



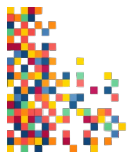
Ontological representation of FAIR principles: A blueprint for FAIRer data sources

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



UNIVERSITY
OF TWENTE.





Open
data
is about
MORE
THAN
DISCLOSURE
it must be
“Fair”

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

FAIR principles

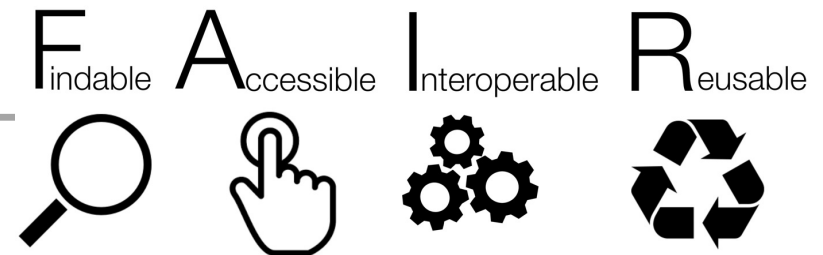
Introduced in 2016 (Wilkinson et al., *Scientific Data*)
as guidelines to improve

the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse
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

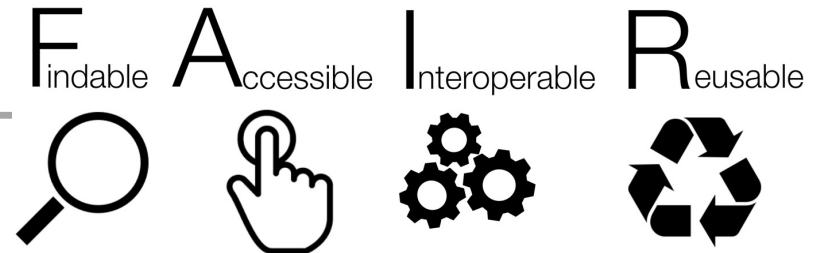
To be Reusable:

- 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)

- 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:

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

- 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

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

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

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. meta(data) have a persistent identifier
 - R1.2. meta(data) are associated with a data or service
 - 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



scientific data

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [scientific data](#) > [articles](#) > [article](#)









Article | [Open Access](#) | [Published: 20 September 2019](#)

Evaluating FAIR maturity through a scalable, automated, community-governed framework

[Mark D. Wilkinson](#) , [Michel Dumontier](#), [Susanna-Assunta Sansone](#) , [Luiz Olavo Bonino da Silva Santos](#), [Mario Prieto](#), [Dominique Batista](#), [Peter McQuilton](#), [Tobias Kuhn](#), [Philippe Rocca-Serra](#), [Mercè Crosas](#) & [Erik Schultes](#) 

[Scientific Data](#) **6**, Article number: 174 (2019) | [Cite this article](#)

A Semi-Automated Workflow for FAIR Maturity Indicators in the Life Sciences

by [Ammar Ammar](#) ^{1,†} , [Serena Bonaretti](#) ^{1,2,†} , [Laurent Winckers](#) ¹ , [Joris Quik](#) ³ , [Martine Bakker](#) ³ , [Dieter Maier](#) ⁴ , [Iseult Lynch](#) ⁵ , [Jeaphianne van Rijn](#) ¹  and [Egon Willighagen](#) ^{1,*} 

¹ Department of Bioinformatics—BiGCaT, NUTRIM, Maastricht University, NL-6200 MD Maastricht, The Netherlands

² Transparent MSK Research, NL-6221 BN Maastricht, The Netherlands

³ National Institute for Public Health and the Environment (RIVM), NL-3720 BA Bilthoven, The Netherlands

⁴ Biomax Informatics AG, 82152 Planegg, Germany

⁵ School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

* Author to whom correspondence should be addressed.

† These authors contributed equally to this work.

Nanomaterials **2020**, *10*(10), 2068; <https://doi.org/10.3390/nano10102068>



FAIR Metrics and Data Quality Task Force

FAIR Assessment Tools: Towards an “Apples to Apples” Comparisons

Authorship Community:

Mark D Wilkinson^{1,3,*}, Susanna-Assunta Sansone^{2,4,*}, Marjan Grootveld^{2,5}, Josefine Nordling^{2,6}, Richard Dennis^{2,7}, David Hecker^{2,8} on behalf of the EOSC FAIR Metrics subgroup



Log in

ARTICLE | VOLUME 2, ISSUE 11, 100370, NOVEMBER 12, 2021 [Download Full Issue](#)

An automated solution for measuring the progress toward FAIR research data

[Anusuriya Devaraju](#) ³  • [Robert Huber](#)  • [Show footnotes](#)

[Open Access](#) • Published: October 29, 2021 • DOI: <https://doi.org/10.1016/j.patter.2021.100370> • [Check for updates](#)



Existing implementation efforts



scientific data

Explore content ▾ About the journal ▾ Publish with us ▾



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.

A Semi-Automated Workflow for FAIR Maturity Indicators in the Life Sciences

by Ammar Ammar ^{1,†} , Serena Bonaretti ^{1,2,†} , Laurent Winckers ¹ , Joris Quik ³ , Marlene Bakker ³ , Dieter Maier ⁴ , Iseult Lynch ⁵ , Jeaphanne van Riin ¹ and and

⁴ Biomax Informatics AG, 82152 Planegg, Germany

⁵ School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham

Nanomaterials 2020, 10(10), 2068; <https://doi.org/10.3390/nano10102068>

FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons

Authorship Community:

Mark D Wilkinson^{1,2}, Susanna-Assunta Sansone^{2,4}, Marjan Grootveld^{2,5}, Josefine Nordling^{2,6}, Richard Dennis^{2,7}, David Hecker^{2,8} on behalf of the EOSC FAIR Metrics subgroup

ARTICLE | VOLUME 2, ISSUE 11, 100370, NOVEMBER 12, 2021 [Download Full Issue](#)

An automated solution for measuring the progress toward FAIR research data

Anusuriya Devaraju ³ • Robert Huber • [Show footnotes](#)

[Open Access](#) • Published: October 29, 2021 • DOI: <https://doi.org/10.1016/j.patter.2021.100370> • [Check for updates](#)





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





- *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:

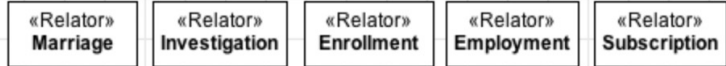

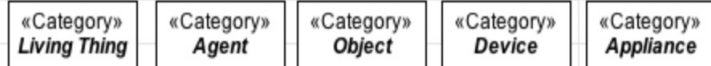

Stereotype	Description	Example
<p>«kind» and «subkind»</p>	<p>1) What the objects in that domain essentially are. 2) Subdivision of a kind.</p>	
<p>«phase»</p>	<p>Type that captures a cluster of change conditions in intrinsic properties</p>	
<p>«role»</p>	<p>Type capturing relational properties shared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



Method: OntoUML (pt. 3)

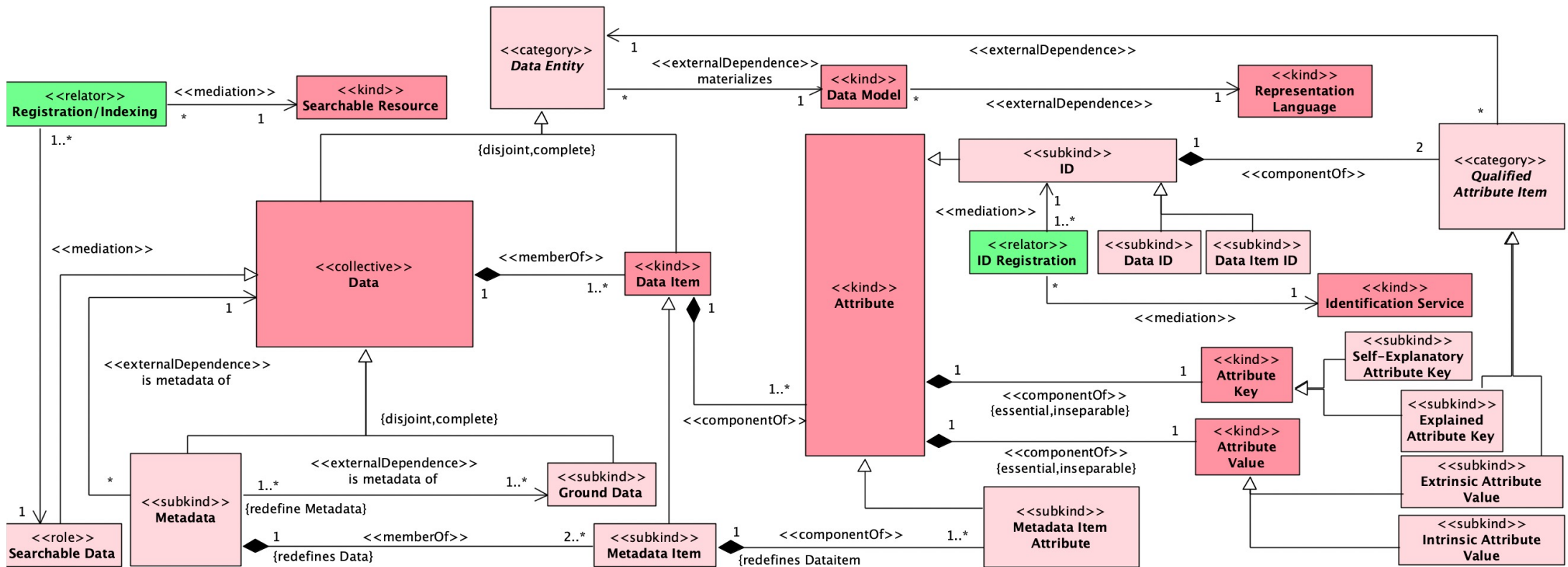


Relevant OntoUML entity stereotypes used by our method:

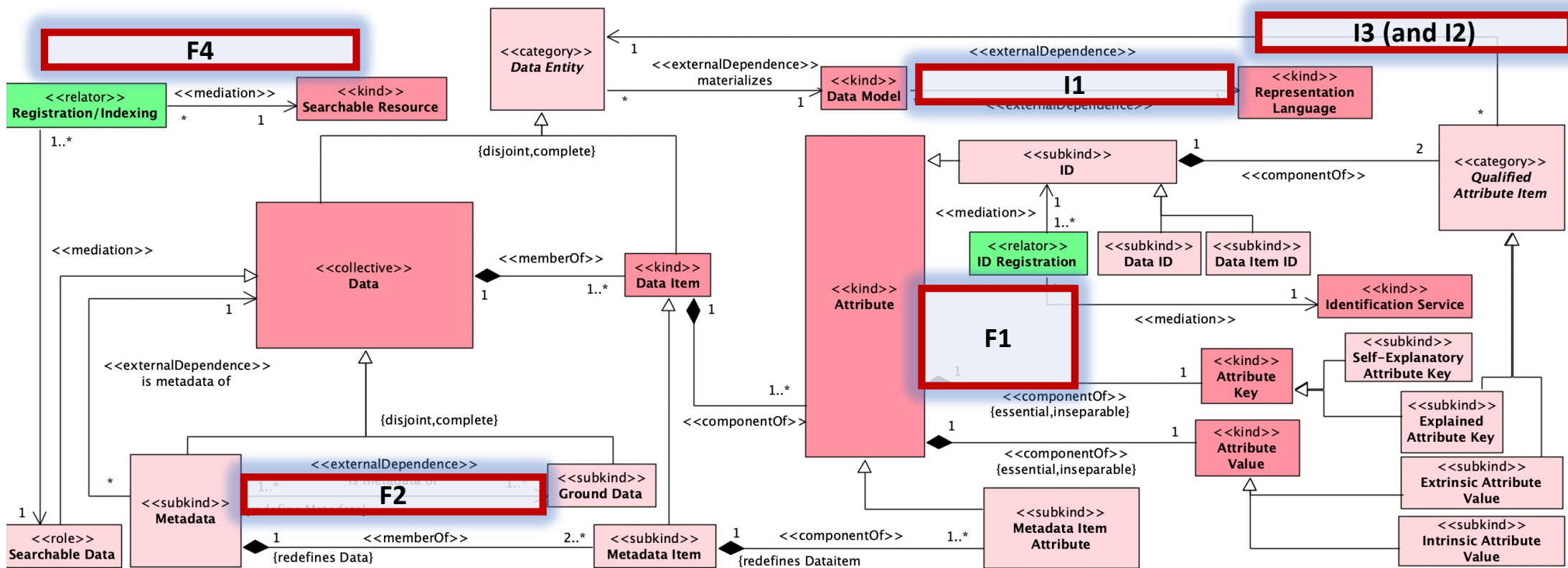
Stereotype	Description	Example
«relator»	Truth-maker of relational propositions. Relations (as classes of n-tuples) can be completely derived from relators.	
«collective»	Plural entity that aggregates parts (members), all of which play the same role with respect to the whole.	
«category»	Necessary properties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



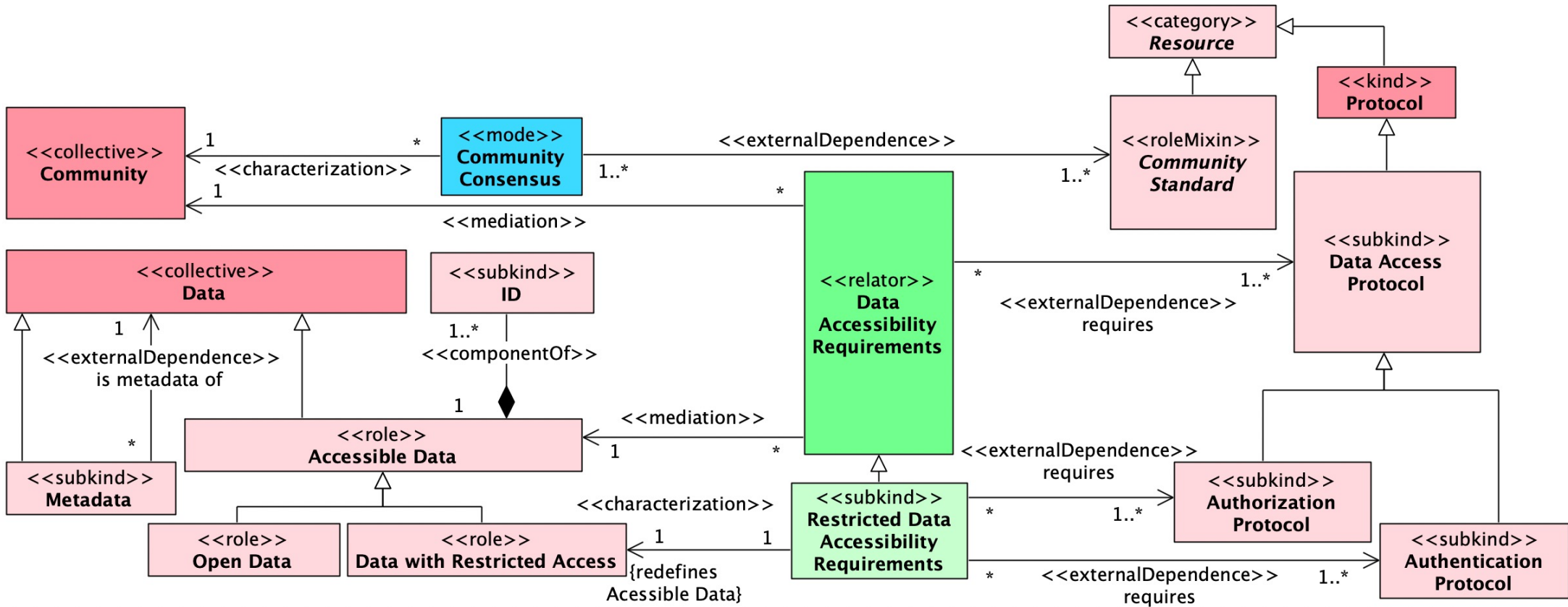
OntoUML FAIR Principles Schema: Findability and Interoperability



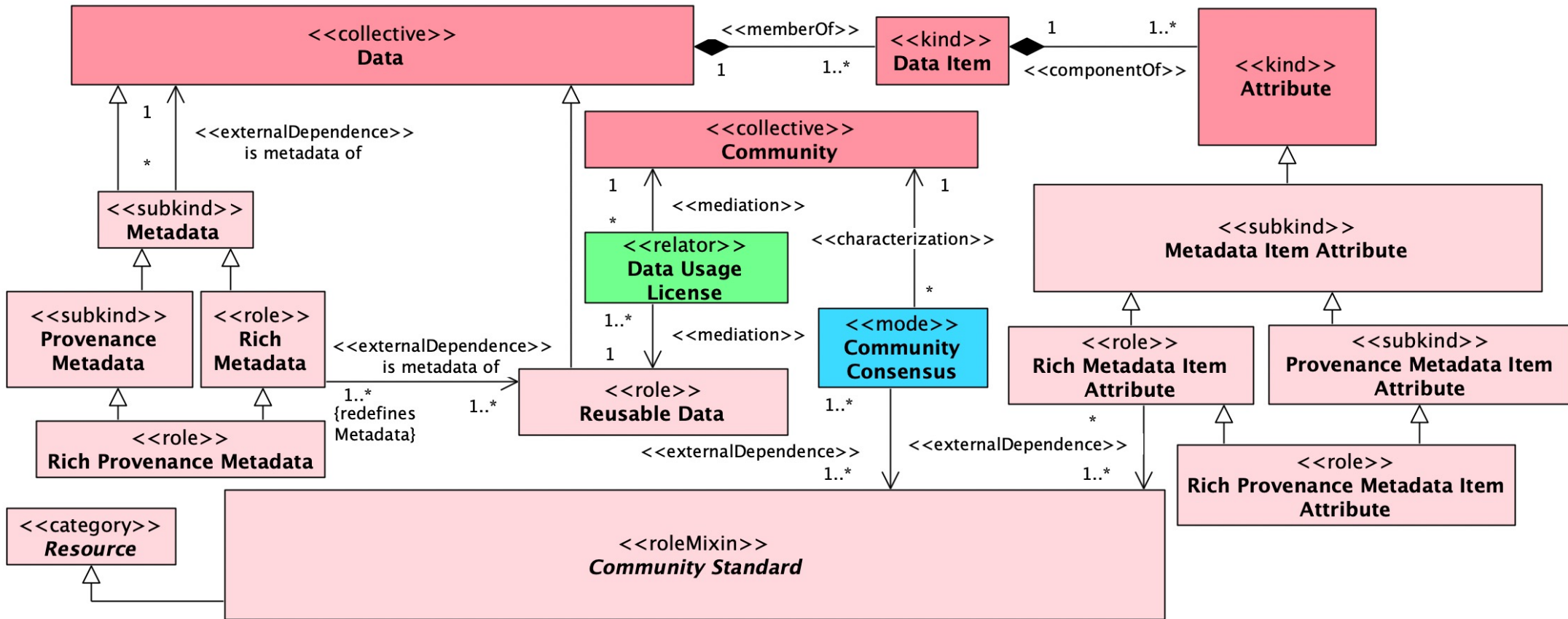
OntoUML FAIR Principles Schema: Findability and Interoperability



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>

NIH National Library of Medicine
National Center for Biotechnology Information

Gene

Full Report ▾

BRCA1 BRCA1 DNA repair associated [*Homo sapiens* (human)]

Gene ID: 672, updated on 11-Jun-2023

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence:

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

NC_000017.11 Find:

Genes, MANE Project (release v1.0)

NCBI RefSeq Annotation GCF_000001405.40-RS_2023_03

Transcript	Accession
NP_001408458.1	NP_001395387.1
NP_001407967.1	NP_001394896.1
NP_001407959.1	NP_001394888.1
NP_001407931.1	NP_001394868.1
NP_001407747.1	NP_001394876.1
NP_001407962.1	NP_001394891.1
NP_001408512.1	NP_001395441.1
NP_001408510.1	NP_001395439.1
NP_001407885.1	NP_001394814.1
NP_001408478.1	NP_001395399.1
NP_001407848.1	NP_001394777.1
NP_001407839.1	NP_001394768.1
NP_001407745.1	NP_001394674.1
NP_001407922.1	NP_001394851.1
NP_001408469.1	NP_001395398.1



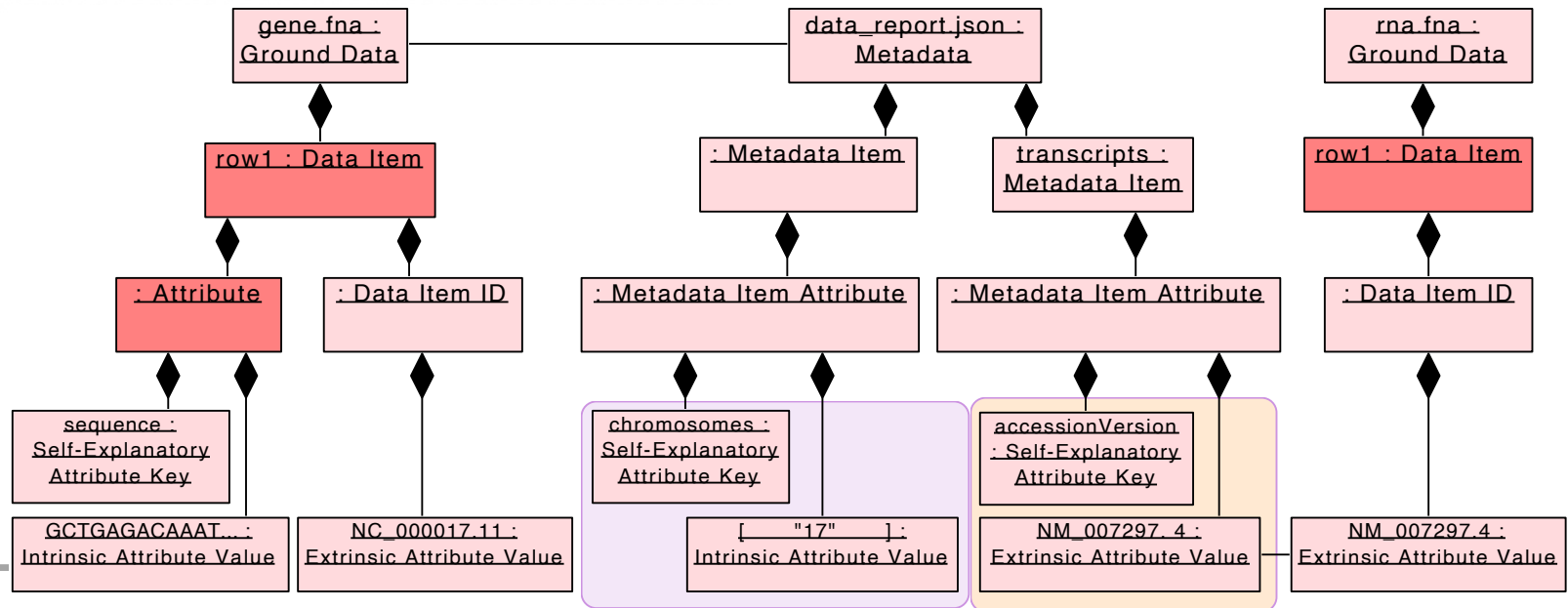
Semantic annotations of (meta)data entities



```

{
  "commonName": "human", "taxId": "9606", "taxname": "Homo sapiens",
  "chromosomes": ["17"],
  "description": "BRCA1 DNA repair associated", "symbol": "BRCA1",
  "geneId": "672", "ensemblGeneIds": ["ENSG0000012048"], "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": {},
  },
},
{}]
}

```





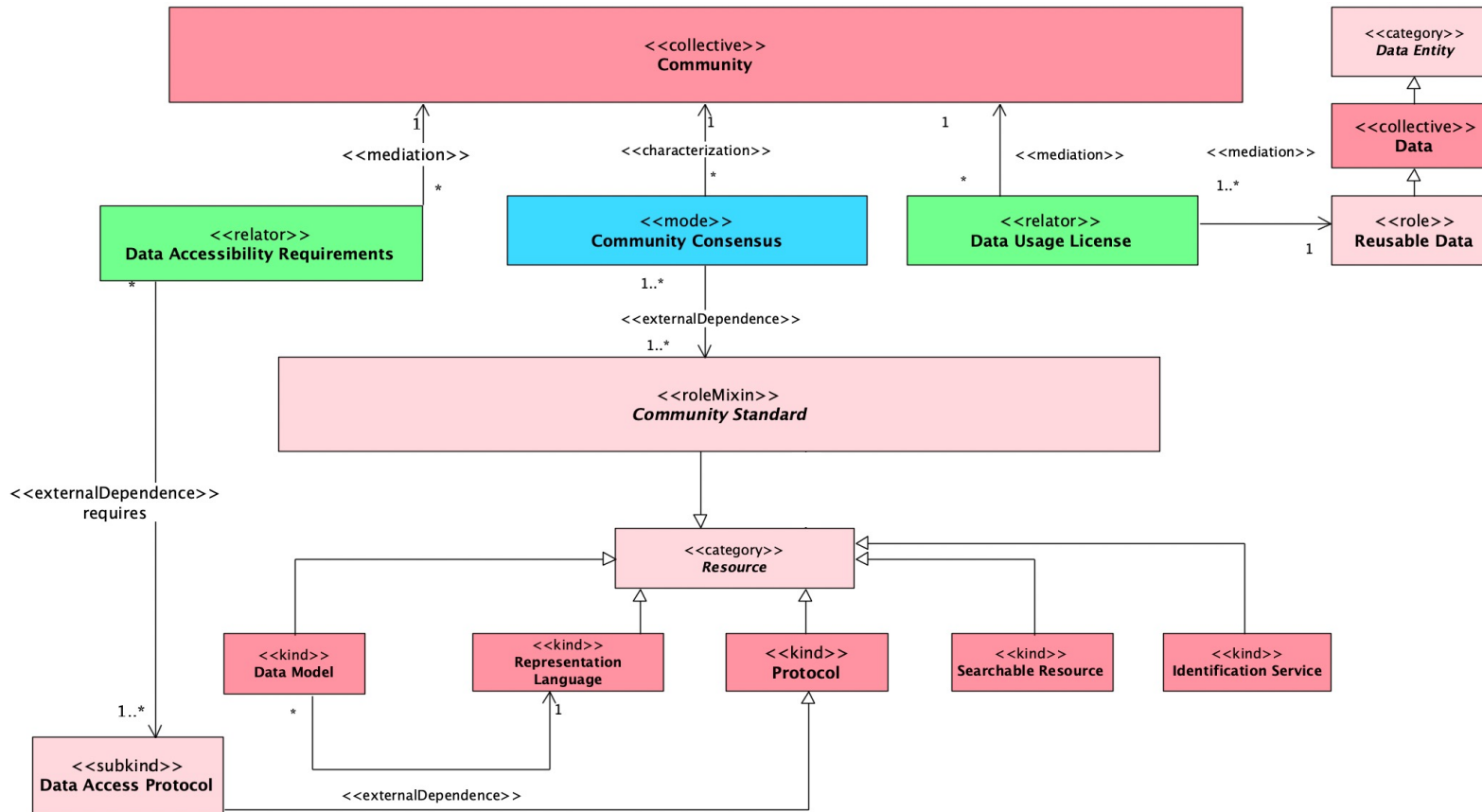
OntoUML FAIR Principles Schema: 4 objectives

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



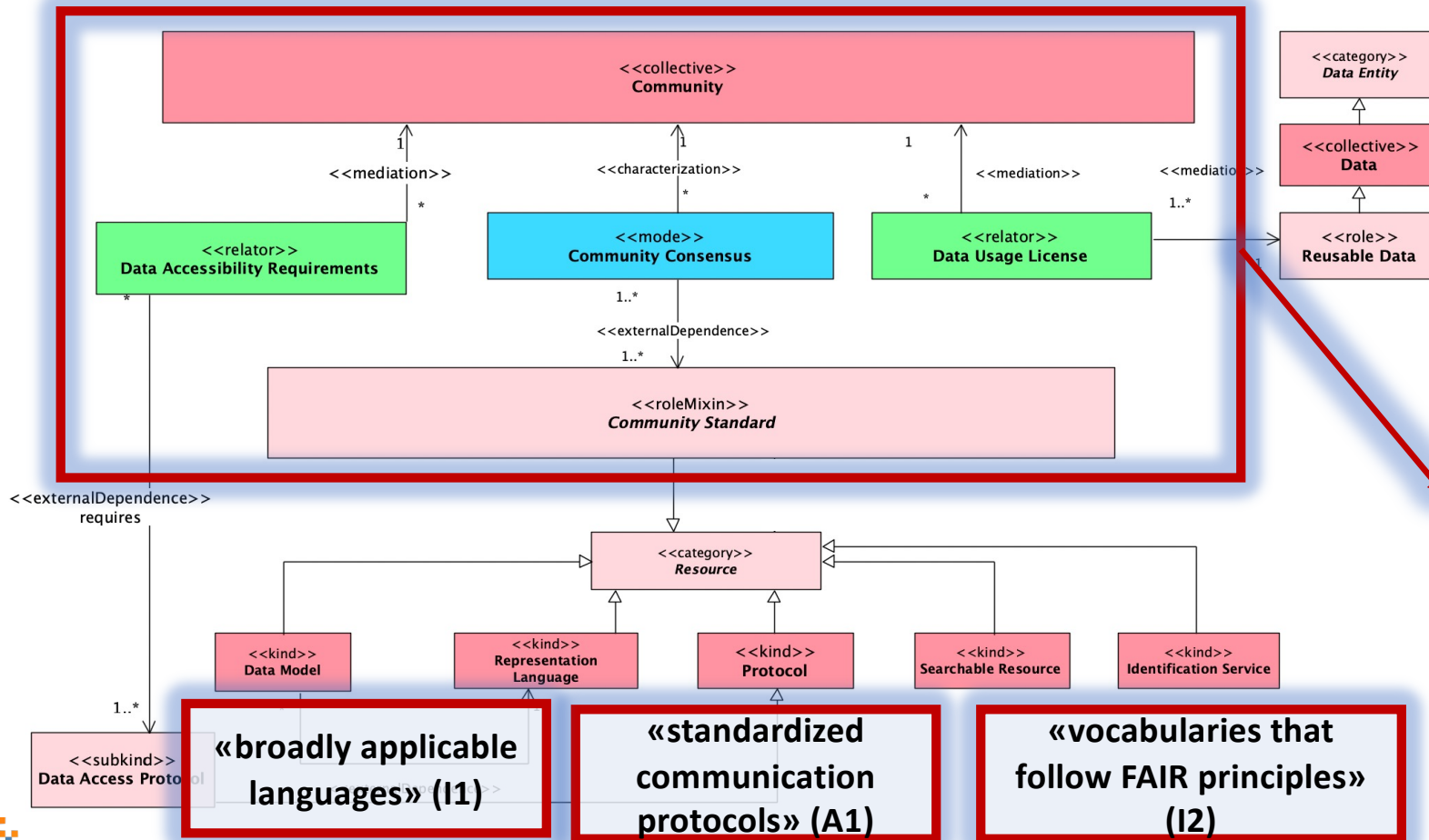


R1.3. (Meta)data meet domain-relevant community standards





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/>

FAIR **OntoUML FAIR Data Point**
Your entrypoint for the OntoUML/UFO Catalog.

Search FAIR Data Point... Log in
Advanced

OntoUML FAIR Data Point / FAIR Model Catalog for Ontology-Driv... / OntoUML FAIR Principles Schema

OntoUML FAIR Principles Schema

Distributions

JSON distribution of OntoUML FAIR Principles Schema
Issued 30-05-2023 Modified 30-05-2023 Media Type
<https://www.iana.org/assignments/media-types/application/json>

PNG distribution of diagram 'Acessibility' from the OntoUML FAIR Principles Schema (original version)
Issued 30-05-2023 Modified 30-05-2023 Media Type
<https://www.iana.org/assignments/media-types/image/png>

PNG distribution of diagram 'Findability Interoperability' from the OntoUML FAIR Principles Schema (original version)
Issued 30-05-2023 Modified 30-05-2023 Media Type
<https://www.iana.org/assignments/media-types/image/png>

PNG distribution of diagram 'Reusability' from the OntoUML FAIR Principles Schema (original version)
Issued 30-05-2023 Modified 30-05-2023 Media Type
<https://www.iana.org/assignments/media-types/image/png>

Conforms to

- [Model Profile](#)

Acronym

- [OntoFAIRPS](#)

Keyword

- [accessibility](#)
- [fair data](#)
- [fair principles](#)

▼ Show more

Theme

[Library of Congress Classification: Class Z](#)

Ontology type

- [Domain](#)
- [Application](#)

International Conference on Conceptual Modeling
↳ ER 2022: [Conceptual Modeling](#) pp 3–17 | [Cite as](#)

Home > [Conceptual Modeling](#) > Conference paper

A FAIR Model Catalog for Ontology-Driven Conceptual Modeling Research

[Pedro Paulo F. Barcelos](#) ✉, [Tiago Prince Sales](#), [Mattia Fumagalli](#), [Claudenir M. Fonseca](#), [Isadora Valle Sousa](#), [Elena Romanenko](#), [Joshua Kritz](#) & [Giancarlo Guizzardi](#)

Conference paper | [First Online: 10 October 2022](#)

658 Accesses | 5 Citations

Part of the [Lecture Notes in Computer Science](#) book series (LNCS, volume 13607)

Abstract

Conceptual models are artifacts representing conceptualizations of particular domains. Hence, multi-domain model catalogs serve as empirical sources of knowledge and insights about specific domains, about the use of a modeling language's constructs, as well as about the patterns and anti-patterns recurrent in the models of that language crosscutting different domains. However, to support domain and language learning, model reuse, knowledge discovery for humans, and reliable automated processing and analysis by machines, these catalogs must be built following generally accepted quality requirements for scientific data





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](#) & [Mireia Costa](#)

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](#)

 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 Model 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 (LNBI, volume 452)

Abstract

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

 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 (LNCS, volume 13811)

Abstract

Conceptual modeling is used to model application domains. It is needed. One of the most complex domains to which conceptual modeling is applied is that of the human genome. Due to its complexity, its use

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

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



**UNIVERSITY
OF TWENTE.**

