### EU One Health 2018 Zoonoses Report

Mirko Rossi

Scientific officer



Trusted science for safe food

# Reported numbers and notificaton rates of confirmed human zoonoses in the EU, 2018





#### Main findings (EUOHZ 2018 report)

Salmonella the most common cause of foodborne outbreaks in the EU

West Nile virus and STEC infections at unusually high levels



- The EU/EEA trend STEC notification has been increasing from 2014 to 2018.
- In 2018 STEC Food-borne (N=43) and waterborne (N=5) with 381 human cases
- Strong-evidence food-borne outbreaks: 2 cheese, 1 milk, 1 'other or mixed red meat and products thereof' and 1 'vegetables and juices and other products thereof'
- Different from 2010-2017 (most by 'bovine meat and products thereof')
- In 2018, 22 MS plus 1 non-MS reported on the presence of STEC in 21,142 food samples while six MS reported on the monitoring of 1,690 animal samples



- 2.4% (485/21,142) STEC-positive food samples in 2018 (= last 4 years)
  - food of animal origin most, fruits and vegetables least contaminated
- Sprouted seeds food category with the highest number of MS reporting
  - 1 STEC-positive retail sample of overall 444 tested
- 58.8% of the strains with serogroup reported belonged to O-groups included in the list of the top 20 STEC serogroups reported to ECDC in the period 2014-2018
- 126 out of the 233 STEC with the stx genes profiles were stx2+, in 27 cases together with the eae gene
  - Information on the eae gene was provided only for 80 of the 126 stx2+ STEC
- Testing of animal samples has been declining over the last years
  - large use of methods that only detect E. coli O157 to test samples from animals

#### EC mandate on 'One Health' system for the collection and analysis of WGS data from food/animal isolates

Mirko Rossi

Scientific officer



Trusted science for safe food





## Two interoperating systems

#### (EFSA and ECDC)

Each system collects and stores the data (i.e. allelic profiles and descriptive data) of the respective data domain.

#### **Cross-sector matches**

#### Machine-to-machine

Databases will be queried, and comparison will be performed live to the data stored, returning any matches (according to business rules) Automatic exchanging of allelic profiles and descriptive data as established in the Collaboration Agreement

### The foreseen 'One Health' system





### EFSA implementation strategy



- Cloud based implementation
  - The EFSA WGS system will be integrated in the EFSA cloud Azure system
  - Azure Data lake located in Western (primary site) and Northern Europe (secondary site)
- User-centered design application for managing data submission
- **Open source** analytical pipeline:
  - Written in nextflow (modular)
  - Open source distribution of the pipeline (GitHub, nf\_core)
- Mirroring ENA/SRA for FastQ submission
  - the system will not allow re-submission of the same raw data (check for MD5 checksum)
- Move-code-to-data → allowing programmatic submission of WGS-based typing data
- Data will be stored and analysed in a system within secured network

### Type of data in the system



- Raw Reads: FASTQ(s) submitted by data-providers or downloaded from public repositories
- EpiData: epidemiological data linked to one Raw Reads submission; <u>always</u> <u>submitted by the data-providers</u>
- Analytical Pipeline Results: containing data extracted from the raw reads by the analytical pipeline in the EFSA system or at the data-provider premises
  - Allelic profile
  - Accessory data statistics of the analysis, QC, Serotype, Pathotype, stx type
  - Assembly
  - **AMR profile** EURL AR official pipeline
  - Virulence profile: BLAST search of dedicated reference datasets



- Data provider is always owner of the submitted private FASTQ
- The data providers remain owners of the transformed FASTQ if it can be related 1:1 with the original raw data
- In case of transformation-results aggregating several records from the second layer, the resulting data is owned by EFSA
- The data providers are allowed to withdraw data for future use (even if the data has indirectly been used in a further downstream analysis)
- Data submitted by data provider (EpiData, FASTQ or both) and EpiData linked to publicly available FASTQ are considered "private".

#### Deletion of the entries



- Physical deletion of FASTQ and assembly → remove the data from the datastore
- Logical deletion of the allelic profile and typing data → data remains in the data-store, but invisible for future research
- EpiData cannot be deleted only update
  - old versions are stored for traceability
  - no check for consistency between versions of EpiData
- In case of ROA → `entries frozen for deletion'
  - during such freeze-period, the owners-organizations will not be capable of deleting data
  - maintained by EFSA administration



- Mirroring ENA submission model
- Raw Reads submitted to Data Lake through SFTP or VPN or WEB-IN
- Raw reads linked to "Local Raw Reads ID" unique for OwnerOrganization
- EpiData can be submitted asynchronously

### ExpData to submit with FASTQ



LocalRawReadsID	data provider's sequence identification code or ENA Accession Number	provided
OwnerOrganization	the organization who owns the RawReads data; it should match to the one derived from the organization according userID who has logged on to provide the data	provided
InstrumentModel	the model of instrument used for the analysis (controlled vocabulary <u>https://ena-docs.readthedocs.io/en/latest/submit/reads/webin-</u>	provided
Species	<ul> <li><u>cli.html#platform</u>)</li> <li>species describing the isolate species: one out of following three possible values are allowed:</li> <li>"Salmonella enterica"</li> <li>"Listeria monocytogenes"</li> <li>"Escherichia coli"</li> </ul>	provided
LibraryLayout	SINGLE or PAIRED	provided
RawReads	Names of the FASTQ-files	provided

### EpiData linked to FASTQ



LocalRawReadsID	data provider's sequence identification code or ENA Accession Number							
OwnerOrganization	The organization who owns the RawReads data; Linker to FASTQ submission	M/Exp						
Sample ID	ID of the sample from which the bacteria has been isolated and sequenced and from which the WGS files have been derived.	М						
Country of sampling	Country of sampling	М						
Date of sampling	Year/month/day	М						
Sample Matrix	Description of the sample taken based on FoodEx2 catalogue	М						
Isolate ID	ID of the bacteria isolate sequenced and from which the WGS files have been derived.	М						
Sampling point	Point where the sample which generates the isolate has been sampled	R						
Country of origin	Country of origin of the sample taken	R						
Date of isolation	Year/month/day of isolation	0						
Area of sampling	Area where the sample was collected							
Programme type	"Outbreak investigation" (K032A) or "RASFF alert notification" (K033A)	0						
Sampler	i.e. "Official sampling" (CX02A)	0						
Additional sampling programme information	RASFF notification number, if available.							

### Submission and management of the data





### Submission and management of the data





#### Interaction with system





- The user can explore the content of EFSA DB based on specific requests
- The user can search for similar strain in EFSA DB and visualize relationship based on open-source ViZ software (e.g. PHYLOViZ or GrapeTree)
- Trough the EFSA system the user can submit request to ECDC DB for searching similar human isolates

### ECDC integration: query to ECDC DB





#### Monitoring calls

 $\rightarrow$  Each user sees own calls; EFSA SO and Admin see all calls

	Queryl D	Sche ma	EFSA user	Time Stam p	Mat ch	Value K	EFSAIDs	ECDC clusterI D	Total matches at ECDC	ECDC entries	ValidTo	ViZ
	F	EC_vI	mario.bianchi@efsa.europa. eu	2021- 01-19 03:14: 07	yes	5	658364503 693435545	EC1412	3	XXxxX YYyyYY ZZzzYY	2021-04-19 03:14:07	Link
	2021000	LM_vI	hans.mustermann@bfr.de	2021- 01-19 09:10: 56	no	7	652667603	-	•	-	-	
÷												

#### Actions:

- Delete result of the query
- Download the ECDC entries and run a ViZ
- Rerun a new query with the same EFSA isolates



### ECDC integration: query from ECDC





#### Monitoring calls

 $\rightarrow$  Only visible to the EFSA SO and Admin

QueryID	Schema	clusterID	Size ECDC	ECDC system user	TimeStamp	Mat ch	Value K	UI	EFSAIDs	Report
						_				$\frown$
20210001	EC_v1	EC1412	14	joe.Smith@ecdc.europa.eu	2021-01-19 03:14:07	yes	5	724	658364503 693435545	link
20210002	LM_v1	LM2534	3	matti.hanninen@thl.fi	2021-01-19 09:10:56	no	7	n.a.	-	-

QUERY EFSAID Country of I ID sampling g		Reportin g year	Year of sampling	Type of matrix	Description of the matrix of the sample taken	Description of the isolate species	Averag e AD	Max Ad	Min Ad	
2020000 I	658364503	UK	2019	2019	FOOD	salmon smocked	Listeria monocytogenes	3	-	4
	693435545	PL	2018	2018	FOOD	salmon smocked	Listeria monocytogenes	5	3	7

### Nextflow pipeline







- Pipeline in EFSA system == open source distribution in GitHub
- We will distribute the code in GitHub and a container in nf\_core (<u>https://nf-co.re/</u>)
  - User needs to install nextflow + docker/singularity (simple!)
  - The pipeline can be run with a single line of command

#### nextflow run nf-core/EFSA\_wgs -r 1.0.0 -params-file params.yaml -profile docker

 This will produce a single JSON file which can be directly submitted to EFSA Azure system using dedicated software

EFSA\_submission.py --userID mirossi --password IlOveWGS \_input /path/to/JSON

### Next steps (by the end of the year)



- Finalization of the design of the front-end app
- Implementation of the FASTQ collection framework
- Usability test in October/November
  - EURL, EFSA internal, invited NRLs
- First draft of the pipeline for PT

#### Stay connected





#### **Subscribe to**

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



#### **Engage with careers**

www.efsa.europa.eu/en/engage/careers



#### **Follow us on Twitter**

@efsa\_eu
@plants\_efsa
@methods\_efsa