1. **WG CHARTER**

A concise articulation of what issues the WG will address within a 12-18 month time frame and what its “deliverables” or outcomes will be.

In May 2020, the Organization for Economic Cooperation and Development (OECD) discussed why and how Open Science is critical to preventing and combating pandemics such as COVID-19 caused by the novel coronavirus, SARS-CoV-2 (OECD 2020). Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks (Vicente-Saez and Martinez-Fuentes 2018). FAIR (findable, accessible, interoperable, and reusable) data principles are an integral part of Open Science. FAIR data principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with no or minimal human intervention) (GoFAIR).

However, there is an urgent need to develop a common standard for reporting communicable disease surveillance data without which Open Science and FAIR data will be difficult to
achieve. Limited by antiquated systems and the lack of an established infrastructure, the tempo of the spread of the disease has outpaced our ability to react and adjust (Austin et al. 2020a,b; Garder et al. 2020).

The need for developing a common standard for reporting epidemiology surveillance data was articulated by the RDA COVID-19 Epidemiology work group (WG) in their recommendations and guidelines, and supporting output (RDA COVID-19 WG 2020; RDA COVID-19 Epidemiology WG 2020).

On October 27, 2020, the WHO, UNESCO, HCHR, and CERN issued a Joint Appeal for Open Science, a call on the international community to take all necessary measures to enable universal access to scientific progress and its applications UNESCO et al. 2020; UNESCO 2020:

"The open science movement aims to make science more accessible, more transparent and thereby more effective. A crisis such as the COVID-19 pandemic demonstrates the urgent need to strengthen scientific cooperation and ensure the fundamental right to universal access to scientific progress and its applications. "Open Science" is about free access to scientific publications, data and infrastructure, as well as open software, open educational resources and open technologies such as tests or vaccines. Open science also promotes trust in science, at a time when rumours and false information abound."

Michelle Bachelet, United Nations High Commissioner for Human Rights stated,

"Data are a vital human rights tool."

A properly aligned framework is needed.

The WG will build upon existing standards and guidelines to develop uniform definitions and data elements to improve data comparability and interoperability.

We will build upon the work begun by the RDA COVID-19 Epidemiology WG, and extend beyond the COVID-19 pandemic to provide an actionable specification for reporting communicable disease surveillance data and metadata, including geospatial data.

This work will be a consensus building effort that contributes to CODATA’s Decadal programme:

- Enabling Technologies and Good Practice for Data-Intensive Science
- Mobilising Domains and Breaking Down Silos
- Advancing Interoperability Through Cross-Domain Case Studies

Outcome
A standard specification for reporting communicable disease surveillance data.
2. VALUE PROPOSITION

A specific description of who will benefit from the adoption or implementation of the WG outcomes and what tangible impacts should result.

Epidemiology surveillance data will enable governments and public health agencies to detect and respond to newly emergent threats of disease. Early detection may prevent development of epidemics and pandemics. It will also enable them to deliver more effective responses at all stages of the threat, from emergence through containment, mitigation, and reopening of society in the case of pandemics. Epidemiology surveillance data and geospatial data are large and varied. Treated as a strategic asset, they have the potential to support evidence-informed policy, stimulate new research areas, expand collaboration opportunities, and increase the health and economic well-being of society. A common standard for reporting epidemiology surveillance data will support these outcomes by improving data and metadata management and provision of findable, accessible, interoperable, reusable, ethical, and reproducible (FAIRER) data.

The common standard for reporting epidemiology surveillance data is intended for implementation by government and international agencies, policy and decision-makers, epidemiologists and public health experts, disaster preparedness and response experts, funders, data providers, teachers, researchers, clinicians, and other potential users.

3. ENGAGEMENT WITH EXISTING WORK IN THE AREA

A brief review of related work and plan for engagement with any other activities in the area.

Nature of the problem to be addressed

The World Health Organization (WHO) defines public health surveillance as, “An ongoing, systematic collection, analysis and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice” (WHO 2020a). The WHO is a source of international standardized COVID-19 data and evidence-based guidelines, and is an invaluable source of technical guidance (WHO 2020b). Available instruments include a case-based reporting form, data dictionary, template, and aggregated weekly reporting form (WHO 2020c). There is also a global COVID-19 clinical data platform for clinical characterization and management of hospitalized patients with suspected or confirmed COVID-19 (WHO 2020d). The WHO (2020e) also notes that continued vigilance is needed to detect the emergence of novel zoonotic viruses affecting humans.

Unfortunately, there are inconsistencies in the manner in which various jurisdictions agencies collect and report their data. This is due to gaps in existing standards, and a failure to comply with those standards that do exist.

COVID-19 threat detection has been slow and ineffective, resulting in rapid development of a pandemic. Countries around the world have implemented a disparate series of public health measures in attempting to suppress and mitigate spread of the disease. The world was not prepared to respond to a novel zoonose that spreads with the tempo and severity of COVID-19 (Greenfield et al. 2020). The pandemic has resulted in serious health and economic consequences for both High Income Countries (HICs) and for Low and Middle Income Countries (LMICs) (Bong et. al, 2020).
The RDA COVID-19 WG recommendations, guidelines, and supporting output highlighted discrepancies in the number of COVID-19 incident and mortality data across data sources which could be directly attributed to varying definitions and reporting protocols (RDA COVID-19 Epidemiology WG, 2020a,b). For example, mortality data from COVID-19 are frequently not comparable between and within jurisdictions due to varying definitions (Dudel, 2020). Variations resulting from discrepancies in official statistics limit effective disease-specific strategies (Modi et al., 2020; Modig & Ebeling, 2020). Other variables (e.g., confirmed cases, probable cases, probable deaths, negative tests, recoveries, critical cases) are also inconsistently defined (Austin et al. 2020b). For example, while the WHO (2020f) defines a confirmed case as "a person with laboratory confirmation of COVID-19 infection, irrespective of clinical signs and symptoms", other datasets report confirmed cases as the number of both laboratory positive subjects and probable cases (JHU, 2020). The US CDC (2020) has amended its previous policy and now reports case counts from commercial and reference laboratories, public health laboratories, and hospital laboratories, but still excludes data from other testing sites within a jurisdiction (e.g. point-of-care test sites). In Turkey, the number of cases published until the end of July represented only symptomatic COVID-19 subjects, excluding asymptomatic laboratory positive individuals (Reuters Editors, 2020). Other issues affecting data accuracy include: duplicate event records, laboratory report delays, missing data, and incorrect dates.

Much of the developed world has a notifiable disease surveillance system for effective and efficient reporting within their national limits, with varying fields of data elements. There exist, also, a large number of international data standards which should be used when reporting epidemiology surveillance data (Table 1). However, these do not address specific requirements that would ensure that epidemiology surveillance data are comparable and interoperable.

**Table 1.** Initial list of types of data standards relevant to notifiable disease surveillance systems (non-exhaustive list) [SOURCE: Haghiri et. al. 2019].

<table>
<thead>
<tr>
<th>TYPE</th>
<th>STANDARD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification system</td>
<td>International classification of disease:</td>
</tr>
<tr>
<td></td>
<td>- ICD, ICD9, ICD9-CM</td>
</tr>
<tr>
<td></td>
<td>Other classification systems:</td>
</tr>
<tr>
<td></td>
<td>- DRG, CPT, ICECI, HCPCS, ICPM, ICF, DSM</td>
</tr>
<tr>
<td>Machine-organizable data</td>
<td>HL7</td>
</tr>
<tr>
<td>Markup language</td>
<td>XML Document Transform (CDT)</td>
</tr>
<tr>
<td>Medical document exchange format</td>
<td>Clinical Document Architecture (CDA)</td>
</tr>
<tr>
<td></td>
<td>Continuity of Care Document (CCD)</td>
</tr>
<tr>
<td></td>
<td>Continuity Care Record (CCR)</td>
</tr>
<tr>
<td>Nomenclature system</td>
<td>LOINC</td>
</tr>
<tr>
<td></td>
<td>SNOMED</td>
</tr>
<tr>
<td></td>
<td>Rx NORM</td>
</tr>
</tbody>
</table>
Disease surveillance systems rely on complex hierarchies for data reporting. Raw data are collected at local level followed by anonymization and aggregation as necessary before sending it up the hierarchy which includes many levels. Even in many of the most developed regions of the world, much of this process continues to be done by hand, although the push to electronic medical records is gaining traction. As a result, most disease surveillance systems across the world experience reporting lags of at least one to two weeks (Fairchild et.al 2018, Janati et.al 2015).

Publicly available data are made available on websites that are often difficult to navigate to find the data and associated definitions.

Historical data are not fixed the first time they are published due to undetected errors, late or missing data, laboratory delays, etc. The dataset is updated when the data becomes available. This problem, often called “backfill,” is due to the complex reporting hierarchy and antiquated systems that disease surveillance systems rely on. Backfill can in some cases drastically affect analyses (Fairchild et.al 2018). The problem is compounded when corrected and missing case counts are added to the date on which the correction was reported, instead of the date on which the event occurred.

Case definitions used in epidemiological surveillance data are not clearly defined. Fairchild et al. (2018) have highlighted the challenges with data reporting and stressed the importance of explicit and clear case definitions. Even with standardized definitions, regions with little support on funding for public health institutions may struggle to adopt a framework of best practices. We will develop guidelines that recognise these limitations and that will support both LMICs and HICs.

Engagement with other related activities

The proposed “Epidemiology common standard for surveillance data reporting WG” will address a high-priority challenge based on assessments of public health needs during a pandemic using COVID-19 as a use case.

The initial WG membership is well connected to various community-based initiatives and WGs that address similar and other relevant topics. The WG will monitor and align its efforts with other related activities, including:

RDA WG and interest groups (IG):
- e-Infrastructure projects e.g. OpenAIRE, EUDAT, European Open Science Cloud (EOSC)
- RDA Birds of a Feather (BoF) - Value of making data F.A.I.R for Infectious Disease Data Management and Analytics (IDDMA)
- RDA Common Standard machine actionable data management plan (maDMP) WG
- RDA COVID-19 WG
- RDA FAIRsharing Registry: connecting data and metadata standards, repositories and policies WG
4. WORK PLAN
A specific and detailed description of how the WG will operate.

**Deliverables**

D1 (months 4-12). **Epidemiology common standard for surveillance data reporting.**
This deliverable will contain the developed common standard specification for reporting epidemiology surveillance data, including variable names, definitions, and rationale.

D2 (months 12-16). **Guidelines for adopting the common standard.**
Guidelines will be based on lessons learned during development of the standard.

**Milestones**

M1 (months 0-1). **Engagement of representatives from prominent stakeholders in public health.**
We will seek engagement with the WHO, eCDC, US-CDC, ICMR etc.

M2 (months 0-3). **Identification of standards gaps and issues concerning data interoperability and comparability across and within jurisdictions.**
- We will use COVID-19 surveillance data as a use-case to identify issues that can be resolved by implementation of a common standard for reporting communicable disease surveillance data.
- Identify related standards and guidelines.
Flag the existing standards that would be cited as mandatory for compliance with the new RDA/CODATA Epidemiology common standard

- Identify standards gaps.

M3 (months 1-3). **Definition of the scope of the standard and detailed objectives.**
We will develop a detailed project management and work plan.

M4 (months 3-6). **Hackathon.**
A hackathon will be conducted for the RDA 17th Plenary in April 2021. The objective will be to combine publicly available COVID-19 related datasets and to present solutions that overcome the barriers encountered. The hackathon will be announced at the 16th Plenary in November 2020 and will be opened in February 2021. Participants will present their results at the 17th Plenary at which time judging will take place and winners announced. Winners will be offered co-authorship on a peer-reviewed publication. From November to January, we will seek sponsors for cash prizes to be awarded to 1st, 2nd, and 3rd place winners.

M5 (months 3-12). **Development of a draft standard for reporting epidemiology surveillance data.**

M6 (months 12-14). **Public review of the draft standard.**

M7 (months 12-16). **Development of guidelines for adoption of the standard**

M8 (month 15-17). **Finalization of the standard.**

M9 (month 1-18). **Dissemination and Communication**
WG activities and outcomes will be disseminated via the RDA website, preprint(s), submission to a peer-reviewed journal, RDA Plenaries, conference presentations, and social media.

**Simplified Gantt Chart**

| Months | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Deliverables |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| D1 draft |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| D2 draft |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| D2 final |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| D1 final |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Milestones | M1 | M2 | M3 | M4 |    |    |    |    |    |    |    |    |    |    |    |    |    |
| M5      |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| M6      |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| M7      |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| M8      |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| M9      |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Plenaries |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| P17     |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| P18     |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| P19     |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

**Work space**
The WG will use the following platforms for communication and development:
- [RDA website](https://www.rda.org)
- Google drive
  - Working documents will be managed on Gdrive to facilitate open collaboration, and to generally make things easier.
- GitHub
  - We will develop a public GitHub repository to host the hackathon material, models, source code, and the proposed common standard, and to raise and resolve issues.
- Zotero
  - We will build a Zotero Library to support this work, based on the RDA COVID-19 WG Zotero Public Library

**Tools**
We will use a variety of tools, for example:

- Visualization
  - Mindmapping
  - Infographics
- Gantt charting for project management
- Voting and consensus building tools

**Meetings**
- Meetings will be held **weekly**.
- An online platform (e.g., GoToMeeting, Zoom, WebEx, MS Teams, or Google Meet) will be used for meetings. Participants will be asked to activate their video to enhance communication effectiveness. The WG will meet at RDA Plenaries, the first such meeting being at the 16th Plenary.
- Agenda, minutes and rolling notes will be circulated via google doc.
- Discussions will be held at the RDA 16th, 17th, and 18th Plenaries, and at other conferences and workshops where possible.

**Consensus**
*A description of how the WG plans to develop consensus, address conflicts, stay on track and within scope, and move forward during operation, and*

Consensus will be achieved mainly through discussions in our regular weekly meetings, where conflicting viewpoints will be identified and openly discussed and debated by group members. If consensus cannot be reached in this manner, the final decision will be taken by the group co-chairs. By setting realistic deadlines and assessing progress on assigned tasks, the co-chairs will keep the WG on track and within scope.

**Community engagement**
*A description of the WG’s planned approach to broader community engagement and participation.*

To encourage broader community engagement and participation in the development of a standard, the WG case statement will be circulated to various public health organizations and epidemiological societies across the globe, and on social media (Linkedin and Twitter). A regular update on events/news related to Epidemiology common standards will be posted on RDA WG webpage to encourage involvement of specialists in the field.

**License**
WG outputs will be published under a [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.
5. **ADOPTION PLAN**

*A specific plan for adoption or implementation of the WG outcomes within the organizations and institutions represented by WG members, as well as plans for adoption more broadly within the community. Such adoption or implementation should start within the 12-18 month timeframe before the WG is complete.*

The WG members will be encouraged to implement the new standard and guidelines within their organizations. We will pursue adoption by a variety of stakeholders and research communities, particularly those involved in public health. The standard will be disseminated via RDA webinars, other scientific presentations and twitter handle. We will follow up with adoption stories.

**Publications**

We will publish preprints and also seek to publish the work in one or more peer-reviewed articles in journal articles in open access journals.

6. **INITIAL WG MEMBERSHIP**

*A specific list of initial members of the WG and a description of initial leadership of the WG.*

**Co-Chairs:** Claire Austin and Rajini Nagrani

**RDA Liaison:** Stefanie Kethers

**Membership roster**

Members/Interested:
- Soegianto Ali
- Anthony Juehne
- Nada El Jundi
- Fotis Georgatos
- Jitendra Jonnagaddala
- Miklós Krész
- Gary Mazzaferro
- Jiban K. Pal
- Carlos Luis Parra-Calderon
- Bonnet Pascal
- Fotis Psomopoulos
- Stefan Sauermann
- Henri Tonnang
- Marcos Roberto Tovani-Palone
- Anna Widyastuti
- Becca Wilson
- Eiko Yoneki

**Current initial membership**

The initial WG includes:
- Cross-domain expertise
  - biostatistics, clinical informatics, computer engineering, data science, epidemiology, global health, health informatics, health sciences, interoperability, IT architecture, mathematics, open science, pathology, predictive modeling, public health, research data management, software development, veterinary medicine
- Experience
- academia, editor of scientific journals, government, international WG leadership, program director, research, standards development.

- Regional representation
  - Africa (sub-saharan), Asia (maritime southeast), Asia (south), Australasia, Europe, North America, and South America.

- Income groups
  - Two lower-middle income, two upper-middle income, and 15 high-income countries.

Initial membership comprises a core group from the RDA-COVID19-Epidemiology WG, and new members who bring additional domain specific expertise. We aim to further strengthen the group to expand global participation (low income, lower-middle income, and upper-middle income countries), interdisciplinary experts, and stakeholder representation to address this pressing common epidemiology surveillance data challenge across the public health domain.

**Actively soliciting WG membership**

The initial membership does not currently include any potential adopters. We will be soliciting the active participation in the WG of representatives from key stakeholders, including the following:

- Official agencies, organizations, and funders having international reach
- Supernational organizations
- European Centers for Disease Control (eCDC)
- Global Early Warning System (GLEWS+)
- Global Health Security Agenda (GHSA)
- Global Influenza Surveillance and Response System (GISRS)
- Global Partnership for Sustainable Development Data (GPSDD)
- GloPID-R
- Indian Council of Medical Research (ICMR)
- Observational Health Data Sciences and Informatics (OHDSI)
- UN Office for Disaster Risk Reduction (UNDRR)
- United Nations Educational, Scientific and Cultural Organization (UNESCO)
- U.S. Centers for Disease Control (CDC)
- Wellcome Trust
- World Data System (WDS)
- World Health Organization (WHO)
- World Bank World Development Indicators (WDI)

**Data aggregators from academia**

- Johns Hopkins University (Killeen et al. 2020)
- University of California, Berkeley (Altieri et al. 2020)
- University of Oxford (Roser et al. 2020)

**Data aggregators from News Outlets**

- The Atlantic
- The Economist
- The Financial Times
- The New York Times

**Communications/graphic artist expertise**
7. REFERENCES


Greenfield J., Tonnang E.Z., Mazzaferro G., Austin, C.C.; and the RDA-COVID19-WG. (2020). Epi-TRACS: Rapid detection and whole system response for emerging pathogens such as SARS-CoV-2 virus and the COVID-19 disease that it causes. IN: COVID-19 Data sharing in epidemiology, version 0.06b. Research Data Alliance RDA-COVID19-Epidemiology WG. https://doi.org/10.15497/rda00049


https://www.who.int/teams/risk-communication/about-epi-win