UPDATE, Fall 2016

COMPARE Quick Facts
Coordinators: Prof. Frank Aarestrup, DTU and Prof. Marion Koopmans, EMC
Project period: 01 December 2014 – 30 November 2019
Funding: €20 million (approximately)
compare@food.dtu.dk
www.compare-europe.eu
@CompareEurope on Twitter

Note from coodinators

Frank Aarestrup

Coordinator of COMPARE
Technical University of Denmark

Marion Koopmans

Co-coordinator of COMPARE
Erasmus Universitair Medisch Centrum Rotterdam

Within the first two years of the project, several infectious disease emergence and transmission events have taken place with potential for regional or global spread (Zika, Ebola, avian influenza, as well as mcr-1 colistin resistance). In all cases, COMPARE partners were involved in analysis of sequenced-based information in combination with relevant epidemiological information, providing insights in sources of infection and modes of spread. However, to a large extend, these analyses became available rather late in the course of the outbreaks, illustrating the need for the platform that COMPARE seeks to deliver. Nevertheless, these outbreaks helped to further define the critical components of the architecture of the COMPARE platform, and all were used as case studies for development of the COMPARE infrastructure.

The first two years of the project have focused on getting started in each of the individual workpackages. In order to avoid duplication of efforts, substantial work has been done to map what is already available in terms of protocols for sampling (WP 1), storage and sequencing (WP 2), bioinformatics workflows, reference sample sets, reference databases for use in clinical, public health and research questions (WPs 3-5). These inventories all were made in the form of databases that can be accessed and searched by the COMPARE partners, and have been shared through the COMPARE Share Site. In the coming period, work will be done to make these resources accessible and searchable for outside users, as resources to potential future users of COMPARE.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

After these foundational studies, collaborative studies have delivered the first versions of protocols for generic risk assessment, sampling, sequencing and data analysis, with potential applicability across pathogens and domains. The basic architecture of data-sharing hubs was designed and piloted, and the first attempts to compare analytic pipelines are on their way. The recent global emergencies have been a good basis for studying communication strategies and barriers for sharing, and plans for economic evaluations will follow.

During the next phase, COMPARE will focus on starting to link the building blocks that are emerging from the individual workpackages into the developing ICT and analysis framework developed in WP9. To stimulate this, cross- workpackage activities have been identified at the annual meeting and by review of the individual workpackage reports. The management team has discussed potential synergies and overlaps on a regular basis with the WP leaders and individual partners.

In addition, we will prepare for an initial pilot study to increase the interaction with the global research and public community. Important steps here will be to ensure the parallel development of collaborative research studies where data are shared in closed private environments and "open science" where raw data and analysis generated by COMPARE partners are immediately made available for both researchers globally and the general public. This latter aspect is expected to provide valuable scientific input from the global research community but also important lessons for our attempts to build easy data- sharing infrastructures as well as learn about the barriers to sharing.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

WP Updates

WP1 – Risk assessment and risk-based strategies for sample and data collection

In WP 1, a generic framework for the risk assessment is being developed. The core question addressed is 'What are the most probable routes of spread for pathogens introduced into Europe, and what are the populations and regions at greatest risk for spread?'

Task 1 work has moved from the developed conceptual frameworks to the selection of case studies to populate the generic spatial risk assessment and microbial risk assessment. For the spatial assessment, this is being completed with a view to modelling three key modes of exposure across different susceptible hosts: animal to animal, vector borne, and human transmission. The proposed pathogens, Aujeszky's disease, Lumpy skin disease, Zika virus and avian influenza virus cover two livestock diseases; one livestock disease with zoonotic potential, and an emerging vector-borne human pathogen. In addition, the feasibility for a generic risk spatial risk assessment based on food trade data is being studied. For the food chain risk assessment, a conceptual risk assessment framework has been established and case studies have been initiated for which sufficient NGS datasets exist across the clinical, food, and veterinary environment for meaningful piloting. Therefore, *Listeria monocytogenes* and *Escherichia coli* were chosen as topics, with a possible future pilot on norovirus – in line with the choice of norovirus as cross-WP target pathogen.

Within Task 2, protocols for risk-based sampling for the early detection of infectious diseases are being developed following an in depth inventory of the (great) diversity of pathogen specific protocols and guidelines that is already available. Generic risk-based sampling protocols are being co-designed across the domains (human, domestic and wildlife animals), with food sampling under discussion. Although a unifying framework is envisaged, each task will tailor the protocols using theme-specific case studies. Close collaboration has been sought with the emerging disease detection capacity building program of FAO, which aims to also develop generic protocols for risk-based sampling.

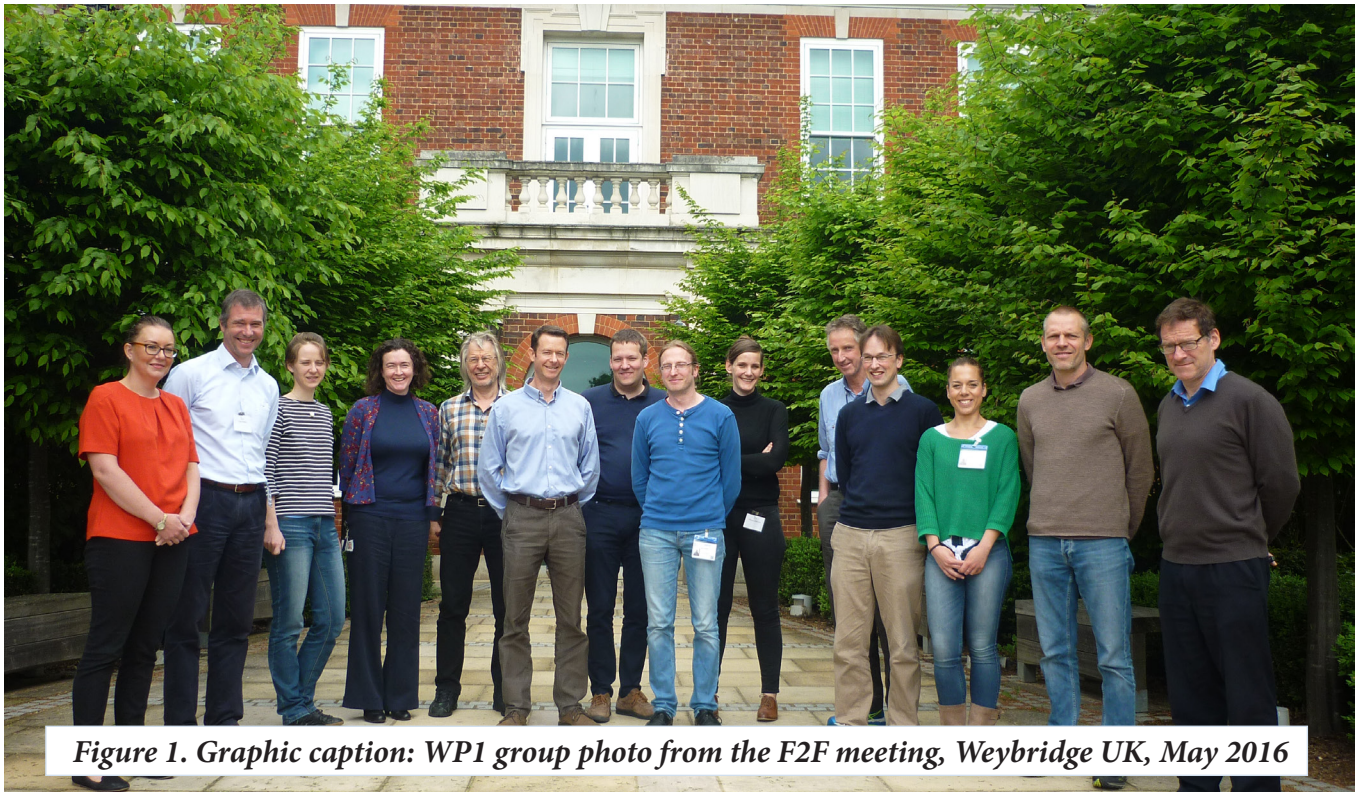


Figure 1. Graphic caption: WP1 group photo from the F2F meeting, Weybridge UK, May 2016

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

WP2 – Harmonized standards for sample processing and sequencing

Progress has been made to unravel the factors that determine the quality of sequence data produced from the broad range of sample types for the different organisms (bacteria, viruses and parasites) and study questions (medical, veterinary, food microbiology, clinical, public health, zoonosis and basic research). In accordance with the results of the survey, sample handling experiments have now been planned and have been initiated.

In parallel, sample-processing pipelines including pathogen inactivation, nucleic acid extraction, and subsequent processing until sequencing were developed and are being intensively tested for different matrices (e.g., tissues, bacterial suspensions, vectors, food samples, see Figure X). Protocols providing best results regarding quality and quantity of extracted nucleic acids as well as sequence reads were already disseminated in form of so-called Laboratory Operating Procedures (LOPs) for their further review and application in laboratories of COMPARE members. These sample handling experiments and sample-processing pipelines will be tested and further refined. The enthusiastic aim of these optimization steps is to develop and provide one protocol for all samples for metagenomics.

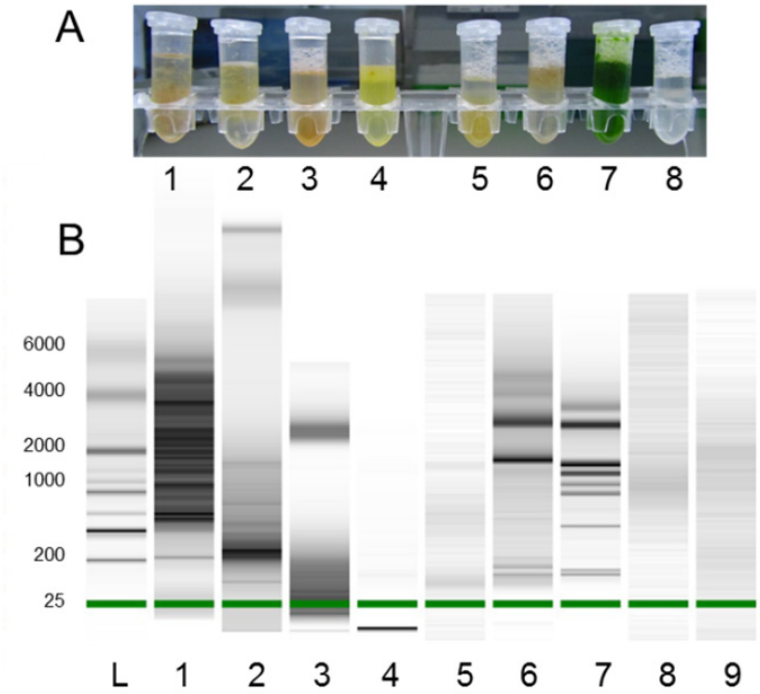


Figure 2. Different food samples (A) after homogenization using the CryoPrep technique and (B) quality control using RNA 6000 Pico Chip (BioAnalyzer, Agilent). Abbreviations: L, Ladder; 1, crude ham; 2, poultry cold cuts; 3, smoked salmon; 4, cheese; 5, oat flakes; 6, mushroom; 7, rocket; 8, coconut milk drink; 9, no sample. The labeling for the ladder is given on the left side.

With respect to the WP2 tasks, task 1 and 2 (harmonization and standardization) have been formally completed, but work in this area will continue in order to collect more protocols and to improve established workflows. This will be realized in close collaboration with the ongoing task 7 (ring trials and quality assurance systems). Additional matrices as well as environmental samples have to be tested. Task 3 (sequencing protocols) is ongoing until year 4 and within this task, newly developed technologies will be assessed. Regarding task 4 (sequence analysis), additional applications will be continuously implemented in the bioinformatics toolbox until the end of the project. Task 5 and task 6 (curation and storage protocols, and biobanks) have been completed during this first WP2 phase. The outcome of the biobank-survey (task 6) is available on the COMPARE Share Site (WP2 /task 6).

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

Task 7 is ongoing until the end of the project. Virtual, technical ring trials are planned as well as protocol comparisons. The Virus Proficiency Test (by RKI) has started in autumn 2016. Further pilot studies are in the planning stage (e.g. to investigate non-suppurative encephalitis cases which might be caused by yet unidentified pathogens).

WP3/6 – Frontline diagnostics

A general framework for applications of NGS to routine clinical microbiology, virology and diagnostics has been developed. For microbiology, the pilot organisms/problems chosen for the initial phase of COMPARE were *E. coli* and urinary tract infections, and antimicrobial resistant bacteria relevant for patient care and hospital epidemiology. Here, there is a need for simplifying the technical algorithms utilized for typing and for antibiotic resistance detection to allow people with no or little knowledge about computational analyses to routinely perform the analysis.

Two workflows, one to utilize NGS in clinical settings to study the epidemiology and population biology of bacterial clones and one for antibiotic resistance gene identification have been developed. These workflows will now be built into the COMPARE platform and evaluated further in WP3/6.

On the virus side, work was started to explore clinical application of NGS metagenomics for evaluation of unusual disease syndromes, with emphasis on severe unexplained illness and/or high -risk patients. Protocols and analytical workflows are operational in several COMPARE partner laboratories, and comparative studies are planned and ongoing to understand the strengths and weaknesses of the different approaches. In addition, this expertise was also used to support recent outbreak investigations (Ebola, Zika, emergence of colistin-resistant *E. coli*).

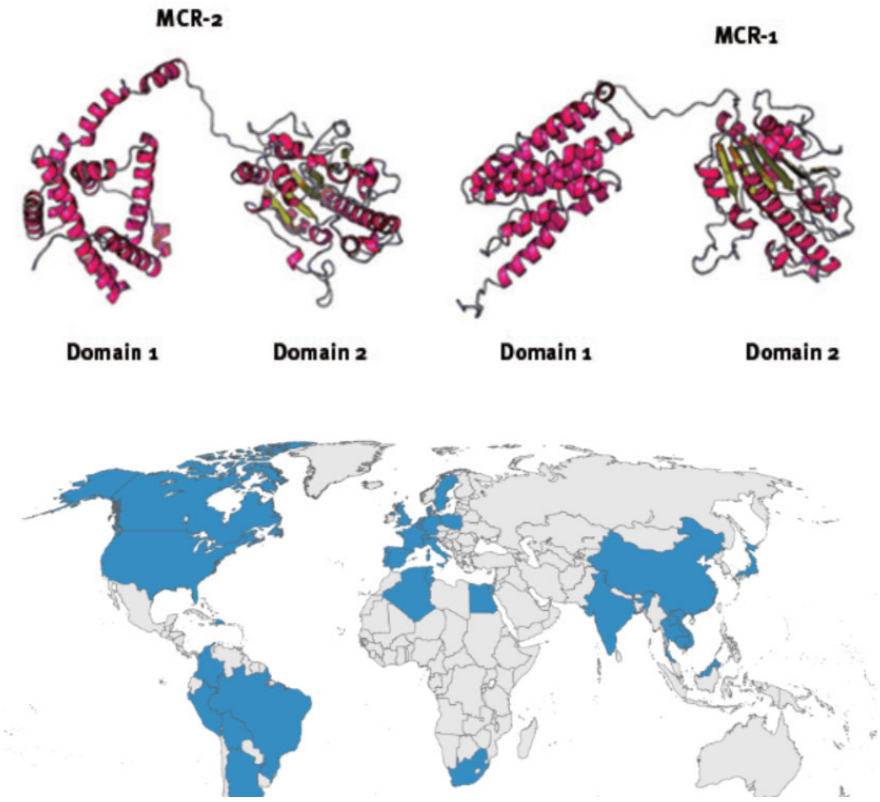


Figure 3. mcr-1 presence map and structural comparison of MCR protein as described in our recent publication Xavier et al 2016. (http://www.eurosurveillance.org/images/dynamic/EE/V21N27/art22525.pdf)

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

WP4/7 – Foodborne pathogen surveillance, outbreak detection and epidemiological analysis

Essential preparatory steps were made to assess the potential added value and challenges of the use of NGS-based techniques in the context of ongoing public health surveillance. For this, a reference genome database was constructed consisting of publically available sequences to cover each of the six pilot foodborne pathogens (*E. coli*, *Salmonella*, *Listeria*, Norovirus, Hepatitis A, *Cryptosporidium*).

These reference genomes were carefully selected to represent the most relevant types to be useful in future sequence-based analysis for outbreak detection and source attribution modelling. This reference genome database is publicly available at http://www.compare-europe.eu/Library/Reference-Genomes together with short descriptions of the disease-causing potential and epidemiology of each organism. In addition, detailed plans have been developed to address the crucial questions of performance of NGS in comparison with current reference methods for pathogen detection and typing, and for cluster analysis and source attribution. The following is a list of the pilot studies being undertaken during the current and next phase.

1) Source attribution modelling of *Salmonella* Typhimurium and the monophasic variant.

2) Long-term evolution study of monophasic *Salmonella* Typhimurium

3) Historical international outbreaks

4) Historical outbreaks

5) Testing and comparing different cluster detection methods

6) Understanding the epidemiology of travel-related *Salmonella* Enteritidis infections

7) Persistent clones of *Listeria monocytogenes*

8) Prospective outbreak investigations

9) Compare protocols for identification and (full) genome characterization of Norovirus and Hepatitis A from a range of sample types (bait-based/targeted and metagenomic)

10) Develop analytical pipelines for recovery of Norovirus and Hepatitis A sequences from metagenomes from different sample types

11) Assess the presence and diversity of Noroviruses and Hepatitis A in the virome content of sewage samples from a global collection sampled and tested in WP8.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

WP5/8 – Detection and response to (re-) emerging diseases

The focus is on harnessing the potential added value of NGS for emerging disease detection and research.

Pathogen discovery pipelines and metagenomics pipelines continue to be developed and optimized in COMPARE by several partners. New analytical methods such as structure-based sequencing matching are under development. A previous version of one of the virus discovery pipelines, which has been installed on the COMPARE Virtual Machine (VM) at EBI, has been rewritten to enable more flexible and modular development in a pipeline development framework. This should make it easier to apply the enhancements needed to address the current issues of the analysis pipeline by implementing novel algorithms developed in the field of bioinformatics.

An analysis test-run has been done on a large set of sequenced diarrheal samples as a baseline assessment of current data throughput and analysis quality of the pipeline. With these preliminary analyses we have started investigating the pitfalls currently present in the use of high-throughput-sequencing data for metagenomic viral surveillance, clinical viral diagnostics and novel virus discovery. Based on these experiences improvements will be made to the analysis pipeline to prepare for the coming prospective high-throughput-sequencing datasets (e.g. sewage water samples).

For the global sewage projects, sewage water samples were collected from 70 major cities around the world. The diversity of pathogens identified in these samples will be indicative for the diversity of pathogens present in the human population in the sampled areas. DNA sequencing for bacterial metagenomics has started on these sample collections and analyses are in progress.

A paper by the Global Consortium for H5N8 and Related Viruses (coordinated by COMPARE partners) has been published in *Science*.

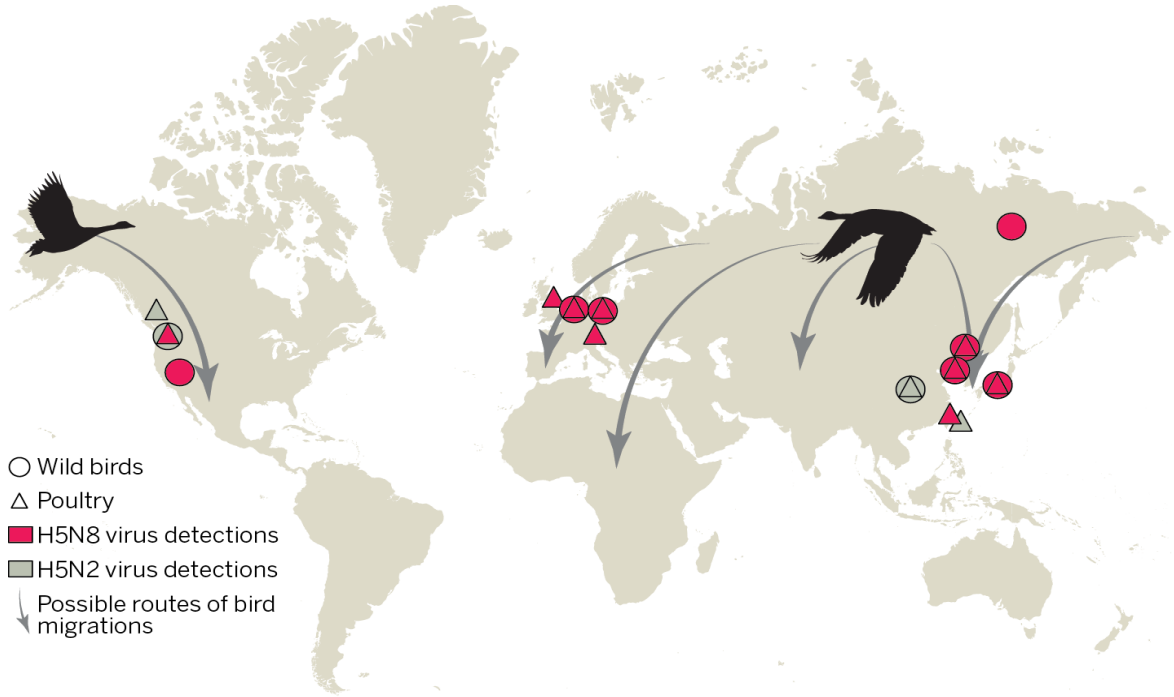


Figure 4. Global map of influenza H5N2 and H5N8 virus detections in poultry and wild birds in 2014, under investigation in COMPARE WP5/8. Circles indicate wild birds, triangles indicate poultry. H5N8 virus detections are in red, H5N2 virus detections in grey. Dark grey arrows indicate possible directions of wild bird migration.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

WP9 – COMPARE platform

Key components of the future COMPARE data infrastructure have been delivered as early usable tools in pathogen data sharing.

Our current frontline service, the COMPARE data hubs, has seen continued uptake and we support a growing number of users around data-sharing initiatives (see table). The data hubs allow pre-publication sharing of structured sequence data and metadata amongst defined groups of collaborators with, at the appropriate time for the data providers, automated full data release. Having enhanced the metadata organisation and structure, with daily metadata reports now presented in the data Hubs in tabular and database-ready formats, these data hubs have grown in utility across bacterial, viral, parasite and metagenomics data sharing initiatives.

Table 1. Currrent Compare Data Hubs

Data Hub	Content
dcc_sibelius	Influenza H5N8 pilot
dcc_berlioz	Ebola pilot
dcc_compare	General COMPARE usage; all bacteria, viruses and some eukaryotic parasites
dcc_liszt	Global Sewage Project
dcc_strauss	Kibera Sewage Project
dcc_handel	Virus Metagenome project (EMC, RIVM, IFREMER, AUTH, FLI)
dcc_puccini	Eukaryotic parasite data
dcc_vivaldi	Salmonella (WP4+7)

Work has continued on the development of the compute environment hosting analytical workflows. This system has been available for some time, but is subject to significant ongoing enhancement and extension. A 'selector' component has been developed that allows rule-based configuration of selection and dispatch of newly reported data into appropriate analysis workflows. The eHive workflow management framework has been installed in the system to support efficient and manageable computation at scale. WP9 participants have continued to port tools, utilities and workflows into the 'Docker' environment such that they become available for installation in the COMPARE compute environment.

We have begun the process of developing the COMPARE data portal. The data portal is expected to provide authenticated user access to authorised pre-publication and all public data for search, browse, geographical visualisation, and navigation of results from analysis workflows in an intuitive web environment and will also include a 'Notebook' scripting environment for data exploration by those with some bioinformatics skills.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

We plan to put together two WP9 'Demonstrator's; one focusing on bacterial genomics, and one focusing on viral genomics. While a number of services are available from WP9 and in use by the COMPARE community, we lack a single example of use that cuts across all of the services. The Demonstrators will comprise specific data sets with appropriate checklist standards, a data hub, at least one analytical workflow and derived analytical results, and will serve as a tool for outreach and user support for our service portfolio as it matures.

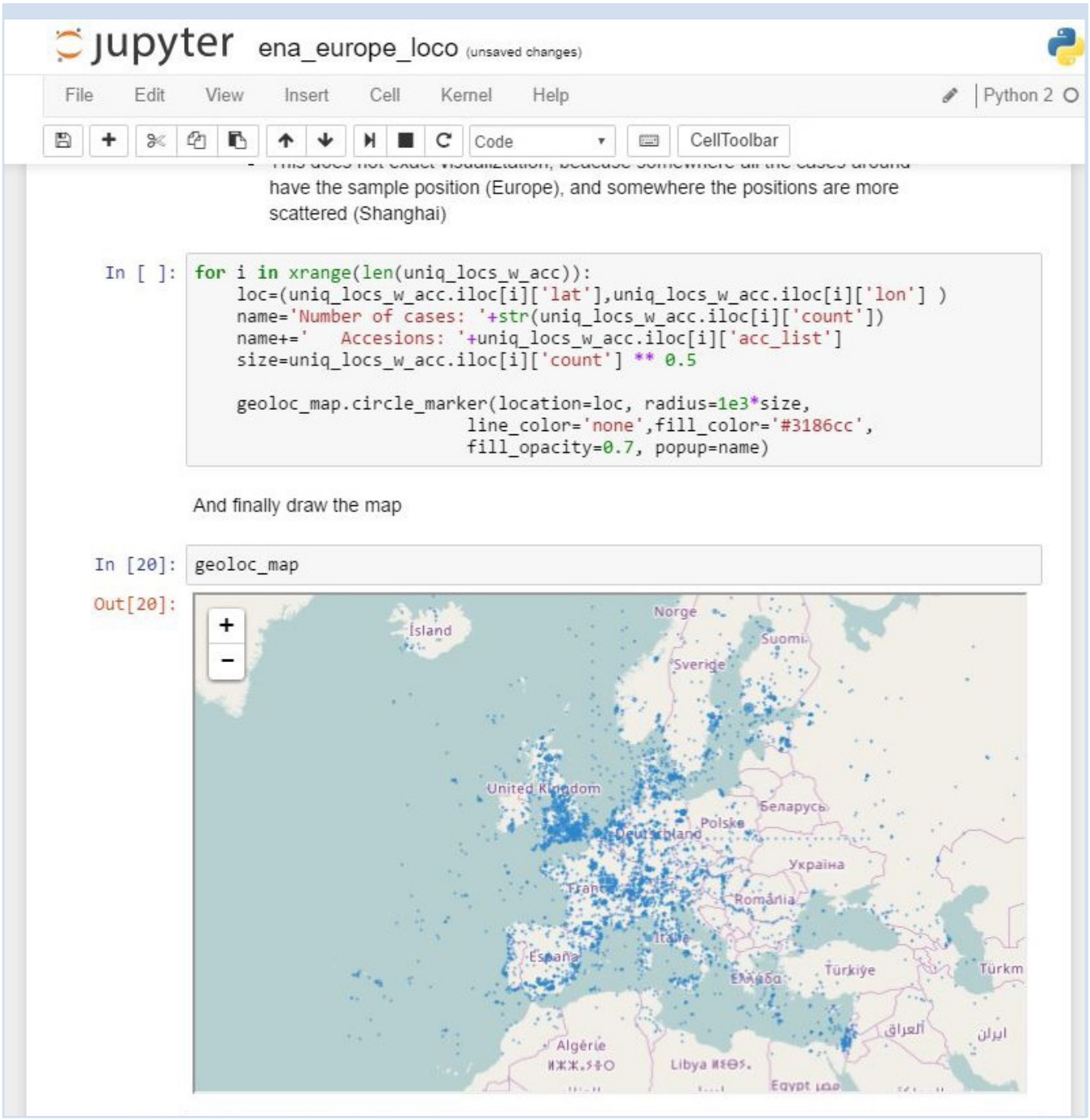


Figure 5. Distribution of geo-tagged ENA datasets in Europe visualized in an interactive IPython notebook.

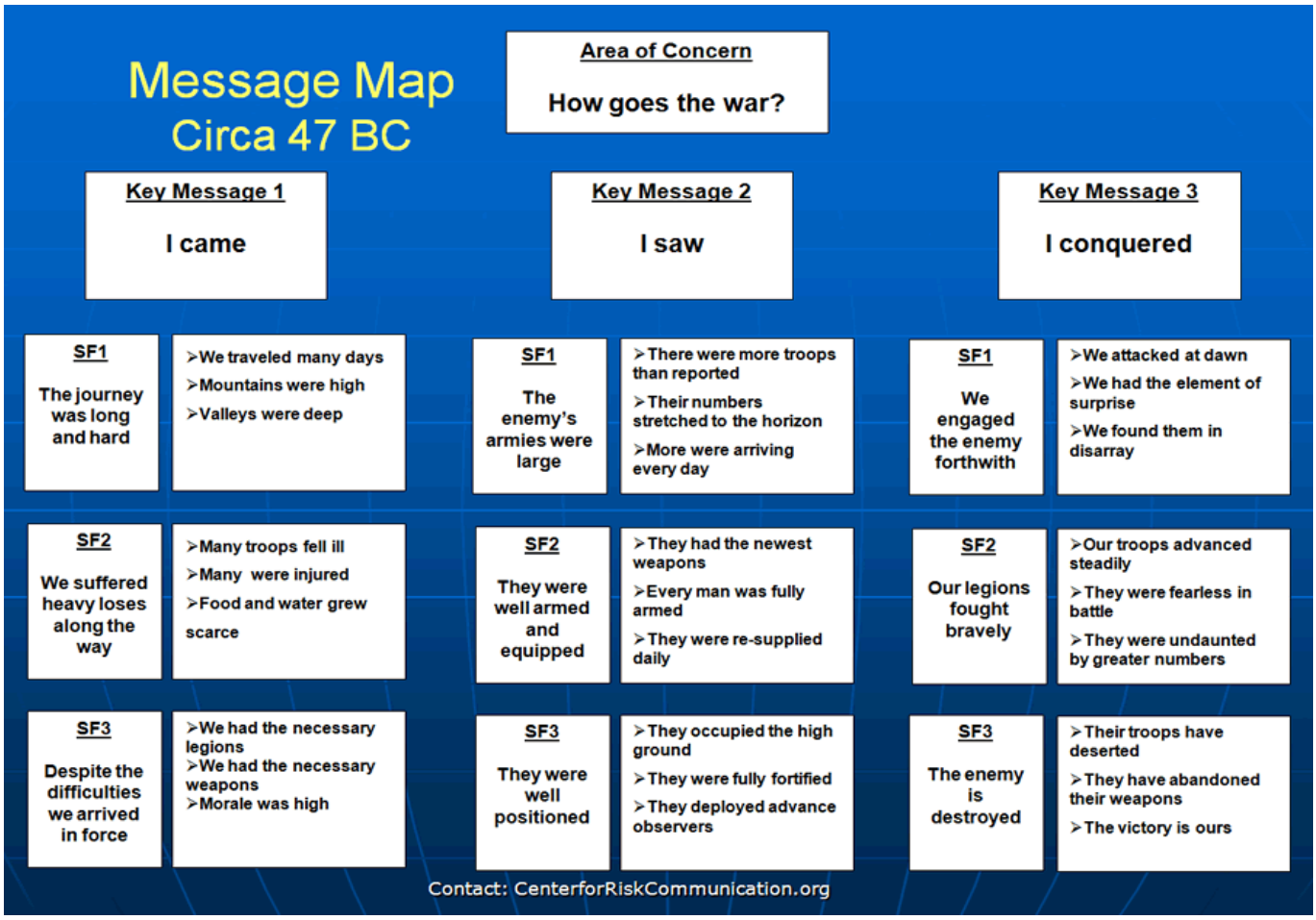
COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

WP10 – Risk communication tools

A comprehensive stakeholder inventory and analysis have been carried out.

The main activities performed most recently chiefly concern, and are related to, the development of the message mapping exercise. A message map is a roadmap for displaying detailed, hierarchically organized responses to anticipated questions or concerns. The final goal is to provide COMPARE with a set of well-focused, clear, concise, messages to be used in risk communication. The first step was to elicit a list of specific questions and concerns. This list has been generated during the reporting period through desk research, including media content analysis, reviews of historical documents (such as public meeting transcripts), book reviews, academic paper review, and alike. The second step, which is currently in progress, consists in putting in relation the list of specific questions and concerns with the stakeholder inventory finalized previously (D.10.1).

In parallel with this main group of activities, WP10 has continued publishing its weekly bulletin on Risk Communication and EIDs (26 issues published form March to August 2016).



COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

WP11 – User consultations

The establishment of the Expert Advisory Panels (EAPs) has been completed. The EAP members have all been briefed on the overall vision for the project and on the structure and work plans.

The EAP members that have provided feedback through EAP round 1 raised many points for discussion; the report of the first round of EAP meetings was shared with the WP leaders (Executive Board) in preparation of the joint session of the COMPARE Executive Board and EAP members for the second round of EAP meetings.

The second round of the COMPARE EAP meetings was held during the annual COMPARE meeting in Copenhagen, Denmark (March 8, 9, and 10, 2016). At the end of the general meeting, a COMPARE Executive Board and EAP member meeting was held. During this meeting, an extensive Q&A session during which the EAP members were invited to provide their feedback as well as any concerns they identified to the Executive Board on the (progress thus far of the) project. In total 17 EAP members attended the second round of EAP meetings, representing the following EAP panels: wildlife (3), public health (2), food safety (5), clinical health (1), domestic animals (3), databases and tools (1), barriers (2).

Furthermore, access to the COMPARE Share Site (read only) was provided to the EAP members to download relevant information related to COMPARE and its project results.

Between May and August 2016, the EAPs participated in a survey performed under WP12 into the barriers to open data sharing that are related to ownership of data and regulations that apply to ownership.

WP12 – Barriers to open data sharing

The focus is on understanding barriers to the development of COMPARE, in terms of legal, ethical, administrative and other considerations that may play a role.

During the COMPARE General Meeting, March 9, 2016, a presentation was given about the focus the WP is developing towards topical and fundamental issues: Towards a Design for an inventory and assessment of key barriers at different community levels.

Between May and August 2016, a qualitative survey was performed of step-limiting barriers through questionnaire and 50 interviews as perceived by scientists, public health officers, and management/ policy makers in the European food-health and human-health domains. This study resulted to date in a concept article: Regulatory and Ownership Barriers for Microbial genomic Data Sharing: Which are the Real Issues?

In the meantime frequent teleconferences and mail discussions took place with and between the WP12 Expert Advisory Panel members on academic studies, the interpretation and consequences of several international treaties that touch on genomic data sharing.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

These studies and discussions resulted in a key note presentation and discussion paper for the annual congress of Global Microbial Identifier at FAO/Rome 23-25 may 2016 (GMI-9): The access and benefit sharing paradigm: Contributing to developing global conditions that support and promote open access to and free exchange of microbial genomic information and related metadata (COMPARE / Haringhuizen).

(Courtesy: Globus Team / Ian Foster / foster@anl.gov)

The collaboration between COMPARE and other EU-funded projects is becoming an issue, as more and more colleagues' projects are looking for a NGS data platforms and tools. The EU project VIROGENESIS was advised on a Code of Conduct compliant to EC conditions and to facilitate sharing of data through a COMPARE provided repository/environment. The European Virus Archive-global (EVAg) project, building a (virtual) biobank library of virus collections, has asked for expertise and input on the analysis of constrains as a consequence of the EU Regulation on the Nagoya Protocol.

WP13 – Dissemination and Training

To ensure that relevant stakeholders of COMPARE are adequately informed about COMPARE's progress and results, many vehicles for dissemination have been established: Twitter account (@CompareEuroe), website (www.compare-europe.eu), web-based news distribution, and a participant portal (members-only, www.share.dtu.dk).

The COMPARE public website is continuously updated with information regarding upcoming events, news of notes and the most recent publications.

The COMPARE Twitter account (@CompareEurope) shares related news about COMAPRE and from the project partners.

Regarding Training, COMPARE will utilize e-learning that can be accessed online via the COMPARE portal. A draft overview of the various training materials is under development. At this time, no e-learning materials have been produced yet.

COMPARE will develop modular workshops for the organizations in the Expert Advisory Panels (EAPs) and others. In these workshops, principal investigators will present the analytical tools and software tailored to the specific needs of the workshop audience. COMPARE still plans a minimum of ten workshops.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

WP14 – Cost-effectiveness framework

The important elements in calculating costs and benefits of COMPARE and related methods and tools have been identified (Deliverable 14.1). This corresponds to the first of the five objectives specified for the activities of Work Package 14, namely "To identify the important elements in calculating costs and benefits of COMPARE and related methods and tools both regarding the system itself and from a societal perspective." Important activities conducted for this deliverable consisted of literature reviews and semi-structured interviews. In parallel, we have prepared (in cooperation with the COMPARE project leaders) an article "Developing a framework for assessing the cost-effectiveness of COMPARE - a global platform for the exchange of sequence-based pathogen data", accepted for publication in Vol. 36 (1), April 2017 of the OIE Scientific and Technical Review.

Furthermore, we have started working on the first case study meant to illustrate the potential costs and benefits of COMPARE. The case study revolves around the Ebola epidemic in Western Africa and aims to estimate the costs and benefits of early and targeted response in an epidemic scenario in low income countries. Currently, we are setting up a model that describes the course of the Ebola epidemic and allows for various policy scenarios.

WP15 – Management

The appropriate organizational structures and processes have been put in place to respond to the EC's as well as partners' needs and to ensure COMPARE's compliance with the EC Grant Agreement and the COMPARE Consortium Agreement.

The major activity in the last year has been to collect and compile WP achievements and financial reports for the Periodic Report for the First Reporting Period. The Periodic Report was submitted on time to the EC and provided the required information and data.

The next tasks include maintaining the reporting structure and documenting minor adjustments to budgets and tasks for partners and workpackages.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.

COMPARE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643476.