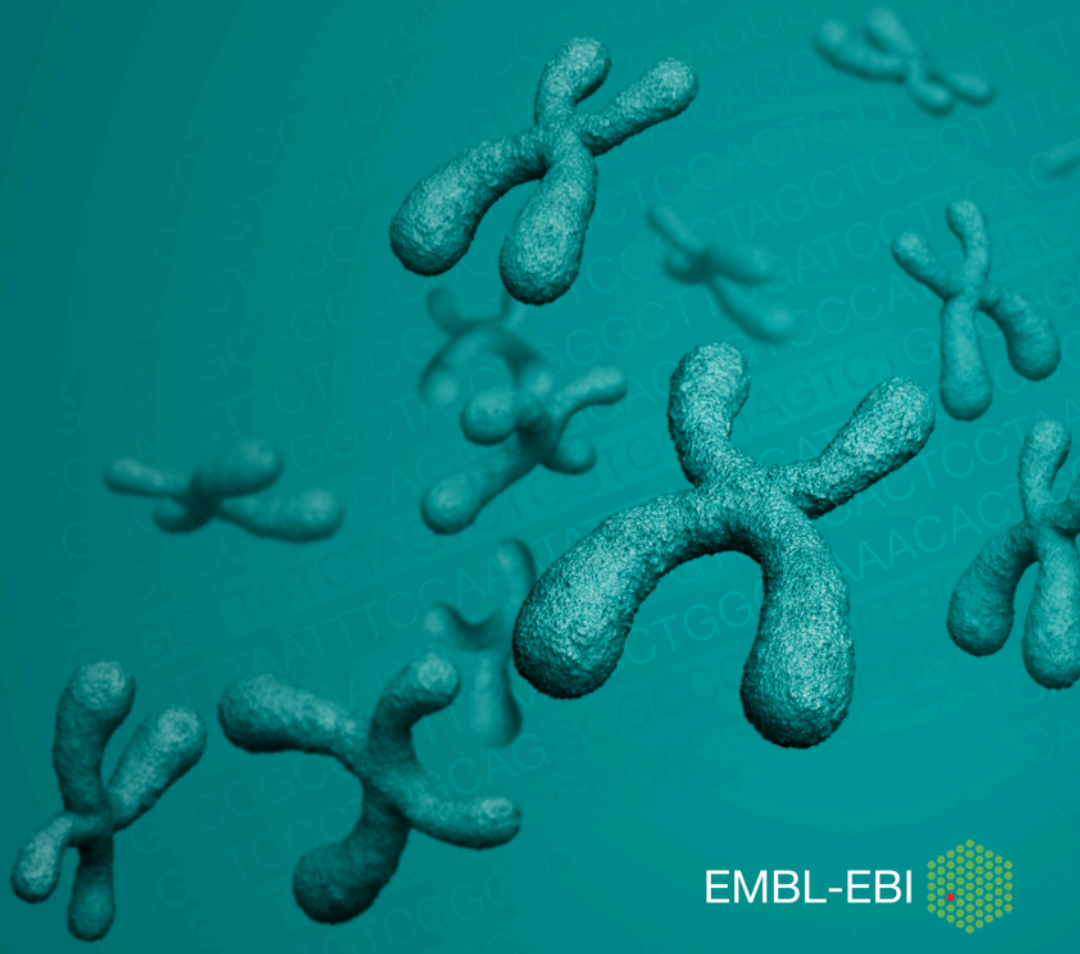
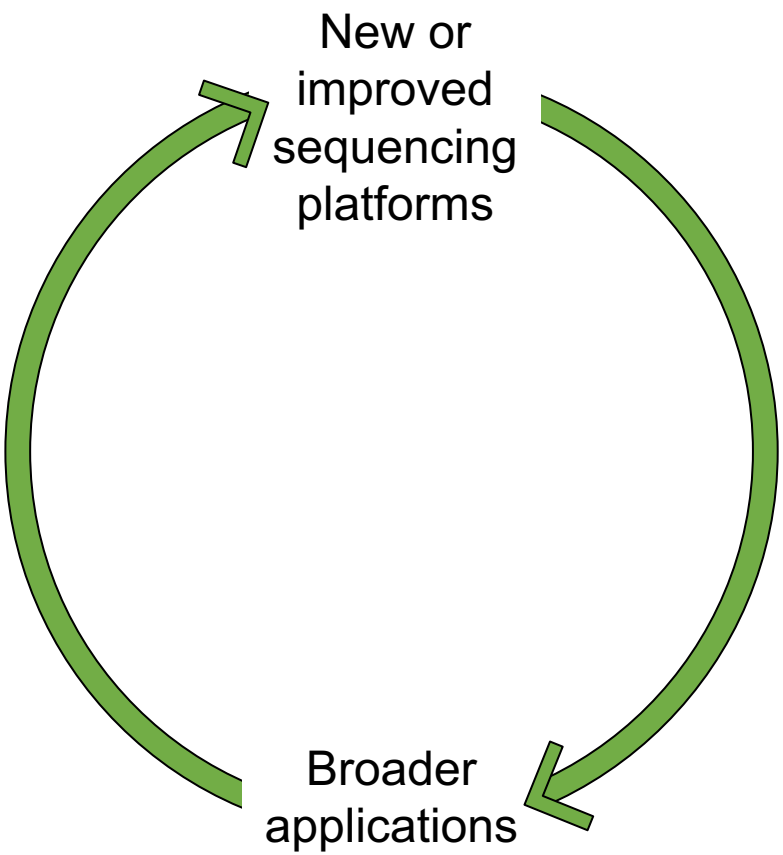
Sequencing communities at scale: dealing with the data

Guy Cochrane



EMBL-EBI



Data growth

Data diversification

Userbase growth

Public environmental sequencing data up to 2002

- Dominant data from small-scale 16S rDNA studies
- Single sequencing platform
- ~35k sequence records
- ~30 kilobase pairs; ~0.5 megabytes

```
ID   AF142985; SV 1; linear; genomic DNA; STD; ENV; 914 BP.
XX
AC   AF142985;
XX
DT   06-JUL-1999 (Rel. 60, Created)
DT   10-MAR-2017 (Rel. 132, Last updated, Version 5)
XX
DE   Uncultured archaeon ORGANIC1_A 16S ribosomal RNA gene, partial sequence.
XX
KW   ENV.
XX
OS   uncultured archaeon ORGANIC1_A
OC   Archaea; Euryarchaeota; Halobacteria; Haloferacales; Halorubraceae;
OC   Halorubrum; environmental samples.
XX
RN   [1]
RP   1-914
RX   DOI; 10.1046/j.1462-2920.2000.00097.x.
RX   PUBMED; 11220308.
RA   Bowman J.P., Rea S.M., McCammon S.A., McMeekin T.A.;
RT   "Diversity and community structure within anoxic sediment from marine
RT   salinity meromictic lakes and a coastal meromictic marine basin, Vestfold
RT   Hilds, Eastern Antarctica";
RL   Environ. Microbiol. 2(2):227-237(2000).
XX
RN   [2]
RP   1-914
RA   Bowman J.P., McCammon S.A., McMeekin T.A.;
RT   ;
RL   Submitted (13-APR-1999) to the INSDC.
RL   School of Agricultural Science, University of Tasmania, GPO Box 252-54,
RL   Hobart, Tasmania 7001, Australia
XX
DR   MD5; 489e036d498b21932c168a5dc2c78a5f.
DR   SILVA-SSU; AF142985.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..914
FT                   /organism="uncultured archaeon ORGANIC1_A"
FT                   /environmental_sample
FT                   /mol_type="genomic DNA"
FT                   /isolation_source="sediment"
FT                   /clone="ORGANIC1_A"
FT                   /db_xref="taxon:98847"
FT   rRNA            <1..>914
FT                   /product="16S ribosomal RNA"
XX
SQ   Sequence 914 BP; 233 A; 209 C; 297 G; 175 T; 0 other;
     attctggttg atcctgccag aggacattgc tattgggatt cgatttagcc atgctagtcg        60
     cacgagttta gactcgtggc agatagctca gtaacacgtg gccaagctac ccttcagacc       120
     agaataacct cgggaaactg aggctaagac tggatacggt atgcacgctg gaatgcagga       180
     tattccaaac gctccggcgc tgaaggatac ggctgcggcc gattaggtag acggtagggt       240
     aacggcccac cgtgccagta atcggtacgg gtcatgagag tgagagcccg gagacggaat       300
     ctgagacaag attccgggcc ctacggggcg cagcaggcgc gaaaccttta cactgcacgc       360
     aagtgcgata ggggaatccc aagttgagag gcatatagcc ctcgcttttg tcgactgtaa       420
     ggaggtcaac gaataagagc tgggcaagac cggtgccagc ggcggcggta ataccggcag       480
     ctcgagtgat gtcgtatttt attgggccta aagcgtcggt agctggccag cgaagttcgt       540
     cgggaaatcc acttgcccaa caggtgggcg tccggcgaaa actgtccggc ttggaaccgg       600
     aaggctcaga gggtacgtct ggggtaggag tgaaatcctg taatcccgga cggacgaccg       660
     atggcgaaag cagtctgaga ggacggatcc gacagtgagg gacgaaagct agggtctcga       720
     accggattag atacccgggt agtcctagct gtaaaggatg ctcgttagat gtggcaccca       780
     ctacgagtgg gtgctgtgtc gtagggaaga agctaaacga gccgcctggg aagtacgtcc       840
     gcaaggatga aacttaaagg aattggcggg ggagcactac aaccggagga gcatgcggtt       900
     taattggatt caac                                                        914
//
```

Public environmental sequencing data in 2017

- 16,000 environmental sequencing studies
- 600,000 metagenomics samples
- diverse data
  - raw data
    - 454, Illumina, PacBio, Ion Torrent, ONT, etc.
    - sequenced libraries (600,000)
    - 30 terabase pairs, ~0.5 petabytes
  - shotgun metagenomics
  - metatranscriptomics
  - metagenome/metatranscriptome assemblies (1,100)
  - single cell genomes
  - reference data (e.g. 100,000 bacterial assemblies)
  - identification data (243)
  - gene catalogue (1)

Context

EMBL European Bioinformatics Institute

Genes, genomes & variation

European Nucleotide Archive
1000 Genomes
Ensembl
Ensembl Genomes
European Genome-phenome Archive
Metagenomics portal

Gene, protein & metabolite expression

ArrayExpress
Expression Atlas
Metabolights
PRIDE

Literature & ontologies

Europe PubMed Central
Gene Ontology
Experimental Factor Ontology

Protein sequences, families & motifs

InterPro
Pfam
UniProt

Molecular structures

Protein Data Bank in Europe
Electron Microscopy Data Bank

Chemical biology

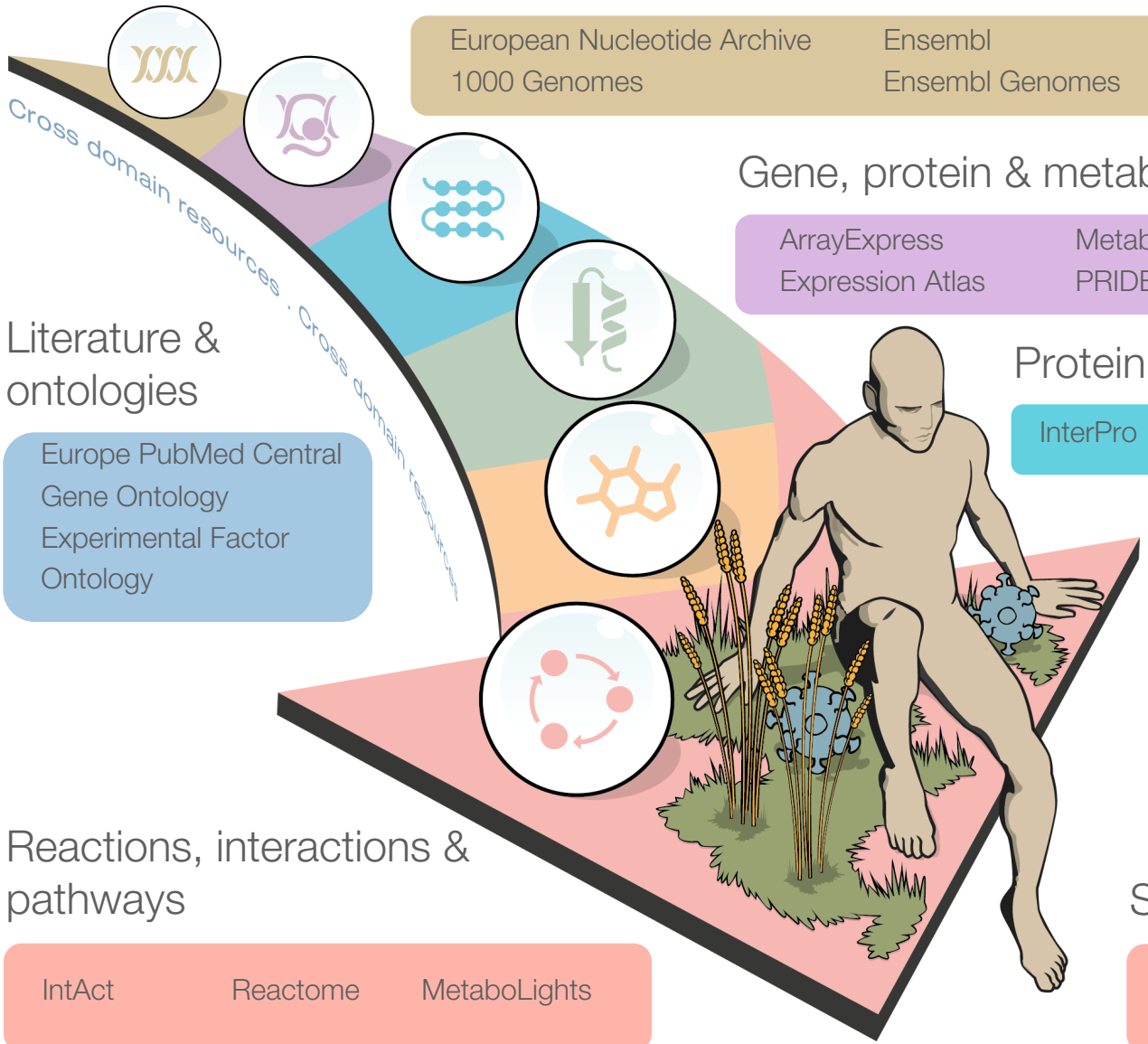
ChEMBL
ChEBI

Reactions, interactions & pathways

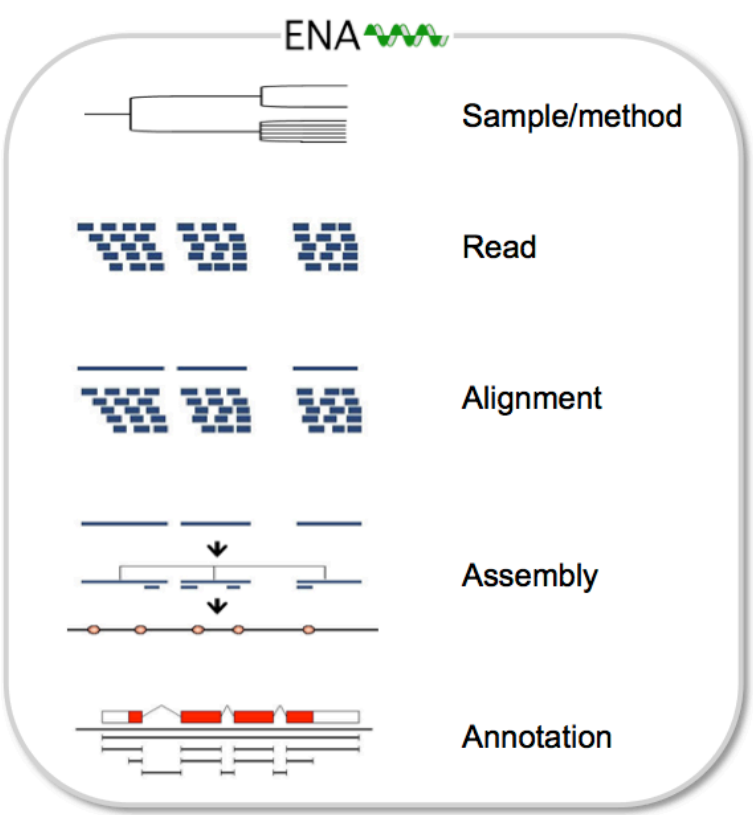
IntAct
Reactome
MetaboLights

Systems

BioModels
BioSamples
Enzyme Portal



European Nucleotide Archive (ENA)

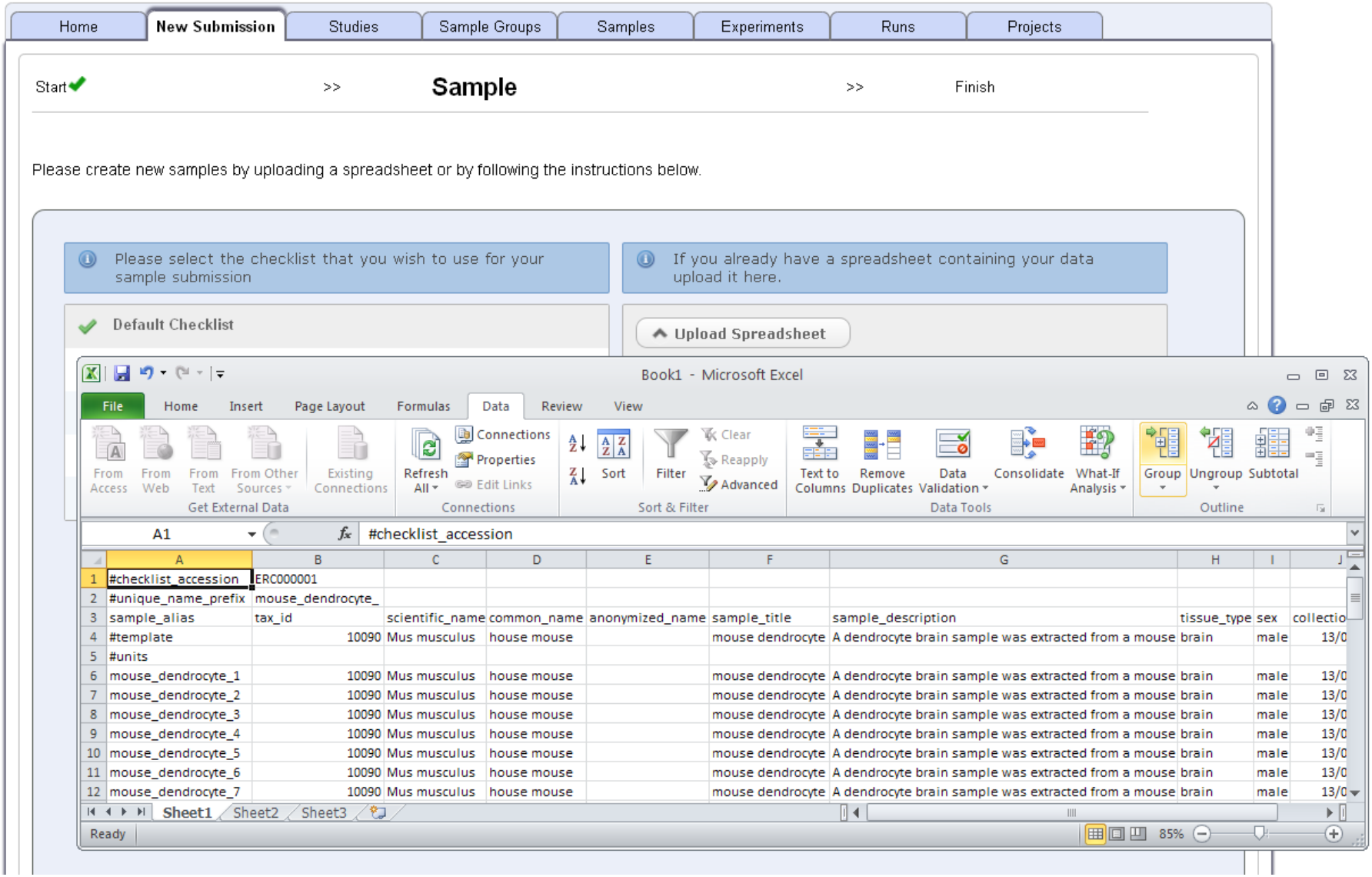


ENA

http://www.ebi.ac.uk/ena/

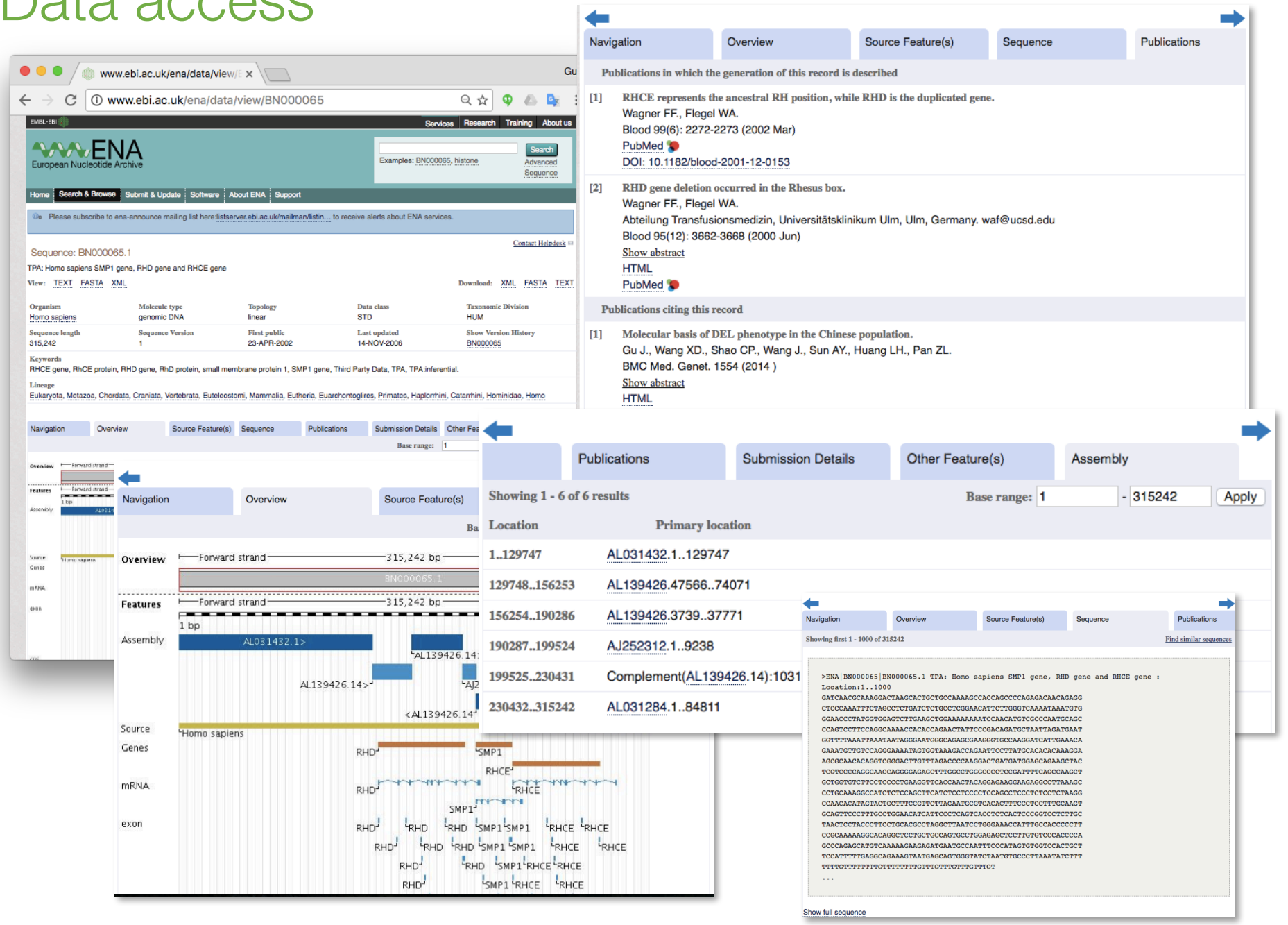
- **Globally comprehensive scientific record** and European node of **INSDC**
- A **broad platform** for the management, sharing, integration and dissemination of sequence data
- Established in the early 1980s, extended for **new technologies and applications**
- **Connectivity** with broader EMBL-EBI resources
- Sequence data **foundation**
- **Sustained** within EMBL-EBI under EMBL funding with additional support from EC, UK Research councils, Wellcome Trust, etc.
- **Substantial scale**: 1 submission every 6 minutes, 1.3 petabase pairs across 1.5 million taxa, 2,000-5,000 active data providers, global consumer userbase
- Rich submission, discovery and retrieval **software, tools and services**

Data validation and submissions

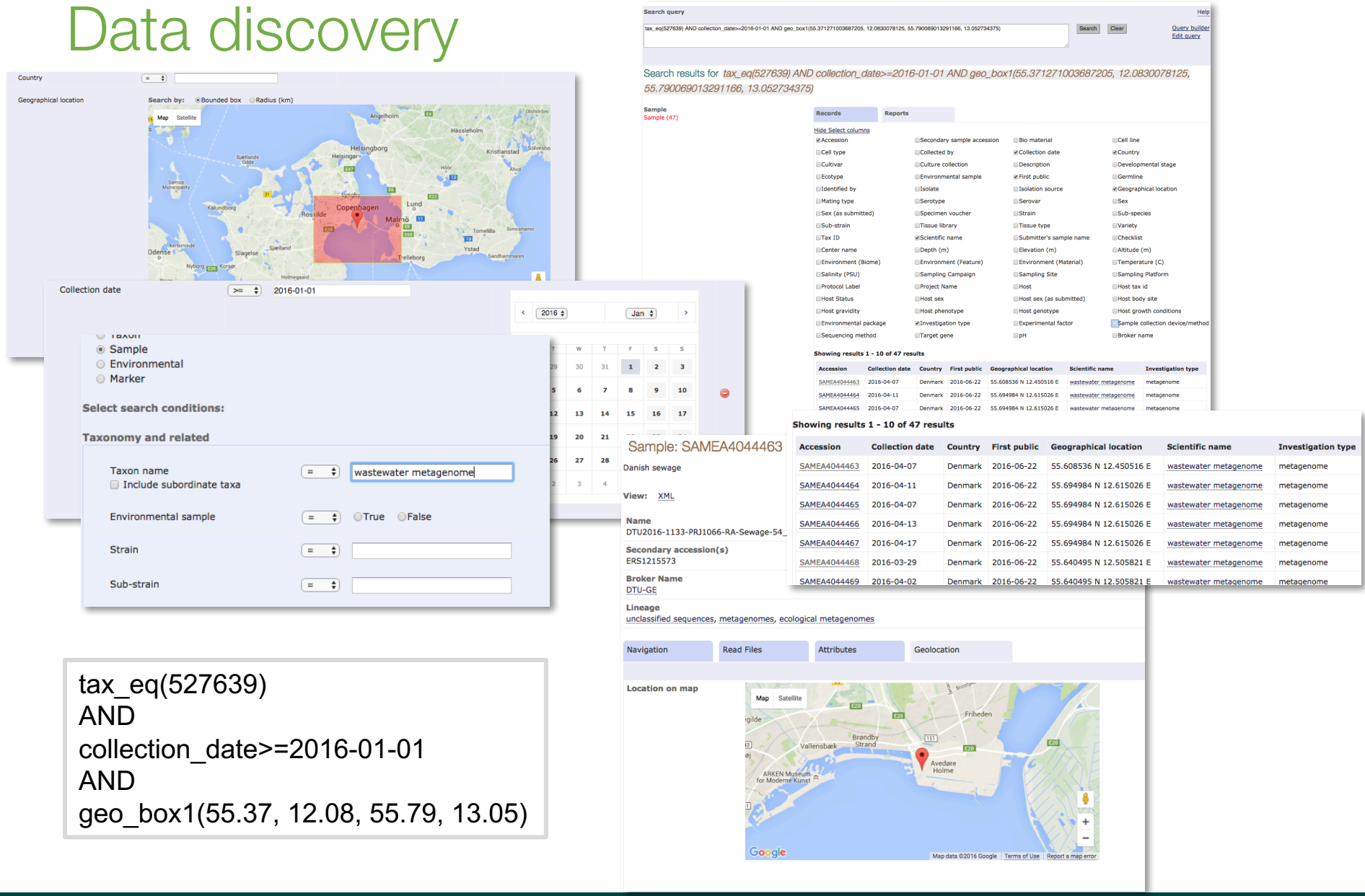


	A	B	C	D	E	F	G	H	I	J
1	#checklist_accession	ERC000001								
2	#unique_name_prefix	mouse_dendrocyte_								
3	sample_alias	tax_id	scientific_name	common_name	anonymized_name	sample_title	sample_description	tissue_type	sex	collectio
4	#template	10090	Mus musculus	house mouse		mouse dendrocyte	A dendrocyte brain sample was extracted from a mouse	brain	male	13/0
5	#units									
6	mouse_dendrocyte_1	10090	Mus musculus	house mouse		mouse dendrocyte	A dendrocyte brain sample was extracted from a mouse	brain	male	13/0
7	mouse_dendrocyte_2	10090	Mus musculus	house mouse		mouse dendrocyte	A dendrocyte brain sample was extracted from a mouse	brain	male	13/0
8	mouse_dendrocyte_3	10090	Mus musculus	house mouse		mouse dendrocyte	A dendrocyte brain sample was extracted from a mouse	brain	male	13/0
9	mouse_dendrocyte_4	10090	Mus musculus	house mouse		mouse dendrocyte	A dendrocyte brain sample was extracted from a mouse	brain	male	13/0
10	mouse_dendrocyte_5	10090	Mus musculus	house mouse		mouse dendrocyte	A dendrocyte brain sample was extracted from a mouse	brain	male	13/0
11	mouse_dendrocyte_6	10090	Mus musculus	house mouse		mouse dendrocyte	A dendrocyte brain sample was extracted from a mouse	brain	male	13/0
12	mouse_dendrocyte_7	10090	Mus musculus	house mouse		mouse dendrocyte	A dendrocyte brain sample was extracted from a mouse	brain	male	13/0

Data access


www.ebi.ac.uk/ena/data/view/BN000065
EMBL-EBI
Services
Research
Training
About us
ENA
European Nucleotide Archive
Search
Examples: BN000065, histone
Advanced Sequence
Home
Search & Browse
Submit & Update
Software
About ENA
Support
Please subscribe to ena-announce mailing list here:listserver.ebi.ac.uk/mailman/listin... to receive alerts about ENA services.
Contact Helpdesk
Sequence: BN000065.1
TPA: Homo sapiens SMP1 gene, RHD gene and RHCE gene
View: TEXT FASTA XML
Download: XML FASTA TEXT
Organism
Homo sapiens
Molecule type
genomic DNA
Topology
linear
Data class
STD
Taxonomic Division
HUM
Sequence length
315,242
Sequence Version
1
First public
23-APR-2002
Last updated
14-NOV-2006
Show Version History
BN000065
Keywords
RHCE gene, RhCE protein, RHD gene, RhD protein, small membrane protein 1, SMP1 gene, Third Party Data, TPA, TPA:inferential.
Lineage
Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia, Eutheria, Euarchontoglires, Primates, Haplorrhini, Catarrhini, Hominidae, Homo
Navigation
Overview
Source Feature(s)
Sequence
Publications
Publications in which the generation of this record is described
[1] RHCE represents the ancestral RH position, while RHD is the duplicated gene.
Wagner FF., Flegel WA.
Blood 99(6): 2272-2273 (2002 Mar)
PubMed
DOI: 10.1182/blood-2001-12-0153
[2] RHD gene deletion occurred in the Rhesus box.
Wagner FF., Flegel WA.
Abteilung Transfusionsmedizin, Universitätsklinikum Ulm, Ulm, Germany. waf@ucsd.edu
Blood 95(12): 3662-3668 (2000 Jun)
Show abstract
HTML
PubMed
Publications citing this record
[1] Molecular basis of DEL phenotype in the Chinese population.
Gu J., Wang XD., Shao CP., Wang J., Sun AY., Huang LH., Pan ZL.
BMC Med. Genet. 1554 (2014)
Show abstract
HTML
Submission Details
Other Feature(s)
Assembly
Base range: 1 - 315242 Apply
Showing 1 - 6 of 6 results
Location
Primary location
1..129747
AL031432.1..129747
129748..156253
AL139426.47566..74071
156254..190286
AL139426.3739..37771
190287..199524
AJ252312.1..9238
199525..230431
Complement(AL139426.14):1031
230432..315242
AL031284.1..84811
Overview
Forward strand
315,242 bp
BN000065.1
Features
1 bp
Assembly
AL031432.1>
AL139426.14>
<AL139426.14
Source
Homo sapiens
Genes
RHD
SMP1
RHCE
mRNA
exon
Showing first 1 - 1000 of 315242
Find similar sequences
>ENA|BN000065|BN000065.1 TPA: Homo sapiens SMP1 gene, RHD gene and RHCE gene :
Location:1..1000
GATCAACGCAAAGGACTAAGCACTGCTGCCAAAAGCCACCAGCCCCAGAGACAACAGAGG
CTCCCAAATTTCTAGCCTCTGATCTCTGCCTCGGAACATTCTTGGGTCAAAATAAATGTG
GGAACCCTATGGTGGAGTCTTGAAGCTGGAAAAAAAATCCAACATGTCGCCCAATGCAGC
CCAGTCCTTCCAGGCAAAACCACACCAGAACTATTCCCGACAGATGCTAATTAGATGAAT
GGTTTTAAATTAAATAATAGGGAATGGGCAGAGCGAAGGGTGCCAAGGATCATTGAAACA
GAAATGTTGTCCAGGGAAAATAGTGGTAAAGACCAGAATTCCTTATGCACACACAAAGGA
AGCGCAACACAGGTCGGGACTTGTTTAGACCCCAAGGACTGATGATGGAGCAGAAGCTAC
TCGTCCCCAGGCAACCAGGGGAGAGCTTTGGCCTGGGCCCCTCCGATTTTCAGCCAAGCT
GCTGGTGTCTTCCTCCCCTGAAGGTTCACCAACTACAGGAGAAGGAAGAGGCCTTAAAGC
CCTGCAAAGGCCATCTCTCCAGCTTCATCTCCTCCCCCTCCAGCCTCCCTCTCCTCTAAGG
CCAACACATAGTACTGCTTTCCGTTCTTAGAATGCGTCACACTTTCCCTCCTTTGCAAGT
GCAGTTCCCTTTGCCTGGAACATCATTCCCTCAGTCACCTCTCACTCCCGGTCCTCTTGC
TAACTCCTACCCTTCCTGCACGCCTAGGCTTAATCCTGGGAAACCATTTGCCACCCCCTT
CCGCAAAAAGGCACAGGCTCCTGCTGCCAGTGCCTGGAGAGCTCCTTGTGTCCCACCCCA
GCCCAGAGCATGTCAAAAAGAAGAGATGAATGCCAATTTCCCATAGTGTGGTCCACTGCT
TCCATTTTTGAGGCAGAAAGTAATGAGCAGTGGGTATCTAATGTGCCCTTAAATATCTTT
TTTTGTTTTTTTTGTTTTTTTTGTTTGTTTGTTTGTTGT
...
Show full sequence

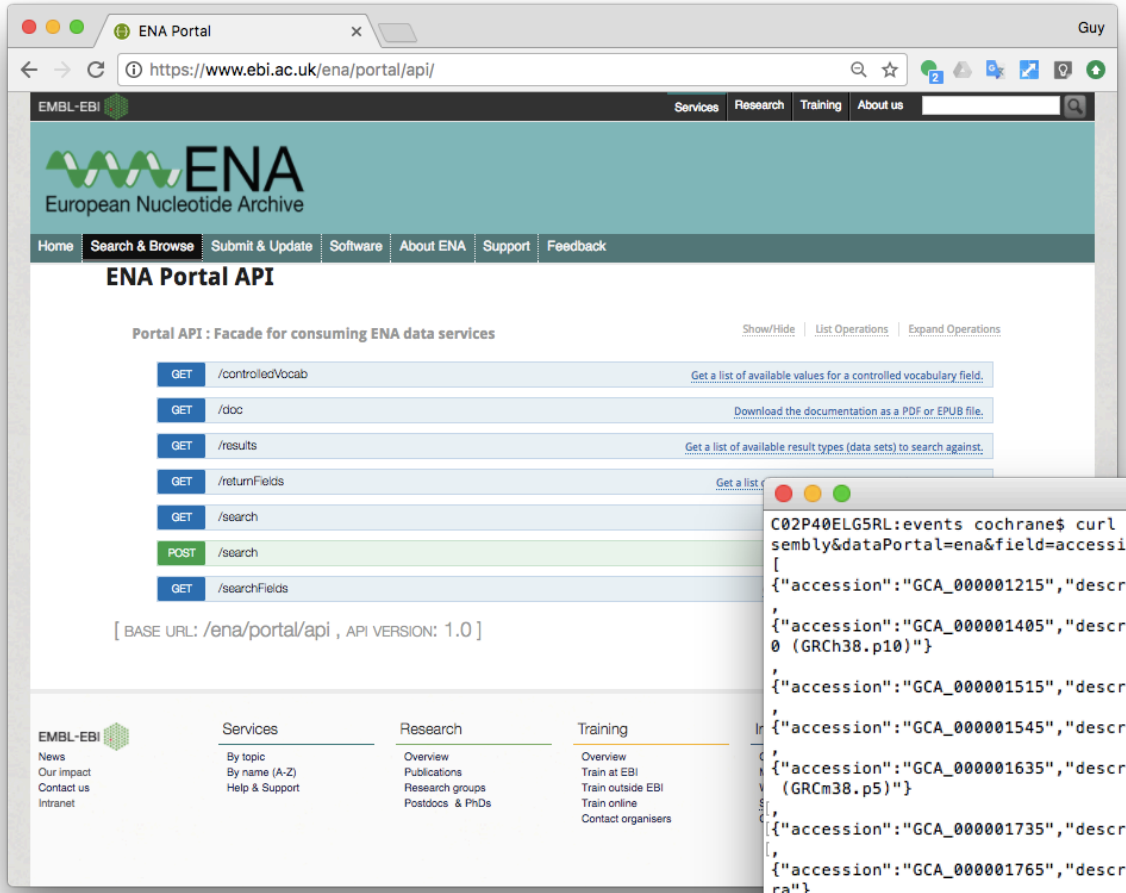
Data discovery

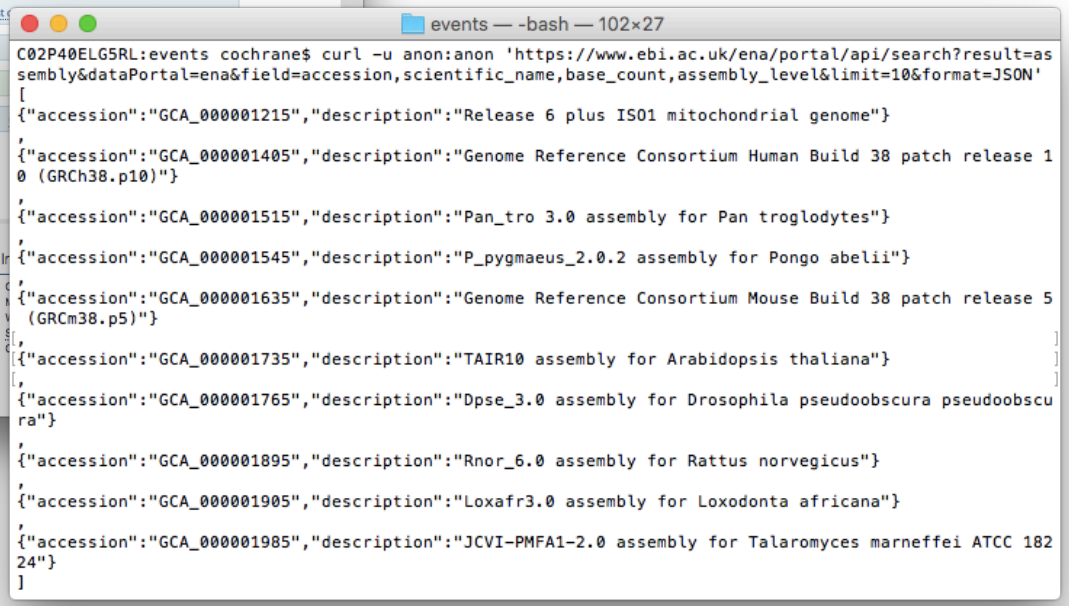


tax_eq(527639)
AND
collection_date>=2016-01-01
AND
geo_box1(55.37, 12.08, 55.79, 13.05)

EMBL-EBI

Programmatic access





- ENA discovery API
  - RESTful
  - Entire ENA content
  - Logical searches across data types
  - TSV/JSON

https://www.ebi.ac.uk/ena/portal/api

Public environmental sequencing data in 2017

- 16,000 environmental sequencing studies
- 600,000 metagenomics samples
- diverse data
  - raw data
    - 454, Illumina, PacBio, Ion Torrent, ONT, etc.
    - sequenced libraries (600,000)
    - 30 terabase pairs, ~0.5 petabytes
  - shotgun metagenomics
  - metatranscriptomics
  - metagenome/metatranscriptome assemblies (1,100)
  - single cell genomes
  - reference data (e.g. 100,000 bacterial assemblies)
  - identification data (243)
  - gene catalogue (1)

ENA as data management platform for metagenomics

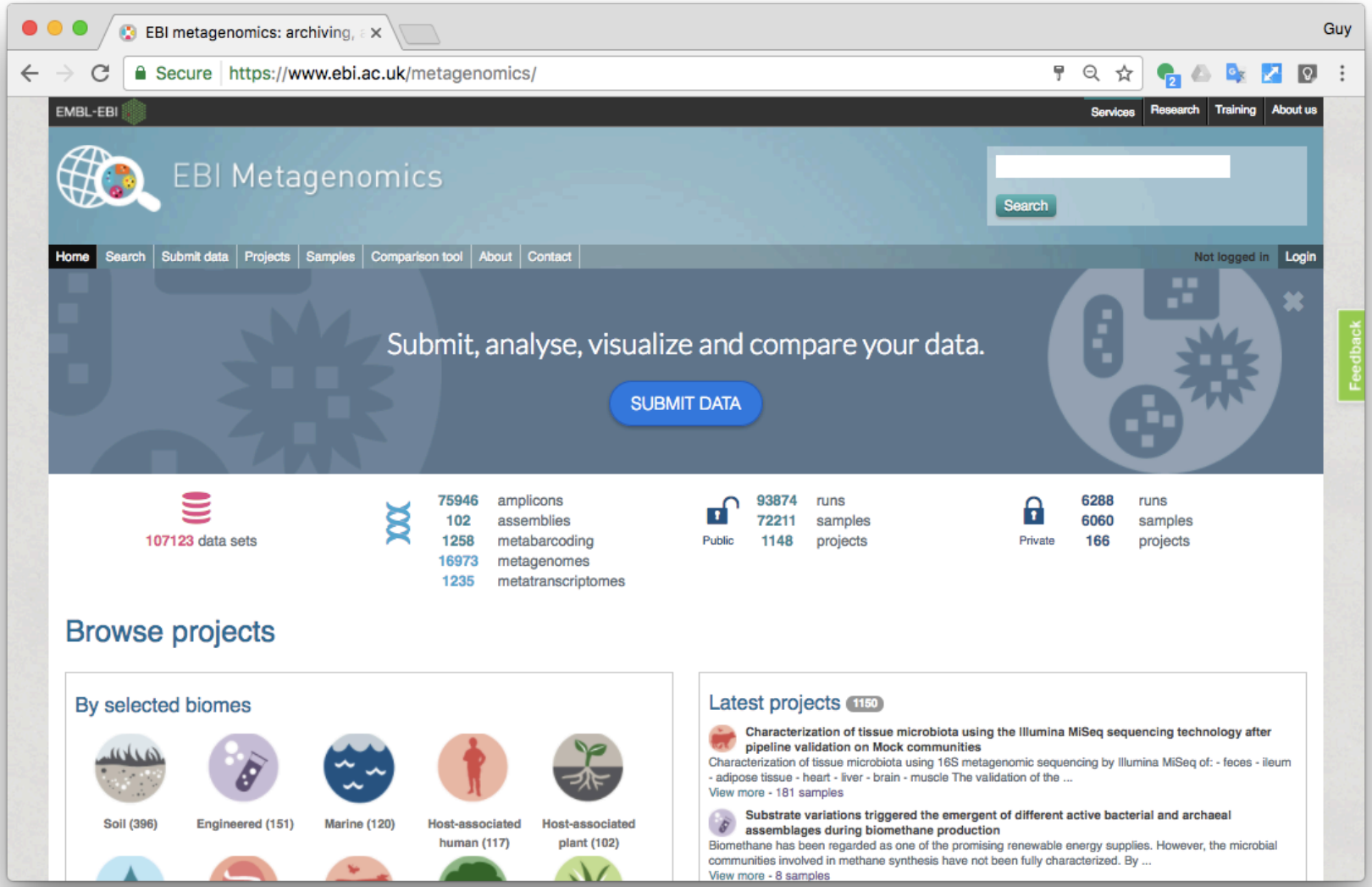
- In-project data sharing
- Data validation
- Dissemination and publishing
- Data coordination

- for customers small....,

- medium...

UniEuk

… and large


EBI metagenomics: archiving,
Guy
Secure
https://www.ebi.ac.uk/metagenomics/
EMBL-EBI
Services
Research
Training
About us
EBI Metagenomics
Search
Home
Search
Submit data
Projects
Samples
Comparison tool
About
Contact
Not logged in
Login
Submit, analyse, visualize and compare your data.
SUBMIT DATA
Feedback
107123 data sets
75946 amplicons
102 assemblies
1258 metabarcoding
16973 metagenomes
1235 metatranscriptomes
Public
93874 runs
72211 samples
1148 projects
Private
6288 runs
6060 samples
166 projects
Browse projects
By selected biomes
Soil (396)
Engineered (151)
Marine (120)
Host-associated human (117)
Host-associated plant (102)
Latest projects
1150
Characterization of tissue microbiota using the Illumina MiSeq sequencing technology after pipeline validation on Mock communities
Characterization of tissue microbiota using 16S metagenomic sequencing by Illumina MiSeq of: - feces - ileum - adipose tissue - heart - liver - brain - muscle The validation of the ...
View more - 181 samples
Substrate variations triggered the emergent of different active bacterial and archaeal assemblages during biomethane production
Biomethane has been regarded as one of the promising renewable energy supplies. However, the microbial communities involved in methane synthesis have not been fully characterized. By ...
View more - 8 samples

Standards

Standards through the data journey

Scientist generating data

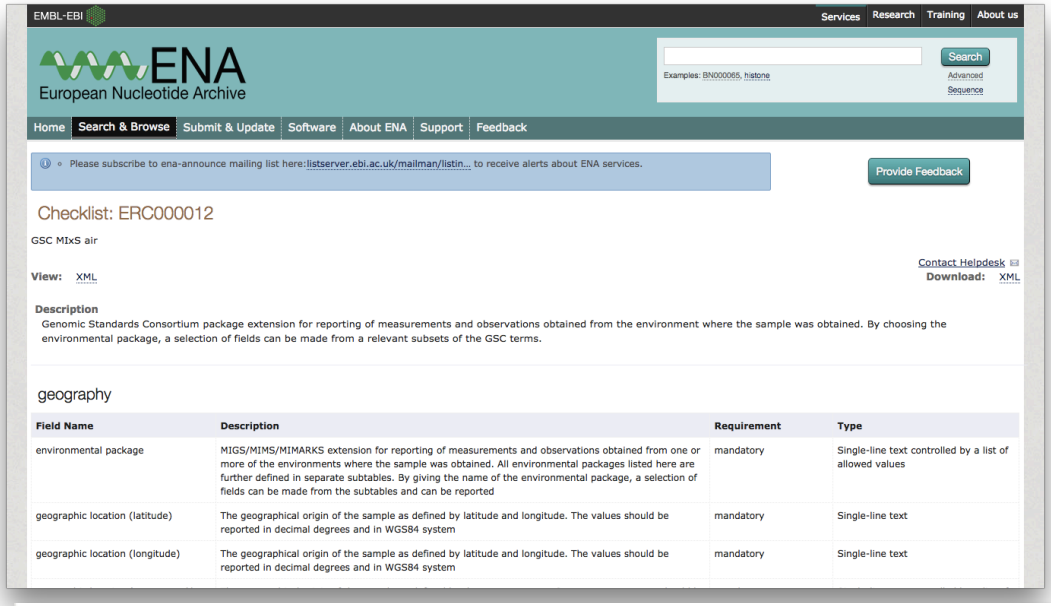
- Study design
- Data capture

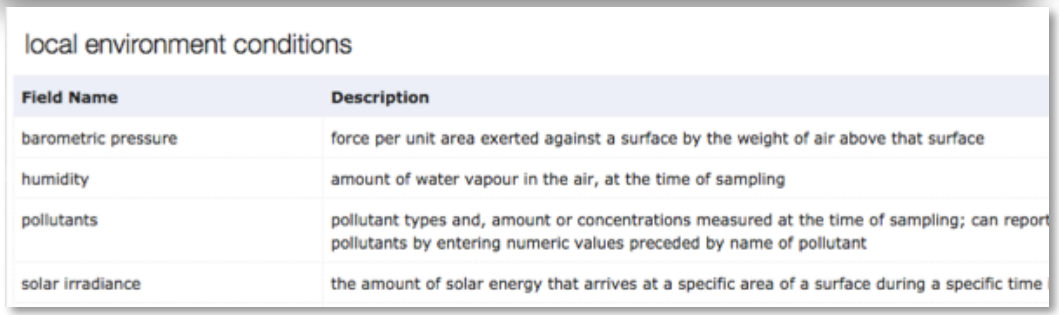
ENA

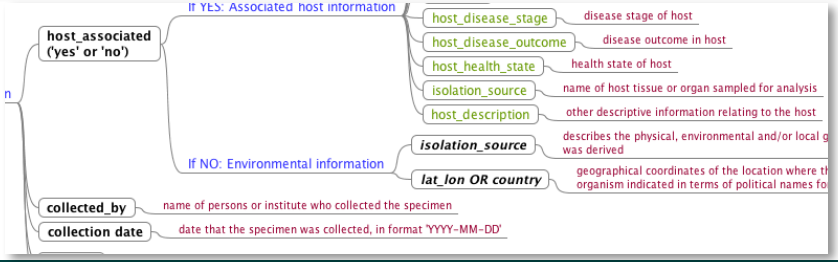
- Submission
- Validation
- Discovery
- Presentation

Scientist consuming data

- Interpretation

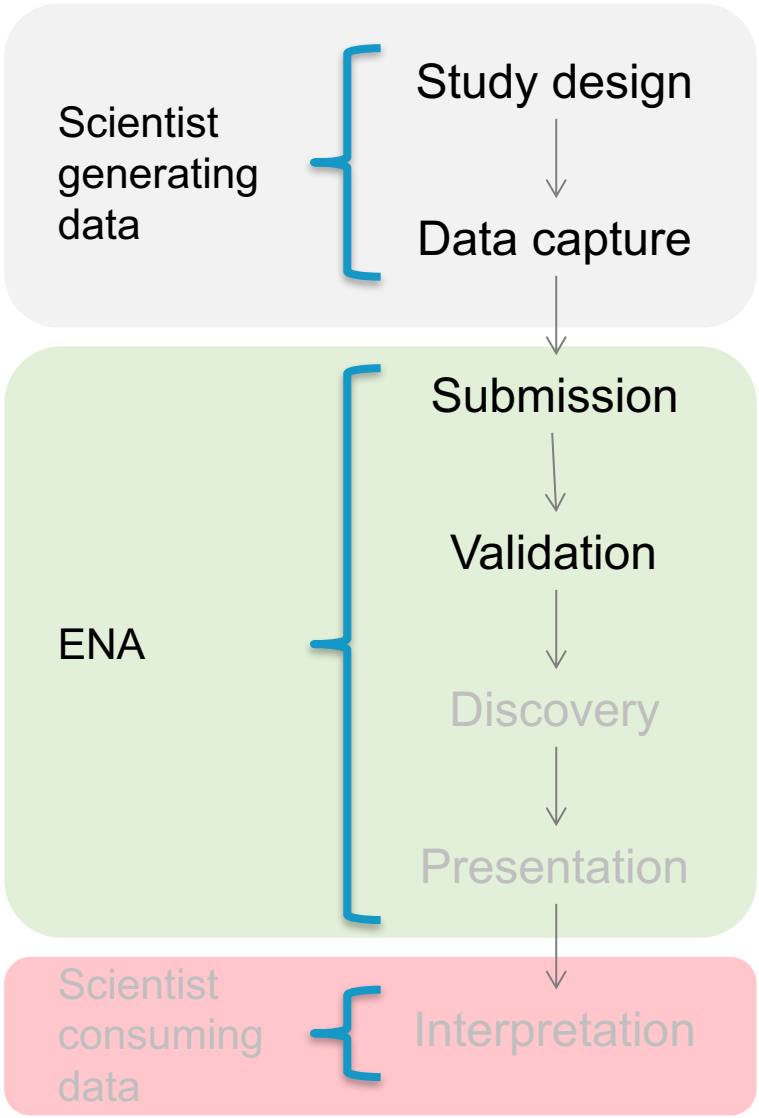


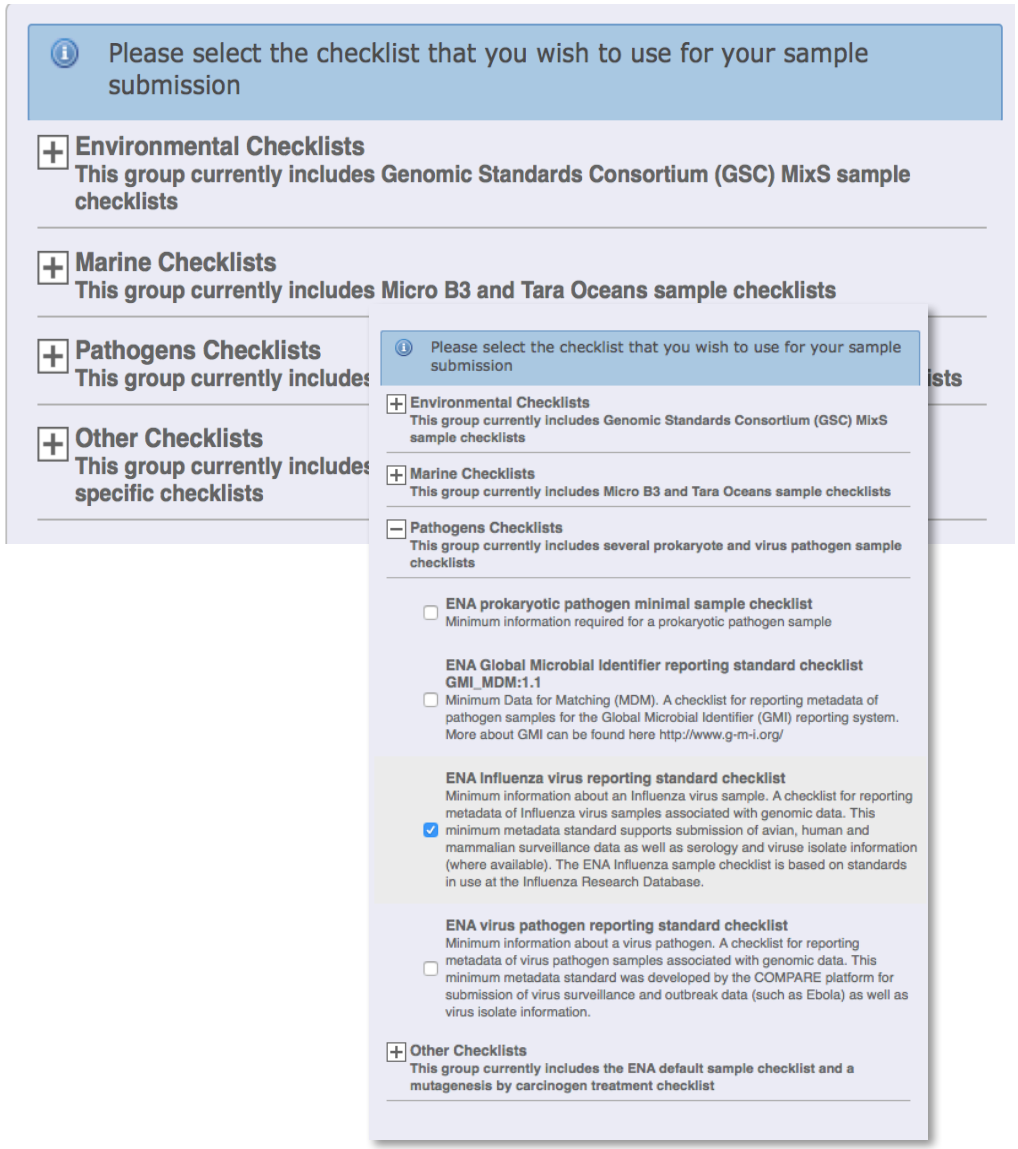




EMBL-EBI

Standards through the data journey





Standards through the data journey

Scientist generating data

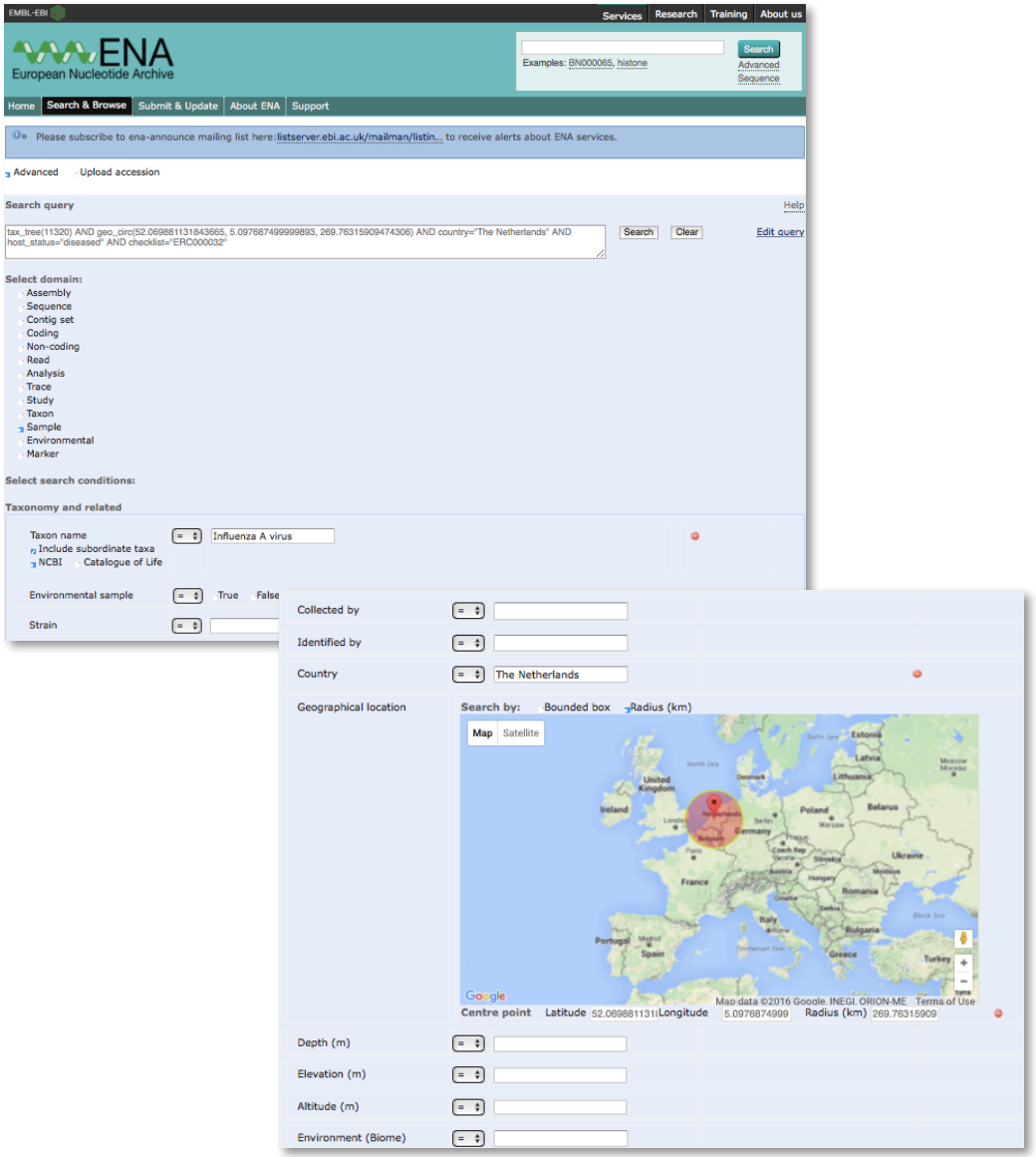
- Study design
- Data capture

ENA

- Submission
- Validation
- Discovery
- Presentation

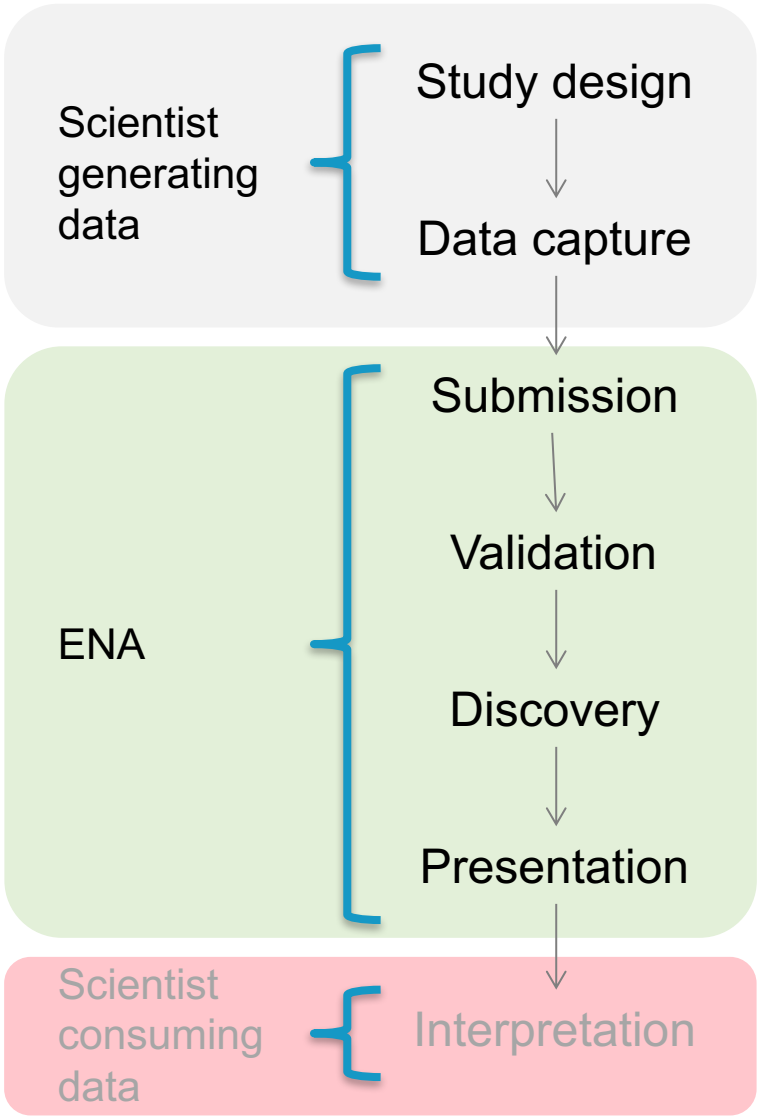
Scientist consuming data

- Interpretation



EMBL-EBI

Standards through the data journey



Please subscribe to ena-announce mailing list here:listserver.ebi.ac.uk/mailman/listin... to

Sample: SAMEA3493153

Influenza A virus; Initial sample from November H5N8 outbreak in UK

View: XML

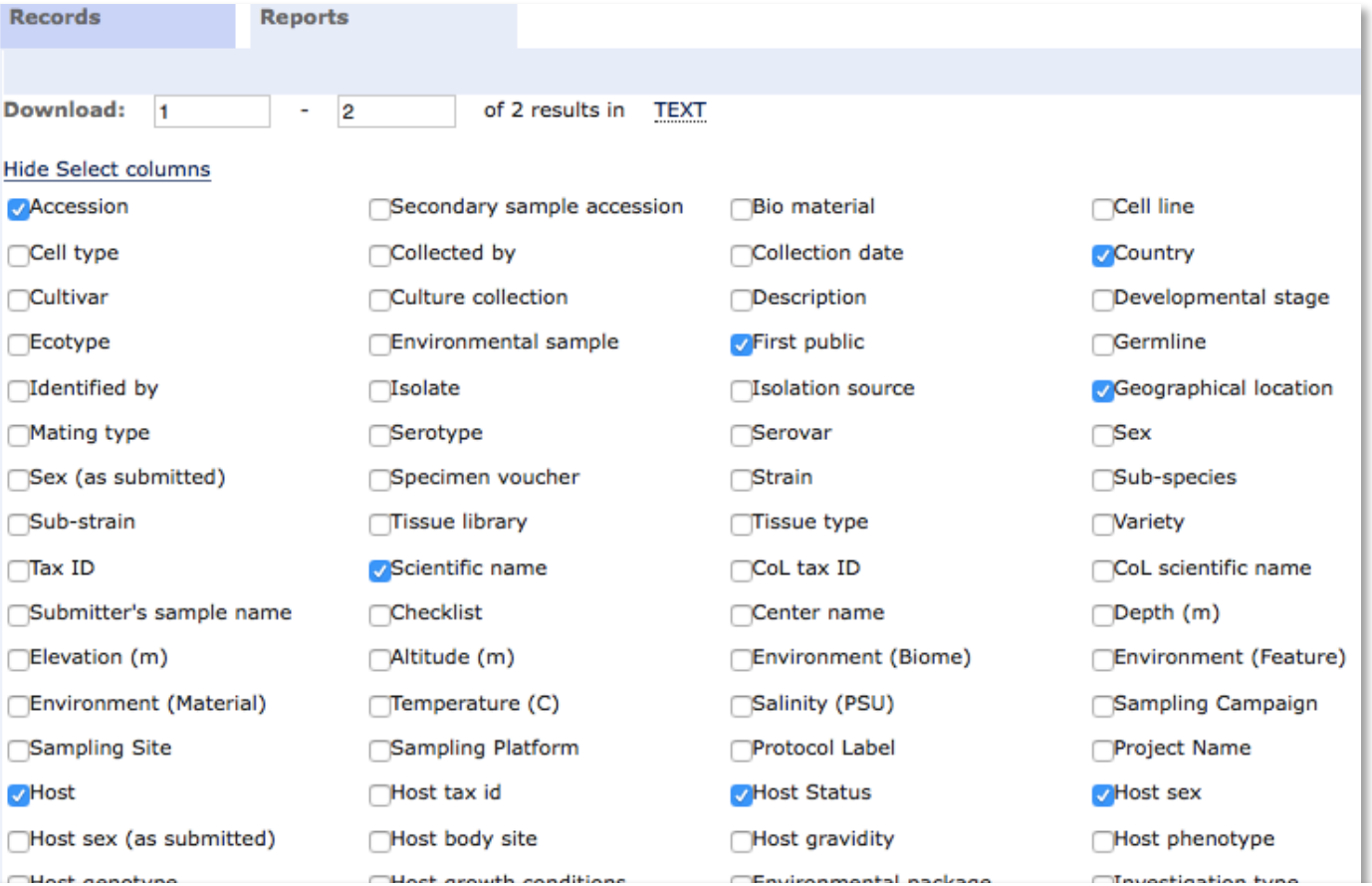
Submitting Centre	Organism
Animal and Plant Health Agency (APHA)	Influenza A virus

Secondary accession(s)
ERS800302

Lineage
Viruses, ssRNA viruses, ssRNA negative-strand viruses, Orthomyxoviridae, Influenzavirus A

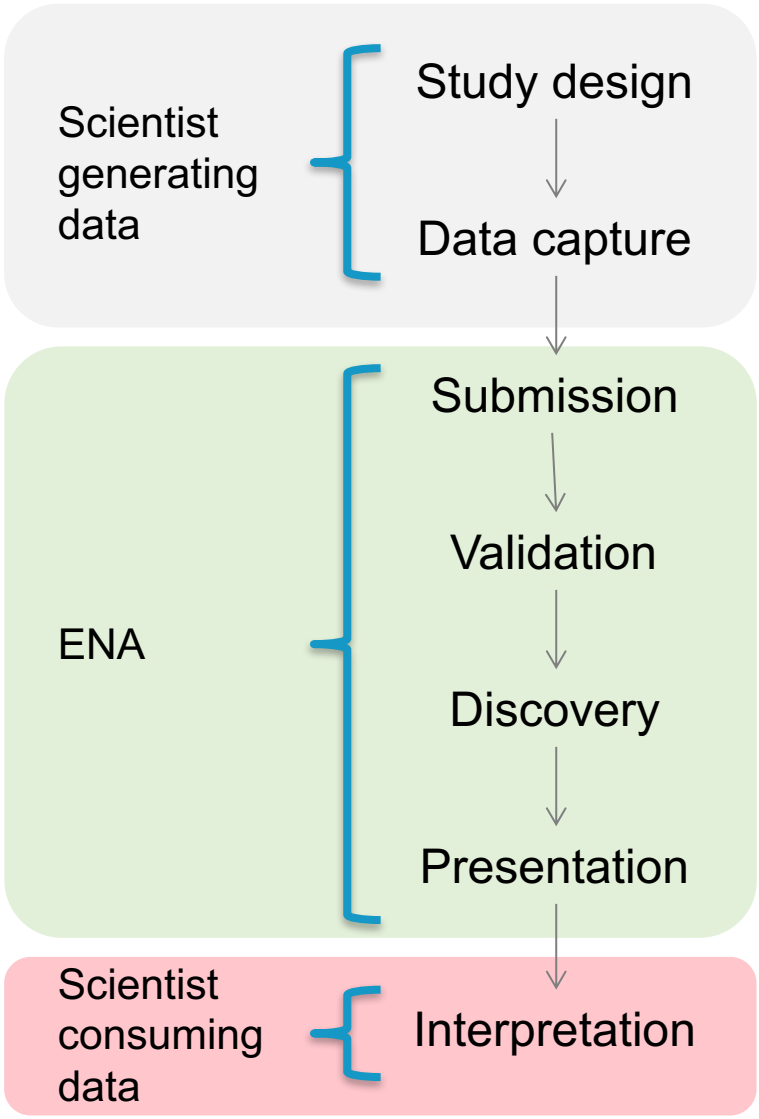
Navigation | Read Files | Attributes | Geolocation

host scientific name	Anas platyrhynchos
host common name	Domestic duck
host subject id	#1
host health state	diseased
host disease outcome	dead
host age	60 weeks
host sex	female
host behavior	domestic
host habitat	domestic:indoor farm
isolation source host-associated	pooled intestines
collector name	APHA



Accession	Collection date	Country	First public	Geographical location	Scientific name	Submitter's sample name	Host	Host Status	Host sex
SAMEA3493153	2014-11-14	United Kingdom	2016-02-17	54.001 N 0.434 W	Influenza A virus	AVP-14-036255	Anas platyrhynchos	diseased	female
SAMEA3860663	2014-11-04	Germany	2016-02-17		Influenza A virus	A/turkey/Germany/AR2485-L01478/2014	Meleagris gallopavo	diseased	

Standards through the data journey



Please subscribe to ena-announce mailing list here:listserver.ebi.ac.uk/mailman/listin... to

Sample: SAMEA3493153

Influenza A virus; Initial sample from November H5N8 outbreak in UK

View: XML

Submitting Centre	Organism
Animal and Plant Health Agency (APHA)	Influenza A virus

Secondary accession(s)
ERS800302

Lineage
Viruses, ssRNA viruses, ssRNA negative-strand viruses, Orthomyxoviridae, Influenzavirus A

Navigation | Read Files | Attributes | Geolocation

host scientific name	Anas platyrhynchos
host common name	Domestic duck
host subject id	#1
host health state	diseased
host disease outcome	dead
host age	60 weeks
host sex	female
host behavior	domestic
host habitat	domestic:indoor farm
isolation source host-associated	pooled intestines
collector name	APHA

Look out for…

ACCEPTED MANUSCRIPT

The metagenomic data life-cycle: standards and best practices

Petra ten Hoopen, Robert D. Finn, Lars Ailo Bongo, Erwan Corre, Bruno Fosso, Folker Meyer, Alex Mitchell, Eric Pelletier, Graziano Pesole, Monica Santamaria, Nils Peder Willassen, Guy Cochrane

Gigascience gix047. **DOI:** https://doi.org/10.1093/gigascience/gix047
Published: 16 June 2017

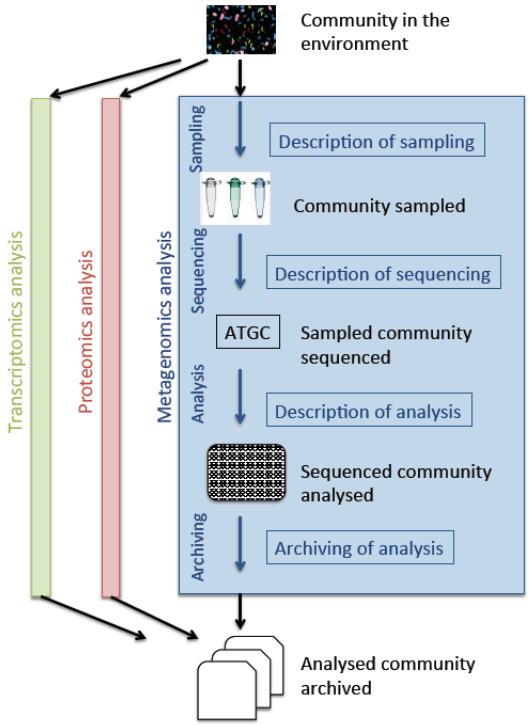
Abstract

Metagenomics data analyses from independent studies can only be compared if the analysis workflows are described in a harmonised way. In this overview, we have mapped the landscape of data standards available for the description of essential steps in metagenomics: (1) material
quencing (3) data analysis and (4) data archiving
arine research, we summarise essential variables
sampling processes and sequencing procedures
iment. These aspects of metagenomics dataset generation have been to some extent addressed by the scientific

OXFORD ACADEMIC

(GIGA)n SCIENCE

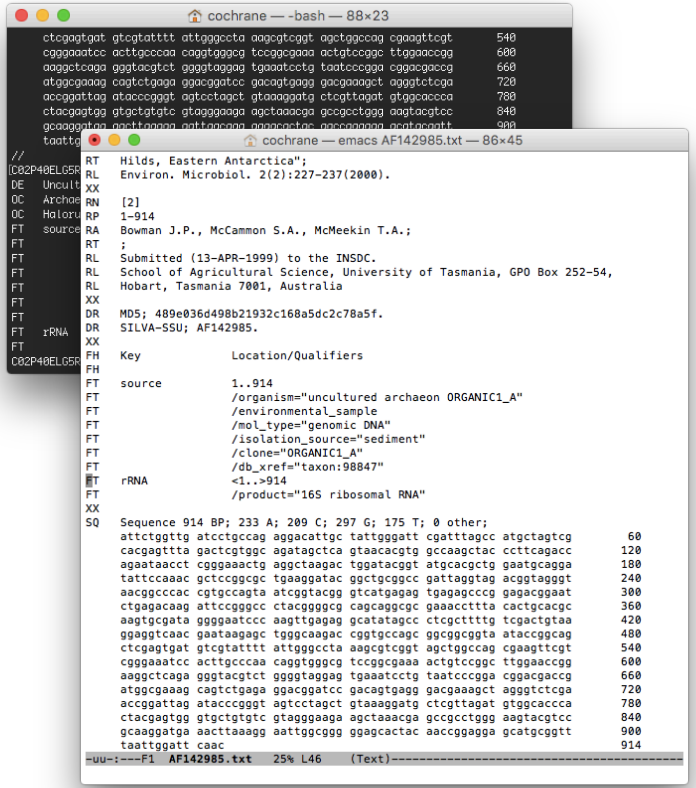
- Authors from 9 institutions, lead author Petra ten Hoopen, input from marine-community@elixir-europe.org
- Covers full workflow from sampling, though sequencing, analysis and result publication
- Promotes systematic archiving of analysis outputs



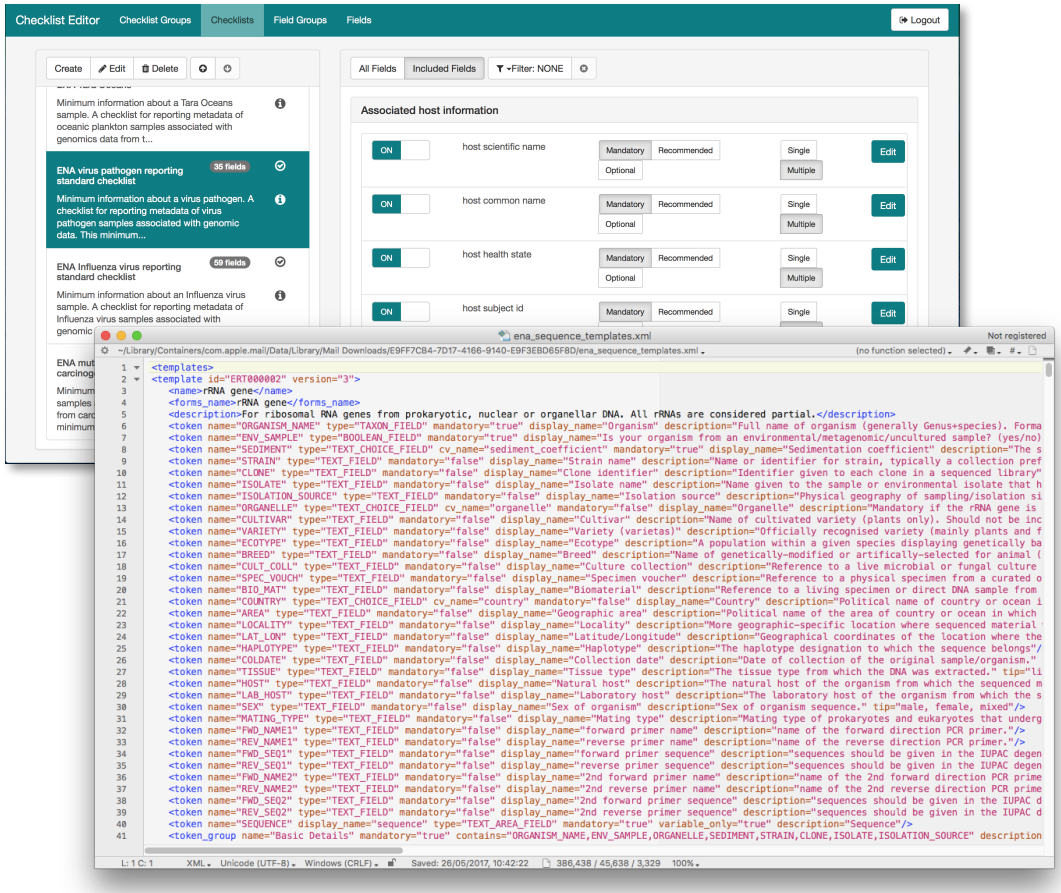
Scaling: internal tools

Sustainable biocuration

From…



… to



- Submission-level processing
- Class-level insertion of scientific organisation

EMBL-EBI

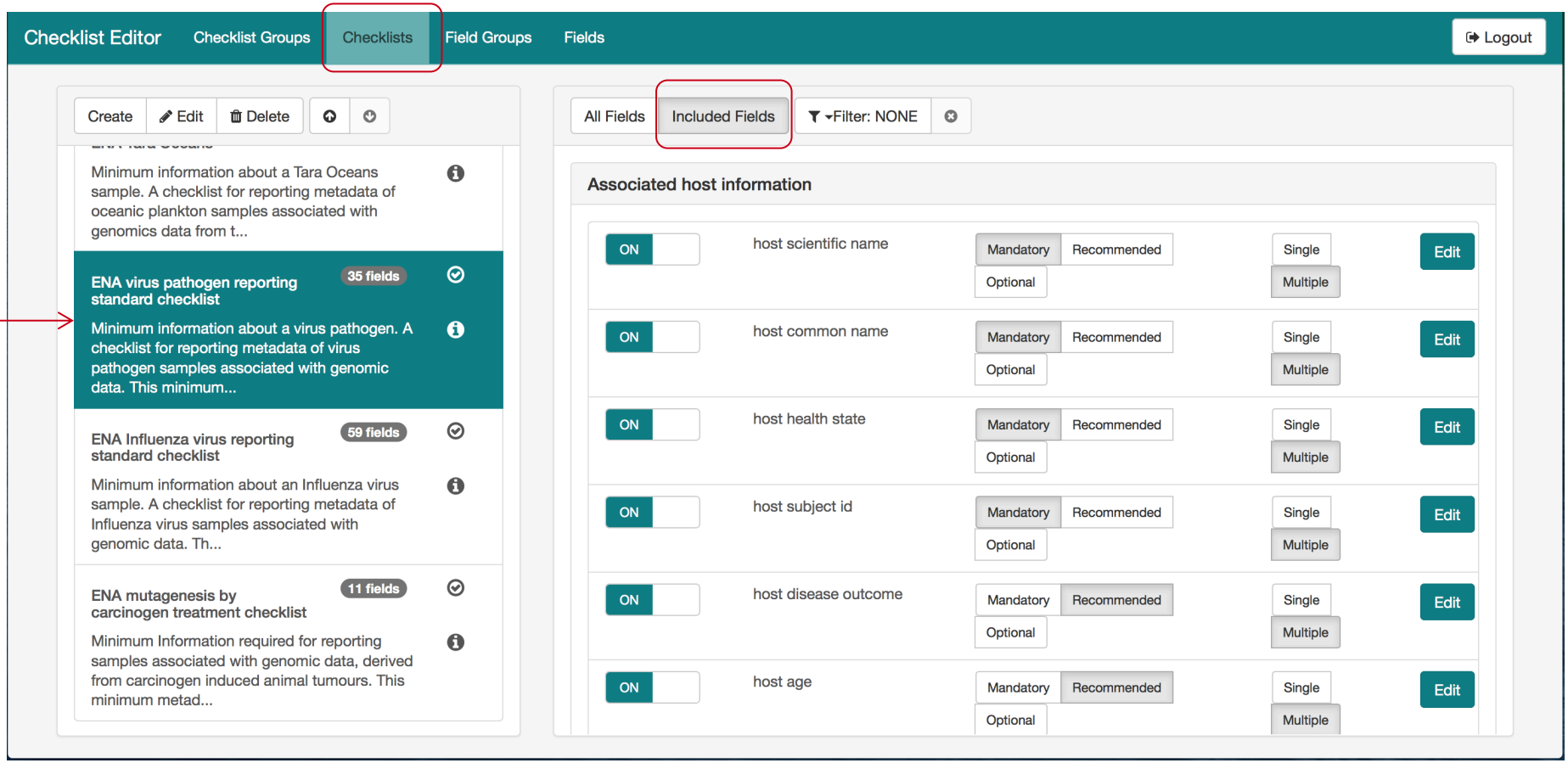
Checklist Editor

fields
- name
- description
- syntax

→ groups
- name
- description

→ checklists
- name
- description

→ implementation
- tools
- services
- content

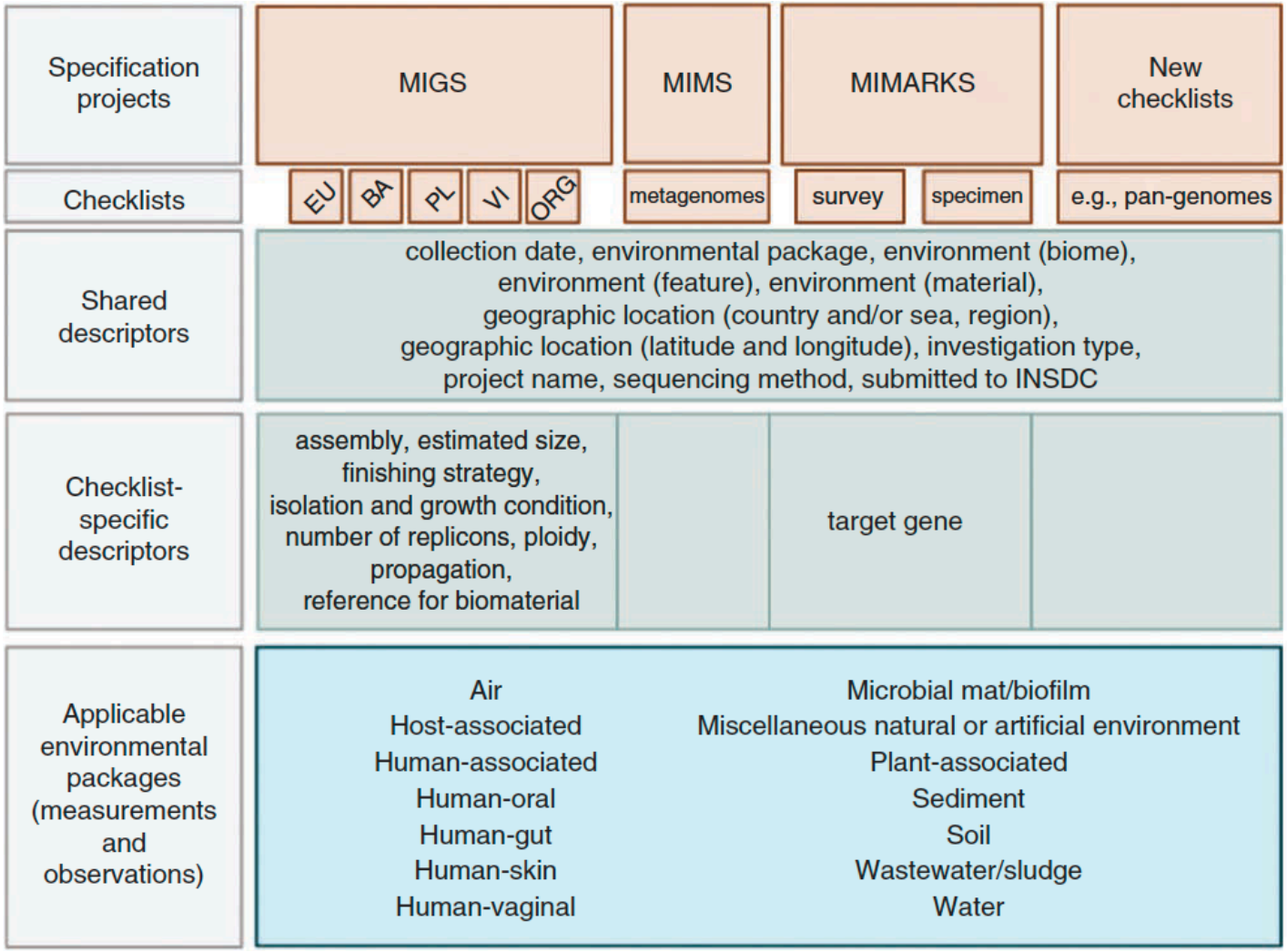


Scaling: community engagement

MIxS family of standards

genomic STANDARDS consortium

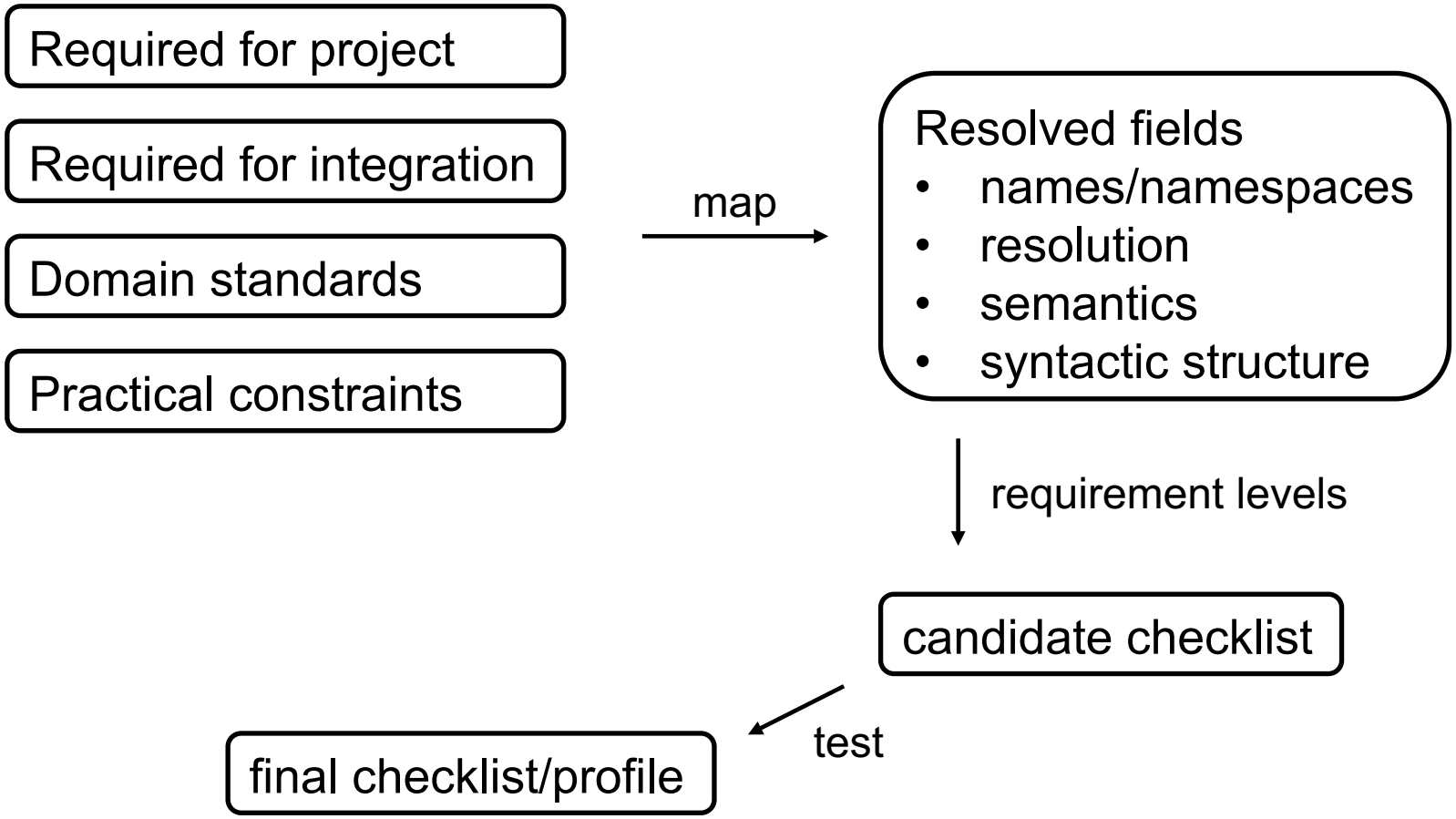
http://gensc.org



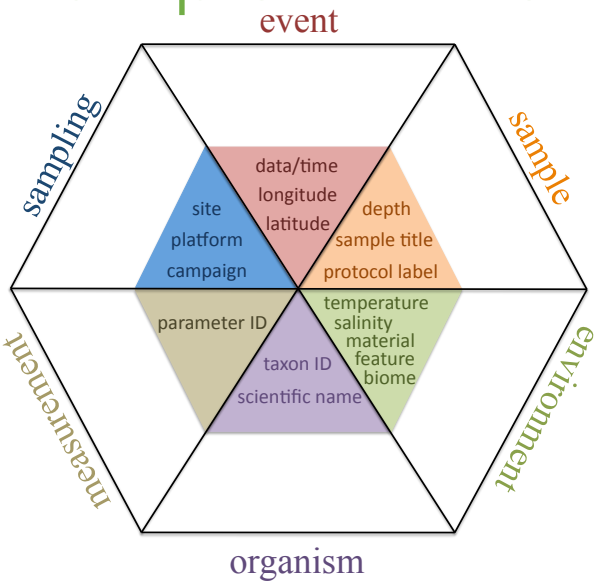
Yilmaz et al.
Nature Biotech.
2011; 29:415-420

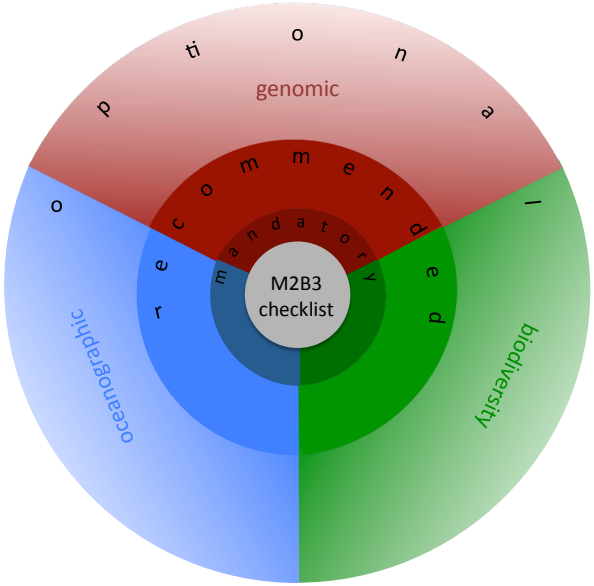
EMBL-EBI

Standards development workflow



Example: M2B3





ten Hoopen et al. Standards in Genomic Sciences (2015) 10:20
DOI 10.1186/s40793-015-0001-5

Standards in Genomic Sciences

COMMENTARY Open Access

Marine microbial biodiversity, bioinformatics and biotechnology (M2B3) data reporting and service standards

Petra ten Hoopen[1], Stéphane Pesant[2], Renzo Kottmann[3], Anna Kopf[3,9], Mesude Bicak[4], Simon Claus[5], Klaas Deneudt[5], Catherine Borremans[6], Peter Thijsse[7], Stefanie Dekeyzer[5], Dick MA Schaap[7], Chris Bowler[8], Frank Oliver Glöckner[3,9] and Guy Cochrane[1*]

Abstract

Contextual data collected concurrently with molecular samples are critical to the use of metagenomics in the fields of marine biodiversity, bioinformatics and biotechnology. We present here Marine Microbial Biodiversity, Bioinformatics and Biotechnology (M2B3) standards for *"Reporting"* and *"Serving"* data. The M2B3 Reporting Standard (1) describes minimal mandatory and recommended contextual information for a marine microbial sample obtained in the epipelagic zone, (2) includes meaningful information for researchers in the oceanographic, biodiversity and molecular disciplines, and (3) can easily be adopted by any marine laboratory with minimum sampling resources. The M2B3 Service Standard defines a software interface through which these data can be discovered and explored in data repositories. The M2B3 Standards were developed by the European project Micro B3, funded under 7th Framework Programme "Ocean of Tomorrow", and were first used with the Ocean Sampling Day initiative. We believe that these standards have value in broader marine science.

Keywords: Data standard, Marine, Molecular, Biodiversity, Microbial, Bioinformatics, Reporting, Interoperability

Background

An immense wealth of genetic, functional and morphological diversity in marine ecosystems remains unexplored, offering the potential for substantial scientific and biotechnological discoveries. Indeed, significant interest in this area has led to large-scale initiatives, such as Tara Oceans [1], the Global Ocean Survey [2] and Malaspina [3], that target the exploration of marine biodiversity on planetary scales. While the shared goal of such initiatives is the development of an understanding of the compos-

biotechnology will derive benefit. Prerequisite for the successful exploitation of acquired data are standards that enable interoperability in the data infrastructure.

Just as marine studies span many disciplines (e.g. biological, oceanographic, molecular), use of data from marine studies requires approaches that traverse the many disciplines, asking questions, for example, of species distribution, physical oceanographic parameters, molecular biology and data licensing. Each discipline has established infrastructure and best practice for the dissemin-

positories,
d analysis
when data
a lack of

explorations and will continue to build a powerful marine data infrastructure from which ecosystems biology and

of a consistent environment for the discovery and retrieval of data.

The Marine Microbial Biodiversity, Bioinformatics, Biotechnology Project (Micro B3) [4] unites intensive oceanographic monitoring, thorough biodiversity studies

* Correspondence: cochrane@ebi.ac.uk
[1]European Nucleotide Archive, EMBL-EBI, Wellcome Trust Genome Campus Hinxton, Cambridge CB10 1SD, UK
Full list of author information is available at the end of the article

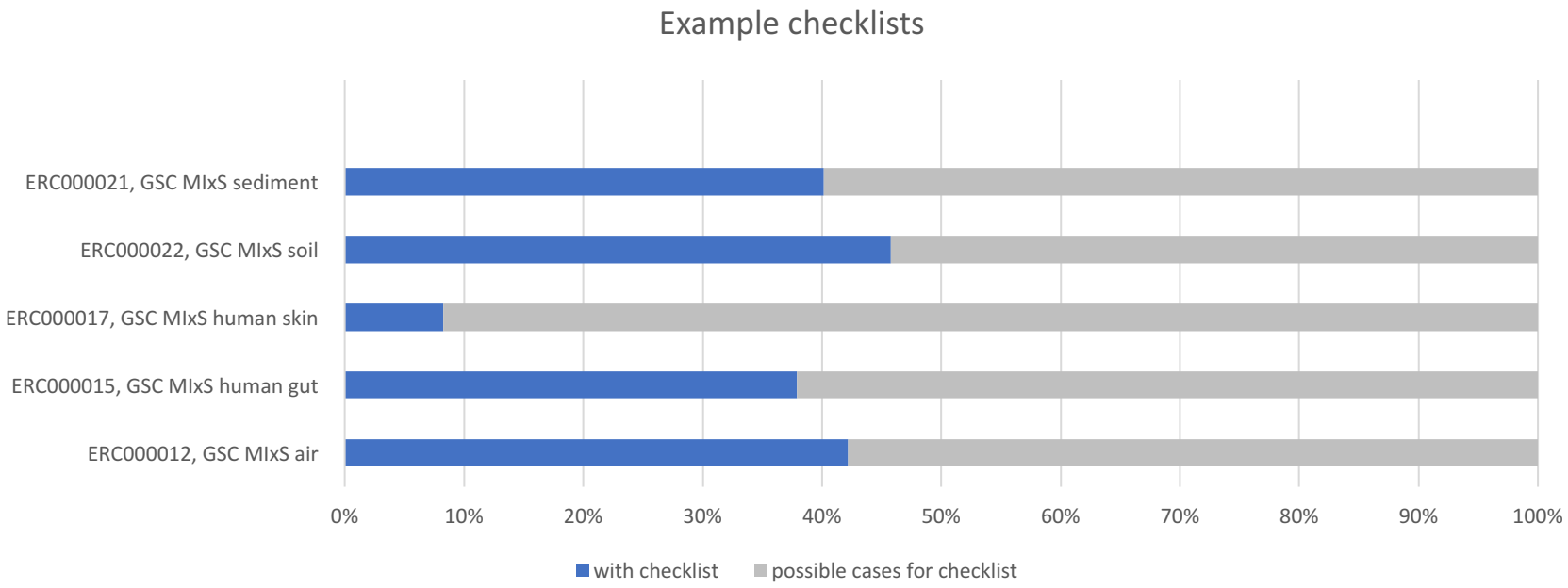
ten Hoopen et al. Standards in Genomic Sciences (2015) 10:20

micro B3 Biodiversity Bioinformatics Biotechnology

EMBL-EBI

Uptake

- **51%** (100,366/195,702) of metagenomics-related samples use appropriate checklists
- (**18%** (188,213/1,059,594) overall usage of MIxS and MIxS-informed checklists)



- Uptake of individual fields may be greater

*includes pre-publication data

Acknowledgements

- **Content & support:** Ana Cerdeño-Tárraga, Ana Luisa Toribio, Petra ten Hoopen, Marc Rosello, Jeena Rajan, Isabel Santos Magalhaes, **Clara Amid**; Ian Streeter, Susan Fairley, David Richardson, Peter Harrison, **Laura Clarke**
- **Technology (presentation & collaborative tools):** Blaise Alako, Simon Kay, **Nima Pakseresht**; Xin Liu, Suran Jayathilaka, **Nicole Silvester**
- **Technology (submissions & data back-end):** Daniel Vaughan, Neil Goodgame, Iain Cleland, Josué Martinez Villacorta, Dmitriy Smirnov, Kethi Reddy, Vadim Zalunin, Rasko Leinonen, **Thomas Keane**
- **Infrastructure:** EMBL-EBI Technical Services Cluster
- **Gigascience publication co-authors**
- **INSDC Partners:** NCBI, DDBJ
- **Funders**