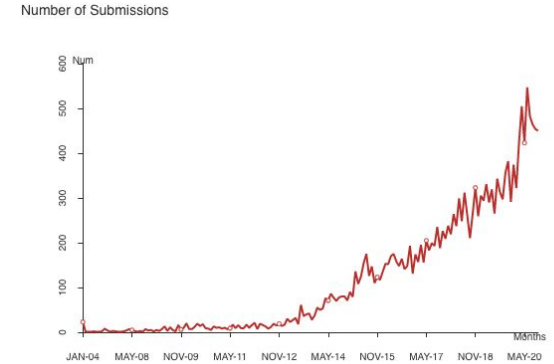
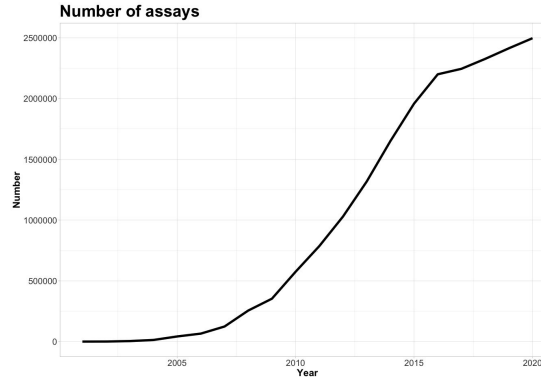
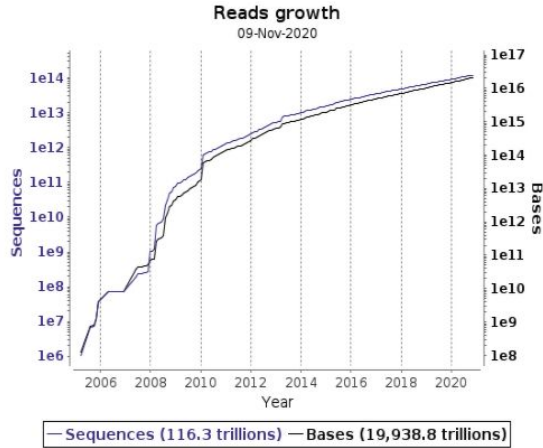


Data brokering

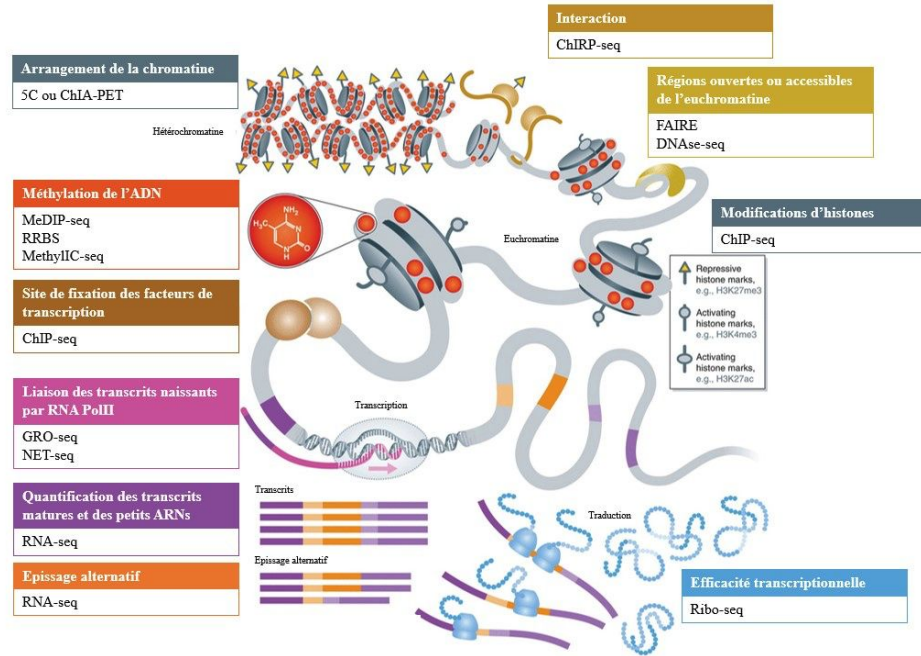
Thomas Denecker, Hélène Chiapello & Jacques van Helden

Contexte

Toujours plus de données...



... de plus en plus hétérogènes ...



(Soon et al. 2013)

... stockées dans des bases de données de plus en plus nombreuses

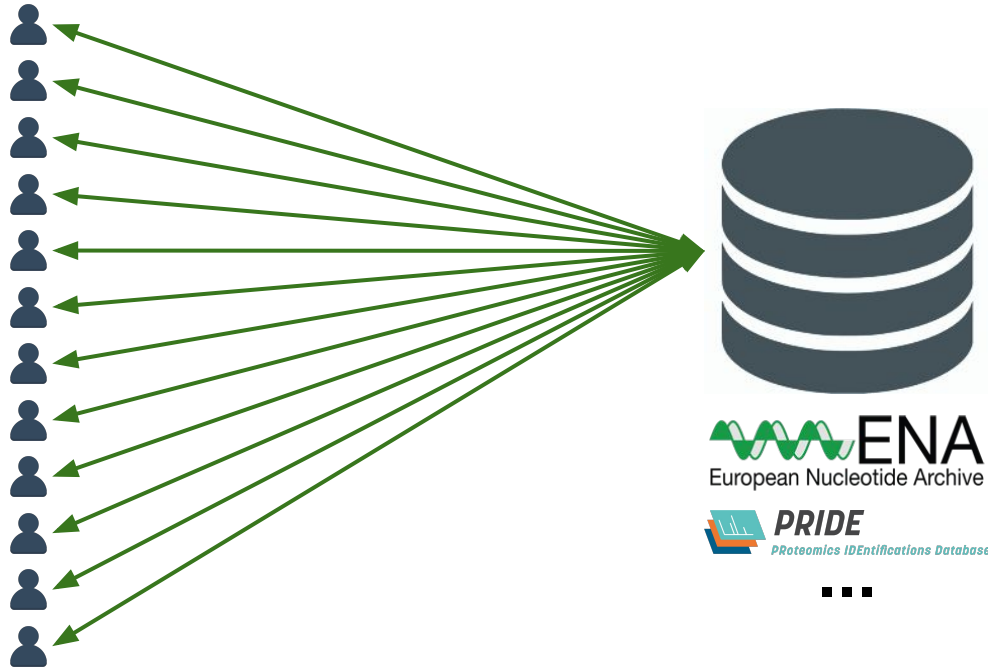


≈ **1700** bases de données
référéncées dans NAR Database en 2018

(Rigden *et al*, 2018)

Soumission

Soumission

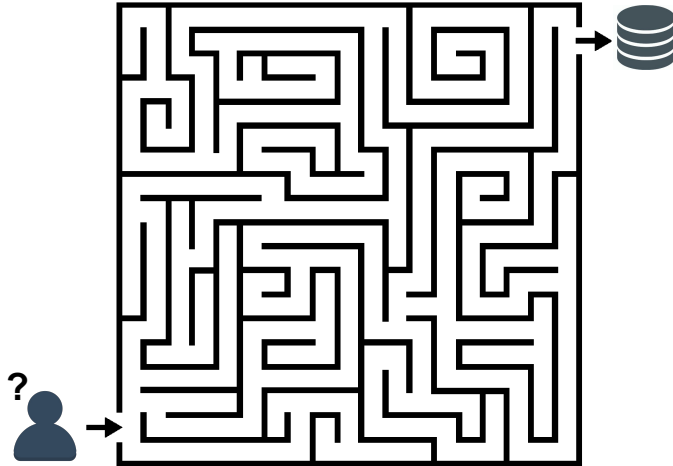


Aujourd'hui

Soumission par l'équipe
expérimentale ou bioinformatique

Allers-retours entre l'utilisateur et
la banque internationale

Avant la soumission



Des questions récurrentes

Dans quelle base de données soumettre ?

Quelles données soumettre ?

Les brutes, les traitées ?

Accompagnées de quelles métadonnées ?

Comment gérer tout ça avant la soumission ?

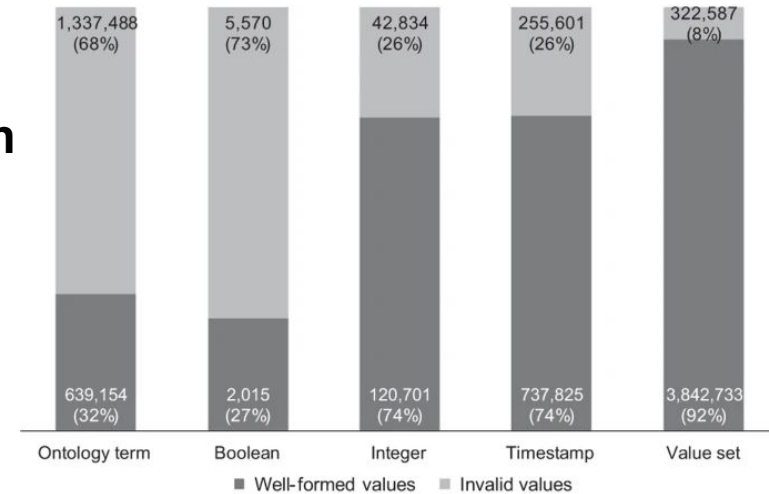
Après soumission

Complexité de la soumission

Hétérogénéité des procédures de soumission

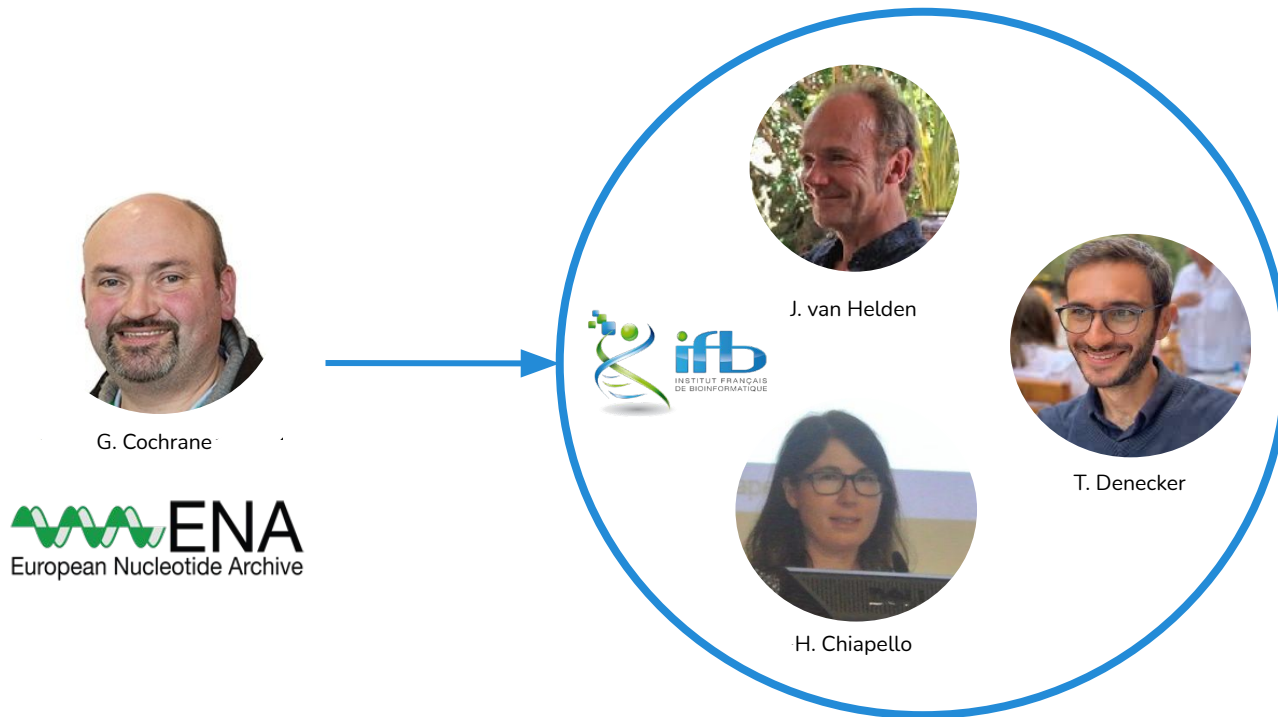
Métadonnées de mauvaise qualité

- Incohérentes
- Incomplètes
- Peu informatives
- Redondantes



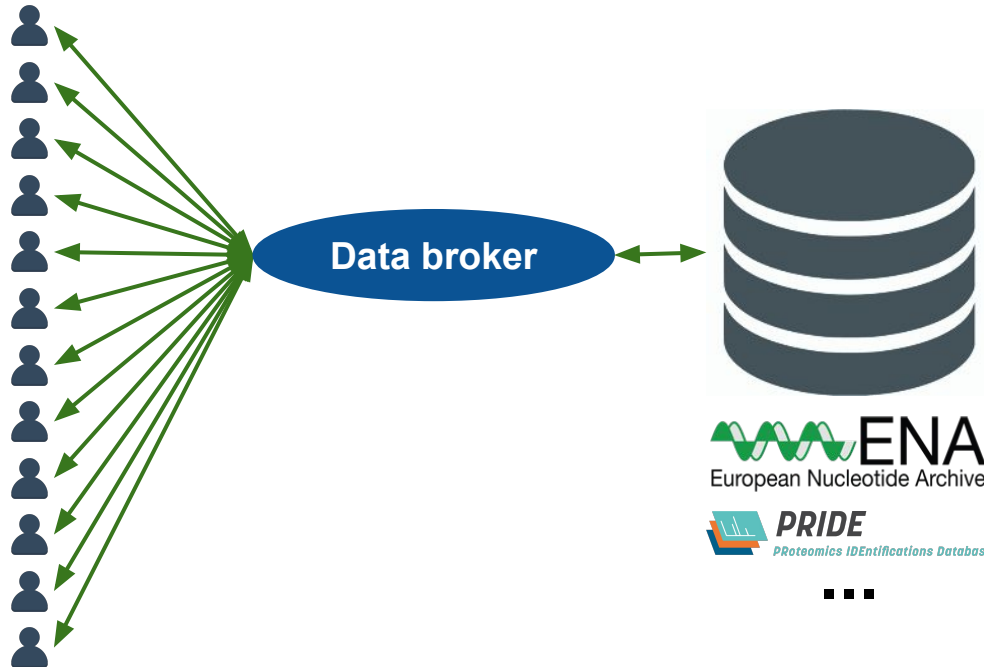
(Gonçalves et al., 2019)

Objectif - Simplifier la soumission des données en France par du *Data brokering*



Data brokering

Un data broker



Intermédiaire entre le producteur de données et la ressource internationale de stockage/archivage

Rationalisation des échanges

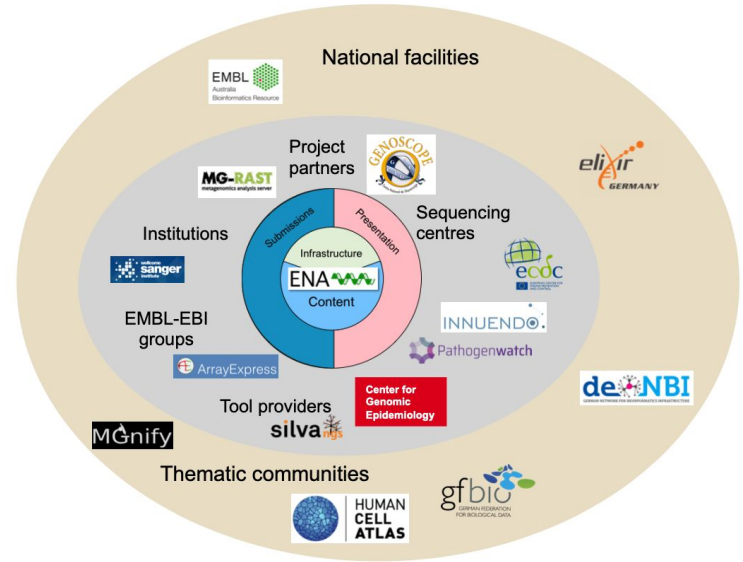
Qui/Que sont les *data broker* ?

Des outils

- METAGENOTE (Quiñones et al. - BMC Bioinformatics 2020)
- gfbio
- ...

Des *national facilities*

- EMBL
- Elixir Germany
- ...



(Cochrane, 2018)

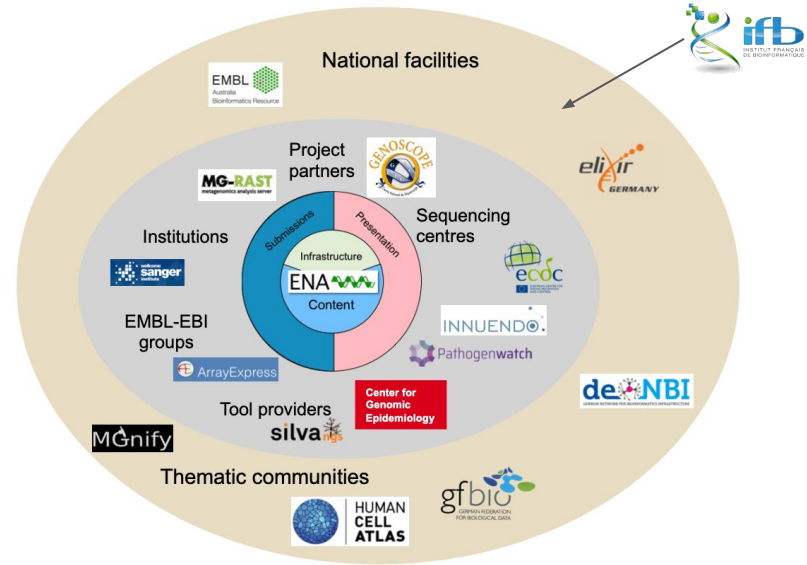
Qui/Que sont les *data broker* ?

Des outils

- METAGENOTE (Quiñones et al. - BMC Bioinformatics 2020)
- gfbio
- ...

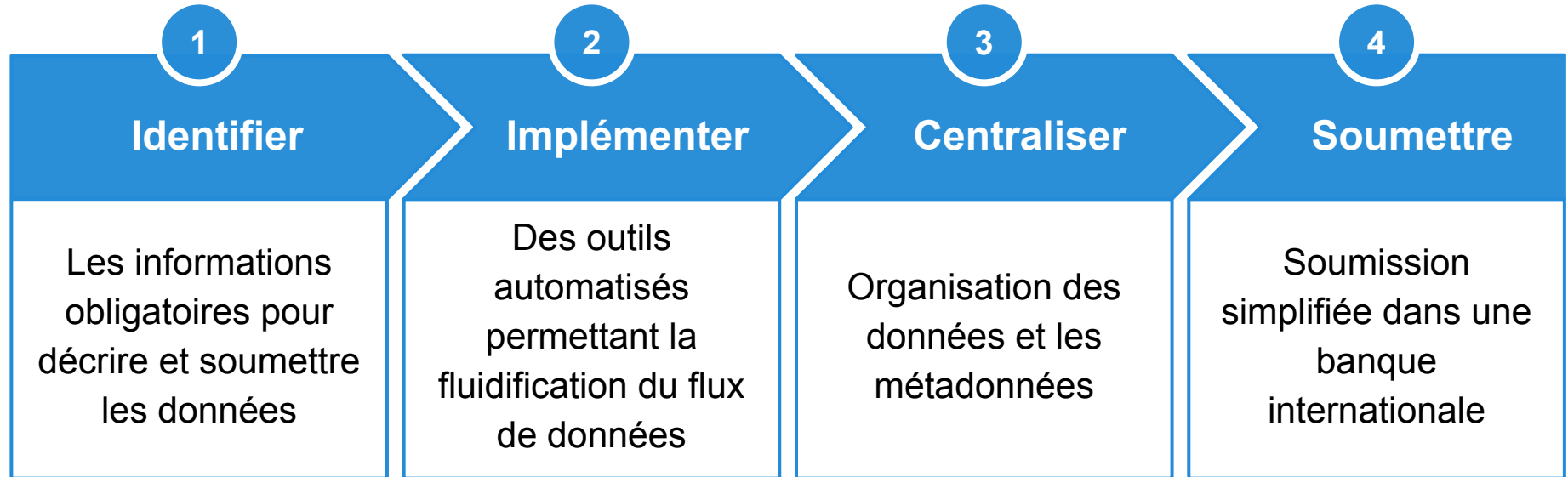
Des *national facilities*

- EMBL
- Elixir Germany
- ...

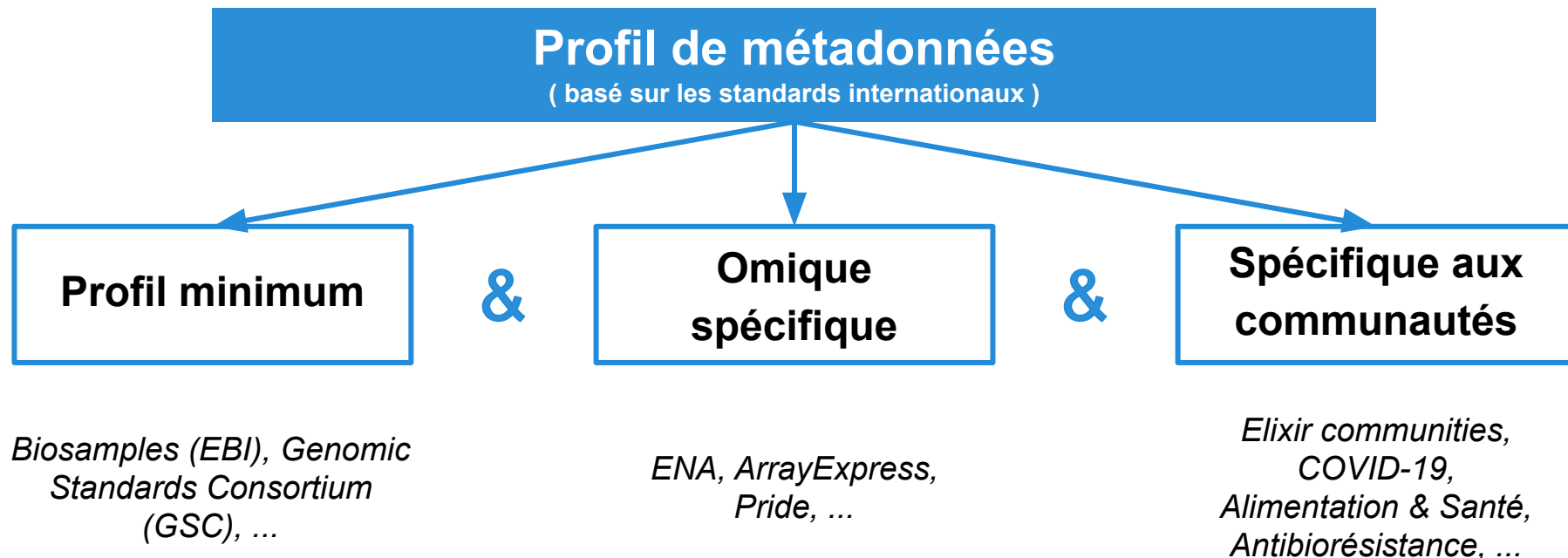


(Cochrane, 2018)

Nos missions pour le data brokering



1. Identifier et collecter les informations obligatoires pour décrire et soumettre les données



2- Implémenter des outils automatisés permettant la fluidification du flux de données

3- Centraliser et organiser les données et les métadonnées

Données

Métadonnées



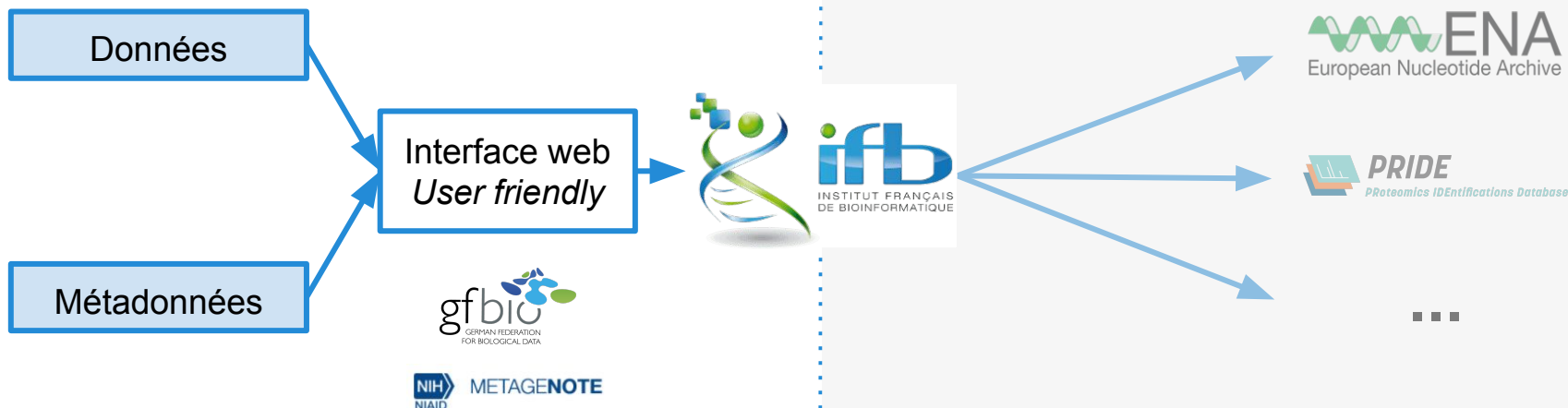
4- Soumettre les données automatiquement et simplement



...

2- Implémenter des outils automatisés permettant la fluidification du flux de données

3- Centraliser et organiser les données et les métadonnées



4- Soumettre les données automatiquement et simplement

2- Implémenter des outils automatisés permettant la fluidification du flux de données

3- Centraliser et organiser les données et les métadonnées



4- Soumettre les données automatiquement et simplement

Mission 1 - Profil de métadonnées

Métadonnées

Les métadonnées sont des « données qui décrivent des données » :

- **Information structurée** associée à un "objet", un document ou un jeu de données ;
- **Documentation** qui permet à l'utilisateur de comprendre, de comparer et d'échanger le contenu du jeu de données décrit.

Un objet sans étiquette n'est connu que de son auteur

Des objets discernés par tous



Métadonnées

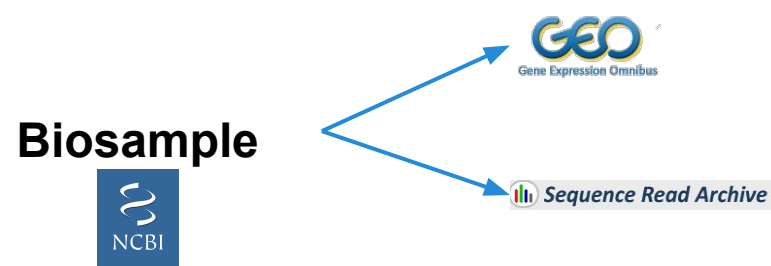
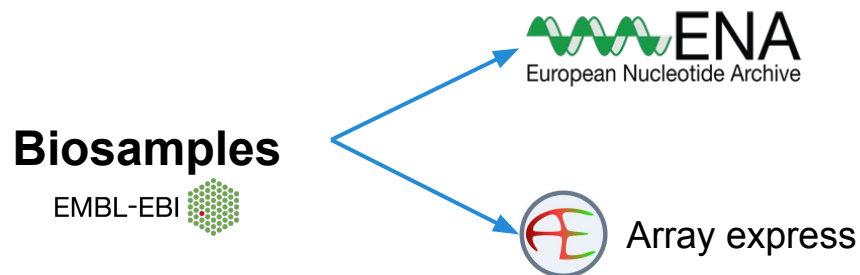


Exemple proposé par
Frédéric de Lamotte
(*data stewardship*)

Des standards



Des bases de données



La qualité actuelle des métadonnées

The variable quality of metadata about biological samples used in biomedical experiments

Rafael S. Gonçalves & Mark A. Musen

Scientific Data volume 6, Article number: 190021 (2019)

11.4 millions de métadonnées testés issues de BioSample (NCBI) et BioSamples (EBI)

Bilan

- Les noms et les valeurs ne sont pas contrôlés ni standardisés
- Les valeurs ne sont pas toujours du bon type (ex : binaire)
- Manque d'outils de validation qui éviteraient des aberrances

Des conseils souvent entendus...

Produire les métadonnées au **moment de la collecte ou de la création des données** plutôt qu'à posteriori.

Les compléter **tout au long du cycle de vie** des données.

... qui entraînent de nombreux questionnements

- Quels métadonnées par rapport aux données ? au projet ?
- Comment les organiser (simplement) ?
- Comment mettre à jour ses métadonnées ?



Création d'un prototype

Nos motivations

1

Identifier

Aider dans le choix des
métadonnées

2

Centraliser

Simplifier la collecte et
l'organisation des
métadonnées

3

Soumettre

Simplifier la soumission
par la génération
automatique des
différents objets
nécessaires

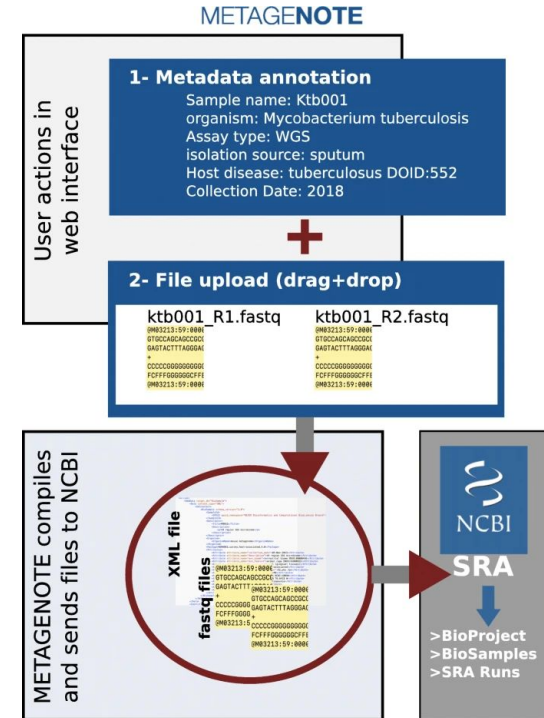
METAGENOTE

Outil de brokering du NCBI qui utilise les *guidelines* du **Genomic Standards Consortium** (GSC) et des **ontologies** (ENVO, FMA, ...)

Interface web + MariaDb

Soumission automatique sur **SRA** via API

(Quiñones et al. - 2020 - <https://doi.org/10.1186/s12859-020-03694-0>)



Avantages et limites de METAGENOTE

Très simple d'utilisation

- Interface web claire
- Aide

Basé sur des standards

Mémoire des projets

- Exportable en excel
- Sauvegarde en ligne

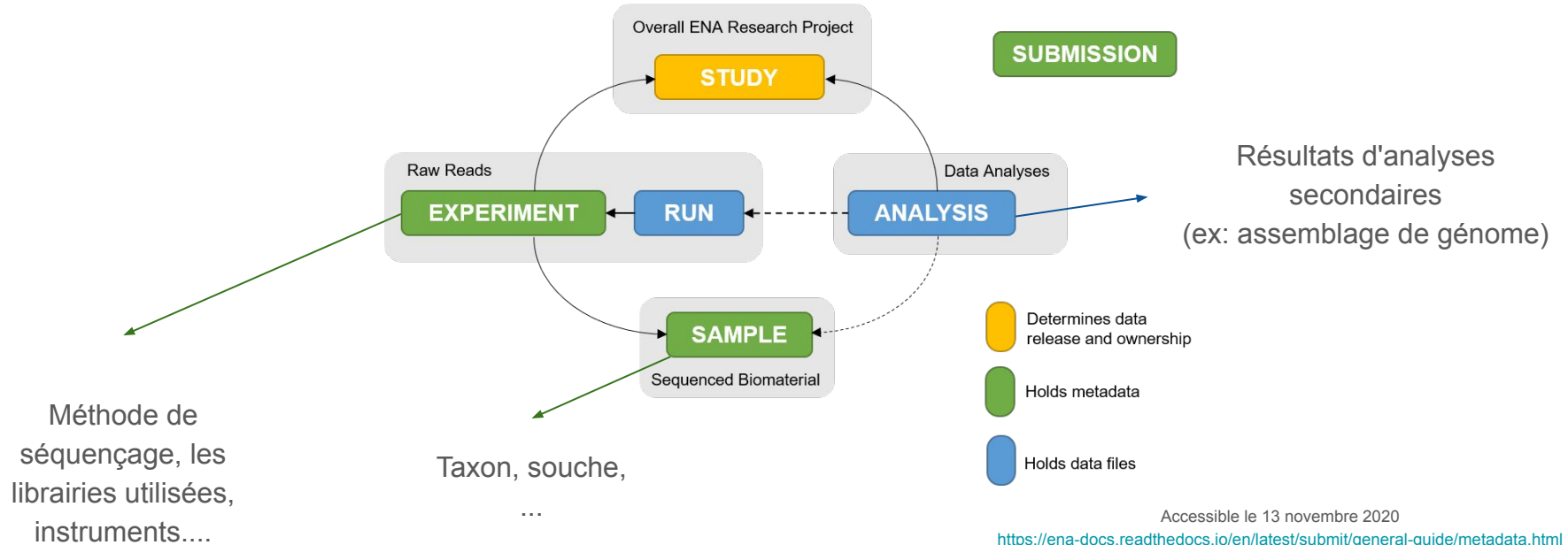
Choix des checklists

- Limité au Genomic Standards Consortium (GSC)
- Pas personnalisable

Soumission uniquement à SRA

Le code n'est pas disponible / modifiable

Modèle des métadonnées à l'ENA



omicsBroker

Prototype fonctionnel (en test)

Le code est disponible sur GitHub

(<https://github.com/IFB-ElixirFr/omicsBroker>)

Plusieurs solutions d'installation :

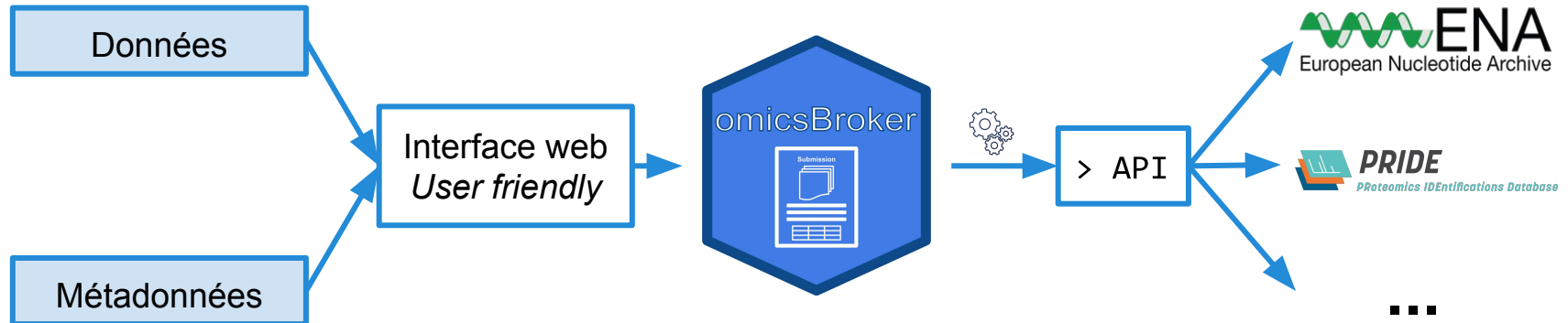
- Package R
- Docker
- shinyProxy

Métadonnées personnalisables

- Checklists de l'ENA
- Manuelle



Où se situe-t-il dans le flux de données?



1- Création d'un projet

omicsBroker

Home

Annotation

Project

Samples

Files

Publish

ENA

Admin

Checklists

Metadata

My projects

More

Session

Help

Project

General

Title *
A short descriptive title for the project.

Alias *
An alias to the project (i.e.: salivarius_dynamic)

Description *
A long description of the scope of the project.

People

Team

Submitter user name

Release date
2020-12-31

Organization

Organization name

Organization role
owner

Organization type
center

Contact

Contact email

Contact first name

Contact last name

DMP - Opidor (experimental)

Import des informations contenues dans un DMP - Opidor (<https://dmp.opidor.fr/>).

Select a PGD - Opidor file

Browse... No file selected

Import PGD

Obligatoire !

2- Choix des métadonnées

omicsBroker

- Home
- Annotation
- Project
- Samples
- Files
- Publish
- ENA
- Admin
- Checklists
- Metadata
- My projects
- More
- Session
- Help

Samples

Database

Select Database: ENA

Filter by mandatory: mandatory

Database description
The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation.

Checklist

Search

	Selected	ACCESSION	LABEL	Mandatory	Recommend	Optional	TYPE	AUTHORITY	DESCRIPTION
	<input type="checkbox"/>	ERC000027	natural or artificial environment ENA Micro B3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Sample	ENA	obtained from the environment where the sample was obtained. By choosing the environmental package, a selection of fields can be made from a relevant subsets of the GSC terms. Minimum information about a Micro B3 sample. A checklist for reporting metadata of marine microbial samples associated with genomics data. NOTE: Non-genomics data, i.e. oceanographic environmental data and morphology-based biodiversity data, should be submitted to the appropriate National Oceanographic Data Centre according to established reporting practices maintained by oceanographic community experts. Major National Oceanographic Data Centres from countries bordering the North-East Atlantic, and its adjacent seas: the Mediterranean, the Black Sea, the Baltic, the North Sea and the Arctic are listed at http://www.seadatanet.org/Overview/Partners . For the Ocean Sampling Day campaign, non-genomics data shall be reported to the PANGAEA (http://www.pangaea.de/submit/).
	<input checked="" type="checkbox"/>	ERC000028	prokaryotic gen minimal sample checklist	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Sample	ENA	Minimum information required for a prokaryotic pathogen sample
	<input type="checkbox"/>		ENA Global Microbial Identifier reporting standard checklist GMI_MDM-1.1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Sample	ENA	Minimum Data for Matching (MDM). A checklist for reporting metadata of pathogen samples for the Global Microbial Identifier (GMI) reporting system. More about GMI can be found here http://www.g-mi.org/
	<input type="checkbox"/>	ERC000030	ENA Tara Oceans	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Sample	ENA	Minimum information about a Tara Oceans sample. A checklist for reporting metadata of oceanic plankton samples associated with genomics data from the Tara Oceans Expedition.
	<input type="checkbox"/>	ERC000031	GSC MixS built environment	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Sample	ENA	Genomic Standards Consortium package extension for reporting of measurements and observations obtained from the environment where the sample was obtained. By choosing the environmental package, a selection of fields can be made from a relevant subsets of the GSC terms.

Metadata table

Excel

	Experience name	Organism	Platform	Instrument	Library layout	Insert size
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

Descriptions

Search

Experience name

Organism

Platform

Instrument

Library layout

Insert size

Library name

Library source

Nominal length

3- Remplissage des métadonnées

Metadata table

Excel

	Experience name	Organism	Platform	Instrument	Library layout	Insert size
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						



Metadata table

Excel

	Experience name	Organism	Platform	Instrument	Library layout	Insert size
1	Sequencing of S. salivarius B35 strain	<i>Streptococcus salivarius</i>	ILLUMINA	Illumina HiSeq 2000	PAIRED	269
2						
3						
4						
5						
6						
7						
8						
9						
10						

Descriptions

Search

Platform

Definition

Platform name. Permitted values : <https://ena-docs.readthedocs.io/en/latest/submit/reads/webin-cl.html#permitted-values-for-platform>

Value

LS454 ; ILLUMINA ; PACBIO_SMRT ; ION_TORRENT ; CAPILLARY ; OXFORD_NANOPORE ; DNBSEQ

Harmonized Name

PLATFORM

* Mandatory

4- Import des données

File Import

Select file(s)

Browse... 2 files

Upload complete

Imported file(s)

	Name	Size (bytes)	md5Server	md5sum	md5sumCheck	Experience name	Mandatory	Recommended	Optional
1	3_R1_001s.fastq.gz	37	1090418039f5338da741538853299c		Fill in an md5sum	0	0	0	
2	3_R2_001s.fastq.gz	37	159719ab322099504d13fbbb37670d		Fill in an md5sum	0	0	0	

4- Import des données

omicsBroker

- Home
- Annotation
- Project
- Samples
- File
- Publish
- ENA
- Admin
 - Checklists
 - Metadata
 - My projects
- More
 - Session
 - Help

File Import

Select file(s)

Upload complete

Imported file(s)

	Name	Size (bytes)	md5Server	md5sumIn	md5sumCheck	Experience name	Mandatory	Recommended	Optional
1	3_R1_001s.fastq.gz	37	1f090418039f533f8da741538853299c	1f090418039f533f8da741538853299c	Check	Sequencing of S. salivarius B35 strain	100	-	0
2	3_R2_001s.fastq.gz	37	159719ab322099504d13fbbba3767cd1	159719ab322099504d13fbbba3767cd1	Check	Sequencing of S. salivarius B35 strain	100	-	0

5- Soumission

omicsBroker
?

- Home
- Annotation
- Project
- Samples
- Files
- Publish
- ENA
- Admin
- Checklists
- Metadata
- My projects
- More
- Session
- Help

Publish in ENA

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed non risus. Suspendisse lectus tortor, dignissim sit amet, adipiscing nec, ultricies sed, dolor. Cras elementum ultrices diam. Maecenas ligula massa, varius a, semper congue, euismod non, mi. Proin porttitor, orci nec nonummy molestie, enim est eleifend mi, non fermentum diam nisi sit amet erat. Duis semper. Duis arcu massa, scelerisque vitae, consequat in, pretium a, enim. Pellentesque congue. Ut in risus volutpat libero pharetra tempor. Cras vestibulum bibendum augue. Praesent blandit odio eu enim. Pellentesque sed dui ut augue blandit sodales. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam nibh. Mauris ac mauris sed pede pellentesque fermentum. Maecenas adipiscing ante non diam sodales hendrerit.

Step 1 - Create ENA account

The first step in making a submission on the ENA is to create a submission account in 'Webin Submissions Portal'. To do so, you will need to create an account at the following address: <https://www.ebi.ac.uk/ena/submit/webin/> Production service or <https://wwwdev.ebi.ac.uk/ena/submit/webin/> Test service .Detailed documentation can be found at : <https://ena-docs.readthedocs.io/en/latest/submit/general-guide/submissions-portal.html>

Step 2 - Fill in your identifiers

Once your account is created, you have a login and a password. Both will be required to submit on the ENA. You must fill them in below

Webin submission account *

Password *

Step 3 - Upload RAW data

It's time to submit DATA on the ENA.

Step 4 - Upload metadata

It's time to submit METADATE on the ENA. A set of checks will be performed to ensure that no information is missing and that the submission is completed correctly. Alerts will be sent to you to guide you in case of problems.

The different files generated

In this section, you can download separately the different files needed for the submission.

6- En test sur l'ENA

Web submissions portal (TEST) Support Logout (Webin-57165)

Sample Report

Shows submitted studies and their release statuses. Search by accession or unique name, or simply click search to show most recent submissions. The results will show the most recently submitted studies to your submission account.

Please click search to see the results.

Search Studies

Accession or Name Release status Search Reset 100 Show unique name

Download of results

Accession	Secondary Accession	Title	Submission date	Release date	Status	Action
PRJNA32313	ERR120131	Characterization of Streptococcus subspecies using gold standards and diversity	31st Dec 2012	31st Dec 2012	Public	🗑️
PRJNA4176	ERR122065	Simultaneous submission	26 Dec 2012		Cancelled	🗑️

EMBL-EBI | [Home](#) | [About](#)

- Services: By topic, By topic (A-Z), Help & Support
- Research: Publications, Research groups, Posters & PreCs
- Training: Train a PhD, Train across EBI, Train online, Contact us
- Industry: Member Area, Workshops, EBI Events, Contact industry programme
- About: Contact us, Search, Jobs, News, People & projects

EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK, +44 (0)1223 80 44 44
Copyright © EMBL 2010 | EMBL-EBI is part of the European Molecular Biology Laboratory | Terms of use

Web submissions portal (TEST) Support Logout (Webin-57165)

Sample Report

Shows submitted samples and their release statuses. Search by accession or unique name, or simply click search to show most recent submissions. The results will show the most recently submitted samples to your submission account.

Please click search to see the results.

Search Samples

Accession or Name Release status Search Reset 100 Show unique name

Download of results

Accession	Biological	Title	Organism	Size	Submission date	Status	Action
ERR106871	SRR1077676	Sequencing of 12 reference H3N2 strains	Streptococcus subspecies	1084	31st Dec 2012	Public	🗑️

EMBL-EBI | [Home](#) | [About](#)

- Services: By topic, By topic (A-Z), Help & Support
- Research: Publications, Research groups, Posters & PreCs
- Training: Train a PhD, Train across EBI, Train online, Contact us
- Industry: Member Area, Workshops, EBI Events, Contact industry programme
- About: Contact us, Search, Jobs, News, People & projects

EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK, +44 (0)1223 80 44 44
Copyright © EMBL 2010 | EMBL-EBI is part of the European Molecular Biology Laboratory | Terms of use

Web submissions portal (TEST) Support Logout (Webin-57165)

Raw Files Report

Shows submitted raw files and their archival statuses. Search by accession or other conditions, or simply click search to show most recent submissions. The results will show the most recently submitted raw files to your submission account.

Please click search to see the results.

Search Raw Files

Accession Archive status Search Reset 100 Show unique name

Download of results

Accession	File name	File format	File size	MIS checksum	Archive status	Action
ERR106871	S_101_S115.fastq.gz	FASTQ	37	193776403226955641390A47576a	File Archived	🗑️
ERR106871	S_101_S115.fastq.gz	FASTQ	37	19364403226955641390A47576a	File Archived	🗑️

EMBL-EBI | [Home](#) | [About](#)

- Services: By topic, By topic (A-Z), Help & Support
- Research: Publications, Research groups, Posters & PreCs
- Training: Train a PhD, Train across EBI, Train online, Contact us
- Industry: Member Area, Workshops, EBI Events, Contact industry programme
- About: Contact us, Search, Jobs, News, People & projects

EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK, +44 (0)1223 80 44 44
Copyright © EMBL 2010 | EMBL-EBI is part of the European Molecular Biology Laboratory | Terms of use

Web submissions portal (TEST) Support Logout (Webin-57165)

Raw Files Report

Shows submitted raw files and their release statuses. Search by accession or unique name, or simply click search to show most recent submissions. The results will show the most recently submitted raw files to your submission account.

Please click search to see the results.

Search Raw Files

Accession Archive status Search Reset 100 Show unique name

Download of results

Accession	Instrument	Study	Sample	Experiment	Submission date	Status	Action
ERR106871	HiSeq 2000	ERR120131	ERR120131	ERR120131	31st Dec 2012	Public	🗑️

EMBL-EBI | [Home](#) | [About](#)

- Services: By topic, By topic (A-Z), Help & Support
- Research: Publications, Research groups, Posters & PreCs
- Training: Train a PhD, Train across EBI, Train online, Contact us
- Industry: Member Area, Workshops, EBI Events, Contact industry programme
- About: Contact us, Search, Jobs, News, People & projects

EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK, +44 (0)1223 80 44 44
Copyright © EMBL 2010 | EMBL-EBI is part of the European Molecular Biology Laboratory | Terms of use

Conclusion & Perspectives

Les actions à venir

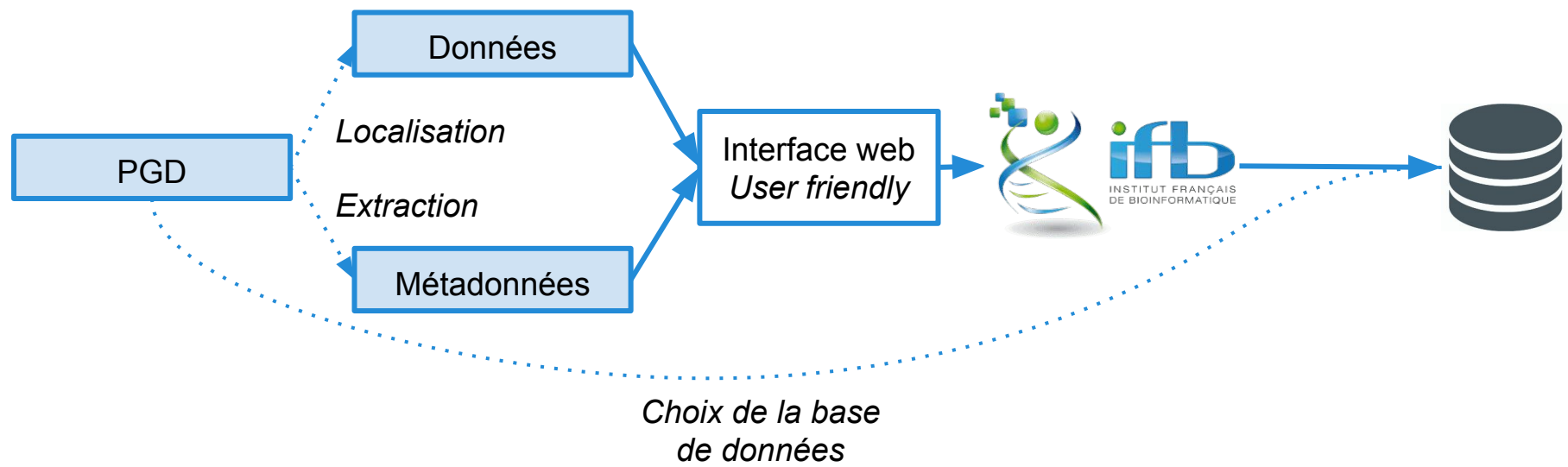
Entre l'utilisateur et le *data broker*

- 1) Établir et collecter un ensemble d'informations nécessaires lors d'une soumission et proposer un **profil de métadonnées**
- 2) Mettre en place un **système *User friendly* de collecte** des métadonnées et des données → **omicsBrokers**

Entre le *data broker* et la base de données

- 3) Mettre en place des **workflows automatiques** de soumission adaptés aux types de données → **omicsBrokers**

Une connexion avec le plan de gestion de données



**Mise à jour de la soumission lors de la mise à jour du PGD
(intégration continue)**

Un réseau en construction



G. Cochrane



C. Pommier



T. Denecker



J. van Helden



H. Chiapello

Data brokering



J. Seiler



F. de Lamotte

PGD



P. Lieby



J.F. Dufayard



*Envie d'en savoir plus sur les métadonnées ? Le FAIR data ?
Les plans de gestion de données ?*

Science ouverte et Plan de Gestion de données

Comment gérer des jeux de données haut-débit en sciences de la vie et de la santé ?

Pour en savoir plus : <https://ifb-elixirfr.github.io/IFB-FAIR-data-training/>

Annexes

Quelques définitions

“Metadata refers to **descriptive information** about the overall study, individual samples, all protocols, and references to processed and raw data file names.”

GEO

“Metadata characterize biological resources by **core information** including a name, a description of its input and its output (parameters or format), its address, and various additional properties.”

Encyclopedia of Database Systems, 2009

“Metadata are the in-depth, **controlled description** of the sample that your sequence was taken from. Essentially, the ‘what, where, how, and when’ of your study from collection to sequence generation, plus contextual data such as environmental conditions (latitude, longitude, temperature) or clinical observations.”

EMBL-EBI

Genomic Standards Consortium (GSC)

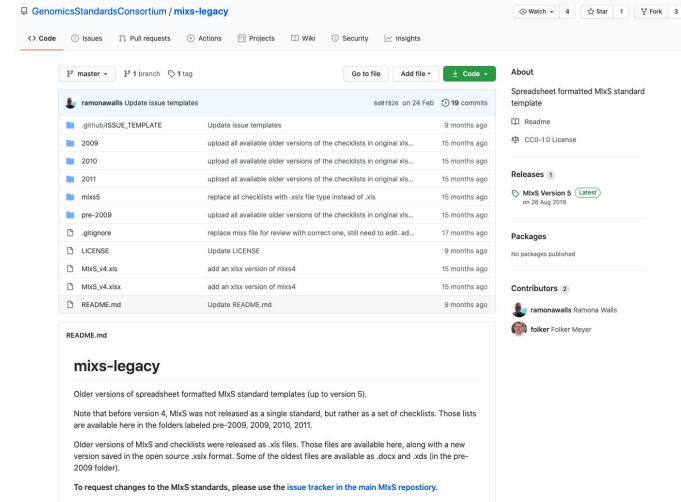


Checklist (obligatoire)

&

Environmental packages (informations complémentaires)

Fichier excel téléchargeable sur GitHub
(dernière MAJ 26 Aug 2019)



Genomic Standards Consortium (GSC)



Checklist - Informations minimales pour les séquences (MIxS)

94 items (\pm obligatoires) décrits dans 22 colonnes :

The MIxS checklist

Column 1 - structured comment name: name of a checklist item as it will appear in GenBank structured comments

Column 2 - item: full name of item as it appears in the publication

Column 3 - definition: a description of the item, including links to ontologies and other resources that can be used to fill in values for the item

Column 4 - expected value: short description and/or expected value of an item

Column 5 - value syntax: the proper syntax for writing the value for a given item

Column 6 - example: examples of values for an item

Column 7 - section: the section of an item

Columns 8 through 18-migs_eu,migs_ba,migs_pl,migs_vi,migs_org,me,mimarks_s,mimarks_c,misag,mimag,miuvig: information about whether an item is mandatory (M), conditional mandatory (C), optional (X), environment-dependent (E) or not applicable (-) for a given checklist type

Column 19 - preferred units: a unit suggestion if a measurement value is given

Column 20 - occurrence: indicates whether a given item can be used only once (1), multiple times (m), or none (0)

Column 21 - position: position of item as it appears in the publication

Column 22 - MIXS ID: a unique of an item

Genomic Standards Consortium (GSC)



Environmental packages - 17 différents

Environment	Structured comment name	Package item	Definition	Expected value	Value syntax	Example	Required	Preferred unit	Occurrence	Position	MIXS ID
air	alt	altitude	Altitude is a term used to identify heights of objects such as airplanes, space shuttles, rockets, atmospheric balloons and heights of places such as atmospheric layers and clouds. It is used to measure the height of an object which is above the earth's surface. In this context, the altitude measurement is the vertical distance between the earth's surface above sea level and the sampled position in the air.	measurement value	[float] [unit]	100 meter	M	meter	1		MIXS:00009 2.4
air	elev	elevation	Elevation of the sampling site is its height above a fixed reference point, most commonly the mean sea level. Elevation is rarely used when referring to points on the earth's surface, while altitude is used for points above the surface, such as an aircraft in flight or a spacecraft in orbit.	measurement value	[float] [unit]	100 meter	C	meter	1		MIXS:00009 0.5
air	barometric_press	barometric pressure	Force per unit area exerted against a surface by the weight of air above that surface.	measurement value	[float] [unit]	5 millibar	X	millibar	1		MIXS:00009 1.7
air	carb_dioxide	carbon dioxide	Carbon dioxide (gas) amount or concentration at the time of sampling.	measurement value	[float] [unit]	410 parts per million	X	micro mole per liter, parts per million	1		MIXS:00009 1.8
air	carb_monoxide	carbon monoxide	Carbon monoxide (gas) amount or concentration at the time of sampling.	measurement value	[float] [unit]	0.1 parts per million	X	micro mole per liter, parts per million	1		MIXS:00009 1.1
air	chem_administration	chemical administration	List of chemical compounds administered to the host or site where sampling occurred, and when (e.g. antibiotics, a herbicide, or herb), can include nucleic compounds. For chemical entities of biological interest ontology (chdbi) (v 103), http://portal.bioontology.org/bioontology/chdbi	CHEBI timestamp	[term:chdbi] [term:chdbi]	ager:2CHEBI:200912016-05-11T20:02Z	X	m	1		MIXS:00075 1.0
air	humidity	humidity	Amount of water vapor in the air, at the time of sampling.	measurement value	[float] [unit]	25 gram per cubic meter	X	gram per cubic meter	1		MIXS:00010 1.1
air	methane	methane	Methane (gas) amount or concentration at the time of sampling.	measurement value	[float] [unit]	1800 parts per billion	X	micro mole per liter, parts per billion, parts per million	1		MIXS:00010 1.2
air	misc_param	miscellaneous parameter	Any other measurement performed or parameter collected, that is not listed here.	parameter name, measurement value	[text:] [float] [unit]	Biomolecules ion concentration:2075 micro mole per kilogram	X	m	1		MIXS:00075 1.3
air	organism_count	organism count	Total cell count of any organism (or group of organisms) per gram, volume or area of sample, should include name of organism followed by count. This method was used for the enumeration (e.g. qPCR, RT-qPCR, etc.) (Should not be provided. (Example: 1000 cells per mL, qPCR).	organism name, measurement value	[text:] [float] [unit]:\$PCRATTYP\$[unit]	total planktonics:5.0e7 cells per milliliter:qPCR	X	number of cells per cubic meter, number of cells per cubic centimeter	1		MIXS:00010 1.4
air	oxygen	Oxygen	Oxygen (gas) amount or concentration at the time of sampling.	measurement value	[float] [unit]	600 parts per million	X	milligram per liter, parts per million	1		MIXS:00075 1.5
air	oxy_sat_satp	oxygenation status of sample	Oxygenation status of sample	enumeration	[enumeration:unavailable:other]	aerobic			1		MIXS:00075 1.6
air	perturbation	perturbation	Type of perturbation, e.g. chemical administration, physical disturbance, etc., associated with perturbation regimen including how many times the perturbation was repeated, how long each perturbation lasted, and the start and end time of the entire perturbation period. Can include multiple perturbation type.	perturbation type name, perturbation interval and duration	[text:] [text:start_time/end_time/duration]	antibiotic: addition:RT2018-05-11T14:30Z2018-05-11T19:30ZP1103M	X		1		MIXS:00075 1.7
air	pollutants	pollutants	Substant types and amount or concentrations measured at the time of sampling; can report multiple pollutants by entering numeric values preceded by name of pollutant.	pollutant name, measurement value	[text:] [float] [unit]	lead:0.15 microgram per cubic meter	X	gram, mole per liter, milligram per liter, microgram per cubic meter	1		MIXS:00010 1.8
air	resp_part_matter	respirable particulate matter	Concentration of substances that remain suspended in the air, and comprise mixtures of organic and inorganic substances (PM10 and PM2.5). Can report multiple PMs by entering numeric values preceded by name of PM.	particulate matter name, measurement value	[text:] [float] [unit]	PM2.5:10 microgram per cubic meter	X	microgram per cubic meter	1		MIXS:00010 1.9
air	samp_safety	sample safety	Safety is the total concentration of all dissolved salts in a liquid or solid (in the form of an extract obtained by centrifugation) sample. While safety can be measured by a complex chemical analysis, this method is difficult and time consuming. More often, it is instead derived from the conductivity measurement. This is known as practical safety. These derivations compare the specific conductance of the sample to a safety standard such as seawater.	measurement value	[float] [unit]	1 milligram per liter	X	milligram per liter, practical safety unit, percentage	1		MIXS:00011 2.0
air	samp_store_dur	sample storage duration	Duration for which the sample was stored	duration	[duration]	P19M			1		MIXS:00011 2.1
air	samp_store_loc	sample storage location	Location at which sample was stored, usually name of a specific freezestorage	location name	[text]	Freezer no.5	X		1		MIXS:00075 2.2
air	samp_store_temp	sample storage temperature	Temperature at which sample was stored, e.g. -80 degrees Celsius	measurement value	[float] [unit]	-80 degree Celsius	X	degree Celsius	1		MIXS:00011 2.3
air	samp_vol_wt_dna_ext	sample volume or weight for DNA extraction	Volume (mL) (mg) of processed sample, or surface area swabbed from sample for DNA extraction	measurement value	[float] [unit]	1500 milliliter	X	milliliter, gram, milligram, square centimeter	1		MIXS:00011 2.4
air	solar_irradiance	solar irradiance	The amount of solar energy that arrives at a specific area of a surface during a specific time interval.	measurement value	[float] [unit]	1.36 kilowatts per square meter per day	X	kilowatts per square meter per day, ergs per square centimeter per second	1		MIXS:00011 2.5
air	temp	temperature	Temperature of the sample at the time of sampling.	measurement value	[float] [unit]	25 degree Celsius	X	degree Celsius	1		MIXS:00011 2.6
air	ventilation_rate	ventilation rate	Ventilation rate of the system in the sampled premises	measurement value	[float] [unit]	750 cubic meter per minute	X	cubic meter per minute, liters per second	1		MIXS:00075 2.7
air	ventilation_type	ventilation type	Ventilation system used in the sampled premises	ventilation type name	[text]	Operable windows	X		1		MIXS:00075 2.8
air	volatile_organic_compounds	volatile organic compounds	Concentration of carbon-based chemicals that easily evaporate at room temperature; can report multiple volatile organic compounds by entering numeric values preceded by name of compound.	volatile organic compound name, measurement value	[text:] [float] [unit]	formaldehyde:500 nanogram per liter	X	microgram per cubic meter, parts per million, nanogram per liter	1		MIXS:00011 2.9
air	wind_direction	wind direction	Wind direction is the direction from which a wind originates	wind direction name	[text]	Northwest	X		1		MIXS:00075 3.0
air	wind_speed	wind speed	Speed of wind measured at the time of sampling	measurement value	[float] [unit]	21 kilometer per hour	X	meter per second, kilometer per hour	1		MIXS:00011 3.1



BioSamples stocke et fournit des descriptions et des métadonnées sur les échantillons biologiques utilisés dans la recherche et le développement par les universités et l'industrie.



Sample

Sample content reference:

Field	Description	Type	Cardinality =====
name	The short name of the sample.	String	Required
release	The date at which the sample was first made public.	Date ISO 8601	Required
update	The date at which the sample was last updated.	Date ISO 8601	System Generated
domain	The AAP domain the sample belongs to.	String	Required
accession	The sample unique identifier in the BioSamples database. If not provided, one will be automatically assigned.	String	Required for [PUT] requests
characteristics	The key-value pairs representing the attributes of the sample.	Object	Optional
externalReferences	A list of links towards external references, such as datasets in other archives.	Array	Optional
relationships	A list of relationships this sample has to other, existing, samples.	Array	Optional
data	A more structured data format to allow submission of tables (eg. antibiogram) in addition to key-value pairs.	Array	Optional, required only for structured data submission using our POST, PUT or PATCH endpoints. Note - You must provide an AAP domain for your data. It can be same as the sample domain if you are the submitter of both the sample metadata and structured data.

https://www.ebi.ac.uk/biosamples/docs/references/api/submit#_submission_minimal_fields

COVID-19

Constat

SCIENCES - MEDECINE

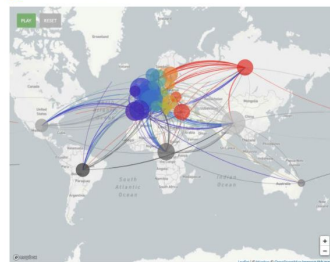
Covid-19 : les chercheurs français peu partageurs des séquences génétiques

La mise en commun massive permet une étude plus précise du virus et de son évolution, mais les scientifiques français y sont réticents.

Par David Larousserie · Publié le 31 août 2020 à 08h30 · Mis à jour le 01 septembre 2020 à 12h47

Lecture 7 min.

Article réservé aux abonnés



Représentation des origines des diverses importations du coronavirus en Europe entre le 7 avril et le 1er juillet 2020, tirée du séquençage de leurs génomes. Image extraite du site Nextstrain

Il n'y a de pire aveugle que celui qui ne veut pas voir. En matière de Covid-19, le dicton s'appliquerait-il à la France ? Notre pays semble en effet peu enclin à utiliser un outil de pointe qui permettrait de répondre à des questions importantes sur l'épidémie, comme

Partage    

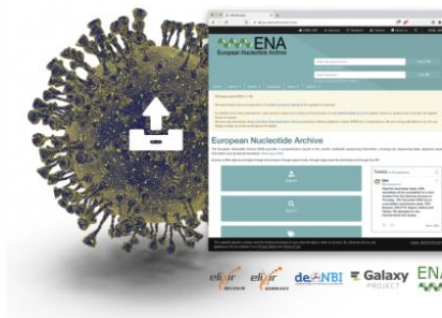
De nouveaux outils pour faciliter la soumission

New tool simplifies the submission of SARS-CoV-2 data to open databases

ELIXIR Belgium and ELIXIR Germany (de.NBI) help researchers share FAIR COVID-19 data

ELIXIR Belgium, in collaboration with ELIXIR Germany and the European COVID-19 Data Platform, have developed a tool to simplify the submission of viral sequencing data to the European Nucleotide Archive (ENA), an ELIXIR Core Data Resource providing open access to nucleotide sequences. The new submission tool offers an easy-to-use interface, guides researchers through the submission process and verifies the data format and description.

Why submit data to ENA?



(Larousserie, Le Monde, 31 août 2020)

(ELIXIR, 17 November 2020)