



EFSA One Health WGS System

Mirko Rossi
BIOHAW Unit

Trusted science for safe food



Objective of the system



To collect **genomic profile of foodborne pathogens and associated epidemiological data** of isolates from food, feed, animals and related environment



To **allow ECDC to query the EFSA database** for finding possible matches between human and non-human isolates



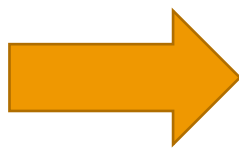
To offer a **set of services through a user-friendly interface** for the analysis and managing of the submitted data

Build a database of genome profiles that can be queried in case of food-borne outbreaks

- EU/EEA countries are invited to submit WGS-based typing data on a voluntary basis at any time throughout the year

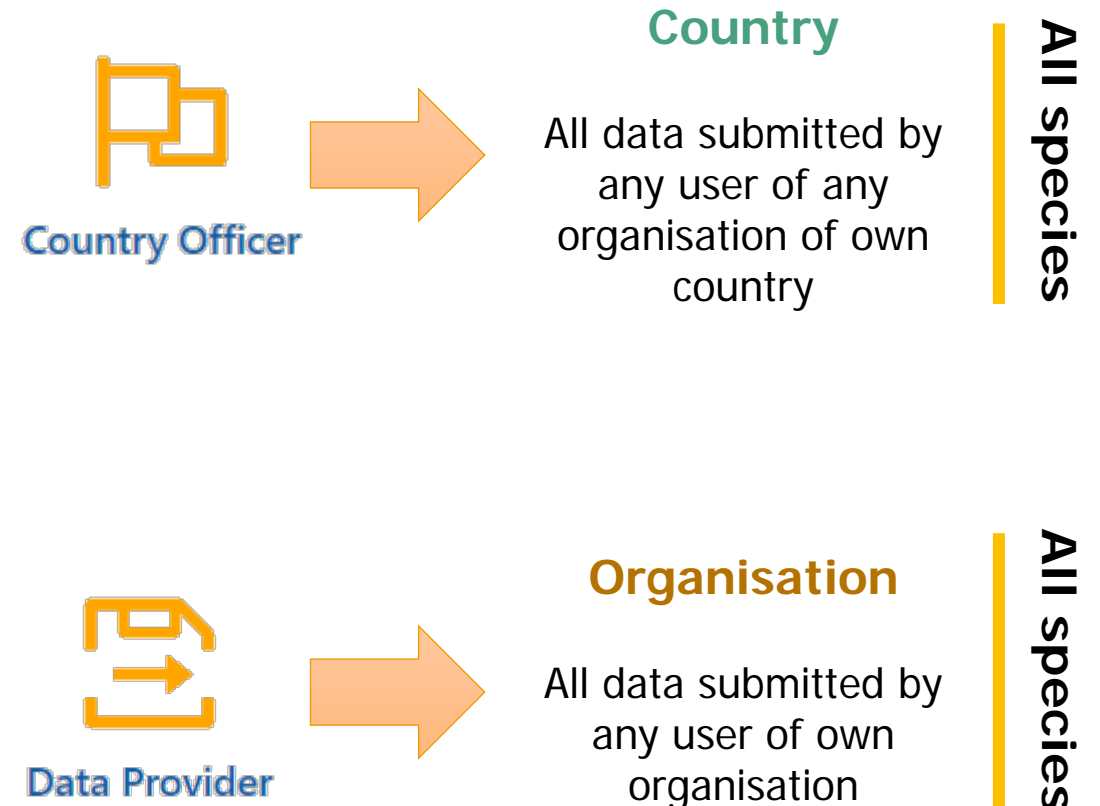
Support the real-time investigation of multi-country food-borne outbreaks

- EU/EEA countries shall provide WGS-based typing data to EFSA related to the specific event



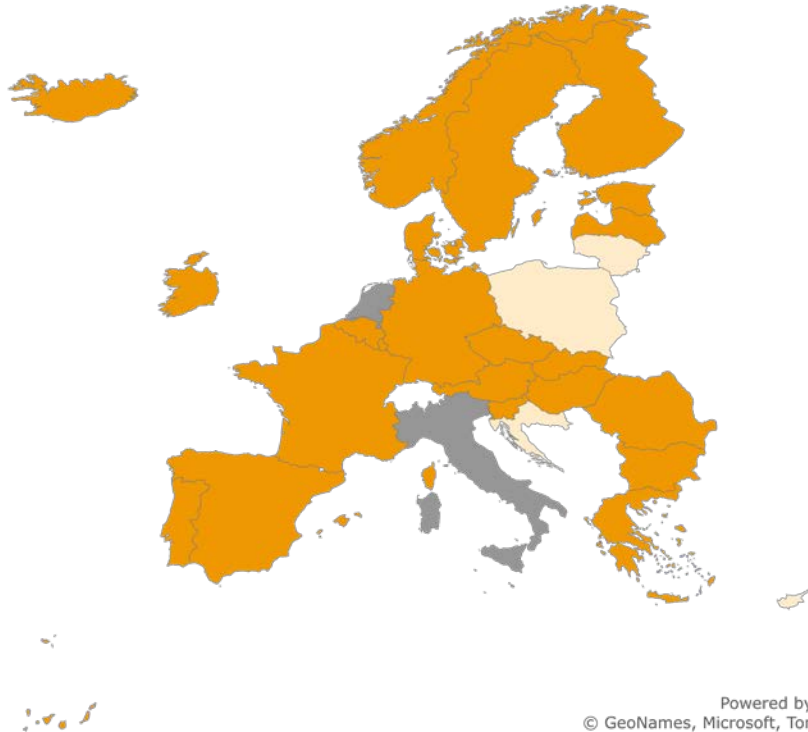
Any S. enterica, L. monocytogenes or E. coli isolates from food, feed, animals and related environments collected sequenced at EU/EEA country level are eligible for this data collection

- The system foresees two type of users at country level (EU/EEA countries) with different functions:
 - **Country Officer**: coordination
 - **Data Provider**: data submission
- Both undergo a formal appointing procedure
- There is no need to sign an agreement between MS and EFSA → users will subscribe a Term of Use before submitting data



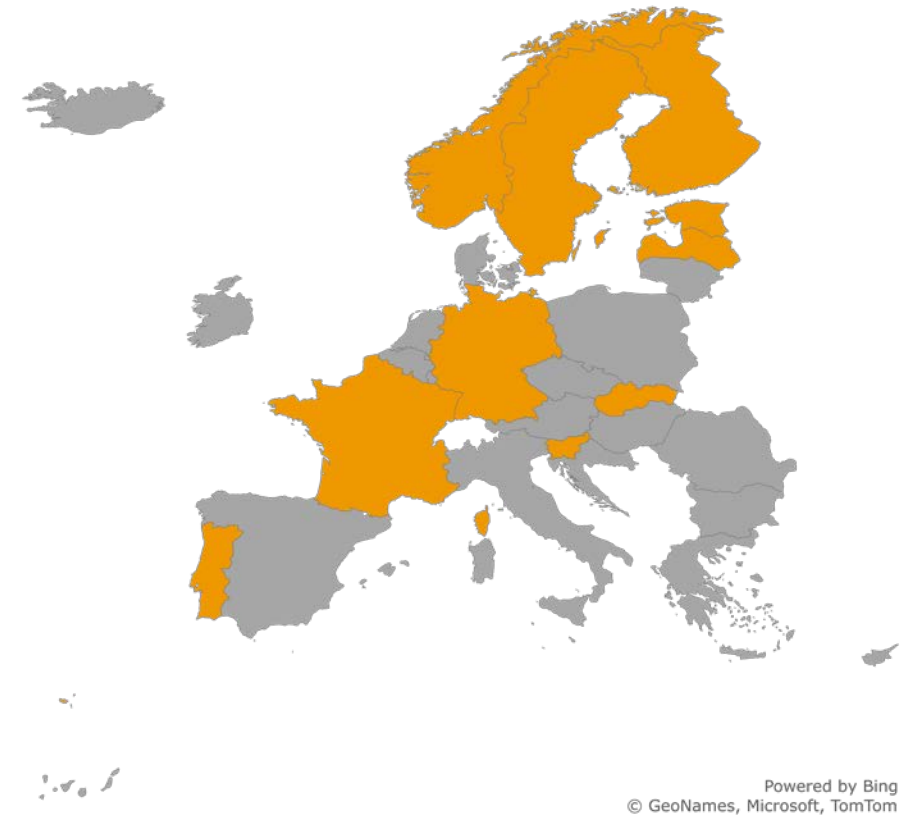
Country officer

- official nominated
- assigned to zoonoses reporting officer
- pending



Data Provider organizations

- pending
- official nominated



*at 16 May 2022



- When relevant EFSA initiates a data call from food sector at Member State level
- EFSA collects the WGS data from the non-human isolates and if matching with the human cluster it shares the data with ECDC
- EFSA contacts the network of Country Officers to initiate the data call
- It is expected that the Country Officer communicates the request to the national data providers and alerts the Food authority

■ EURL users

(Limited to the bacterial species of their interest)

- able to view and analyse all data from **all users of all organisations**
- **not able to formally submit data** to the EFSA or to edit data already submitted
- provide **support to Data Providers** on
 - technical aspect related to the upload of the WGS data to the system
 - interpretation of the microbiological clusters identified by the Data Provider users

■ EFSA Data Steward

- can perform any action on any data from **all users of all organisations**
- support **outbreak investigation and Data Provider users in complying with the data collection rules**
- in case of multi-country outbreak investigation **can “lock” a specific Entry** which is part of an investigation.
 - when an Entry is locked no changes are possible by any user of the system



Experimental data: information related to the experiment (raw sequencing reads) - **Mandatory**



Typing data: genomic profile and other typing data extracted from the raw sequencing reads - **Mandatory**



Epidemiological data: information related to the food, feed and animal samples from which the pathogen isolates linked to genomic profiles originated – **Recommended**

- **Experimental and typing data** are stored simultaneously in the EFSA DB
- Each **Experimental and typing data** are **uniquely linked to an Entry ID** automatically assigned by the system upon submission of the Experimental data
- System does **not allow the multiple submission** of the same experiment
- Experiment is **univocally identified with the MD5 Checksum** of the FASTQ file

- **Typing data** are set of information extracted from the experiment by the EFSA One Health WGS analytical pipeline or by comparable bioinformatic pipelines.
- They are composed by
 - the **cgMLST allelic profile** of the isolate
 - **other elements** describing the quality of the sequence and other typing information (other than the cgMLST profile) related to the sequenced isolate
- cgMLST profile is considered valid only if the allele calling is performed on genome assemblies using chewBBACA v >2.8.4 with schemas downloaded from chewieNS

Experimental data

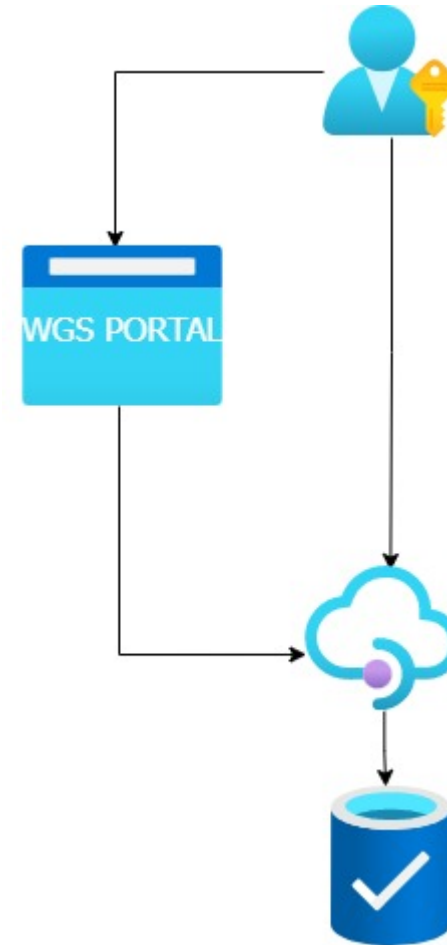
- Local RawReads ID
- Isolate species
- Layout
- FASTQ name
- FASTQ MD5 checksum

Typing Data

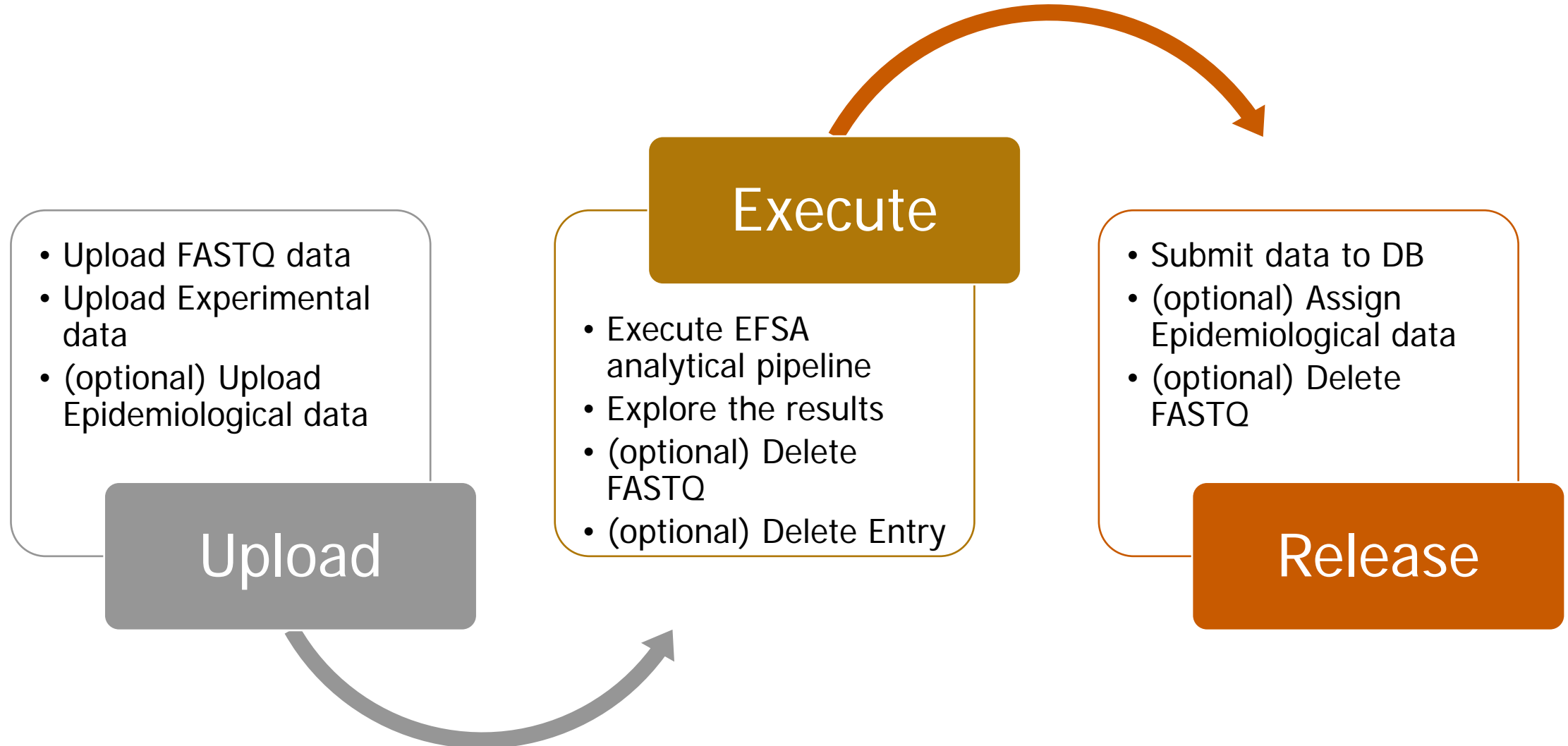
- cgMLST profile
- FASTQ quality
- Assembly quality
- MLST
- Serotype
- Pathotype (*E. coli*)

- **Epidemiological data** can be submitted at the same time with the Experimental and typing data or in a different moment by the any data provider user of the same organization
- **Epidemiological data** can be linked to **multiple Entry ID**
 - They cannot be deleted when assigned
- **Mandatory elements**
 - Sampling ID
 - Country of sampling
 - Coded description of the matrix of the sample taken
 - Additional information of the matrix of the sample
 - Sampling year

- **Through WGS PORTAL:** use cloud resources of EFSA for extracting Typing data from sequences and submit data to DB
- **Programmatically:** submit to DB precalculated typing data accessing partner API



Submission process through the WGS portal



- FASTQ files remain stored in the EFSA system on discretion of the Data Provider
 - They can be deleted at any time after execution of the pipeline
- Before releasing data to DB, Data Provider can delete Entry from system at any time
 - Uploaded not released data are not formally submitted to EFSA
- Epidemiological data can be assigned to any entries regardless the release status

- Data Provider should prepare Experimental, Typing data and Epidemiological data in specific file format
- Uploading the files to the EFSA system programmatically
 - Using a Command Line Interface or accessing directly the partner API
- Upon successful validation of the files a new Entry ID will be generated and data are released in the EFSA DB
- The Entry will be visitable in the WGS portal
- The WGS portal remains the single point for reading data from the database and to perform cluster analysis

- All data is accessible for query from other users (including ECDC)
- Data Provider can “unrelease” Entries at any time blocking the sharing of data (if not locked by EFSA Steward)
- Data Provider cannot delete data from the system
- Data Provider can specify if country information can be share or not

- **Detection of joint microbiological clusters of human and non-human** food-borne pathogens isolates,
- **Automatic exchange of WGS-based typing data and epidemiological data** between the two systems
 - limited to cgMLST allelic profiles where for each locus the allele sequence is reported as a CRC32 hash code
 - limited set of Epidemiological data is shared between EFSA and ECDC
- **Both systems produce a cluster analysis**

data with no restriction

- cgMLST profile
- other Typing data including serotype, pathotype, MLST
- cluster type
- sampling date (if submitted)
- sample category (if submitted) => HUMAN, FOOD categories (e.g. Milk and milk products (dairy), Eggs and egg products, etc.) , FEED, NON-FOOD matrices

data with restricted access

- country of the Data Provider => Country of origin of the Data Provider organization
- reporting country => Country where the sampling was performed

■ Non-human data

ECDC system

- Visible to user having isolate in the cluster => regardless the country of origin of the organisation of the user
- EFSA data provider user can decide if shared this data with ECDC

 Share country info

 Hide country info

EFSA system

- Visible to EFSA user of the same organisation
- Visible to EFSA user of organizations of the same country

■ Human data in EFSA system

- Visible to EFSA user of organisations of the same country

- Submission of the data doesn't affect the ownership, existing IPRs and the copyrights
- EFSA and ECDC do not become owner of the data, neither the data submitted by the respective data providers nor the data exchange between the systems
- Neither EFSA, ECDC nor any users can claim any IPR on data as well as on discoveries, detections or inventions stemming from the use of the data
- EFSA and ECDC have the right to use the data performing analyses in support of assessment of food-borne outbreaks

- Without prejudice to rules on public access to documents, EFSA considers the data stored in its database as **sensitive non-classified information**
- EFSA might share information to third parties only prior agreement with the Data Provider
- According to Art 4(1) of GDPR and Art 3(1) of EDPR any data from human (including the cgMLST profiles) are qualified as “personal”
- Data from human are duly pseudonymized by ECDC
- **EFSA users are by no means able to make any personal identification**



Subscribe to

efsa.europa.eu/en/news/newsletters
efsa.europa.eu/en/rss



Receive job alerts

careers.efsa.europa.eu – job alerts



Follow us on Twitter

[@efsa_eu](https://twitter.com/efsa_eu)
[@plants_efsa](https://twitter.com/plants_efsa)
[@methods_efsa](https://twitter.com/methods_efsa)
[@animals_efsa](https://twitter.com/animals_efsa)



Follow us Linked in

[Linkedin.com/company/efsa](https://linkedin.com/company/efsa)



Contact us

efsa.europa.eu/en/contact/askefsa