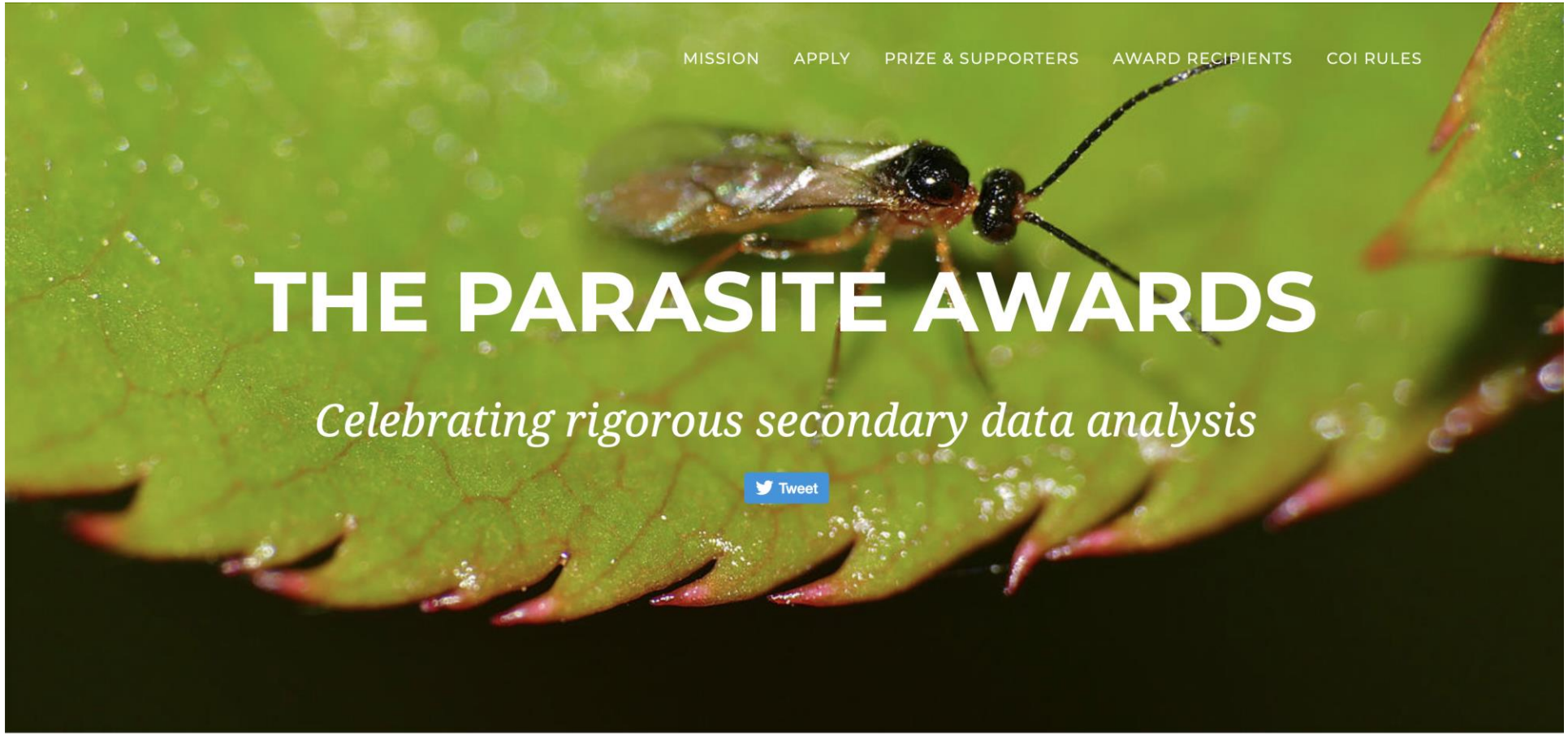
It’s All in the Metadata: Making Datasets FAIR

Mark A. Musen, M.D., Ph.D.
Stanford Center for Biomedical Informatics Research
Stanford University
musen@stanford.edu

CEDAR
CENTER FOR EXPANDED DATA ANNOTATION AND RETRIEVAL


The Economist
OCTOBER 19TH–25TH 2013
Economist.com
Washington's lawyer surplus
How to do a nuclear deal with Iran
Investment tips from Nobel economists
Junk bonds are back
The meaning of Sachin Tendulkar
HOW SCIENCE GOES WRONG.
99
Einsteinium



THE "PARASITES"

PSB Awards for rigorous secondary data analysis. A companion to the Research Symbiont Awards.

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES

» Research data

» Publication characteristics

Received: 10 December 2015

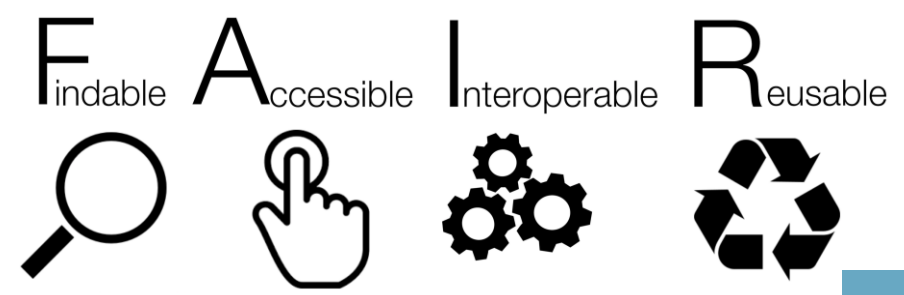
Accepted: 12 February 2016

Published: 15 March 2016

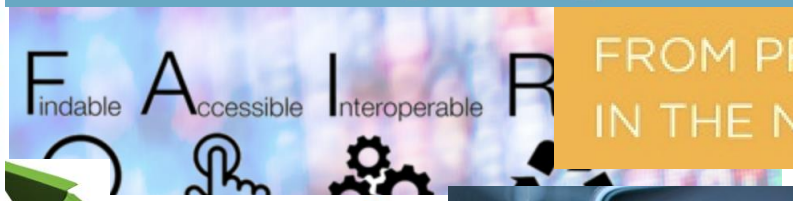
Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have 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 reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.



Why we chose Fair Data



FROM PRINCIPLES TO
IN THE NETHERLANDS

The FAIR Guiding Principles

F1: (Meta) data are assigned globally unique and persistent identifiers

F2: Data are described with rich metadata

F3: Metadata clearly and explicitly include the identifier of the data they describe

F4: (Meta)data are registered or indexed in a searchable resource

A1: (Meta)data are retrievable by their identifier using a standardised communication protocol

A1.1: The protocol is open, free and universally implementable

A1.2: The protocol allows for an authentication and authorisation where necessary

A2: Metadata should be accessible even when the data is no longer available

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

I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data

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

FAIR Evaluator API V.0.3.0 OAS3

The **FAIR Evaluator** - automated testing of Web resources for their compliance

FAIR Metrics Evaluation

HIDE DETAILS

Registry URL	http://smart-api.info/registry?q=ad830426bed193d36838091ef5d14407
SmartAPI ID	ad830426bed193d36838091ef5d14407 Copy
Metadata URL	https://w3id.org/FAIR_Evalu...
Version	V 0.3.0
Contact	Mark Wilkinson

NIH LINCS PROGRAM

FAIRness of LINCS Datasets and Tools project

FAIR evaluation of the LINCS NIH Program tools and datasets

Tags: DCPPC

URL(s):

http://lincsproject.org/

View Analytics View Assessments

Associated Digital Objects (81)

Assess

L1000CDS2 tool

An ultra-fast LINCS L1000 Characteristic Direction signature search engine

test

Assess

iLINCS tool

An integrative web platform for analysis of LINCS data and signatures

test

Assess

Query App tool

Connections to user-defined signatures

test

Assess

Drug-Pathway Browser tool

Interactive map of key signal transduction pathways and drug-target data

test

Swagger
Supported by SMARTBEAR

/fuji/api/v1/openapi.json

Explore

F-UJI 1.0.0 OAS3

/fuji/api/v1/openapi.json

A Service for Evaluating Research Data Objects Based on FAIRsFAIR Metrics.

This work was supported by the FAIRsFAIR project (H2020-INFRAEOSC-2018-2020 Grant Agreement 831558).

Contact the developer

MIT License

Find out more about F-UJI

Servers

/fuji/api/v1

Authorize

FAIR object FAIRness assessment of a data object

POST /evaluate

FAIR metric FAIRsFAIR assessment metrics

GET /metrics Return all metrics and their definitions

All these systems to evaluate FAIRness have struggled to find an audience

- Scientists really don’t want a FAIR “report card”
- No one wants to hear about problems with datasets that have *already* been uploaded to a repository
- There is no fully computable solution to the question of whether a dataset is FAIR in the first place

The FAIR Guiding Principles

F1: (Meta) data are assigned globally unique and persistent identifiers

F2: Data are described with rich metadata

F3: Metadata clearly and explicitly include the identifier of the data they describe

F4: (Meta)data are registered or indexed in a searchable resource

A1: (Meta)data are retrievable by their identifier using a standardised communication protocol

A1.1: The protocol is open, free and universally implementable

A1.2: The protocol allows for an authentication and authorisation where necessary

A2: Metadata should be accessible even when the data is no longer available

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

I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data

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

Most FAIR principles are about *metadata*

F1: (Meta) data are assigned globally unique and persistent identifiers

F2: Data are described with rich metadata

F3: Metadata clearly and explicitly include the identifier of the data they describe

F4: (Meta)data are registered or indexed in a searchable resource

A1: (Meta)data are retrievable by their identifier using a standardised communication protocol

A1.1: The protocol is open, free and universally implementable

A1.2: The protocol allows for an authentication and authorisation where necessary

A2: Metadata should be accessible even when the data is no longer available

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

I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data

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

Scientists have no direct control over repository infrastructure

F1: ~~(Meta)data are assigned globally unique and persistent identifiers~~

F2: Data are described with rich metadata

F3: ~~Metadata clearly and explicitly include the identifier of the data they describe~~

F4: ~~(Meta)data are registered or indexed in a searchable resource~~

A1: ~~(Meta)data are retrievable by their identifier using a standardised communication protocol~~

A1.1: ~~The protocol is open, free and universally implementable~~

A1.2: ~~The protocol allows for an authentication and authorisation where necessary~~

A2: ~~Metadata should be accessible even when the data is no longer available~~

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

I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data

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

Many FAIR principles depend on community standards for metadata and are not objectively computable

F1: ~~(Meta)data are assigned globally unique and persistent identifiers~~

F2: Data are described with rich metadata

F3: ~~Metadata clearly and explicitly include the identifier of the data they describe~~

F4: ~~(Meta)data are registered or indexed in a searchable resource~~

A1: ~~(Meta)data are retrievable by their identifier using a standardised communication protocol~~

A1.1: ~~The protocol is open, free and universally implementable~~

A1.2: ~~The protocol allows for an authentication and authorisation where necessary~~

A2: ~~Metadata should be accessible even when the data is no longer available~~

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

I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data

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

Metadata in public repositories are a mess!

- Investigators view their work as publishing papers, not leaving a legacy of reusable data
- Sponsors may require data sharing, but they do not encourage the use of grant funds to pay for it
- Creating the metadata to describe data sets is unbearably hard

	A	B	C	E	F	G
1	# Use this template for 3' or whole Gene expression studies when summarization probe set data will be provided as **CHP files**.					
2	# Do **NOT** submit CHP files unless they are relevant to your analysis (instead, use the Matrix table option to submit the relevant data, e.g. **Bioconduct**					
3	# Incomplete submissions will be returned. Click the **Metadata Example** tab below to view a completed worksheet					
4	# A complete submission will consist of: (1) a completed metadata worksheet, (2) the CHP files, and (3) the original CEL files.					
5	**# Field names (in blue on this page) should not be edited. Hover over cells containing field names to view field content guidelines or,**					
6	**# CLICK HERE for Field Content Guidelines Web page.**					
7						
8	**SERIES**					
9	# This section describes the overall					
10						
11	**title**					
12	**summary**					
13	**summary**					
14	**overall design**					
15	**contributor**					
16	**contributor**					
17						
18	**SAMPLES**					
19	**# The Sample names in the first column are arbitrary but they must match the column headers of the Matrix table (see next worksheet).**					
20						
21	**Sample name**	**title**	**CHP file**	**source name**	**organism**	**characteristics: tag**
22	**SAMPLE 1**					
23	**SAMPLE 2**					
24	**SAMPLE 3**					
25	**SAMPLE 4**					
26	**SAMPLE 5**					
27	**SAMPLE 6**					
28	**SAMPLE 7**					
29	**SAMPLE 8**					
30	**SAMPLE 9**					
31	**SAMPLE X**					
32						
33						
34	**PROTOCOLS**					
35	# This section includes protocols and fields which are common to all Samples.					
36	# Protocols which are applicable to specific Samples or specific channels should be included in additional columns of the **SAMPLES** section instead.					
37						
38	**growth protocol**					
39	**treatment protocol**					
40	**extract protocol**					
41	**label protocol**					
42	**hyb protocol**					

Unique title (less than 120 characters) that describes the overall study.

"Firstname,Initial,Lastname". Example: "John,H,Smith" or "Jane,Doe".

Unique title that describes the Sample. We suggest that you use the convention: [biomaterial]-[condition(s)]-[replicate number], e.g., Muscle_exercised_60min_rep2.

Replace 'tag' with a biosource characteristic (e.g. "gender", "strain", "tissue", "developmental stage", "tumor stage", etc), and then enter the value for each sample beneath (e.g. "female", "129SV", "brain", "embryo", etc). You may add additional characteristics columns to this template (see 'Metadata Example' spreadsheet).

[Optional] Describe the conditions that were used to grow or maintain organisms or cells prior to extract preparation.

Full Send to:

JHH-2, human cell line STR and SNP profiles from GNE, Genentech

Identifiers BioSample: SAMN03473249; GNE: GNE Tracking ID: 586138

Organism Homo sapiens (human)

Attributes

cell line	JHH-2
culture collection	GNE:586138
repository	Genentech (GNE)
tissue	liver
disease	carcinoma hepatocellular
sex	male
ethnicity	japanese
age	57 year
development stage	adult
canonical name	JHH-2
human cell line STR profile	yes
human cell line STR profile status	repository authenticated
human cell line SNP profile	yes

BioProject PRJNA271020 Homo sapiens
Retrieve all samples from this project

Related information

BioProject

BioCollections

Taxonomy

Recent activity

Turn Off Clear

JHH-2, human cell line STR and SNP profiles from GNE, sut biosample

Human sample from Homo sapiens biosample

"disease=carcinoma hepatocellular"[attr] (28) BioSample

Con rep1 biosample

carcinoma hepatocellular (11749) BioSample

See more...

LinkOut to external resources

JHH-2 (CVCL_2786)

NCBI Resources How To Sign in to NCBI

Full Send to:

JHH-2, human cell line STR and SNP profiles from GNE, Genentech

Identifiers BioSample: SAMN03473249; GNE: GNE Tracking ID: 586138

Organism Homo sapiens (human)

Attributes

cell line	JHH-2
culture collection	GNE:586138
repository	Genentech (GNE)
tissue	liver
disease	carcinoma hepatocellular
sex	male
ethnicity	japanese
age	57 year
development stage	adult
canonical name	JHH-2
human cell line STR profile	yes
human cell line STR profile status	repository authenticated
human cell line SNP profile	yes

BioProject PRJNA271020 Homo sapiens
Retrieve all samples from this project

Related information

BioProject

BioCollections

Taxonomy

Recent activity

Turn Off Clear

JHH-2, human cell line STR and SNP profiles from GNE, sul biosample

Human sample from Homo sapiens biosample

"disease=carcinoma hepatocellular"[attr] (28) BioSample

Con rep1 biosample

carcinoma hepatocellular (11749) BioSample

See more...

LinkOut to external resources

JHH-2 (CVCL_2786)

NCBI *BioSample* Metadata are Dreadful!

- 73% of “Boolean” metadata values are not actually *true* or *false*
  - *nonsmoker*, *former-smoker*
- 26% of “integer” metadata values cannot be parsed into integers
  - *JM52*, *UVPgt59.4*, *pig*
- 68% of metadata entries that are supposed to represent terms from biomedical ontologies do not actually do so
  - *presumed normal*, *wild_type*

Metadata authors need to use controlled terms!

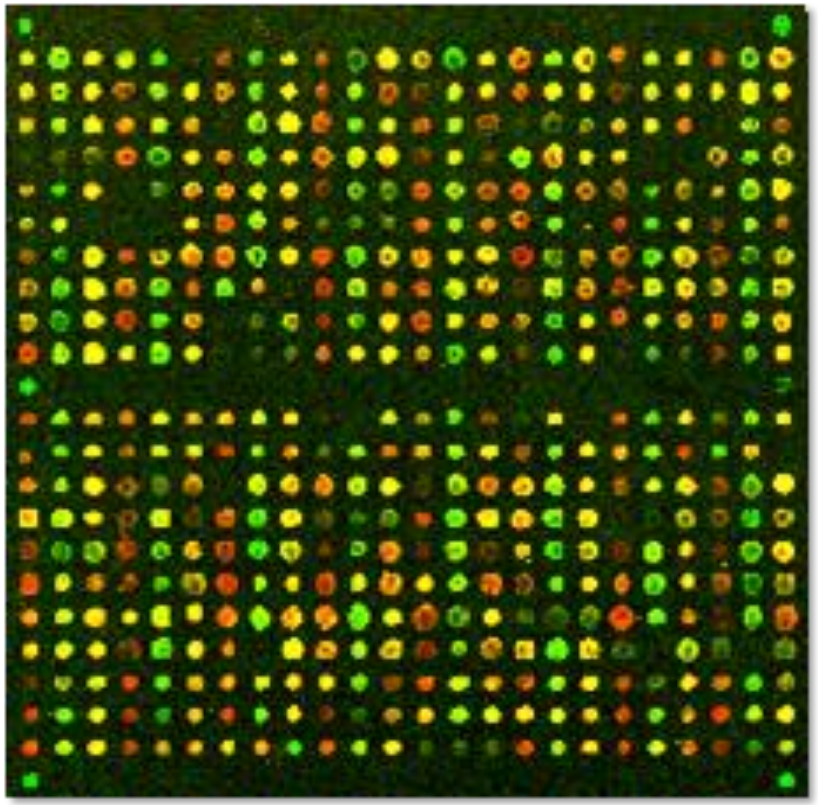
age
Age
AGE
`Age
age (after birth)
age (in years)
age (y)
age (year)
age (years)
Age (years)
Age (Years)
age (yr)
age (yr-old)
age (yrs)
Age (yrs)

age [y]
age [year]
age [years]
age in years
age of patient
Age of patient
age of subjects
age(years)
Age(years)
Age(yrs.)
Age, year
age, years
age, yrs
age.year
age_years

Gene Expression Omnibus

The microarray community took the lead in standardizing metadata **reporting guidelines**

- What was the substrate of the experiment?
- What array platform was used?
- What were the experimental conditions?



DNA Microarray

Minimum Information About a Microarray Experiment - MIAME

MIAME describes the **Minimum Information About a Microarray Experiment** that is needed to enable the interpretation of the results of the experiment unambiguously and potentially to reproduce the experiment. [Brazma et al., Nature Genetics]

The six most critical elements contributing towards MIAME are:

1. The raw data for each hybridisation (e.g., CEL or GPR files)
2. The final processed (normalised) data for the set of hybridisations in the experiment (study) (e.g., the gene expression data matrix used to draw the conclusions from the study)
3. The essential sample annotation including experimental factors and their values (e.g., compound and dose in a dose response experiment)
4. The experimental design including sample data relationships (e.g., which raw data file relates to which sample, which hybridisations are technical, which are biological replicates)
5. Sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe oligonucleotide sequences or reference commercial array catalog number)
6. The essential laboratory and data processing protocols (e.g., what normalisation method has been used to obtain the final processed data)

For more details, see MIAME 2.0.

But it didn’t stop with MIAME!

- Minimal Information About T Cell Assays (MIATA)
- Minimal Information Required in the Annotation of biochemical Models (MIRIAM)
- MINImal MEtagemome Sequence analysis Standard (MINIMESS)
- Minimal Information Specification For In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE)

A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

HOW CAN WE HELP?

We guide consumers to discover, select and use these resources with confidence, and producers to make their resource more discoverable, more widely adopted and cited.

Researchers in academia, industry and government

Identify and cite the standards, databases or repositories that exist for your discipline when creating a data management plan, releasing data or submitting a manuscript to a journal...
[read more]

If we want to have FAIR data, we need good metadata. Good metadata need:

- **Ontologies** to provide controlled terms
- **Reporting guidelines**—like MIAME—to provide a standardized structure for the metadata components
- **Technology** to make it easy to author good metadata in the first place

Our approach in CEDAR

- Encode standard, community-endorsed *reporting guidelines* as **templates** that offer fill-in-the-blank authoring opportunities
- Use selections from *ontologies* whenever possible to provide **standardized values** for the template fields

CEDAR

CENTER FOR EXPANDED DATA ANNOTATION AND RETRIEVAL

Some FAIR principles can be addressed only by domain experts themselves using custom metadata templates

F1: ~~(Meta)data are assigned globally unique and persistent identifiers~~

F2: Data are described with rich metadata

F3: ~~Metadata clearly and explicitly include the identifier of the data they describe~~

F4: ~~(Meta)data are registered or indexed in a searchable resource~~

A1: ~~(Meta)data are retrievable by their identifier using a standardised communication protocol~~

A1.1: ~~The protocol is open, free and universally implementable~~

A1.2: ~~The protocol allows for an authentication and authorisation where necessary~~

A2: ~~Metadata should be accessible even when the data is no longer available~~

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

I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data

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

A **metadata template** can ensure compliance with all investigator-controlled FAIR principles, including:

- Making metadata “rich”
- Using metadata vocabularies that follow the FAIR principles
- Meeting domain-relevant community metadata standards

BioSample Human

BioSample Human

* Sample Name	056
* Organism	Homo sapiens
* Tissue	skin of body
* Sex	Male
* Isolate	N/A
* Age	74
* Biomaterial Provider	Life Technologies
Attribute (1)	
Name	disease
Value	dermatitis
Attribute (2)	
Name	description
Value	Cell line was cultured until the 5th passage
Attribute (3)	
Name	treatment
Value	350mg brodalumab

F1: ~~(Meta)data are assigned globally unique and persistent identifiers~~

F2: Data are described with rich metadata

A2: ~~Metadata should be accessible even when the data is no longer available~~

I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for

Don’t even try to measure FAIRness. Make data FAIR from the beginning!

A1: ~~(Meta)data are retrievable by their identifier using a standardised communication protocol~~

A1.1: ~~The protocol is open, free and universally implementable~~

A1.2: ~~The protocol allows for an authentication and authorisation where necessary~~

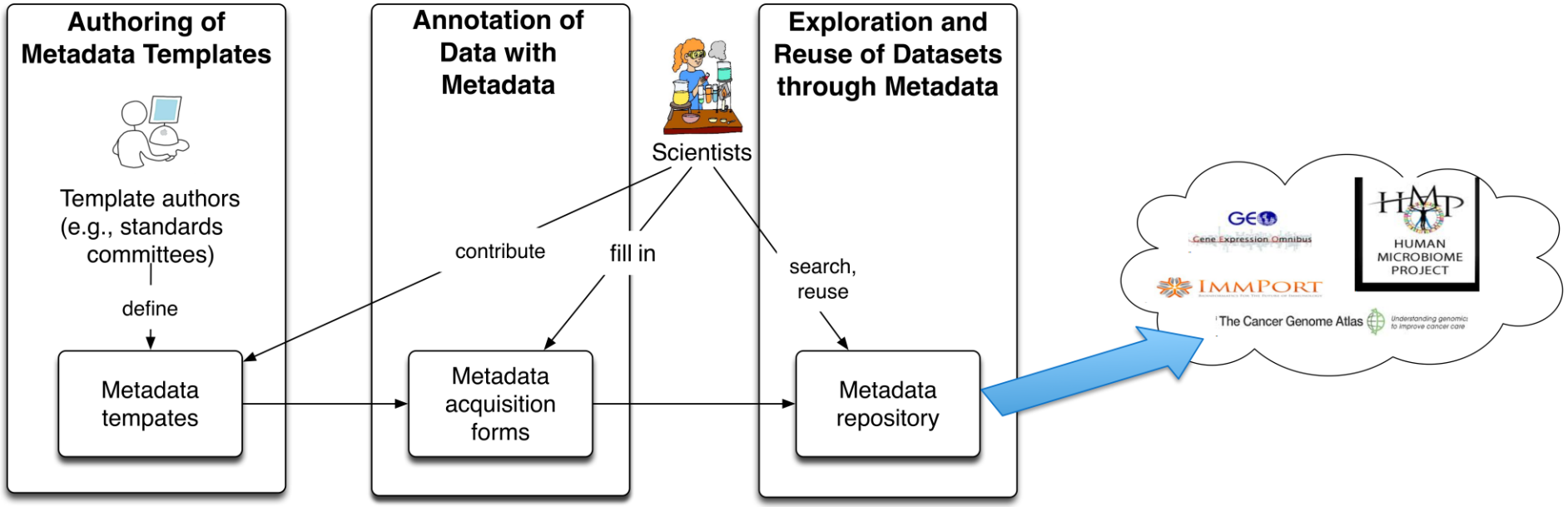
R1: (Meta)data are richly described with a plurality of accurate and relevant attributes

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

The CEDAR Workbench



CEDAR

Search

All / Users / Mark A. Musen

Workspace

Shared with Me

FILTER RESET

TYPE

Title	Created	Modified
GEO	9/5/17 9:48 AM	9/5/17 10:24 AM
BioCADDIE	9/5/17 9:48 AM	9/5/17 10:24 AM
BioSample Human	9/5/17 9:49 AM	9/5/17 11:28 AM
Optional Attribute	9/5/17 10:38 AM	9/5/17 10:38 AM
ImmPort Investigation	9/5/17 9:49 AM	9/5/17 10:21 AM
LINCS Cell Line	9/5/17 9:49 AM	9/5/17 9:49 AM
LINCS Antibody	9/5/17 9:49 AM	9/5/17 9:49 AM
ImmPort Study	9/5/17 9:49 AM	9/5/17 9:49 AM

CEDAR

Search

All / Users / Mark A. Musen

Workspace

Shared with Me

FILTER RESET

TYPE

Title	Created	Modified
GEO	9/5/17 9:48 AM	9/5/17 10:24 AM
BioCADDIE	9/5/17 9:48 AM	9/5/17 10:24 AM
BioSample Human	17 9:49 AM	9/5/17 11:28 AM
Optional Attribute	17 10:38 AM	9/5/17 10:38 AM
ImmPort Investigation	17 9:49 AM	9/5/17 10:21 AM
LINCS Cell Line	17 9:49 AM	9/5/17 9:49 AM
LINCS Antibody	9/5/17 9:49 AM	9/5/17 9:49 AM
ImmPort Study	9/5/17 9:49 AM	9/5/17 9:49 AM

Open
Populate
Share...
Copy to...
Move to...
Rename...
Delete

BioSample Human

BioSample Human

- * Sample Name
- * Organism
- * Tissue
- * Sex
- * Isolate
- * Age
- * Biomaterial Provider
- Attribute
  - Name
  - Value

CANCEL VALIDATE SAVE

BioSample Human

BioSample Human

* Sample Name 056
* Organism Homo sapiens
* Tissue
  - blood (UBERON) (50%)
  - liver (UBERON) (9%)
  - bone marrow (UBERON) 6%)
  - breast (UBERON) (6%)
  - lymph node (UBERON) (6%)
  - lung (UBERON) (6%)
  - colon (UBERON) (6%)
* Sex
* Isolate
* Age
* Biomaterial Provider
* Attribute
  - Name
  - Value

BioSample Human

BioSample Human

* Sample Name	056
* Organism	Homo sapiens
* Tissue	lung
* Sex	Male
* Isolate	N/A
* Age	74
* Biomaterial Provider	Life Technologies

Attribute

Name	disease
Value	

- lung cancer (DOID) (61%)
- chronic obstructive pulmonary disease (DOID) (31%)
- lung squamous cell carcinoma (DOID) (5%)
- idiopathic pulmonary fibrosis (DOID) (4%)
- lung adenocarcinoma (DOID) (4%)
- adenocarcinoma (DOID) (3%)
- carcinoma (DOID) (2%)

BioSample Human

BioSample Human

* Sample Name	056
* Organism	Homo sapiens
* Tissue	brain
* Sex	Male
* Isolate	N/A
* Age	74
* Biomaterial Provider	Life Technologies
Attribute	
Name	disease
Value	

Parkinson's disease (DOID) (39%)
central nervous system lymphoma (DOID) (27%)
autistic disorder (DOID) (22%)
melanoma (DOID) (5%)
Edwards syndrome (DOID) (2%)
schizophrenia (DOID) (1%)

If we want to have FAIR data, we need good metadata. Good metadata need:

- **Ontologies** to provide controlled terms
- **Reporting guidelines**—like MIAME—to provide a uniform structure
- **Technology** to make it easy to author good metadata in the first place

	A	B	C	E	F	G
1	# Use this template for 3' or whole Gene expression studies when summarization probe set data will be provided as **CHP files**.					
2	# Do **NOT** submit CHP files unless they are relevant to your analysis (instead, use the Matrix table option to submit the relevant data, e.g. **Bioconduct**					
3	# Incomplete submissions will be returned. Click the **Metadata Example** tab below to view a completed worksheet					
4	# A complete submission will consist of: (1) a completed metadata worksheet, (2) the CHP files, and (3) the original CEL files.					
5	**# Field names (in blue on this page) should not be edited. Hover over cells containing field names to view field content guidelines or,**					
6	**# CLICK HERE for Field Content Guidelines Web page.**					
7						
8	**SERIES**					
9	# This section describes the overall					
10						
11	**title**					
12	**summary**					
13	**summary**					
14	**overall design**					
15	**contributor**					
16	**contributor**					
17						
18	**SAMPLES**					
19	**# The Sample names in the first column are arbitrary but they must match the column headers of the Matrix table (see next worksheet).**					
20						
21	**Sample name**	**title**	**CHP file**	**source name**	**organism**	**characteristics: tag**
22	**SAMPLE 1**					
23	**SAMPLE 2**					
24	**SAMPLE 3**					
25	**SAMPLE 4**					
26	**SAMPLE 5**					
27	**SAMPLE 6**					
28	**SAMPLE 7**					
29	**SAMPLE 8**					
30	**SAMPLE 9**					
31	**SAMPLE X**					
32						
33						
34	**PROTOCOLS**					
35	# This section includes protocols and fields which are common to all Samples.					
36	# Protocols which are applicable to specific Samples or specific channels should be included in additional columns of the **SAMPLES** section instead.					
37						
38	**growth protocol**					
39	**treatment protocol**					
40	**extract protocol**					
41	**label protocol**					
42	**hyb protocol**					

Unique title (less than 120 characters) that describes the overall study.

"Firstname,Initial,Lastname". Example: "John,H,Smith" or "Jane,Doe".

Unique title that describes the Sample. We suggest that you use the convention: [biomaterial]-[condition(s)]-[replicate number], e.g., Muscle_exercised_60min_rep2.

Replace 'tag' with a biosource characteristic (e.g. "gender", "strain", "tissue", "developmental stage", "tumor stage", etc), and then enter the value for each sample beneath (e.g. "female", "129SV", "brain", "embryo", etc). You may add additional characteristics columns to this template (see 'Metadata Example' spreadsheet).

[Optional] Describe the conditions that were used to grow or maintain organisms or cells prior to extract preparation.

age
Age
AGE
`Age
age (after birth)
age (in years)
age (y)
age (year)
age (years)
Age (years)
Age (Years)
age (yr)
age (yr-old)
age (yrs)
Age (yrs)

age [y]
age [year]
age [years]
age in years
age of patient
Age of patient
age of subjects
age(years)
Age(years)
Age(yrs.)
Age, year
age, years
age, yrs
age.year
age_years

FAIR Evaluator API V.0.3.0 OAS3

The **FAIR Evaluator** - automated testing of Web resources for their compliance

FAIR Metrics Evaluation

HIDE DETAILS

Registry URL	http://smart-api.info/registry?q=ad830426bed193d36838091ef5d14407
SmartAPI ID	ad830426bed193d36838091ef5d14407 Copy
Metadata URL	https://w3id.org/FAIR_Evalu...
Version	V 0.3.0
Contact	Mark Wilkinson

Some FAIR principles can be addressed only by domain experts themselves using custom metadata templates

F1: ~~(Meta)data are assigned globally unique and persistent identifiers~~

F2: Data are described with rich metadata

F3: ~~Metadata clearly and explicitly include the identifier of the data they describe~~

F4: ~~(Meta)data are registered or indexed in a searchable resource~~

A1: ~~(Meta)data are retrievable by their identifier using a standardised communication protocol~~

A1.1: ~~The protocol is open, free and universally implementable~~

A1.2: ~~The protocol allows for an authentication and authorisation where necessary~~

A2: ~~Metadata should be accessible even when the data is no longer available~~

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

I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data

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

Metadata for Machines Workshops

- Are intensive 2–3 day invited, highly participatory sessions
- Historically, have been hosted by GO FAIR Organization
- Lead groups of scientists to consensus regarding essential metadata fields
  - for different areas of science
  - for different kinds of experiments
- Ultimately result in new CEDAR metadata templates

M4M for the Danish e-infrastructure Cooperation

POSTED ON 8 JULY 2020

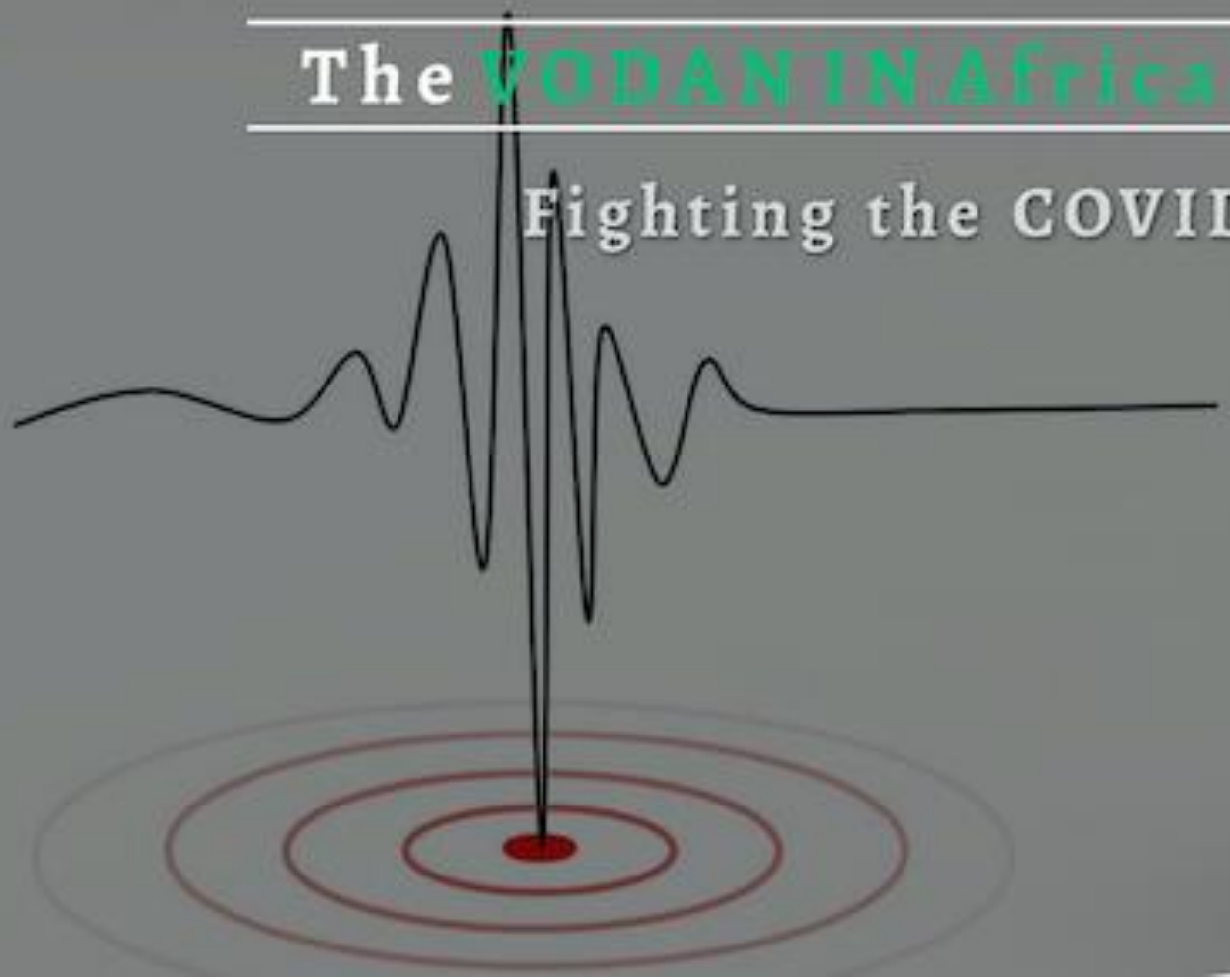
Making it easy for humans to make metadata for machines

On June 26, the **Danish e-Infrastructure Cooperation** (DeiC), in cooperation with the **GO FAIR Foundation**, launched two Metadata for Machine workshops on behalf of two research communities seeking to upgrade the FAIRness of research data. The workshops are conducted via teleconference in a series of five modules to be completed in mid-September.

Participants include: members from the **AnaEE research infrastructure** (M4M.5), the National Energy System Transition Facilities project represented by the **Wind Energy department** at Danmarks Tekniske Universitet (M4M.6), and John Graybeal from Stanford University's **Center for Expanded Data Annotation and Retrieval** (CEDAR). The workshop is co-directed by Erik Schultes from the GO FAIR International Support and Coordination Office, and Diba Terese Markus, **Aalborg University Library**.

The VODAN IN Africa Training the Trainers

Fighting the COVID-19 with FAIR Data



The Netherlands Organization for Health Research and Development

- Has hosted Metadata for Machines workshops to develop metadata templates and controlled terminologies needed for all its funded research related to COVID
- Uses CEDAR to create the metadata templates during the workshops
- Mandates the use of these metadata templates *as a condition of funding*
- Is now expanding the use of M4Ms and standardized metadata into other areas of research that it supports

ZonMw

Online Data Will Never Be FAIR

- Until we standardize metadata structure using common **templates**
- Until we can fill in those templates with **controlled terms** whenever possible
- Until we create **technology** that will make it easy for investigators to annotate their datasets in standardized, searchable ways
- Until we recognize that we can't solve the problem of data FAIRness by trying to evaluate FAIRness when it's already too late

CEDAR
CENTER FOR EXPANDED DATA ANNOTATION AND RETRIEVAL