VODAN IN Introduction

Data Stewardship Cluster

Barend Mons, VODAN IN Interim Coordinator

April 15, 2020
Launched 12 March, 2020

VODAN IN Goals

Virus Outbreak Data Network (VODAN)

The VODAN Implementation Network is one of the joint activities carried out by CODATA, RDA, WDS, and GO FAIR. Read the full statement on Data Together COVID-19 Appeal and Actions.

Active GO FAIR Implementation Network

The spread of the virus causing the COVID-19 outbreak is far from over. 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. 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 Federated, AI-Ready).

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 ready state for potential future infectious disease outbreaks.

We started this implementation network with a very narrow focus, based on seed funding from co-founding partners, ZonMW and the Philips Foundation (see the manifesto), namely to make source data FAIR and make them available for reuse in a distributed manner. With a sense of urgency driven by the rapid developments on COVID-19 we came together to launch a GO FAIR Implementation Network to address the 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.

Trains (VMs) can visit multiple local FAIR Data Points to get their questions answered. For more information on the underlying technological approach please visit the Personal Health Train IN pages. The data stewardship aspects of FAIR data will be addressed wherever possible with the Data Stewardship Competence Centres IN.

Targeted Objectives

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

- Ensure that the WHO-CRF(s) and other input forms for Corona data (and similar outbreaks in general) are properly managed to facilitate better FAIR data sharing & reuse.
VODAN IN Goals

Subscribed Students

 Decision support
- Trusted World of Corona
- Multiple visualisations
- Vettel new assertions

Common Goals

1. New concepts (subject to review)
2. New assertions (subject to review)

F-DP’s

Basic (student) checks (concepts/triples)
annotation client
web service
RDF nanopublication annotator

E Harvester

F

1. Immunology
2. Epidemiology
3. Drug Repurposing
4. Diagnostics
5. Life Style
6. Social aspects

Publication Records (ORCID)

For now I would go for the English version of the maintained UMLS 2019AB

Literature Ingestor A
Literature Ingestor B
Medline Ingestor
VODAN IN Organisational Structure

Fig 1. The rapid growth of partners in the first weeks of the Virus Outbreak Data Network.

--> Need for structure and introduction of clusters within the IN
VODAN IN Organisational Structure

IN Coordination

• Overall VODAN IN coordination by Barend Mons, supported by GFISCO
• Projects handled under GO FAIR Foundation coordinated by Bert Meerman
• Individual cluster leads identified -> DS Cluster (co)lead: Laurents Sesink

VODAN Clusters

<table>
<thead>
<tr>
<th>E-INFRA CLUSTERS</th>
<th>APPLICATIONS CLUSTERS</th>
<th>DATA SOURCES CLUSTERS</th>
<th>OVERARCHING CLUSTERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIR Core Specs</td>
<td>Data Stewardship</td>
<td>New Publications (FAIR Annotations)</td>
<td>Advice/Overview</td>
</tr>
<tr>
<td>Compute – Routing Infra</td>
<td>Community Annotation</td>
<td>FAIR Sensitive Citizen Data</td>
<td>International Support and Coordination Office</td>
</tr>
<tr>
<td>Compute – AAI, Privacy and Security</td>
<td>Conceptual Modelling</td>
<td>FAIR Meta-Analysed Citizen Data</td>
<td></td>
</tr>
<tr>
<td>Compute – CPU-provisioning/HPC Capacity</td>
<td>Machine Learning and Bioinformatics</td>
<td>Data Entry</td>
<td></td>
</tr>
<tr>
<td>Online Training and Instructions</td>
<td>Established Knowledge</td>
<td>Visualisation and Intellectual Support</td>
<td>International Coverage/Policy</td>
</tr>
</tbody>
</table>
The key take-home messages:

**In the short term**

- Open access to relevant scientific information and datasets should be ensured through the creation of a global open access platform/shared repository for scientific data on COVID-19 and other pandemic research, treatment and prevention.
- UNESCO was called on to bolster a centralized open access portal (repository) for the sharing of scientific data and knowledge.
- Unanimous support was given to UNESCO to coordinate an international Open Science coalition to respond to pandemics, environmental crises and other global threats, and to reduce the knowledge gap.
- UNESCO was called on to establish a forum/platform/network to support cooperation between science ministers by sharing experiences on the current crisis and ways in which to address future crises through scientific knowledge.
- Additional investment in science and innovation and international scientific collaboration should be an imperative for a multidisciplinary approach to the COVID-19 and beyond;
- Open access to relevant digital tools and open source products should be encouraged to enable all segments of the population to obtain the information needed to combat the crisis.
- Policy incentives should be shared and measures taken to ensure science-based decision-making and sharing of information in the event of pandemics.
- UNESCO was called on to launch a multi-sectoral multi-dimensional initiative covering all of the Organization’s fields of competence, including education, the social and human sciences and the natural sciences, to assist Member States in fighting the pandemic.
VODAN IN Internal Communications

- Overall VODAN IN Communications handled Barend and/or GFISCO
- Calls of clusters leads will be organised by GFISCo as needed
- Expectation that the clusters are fully self-organised, with support of GFISCO

GFISCO Support
- Advice and consultation on request -> the cluster lead can contact GFISCO
- **Collaboration space**
  - DS Cluster space in OSF can be used, if desired: [https://osf.io/89eqg/](https://osf.io/89eqg/)
  - GitHub for INs: [https://github.com/go-fair-ins](https://github.com/go-fair-ins)
- **Communications**
  - Member list (incl. emails) maintained by GFISCO and shared with the cluster lead(s)
  - DS Cluster Slack channel available, if the cluster wishes to use it
  - Zoom channel available to INs: link allowing to book Zoom for the cluster shared with the the cluster lead(s)
SARS-CoV-2/human proteome

ACE2/ACE

Cytokine Storm (CRS)

Renin-angiotensin system

Further sub-hypotheses

Greatest Common Denominator (GDC) ‘Connectome’

Phasing Experts

FAIR abstract Paper 1

FAIR abstract Paper 2

FAIR abstract Paper 3

Paper 4

…………..
Example of ‘connectome’ (ACE2 with the semantic type ‘disorders’)

- pulmonary hypertension
- severe acute respiratory...
- sarcoïdosis
- pulmonary fibrosis, non-inflam...
- kidney failure, chronic
- diabetes mellitus, insulin...
- vascular diseases
- hypertensive disease
- sarcoidosis, pulmonary
- chronic heart failure
- myocardial infarction
- ischemia
- congestive heart failure
- obesity
- kidney disease, cerebrovascular accident
- hypertension, renovascular...
- essential hypertension
- renal glomerular disease
- Alzheimer's disease
- renal hypertension
- ACE2 (homo sapiens)
- arteriosclerosis
- endothelial dysfunction
- coronary arteriosclerosis
- acute myocardial infarction...
**CT:** negative / Perfusion deficits / hyperemia / GGO / Vascular thickening / Interlobular septa / GGO >> / >> Consolidations / embolism /

**PA:** Endotheliitis / DAD / MVO / multiple MVO / Venous thrombosis / Infarctions / Organizing Pneumonia / Macro-thrombi / heart failure / death

**Non-survivors:** progression after 10 days / D-dimer >>>1000

**Survivors:** repair after 10 days / D-dimer < 1000
<table>
<thead>
<tr>
<th>CT characteristics</th>
<th>Normal</th>
<th>Normal</th>
<th>Perfusion defects/GGO/hyperemia</th>
<th>Infarction/GGO (increase)</th>
<th>Interlobular septa, Pulmonary embolism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase</td>
<td>Phase 1 (mild symptoms)</td>
<td>Phase II (onset Cytokine storm)</td>
<td>Phase III Early endothelial/vascular disorders</td>
<td>Phase IV Vascular and thrombotic disorders (severe)</td>
<td>Phase V Severe, life threatening manifestations</td>
</tr>
<tr>
<td>Lung</td>
<td>Cytokine release syndrome, shortness of breath</td>
<td>endotheliits/Vasculitis,</td>
<td>Diffuse alveolar damage, microvascular obstruction, Hyperemia, obstruction, Capillary leak syndrome,</td>
<td>Venous thrombosis/permeability disorders, extravasation</td>
<td>Organised pneumonia, pulmonary embolism.</td>
</tr>
<tr>
<td>Heart</td>
<td>arrhythmia, tachycardia</td>
<td></td>
<td></td>
<td>Myocardial dysfunction</td>
<td>Heart failure</td>
</tr>
<tr>
<td>Brain</td>
<td>Stroke</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acute renal failure</td>
</tr>
<tr>
<td>General symptoms/manifestations</td>
<td>Nausea, shortness of breath, dyspnea, headache, tachycardia, exanthema, coughing, fever, hypotension</td>
<td>Macrophage activation syndrome, cell injury, Histiocytosis, Hematophagocytic, reactive hemaphagocytic syndrome, Lymphohistiocytosis, ham...</td>
<td>Acute inflammatory disease, systemic inflammatory response</td>
<td>Thrombus, blood coagulation disorders, hemorrhages</td>
<td>Septicemia, hemodynamic instability, DIC, diffuse hemorrhagic shock, toxic shock syndrome, anaphylaxis, multiple organ failure</td>
</tr>
<tr>
<td>Direct interaction SARS-COV-2 proteome-human proteome</td>
<td>GGCX, SLC9A3R1, HDCA2</td>
<td>GNG5, RHOA, FGFR1OP, CNTRL, PKRCA, TTGB1, HMOX1, IL17RA, PABPC1, LOX</td>
<td>PTPN11, CD14, CD28, TLR4, MPO, C5, IFNG, TLR3</td>
<td>CCL3, CSF3, CXCL10, GPX1, KNG1, IKL6R, HMOX1, PPARA, VEGFA, TGFß1, APP</td>
<td></td>
</tr>
<tr>
<td>Genes affected in human proteome</td>
<td>CRP, CSF2, CD19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processes</td>
<td>Viral induction?</td>
<td>Cytokine release syndrome</td>
<td>Transition</td>
<td>Onset cytokine storm</td>
<td>downstream effects renin-angiotensin disturbance</td>
</tr>
<tr>
<td>Small molecules</td>
<td>Glutathione</td>
<td>Myeloperoxidase</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>