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

VEO Kick-off Meeting Report

04-05 June 2020

Contents

The VEO Kick-off Meeting was an online meeting held 04-05 June 2020. This online meeting brought together the Principle Investigators, WP Leaders and Co-leaders, participants, as well as stakeholders, EC Officers and other guests, such representatives from other relevant research projects. The first day, 04 June 2020, was dedicated to the VEO COVID-19 Response. Many tasks in VEO relate to data gathering, sharing and analysis surrounding a pandemic such as COVID-19. The second day, 05 June 2020, focused on progress reports from the other WPs in the project.

04 June 2020, The VEO Project and COVID-19

Evelyn Depoortere, EC Scientific Officer

The basis for the call behind VEO was the ZIKA outbreak. Data are/were being collected, but held in silos. There was no sharing of the data to get a broader understanding of the disease and outbreak. What we hope to achieve is to collect and share data. COMPARE had success in this area, and we want VEO to build upon that success.

Now with the COVID-19 outbreak, hopefully RECODID, MOOD and VEO can build upon the infrastructures from PREPARE, COMPARE and other projects.

Marion Koopmans, EMC, VEO – the project

From a One Health perspective, global changes in global trends can act as drivers in emerging infectious diseases (EIDs). With EIDs, usually only a small percentage seek medical care, and thus are sampled.

VEO aims to answer the following key questions:

- How can data science and technology innovations improve infectious disease preparedness and response?
- Can we predict outbreaks? How can we use the novel tools to be better prepared? How can we bring outbreaks under control faster?
- Which are the barriers to the use of big data combined with citizen science? How do we resolve potential barriers?

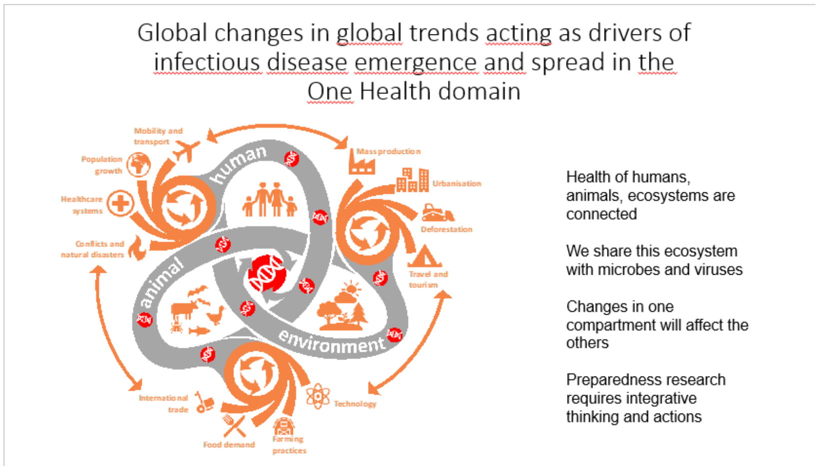
Regarding barriers, we can look to previous projects (COMPARE, PREPARE and Global exchange of viral sequences to underpin response to health threats) as well as current projects (RECODID) to continue to resolve issues. But big data has other challenges as well: platforms, methods, storage issues, carrying rules, etc.

What VEO hopes to achieve, is to take the data types that were brought together in COMPARE (sample, genomic, clinical, laboratory/microbiological and epidemiological data) and put it with other biodata and contextual data: novel lab and field phenotypic data, notifications, climate and environmental, citizen science, geospatial, macro-economic/demographic, and socio-cultural data.

VEO will use five scenarios and the key data drivers to develop the VEO platform.

BUT, at this point, we will incorporate lessons from COVID-19 (modelling abrupt change in contact patterns).

The presentations on 04 June will review COVID-19 as a disease X scenario, explore how VEO resources are being redirected within the existing WPs, what VEO work plans have been accelerated due to the outbreak, and how VEO is supporting the EU COVID data initiative.



Global changes in global trends acting as drivers of infectious disease emergence and spread in the One Health domain

Health of humans, animals, ecosystems are connected

We share this ecosystem with microbes and viruses

Changes in one compartment will affect the others

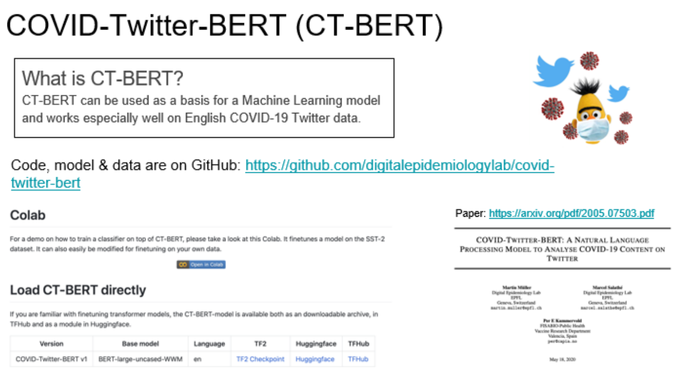
Preparedness research requires integrative thinking and actions

VEO & COVID: Unravelling the infodemic

Marcel Salathé, EPFL, Digital epi and COVID

Marcel Salathé, is leading efforts to use Twitter data as an additional type of data to indicate outbreaks. In addition, Marcel Salathé has been active in the digital contact tracing efforts related to COVID-19 as part of the DP-3T group.

From January 2020, there have been 160 million tweets related to COVID. BERT (Bidirectional Encoder Representations from Transformers) is increasingly the preferred NLP technique for this type of data. COVID-Twitter-BERT (CT-BERT) works with English COVID-19 Twitter data. See article: https://arxiv.org/pdf/2005.07503.pdf . With the data the VEO Consortium members gathered and annotated, Twitter took notice and offered additional data (the full COVID stream) https://developer.twitter.com/en/docs/labs/covid19-stream/overview .



COVID-Twitter-BERT (CT-BERT)

What is CT-BERT?
CT-BERT can be used as a basis for a Machine Learning model and works especially well on English COVID-19 Twitter data.

Code, model & data are on GitHub: https://github.com/digitalepidemiologylab/covid-twitter-bert

Colab

For a demo on how to train a classifier on top of CT-BERT, please take a look at this Colab. It finetunes a model on the SST-2 dataset. It can also easily be modified for finetuning on your own data.

Load CT-BERT directly

If you are familiar with finetuning transformer models, the CT-BERT-model is available both as an downloadable archive, in TFHub and as a module in Huggingface.

Version	Base model	Language	TF2	Huggingface	TFHub
COVID-Twitter-BERT v1	BERT-large-uncased-WWM	en	TF2 Checkpoint	Huggingface	TFHub

Paper: https://arxiv.org/pdf/2005.07503.pdf

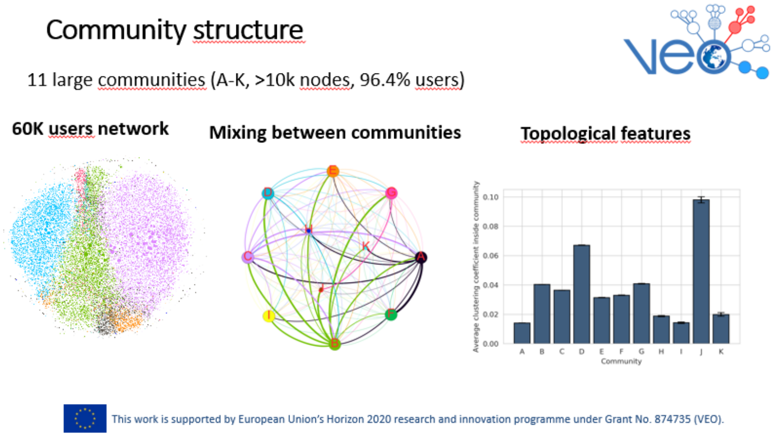
COVID-TWITTER-BERT: A NATURAL LANGUAGE PROCESSING MODEL TO ANALYSE COVID-19 CONTENT ON TWITTER

Daniel Remondini, UNIBO, Network Analysis of social media data

Daniel Remondini, is undertaking a network analysis of social media data.

The aim of the analysis is to characterize Twitter network related to COVID-19 (English). They have the tweets, the users/nodes and weighted links (retweets).

Clear communities emerge, related to countries or to specific topics, with leading influencers. One of the aims is to characterize network evolution over time. How do communities, leading users and trend topics change, in relation to main COVID-19 events worldwide? Which picture is the social network structure providing about the pandemics concern?



Community structure

11 large communities (A-K, >10k nodes, 96.4% users)

60K users network

Mixing between communities

Topological features

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

VEO & COVID: Forecasting, nowcasting and tracking

Luisa Barzon, UNIPD, The COVID epidemic in northern Italy

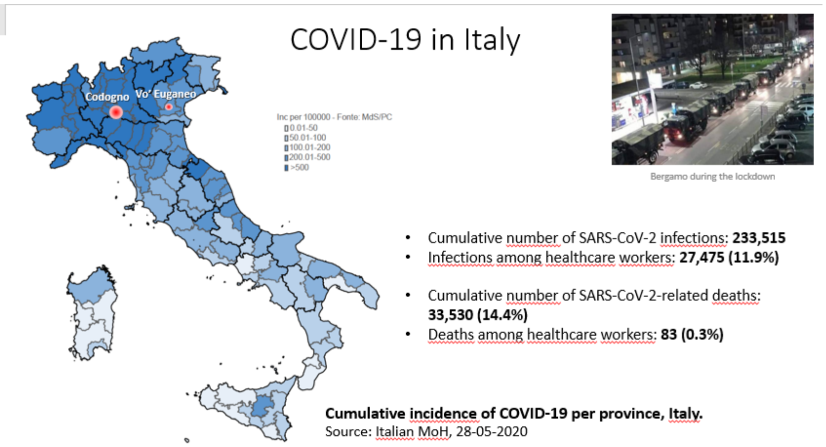
Luisa Barzon is at the center of the European outbreak of COVID-19 in northern Italy.

During the COVID-19 emergency, Dr. Barzon and her team were on the frontlines of the pandemic, and could gather and share firsthand knowledge of the spread, containment and related outcomes from the outbreak.

How did the COVID-19 pandemic develop in northern Italy? What was the epidemiology of COVID-19, what were the public health strategies?

Extensive case and contact investigation in the early epidemic phase provided relevant information on COVID-19 epidemiology and transmission dynamics that will be useful to WP 09, Disease X scenario.

At an initial glance, we can carry forward the following lessons: proactive surveillance and tracing, laboratory preparedness and capacity, and timely and extensive testing as crucial to prevent and control epidemics.



COVID-19 in Italy

Bergamo during the lockdown

- Cumulative number of SARS-CoV-2 infections: **233,515**
- Infections among healthcare workers: **27,475 (11.9%)**

- Cumulative number of SARS-CoV-2-related deaths: **33,530 (14.4%)**
- Deaths among healthcare workers: **83 (0.3%)**

Cumulative incidence of COVID-19 per province, Italy.
Source: Italian MoH, 28-05-2020

Simon Cauchemez, IP, Estimating epidemic size and trajectories: the challenges

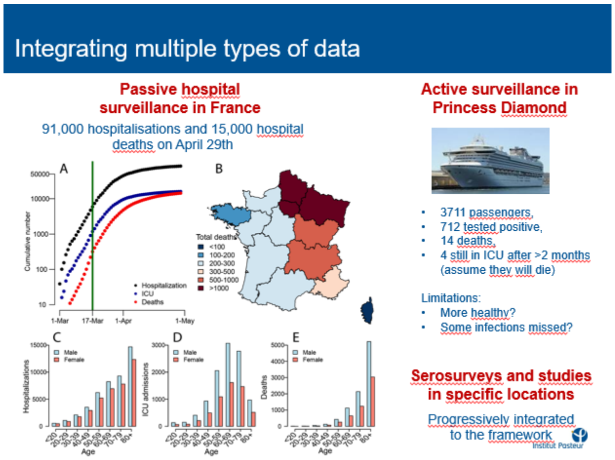
Simon Cauchemez is exploring estimating the burden of COVID-19 in France by integrating multiple data types.

The primary questions are:

- What has been the impact of the French lockdown on COVID-19 transmission?
- What is the proportion of the French population that has been infected by COVID-19 (national and regional)?
- Can we anticipate ICU admissions and bed occupancy to support planning?

Multiple types of data can contribute to answering these questions: passive hospital surveillance in France, active surveillance on the Princess Diamond cruise ship, serosurveys and specific location studies.

We need to explore limits of specific data and data types. The next steps include integrating serological data, what are the changes in transmission after lockdown, can we make short-term predictions based on this.



Integrating multiple types of data

Passive hospital surveillance in France

91,000 hospitalisations and 15,000 hospital deaths on April 29th

Active surveillance in Princess Diamond

- 3711 passengers,
- 712 tested positive,
- 14 deaths,
- 4 still in ICU after >2 months (assume they will die)

Limitations:

- More healthy?
- Some infections missed?

Serosurveys and studies in specific locations

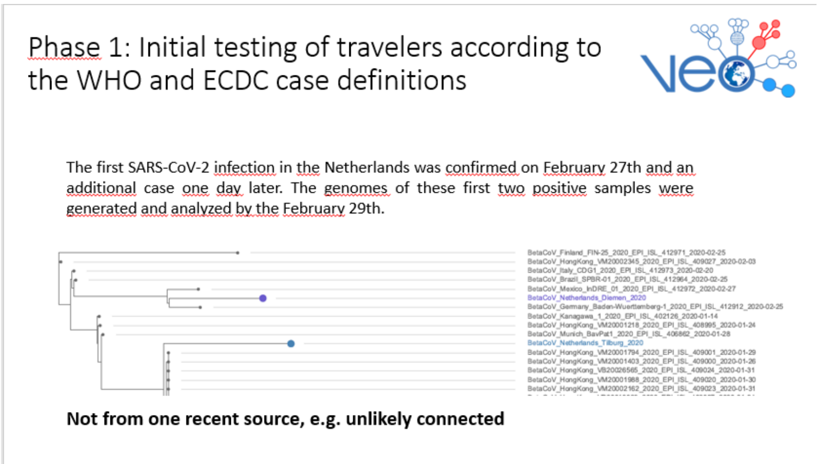
Progressively integrated to the framework

Bas Oude Munnink, EMC, Precision public health, real-time sequencing to track and support public health investigations

Bas Oude Munnink is part of the team at EMC who conducts the rapid SARS-CoV-2 whole genome sequencing for informed public health decision making in the Netherlands (https://doi.org/10.1101/2020.04.21.050633).

In the Netherlands, WGS was used during to facilitate informed public health decision making during three different diagnostic phases: Initial testing of travelers, screening high-risk contacts, and systematic sequencing during exponential growth phase.

There are additional outbreak investigations that are ongoing, for example: mink farms, hospitals, nursing homes, schools and slaughterhouses.



Phase 1: Initial testing of travelers according to the WHO and ECDC case definitions

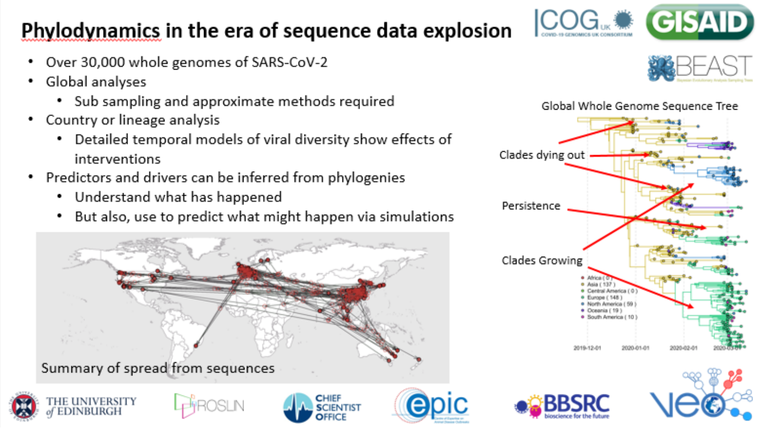
The first SARS-CoV-2 infection in the Netherlands was confirmed on February 27th and an additional case one day later. The genomes of these first two positive samples were generated and analyzed by the February 29th.

Not from one recent source, e.g. unlikely connected

Sam Lycett, UEDIN, Phylodynamics in the era of sequence data explosion

Sequence data collected globally are shared via GISAID. This platform was originally set up for influenza, but now there are over 30,000 whole genomes of SARS-CoV-2. In the UK, COVID-19 Genomics Consortium has sequenced thousands of UK samples, and these data are organized into Global and UK lineages with the Pangolin lineage assigner. The team has developed timeline and trees for early introductions and spread of COVID-19 outside of Asia.

Large data sets of thousands of sequences cannot be processed using BEAST (the phylodynamic inference program of choice), so the data needs to be subsampled for using a suitable stratified method for global analysis. However, for a single country (with few introductions) or per single lineage, it is possible to generate time-resolved phylogenies and estimate viral effective population sizes over time, which can show the bottleneck or suppression effects from interventions. Predictors and drivers for outbreaks can also be inferred from phylogenies.



Phylodynamics in the era of sequence data explosion

- Over 30,000 whole genomes of SARS-CoV-2
- Global analyses
  - Sub sampling and approximate methods required
- Country or lineage analysis
  - Detailed temporal models of viral diversity show effects of interventions
- Predictors and drivers can be inferred from phylogenies
  - Understand what has happened
  - But also, use to predict what might happen via simulations

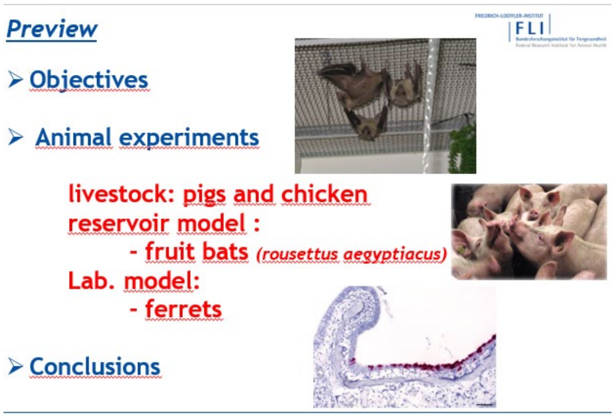
One Health: The human animal interface

Martin Beer, FLI, Host range of SARS-CoV-19

Martin Beer, FLI, is now looking into the host range of COVID-19 via fruit bats, swine, chickens and ferrets and further animal models will be studied.

Work will now include zoonotic SARS-CoV-2 scenarios; including animal reservoirs, zoonotic spillover (e.g., mink-human, cat-human), and fieldable sequencing. The work included the following questions:

- Are important livestock animals (pigs, chicken) susceptible to SARS-CoV-2-infection?
- Infection of fruit bats as model for a SARS-CoV-2 reservoir host
- Laboratory animal model – are ferrets suitable? (infection, replication, transmission)



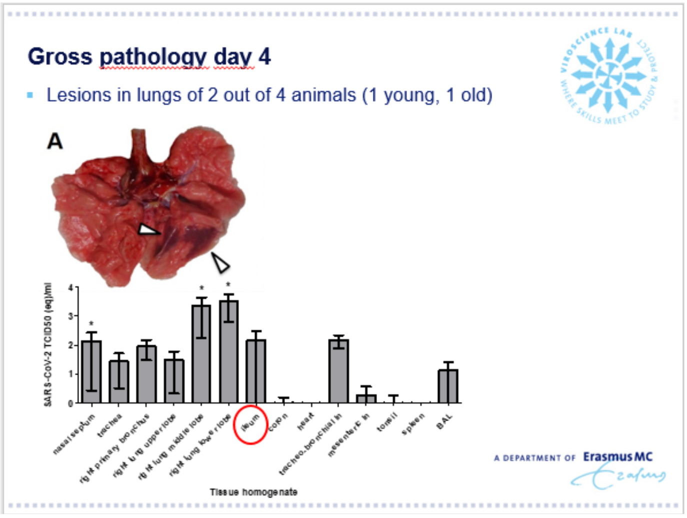
Preview

- **Objectives**
- **Animal experiments**
  - **livestock: pigs and chicken**
  - **reservoir model :**
    - **- fruit bats** ***(rousettus aegyptiacus)***
  - **Lab. model:**
    - **- ferrets**
- **Conclusions**

Bart Haagmans, EMC, Animal models of disease and transmission

Bart Haagmans’ work starts with studies on the COVID-19 interaction with the ACE-2 viral receptor. Given the high homology between human and non-human primate ACE-2, the pathogenesis of SARS-CoV-2 was studied in NHP and compared to results obtained previously with other coronaviruses such as MERS and SARS coronavirus. The team is also using the ferret model to study transmission of coronaviruses in more detail.

The next steps include integrating knowledge on pathogenesis and transmission of coronaviruses in humans, animals, and in vitro systems (e.g., organoids).



Gross pathology day 4

- Lesions in lungs of 2 out of 4 animals (1 young, 1 old)

A DEPARTMENT OF Erasmus MC

The COVID data initiative

Guy Cochrane, EMBL, The COVID-19 data hub

Due to the COVID-19 pandemic, work under WP 01 has been refocused to develop the SARS-CoV-2 Data Hubs; a new, specific WP (WP 15 COVID-19 Response) was initiated.

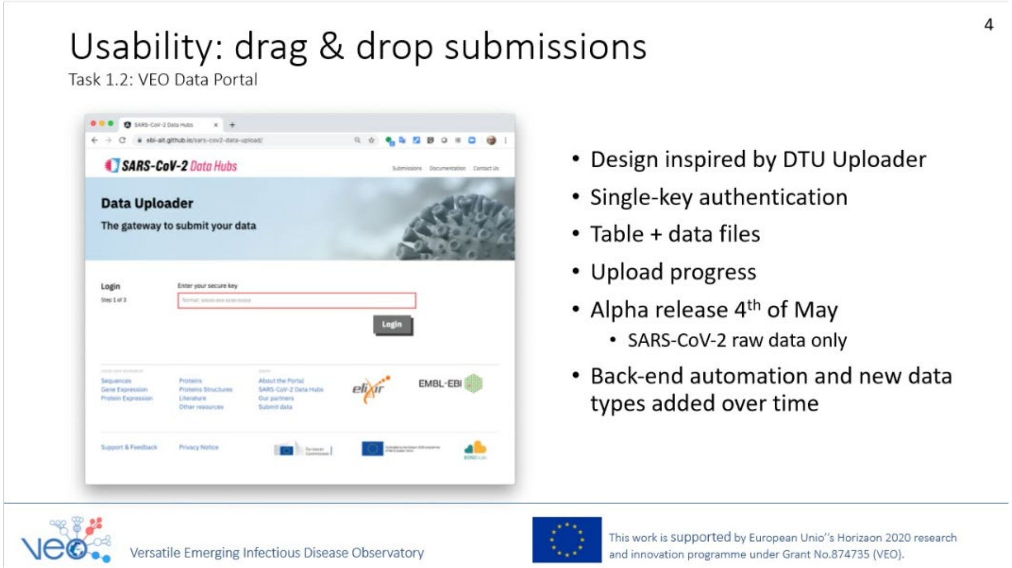
There is a need for a suite of analytical tools, storage, and a data sharing workspace to facilitate the sharing, analysis and reuse of raw and annotated SARS-CoV-2 genomic data. Sequences from around the world were being shared via other platforms, but none included any linked metadata or offered a way of sharing data and common analysis pipelines.

The [European COVID-19 Data Platform]was started in March 2020 and launched 20 April 2020 and includes the SARS-CoV-2 Data Hubs and the COVID-19 Data Portal components. https://www.eosc-life.eu/news/european-commission-launches-data-sharing-platform-for-covid-19/

Within the SARS-CoV-2 Data Hubs, work from WP 01 was accelerated to respond to the need of immediate international use. One of these elements includes adding a simplified data submission tool (drag & drop), installation of newly emerging analytical workflows, data visualization systems and substantial data mobilisation. In addition, national portals will be added to the COVID-19 data hub

(Sweden is available now) to provide access to a number of national-level functions including support and links to national public health and clinical data systems.

Further improvements and additional tools will be added to the SARS-CoV-2 Data Hubs.



4

Usability: drag & drop submissions

Task 1.2: VEO Data Portal

- Design inspired by DTU Uploader
- Single-key authentication
- Table + data files
- Upload progress
- Alpha release 4th of May
  - SARS-CoV-2 raw data only
- Back-end automation and new data types added over time

Versatile Emerging Infectious Disease Observatory

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

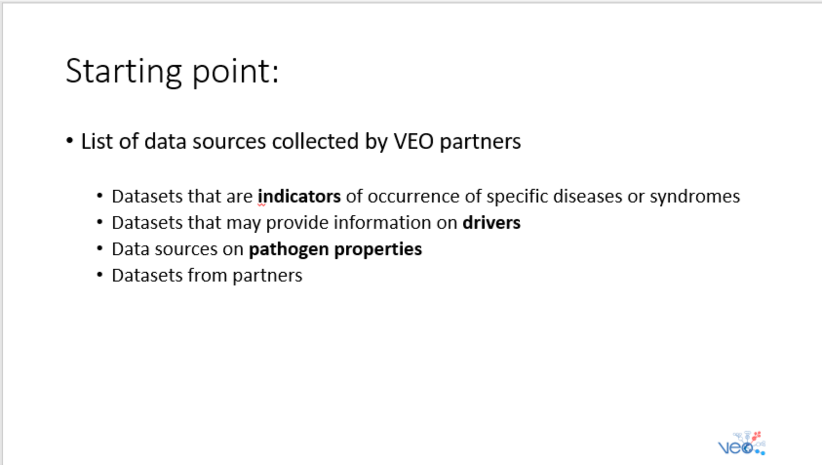
05 June 2020, VEO In-depth

Marion Koopmans, EMC, Welcome back and Scoping Workshops (WP 11)

We will hold VEO scoping workshops (WP 11) for a structured discussion of the use case scenarios. This will bring the biology and data people together, develop a common understanding.

Five key questions:

- What are the key drivers for disease emergence/spread for this scenario that you would like to have information on?
- What data types do you think would be interesting to look into/use, if you ignore possible hurdles (data quality, access, formats)?
- Additional data types to add to the inventory?
- What criteria are needed to assess suitability of data types from the attached table?
- What are critical gaps and (how) can and should we fill them?



Starting point:

- List of data sources collected by VEO partners
  - Datasets that are **indicators** of occurrence of specific diseases or syndromes
  - Datasets that may provide information on **drivers**
  - Data sources on **pathogen properties**
  - Datasets from partners

Frederic Bartumeus, CSIC, Mosquito-borne scenario (WP 05)

Bartumeus started by describing a mobility and social distancing research app called Space Mapper. Space Mapper could be related to the work in WP0 09, Disease X.

General information about WP 05 was presented.

Using citizen science for tracking disease-carrying mosquitoes, data are being modeled but looking for further modelling through WP 02. This summer, a few ‘smart traps’ (automatically counting and differentiating mosquito species) will be used to calibrate citizen data quality and abundance estimates within the city of Barcelona..

We will launch a new multilingual app for citizen science to identify mosquitoes and gather information related to breeding sites and bites in July 2020. The up-scale to Europe will leverage work from the AIM-COST network project.

There is ongoing work with WP 02 regarding identification of mosquito species based on machine learning algorithms. A tutorial and webinar sessions have been launched for VEO WP 05 partners learn how to use the new EU digital EntoLab.

New app for citizen science surveillance and monitoring of mosquitoes

Multi-lingual App 13 EU languages

Notification system (direct contact with user)

Gamification improved

New biting report new

New species added new

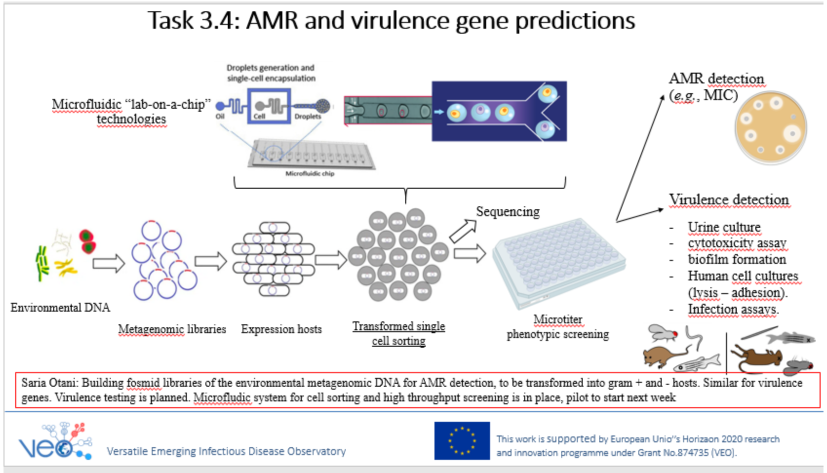
Breeding sites report

Reports of the user improved

Citizen validation new flow

Ron Fouchier, EMC, From genotyping to phenotyping (WP 03)

Presentation of the WP and Tasks. Notes that Tasks 3.2 and 3.4 have started with EMC and DTU lab work. As there is confusion on the scope of Task 3.3, this was narrowed down to host-range profiling using reverse genetics, and to include COVID-19 work. Ron referred to the COVID-19 in animals presentation from the previous day, in relation to Task 3.1, providing real-life opportunities for research on host-switching. Interest of partners in "interactomics" was noted when appropriate datasets become available.



Task 3.4: AMR and virulence gene predictions

Saria Otani: Building fosmid libraries of the environmental metagenomic DNA for AMR detection, to be transformed into gram + and - hosts. Similar for virulence genes. Virulence testing is planned. Microfludic system for cell sorting and high throughput screening is in place, pilot to start next week

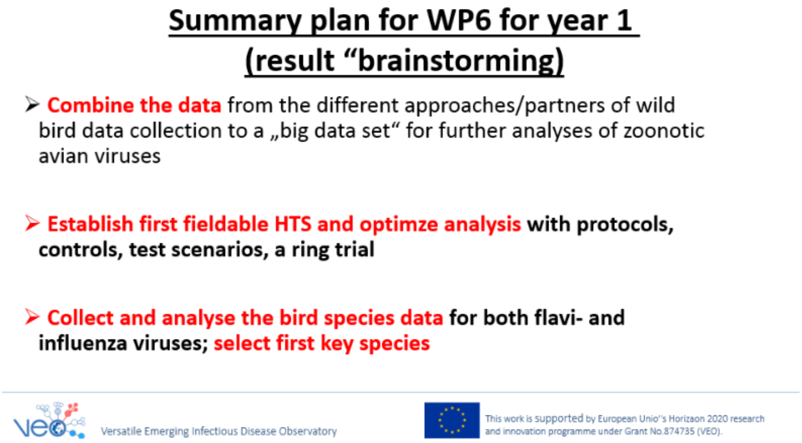
Versatile Emerging Infectious Disease Observatory

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

Martin Beer, FLI, Zoonotic (wildlife) use case scenario (WP 06)

A summary of the first-year plan was presented, however, COVID-19 zoonotic scenarios need to be added in; animal reservoirs, zoonotic spillover (e.g., mink-human, cat-human), and fieldable sequencing.

Novel HPAIV H5N8 reassortments are detected throughout Europe. The various inputs to this will be studied in WP 06; big data analyses, fieldable high-throughput sequencing, and reservoir surveillance.



Summary plan for WP6 for year 1 (result “brainstorming)

- **Combine the data** from the different approaches/partners of wild bird data collection to a „big data set“ for further analyses of zoonotic avian viruses
- **Establish first fieldable HTS and optimze analysis with protocols, controls, test scenarios, a ring trial**
- **Collect and analyse the bird species data for both flavi- and influenza viruses; select first key species**

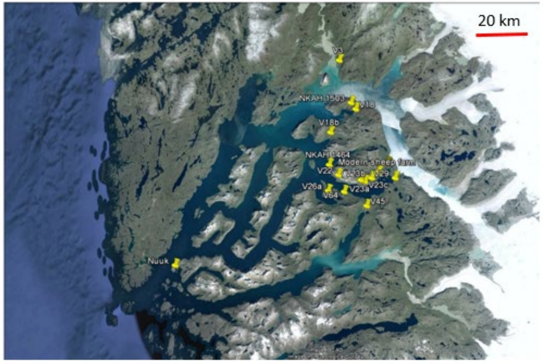
Versatile Emerging Infectious Disease Observatory

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

Anders Prieme, UC, The changing cryosphere scenario (WP 07)

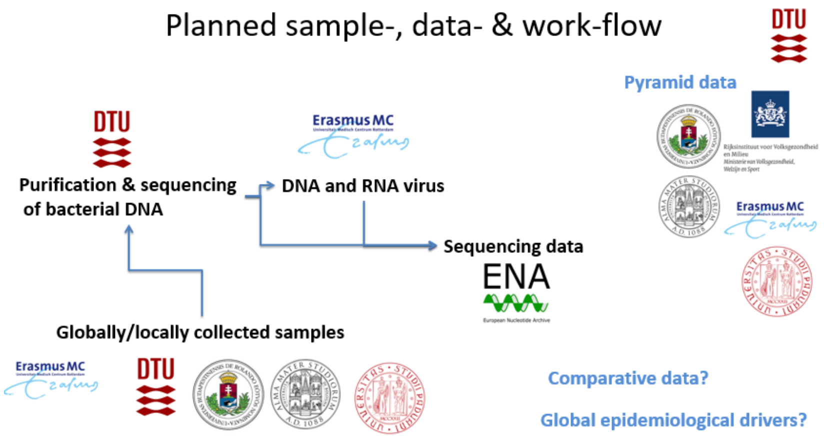
The arctic cryosphere is changing fast. We are looking at the Artic as a possible threat to humans. In 2020, we will sample Eskimo and Norse (Viking descendants) middens and farms in Greenland, as well as meltwater from Greenlandic ice sheets to look for ancient pathogenic agents on humans or livestock. In addition, we will initiate experiments to unravel the transport of bacteria and viruses in thawed arctic soils. In 2021, we will sample two species of geese with overlapping summer grounds in Greenland, but contrasting migratory routes (wintering in NE USA or Ireland/UK) to investigate if changing routes of migratory birds may influence the spread of infectious diseases.

Sites selected for sampling Norse middens and farms in the Nuuk fjord system, 2020



Frank Aarestrup, DTU, Urban epidemic scenario (WP 08)

Presenting the surveillance pyramid, very limited reporting, interested in getting the complete pyramid of data. Building on data and collections from COMPARE. The original plan for global and local studies has been adjusted due to COVID-19. Global studies will look at the major intervention of COVID-19. Local studies (4 cities), collecting data, including clinical, test positive and negative, social media, sewage and phylogenies to compare all the different possible ways of describing and modelling the outbreak..



Planned sample-, data- & work-flow

Pyramid data

Purification & sequencing of bacterial DNA

DNA and RNA virus

Sequencing data

ENA

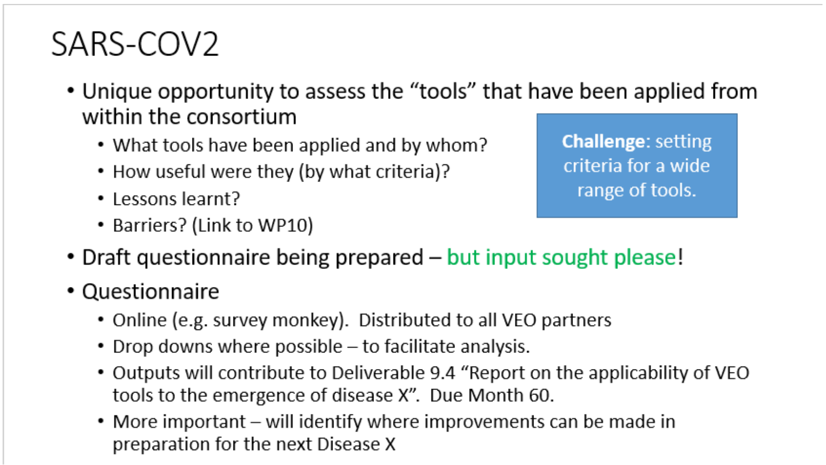
Globally/locally collected samples

Comparative data?

Global epidemiological drivers?

Emma Snary, DEFRA, Disease X (WP 09)

Disease X is a fast spreading emerging infectious disease, with a high case fatality by an unknown pathogen. Reviewing the tasks within WP 09 we can summarize them as the following: can we predict an outbreak; can we respond to an outbreak; can we do any better? COVID-19 has given us a unique opportunity, and therefore, Task 9.4, 'Pandemic response-assessing the applicability of all VEO tools', has moved forward sooner than was originally intended. This task focuses on the latter theme of 'can we do any better' and in collaboration with WP 10, a draft questionnaire has been developed to identify lessons learnt and barriers experienced when applying 'tools' to COVID-19. The draft questionnaire was presented, which included a case study on sequence analysis and sharing.



SARS-COV2

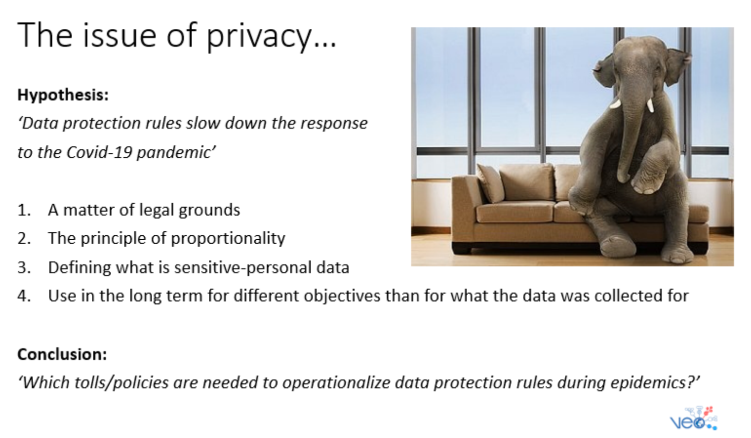
- Unique opportunity to assess the "tools" that have been applied from within the consortium
  - What tools have been applied and by whom?
  - How useful were they (by what criteria)?
  - Lessons learnt?
  - Barriers? (Link to WP10)

Challenge: setting criteria for a wide range of tools.

- Draft questionnaire being prepared – but input sought please!
- Questionnaire
  - Online (e.g. survey monkey). Distributed to all VEO partners
  - Drop downs where possible – to facilitate analysis.
  - Outputs will contribute to Deliverable 9.4 "Report on the applicability of VEO tools to the emergence of disease X". Due Month 60.
  - More important – will identify where improvements can be made in preparation for the next Disease X

Carolina Ribeiro and George Haringhuizen, RIVM, Barriers inventory – ethical issues in a pandemic (WP 10)

While past epidemics have provided experience to deal with the current pandemic, some barriers for collaboration data sharing persist. What lessons have we learned, what is different now? Technological and infrastructural developments have helped to decrease timelines in data sharing and use, but the politics and conflicts of interest can stand on the way. The ethics of sharing also move into scientific ownership and business. During a pandemic, sharing genetic resources takes on an urgency, but it exposes flaws in international coordination, rules and regulations. There are opportunities and limitations of a Consortium like VEO.



The issue of privacy…

Hypothesis:
'Data protection rules slow down the response to the Covid-19 pandemic'

1. A matter of legal grounds
2. The principle of proportionality
3. Defining what is sensitive-personal data
4. Use in the long term for different objectives than for what the data was collected for

Conclusion:
'Which tolls/policies are needed to operationalize data protection rules during epidemics?'

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