Declaration: Virus Outbreak Data Network (VODAN) GO FAIR Implementation Network

Introduction:

The spread of the virus causing the COVID-19 outbreak is far from over. As of March 4, it has been detected in over 75 countries. Fatality rates range from 0% (so far) in some countries to 3-4% (China, Italy) and up to 7% in some other countries (USA).

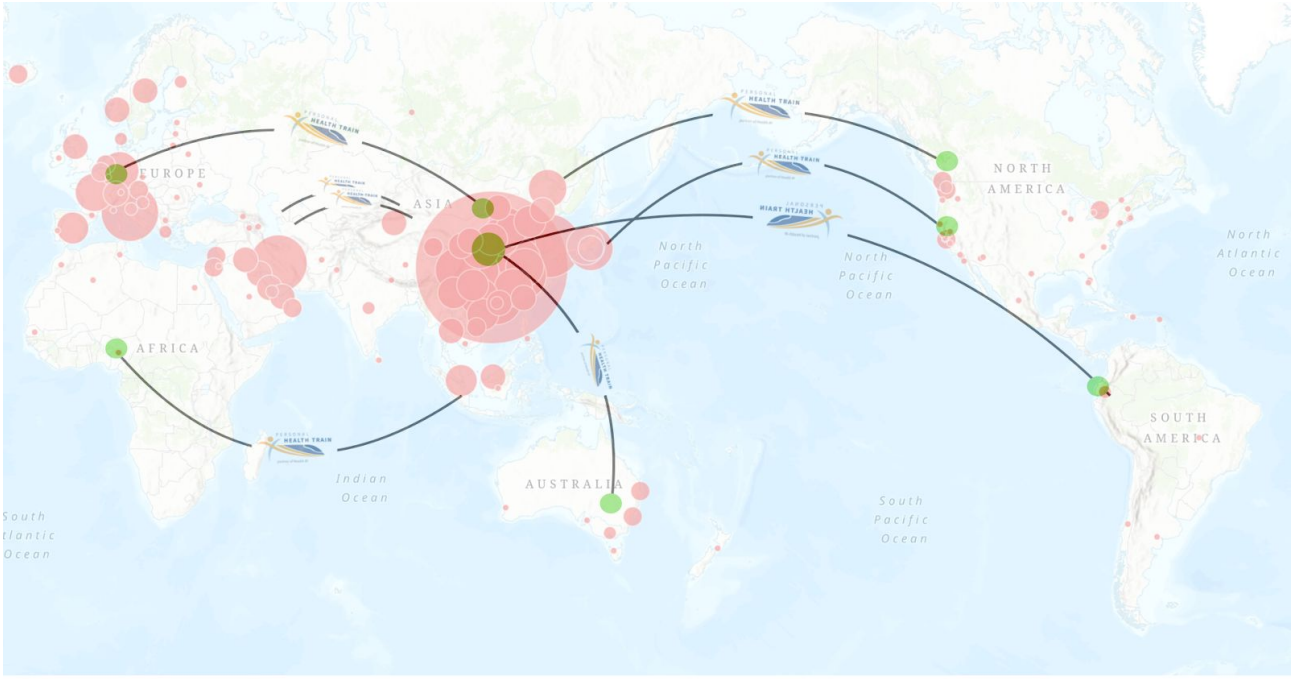
During this epidemic and in earlier occasions, we have seen severely suboptimal data management and data reuse. Moreover, access to the immensely valuable data of past and current epidemics is not always equally accessible for different affected populations and countries. For instance, the data from the past Ebola epidemics are very difficult to find, to access, and if accessible, they are not interoperable, *let alone reusable*. In the case of Ebola this is even more harrowing and ironic as the data are *least available* to the population that were *most affected* by the disaster. Under the urgent need to harness machine-learning and future AI approaches to discover meaningful patterns in epidemic outbreaks, we need to do better and ensure that data are FAIR (in this sense also meaning **F**ederated, **AI**-**R**eady).

Purpose of the Implementation Network:

This time, we can do better. We now have the technical ability, as well as the commitment from experts in a series of affected countries, to make the SARS CoV-2 virus data FAIR, meaning that they are Findable, Accessible, Interoperable and thus Reusable by both humans and machines, during this epidemic of COVID-19. The technical components that make this possible can remain in place, waiting in the ready state *for potential future infectious disease outbreaks*.

With a sense of urgency driven by the rapid developments on COVID-19 we come together to launch a GO FAIR Implementation Network to address the current and immediate challenges. For this epidemic, unfortunately, we have to 'FAIRify' COVID-19 data '*after the fact*' and use Chinese, Dutch, Swedish, etc. and English electronic (or even hand-written) health records to create proper FAIR data. The FAIRification will initially focus on the Clinical Research Form (CRF) model following the WHO standards. Multiple IN partners will create input forms that make it easy for local caregivers to create FAIR-CRF data in real time as a first step. As a second step, we will jointly develop (via online work sessions) localized FAIR Data Points (FDP). FDP is a FAIR data repository with 'docking' capabilities as a 'station' for 'trains' (virtual machines (VMs)) that come to 'visit' the data locally, with a specific question to ask. The local data custodian (frequently a hospital or centre for disease control and prevention type of institution) grants permission to VMs to ask the question / run analyses. As the personal data of patients never leaves the underlying database of the local institution, GDPR issues are largely accommodated and in this way data can be 'shared' or rather 'visited' without violating any patient rights and, in the case of a disease outbreak, also governed by the laws and policies of the individual jurisdictions in which the outbreak manifests.

Fig 1. below gives an impression of the desired situation: Trains (VMs) can visit multiple local FAIR data points to get their questions answered. For more information on Personal Health Train IN see here.



Overarching Principle of Operation

"We commit to comply with the Rules of Engagement of GO FAIR Implementation Networks".

Members of this rapid response VODAN IN will work primarily remotely in collaborative online work sessions to refine standards, build tooling and deploy FAIR Data Points.

Targeted Objectives and Tasks

The **VODAN IN consortium** is a light-weight public private partnership (members listed below) that will jointly address in a stepwise fashion, the following issues:

1. **Ensure** that the WHO-CRF(s) for Corona (and later viral outbreaks in general) are properly mapped to a machine readable (RDF) format, so that any stakeholder can create input forms that lead to the resulting data being a machine actionable (FAIR) digital objects (a machine actionable and readable file, which can interoperate between human languages and local data models).
2. **Create** multiple user-friendly input systems (e.g., web forms) that create interoperable (FAIR) data 'upon save'. Castor, one of the IN members, already took a first step[1].
3. **Assist** partners in affected regions to use traditional source files to create FAIR versions of selected data available in their country with local experts (domain experts, EHR experts and Semantic Data experts), mainly with Online Collaboration sessions.
4. **Install**, jointly with the local partners, a local FAIR data point (FDP) -or multiple points, in case the data are not centrally collected in the country. Again, this is a remote working session, as an FDP is a server application that can be installed in partnership with local institutions, according to the local specifications and in compliance with regional laws.
5. **Deliver** (as the deliverable of phase 1), with the partners in the participating countries, a series of FDPs that can be 'visited' under well-defined conditions by VMs (trains) to answer questions and discover patterns in these 'real world observation data'. (note: this will not yet be a fully automated distributed learning environment).
6. **Starting point of phase 2** is to **demonstrate** the value of this approach to WHO and other (national and international authorities and initiatives such as GLOPID-R) and seek certification and WHO approval for FAIR compliant CRF forms, FDPs and access protocols that do not violate personal privacy and/or national legislation.

[1] https://www.castoredc.com/blog/castor-vs-the-coronavirus-covid19/

7. Advanced stage: **Develop** a FAIR conceptual model for viruses and viral outbreaks originally based on CRFs and the sort of data CDC/RIVM-type institutions typically collect. The conceptual model will effectively create a FAIR digital twin (FDT) of individual patients (which can be anonymised), of the Virus (variant), sub-populations, etc.
8. **Offer** the 'Real World Data' FDPs under **agreed conditions** to qualified research groups, institutions and private companies to use the data (by controlled querying, ***not downloading***) to answer questions that may lead to the discovery of patterns in real world and established knowledge data, which in turn may lead to new prevention and intervention options.
9. **As a final step** before the IN will **dissolve and hand over** its assets to either **standing organisations** or a new, **larger IN** dealing with a wider approach to real world data FAIRness (in a preparatory phase now), including registries of vaccines drugs, side effects etc., the consortium will **document and publish** in Open Access and under CC-by license, all **specifications** that allow future repetition of this process for new epidemics that will undoubtedly confront society with similar challenges.

EXPECTATION MANAGEMENT statement: This IN will **NOT include** the actual research projects that may make use of the FAIR data point network created. Third parties (that may or may not include partners also participating in the VODAN IN) will be able to gain access to the data under the conditions set by the data custodians and/or WHO. FAIR data form a **substrate** for machine-assisted research, and are not a solution or a goal in themselves. In addition, the IN itself will not be able to sustain, or expand the service beyond its initial activity and funding cycle. It is therefore crucial that on 'day one' a scalability and sustainability model of the services will be developed by the partners (including the private service providers in the IN, but this is the concern of a subgroup and not a stated aim or deliverable of the IN. It will be also crucial for the Real World Observation IN that is under development in GO FAIR).

Membership list:

We consider this Manifesto to be one way by which the undersigned stakeholders can **speak with one voice** on a number of critical issues that are of generic importance to the objectives of FAIR, and on which we feel we have reached consensus.

Castor
University of Twente
GO FAIR Foundation
ZonMW (Netherlands Organization for Health Research and Development)
Databiology; (Luke Smith, Folkert Saathoff and Les Mara)
FAIR Solutions
Zorginstituut Nederland
University of California
Indiana University
Leiden University Medical Center
Chinese Academy of Sciences Computer Network Information Center (other Chinese partners may be advised by CAS and asked to join)
Euretos
Codevence
Czech Technical University
Phortos Consultants
...
Note: more partners are expected to join the IN over next few weeks.

Date: 10 March 2020