Ontological representation of FAIR principles: A blueprint for FAIRer data sources

Anna Bernasconi, Alberto García S.,
Giancarlo Guizzardi, Luiz Olavo Bonino da Silva Santos, Veda C. Storey

35th International Conference on Advanced Information Systems Engineering
Zaragoza, Spain – June 15th, 2023

POLITECNICO MILANO 1863
UNIVERSITAT POLITÈCNICA DE VALÈNCIA
UNIVERSITY OF TWENTE.
LUMC Leiden University Medical Center
Georgia State University

FAIR: where it all started

Workshop in Leiden, Netherlands, 2014:
"Jointly Designing a Data Fairport"

FAIR principles

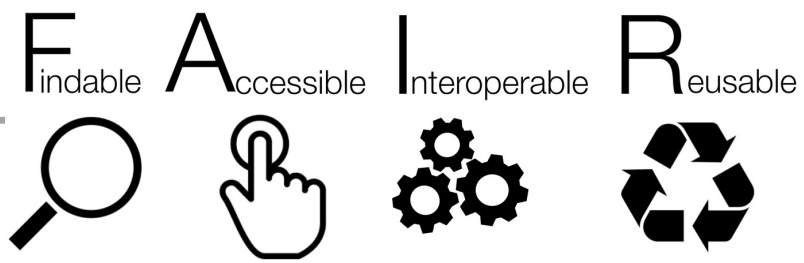
Introduced in 2016 (Wilkinson et al., *Scientific Data*)

as guidelines to improve

the Findability, Accessibility, Interoperability, and Reuse of datasets,

i.e., enabling machines to perform automatic actions on them

FAIR: for whom forgot



The FAIR Guiding Principles

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

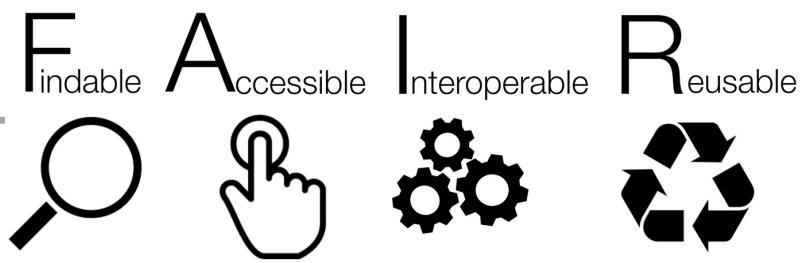
To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even 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

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.

FAIR: for whom forgot



The FAIR Guiding Principles

To be Findable:

F2. data are described with rich metadata (defined by R1 below)

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

I3. (meta)data include qualified references to other (meta)data

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

R1.3. (meta)data meet domain-relevant community standards

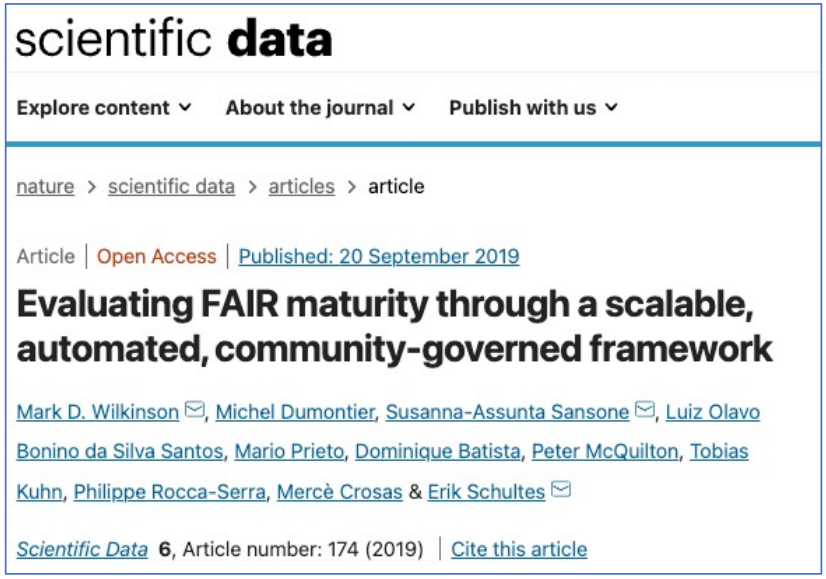
Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.

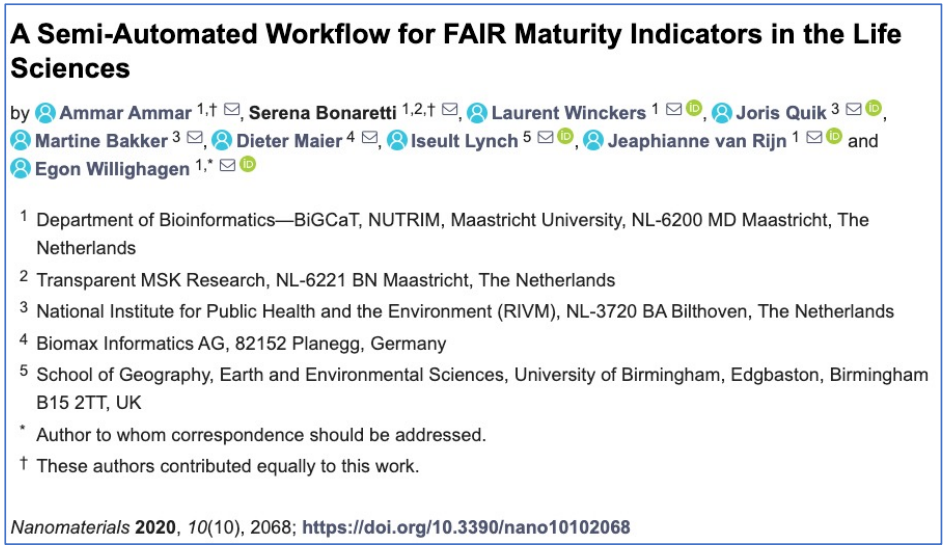
Matters arising

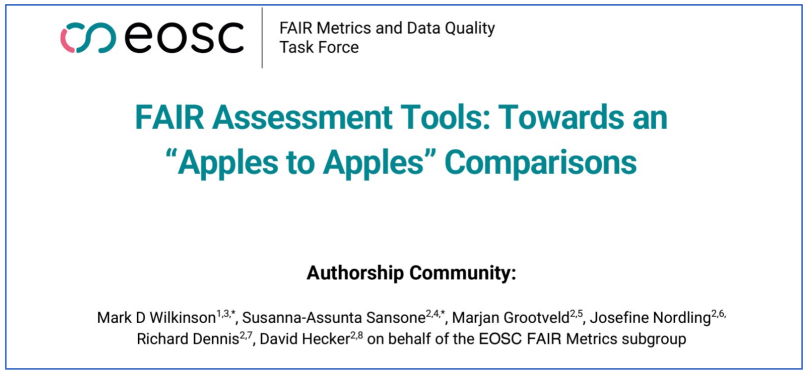
- The original FAIR paper (Wilkinson et al., 2016, *Scientific Data*) did not explain in detail the intentions behind the principles and related consequences

- By design, principles do not provide specific implementation definitions (→ potentially incompatible implementations)

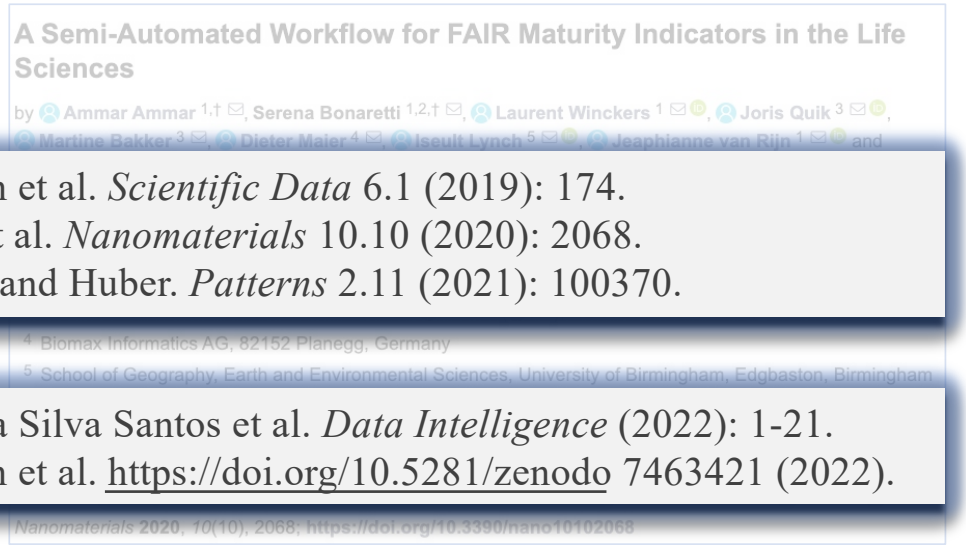
Existing implementation efforts

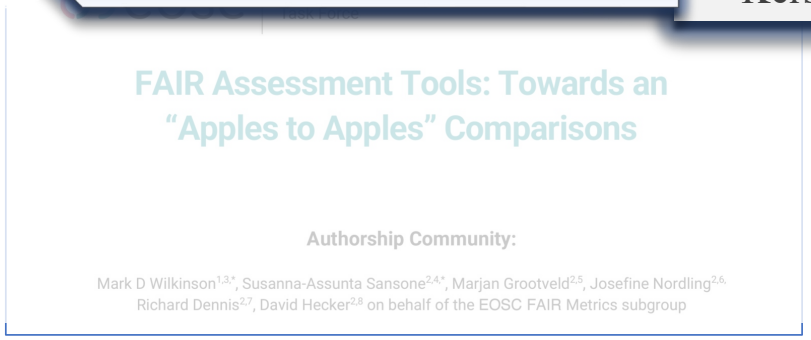






Existing implementation efforts





FAIRness assessment

- Wilkinson et al. *Scientific Data* 6.1 (2019): 174.
- Ammar et al. *Nanomaterials* 10.10 (2020): 2068.
- Devaraju and Huber. *Patterns* 2.11 (2021): 100370.

FAIR tooling

- Bonino da Silva Santos et al. *Data Intelligence* (2022): 1-21.
- Wilkinson et al. https://doi.org/10.5281/zenodo 7463421 (2022).

FAIR service support

- Sansone et al. *Nature Biotechnology* 37.4 (2019): 358-367.
- Kersloot et al. *Scientific Data* 9.1 (2022): 241.

Information integration & Interoperability

FAIR principles can facilitate the development of

1) **integrated data ecosystems**
2) **interoperable information ecosystems**

What is missing?

A consistent interpretation of the FAIR principles

How do I make my data ‘FAIR’?

[1]

What is the **current FAIRness level** of my dataset?

[2]

What is the **intended FAIRness level** that I want it to reach?

[3]

How can I improve **from the current-level to the intended-level FAIRness**?

To answer these questions, one must hold a good understanding of the principles.

What we propose

Use of ontological models

- To provide clear and precise *explanations*
- To enforce shared understanding among stakeholders

The **OntoUML FAIR Principles Schema**
=
a blueprint for designing a dataset FAIRness strategy

Benefits of using ontology-based model to present FAIR principles

1. Explicit **representation** of a shared interpretation of the principles in a **concrete artifact**

2. Controlled **vocabulary** for use in **semantic annotations** of entities

3. **Rationale** for deriving FAIR **evaluation metrics**

4. Prescriptive **guidelines to operationalize** the abstract guiding principles

Method: OntoUML (pt. 1)

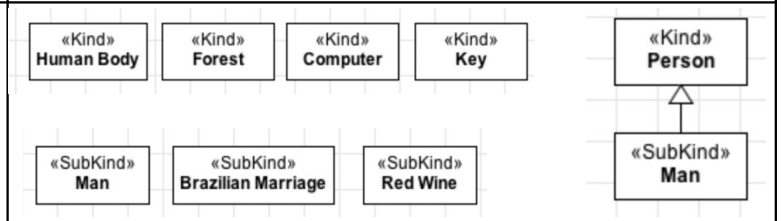
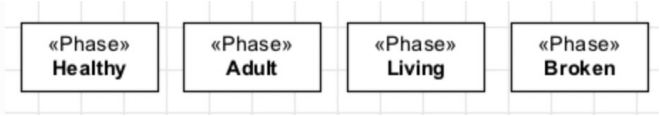
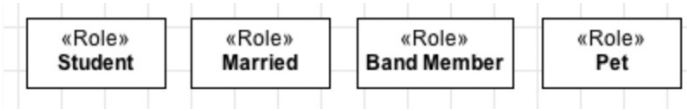
- *Ontologically well-founded language* for Ontology-driven Conceptual Modeling

- Its metamodel reflects the ontological distinctions of *Unified Foundational Ontology* (*UFO*) [Guizzardi, 2005]

- A proxy for *ontological analysis* effective in supporting the explanation of complex domains

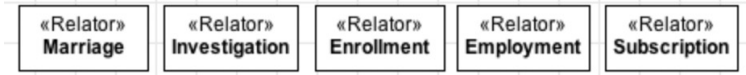
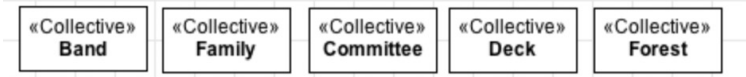
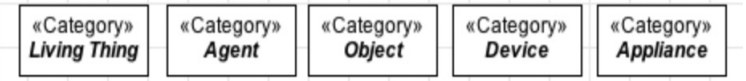
Method: OntoUML (pt. 2)

Relevant OntoUML entity stereotypes used by our method:

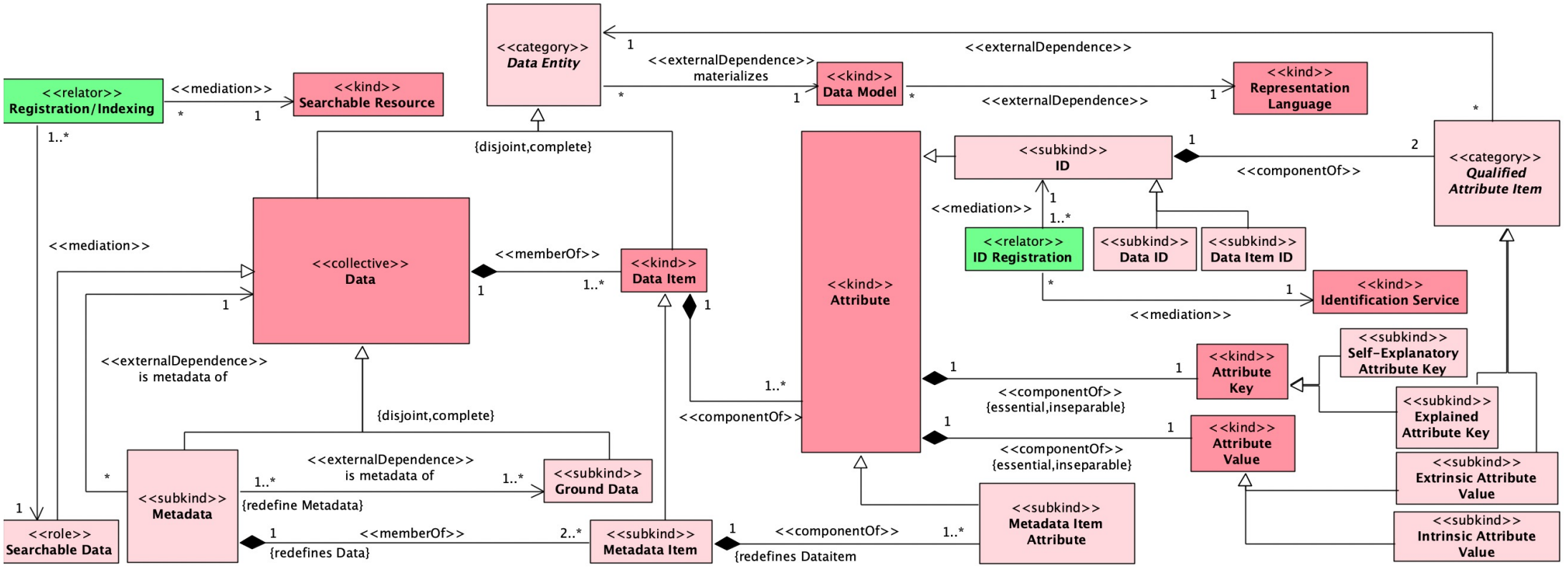
Stereotype	Description	Example
«kind» and «subkind»	1) What the objects in that domain essentially are. 2) Subdivision of a kind.	
«phase»	Type that captures a cluster of change conditions in intrinsic properties	
«role»	Type capturing relational properties shared by instances of a given kind	

Method: OntoUML (pt. 3)

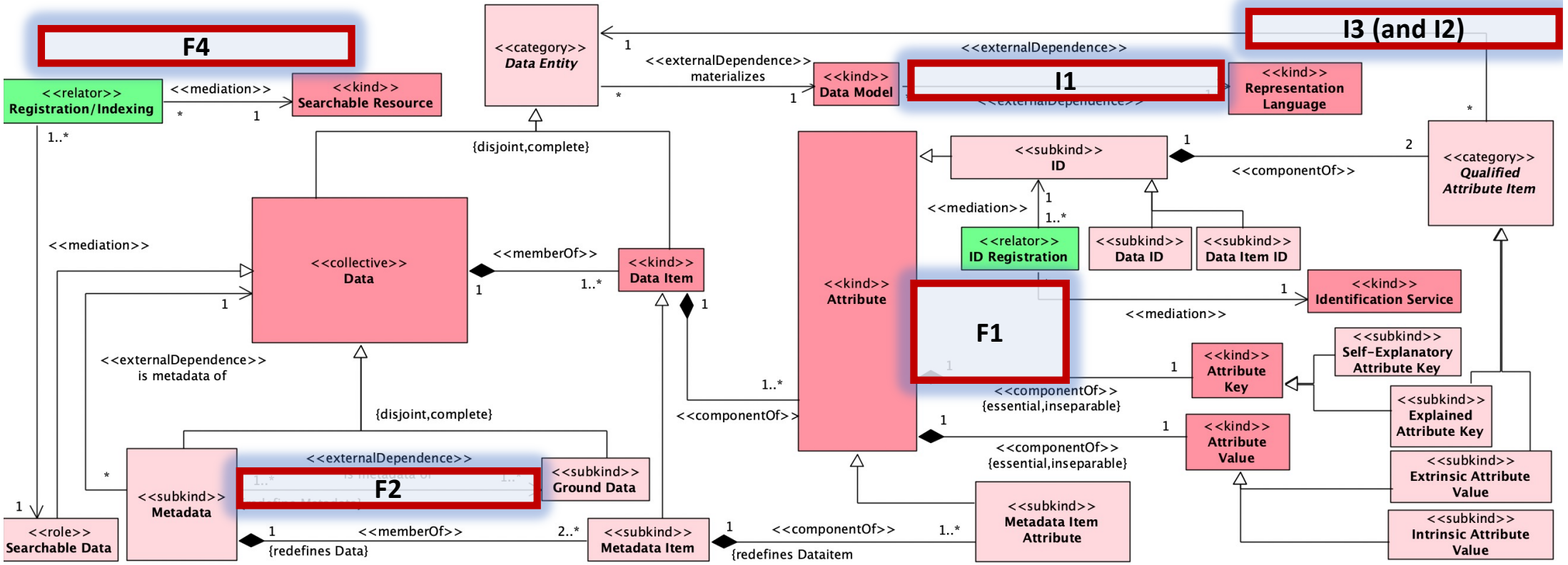
Relevant OntoUML entity stereotypes used by our method:

Stereotype	Description	Example
«relator»	Truth-maker of relational propositions. Relations (as classes of n-tuples) can be completely derived from relators.	
«collective»	Plural entity that aggregates parts (members), all of which play the same role with respect to the whole.	
«category»	Necessary properties that are shared by entities of multiple kinds.	
«mode»	Particular type of intrinsic property that has no structured value.	

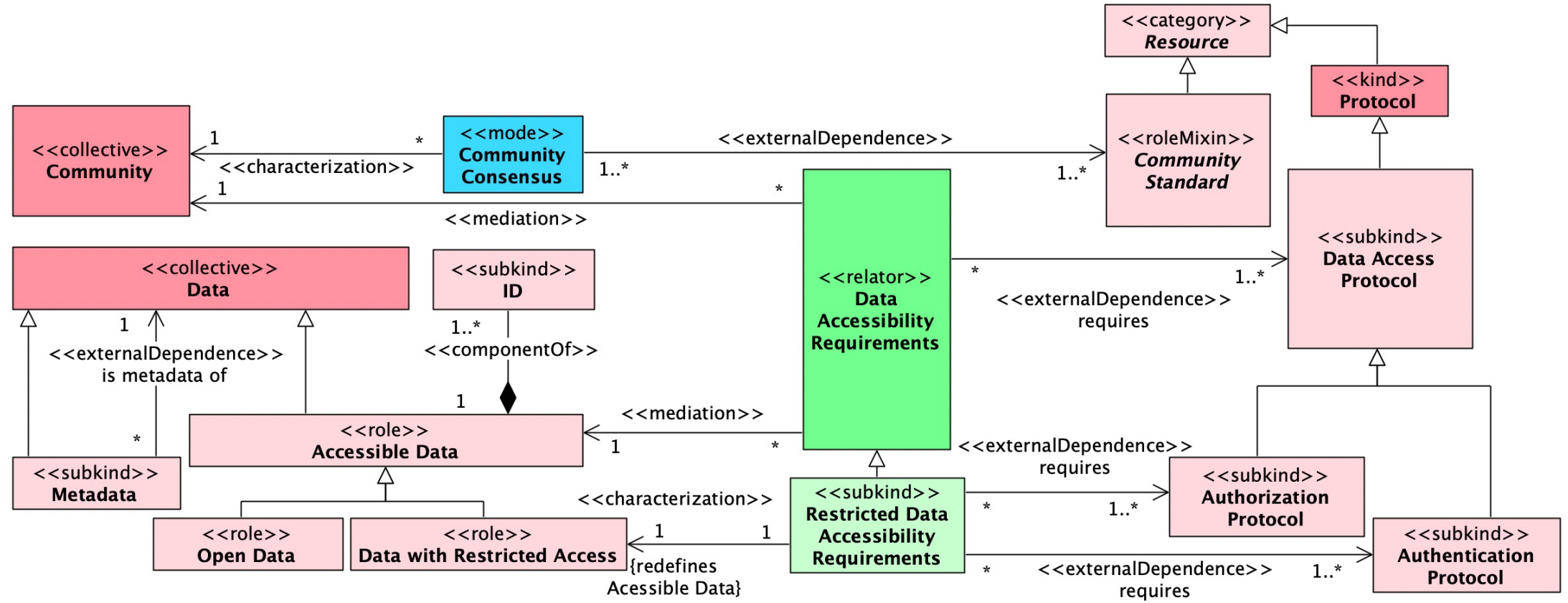
OntoUML FAIR Principles Schema: Findability and Interoperability



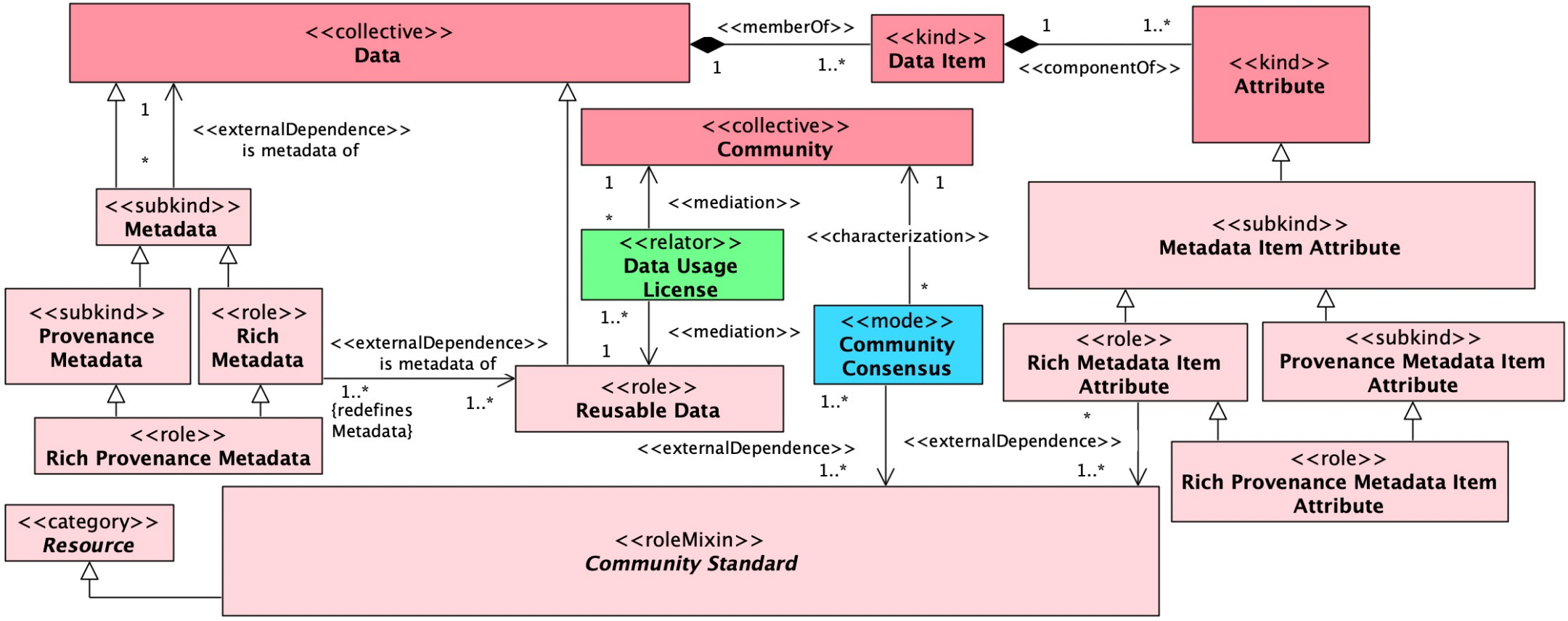
OntoUML FAIR Principles Schema: Findability and Interoperability



OntoUML FAIR Principles Schema: Accessibility



OntoUML FAIR Principles Schema: Reusability



OntoUML FAIR Principles Schema: 4 objectives

1. Explicit representation of a
   shared interpretation of the principles in a concrete artifact

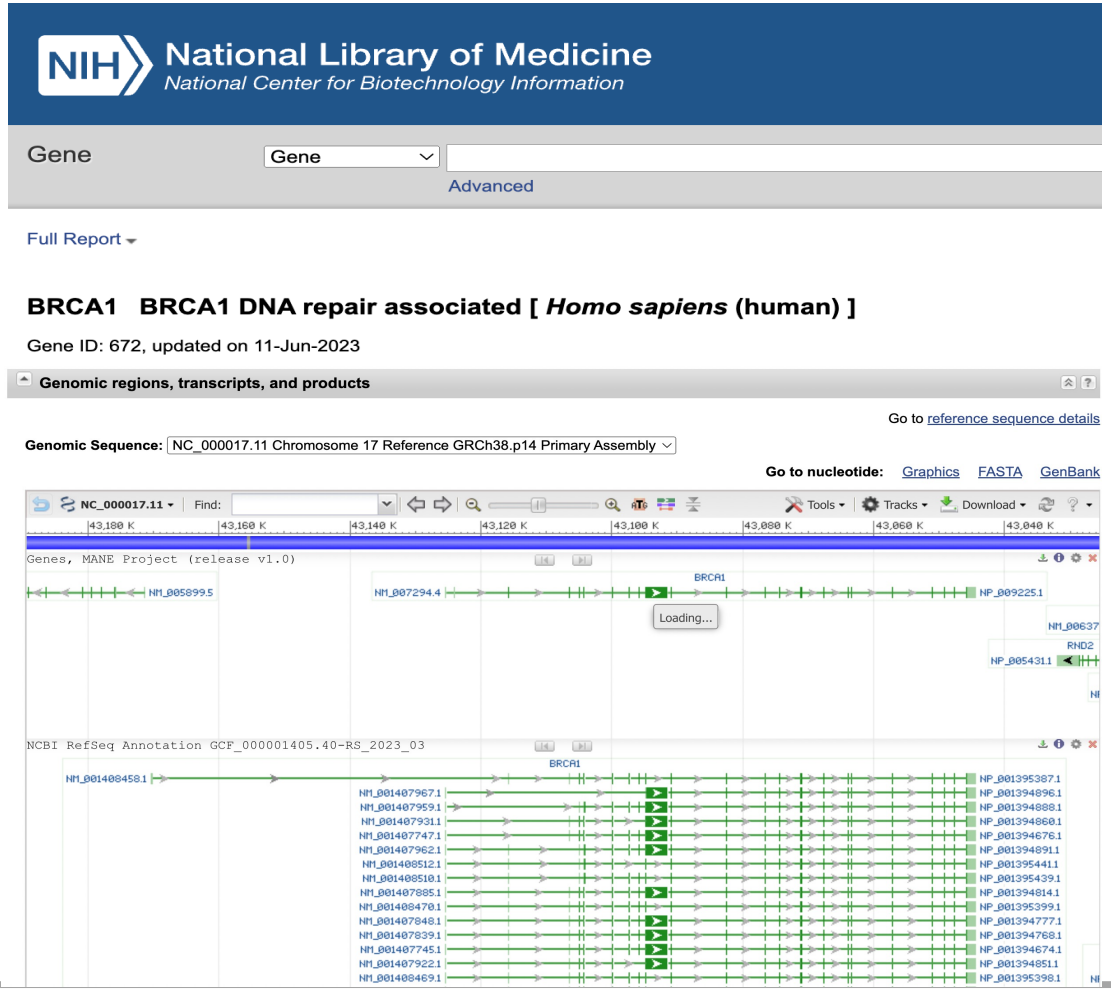
2. Controlled vocabulary for use in semantic annotations of entities

3. Rationale for deriving FAIR evaluation metrics

4. Prescriptive guidelines to operationalize the abstract guiding principles

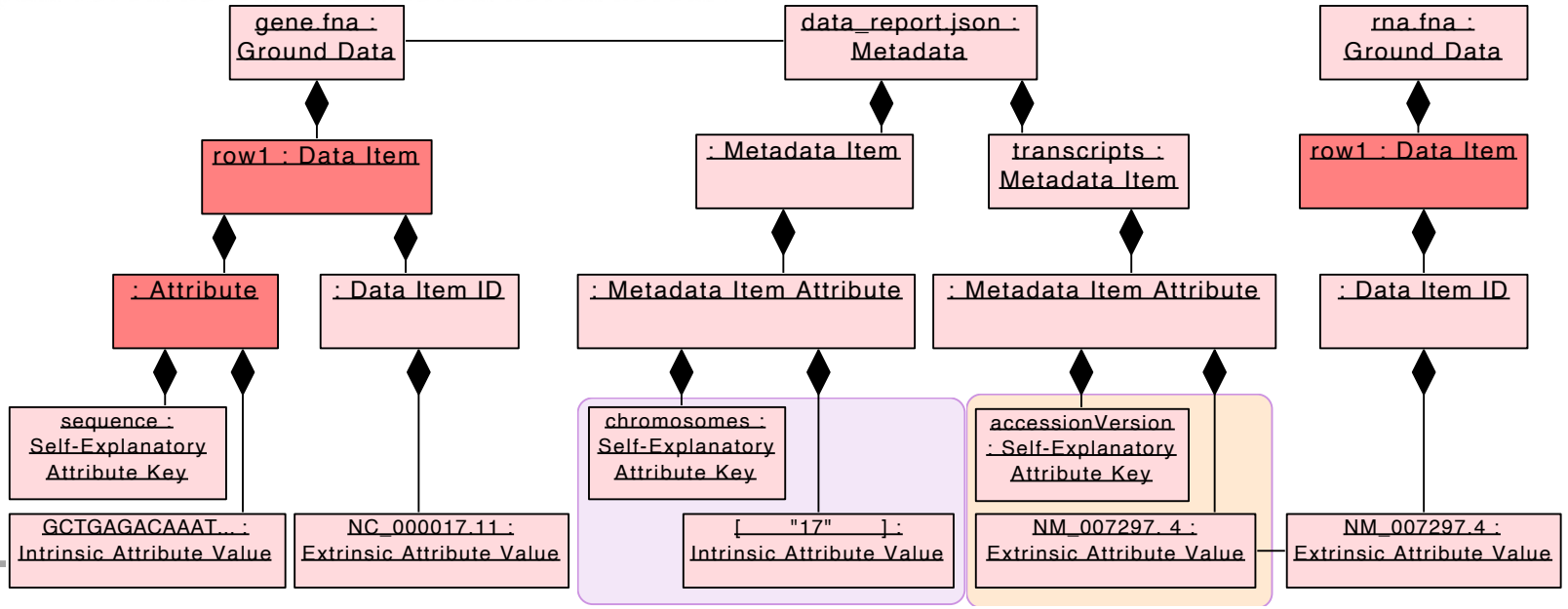
Semantic annotations of (meta)data entities

https://www.ncbi.nlm.nih.gov/gene/672



Semantic annotations of (meta)data entities


```
{
    "commonName": "human", "taxId": "9606", "taxname": "Homo sapiens",
    "chromosomes": ["17"],
    "description": "BRCA1 DNA repair associated", "symbol": "BRCA1",
    "geneId": "672", "ensemblGeneIds": ["ENSG00000012048"], "swissProtAccessions": ["P38398"],
    "nomenclatureAuthority": {"authority": "HGNC", "identifier": "HGNC:1100"},
    "genomicRanges": [{"accessionVersion": "NC_000017.11",
                       "range": [{"begin": "43044295", "end": "43125364", "orientation": "minus"}]},
                      {…}],
    "transcripts": [{…},
                    {
                        "accessionVersion": "NM_007297.4",
                        "name": "transcript variant 3",
                        "length": 7028,
                        "ensemblTranscript": "ENST00000493795.5",
                        "cds": {…}, "exons": {…}, "genomicLocations": […], "genomicRange": {…},
                        },
                    },
                    {…}]
}
```

OntoUML FAIR Principles Schema: 4 objectives

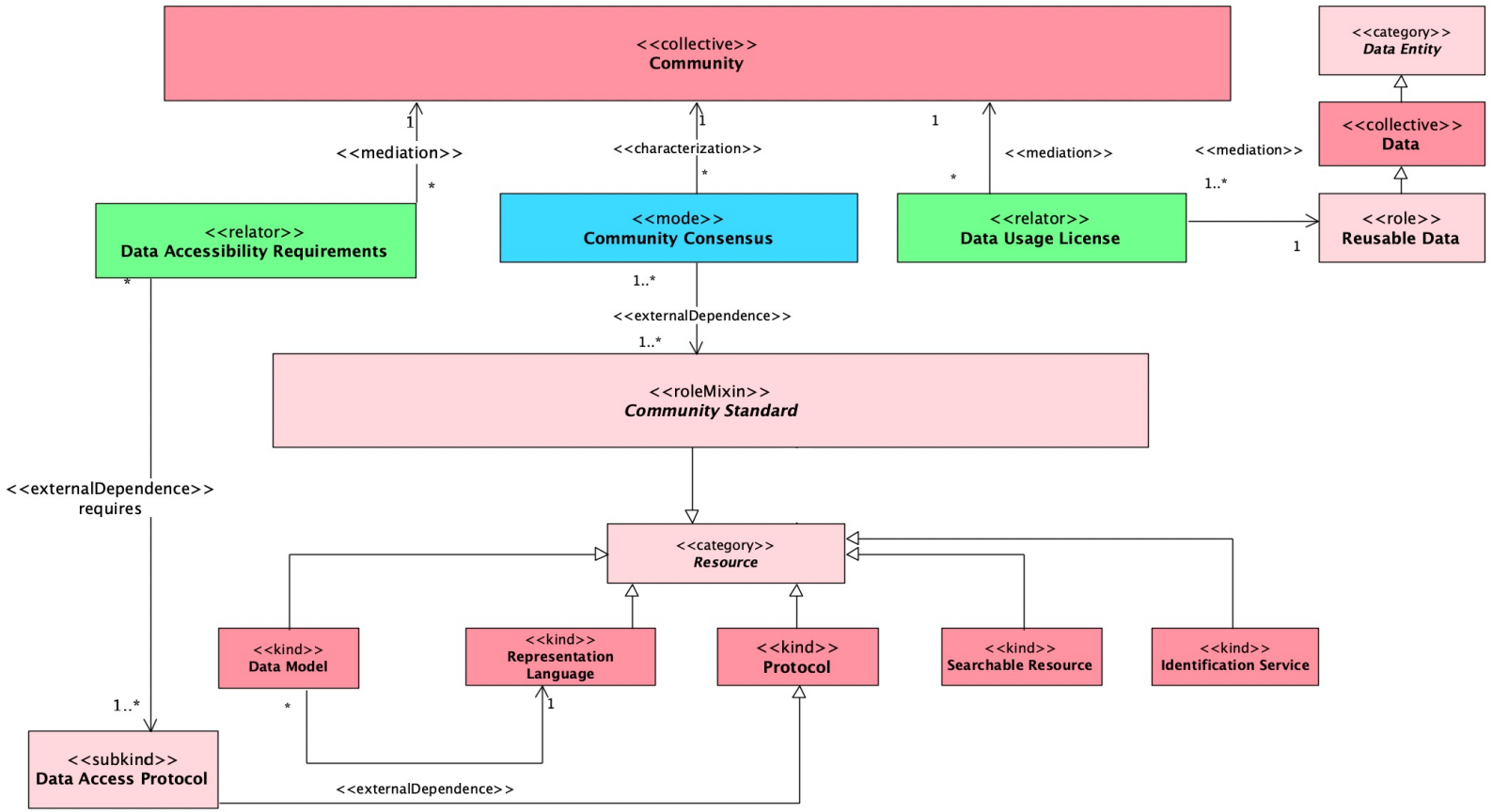
1. Explicit representation of a
   shared interpretation of the principles in a concrete artifact

2. Controlled vocabulary for use in semantic annotations of entities

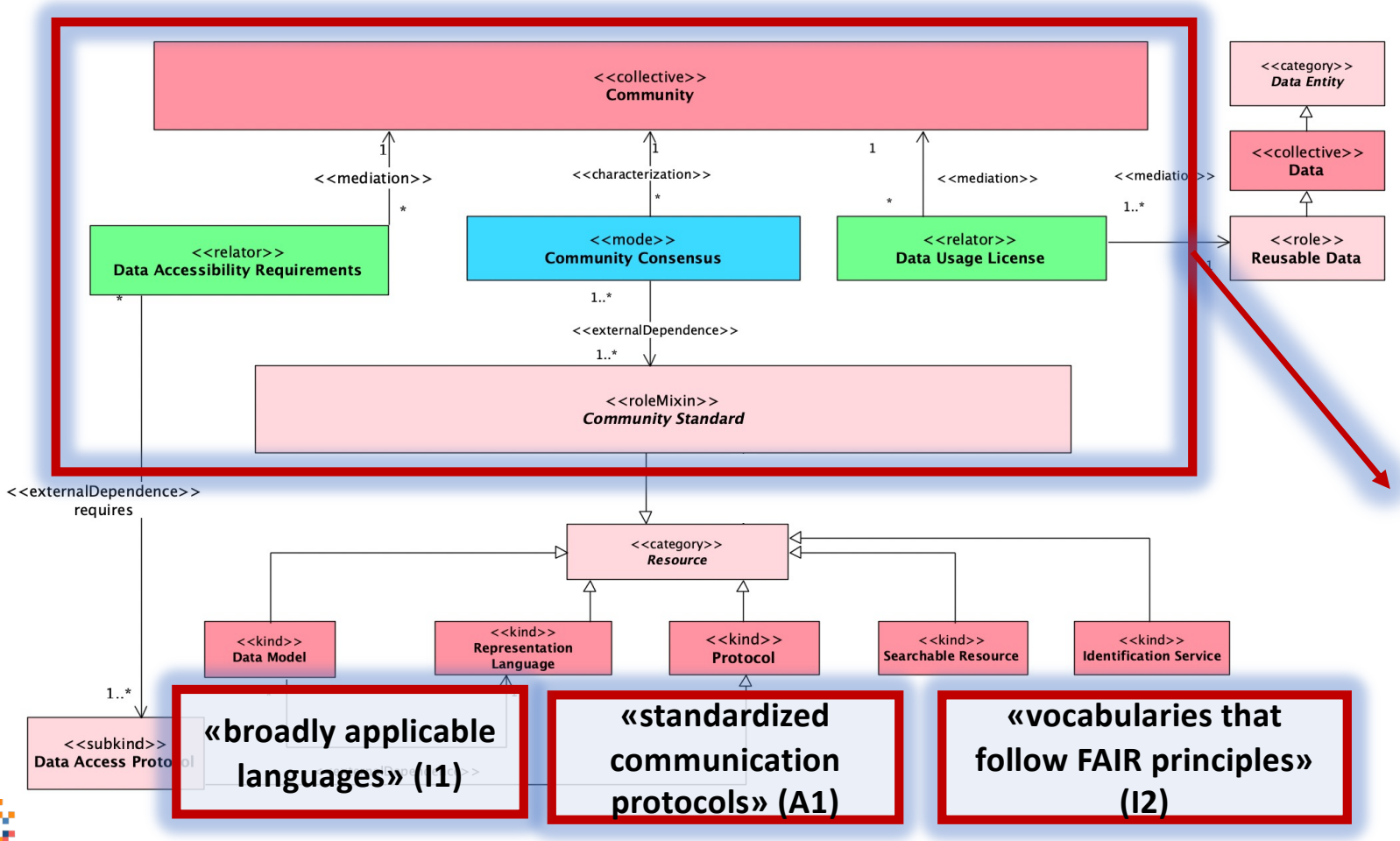
3. Rationale for deriving FAIR evaluation metrics

4. Prescriptive guidelines to operationalize the abstract guiding principles

R1.3. (Meta)data meet domain-relevant community standards



R1.3. (Meta)data meet domain-relevant community standards



***Our analysis makes the centrality of communities and their collective beliefs clear**, (even if "community" is used only once in the principles, **R1.3**)*

Summary and future directions

OntoUML FAIR Principles Schema extends work on the adoption of FAIR principles

The following aspects were found to be critical:

- *Under specified areas.*
- *Asymmetric definitions*
- *Unclear community roles.*
- *How rich is 'rich'?*

Metadata of data… is also data

'F' fundamentally based on technology 'I' is conceptual

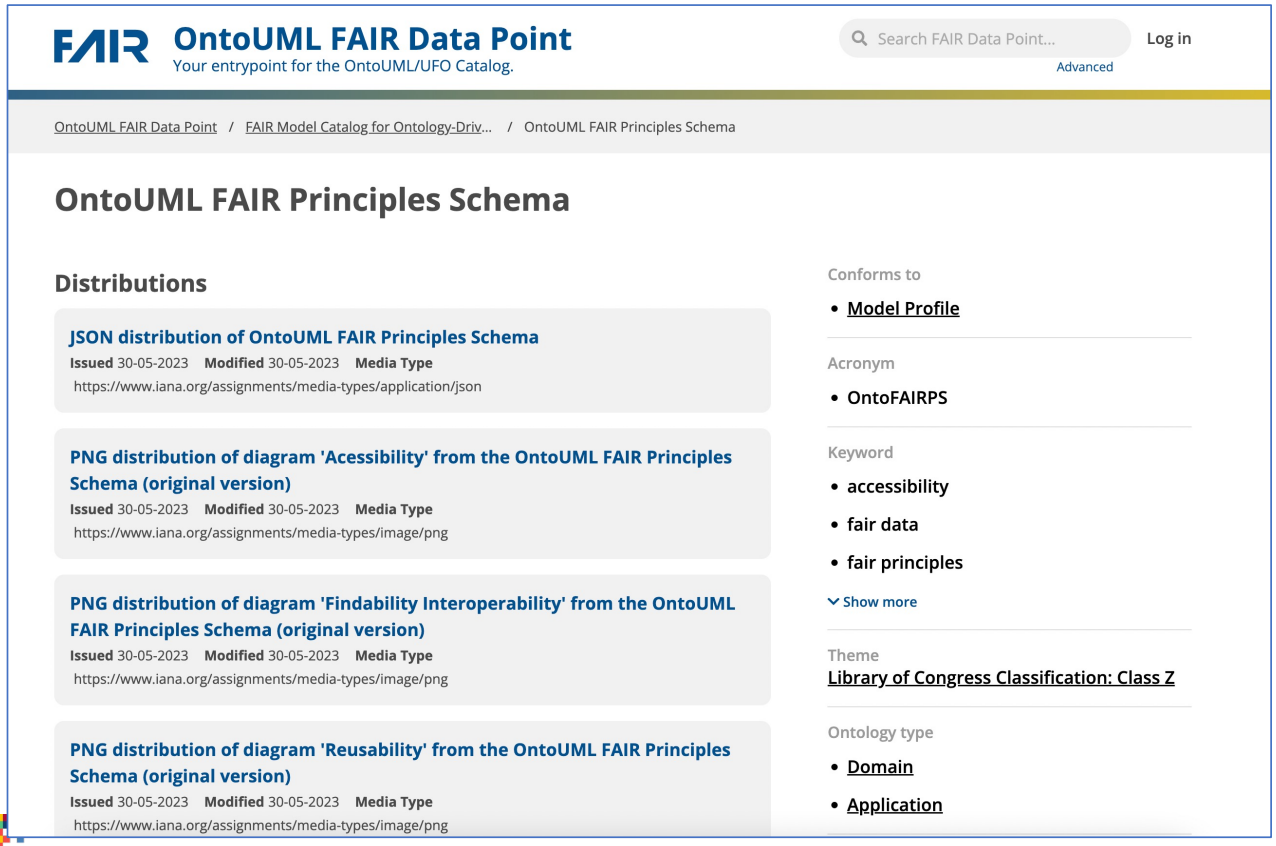
'unFAIR' situation: no community standards or not possible to clearly and unambiguously identify them

What is next

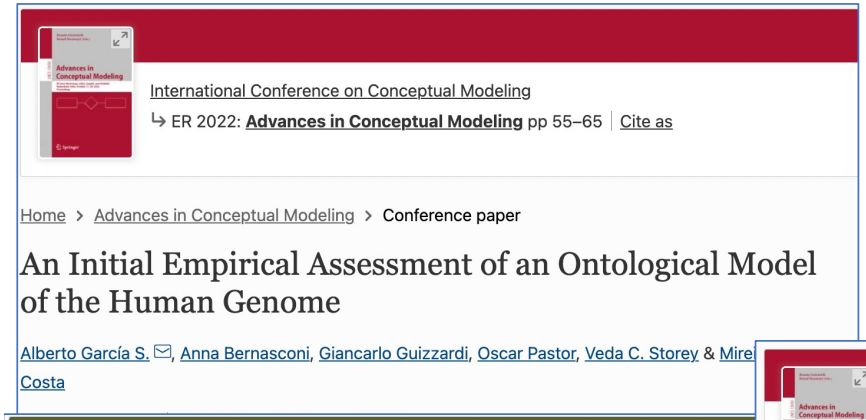
Refining the schema and applying it to *FAIR Digital Objects* (not only datasets!)

Putting the FAIRness forward

https://scs-ontouml.eemcs.utwente.nl/model/648b1a98-41f6-49b9-be93-6b012d52593c/



Our work on *ontological unpacking* for explanation of complex domains

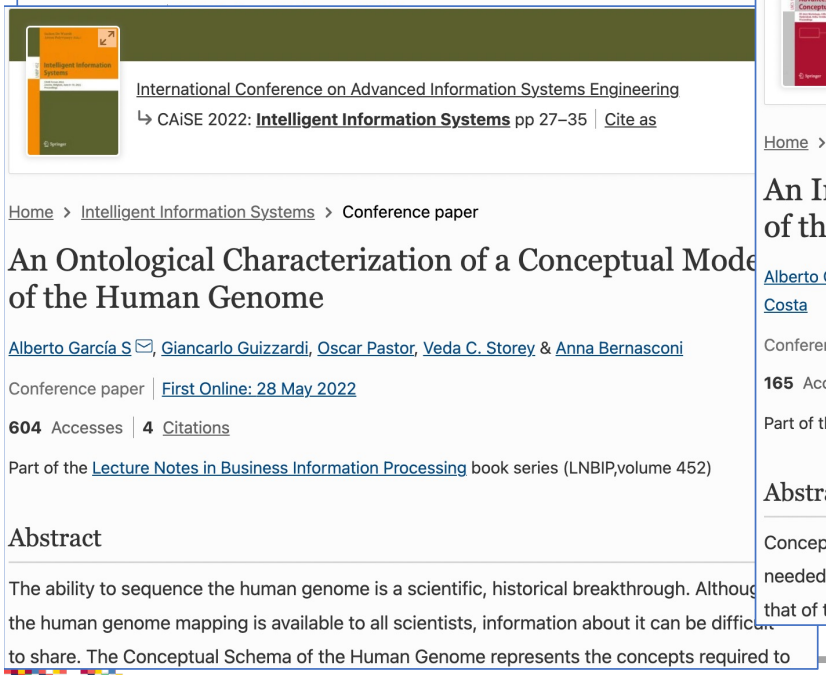


International Conference on Conceptual Modeling
↳ ER 2022: **Advances in Conceptual Modeling** pp 55–65 | Cite as

Home > Advances in Conceptual Modeling > Conference paper

An Initial Empirical Assessment of an Ontological Model of the Human Genome

Alberto García S., Anna Bernasconi, Giancarlo Guizzardi, Oscar Pastor, Veda C. Storey & Mirei Costa



International Conference on Advanced Information Systems Engineering
↳ CAiSE 2022: **Intelligent Information Systems** pp 27–35 | Cite as

Home > Intelligent Information Systems > Conference paper

An Ontological Characterization of a Conceptual Mode of the Human Genome

Alberto García S, Giancarlo Guizzardi, Oscar Pastor, Veda C. Storey & Anna Bernasconi

Conference paper | First Online: 28 May 2022

604 Accesses | **4** Citations

Part of the Lecture Notes in Business Information Processing book series (LNBIP,volume 452)

Abstract

The ability to sequence the human genome is a scientific, historical breakthrough. Althoug the human genome mapping is available to all scientists, information about it can be difficu to share. The Conceptual Schema of the Human Genome represents the concepts required to

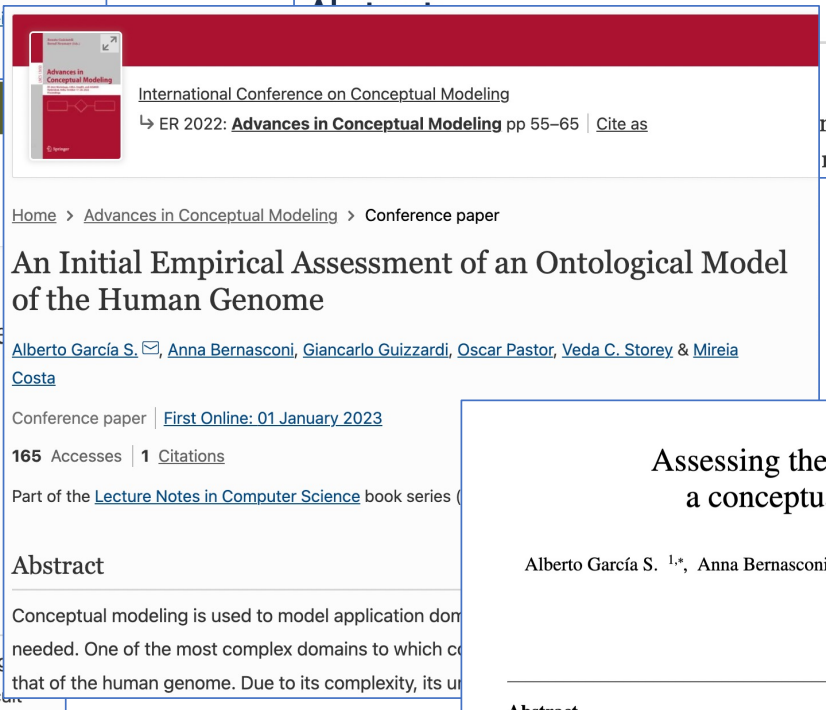
Semantic interoperability: ontological unpacking of a viral conceptual model

Anna Bernasconi, Giancarlo Guizzardi, Oscar Pastor & Veda C. Storey

BMC Bioinformatics **23**, Article number: 491 (2022) | Cite this article

1404 Accesses | **1** Citations | **3** Altmetric | Metrics

rtant, but complex, domains being
need to facilitate and support work within



International Conference on Conceptual Modeling
↳ ER 2022: **Advances in Conceptual Modeling** pp 55–65 | Cite as

Home > Advances in Conceptual Modeling > Conference paper

An Initial Empirical Assessment of an Ontological Model of the Human Genome

Alberto García S., Anna Bernasconi, Giancarlo Guizzardi, Oscar Pastor, Veda C. Storey & Mireia Costa

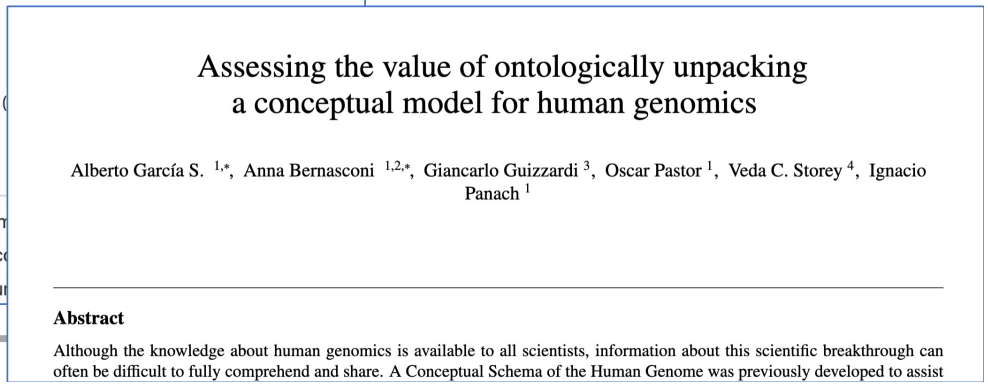
Conference paper | First Online: 01 January 2023

165 Accesses | **1** Citations

Part of the Lecture Notes in Computer Science book series (

Abstract

Conceptual modeling is used to model application dom
needed. One of the most complex domains to which c
that of the human genome. Due to its complexity, its u



Assessing the value of ontologically unpacking a conceptual model for human genomics

Alberto García S. [1,*], Anna Bernasconi [1,2,*], Giancarlo Guizzardi [3], Oscar Pastor [1], Veda C. Storey [4], Ignacio Panach [1]

Abstract

Although the knowledge about human genomics is available to all scientists, information about this scientific breakthrough can often be difficult to fully comprehend and share. A Conceptual Schema of the Human Genome was previously developed to assist

Ontological representation of FAIR principles: A blueprint for FAIRer data sources

Anna Bernasconi*, Alberto García S.*, Giancarlo Guizzardi, Luiz Olavo Bonino da Silva Santos, Veda C. Storey

*co-first authors

Thank you all for your kind attention … for more information, do not hesitate to contact us!

anna.bernasconi@polimi.it - algarsi3@inf.upv.es

ANNA BERNASCONI

ALBERTO GARCÍA S.

GIANCARLO GUIZZARDI

LUIZ OLAVO BONINO DA SILVA SANTOS

VEDA C. STOREY

POLITECNICO MILANO 1863

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITY OF TWENTE.

Leiden University Medical Center

Georgia State University