

Samplesheets & Assay plugins

A crash course for working with metadata

Thomas Sell

Core Unit for Bioinformatics, BIH@Charité

21 October 2024

What is FAIR data?

- Findable
 - Accessible
 - Interoperable
 - Reusable
- Gold standard of science data management.
 - Ensures quality and reusability of data.
 - Increasingly required by funding bodies.

SCIENTIFIC DATA 

Amended: Addendum

OPEN

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

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

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

DOI: 10.1038/sdata.2016.18

What is metadata?



Patient ID Mutation Source material Sample ID
Experimental question Characteristics Who Perturbation
Treatment
Genetic background Machine Multiplex
Procedure
When

ISA-tabs

- **I**nvestigation → Project
- **S**tudy → Sample characteristics
- **A**ssay → Measurement technique



<https://isa-specs.readthedocs.io/>

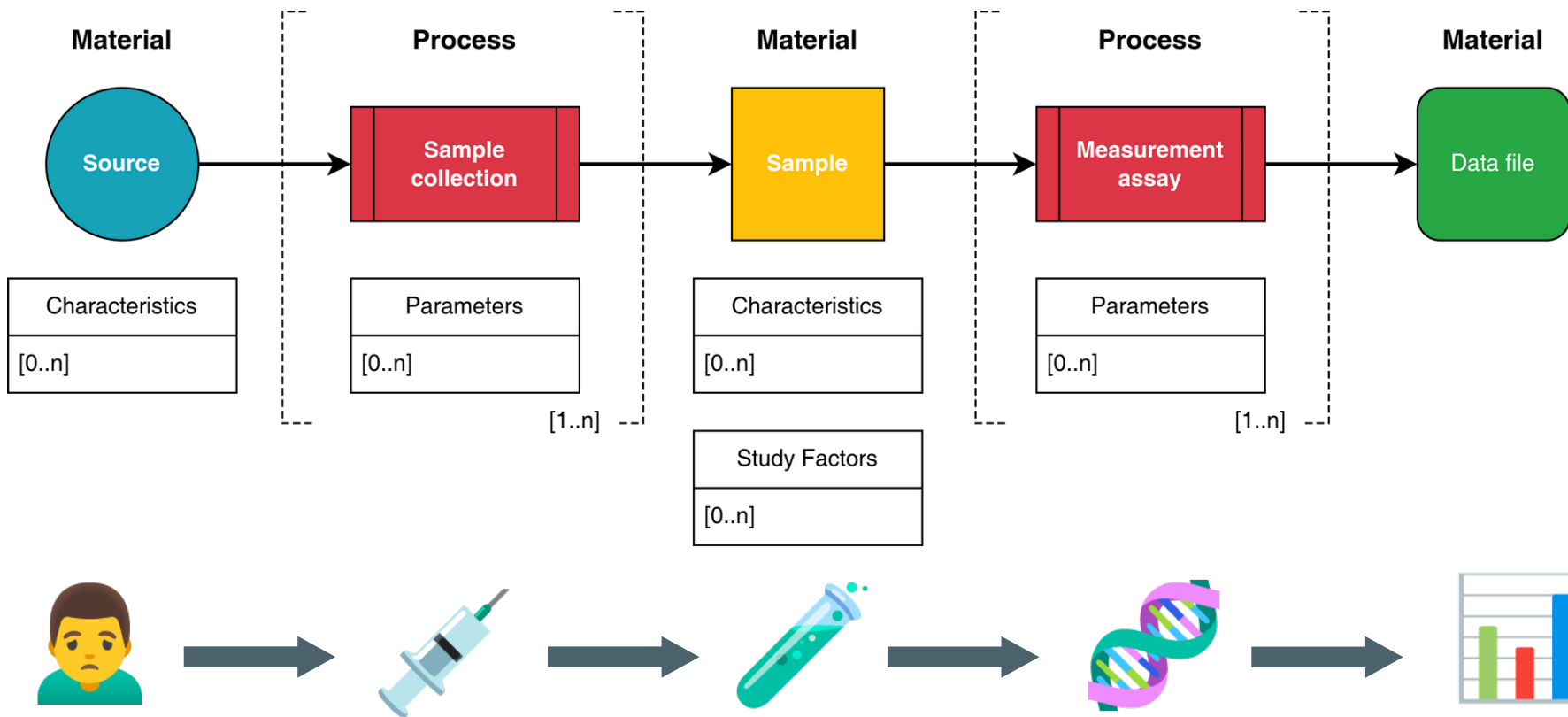


Open Science for Life in Space

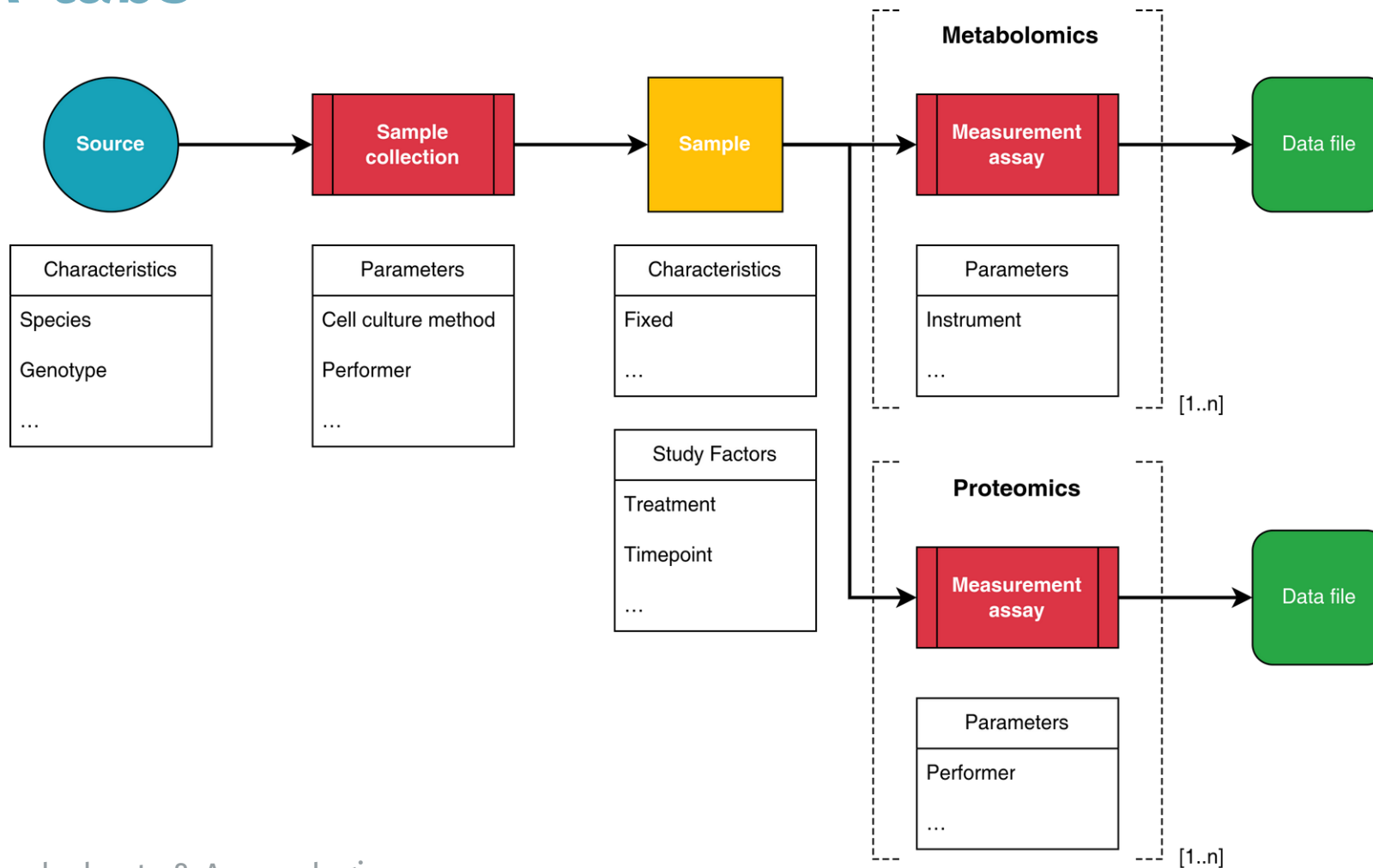


MetaboLights

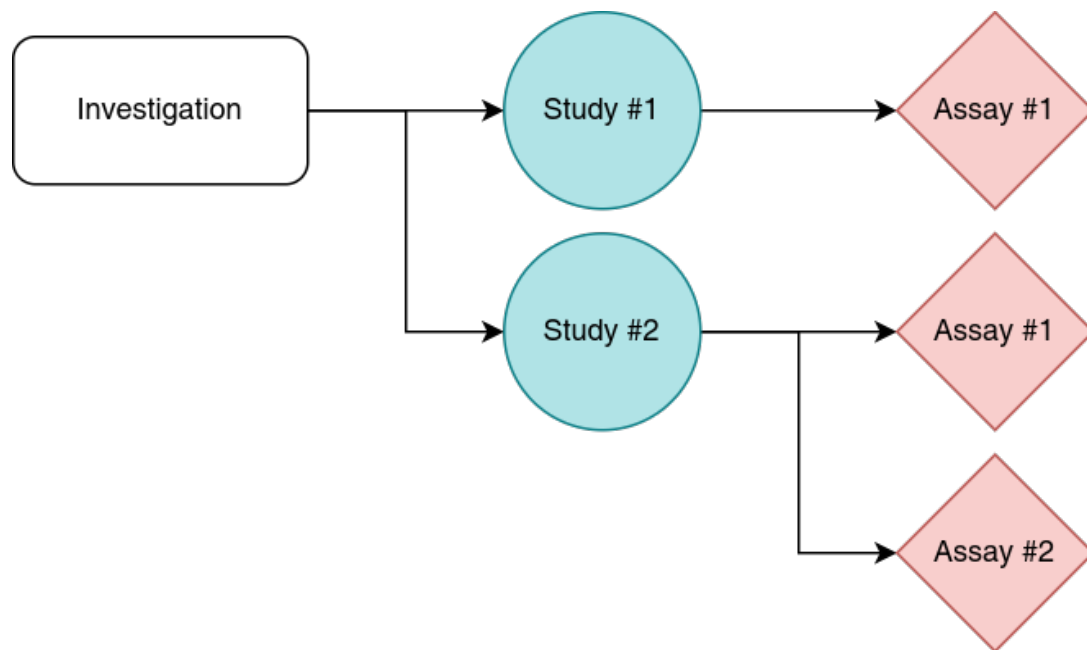
ISA-tabs



ISA-tabs

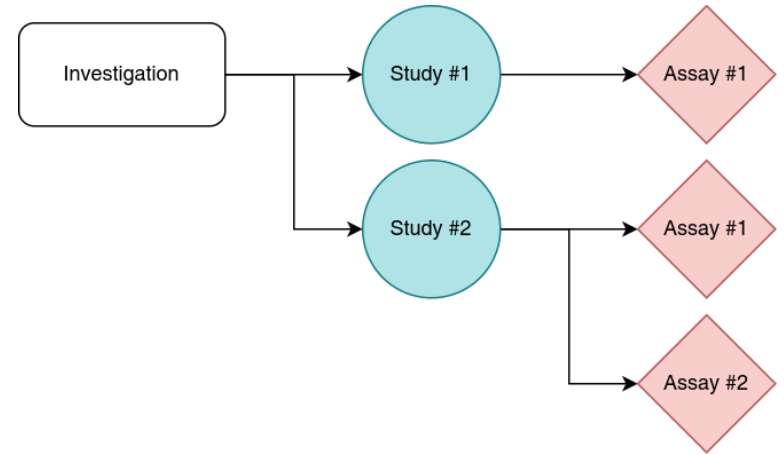


ISA-tabs



File structure

- i_Investigation.txt
 - s_<study1>.txt
 - a_<study1>_<assay1>.txt
 - s_<study2>.txt
 - a_<study2>_<assay1>.txt
 - a_<study2>_<assay2>.txt



Summary 1

- ISA-tabs give metadata structure
- Samplesheets should mirror real-life workflows
 - <https://isa-specs.readthedocs.io/>

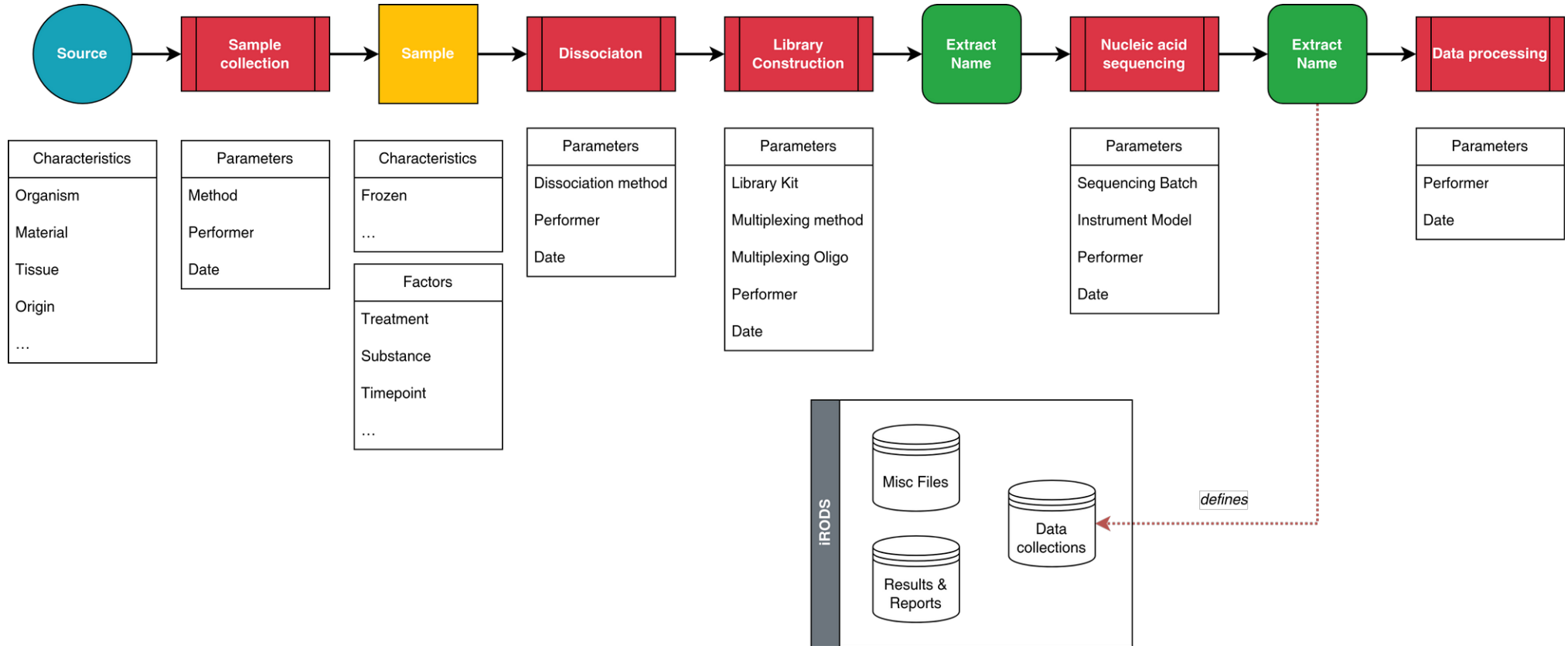
SODAR assay plugins

- Link between samplesheets & iRODS
- Creation of pre-defined data collections
- Make accessing data easier
- Automatically selected via “measurement type”

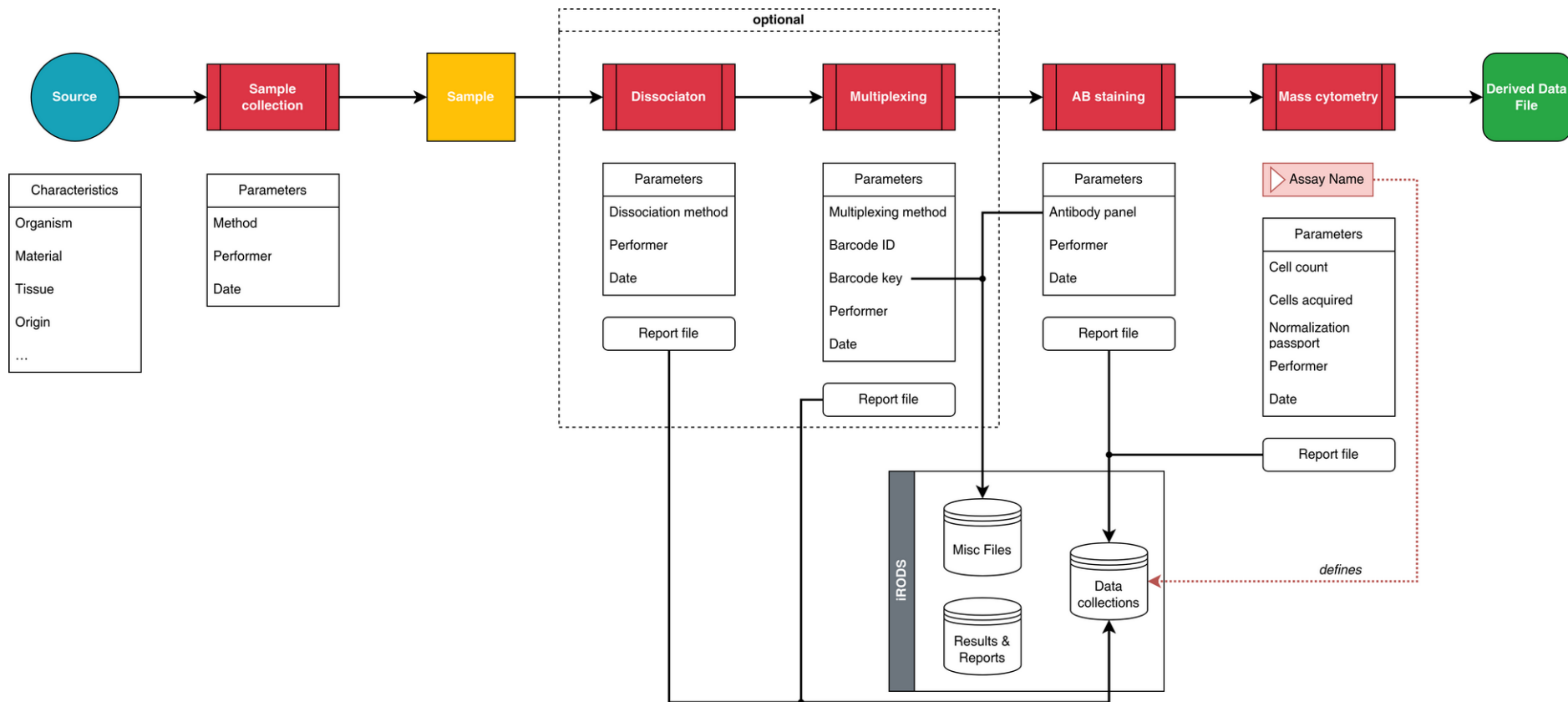
SODAR assay plugins

- **DNA Sequencing**
 - Last material name defines data collections
 - Library Name / Extract Name
- **Mass Cytometry**
 - Assay Name parameter of “mass cytometry” process defines data collections
 - Creates simple links for additional metadata files & data

Example: DNA Sequencing



Example: Mass cytometry



Example: Mass cytometry

Assay: Mass Cytometry



Assay Shortcuts

Results and Reports



Misc Files



Assay Data

Filter

Row	Source	Process						Proc	iRODS
#	Name	Protocol	Antibody Panel	Report File	Performer	Perform Date	Labfolder Entry	Prot	Links
1	NCO	Antibody staining	HYDRA_panel1.tsv	protocol_20191014.pdf	Thomas Sell	2019-10-15	Media conditions – CyTOF measur...	Mass	
2	NCO	Antibody staining	HYDRA_panel1.tsv	protocol_20191014.pdf	Thomas Sell	2019-10-15	Media conditions – CyTOF measur...	Mass	
3	NCO	Antibody staining	HYDRA_panel1.tsv	protocol_20191014.pdf	Thomas Sell	2019-10-15	Media conditions – CyTOF measur...	Mass	
4	NCO	Antibody staining	HYDRA_panel1.tsv	protocol_20191014.pdf	Thomas Sell	2019-10-15	Media conditions – CyTOF measur...	Mass	
5	NCO	Antibody staining	HYDRA_panel2.tsv	protocol_20200505.pdf	Thomas Sell	2020-05-05	medium conditions – replicate	Mass	
6	NCO	Antibody staining	HYDRA_panel2.tsv	protocol_20200505.pdf	Thomas Sell	2020-05-05	medium conditions – replicate	Mass	
7	NCO	Antibody staining	HYDRA_panel2.tsv	protocol_20200505.pdf	Thomas Sell	2020-05-05	medium conditions – replicate	Mass	
8	A	Antibody staining	HYDRA_panel1.tsv	protocol_20191014.pdf	Thomas Sell	2019-10-15	Media conditions – CyTOF measur...	Mass	
9	A	Antibody staining	HYDRA_panel1.tsv	protocol_20191014.pdf	Thomas Sell	2019-10-15	Media conditions – CyTOF measur...	Mass	
10	A	Antibody staining	HYDRA_panel1.tsv	protocol_20191014.pdf	Thomas Sell	2019-10-15	Media conditions – CyTOF measur...	Mass	
11	A	Antibody staining	HYDRA_panel1.tsv	protocol_20191014.pdf	Thomas Sell	2019-10-15	Media conditions – CyTOF measur...	Mass	
12	A	Antibody staining	HYDRA_panel2.tsv	protocol_20200505.pdf	Thomas Sell	2020-05-05	medium conditions – replicate	Mass	
13	A	Antibody staining	HYDRA_panel2.tsv	protocol_20200505.pdf	Thomas Sell	2020-05-05	medium conditions – replicate	Mass	
14	A	Antibody staining	HYDRA_panel2.tsv	protocol_20200505.pdf	Thomas Sell	2020-05-05	medium conditions – replicate	Mass	

Example: Mass cytometry

Assay: Mass Cytometry



Assay Shortcuts

Results and Reports



Misc Files



Assay Data

Filter

Row	Source	Raw Data File	Process				Derived Data File	iRODS
#	Name	Name	Protocol	Method	Performer	Perform Date	Name	Links
1	NCO	20191015.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2019-10-15	20191015_HYDRA_media.fcs	
2	NCO	20191015.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2019-10-15	20191015_HYDRA_media.fcs	
3	NCO	20191015.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2019-10-15	20191015_HYDRA_media.fcs	
4	NCO	20191015.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2019-10-15	20191015_HYDRA_media.fcs	
5	NCO	20200506.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2020-05-06	20200506_HYDRA_media.fcs	
6	NCO	20200506.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2020-05-06	20200506_HYDRA_media.fcs	
7	NCO	20200506.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2020-05-06	20200506_HYDRA_media.fcs	
8	A	20191015.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2019-10-15	20191015_HYDRA_media.fcs	
9	A	20191015.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2019-10-15	20191015_HYDRA_media.fcs	
10	A	20191015.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2019-10-15	20191015_HYDRA_media.fcs	
11	A	20191015.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2019-10-15	20191015_HYDRA_media.fcs	
12	A	20200506.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2020-05-06	20200506_HYDRA_media.fcs	
13	A	20200506.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2020-05-06	20200506_HYDRA_media.fcs	
14	A	20200506.7z	Data normalization	EQ four element calibration beads	BIH Cytometry Core	2020-05-06	20200506_HYDRA_media.fcs	

Summary 2

- ISA-tabs give metadata structure
- Samplesheets should mirror real-life workflows
 - <https://isa-specs.readthedocs.io/>
- Assay plugins link metadata and main data
 - SODAR Manual -> Advanced Metadata Topics
- One more thing ...

Generic Assay plugin

- User-defined data linking
- Via investigation file comments
- Measurement type agnostic
- Can be used with arbitrary assays/data

Generic Assay plugin

