Initiative for a new Research Data Alliance (RDA) working group:

FAIRification of Genomic Tracks WG

- Enabling data-driven life science with sub-dataset granularity through uniform discovery and access of sequence annotation data



Building on: FAIRtracks + Omnipy

Sveinung Gundersen, ELIXIR Norway / Centre for Bioinformatics, University of Oslo







A selection of major consortia producing genomic data over the last 20 years



Public data costing billions to produce!

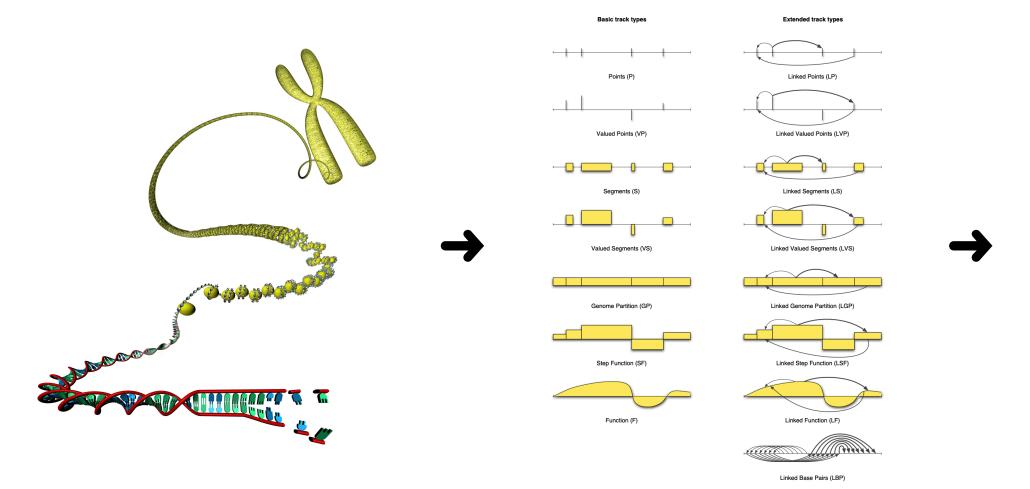
- split into separate data portals with largely incompatible metadata models

... plus 1000s of smaller scale datasets, many with little metadata



Genomic Tracks

– unified model for data analysis



The mental model of data elements positioned along a reference genome is essential also for non-visual analysis

"Tracks" in this broader sense covers many common file formats, such as BED/BigBED, WIG/BigWIG, SAM/BAM, VCF, GFF, ...





Genome browser meme by Jedidiah Carlson

Variation of data content in genomic tracks

Experiment Matrix

Clear all selections \boldsymbol{x}																	
Q Filter the experiments inc	luded in																
the matrix:		Showin	g 16574 results														
Enter search term(s)		List	-	ualize												(;	}
Assay •	_	ASSA	ightarrow														
		BIOSAMPLE															
Assay type	<u> </u>	SAI															
DNA binding	8917	BIO			-				seq								
Transcription	3568				-sec		-		NA-5		σ					σ	
DNA accessibility Single cell	2539 749	\downarrow		ba	hP	σ	-sec	-Se	sR		I-Se		p.	ray		-se	
DNA methylation	566			P-S	e C	-Se	NA	ald.	nlq	seq	RN/	-Se	C-S	e al		RN/	
Assay title				TF ChIP-seq	Histone ChIP-seq	DNase-seq	total RNA-seq	Mint-ChIP-seq	polyA plus RNA-seq	ATAC-seq	microRNA-seq	scRNA-seq	snATAC-seq	DNAme array	eCLIP	small RNA-seq	DICIDIC
-	^							Σ									_
Search		\sim	tissue	443	2182	933	415		399	225	305	385	275	121	2	67	1
TF ChIP-seq	4321		middle frontal area 46	59	188	56	120				43						
Histone ChIP-seq	3701		heart left ventricle	16	84	10	5		2	15	5	54	65	3			
DNase-seq	1700		heart	7	83	39	19		16	7	25	40	8	-		1	
total RNA-seq Mint-ChIP-seq	896 895		adrenal gland	8	39	32	23		11	8	23	59	15	5	2	2	
polyA plus RNA-seq	753		liver	42	91	15	3		20	8	7			1		1	
ATAC-seq	460																
microRNA-seq	412	\sim	cell line	2753	773	180	136	54	155	124	39	6	32	87	238	110	1
scRNA-seq	399		K562	735	19	4	17		15	4	2		3	3	135	7	
snATAC-seq	347		HepG2	764	15	2	6		11	2	2		3	3	103	3	
			GM12878	190	15	1	5		13	2	2	2	3	3		6	
Target category	~	•	MCF-7 HEK293	150 203	18 6	4	1		4	1	2		3	2		7	
Target of assay	~			203	0									2			
Search	-		primary cell	74	506	547	223	715	136	71	52	3	29	38		24	
H3K4me3	700		naive thymus-derived CD4-		16	16	5	74	8	2	2		2				
H3K27ac	790 700		positive, alpha-beta T cell								_		-				
H3K27me3	657		T-cell		11	60	5	18	1	6							
H3K4me1	636		naive thymus-derived CD8- positive, alpha-beta T cell			13	2	66	6	3	1		1				
H3K36me3	626		CD14-positive monocyte	2	21	8		41	8				2			1	
H3K9me3	589		macrophage	2	4	0	78	41	1				2			1	
CTCF	528				7		10		1								
POLR2A H3K9ac	187 186			997			68		41								
H3K4me2	132	W	hole organisms whole organism	997			64		33								
10111112	101		carcass	551			4		8								
Hide control experiments	~		carcass				7		0								
	16574		in vitro differentiated cells	47	240	38	53	78	22	40	16	5	11	5		11	2
🔿 no	19926		dendritic cell		11		25	10	~~	30	10	3		Ŭ			~
			motor neuron		11		9			10							1
Biosample •		-	endodermal cell	1	8	2	3	12	2	10	2		1				
			neuronal stem cell		27	1			2							1	
Library		-	mesenchymal stem cell		27	1			2							1	
Analysis		Ce	ll-free sample														
Provenance		-	cell-free sample	_													
		or	ganoid	7		2		48									
Quality •		-	brain	4		2		30									
Other filters			nephron	3				18									
Other filters •															I		
		te	chnical sample				1										
			technical sample				1										

Gene regions, repeating elements, conserved regions

- Chromatin accessibility (e.g., DNase I Hypersensitivity)
- **Binding of Transcription Factors to DNA**
- Histone modifications along DNA
- Gene expression, Gene fusion, Transcription start sites (TSS)
- Cis-regulatory elements (promoters, enhancers...)
- **DNA** methylation
- 3D genome structure
- Variation: GWAS SNPs, e-QTLs, SNVs, CNVs, ...
- Virus insertion sites
- = Any "omics" data file with reference genome positions!

What is special about track files?

- A. Routinely generated through standard pipelines/tools and typically **stored** alongside raw sequence files in larger datasets
- B. Represent *summaries of the raw data*:

- A genomic track relates to the raw data much like an abstract describes a scientific publication!

These two properties together makes track files uniquely suited for:

Data-driven discovery of genomic datasets and their relationships – with sub-dataset granularity (at the experiment level)!

FAIR aspects of Genomic Tracks – Potential for Improvements

Findable

Accessible

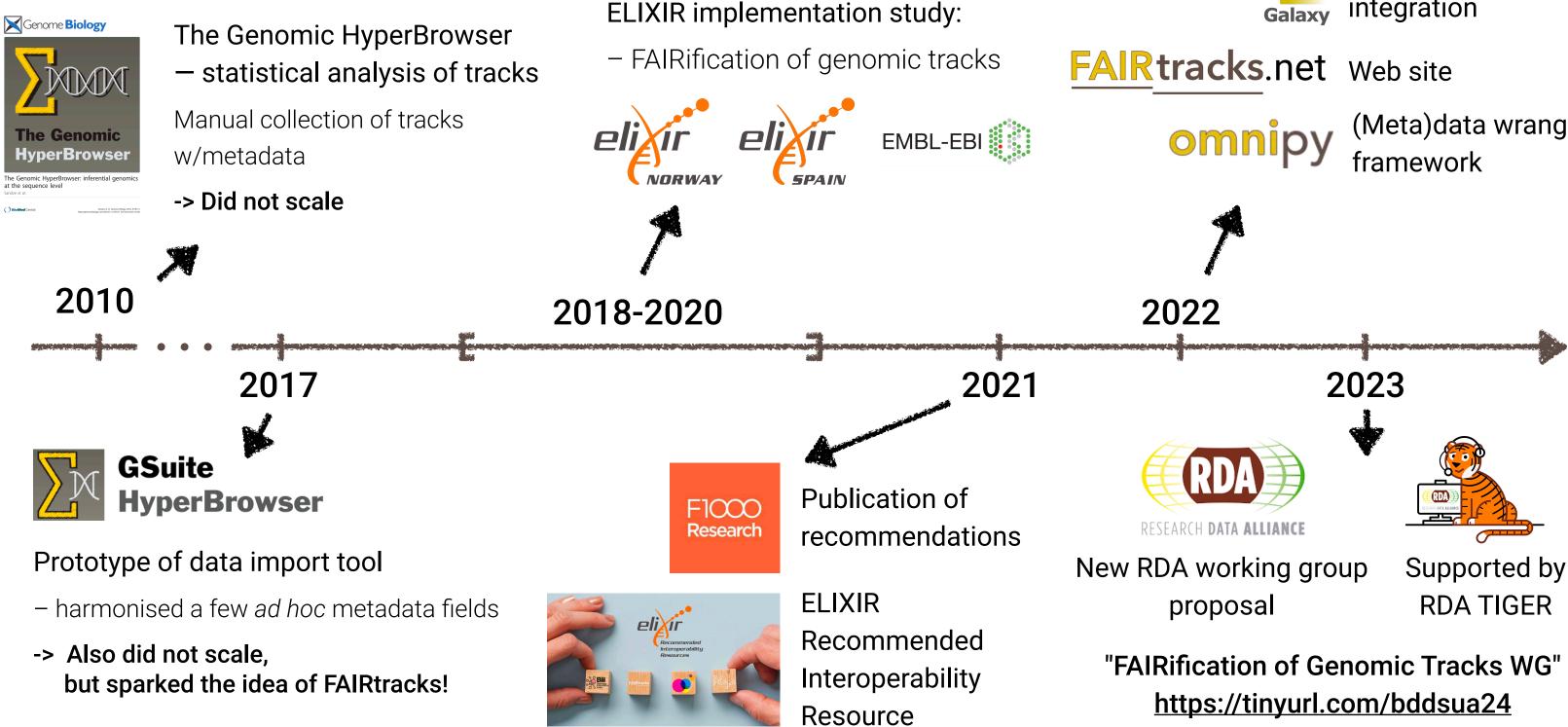
Interoperable

Reusable

- Global identifiers for track files, as well as track collections, studies, samples, and experiments
- Search and import of individual track files across repositories, also repositories not supported by consortia data portals
- Search using formal (non-free-text) queries
- Easy (automated) retrieval of track data
- Persistence of track data and versioned metadata
- Lack of standard metadata model with practically useful attributes
- Annotation using community-accepted vocabularies/ontologies
- Cross-references to records in relevant (meta)data repositories
- Support for detailed context-specific metadata content together with standardised summary attributes
- Simple process for data providers to submit data and metadata that at the same time accommodates the required stringency for (automated) downstream usability
- Easily available data usage licenses
- Detailed provenance of experimental and *in silico* analysis steps

Gundersen S, Boddu S, Capella-Gutierrez S et al. Recommendations for the FAIRification of genomic track metadata [version 1; peer review: 1 approved]. F1000Research 2021, 10(ELIXIR):268 (https://doi.org/10.12688/f1000research.28449.1)

Timeline of the FAIRtracks project





Galaxy integration

(Meta)data wrangling

FAIRification of Genomic Tracks WG Main Goals

Allow researchers to:

- Discover and access genomic data
- ... from disparate repositories
- ... from different types of experiments
- ... at a sub-dataset granularity
- ... through categorical search in uniformly FAIRified metadata that relate to "genomic tracks"

Ulterior goal

Enable data-driven discoveries through analysis of condensed and heterogeneous genomic data files from disparate repositories

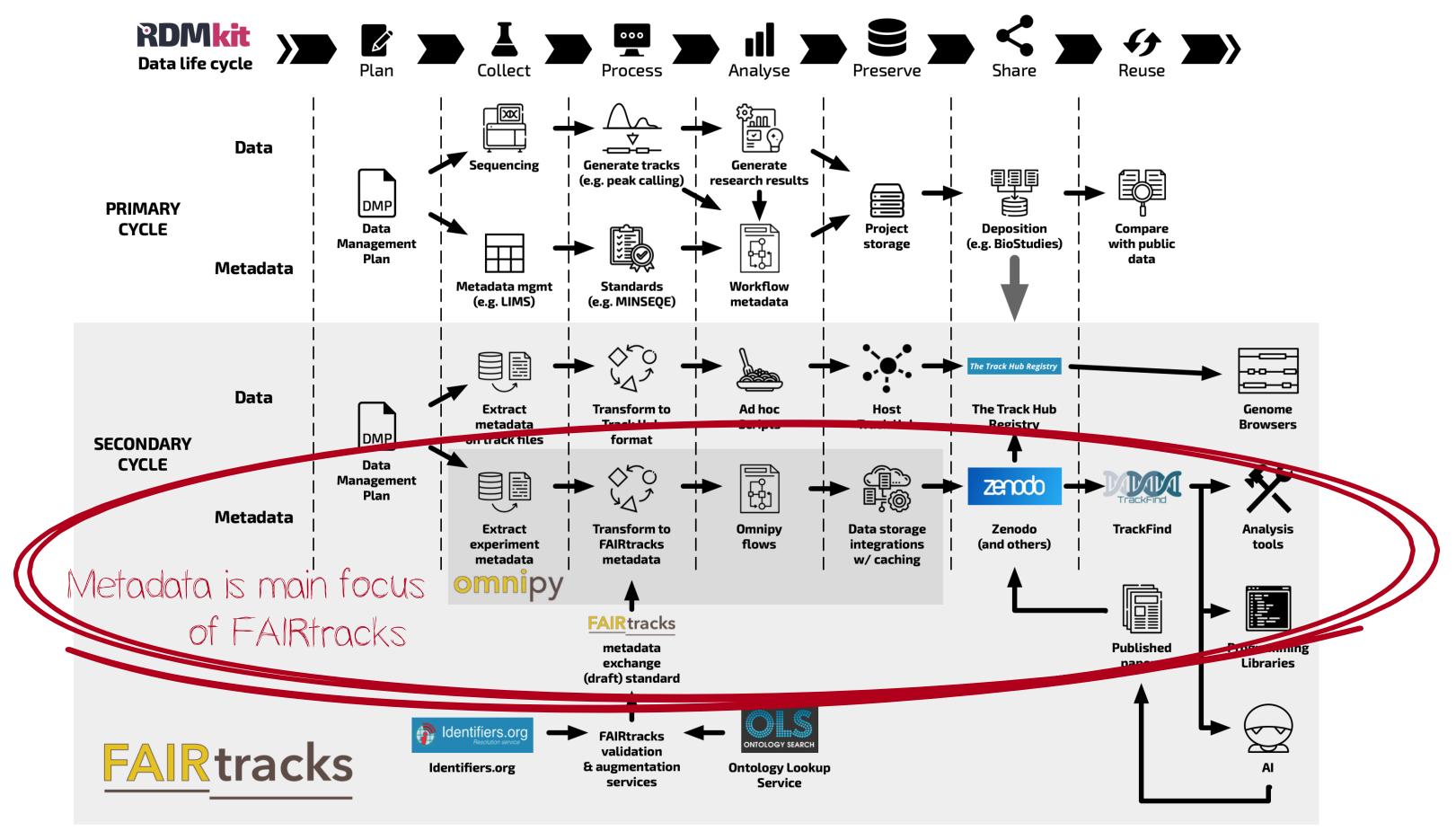
FAIRification of Genomic Tracks WG **Detailed Goals**

- Build a global community, bringing in key people and entities to build a solid foundation for a truly global standard
- Further development of recommendations / metadata schemas
 - Refine ontology selections, metadata fields
 - Support more experiment types, species, biospecimen types etc. based on expert knowledge
- Integrate with new data repositories or other data sources
 - Implement continuous metadata transformation pipelines from different collections containing track files
- Integrate with downstream software tools and frameworks
- Support use cases from analytical end users, e.g. Al-driven methodologies
- Build on other efforts:
 - Adopt recommendations/standards from RDA, GA4GH etc.
 - Integrate with relevant solutions: metadata standards, metadata and ontology term mapping, etc.

FAIRification of Genomic Tracks WG What we have (1/4)

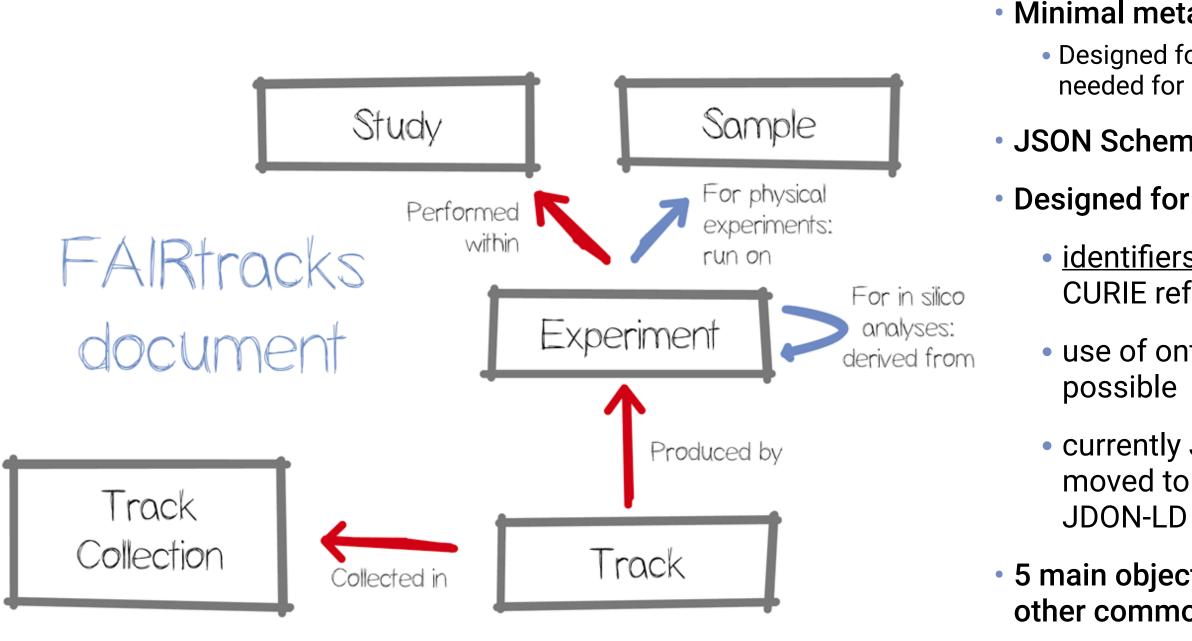
FAIRtracks

- FAIRtracks draft metadata exchange standard
- Services for augmentation and validation of FAIRtracks metadata
- TrackFind
 - categorical search of track metadata (web + REST API)
- User–friendly data import tool in Galaxy
 - using TrackFind as backend



Icons: RDMkit, Google Material Symbols, Flaticon.com

FARtracks draft metadata standard



Minimal metadata exchange format

 Designed for minimal information needed for downstream usability

JSON Schema-based validation

Designed for interoperability:

• <u>identifiers.org</u>/<u>N2T.net</u> – resolvable CURIE references to external records

• use of ontology terms as much as

currently JSON, but will probably be

JDON-LD for semantic web, RO-Crate, etc.

 5 main object types, mappable to other common standards

<u>https://fairtracks.net/standards</u>

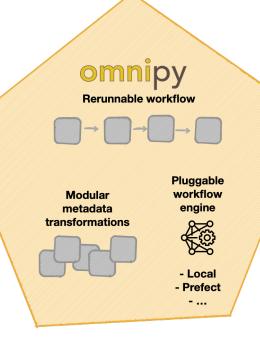
FAIRification of Genomic Tracks WG What we have (2/4)

omnipy

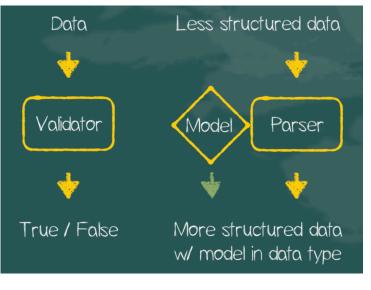
- High-level Python library for type-driven data wrangling and scalable data flow orchestration
- **Developed to allow for rerunnable metadata transformation flows that:**
 - Help researchers extract, manipulate and integrate data and/or metadata from different sources
 - Supports conversion of existing track metadata to the FAIRtracks standard
 - Is general enough to support other data transformation flows, even in other domains

omnipy

- Import (meta)data in almost any shape or form
 - nested JSON
 - tabular (relational) data
 - binary streams
 - other data structures
- Set up a step-by-step data flow that:
 - Reformats data structures
 - Cleans up errors
 - Removes duplicate data
 - Maps and integrates dataset fields
- Provide a catalog of generic task and flow templates that the researcher can refine according to the use case
- For large datasets:
 - Omnipy allows local test jobs to be seamlessly scaled up to the full dataset and offloaded to external compute resources



Highly modular software architecture



"Parse, don't validate"

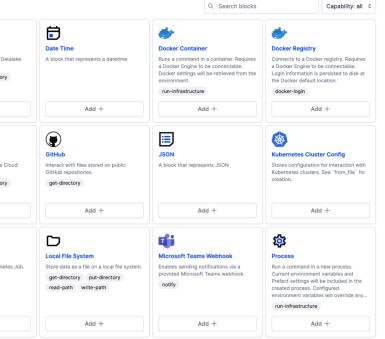
■ Project -	ill m	y_table.csv ×				
monipy_examples - ~/PycharmProjects/omnipy_examples master / Ø	Q	۹				csv 🗸 🛨 🛧 🥠
🗠 🖿 data		i ≣synomym :	ategory :	name ÷	primaryAccession	: ∎uniProtkbId :
✓ ■ 2023_02_27-13_52_44		CDC7L1	Technical term	3D-structure	000311	CDC7 HUMAN
> 10 00_task_import_uniprot.tar.gz		CDC7L1	Coding sequence diversity		000311	CDC7_HUMAN
 I 01_task_transpose_dataset_of_dicts_to_lists.tar.gz I 02_func_flow_flatten_nested_json.tar.gz 	3	CDC7L1	Ligand	ATP-binding	000311	CDC7 HUMAN
 Image: Solution of the state of	4	CDC7L1	Biological process	Cell cvcle	000311	CDC7 HUMAN
3 04_linear_flow_import_and_flatten_uniprot.tar.gz		CDC7L1	Biological process	Cell division	000311	CDC7_HUMAN
III 05_task_pandas_magic.tar.gz	6	CDC7L1	PTM	Isopeptide bond	000311	CDC7_HUMAN
Ill 06_func_flow_import_and_flatten_uniprot_with_magic.tar.gz		CDC7L1	Molecular function	Kinase	000311	CDC7_HUMAN
d my_table.csv		CDC7L1	Ligand	Magnesium	000311	CDC7_HUMAN
> 🖿 dist	9	CDC7L1	Ligand	Metal-binding	000311	CDC7_HUMAN
> 🖿 logs	10	CDC7L1		Nucleotide-binding	000311	
> Contraction of the second se		CDC7L1 CDC7L1	Ligand	Nucleus	000311	CDC7_HUMAN
> src			Cellular component			CDC7_HUMAN
> workshops		CDC7L1	РТМ	Phosphoprotein	000311	CDC7_HUMAN
₫.flake8		CDC7L1	Technical term	Reference proteome	000311	CDC7_HUMAN
🚽 .gitignore		CDC7L1	Molecular function	Serine/threonine-protein kinase		CDC7_HUMAN
🚽 .prefectignore		CDC7L1	Molecular function	Transferase	000311	CDC7_HUMAN
🏭 isajson-deployment.yaml		CDC7L1	РТМ	Ubl conjugation	000311	CDC7_HUMAN
🛗 isajson_prefect-deployment.yaml		Cdc7l1	Coding sequence diversity	Alternative splicing	Q9Z0H0	CDC7_MOUSE
		Cdc7l1	Ligand	ATP-binding	Q9Z0H0	CDC7_MOUSE
間 poetry.lock 的pyproject.tom		Cdc7l1	Biological process	Cell cvcle	Q9Z0H0	CDC7 MOUSE
Run: 🗼 uniprot ×	Te	ct Data				\$
OMNIPY] Mon Feb 27 13:52:44 2023 - INFO: Fini [OMNIPY] Mon Feb 27 13:52:44 2023 - INFO: Writ [OMNIPY] Image: State St	ing (iali: ted : shed ing (shed	dataset as a red "task-pan running "task running "tas dataset as a running "fun	gzipped tarpack to "/Users/ das-magic-qualified-zebra" -pandas-magic-qualified-zeb k-pandas-magic-qualified-ze gzipped tarpack to "/Users/ c-flow-import-and-flatten-u	sveinugu/PycharmProjects/omnipy_ [omnipy.log.registry.RunStateRegi ra* [omnipy.log.registry.RunSt bra*! [omnipy.log.registry.RunSt sveinugu/PycharmProjects/omnipy_ niprot-with-magic-uppish-pudu"! [xamples/data/2023_02 stry] ateRegistry] teRegistry] xamples/data/2023_02 omnipy.log.registry.	_27-13_52_44/04_lin _27-13_52_44/05_tas RunStateRegistry]
Process finished with exit code 0						

Develop, inspect and deploy directly from IDE

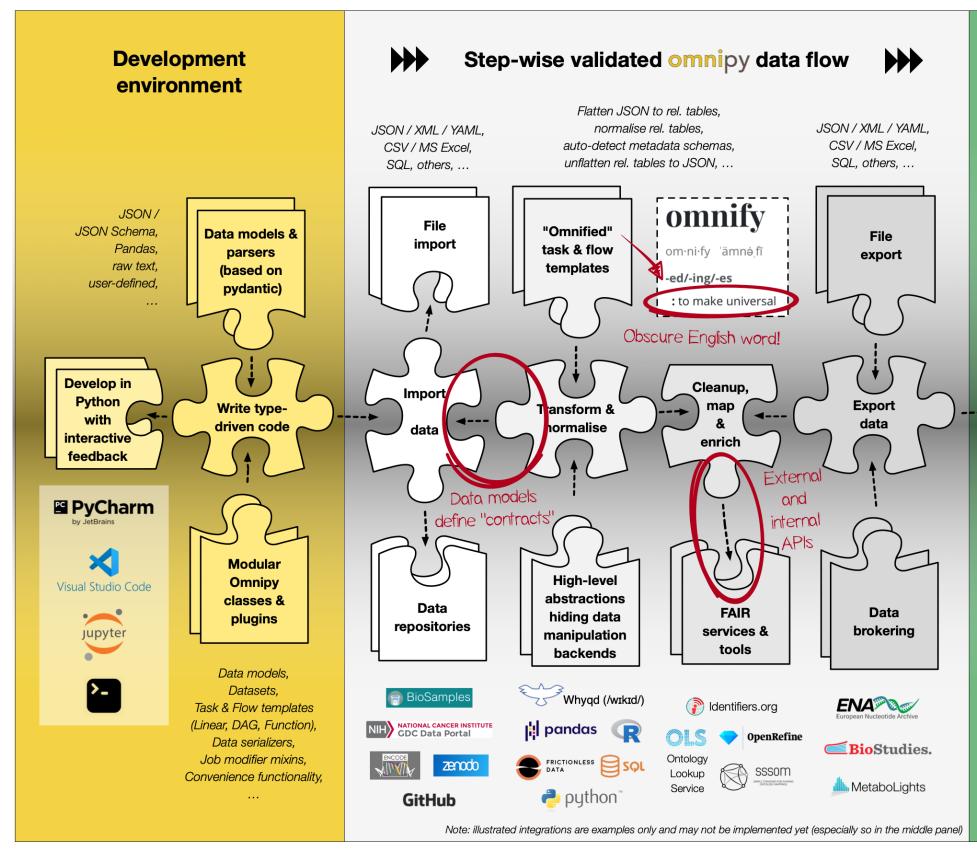
¢	Flow Runs	Blocks / Choose a Bloc
\$	Flows	20 Blocks
0	Deployments	
8	Work Queues	A
ଚ	Blocks	Azure
111	Notifications	Store data as a file on Azure D and Azure Blob Storage.
		get-directory put-director
		read-path write-path
		Add +
		Couproted GCS Store data as a file on Google
		Storage.
		get-directory put-director
		read-path write-path
		+ bbA
		()
		Kubernetes Job
		Runs a command as a Kuberne
		run-infrastructure
ø	Settings	Add +

Prefect orchestration GUI for local/remote deployment

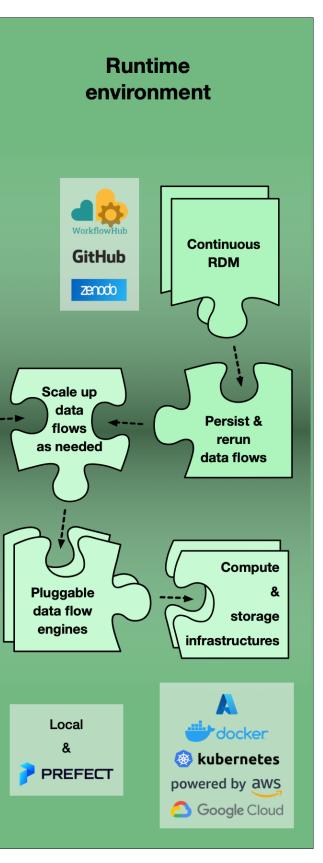
ock



omnipy — The interoperability layer of data wrangling!



Modular Python library for developing and orchestrating scalable (meta)data flows



FAIRification of Genomic Tracks WG What we have (3/4)

FAIRtracks.net website

- Information on:
 - Genome browsers and genomic tracks
 - FAIR interoperability solutions
 - FAIRtracks-related services
 - Background material
- Overview tables:
 - Genome browsers services and software
 - Data portals and other repositories hosting track data
 - Ontologies, datasets, and controlled vocuabularies
- ... plus more to come
- Aims to be a hub for community involvement

Collaboration between

- ELIXIR Norway
- ELIXIR Spain
- EMBL/EBI
- efforts with:
 - Genomic HyperBrowser
 - EPICO (BLUEPRINT data analysis portal)
 - Track Hub Registry
 - ENSEMBL Genome Browser
 - Euro-FAANG project as use case through Peter Harrison (EMBL/EBI)

Existing and planned integration

FAIRification of Genomic Tracks WG What we have (4/4)

- Support from RDA TIGER project
 - "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions"
 - Kick-off Jan 2023
 - Funded through HORIZON-INFRA-2022-EOSC-01-04 call
- FAIRification of Genomic Tracks WG has been selected as a pilot to receive support/facilitation
- We still need to write a case statement to be evaluated by the RDA Technical Advisory Board!



Facilitation services from Group Inception to wrap-up, covering case statement support, meeting management, finalisation support and access to grants.



Output Services

Output support and guidance, including the creation of a maintenance facility.

Communications

Disseminating Group progress and updates at each stage of Group lifecycle to existing and potential stakeholders in the RDA community and beyond.



Landscape & Engagement

Landscape monitoring and knowledge building of potential WG-relevant developments globally.



FAIRification of Genomic Tracks WG What we wish for

- Global members of a FAIRtracks WG, preferably:
 - Data producers
 - FAIR / RDA community
 - Tool developers
 - Analytical end-users / Domain experts
- WG Chairs outside of Europe!
- Help developing the WG case statement
- Lots of members and lots of input!
- FAIRtracks WG kick-off in Salzburg, Oct 2023?

FAIRification of Genomic Tracks WG Interested in contributing to the working group?

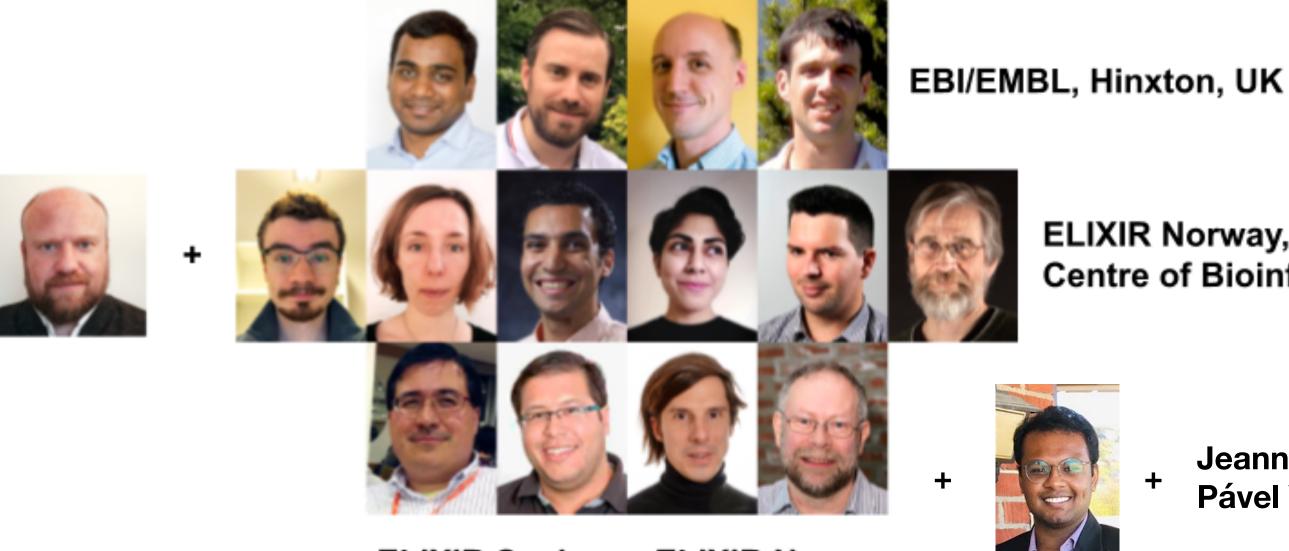
Register you interest by

- Filling out our survey:
 - FAIRification of Genomic Tracks: Community interest for an RDA working group (https://nettskjema.no/a/fairtracks-wg)
- and/or send an email to:
 - <u>fairtracks@elixir.no</u>
- Working document for the "FAIRification of Genomic Tracks WG" initiative
 - <u>https://tinyurl.com/bddsua24</u>
- Shareable link to this presentation
 - <u>https://tinyurl.com/2xhhwsb8</u>

Acknowledgements



The FAIRtracks team



ELIXIR Spain, Barcelona Supercomputing Centre

ELIXIR Norway, NTNU & UIB

+

ELIXIR Norway, Centre of Bioinformatics, UiO

Jeanne Cheneby, Pável Vázquez

Acknowledgements





With funding from **The Research** Council of Norway



Barcelona Supercomputing Center Centro Nacional de Supercomputación









meosc



Norwegian University of Science and Technology



UiO **University of Oslo**



Extra slides

Relevant metadata standards

• We have considered important existing standards, authorities, and consumers of track (and related) metadata:

 INSDC 	 DATS (BioCADDIE, 	• Tra
 IHEC 	GDC)	 Zer
• FAANG	 ICGC/PCAWG 	• GSI
	 ISA Model 	•

- In contrast to many other standards, FAIRtracks have been driven mainly by end-user functionality
- Our focus has *not* been to make sure all important metadata is archived properly
- Rather, FAIRtracks is a **metadata exchange** standard!

ackHub file format nbu Suite HyperBrowser



Identifiers for genomic tracks?

- Should there be globally unique, persistent identifiers for genomic track files (e.g. BigBED/BigWIG, VCF, etc)
- No such thing exists, but we highly recommend that they should be created and indexed
- Also, identifiers for track collections would be very useful
 - Would easily FAIRify "Mix-and-match" track file collections analysed in particular research papers, if the track files are already FAIRified



Use of ontology terms

Advantages:

- Interoperability
- Make use of existing expert knowledge

Disadvantages:

- Selecting proper ontologies
- Missing terms need to coordinate with ontologies
 - But can create simple vocabulary in the meantime
- Ontology versioning
 - How to update metadata when ontologies change?
 - Our solution: data flows that can be automatically in a CI/CD framework

Metadata augmentation

- Makes machinereadable metadata understandable by humans
 - Fills out ontology labels from term IDs
 - Fetches info by ID in other databases
 - Creates summary fields
 - Other housekeeping tasks

README.md

The FAIRtracks augmentation service

The FAIRtracks draft standard is a set of JSON Schemas that define a minimal standard for genomic track metadata. The FAIRtracks augmentation service is a simple Flask-based service written in Python that provides a HTTP-based API for augmenting a minimal FAIRtracks JSON document with automatically generated content for all the properties defined in the FAIRtracks JSON schemas with augmented=true.

The FAIRtracks augmentation service fills out the following fields:

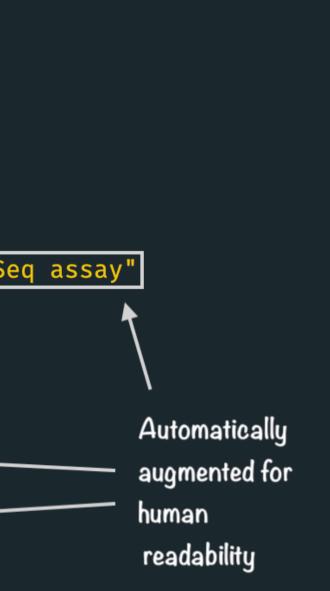
- The newest versions of all ontologies required by the FAIRtracks standard are fetched and the versioned URLs are added to the doc_info -> doc_ontology_versions object.
- All term_value properties are generated by the related term_id property by lookup in the relevant ontology.
- track -> file_name is generated from the track -> file_url property.a
- The two properties sample -> sample_type -> summary and experiment -> target -> summary are generated based upon the relevant rule as defined by sample -> biospecimen_class -> term_id and target -> technique -> term_id , respectively, as described in the top-level FAIRtracks schema.
- sample -> species_name is generated by a lookup in the NCBI Taxonomy Database



0

Experiment object example – augmented version

```
"global_id": "ega.experiment:EGAX00001215632",
"local_id": "ERX547964",
"study_ref": "EGAS00001000326",
"sample_ref": "S00B1LH1",
"technique": {
    "term_id": "http://purl.obolibrary.org/obo/OBI_0002017",
   "term_label": "histone modification identification by ChIP-Seq assay"
},
"target": {
    "sequence_feature": {
        "term_id": "http://purl.obolibrary.org/obo/S0_0001410",
        "term_label": "experimental_feature"
    },
    "summary": "experimental_feature (Input)",
    "details": "Input"
},
"lab_protocol_description":
  "http://www.blueprint-epigenome.eu/index.cfm?p=7BF8A4B6-F4FE-861A-2AD57A08D63D0B58",
"compute_protocol_description":
  "http://dcc.blueprint-epigenome.eu/#/md/chip_seq_grch38"
```



Experiment object example – minimal version

```
"global_id": "ega.experiment:EGAX00001215632",
   "local_id": "ERX547964",
   "study_ref": "EGAS00001000326",
   "sample_ref": "S00B1LH1",
   "technique": {
       "term_id": "http://purl.obolibrary.org/obo/OBI_0002017"
   },
   "target": {
       "sequence_feature": {
            "term_id": "http://purl.obolibrary.org/obo/S0_0001410"
       },
       "details": "Input"
   },
   "lab_protocol_description":
     "http://www.blueprint-epigenome.eu/index.cfm?p=7BF8A4B6-F4FE-861A-2AD57A08D63D0B58",
   "compute_protocol_description":
     "http://dcc.blueprint-epigenome.eu/#/md/chip_seq_grch38"
}.
```



Minimal version only requires ontology term identifiers

Overview of metadata fields for Example schema

Experiment

Droporty	Turne	Dequired	P
Property	Туре	Required	te
@schema	string	Optional	
global_id	string	Optional	te
local_id	string	Required	
study_ref	string	Required	
sample_ref	string	Optional	λ συρ
aggregated_from	array	Optional 🔺	AnyOf
technique	object	Required	
target	object	Required	
lab_protocol_description	string	Optional	
compute_protocol_description	string	Optional	
Additional Properties	Any	Optional	

Technique

Property	Туре	Required
term_id	string	Required
term_label	string	Optional

Target

Property	Туре	Required
sequence_feature	object	Optional
gene_id	string	Optional
gene_product_type	object	Optional
macromolecular_structure	object	Optional
phenotype	object	Optional
details	string	Optional
summary	string	Optional

- Subclass of:
- "Planned process" (OBI ontology) or
- "Operation" (EDAM ontology)

Example schema: technique -> target dependencies

technique	target sequence_feature	aono product typo	gono id	maaram
technique	sequence_reature	gene_product_type	gene_id	macrom
bisulfite sequencing assay	open_chromatin_region			
DNase I hypersensitive sites sequencing assay	DNasel_hypersensitive_site			
histone modification identification by ChIP-Seq assay	(any)			
microRNA profiling by high throughput sequencing assay		MicroRNA		
transcription factor binding site identification by ChIP-Seq assay		Transcription Factor	(any)	
RNA-seq assay		Messenger RNA		
Hi-C assay				Chromati
GWAS study				

In addition:

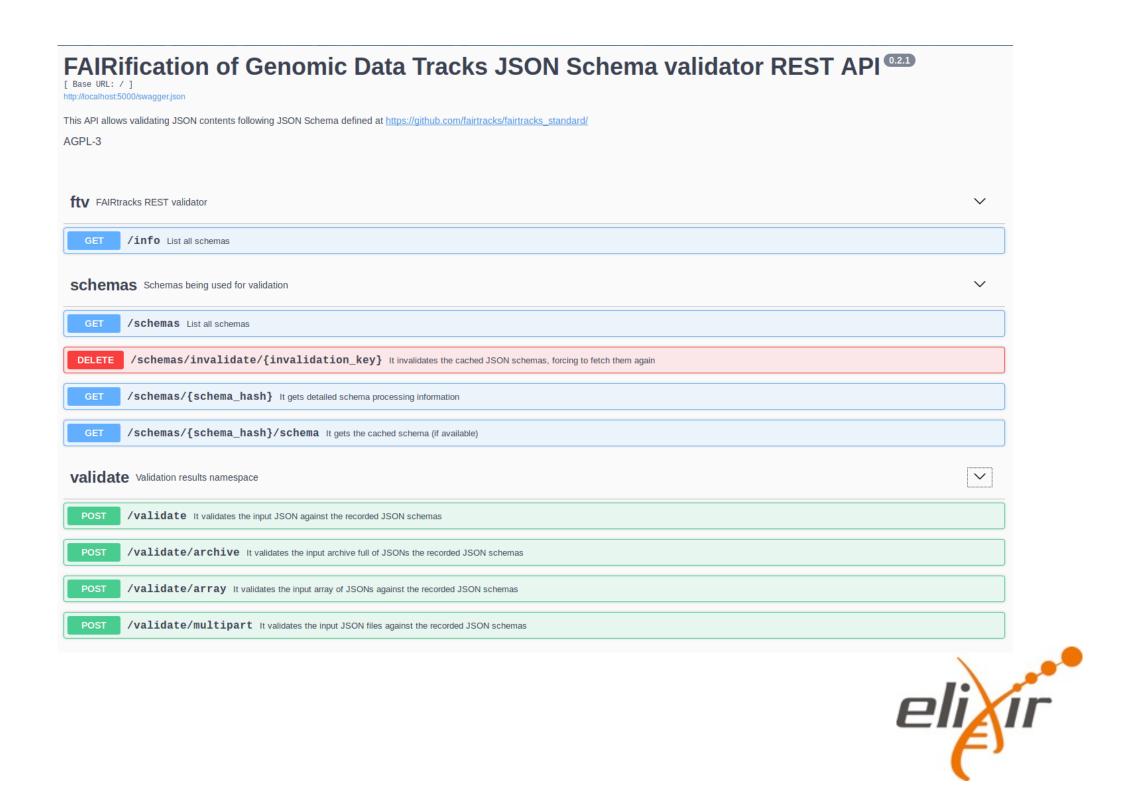
- details text field
- summary autogenerated based on other fields (incl details)

molecular_structure	phenotype
tin Structure	
	(any)

Metadata validation

 Extension of JSON Schema validator developed through OpenEBench

- Validates:
 - identifier.org CURIEs
 - ontology terms
 - duplicate records in relational schemas



TrackFind



experiments.content->'target'->'term_value' ?

Clear all

Search 🗷

AND experiments.content->'technique'-

AND samples.content->'sample_type'-

>'term_value' ? 'ChIP-seq assay'

>'term_value' ? 'bone marrow'

Search query

Categories

tracks

studies

Limit 10

experiments

'H3K4_trimethylation'

Blueprint	IHEC	< >
	CD4-positive, alpha-beta T cell	
	CD4-positive, alpha-beta thym	ocyte
	CD8-positive, alpha-beta T cell	
	CD8-positive, alpha-beta thym	ocyte
	Kit-negative, Ly-76 high polych	romatop
	adult endothelial progenitor ce	ell
	alternatively activated macrop	hage
	band form neutrophil	
	bone marrow	
	capillary blood	
	central memory CD4-positive,	alpha-be
	central memory CD8-positive,	alpha-be
Show o	nly FAIRtracks attributes	
- ilter values		
^	dd to query 🕶 (乀: OR, 企: NOT)	



elar Login

content_type

cs_hash

Results

- experiment_ref
- genome_assembly

Export (6) entries as GSuite file

Export (6) entries as JSON file



GDPR Privacy Policy Terms of Use



Tool integration

- TrackFind client implemented in GSuite HyperBrowser:
 - <u>https://hyperbrowser.uio.no/trackfind_test</u> (search for tool "trackfind")
- JSON and GSuite (<u>http://gtrack.no</u>) formats as metadata / search result exchange format
- Search results can be transferred to the HyperBrowser server, preprocessed, and used in statistical analyses
- EPICO integration still in development

TrackFind client		
Select repository: Blueprint – Blu	eprint 🕈	
Select attribute: samples 🕈	Ð	
_ sample_type	e 🗘	
_ (term_value	e 🗘	
Selection type: Single selection	\$	
Select value: naive B cell		
Select attribute: (experiments 🕈		
_ (tech_type	•	
_ (term_value	e \$)	
Selection type: Single selection		
Select value: ChIP-seq assay	•	
Check all Uncheck all	ound)	
Select tracks Keep all tracks	\$	
Track title	Type of data	Cell/tis
H3K27me3 on naive B cell (16)	Annotation track	naive B
H3K36me3 on naive B cell (6)	Annotation track	naive B
H3K4me1 on naive B cell (7)	Annotation track	naive B
H3K4me1 on naive B cell (13)	Annotation track	naive B
H3K4me3 on naive B cell (8)	Annotation track	naive B
H3K36me3 on naive B cell (10)	Annotation track	naive B
Expand table (now showing 6 of	107 rows)	

Execute

ype	Target	Genome build	File format
	H3K27_trimethylation_site	GRCh38	bigWig
	H3K36_trimethylation_site	GRCh38	bigWig
	H3K4_monomethylation_site	GRCh38	bigBed
	H3K4_monomethylation_site	GRCh38	bigBed
	H3K4_trimethylation	GRCh38	bigBed
	H3K36_trimethylation_site	GRCh38	bigWig

