Versatile Emerging infectious disease Observatory
Forecasting, nowcasting and tracking in a changing world

VEO Symposium 2022 Report

17-18 May 2022

This work is supported by European Union's Horizon 2020 research and innovation programme under Grant No. 874735 (VEO).

Contents

Overview

The VEO Symposium 2022 was held in The Hague, The Netherlands, 17-18 May 2022. An online link was shared with participants who could not attend. Approximately 110 people participated in the Symposium.

On 16 May 2022, 'Young VEO' held its own meeting. There were oral presentations as well as posters. The posters were available during the VEO Symposium as well. Sidsel Nag, DTU, won best poster.

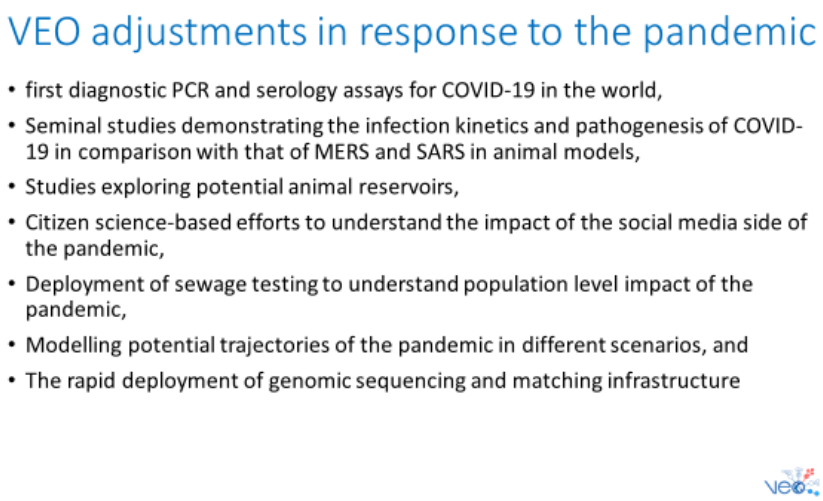
This was the first in-person meeting of the VEO Consortium. The meeting brought together the Principle Investigators, WP Leaders and Co-leaders, participants, as well as stakeholders, EC Officer and other guests, such representatives from other relevant research projects.

Marion Koopmans, EMC, VEO – Welcome and Recap of the Project

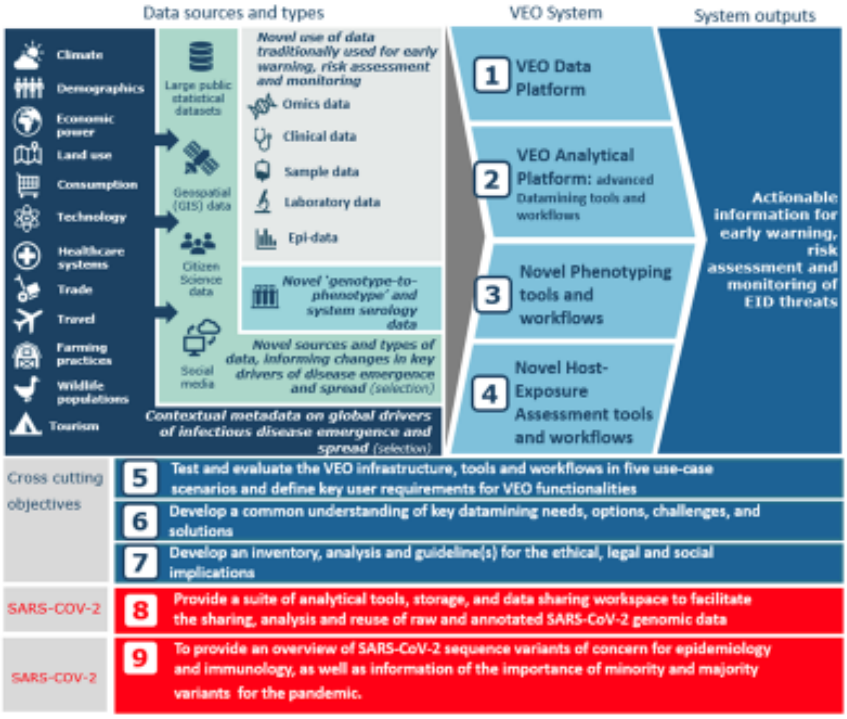
The question presented was '(How) can data science and technology innovations improve infectious disease preparedness and response?'

VEO aims to answer the following key questions:

- Technology perspective
  - Can we use the vast amounts of public data for prediction, early warning?
- User perspective
  - Can we predict outbreaks? How can we use the novel tools to be better prepared?
- Ethical, legal & societal perspective
  - Which are the barriers to the use of big data combined with citizen science? How do we resolve potential barriers?



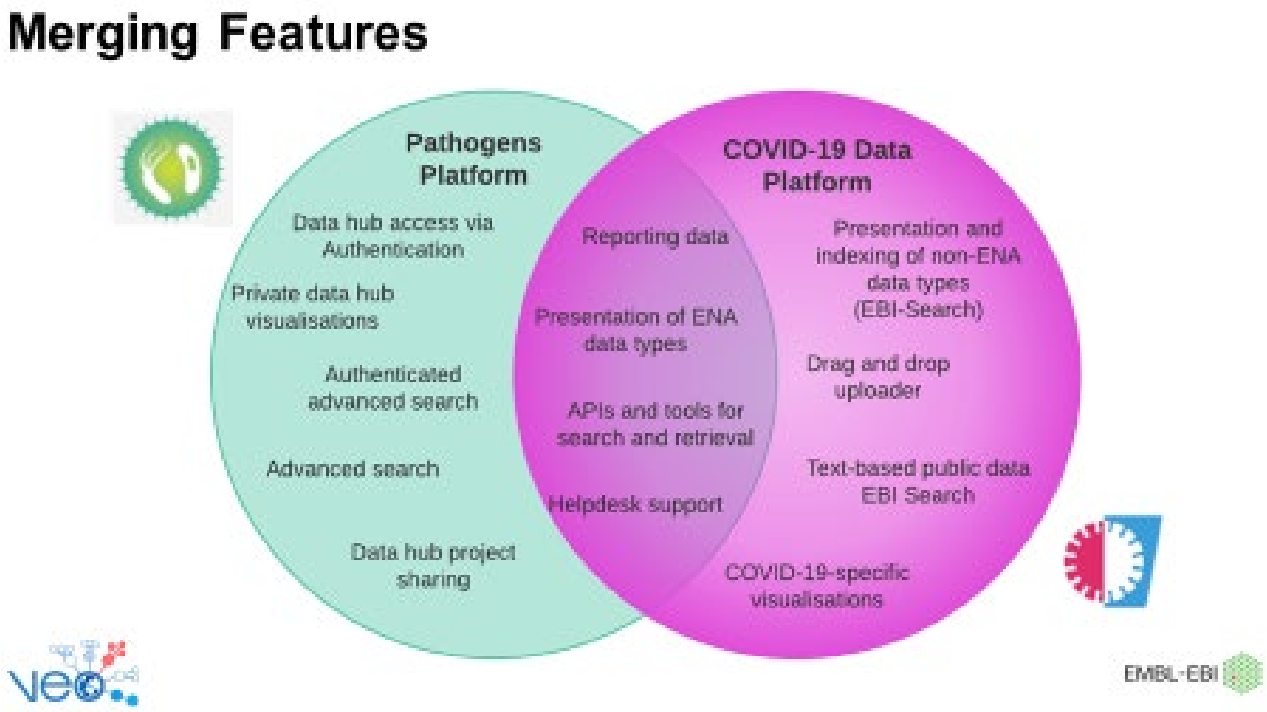
Though VEO made adjustments to its research plan due to the COVID-19 pandemic (at the request of the EC), VEO has had many results over the past year (+) from of the original VEO research plan. These results were reviewed in the Welcome presentation.


Data sources and types
VEO System
System outputs
Climate
Demographics
Economic power
Land use
Consumption
Technology
Healthcare systems
Trade
Travel
Farming practices
Wildlife populations
Tourism
Large public statistical datasets
Geospatial (GIS) data
Citizen Science data
Social media
Novel use of data traditionally used for early warning, risk assessment and monitoring
Omics data
Clinical data
Sample data
Laboratory data
Epi-data
Novel 'genotype-to-phenotype' and system serology data
Novel sources and types of data, informing changes in key drivers of disease emergence and spread (selection)
Contextual metadata on global drivers of infectious disease emergence and spread (selection)
1 VEO Data Platform
2 VEO Analytical Platform: advanced Datamining tools and workflows
3 Novel Phenotyping tools and workflows
4 Novel Host-Exposure Assessment tools and workflows
Actionable information for early warning, risk assessment and monitoring of EID threats
Cross cutting objectives
5 Test and evaluate the VEO infrastructure, tools and workflows in five use-case scenarios and define key user requirements for VEO functionalities
6 Develop a common understanding of key datamining needs, options, challenges, and solutions
7 Develop an inventory, analysis and guideline(s) for the ethical, legal and social implications
SARS-COV-2
8 Provide a suite of analytical tools, storage, and data sharing workspace to facilitate the sharing, analysis and reuse of raw and annotated SARS-CoV-2 genomic data
SARS-COV-2
9 To provide an overview of SARS-CoV-2 sequence variants of concern for epidemiology and immunology, as well as information of the importance of minority and majority variants for the pandemic.

VEO Cross WP – Science Update: News, Achievements and Future Plans Part I

Nadim Rahman, EMBL-EBI, WP01 VEO Data Platform

Nadim presented the pathogens platform, including the COVID-19 Portal and how they are merging features from the existing infrastructures. The integration of datasets was presented, with Mosquito Alert data as an example. The concept of 'data hubs' is being extended to other data types. This requires new data examples.



Ron Fouchier, EMC, WP03 Novel Phenotyping Tools and Workflows

The presentations were broken down as follows:

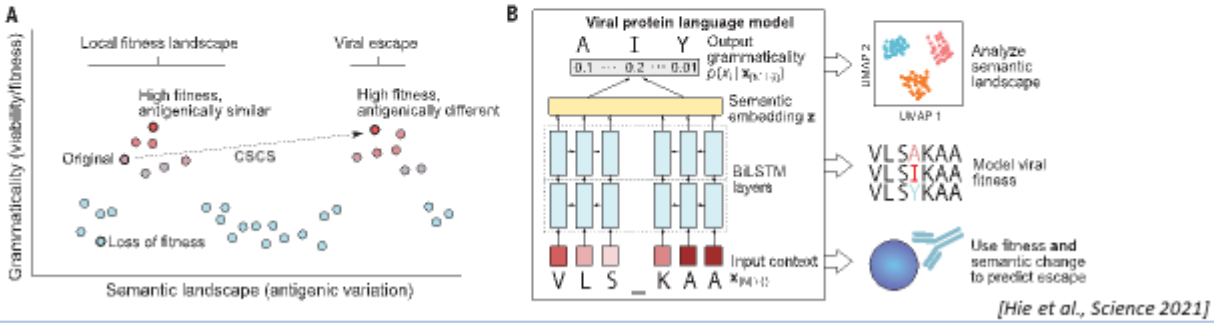
- Miranda de Graaf, EMC: Norovirus polymerase mutations
  - Test more GII.P16 polymerases
  - Work on reproducibility of the assays
  - Replicate GII.P16 viruses in organoid cultures
  - Measure polymerase activity with full ORF1
- Saria Otani, DTU: Microbial virulence; cloning the Global Sewage Metagenome
  - Expand on fosmid libraries - Test other putative bacterial hosts (than *E. coli*). Different hosts have impact on expression levels – phenotypic characteristics.
  - Identify the potentially novel AMR genes based on functional cloning (not in sequencing data).
  - All are plasmid-derived clones. Genomic DNA based functional cloning currently in progress.
  - Phage-derived AMR/virulence genes in sewage and middens.
    Phage isolation from natural source – incubate with bacterial hosts – test for haemolysis
- Daniel Remondini, UNIBO: Artificial Intelligence on viral protein sequences
  - Which informative content do they have (e.g. "neutral" proteins)?
  - How much the AI model is protein/virus specific? Can we use the model trained one protein/virus to extract information on about protein/virus?

Besides the presentation, other interesting updates from the WP include the following:

- Flavivirus (TBEV-Eur&Sib; WNV-lin1&2; ZIKA); molecular clones & phenotyping (ZIKA)
- SARS-CoV-2 fitness studies
- Antigenic variation of influenza virus, norovirus & SARS-CoV-2
- Carbapenem-resistance of *Klebsiella pneumoniae* in Greece; 66.3%!
- Avian influenza; reassortment compatibility (HA, NA); host-range barriers; H5 virulence/fitness
- Cell culture systems (e.g. neural, respiratory) and organoids (brain, respiratory tract, intestinal tract)
- Daniel Remondini is undertaking a network analysis of social media data.

Protein sequence "embedding" in vector space

- Each protein sequence (eg HA1) is "transformed" into an N-dim (1024) vector
- Training: reconstruct sequence estimating missing aminoacid in the sequence p(AA)
- Vector space: calculate sequence "distances" (different from sequence alignment)
- Sequence "grammaticality" (AA probability) and "immune escape" (vector distance)



[Hie et al., Science 2021]

Versatile Emerging Infectious Disease Observatory

This work is supported by European Union's Horizon 2020 research and innovation programme under Grant No.874735 (VEO).

Marion Koopmans, EMC, WP04 Novel Host Exposure Assessment Tools and Workflows

Marion presented the serology story covering SARS-CoV-2 and Norovirus in VEO.

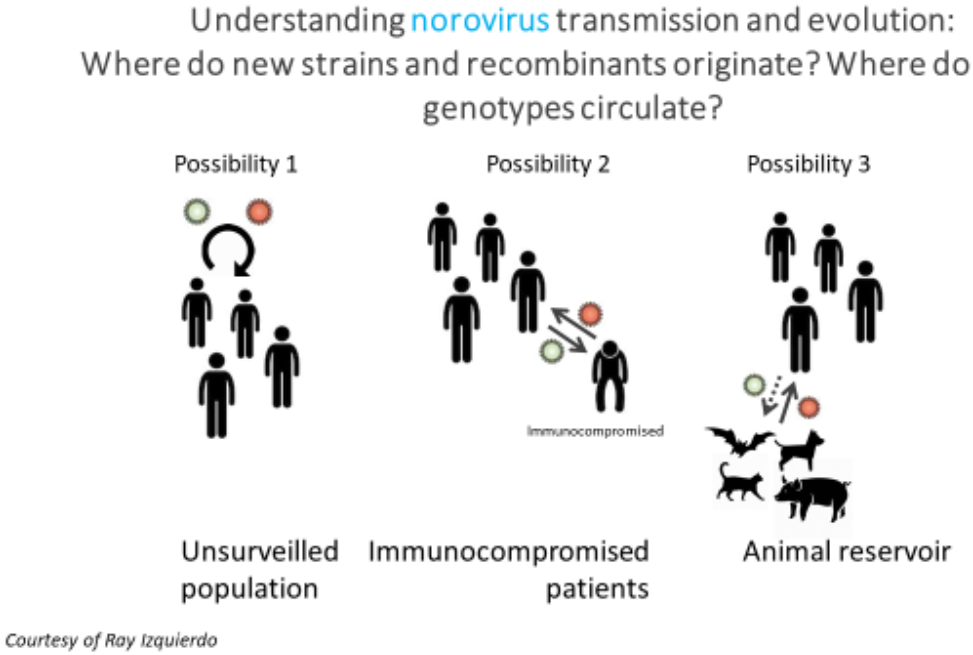
Serology has a crucial role in assessing key properties of an emerging infection. There are several key questions to ask:

- How can I detect it?
- What is the reservoir?
- Who can be infected?
- What is the mode of transmission?
- How contagious is it?
- How widespread is it?
- How severe is it?

Screening tools for SARS-CoV-2 developed (ELISA and generic homogenous assay for antibodies).

Next steps in the WP include the following:

- Arbovirus and influenza virus serology in Greenland expedition
- Screening of animal reservoirs for (exposure to) human noroviruses
- A range of ongoing long term follow-up studies of SARS-CoV-2 antibody responses, quality and longevity
- Inclusion of serology in West Nile emergence pathway assessment
- Differential serology West Nile – Usutu in wild birds
- Exploring antibody landscapes for arboviruses

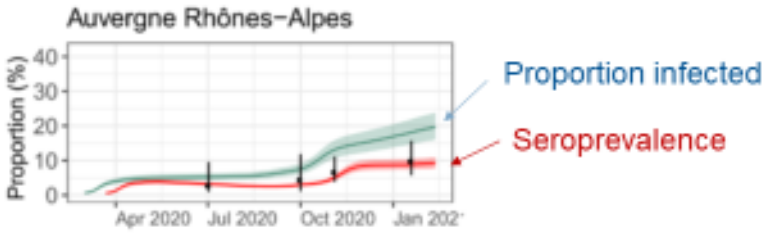


Courtesy of Ray Izquierdo

Simon Cauchemez, IP, WP04 Using serology to assess age-specific mortality and immunity patterns of SARS-CoV-2: Then and now

Use of serology to track levels of infection in population later on

- Interpretation of seroprevalence studies quickly became more difficult due to antibody decay.



[Gallian et al, submitted]

- SARS-CoV-2 immunity has become very complex
  - A lot of different immunity profiles depending on: vaccination status, past infections and infecting variants, age…
  - We want to reconstruct the diversity of immunity profiles for France:
    - E.g. What is the proportion of individuals aged 60-70 y.o. that received a vaccine boost and were infected by Omicron?
    - So far, assessment based on case, hospitalisation, death and vaccination data.
    - Could serology help? Can it provide more granular characterization than just seropositive/seronegative? SARS-CoV-2 antibody landscapes?

Institut Pasteur

Frederic Bartumeus, CSIC, WP05 Mosquito Borne Diseases

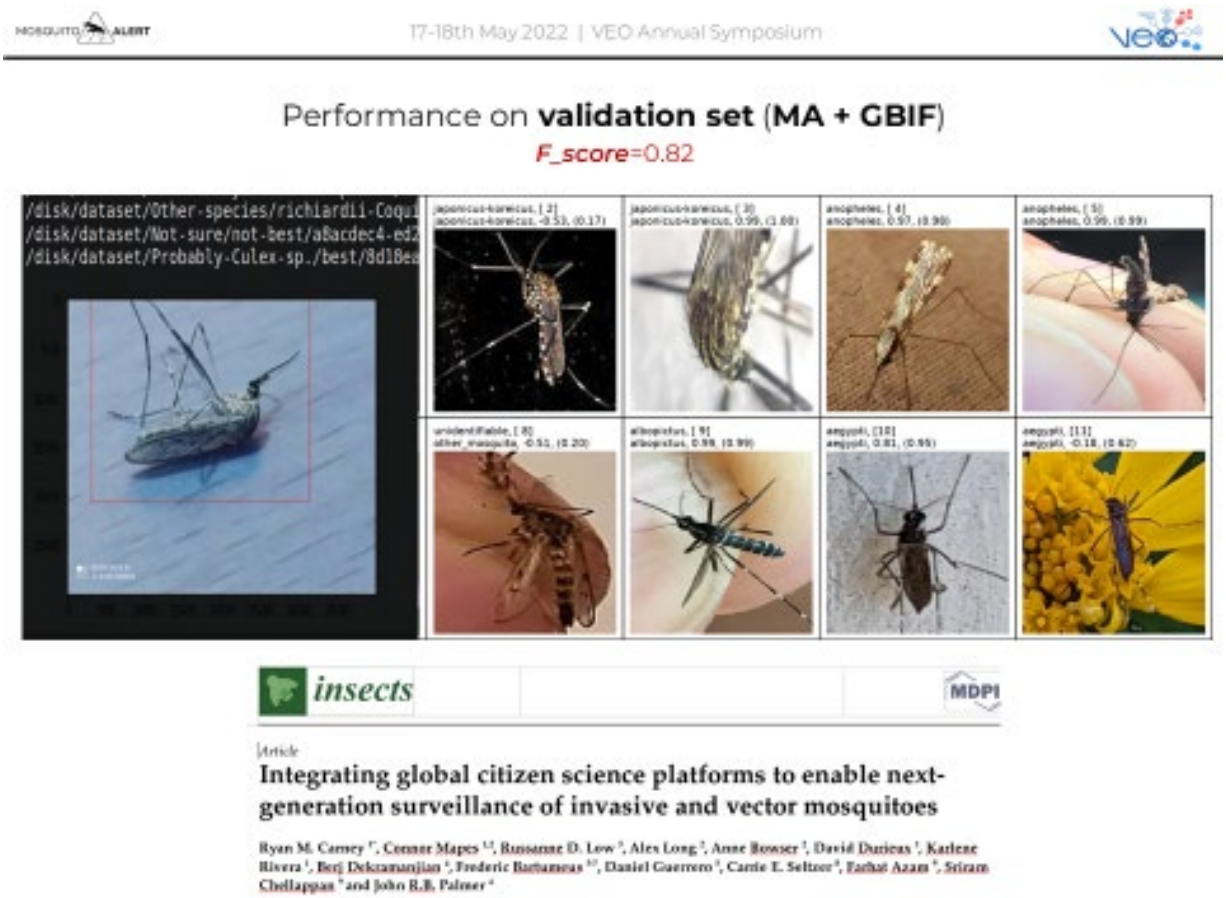
The work includes vector field sampling, pathogen field sampling and citizen-based and digital sampling.

The WP consists of the following studies:

- Vector and pathogen surveillance in Italy: entomological surveillance of WNV and USUV and surveillance of invasive mosquitoes
- Vector and pathogen surveillance in Greece: Study WNV+ rates in *Culex* spp. in EU, Metagenomics study on vectors, WNV Phylogeny Group (VEO partners), WNV sampled & tested in *Culex pipiens*
- Vector surveillance in Spain: Catalonia, NE Spain, *Aedes albopictus,* South Spain, WNV-*Culex* sp.

The WP also plans to start bird data analysis: extend WNV model running at APHA with novel data streams and bird data

In addition, an update on the Mosquito Alert app and AI identification of mosquito species was given.

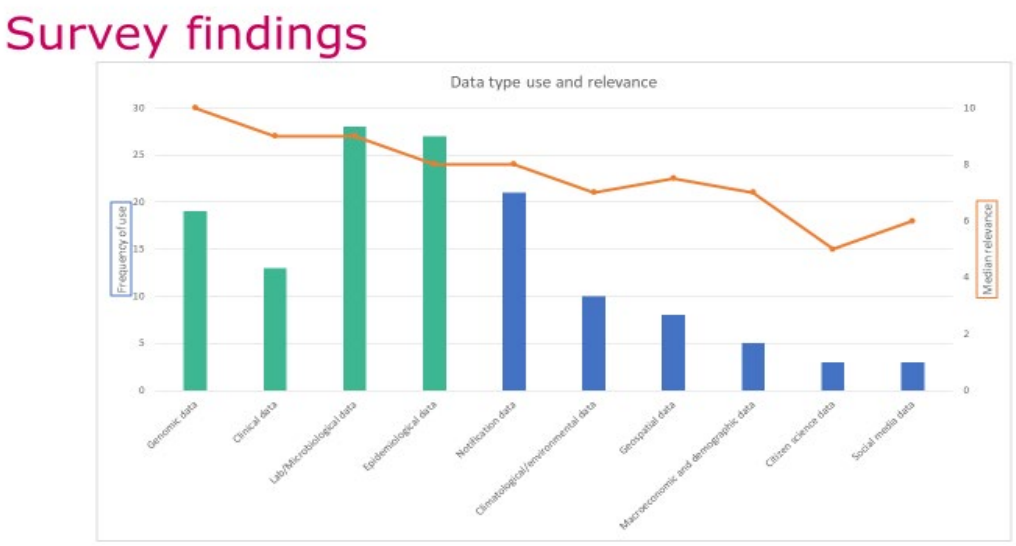


A Way with Cases about: COVID-19, Privacy and Twitter Infodemic

George Haringhuizen, RIVM, WP10 Timeline and ELSI Survey Findings

The development and results of the survey conducted by WP10 were presented.

The extended description of the study, analysis and results are part of the "Report on ELSI constraints and opportunities for processing diverse datasets within VEO" (EC-portal Deliverable 10.1 Part I, chapter 2 and 3 + Annex II).



- Generally, barriers were reported for all data types, and at every stage of working with data but by minority of respondents
- Exceptions, i.e. situations where a majority encountered barriers, were:
    - *Accessing clinical data: 69% encountered barriers (n=13)*
    - *Sharing genomic data: 57% encountered barriers (n=19)*
- *Privacy*, GDPR and related frameworks, and institutional policies were more frequently reported as barrier types

Conclusions

- ELSI barriers are ubiquitous, but not there all the time
- Barriers can be embedded in the system and hard to solve due to legislation (GDPR, agreements, consent, etc),
- OR more practical with easier solutions (lack of knowledge, access to services, bureaucratic or technical processes)

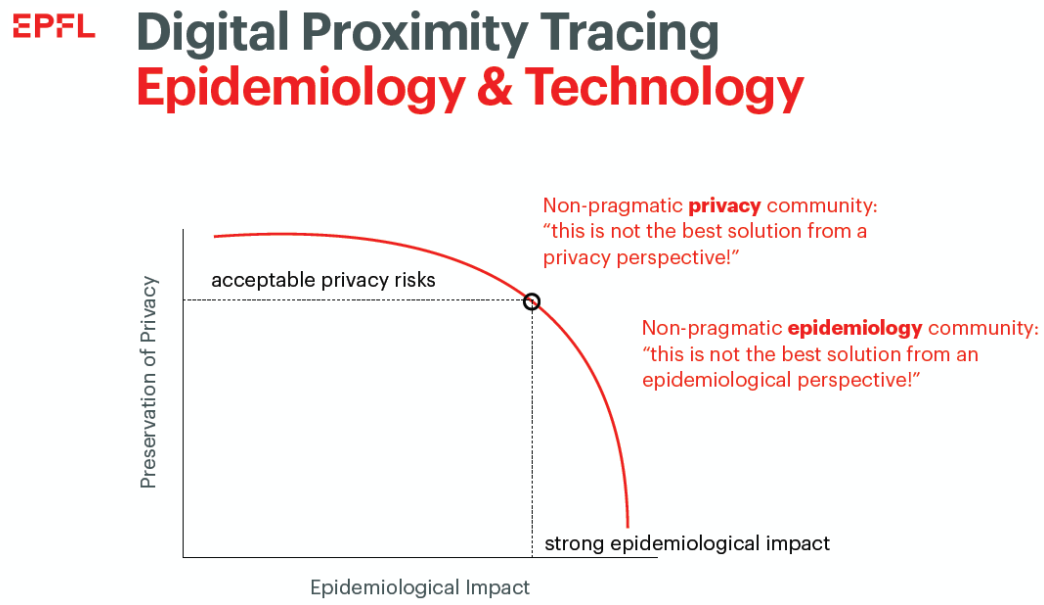
- Strength: coverage of work backgrounds, use of data types
- Limitation: *lack of in-depth description of barriers in this survey, less representation of contextual data types [outside the bio/medical inf.-disease domain] to be merged with bio-data.*

14

Marcel Salathé, EPFL, WP02 Digital Epidemiology: Data & Privacy

Looking at the 'modern' approach to epidemiology, what modern tools do we use: mobile phones to track disease outbreaks; Wikipedia for influenza nowcasting; and Twitter for vaccine sentiment and uptake.

For COVID-19 pandemic, digital proximity tracing (digital contact tracing) was used to track the spread, but also to notify the public of exposure. But what about privacy and digital proximity tracing?



VEO & COVID-19 – Genomics and Data Sharing (WP15 & WP16) Part I

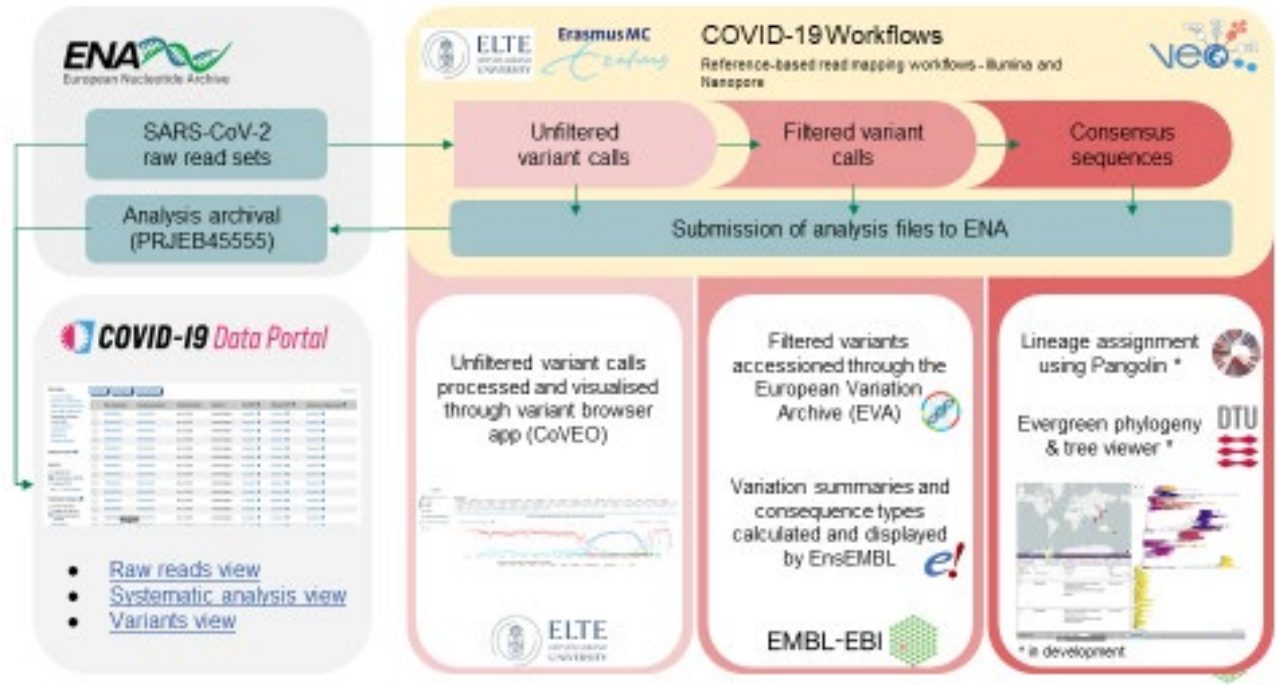
Nadim Rahman, EMBL-EBI, WP15 VEO Covid-19 Response

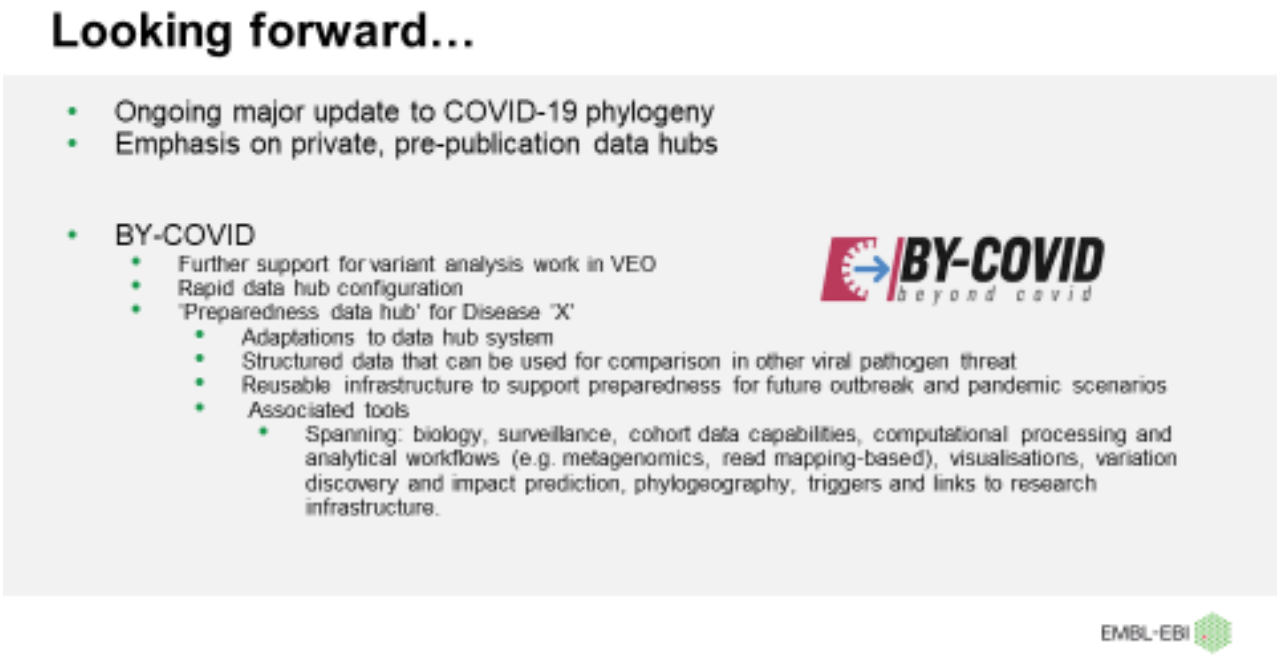
WP15 should mobilize analysis, mobilize data and enhance access.

The European COVID-19 Data Portal allows systematic analysis of raw datasets, but there are challenges to the analysis of data at scale.

Nadim presented an overview of activities within WP15, including tasks - mobilize analysis, mobilize data and enhance access. There was particular focus on the systematic analysis of public raw datasets in the COVID-19 Data Portal, presenting also the challenges of analyzing data at scale.'

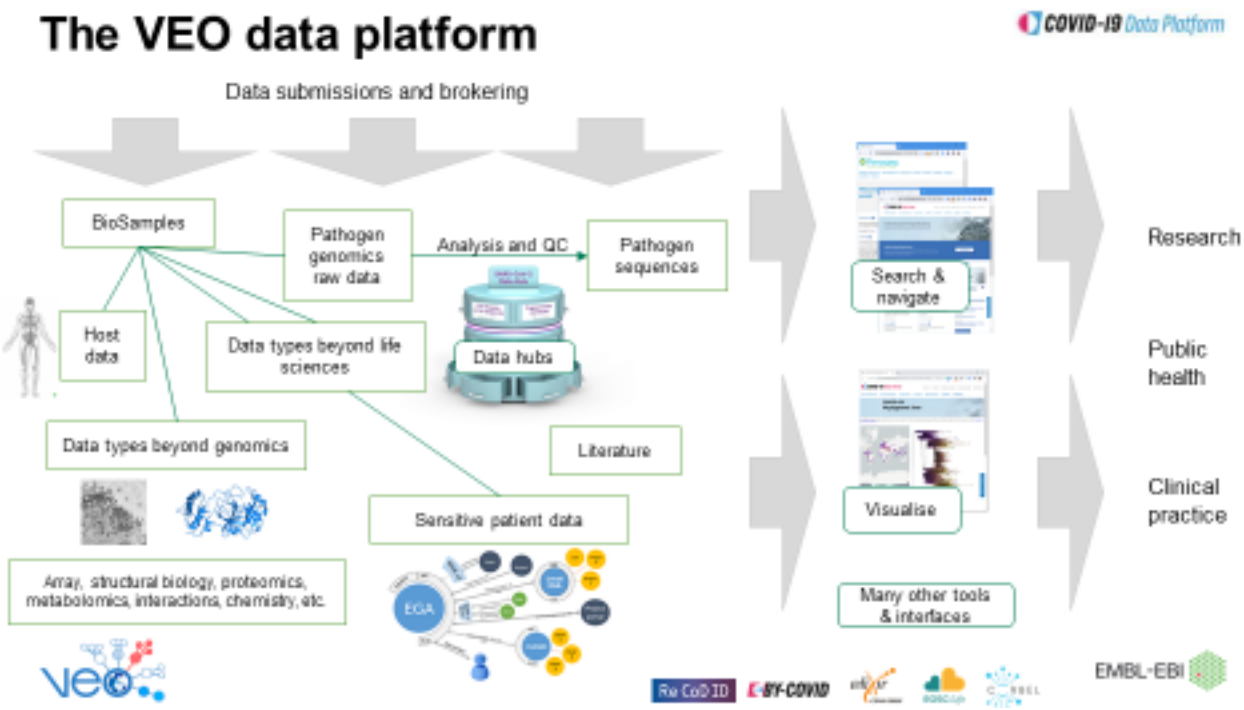
The presentation included how the European COVID-19 Data Portal and activates within WP15 of VEO also supports the BY-COVID project with future focus.





Guy Cochrane, EMBL-EBI, WP15 Sharing and Connecting Data

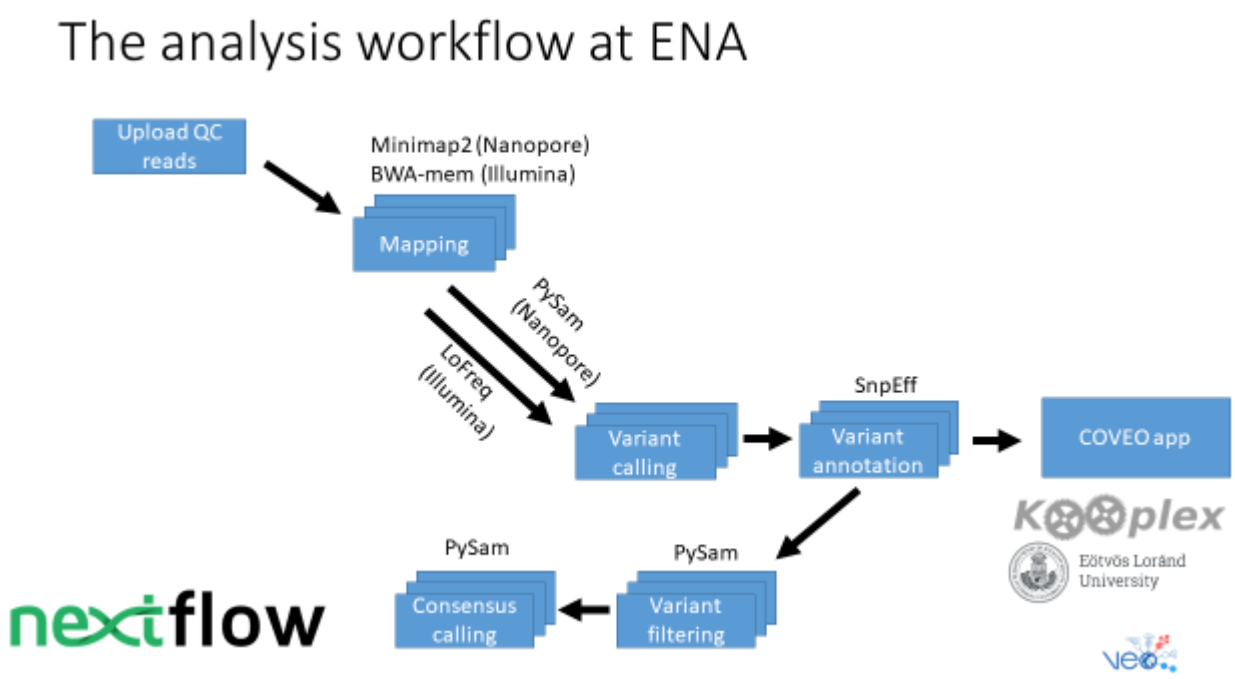
Looking at the how data are shared. There are reference data sets, tools, contextual data and human sensitivity and response data.



Nathalie Worp and David Nieuwenhuijse, EMC, WP15 SARS-CoV-2 Variant Tracking and Real-time Data Analysis

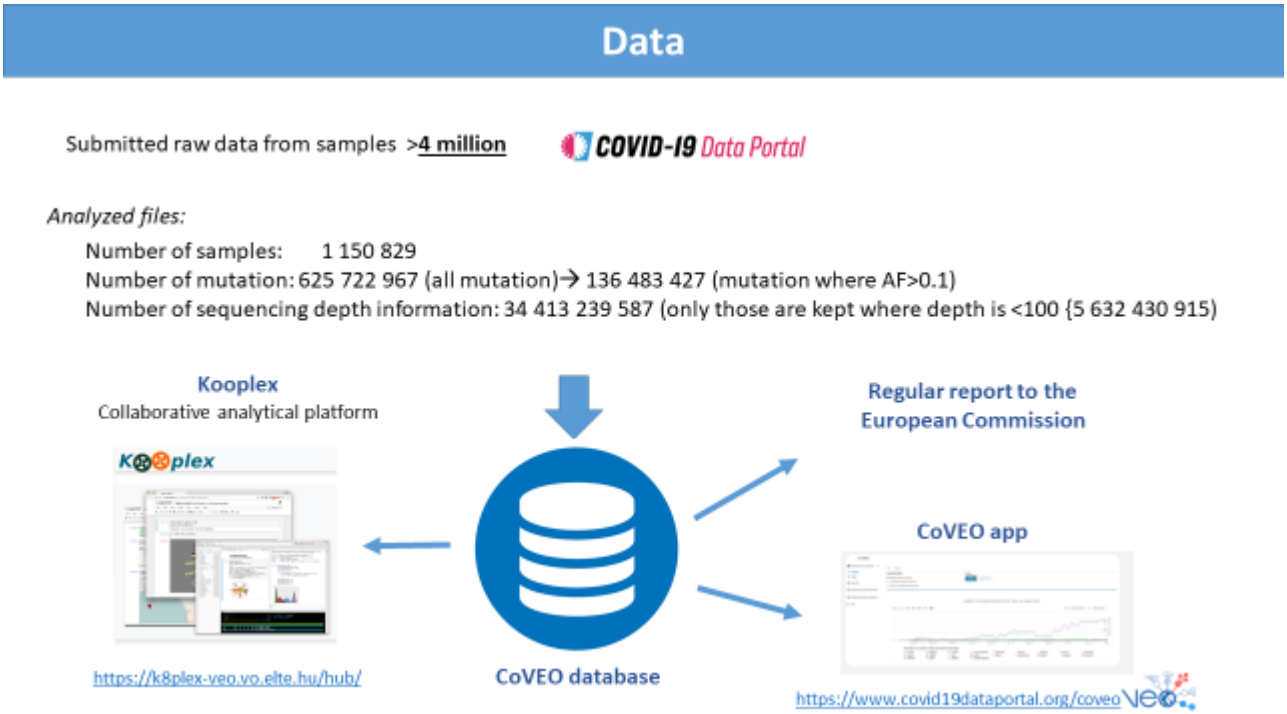
All viruses mutate and change over time. The emergence of new SARS-CoV-2 variants was expected; the more animal and human infections, the more mutations. In VEO, we used different studies to generate data for genotype to phenotype prediction:

- Comparative VNT
- Comparative organoid infection
- Competition experiments
- Animal infection experiments
- Antigenic Cartography
- T-cell response
- Vaccination response



Krisztián Papp, ELTE, WP16 CoVEO SARS-CoV-2 Variant Browser

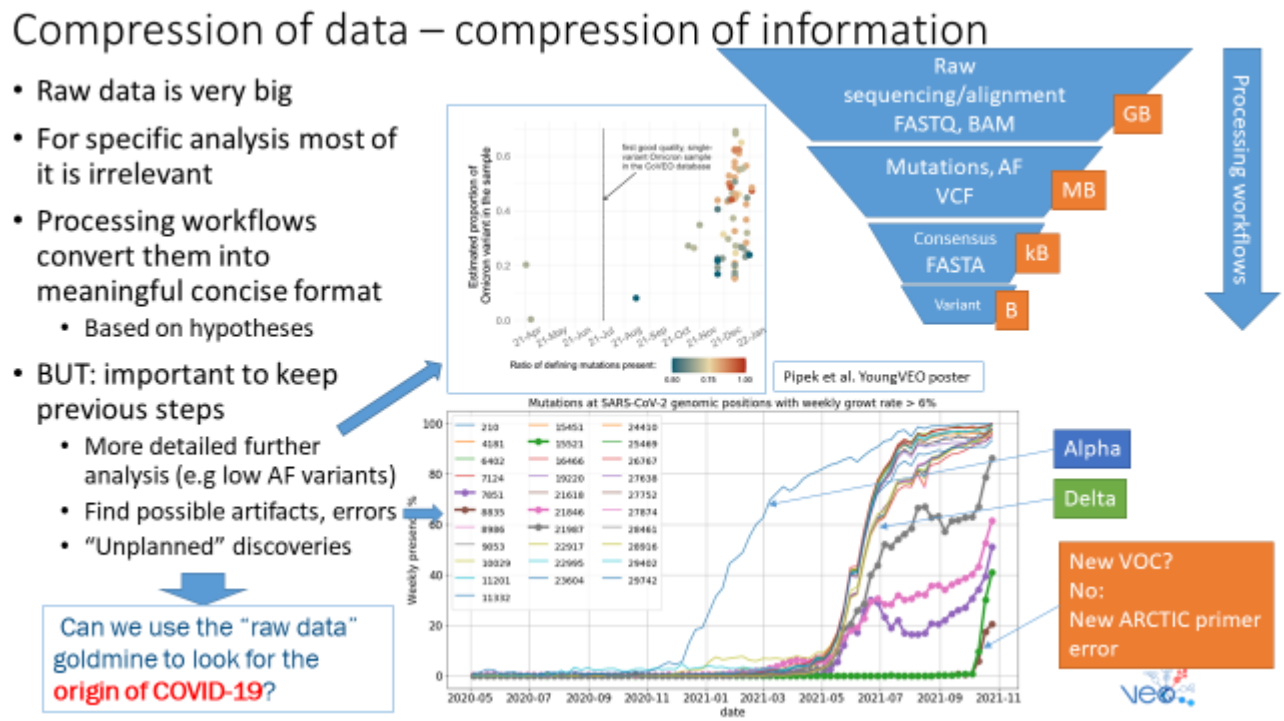
The CoVEO app interprets and summarizes the variation data produced by pipelines. Here, users can explore the emergence, spread and incidence of SARS-CoV-2 variants across the globe to give a view of the status of the pandemic. This app can be accessed by clicking the 'Variant Browser' links throughout the COVID-19 Data Portal, or by visiting: https://covid19dataportal.org/coveo.



VEO & COVID-19 – Genomics and Data Sharing (WP15 & WP16) Part II

István Csabai, ELTE, WP15 The Power of Raw Data

There are many types of SARS-CoV-2 data archives: SARS-CoV-2 consensus genome sequences, sequencing raw data, and phylogeny and epidemiology data archives. With raw data sets, usually they include indexed metadata.



Judit Szarvas, DTU, WP15 Data Visualisation with PhyloDash

SARS-CoV-2 is a slowly evolving RNA virus and whole-genome sequencing of its genetic material allows the prediction of the evolutionary path of the virus. An analysis workflow has been developed to infer the genetic relatedness of the submitted consensus sequences.

The result of this analysis is supplemented with user-submitted metadata, PANGO-lineage classification and variants detected in the N and S genes of SARS-CoV-2, and displayed in an interactive app in the COVID-19 Data Platform under Phylogeny. Through this app, called PhyloDash, (https://bitbucket.org/genomicepidemiology/phylodash), the users can search for their sample(s) of interest, or filter on metadata, variant(s) or lineage, and highlight these samples in the phylogenetic tree. By default, the view is of the global tree, but regional trees are available with the drop down menu in the footer of PhyloDash.

Tree visualisation with PhyloDash


Martin Koliba

Phylogenetic analysis pipeline publishes the results in a web-application, built with Next.js
https://bitbucket.org/genomicepidemiology/phylodash

European WHO-region,
May 2022

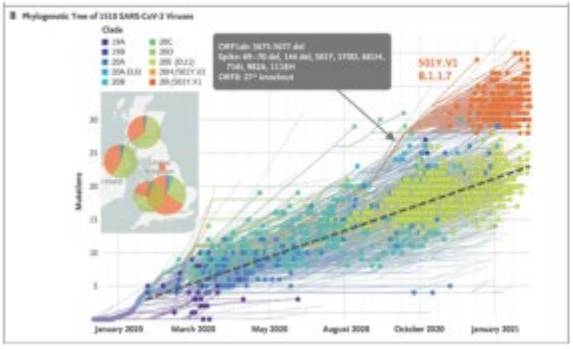
Bas Oude Munnink, EMC, WP16 Monitoring the Emergency of Novel SARS.CoV-2 Variants

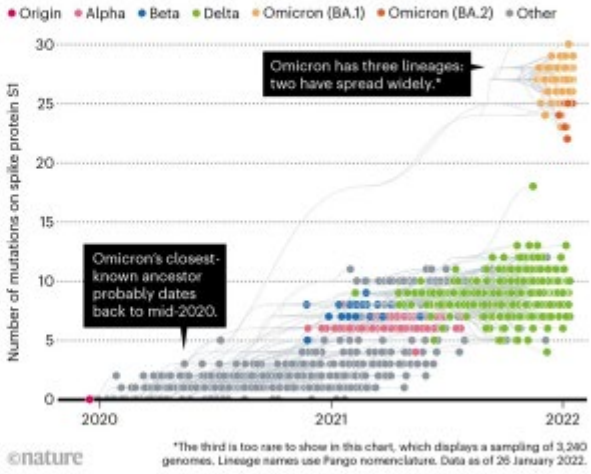
The COVID-19 pandemic is a breakthrough for the use of virus genetics. Currently around 10,000,000 sequences are generated and shared on GISAID.

There are three main theories on the emergence of SARS-CoV-2 variants of concern

- The virus variant has evolved gradually in parts of the world where there is less genomic surveillance, widespread circulation and limited travel.
- (Unknown) animal reservoir.
- Persistent infection of SARS-CoV-2 in immunocompromised individuals.

Emergence of VOC/VOI

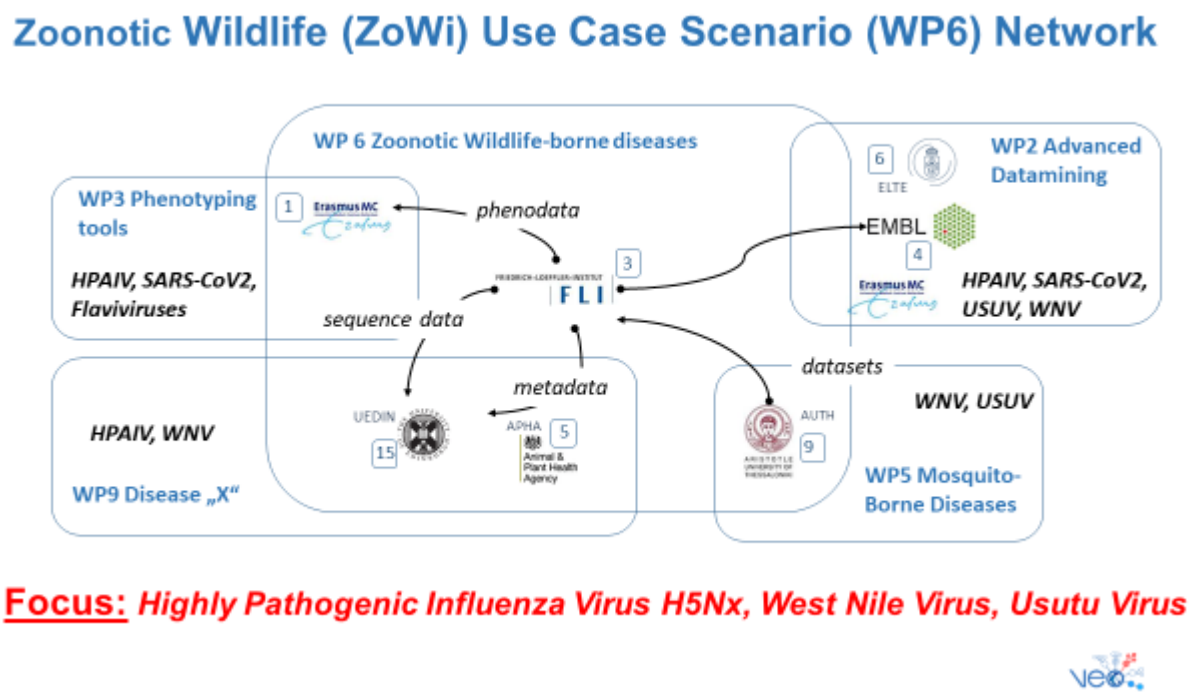




https://www.nejm.org/doi/full/10.1056/NEJMsb2104756 https://www.nature.com/articles/d41586-022-00215-2

VEO Cross WP – Science Update: News, Achievements and Future Plans Part II

Martin Beer, FLI, WP06 Zoonotic Wildlife



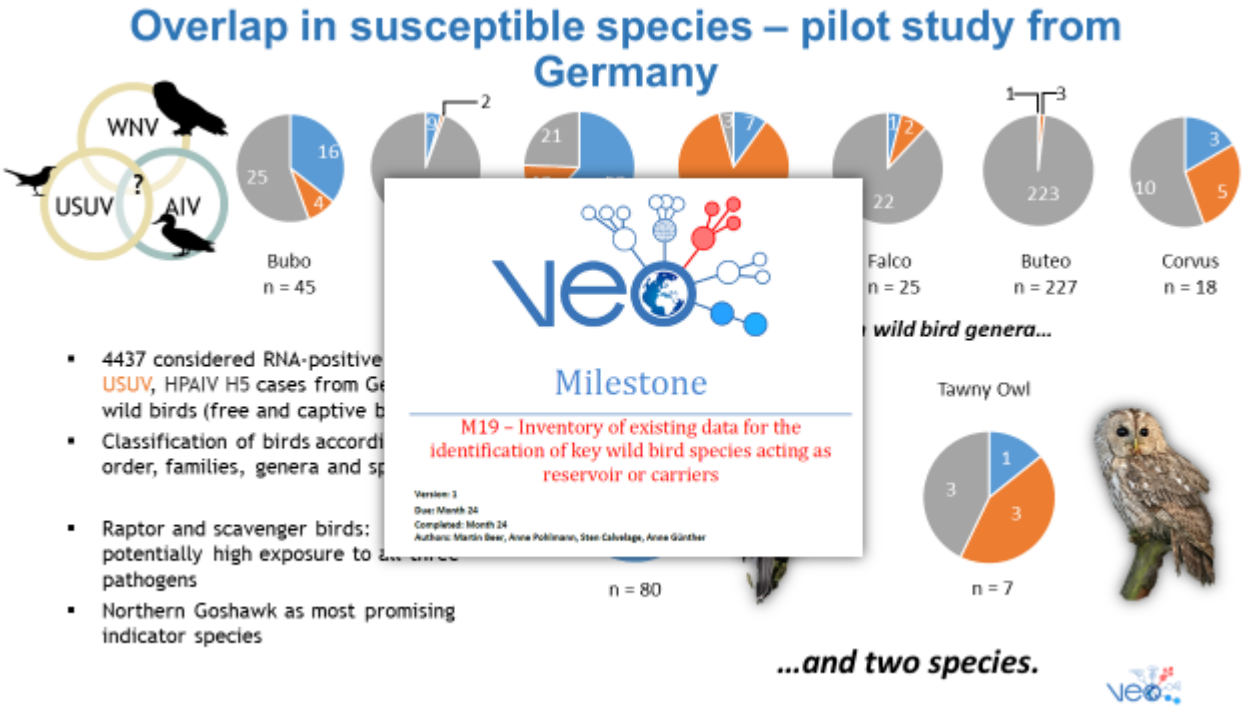
The massive bird die-off in Germany and the Netherlands in 2021, was linked to HPAIV H5, also in other European countries.

The initial results for 2020-2022 H5NX using predictive factors including the following:

- Correlate phylodynamic dispersion with risk factors as gridded datasets (0.5 deg)
- Test virus remain in, and/or disperse towards
- Consider 33 unique risk factors in 10 groups
    - Biodiversity, Bird flyway, Climatic, Elevation, Forest, Land use, Socio-economic, Vegetation and Water
- AIV tended to remain in and to leave from areas with higher poultry and human density
- Other significant predictors:
    - Flyways of anseriformes and passeriformes, Vegetation, cropland use, urban land, broadleaf trees, and wetlands

The first summary on suitable point-of-care high-throughput sequencing methods has been written up in a Deliverable.

One goal was to define overlapping wild bird species for surveillance of WNV and HPAIV H5. A summary report can be found in the milestone 19 report showing raptors as a good choice



Frank Aarestrup, DTU, WP08 Silent Epidemics

In WP08, we aim to understand the full burden of these 'silent epidemics', the circulating strains, hotspots of occurrence or emergence, and associated drivers/risk factors.

We do this mainly using sewage

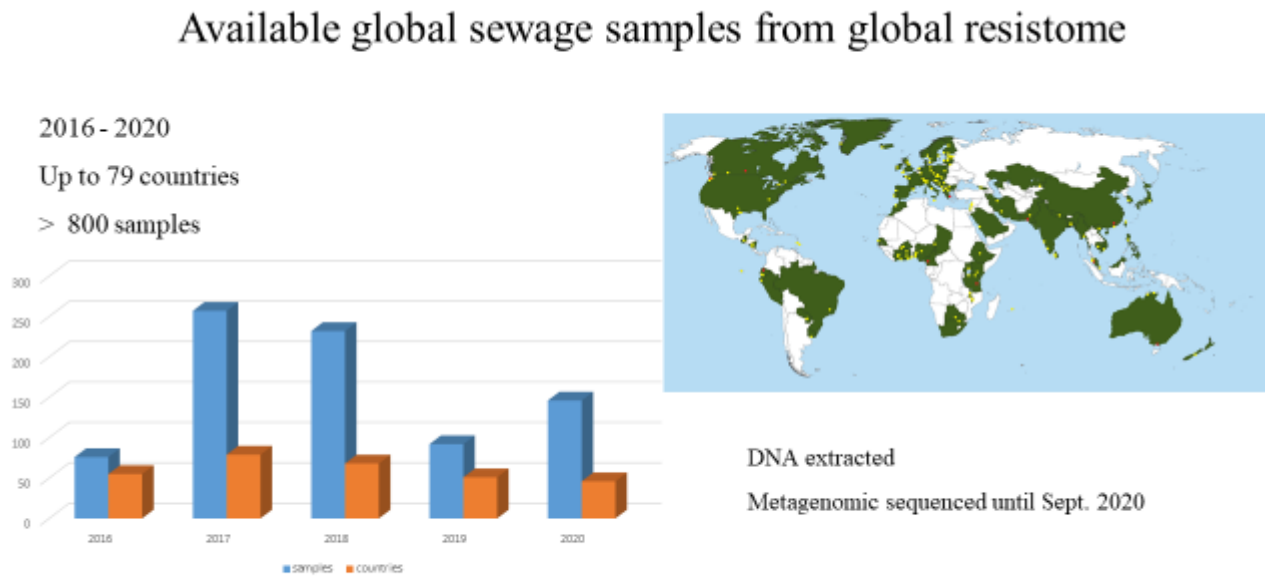
- Can temporal dynamics be predicted by sewage surveillance?
- Can sewage metagenomic surveillance be used to elucidate the burden of silent epidemics?
- Can transmission networks and pathogen phylogeny be reconstructed from metagenomic sewage surveillance?
- Can we predict silent epidemics by epidemiological drivers?

The WP reviewed the rise and fall of SARS-CoV-2 variants in Rotterdam via a comparison of wastewater and clinical surveillance.

We have generated a lot of (sewage) metagenomics data:

- Global point prevalence – sewage and clinical isolates
- Longitudinal in selected cities
- Mapping of all public metagenomes
- Middens in Greenland

How can we further use this data?



Emma Snary, APHA, and Reina Sikkema, EMC, WP09 Disease X

WP09 is aimed at testing and evaluating the VEO system on their applicability in a Disease X use-case scenario. Disease X: a fast spreading EID, with high case fatality by an unknown pathogen.

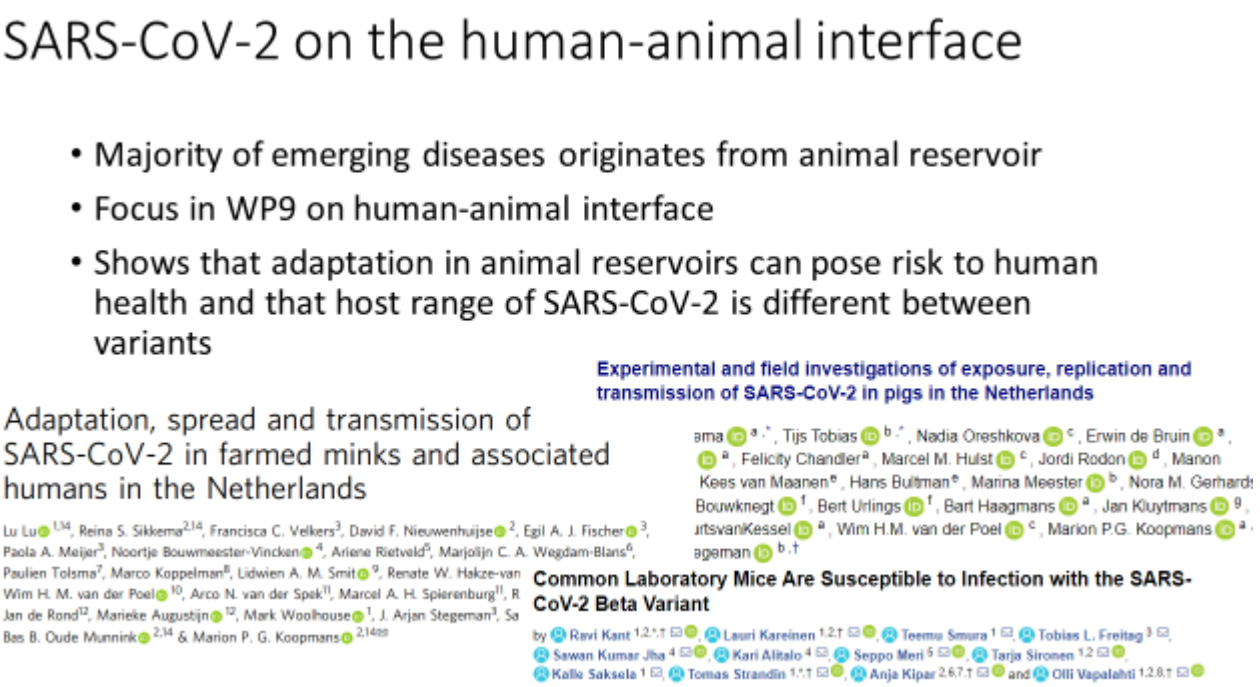
Since VEO started with the COVID-19 pandemic, we have learned some lessons from COVID 19:

- Survey - to identify the development and implementation of new and improved tools/techniques/activities to control and investigate SARS-CoV-2 and lessons learnt in the process
- Assess preparedness for Disease X
- VEO, 13 institutes, 39 "tools"
- Summary published on VEO website

The WP has looked at the human-animal interface of SARS-CoV-2. It also is looking at Disease X scenarios in swine. The WP is reviewing the drivers of disease emergence by assessing key datasets and combining datasets.

In summary:

- Initial stages dominated by COVID
  - VEO contribution to the pandemic was substantial
  - Provided lessons learnt / preparedness for a future Disease X scenario
- WNV – combination of sequencing data, surveillance and driver data
- New focus - Disease X in swine



Workshops: From Data Platform and Data Sets to Addressing Relevant Questions

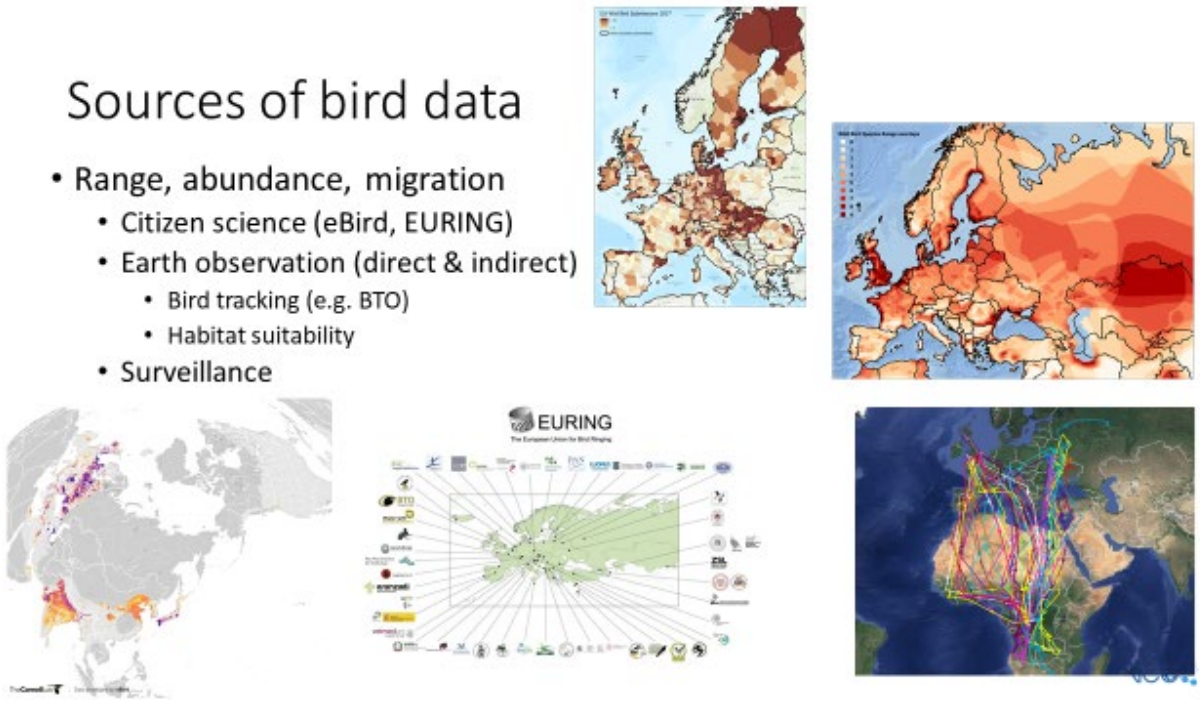
What are key (and interesting) questions to address in the different scenario's based on the data gathered?

- What data do we need / what analysis needs to be done?
- What are the key data challenges?
- Are there data gaps? Could VEO help to address this?
- Can it be used for another emerging disease?
- Who do I collaborate with (multi-disciplinary, across WPs, across institutes)

Emma Snary, APHA, and Martin Beer, FLI, Bird and Mosquito Borne Data

The following research questions were provided from the workshop participants:

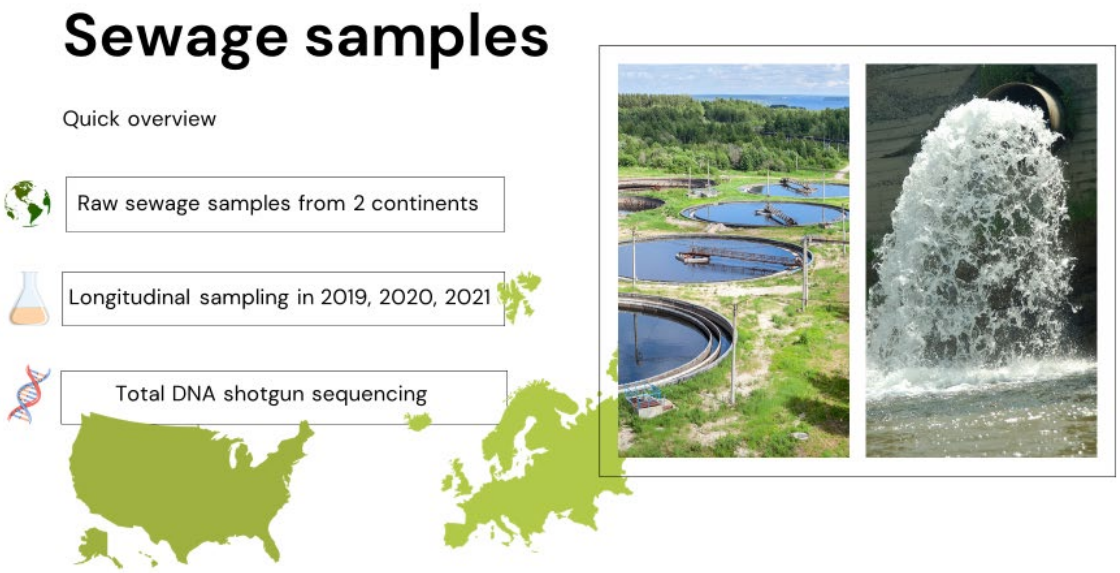
- How do we improve surveillance across the world?
- What would the upcoming risk map look like?
- How does Usuto / WNV over-winter in Europe?
- Development of a spatial-temporal model on WNV / AI / Other pathogens (use of nowcasting)
- How to respond to a new bird / mosquito borne disease?
- What makes a virus stay in a particular region?
- Where is WNV going next?



The data required for each of the research questions was considered. It was widely acknowledged that there is a lot of data already available, including data for birds, mosquitoes, drivers of disease occurrence and also the diseases themselves. These data sets have been collated via active and passive surveillance.

A report from this workshop is available on the VEO Share Site.

Frank Aarestrup, DTU, and Istvan Csabai, ELTE, Sewage Surveillance



- What questions do you think of that may be answered by using the sewage sequencing datasets? List at least 5 questions!
- Which tool or analytical approach would help you answer these questions?
- What data/metadata would be important to combine this data with? Is it freely available? In a usable format?
- Beyond sewage: What other data already on hand in VEO will you use? What other data now outside VEO would you like to see in VEO? Would they benefit from data curation, organization? If yes, specify the need!

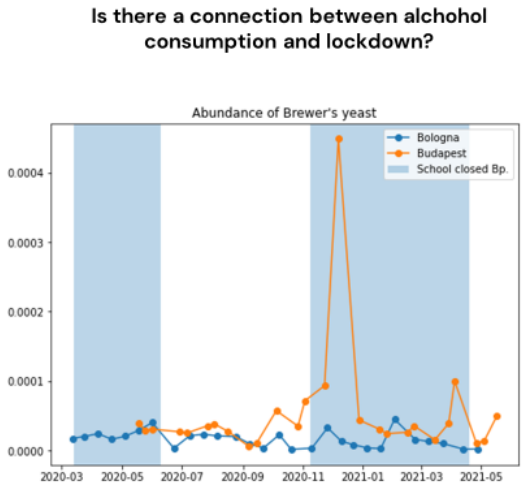
Based on the work of the workshop participants, the following research ideas/questions were formed:

- AMR floating out of a country are contaminating fish/ shellfish and getting back to humans
- Ethical implications – group profiling based on sewage – certain diseases – issues?
- Missing bacterial and viral taxa and eucaryotes – microbial dark matter
- Finding SARS variants in RNA data
- New pipeline for DNA-virus for comparison

- Confirming findings based on demographic data – combining data
- Pre-build rapid pipelines for screening disease X (BY-COVID)
- Relation between sewage and above ground – other data (four cities)
- Diet influence on diversity – detection of changes (beer and alcohol)
- Ease of detecting bacteria AND IMPORTANCE OF GROWTH IN SEWER (Release, in vitro growth, sampling)
- Parasites – who knows something?
- Disinfectants – alcohol resistance
- Co-abundance binning of genomes
- Network to look at gene content within MACs
- Animal DNA (microbiome)
- Where and when for global surveillance to keep standardization
- Citizen sewage

Traces of food/drink in the sewage?

Question: Is abundance of brewer's yeast changing in the sewage over time?
Method: We extracted the normalized fragment counts which were classified as Saccharomyces cerevisie.
First results: Proportion of S. cereviesiae slightly increased in one of the samples in Budapest.
Whats next? We can look into the sewage samples from other cities or check out the changes of other food related microbes.



The two sets of workshop participants came together in plenum to discuss the outcomes from each workshop. The ideas generated from each workshop will be taken forward over the second half of VEO.

This work is supported by European Union's Horizon 2020 research and innovation programme under Grant No. 874735 (VEO).