# WELCOME

**SESSION 4** 

# **SPONSORED BY**



# FAIR DATA STEWARDSHIP AWARENESS COURSE

FAIR DATA STEWARDSHIP: PRACTISING FAIR DATA





### PRACTICING FAIR DATA STEWARDSHIP

### THE STAGES OF KNOWING

There are four stages of "knowing" things. Like car drivers of 12yo, 17yo, 19yo, 40yo

We want the researchers to become at least consciously incompetent.

They need help for their projects from people that are at least consciously competent.



https://en.wikipedia.org/wiki/Four\_stages\_of\_competence









Distinguish: Information Data Knowledge

Consider:

**Semantics** is the <u>linguistic</u> and <u>philosophical</u> study of <u>meaning</u>.

https://en.wikipedia.org/wiki/Semantics

# **Semantic Data** is, data with meaning.

**4a** Find/Build semantic data model

#### 5a Transform data records into RDF

# **Data:** controlled information (referenced, calibrated, normalized)

•••	•			_				I ne top 1	60 parasite sec	quences heat map.	XISX						
2 🛍	3 🗊 日 📾 📈	h 🖺 💰 🖄	• 🖂 • 🛛 • 🛣 • 搔 • 搔 • 🕻	📧 🎦 🛃 200% 📼 🕜													
	Home Layout T	ables Charts	SmartArt Formulas D	ata Review													
Sort	& Filter Analysi	s	External Data Sources	Tools		Group & Outline											
ZU.	· 💎 · 🚺 ·	9. 🗳 . 1	à tà tà tà.	Text to Columns	• • • •	, 🚼 🗸 📜 Sho	w Detail										
Sort	Filter PivotTable	What-If Refresh T	ext Database HTML FileMaker	Consolidate Validate	temove Group	Ungroup Hide	Detail										
	A2 🛟 😳 🔿	( fx WPNLRTL			pricetes -												
	A		C D			G	H				К		Μ	N	0	P	Q
1	sequence	structureid	naiveC count	AA	HEL1	HEL2	HEL3	HEL4	HEI	L5 HEL	6 HEL	7 HEI	L8	HEL9 I	HEL10	HEL11	HEL12
2	WPNLRTL	1414	15	WPNLRTL		0	0	0	1,4	1,4	1,4	1,4	0,6	0.2		0 (	0 0.72
3	WPNLRTL	1414	15	WPNLRTL		0	0	0	0,9	0,9	0,9	0,9	0,4	0,1		0 (	0 0.48
4	NYMEYMS	888629	9 15	NYMEYMS	-	0 0	,7	0,7	0,7	0,7	0,5	0,2	0,1	0		0 (	0 0,4
5	FDRNEKY	42935	5 16	FDRNEKY		0 0	.1	0,7	0,7	0,7	0,7	0,3	0,2	0		0 (	0 0,38
6	WPNLRTL	1414	15	WPNLRTL		0	0	0	0.7	0.7	0.7	0.7	0.3	0.1		0 (	0 0.34
7	IPRSALA	20609	9 19	IPRSALA	-	0 0	.1	0.3	0,4	0.6	0.6	0.5	0.3	0		0 (	0 0.31
8	FDRNEKY	42935	5 16	FDRNEKY		0	0	0,6	0,6	0,6	0,6	0,2	0,1	0		0 (	0 0,31
9	NYMEYMS	888629	9 15	NYMEYMS		0 0	,5	0,5	0,5	0,5	0,3	0,1	0,1	0		0 (	0 0,28
10	TSPLRAL	9524	15	TSPLRAL		0	0	0.4	0.5	0.5	0.5	0.4	0.2	0		0 (	0 0.27
11	FDRNEKY	42935	5 16	FDRNEKY		0	0	0.5	0.5	0.5	0.5	0.2	0.1	0		0 (	0 0.26
12	IPRSALA	20609	9 19	IPRSALA		0	0	0.3	0.3	0,4	0,4	0.4	0.2	0		0 (	0 0.24
13	VGYRLME	20311	25	VGYRLME		0	0	0,4	0,4	0,4	0,4	0,3	0,2	0		0 (	0 0,23
14	NYMEYMS	888629	9 15	NYMEYMS		0 0	,4	0,4	0,4	0,4	0,2	0,1	0,1	0		0 (	0 0,22
15	HSSSRAI	220	) 15	HSSSRAI		0	0	0,1	0,2	0,4	0,4	0,4	0,3	0,1		0 (	0 0,2
16	IPRSALA	20609	9 19	IPRSALA		0	0	0,2	0,3	0,4	0,4	0,3	0,2	0		0 (	0,19
17	GLGLTQK	237735	5 17	GLGLTQK		0 0	,2	0,2	0,3	0,3	0,3	0,2	0,1	0		0 (	0,19
18	VGYRLME	20311	25	VGYRLME		0	0	0,3	0,3	0,3	0,3	0,2	0,1	0		0 (	0,17
19	STRALHT	237729	9 15	STRALHT		0 0	,2	0,3	0,3	0,3	0,1	0,1	0,1	0		0 (	0,17
20	TSPLRAL	9524	15	TSPLRAL		0	0	0,2	0,3	0,3	0,3	0,2	0,1	0		0 (	0,17
21	LPSLALT	9524	17	LPSLALT		0	0	0,1	0,3	0,3	0,3	0,3	0,1	0		0 (	0,15
22	LTQADSL	20311	15	LTQADSL		0	0	0,3	0,3	0,3	0,3	0,2	0,1	0		0 (	0,15
23	GLGLTQK	237735	5 17	GLGLTQK		0 0	,1	0,1	0,2	0,2	0,3	0,1	0,1	0		0 (	0,14
24	HSSSRAI	220	) 15	HSSSRAI		0	0	0	0,2	0,3	0,3	0,3	0,2	0,1		0 (	0,14
25	VGYRLME	20311	25	VGYRLME		0	0	0,2	0,2	0,2	0,2	0,1	0,1	0		0 (	0,13
26	TTLTNLW	9319	9 16	TTLTNLW		0 0	,1	0,2	0,2	0,2	0,2	0,2	0,1	0		0 (	0,12
27	SLPSSRA	207	7 15	SLPSSRA		0	0	0	0,1	0,2	0,2	0,2	0,2	0,1		0 (	0,12
28	LPSLALT	9524	17	LPSLALT		0	0	0,1	0,2	0,2	0,2	0,2	0,1	0		0 (	0,11
29	GYAHLRD	237569	9 15	GYAHLRD		0 0	,2	0,2	0,2	0,2	0,2	0,1	0	0		0 (	0,11
30	LTQADSL	20311	15	LTQADSL		0	0	0,2	0,2	0,2	0,2	0,1	0,1	0		0 (	0,11
31	STRALHT	237729	9 15	STRALHT		0 0	,2	0,2	0,2	0,2	0,1	0,1	0	0		0 (	0,11
32	TSPLRAL	9524	15	TSPLRAL		0	0	0,1	0,2	0,2	0,2	0,2	0,1	0		0 (	0,11
33	GLGLTQK	237735	5 17	GLGLTQK		0 0	,1	0,1	0,2	0,2	0,2	0,1	0,1	0		0 (	0,1
34	HSSSRAI	220	) 15	HSSSRAI		0	0	0	0,1	0,2	0,2	0,2	0,1	0		0 (	0,1
35	LSSLSLS	9319	9 15	LSSLSLS		0	0	0,1	0,2	0,2	0,2	0,1	0	0		0 (	0,1
36	SSPNTYE	9318	3 15	SSPNTYE		0	0	0,2	0,2	0,2	0,2	0,1	0,1	0		0 (	0,1
37	TPSTSTL	198	3 15	TPSTSTL		0	0	0	0,2	0,2	0,2	0,2	0,1	0		0 (	0,1
38	LPSLALT	9524	17	LPSLALT		0	0	0,1	0,2	0,2	0,2	0,2	0,1	0		0 (	0,09
39	AAPMLRY	9318	3 16	AAPMLRY		0	0	0,1	0,2	0,2	0,2	0,1	0	0		0 (	0,09
40	LTQADSL	20311	15	LTQADSL		0	0	0,2	0,2	0,2	0,2	0,1	0,1	0		0 (	0,09
41	SLPSSRA	207	7 15	SLPSSRA		0	0	0	0,1	0,2	0,2	0,2	0,1	0		0 (	0,09
42	TAVDIVT	0246	45	TAVDIVT		0	4	0.0	0.0	0.0	0.1	0.4	0	0		0	0.00



**Knowledge:** associations implied by data (an understanding about a subject)

Malaria is transmitted by mosquitoes.



ntologies

**Knowledge:** associations implied by data (an understanding about a subject)

Malaria is transmitted by mosquitoes. amantictriple

Subject - Predicate - Object

# ORPHA673 - Do18562 - Anopheles gambiae

http://www.orpha.net/consor/cgi-bin/OC Exp.php?Lng=GB&Expert=673 -

http://bioportal.bioontology.org/ontologies/MESH?p=classes&conceptid=Do18562 -

https://www.vectorbase.org/organisms/anopheles-gambiae

SUPP



# **Knowledge:** associations implied by data (an understanding about a subject)

Data models







20% of known disease causing variants map to TSS



**Knowledge:** associations implied by data (an understanding about a subject)

# **Generation Semantic Data Modeling**

• Deep dive course

**General Common Data Model templates** 

• Libraries of FAIR Data Models (F1000)

• Peer review (Myles Axton)



**4b** Find/Build semantic metadata model **5b** Transform metadata records into RDF



come together to design and jointly endorse a concise and measureable

set of principles that we refer to as the FAIR Data Principles. The intent

is that these may act as a guideline for those wishing to enhance the

**4b** Find/Build semantic metadata model

5b Transform metadata records into RDF

## Metadata: Data about data

#### Acknowledgements

The original Lorentz Workshop 'Jointly Designing a Data FAIRport' was organized by Barend Mons in collaboration with and co-sponsored by the Lorentz center, The Dutch Techcenter for the Life Sciences and the Netherlands eScience Center. The principles and themes described in this manuscript represent the significant voluntary contributions and participation of the authors at, and/or subsequent to, this workshop and from the wider Force11, BD2K and ELIXIR communities. We also acknowledge and thank the organizers and backers of the NBDC/DBCLS BioHackathon 2015, where several of the authors made

### The FAIR Guiding Principles for scientific data management and stewardship

MD Wilkinson, M Dumontier, IJJ Aalbersberg, G Appleton, M Axton, ... Scientific data 3

#### Ασαιτιοπαι Ιπτοrmation

**Competing financial interests:** M.A. is the *Nature Genetics*' Editor in Chief, S.A.S. is *Scientific Data*'s Honorary Academic Editor and consultant.

How to cite this article: Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0

823

5b

Transform

metadata

records

into RDF

**4b** Find/Build semantic metadata model

### Metadata: Data about data

Mark D. Wilkinson<sup>1</sup>, Michel Dumontier<sup>2</sup>, IJsbrand Jan Aalbersberg<sup>3</sup>, Gabrielle Appleton<sup>3</sup>, Myles Axton<sup>4</sup>, Arie Baak<sup>5</sup>, Niklas Blomberg<sup>6</sup>, Jan-Willem Boiten<sup>7</sup>, Luiz Bonino da Silva Santos<sup>8</sup>, Philip E. Bourne<sup>9</sup>, Jildau Bouwman<sup>10</sup>, Anthony J. Brookes<sup>11</sup>, Tim Clark<sup>12</sup>, Mercè Crosas<sup>13</sup>, Ingrid Dillo<sup>14</sup>, Olivier Dumon<sup>3</sup>, Scott Edmunds<sup>15</sup>, Chris T. Evelo<sup>16</sup>, Richard Finkers<sup>17</sup>, Alejandra Gonzalez-Beltran<sup>18</sup>, Alasdair J.G. Gray<sup>19</sup>, Paul Groth<sup>3</sup>, Carole Goble<sup>20</sup>, Jeffrey S. Grethe<sup>21</sup>, Jaap Heringa<sup>22</sup>, Peter A.C. 't Hoen<sup>23</sup>, Rob Hooft<sup>24</sup>, Tobias Kuhn<sup>25</sup>, Ruben Kok<sup>22</sup>, Joost Kok<sup>26</sup>, Scott J. Lusher<sup>27</sup>, Maryann E. Martone<sup>28</sup>, Albert Mons<sup>29</sup>, Abel L. Packer<sup>30</sup>, Bengt Persson<sup>31</sup>, Philippe Rocca-Serra<sup>18</sup>, Marco Roos<sup>32</sup>, Rene van Schaik<sup>33</sup>, Susanna-Assunta Sansone<sup>18</sup>, Erik Schultes<sup>34</sup>, Thierry Sengstag<sup>35</sup>, Ted Slater<sup>36</sup>, George Strawn<sup>37</sup>, Morris A. Swertz<sup>38</sup>, Mark Thompson<sup>32</sup>, Johan van der Lei<sup>39</sup>, Erik van Mulligen<sup>39</sup>, Jan Velterop<sup>40</sup>, Andra Waagmeester<sup>41</sup>, Peter Wittenburg<sup>42</sup>, Katherine Wolstencroft<sup>43</sup>, Jun Zhao<sup>44</sup>

## 53 Authors

### 47 Affiliations

<sup>1</sup>Center for Plant Biotechnology and Genomics, Universidad Politécnica de Madrid, Madrid 28223, Spain.
<sup>2</sup>Stanford University, Stanford 94305-5411, USA. <sup>3</sup>Elsevier, Amsterdam 1043 NX, The Netherlands. <sup>4</sup>Nature Genetics, New York 10004-1562, USA. <sup>5</sup>Euretos and Phortos Consultants, Rotterdam 2741 CA, The Netherlands.
<sup>6</sup>ELIXIR, Wellcome Genome Campus, Hinxton CB10 1SA, UK. <sup>7</sup>Lygature, Eindhoven 5656 AG, The Netherlands.
<sup>8</sup>Vrije Universiteit Amsterdam, Dutch Techcenter for Life Sciences, Amsterdam 1081 HV, The Netherlands.
<sup>9</sup>Office of the Director, National Institutes of Health, Rockville 20892, USA. <sup>10</sup>TNO, Zeist 3700 AJ, The Netherlands. <sup>11</sup>Department of Genetics, University of Leicester, Leicester LE1 7RH, UK. <sup>12</sup>Harvard Medical School, Boston, Massachusetts MA 02115, USA. <sup>13</sup>Harvard University, Cambridge, Massachusetts MA 02138, USA. <sup>14</sup>Data Archiving and Networked Services (DANS), The Hague 2593 HW, The Netherlands. <sup>15</sup>GigaScience, Beijing Genomics Institute, Shenzhen 518083, China. <sup>16</sup>Department of Bioinformatics, Maastricht University, Maastricht 6200 MD, The Netherlands. <sup>17</sup>Wageningen UR Plant Breeding, Wageningen 6708 PB, The

#### GUPRI

ORCID

metadata profile machine-actionable



Metadata: Data about data

# The Future of Semantic Data Modeling

• Metadata Editors:

**Genter for Expanded Data Annotation and Retrieval (CEDAR)** <u>https://metadatacenter.org</u>

Generation DTL https://fair-course.fair-dtls.surf-hosted.nl/editor/#!/

- Common Metadata templates:
  - FAIRSharing <u>https://fairsharing.org</u>
  - People: ORCID profiles <a href="https://orcid.org">https://orcid.org</a>
  - Institutions:

• your institution 😉 https://www.grid.ac

- funding organisations <a href="https://www.crossref.org/services/funder-registry/">https://www.crossref.org/services/funder-registry/</a>
- Libraries of FAIR MetaData Model (F1000)





6 Push to FDP & Certified **Metrics Score** 

# **FAIR Metrics:** http://fairmetrics.org



created by community members themselves, rather than attempting to create a set of one-sizefits-all metrics to apply to every resource.

With a mechanism in-place to design metrics, we now open the process of generating metrics to with most in the Advantage of the second second

**6** Push to FDP & Certified Metrics Score

# FAIR Metrics: http://fairmetrics.org



With a mechanism in-place to design metrics, we now open the process of generating metrics to

**6** Push to FDP & Certified Metrics Score

# FAIR Metrics: http://fairmetrics.org



With a mechanism in-place to design metrics, we now open the process of generating metrics to

### **14 CORE METRICS**

# Findable:

F1 (meta)data are assigned a globally unique and persistent identifier;

F2 data are described with rich metadata;

F3 metadata clearly and explicitly include the identifier of the data it describes;

F4 (meta)data are registered or indexed in a searchable resource;

## **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;

# Interoperable:

I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

12 (meta)data use vocabularies that follow FAIR principles;

13 (meta)data include qualified references to other (meta)data;

# **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;

Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016) http://fairmetrics.org

### **14 CORE METRICS**

# Findable:

F1 (meta)data are assigned a globally unique and persistent identifier;

FM-F1B

FM-F3

F2 data are described with rich metadata;

F3 metadata clearly and explicitly include the identifier of the data it describes;

FM-F4 F4 (meta)data are registered or indexed in a searchable resource;

# Interoperable:

I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge FM-I1 representation.

12 (meta)data use vocabularies that follow FAIR principles,

I3 (meta)data include qualified references to other (meta)data; FM-I3

Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016) http://fairmetrics.org https://github.com/FAIRMetrics/Metrics/blob/master/ALL.pdf

# 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 Fata are no longer available;

# **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; FM-R1.1

R1.2 (meta)data are associated with detailed provenance;

#### **FM-R1.2**

R1.3 (meta)data meet domain-relevant community standards; FM-R1.3 **FAIR Principle F1**: (meta) data are assigned globally unique and persistent identifiers. Fundamental requirement for accurate and reproducible machine actionability. Examples: Universally unique identifier (UUID): <u>https://en.wikipedia.org/wiki/Universally\_unique\_identifier</u>; Digital Object Identifier (DOI): <u>http://www.doi.org</u>

### 1.FAIR Metric F1A:

Question 1: Provide an URL to a registered scheme that defines the globally-unique structure of the identifier(s) for your digital resource.

#### 2.FAIR Metric F1B:

Question 2: Provide an URL to a document that defines the persistence policy of your identifier(s).

### Community challenge:

(1) What are your required (or preferred) identifier registration services ?

(2) What is your minimal persistence policy?

(3) Can you make your persistence policy machine-readable?

### **FAIR Principle F2:** F2: Data are described with rich metadata

Data should be machine-discoverable by the widest range of stakeholders possible. That is, you should not presume who will want to use your data, or for what purpose. Resource providers should be generous and expansive with their metadata (see also R1). Exemplar rich metadata frameworks: DCAT; ISA framework.

FAIR Metric F2: Question 3.Provide the URL to a document that contains machine-readable metadata for the digital resource.

Question 4. Provide the URL for the file format of this metadata.

Community challenge: (4) Can you define a minimal set of metadata for your community? (5) Can you make your metadata machine-readable? **FAIR Principle F3:** Metadata clearly & explicitly include the identifier of the data it describes Metadata and the dataset they describe are often separate (and probably should be, v.v. Principle A2). Because most data formats are not extendible, and therefore cannot always refer to the metadata, the association between metadata and the data can often only be achieved by explicit references to the data's globally unique identifier within the metadata record, thus facilitating ID-based search.

**FAIR Metric F3**: https://purl.org/fair-metrics/FM\_F3

Question 5. Provide an URL to the metadata document that contains the globally unique and persistent identifier for the digital resource.

Question 6. Provide the URL to the data described by in that metadata document.

Community challenge:

(6) Can you define the metadata model that explicitly links data and metadata?(7) Can you make this metadata model machine-readable?

<ul> <li>(3) Can you make your persistence policy machine-readable?</li> <li>(4) Can you define a minimal set of metadata for your community?</li> <li>(5) Can you make your metadata machine-readable?</li> </ul>
(4) Can you define a minimal set of metadata for your community? (5) Can you make your metadata machine-readable?
<b>FINC</b> (5) Can you make your metadata machine-readable?
()) can you make your metadata machine readable.
(6) Can you define the metadata model that explicitly links data and metadata?
(7) Can you make this metadata model machine-readable?
(8) What is the required (preferred) search engine for your community ?
(9) What is the required (preferred) communication protocol for your community
(10) What is your required (preferred) protocols for restricting access to data ?
(11) Can you make this protocol machine-readable?
(12) What is your minimal persistence policy for metadata?
(13) Can you make this persistence policy machine-readable?
(14) What is your required (preferred) standards in knowledge representation ?
(15) What are your required (preferred) vocabularies ?
(16) What is your required LinkSet ?
(17) What is your required (preferred) usage license framework?
(18) Can you make these usage licenses machine-readable?
(19) What is your required (preferred) provenance metadata descriptions?
<b>Keuse</b> (20) Can you make this provenance metadata machine-readable?
(21) What are your certification criteria for data & metadata?
(22) What is your machine-actionable validation-certification system ?

### FAIR METRICS COMMUNITY CHALLENGES

#### Community Challenges toward increased FAIRness

The metrics framework provides not only Metric Tests that deliver FAIRness scores for a digital resource, but more importantly, an indication about where the digital resource is failing to comply with the FAIR Principles and thus a clear directive as to how compliance may be remedied. Each stakeholder community that aspires to high levels of FAIRness has the obligation to commit to, develop and deploy the standards and resources necessary to comply with the 15 principles in ways that are most relevant to them. The 14 Core Metrics indicate (at least) 29 Community Challenges. Many of the Community Challenges listed here may be generic to numerous communities, so that when solutions are found in one community, they could be widely re-used by others.

#### References:

FAIR Principles paper: https://www.nature.com/articles/sdata201618 FAIR Principles explained: https://www.go-fair.org/fair-principles/ FAIR Metrics paper: https://www.nature.com/articles/sdata2018118 FAIR Metrics explained: https://github.com/FAIRMetrics/Metrics/blob/master/ALL.pdf

#### FAIR Metric F1A

1. The community should choose what are preferred (or required) identifier registration services, for its own purposes.

2. The community should define how to reference in a machine-readable manner, the preferred (or required) identifier registration services.

#### FAIR Metric F1B

3. The community should define minimal persistence policy requirements for its chosen identifier registration services.

4. The identifier registration services should define, or preferably, re-use existing machine-readable templates for persistence policy documents.

#### FAIR Metric F2

The community should define a minimal set of required metadata elements to optimize machine Findability for its own purposes.
 The community should define, or preferably, re-use existing machine-readable templates for Findability-related metadata.

#### FAIR Metric F3

. The community should define or preferably, re-use a machine-readable metadata model that explicitly links metadata to data.

#### FAIR Metric F4

8. The community should choose what are preferred (or required) search engines for its own purposes.9. The community should define how to reference in a machine-readable manner, the preferred (or required) search engines.

#### FAIR Metric A1.1

10. The community should choose what are preferred (or required) communication protocols for for its own purposes.11. The community should define how to reference in a machine-readable manner, the preferred (or required) communication protocols.

#### FAIR Metric A1.2

12. The community should choose what are preferred (or required) protocols when restricting access to data.

13. The community should define how to reference in a machine-readable manner, the preferred (or required) communication protocols.

#### FAIR Metric A2

14. The community should define what are preferred (or required) longevity plan (persistence policy) for metadata?

15. The community should define, or preferably, re-use existing machine-readable templates for metadata-related persistence policy documents.

### FAIR METRICS COMMUNITY CHALLENGES

#### Community Challenges toward increased FAIRness

The metrics framework provides not only Metric Tests that deliver FAIRness scores for a digital resource, but more importantly, an indication about where the digital resource is failing to comply with the FAIR Principles and thus a clear directive as to how compliance may be remedied. Each stakeholder community that aspires to high levels of FAIRness has the obligation to commit to, develop and deploy the standards and resources necessary to comply with the 15 principles in ways that are most relevant to them. The 14 Core Metrics indicate (at least) 29 Community Challenges. Many of the Community Challenges listed here may be generic to numerous communities, so that when solutions are found in one community, they could be widely re-used by others.

#### References:

FAIR Principles paper: https://www.nature.com/articles/sdata201618 FAIR Principles explained: https://www.go-fair.org/fair-principles/ FAIR Metrics paper: https://www.nature.com/articles/sdata2018118 FAIR Metrics explained: https://github.com/FAIRMetrics/Metrics/blob/master/ALL.pdf

#### FAIR Metric F1A

1. The community should choose what are preferred (or required) identifier registration services, for its own purposes.

2. The community should define how to reference in a machine-readable manner, the preferred (or required) identifier registration services.

#### FAIR Metric F1B

3. The community should define minimal **persistence policy requirements** for its chosen identifier registration services.

4. The identifier registration services should define, or preferably, re-use existing machine-readable templates for persistence policy documents.

#### FAIR Metric F2

The community should define a minimal set of required metadata elements to optimize machine Findability for its own purposes.
 The community should define, or preferably, re-use existing machine-readable templates for Findability-related metadata.

#### FAIR Metric F3

. The community should define or preferably, re-use a machine-readable metadata model that explicitly links metadata to data.

#### FAIR Metric F4

8. The community should choose what are preferred (or required) search engines for its own purposes.9. The community should define how to reference in a machine-readable manner, the preferred (or required) search engines.

#### FAIR Metric A1.1

10. The community should choose what are preferred (or required) communication protocols for for its own purposes.11. The community should define how to reference in a machine-readable manner, the preferred (or required) communication protocols.

#### FAIR Metric A1.2

12. The community should choose what are preferred (or required) protocols when restricting access to data.

13. The community should define how to reference in a machine-readable manner, the preferred (or required) communication protocols.

#### FAIR Metric A2

14. The community should define what are preferred (or required) longevity plan (persistence policy) for metadata?

15. The community should define, or preferably, re-use existing machine-readable templates for metadata-related persistence policy documents.

#### FAIR Metric I1

16. The community should choose what is its preferred (or required) language for knowledge representation.

17. The community should define how to reference in a machine-readable manner, the preferred (or required) language for knowledge representation.

#### FAIR Metric I2

18. The community should choose what is its preferred (or required) units of measure, vocabularies, ontologies, and conceptual mappings.

19. The community should define how to reference in a machine-readable manner, the preferred (or required) units of measure, vocabularies, ontologies, and conceptual mappings.

#### FAIR Metric 13

20. The community should define what is its preferred (or required) formal LinkSet.

21. The community should define how to reference in a machine-readable manner, the preferred (or required) formal LinkSet.

#### FAIR Metric R1.1

22. The community should choose or define what is its preferred (or required) usage license or licensing requirements.

23. The community should define, or preferably, re-use existing machine-readable templates for licenses.

24. The community should define how to reference in a machine-readable manner, the preferred (or required) usage license.

#### FAIR Metric R1.2

25. The community should define what is its preferred (or required) provenance metadata descriptions.

26. The community should define, or preferably, the re-use existing machine-readable templates for provenance metadata descriptions.

27. The community should define how to reference in a machine-readable manner, the preferred (or required) provenance metadata descriptions.

#### FAIR Metric R1.3

28. The community should define what is its preferred (or required) certification criteria for data & metadata. [Comments here about what the process is...where is authority derived from] 29. The community should define a machine-actionable validation and certification system for data & metadata compliance.

#### FAIR Metric I1

16. The community should choose what is its preferred (or required) language for knowledge representation.

17. The community should define how to reference in a machine-readable manner, the preferred (or required) language for knowledge representation.

#### FAIR Metric I2

18. The community should choose what is its preferred (or required) units of measure, vocabularies, ontologies, and conceptual mappings.

19. The community should define how to reference in a machine-readable manner, the preferred (or required) units of measure, vocabularies, ontologies, and conceptual mappings.

#### FAIR Metric 13

20. The community should define what is its preferred (or required) formal LinkSet.

21. The community should define how to reference in a machine-readable manner, the preferred (or required) formal LinkSet.

#### FAIR Metric R1.1

22. The community should choose or define what is its preferred (or required) usage license or licensing requirements.

23. The community should define, or preferably, re-use existing machine-readable templates for licenses.

24. The community should define how to reference in a machine-readable manner, the preferred (or required) usage license.

#### FAIR Metric R1.2

25. The community should define what is its preferred (or required) provenance metadata descriptions.

26. The community should define, or preferably, the **re-use existing machine-readable templates** for provenance metadata descriptions.

27. The community should define how to reference in a machine-readable manner, the preferred (or required) provenance metadata descriptions.

#### FAIR Metric R1.3

28. The community should define what is its preferred (or required) certification criteria for data & metadata. [Comments here about what the process is...where is authority derived from] 29. The community should define a machine-actionable validation and certification system for data & metadata compliance.

### FAIR METRICS COMMUNITY CHALLENGES 2.0.







DANS



@prefix this: <http://purl.org/np/RAbMNmYJ5wJEyJ-EymT-f67hoLUm1Xj0tQEaApAWmiEOE> . @prefix sub: <http://purl.org/np/RAbMNmYJ5wJEyJ-EymT-f67hoLUm1Xj0tQEaApAWmiEOE#> . @prefix xsd: <http://www.w3.org/2001/XMLSchema#> . @prefix dct: <http://purl.org/dc/terms/> . @prefix pav: <http://purl.org/pav/> . @prefix np: <http://www.nanopub.org/nschema#> . @prefix npx: <http://purl.org/nanopub/x/> . @prefix orcid: <https://orcid.org/> . @prefix fairv: <http://www.nanopubs.d2s.labs.vu.nl/fair/fair-vocabulary.html#> . @prefix fairp: <http://www.nanopubs.d2s.labs.vu.nl/fair/fair-principles.html#> . @prefix faircc: <http://www.nanopubs.d2s.labs.vu.nl/fair/fair-challenges.html#> . @prefix fairm: <http://www.nanopubs.d2s.labs.vu.nl/fair/fair-metrics.html#> . sub:Head { this: np:hasAssertion sub:assertion ; np:hasProvenance sub:provenance ; np:hasPublicationInfo sub:pubinfo ; a np:Nanopublication . sub:assertion { faircc:Challenge1 dct:description "The community should choose what are preferred (or required) identifier registration services, for its own purposes."; fairv:addresses fairm:Metric1 , fairp:Findable . sub:provenance { sub:assertion pav:authoredBy orcid:0000-0001-8888-635X , orcid:0000-0003-4818-2360 sub:pubinfo { this: dct:created "2018-10-15T07:22:11+01:00"^^xsd:dateTime ;

pav:createdBy orcid:0000-0002-1267-0234 ; a npx:ExampleNanopub .

### THE M4M PUBLIC PRIVATE PARTNERSHIP PILOT



### **RESOURCES FOR THE WEEK**

### https://fair-course.fair-dtls.surf-hosted.nl

# **Data FAIRification for Data Stewards**

### Links

- What is FAIR? / What FAIR is not
- FAIR Principles explained
- FAIR Metrics

### Resources

• Course information

### Tools

- FAIR Metadata Editor
- FAIRifier
- FAIR DataPoint
- FAIR Search Engine

