



Editorial: Science, data and society

David Carlson^{1,☆,i} and Hans Pfeiffenberger^{2,☆,*}

¹formerly at: World Climate Research Programme, Geneva, Switzerland

²formerly at: Alfred Wegener Institute, Bremerhaven, Germany

☆retired

ⁱESSD senior chief editor

*ESSD advisory board member

Correspondence: David Carlson (ipy.djc@gmail.com)

Published: 8 February 2023

Abstract. Quality data remain elusive while data access freedoms disappear. Serious mis-matches between data availability and human need should attract societal attention.

The journal we founded, *Earth System Science Data* (*ESSD*, Copernicus), developed into the first high-impact mechanism to facilitate free exchange of reliable research data. *ESSD* now presents a remarkable variety and volume of openly accessible quality-certified data products covering many aspects of environmental, geophysical and biogeochemical science. Despite these positive developments, we hesitate to celebrate because we detect clear evidence that present deluges of information remain largely untrustworthy and/or not accessible. Science and the society it serves grow increasingly vulnerable to mistakes and mis-steps as a consequence of limitations on open sharing of trustworthy data.

As instantaneous access to geographic, economic and social information grows, we remind ourselves that contemporary world views, shaped by this information, remain subservient to imposed perspectives and interests. Governments and corporations compile and exploit datasets and models, computer-based visualisations, data analytics and machine learning, and other artificial intelligence tools to shape citizen and consumer perspectives. “Customary” interpolations and extrapolations of patchy data mislead or support misconceptions about under-served parts of the world. Emerging usage of cloud storage and computing by countries, institutions and individuals, promoted as useful, affordable and convenient, introduces new barriers to data exchange while eroding essential concepts of reproducibility: when commercial entities control – and modify at will – provenance of global data products, science loses necessary traceability. Rapid exchanges of forecast products or satellite images via high-

bandwidth connections, exploited between prominent centres, fail in many cases to ensure equitable global access.

Data-dependent society faces unprecedented challenges. Global pandemics threaten human health. Storms, flash floods, droughts and heatwaves exacerbated by climate change assault human and planetary well-being. Biodiversity, when measurable, deteriorates on local as well as global scales. These challenges occur amidst society-wide deficiencies in equity and justice. At the same time, data journals such as Copernicus’ *ESSD* or Nature Publications’ *Scientific Data* receive an increasing volume of diverse submissions. Do more authors seek high impact factors of newly successful data journals? Or do we observe increasing recognition of individual and collective benefit from openly shared data? We highlight three examples – COVID-19 infections (public health), CO₂ emissions (climate), and species abundance and distributions (biodiversity) – wherein current data practices seem insufficient to support useful science-based social responses.

1 Infections

Countries accumulate genetic data to exploit commercial aspects, establish or maintain global pre-eminence, or protect national security (New York Times, 2021). Calls for “coordination and standardisation of data collection, data quality, monitoring, and reporting” to serve public health needs (Sachs et al., 2022) clearly conflict with plans to exploit data for national or monopolistic commercial benefit. Fail-

ure to define, collect, and share reliable COVID-19 infection and health data in a timely manner inhibits effective global, national and local responses (SPIEGEL, 2021; Washington Post, 2022). As this particular virus proves more persistent and more flexible than expected, accurate timely tracking of infections seems to recede. Do such outcomes represent public health success or public information failures? How will our global public health community develop necessary accessible tools to track and respond to the next pandemic? How will global societies share trustworthy warnings or efficacy assessments? Will data journals play useful roles?

2 Emissions

Countries report greenhouse gas emissions by territory and sector to UNFCCC, but many emissions reports arrive late, lack necessary detail, and, rarely, exploit loopholes or otherwise manipulate data to hide non-compliance or to project favourable impressions (Washington Post, 2021). Emissions data from military operations remain largely “off the books”. As a consequence, countries debate international climate policies based on flawed emissions accounting. A global research community, evaluating shared up-to-date data from remote sensing, ground-based networks and advanced models, expends increasing time and effort to identify and resolve discrepancies (e.g. Deng et al., 2022; Grassi et al., 2022). Amidst unfortunate uncertainty about data veracity, how will society develop, certify and apply high-resolution openly shared reliable global CO₂ emission data?

3 Species distributions

Biodiversity data remain restricted. Habits and policies (e.g. the Nagoya protocol promoted by the Convention on Biological Diversity, CBD, 2022) imposed nationally or adopted among international scientific communities, despite positive intentions, make no progress toward convenient global access with the effect of inhibiting systematic analyses. Countries and communities have not yet agreed on definitions of “key” or essential information. While abundance data emerge for some species, data for other species or same species in other regions remain hidden or blocked. Cross-border approaches remain rare: compiling sufficient data to plan or evaluate ecosystem-wide management options (e.g. protected areas, migration corridors) remains extraordinarily difficult. Ecologically relevant information remains largely unavailable in quality-certified open-access formats as practised by data journals. As land use and land use changes become more evident via remote sensing, more important for monitoring biodiversity as well as emissions, and more subject to national manipulation, will biodiversity communities agree on terms and issues? In ocean ecosystems largely hidden from remote sensing, will resource exploitation data

remain subject to proprietary national and commercial policies?

We repeat our initial assessment: severe deficiencies in how science and society develop, share, validate and use data leave us increasingly vulnerable to mistakes and mis-steps as we confront planetary challenges. We note recent admonitions (Anthony Fauci, New York Times, 2022) that reinforce concerns: “It is our collective responsibility to ensure that public health policy decisions are driven by the best available data.” Open access to quality data via recently successful data journals represents a positive but meagre response, not scalable to vast varieties or volumes of data. We proclaim three urgent recommendations: (a) all publicly funded data, and all other data necessary to inform public policies, must be open; (b) we must train society to expect, discover and use open trustworthy data; (c) we insist on open data as the international, national, commercial, environmental and economic default rather than exception. Meeting these challenges presumes functional funded data infrastructure.

- A. All data from all sources must emerge and remain as freely and openly accessible as technically and ethically feasible. We particularly urge that all data, of any form or format, used in, applied to, or serving as a basis for public environmental, economic, security and health policies reside in free well-documented accessible public repositories. Open data access as practised at the moment by data journals will represent a powerful antidote to inadvertent or deliberate biases, allowing and encouraging evaluation of completeness, accuracy and trustworthiness.
- B. A world of free access to open data will only develop in parallel with a data-smart society. We call for focus on data availability, reliability and use as a highly relevant feature of “higher” and vocational education. We urge broad exposure to global health, biodiversity and environmental data issues for every student regardless of intended specialisation. We must offer next generations knowledge and tools to challenge access barriers, assimilate disparate data, produce skilful analyses, and scrutinise sources and biases. We identify clear roles for data journals: Borduas-Dedekind et al. (2022) report positive outcomes when students review datasets.
- C. Beyond recommendations for data as a free open asset available to all citizens, we argue for philosophical and technical change of direction: open data to exist as the default. Citizens should govern all data with societal relevance, except in rare cases when they have, a priori, agreed to exceptions. Mindful that a single researcher apparently initiated the rapid exchange of sequence data for the recent corona virus that, within weeks, led to specific PCR tests and mRNA vaccines, we wish to see such brave decisions become common rather than rare, expected rather than serendipitous. A society that enjoys

open access to data gains at least a science-based chance to improve its response to health, climate and biodiversity issues. Next generations, expecting open access and trained to evaluate data quality, will take positive steps toward understanding how society in general, and politicians specifically, expose or hide and use or misuse their data.

We do not underestimate the magnitude of change needed to confront global environmental and equity issues. We do not call for more data. Current or future open data practices within scientific journals may have negligible impact on larger social and political issues, but they set good examples and occasionally provide benchmarks for plausibility checks. As countries cooperate to identify and implement persistent coherent responses to health, climate and biodiversity issues, they will quickly confront issues of data access and reliability. We advocate for substantial improvement in how society interacts with information and how cultures interact with each other through data exchange, based on positive community experience with *ESSD*. We anticipate vociferous objections based on national, commercial, military and privacy interests. We do not discount those concerns but prefer them as rare exceptions rather than broad justification. We contend that humanity's necessary successful response to health, climate and biodiversity challenges will require careful competent open access to reliable data.

Unfortunately, data access freedoms, dependent on careful collection, responsible development of algorithms and code, and effective quality assurance, disappear faster than researchers know, beyond the ken of most citizens. Serious mis-matches between data availability and human need should alarm researchers, data managers and journalists and should attract societal attention. During the most recent (2007–2008) International Polar Year (IPY), one of us (Carlson, 2011) reported “inadequate services, almost no international support, and few solutions”. We suspect this dismal situation has yet to improve for future international efforts; recovering IPY data after-the-fact required substantial effort (Driemel et al., 2015). To meet impending challenges, society must provide urgent attention to openly shared trustworthy data.

Acknowledgements. We thank Falk Huettmann (long-time advocate of open data policies and practices) and Andrew Hufton (founding chief editor of *Scientific Data*) for useful discussions. We thank Francesco Tubiello (frequent *ESSD* contributor) for helpful comments. We thank a large (and still growing) community of editors, reviewers, authors and data providers for remarkable efforts on behalf of *ESSD*. Finally, we acknowledge with deep thanks enduring support (at all levels) from Copernicus Publications.

References

- Borduas-Dedekind, N., Short, K., and Carlson, S.: Peer-reviewing for Earth System Science Data as a student training exercise, *Earth Syst. Sci. Data Discuss.* [preprint], <https://doi.org/10.5194/essd-2022-387>, in review, 2022.
- Carlson, D.: A lesson in sharing, *Nature*, 469, 293, <https://doi.org/10.1038/469293a>, 2011.
- CBD (Convention on Biological Diversity): The Nagoya Protocol on Access and Benefit-sharing, <https://www.cbd.int/abs/>, last access: 5 May 2022.
- Deng, Z., Ciaia, P., Tzompa-Sosa, Z. A., Saunio, M., Qiu, C., Tan, C., Sun, T., Ke, P., Cui, Y., Tanaka, K., Lin, X., Thompson, R. L., Tian, H., Yao, Y., Huang, Y., Lauerwald, R., Jain, A. K., Xu, X., Bastos, A., Sitch, S., Palmer, P. I., Lauvaux, T., d'Aspremont, A., Giron, C., Benoit, A., Poulter, B., Chang, J., Petrescu, A. M. R., Davis, S. J., Liu, Z., Grassi, G., Albergel, C., Tubiello, F. N., Perugini, L., Peters, W., and Chevallier, F.: Comparing national greenhouse gas budgets reported in UNFCCC inventories against atmospheric inversions, *Earth Syst. Sci. Data*, 14, 1639–1675, <https://doi.org/10.5194/essd-14-1639-2022>, 2022.
- Driemel, A., Grobe, H., Diepenbroek, M., Grüttemeier, H., Schumacher, S., and Sieger, R.: The IPY 2007–2008 data legacy – creating open data from IPY publications, *Earth Syst. Sci. Data*, 7, 239–244, <https://doi.org/10.5194/essd-7-239-2015>, 2015.
- Grassi, G., Conchedda, G., Federici, S., Abad Viñas, R., Korosuo, A., Melo, J., Rossi, S., Sandker, M., Somogyi, Z., Vizzarri, M., and Tubiello, F. N.: Carbon fluxes from land 2000–2020: bringing clarity to countries' reporting, *Earth Syst. Sci. Data*, 14, 4643–4666, <https://doi.org/10.5194/essd-14-4643-2022>, 2022.
- New York Times: U.S. Warns of Effort by China to Collect Genetic Data, <https://www.nytimes.com/2021/10/22/us/politics/china-genetic-data-collection.html> (last access: 9 May 2022), 2021.
- New York Times: Anthony Fauci: A Message to the Next Generation of Scientists, <https://www.nytimes.com/2022/12/10/opinion/anthony-fauci-retirement.html> (last access: 17 January 2023), 2022.
- Sachs, J. D., Karim, S. S. A., Akin, L., et al.: The Lancet Commission on lessons for the future from the COVID-19 pandemic, *Lancet*, 400, 1224–1280, [https://doi.org/10.1016/S0140-6736\(22\)01585-9](https://doi.org/10.1016/S0140-6736(22)01585-9), 2022.
- Spiegel: Wann kommt die fünfte Welle?, <https://www.spiegel.de/wissenschaft/medizin/coronavirus-fachleute-fordern-neuen-fruehwarnindikator-a-bc1631ca-66b1-4271-ada2-78178eabe65d> (last access: 14 November 2021), 2021.
- Washington Post: Climate pledges built on flawed emissions data, <https://www.washingtonpost.com/climate-environment/interactive/2021/greenhouse-gas-emissions-pledges-data/> (last access: 9 May 2022), 2021.
- Washington Post: The pandemic's latest unexpected turn: A lack of test result data, <https://www.washingtonpost.com/opinions/2022/04/15/covid-test-data-shortage-clouds-pandemic-future/> (last access: 9 May 2022), 2022.