

Big Data at the FDA Center for Veterinary Medicine

Errol Strain, PhD

FDA Center for Veterinary Medicine

Office of Research

1/6/2021

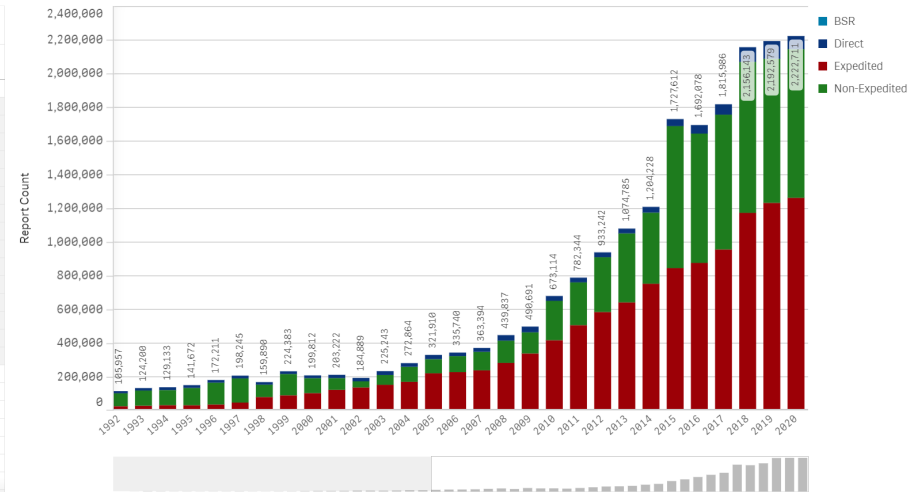
EMA Veterinary Big Data Stakeholder Forum

Examples of Big Data at the FDA

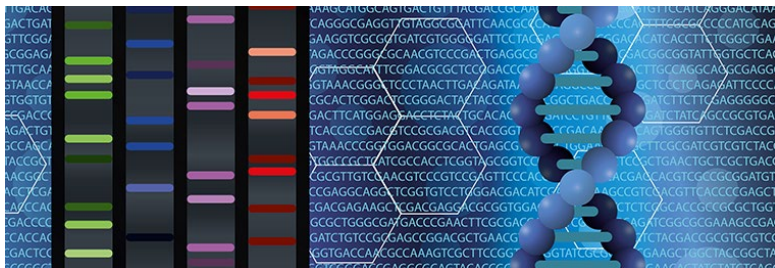
Safety Reports (kB to MB)

- MedWatch/Safety Reporting Portal
- FAERS ~ 2M reports/year

Reports received by Report Type

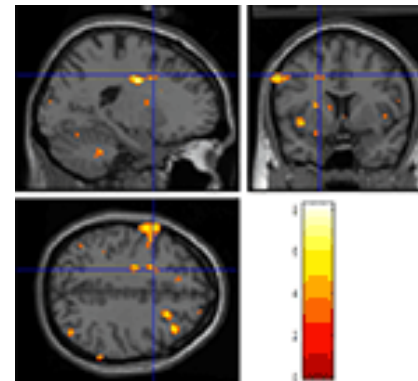


Sequence (GB to TB)



<https://www.dlapiper.com/~media/images/insights/publications/2016/nextgensequencingwebheader.jpg?h=257&la=en&w=759&hash=6E70B72BE6456CBD96EDAD0AD13A502EE598F933>

Images/Sensors (TB+)



https://www.frontiersin.org/files/Articles/1482/fpsyg-01-00035-HTML/image_m/fpsyg-01-00035-g001.jpg

Receiving and Storing Big Data

Electronic Submission Gateway

Business Hours Submissions

Application Type	Submission Size Range	Avg Submission Size (compressed)	Upload Duration (Avg; Target: 6 Hrs)	Deliver Receipt to User (Avg; Target: 2 Hrs)	Delivery to Centers (Avg; Ack2 to User Target: 2 Hrs 55 Mins)	Delivery Ack3 to User (Avg; Target: 5 Mins)	Delivery Ack3 to User (Avg; Target: 1 Hrs)	Total Duration (Avg; Target: 12 Hrs)
BLA	Up to 1 GB	0.03 GB	24 sec	2 sec	2 sec	7 mins	30 mins	37 mins
	1 GB to 5 GB	1.95 GB	40 mins	2 mins	1 mins	9 mins	24 mins	1.2 hrs
	5 GB to 10 GB	7.95 GB	3.8 hrs	3 mins	3 mins	13 mins	27 mins	4.9 hrs
	10 GB to 20 GB	18.51 GB	1.2 hrs	37 mins	14 mins	36 mins	2.8 hrs	5.4 hrs
	20 GB to 30 GB	No Data Available						
	30 GB to 40 GB	No Data Available						
DMF	Up to 1 GB	0.04 GB	3 mins	8 sec	2 sec	8 mins	32 mins	43 mins
	1 GB to 5 GB	1.39 GB	15 mins	39 sec	42 sec	7 mins	11 mins	33 mins
	5 GB to 10 GB	No Data Available						
	10 GB to 20 GB	No Data Available						
	20 GB to 30 GB	No Data Available						
	30 GB to 40 GB	No Data Available						
NDA	Up to 1 GB	0.03 GB	1 mins	7 sec	2 sec	7 mins	21 mins	30 mins
	1 GB to 5 GB	2.38 GB	32 mins	2 mins	59 sec	10 mins	47 mins	1.5 hrs
	5 GB to 10 GB	7.37 GB	1.2 hrs	10 mins	4 mins	16 mins	1.1 hrs	2.8 hrs
	10 GB to 20 GB	11.52 GB	2.8 hrs	20 mins	7 mins	18 mins	15 mins	4.1 hrs
	20 GB to 30 GB	No Data Available						
		30 GB to 40 GB	No Data Available					
	40 GB to 100 GB	61.67 GB	4.2 hrs	1.8 hrs	1.3 hrs	5.8 hrs	7 hrs	19.3 hrs
	100 GB and Above	118 GB	8.8 hrs	18.3 hrs	2.3 hrs	1.2 hrs	9.1 hrs	39.4 hrs

Genomics/Precision Medicine



Hard Drives



Computing on Big Data

High-Performance Computing

- Support key Priorities:
 - Precision Medicine Initiative
 - Cancer Moonshot
 - Combating Antibiotic Resistance Bacteria Initiative
 - Food Safety Initiative
- Support priority FDA programs:
 - Sentinel Initiative, Modernizing Toxicology, Improving Clinical Trial Design, Real-world evidence
- Support research to prepare for future challenges facing FDA

On-Premise (100s-1000s CPUs, GPUs)

Raven2



Center for Food Safety and Applied Nutrition

Betsy/Bluefin



HIVE



Cloud



Big Data at the Center for Veterinary Medicine

Sample of Regulatory Applications

1. * Review of Intentional Genomic Alterations (IGAs) in Animals
2. * Genomics/Metagenomics for Surveillance of Antimicrobial Resistance
3. Veterinary Environmental Transport and Fate Models for Predicting Environmental Concentrations (VETPEC)
4. Pharmacovigilance - monitor adverse events for approved drugs, pet foods, unapproved drugs, and veterinary devices
(<https://open.fda.gov/data/downloads/>) - JSON

1. Heritable Intentional Genomic Alterations in Animals

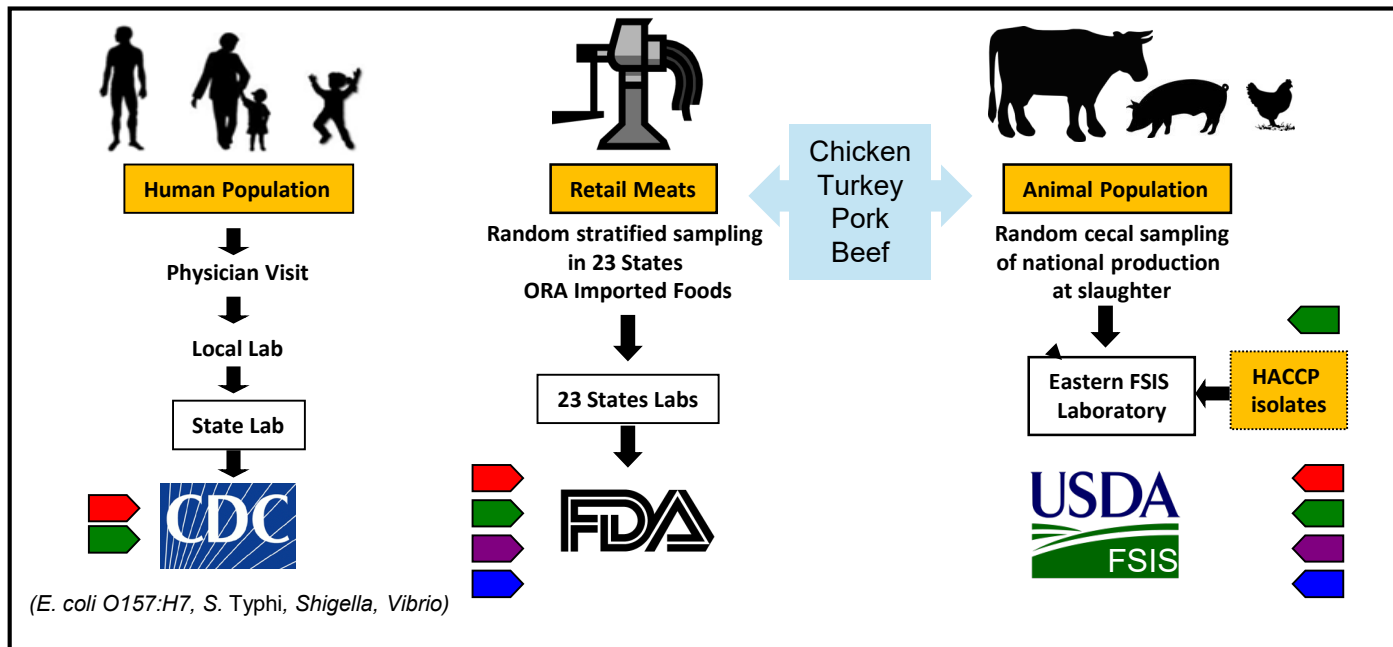
- Intentional Genomic Alterations (IGAs) in Animals
 - Generated using modern molecular technologies (*e.g.*, genetic engineering or genome editing)
 - Random or targeted genomic changes
- Types of big data that could be used to support molecular characterization include
 - Whole genome sequencing
 - Targeted sequencing

Molecular Characterization of IGAs in Animals

- Steps
 1. Look at the target site → Does it have the intended alteration?
 2. Screen for unintended alteration(s) → any on- or off- target alterations?
- Analysis
 - Compare to reference genome/sequences or *de novo* genome assembly (long read datasets)
 - Compare to non-IGA controls

Confirms intended alteration is present; identifies any unintended alterations that, if present, would need further analysis to determine if there is a safety concern

2. Genomics/Metagenomics for Surveillance of Antimicrobial Resistance



- Campylobacter*
- Salmonella*
- Enterococcus*
- E. coli*



NARMS Now: Integrated Data

National Antimicrobial Resistance Monitoring System

Home Antimicrobial Resistance by Year Resistance to Multiple Antimic... Multidrug Resistance Map of Resistance Number of Resistant >

Antimicrobial Resistance by Year ? [Back to Table of Contents](#)

1. Select a bacterium:

[Salmonella](#)
[Campylobacter coli](#)
[Campylobacter jejuni](#)
[Escherichia coli](#)
[Enterococcus faecalis](#)
[Enterococcus faecium](#)
[Enterococcus hirae](#)
[Enterococcus other](#)

2. Select a host:



3. Select a source and antimicrobial agent(s): ?

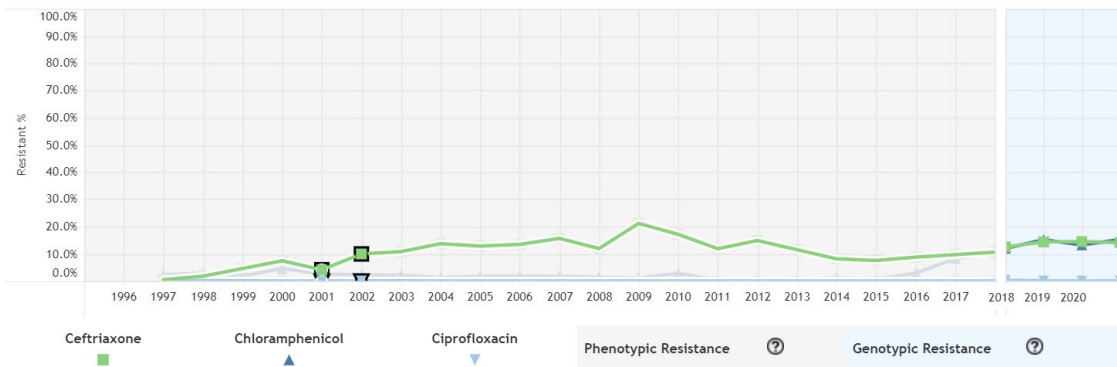
[Retail Chickens](#)
[Chickens \(Cecal\)](#)
[Chickens \(HACCP\)](#)

(Multiple values) ?

4. Select a serotype: (Only for Salmonella. No selection defaults to all serotypes) ?

[Anatum](#)
[Derby](#)
[Enteritidis](#)
[Hadar](#)
[Heidelbe...](#)
[4,\[5\],12...](#)
[Infantis](#)
[Johanne...](#)
[Kentucky](#)
[Montevid...](#)
[Newport](#)
[other](#)
[Reading](#)
[Typhimu...](#)

Resistance in *Salmonella* from all sources from chicken host



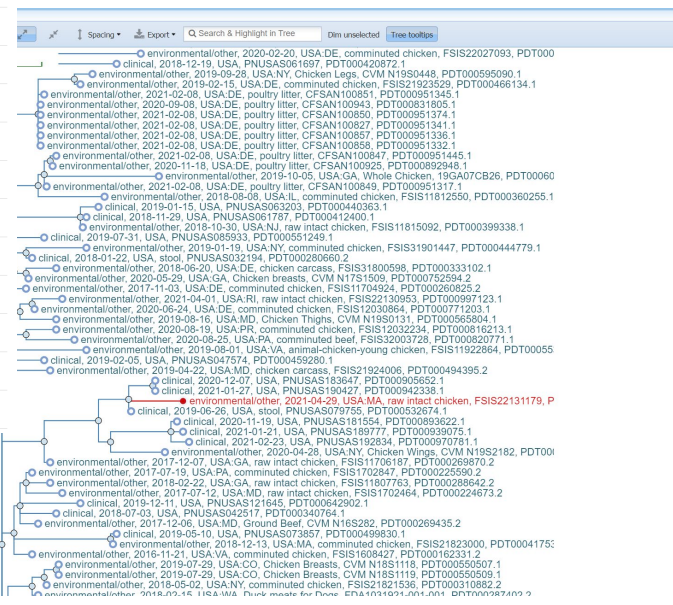
Serotype Distribution in 2002



NCBI Pathogen Detection

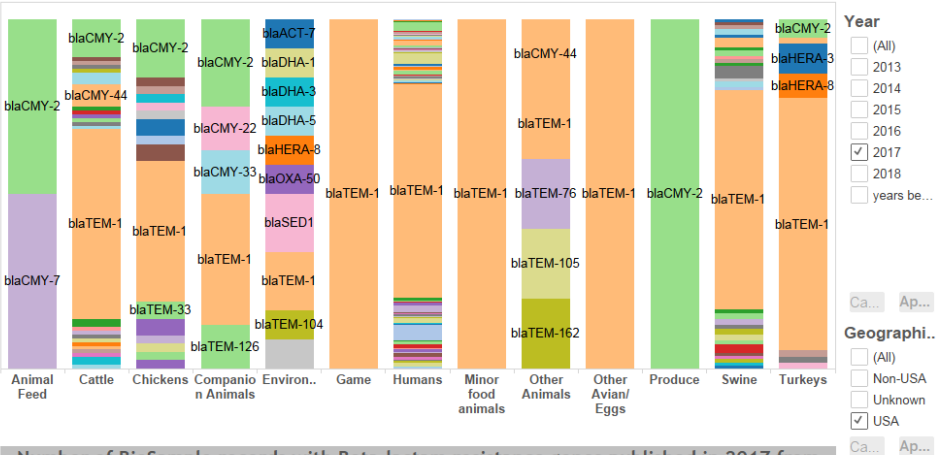
Organism Groups

Species	Version	Publication Date*	Latest Isolate Creation Date*	Clusters	New Isolates	New Clinical Isolates	New Environmental Isolates	Total Isolates
Salmonella enterica	PDG000000002.2196	2021-05-01 15:45	2021-04-29 18:28	17,711	546	497	49	358,597
E.coli and Shigella	PDG000000004.2611	2021-05-01 16:06	2021-05-01 01:15	17,173	101	39	62	175,489
Campylobacter jejuni	PDG000000003.1382	2021-05-01 09:59	2021-04-30 12:27	4,676	36	1	35	62,095
Listeria monocytogenes	PDG000000001.2196	2021-05-02 06:48	2021-05-02 00:28	3,684	1	1	0	42,754
Klebsiella pneumoniae	PDG000000012.914	2021-05-01 08:53	2021-04-30 16:00	3,676	22	20	2	40,242
Staphylococcus aureus	PDG000000073.106	2021-05-01 08:39	2021-04-29 18:02	2,731	314	313	1	26,404
Enterococcus faecium	PDG000000071.152	2021-05-01 08:06	2021-04-30 11:27	903	7	4	3	19,290
Neisseria	PDG000000032.212	2021-04-26 06:26	2021-04-25 09:43	1,228	1	1	0	15,907
Acinetobacter baumannii	PDG000000010.555	2021-05-02 06:11	2021-04-30 15:49	691	0	0	0	12,050
Pseudomonas aeruginosa	PDG000000036.446	2021-05-01 07:57	2021-04-28 10:16	947	4	4	0	11,844
Mycobacterium tuberculosis	PDG000000034.205	2021-04-26 06:33	2021-04-24 22:06	764	2	2	0	8,484
Enterococcus faecalis	PDG000000072.123	2021-05-01 07:50	2021-04-30 12:27	955	14	0	14	6,959
Vibrio cholerae	PDG000000055.187	2021-04-22 05:55	2021-04-21 09:42	202	5	0	5	5,552
Enterobacter	PDG000000028.449	2021-05-01 06:02	2021-04-28 10:45	539	2	0	2	5,120
Clostridioides difficile	PDG000000045.186	2021-04-12 19:45	2021-03-27 22:07	188	0	0	0	4,468
Vibrio parahaemolyticus	PDG000000023.439	2021-04-24 07:05	2021-04-23 06:54	371	1	0	1	4,274
Legionella pneumophila	PDG000000026.120	2021-04-02 08:04	2021-04-01 09:20	181	12	0	12	4,098
Staphylococcus pseudintermedius	PDG000000042.201	2021-04-25 06:54	2021-04-24 07:24	103	23	0	23	2,126



RESISTOME TRACKER

Salmonella



Number of BioSample records with Beta-lactam resistance genes published in 2017 from USA regions

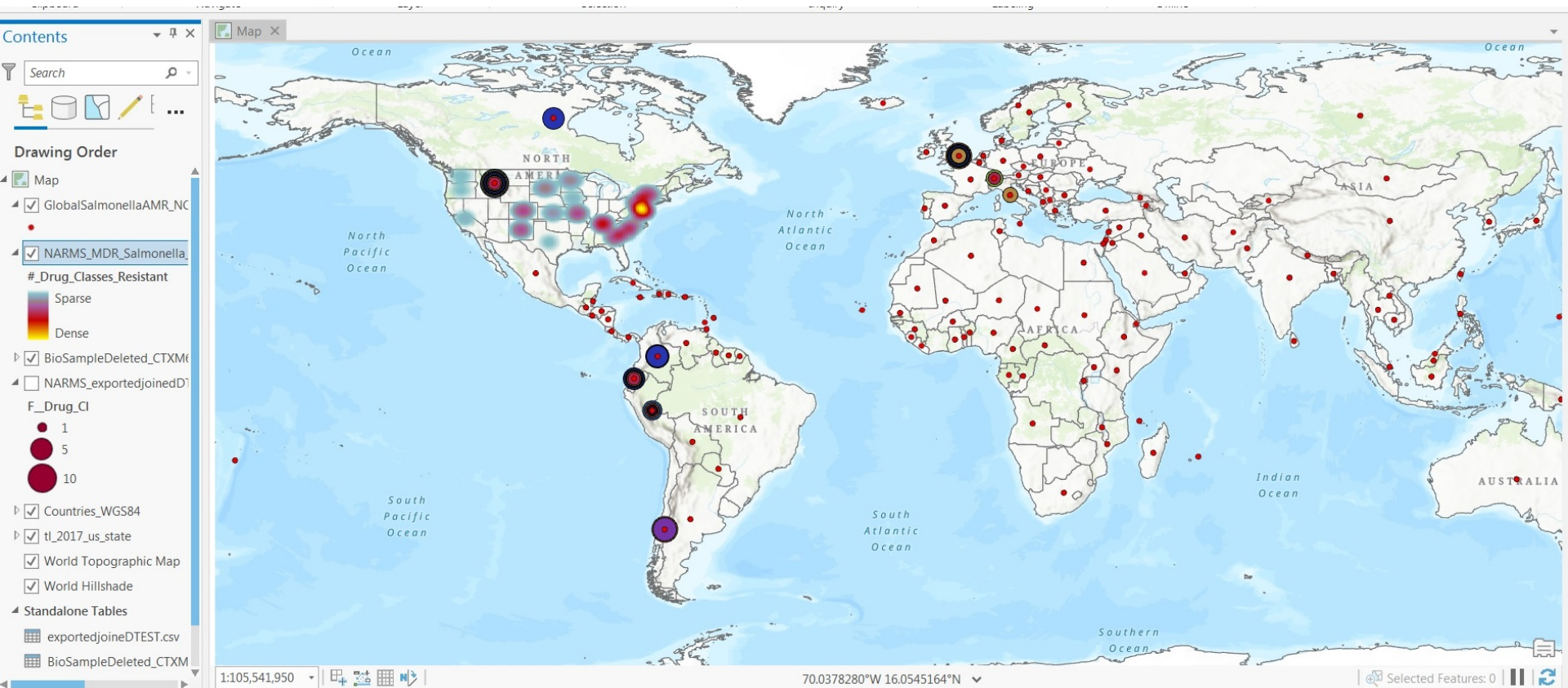
Animal Feed	Cattle	Chickens	Companion Animals	Environmental	Game	Humans	Minor food animals	Other Animals	Other Avian/Eggs	Produce	Swine	Turkeys
2	83	42	7	12	1	591	1	5	4	1	103	57

Total number of distinct Biosample records published in 2017 from USA regions

Animal Feed	Cattle	Cheese/Dairy	Chickens	Companion Animals	Environmental	Game	Humans	Minor food animals	Nuts	Other Animals	Other Avian/Eggs	Produce	Reptiles	Seafood	Seeds, Spices, Herbs	Swine	Turkeys
200	1,456	10	3,942	110	1,806	7	18,816	10	105	61	80	186	54	20	35	1,704	522

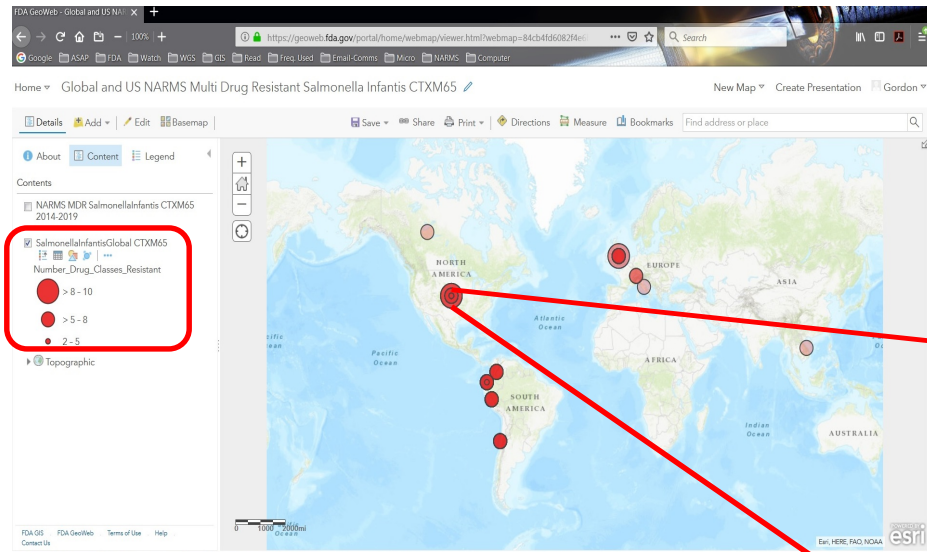
Pettengill JB, et al. Distribution of antimicrobial resistance genes across *Salmonella enterica* isolates from animal and non-animal foods. *Submitted*.

Emergence and Geographic Distribution of Multi Drug Resistant S. Infantis CTX-M-65



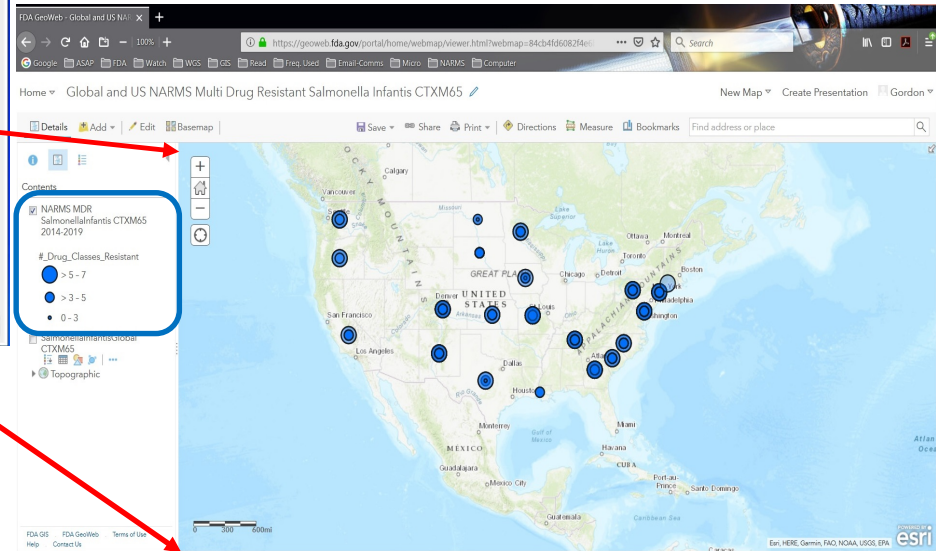
Multidrug Resistant Salmonella Infantis with beta lactamase resistant CTX-M-65 gene 2010 - 2019

Global 2010 - 2019



Zoom in to visualize and analyze data of interest at different levels

NARMS Retail Meat 2014 - 2019



Metadata is available for isolates of interest

The screenshot displays the ArcGIS Pro interface with a map of the United States. A red arrow points from a specific location in Maryland on the map to a pop-up window. The pop-up window shows the following metadata for the isolate:

Attribute	Value
OBJECTID	11
#Organism Group	Salmonella enterica
Serovar	<Null>
geoloc	USA
geoloc_state	MD
geoloc_city	<Null>
Isolation_Source	Turkey
Isolation_Source_Part	Ground Turkey
Collection_Date	2019
Collection_Year	2019
AMR genotypes	aac(3)-IV,aadA1,aph(4)-Ia,blaCTX-M-65,qacEdelta1,sul1,tet(A)
aac(3)-IV	aac(3)-IV
aadA1	aadA1
aph(3')-Ia	<Null>
aph(4)-Ia	aph(4)-Ia
Amino_resistant	1
blaCTX-M	<Null>
blaCTX-M-65	blaCTX-M-65
blaTEM-1	<Null>
blaCTX-M65_present	1
Betalactam_resistant	1
trimethoprim	<Null>
trim_resistant	0
phenicol	<Null>
phenicol_resistant	0
fosfomycin	<Null>
fos_resistant	0
sulphonamide	sul1
sul_resistant	1
tetracycline	tet(A)
tet_resistant	1
#_Drug_Classes_Resistant	4
MDR_greater-equal_5	0
TOTAL #_AMR_Genes	6

Thank You

