

# Bioinformatics carpentry utilizing Galaxy

**Day 1 - Introduction to Galaxy**

April 12-16, 2021, 09:00 - 12:00 (MEZ)

Tutor - Steffen C. Lott

# Course Overview - Tutors

## - Rostock -

David Brauer

**Data formatting**

Markus Wolfien

**Transcriptomics**

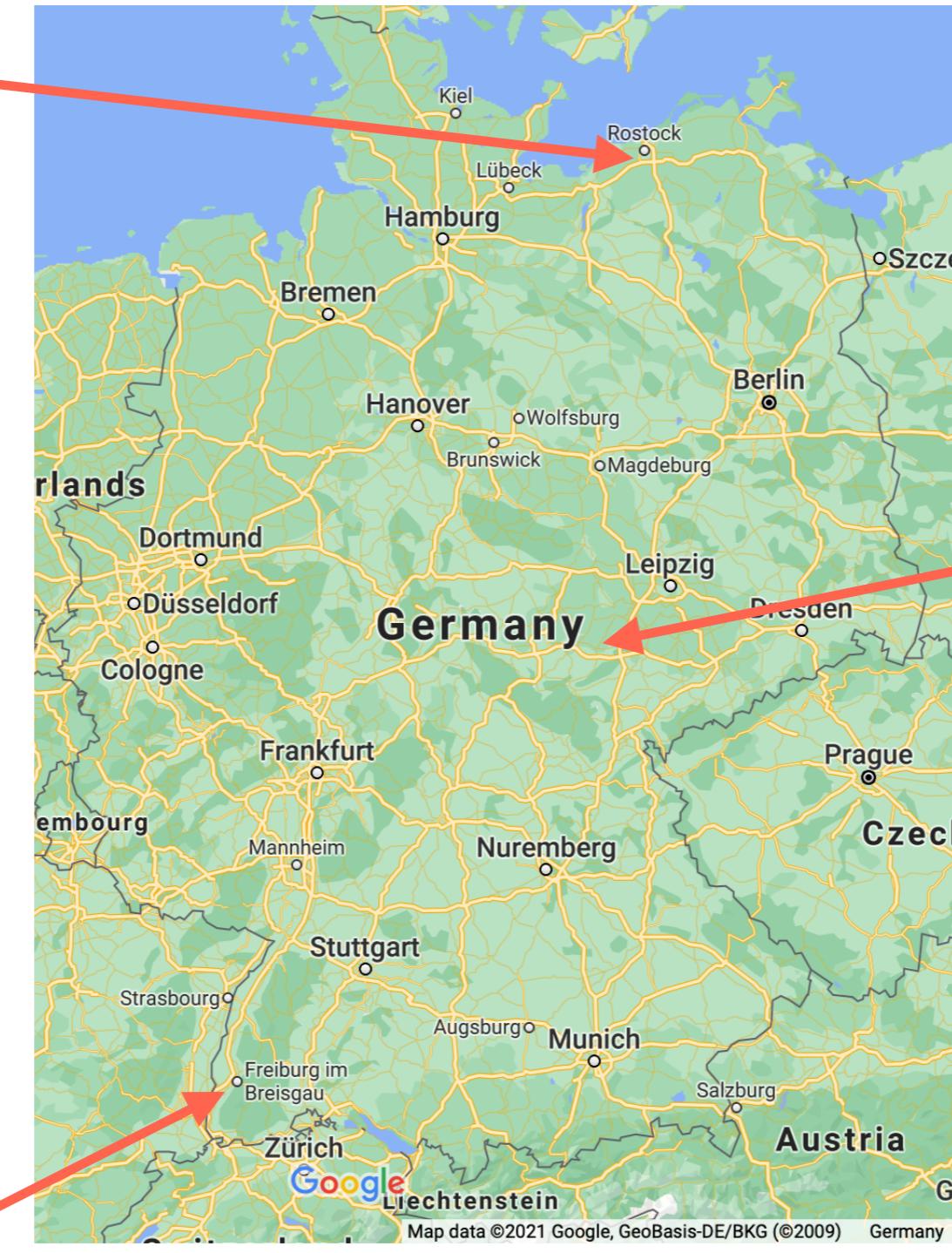
Maximilian Hillemanns

**Machine learning**

## - Freiburg -

Steffen C. Lott

**Introduction to Galaxy**

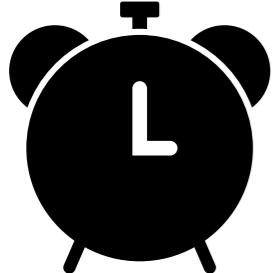


## - Jena -

Konstantin Riege

**Workflow Generator**

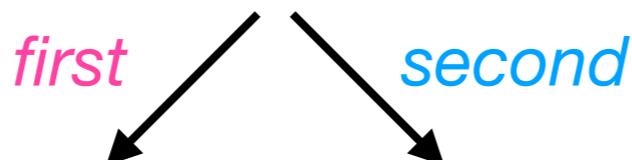
# Course Overview



The course starts at 9 o'clock and will end at '12 o'clock'

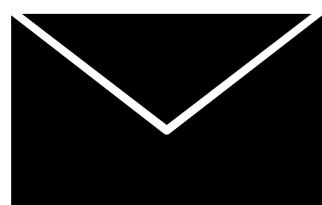
From Tuesday - Friday we start with a question and answer session

The course is divided into two parts



**Lecture Session      Hands-On Session**

*questions ?  
not later than 5 p.m.*



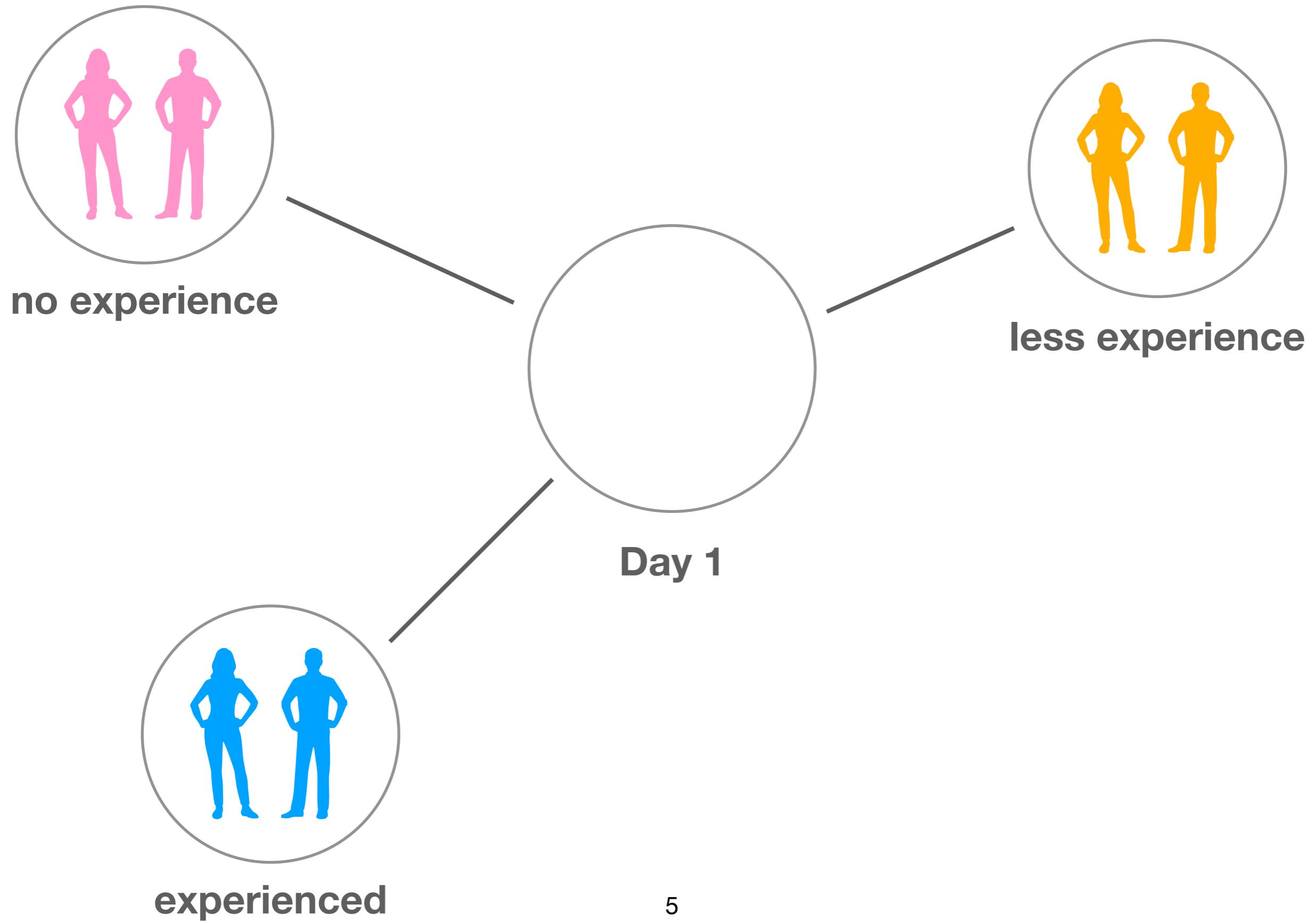
[steffen.lott@biologie.uni-freiburg.de](mailto:steffen.lott@biologie.uni-freiburg.de)  
[david.brauer@uni-rostock.de](mailto:david.brauer@uni-rostock.de)  
[markus.wolfien@uni-rostock.de](mailto:markus.wolfien@uni-rostock.de)  
[maximilian.hillemanns@uni-rostock.de](mailto:maximilian.hillemanns@uni-rostock.de)  
[konstantin.riege@leibniz-fli.de](mailto:konstantin.riege@leibniz-fli.de)

day 1  
day 2  
day 3  
day 4  
day 5

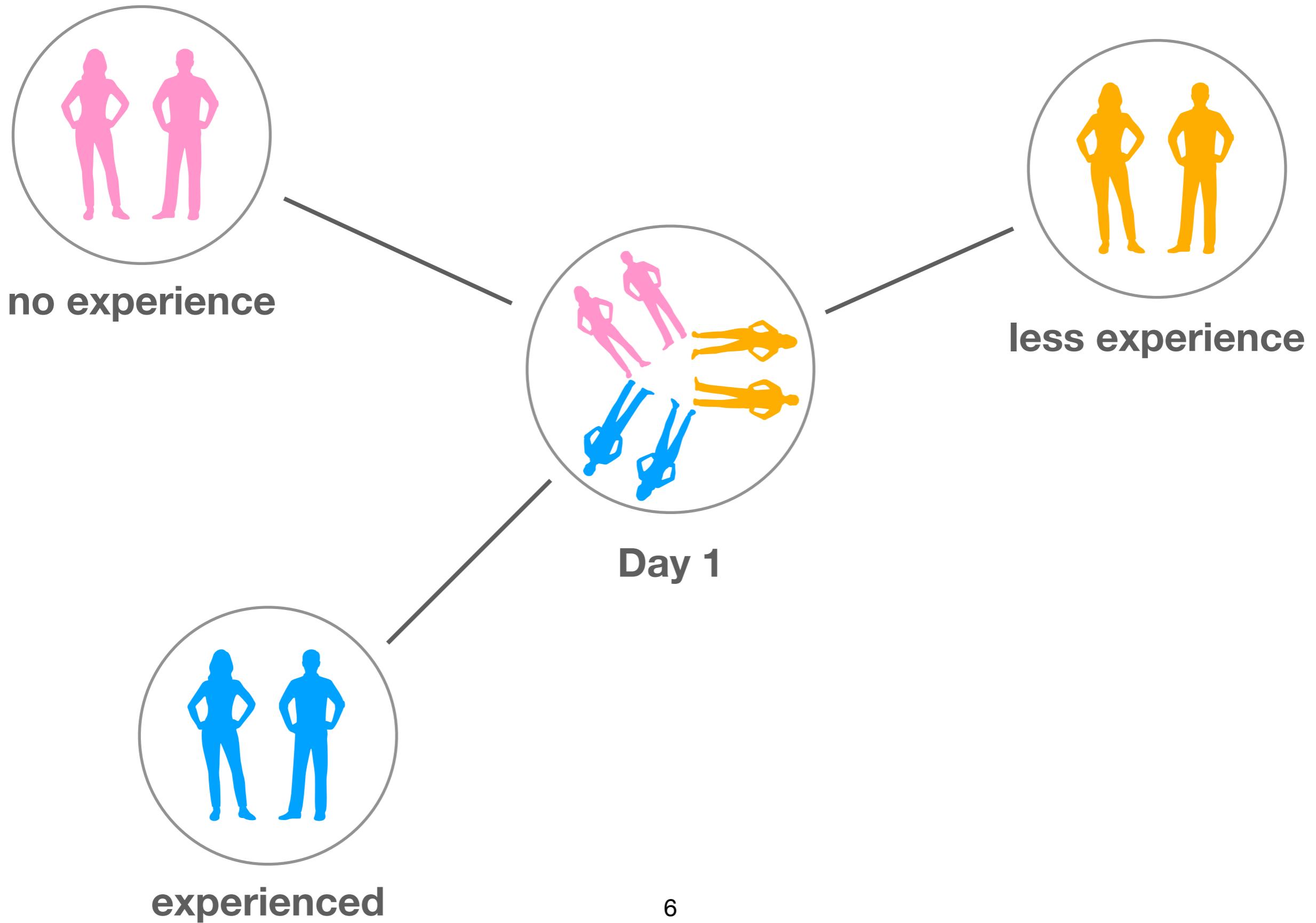
# Course Overview

## Evaluation

# Participants - Galaxy



# Participants - Galaxy



# Upcoming Events 2021



June / July

<https://www.vibconferences.be/events/gcc2021-virtual-edition>



September / October

de.NBI Summer School 2021 - Proteomics & Lipidomics

<https://www.denbi.de/training/1150-de-nbi-summer-school-2021-proteomics-lipidomics>



<https://usegalaxy.eu>

## Events

Oct 19, 2021

Analyse avancée de séquences

Oct 11, 2021

Workflow4Experimenters (W4E) international course 2021

Sep 14, 2021

Analyse de données de métabarcoding

Jul 25, 2021 - Jul 31, 2021

ISMB/ECCB 2021 & BOSC 2021

Jun 28, 2021 - Jul 10, 2021

Galaxy Community Conference 2021

Jun 7, 2021

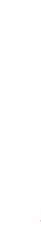
Galaxy for microbiome via ASaIM story – From a local project to a global effort to support microbiome data analysis

# Course Overview - Day 1

Data Handling / Analysis without Galaxy



FAIR Principles



Galaxy



# Data Analysis without Galaxy

# Data Analysis without Galaxy



<https://freesvg.org>

# Data Analysis without Galaxy



<https://freesvg.org>

- developing software
- uses Snakemake or Nextflow
- working with servers, ..., Slurm, and Conda
- ...

# Data Analysis without Galaxy



<https://freesvg.org>



<https://freesvg.org>



# Data Analysis without Galaxy



<https://freesvg.org>

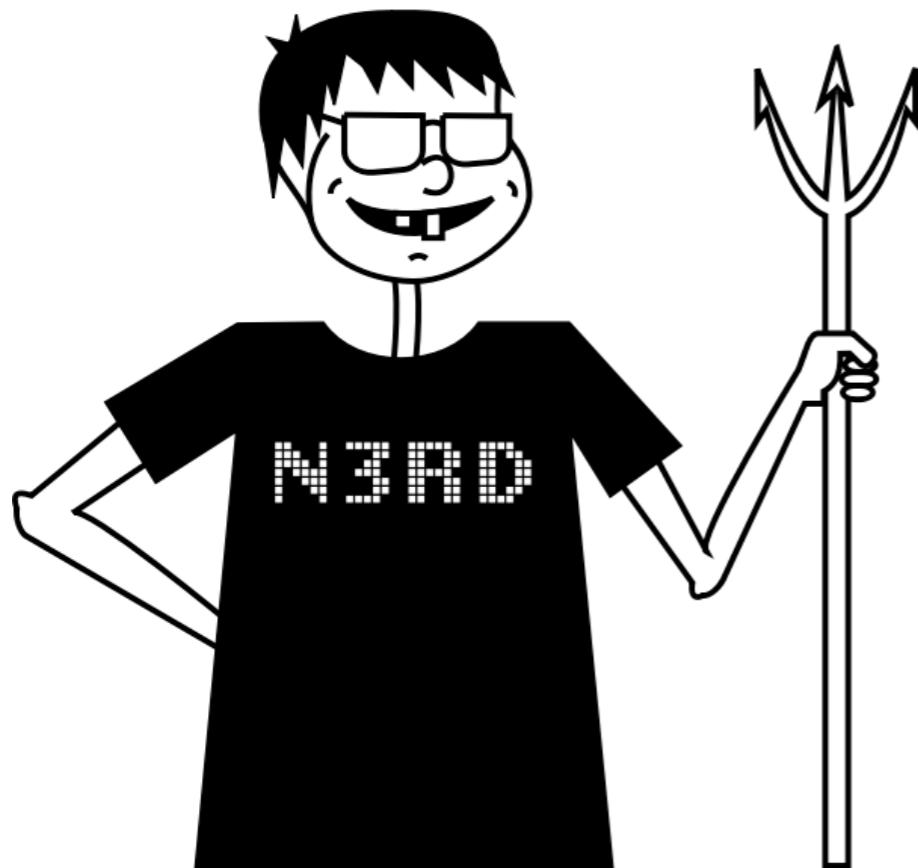


<https://freesvg.org>

*communicate*

- single Tasks e.g. *BLAST* search
- establishing a new pipeline
- applying the same pipeline again and again

# Data Analysis without Galaxy



<https://freesvg.org>



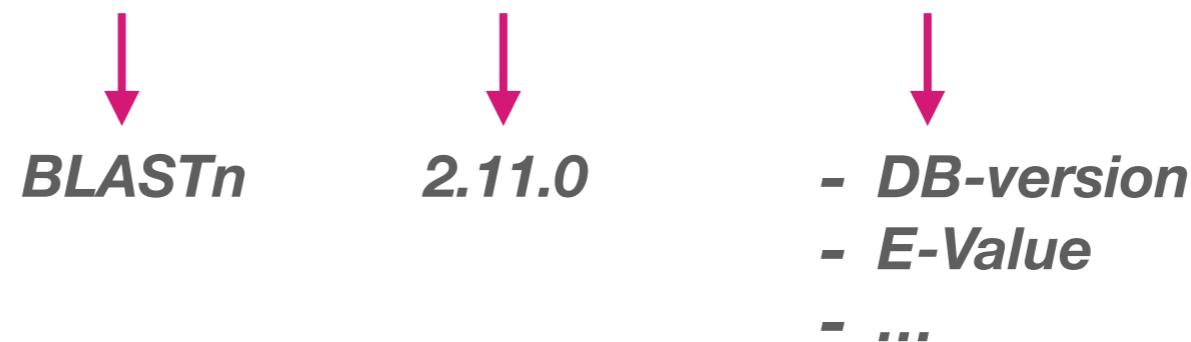
<https://freesvg.org>



# Data Analysis without Galaxy

## Possible problems that may occur

- Which tool was used (tool name, tool version, tool settings)?



# Data Analysis without Galaxy

## Possible problems that may occur

- Which tool was used (tool name, tool version, tool settings)?



- Order of tools (RNA-Seq)?



# Data Analysis without Galaxy

## Possible problems that may occur

- Which tool was used (tool name, tool version, tool settings)?



- Order of tools (RNA-Seq)?



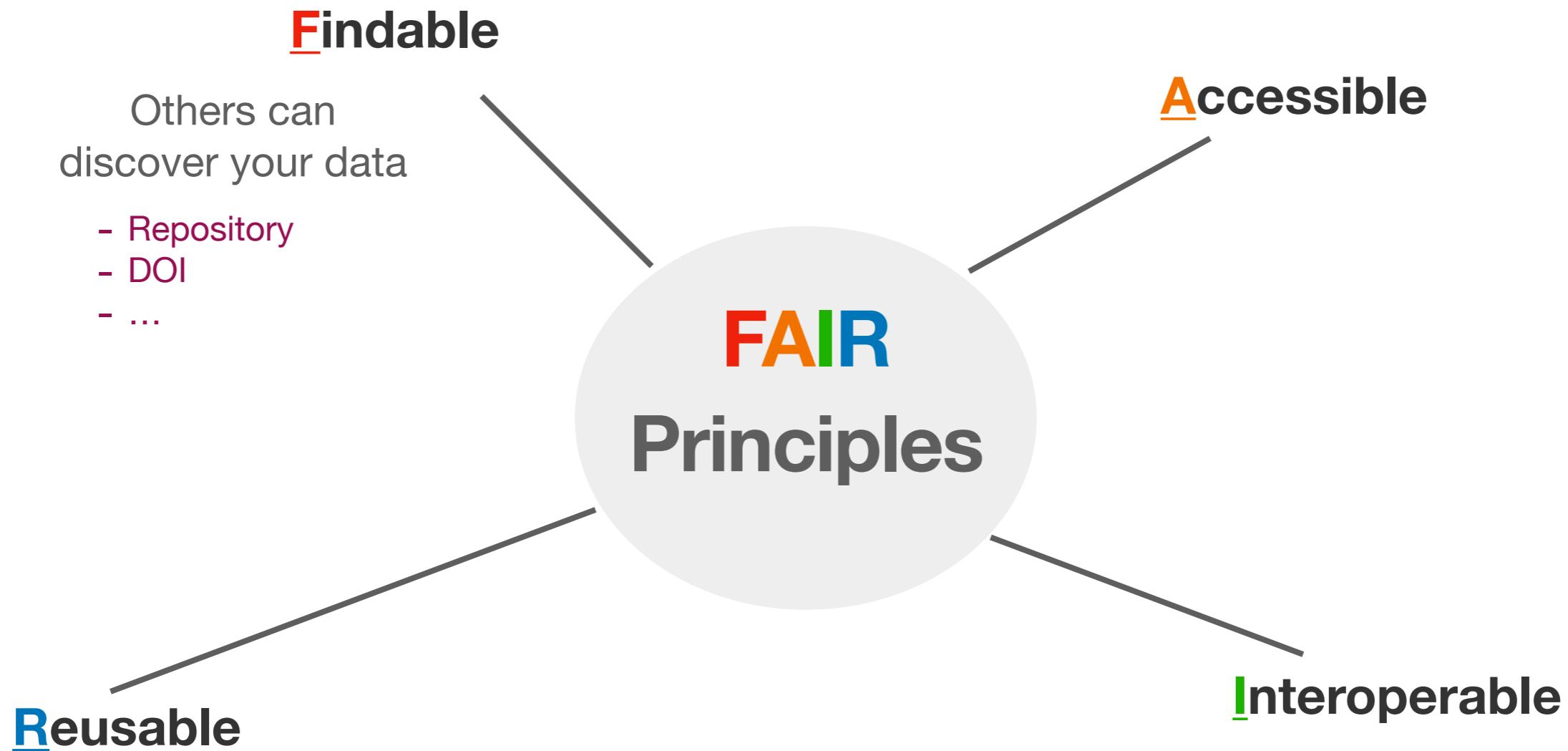
- Data location and how can the data be shared?



# FAIR

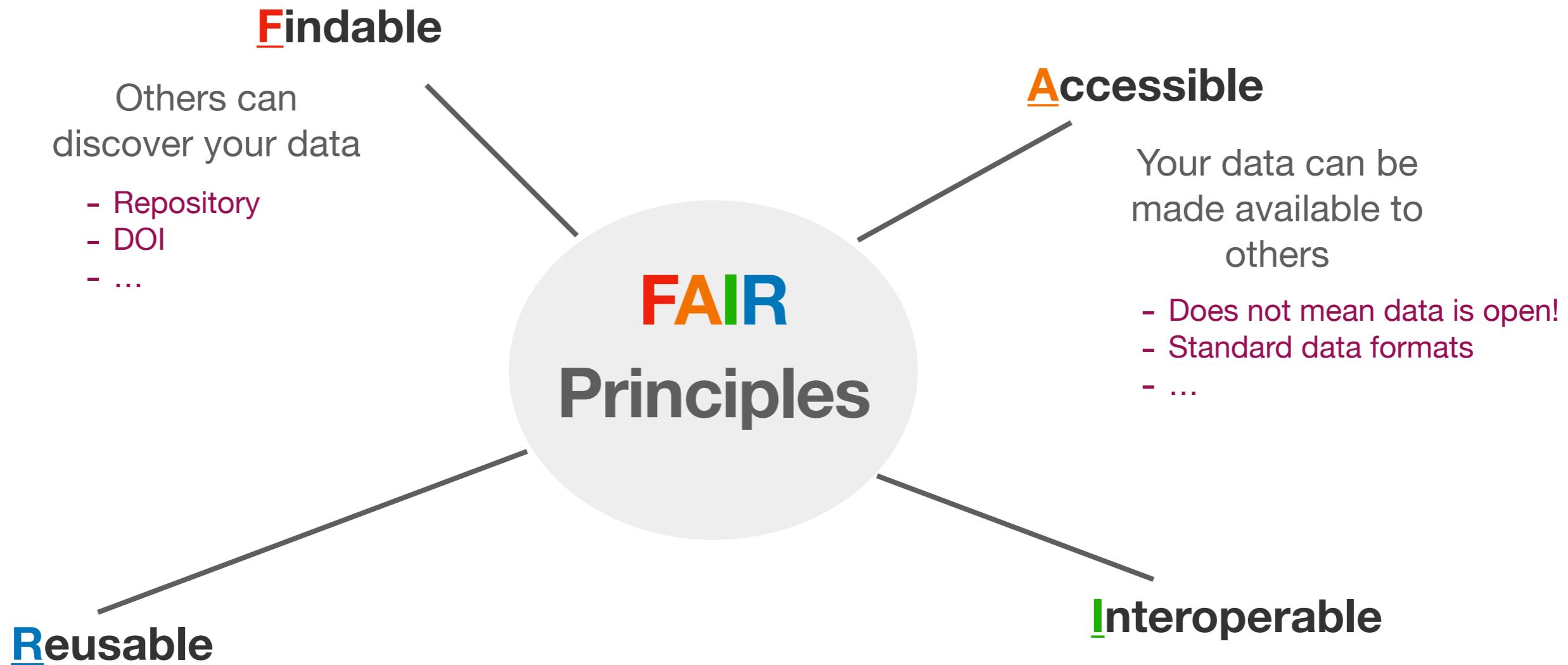
# FAIR Principles

*“The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention)”* <https://www.go-fair.org/fair-principles/>



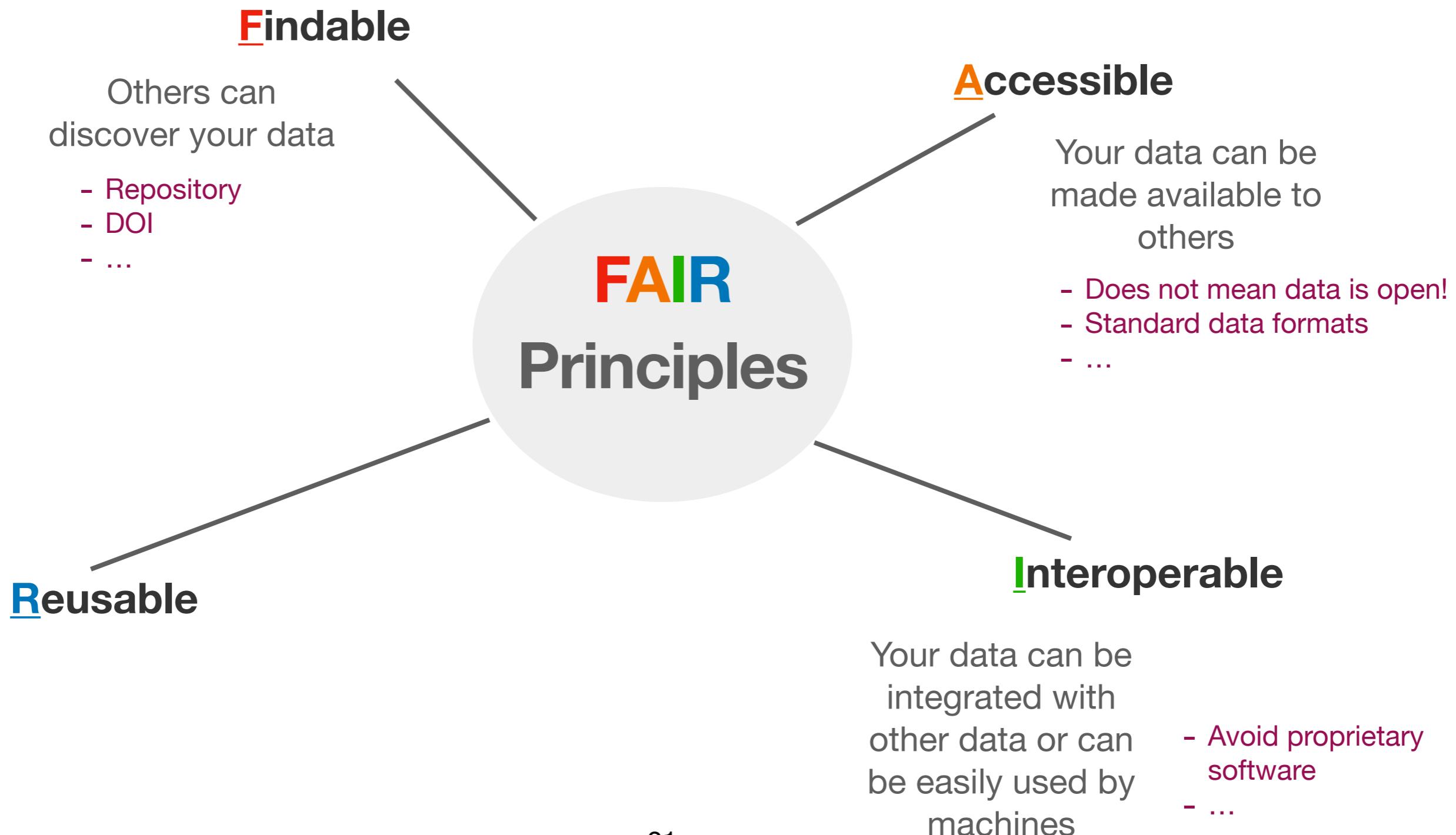
# FAIR Principles

*“The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention)”* <https://www.go-fair.org/fair-principles/>



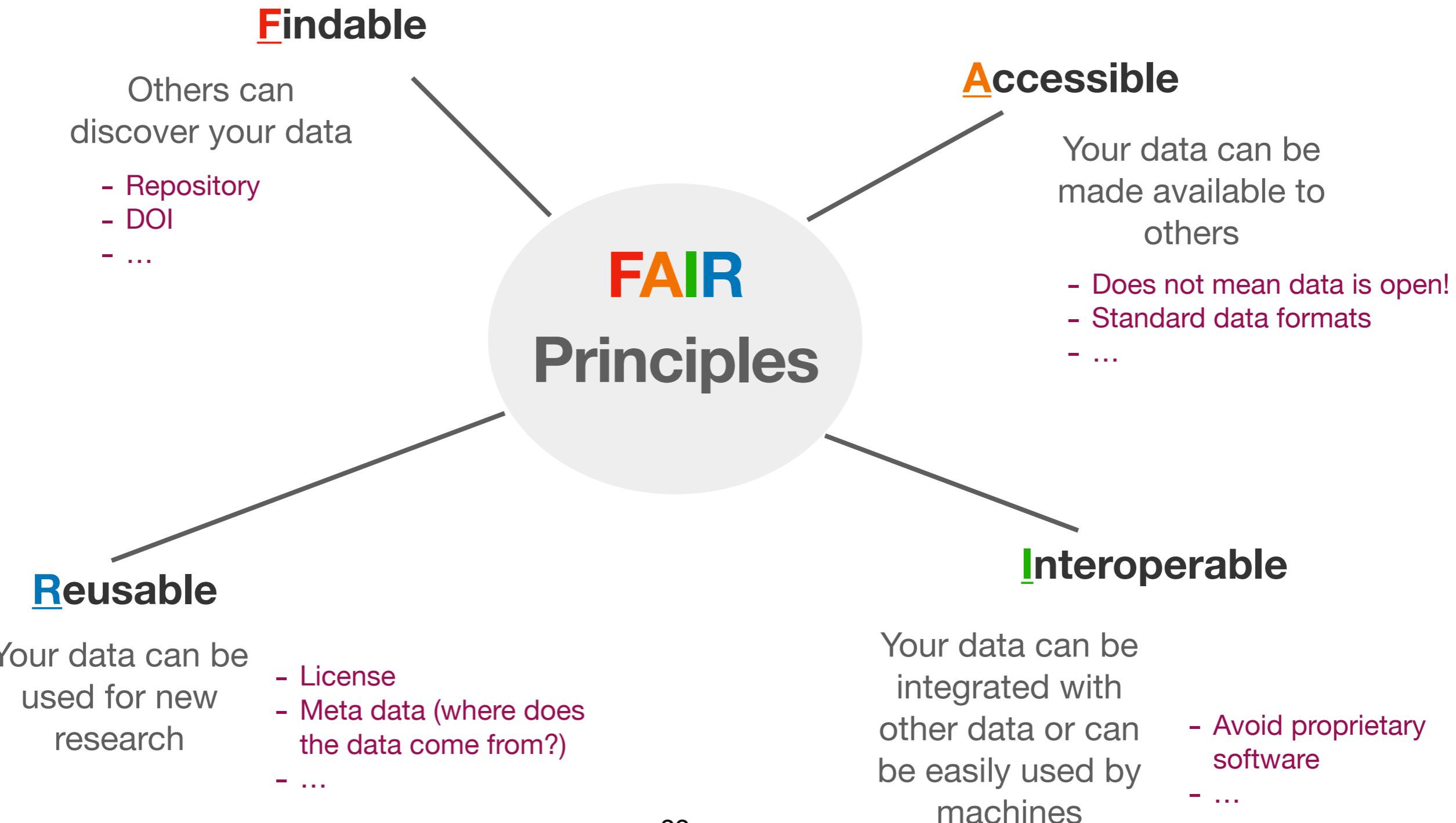
# FAIR Principles

*“The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention)”* <https://www.go-fair.org/fair-principles/>



# FAIR Principles

*“The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention)”* <https://www.go-fair.org/fair-principles/>



# Data Analysis with Galaxy

# Galaxy

## What is Galaxy?

# Galaxy

## What is Galaxy?

- Galaxy is an open source, web-based platform for data intensive biomedical research.

# Galaxy

## What is Galaxy?

- Galaxy is an open source, web-based platform for data intensive biomedical research.



- Galaxy is an open source, web-based platform for data intensive *biomedical* research.

# Galaxy - Core values

## Accessibility

Users without computer science background can easily upload / retrieve data, run complex tools and workflows

## Reproducibility

Galaxy keeps tracking of used tools, tool versions, parameter settings, and the data itself. Therefore, the user can easily understand and repeat a completed analysis.

## Transparency

Users can easily share / publish their histories or workflows.

# Galaxy - Interface

 Galaxy Europe Analyze Data Workflow Visualize ▾ Shared Data ▾ Help ▾ User ▾   Using 77.8 GB

**Tools** 

search tools 

 Upload Data

**Get Data**

**Send Data**

**Collection Operations**

**GENERAL TEXT TOOLS**

**Text Manipulation**

**Filter and Sort**

**Join, Subtract and Group**

**GENOMIC FILE MANIPULATION**

**Convert Formats**

**FASTA/FASTQ**

**FASTQ Quality Control**

<  >

**COVID-19 Research!**

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror **all public SARS-CoV-2 data from ENA** in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our [recent activities](#) for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

**Limited computing capacity on next April 14th and 15th, 2021**

Due to upgrade work to the de.NBI cloud infrastructure, we need to shut down most of our computing resources next April 14th and 15th, 2021. The shutdown will start at **17:00 CEST of April 14th** and everything should return to normal operation by the end of the next day. We will keep the web server running, accepting jobs and workflows, but there will be a delay in processing them. The runn... 

**History**   

search datasets  

**Unnamed history**

(empty)  

 This history is empty. You can [load your own data](#) or [get data from an external source](#)

# Galaxy - Interface

## Menu bar

The screenshot shows the Galaxy Europe interface with a blue header bar. The header includes the "Galaxy Europe" logo, navigation links for "Analyze Data", "Workflow", "Visualize", "Shared Data", "Help", "User", and icons for a graduation cap and a grid. On the right, it shows "Using 77.8 GB".

**Tools** sidebar (left):

- search tools
- Upload Data**

**Get Data**, **Send Data**, **Collection Operations**, **GENERAL TEXT TOOLS**, **Text Manipulation**, **Filter and Sort**, **Join, Subtract and Group**, **GENOMIC FILE MANIPULATION**, **Convert Formats**, **FASTA/FASTQ**, **FASTQ Quality Control**.

**COVID-19 Research!** banner (center-left):

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror **all public SARS-CoV-2 data from ENA** in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our [recent activities](#) for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

**Limited computing capacity on next April 14th and 15th, 2021** banner (center-right):

Due to upgrade work to the de.NBI cloud infrastructure, we need to shut down most of our computing resources next April 14th and 15th, 2021. The shutdown will start at **17:00 CEST of April 14th** and everything should return to normal operation by the end of the next day. We will keep the web server running, accepting jobs and workflows, but there will be a delay in processing them. The runn... [OPEN CHAT](#)

**History** panel (right):

search datasets

**Unnamed history** (empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

# Galaxy - Interface

The screenshot shows the Galaxy Europe interface with a red box highlighting the 'Tools' sidebar.

**Galaxy Europe** Analyze Data Workflow Visualize ▾ Shared Data ▾ Help ▾ User ▾ Using 77.8 GB

**Tools**

- search tools
- Upload Data**

Get Data  
Send Data  
Collection Operations  
**GENERAL TEXT TOOLS**  
Text Manipulation  
Filter and Sort  
Join, Subtract and Group  
**GENOMIC FILE MANIPULATION**  
Convert Formats  
FASTA/FASTQ  
FASTQ Quality Control

**COVID-19 Research!**

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror **all public SARS-CoV-2 data from ENA** in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our [recent activities](#) for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

**Limited computing capacity on next April 14th and 15th, 2021**

Due to upgrade work to the de.NBI cloud infrastructure, we need to shut down most of our computing resources next April 14th and 15th, 2021. The shutdown will start at **17:00 CEST of April 14th** and everything should return to normal operation by the end of the next day. We will keep the web server running, accepting jobs and workflows, but there will be a delay in processing them. The runn...

**History**

search datasets

**Unnamed history**  
(empty)

**i** This history is empty. You can [load your own data](#) or [get data from an external source](#)

Tool bar

# Galaxy - Interface

Galaxy Europe Analyze Data Workflow Visualize ▾ Shared Data ▾ Help ▾ User ▾ Using 77.8 GB

Tools

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

FASTQ Quality Control

COVID-19 Research!

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror **all public SARS-CoV-2 data from ENA** in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our [recent activities](#) for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

Limited computing capacity on next April 14th and 15th, 2021

Due to upgrade work to the de.NBI cloud infrastructure, we need to shut down most of our computing resources next April 14th and 15th, 2021. The shutdown will start at **17:00 CEST of April 14th** and everything should return to normal operation by the end of the next day. We will keep the web server running, accepting jobs and workflows, but there will be a delay in processing them. The runn...

History

Unnamed history  
(empty)

This history is empty. You can load your own data or get data from an external source

Main section

# Galaxy - Interface

states ➔ waiting - running - success - failed

The screenshot shows the Galaxy Europe interface. At the top, there's a navigation bar with links for Analyze Data, Workflow, Visualize, Shared Data, Help, User, and a grid icon. A message at the top right says "Using 77.8 GB". Below the navigation is a "Tools" sidebar with sections for search tools, upload data, Get Data, Send Data, Collection Operations, and GENERAL TEXT TOOLS (which is currently selected). A "COVID-19 Research!" banner provides information about SARS-CoV-2 analysis and training materials. Another banner at the bottom left discusses limited computing capacity on April 14th and 15th, 2021, due to infrastructure upgrade work. On the right side, a red-bordered "History" bar is shown, which is currently empty. It includes a search datasets field, a help icon, and a delete icon.

Galaxy Europe Analyze Data Workflow Visualize ▾ Shared Data ▾ Help ▾ User ▾ Using 77.8 GB

Tools

search tools ✖

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

FASTQ Quality Control

COVID-19 Research!

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror **all public SARS-CoV-2 data from ENA** in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our [recent activities](#) for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

Limited computing capacity on next April 14th and 15th, 2021

Due to upgrade work to the de.NBI cloud infrastructure, we need to shut down most of our computing resources next April 14th and 15th, 2021. The shutdown will start at **17:00 CEST of April 14th** and everything should return to normal operation by the end of the next day. We will keep the web server running, accepting jobs and workflows, but there will be a delay in processing them. The running data will be interrupted and rescheduled.

OPEN CHAT

History

search datasets ✖

Unnamed history

(empty)

This history is empty. You can load your own data or get data from an external source

History bar

# Galaxy - Tool Shed

## App-Store for Galaxy

**Galaxy Tool Shed**    Repositories Groups Help ▾ User ▾

8140 valid tools on Apr 02, 2021

**Search**  
Search for valid tools

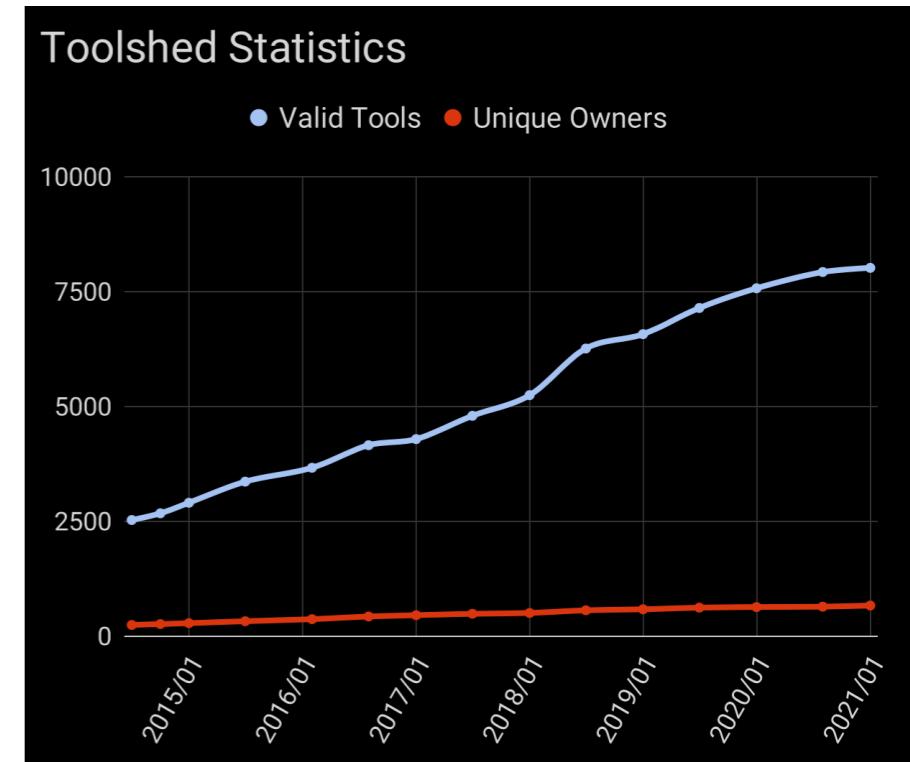
**Valid Galaxy Utilities**  
Tools  
Custom datatypes  
Repository dependency definitions  
Tool dependency definitions

**All Repositories**  
Browse by category

**Available Actions**  
Login to create a repository

Repositories by Category

Name	Description	Repositories
Assembly	Tools for working with assemblies	161
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	75
Climate Analysis	Tools for analyzing climate data	5
Combinatorial Selections	Tools for combinatorial selection	9
Computational chemistry	Tools for use in computational chemistry	161
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	11
Convert Formats	Tools for converting data formats	124
Data Export	Tools for exporting data to various destinations	10
Data Managers	Utilities for Managing Galaxy's built-in data cache	79
Data Source	Tools for retrieving data from external data sources	89
Ecology	Tools related to ecological studies	46



<https://galaxyproject.org/galaxy-project/statistics/>

# Galaxy - Tools

- Proteomics
- Cheminformatics
- Exom-Seq
- BLAST+
- Imaging (ImageJ)
- Metabolomics
- Genome Annotations
- RNA workbench
- EMBOSS
- Phylogeny
- RNA-seq
- ChIP-Seq
- Text Manipulation
- Format Converters
- Filtering and Sorting
- Join, Group
- Sequence Tools
- Multi-Species Alignment Tools
- Summary Statistics
- Graphing / Plotting
- ...

# Galaxy - Flavors

<https://usegalaxy.eu>

The screenshot shows the Galaxy Europe interface. On the left, a sidebar lists various tools under categories like GENERAL TEXT TOOLS and GENOMIC FILE MANIPULATION. The main content area features a green box titled "COVID-19 Research!" with information about SARS-CoV-2 analysis and training materials. Below it is a blue box about computing capacity on April 14th and 15th, 2021. At the bottom, there are news and events sections. On the right, a "History" panel shows an empty history with a message encouraging users to load their own data or get data from an external source.

- > 2000 bioinformatics tools
- Reference Genomes
- 250 GB per user
- Free registration
- Free compute and storage resources

<https://rna.usegalaxy.eu>

The screenshot shows the Galaxy / RNA interface. The sidebar is identical to the Galaxy Europe version. The main content area includes the same COVID-19 Research and computing capacity boxes. At the bottom, a "Welcome to the RNA Galaxy workbench 2.0" message is displayed. The history panel on the right is also empty, with the same encouragement message.

- > 700 bioinformatics tools focused on RNA-Biology
  - RNA structure analysis
  - RNA alignment
  - RNA annotation
  - RNA-protein interaction
  - ribosome profiling
  - RNA-Seq analysis
  - RNA target prediction

# Galaxy - Important Sources

- Galaxy instance, which we will use !  
<https://usegalaxy.eu>
- Source for building a Galaxy instance e.g. run Galaxy on your own computer  
<https://github.com/bgruening/docker-galaxy-stable>
- Collection of Galaxy related tutorials  
<https://galaxyproject.github.io/training-material/>
- de.STAIR GitHub account  
<https://github.com/destairdenbi>
  - **training-material**
  - **galaxy-workflow-generator**
  - **RNA-Seq-scripts**
  - **galaxy-workflows**
  - ...
- de.STAIR readme  
<https://destair.readthedocs.io/en/latest/>

# FASTQ / FASTA

# FASTQ / FASTA

- Single FASTA file

```
>ID Description  
ATTATATATATATCGGGATA
```

**Line 1** begins with a '>' character and is followed by a sequence identifier and an *optional* description  
**Line 2** is the raw sequence letters.

# FASTQ / FASTA

- Single FASTA file

```
>ID Description  
ATTATATATATATCGGGATA
```

**Line 1** begins with a '>' character and is followed by a sequence identifier and an *optional* description  
**Line 2** is the raw sequence letters.

- Multi FASTA file

```
>ID Description  
ATTATATATATATCGGGATA  
>ID Description  
ATTATATATATATCGGGATA  
>ID Description  
ATTATATATATATCGGGATA
```

# FASTQ / FASTA

- Single FASTA file

>ID Description —  
ATTATATATATATATCGGGATA

**Line 1** begins with a '>' character and is followed by a sequence ID and an *optional* description  
**Line 2** is the raw sequence letters.

### o Multi FASTA file

```
>ID Description
ATTATATATATATCGGGATA
>ID Description
ATTATATATATATCGGGATA
>ID Description
ATTATATATATATCGGGATA
```

- Single FASTQ file

- Line 1 begins with an '@' character and is followed by a sequence ID and an *optional* description
- Line 2 is the raw sequence letters.
- Line 3 begins with a '+' character and is *optionally* followed by ID again.
- Line 4 encodes the quality values

# FASTQ / FASTA

- Quality values are encoded in ASCII code

ASCII control characters			ASCII printable characters					Extended ASCII characters								
00	NULL	(Null character)	32	space	64	@	96	`	128	Ç	160	á	192	Ł	224	Ó
01	SOH	(Start of Header)	33	!	65	A	97	a	129	ü	161	í	193	ł	225	ß
02	STX	(Start of Text)	34	"	66	B	98	b	130	é	162	ó	194	ł	226	ö
03	ETX	(End of Text)	35	#	67	C	99	c	131	â	163	ú	195	ł	227	ő
04	EOT	(End of Trans.)	36	\$	68	D	100	d	132	ä	164	ñ	196	—	228	ő
05	ENQ	(Enquiry)	37	%	69	E	101	e	133	à	165	Ñ	197	+	229	ő
06	ACK	(Acknowledgement)	38	&	70	F	102	f	134	â	166	¤	198	ã	230	µ
07	BEL	(Bell)	39	'	71	G	103	g	135	ç	167	°	199	Ã	231	þ
08	BS	(Backspace)	40	(	72	H	104	h	136	ê	168	¿	200	Ł	232	þ
09	HT	(Horizontal Tab)	41	)	73	I	105	i	137	ë	169	®	201	Ł	233	ú
10	LF	(Line feed)	42	*	74	J	106	j	138	è	170	¬	202	Ł	234	ó
11	VT	(Vertical Tab)	43	+	75	K	107	k	139	î	171	½	203	Ł	235	ú
12	FF	(Form feed)	44	,	76	L	108	l	140	†	172	¼	204	Ł	236	ý
13	CR	(Carriage return)	45	-	77	M	109	m	141	ı	173	ı	205	=	237	Ý
14	SO	(Shift Out)	46	.	78	N	110	n	142	Ä	174	«	206	+	238	—
15	SI	(Shift In)	47	/	79	O	111	o	143	Å	175	»	207	¤	239	‘
16	DLE	(Data link escape)	48	0	80	P	112	p	144	É	176	„	208	ð	240	≡
17	DC1	(Device control 1)	49	1	81	Q	113	q	145	æ	177	„	209	Đ	241	±
18	DC2	(Device control 2)	50	2	82	R	114	r	146	Æ	178	„	210	È	242	„
19	DC3	(Device control 3)	51	3	83	S	115	s	147	ô	179	—	211	È	243	¾
20	DC4	(Device control 4)	52	4	84	T	116	t	148	ö	180	—	212	È	244	¶
21	NAK	(Negative acknowl.)	53	5	85	U	117	u	149	ò	181	À	213	ı	245	§
22	SYN	(Synchronous idle)	54	6	86	V	118	v	150	û	182	À	214	ı	246	+
23	ETB	(End of trans. block)	55	7	87	W	119	w	151	ù	183	À	215	ı	247	.
24	CAN	(Cancel)	56	8	88	X	120	x	152	ÿ	184	©	216	ı	248	°
25	EM	(End of medium)	57	9	89	Y	121	y	153	Ö	185	‡	217	ı	249	“
26	SUB	(Substitute)	58	:	90	Z	122	z	154	Ù	186	‡	218	Ł	250	‘
27	ESC	(Escape)	59	;	91	[	123	{	155	ø	187	Ł	219	Ł	251	‘
28	FS	(File separator)	60	<	92	\	124		156	£	188	Ł	220	Ł	252	‘
29	GS	(Group separator)	61	=	93	]	125	}	157	Ø	189	¢	221	Ł	253	‘
30	RS	(Record separator)	62	>	94	^	126	~	158	×	190	¥	222	Ł	254	■
31	US	(Unit separator)	63	?	95	_			159	f	191	ł	223	Ł	255	nbsp
127	DEL	(Delete)														

[https://computersciencewiki.org/images/3/3d/Ascii\\_table.png](https://computersciencewiki.org/images/3/3d/Ascii_table.png)

# FASTQ / FASTA

- Quality values are encoded in ASCII code

## ASCII printable characters

32	space	64	@	96	`
33	!	65	A	97	a
34	"	66	B	98	b
35	#	67	C	99	c
36	\$	68	D	100	d
37	%	69	E	101	e
38	&	70	F	102	f
39	'	71	G	103	g
40	(	72	H	104	h
41	)	73	I	105	i
42	*	74	J	106	j
43	+	75	K	107	k
44	,	76	L	108	l
45	-	77	M	109	m
46	.	78	N	110	n
47	/	79	O	111	o
48	0	80	P	112	p
49	1	81	Q	113	q
50	2	82	R	114	r
51	3	83	S	115	s
52	4	84	T	116	t
53	5	85	U	117	u
54	6	86	V	118	v
55	7	87	W	119	w
56	8	88	X	120	x
57	9	89	Y	121	y
58	:	90	Z	122	z
59	;	91	[	123	{
60	<	92	\	124	
61	=	93	]	125	}
62	>	94	^	126	~
63	?	95	_		

# FASTQ →

ASCII Character	ASCII Decimal	Offset	Phred-Score
A	65	33	32
#	35	33	2
A	65	33	32
F	70	33	37
F	70	33	37
J	74	33	41
J	74	33	41
J	74	33	41
J	74	33	41
J	74	33	41
J	74	33	41
J	74	33	41
J	74	33	41
J	74	33	41
J	74	33	41
<	60	33	27
J	74	33	41

# Offset “33”

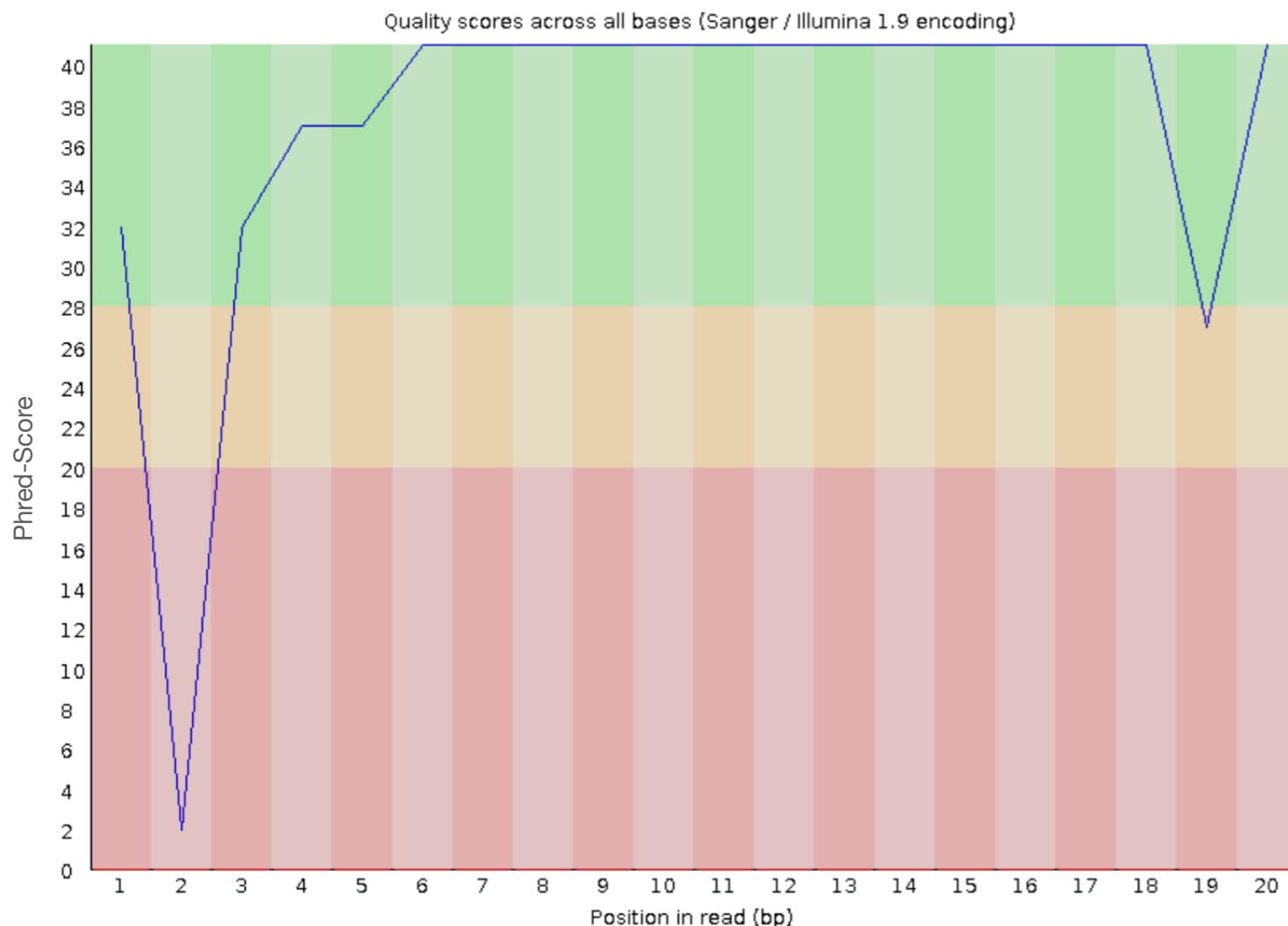
## Since Illumina 1.8

# FASTQ / FASTA

FASTQ →

```
@ID Description
ATTATATATATATCGGGATA
+
A#AFFJJJJJJJJJJJJJJJJ<J
```

- FASTQC Report - Per base sequence quality

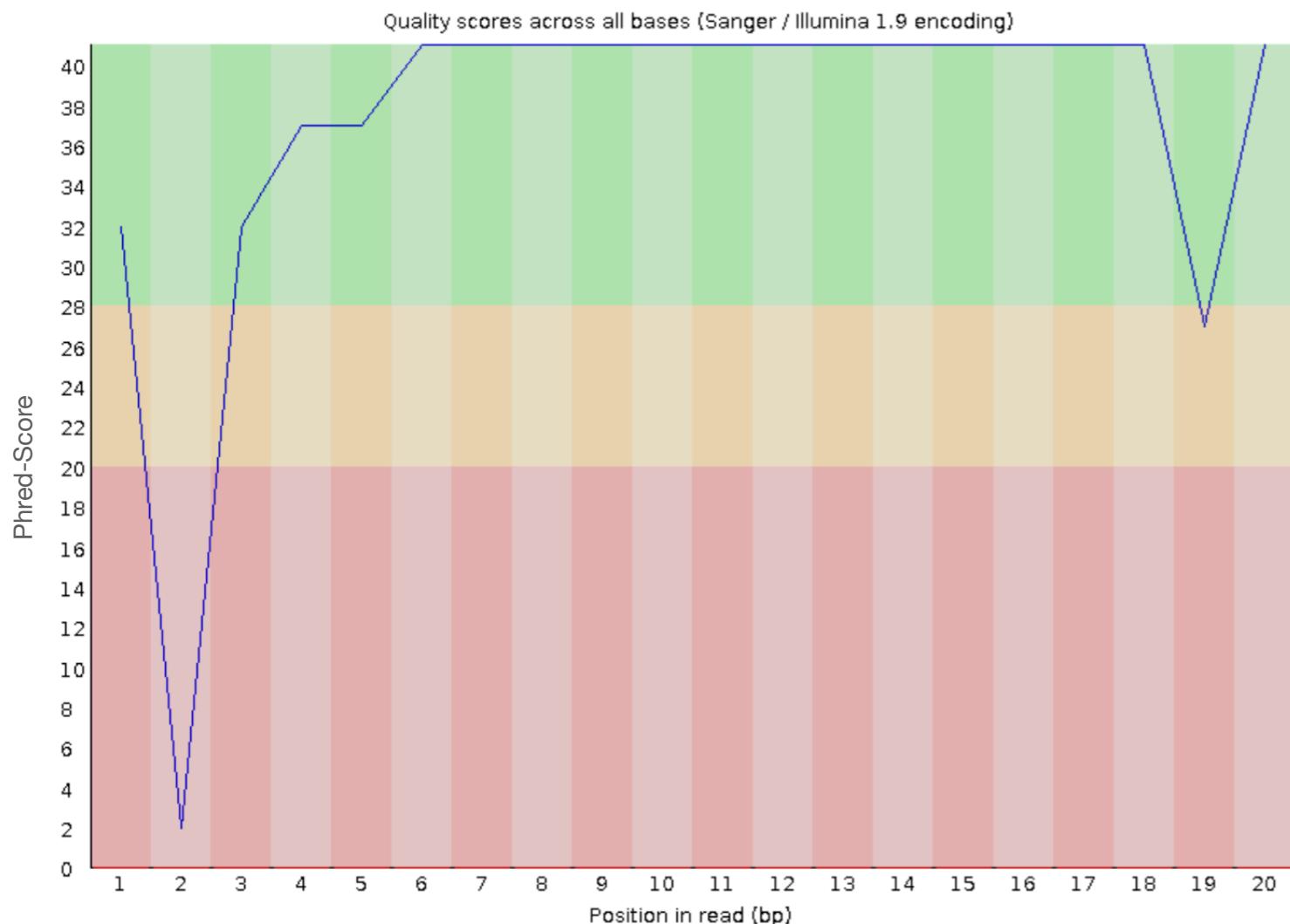


# FASTQ / FASTA

FASTQ →

```
@ID Description
ATTATATATATATCGGGATA
+
A#AFFJJJJJJJJJJJJJJJJ<J
```

- FASTQC Report - Per base sequence quality



$$Q = -10 \log_{10} P$$
$$P = 10^{\frac{-Q}{10}}$$

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

[https://en.wikipedia.org/wiki/Phred\\_quality\\_score](https://en.wikipedia.org/wiki/Phred_quality_score)

# DEMO

# Galaxy - The big problem!



# Galaxy - The big problem!

What are the pro and cons of  
the selected algorithm?

Parameter Settings ?

**Easy to use!**

Apply workflows from others

Data

# Summary

- Galaxy is free to use (open source)
- Galaxy can be installed on your local machine or server infrastructure
- top-performing Galaxy instances are available (e.g. <https://usegalaxy.eu>)
  - Free registration, compute, and storage resources
  - > 2000 CPU cores
  - > 20 TB RAM
  - > 1.5 PB storage
- Galaxy can improve your work in case of
  - ***accessibility***
  - ***reproducibility***
  - ***transparency***

# Hands-On Session

**Get your own account**



<https://usegalaxy.eu>

**Hands-On instructions**



<https://github.com/destairdenbi/training-material/tree/master/trainings/2021-04-12-Online/Day1>

**or**

[https://www.dropbox.com/sh/6gkfwhb3ux3hipl/AAD\\_x7LBRp55u8HNkxa1QhaUa?dl=0](https://www.dropbox.com/sh/6gkfwhb3ux3hipl/AAD_x7LBRp55u8HNkxa1QhaUa?dl=0)

- 1. Uploading Data to Galaxy**
- 2. Build a small Workflow**
- 3. Galaxy's History system**
- 4. Extracting Workflows from Histories**

**THANK  
YOU!**

