Operationalizing FAIR

at the Common Fund (Data Ecosystem)

Dr. Amanda Charbonneau

Dr. Amanda Charbonneau

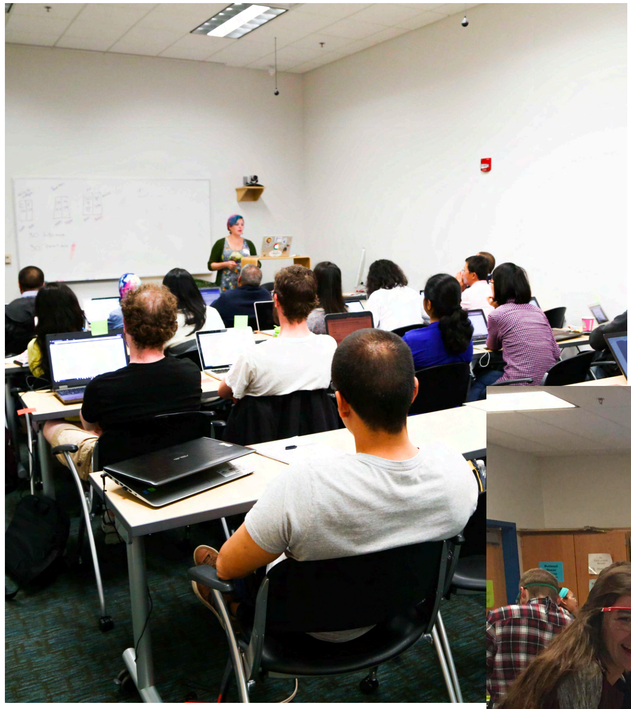
A brief history

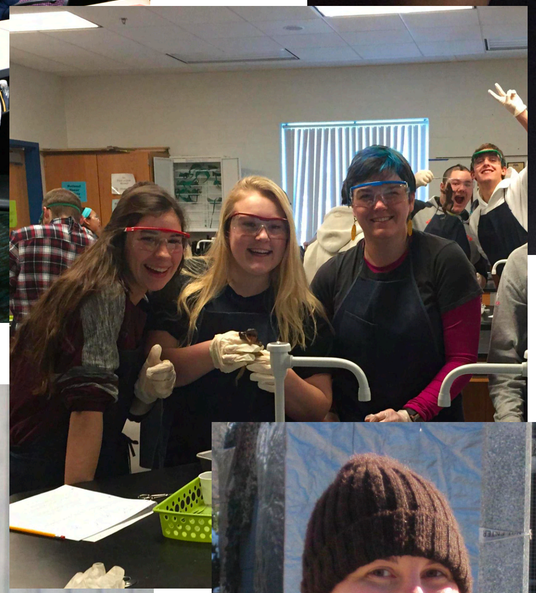
- BA in Criminal Justice & BA in Biochemistry
- Landscape Entomology
- Experimental tank armor and plasma chemistry
- Large scale mass spectrometry and genetic screens
- Phd in (population) Genetics

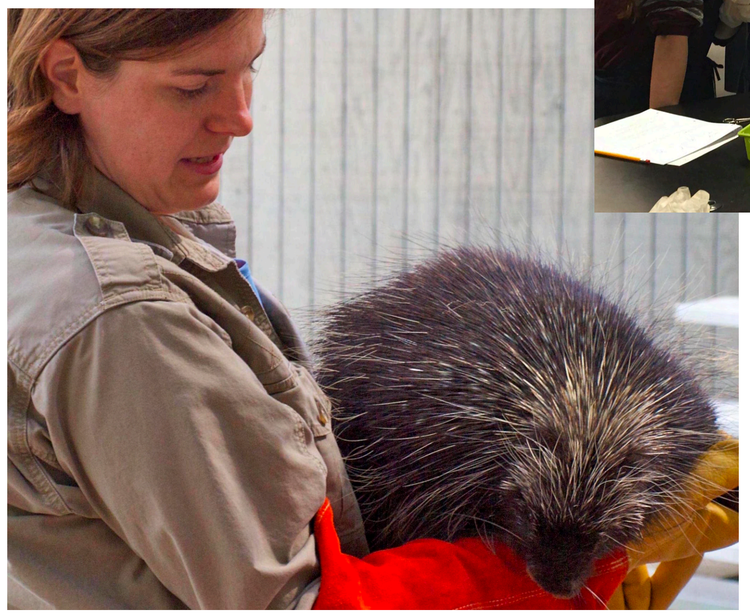
Dr. Amanda Charbonneau

A brief history

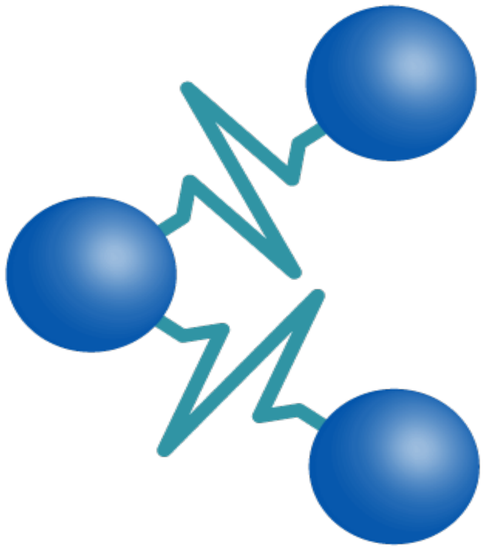
- BA in Criminal Justice & BA in Biochemistry
- Landscape Entomology
- Experimental tank armor and plasma chemistry
- Large scale mass spectrometry and genetic screens
- Phd in (population) Genetics
- Lots and lots of teaching







- Enhance the ability to ask scientific questions across data sets
- Enable the uptake, reuse, and addition of Common Fund data and tools
- Support the storage, sharing, and sustainability of Common Fund data sets
- Provide training that maximizes scientists' ability to upload data and use Common Fund data and other resources



COMMON FUND DATA ECOSYSTEM

NIH is made up of 27 Institutes and Centers, each with a specific research agenda, often focusing on particular diseases or body systems.

National Cancer Institute

National Institute of Allergy and Infectious Diseases

National Institute on Aging

National Human Genome Research Institute

National Library of Medicine

National Eye Institute
National Institutes of Health

National Heart Lung and Blood Institute

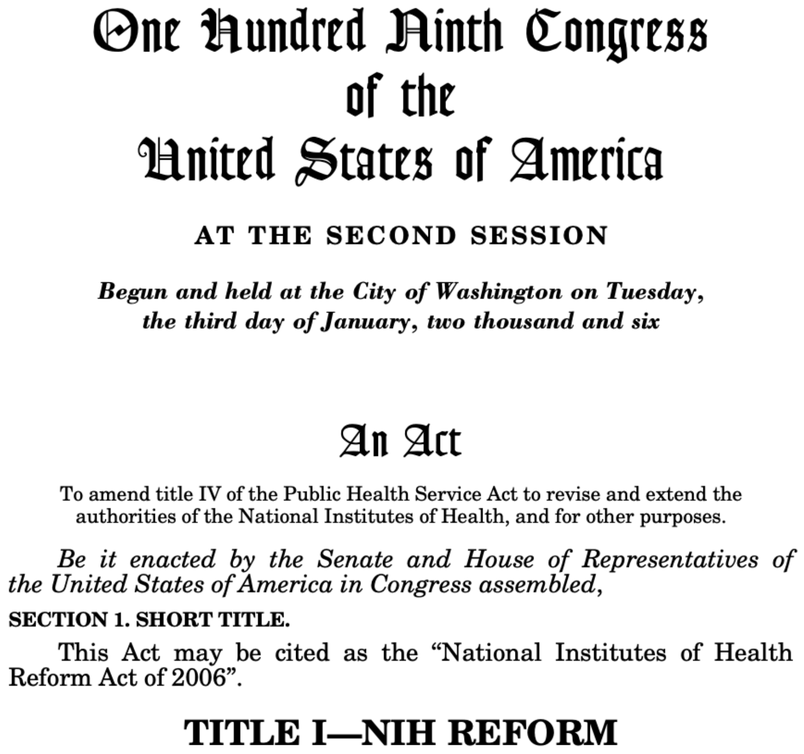
Etc.

The Common Fund

Common Fund programs must meet five overarching criteria. Programs must be:

- **Transformative**: Must have high potential to dramatically affect biomedical and/or behavioral research over the next decade
- **Catalytic**: Must achieve a defined set of high impact goals within a defined period of time
- **Synergistic**: Outcomes must synergistically promote and advance individual missions of NIH ICs to benefit health
- **Cross-cutting**: Program areas must cut across missions of multiple NIH ICs, be relevant to multiple diseases or conditions, and be sufficiently complex to require a coordinated, trans-NIH approach
- **Unique**: Must be something no other entity is likely or able to do

Common Fund programs are intended to change paradigms, develop innovative tools and technologies, and/or provide fundamental foundations for research that can be used by the broad biomedical research community.



One Hundred Ninth Congress
of the
United States of America

AT THE SECOND SESSION

Begun and held at the City of Washington on Tuesday, the third day of January, two thousand and six

An Act

To amend title IV of the Public Health Service Act to revise and extend the authorities of the National Institutes of Health, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

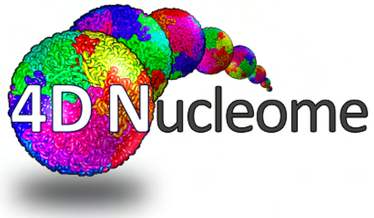
SECTION 1. SHORT TITLE.

This Act may be cited as the "National Institutes of Health Reform Act of 2006".

TITLE I—NIH REFORM

Common Fund was later given a separate appropriation line that is 1.7% of the total NIH budget

Some Common Fund Programs

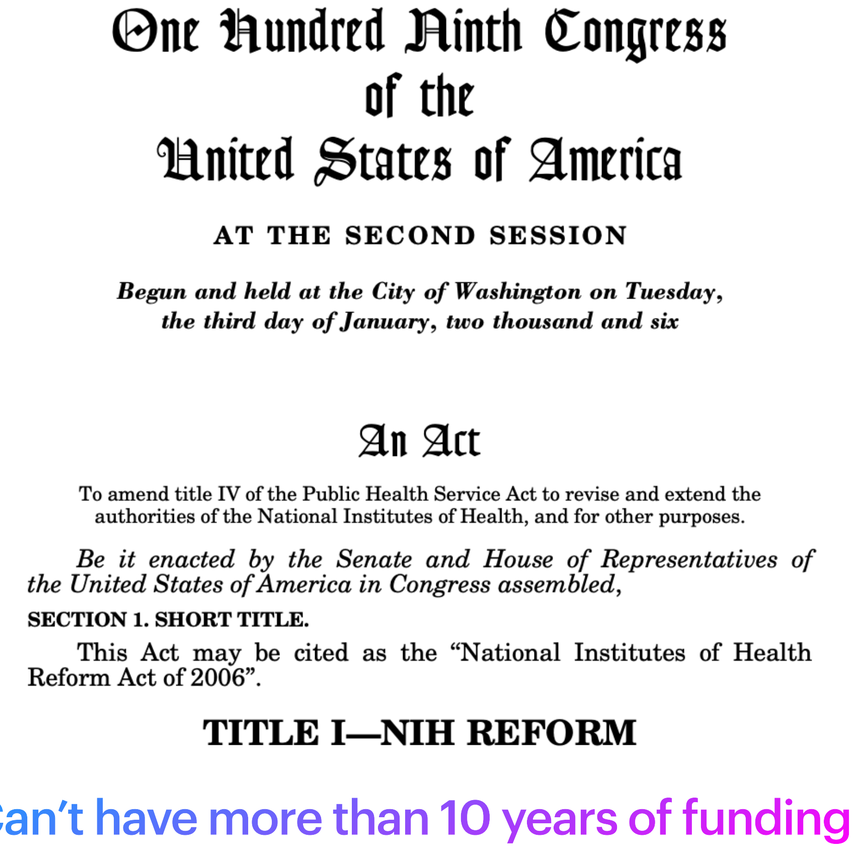


The Common Fund

Common Fund programs must meet five overarching criteria. Programs must be:

- **Transformative**: Must have high potential to dramatically affect biomedical and/or behavioral research over the next decade
- **Catalytic**: Must achieve a defined set of high impact goals within a defined period of time
- **Synergistic**: Outcomes must synergistically promote and advance individual missions of NIH ICs to benefit health
- **Cross-cutting**: Program areas must cut across missions of multiple NIH ICs, be relevant to multiple diseases or conditions, and be sufficiently complex to require a coordinated, trans-NIH approach
- **Unique**: Must be something no other entity is likely or able to do

Common Fund programs are intended to change paradigms, develop innovative tools and technologies, and/or provide fundamental foundations for research that can be used by the broad biomedical research community.



One Hundred Ninth Congress
of the
United States of America

AT THE SECOND SESSION

Begun and held at the City of Washington on Tuesday, the third day of January, two thousand and six

An Act

To amend title IV of the Public Health Service Act to revise and extend the authorities of the National Institutes of Health, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

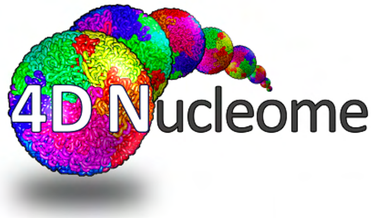
This Act may be cited as the "National Institutes of Health Reform Act of 2006".

TITLE I—NIH REFORM

Can't have more than 10 years of funding

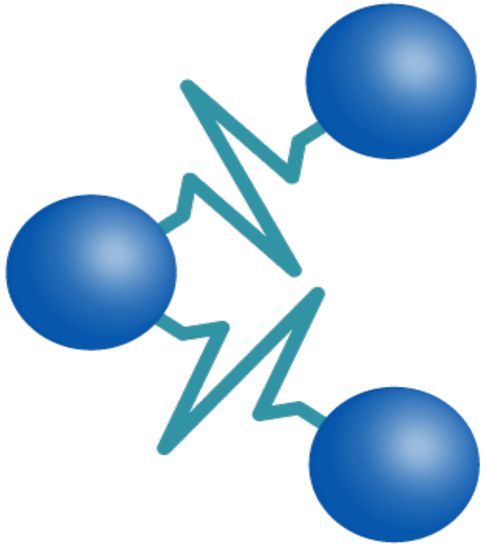
Common Fund was later given a separate appropriation line that is 1.7% of the total NIH budget

Some Common Fund Programs



Current CFDE Goals

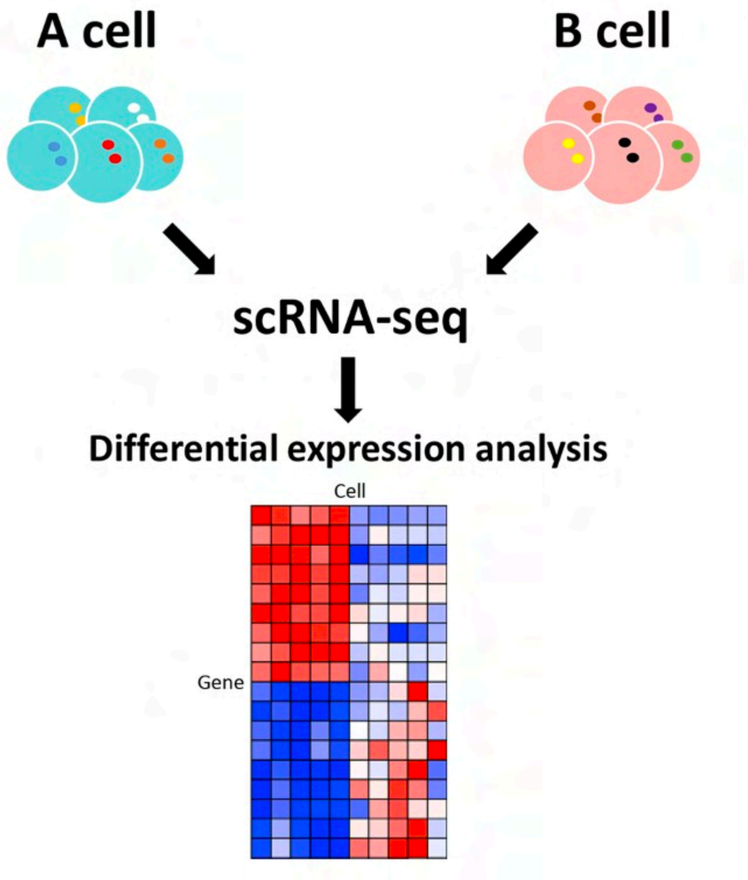
- Increase data reuse and interoperability for existing and future data sets
- Move data analysis into the cloud (for many reasons!)
- Track data usage and help prioritize sustainability efforts
- Improve sunset and sunrise for CF projects (10 year funding limit)



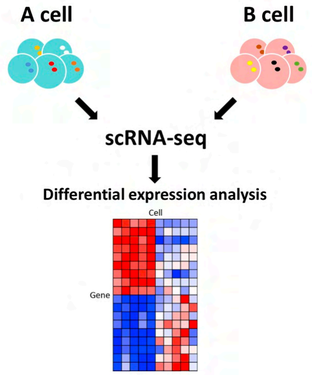
COMMON
FUND
DATA
ECOSYSTEM

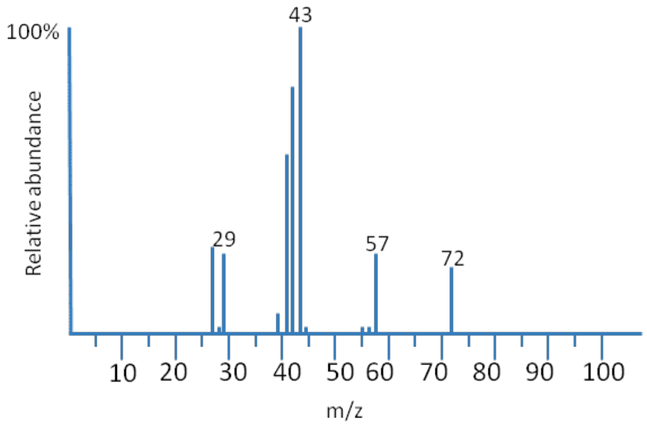
Two ways to combine data

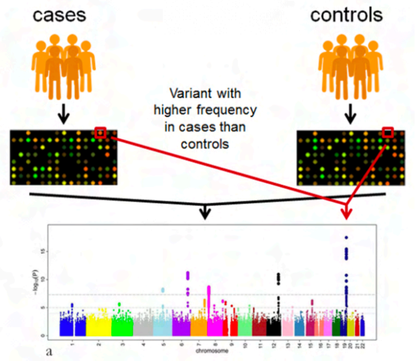
Munging



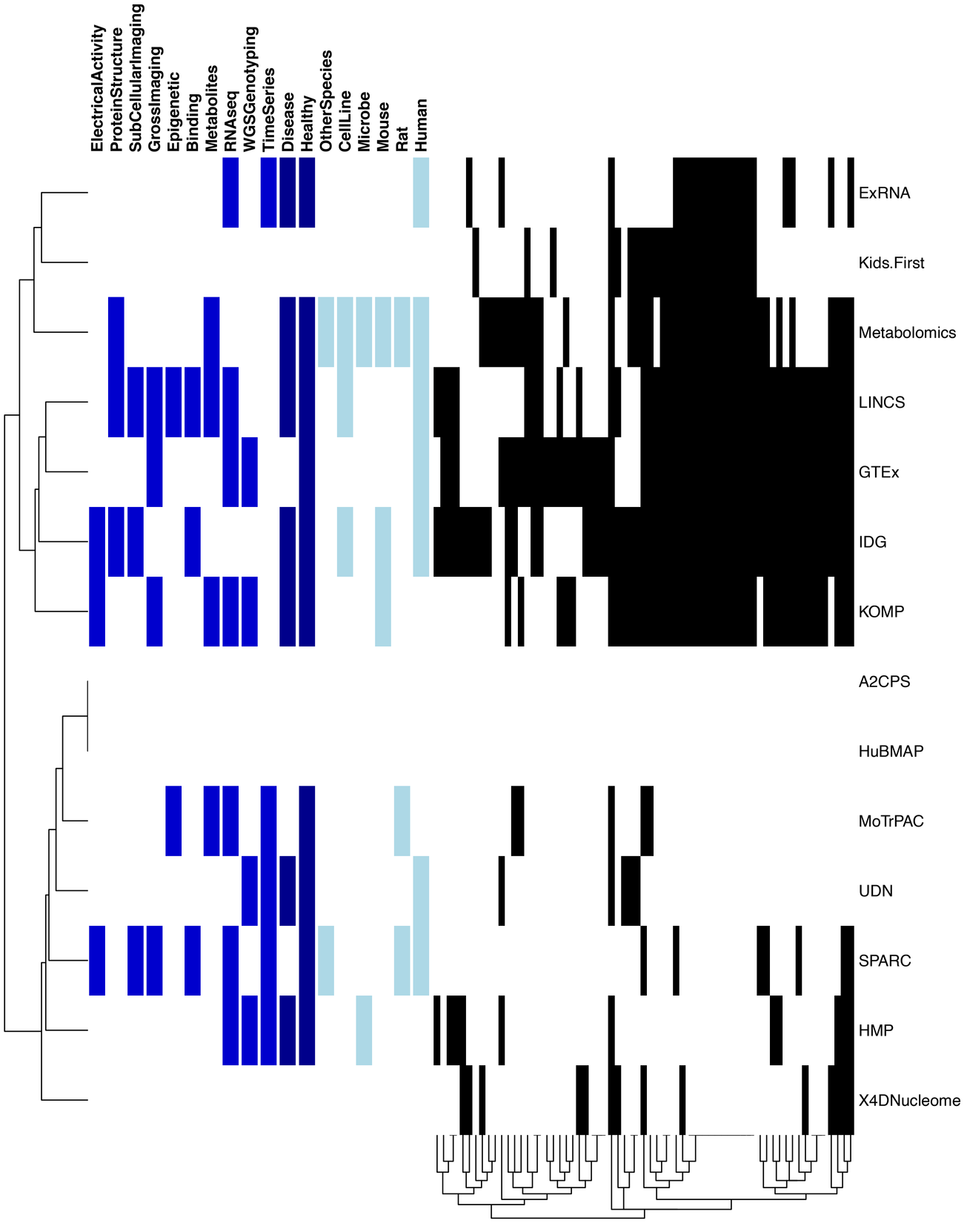
Layering



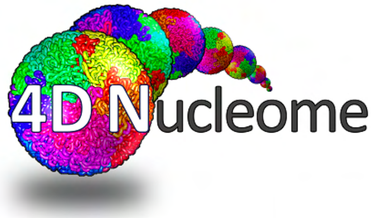




https://link.springer.com/article/10.1007/s40484-016-0089-7

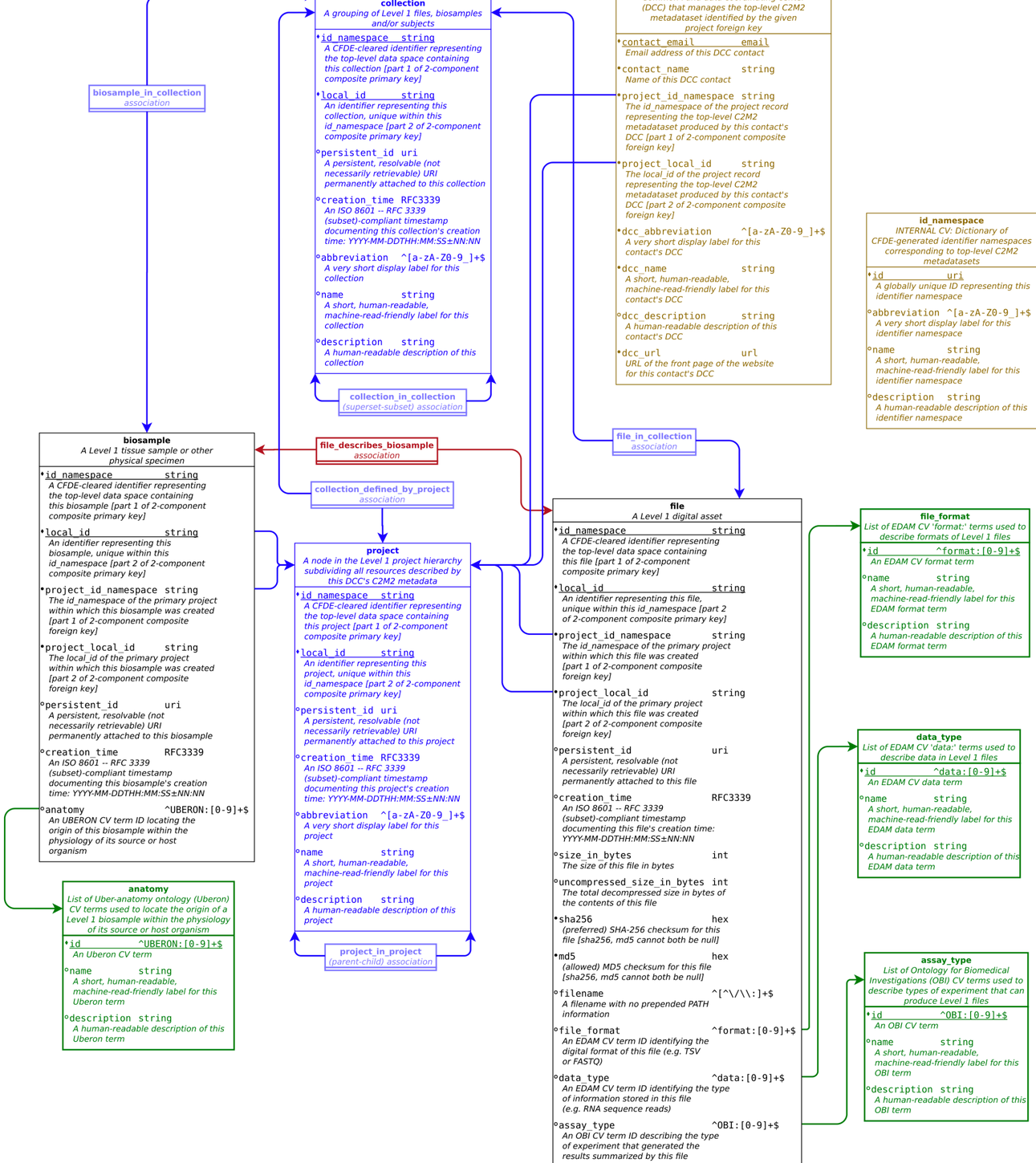

ElectricalActivity
ProteinStructure
SubCellularImaging
GrossImaging
Epigenetic
Binding
Metabolites
RNAseq
WGSGenotyping
TimeSeries
Disease
Healthy
OtherSpecies
CellLine
Microbe
Mouse
Rat
Human
ExRNA
Kids.First
Metabolomics
LINCS
GTEx
IDG
KOMP
A2CPS
HuBMAP
MoTrPAC
UDN
SPARC
HMP
X4DNucleome

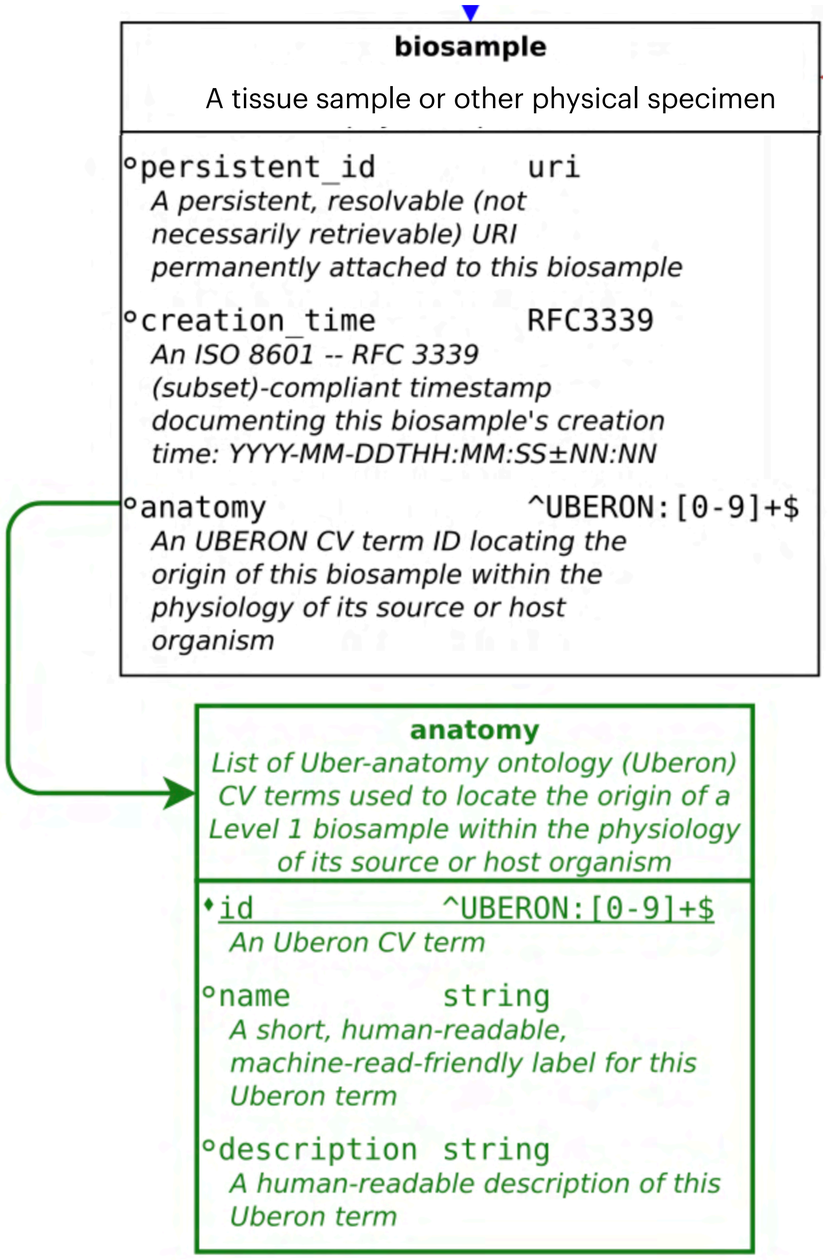
Some Common Fund Programs



How to FAIR in four easy steps

- Interview some of the programs about their data
- Build a model that all of them can use to describe their data assets
- Build a portal to host that modeled metadata
- Profit





Box 2: The FAIR Guiding Principles

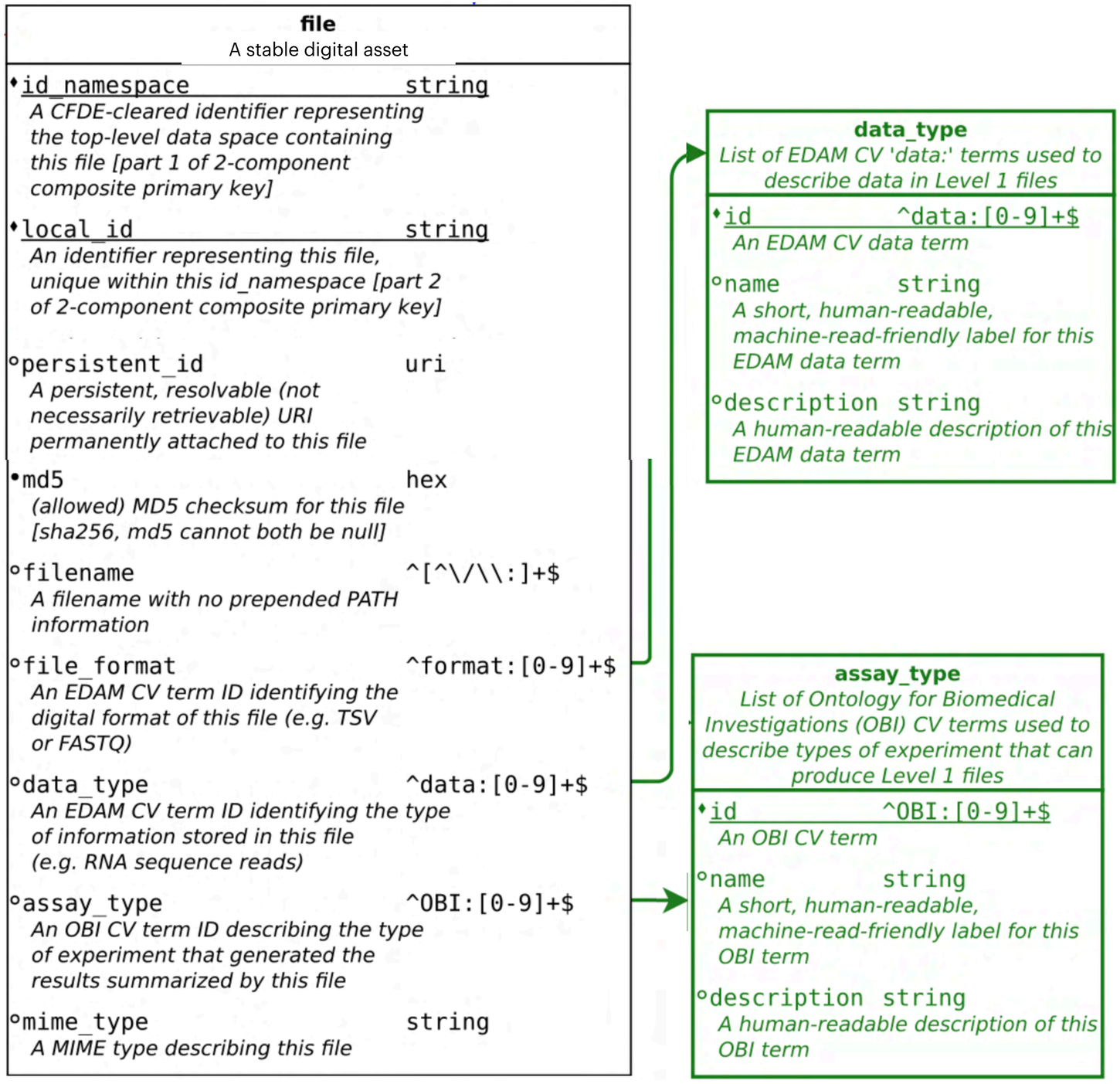
To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

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

https://doi.org/10.1038/sdata.2016.18



Box 2: 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

https://doi.org/10.1038/sdata.2016.18

Technology is important, but technology alone doesn’t make data FAIR.

People make data FAIR.

CFDE Cross-Pollination Meetings

Cross-Pollination Meetings are Tuesdays at 11AM PST / 2PM EST

Join our mailing list to receive updates, invitations, and reminders : https://groups.io/g/CrossPollinationEvents

Join our Slack to chat during and after the meetings: https://join.slack.com/t/cfdeworkspace/shared_invite/zt-hupdgmhw-ZzSUc8Oau3DTpfBr4PccKg

Need Help? Visit our Resources for Attendees for details on how to participate in these meetings.

Working Group meetings take place as noted in the schedule by agreement within each group.

Date	Agenda Link	Event description
Working Group Meetings		
4/7/2021	**Ontology Working Group**	Bi-Weekly on Wednesdays - 8am PST / 11am EST - Ontology WG Agenda
4/7/2021	**Gene Working Group**	Scheduled monthly on first Wednesday - Gene WG Agenda
4/20/2021	**Anatomy Working Group**	Scheduled Tuesdays - 12pm PST / 3pm EST - cAWG Agenda
Upcoming Cross-Pollination Meetings		
4/6/2021	**HuBMAP & ExRNA**	**ExRNA**: Integration of exRNA Atlas Data and Genetic Variant Information into the CFDE portal **HuBMAP**: Anatomical Structure, Cell Types, and Biomarkers Tables: Construction and Usage
5/4/2021	**IDG & SPARC**	**IDG:** Mapping and Matching DataSets: TCRD Thus Far **SPARC:** Interlex and Term Mapping
6/1/2021	**CFDE & SPARC**	**CFDE-CC:** CurIndex: More than a disease simlilarity network **SPARC:** Connectivity Knowledge Base

Thanks for listening!

COMMON FUND DATA ECOSYSTEM

CFDE - CC
Owen White, PI

UMB
Owen White, PI

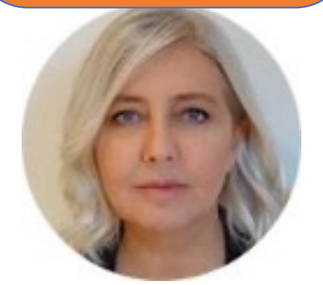
USC/ISI
Carl Kesselman, PI

UCHI
Ian Foster, PI

RTI
Becky Boyles, PI

UMB
Bob Carter, PM

Oxford
Susanna Sansone, PI



ICAHN/Mt. Sinai
Avi Ma'ayan, PI

UCD
Titus Brown, PI

UCD
Amanda Charbonneau, PM