
Data Visualization during an outbreak like COVID-19

Release v2020.05.13 14.33

Abdelkrim B.

May 13, 2020

Contents:

1	Introduction	1
2	REST APIs	2
2.1	Blogs	2
3	News	3
4	Articles and Blogs	4
5	Dashboards	5
6	Datasources	6
6.1	Lockdowns by country	9
6.2	Tests conducted by country	9
6.3	Flu Tracking	9
6.4	clinical trials	9
7	Widgets anyone can embed into a html page	10
7.1	COVID-19 on Bing	10
8	Data formats	11
8.1	List of locations in a TXT file	11
9	Data models	12
9.1	Data model	12
9.2	Multilingual applications	12
9.3	List of locations	13
10	Metadata	14
11	Projects	15
11.1	Comparison of COVID-19 case reporting from different sources	15
12	Software Architecture	16
13	Vocabulary	17
14	RDA COVID-19 Recommendations and guidelines	18
14.1	FAIR and timely	18
14.2	Metadata	18
14.3	Documentation	19
14.4	Use of Trustworthy Repositories	19
14.5	Ethics and Privacy	19

14.6	Legal	19
14.7	What are the blocking factors for sharing data?	19
14.8	Major difficulty	20
14.9	Glossary	20
14.10	Citations	20
15	Glossary for Data Visualization	21
16	Citations	22
17	To do list	23
18	Indices and tables	24
	Bibliography	25
	Index	26

CHAPTER 1

Introduction

Hundreds of initiatives have been initiated to display key indicators related to the spread of COVID-19; individuals, non-profit companies, for-profit companies and public administrations published dashboards all over the world.

The dashboards were built by software like Qlik or relied on private or open-source code.

Centers for Disease Control and Prevention became famous : *CDC* in the US, *ECDC* in Europe became overloaded by requests.

Data collected by public administrations were made publicly available and the first dashboards came up with its share of critics.

CHAPTER 2

REST APIs

Access data on COVID19 through an easy API for free. Build dashboards, mobile apps or integrate in to other applications.

<https://covid19api.com/> with 34,642,916 requests served by the API

<https://rapidapi.com/collection/coronavirus-covid-19>

2.1 Blogs

Developers Respond to COVID-19 <https://rapidapi.com/blog/developers-respond-to-covid-19>

CHAPTER 3

News

The Latest News On The API Economy https://www.programmableweb.com/category/coronavirus%2Bcovid-19/news?category=29999%2C30105&api_videos=0

CHAPTER 4

Articles and Blogs

Articles and Blog posts describing how to increase the quality of your dashboards

Open collaboration on COVID-19 <https://github.blog/2020-03-23-open-collaboration-on-covid-19> By Martin Woodward

Live tracker: How many coronavirus cases have been reported in each U.S. state? <https://www.politico.com/interactives/2020/coronavirus-testing-by-state-chart-of-new-cases/> By Beatrice Jin | 03/16/2020 8:10 PM EDT | Updated 04/22/2020 7:10 PM EDT

Closing the data divide: the need for open data, Apr 21, 2020 | Jennifer Yokoyama - Chief IP Counsel <https://blogs.microsoft.com/on-the-issues/2020/04/21/open-data-campaign-divide/>

Bing delivers new COVID-19 experiences including partnership with GoFundMe to help affected businesses https://blogs.bing.com/search/2020_04/Bing-delivers-new-COVID-19-experiences-including-partnership-with-GoFundMe-

17 (or so) responsible live visualizations about the coronavirus, for you to use <https://blog.datawrapper.de/coronaviruscharts>

Epidemic Modeling 101: Or why your CoVID19 exponential fits are wrong <https://github.com/DataForScience/Epidemiology101>

Charting new territory; How The Economist designs charts for Instagram <https://medium.economist.com/charting-new-territory-7f5afb293270>

CHAPTER 5

Dashboards

Table 1: Dashboards

description	dashboard url	source code	
WHO Coronavirus (COVID-19)	https://covid19.who.int	none	
WHO - Explore the Data	https://covid19.who.int/explorer	none	
Healthmap (made by schools and hospitals)	https://healthmap.org/covid-19	none	
COVID-19 India	https://www.covid19india.org	https://github.com/covid19india/covid19india-react	
COVID-19 Scenarios	https://covid19-scenarios.org	https://github.com/neherlab/covid19_scenarios	
List of dashboards WW	https://covid19dashboards.com	https://github.com/github/covid19-dashboard	
How many days each country's outbreak is behind or ahead of the United States	https://predictcovid.com	https://github.com/lachlanjc/covid19	
COVID-19 Italia - Monitoraggio situazione (Desktop app)	https://github.com/pcm-dpc/COVID-19	http://arcg.is/C1unv	
COVID-19 Italia - Monitoraggio situazione (Mobile app)	https://github.com/pcm-dpc/COVID-19	http://arcg.is/081a51	
Latest updates on COVID-19 in Tokyo	https://stopcovid19.metro.tokyo.lg.jp	https://github.com/tokyo-metropolitan-gov/covid19	
Official World Health Organization COVID-19 App	https://github.com/WorldHealthOrganization/app		
Real-time tracking of pathogen evolution	https://nextstrain.org	https://github.com/nextstrain	
16.000 viral genomic sequences of hCoV-19 shared with unprecedented speed via GISAID	https://www.gisaid.org	https://www.gisaid.org/epiflu-applications/next-hcov-19-app	
Dashboard of the COVID-19 Virus Outbreak	https://co.vid19.sg/singapore	none	

CHAPTER 6

Datasources

72h nonprofit online hackathon is to develop open-source prototypes, which contribute to solving the most pressing challenges in the current crisis.

<https://www.codevscovid19.org/>

Tens of thousands of volunteers build solutions for the Corona pandemic. To avoid reinventing the wheel, we created a central place to learn about existing projects and add new ideas.

<https://airtable.com/shrPm5L5I76Djdu9B/tbl6pY6HtSZvSE6rJ/viwbIjyehBIOKYYt1?blocks=bipjdZOhKwkQnH1tV>

Location for summaries and analysis of data related to n-CoV 2019, first reported in Wuhan, China

<https://github.com/beoutbreakprepared/nCoV2019>

The Institute for Health Metrics and Evaluation (IHME) is an independent population health research center at UW Medicine, part of the University of Washington, that provides rigorous and comparable measurement of the world's most important health problems and evaluates the strategies used to address them

<http://www.healthdata.org/>

COVID-19 Projections worldwide

<https://covid19.healthdata.org/united-states-of-america>

NIH National Institute of Health, Open-Access Data and Computational Resources to Address COVID-19:

<https://datascience.nih.gov/covid-19-open-access-resources>

The COVID Tracking Project collects and publishes the most complete testing data available for US states and territories

<https://covidtracking.com>

Bing COVID-19 data sources

<https://help.bing.microsoft.com/#apex/18/en-us/10024>

A repo for coronavirus related case count data from around the world. The repo will be regularly updated

<https://github.com/microsoft/Bing-COVID-19-Data>

We are building an open database of COVID-19 cases with chest X-ray or CT images

<https://github.com/ieee8023/covid-chestxray-dataset>

Data in time of COVID-19

<https://opendatawatch.com/what-is-being-said/data-in-the-time-of-covid-19/>

Hack for Wuhan

https://github.com/wuhan2020/Hackathon/blob/master/README_EN.md

#WirVsVirus Hackathon

<https://wirvsvirushackathon.org/>

Overview of topics and challenges: <https://airtable.com/shrPm5L5I76Djdu9B>

Overview of submissions: <https://wirvsvirushackathon.devpost.com/submissions>

Data Repository by Johns Hopkins

<https://github.com/CSSEGISandData/COVID-19>

Open Source Know-How

<https://www.healthcare-computing.de/know-how-zum-coronavirus-fuer-wissenschaftscommunity-a-909391/>

White House Dataset

<https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/>

COVID-19 Open Research Dataset (CORD-19) by Microsoft Research

Details: <https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-resources-and-their-application-to-covid-19-research/>

Access: <https://pages.semanticscholar.org/coronavirus-research>

COVID-19 Data

<https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data/resource/62eb477f-be00-462a-831a-594095f7306a>

COVID-19 Open Research Dataset Challenge

<https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks>

Crowdbreaks Data

<https://www.crowdbreaks.org/en/projects/covid>

COVID-19 Cases Switzerland

<https://github.com/daenuprobst/covid19-cases-switzerland>

Project of Daniel Probst: <https://www.corona-data.ch/>

COVID-19 case numbers communicated by official Swiss Canton's and FL's sources

https://github.com/openZH/covid_19/blob/master/README.md

Swiss Hospital Data

<https://github.com/schoolofdata-ch/swiss-hospital-data>

Swiss Federal Railways Data

Open Data: <https://opentransportdata.swiss/> and <https://data.sbb.ch/>

Open Journey Planner (OJP): <https://opentransportdata.swiss/de/cookbook/open-journey-planner-ojp/>

Open Data City of Zurich

Air Quality: <https://data.stadt-zuerich.ch/dataset/luftqualitaet-tages-aktuelle-messungen>

Traffic count data for motorized private transport (hourly values): https://data.stadt-zuerich.ch/dataset/sid_dav_verkehrszaehlung_miv_od2031

Parking guidance system: <https://data.stadt-zuerich.ch/dataset/parkleitsystem>

Zurich Tourism Open Data

<https://www.zuerich.com/open-data>

Reddit Resources

https://www.reddit.com/r/MachineLearning/comments/fks234/nd_resources_and_channels_to_help_with_covid19/

Nth Opinion

<https://github.com/nthopinion/covid19>

World Health Organization App <https://github.com/WorldHealthOrganization/app>

European CDC Data

<https://www.ecdc.europa.eu/en>

ACAPS Resources

<https://www.acaps.org/projects/covid19>

Real-time tracking of pathogen evolution

nextstrain.org

COVID Tracker

<https://github.com/sagarkarira/coronavirus-tracker-cli>

COVID-19 Hospital Impact Model for Epidemics

<https://github.com/CodeForPhilly/chime>

NVIDIA Parabricks Genomics Analysis Toolkit

<https://www.developer.nvidia.com/nvidia-parabricks>

Electricity Generation, Transportation and Consumption Data

<https://transparency.entsoe.eu/dashboard/show>

COVID-19 API Initiative

<https://www.xapix.io/covid-19-initiative>

Facebook API & SDK

Messenger: <https://developers.facebook.com/docs/messenger-platform/>

Instagram & Facebook Stories: <https://developers.facebook.com/docs/instagram-api/reference/user/stories/>

Proximity Tech to fight COVID-19 by Uepaa

<http://blog.p2pkit.io/proximity-tech-to-fight-covid-19/>

Scandit SDK

For barcode, text and ID document scanning capabilities

<https://www.scandit.com/developers/>

Swiss Radio and Television API

<https://developer.srgssr.ch/>

<https://developer.srgssr.ch/apis>

MongoDB's flexible data model

Go and try it for free at <https://cloud.mongodb.com> and if you require more credits, please fill in this form: <https://forms.gle/R4zLtWurWkNFMozq9>

IBM Cloud and Watson APIs

<https://ibm.biz/cloud-vs-covid19> Please ping IBM on slack to get a unique Feature Code

<https://developer.ibm.com/blogs/the-2020-call-for-code-global-challenge-takes-on-covid-19/>

Postman COVID-19 API Resource Center

<https://covid-19-apis.postman.com/>

Google Maps, Cloud and G Suite

<https://docs.google.com/document/d/17eODy-Jq7yPE8PiVXKWK7pUePc4cOOG0YJ-Rhc7sdyg/edit>

Maps: <https://developers.google.com/maps/covid19>

Fossil.com - Daily Archives of COVID-19 relevant pages

<https://www.fossil.com/covid-19-offline>

6.1 Lockdowns by country

COVID-19 Lockdown dates by country; A list of countries and the dates that each country went into lockdown.

<https://www.kaggle.com/jcyzag/covid19-lockdown-dates-by-country>

6.2 Tests conducted by country

Covid19 Tests Conducted by Country; Captures the number of tests conducted in any country/region

<https://www.kaggle.com/skylord/covid19-tests-conducted-by-country>

6.3 Flu Tracking

Belgium

<https://www.covid19stop.org>

Australia

<https://www.flutracking.net>

New Zealand

<https://www.flutracking.net>

6.4 clinical trials

Vivli A global clinical research data sharing platform. The Vivli team is dedicated to helping researchers share and access data from clinical trials to advance science

<https://vivli.org/>

Todo: continue à partir de data vizualization

Widgets anyone can embed into a html page

7.1 COVID-19 on Bing

Data is collected from multiple sources (CDC, WHO, ECDC, Wikipedia, 24/7 Wall St., BNO News) that update at different intervals. <https://bing.com/covid>

```
<div class="bingwidget" data-type="covid19" data-market="en-us" data-language="en-us" data-app="bingwidget"></div>

<script src="//www.bing.com/widget/bootstrap.answer.js" async=""></script>
```

How to format the data your are exposing: text, numeric, complex object using XML or JSON?

8.1 List of locations in a TXT file

Source : <https://github.com/microsoft/COVID-19-Widget/blob/master/AllLocations.txt>

Table 1: All Locations.txt

Location: Country/State
/
/United States
/United States/Alabama
/United States/Alaska
/Afghanistan
/Albania
/Algeria

How should you design your data model? Which level of details do you plan?

9.1 Data model

Table 1: Data models

Record name
Total world statistics
Cases by country
Cases by country by day
Data for all countries
Mask usage instructions
Infection histories by country
India-specific figures
North American figures
Data by ISO code
Testing statistics
Transportation infection data

9.2 Multilingual applications

Twenty percent of the population understand English, sharing a dashboard in English describing an outbreak is of little help for the vast majority of the population.

Note: Build a multilanguage application from day 1.

[WikiLangSpok20] indicates an approximate list of languages by the total number of speakers.

Table 2: List of languages by total number of speakers

Rank	Language	Speakers (millions)
1	English	1.268
2	Mandarin Chinese	1.120
3	Hindi	637.2
4	Spanish	537.9
5	French	276.6
6	Standard Arabic	274.0
7	Bengali	265.2
8	Russian	258.0
9	Portuguese	252.2
10	Indonesian	199.0
11	Urdu	170.6
12	German	131.6
13	Japanese	126.4
14	Swahili	98.5
15	Marathi	95.3

9.3 List of locations

List of locations consist of several information; e.g., country name, state, county, province, commune, longitude, latitude, post box

Bing lists countries and states names in a text file; US States are separated by a slash '/' <https://github.com/microsoft/COVID-19-Widget/blob/master/AllLocations.txt>

CHAPTER 10

Metadata

- Hospital
 - All beds needed is the total number of beds needed exclusively for COVID patients, and includes ICU beds needed for COVID patients. covid19.healthdata.org
 - Bed shortage
 - ICU beds needed is the total number of ICU beds needed exclusively for COVID patients. covid19.healthdata.org
 - ICU bed shortage
 - Invasive ventilators needed
- Uncertainty is the range of values that is likely to include the correct projected estimate for a given data category. Larger uncertainty intervals can result from limited data availability, small studies, and conflicting data, while smaller uncertainty intervals can result from extensive data availability, large studies, and data that are consistent across sources. The model presented in this tool has a 95% uncertainty interval and is represented by the shaded area(s) on each chart. covid19.healthdata.org
- correlation between metadata
 - All beds needed -> Bed shortage
 - ICU beds needed -> ICU bed shortage

11.1 Comparison of COVID-19 case reporting from different sources

Daily cumulative case numbers (starting Jan 22, 2020) reported by the Johns Hopkins University Center for Systems Science and Engineering (CSSE), WHO situation reports, and the Chinese Center for Disease Control and Prevention (Chinese CDC) for within (A) and outside (B) mainland China.

Source : An interactive web-based dashboard to track COVID-19 in real time: <https://www.thelancet.com/journals/laninf/article/PIIS1473-3099%2820%2930120-1/fulltext>

CHAPTER 12

Software Architecture

Gatsby is a free and open source framework based on React that helps developers build blazing fast websites and apps:

<https://www.gatsbyjs.org/>

urlwatch configuration to monitor state COVID-19 data:

<https://github.com/COVID19Tracking/covid-tracking>

urlwatch monitors webpages for you:

<https://github.com/thp/urlwatch>

The COVID Tracking Project collects and publishes the most complete testing data available for US states and territories.

<https://github.com/COVID19Tracking/website/tree/master/build>

Scan/Trim/Extra Pipeline for State Coronavirus Site

<https://github.com/COVID19Tracking/covid-data-pipeline>

The crawler and parsers are there for the 50 US states and DC. The focus is to collect official published COVID-19 statistics.

<https://github.com/coronavirusapi/crawl-and-parse>

<https://coronavirusapi.com/>

Provide embed code to integrate your dashboards easily

<https://github.com/microsoft/COVID-19-Widget>

CHAPTER 13

Vocabulary

Use **fatal** instead of death

Use **Lives Lost** instead of death

RDA COVID-19 Recommendations and guidelines

The objectives of the [RDA](#) COVID-19 Working Group (CWG) are:

to clearly define detailed guidelines on data sharing under the present COVID-19 circumstances to help stakeholders follow best practices to maximize the efficiency of their work, and to act as a blueprint for future emergencies; to develop guidelines for policymakers to maximise timely data sharing and appropriate responses in such health emergencies; to address the interests of researchers, policy makers, funders, publishers, and providers of data sharing infrastructures.

Source : RDA COVID-19 Recommendations and guidelines 1st release
 - open for comments <https://www.eoscsecretariat.eu/eosc-liaison-platform/post/rda-covid-19-recommendations-and-guidelines-1st-release-open-comments>

14.1 FAIR and timely

“FAIR principles [means] that data, software, models and other outputs should be Findable, Accessible, Interoperable and Reusable”

“A balance between achieving ‘perfectly’ FAIR outputs and timely sharing is necessary with the keygoal of immediate and open sharing as a driver.”

14.2 Metadata

“While rich metadata is desirable, even a minimum set of key fields/descriptors is valuable” [RDA20]

“The use of common metadata standards, as adopted by one’s relevant discipline, as well as vocabularies, are highly recommended”

“metadata should describe the data as well as the terms under which it can be accessed and reused.”

“Ideally, data and metadata should be exposed via machine readable endpoints (e.g. RDF, APIs)”

“Where there are restrictions on accessing or using datasets, metadata should be shared openly to enable discovery (e.g. CC0 or CC-BY licenses).”

14.3 Documentation

“Research outputs need to be documented, which includes documentation of methodologies used to define and construct data, data cleaning, data imputation, data provenance and so on”

“Software should provide documentation that describes at least the libraries, algorithms, assumptions and parameters used”

“Equally, research context, methods used to collect data, and quality-assurance steps taken are important”

“When sharing datasets, other relevant outputs (or documents) should also be made available, such as codebooks, lab journals, or informed consent form templates,”

14.4 Use of Trustworthy Repositories

“To facilitate data quality control, timely sharing and sustained access, data should be deposited in trustworthy data repositories (TDRs)”

“Whenever possible, these should be trustworthy data repositories (TDRs) that have been certified, subject to rigorous governance, and committed to longer-term preservation of their data holdings.”

“As the first choice, widely used disciplinary repositories are recommended for maximum accessibility and assessability of the data, followed by general or institutional repositories”

“Using existing open repositories is better than starting new resources.”

“By providing persistent identifiers, demanding preferred formats, rich metadata, etc., certified trustworthy repositories already guarantee a baseline FAIRness of and sustained access to the data, as well as citation.”

14.5 Ethics and Privacy

“Access to individual participant data and trial documents should be as open as possible and as closed as necessary, to protect participant privacy and reduce the risk of data misuse”

14.6 Legal

“Technical solutions that ensure anonymisation, encryption, privacy protection, and data de-identification will increase trust in data sharing.”

“Emergency data legislation activated during a pandemic needs to clearly outline data custodianship/ownership, publication rights and arrangements, consent models, and permissions around sharing data and exemptions.”

14.7 What are the blocking factors for sharing data?

- non-machine-readable data (e.g., PDF)
- heterogeneous measurement standards
- divergent metadata formats
- lack of version control
- fragmented datasets
- delays in releasing data
- non-standard definitions and reporting parameters

- lack of metadata
- unavailable or undocumented computer code
- frequently changing web addresses
- copyright and usage conditions
- translation requirements
- consents
- approvals
- legal restrictions
- lack or no integration of clinical, eHealth, surveillance, and research systems within and across jurisdictions or providers

14.8 Major difficulty

Lack of contextual data needed to study the evolution of disease in sub-populations. e.g.,

- healthy sub-populations that are vulnerable to serious long term effects following recovery that we don't know about yet because we don't have the data and because we are focusing on deaths
- age-specific vulnerabilities
- disadvantaged sub-populations with limited health care
- vulnerabilities evident in severe disease associated with comorbidities
- vulnerabilities due to environmental conditions
- vulnerabilities due to social and cultural norms
- following sequelae and immunity

14.9 Glossary

FAIR

FAIR principles Data, software, models and other outputs should be Findable, Accessible, Interoperable and Reusable

[WI16]

FAIRER

FAIRER principles Findable, Accessible, Interoperable, Reusable, Ethical, and Reproducible [RDA20]

OMICS Omics are defined as data from cell and molecular biology [RDA20]

RDA Reasearch Data alliance <https://www.rd-alliance.org>

SPIRIT Standard Protocol Items: Recommendations for Interventional Trials

TDR

Trustworthy Data Repositories Trustworthy Data Repositories (TDRs) are repositories that have been certified, subject to rigorous governance, and committed to longer-term preservation of their data holdings. [RDA20]

14.10 Citations

CHAPTER 15

Glossary for Data Visualization

CDC Center for Disease Control and Prevention

<https://www.cdc.gov>

ECDC European Center for Disease Control and Prevention

<https://ecdc.europa.eu>

CHAPTER 16

Citations

CHAPTER 17

To do list

Todo: continue à partir de data vizualization

(The [original entry](#) is located in `/home/docs/checkouts/readthedocs.org/user_builds/data-visualization/checkouts/latest/source/dataviz/datasources.rst`, line 251.)

CHAPTER 18

Indices and tables

- `genindex`
- `modindex`
- `search`

Bibliography

- [RDA20] <https://www.rd-alliance.org/system/files/RDA%20COVID-19%3B%20recommendations%20and%20guidelines%2C%201st%20release%2024%20April%202020.pdf>
- [WI16] The FAIR Guiding Principles for scientific data management and stewardship. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- [WikiLangSpok20] Wikipedia contributors. (2020, April 28). List of languages by total number of speakers. In Wikipedia, The Free Encyclopedia. Retrieved 17:52, April 30, 2020, from https://en.wikipedia.org/w/index.php?title=List_of_languages_by_total_number_of_speakers&oldid=953636952

C

CDC, [21](#)

E

ECDC, [21](#)

F

FAIR, [20](#)

FAIR principles, [20](#)

FAIRER, [20](#)

FAIRER principles, [20](#)

O

OMICS, [20](#)

R

RDA, [20](#)

S

SPIRIT, [20](#)

T

TDR, [20](#)

Trustworthy Data Repositories, [20](#)