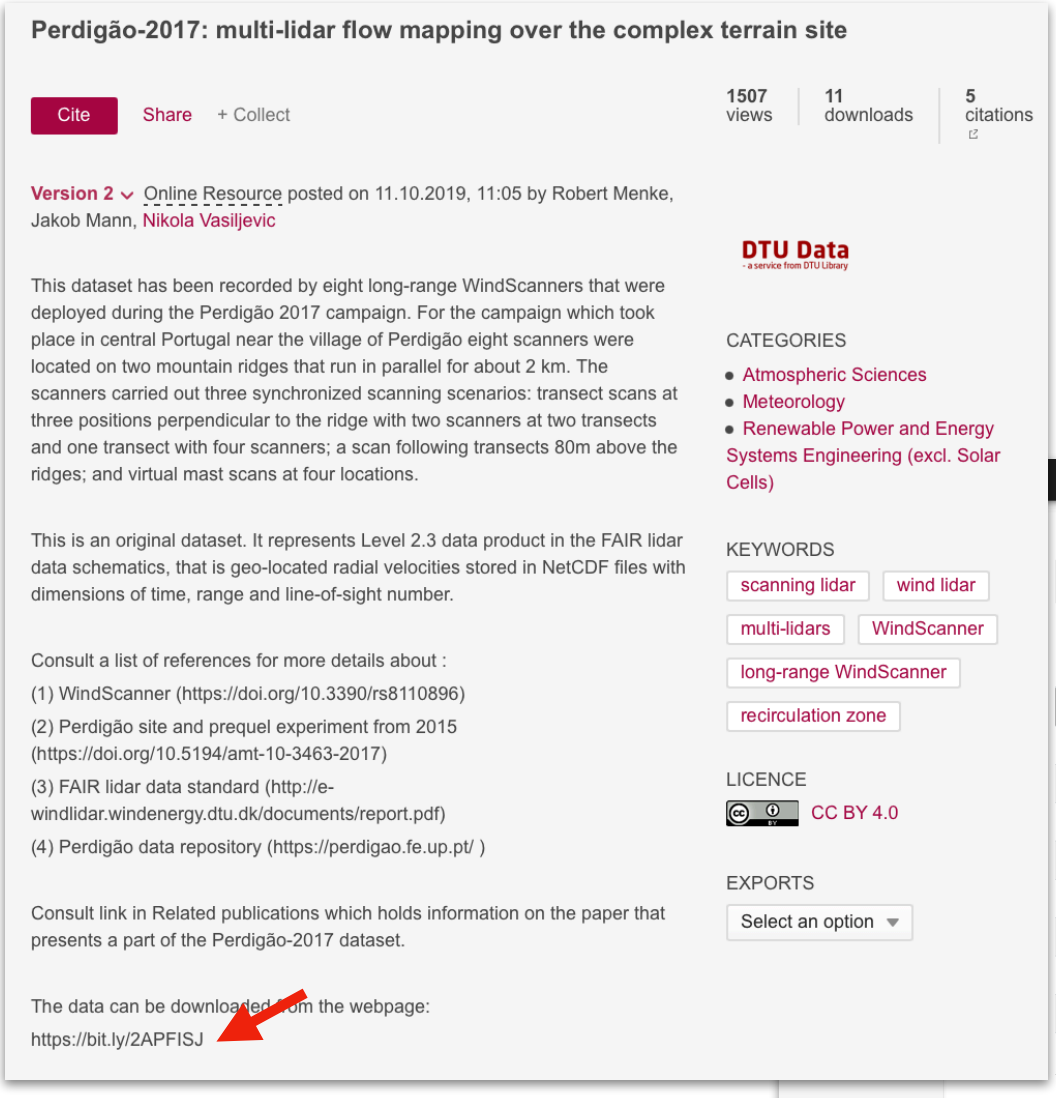
- Separation of concerns and division of responsibilitie
  Often researchers can only publish metadata (data access remains restricted).

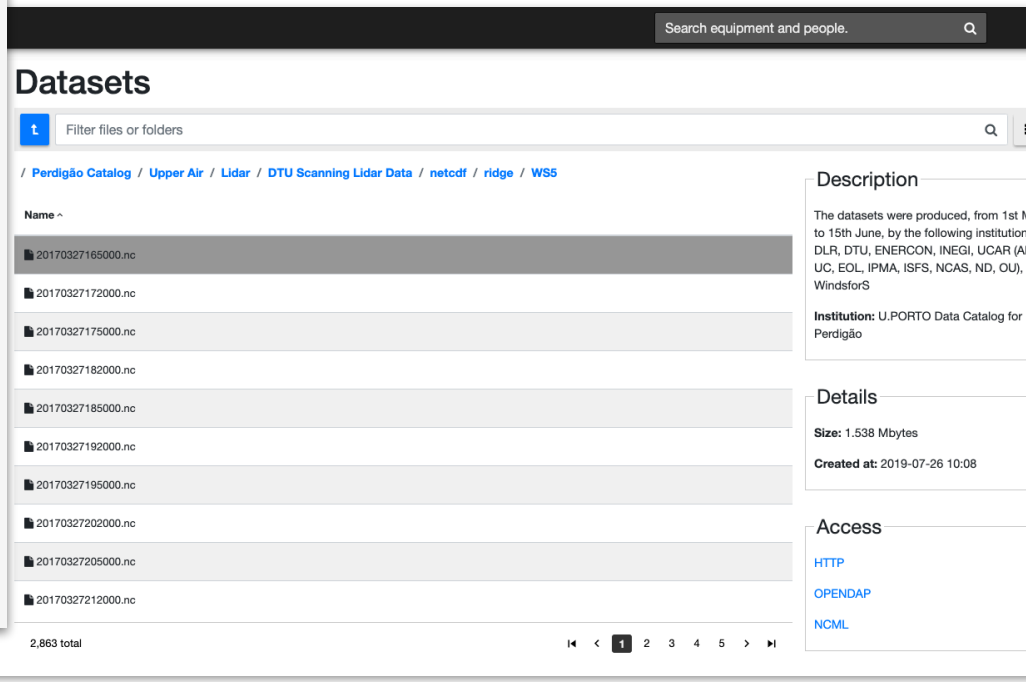
The metadata of most repositories clearly and explicitly include the identifier of the data they describe, but still fail Principle F3.

DTU Data
- a service from DTU Library

https://data.dtu.dk/articles/online_resource/Perdig_o-2017_multi-lidar_flow_mapping_over_the_complex_terrain_site/7228544

The metadata of most repositories clearly and explicitly include the identifier of the data they describe, but still fail Principle F3.





The metadata of most repositories clearly and explicitly include the identifier of the data they describe, but still fail Principle F3.

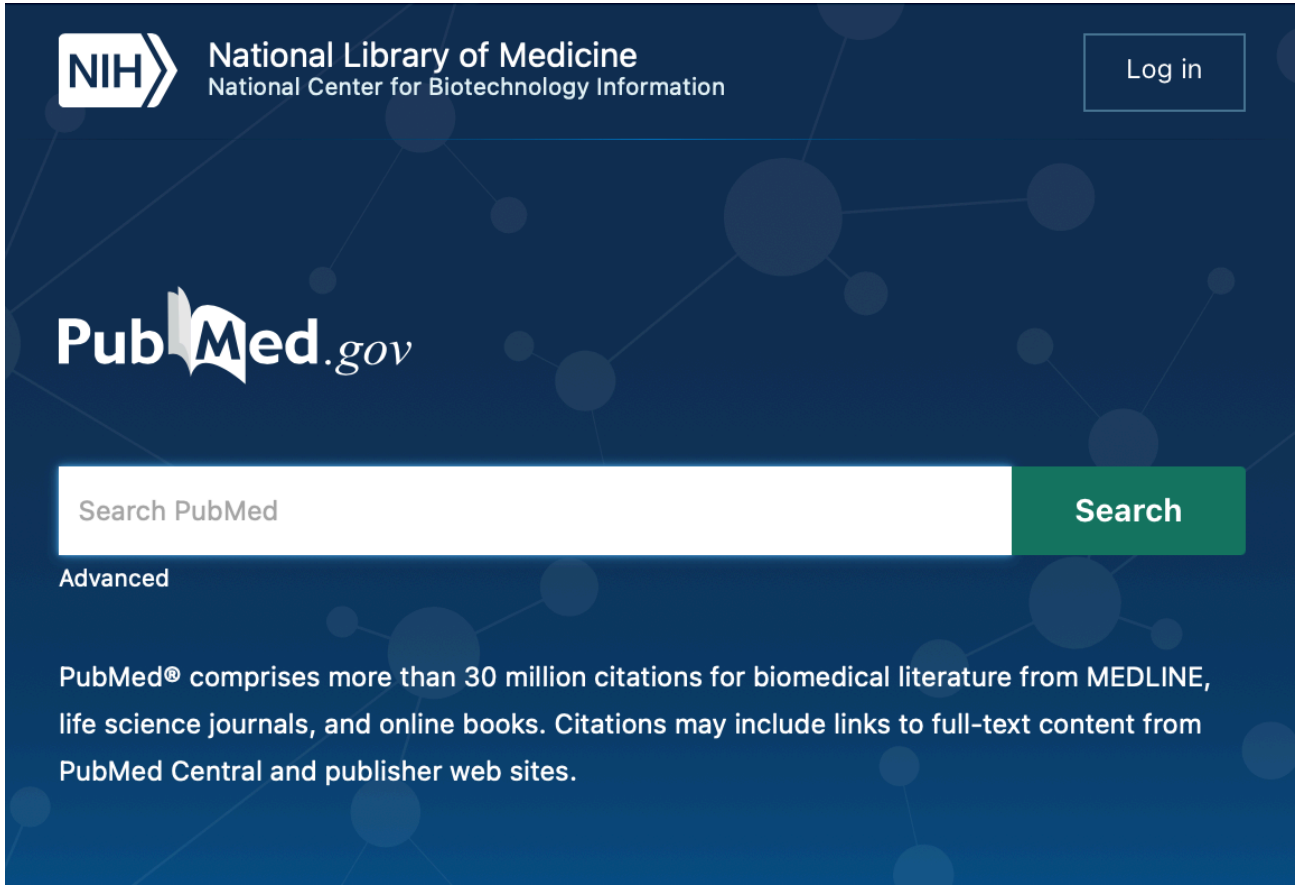
```
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/
schema/kernel-4" xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://
schema.datacite.org/meta/kernel-4.3/metadata.xsd">
<identifier identifierType="DOI">10.11583/DTU.7228544.v2</identifier>
<creators>
  <creator>
    <creatorName nameType="Personal">Robert Menke</creatorName>
    <givenName>Robert</givenName>
    <familyName>Menke</familyName>
  </creator>
  <creator>
    <creatorName nameType="Personal">Jakob Mann</creatorName>
    <givenName>Jakob</givenName>
    <familyName>Mann</familyName>
  </creator>
  <creator>
    <creatorName nameType="Personal">Nikola Vasiljevic</creatorName>
    <givenName>Nikola</givenName>
    <familyName>Vasiljevic</familyName>
  </creator>
</creators>
<titles><title>Perdigão-2017: multi-lidar flow mapping over the complex terrain site</title></
titles>
<publisher>Technical University of Denmark</publisher>
<publicationYear>2019</publicationYear>
<subjects>
  <subject>scanning lidar</subject>
  <subject>wind lidar</subject>
  <subject>multi-lidars</subject>
  <subject>WindScanner</subject>
  <subject>long-range WindScanner</subject>
```

The metadata of most repositories clearly and explicitly include the identifier of the data they describe, but still fail Principle F3.

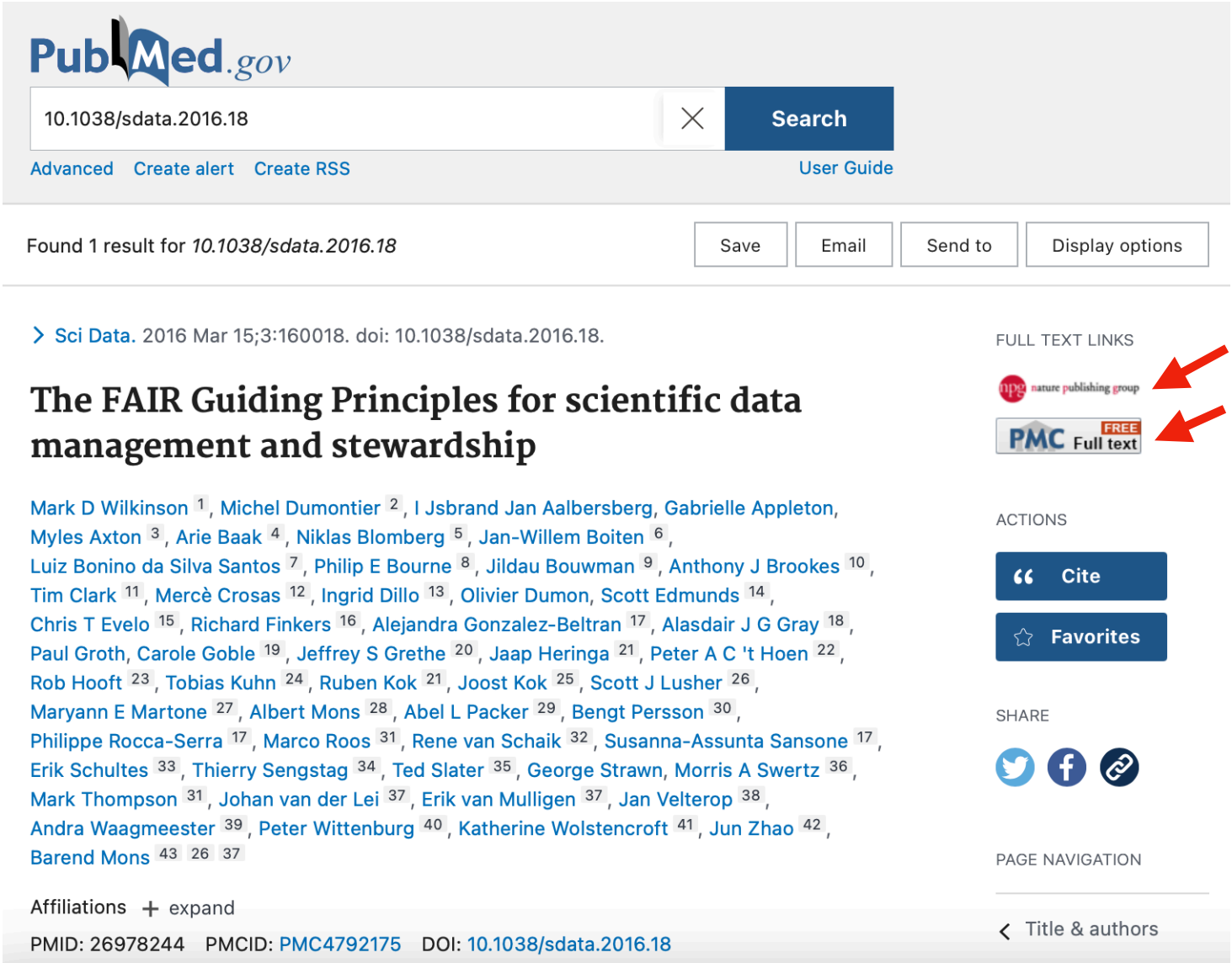
F1

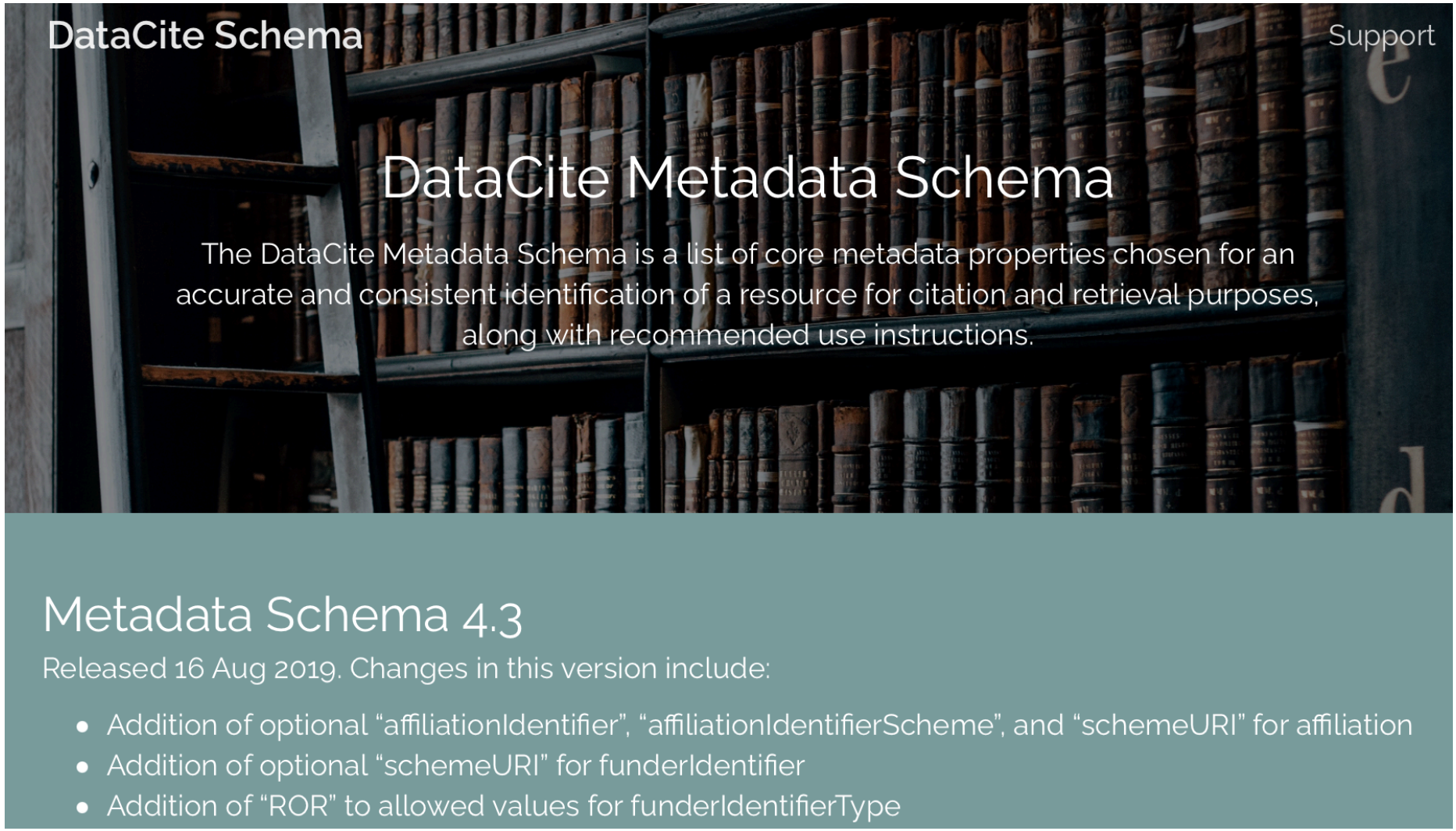
DOI: 10.1038/sdata.2016.18

F4



The metadata of most repositories clearly and explicitly include the identifier of the data they describe, but still fail Principle F3.

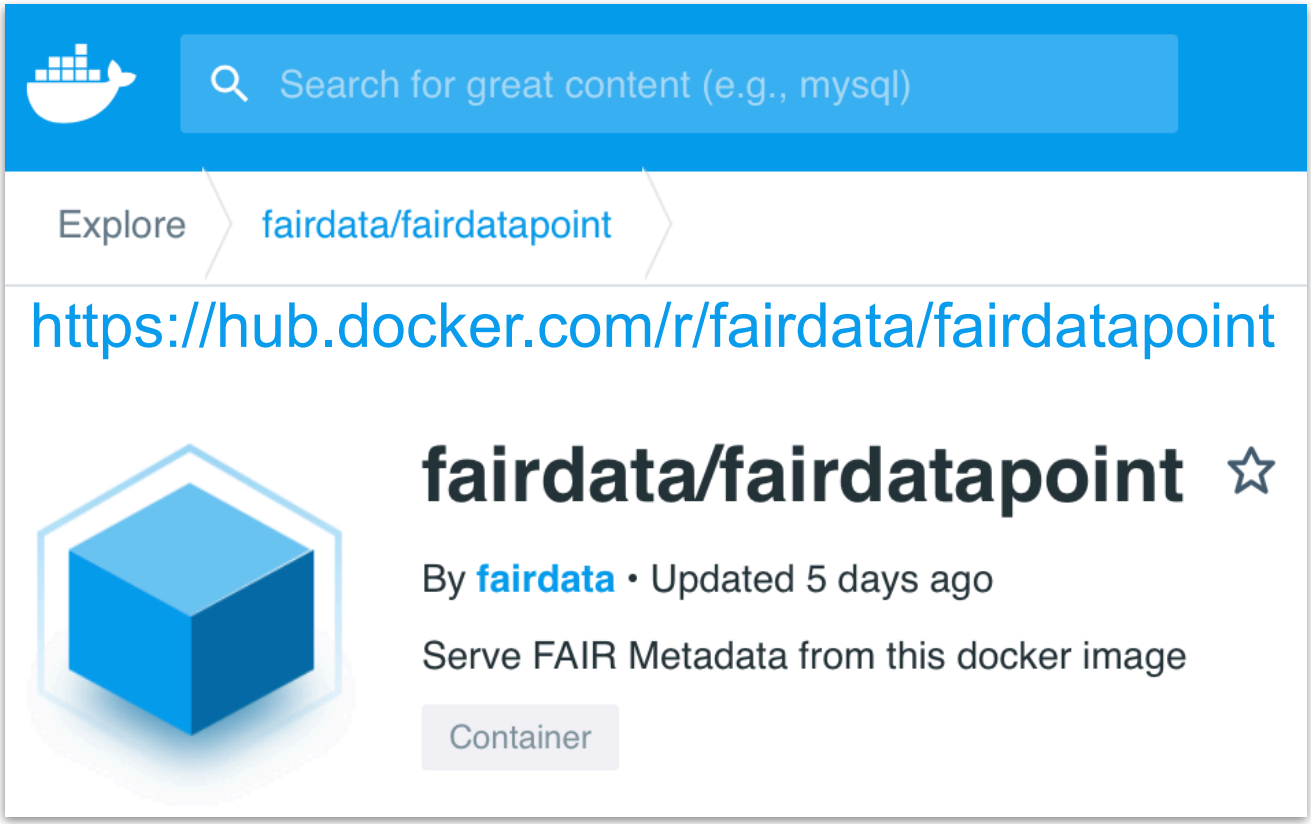




Yes please feel free to share our plans. [F3] will be implemented through our metadata working group who are currently getting our schema 4.4 released and then they will move onto these discussions.

- Matt Buys, Executive Director of Datacite, February 2 2021

What does F3 look like?



FDP:

- Metadata publication platform
- LOD
- Follows DCAT
- Adheres to F3
- Adheres to I3 (qualified links)

	Repository	Catalog	Data	Distribution
rdf:type	x	x	x	x
dct:title	x	x	x	x
dct:hasVersion		x	x	x
dct:description	x	x	x	x
dct:publisher	x	x	x	
dct:language	x	x	x	
dct:license	x	x	x	x
dct:issued		x	x	x
dct:modified		x	x	x
dct:conformsTo	x	x	x	x
dct:rights	x	x	x	x
dct:accessRights	x	x	x	x
dct:hasPart		x		
dct:isPartOf		x	x	x
dcat:themeTaxonomy		x		
dcat:contactPoint	x			
dcat:keyword	x			
dcat:theme	x			
dcat:endPointURL	x			
dcat:endPointDescription	x			
fdp:metadataIdentifier	x	x	x	x
fdp:metadataIssued	x	x		
fdp:metadataModified	x	x	x	
fdp-o:startDate	x			
fdp-o:endDate	x			
fdp-o:repositoryLanguage	x			
fdp-o:hasSoftwareVersion	x			
fdp-o:conformsToFdpSpec	x			
ldp:DirectContainer	x			
foaf:homepage		x		
rdfs:label			x	
dcat:distribution			x	
dcat:theme			x	
dcat:contactPoint			x	
dcat:keyword			x	
dcat:landingPage			x	
dcat:accessURL				x
dcat:downloadURL				x
dcat:mediaType				x
dcat:format				x
dcat:byteSize				x

https://fdp.kiu.ac.ug

FAIR Data Point

Log in

KIU FAIR Data Point

This FAIR Data Point contains the metadata of datasets and other artefacts related to the COVID-19 pandemic and other health data in Kampala International University and the rest of Uganda.

Catalogs

Covid-19 Case Report Form

Covid-19 case report forms following WHO standard

Issued 13-07-2020 **Modified** 13-07-2020

Migrant Media Reports

These are national and international media reports published during the covid-19 crisis.

Issued 13-07-2020 **Modified** 13-07-2020

Metadata Issued
01-07-2020

Metadata Modified
13-07-2020

Version
1.0

Language
en

License
cc-by-nc-nd3.0

Institution
kiu.ac.ug

Start date
03-06-2020

Last update
13-07-2020

Institution country
Q1036

https://fdp.kiu.ac.ug

FDP FAIR FAIR Data Point

Log in

KIU FAIR Data Point

This FAIR Data Point contains the metadata of datasets and other artefacts related to the COVID-19 pandemic and other health data in Kampala International University and the rest of Uganda.

Catalogs

Covid-19 Case Report Form

Covid-19 case report forms following WHO standard

Issued 13-07-2020 **Modified** 13-07-2020

Migrant Media Reports

These are national and international media reports published during the covid-19 crisis.

Issued 13-07-2020 **Modified** 13-07-2020

Metadata Issued
01-07-2020

Metadata Modified
13-07-2020

Version
1.0

Language
en

License
cc-by-nc-nd3.0

Institution
kiu.ac.ug

Start date
03-06-2020

Last update
13-07-2020

Institution country
Q1036

https://fdp.kiu.ac.ug

FAIR Data Point

Log in

Catolog

KIU FAIR Data Point / Migrant Media Reports

Migrant Media Reports

These are national and international media reports published during the covid-19 crisis.

Datasets

Media Report Dataset

These datasets are media reports concerning events related to the impact of covid-19 crisis on immigrants in Tunisia.

Q38926 Q57979909

Issued 13-07-2020 **Modified** 13-07-2020

Metadata Issued
13-07-2020

Metadata Modified
13-07-2020

Version
1.0

Language
en

License

Issued
13-07-2020

Modified
13-07-2020

Theme taxonomy

- Q38926
- Q57979909

https://fdp.kiu.ac.ug

FAIR Data Point

Log in

Dataset

KIU FAIR Data Point / Migrant Media Reports / Media Report Dataset

Media Report Dataset

These datasets are media reports concerning events related to the impact of covid-19 crisis on immigrants in Tunisia.

Distributions

Media Reports listed in a google document

These are internet links to media reports listed in a google document.

Issued 13-07-2020 **Modified** 13-07-2020 **Media Type** https://w3id.org/mediatype/text/html

Metadata Issued
13-07-2020

Metadata Modified
13-07-2020

Version
1.0

Language
en

License

Issued
13-07-2020

Modified
13-07-2020

Theme

- Q38926
- Q57979909

https://fdp.kiu.ac.ug

FAIR Data Point

Log in

KIU FAIR Data Point / Migrant Media Reports / Media Report Dataset / Media Reports listed in a google do...

Media Reports listed in a google document

These are internet links to media reports listed in a google document.

Access online

Access to the data

Metadata Issued
13-07-2020

Metadata Modified
13-07-2020

Version
1.0

Language
en

License

Issued
13-07-2020

Modified
13-07-2020

Media type
https://w3id.org/mediatype/text/html

Tunisie

- **Facts**

Reportages sur la situation des migrants
https://www.theguardian.com/global-development/2020/apr/16/we-have-nothing-as-lockdown-bites-migrants-in-tunisia-feel-the-pinch

Yossra Kallali: Réflexion sur l'impact de Covid-19 sur les migrants, et la migration

https://www.leaders.com.tn/article/30036-yossra-kallali-reflexion-sur-l-impact-de-covid-19-sur-les-migrants-et-la-migration

https://nawaat.org/portail/2020/04/23/confinement-la-detresse-des-migrants-subsahariens-en-tunisie/

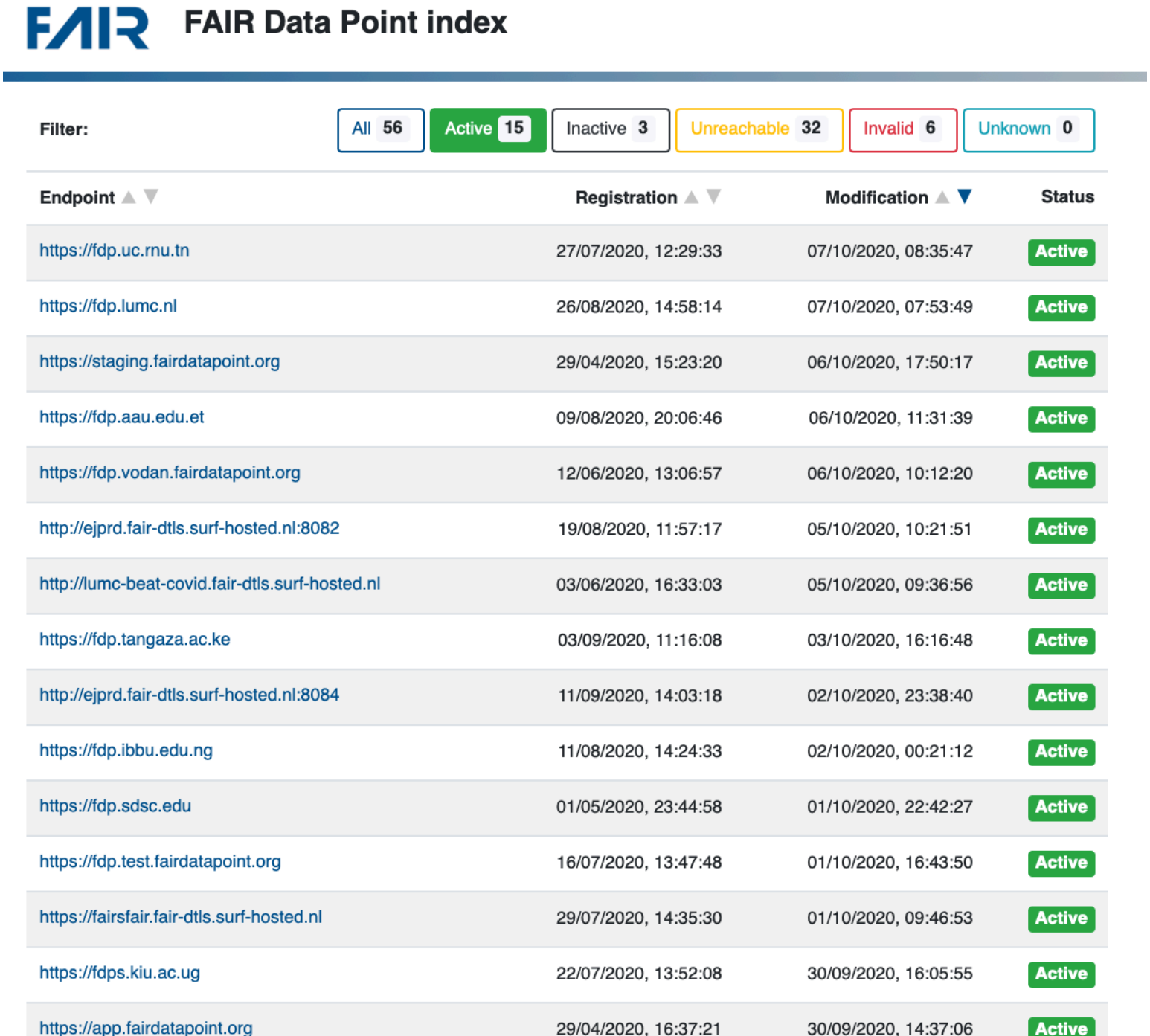
Les étudiants Tchadiens quittent le pays
https://www.afrikyes.com/actualite/vie-associative/depart-des-etudiants-tchadiens-de-la-tunisie/

Les subsahriens prient pour la Tunisie
https://www.youtube.com/watch?v=Pu3V-gZ0Cy8&feature=youtu.be

En Tunisie, les migrants subsahariens démunis face à la pandémie de coronavirus

https://www.jeuneafrique.com/915161/societe/en-tunisie-les-migrants-subsahariens-demunis-face-a-la-pandemie-de-coronavirus/

https://home.fairdatapoint.org



Thank you
&
Questions?