FAIRify Your Data

Jessica Rex, Cora Assmann

LOVE DATA WEEK 14.02.2023

Herzlich Willkommen!

1. Begrüßung und Vorstellung des Netzwerks
2. Die FAIR Data Prinzipien
3. Einführung FAIR Assessment Tools
4. FAIR Evaluierung mit dem ARDC Tool
5. FAIR Evaluierung mit dem F-UJI Tool
6. FAIR Evaluierung eines eigenen Datensatzes
7. Diskussion und Wrap-Up

TKFDM - Eckdaten

Roman Gerlach
Friedrich-Schiller-Universität Jena

Dr. Nadine Neute
Universität Erfurt

Dr. rer. nat. Cora Assmann
Friedrich-Schiller-Universität Jena

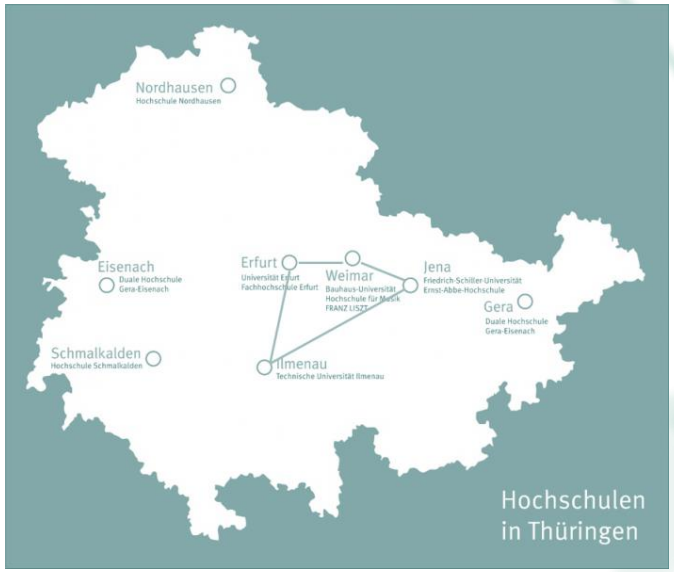
Jessica Rex
Technische Universität Ilmenau

Kevin Lang
Bauhaus-Universität Weimar

Dr.-Ing. Steve Göring
Technische Universität Ilmenau

Stefan Kirsch
Ernst-Abbe-Hochschule Jena

- An 4 Universitäten und 4 Fachhochschulen und HAWs (für die 10 Hochschulen in Thüringen)
- 4 Kontaktstellen + 2 Data Stewards + 1 Koordinationsstelle an der FSU Jena
- seit Oktober 2018 (Digitalisierungsstrategie im Hochschulbereich) und bis 2025 fortfinanziert (Fortschreibung 2021-2025)
- unsere Webseite: forschungsdaten-thueringen.de



TKFDM - Arbeitsfelder

- **Beratung**
  - FDM Unterstützung und Beratung für alle Thüringer Forschenden
- **Schulung**
  - Informations- und Outreachveranstaltungen, Workshops und Coffee Lectures
  - Informations- und Schulungsmaterialien
- **Vernetzung**
  - zwischen den Universitäten, Forschungsinstituten, Infrastruktureinrichtungen, den NFDI Konsortien und der FDM Community

TKFDM - Herausforderungen

Woran wir in Thüringen gerade arbeiten:

- Konzeption und Etablierung eines Repositorium sowie eines Langzeitarchivs für Forschungsdaten
  - *Große Datenmengen, Metadaten, Langzeitarchivierung, etc.*
- Neues Data Steward Pilotprojekt
- Integration neuer Kolleg:innen an den Fachhochschulen
- Rolle der Landesinitiativen in der NFDI
- Veranstaltungen und Materialien

Materialien und Veranstaltungen

Veranstaltungen

- Thüringer FDM-Tage
- FAIRest Dataset Award
- Coffee Lectures
- Data Horror Week

Materialien

- Handreichungen & Best-Practices
- Elektronische Laborbücher
- 23 Dinge zum Forschungsdatenmanagement
- Kartenspiel „Scary Tales“
- 5S Data

Die FAIR Data Prinzipien

https://www.go-fair.org/fair-principles/

Daten sollten

Findable **einfach auffindbar**
Accessible **leicht zugänglich**
Interoperable **mit anderen Systemen interoperabel**
Reusable **nachnutzbar**

sein

Die FAIR Data Prinzipien

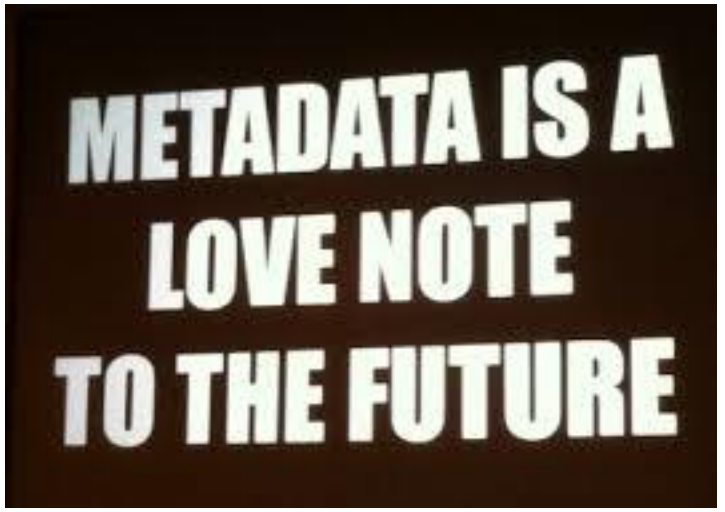
Findable:

- Daten und Metadaten sind sowohl menschen- als auch maschinenlesbar
- Metadaten sind aussagekräftig
- die(Meta)Daten haben einen persistenten Identifikator

Die FAIR Data Prinzipien

Accessible:

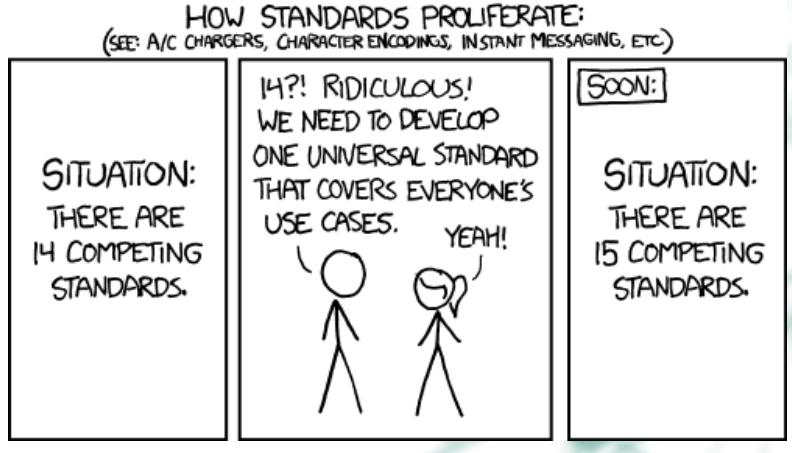
- Daten sind über ein Standardprotokoll (bspw. https) online verfügbar und abrufbar
- wenn die Daten nicht online gestellt werden können, sind zumindest die Metadaten offen verfügbar



Die FAIR Data Prinzipien

Interoperable:

- Verwendung von Ontologien, Thesauri und kontrollierten Vokabularien
- Verwendung von möglichst standardisierten, offenen Datenformaten
- Verwendung standardisierter, maschinenlesbarer Formate für Metadaten (XML, JSON)
- Verknüpfung verwandter Datensätze in den Metadaten

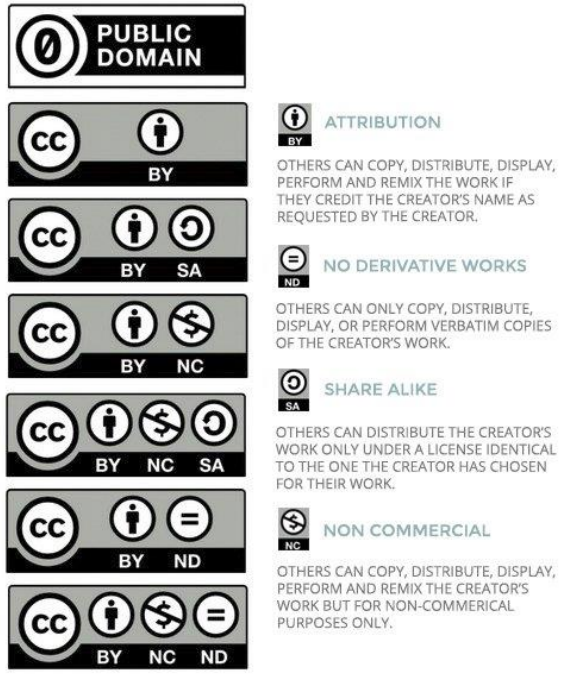


https://xkcd.com/927/

Die FAIR Data Prinzipien

Reusable:

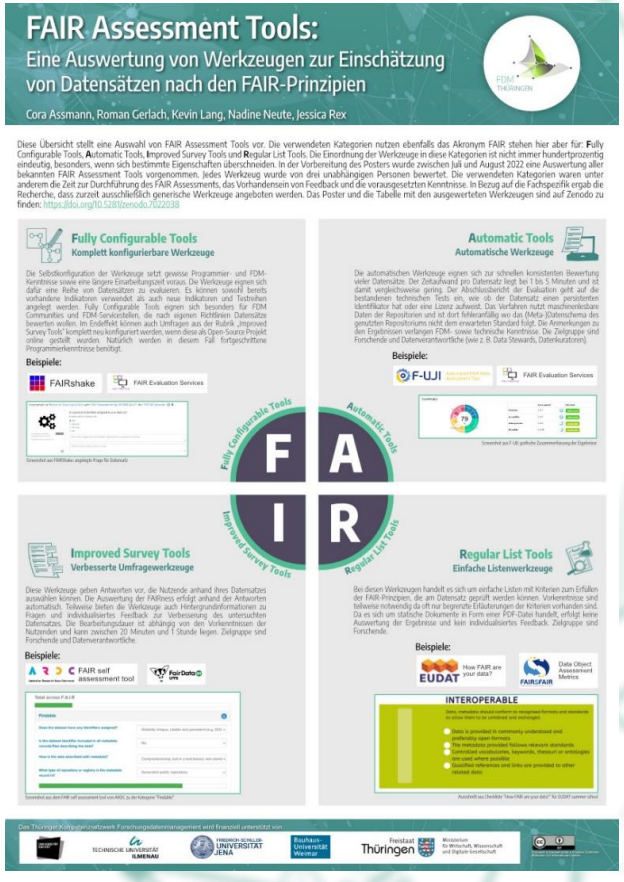
- ausführliche Datendokumentation
- Angabe von Provenienzdaten
- Angabe von Lizenzen und Nachnutzungsinformationen



Die FAIR Data Prinzipien

Gut zu wissen:

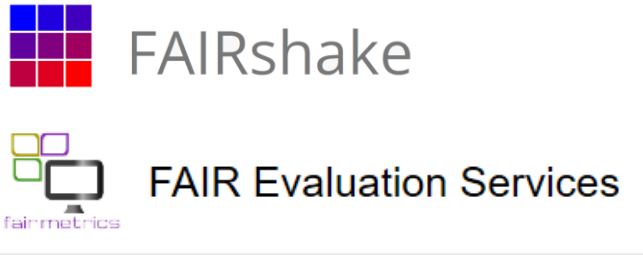
- FAIR ≠ OPEN
- Einhaltung der FAIR-Prinzipien ist mittlerweile oft obligatorisch – relevante Leitlinien, Policies und Förderkriterien checken!
- Die Buchstaben F und A sind einfacher zu erfüllen – I und R benötigen meist etwas mehr Planung und Aufwand
- für Überprüfung der FAIRness der eigenen Daten gibt es FAIR (Self) Assessment Tools
- [FAIR Assessment Tools Poster des TKFDM]



FAIR Assessment Tools

Komplett konfigurierbare Werkzeuge

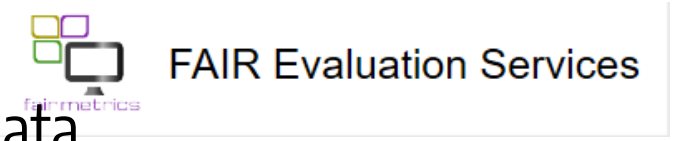
- Beispiele:
  - FAIRshake
  - FAIR Evaluation Services
- eignen sich besonders für die FDM Community und FDM Servicestellen
- Vorteile:
  - Evaluation einer Reihe von Datensätzen möglich
  - eigene Indikatoren und Testreihen können angelegt werden
- Nachteile:
  - setzen oft Programmier- und FDM-Kenntnisse voraus



FAIR Assessment Tools

Automatische Werkzeuge

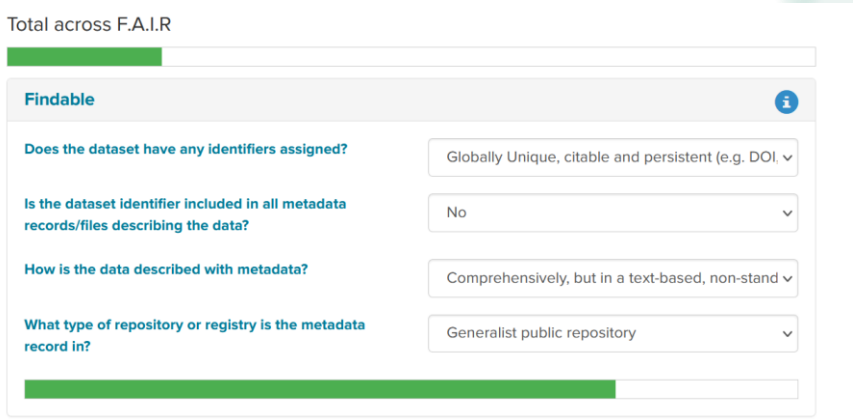
- Beispiele:
  - F-UJI Automated Fair Data Assessment Tool
  - FAIR Evaluation Services
- Zielgruppe: Forschende und Datenverantwortliche wie Data Stewards
- Vorteile:
  - Schnelle und konsistente Bewertung vieler Datensätze
  - geringer Aufwand
- Nachteile:
  - fehleranfällig bei Repositorien, die keine standardisierten Metadatenschemata nutzen



FAIR Assessment Tools

Verbesserte Umfragewerkzeuge

- Beispiele:
  - ARDC FAIR Self Assessment Tool
  - FairDataBR
- Zielgruppe: Forschende und Datenverantwortliche
- Vorteile:
  - bieten teilweise Hintergrundinformationen zu den Fragen und individualisiertes Feedback
  - Auswertung erfolgt anhand der Antworten automatisch
- Nachteile:
  - längere Bearbeitungsdauer zwischen 20 min und 1h (abhängig von Vorkenntnissen)



Screenshot aus dem ARDC Tool

FAIR Assessment Tools

Einfache Listenwerkzeuge

- Einfache Listen mit Kriterien
- Beispiele:
  - EUDAT How fair are your data? Checkliste
  - FAIRsFAIR Data Object Assessment Metrics
- Zielgruppe: Forschende
- Vorteile
  - keine Programmierkenntnisse erforderlich
- Nachteile:
  - keine automatisierte Auswertung/individualisiertes Feedback
  - begrenzte Erläuterung der Kriterien - FDM-Vorkenntnisse erforderlich

FAIR Bewertung mit dem ARDC Tool

Evaluations Tool

Australian Research Data Commons

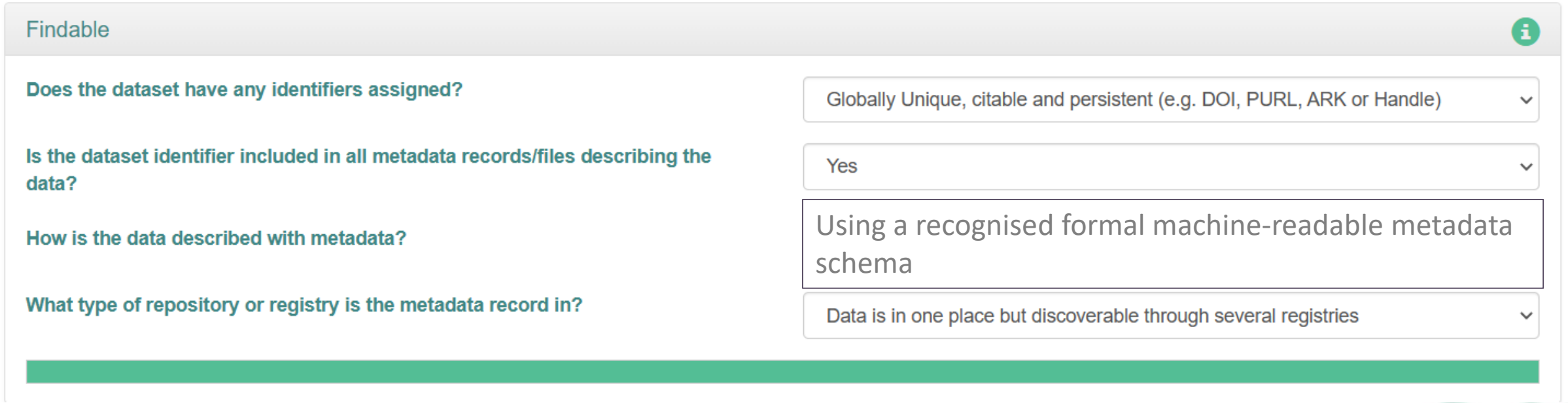
Offizielle Seite:
https://ardc.edu.au/resource/fair-data-self-assessment-tool/

Publizierter Datensatz

Link zum Datensatz:
https://zenodo.org/record/4064019#.Y85M_3bMJPa

DOI: https://doi.org/10.5281/zenodo.4064019

Findable


Findable
Does the dataset have any identifiers assigned?
Globally Unique, citable and persistent (e.g. DOI, PURL, ARK or Handle)
Is the dataset identifier included in all metadata records/files describing the data?
Yes
How is the data described with metadata?
Using a recognised formal machine-readable metadata schema
What type of repository or registry is the metadata record in?
Data is in one place but discoverable through several registries

JSON Export der Metadaten

JSON Export

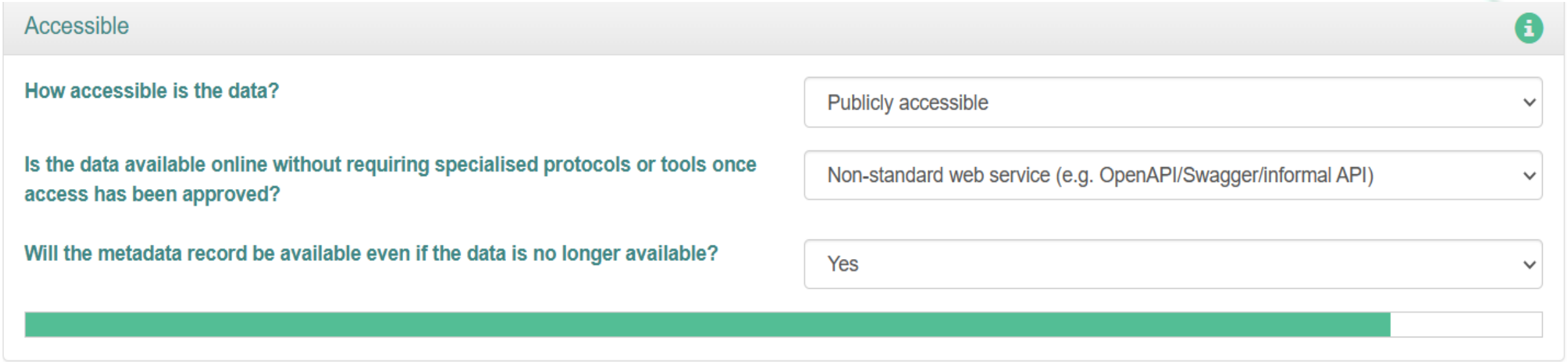

```
{
  "files": [
    {
      "links": {
        "self": "https://zenodo.org/api/files/f46cd659-7a0c-42fa-8811-97d11c61a87f/page-quality-interf
```

```
        "relation": "isDocumentedBy",
        "resource_type": "publication-article"
      },
      {
        "scheme": "doi",
        "identifier": "10.5281/zenodo.1002203",
        "relation": "isVersionOf"
      }
    ]
  }
}
```

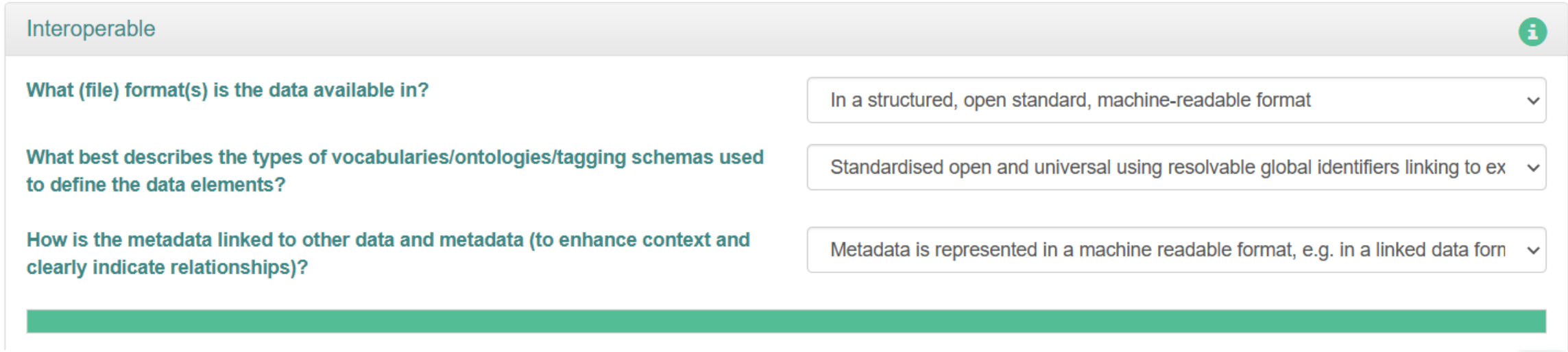

Export

BibTeX CSL DataCite Dublin Core DCAT
JSON JSON-LD GeoJSON MARCXML
Mendeley

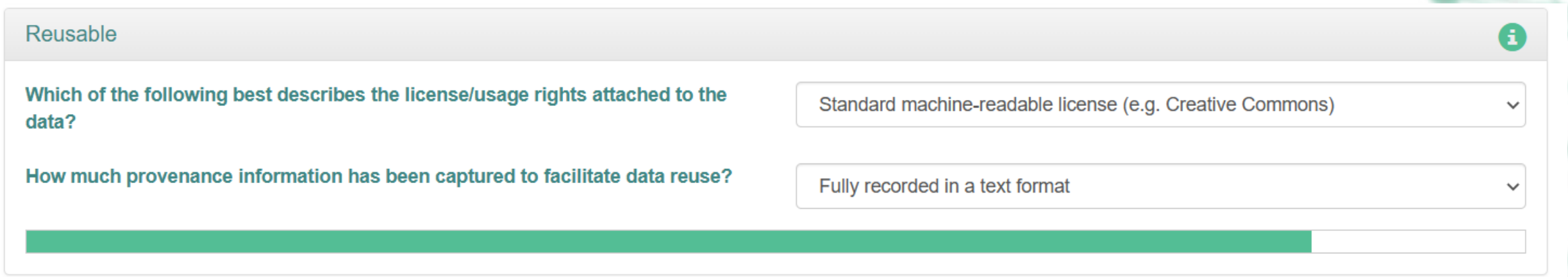
Accessible


Accessible
How accessible is the data?
Publicly accessible
Is the data available online without requiring specialised protocols or tools once access has been approved?
Non-standard web service (e.g. OpenAPI/Swagger/informal API)
Will the metadata record be available even if the data is no longer available?
Yes

Interoperable



Reusable


Reusable
Which of the following best describes the license/usage rights attached to the data?
Standard machine-readable license (e.g. Creative Commons)
How much provenance information has been captured to facilitate data reuse?
Fully recorded in a text format

Endergebnis

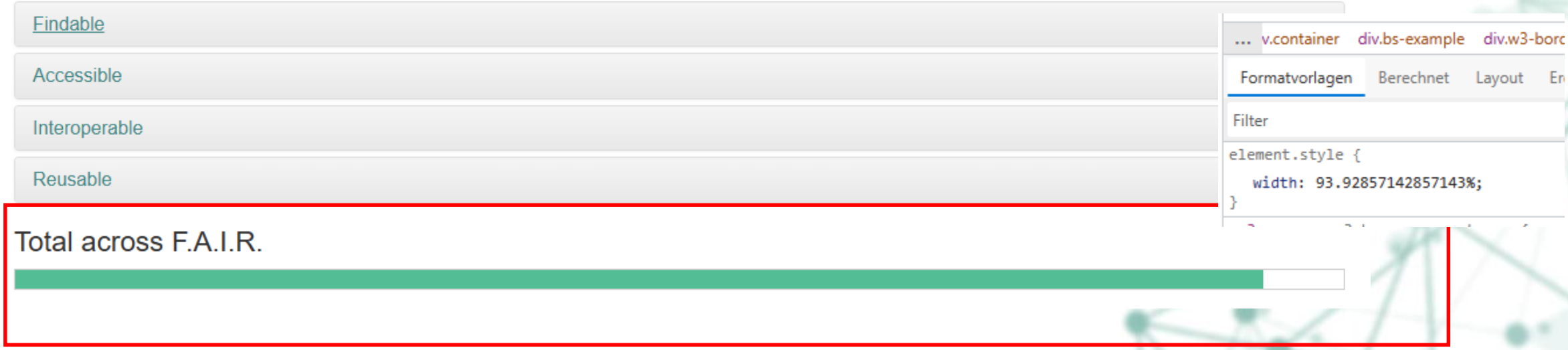
FAIR Assessment Tool

Welcome to the ARDC FAIR Data self-assessment tool. Using this tool you will be able to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).

This self-assessment tool has been designed predominantly for data librarians and IT staff, but could be used by software engineers developing FAIR Data tools and services, and researchers provided they have assistance from research support staff.

You will be asked questions related to the principles underpinning Findable, Accessible, Interoperable and Reusable. Once you have answered all the questions in each section you will be given a 'green bar' indicator based on your answers in that section, and when all sections are completed, an overall 'FAIRness' indicator is provided.

Please be aware that additional explanatory information is provided within the tool. The (i) information button provides an overview of each of the FAIR high-level elements (Findable, Accessible, Interoperable and Reusable). Additionally, each question is hyperlinked, leading users to explanatory information and links to wider resources on related topics.



94 %

FAIR Bewertung mit dem F-UJI Tool

Publizierter Dataensatz

https://doi.org/10.5281/zenodo.4064019

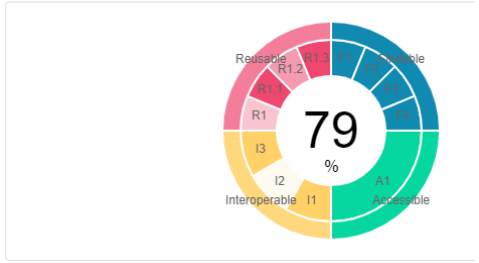
Bewertungs Tool

F-UJI

F-UJI

Bewertungsergebnis

Summary:



Begleitdokument für F-UJI „Data Object Assessment Metrics“
https://doi.org/10.5281/zenodo.6461229

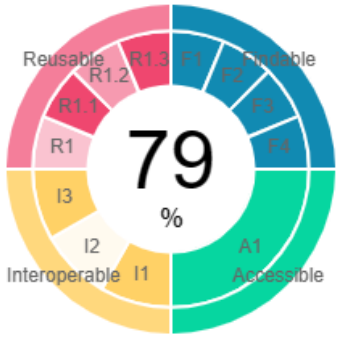
FAIR Bewertung mit dem F-UJI Tool

Webis-Web-Archive-17

Kiesel, Johannes; Potthast, Martin; Hagen, Matthias; Kneist, Florian; Stein, Benno

The Webis-Web-Archive-17 comprises a total of 10,000 web page archives from mid-2017 that were carefully sampled from the Common Crawl to involve a mixture of high-ranking and low-ranking web pages. The dataset contains the web archive files, HTML DOM, and screenshots of each web page, as well as per-page annotations of visual web archive quality. See this overview for all datasets that built upon this one. If you use this dataset in your research, please cite it using this paper.

Summary:



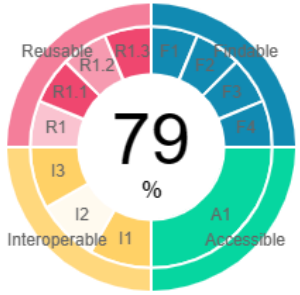
	Score earned:	Fair level:
Findable:	7 of 7	advanced
Accessible:	3 of 3	advanced
Interoperable:	3 of 4	moderate
Reusable:	6 of 10	moderate

Vergleich F-UJI und ARDC Bewertung

F-UJI

Findable	7 of 7	advanced
Accessible	3 of 3	advanced
Interoperable	3 of 4	moderate
Reusable	6 of 10	moderate

Endergebnis



ARDC

Findable	100 %
Accessible	90 %
Interoperable	100 %
Reusable	86 %

Endergebnis 94%

Teste deinen eigenen Datensatz – 30 min Zeit

eigener publizierter Datensatz?
oder
eigener Testdatensatz

oder
Testdatensatz
von uns

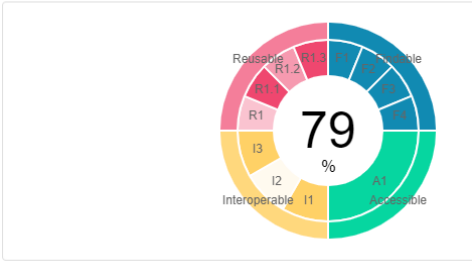
Tools

Australian Research Data Commons

F-UJI

Ergebnis

Summary:



Diskussion

ARDC: https://ardc.edu.au/resource/fair-data-self-assessment-tool/
F-UJI: https://www.f-uji.net/

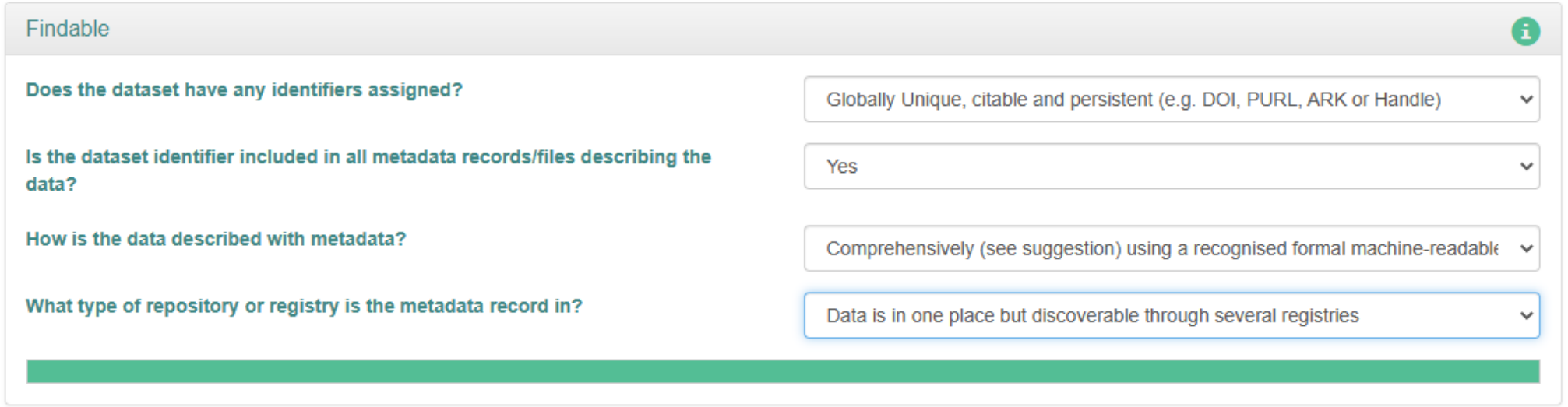
Testdatensatz

Web Data Commons (December 2020) Property and Datatype Usage Dataset

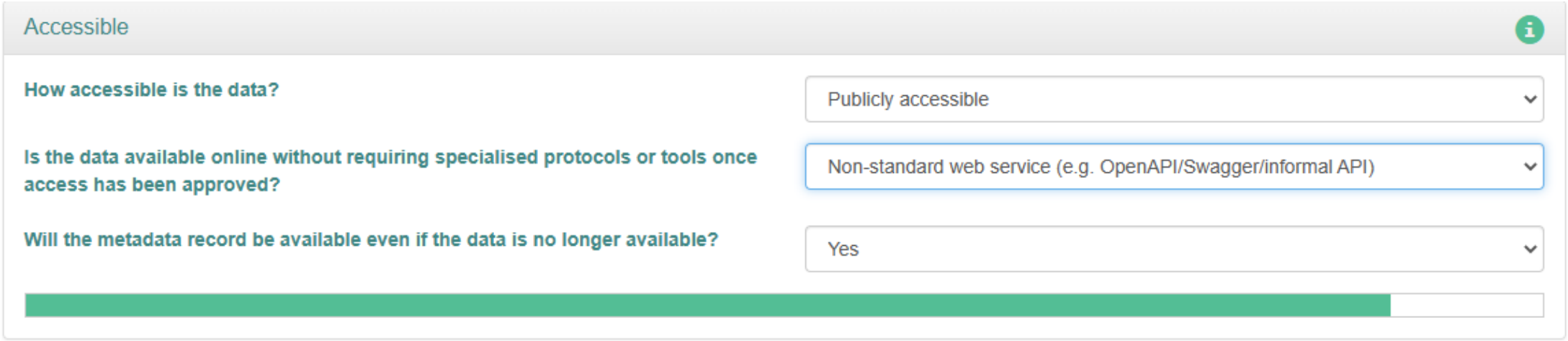
Jan Martin Keil; Merle Gänßinger

https://doi.org/10.5281/zenodo.6337661
https://zenodo.org/record/6337661#.Y-9IBHbMJPY

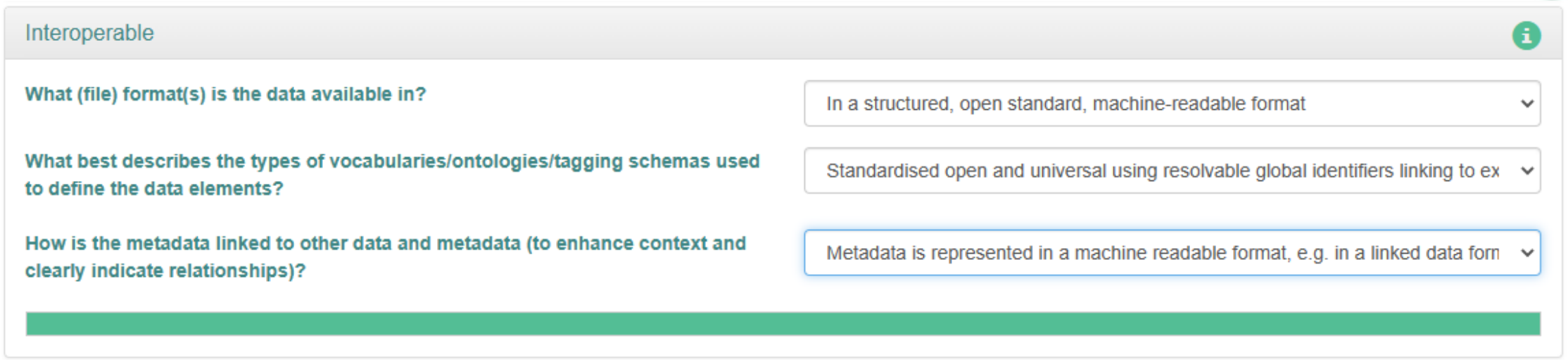
Findable


Findable
Does the dataset have any identifiers assigned?
Globally Unique, citable and persistent (e.g. DOI, PURL, ARK or Handle)
Is the dataset identifier included in all metadata records/files describing the data?
Yes
How is the data described with metadata?
Comprehensively (see suggestion) using a recognised formal machine-readable
What type of repository or registry is the metadata record in?
Data is in one place but discoverable through several registries

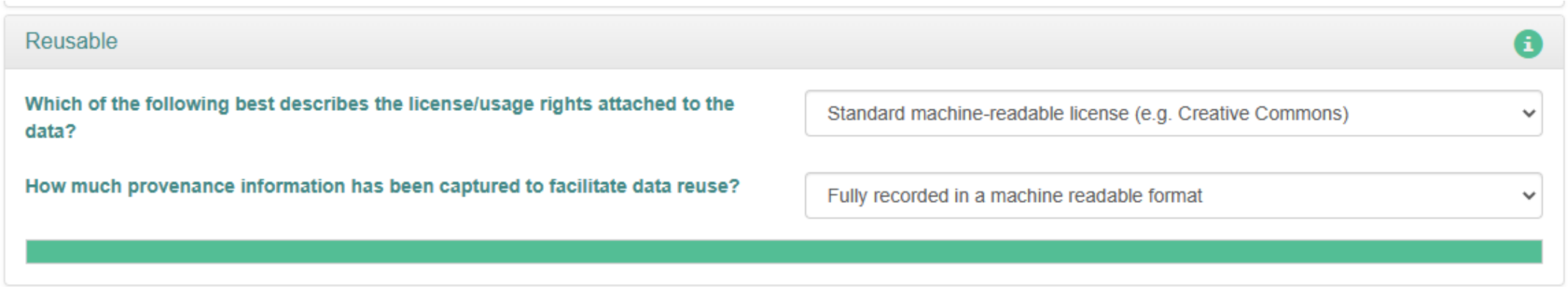
Accessible


Accessible
How accessible is the data?
Publicly accessible
Is the data available online without requiring specialised protocols or tools once access has been approved?
Non-standard web service (e.g. OpenAPI/Swagger/informal API)
Will the metadata record be available even if the data is no longer available?
Yes

Interoperable



Reusable



Total across F.A.I.R.

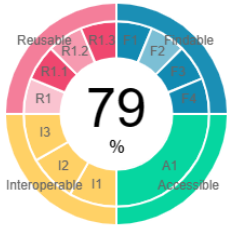
97%

Vergleich F-UJI und ARDC Bewertung

F-UJI

	Score earned:	Fair level:
Findable	6 of 7	advanced
Accessible	3 of 3	advanced
Interoperable	4 of 4	advanced
Reusable	6 of 10	moderate

Endergebnis



ARDC

Findable	100 %
Accessible	90 %
Interoperable	100 %
Reusable	100 %

Endergebnis 97 %

Ankündigung FAIRest Dataset Award 2023 des TKFDMs

"FAIRest Dataset" Award 2023

bis zu **2.000 Euro** für den
FAIResten Datensatzt Thüringens
Bewerbungszeit:
Mitte Februar bis Ende April 2023

Mehr Informationen gibt es hier:
https://forschungsdaten-thueringen.de/nachricht/basically-fair-tkfdm-schreibt-4-fairest-dataset-wettbewerb-aus.html

FDM THÜRINGEN

Kommende Veranstaltungen des TKFDMs

TKFDM Coffee Lecture , 22. Februar um 14- 14:30 Uhr

Thema: ***Forschungsdaten publizieren, lizenzieren und nachnutzen***

Anmeldung: https://forschungsdaten-thueringen.de/veranstaltung/coffee-lecture-publizieren-lizenzieren-nachnutzen-de.html

Vielen Dank für Ihre Aufmerksamkeit.