

FAIR-Checker

<https://frama.link/ifb-ag20-modele>

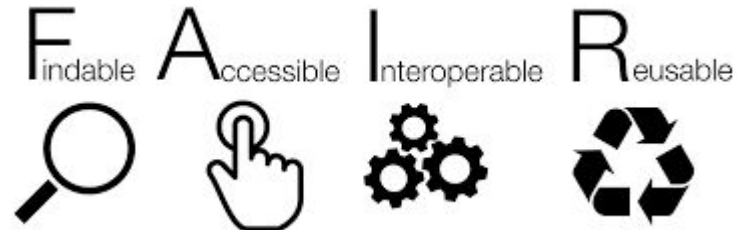
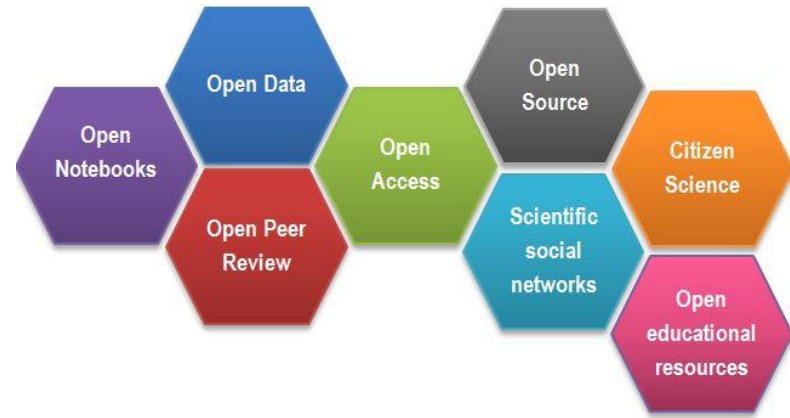
Thomas Rosnet, Marie-Dominique Devignes, Alban Gaignard

Open science allow **free** access to experimental **data** and **metadata** for anyone.

- Better reactivity to the latest breakthrough by the scientific community (e.g. COVID-19).
- Allow large scale meta-analysis for more accurate results.
- etc...

Need for standardize and structured data to ease:

- Findability
- Accessibility
- Interoperability
- Reusability



FAIR principles : how to implement guidelines ?



From Australian Research Data
Commons

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

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

FAIR principles

- many guidelines ...
- ... but no recommended technologies !

How to implement the principles and go
beyond checklists ?

Resource providers & developers need
help, and **tooling** !

- I'm a **software developer**, my source code is on GitHub, but not mature enough to be part of a registry. How to check if my tool is **findable** enough ?
 - Which kind of metadata should I advertise ?
 - Which vocabulary should I use ?
 - Where should I request a permanent ID ?

- I'm a **data producer**, I published my dataset through an online repository, does it provides rich metadata ? are these metadata **interoperable** ?

Issue: How to help researchers with technical tips to enhance the FAIRness of their resource -> data, softwares, workflows ?



Goal: Web tool **FAIR-Checker** - Monitoring progress in FAIRification through self-assessment of resources maturity indicators.

Based on the **FAIRMetrics** framework/API [Wilkinson, Dumontier et al., Scientific Data 6:174] to **automatically** evaluate **FAIR Principles** and provide hints to **progress** in the **FAIRification** process.

The screenshot shows the homepage of the FAIR Evaluation Services website. At the top, there is a navigation bar with links for HOME, EVALUATIONS, MI TESTS, COLLECTIONS, and ABOUT US, along with a search bar. The main header features the title "FAIR Evaluation Services" and a subtitle "Resources and guidelines to assess the FAIRness of digital resources." Below this, a green banner contains a message: "We are back online ! Thank you for your patience ! If you notice any unexpected failures in the tests, please report them to mark.wilkinson@upm.es". The main content area is divided into three columns, each with an icon and a "Get started" button: "Import MI Tests" (gear icon), "Create collections" (list icon), and "Evaluate resources" (scales icon). At the bottom, there is a section titled "Philosophy of FAIR testing" with a paragraph of text.

FAIR Evaluation Services

- Computable FAIR metrics
- Open to contributions (new metrics)
- **Technical expertise needed to interpret logs**

<https://fairsharing.github.io/FAIR-Evaluator-FrontEnd>

<https://fair-checker.france-bioinformatique.fr>

1. Submission of a resource (URL / DOI) with minimal information
2. Test of the resource against the Maturity Indicators (F1, F2..., A1, A2..., I1, I2..., R1, R2,... = “FAIRMetrics”)
3. Results and explanation in case of failure : provide suggestions on how to validate this Maturity Indicator.

Monitoring progress in FAIRification through self-assessment of resources maturity indicators

This demo is based on the FAIRMetrics framework [Wilkinson, Dumontier et al., Scientific Data 6:174] [GitHub] that is composed of Maturity Indicators (MI), compliance tests and the evaluator application itself. For now, few efforts have been done so far to take advantage from their concrete implementation, in the process of improving FAIRness of users/community resources. Furthermore, existing interfaces do not provide concrete help or guidelines to developers for better sharing their published works. In this work we propose a web demonstrator, leveraging existing web APIs, aimed at i) evaluating FAIR maturity indicators and ii) providing hints to progress in the FAIRification process.

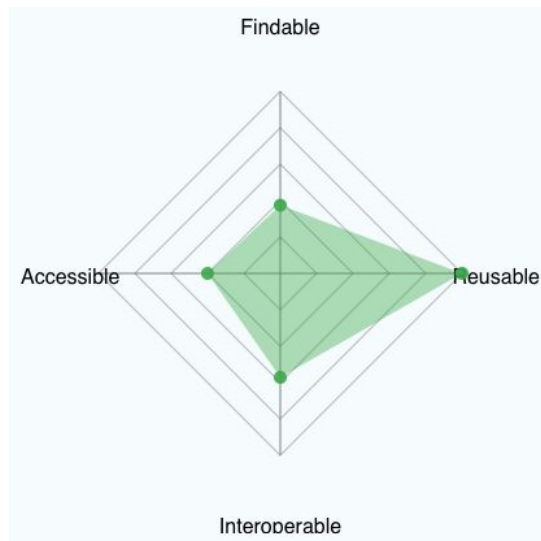
Enter or select a sample resource identifier

input_data <input type="text"/>	input_software <input type="text"/>	input_database <input type="text"/>
input_ontology <input type="text"/>	input_publication <input type="text"/>	input_training <input type="text"/>
input_elixir-fr_SDP <input type="text" value="http://lifemap.univ-lyon1.fr"/>	<input type="text" value="http://lifemap.univ-lyon1.fr"/>	

Start basic or advanced test (select metrics using checkboxes below)

Select the metrics you want to test

<input type="checkbox"/> F - All	<input type="checkbox"/> A - All	<input type="checkbox"/> I - All	<input type="checkbox"/> R - All
<input checked="" type="checkbox"/> F1 - Unique Identifier	<input type="checkbox"/> A1.1 - Uses open free protocol for data retrieval	<input checked="" type="checkbox"/> I1 - Metadata Knowledge Representation Language (weak)	<input type="checkbox"/> R1.1 - Metadata Includes License (strong)
<input type="checkbox"/> F1 - Identifier Persistence	<input checked="" type="checkbox"/> A1.1 - Uses open free protocol for metadata retrieval	<input checked="" type="checkbox"/> I1 - Metadata Knowledge Representation Language (strong)	<input type="checkbox"/> R1.1 - Metadata Includes License (weak)
<input type="checkbox"/> F1 - Data Identifier Persistence	<input type="checkbox"/> A1.2 - Data authentication and authorization	<input type="checkbox"/> I1 - Data Knowledge Representation Language (weak)	
<input checked="" type="checkbox"/> F2 - Structured Metadata	<input checked="" type="checkbox"/> A1.2 - Metadata authentication and authorization	<input type="checkbox"/> I1 - Data Knowledge Representation Language (strong)	
<input checked="" type="checkbox"/> F2 - Grounded Metadata			
<input type="checkbox"/> F3 - Data Identifier Explicitly in			



F1 - Unique Identifier

Status

Description

Metric to test if the metadata resource has a unique identifier. This is done by comparing the GUID to the patterns (by regexp) of known GUID schemas such as URLs and DOIs. Known schema are registered in FAIRSharing (https://fairsharing.org/standards?q=&selected_facets=type_exact:identifier%20schema)

<https://w3id.org/fair/principles/terms/F1>

Result

Score	1
Time	0:00:04
Comment	SUCCESS: Found an identifier of type 'uri'

Check

F1 - Identifier Persistence

Status

Description

Metric to test if the unique identifier of the metadata resource is likely to be persistent. Known schema are registered in FAIRSharing (https://fairsharing.org/standards?q=&selected_facets=type_exact:identifier%20schema). For URLs that don't follow a schema in FAIRSharing we test known URL persistence schemas (purl, oclc, fdlp, purlz, w3id, ark).

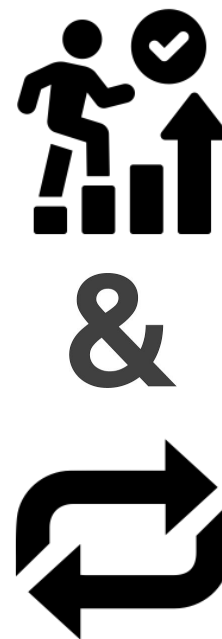
<https://w3id.org/fair/principles/terms/F1>

Result

Score	0
Time	0:00:02
Comment	FAILURE: The metadata GUID does not conform with any known permanent-URL system.
Recommendation	Ensure that meta-data describing your resource use permanent and well formed identifiers (PURLs, DOIs, etc.)

Check

4. Improve the FAIRness of his ressource and come back to step 1.



- Keeping autonomy while providing open access and evaluation transparency / explainability
- Python framework to **draft / extend / implement / propose new** metrics:
<https://github.com/IFB-ElixirFr/fair-checker>
 - E.g. targeting bioinformatics softwares (tools, workflows) ?
 - E.g. targeting specific community needs ?
(challenging due to domain-agnosticity of FAIR metrics)
- Focus on technical “help” for developers and resource providers with external references (e.g. FAIR Cookbook, etc)
- Ongoing work:
 - Opportunity to **rank** BioSchemas annotated resources based on information profiles (minimal / recommended / optional)
 - On **augmenting** metadata base on online Knowledge Graphs (e.g. WikiData, BioPortal, LOV, OLS, DataCite, OpenAire)
 - Aggregate **usage** statistics to monitor progress on FAIR implementation



First feedbacks from some **SDP Elixir-fr** resources:

Ergonomie	Features	Metrics
<ul style="list-style-type: none">• Button for reduce/expand some elements• Re-organize placement of some elements	<ul style="list-style-type: none">• Add CSV, Radarchart SVG, et table export	<ul style="list-style-type: none">• FAIRMetrics caching causing problems on the dynamic side

Perspectives:

- Actually **use** FAIR-Checker as a **FAIRification** assist tool
- New **metrics** with **knowledge graph** focus (ontologies, vocabularies, etc...)
- Provide **templates/tools** to FAIRly **annotate a web page** with metadata (code snippets)
- Usage statistics

Acknowledgments

Thanks to:

- First beta-tester from Elixir-FR SDP
- IFB and Interoperability Action
- Users and future users



We need you to test, contribute on
feedbacks and github !



URL of the demo:

- <https://fair-checker.france-bioinformatique.fr>

URL of the github:

- <https://github.com/IFB-ElixirFr/fair-checker>

URL for feedback-issues:

- <https://github.com/IFB-ElixirFr/fair-checker/issues>



Backup slides

Premiers retours de certaines ressources du **SDP Elixir-fr**

- Besoin d'export CSV, et SVG pour le RadarChart et le tableau
- Intuitif
- Ajouter bouton réduire/agrandir pour certains éléments
- Besoin de réorganiser l'emplacement de certains éléments
- Mise en cache côté FAIRMetrics (rend l'utilisation de FAIRChecker moins dynamique/itératif)

Développements à venir:

- Amélioration de l'ergonomie de l'interface
- Amélioration des recommandations avec exemples de code
- Ajout de nouvelles features (export de résultats)
- Implémentation de métriques supplémentaires/étendre celles existantes
- Augmentation du graphe de connaissance avec requêtes SPARQL
- Statistiques d'utilisation de FAIR-Checker
- Travail avec FAIRMetrics pour la mise en cache

- Propose a (simple) web interface for developers to interact with “FAIR Evaluation Services” APIs and **make progress on FAIRification** (iterative testing) through **recommendations** and **technical implementation** examples (e.g. FAIR Cookbook, etc)
- Evaluate **semantic web technologies** (RDF, SPARQL, **SHACL**) for Findability, Interoperability and some of Reusability principles (e.g. provenance)
- Implement, extend metrics, and **contribute** to FAIR evaluation initiatives



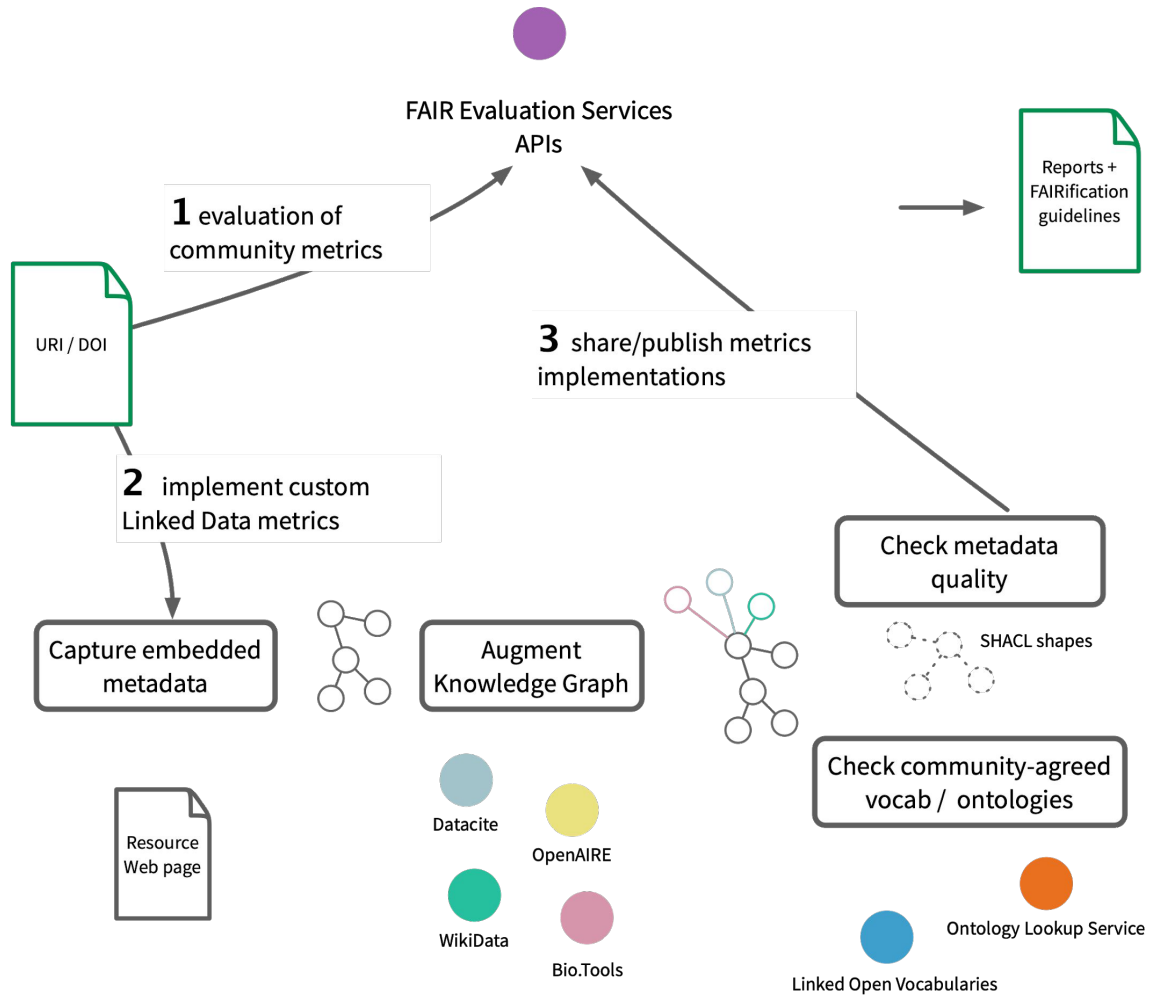
URIs to identify things on the web

RDF to link things on the web
→ graphs

OWL / RDFS to represent knowledge
→ ontologies / vocabularies

SPARQL to query knowledge graphs

SHACL to create graphs shapesconstraint



Custom LD metrics (1): checking FAIR vocabularies

[I2: \(Meta\)data use vocabularies that follow the FAIR principles](#)
[F4: \(Meta\)data are registered or indexed in a searchable resource](#)

Informing on semantic **concepts** and **relations** in metadata.

Querying online registries to check that ontologies / vocabularies are “recognized”:

- **LOV, Linked Open Vocabularies** (general purpose, SPARQL endpoint)
- **OLS, Ontology Lookup Service** (life sciences, REST API)
- *BioPortal (life sciences) SPARQL + REST, access controlled (API-key)*

```
ns11:rsat_peak-motifs a ns2:SoftwareApplication ;
ns2:additionalType "Web service" ;
ns2:citation "pubmed:21715389" ;
ns2:description "A workflow combining a series of time- and memory-efficient motif analysis tools to extract motifs from full-size collections of peaks as generated by ChIP-seq, ChIP-chip or other ChIP-X technologies." ;
ns2:name "RSAT peak-motifs" ;
ns2:operatingSystem "Linux"
```

Classes

<http://schema.org/SoftwareApplication>

LOV

Properties

<http://ogp.me/ns#description>

<http://ogp.me/ns#image>

LOV

<http://ogp.me/ns#title>

<http://schema.org/additionalType>

<http://schema.org/citation>

<http://schema.org/description>

LOV

<http://schema.org/name>

LOV

<http://schema.org/operatingSystem>

<http://schema.org/url>

LOV

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

LOV

<https://bio.tools/ontology/primaryContact>

Custom LD metrics (2): checking metadata quality

[R1.3: \(Meta\)data meet domain-relevant community standards](#)

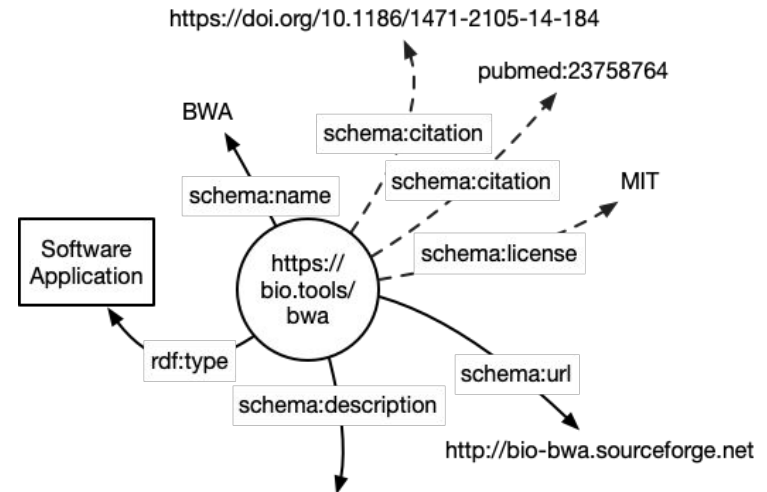
SHACL : RDF Shape Constraint Language

- Declare constraints/rules for knowledge graphs
- Matched against **domain-specific** node or edge types (tools, workflows, datasets, articles, etc.)

Template engine to generate SHACL shapes

Results of the shape validation constitute a knowledge graph:

- Queries to filter severity (warnings or errors)
- Queries to produce user-oriented messages and provide **recommendation**



Fast, accurate, memory-efficient aligner for short and long sequencing reads

—> : mandatory
- - -> : recommended

- **Pourquoi c'est important !** (cf plan sciences ouvertes CNRS ? autre)
- **Challenge de l'interopérabilité** : action transverse / disséminée dans les plateformes / dans les communautés
- Bilan actions 2019 - 2020

Développement d'outils + Animation communauté FR + Animation Elixir

@Cyril @David S. @David L. @Eric P.

- Montrer que l'action est vivante (combien de réunion, combien de personnes, représentativité des communautés FR)
- **Participation active au catalogue** (décembre-février + septembre = 4 mois) : import de données suivant le schéma le nouveau schéma de la BDD (outils, plateformes, compétences, training, etc.). Déploiement du catalogue Docker + VM cloud IFB.
- **[X] FAIR Checker** short comm au All Hands, mise en ligne d'une version demo, travail sur les recommandations
- Enregistré sur FAIRAssist.org
- **[X] Participation au BioHackathon** : EDAM / profils d'information pour Bio.Tools / Orphanet /Orphadata (bioschemas / schema.org (findability)) + Interop entre TeSS et Bio.Tools via Bioschemas.
- **[X] Bio.Tools** : export contenu en Bioschemas sur GitHub (Elixir Tools Ecosystem) hackathon plateforme tool + Annotathon bio.tools
- [] RIR : Waves / AgroLD / AgroPortal
- [] F2F Elixir InteropPlatform orienté COVID-19
- Covid 19 biohackathon → "workflow hub" provenance / RO-crate ([VIPER](#) workflow)

○ AG IFB: 06/11/2020, prêt mais à lancer *oui mais l'IFB pour tous*

Pourquoi FAIR-Checker ?

<https://github.com/IFB-ElisirFr/fair-checker> ; <https://fair-checker.france-bioinformatique.fr>

1. Submission of a resource (URL / DOI) with minimal information
2. Test of the resource against the Maturity Indicators (F1, F2..., A1, A2..., I1, I2..., R1, R2,... = “FAIRMetrics”)
3. Results and explanation in case of failure : provide suggestions on how to validate this Maturity Indicator.

Monitoring progress in FAIRification through self-assessment of resources maturity indicators

This demo is based on the FAIRMetrics framework [Wilkinson, Dumontier et al., Scientific Data 6:174] [GitHub] that is composed of Maturity Indicators (MI), compliance tests and the evaluator application itself. For now, few efforts have been done so far to take advantage from their concrete implementation, in the process of improving FAIRness of users/community resources. Furthermore, existing interfaces do not provide concrete help or guidelines to developers for better sharing their published works. In this work we propose a web demonstrator, leveraging existing web APIs, aimed at i) evaluating FAIR maturity indicators and ii) providing hints to progress in the FAIRification process.

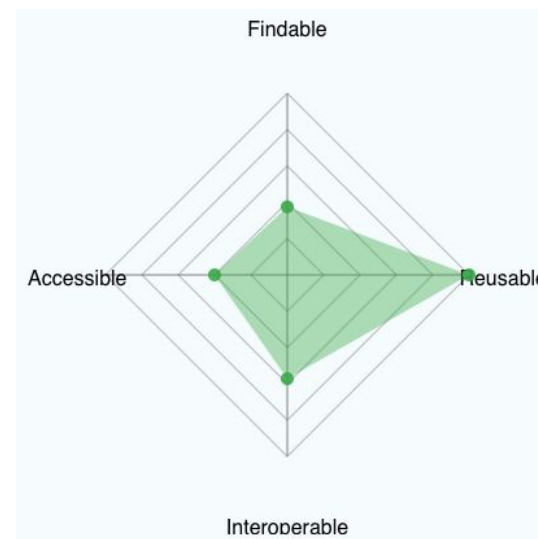
Enter or select a sample resource identifier

input_data	input_software	input_database
<input type="text"/>	<input type="text"/>	<input type="text"/>
input_ontology	input_publication	input_training
<input type="text"/>	<input type="text"/>	<input type="text"/>
input_elixir-fr_SDP	<input type="text"/>	
<input type="text"/>	<input type="text"/>	

Start basic or advanced test (select metrics using checkboxes below)

Select the metrics you want to test

<input type="checkbox"/> F - All	<input type="checkbox"/> A - All	<input type="checkbox"/> I - All	<input type="checkbox"/> R - All
<input checked="" type="checkbox"/> F1 - Unique Identifier	<input type="checkbox"/> A1.1 - Uses open free protocol for data retrieval	<input checked="" type="checkbox"/> I1 - Metadata Knowledge Representation Language (weak)	<input type="checkbox"/> R1.1 - Metadata Includes License (strong)
<input type="checkbox"/> F1 - Identifier Persistence	<input checked="" type="checkbox"/> A1.1 - Uses open free protocol for metadata retrieval	<input checked="" type="checkbox"/> I1 - Metadata Knowledge Representation Language (strong)	<input type="checkbox"/> R1.1 - Metadata Includes License (weak)
<input type="checkbox"/> F1 - Data Identifier Persistence	<input type="checkbox"/> A1.2 - Data authentication and authorization	<input type="checkbox"/> I1 - Data Knowledge Representation Language (weak)	
<input checked="" type="checkbox"/> F2 - Structured Metadata	<input checked="" type="checkbox"/> A1.2 - Metadata authentication and authorization	<input type="checkbox"/> I1 - Data Knowledge Representation Language (strong)	
<input checked="" type="checkbox"/> F2 - Grounded Metadata			
<input type="checkbox"/> F3 - Data Identifier Explicitly in			



4. Improve the FAIRness of his ressource and come back to step 1.

Premiers contacts avec ressources du SDP

Hervé / instable / outil disponible pour la communauté (au delà IFB)

TODO

Ergonomie interface utilisateur (Nicolas / Suzanne)

Statistiques / évolution de la FAIRification des ressources ..

Nouvelles métriques

Use case Data Brokering

Recommandation techniques

- **Projet 2020 - 2021**
 - FAIR checker : message “**communauté IFB ou autre, on cherche des utilisateurs, contacter Thomas !!**”
 - utiliser FAIR checker comme outil d'assistance à la fair-ification
 - Nouvelles métriques transparentes et efficaces
 - Focus sur les knowledge graphs (ontologies, vocabulaires, etc.)
 - template / outils pour l'annotation de pages (code snippets for developers)
 - Statistiques
 - Soutien aux communautés (plantes / mer / santé (catalogue use case COVID-19? autre ?))
 - Représentant par communauté
 - besoin IS Bioschemas plantes
 - Projets pilotes
 - ? participation aux communautés structurées (IFB)
 - Actions concertées avec la plateforme tools (workflows et provenance ?)
 - Actions en direction des DMPs en lien avec projet PIA3 et ANR Flash OpenScience
 - Animation BioSchemas (Tools/Workflows)
 - Développer l'impact de l'”interop” au sens large pour les **sciences ouvertes**
 - **Levier = projets DMPs en cours (maDMP) / DMP (Elixir Converge)**
 - Questionnaire pratiques FAIR : Action coordonnée OpenScience / F. Delamotte / J.-F. Dufayard / P. Lieby
 - Use case Data brokering
 - Bien se phaser avec Elixir : EIP / SDP / RIR
 -

Programming language:

- Python 3



Web Framework:

- Flask
- Use of socket-io for async handling



Environment/deployment:

- Docker



Web server:

- Gunicorn + Nginx

