GO FAIR US Webinar Series: FAIR Data & COVID-19

Prof. Dr. Mirjam van Reisen
VODAN-Africa / Training of Trainers (ToT)

8 April 2020

#VODAN
Webinar Series Overview

1 April
- Intro to VODAN
  • Prof. Dr. Barend Mons
  Recording at https://tinyurl.com/rds-gofair

8 April
- VODAN-Africa / Training of Trainers
  • Prof. Mirjam van Reisen

15 April
- FAIR Data Work in Action
  • Albert Mons, Luiz Bonino

22 April
- Fighting COVID-19 by Mining Insights from Heterogeneous Datasets
  • Peter Rose & Ilya Zaslavsky, Iris Shen, Natalie Meyers & Eric Morgan
Virus Outbreak Data Network: Training of Trainers

Key Messages

• By Prof. Dr Mirjam van Reisen, Professor ‘Computing for Society’, Leiden Institute for Advanced Computer Science, Leiden University

• In collaboration with Kampala International University, Mekelle University, Great Zimbabwe University, Olabiso Olabanjo University, Leiden University, GO FAIR Foundation and Philips Foundation

• April 8, 2020 18:00 CET, Webinar @ Centre for Super Computers, San Diego Super Computer Centre
Outline

Objectives
Training of Trainers
Next steps
I. Objectives
Coronavirus exposes lack of common data approach

"One of the main problems is the duplication of data and lack of coordination between countries", said Mirjam van Reisen, who is part of the Virus Outbreak Data Network (VODAN) at the University of Leiden.

Some experts believe that the EU’s current approach to data collection may be missing an opportunity to fight the coronavirus outbreak (Photo: Christiaan Colen)

By ELENA SÁNCHEZ NICOLÁS
BRUSSELS, TODAY, 07:21

The enormous differences between coronavirus cases reported worldwide raises questions on how countries are tracking their coronavirus outbreaks - or deliberately underreporting them.
Lack of common data approach: defining the problem

- Increasing (unconnected) digital health data:
  - Patient data
  - Research data
  - Published articles

- CORD-19 is a resource of over 47,000 scholarly articles, including over 36,000 with full text, about COVID-19, SARS-CoV-2, and related coronaviruses

- Too much for manual handling, but current internet not equipped for *machine*-readability of data

- Data need to be prepared for human- and machine-readable Findability, Accessibility, Interoperability, Re-usability (FAIR)

- This constitutes the basis for the Internet of FAIR Data And Services
Internet of FAIR Data and Services - Connecting Principles

Australian Research Data Commons
Data Science meets AI:

Quality data for Federated AI Ready (FAIR) solutions

- Distributed data
- Data remain in their place of origin
- Data are managed and stewarded in their place of origin
- Allows contextualized data analytics
- Fake data are detectable and traceable
- Allows Data sovereignty within legal and policy framework of the location where data is produced
- Connects fragmented data depositories
- Algorithms (tools) visit the data
- Computes and analytics on the basis of data visiting
Specific benefits of VODAN for Africa

- Avoid digital data removal to warehouses elsewhere
- Strengthen data-informed Health Systems
- Ensure data ownership and handling
- Strengthen digital data stewardship and tooling
II. Training for Trainers:
Building a Community fit for purpose
VODAN: Establishing a solid network of trust and purpose

- **Step 1**: Decide
  - countries for FDPs: Uganda, Ethiopia, Zimbabwe and Nigeria

- **Step 2**: Identify
  - Training of Trainers' hubs

- **Step 3**: Inform
  - Bureau of Health and or Ministry of Information and build a contact with them

- **Step 4**: Contact
  - the hospital/clinic to work with in your environment

- **Step 5**: Explain
  - What the project is about

- **Step 6**: Establish
  - A working relation with the hospital/clinic
Set up information channel with MoH / Bureau of Health

- Approach the MoH / Bureau of Health
- Provide information on VODAN and who are involved
- Arrange how you will communicate progress
- Ask for their feedback
- Request endorsement
## Determining the governance & regulatory framework

<table>
<thead>
<tr>
<th>Identiﬁe</th>
<th>the relevant policy documents and regulations on medical data handling;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorise</td>
<td>categorise the do’s and don'ts of the medical data regulatory framework;</td>
</tr>
<tr>
<td>Involve</td>
<td>involve relevant people who understand laws and regulations on medical data handling;</td>
</tr>
<tr>
<td>Explain</td>
<td>the purpose of the task to the experts in MoH and hospital/clinic;</td>
</tr>
<tr>
<td>Ask</td>
<td>ask colleagues in MoH &amp; hospital &amp; clinics to review the document and improve it;</td>
</tr>
<tr>
<td>Check</td>
<td>if you have all the right people on board to determine the regulatory framework;</td>
</tr>
<tr>
<td>Involve</td>
<td>any senior people if necessary to do final check;</td>
</tr>
<tr>
<td>Present</td>
<td>a regulatory framework</td>
</tr>
</tbody>
</table>
Training of Trainers:

Questions to consider for identification of the hospital or clinic (health data producer)

• What is the hospital/clinic you will work with?
  
  • Preferably there is a working relationship and trust with the hospital management
  
  • Ensure clear communication on the reason for the involvement of the hospital
  
  • The data-manager of the hospital/clinic is invited to the training so that knowledge is transferred
  
  • Alternatively follow up training is arranged with the data-manager of the hospital/clinic
  
  • Multiple hospitals/clinics can participate, but for the first step it should be a manageable number of participating outfits – expansion will happen after the first successful steps are implemented
Compilation of information flow charts

- **Ask**
  - permission from hospital/clinic to compile an information flow chart on clinical data (written permission would be best)

- **Explain**
  - that you do not need to have access to the data itself (only the forms)

- **Request**
  - access to key resource persons handling information in the hospital/clinic

- **Interview**
  - the key resource persons (this can be by phone) and ask how information is compiled

- **Identify**
  - the different forms/templates for information

- **Understand**
  - where the repository of the information is
Understanding the information flow
Training of a COVID-19 Data Community

1. To establish the basics of the FAIR Data Point:
   - Officials in health administration (governance, regulatory framework)
   - Health professionals (to work with the information flow and data handling in hospitals and clinics)
   - Data scientists, semantic data modelers, interdisciplinary data stewards (to define and model the data in human and machine-readable ontologies and determine informatical models)

2. In subsequent phase:
   - Experts in Machine Learning and Artificial Intelligence, text mining tooling
   - Experts in Algorithm auditing
   - Experts with Legal knowledge for regulatory embedding of the use of the FAIR Data Point
   - Clinicians and health analysts for health data analytics
   - Researchers, experts for the formulation of hypotheses and judging cardinal assertions derived from data analytics
III. Next Steps: implementation
Decision-making in FAIRification of the COVID-19 Data Community

- Application of FAIR Principles based on:
  - Technical specifics (infrastructure) – code red
  - local authorities (governance and regulations) and domain specialists (doctors, virologists, researchers) – code blue

The COVID-19 Data Community determines:
Data as Open as Possible and as Closed as Necessary
As Distributed as Possible, as Centralised as Necessary

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier.
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
   A1.1 the protocol is open, free, and universally implementable
   A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. (meta)data are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards
Information structure of COVID-19 semantic model of patient data:

1. Digital Data Entry at hospital/clinic
2. Machine Readable interoperability of existing digital entries (ensuring interoperability and reusability)
3. Defining metadata (ensuring findability and establishes conditions for accessibility)
VODAN Project: Social-Technical Specification

Phase 1
Phase 2
Phase 3

Regional FAIR Data Point

FAIR-at-source
Data / Metadata capture tools

Information Stream:
Local process of recording patient data, often with paper and pencil

Metadata pointing to eCRFs (VODAN standard as RDF)

Designation 1
Designation 2
Designation 3...

FIP development:
- Community schema
- FIP RDF
- FIP presentation schema

FAIRification of patient records
FAIRification of patient records
FAIRification of patient records
FAIRification of patient records

FDP development
- search
- Access

DSW castor CEDAR
There is no FAIR Data without Machine-actionable Metadata
Metadata for Machines Workshops

Domain Experts + Metadata Experts = Machine-actionable Metadata

The creation of community specified, reusable metadata templates that prevent the reinvention of the wheel and drive convergence and interoperability.
## Metadata for Machines Workshops

<table>
<thead>
<tr>
<th>Question</th>
<th>FAIR Principle</th>
<th>Metadata decisions</th>
<th>National</th>
<th>Regional</th>
<th>Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F1</td>
<td>What is the persistence policy for identifier systems used for data elements?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>F2</td>
<td>What are the minimal data elements needed to ensure Findability?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A2</td>
<td>What is the persistence policy for the metadata?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>R1.1</td>
<td>What usage license(s) will be used?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>R1.2</td>
<td>What are the minimal provenance metadata needed to ensure reuse?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>R1.3</td>
<td>Give a machine-actionable FIP.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Key Aims of the FAIR Principles

- Data to be as open as possible and as closed as necessary
- Protect personal data! Always!
- Embedded in local regulatory frameworks
- Maximize freedom to operate: Do not generate more work for the hospital/clinic
- Use the data recording mechanisms in place in the medical facility
- Build on any tool already available
- Ensure transparency of algorithmic tools
- Develop new Tools
- Triangulate cardinal assertions!
Virus Outbreak Data Network – Africa and Ambassadors

• PI: Mirjam van Reisen, Leiden University
  mirjamvanreisen@gmail.com

• Coordinator: Francisca Oladipo, Kampala International University
  francisca.oladipo@kiu.ac.ug

• Technical Support Group lead: Mariam Basajja, Leiden University/Kampala International University
  mariam.basajja@gmail.com

• Support ambassadors Implementation Network: Aliya Aktau, Leiden University Aliya Aktau
  aleka.aktau@gmail.com

• GO FAIR International Support and Coordination Office: Erik Schultes
Join and stay tuned:

Tune in next week for: VODAN, FAIR data in action

Join VODAN: tinyurl.com/join-vodan

Welcome at VODAN Training of Trainers: https://www.vodan-totafrica.info/

Join RDA-COVID19 WG:
www.rd-alliance.org/groups/rda-covid19

Data (Science) COVID-19 Resources:
https://www.academicdatascience.org/covid

Join the conversation: gofair.slack.com
Thank you!