Open Weather and Climate Science in the Digital Era

Martine G. de Vos[1,2], Wilco Hazeleger[1,3], Driss Bari[4], Joerg Behrens[5], Sofiane Bendoukha[5], Irene Garcia-Marti[6], Ronald van Haren[1], Sue Ellen Haupt[7], Rolf Hut[8], Fredrik Jansson[9], Andreas Mueller[10], Peter Neilley[11], Gijs van den Oord[1], Inti Pelupessy[1], Paolo Ruti[12], Martin G. Schultz[13], and Jeremy Walton[14]

[1]Netherlands eScience center, Amsterdam, the Netherlands
[2]Information and Technology Services, Utrecht University, Utrecht, the Netherlands
[3]Geosciences, Utrecht University, Utrecht, the Netherlands
[4]CNRMSI/SMN, Direction de la Meteorologie Nationale Casablanca, Morocco
[5]German Climate Computing Centre (DKRZ), Hamburg, Germany
[6]Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands
[7]Research Applications Laboratory, National Center for Atmopsheric Research, Boulder, USA
[8]Water Resources Management, Delft University of Technology, Delft, the Netherlands
[9]Centrum Wiskunde & Informatica, Amsterdam, the Netherlands
[10]Numerical methods, European Centre for Medium-Range Weather Forecasts, Reading, UK
[11]The Weather Company/IBM, Boston MA, USA
[12]World Weather Research Division, World Meteorological Organization, Geneva, Switzerland
[13]Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany
[14]Hadley Centre for Climate Science, Met Office, Exeter, UK

Correspondence: Martine G. de Vos (m.g.devos@uu.nl)

Abstract. The need for open science has been recognized by the communities of meteorology and climate science. While these domains are mature in terms of applying digital technologies, the implementation of open science methodologies is less advanced. In a session on "Weather and Climate Science in the Digital Era" at the 14th IEEE International eScience Conference domain specialists and data and computer scientists discussed the road towards open weather and climate science.

Roughly 80% of the studies presented in the conference session showed the added value of open data and software. These studies included open datasets from disparate sources in their analyses, or developed tools and approaches that were made openly available to the research community. Furthermore, shared software is a prerequisite for the studies which presented systems like a model coupling framework or digital collaboration platform. Although these studies showed that sharing code and data is important, the consensus among the participants was that this is not sufficient to achieve open weather and climate science and that there are important issues to address.

At the level of technology, the application of the FAIR principles to many datasets used in weather and climate science remains a challenge. This may be due to scalability (in the case of high-resolution climate model data, for example), legal barriers such as those encountered in using weather forecast data, or issues with heterogeneity (for example, when trying to make use of citizen data). In addition, the complexity of current software platforms often limits collaboration between researchers and the optimal use of open science tools and methods.

The main challenges we observed, however, were non-technical and impact the practice of science as a whole. There is a need for new roles and responsibilities in the scientific process. People working at the interface of science and digital technology - e.g., data stewards and research software engineers - should collaborate with domain researchers to ensure the optimal use of open science tools and methods. In order to remove legal boundaries on sharing data, non-academic parties such as meteorological institutes should be allowed to act as trusted agents. Besides the creation of these new roles, novel policies regarding open weather and climate science should be developed in an inclusive way in order to engage all stakeholders.

Alongside the issues and challenges regarding open weather and climate science, we discussed opportunities and possible solutions. We have compiled these into a list of concrete recommendations that could bring us closer to open weather and climate science. We acknowledge that the development of open weather and climate science requires effort to change, but the benefits are large. We have observed these benefits directly in the studies presented in the conference and believe that it leads to much faster progress in understanding our complex world.

1 INTRODUCTION

In this article we describe the main findings of a conference session on "Weather and Climate Science in the Digital Era" with a special focus on the implementation of open science methodologies.

Meteorology and climate sciences are data- and computationally-intensive areas of research by tradition. Being primarily a physical science, empirical data collection has always been important and meteorology was one of the first fields that standardized data collection from the advent of systematic instrumental observations in the mid-1800s (e.g. Maury, 1853; Quetelet, 1874). In addition, the production of meteorological forecasts was one of the first applications to be developed for electronic computers, following decades during which the calculations were performed by hand (we recall that "computer" originally meant "one who computes", and that the adjective "electronic" was introduced to distinguish the machine from the human). Numerical weather prediction (NWP) has advanced from the first operational predictions in the 1950s (Charney et al., 1950), aided by increased computing capability and the growing supply of observational data to generate initial conditions for assimilation into the model state. Climate research has benefitted from the same developments (see e.g. Lynch, 2008, for an overview).The assimilation of observational data into NWP models has been a turning point for the development of high-resolution gridded information of the atmosphere and ocean state (e.g. Kalnay et al., 1996; Dee et al., 2011). The use of this methodology for reanalysis - that is, generating a comprehensive and physically consistent record of how the weather is changing over time - has ensured a baseline for climate research and triggered the development of downstream climate services.

Meteorologists have been using machine learning to post-process model output, blend multiple models, and optimize the weighting of those multiple models for over 20 years (Haupt et al., 2018). Neural nets were used in the 90s to speed up the calculation of outgoing longwave radiation in climate models (Chevallier et al., 1999), and for both short- and long-wave radiation parameterization in the National Center for Atmospheric Research (NCAR) Community Atmospheric Model (CAM) (Krasnopolsky et al., 2007). Present and future strategies feature an Earth System approach for assimilating environmental data into a more comprehensive coupled system including the atmosphere, ocean, biosphere and sea-ice (Penny and Hamill, 2017).

The influence and application of digital technologies has shown no sign of abatement in recent times. Three technological developments are having a strong effect on meteorology and climate research (Ruti et al., 2019). First, the increase of computing power. This is currently approaching exascale (i.e., 10^{18} operations per second), which is three orders of magnitude greater than the speed of current machines (Reed and Dongarra, 2015) and provides unprecedented opportunities with regard to the finer resolution of scales in time and space, and/or the coupling of more components that represent different parts of the Earth system. However, it also poses large software development and data management challenges, such as the impact of increasing numerical model resolution, increasing code complexity, and the volumes of data that are handled (Bauer et al., 2015; Sellar et al., 2020). A second development concerns the open availability of standard meteorological data and data from a variety of sources, including citizen science projects and low-cost sensors. Modern data management tools enable handling these data sources. Thirdly, there has been increasing use of machine learning, in particular so-called deep learning. A plethora of machine learning methods have been and are being applied to problems of weather and climate prediction, from emulating unresolved processes in numerical models to calibrating forecasts produced with numerical models and the production of forecasts based on data and machine learning methods only (Huntingford et al., 2019; Schneider et al., 2017; Reichstein et al., 2019).

Digital technologies enable new research methods, accelerate the growth of knowledge, and spur the creation of new means of communicating such knowledge amongst researchers and within the broader scientific community. As such, these technologies have reshaped the scientific enterprise and are strongly connected to open science (OECD, 2015; Bourne et al., 2012). Open science methodologies such as open access publications, open source software development and FAIR data (see below) stimulate the use of data and software resources and lead to more reproducible research (Wilkinson et al, 2016; Munafò et al., 2017). The need for open research practices has been recognized by the communities of meteorology and climate science. Nonetheless, whilst these domains are mature in terms of the application of digital technologies, the implementation of open science methodologies is less advanced.

In a session on "Weather and Climate Science in the Digital Era" at the 14th IEEE International eScience Conference, domain specialists and data and computer scientists discussed the road towards open weather and climate science. This paper describes the main findings and insights from this conference session.

The remainder of this paper is organized as follows: In the Methods section we describe the set-up of the conference session in detail, since the insights and claims in this paper are based on the observations made during the session. The Open Science section contains a small literature review which describes the progress of open weather and climate science in the context of open science developments in general. In the section Towards Open Weather and Climate Science we discuss the challenges and opportunities of open data and open software. The last section provides a synthesis of the issues that should be addressed in order to achieve open weather and climate science.

2   OPEN SCIENCE

Based on a small literature review, this section describes the progress of open weather and climate science in the context of open science developments in general.

Open science refers to open research practices, and includes but is not limited to public access to the academic literature, sharing of data and code (Mckiernan et al., 2016). However, the interpretation of the concept of open science varies between different schools of thought (Fecher and Friesike, 2014). In general, open science concerns various stakeholders: besides scholars, these include institutes, research funders, librarians and archivists, publishers and decision makers (Bourne et al., 2012; OECD, 2015; Fecher and Friesike, 2014).

It has been shown that the adoption of open research practices leads to significant benefits for researchers: specifically, increases in citations, media attention, potential collaborators, job opportunities and funding opportunities (Mckiernan et al., 2016). Europe and the United States have made efforts to adapt legal frameworks and implement policy initiatives for greater openness in scientific research (OECD, 2015; National Science Foundation, 2018). Several countries provide digital infrastructure based on rich metadata that support the optimal re-use of resources in the research environment (Mons et al., 2017). Examples include the European Open Science Cloud in Europe (Directorate-General for Research and Innovation, 2018), NIH Data Commons projects in the United States, AARnet in Australia (AARNet, 2018) and the African Data Intensive Research Cloud in South Africa (R. Simmonds et al., 2016). Funders and research institutes have announced policies encouraging, mandating, or specifically financing open research practices (Mckiernan et al., 2016; Wilkinson et al, 2016) - for example, the National Science Foundation in the United States (National Science Board, 2011), CERN in Switzerland (CERN-OPEN-2014-049, 2014), the Netherlands Organization for Scientific Research (executive board, 2019) and the United Nations Educational, Scientific and Cultural Organization (board, 2013).

The need for open research practices has been recognized by the communities of meteorology and climate science and has even entered into the political arena. For instance, in its report on the so-called "Climatic Research Unit email controversy" in 2009 the Science and Technology Committee of the UK House of Commons stated that climate science is a matter of great importance and that the quality of the science should be irreproachable. The committee called for the climate science community to become more transparent by publishing raw data and detailed methodologies (House of Commons, 2010).

There are many examples of open access, open data and open source software in meteorology and climate science. The United States has a long history of making meteorological observations, model source codes and model output an open public commodity, available to all. The WRF regional model, MPAS global model, and the CESM climate model (Skamarock et al., 2019; Hurrell, J.W. et al., 2013) are good examples of shared numerical weather and climate model codes. Output from NOAA weather and climate prediction models are freely available. The European Center for Medium-range Weather Forecasts (ECMWF) provides researchers with a free, and easy-to-use version of the Integrated Forecasting System (IFS), which is one of the main global NWP systems (Carver, 2019). It allows IFS to be used by a much wider community and the academic community contributes to improving the forecast model with new developments. The UK Earth System model (Sellar et al., 2019), a joint development between the National Environment Research Council (NERC) and the UK Met Office, has been made available to the research community in a similar fashion.In addition, co-ordinated coupled model intercomparison projects (CMIP) (Taylor et al., 2012; Eyring et al., 2016) are excellent examples of the climate modeling community working together. The construction of multi-model comparisons and statistics forces research groups to accept common input forcings, provide detailed documentation of the numerical schemes in their model and produce open, standardized output data (Sellar

et al., 2020, see e.g.). The result is a better understanding of climate change arising from natural, unforced variability or in response to changes in radiative forcing in a multi-model context.

The international meteorological and climate research communities have been sharing data since the 1990s, using common file and metadata formats. Besides CMIP (Taylor et al., 2012), examples include the sharing of reanalysis data, starting with NCEP/NCAR reanalysis and ECMWFs ERA reanalysis data products (Dee et al., 2011; Kalnay et al., 1996, e.g.).

The examples described above show that open research practices are growing in popularity and necessity. However, widespread adoption of these practices has not yet been achieved, which is also true for meteorology and climate science. In fact, sharing of data, software and vocabularies is only common practice in a few fields such as astronomy and genomics (Consortium, 2004; Borgman, 2012; Shamir et al., 2013, e.g.). Recent studies show that transparency and reproducibility are still a matter of concern to the scientific community as a whole. It requires that all stakeholders work together to create a more open and robust system (Baker, 2016; Munafò et al., 2017; Gil et al., 2016).

3 TOWARDS OPEN WEATHER AND CLIMATE SCIENCE

In the following section we present our perspective on the challenges and opportunities regarding open weather and climate science.

3.1 OPEN DATA

About 50% of the studies reported in the proceedings of the conference session include open data from different sources in their analyses. Examples include the use of open satellite data, geolocated data via OpenStreetMap and openly available in-situ meteorological observations (Haupt et al., 2018; Garcia-Marti et al., 2018; Bari, 2018; Schultz et al., 2018, and references therein). Two studies include data that are not common in meteorological or climate research. Citizen data such as social media posts (Brangbour et al., 2018) and observations from amateur weather stations (van Haren et al., 2018) can lead to new perspectives on local conditions beyond data from traditional meteorological stations.

At least 50% of the studies use common file formats and standard protocols to facilitate the exchange and use of data. Van den Oord et al. 2018 use CF-netCDF formats. The CF conventions provide guidelines for the use of metadata in the netCDF file and are increasingly used in climate studies. Behrens et al. 2018, Pelupessy et al. 2018, Schultz et al. 2018 and Stringer et al. (2018) all use standard protocols for inter-process communication (like MPI and REST) in their numerical codes. Furthermore, the use of common file formats and standard protocols is a prerequisite for the digital collaboration platforms which were presented in the session (Ramamurthy, 2018; Hut et al., 2018; Bendoukha, 2018).

The session participants recognized that in the current weather and climate science community the focus is primarily on making data and software findable and accessible, often via web portals. Although these are necessary first steps towards open science, we acknowledge that these steps are not sufficient. Data and software that are findable and accessible may still be hard to obtain in practice or may be disseminated in a way that it is still difficult to interpret and use. Wilkinson and colleagues (2016) defined guidelines to ensure the transparency, reproducibility, and reusability of scientific data. These state that data

- and also the algorithms, tools, and workflows that led to these data - should be Findable, Accessible, Interoperable and Reusable (FAIR). The FAIR guidelines put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.

In order to make the output from weather and climate models open and interoperable, i.e. formatted according to standards such as CF-netCDF,, including all necessary metadata, we consider performance scalability as the foremost technological challenge. Whereas the simulation models are predominantly run on large clusters using many compute nodes, subsequent processing and analysis of the output is often still confined to a single CPU and does not scale easily with (say) increased model resolution. Thus, producing FAIR model output via traditional post-processing pipelines is quickly becoming infeasible for advanced simulation models due to the sheer volume and complexity of their output.

For simulation models, this trend is a consequence of the advance of processor speed and model scalability compared to storage bandwidth, and can be countered with two strategies. The first is removing the need for post-processing by incorporating as many steps as possible within the application itself. This will make the model more expensive, especially in terms of memory usage, but the overhead may often be mitigated by offloading the post-processing to a small extra set of dedicated high-memory compute nodes. This approach requires a technical effort from the data providers in the community, and it can only solve the data problem to a limited extent, since there will always be extra manipulations required for many scientific analyses. Hence we need a second strategy on the data users' side to increase parallelism in the climate data processing toolchain. Existing cloud computing technologies, like Apache SPARK (Zaharia et al., 2016) or Dask (Team, 2016), may provide a suitable basis, since data processing and analysis pipelines can usually be represented by task graphs with a large degree of parallelism (over grid points, over multiple variables, over ensemble members, etc.). One of the key aspects, however, is the capability of the developer, usually a meteorologist or climate scientist, to adopt a new programming paradigm which facilitates the parallel execution of the workflow on cloud infrastructure. Here, research software engineers may play a key role by - for instance - developing higher-complexity algorithms for efficient processing of distributed climate data and adopting tools like xarray (Hoyer and Hamman, 2017) and Iris (Office, 2010).

In addition to these technological issues, we observe that some important challenges for open data arise from the political or legal context, and as such require additional efforts beyond the scientific domain. Weather services and commercial entities can see their data as a business advantage and be reluctant to make these open. Various resolutions by the World Meteorological Organisation (e.g. Resolution 40, 25 and 60) promote open access and exchange of data in order to better manage the risks from weather and climate-related hazards, but leave room for additional conditions. These resolutions have no legal status and national legislation may lead to restricted access to data and charges (Sylla, 2018). Also, policies to promote open data are less mature than those to promote open access to scientific publications (OECD, 2015). Another way to solve these issues is by signing nondisclosure agreements and allow the weather services to act as trusted agents who use the data for the public good without disclosing their details. These trusted agents should be considered as occupying a new role in the scientific process.

Furthermore, data need to be hosted and maintained, and their quality should be ensured. These requirements are well-addressed for large operational data services, such as the European Copernicus program, but this is not usually the case for research data of individual scientists, despite the increasing attention being paid to data management. Currently, data providers

have no clear policy (such as - for example - the FAIR principles) to follow in their hosting and management of data. Publications such as Geoscience Data Journal, Scientific Data and Earth System Data, are a partial remedy as these provide open access platforms where scientific data can be peer-reviewed and formally published. Some funding agencies - for example NWO in the Netherlands - are now requiring that, for all projects they fund, software becomes open source and the data are archived and findable unless there are strong reasons not to do so (e.g. privacy). Also, research funded by the European Commission should adhere to FAIR principles and data management plans need to be in place.

3.2 OPEN SOFTWARE

The conference session provided excellent examples of tools and approaches that were developed and made openly available to the research community. For example, approaches to reduce the computational or post processing costs of existing simulation models (Stringer et al., 2018; Behrens et al., 2018; van den Oord et al., 2018; Jansson et al., 2018) and approaches to integrate data sets from different sources (van Haren et al., 2018; Schultz et al., 2018). Four studies in the session presented an approach for which open data and software is a prerequisite, as these comprise a model coupling framework or a digital collaboration platform (Pelupessy et al., 2018; Jansson et al., 2018; Ramamurthy, 2018; Hut et al., 2018; Bendoukha, 2018).

We strongly support open publication of code, even if this code is under development, and especially when this code is used in a paper to support research findings. Open code can be inspected and reused by peers; this improves the reproducibility and quality of the corresponding research. Code sharing is crucial to science and to climate research in particular, since local and global policies depend on the scientific results. Open publication, however, requires the code to be documented and tested, which is a time-consuming effort. This level of documentation and testing is not yet standard practice, partially because there is no incentive to do so. There is a need for open science practices where incentives are developed to share scientific information beyond the final result in a scientific paper. Agile (Fowler and Highsmith, 2001) is a well-known approach in the software engineering community, and may provide a means to achieve open scientific software in a feasible way. According to the Agile approach, software is developed in small increments every few weeks, which makes it possible to provide continuous feedback to the developers. With its focus on flexibility and communication, Agile lends itself naturally to scientific software projects which are characterized by frequent code alterations due to changing requirements, tight collaboration in small teams, and short planning horizons (Sletholt et al., 2012). Agile practices are used, for example, by the ECMWF to develop the Climate Data Store (Raoult et al., 2017) and the Met Office Hadley Centre to develop climate models (Easterbrook and Johns, 2009).

In four studies that were presented in the conference, machine learning technologies are used for data analysis and prediction (Haupt et al., 2018; Garcia-Marti et al., 2018; Bari, 2018; Schultz et al., 2018). Besides using standard meteorological datasets, these studies employed additional data to infer relationships that are relevant to the end user. For example, prediction of solar power output over a future time period requires the inclusion of historical and real-time solar energy production data (Haupt et al., 2018). It was observed that the use of machine learning approaches in weather and climate science is increasing. These approaches are powerful, for instance, in emulating processes that are not resolved in simulation models (because of computational costs), in calibrating or post-processing simulation results and in building models to describe or forecast meteorological and climatological events. The caveats, on the other hand, are that trained models are not transparent as models

based on laws of physics and their results can be hard to interpret. Following the open science principle, machine learning approaches should be understandable and reusable by other researchers. Emerging fields like Explainable AI and knowledge based machine learning may provide approaches that help humans experts to understand how machine learning results are produced (Adadi and Berrada, 2018). Data-driven machine learning approaches should be combined with knowledge on physical processes (Dueben and Bauer, 2018; Reichstein et al., 2019) to gain further understanding of Earth system science problems. More broadly, machine learning methods should be accompanied by proper validation and verification.

This use of software, motivated by open science principles, requires a suitable digital infrastructure. The cloud appears to be a potential avenue as it enables individual researchers to gain access to high computing resources, vast amounts of storage and suites of software tools. In our session, three digital platforms were presented that use cloud technologies to create a virtual research environment in which scientific end-users can store, analyze and share their data (Ramamurthy, 2018; Hut et al., 2018; Bendoukha, 2018).The session participants also observed, however, that current platforms such as the Open Geospatial Consortium (D. Maidment et al., 2011) and JRC Earth Observation Data and Processing Platform (Soille et al., 2017), do not seem to increase the extent of scientific collaboration, particularly across disciplines. This may be partly due to the fact that these platforms have each implemented their own set of standards both for data formats and interfaces to access these data. Since scientists are required to invest time and effort in working with a specific platform, this heterogeneity can pose obstacles to their collaboration with researchers on another platform.

4   DISCUSSION

In this paper we describe the latest developments as well as future challenges and opportunities regarding open weather and climate science. We are basing our claims on the insights and observations made during the conference session on "Weather and Climate Science in the Digital Era". These observations are representative of what we are seeing in the field, although we recognize that our analysis is not complete. However, we believe that, given our experience, we have a solid view of the accomplishments of open science along with what still needs to be implemented.

The studies presented in the session show the value of sharing open data, and using and developing open source software and open platforms. Scientific advances are shown, for instance, through combining data sets and including non-standard meteorological data such as social media posts and observations from amateur weather stations.The increase in accuracy and skill of forecasts at local scales show improved consistency of data products and improved efficiency and skill of simulations, often crossing different disciplines. The utilisation of machine learning and increased computational capabilities have facilitated the use of disparate sources of data. In our conference session we concluded that sharing data and code offers many opportunities for scientific progress, leads to better reproducible science and vastly enhances the user base. However, we realized that open publication of data and code is not sufficient to achieve open weather and climate science and that there are important issues to address, which are described below.

The findability and accessibility of data increasingly receives attention in weather and climate research, and common file and metadata formats increase interoperability. However, for many data sets the implementation of the FAIR principles remains

a challenge due to their origin, scalability issues or legal barriers. We also acknowledge that data quality can be difficult to judge, depending as it does on its intended use, or the reason for its generation. Addressing this data quality challenge requires continued discussion on what aspects of open data can be implemented generically and what aspects are specific.

Technologically, the promise of using modern digital technologies is not always met due to the complexity of software platforms. While this paper does not address hardware, this is true for hardware and software-hardware interaction as well. A further development of platforms should facilitate the ease-of-use and provenance. This also calls for more attention to research software engineering where collaboration and interaction between software engineers and domain researchers can lead to optimal use of open science tools and methods.

As mentioned before, open science concerns various stakeholders in addition to scholars. Data management and programming have become an integral part of current research practice, and these activities require specific digital skills (Akhmerov et al., 2019). It is therefore important to acknowledge and define roles, responsibilities and mandates concerning data stewardship and research software engineering. This requires institutional change as the personnel portfolio of academic institutions needs to become more diverse, and in addition, a broader consideration of the impact of academic work beyond scientific publications and teaching.

In order to remove legal boundaries on sharing data, it is important to also engage non-academic parties such as operational and commercial meteorological institutions in open science. New policies regarding open science should be developed in an inclusive way to engage all stakeholders. Open science strategies and policies facilitate a higher quality of scientific research, increased collaboration, and engagement between research and society, which in turn can lead to higher social and economic impacts of public research (OECD, 2015).

5 CONCLUSION AND RECOMMENDATIONS

Alongside the issues and challenges regarding open weather and climate science, this paper also discusses opportunities and possible solutions for these issues. We have compiled these into the following list of concrete recommendations which will bring us closer to open weather and climate science: Regarding data:

- Developers should include post-processing steps in their simulation models. This requires additional compute and memory.

- Researchers using data from simulation models should increase parallelism in the data processing tool chain. This requires additional expertise in cloud computing, parallel and distributed computing.

- Individual researchers should be encouraged to publish scientific data in dedicated data journals.

Regarding software and infrastructure:

- Cloud technologies provide a suitable digital infrastructure for individual researchers to gain access to resources and tools and to collaborate with colleagues.

– Platforms for scientific collaboration should consider interoperability and user friendliness.

Regarding stakeholders and context:

– Nondisclosure agreements should be signed between scientific partners and weather services and the latter should be allowed to act as trusted agents. This requires including trusted agents as new roles in the scientific process and engaging them as stakeholders in new policies regarding open weather and climate science.

– Funders should request researchers to adhere to FAIR principles.

– All stakeholders should acknowledge and define roles, responsibilities and mandates concerning data stewardship and research software engineering. This requires both institutional change and a broader consideration of the impact of academic work.

Open science has implications for the stakeholders, the institutions and the system of science as a whole. It requires effort to change, but the benefits are large. Openly sharing data, code, and knowledge vastly enhances the user base, which means manifold growth of opportunities for new discoveries. As we observed from our conference session, this can lead to an improved understanding of our complex world.

References

AARNet: ANNUAL REPORT / 2018 DATA CONNECTOR FOR THE FUTURE, Tech. rep., Australia's Academic and Research Network, Chatswood, Australia, 2018.

Adadi, A. and Berrada, M.: Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), IEEE Access, 6, 52 138–52 160, https://doi.org/10.1109/ACCESS.2018.2870052, 2018.

Akhmerov, A., Cruz, M., Drost, N., Hof, C., Knapen, T., Kuzak, M., Martinez-Ortiz, C., Turkyilmaz-van der Velden, Y., and Van Werkhoven, B.: Raising the Profile of Research Software : Recommendations for Funding Agencies and Research Institutions, Tech. rep., Netherlands eScience Center, Amsterdam, Netherlands, https://doi.org/10.5281/zenodo.3378572, 2019.

Baker, M.: Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the 'crisis rocking science and what they think will help., Nature, 533, 452+, 2016.

Bari, D.: Visibility Prediction based on kilometric NWP Model Outputs using Machine-learning Regression, in: IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00048, 2018.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, https://doi.org/10.1038/nature14956, 2015.

Behrens, J., Biercamp, J., Bockelmann, H., and Neumann, P.: Increasing parallelism in climate models via additional component concurrency, in: IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00044, 2018.

Bendoukha, S.: Towards a new Big Data Analytics Platform for Climate Community, in: IEEE 14th International Conference on e-Science, 2018.

board, E.: Open Access Policy concerning UNESCO publications, Tech. rep., United Nations Educational, Scientific and Cultural Organization, 2013.

Borgman, C. L.: The conundrum of sharing research data, Journal of the American Society for Information Science and Technology, 63, 1059–1078, https://doi.org/https://doi.org/10.1002/asi.22634, 2012.

Bourne, P. E., Clark, T., de Ward, D. R., Herman, I., Hovy, E., and Shotton, D.: Force 11 White Paper: Improving the future of research communication and e-scholarship, Tech. rep., Force11, https://doi.org/10.4230/DagMan.1.1.41, https://www.force11.org/about/manifesto, 2012.

Brangbour, E., Bruneau, P., and Marchand-Maillet, S.: Extracting Flood Maps from Social Media for Assimilation, in: IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00045, 2018.

Carver, G.: The ECMWF OpenIFS numerical weather prediction model release cycle 40r1: description and use cases", in preparation to be submitted to GMDD, 2019.

CERN-OPEN-2014-049: Open Access Policy for CERN Physics Publication, Tech. rep., CERN, 2014.

Charney, J. G., FjÖrtoft, R., and Neumann, J. V.: Numerical Integration of the Barotropic Vorticity Equation, Tellus, 2, 237–254, https://doi.org/10.3402/tellusa.v2i4.8607, 1950.

Chevallier, F., Cheruy, F., Scott, N. A., and Chedin, A.: A neural network approach for a fast and accurate computation of a longwave radiative budget, Journal of Applied Meteorology, 37, 1385–1397, https://doi.org/10.1175/1520-0450(1998)037<1385:annafa>2.0.co;2, 1999.

Consortium, G. O.: The Gene Ontology (GO) database and informatics resource, Nucleic Acids Research, 32, 258–261, https://doi.org/10.1093/nar/gkh036, 2004.

340   D. Maidment, Domenico, B., Gemmell, A., Lehnert, K., Tarboton, D., and Zaslavsky, I.: The open geospatialconsortium and earthcube, Tech. rep., EarthCubeTechnology, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-Sanz,

345   B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Directorate-General for Research and Innovation: Prompting an EOSC in practice. Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC), Tech. rep., European Commission,

350   https://doi.org/10.2777/112658, 2018.

Dueben, P. D. and Bauer, P.: Challenges and design choices for global weather and climate models based on machine learning, Geoscientific Model Development, 11, 3999–4009, https://doi.org/10.5194/gmd-11-3999-2018, 2018.

Easterbrook, S. M. and Johns, T. C.: Engineering the software for understanding climate change, Computing in Science and Engineering, 11, 64–74, https://doi.org/10.1109/MCSE.2009.193, 2009.

355   executive board, N.: Connecting Science and Society - NWO strategy 2019-2022, Tech. rep., Netherlands Organisation for Scientific Research, 2019.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled MOdel Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization, Geoscientific Model Development, 9, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

360   Fecher, B. and Friesike, S.: Open Science: One Term, Five Schools of Thought, in: Opening Science, edited by Bartling, S. and Friesike, S., 1, pp. 1–7, The Author(s), https://doi.org/10.1007/978-3-319-00026-8_2, 2014.

Fowler, M. and Highsmith, J.: The agile manifesto, Software Development, 9, 28–35, 2001.

Garcia-Marti, I., Noteboom, J. W., and Diks, P.: Detecting probability of ice formation on overhead lines of the Dutch railway network, in: IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00050, 2018.

365   Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., Karlstrom, L., Lee, H., Mills, H. J., Oh, J. H., Pierce, S. A., Pope, A., Tzeng, M. W., Villamizar, S. R., and Yu, X.: Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance, Earth and Space Science, 3, 388–415, https://doi.org/10.1002/2015EA000136, 2016.

Haupt, S. E., Cowie, J., Linden, S., Mccandless, T., Kosovic, B., and Alessandrini, S.: Machine Learning for Applied Weather Prediction, in:

370   IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00047, 2018.

House of Commons: The disclosure of climate data from the Climatic Research Unit at the University of East Anglia, Science and Technology Committee, https://doi.org/citeulike-article-id:11615640, http://www.publications.parliament.uk/pa/cm200910/cmselect/cmsctech/387/387i.pdf, 2010.

Hoyer, S. and Hamman, J.: xarray: N-D labeled Arrays and Datasets in Python, Journal of Open Research Software, 5,

375   https://doi.org/10.5334/jors.148, 2017.

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., and Yang, H.: Machine learning and artificial intelligence to aid climate change research and preparedness, Environmental Research Letters, 14, 124 007, https://doi.org/10.1088/1748-9326/ab4e55, 2019.

Hurrell, J.W., Holland, M., Gent, P., Ghan, S., Kay, J., Kushner, P., Lamarque, J., Large, W., Lawrence, D., Lindsay, K., and Lipscomb,

380   W.: The Community Earth System Model: A framework for collaborative research, Bulletin of the American meteorological Society, 94, https://doi.org/10.1175/BAMS-D-12-00121.1., 2013.

Hut, R., Drost, N., van Hage, W., and van de Giesen, N.: eWaterCycle II, in: IEEE 14th International Conference on e-Science, 2018.

Jansson, F., van den Oord, G., Siebesma, P., and Crommelin, D.: Resolving clouds in a global atmosphere model - a multiscale approach with nested models, in: IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00043, 2018.

385   Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., and Zhu, Y.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American meteorological Society, 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Belochitski, A. A.: Accurate and fast neural network emulation of full, long-and short wave, model radiation used for decadal climate simulations with NCAR CAM, in: 19th conference on climate variability and change/fifth

390   conference on artificial intelligence applications to environmental science, 87th AMS Annual Meeting, 2007.

Lynch, P.: The origins of computer weather prediction and climate modeling, Journal of Computational Physics, 227, 3431–3444, https://doi.org/10.1016/j.jcp.2007.02.034, 2008.

Maury, M. F.: Explanations and Sailing Directions to Accompany the Wind and Current Charts, in: First International Maritime Conference Held for Devising an Uniform System of Meteorological Observations at Sea, Brussels, 1853.

395   Mckiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Mcdougall, D., Nosek, B. A., Ram, K., and Soderberg, C. K.: How open science helps researchers succeed, Elife, 5, 1–26, https://doi.org/10.7554/eLife.16800, 2016.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., Da Silva Santos, L. O. B., and Wilkinson, M. D.: Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud, Information Services and Use, 37, 49–56, https://doi.org/10.3233/ISU-170824, 2017.

400   Munafò, M. R., Nosek, B. A., Dorothy V. M. Bishop, K. S. B., Christopher D. Chambers, N. P. d. S., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., and Ioannidis, J. P. A.: A manifesto for reproducible science, Nature Human Behaviour, 1, https://doi.org/10.1038/s41562-016-0021, 2017.

National Science Board: Digital Research Data Sharing and Management, Tech. rep., National Science Foundation, 2011.

National Science Foundation: PROPOSAL & AWARD POLICIES AND PROCEDURES GUIDE (PAPPG), Tech. Rep. OMB Control Num-

405   ber 3145-0058, National Science Foundation, 2018.

OECD: Making open science a reality, OECD Science, Technology and Industry Policy Papers, p. 112, https://doi.org/http://dx.doi.org/10.1787/5jrs2f963zs1-en, 2015.

Office, U. M.: Iris: A Python library for analysing and visualising meteorological and oceanographic data sets, https://scitools.org.uk/iris/docs/latest/, 2010.

410   Pelupessy, I., Werkhoven, B. V., van den Oord, G., Zwart, S. P., van Elteren, A., and Dijkstra, H.: Development of the OMUSE / AMUSE modelling system, in: IEEE 14th International Conference on e-Science, 2018.

Penny, S. G. and Hamill, T. M.: Coupled Data Assimilation for Integrated Earth System Analysis and Prediction, Bulletin of the American Meteorological Society, 98, ES169–ES172, https://doi.org/10.1175/BAMS-D-17-0036.1, http://journals.ametsoc.org/doi/10.1175/BAMS-D-17-0036.1, 2017.

415 Quetelet, A.: Notice sur Le Capitaine M. F. Maury, in: Associé de l'Académie Royale de Belgique, published by the Academy, Brussels, 1874.

R. Simmonds, Taylor, R., Horrell, J., Fanaroff, B., Sithole, H., van Rensburg, S., and Al., E.: The African data intensive research cloud, in: IST - Africa Week Conference, 2016.

Ramamurthy, M.: Toward a Cloud Ecosystem for Modeling as a Service, in: IEEE 14th International Conference on e-Science, 420 https://doi.org/10.1109/eScience.2018.00046, 2018.

Raoult, B., Bergeron, C., Lopez Alos, A., Thepaut, J.-N., and Dee, D.: Climate service develops user-friendly data store, ECMWF Newsletter, pp. 22–27, https://doi.org/10.21957/p3c285, https://www.ecmwf.int/en/newsletter/151/meteorology/climate-service-develops-user-friendly-data-store, 2017.

Reed, D. A. and Dongarra, J.: Exascale computing and big data, Communications of the ACM, 58, 56–68, https://doi.org/10.1145/2699414, 425 2015.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Ruti, P., Tarasova, O., Keller, J., Carmichael, G., Hov, Ø., Jones, S., Terblanche, D., Anderson-Lefale, C., Barros, A., Bauer, P., Bouchet, 430 V., Brasseur, G., Brunet, G., DeCola, P., Dike, V., Kane, M. D., Gan, C., Gurney, K., Hamburg, S., Hazeleger, W., Jean, M., Johnston, D., Lewis, A., Li, P., Liang, X., Lucarini, V., Lynch, A., Manaenkova, E., Jae-Cheol, N., Ohtake, S., Pinardi, N., Polcher, J., Ritchie, E., Sakya, A. E., Saulo, C., Singhee, A., Sopaheluwakan, A., Steiner, A., Thorpe, A., and Yamaji, M.: Advancing Research for Seamless Earth System Prediction, Bulletin of the American Meteorological Society, pp. 23–35, https://doi.org/10.1175/bams-d-17-0302.1, 2019.

Schneider, T., Lan, S., Stuart, A., and Teixeira, J.: Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and 435 Targeted High-Resolution Simulations, Geophysical Research Letters, 44, 12,396–12,417, https://doi.org/10.1002/2017GL076101, 2017.

Schultz, M. G., Apweiler, S., Vogelsang, J., Kleinert, F., and Mallmann, D.: A web service architecture for objective station classification purposes, in: IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00051, 2018.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., 440 Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modeling Earth Systems, 11, 4513–4558, https://doi.org/10.1029/2019MS001739, 2019.

445 Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Abraham, N. L., Andrejczuk, M., Andrews, M. B., Andrews, T., Archibald, A. T., Mora, L., Dyson, H., Elkington, M., Ellis, R., Florek, P., Good, P., Gohar, L., Haddad, S., Hardiman, S. C., Hogan, E., Iwi, A., Jones, C. D., Johnson, B., Kelley, D. I., Kettleborough, J., Knight, J. R., Köhler, M. O., Kuhlbrodt, T., Liddicoat, S., Linova-Pavlova, I., Mizielinski, M. S., Morgenstern, O., Mulcahy, J., Neininger, E., O'Connor, F. M., Petrie, R., Ridley, J., Rioual, J., Roberts, M., Robertson, E., Rumbold, S.,

Seddon, J., Shepherd, H., Shim, S., Stephens, A., Teixeira, J. C., Tang, Y., Williams, J., and Wiltshire, A.: Implementation of UK Earth system models for CMIP6, Journal of Advances in Modeling Earth Systems, pp. 1–27, https://doi.org/10.1029/2019ms001946, 2020.

Shamir, L., Wallin, J. F., Allen, A., Berriman, B., Teuben, P., Robert J. Nemiroff, J. M., Hanisch, R. J., and DuPrie, K.: Practices in source code sharing in astrophysics., Astronomy and Computing, 1, 54–58., 2013.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., and Huang, X.-Y.: A Description of the Advanced Research WRF Version 4. NCAR Tech. Note NCAR/TN-556+STR, Tech. rep., NCAR, https://doi.org/10.5065/1dfh-6p97, 2019.

Sletholt, M. T., Hannay, J. E., Pfahl, D., and Langtangen, H. P.: What do we know about scientific software development's agile practices?, Computing in Science and Engineering, 14, 24–36, https://doi.org/10.1109/MCSE.2011.113, 2012.

Soille, P. &., Burger, A. &., Hasenohr, P. &., Kempeneers, Pieter & Rodriguez Aseretto, Dario & Syrris, V. &., Vasilev, V. &., and Marchi, D.: THE JRC EARTH OBSERVATION DATA AND PROCESSING PLATFORM, in: Big Data From Space, Toulouse, France, 2017.

Stringer, M., Jones, C., Hill, R., Dalvi, M., Johnson, C., and Walton, J.: A Hybrid-Resolution Earth System Model, in: IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00042, 2018.

Sylla, M. B.: Review of meteorological / climate data sharing policy (WMO Resolution 40) to promote their use to support Climate Information Services uptake in the African continent, in: Expert Group Meeting on data sharing policy in Africa, July, pp. 10–11, Dakar, Senegal, 2018.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Team, D. D.: Dask: Library for dynamic task scheduling, https://dask.org, 2016.

van den Oord, G., Yepes, X., and Acosta, M.: Post-processing strategies for the ECMWF model, in: IEEE 14th International Conference on e-Science, 2018.

van Haren, R., Koopmans, S., Steeneveld, G.-J., Theeuwes, N., Uijlenhoet, R., and Holtslag, A. A. M.: Weather reanalysis on an urban scale using WRF, in: IEEE 14th International Conference on e-Science, https://doi.org/10.1109/eScience.2018.00049, http://www2.mmm.ucar.edu/wrf/users/docs/arw{_}v3.pdf, 2018.

Wilkinson et al, M. D.: Comment: The FAIR Guiding Principles for scientific data management and stewardship, Scientific data, 3, https://doi.org/DOI: 10.1038/sdata.2016.18, 2016.

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., and Stoica, I.: Apache spark: A unified engine for big data processing, Communications of the ACM, 59, 56–65, https://doi.org/10.1145/2934664, 2016.