unesco

Open data for AI

What now?

Published in 2023 by the United Nations Educational, Scientific and Cultural Organization,
7 place de Fontenoy, 75352 Paris 07 SP, France
© UNESCO, 2023
ISBN: 978-92-3-100600-5

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area of its authorities, or concerning the delimitation of its frontiers or boundaries.

The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

This publication is authored by Soenke Ziesche.
Graphic design, covers design, and typeset: Cromatik Ltd.
ePub: Cromatik Ltd.
Printed by UNESCO in Paris, France

SHORT SUMMARY

OPEN DATA, A MUST FOR AI

In order to fight the epidemic, timely, relevant, and high-quality open data were globally shared. This cooperative and efficient undertaking could serve as a blueprint for a better and more sustainable future for all.

Having raw data is the first essential step in processing and transforming them into actionable information. But the required data need to have certain characteristics, such as being accurate, timely and reliable. **Further critical aspects of data are that they are findable, accessible, interoperable and reusable (FAIR) by anyone for any purpose.**

The aim of these guidelines is to apprise Member States of the value of open data, and to outline how data are curated and opened. These guidelines specify concrete steps that Member States can take to open their data in three phases - Prepare, Open, Follow-up

These guidelines follow up on the **UNESCO Recommendation on the Ethics of Artificial Intelligence**, which, among other topics, includes a call for open data for AI. These guidelines will also play a crucial role in supporting the **UNESCO Recommendations on Open Science** by facilitating data sharing, enhancing reproducibility and transparency, promoting data interoperability and standards, supporting data preservation and long-term access.

120

Countries without Open Data policies

"Since wars begin in the minds of men and women it is in the minds of men and women that the defences of peace must be constructed"

Open data for AI

What now?

TABLE OF CONTENTS

PREFACE 5

ACKNOWLEDGEMENTS 6

EXECUTIVE SUMMARY 7

1. DATA IN THE FIGHT OF A GLOBAL PANDEMIC - A CASE STUDY 11

1.1 INTRODUCTION 12

1.2 AI CHALLENGES 13

1.3 DATA CHALLENGES 15

1.4 WAY FORWARD 16

1.5 SUMMARIZED LESSONS LEARNED TOWARDS OPEN DATA 17

2. VISION 21

2.1 INTRODUCTION 22

2.2 OPEN DATA 25

2.2.1 BACKGROUND 25

2.2.2 DEFINITION, PRINCIPLES, FIVE STAR PLAN AND LICENCES 27

2.2.3 CATEGORIZATIONS OF OPEN DATA 29

2.2.4 PROS AND CONS OF OPEN DATA 31

2.2.5 EXISTING GENERAL OPEN DATA INITIATIVES AND RESOURCES 32

2.2.5.1 SDG RELATED OPEN DATA INITIATIVES 34

2.2.5.2 GOVERNMENT OPEN DATA INITIATIVES 35

2.2.5.3 SCIENCE OPEN DATA INITIATIVES 35

2.2.5.4 PRIVATE OPEN DATA INITIATIVES 35

2.2.5.5 COVID-19 OPEN DATA INITIATIVES 36

2.2.6 INDICES AND BAROMETERS 38

2.3 AI AND OPEN DATA 39

2.3.1 BACKGROUND 39

2.3.2 SYNERGIES BETWEEN AI AND OPEN DATA TO TACKLE GLOBAL CHALLENGES 40

2.3.3 POTENTIAL INCREASED AI RISKS DUE TO OPEN DATA 42

FURTHER TOPICS 43

3. GUIDELINES 47

3.1 PREAMBLE 48

3.2 PREPARATION 49

3.3 OPEN THE DATA 51

3.4 FOLLOW UP FOR REUSE AND SUSTAINABILITY 53

EPILOGUE 56

4. REFERENCES 57

PREFACE

Processes, production cycles and associated research studies generate large amounts of data in diverse formats and sequences. These factors lead to complex issues in handling data and using data for generative AI systems. As we increasingly recognize the role of Artificial Intelligence (AI), availability and access to data are more crucial than ever.

UNESCO executes research in AI to provide insight and practical solutions to foster digital transformation and to build inclusive Knowledge Societies. We created these guidelines to promote a shared understanding of data openness, its benefits and challenges. The aim is to inform readers, particularly the UNESCO Member States, about the importance of open data for AI and supporting our Member States in taking action through a series of steps.

The guidelines outline the type and scope of a data ecosystem and reframes FAIR (Findable, Accessible, Interoperable and Reusable) data concepts for AI.

The guidelines then list twelve concrete steps that Member States can take to make their data open and available in a FAIR format. These instructions are broken down into three phases (i) preparation; (ii) data opening; and (iii) follow-up for reuse and sustainability, each including four steps.

I hope that the publication will act as a call to action, encouraging Member States not only to support openness of high-quality data, but also to adopt a data-openness culture to fuel AI applications.

Tawfik Jelassi

Assistant Director-General for Communication and Information

UNESCO

ACKNOWLEDGMENTS

UNESCO thanks the lead author Mr. Soenke Ziesche for taking up the task of drafting these guidelines.

UNESCO also sincerely thanks the core group that oversaw the preparation of this document, namely

Prof. John Shawe-Taylor,

Prof. Maria Fasli,

Mr. Davor Orlic,

Mr. Aleš Veršič

Mr. Bhanu Neupane

Peer reviewers for this publication included:

Ms. Marielza Oliveira, Director for Partnerships and Operational Programme Monitoring, Communication and Information Sector, UNESCO

Mr. Jaco du Toit, Chief of Universal Access to Information Section, Communication and Information Sector, UNESCO

Ms. Dorothy Gordon, Chair, IFAP, UNESCO

Ms. Elsa Estevez, UNESCO Chair on Knowledge Societies and Digital Governance at Universidad Nacional del Sur, and Full Professor at the National University of La Plata, Argentina

Ms. Iryna Kuchma, the OA programme manager at Stichting EIFL.net

UNESCO would also like to thank colleagues from the Communication and Information Sector, the Science Sector, the Culture Sector, the Education Sector, the Social and Human Sciences Sector, the Division for Gender Equality and the Africa Department who are part of UNESCO's Task Team on AI for their support in framing the document.

EXECUTIVE SUMMARY

While the COVID-19 pandemic has been a global crisis bringing the world almost to a standstill, it has also brought national and international efforts together, from governments and policy makers, to the academic and research community and the public to address its challenges. To tackle the pandemic from the onset not only timely, relevant and quality data, but also open data were essential, which could be deployed for scientific research and to inform interventions and policy. Large amounts of open data related to COVID-19 were shared and had significant impact, including fuelling Artificial Intelligence (AI) applications and demonstrating the potential for the use of data to address other global challenges. Indeed, this cooperative and efficient undertaking could serve as a blueprint to gather reliable data as well as the proposal set out in the United Nations Secretary General's Our Common Agenda that calls for an upgrade of the United Nations with a call of change including data and strategic foresight.

Having raw data is the first essential step in processing and transforming them into actionable information. But the required data need to have certain characteristics, such as being accurate, timely and reliable. Further critical aspects of data are that they are findable, accessible, interoperable and reusable (FAIR) by anyone for any purpose. According to the UNESCO Recommendation on Open Science (2021b), these data comprise "open research data that include, among others, digital and analogue data, both raw and processed, and the accompanying metadata, as well as numerical scores, textual records, images and sounds, protocols, analysis code and workflows that can be openly used, reused, retained and redistributed by anyone, subject to acknowledgement" (UNESCO, 2021b, p. 9). Beyond that open data can be freely used, modified, and shared by anyone for any purpose[1], it can help to build trust, enable reuse and innovation, and accountability.

A vast amount of data on environment, industry, agriculture health about the world is now being collected through automatic processes, including sensors. Such data may be readily available, but also are potentially too big for humans to handle or analyse effectively, nonetheless they could serve as input to AI systems. AI and data science techniques have demonstrated great capacity to analyse large amounts of data, as currently illustrated by generative AI systems, and help uncover formerly unknown hidden patterns to deliver actionable information in real-time. However, many contemporary AI systems run on proprietary datasets, but data that fulfil the criteria of open data would benefit AI systems further and mitigate potential hazards of the systems such as lacking fairness, accountability, and transparency.

[1] https://opendefinition.org/

The aim of these guidelines is to apprise Member States of the value of open data, and to outline how data are curated and opened. Member States are encouraged not only to support openness of high-quality data, but also to embrace the use of AI technologies and facilitate capacity building, training and education in this regard, including inclusive open data as well as AI literacy.

The report has been produced through an extensive literature review and consultations with stakeholders, followed by a peer review process. It outlines concrete steps that can assist Member States in opening up their data, divided into three phases: (i) preparation; (ii) opening of the data; and (iii) follow up for reuse and sustainability; with each phase consisting of four steps. The preparation phase guides Member States in preparing for opening their data, and includes the following suggested steps: drafting an open data policy, gathering and collecting high quality data, developing open data capacities and making the data AI-ready. The opening of the data phase consists of the following steps: selecting datasets to be opened, opening the datasets legally, opening the datasets technically, and creating an open-data-driven culture. The follow-up for reuse and sustainability phase consists of the following steps: supporting citizen engagement, supporting international engagement, supporting beneficial AI engagement, and maintaining high quality data.

These guidelines follow up on the UNESCO Recommendation on the Ethics of Artificial Intelligence, which, among other topics, includes a call for open data (UNESCO, 2021d). If UNESCO Member States follow these guidelines and open their data in a sustainable manner, and create capacities as well as an open-data-driven culture, initiatives, while Member States can also use the same principles to tackle other global, regional and national challenges. Although these guidelines provide concrete steps that can aid Member States in their efforts to create, open up, make available and use data, the use of data and AI require addressing challenges and issues that arise in a range of areas including privacy, ethics and capacity building. UNESCO will provide further support to Member States to implement these guidelines, and specialist reports covering specific areas.

©ImageFlow/Shutterstock.com

1. DATA IN THE FIGHT OF A GLOBAL PANDEMIC - A CASE STUDY

1.1 INTRODUCTION

The COVID-19 pandemic can be described as "the most consequential global health crisis since the era of the influenza pandemic of 1918" (Cascella et al., 2022). By November 2021, there were over 251 million confirmed cases globally. Five million people had succumbed to COVID-19[2] and a significant number suffer from persistent symptoms (Raveendran et al., 2021). Almost the entire population of the world has been affected by, partly protracted, interventions such as lockdowns, which have also had psychological effects (Odriozola-González et al., 2020; Panchal et al., 2020). Moreover, the pandemic has considerably impacted education systems (UNESCO, 2021a) as well as the economy.[3]

While the WHO declared a Public Health Emergency of International Concern related to COVID-19 on 30 January 2020 and a pandemic on 11 March 2020, scientists had issued a letter and called for open sharing of SARS-CoV-2 sequence data already on 29 January 2020.[4] On 16 March 2020 data practitioners called for data infrastructure and a data ecosystem suitable to tackle pandemics.[5] On 30 March 2020, the UNESCO Director-General, Ms Audrey Azoulay, called on governments and advocated for scientific cooperation and the integration of Open Science in their research programmes.[6] In April 2020, the Open COVID Pledge was launched, which calls on organizations around the world to make their patents and copyrights freely available, which received many signatories, including from major corporations.[7]

The COVID-19 pandemic also served as a background for the development of an international standard-setting instrument on Open Science in the form of a UNESCO Recommendation (UNESCO, 2021b). The Recommendation urges Governments to make efforts to foster capabilities towards open science and open research data, both raw and processed, and the accompanying metadata as well as numerical scores, textual records, images and sounds, protocols, analysis code and workflows.

In order to tackle the COVID-19 pandemic from the onset timely, relevant and quality data were essential. This case study illustrates, in addition to the mentioned features also openness of data is critical. Secondly, indeed large amounts of open data related to COVID-19 were shared and had a significant impact. A distinction can be made as to whether the available data are harnessed by humans or to fuel AI applications. Below some AI applications, challenges as well as lessons for the overall open data movement are outlined.

[2] https://covid19.who.int/

[3] https://www.statista.com/topics/6139/covid-19-impact-on-the-global-economy/#dossierKeyfigures

[4] https://www.covid19dataportal.org/support-data-sharing-covid19

[5] https://thegovlab.org/static/files/publications/ACallForActionCOVID19.pdf

[6] https://en.unesco.org/news/unesco-mobilizes-122-countries-promote-open-science-and-reinforced-cooperation-face-covid-19

[7] https://opencovidpledge.org

©Billion Photos/Shutterstock.com

1.2 AI CHALLENGES

AI machine learning techniques[8] were developed to support tackling the pandemic in various aspects. a data-driven AI model has been developed for the city of Valencia in Spain by a team of researchers, which won the 500k XPRIZE Pandemic Response Challenge (Lozano et al., 2021). The computational epidemiological model is based on open data and aims to predict COVID-19 infection rates as

[8] Machine learning refers to the capabilities of a computer to adapt to new circumstances and to detect and extrapolate patterns (Russell and Norvig, 2015).

well as prescribes non-pharmaceutical intervention plans. Further AI systems have been applied to available COVID-19 data to support areas such as diagnosis, prognostication, containment and monitoring, drug and vaccine development and treatments, as well as forecasting.[9] The WHO Hub for Pandemic and Epidemic Intelligence intends to use AI to expand forecasting of and early warning for future epidemics and pandemics.[10]

AI systems were instrumental in some areas, such as vaccine development.[11] However, despite a large number of diagnostic and predictive tools most of them were flawed for various reasons.[12] Most of these issues were related to the data, which were of poor quality, not standardized, mislabelled and frequently from unknown sources.[13] Moreover, the data were often not inclusive, e.g., by not representing minorities adequately, thus biased. All this combined let models, which were trained on these data, fail to produce accurate results. Another issue was that frequently neither data nor training models were shared since academic researchers have commonly few career incentives to do so.

©Jirsak/Shutterstock.com

[9] See for overviews: Harrus & Wyndham (2021), Hussain et al. (2020) and Khemasuwan & Colt (2021).

[10] https://pandemichub.who.int/

[11] See e.g.: Ong et al. (2020).

[12] See for overviews: Heaven (2021), von Borzyskowski et al. (2021), Wynants et al. (2020) and Roberts et al. (2021).

[13] See for overviews: OECD (2020) and https://opendatawatch.com/whats-being-said-resource/data-in-the-time-of-covid-19/

1.3 DATA CHALLENGES

There were data gaps, especially in developing countries and among certain at-risk and vulnerable populations (Milan & Treré, 2020),[14] and also issues in regard to lack of data disaggregation, especially by sex.[15] These challenges are, among other problems, related to the funding gaps that over 60 percent of low-income and lower-middle-income countries have been facing in financing their COVID-19 data and statistics.[16]

Overall, the data often do not fulfil the so-called FAIR principles, and are not findable, accessible, interoperable and reusable, while the acronym also has been interpreted as "Federated, AI–Ready" to illustrate the importance that the data can be used by AI systems.[17] There is mostly no clear distinction if the data are intended to feed AI processes designed to address the pandemic or not. In other words: Are open data and open data for AI interchangeable? If so, what are the consequences?

©Miha Creative/Shutterstock.com

[14] Also: https://www.npr.org/sections/goatsandsoda/2021/03/15/977455005/covid-19-data-is-missing-a-lot-of-people-and-raising-a-lot-of-questions?t=1622558115396.

[15] https://data2x.org/tracking-the-gender-impact-of-covid-19/

[16] https://paris21.org/news-center/news/press2020-under-covid-19-worrying-stagnation-funding-despite-growing-data-demand

[17] https://www.go-fair.org/implementation-networks/overview/vodan/

Another challenge is the balance between open COVID-19 data and the right to privacy,[18] as there have been instances of leaked data of COVID-19 patients.[19]

The COVID-19 pandemic has also been accompanied by disinformation; a phenomenon, which has been dubbed by UNESCO (2020) as "disinfodemic". One aspect of this is that COVID-19 sceptics created data visualizations allegedly showing that Governments' pandemic responses were disproportionate and that the facts and severity related to the pandemic were either under- or over-reported. This disinformation was often spread on social media (Lee et al., 2021).

1.4 WAY FORWARD

On the one hand, the COVID-19 pandemic has brought the world together to address this global challenge. There were breakthroughs owing to timely open data sharing, collaboration, and forecasting, enabling diagnosis tools as well as vaccines developed in record times. On the other hand, the pandemic aggravated polarization not only related to open data sharing,[20] but also related to fair vaccine distribution as well as vaccination hesitancy; issues, which go beyond the scope of this report. Overall, the positive aspects outlined above in terms of cooperation and efficiency should be taken forward and should serve as a role model to gather reliable data.

However, it also has to be noted that most of the open data initiatives were ad hoc and not well coordinated, since the world was unprepared for a pandemic. Therefore, as one lesson learned, regulatory frameworks and data governance models should be developed, supported by sufficient infrastructure, human resources and institutional capabilities to address the challenges related to open data outlined above, in order to be better prepared for pandemics and other global challenges (OECD, 2020). Moreover, the relationship between open data and AI needs to be further specified, including what features of open data are required so that they are "AI-Ready".

As will be further described below, this case study is useful since many opportunities and challenges briefly illustrated here apply also to open data in general.

[18] https://en.unesco.org/covid19/communicationinformationresponse/opensolutions

[19] https://www.hrw.org/news/2020/12/15/personal-data-thousands-covid-19-patients-leaked-moscow

[20] See e.g.: Waltman et al. (2021) for those in favour of data sharing and those emphasising the importance of patenting.

©Photo Smoothies/Shutterstock.com

1.5 SUMMARIZED LESSONS LEARNED TOWARDS OPEN DATA

❑ A data management, collaboration and sharing policy should be in place for research as well as for Government institutions holding or processing health-related data, while ensuring data privacy through anonymization. This should include especially collaboration between AI researchers and clinicians.

❑ Government officials handling data, which are or may become pertinent for pandemics, may require training to recognize the importance of such data as well as the imperative to share them.

❑ As much high-quality data as possible should be gathered and collected. Features of "high-quality" data are accurate, not outdated and comprehensive; thus, the data have to be from a variety of credible sources, which, however, also must be of ethical, i.e. must not include datasets with biases and harmful content and have to be collected with consent only and not in privacy-invasive ways. Data about people should be disaggregated where relevant, ideally by income, sex, age, race, ethnicity, migratory status, disability and geographic location (Inter-Agency and Expert Group on Sustainable Development Goal Indicators, 2017). Furthermore, pandemics are usually fast evolving processes; thus, continuous updating of data is essential.

- □ Such data features are especially mandatory to improve in the future inadequate AI diagnostic and predictive tools. Also, data may be included, which appear negligible, since it is a strength of machine learning AI systems to discover pertinent patterns and correlations within data, which humans are not able to recognize.

- □ Efforts are required to convert the relevant data in a machine-readable format, which involves curating of the collected data, i.e., cleaning and labelling.

- □ A wide range of data related to pandemics should be opened, adhering to the FAIR principles, which are findable, accessible, interoperable and reusable, while the acronym also has been interpreted as "Federated, AI–Ready"

- □ The target audience of pandemic-related open data includes research and academia, decision makers in Governments, the private sector for the development of relevant products, but also the public, all of whom should be informed about the available data. The public plays an important role in the open data movement, and was involved in the COVID-19 pandemic through initiatives under the umbrella term "citizen science".[21]

- □ Pandemic-related open data initiatives should be institutionalized instead of forming ad hoc, and thus should be put in place for the preparedness for further pandemics. Such initiatives should also be inclusive, bringing together different types of data producers and users.

- □ Also, the beneficial use of pandemic-related data for AI machine learning techniques should be regulated with the aim to prevent the misuse for the development of engineered pandemics, i.e., bioweapons, with the help of AI systems.

[21] See e.g.: https://citizenscience.org/covid-19/

©Wright Studio/Shutterstock.com

2. VISION

2.1 INTRODUCTION

The world is facing several interconnected crises. In addition to the on-going COVID-19 pandemic and devastating disasters, environmental degradation and loss of biodiversity, there are various other natural and anthropogenic hazards such as technological, including nuclear, accidents, as well as space-based threats, which create compounding, cascading or even systemic disaster risks. At the same time inequality, poverty and hunger prevail in the world, and are being exacerbated by the above crises and hazards, particularly the climate emergency.

To ensure sustainable development, it is important to continuously survey the world as well as measure and verify progress. In order to continuously survey the world and to measure and verify progress, data about the world is needed.

This has been illustrated by the case study above: Progress in tackling the COVID-19 pandemic has been significantly facilitated due to the availability of data. The

©metamorworks/Shutterstock.com

data have supported the interpretation of scenarios, the development and subsequent adjustment of policies as well as effective non-pharmaceutical, medical and vaccine interventions. Moreover, the data have enabled the scrutiny of Government policies leading to improved transparency and accountability.

If relevant data were available, similar effects can be anticipated, thus ultimately for tackling the global crises. Societies are at a pivotal point in time when the world has come together to address one crisis, the COVID-19 pandemic, and must overcome various dimensions of polarization in order to seize another historical opportunity by transferring and extending the good practices applied during the pandemic, especially those related to open data, towards other challenges.

The availability of raw data is only the first step. A second step involves cleaning the data and separating biased, unconsented and harmful data, which cannot be used for ethical reasons. During a third step the data are transformed into actionable information. Furthermore, as outlined in the case study above, the data could also serve as input to AI systems, which may identify unbeknown patterns.

Owing to significant developments in the field of sensors and the so-called Internet of Things (International Telecommunication Union & CISCO Systems, 2016) more and more metrics about the world, processes in nature as well as human activities, are being measured, which creates big data; too big for humans to handle or analyse without the help of automatic systems, including data science and AI systems.

At the same time many features of the world are not measured because either sensors for that particular feature have not been developed or the sensors exist, but are not deployed everywhere or there are other issues such as data privacy. This is a challenge if no one is to be left behind. In addition, much of the data collected is not accessible to everybody, for a range of reasons.[22]

Imagine a scenario where data for all 231 SDG indicators, which monitor the achievement of the 169 targets of the 17 SDGs, are available in a transparent manner and in real-time for everybody to check anytime through an online platform.[23] This would be a great example for availability of relevant data.

In order to tackle global problems, it is clear that the required data and their sources need to have certain characteristics such as being accurate, timely, reliable, comprehensive and inclusive. Yet, these characteristics on their own may not suffice to address the challenges. Further critical aspects of data are that they are findable, accessible, interoperable and reusable by anyone for any purpose. Such data are termed FAIR data, while open data require the characteristics to "be freely used, modified, and shared by anyone for any purpose".[24]

Open data is the main focus of this report since open data are considered as a prerequisite for informed plans, decisions and interventions. Therefore, the report asserts that the Member States ought to share data and insights, providing for transparency and accountability as well the opportunities for anybody to make use of the data.

The key objective of these guidelines is to foster universal access to information and knowledge through Open Solutions for inclusive digital transformation and AI development. At the same time there are valid AI safety concerns, which must not be neglected, but are because of their complexity beyond the scope of these guidelines (e.g., Bostrom, 2014; Yampolskiy, 2018). Moreover, the report aims to raise awareness of the technologies and tools that enable informed planning and decisions towards different interventions as well as monitoring and independent verification of progress.

[22] See also the "System-wide Road Map for Innovating UN Data and Statistics", developed by the Committee of the Chief Statisticians of the United Nations System through the High-level Committee on Programmes: https://unstats.un.org/unsd/unsystem/documents/Roadmap-Innovating%20UN%20Data%20and%20Statistic.pdf

[23] While platforms in this regard have been developed (https://unstats-undesa.opendata.arcgis.com/, https://sdg.tracking-progress.org/, https://unstats.un.org/sdgs/dataportal), they are limited by significant data gaps and time lags since countries lack the capacities to collect proxies in real time (e.g., Sachs et al., 2021).

[24] https://opendefinition.org/

It is argued here that the main benefit of open data, out of many, is that it both exemplifies and builds trust and solidarity. For example, tackling climate change requires a collaborative effort of many stakeholders, including the citizens of the world. The precondition for this endeavour is that all stakeholders not only have access to, but also trust climate change related information, which is compromised

by the menace of fake news. Yet, if Governments do not demonstrate full transparency about climate-related information in their countries,[25] neither will trust prevail nor will buy-in and ownership be created. Hence, the potential of data will not be harnessed to its full extent by all stakeholders; which is an imperative, given the enormous challenge and impact of climate change (e.g., Ratcliffe & Tuzeneu, 2019; Bodor et al., 2020).

The vision of these guidelines is to assist UNESCO Member States not only in measuring those aspects of the world, but also in making openly available the data that can contribute to solutions openly available (while, at the same time, respecting the human right to privacy) so that they can be processed and analyzed further for societal good by anybody, including through the use of AI systems.

The methodology used for these guidelines consisted of a thorough literature review as well as consultations with stakeholders, which was followed by a peer review process.

The remainder of these guidelines is structured as follows: in the next section, open data are introduced in more detail. This is followed by a section on AI and why open data are particularly relevant for AI. Subsequently, related topics, which cannot be addressed in this report, but are nonetheless critical, are outlined. The second part of this report presents critical steps for open data, which serve as a call to action.

2.2 OPEN DATA

2.2.1 Background

Technological developments in recent decades, subsequently also dubbed as the “Fourth Industrial Revolution”, have led to an unprecedented increase in the volume and complexity of data being generated, as well as being collected in large datasets (also named big data). It has been estimated that in 2020 the digital universe contained 44 zettabytes[26] of data. Yet, against the backdrop of an ever-

[25] An additional challenge is the so-called “spillover effect” if actions of one country have positive or negative effects on other countries’ abilities to achieve the SDGs, especially those related to climate. See https://www.unsdsn.org/spillover-effects and https://dashboards.sdgindex.org/map/spillovers.

[26] One zettabyte = 1021. See https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

increasing deluge of data, there are several challenges related to the availability of reliable and timely data for development, such as:

- ❑ Data have not been collected.
- ❑ Data have been collected, but are not shared.
- ❑ Data have been collected, but are siloed unbeknown to others.
- ❑ Data have been collected, but are outdated, unreliable or inaccurate.
- ❑ Data have been collected, but are not in interoperable formats, including, for example, in hardcopy format.
- ❑ Data have been collected, but are not labelled with an appropriate licence.

The immense amount of data offers opportunities as such data, for example, can be fed into AI systems for analysis and use. But even data in large amounts may not provide for an adequate mapping of the world as it would be desirable, since many relevant data are not collected, and many data which have been collected are not accessible.

It has become evident that data must have certain properties that makes them useful to address the global challenges outlined in the introduction, such as availability, findability, accessibility, interoperability, modifiability and reusability, which will be further described below. Therefore, stakeholders have been called upon to provide open data, especially Governments, but also the scientific community and the private sector. In 2012 the inventor of the World Wide Web Berners-Lee requested in an article "Raw data, now!" and outlined opportunities that additional data would materialize, such as "innovation, transparency, accountability, better governance and economic growth" (Berners-Lee, 2012). Also, former UN Secretary-General Ban Ki-moon acknowledged the relevance of data and commissioned a report called "A world that counts", which also stressed the value of open data and recommended partnerships between international organizations, governments, private companies and civil society organizations for data sharing and monitoring (Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014).

Currently, the Third Wave of open data is emerging, characterized by a focus on impactful reuse of data[27] as it also stressed in the draft UNESCO Recommendation on Open Science (UNESCO, 2021b). The currently proposed EU Data Governance Act supports the enhanced reuse of public sector data as well.[28] Further good practices are, for example, the Big Data for Migration Alliance, which organized problem-led virtual workshops for West Africa where stakeholders in government,

[27] The First Wave focused on regulation and legislation, while the Second Wave advocated to open data by default, but without specific purpose. See https://opendatapolicylab.org/images/odpl/third-wave-of-opendata.pdf

[28] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0767&from=EN

international organizations, research, civil society and the public sector participated and identified collaborative approaches to data innovation (Big Data for Migration Alliance, 2021), or the EU INSPIRE Directive, which led to unified data models and specifications for 34 spatial data themes.[29]

2.2.2 Definition, principles, five-star plan and licences

While the Recommendation on Open Science did not define open data, but noted what it contains and clarified the terms of its usage (UNESCO, 2021b), the definition by the Open Knowledge Foundation for open data is more precise and states: "Open data and content can be freely used, modified, and shared by anyone for any purpose".[30] As the focus of this guideline is on Open Government Data (OGD), which is "a philosophy – and increasingly a set of policies – that promotes transparency, accountability and value creation by making government data available to all",[31] it is worthwhile to posit that Governments do not only collect data and develop norms for data sharing, but also serve as the data providers.

Location
Item
Pieces
SG01 0005 05
SG02 0089 07
SG03 0195 09
SG04 0250 10
SG05 995 12
ORDER PICK TIME 00 : 02 : 16
ORDER PICK TIME 00 : 30 : 00
ORDER PICK TIME 04 : 55 : 00

©Blue Planet Studio/Shutterstock.com

[29] https://inspire.ec.europa.eu/inspire-directive/2

[30] https://opendefinition.org/

[31] https://www.oecd.org/gov/digital-government/open-government-data.htm

As an important supplement to this definition, the six principles by the Open Data Charter represent "a globally-agreed set of aspirational norms for how to publish data": 1) open by default, 2) timely and comprehensive, 3) accessible and usable, 4) comparable and interoperable, 5) for improved governance and citizen engagement and 6) for inclusive development and innovation.[32]

A further important set of principles referred to as FAIR has been mentioned already, which stands for findable, accessible, interoperable and reusable data. Guidance to implement the FAIR principles has been outlined in a so-called FAIRification Framework:[33]

- For findability by humans as well as by machines it is essential to assign globally unique and persistent identifiers to the data, to develop rich metadata for them and to ensure that they are in a searchable resource.
- For accessibility it is critical that the data can be retrieved by their identifier through a standardized communications protocol, which is open, free and universally implementable and potentially includes an authentication and authorization procedure.
- For interoperability it is important that the data are represented in a formal, accessible, shared and broadly applicable language and are integrated with other data through qualified references.
- For reusability it is relevant that the data are well described by metadata and meet community standards and that a clear and accessible data usage licence as well as information about the provenance of the data, their collection method and their maintenance are provided.

The set of FAIR principles has been complemented by the CARE Principles for Indigenous Data Governance, which stand for collective benefit, authority to control, responsibility and ethics.[34] The usage of FAIR and CARE principles form the cornerstone of all data governance actions that are aimed at setting standards for data sharing and ethical usage.

When it comes to opening the data technically, Berners-Lee developed a five-star open data plan:[35]

- One star: Make data available online in any format;
- Two stars: Make data available in structured format (e.g., MS Excel instead of a scanned table);

[32] https://opendatacharter.net/principles/

[33] https://www.go-fair.org/fair-principles/

[34] https://www.gida-global.org/care

[35] https://5stardata.info/en/ and https://www.w3.org/DesignIssues/LinkedData.html

©Monster Ztudio/Shutterstock.com

- ❑ Three stars: Make data available in a non-proprietary open format (e.g., CSV instead of MS Excel);
- ❑ Four stars: Use URIs to denote things,[36] so that other users can point at the data;
- ❑ Five stars: Link the data to other data to provide context.

There is a distinction between technically open and legally open data. On the legal side, it has to be specified how the data can be (re)used, and if attribution is required. A range of open data licences exists, of which Creative Commons licences are a prominent example.[37] Regarding the important question of if and to what extent open data can be used to train AI systems, there is an ongoing debate.[38]

2.2.3 Categorizations of open data

In addition to the five-star plan above, open data can be also categorized according to other criteria, which are illustrated below in the context of open COVID-19-related data:

- ❑ Data types: Predominant are numeric (geotagged) data, but also other data types are relevant for COVID-19 research, thus desirable to be shared, such as images (CT scans and X-ray) and sounds (breath and cough) (Shuja et al., 2021).[39]

[36] URI stands for Universal Resource Identifier, which identifies universally a physical or logical resource in the world. Examples for resources are real-world objects, including people, locations and concepts, or information resources, including web pages, documents and books.

[37] https://creativecommons.org/about/

[38] https://creativecommons.org/2021/03/04/should-cc-licensed-content-be-used-to-train-ai-it-depends/

[39] Further initiatives offered open access to scholarly articles, which is not the focus of this case study. See for example: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov and https://www.semanticscholar.org/cord19

- ❑ Data sources: The main sources of open COVID-19-related data are open government data and data from academia and research. While this was hardly the case for the COVID-19 pandemic, it is desirable for the future that also the private sector opens up pertinent data.
- ❑ Data use: The intended use of the data depends on the context and the specifications of a particular project, which can be manifold. For the COVID-19 pandemic data were required, which support the following two utilizations: 1) to understand how the coronavirus spreads, thus, to inform public health measures towards containment, and 2) to fight the virus, thus, to enable more efficient research towards a vaccine.
- ❑ Data products: Different formats can determine how the data are shared, such as data repositories, data dashboards, which allow interactions, or data visualizations.
- ❑ Target audiences: It refers to the actors who are expected to use the data, which have been analyzed by OECD and GovLab (2021) as follows: General public (71%), civil society (68%), government (41%), research/ academia (38%) and private sector (22%).

©LookerStudio/Shutterstock.com

2.2.4 Pros and cons of open data

There are arguments for and against open data. For example, one argument in favour[40] of open data, is that there cannot be a copyright on factual data anyway and that everyone should have the right to access them. Furthermore, it should be a feature of a democracy that activities of Governments are transparent through open data, which also helps to build trust in Governments. In addition, the opportunity to reuse, rearrange and combine them and to potentially gain new (scientific) insights from them enables citizens' engagement as well as creates new innovative services and products, thus contributing to deliver social, economic and environmental value.[41] Not only citizens may produce new knowledge from open data, but also AI systems, which are, as argued here, potential contributors to tackling the crises of the world.

©Mangostar/Shutterstock.com

Opponents of making data openly available argue, for example, that the data may violate the privacy of concerned individuals whose right it is to control what actively- or passively-collected data about them are disclosed. In certain cases, it may also violate copyrights and/or intellectual property rights, if collected and used without consent, attribution or compensation. Moreover, in many cases the collection, cleaning and dissemination of data is both labour and cost-intensive, which deserves financial compensation and which is omitted if these data are openly shared. In addition, datasets may present data bias arising at various stages, including during human reporting, data selection, data labelling and classification, all the way to interpretation of model results. There is also the concern that data may be misused with malicious intent, especially given that many new technolo-

[40] See e.g. https://okfn.org/opendata/why-open-data/, https://opendatahandbook.org/guide/en/why-open-data/

[41] See e.g. http://opendatatoolkit.worldbank.org/en/essentials.html#uses

gies, including AI, could be of dual use. In this regard, Bostrom (2011, p.3) defined a “data hazard” as follows: “Specific data, such as the genetic sequence of a lethal pathogen or a blueprint for making a thermonuclear weapon, if disseminated, create risk.”

2.2.5 Existing general open data initiatives and resources

In the last few years, there have been several open data initiatives and resources created, which will be introduced briefly through a timeline and by topic:

In 2012, Berners-Lee and Nigel Shadbolt founded the Open Data Institute towards an open and trustworthy data ecosystem.[42]

In 2013, the United Nations Department for Economic and Social Affairs published “Guidelines on Open Government Data for Citizen Engagement”.[43]

Also in 2013, the G8 leaders endorsed the G8 Open Data Charter advocating for data to support transparency, innovation and accountability. In 2015 a refined and more inclusive version of the now called International Open Data Charter was launched at the margins of the UN General Assembly (Open Data Charter, 2015).

Furthermore in 2013, the Research Data Alliance was launched, which is a community-driven initiative by the European Commission, the United States Government’s National Science Foundation and National Institute of Standards and Technology and the Australian Government’s Department of Innovation.[44]

In 2014, the global network Open Data for Development (OD4D) was founded to promote open data ecosystems in developing countries to support social change.[45]

In 2015, the Anti-corruption Working Group (ACWG) of the G20 published anti-corruption open data principles linked to the role of open data as a tool for transparency, accountability and access to information.[46]

In 2016, an article was published in Scientific Data advocating for findable, accessible, interoperable and reusable data (Wilkinson et al., 2016). The G20 leaders endorsed these so-called FAIR principles, which were described in detail above, at their summit in the same year.

[42] https://theodi.org/

[43] https://publicadministration.un.org/en/ogd

[44] https://www.rd-alliance.org/about-rda

[45] https://www.od4d.net/

[46] http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf

©ronstik/Shutterstock.com

In 2018, the Statistics Commission of the UN Department of Economic and Social Affairs (DESA) established a "Working Group on Open Data" focusing on principles, guidance and support for the implementation of open data in Member States.[47]

In 2020, the UN Secretary-General presented a "Roadmap for Digital Cooperation" to address challenges of new technologies towards a safer, more equitable digital world (UN General Assembly, 2020).

Also in 2020, the UN System Chief Executives Board for Coordination (2020) endorsed the "System-wide Road Map for Innovating UN Data and Statistics", which has the vision "to contribute to a better world through timely, trusted data and statistics for everyone".

In 2021, the 193 Member States of UNESCO adopted the 'Recommendation on the Ethics of Artificial Intelligence', which was not only a milestone because it was the first global standard-setting instrument on the ethics of AI, but the document also encouraged the Member States to promote open data (UNESCO, 2021d).

[47] https://unstats.un.org/open-data/

➢ 2.2.5.1 SDG related open data initiatives

The global indicator framework for the SDGs comprises 17 goals, 169 associated targets and 231 indicators, which rely heavily for their implementation and for the monitoring on data. A challenge is that many of the required data are currently not regularly collected by countries, which applies to 95 indicators as of February 2022.[48] In addition, Sachs et al. (2021) pointed out significant data gaps, especially for SDGs 4 (Quality Education), 5 (Gender Equality), 12 (Responsible Consumption and Production), 13 (Climate Action) and SDG (Life Below Water) and urged further investments to strengthen statistical capacities in concerned countries. Besides features such as accessible, accurate, timely and disaggregated,[49] it has been claimed that openness of data also supports the achievement of the SDGs (e.g., Petrov et al., 2016). The Statistics Commission of the UN DESA established an "Open SDG Data Hub" in this regard.[50] Furthermore, the UN Sustainable Development Solutions Network maintains a platform to monitor the SDGs.[51]

©Panchenko Vladimir/Shutterstock.com

[48] https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/

[49] While not in the focus of this document, disaggregation of data is highly relevant for the achievement of the SDGs. The Inter-Agency and Expert Group on Sustainable Development Goal Indicators (2017) calls for SDG indicators to "be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics."

[50] https://unstats-undesa.opendata.arcgis.com/ and also https://sdg.tracking-progress.org/

[51] https://www.unsdsn.org/sdg-index-and-monitoring

➢ 2.2.5.2 Government open data initiatives

Governments produce large quantities of data, which in the past were hardly shared, let alone for reuse in interoperable format. Many open data initiatives by national Governments, intergovernmental organizations, subnational regions as well as municipalities exist already.[52] The Open Government Partnership is an initiative that supports countries, including local Governments, towards transparency including provision of open data.[53] There are also regional initiatives such as openAFRICA, "Africa's Largest Volunteer Driven Open Data Platform",[54] the Asia Open Data Partnership,[55] the Latin American Open Data Initiative[56] and the official portal for European data.[57] In 2021 the second phase of the G20 Data Gaps Initiative came to an end (International Monetary Fund and Financial Stability Board, 2021).

➢ 2.2.5.3 Science open data initiatives

Also, scientific activities produce large amounts of data. In the spirit of science as well as in the interest of progress, these data should be shared particularly if generated with public funding, while keeping in mind the issue of data hazards introduced above. Advocacy and initiatives towards open data in the field of science are not new. For example, the Committee on Data of the International Science Council was already established in 1966.[58] Although there is agreement on sharing scientific data and various platforms exist,[59] their findability and accessibility is challenged by the large amount of data as well as inadequate management. This issue is addressed by FAIR Guiding Principles, an initiative for better data management and stewardship, with an emphasis on involving machines to automatically find data (Wilkinson et al., 2016). Another milestone is the UNESCO Recommendation on Open Science, which was adopted by the UNESCO General Conference in November 2021 (UNESCO, 2021b).

➢ 2.2.5.4 Private open data initiatives

Corporations collect and own large amounts of data too, usually with the sole purpose to increase their revenues. However, a significant subset of these data is pertinent for sustainable development. Although this used to be a challenge, it has now been turned into a win-win opportunity: several corporations have

[52] See here for a (dynamic) overview: https://en.wikipedia.org/wiki/List_of_open_government_data_sites

[53] https://www.opengovpartnership.org/

[54] https://africaopendata.org/

[55] https://www.linkedin.com/company/aodp/

[56] https://idatosabiertos.org/en/

[57] https://data.europa.eu/en

[58] https://codata.org/

[59] See e.g.: https://opendatascience.com/

joined an initiative by UN Global Pulse called "data philanthropy" and committed themselves to share their data for public benefit.[60] Many others work with the UN Institute for Training and Research, under the UN Operational Satellite Applications Programme initiative, to share satellite imagery for humanitarian purposes.[61] While these data can be now used for the implementation and monitoring of sustainable development there are also a variety of incentives for corporations to provide access to their data such as reciprocity; research, recruitment and insights; reputation and public relations; increasing revenue; regulatory compliance; and responsibility and corporate philanthropy (Klein & Verhulst, 2017). In the currently proposed EU Data Governance Act, a similar mechanism, called "data altruism", is introduced to encourage and enable individuals and companies to donate data for the public good.[62]

➢ 2.2.5.5 COVID-19 open data initiatives

Calls for open data initiatives related to COVID-19 were answered by a variety of data sharing initiatives, ranging from open government data to private projects.[63] Some examples illustrating the diversity of the initiatives include:

- The COVID-19 Dashboard by the Center for Systems Science and Engineering at Johns Hopkins University shows the number of cases, deaths and vaccine doses administered by country and is a resource for the public as well as for policymakers.[64]
- Worldometer provides global COVID-19 live statistics by manually analyzing, validating and aggregating data from over 5,000 sources in real time.[65]
- The European COVID-19 Data Platform facilitates data sharing and analysis in order to accelerate coronavirus research.[66]
- The GitHub COVID-19 Open-Data repository attempts to become the largest COVID-19 epidemiological database.[67]

[60] https://www.unglobalpulse.org/2011/09/data-philanthropy-public-private-sector-data-sharing-for-global-resilience/

[61] https://www.unitar.org/sustainable-development-goals/united-nations-satellite-centre-UNOSAT

[62] https://www.consilium.europa.eu/en/press/press-releases/2021/10/01/eu-looks-to-make-data-sharing-easier-council-agrees-position-on-data-governance-act

[63] See for an overview: Alamo et al. (2020).

[64] https://coronavirus.jhu.edu/map.html

[65] https://www.worldometers.info/coronavirus/

[66] https://www.covid19dataportal.org/

[67] https://github.com/GoogleCloudPlatform/covid-19-open-data

- The GISAID platform hosts more than 450,000 viral genomes,[68] but has been criticized that users cannot republish the genomes without permission (Van Noorden, 2021).

- Nextstrain is an open-source project for real-time tracking of the pathogen evolution.[69]

- Global.health is a data repository and visualization platform that enables open access to real-time epidemiological anonymized line-list data.[70]

- Vivli is a platform that shares anonymized participant-level data from completed clinical trials.[71]

- Coronavirus Watch is a platform that provides information on emerging trends related to COVID-19 around the world in the forms of maps, visualizations of the disease status and dashboards.[72]

- Covid-19 Universal Resource gateway (CURE) is supported by UNESCO and is a resource portal for open access resources including recent articles, data dashboards and educational and training on COVID-19.[73]

©tirachardz/Freepik.com

[68] https://www.gisaid.org/

[69] https://nextstrain.org/

[70] https://global.health/

[71] https://search.vivli.org/

[72] https://www.unesco.org/en/articles/corona-virus-media-watch-launched-unescos-international-research-centre-artificial-intelligence

[73] https://www.goap.info/cure

2.2.6 Indices and barometers

In addition, several indices and barometer exist, which assess and rank countries according to their open data efforts, such as the following ones:

- □ Open Data Barometer,[74]
- □ Global Data Barometer,[75]
- □ Open Data Inventory / Watch,[76]
- □ Global Open Data Index,[77]
- □ OECD Open Government Data / OURdata Index,[78]
- □ Open Data Maturity in Europe.[79]

©Chutima Chaochaiya/Shutterstock.com

[74] https://opendatabarometer.org

[75] https://globaldatabarometer.org/

[76] https://opendatawatch.com/

[77] http://index.okfn.org/

[78] https://www.oecd.org/gov/digital-government/open-government-data.htm

[79] https://data.europa.eu/en/impact-studies/open-data-maturity

2.3 AI AND OPEN DATA

2.3.1 Background

While open data are applicable in many fields, they are critical in particular for the field of AI, which will be motivated and explained in this part.

There are various definitions of AI owing to the fact that the definition of intelligence itself is not straightforward. Legg and Hutter (2007, p.12) provide the following general definition: “Intelligence measures an agent’s ability to achieve goals in a wide range of environments”. If the “agent” in this definition is a human being or a non-human animal it is regular intelligence, while it is AI if the agent is a machine. The field of AI can be divided in interacting sub-disciplines, of which Russell and Norvig (2015) list the following ones: natural language processing, knowledge representation, automated reasoning, machine learning, computer vision and robotics.

The recent boom in AI is largely based on massive advancements in machine (and particularly deep) learning, which necessitates large data sets, while some of the other sub-disciplines do not require that much data. Generative AI is the umbrella term for currently very successful machine learning algorithms, which also received much media attention, and which generate artificial digital content such as text, images, audio and video content, based on large amounts of training data. Rather than merely processing existing data, generative AI entails the creation of machines or models that can generate new content, such as images, videos, music, or text. Deep learning is a type of machine learning that employs neural networks to understand patterns and generate new data. The applications of generative AI include the creation of realistic images or videos, the generation of natural language responses, the design of new products, and the generation of music. Concerns exist regarding the possible misuse of generative AI, such as the production of false news or deepfakes. Consequently, it is essential to thoroughly consider the ethical implications of generative AI and to develop responsible usage practices. Not least, the quality of this content is increasingly of an extent that humans cannot distinguish whether the content has been created by a machine or a human, which is causing concerns and controversies.

Therefore, progress in machine learning is on the one hand based on the development of innovative techniques, but on the other hand also on the ever-increasing abundance of accessible data, while their lack in previous decades stifled AI.

However, when it comes to AI and data, there are currently several bottlenecks:

- ❑ There is such a deluge of data that much data cannot be analyzed.
- ❑ There are also available AI tools which are lacking data to work on, since the relevant data have not been collected, or have not been shared, or are siloed, or are out-dated, unreliable or inaccurate, or are in inoperable formats, or are not labelled.

❑ Moreover, for some fields limited, yet critical data are available, but there is a lack of AI tools that can extract findings from such small data.

It has been said that organizations run at the speed of their data. However, if that speed of data is hampered by availability, throughput and latency challenges resulting from legacy limitations, organizations may become hard pressed to achieve the desired insights leading to competitive advantage. Moreover, the deluge of potentially available raw data is not of much use as long as the data are not transformed from various formats into actionable information that informs intelligent decisions.

AI has the capacity to analyse large (or, if there is no other option, small) amounts of data, to uncover formerly unknown or hidden patterns, and to deliver actionable information in real-time. This includes also actionable and corrective information for the implementation of the 2030 Agenda for Sustainable Development as well as for progress towards solutions to further global challenges. This has been recognized in the UN Secretary General's Roadmap for Digital Cooperation, which calls to strengthen digital capacity for enhanced digital cooperation and inclusion (Goal 4) as well as to support global cooperation on AI (Goal 6) (UN General Assembly, 2020).

2.3.2 Synergies between AI and open data to tackle global challenges

Many AI systems run on proprietary data. However, it has been acknowledged that data, which fulfil the criteria of open data, would benefit AI systems further and mitigate potential hazards of the systems such as lacking fairness, accountability and transparency (Davies, 2019). This insight is reflected by the UNESCO Recommendation on Open Science (UNESCO, 2021b), the UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021d) as well as with the theme of the current Third Wave of open data since both encourage impactful reuse of data.

It is also reflected in a recommendation in a report by the Executive Office of the US President – National Science and Technology Council, Committee on Technology, called "Preparing for a Future of Artificial Intelligence", which states that "Federal agencies should prioritize open training data and open data standards in AI. The government should emphasize the release of datasets that enable the use of AI to address social challenges. Potential steps may include developing an "Open Data for AI" initiative with the objective of releasing a significant number of government data sets to accelerate AI research and galvanize the use of open data standards and best practises across government, academia, and the private sector." (Executive Office of the President of the United States of America – National Science and Technology Council, Committee on Technology, 2016, p.14)

However, it has been also recognized that sharing of existing data in an accessible, interoperable and reusable manner is not sufficient, but that also the other issues listed earlier have to be addressed, such as that data are lacking, outdated, unreliable, inaccurate or not labelled. Of concern are, for example, lacking data regarding marginalized communities or the informal economy, which is according to Access Partnership (2018) a political problem since AI systems would have no knowledge of these aspects of society.

This can be also illustrated by the following example: translation systems between English and French are quite advanced nowadays because there are large corpora of both languages for the training of these systems available, while this is not the case for a variety of other languages. In other words, the development of successful AI machine learning depends on the available data, which should be seen as an opportunity for all Member States: In order to develop AI applications for fair decision making and societal good etc., data must not be biased or discriminate against certain countries, cultures or groups Therefore, comprehensive data of the neglected countries, cultures or groups have to be available and included.[80]

©Studio Romantic/Shutterstock.com

There are initiatives to address these challenges, such as by FAIR Forward[81] and the Open for Good Alliance.[82] These initiatives seek to improve localized AI training data of sufficient quality for localized AI innovation. For example, FAIR Forward has gathered 1,200 hours of voice recordings of the Rwandan language Kinyarwanda. FAIR Forward also looked into Indian languages and noted challenges to find general-purpose speech datasets for them in good quality and in standardized format (GIZ, 2020). While such specific initiatives are much welcomed, open training data portals dedicated to AI are scarce and an overall data strategy is missing,

[80] It has to be noted that while more diverse data are necessary to reduce bias and discrimination of AI systems, more diverse data alone are not a sufficient solution to this complex problem.

[81] https://www.giz.de/expertise/html/61982.html

[82] https://www.openforgood.info

which addresses both: How can data be made available to everyone and how can small available datasets lead to actionable insights?

AI-powered technologies are increasingly permeating all aspects of societies. Thus, AI's links to gender issues have become progressively more important in the struggle for gender equality. As there are concerns that gender and open data efforts are mostly siloed (GODAN, 2018), primarily because the constituencies that discuss concerns for women and those that discuss openness of data rarely interact. Thus, the need to open data to achieve can serve as an important step towards gender equality and empowerment. In 2022 to 2023, UNESCO has already pledged to spearhead an inclusive and integrated approach to the development digital competencies, including for AI and bridge digital and knowledge divides. UNESCO's framework of Internet Universality (ROAM-X) advocates for openness of data to address the gender digital divide.

In summary, there are many opportunities for AI towards the measurement and the achievement of sustainable development, provided relevant data are collected as well as openly shared. As outlined, biases are a risk for AI applications, which needs to be mitigated by providing as inclusive datasets as possible. This leads to a call on Member States to not only support openness of high-quality data, but also to embrace the use of AI technologies and facilitate capacity building, training and education in this regard, including AI-inclusive open data literacy.

2.3.3 Potential increased AI risks due to open data

AI is a dual-use technology, thus involves not only opportunities, but also risks. It is important not to only look at open data as a driver for AI risks. The datasets, which are fed into AI systems, may pose a security threat as they may lead to unwanted outputs or reactions of the AI system. While the issues related to bias and discrimination mentioned above are also unwanted, but largely unintentional, there are also malicious acts towards AI systems conceivable through misuse of open data. Such attacks have been named "poison in the well" or "evasion attacks" and comprise different methods (Lohn, 2021). One way would be, for instance, to just replace labels in datasets, while another more sophisticated method targets images through adversarial examples, which are carefully crafted and for humans hardly recognizable perturbations and which cause AI recognition systems to fail (Biggio & Roli, 2018). Openly available datasets unintentionally increase the risk for such attacks since this means an increased volume of potentially attackable data, and at the same time the large amount of data reduces the chances to identify manipulated data.

FURTHER TOPICS

Open data and AI are closely linked to a number of other topics as well as challenges, which cannot be covered in detail in this report, but are nevertheless relevant and listed below to provide an overview:

- ❑ **Data privacy**: Many valuable Government data are about citizens. Therefore, the balance between releasing these data and the privacy of the citizens is an ongoing debate (e.g., Scassa, 2019).
- ❑ **Data ethics**: This topic is related to privacy and "evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use".[83]
- ❑ **Legal issues**: The use of data is typically covered by legislation. For example, article 8 of the Charter of Fundamental Rights of the European Union (2012) states "Everyone has the right to the protection of personal data concerning him or her."[84]

©Brian A Jackson/Shutterstock.com

[83] https://theodi.org/service/consultancy/data-ethics/ See, e.g., also: https://dataethics.eu/

[84] See also European Parliament and Council (2016).

©metamorworks/Shutterstock.com

- **Data governance:** This topic is, for example, addressed by a working group[85] within the Global Partnership on Artificial Intelligence (GPAI), which is a "multi-stakeholder initiative which aims to bridge the gap between theory and practice on AI".[86]

- **Data literacy**: Data literacy is a prerequisite for the use of (open) data. However, there are indications that data literacy levels are low within certain groups. Therefore, capacity building is very much desirable (e.g., Montes & Slater, 2019).

- **Data infrastructure**: Another prerequisite for the provision as well as use of open data is the availability of the hardware component of data infrastructure, which is not evenly distributed across the world (Dodds & Wells 2019).

- **Gender**: As indicated above, gender equality is related to aspects of society, thus also to (open) data. Imbalance of data about women affects AI systems in particular, which mirror this situation in their conclusions.[87]

[85] https://gpai.ai/projects/data-governance/

[86] https://gpai.ai

[87] https://aplusalliance.org

- ❑ **Indigenous data**: Indigenous peoples request additional considerations regarding their data, including sovereignty, summarized as CARE: collective benefit, authority to control, responsibility, ethics.[88]

- ❑ **Small data**: As mentioned before, especially from marginalized groups often only limited data are available. However, it is still important that AI nevertheless can work with them, such as the "Few-Shot Learning" approach (Wang et al., 2020).

- ❑ **Semantic Web**: The Semantic Web is an approach to make Internet data machine-readable. To connect open data to this approach the concept of Linked Open Data has been contrived and formalized (e.g., Bauer & Kaltenböck, 2011).

[88] https://www.gida-global.org/care

3. GUIDELINES

3.1 PREAMBLE

The purpose of the guidelines is for Member States to establish an action plan as well as an open data policy for sustainable development. While the above vision part is about the big picture, the guidelines provide a sharper picture for experts to "dive" into specific topics.

What follows are hands-on yet high-level steps on how to open data, drawing on existing guidelines. Three phases, preparation, opening of the data and follow up for reuse and sustainability, are distinguished and four steps for each phase are presented.

It is important to note that several of the steps can be performed simultaneously, i.e., not necessarily consecutively.

©Vasin Lee/Shutterstock.com

3.2 PREPARATION

➢ Draft a data management and sharing policy

A data management and sharing policy is an important prerequisite before opening data, since such a policy defines the commitment of governments to share data. The Open Data Institute suggests the following elements of an open data policy: a definition of open data, a general declaration of principles, an outline of the types of data and references to any relevant legislation, policies or other guidance.[89] Governments are encouraged to adhere to the principle "as open as possible, as closed as necessary" (Landi et al., 2020). If data cannot be opened due to legal, privacy or other concerns, e.g., personal or sensitive data, this should be clearly explained. The data management and sharing policy should also follow the FAIR as well as the CARE Principles for indigenous data governance. Moreover, Governments should also encourage researchers as well as the private sector in their countries to develop data management and sharing policies adhering to the same principles.

➢ Gather and collect high quality data

Existing data have to be gathered and stored in the same repository, e.g., from various government departments where they may have been stored in silos. Moreover, data gaps have to be filled through new data collection. The data should be accurate and not outdated. Moreover, data should be comprehensive and should not, for example, neglect minorities or the informal economy. Data about people should be disaggregated where relevant, including by income, sex, age, race, ethnicity, migratory status, disability and geographic location (Inter-Agency and Expert Group on Sustainable Development Goal Indicators, 2017).

➢ Develop open data capacities

Capacity building targets two groups: Government officials as well as potential users of the data. The keyword is "data literacy", which falls under the competencies related to media and information literacy proposed by UNESCO[90] and for which courses and curricula exist. For the Government officials, capacity building includes understanding of the benefits of open data and overcoming of common fears and misunderstandings. In this regard also the term data stewardship has been

[89] https://theodi.org/article/how-to-write-a-good-open-data-policy/

[90] https://iite.unesco.org/mil/

coined, which stands for "the systematic, sustainable and responsible management of data for public benefit" (GovLab, 2020). For the potential users, capacity building includes demonstrating opportunities of open data, such as for reusing them, and how to make informed choices.

➢ Make the data AI ready

If the data are not to be used only by humans, but they can also be fed into AI systems, the data have to fulfil a few more criteria to be "AI ready". The first step in this regard is to prepare the data in a machine-readable format. Some formats support readability by AI systems more than others.[91] If the data are, in addition, to be used as training data for supervised learning[92] (as opposed to unsupervised[93] or reinforcement[94] learning) the data also need to be cleaned and labelled, which is often time-consuming, thus costly. The success of an AI system depends on the quality of the training data, including their consistency and relevance. The required amount of training data is hard to know beforehand and has to be monitored through performance checks. The data should cover all scenarios the AI system has been built to work on.

©Peshkova/Shutterstock.com

[91] See for an overview: http://opendatahandbook.org/guide/en/appendices/file-formats/

[92] Supervised learning is based on labelled data, e.g., photographs of people that have been labelled by humans. Then a model is built that can be applied to similar data, e.g., to automatically identify the same people in new photographs (UNESCO, 2021c).

[93] Unsupervised learning is based on data that has not been categorized or labelled. The aim is to uncover hidden patterns in the data, which are used to classify new data. An example would be to automatically identify letters and numbers in handwriting by looking for patterns in a large number of samples (UNESCO, 2021c).

[94] Reinforcement learning involves continuous improvement of a model through feedback. The AI derives from initial data a model, which is assessed as correct or incorrect and rewarded or punished accordingly. Then, this reinforcement is used to update the model, which is iteratively repeated over time (UNESCO, 2021c).

3.3 OPEN THE DATA

➢ Select datasets to be opened

The first step in opening the data is to decide which datasets are to be open. Criteria in favour of opening are: whether there have been requests before for opening these data or whether other Governments have opened these data and whether it resulted in beneficial uses of the data. Criteria against opening are issues related to privacy, intellectual property rights or national security. In other words, the opening of the data must not violate national laws such as data privacy laws.

➢ Open the datasets legally

Before opening the datasets, the respective Government has to specify exactly under which conditions, if any, the data can be used. A variety of open data licences have been developed and the Government can choose which one suits its goals best. Well known licences are Creative Commons and Open licences "that give every person and organization in the world a free, simple, and standardized way to grant copyright permissions for creative and academic works; ensure proper attribution; and allow others to copy, distribute, and make use of those works".[95] Apart from waiving all rights to data the following restrictions may be considered, for all of which Creative Commons licences are available:

- Attribution: When using the data credit to the source must be given.
- Non-Commercial: The data must not be used for commercial purposes.
- No Derivatives: If the data are modified they may not be shared.
- Share Alike: Modified data may be shared, but under the same licence.

Another option is for Governments to develop their own licences as for example the UK Government has done.[96]

[95] https://creativecommons.org/about/

[96] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

➢ Open the datasets technically

The most common way to open data is to publish them in electronic format for download on a website, either the Government's own website or a third-party website. In addition, the data could be also published in a data archive or repository, which should be, according to the UNESCO Recommendation on Open Science, "supported and maintained by an academic institution, scholarly society, government agency or other well-established not-for-profit organization devoted to common good that enables open access, unrestricted distribution, interoperability and long-term digital preservation and archiving" (UNESCO, 2021b, p.9). The data should be presented in a format that they are findable, accessible, interoperable and reusable, thus complying with the FAIR principles as introduced above. For the findability and the accessibility are various open-source data management systems available to establish the open data portal as a one-stop shop.[97] The datasets should be accompanied by metadata, which further specify the data and for which four generic standards exist: Dublin Core,[98] Data Catalog Vocabulary,[99] DataCite[100] and Schema.org.[101] Regarding specific data formats, the five-star open data plan was introduced earlier in this report. The higher the number of stars the more open are the data.[102] The data formats should support interoperability, e.g., with application programming interfaces, and reusability, e.g., by other users as well as by machines.[103] In addition, the AI readability, discussed in step 4, must be kept in mind. Moreover, security measures are critical to prevent data from being manipulated by unauthorized users.

©one photo/Shutterstock.com

[97] See e.g.: https://ckan.org/, https://magda.io/ or https://www.re3data.org/

[98] http://dublincore.org/

[99] https://www.w3.org/TR/vocab-dcat-2/

[100] https://datacite.org/

[101] https://schema.org/

[102] https://5stardata.info/en/ and https://www.w3.org/DesignIssues/LinkedData.html

[103] See for an overview: http://opendatahandbook.org/guide/en/appendices/file-formats/

➢ Create an open-data-driven culture

Experience has shown that in addition to opening data legally and technically at least two more things have to be accomplished to reach an open-data-driven culture: Often staff, especially in Government departments, are not used to sharing data, but rather have been trained to keep a silo mentality. Moreover, data should become, if possible, the exclusive basis for decision-making; in other words, decisions should be anchored in data. Both topics are also raised by "Step 3: Develop open data capacities", yet in addition to capacities cultural changes are required by all involved staff. There is a need to foster proactive data disclosure, which can guarantee that data is made available even before a demand for the data is placed. This can be ensured through the inclusion of such provisions in the Access to Information Laws.[104]

3.4 FOLLOW UP FOR REUSE AND SUSTAINABILITY

➢ Support citizen engagement

After opening the data, they must be discovered by potential users. For this an advocacy strategy has to be developed, which may comprise announcing the opening of the data in open data communities and relevant social media channels, for example by using the hashtag #opengov. Another important activity is early consultation and engagement with the potential users. In addition to informing potential users about the open data they should be also encouraged to use and reuse the data and to remain further engaged. Mechanisms, which have proven to be successful, are hackathons or other competitions to use open data to address specific challenges and problems.[105]

[104] See e.g.: https://en.unesco.org/themes/access-information-laws

[105] http://opendatatoolkit.worldbank.org/en/demand.html#stage-3

©Miha Creative/Shutterstock.com

➢ Support international engagement

International partnerships would increase the benefits of the open data even further, such as through south-south and north-south collaboration. Especially partnerships are important, which support and build capacities towards the impactful reuse of the data, either through the use of AI (see below) or without. The Open for Good Alliance, Global Partnership on Artificial Intelligence in particular as well as the intergovernmental and regional initiatives, which were introduced above, are commendable.

➢ Support beneficial AI engagement

Open data provide many opportunities for AI systems. To achieve the data's full potential, this requires advocacy for developers to make use of the data and develop AI systems accordingly. At the same time, the abuse of the open data for irresponsible and harmful AI applications has to be prevented. One example is Responsible AI Licences (RAIL).[106] This step is linked to the previous one. Another recommended practice is to keep a public record, which data have been used by AI systems and how.

[106] https://www.licenses.ai/about

➢ Maintain high quality data

Many data are quickly outdated. Therefore, the datasets have to be updated regularly.

The step “Maintain high quality data” makes this guideline a loop as it links back to the step “Gather and collect high quality data”.

©UnderhilStudio/Shutterstock.com

EPILOGUE

These guidelines serve as a call to action in accordance with the UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021d). If UNESCO Member States follow these guidelines and open their data in a sustainable manner, and create capacities as well as an open-data-driven culture accordingly, the vision outlined in the previous part with all its benefits could become reality.

Open data are a necessary prerequisite for the monitoring as well as the achievement of sustainable development. Due to the magnitude of the tasks Governments should not only embrace opening the data, but also create favourable conditions for beneficial AI engagement that creates new knowledge out of the open data, for evidence-based decision-making.

In a nutshell: Make your data findable, accessible, interoperable and reusable as well as AI-ready, in short FAIR, so that they can be processed and analyzed further for societal good by anybody.

©3rdtimeluckystudio/Shutterstock.com

4. REFERENCES

Access Partnership (2018). Artificial Intelligence for Africa: An opportunity for growth, development and democratisation. South Africa: University of Pretoria.

https://www.up.ac.za/media/shared/7/ZP_Files/ai-for-africa.zp165664.pdf

Alamo, T., Reina, D. G., Mammarella, M., & Abella, A. (2020). Covid-19: Open-data resources for monitoring, modeling, and forecasting the epidemic. Electronics, 9(5), 827.

https://www.mdpi.com/2079-9292/9/5/827/htm

Bauer, F., & Kaltenböck, M. (2011). Linked open data: The Essentials. *Edition mono/monochrom, Vienna*, *710*.

https://www.reeep.org/LOD-the-Essentials.pdf

Berners-Lee, T. (2012). Raw data, now! Wired.

https://www.wired.co.uk/article/raw-data

Big Data for Migration Alliance (2021). Designing Data Collaboratives to Better Understand Human Mobility and Migration in West Africa.

https://odpl.thegovlab.com/uploads/ccm-directus/originals/aa45abf2-743d-4d58-8471-e25bff956ba5.pdf

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, *84*, 317-331.

https://arxiv.org/pdf/1712.03141.pdf

Bodor, Á., Varjú, V., & Grünhut, Z. (2020). The effect of trust on the various dimensions of climate change attitudes. *Sustainability*, *12*(23), 10200.

https://www.mdpi.com/2071-1050/12/23/10200

von Borzyskowski, I., Mazumder, A., Mateen, B., & Wooldridge, M. (2021). Data science and AI in the age of COVID-19. The Alan Turing Institute.

https://www.turing.ac.uk/research/publications/data-science-and-ai-age-covid-19-report

Bostrom, N. (2011). Information hazards: A typology of potential harms from knowledge. *Review of Contemporary Philosophy*, (10), 44-79.

https://www.nickbostrom.com/information-hazards.pdf

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.

Cascella, M., Rajnik, M., Aleem, A., Dulebohn, S. C., & Di Napoli, R. (2022). Features, evaluation, and treatment of coronavirus (COVID-19). Statpearls [internet].

https://www.ncbi.nlm.nih.gov/books/NBK554776/

Davies, T. (2019). Issues in Open Data – Algorithms and AI. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre.

https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/57585/The%20State%20of%20Open%20Data.pdf

Dodds, L. & Wells, P. (2019). Issues in Open Data – Data Infrastructure. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre.

https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/57585/The%20State%20of%20Open%20Data.pdf

European Parliament and Council (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN

European Union (2012). Charter of Fundamental Rights of the European Union. Official Journal of the European Union, C 326/391.

https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT&from=EN

Executive Office of the President of the United States of America – National Science and Technology Council Committee on Technology (2016). Preparing for a Future of Artificial Intelligence.

https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Giovannini, G., Haick, H., & Garoli, D. (2021). Detecting COVID-19 from breath: A game changer for a big challenge. ACS sensors, 6(4), 1408-1417.

https://pubs.acs.org/doi/10.1021/acssensors.1c00312

GIZ (2020). A Study on Open Voice Data in Indian Languages.

https://toolkit-digitalisierung.de/app/uploads/2021/02/Study-on-Open-Voice-Data-in-Indian-Languages_GIZ-BizAugmentor.pdf

GovLab (2020). Wanted: Data Stewards.

https://thegovlab.org/static/files/publications/wanted-data-stewards.pdf

Harrus, I., & Wyndham, J. (2021). Artificial intelligence and COVID-19: applications and impact assessment. AAAS AI Report.

https://www.aaas.org/sites/default/files/2021-05/AIandCOVID19_2021_FINAL.pdf

Heaven, W.D. (2021). Hundreds of AI tools have been built to catch covid. None of them helped. MIT Technology Review.

https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/

Hussain, A. A., Bouachir, O., Al-Turjman, F., & Aloqaily, M. (2020). AI techniques for COVID-19. IEEE access, 8, 128776-128795.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9136710

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014). *A world that counts: Mobilizing a Data Revolution for Sustainable Development*.

http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf

Inter-Agency and Expert Group on Sustainable Development Goal Indicators (2017). Revised list of global Sustainable Development Goal indicators. Annex III. E/CN.3/2017/2.

https://unstats.un.org/sdgs/indicators/official%20revised%20list%20of%20global%20sdg%20indicators.pdf

International Monetary Fund and Financial Stability Board (2021). G20 Data Gaps Initiative (DGI-2). The Sixth Progress Report — Countdown to December 2021.

https://www.fsb.org/wp-content/uploads/P081021-1.pdf

International Telecommunication Union and CISCO Systems (2016). Harnessing the Internet of Things for Global Development.

http://www.itu.int/en/action/broadband/Documents/Harnessing-IoT-Global-Development.pdf

Khemasuwan, D., & Colt, H. G. (2021). Applications and challenges of AI-based algorithms in the COVID-19 pandemic. BMJ Innovations, 7(2).

https://innovations.bmj.com/content/bmjinnov/7/2/387.full.pdf

Klein, T., & Verhulst, S. (2017). Access to new data sources for statistics: Business models and incentives for the corporate sector.

https://www.thegovlab.org/static/files/publications/paris-21.pdf

Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., da Silva Santos, L. O. B., & Roos, M. (2020). The "A" of FAIR–as open as possible, as closed, as necessary. *Data Intelligence*, *2*(1-2), 47-55.

https://direct.mit.edu/dint/article/2/1-2/47/9998/The-A-of-FAIR-As-Open-as-Possible-as-Closed-as

Lee, C., Yang, T., Inchoco, G. D., Jones, G. M., & Satyanarayan, A. (2021). Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1-18).

https://arxiv.org/pdf/2101.07993.pdf

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines*, *17*(4).

https://arxiv.org/pdf/0712.3329.pdf

Lohn, A.J. (2021). Poison in the Well - Securing the Shared Resources of Machine Learning. CSET Policy Brief.

https://cset.georgetown.edu/publication/poison-in-the-well/

Lozano, M. A., Piñol, E., Rebollo, M., Polotskaya, K., Garcia-March, M. A., Conejero, J. A., ... & Oliver, N. (2021). Open Data Science to Fight COVID-19: Winning the 500k XPRIZE Pandemic Response Challenge. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 384-399). Springer, Cham.

https://2021.ecmlpkdd.org/wp-content/uploads/2021/07/sub_638.pdf

Milan, S., & Treré, E. (2020). The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins. Social Media+ Society, 6(3), 2056305120948233.

https://journals.sagepub.com/doi/full/10.1177/2056305120948233?fbclid=IwAR0jUNmHgCSSZmOXaByn3X4EyKoWRDCxM6qXnCA1FuN-yrgF_s8rnoCgVqM&

Montes, M. & Slater, D. (2019). Issues in Open Data - Data Literacy. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre.

https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/57585/The%20State%20of%20Open%20Data.pdf

Odriozola-González, P., Planchuelo-Gómez, Á., Irurtia, M. J., & de Luis-García, R. (2020). Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. *Psychiatry research*, *290*, 113108.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7236679/

OECD (2020). Why open science is critical to combatting COVID-19. OECD Policy Responses to Coronavirus (COVID-19).

https://read.oecd-ilibrary.org/view/?ref=129_129916-31pgjnl6cb&title=Why-open-science-is-critical-to-combatting-COVID-19

OECD & GovLab (2021). Open Data in action. Initiatives during the initial stage of the COVID-19 pandemic.

https://www.oecd.org/gov/digital-government/open-data-in-action-initiatives-during-the-initial-stage-of-the-covid-19-pandemic.pdf

Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. Frontiers in immunology, 11, 1581.

https://www.frontiersin.org/articles/10.3389/fimmu.2020.01581/full

Open Data Charter (2015). International Open Data Charter.

https://opendatacharter.net/wp-content/uploads/2015/10/opendatacharter-charter_F.pdf

Pahar, M., Klopper, M., Warren, R., & Niesler, T. (2021). COVID-19 cough classification using machine learning and global smartphone recordings. Computers in Biology and Medicine, 135, 104572.

https://arxiv.org/pdf/2012.01926.pdf

Panchal, N., Kamal, R., Orgera, K., Cox, C., Garfield, R., Hamel, L., & Chidambaram, P. (2020). The implications of COVID-19 for mental health and substance use. Kaiser family foundation, 21.

https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/

Petrov, O., Gurin, J., & Manley, L. (2016). Open data for sustainable development.

https://openknowledge.worldbank.org/bitstream/handle/10986/24017/Open0data0for0sustainable0development.pdf?sequence=4&isAllowed=y

Ratcliffe, J. & Tuzeneu, M.-C. (2019). Maximizing the Power of Climate Data by Building Trust. Blog post.

https://www.climatelinks.org/blog/maximizing-power-climate-data-building-trust

Raveendran, A. V., Jayadevan, R., & Sashidharan, S. (2021). Long COVID: an overview. Diabetes & Metabolic Syndrome: Clinical Research & Reviews.

https://www.sciencedirect.com/science/article/pii/S1871402121001193

Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., ... & Schönlieb, C. B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence, 3(3), 199-217.

https://www.nature.com/articles/s42256-021-00307-0

Russell, S. J. & Norvig, P. (2015). Artificial Intelligence-A Modern Approach, Third Edition. Upper Saddle River: Pearson.

Sachs, J. D., Kroll, C., Lafortune, G., Fuller, G. & Woelm, F. (2021). Sustainable Development Report 2021. Cambridge: Cambridge University Press.

https://s3.amazonaws.com/sustainabledevelopment.report/2021/2021-sustainable-development-report.pdf

Scassa, T. (2019). Issues in Open Data - Privacy. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre.

https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/57585/The%20State%20of%20Open%20Data.pdf

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., ... & Shen, D. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE reviews in biomedical engineering, 14, 4-15.

https://arxiv.org/ftp/arxiv/papers/2004/2004.02731.pdf

Shuja, J., Alanazi, E., Alasmary, W., & Alashaikh, A. (2021). COVID-19 open source data sets: a comprehensive survey. Applied Intelligence, 51(3), 1296-1325.

https://link.springer.com/article/10.1007/s10489-020-01862-6

UN General Assembly (2020). Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation. Report of the Secretary-General. A/74/821.

https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/102/51/PDF/N2010251.pdf?OpenElement

UNESCO (2020). Disinfodemic. Deciphering COVID-19 disinformation. Policy brief 1.

https://en.unesco.org/sites/default/files/disinfodemic_deciphering_covid19_disinformation.pdf

UNESCO (2021a). One year into COVID: prioritizing education recovery to avoid a generational catastrophe. ED/ADG/2021/01

https://unesdoc.unesco.org/ark:/48223/pf0000376984

UNESCO (2021b). Recommendation on Open Science. SC-PCB-SPP/2021/OS/UROS.

https://unesdoc.unesco.org/ark:/48223/pf0000379949

UNESCO (2021c). AI and Education Guidance for Policy-makers. Paris, UNESCO.

https://unesdoc.unesco.org/ark:/48223/pf0000376709

UNESCO (2021d). Recommendation on the Ethics of Artificial Intelligence. Paris, UNESCO.

https://unesdoc.unesco.org/ark:/48223/pf0000381137

UN System Chief Executives Board for Coordination (2020). Summary of deliberations. Addendum. System-wide Road Map for Innovating United Nations Data and Statistics. CEB/2020/1/Add.1

https://unsceb.org/sites/default/files/2021-01/CEB_2020_1_Add1.pdf

Van Noorden, R. (2021). Scientists call for fully open sharing of coronavirus genome data. Nature, 590(7845), 195-197.

https://www.nature.com/articles/d41586-021-00305-7

Waltman, L., Pinfield, S., Rzayeva, N., Henriques, S. O., Fang, Z., Brumberg, J., ... & Swaminathan, S. (2021). Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic.

https://rori.figshare.com/articles/report/_/17125394

Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. ACM Computing Surveys (CSUR), 53(3), 1-34.

https://arxiv.org/pdf/1904.05046.pdf

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., … & van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. bmj, 369.

https://www.bmj.com/content/369/bmj.m1328

Yampolskiy, R. V. (Ed.). (2018). *Artificial intelligence safety and security.* CRC Press.

Yan, L., Zhang, H. T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., … & Yuan, Y. (2020). An interpretable mortality prediction model for COVID-19 patients. Nature machine intelligence, 2(5), 283-288.

https://www.nature.com/articles/s42256-020-0180-7.pdf

Ziesche, S. (2017). Innovative big data approaches for capturing and analyzing data to monitor and achieve the SDGs. *Report of the United Nations Economic and Social Commission for Asia and the Pacific: Subregional Office for East and North-East Asia (ESCAP-ENEA).*

https://www.unescap.org/publications/innovative-big-data-approaches-capturing-and-analyzing-data-monitor-and-achieve-sdgs

unesco
Centre
Under the auspices
of UNESCO

IRCAI
International Research Centre
on Artificial Intelligence
under the auspices of UNESCO

open
for
good

UniTwin
unesco
University Twinning and
Networking Programme

UNESCO Chair on Artificial Intelligence (2019), University College of London, United Kingdom of Great Britain and Northern Ireland
UNESCO Chair on Analytics and Data Science (2016) University of Essex, United Kingdom of Great Britain and Northern Ireland
UNESCO Chair on Artificial Intelligence and Data Science for Society (2022), The Hague University of Applied Sciences (THUAS)

9 789231 006005