Geotechnical Data Management for Enhanced Accessibility and Reusability

Briefing Book

Keck Center of The National Academies of
Sciences, Engineering, and Medicine
500 5th Street, N.W.
Washington, DC 20001
Room 103

November 2nd, 2023

Geotechnical Data Management for Enhanced Accessibility and Reusability

Keck Center of the National Academies Building Information

Address
500 5th Street NW,
Washington, DC 20001

Effective March 20, 2023

All facilities of the National Academies of Sciences, Engineering, and Medicine are **open**.

Visitors must show a valid government ID (or a digital photo of the card) to the security staff at the Keck Center or the NAS Building, or to the management staff at the Beckman Center, when they enter the facility.

COVID-19 Updates and Guidance

As of March 20, 2023, COVID-19 proof of vaccination requirements are no longer in effect. Wearing a mask or respirator is not a requirement, but anyone who chooses to wear one should feel free and comfortable to do so.

Parking

Limited parking is available for meeting participants in the visitors parking area of the Keck building.

The public parking facilities closest to the Keck Center are Colonial Parking 6th & Pennsylvania Avenue, N.W., Diplomat 932 F St. & E Streets, N.W. and Carr Park 601 F St. Parking is available for meeting participants in Visitor's Parking on the P1 level of the garage.

GEOTECHNICAL DATA MANAGEMENT FOR ENHANCED ACCESSIBILITY AND REUSABILITY

Keck Center
500 5th ST NW
Washington, DC 20001
Room 103

Increasingly, technical decisions in many fields will rely on the capacity of computational systems to locate and analyze appropriate data. To optimize data use, data will need to be findable, accessible, interoperable, and reusable (FAIR). Public funds for data collection and research are provided increasingly with the caveat that resulting data are made available publicly. Various disciplines are embracing FAIR data principles when collecting and storing data to optimize their future use. This meeting of the Committee on Geological and Geotechnical Engineering (COGGE) will explore the principles of FAIR data, how state-of-the-art data management looks like in other fields, and how FAIR data management could benefit the geo-professional community. The goal of this meeting is to determine how COGGE and the National Academies of Sciences, Engineering, and Medicine can serve the broader geo-professional community to further practices in data management.

Full Agenda for Committee

November 2, 2023
ALL TIMES EASTERN
Join from PC, Mac, Linux, iOS or Android:
LINK: https://nasem.zoom.us/j/95601529229?pwd=NklWcFQ4QU1TZmF6M1paTUMya01vUT09
Password: 602133

8:30-9:30	**Closed Session** **Working breakfast, description of meeting objectives** • Review meeting objectives • Review open session agenda and committee member roles
9:30	**Open session: 9:30 – 3:00**
9:45	**Welcome** Allen Marr, COGGE chair Introductions, statement of objectives
10:00	**Keynote: What is FAIR data and who needs it?** *Philip Bourne, Dean of the School of Data Science, University of Virginia*
10:40	**Current geotechnical data management** *Scott Brandenberg, UCLA Samueli School of Engineering*

11:00	**Break**
11:15	**Panel discussion: How can we manage data to meet technical community needs? Data management panel discussion? What can FAIR data do for Geotech?** *Steve Diggs, University of California Office of the President* *Ge Peng, University of Alabama in Huntsville* *Mark Parsons, NASA*
12:30	**Working Lunch** Continued discussion
1:30	**Group discussion: How can COGGE help make progress on this topic?**
2:45	**Concluding remarks**
3:00	**Open session adjourns**

Speaker Biographies

Philip E. Bourne, PhD, FACMI is the Stephenson Founding Dean of the School of Data Science, Professor of Data Science and Biomedical Engineering at the University of Virginia, USA. Prior to that he was the Associate Director for Data Science (ADDS; aka Chief Data Scientist) for the US National Institutes of Health (NIH) and a Senior Investigator at the National Center for Biotechnology Information (NCBI). In his role as ADDS he led the trans NIH US $110M per year Big Data to Knowledge (BD2K) research initiative and contributed to data policies and infrastructure aimed at accelerating biomedical discovery. Examples include: establishing the NIH Commons, support for data and software citation and establishing preprints as a supported form of research. Prior to joining NIH, Dr. Bourne was Associate Vice Chancellor for Innovation and Industry Alliances in the Office of Research Affairs and a Professor in the School of Pharmacy and Pharmaceutical Sciences at the University of California San Diego (UCSD). His current research focuses on data science methods applied to systems pharmacology structural bioinformatics and scholarly communication. He has a strong interest in helping the next generation through the Ten Simple Rules series of professional development articles and his work as Dean of one of the few data science schools worldwide where new models of higher education are being emphasized.

Scott J. Brandenberg is a Professor in the Department of Civil and Environmental Engineering at UCLA. His research focuses on geotechnical earthquake engineering, geophysical imaging, data acquisition and signal processing, and numerical analysis with current research projects in developing fragility functions for bridges in liquefied and laterally spreading ground; developing design guidelines for pile foundations in liquefied ground; evaluation of the seismic levee deformation by destructive cyclic field testing; centrifuge modeling of the large-strain site response behavior of soft clays; CPT-based ultrasonic imaging of jet grout columns; and developing correlations between shear wave velocity and penetration resistance at Caltrans bridge sites. He is the 2022 recipient of the American Society of Civil Engineers Thomas a. Middlebrooks Award. Brandenberg earned his BS from Cal Poly, San Luis Obispo in 2000, and his MS and PhD from UC Davis in 2002 and 2005, respectively.

Steve Diggs is the University of California Curation Center (UC3)'s Senior Product Manager for the California Digital Library (CDL). Prior to his work at CDL, he served as the Technical Director of the Hydrographic Data Office at the Scripps Institute of Oceanography, and spent a decade working as a Department of Defense contractor. His career has evolved into design and maintenance of geoscience data resources. In addition to actively managing data, Steve is a member of numerous international marine data and data science teams including Argo and OceanSITES and is currently the chair of the data subcommittee for the Southern Ocean Observing System. Along with his deep experience and applying the same to new and ongoing innovations, Steve brings his years working as part of and with the research community in situ and in the lab and understands well the challenges and opportunities they face with managing their data and accessing that which will advance their own work.

Ge Peng is a Senior Principal Research Scientist at the Earth System Science Center/NASA MSFC IMPACT at the University of Alabama in Huntsville. Dr. Peng has over twenty years of technical experience in data analysis, ocean and atmosphere modeling including model development, implementation, application, process study, and validation. She specializes in applying statistical analysis and numerical models to both atmospheric and oceanic systems to examine seasonal, interannual, and decadal variability. Dr.

Peng has knowledge and experience in assimilating in situ and satellite measurements into models and assessing impact of data assimilation on model weather and climate predictions. She also has knowledge of NOAA scientific data appraisal and archive procedures and stewardship, national and international metadata standards, and life cycle of archiving scientific data and has been leading cutting-edge research on effective long-term scientific stewardship for Earth Science data products and evaluating satellite climate data products. Dr. Peng holds a Ph.D in Meteorology from Rosenstiel School of Marine & Atmospheric Science (RSMAS), University of Miami, with extensive training and working experience in Physical Oceanography.

Mark A. Parsons is a Research Scientist and geographer at the University of Alabama in Huntsville working to help align data, software, and information standards and processes across NASA's science divisions. Mark has more than 25 years of experience in researching and developing data stewardship policies, practices, and systems. He has repeatedly and effectively built dynamic, functional collaborations across all sorts of differences in language and professional cultures. Mark was the first Secretary General of the Research Data Alliance. He has helped coordinate stewardship of a broad range of data from satellite remote sensing to Indigenous knowledge of Arctic change. He led the data management effort for the International Polar Year and helped establish the Exchange for Local Observations and Knowledge of the Arctic (ELOKA). His published work has guided national data policies and practice and has contributed to educational programs. Mark lives in Colorado and likes to ride bicycles, bake bread, and play outside.

National Academies of Sciences, Engineering, and Medicine
Division on Earth and Life Studies
Board on Earth Sciences and Resources
Committee on Geological and Geotechnical Engineering

Advancing Geotechnical Data Access, Curation, and Use

Geotechnical and Geological engineers collect numerous types of geospatially related data during site characterization, infrastructure maintenance and management, and post-event response. Other information and data are generated as a result of modeling and site simulation. The volumes of data collected will increase as technology driven advances in characterization and testing approaches and data science capabilities improve. Some standards exist within different industries regarding data collection, but few data management decisions are being driven by long-term data-related needs within and across geo-related industries, and the value propositions for making data findable, accessible, interoperable, and reusable (i.e., FAIR data; Wilkinson, et al., 2016) as well as ethical and of high quality are not understood. A new study is proposed to consider the value propositions and crucial issues associated with long-term data curation and analyses that need to be addressed within the geo-related industries to optimize the sustainability, usability, and utility of geo-related data. The study will incorporate knowledge and experience from other disciplines that are more mature in their approaches to modern data maintenance and analytical techniques.

Statement of Task

An ad hoc committee of the National Academies of Sciences, Engineering, and Medicine will conduct a study and solicit input from data scientists, informaticians, and geo-professional experts to critically examine models for management of findable, accessible, interoperable, and reusable (FAIR) geospatially related engineering data. The study committee consider

- Advances in and tools for data collection, management, and analyses that allow for the creation of sustainable and FAIR datasets and repositories and the ways in which the geo-professional community could benefit from those advances;
- Lessons learned from other technical communities regarding the sustainable management and integration of large datasets; and
- Factors critical for successful accessioning of geo-related FAIR, high-quality, and ethical data and repositories.

The committee will describe assumptions built into successful data collection and modeling processes the benefits of pursuing FAIR data management, and will identify potential technological disruptors and future developments that could affect future data needs and accessibility in a 5- to 10-year horizon. The committee's report will identify elements of a roadmap for integrating the knowledge from other technical communities that manage large data collections for research and decision making and will recommend next steps for creating broad access to high-quality FAIR data for the geo-professions.

Reference

Wilkinson, M.D., M. Dumontier, I. Jan Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018.

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

August 25, 2022

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Dr. Alondra Nelson
Deputy Assistant to the President and Deputy Director for Science and Society
Performing the Duties of Director
Office of Science and Technology Policy (OSTP)

SUBJECT: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

This memorandum provides policy guidance to federal agencies with research and development expenditures on updating their public access policies. In accordance with this memorandum, OSTP recommends that federal agencies, to the extent consistent with applicable law:

1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.

1. Background and Policy Principles

Since February 2013, federal public access policy has been guided by the *Memorandum on Increasing Access to the Results of Federally Funded Research* (2013 Memorandum).[1] Issued by the White House Office of Science and Technology Policy (OSTP), the 2013 Memorandum directed all federal departments and agencies (agencies) with more than $100 million in annual research and development expenditures to develop a plan to support increased public access to the results of federally funded research, with specific focus on access to scholarly publications and digital data resulting from such research.

Nearly ten years later, every federal agency subject to the 2013 Memorandum has developed and implemented a public access policy in accordance with its guidance.[2] As a result, the American public has experienced great benefits: more than 8 million scholarly publications have become accessible to the public. Over 3 million people read these articles for free every day. The 2013 federal public access policy set the stage for a paradigm shift away from research silos and

[1] See the 2013 Memorandum: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
[2] See the 2021 OSTP Public Access Congressional Report: https://www.whitehouse.gov/wp-content/uploads/2022/02/2021-Public-Access-Congressional-Report_OSTP.pdf

toward a scientific culture that values collaboration and data sharing. The 2013 Memorandum helped to reshape the landscape for data and research by sharing results freely and openly with the public and the scientific community.

Building on these important advances, the policy guidance laid out in the 2013 Memorandum can be improved to achieve delivery of federally funded research results and data to *all* of America. Years of public feedback have indicated that the primary limitation of the 2013 Memorandum is the optional 12-month embargo from public access of any publication resulting from federally funded research. This provision has limited immediate access of federally funded research results to only those able to pay for it or who have privileged access through libraries or other institutions. Financial means and privileged access must never be the pre-requisites to realizing the benefits of federally funded research that the American public deserves.

A federal public access policy consistent with our values of equal opportunity must allow for broad and expeditious sharing of federally funded research—and must allow all Americans to benefit from the returns on our research and development investments without delay. Upholding these core U.S. principles in our public access policy also strengthens our ability to be a critical leader and partner on issues of open science around the world. The U.S. is committed to the ideas that openness in science is fundamental, security is essential, and freedom and integrity are crucial.[3] Improving public access policies across the U.S. government to promote the rapid sharing of federally funded research data with appropriate protections and accountability measures will allow for greater validity of research results and more equitable access to data resources aligned with these ideals. To promote equity and advance the work of restoring the public's trust in Government science, and to advance American scientific leadership, now is the time to amend federal policy to deliver immediate public access to federally funded research.

2. Learning from the Lessons of COVID-19

When federally funded research is available to the public, it can improve lives, provide policymakers with important evidence with which to make critical decisions, accelerate the rates of discovery and translation, and drive more equitable outcomes across every sector of society.

Americans were offered a window into the great benefits of immediate public access to federally funded research at the outset of the COVID-19 pandemic. In the wake of the public health crisis, government, industry, and scientists voluntarily worked together to adopt an immediate public access policy, which yielded powerful results: research and data flowed effectively, new accessible insights super-charged the rate of discovery, and translation of science soared. The shift in practice during COVID-19 demonstrated how delivering immediate public access to federally funded research publications and data can provide near real-time returns on American taxpayer investments in science and technology.

Immediate public access to COVID-19 research is a powerful case study on the benefits of delivering research results and data rapidly to the people. The insights of new and cutting-edge research stemming from the support of federal agencies should be immediately available—not

[3] See: https://www.whitehouse.gov/ostp/news-updates/2022/06/21/readout-of-dr-alondra-nelsons-participation-in-the-g7-science-ministerial-progress-toward-a-more-open-and-equitable-world/

just in moments of crisis, but in every moment. Not only to fight a pandemic, but to advance all areas of study, including urgent issues such as cancer, clean energy, economic disparities, and climate change. American investment in such research is essential to the health, economic prosperity, and well-being of the Nation. There should be no delay between taxpayers and the returns on their investments in research.

3. Updates to Policy Guidance on Increasing Equitable Access to Federally Funded Research Results

To meet these core commitments, OSTP is updating policy guidance to promote improved public access to federally funded research results. **In accordance with the provisions listed in Section 3, Federal agencies should develop new, or update existing, public access plans as soon as possible, and submit them to OSTP and the Office of Management and Budget (OMB) no later than:**

(1) 180 days after the date of this memorandum for federal agencies *with more* than $100 million in annual research and development (R&D) expenditures; and
(2) 360 days after the date of this memorandum for federal agencies with $100 million *or le*ss in annual R&D expenditures. This extended deadline is designed to accommodate a longer lead time for federal agencies who were not subject to the 2013 Memorandum.

Agencies should complete and publish full policy development for plans implementing provisions in Section 3 by December 31st, 2024, with an effective date no later than one year after the publication of the agency plan. The timeline is designed to accommodate the items identified in Section 5 of this memorandum, including interagency collaboration, public engagement with those impacted by the change in policy, and OSTP feedback on agency drafts.

a) **Peer Reviewed Scholarly Publications**:
Federal agencies should update or develop new public access plans for ensuring, as appropriate and consistent with applicable law, that all peer-reviewed scholarly publications[4] authored or co-authored by individuals or institutions resulting from federally funded research **are made freely available and publicly accessible by default in agency-designated repositories without any embargo or delay after publication**.

Plans should describe:

i. How peer-reviewed scholarly publications should be made publicly accessible;
ii. How to maximize equitable reach of public access to peer-reviewed scholarly publications, including by providing free online access to peer-reviewed scholarly

[4] Such scholarly publications always include peer-reviewed research articles or final manuscripts published in scholarly journals, and may include peer-reviewed book chapters, editorials, and peer-reviewed conference proceedings published in other scholarly outlets that result from federally funded research.

publications in formats that allow for machine-readability[5] and enabling broad accessibility through assistive devices; and,

iii. The circumstances or prerequisites needed to make the publications freely and publicly available by default, including any use and re-use rights, and which restrictions, including attribution, may apply.

b) **Scientific Data**

i. Scientific data[6] underlying peer-reviewed scholarly publications resulting from federally funded research should be made freely available and publicly accessible by default at the time of publication, unless subject to limitations as described in Section 3(c)(i) and should be subject to federal agency guidelines for researcher responsibilities regarding data management and sharing plans, consistent with Section 3(c) of this memorandum.

ii. Federal agencies should develop approaches and timelines for sharing other federally funded scientific data that are not associated with peer-reviewed scholarly publications.

iii. Federal agencies should also provide guidance to researchers that ensures the digital repositories used align, to the extent practicable,[7] with the National Science and Technology Council document entitled "Desirable Characteristics of Data Repositories for Federally Funded Research."[8]

iv. Federal agency research: Agency public access plans and policies should clarify that federal researchers must follow federal laws and OMB policies that govern federal agencies' information management practices and protect certain types of data,[9] to the extent that the scientific data created by, collected by, under the control or direction of, or maintained by the federal researchers is subject to those laws and policies.

---

[5] "Machine readability" refers to a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost (such as the NISO Z39.96-2015 JATS XML standard currently used by PubMed Central).

[6] For the purposes of this memorandum, "scientific data" include the recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings. Such scientific data do not include laboratory notebooks, preliminary analyses, case report forms, drafts of scientific papers, plans for future research, peer-reviews, communications with colleagues, or physical objects and materials, such as laboratory specimens, artifacts, or field notes. The definition of "scientific data" is similar to but broader than the term "research data" defined by 2 CFR 200.315 (e) and 45 CFR 75.322 (e).

[7] The term "extent practicable" is used to signal that suitable repositories for all types of data may not be available within the timeframe provided.

[8] See the 2022 NSTC Subcommittee on Open Science guidance: https://doi.org/10.5479/10088/113528

[9] For instance, the Paperwork Reduction Act, E-Government Act, Freedom of Information Act, Federal Information Security Management Act, Privacy Act, Health Information Technology for Economic and Clinical Health Act, Information Quality Act, Foundations for Evidence-Based Policymaking Act, Confidential Information Protection and Statistical Efficiency Act, Federal Policy for the Protection of Human Subjects, Federal Records Act, and OMB guidance under OMB M-13-13 and subsequent open data policies (e.g., those to be promulgated under the -OPEN Government Data Act and Pub. L. No. 115-435), OMB Circular A-130, and other laws and policies that require federal agencies to protect trade secrets, confidential commercial information, personally identifiable information, and other information which is protected under law or policy. See also, language from OMB M-19-15 with respect to maximizing the amount of data that can be made public using cutting-edge technologies to provide secure access to confidential data while reducing the risk of re-identification.

c) Public access plans should outline the policies that federal agencies will use to establish researcher responsibilities on how federally funded scientific data will be managed and shared, including:
    i) Details describing any potential legal, privacy, ethical, technical, intellectual property, or security limitations, [10] and/or any other potential restrictions or limitations on data access, use, and disclosure, including those defined in terms and conditions of funding agreement or award or that convey from a data use agreement or stipulations of an Institutional Review Board;
    ii) Plans to maximize appropriate[11] sharing of the federally funded scientific data identified in Section 3(a) of this memorandum, such as providing risk-mitigated opportunities for limited data access;[12] and,
    iii) The specific online digital repository or repositories where the researcher expects to deposit their relevant data, consistent with the federal agency's guidelines.
d) In consultation with OMB, federal agencies should allow researchers to include reasonable publication costs and costs associated with submission, curation, management of data, and special handling instructions as allowable expenses in all research budgets.
e) Federal agencies should report to OSTP, when requested, on the status of their public access plans and policy implementation, including the number of all scholarly publications funded by the federal agencies and any other relevant statistics collected by the agency.

4. Ensuring Scientific and Research Integrity in Agency Public Access Policies

Public access policies that deliver transparent, open, secure, and free communication of federally funded research and activities in an expeditious manner are an important tool to uphold scientific[13] and research[14] integrity. Federal agencies should take steps to ensure that public access policies support scientific and research integrity by transparently communicating to the public critical information, including that which is related to the authorship, funding, affiliations, and development status of federally funded research. The public should be able to identify which federal agencies support given investments in science, the scientists who conduct that research, and the extent to which peer-review was conducted. These actions support the value that maintaining and restoring public trust in science requires openness, security, freedom, and integrity. **Federal agencies should take actions to ensure that these elements of scientific and**

[10] Including national security concerns.
[11] The term "appropriate" is used to signal that public access to federally funded research results and data should be maximized in a manner that protects confidentiality, privacy, business confidential information, and security, avoids negative impact on intellectual property rights, innovation, program and operational improvements, and U.S. competitiveness, and preserves the balance between the relative value of long-term preservation and access and the associated cost and administrative burden.
[12] For example, secure research data centers, data use agreements, perturbing identifiable information, or excluding sensitive variables.
[13] See the 2022 NSTC Report "Protecting the Integrity of Government Science": https://www.whitehouse.gov/wp-content/uploads/2022/01/01-22-Protecting_the_Integrity_of_Government_Science.pdf
[14] See the 2022 NSTC "Guidance for Implementing National Security Presidential Memorandum 33 (NSPM-33) on National Security Strategy for United States Government-Supported Research and Development" (NSPM-33 Implementation Guidance): https://www.whitehouse.gov/wp-content/uploads/2022/01/010422-NSPM-33-Implementation-Guidance.pdf

research integrity are in place in order to strengthen public trust in federally funded science.

To achieve these goals, the following steps should be taken by federal agencies, as appropriate and consistent with their missions. By December 31st, 2024, federal agencies should submit to OSTP and OMB a second update to their public access plans specifying approaches taken to implement the provisions in this Section 4. Agencies should complete and publish full policy development for plans implementing these provisions by December 31st, 2026, with an effective date no later than one year after the publication of the agency plan. Federal agencies should, consistent with applicable law:

a) Collect and make publicly available appropriate metadata[15] associated with scholarly publications and data resulting from federally funded research, to the extent possible at the time of deposit in a public access repository. Such metadata should include at minimum:
   i) all author and co-author names, affiliations, and sources of funding, referencing digital persistent identifiers,[16] as appropriate;
   ii) the date of publication; and,
   iii) a unique digital persistent identifier for the research output;
b) Instruct federally funded researchers to obtain a digital persistent identifier that meets the common/core standards of a digital persistent identifier service defined in the NSPM-33 Implementation Guidance,[17] include it in published research outputs when available, and provide federal agencies with the metadata associated with all published research outputs they produce, consistent with the law, privacy, and security considerations.
c) Assign unique digital persistent identifiers[18] to all scientific research and development awards[19] and intramural research protocols that have appropriate metadata linking the funding agency and their awardees through their digital persistent identifiers.

5. Public Access Plan Coordination Among Federal Agencies

Coordination among federal science agencies[20] is critical for the success of delivering America's research to the public. The National Science and Technology Council Subcommittee on Open Science was chartered to facilitate such coordination between federal science agencies in conjunction with OSTP. Concurrent with and following the development of agency plans described Section 3 and Section 4 of this memorandum, the Subcommittee on Open Science will:

[15] For the purposes of this memorandum, metadata include information conveyed with the publications and data upon deposit in a public access repository to ensure proper attribution and versioning.
[16] See the NSPM-33 Implementation Guidance for definition: A digital identifier that is globally unique, persistent, machine resolvable and processable, and has an associated metadata schema.
[17] See Point 5 in the Digital Persistent Identifiers section of the NSPM-33 Implementation Guidance
[18] As a complement to implementation of the Federal Funding and Accountability Transparency Act
[19] Consistent with NSPM-33 Implementation Guidance, a research and development award refers to support provided to an individual or entity by a federal research agency to carry out research and development activities, which may include support in the form of a grant, contract, cooperative agreement, or other such transaction.
[20] Federal science agencies here are defined as any federal agency with an annual extramural research expenditure of over $100,000,000 per 42 USC § 6623(f).

a) coordinate between federal science agencies to enhance efficiency and reduce redundancy in public access plans and policies, including as it relates to digital repository access;
b) improve awareness of federally funded research results by all potential users and communities;
c) consider measures to reduce inequities in publishing of, and access to, federally funded research and data, especially among individuals from underserved backgrounds and those who are early in their careers;
d) develop procedures and practices to reduce the burden on federally funded researchers in complying with public access requirements;
e) recommend standard consistent benchmarks and metrics to monitor and assess implementation and iterative improvement of public access policies over time;
f) improve monitoring and encourage compliance with public access policies and plans;
g) coordinate engagement with stakeholders, including but not limited to publishers, libraries, museums, professional societies, researchers, and other interested non-governmental parties on federal agency public access efforts;
h) develop guidance on desirable characteristics of, and best practices for sharing in, online digital publication repositories;
i) identify the key parameters that must be considered in planning how to maximize appropriate sharing of federally funded scientific data that have not been used to support scholarly publications; and,
j) develop strategies to make federally funded publications, data, and other such research outputs and their metadata are findable, accessible, interoperable, and re-useable, to the American public and the scientific community in an equitable and secure manner.

6. General Provisions

Nothing in this memorandum shall be construed to impair or otherwise affect authority granted by law to an executive department, agency, or the head thereof; or functions of the Director of OMB.

Nothing in this memorandum, or the agency plans developed pursuant to it, shall be construed to authorize or require federal agencies to undermine any right under the provisions of Title 17, 18, or 35 of the United States Code, or to violate the international obligations of the United States.

Provisions of this memorandum should be implemented to the extent feasible and consistent with applicable law, privacy, indigenous rights, foreign policy and international development objectives, and national security considerations. Any provisions of the 2013 Memorandum that are not updated or superseded by this new policy guidance are maintained. Provisions of this memorandum should be implemented consistent with law, OMB Guidance, and the Uniform Guidance 2 CFR 200.

This memorandum is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity, by any party against the United States; its departments, agencies, or entities; its officers, employees, or agents; or any other person.

7. Taking Next Steps Together

The extraordinary progress in open science and public access led by federal agencies has laid the foundation for these critical next steps. As we move forward together in implementing these critical actions, we will do so in partnership and with a shared vision for an ever-stronger and more equitable federal scientific ecosystem.

Immediate public access to America's research publications and data will serve our collective goals of accelerating scientific discovery, strengthening translation and policymaking, and lowering the barriers of access to science for all of America.

As we move forward, OSTP will establish a process for supporting the implementation of these updates. We are grateful to you and your dedicated staff for your valued contributions to strengthening public access and supporting the advancement of health, safety, security, and equity.

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects[2]—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

There are numerous and diverse stakeholders who stand to benefit from overcoming these obstacles: researchers wanting to share, get credit, and reuse each other's data and interpretations; professional data publishers offering their services; software and tool-builders providing data analysis and processing services such as reusable workflows; funding agencies (private and public) increasingly

Correspondence and requests for materials should be addressed to B.M. (email: barend.mons@dtls.nl).
#A full list of authors and their affiliations appears at the end of the paper.

Box 1 | Terms and Abbreviations

BD2K—Big Data 2 Knowledge, is a trans-NIH initiative established to enable biomedical research as a digital research enterprise, to facilitate discovery and support new knowledge, and to maximise community engagement.

DOI—Digital Object Identifier; a code used to permanently and stably identify (usually digital) objects. DOIs provide a standard mechanism for retrieval of metadata about the object, and generally a means to access the data object itself.

FAIR—Findable, Accessible, Interoperable, Reusable.

FORCE11—The Future of Research Communications and e-Scholarship; a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing, initiated in 2011.

Interoperability—the ability of data or tools from non-cooperating resources to integrate or work together with minimal effort.

JDDCP—Joint Declaration of Data Citation Principles; Acknowledging data as a first-class research output, and to support good research practices around data re-use, JDDCP proposes a set of guiding principles for citation of data within scholarly literature, another dataset, or any other research object.

RDF—Resource Description Framework; a globally-accepted framework for data and knowledge representation that is intended to be read and interpreted by machines.

concerned with long-term data stewardship; and a data science community mining, integrating and analysing new and existing data to advance discovery. To facilitate the reading of this manuscript by these diverse stakeholders, we provide definitions for common abbreviations in Box 1. Humans, however, are not the only critical stakeholders in the milieu of scientific data. Similar problems are encountered by the applications and computational agents that we task to undertake data retrieval and analysis on our behalf. These 'computational stakeholders' are increasingly relevant, and demand as much, or more, attention as their importance grows. One of the grand challenges of data-intensive science, therefore, is to improve knowledge discovery through assisting both humans, and their computational agents, in the discovery of, access to, and integration and analysis of, task-appropriate scientific data and other scholarly digital objects.

For certain types of important digital objects, there are well-curated, deeply-integrated, special-purpose repositories such as Genbank[3], Worldwide Protein Data Bank (wwPDB[4]), and UniProt[5] in the life sciences; Space Physics Data Facility (SPDF; http://spdf.gsfc.nasa.gov/) and Set of Identifications, Measurements and Bibliography for Astronomical Data (SIMBAD[6]) in the space sciences. These foundational and critical core resources are continuously curating and capturing high-value reference datasets and fine-tuning them to enhance scholarly output, provide support for both human and mechanical users, and provide extensive tooling to access their content in rich, dynamic ways. However, not all datasets or even data types can be captured by, or submitted to, these repositories. Many important datasets emerging from traditional, low-throughput bench science don't fit in the data models of these special-purpose repositories, yet these datasets are no less important with respect to integrative research, reproducibility, and reuse in general. Apparently in response to this, we see the emergence of numerous general-purpose data repositories, at scales ranging from institutional (for example, a single university), to open globally-scoped repositories such as Dataverse[7], FigShare (http://figshare.com), Dryad[8], Mendeley Data (https://data.mendeley.com/), Zenodo (http://zenodo.org/), DataHub (http://datahub.io), DANS (http://www.dans.knaw.nl/), and EUDat[9]. Such repositories accept a wide range of data types in a wide variety of formats, generally do not attempt to integrate or harmonize the deposited data, and place few restrictions (or requirements) on the descriptors of the data deposition. The resulting data ecosystem, therefore, appears to be moving away from centralization, is becoming more diverse, and less integrated, thereby exacerbating the discovery and re-usability problem for both human and computational stakeholders.

A specific example of these obstacles could be imagined in the domain of gene regulation and expression analysis. Suppose a researcher has generated a dataset of differentially-selected polyadenylation sites in a non-model pathogenic organism grown under a variety of environmental conditions that stimulate its pathogenic state. The researcher is interested in comparing the alternatively-polyadenylated genes in this local dataset, to other examples of alternative-polyadenylation, and the expression levels of these genes—both in this organism and related model organisms—during the infection process. Given that there is no special-purpose archive for differential polyadenylation data, and no model organism database for this pathogen, where does the researcher begin?

We will consider the current approach to this problem from a variety of data discovery and integration perspectives. If the desired datasets existed, where might they have been published, and how would one begin to search for them, using what search tools? The desired search would need to filter based on specific species, specific tissues, specific types of data (Poly-A, microarray, NGS), specific conditions (infection), and specific genes—is that information ('metadata') captured by the repositories, and if so, what formats is it in, is it searchable, and how? Once the data is discovered, can it be downloaded? In what format(s)? Can that format be easily integrated with private in-house data (the local dataset of alternative polyadenylation sites) as well as other data publications from third-parties and with the community's core gene/protein data repositories? Can this integration be

done automatically to save time and avoid copy/paste errors? Does the researcher have permission to use the data from these third-party researchers, under what license conditions, and who should be cited if a data-point is re-used?

Questions such as these highlight some of the barriers to data discovery and reuse, not only for humans, but even more so for machines; yet it is precisely these kinds of deeply and broadly integrative analyses that constitute the bulk of contemporary e-Science. The reason that we often need several weeks (or months) of specialist technical effort to gather the data necessary to answer such research questions is not the lack of appropriate technology; the reason is, that we do not pay our valuable digital objects the careful attention they deserve when we create and preserve them. Overcoming these barriers, therefore, necessitates that all stakeholders—including researchers, special-purpose, and general-purpose repositories—evolve to meet the emergent challenges described above. The goal is for scholarly digital objects of all kinds to become 'first class citizens' in the scientific publication ecosystem, where the quality of the publication—and more importantly, the impact of the publication—is a function of its ability to be accurately and appropriately found, re-used, and cited over time, by all stakeholders, both human and mechanical.

With this goal in-mind, a workshop was held in Leiden, Netherlands, in 2014, named 'Jointly Designing a Data Fairport'. This workshop brought together a wide group of academic and private stakeholders all of whom had an interest in overcoming data discovery and reuse obstacles. From the deliberations at the workshop the notion emerged that, through the definition of, and widespread support for, a minimal set of community-agreed guiding principles and practices, all stakeholders could more easily discover, access, appropriately integrate and re-use, and adequately cite, the vast quantities of information being generated by contemporary data-intensive science. The meeting concluded with a draft formulation of a set of foundational principles that were subsequently elaborated in greater detail—namely, that all research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people. These are now referred to as the FAIR Guiding Principles. Subsequently, a dedicated FAIR working group, established by several members of the FORCE11 community[10] fine-tuned and improved the Principles. The results of these efforts are reported here.

The significance of machines in data-rich research environments

The emphasis placed on FAIRness being applied to both human-driven and machine-driven activities, is a specific focus of the FAIR Guiding Principles that distinguishes them from many peer initiatives (discussed in the subsequent section). Humans and machines often face distinct barriers when attempting to find and process data on the Web. Humans have an intuitive sense of 'semantics' (the meaning or intent of a digital object) because we are capable of identifying and interpreting a wide variety of contextual cues, whether those take the form of structural/visual/iconic cues in the layout of a Web page, or the content of narrative notes. As such, we are less likely to make errors in the selection of appropriate data or other digital objects, although humans will face similar difficulties if sufficient contextual metadata is lacking. The primary limitation of humans, however, is that we are unable to operate at the scope, scale, and speed necessitated by the scale of contemporary scientific data and complexity of e-Science. It is for this reason that humans increasingly rely on computational agents to undertake discovery and integration tasks on their behalf. This necessitates machines to be capable of autonomously and appropriately acting when faced with the wide range of types, formats, and access-mechanisms/protocols that will be encountered during their self-guided exploration of the global data ecosystem. It also necessitates that the machines keep an exquisite record of provenance such that the data they are collecting can be accurately and adequately cited. Assisting these agents, therefore, is a critical consideration for all participants in the data management and stewardship process—from researchers and data producers to data repository hosts.

Throughout this paper, we use the phrase 'machine actionable' to indicate a continuum of possible states wherein a digital object provides increasingly more detailed information to an autonomously-acting, computational data explorer. This information enables the agent—to a degree dependent on the amount of detail provided—to have the capacity, when faced with a digital object never encountered before, to: a) identify the type of object (with respect to both structure and intent), b) determine if it is useful within the context of the agent's current task by interrogating metadata and/or data elements, c) determine if it is usable, with respect to license, consent, or other accessibility or use constraints, and d) take appropriate action, in much the same manner that a human would.

For example, a machine may be capable of determining the data-type of a discovered digital object, but not capable of parsing it due to it being in an unknown format; or it may be capable of processing the contained data, but not capable of determining the licensing requirements related to the retrieval and/or use of that data. The optimal state—where machines fully 'understand' and can autonomously and correctly operate-on a digital object—may rarely be achieved. Nevertheless, the FAIR principles provide 'steps along a path' toward machine-actionability; adopting, in whole or in part, the FAIR

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

principles, leads the resource along the continuum towards this optimal state. In addition, the idea of being machine-actionable applies in two contexts—first, when referring to the contextual metadata surrounding a digital object ('what is it?'), and second, when referring to the content of the digital object itself ('how do I process it/integrate it?'). Either, or both of these may be machine-actionable, and each forms its own continuum of actionability.

Finally, we wish to draw a distinction between data that is machine-actionable as a result of specific investment in software supporting that data-type, for example, bespoke parsers that understand life science wwPDB files or space science Space Physics Archive Search and Extract (SPASE) files, and data that is machine-actionable exclusively through the utilization of general-purpose, open technologies. To reiterate the earlier point—ultimate machine-actionability occurs when a machine can make a useful decision regarding data that it has not encountered before. This distinction is important when considering both (a) the rapidly growing and evolving data environment, with new technologies and new, more complex data-types continuously being developed, and (b) the growth of general-purpose repositories, where the data-types likely to be encountered by an agent are unpredictable. Creating bespoke parsers, in all computer languages, for all data-types and all analytical tools that require those data-types, is not a sustainable activity. As such, the focus on assisting machines in their discovery and exploration of data through application of more generalized interoperability technologies and standards at the data/repository level, becomes a first-priority for good data stewardship.

The FAIR Guiding Principles in detail

Representatives of the interested stakeholder-groups, discussed above, coalesced around four core desiderata—the FAIR Guiding Principles—and limited elaboration of these, which have been refined (Box 2) from the meeting's original draft, available at (https://www.force11.org/node/6062). A separate document that dynamically addresses community discussion relating to clarifications and explanations of the principles, and detailed guidelines for and examples of FAIR implementations, is currently being constructed (http://datafairport.org/fair-principles-living-document-menu). The FAIR Guiding Principles describe distinct considerations for contemporary data publishing environments with respect to supporting both manual and automated deposition, exploration, sharing, and reuse. While there have been a number of recent, often domain-focused publications advocating for specific improvements in practices relating to data management and archival[1,11,12], FAIR differs in that it describes concise, domain-independent, high-level principles that can be applied to a wide range of scholarly outputs. Throughout the Principles, we use the phrase '(meta)data' in cases where the Principle should be applied to both metadata and data.

The elements of the FAIR Principles are related, but independent and separable. The Principles define characteristics that contemporary data resources, tools, vocabularies and infrastructures should exhibit to assist discovery and reuse by third-parties. By minimally defining each guiding principle, the barrier-to-entry for data producers, publishers and stewards who wish to make their data holdings FAIR is purposely maintained as low as possible. The Principles may be adhered to in any combination and incrementally, as data providers' publishing environments evolve to increasing degrees of 'FAIRness'. Moreover, the modularity of the Principles, and their distinction between data and metadata, explicitly support a wide range of special circumstances. One such example is highly sensitive or personally-identifiable data, where publication of rich metadata to facilitate discovery, including clear rules regarding the process for accessing the data, provides a high degree of 'FAIRness' even in the absence of FAIR publication of the data itself. A second example involves the publication

of non-data research objects. Analytical workflows, for example, are a critical component of the scholarly ecosystem, and their formal publication is necessary to achieve both transparency and scientific reproducibility. The FAIR principles can equally be applied to these non-data assets, which need to be identified, described, discovered, and reused in much the same manner as data.

Specific exemplar efforts that provide varying levels of FAIRness are detailed later in this document. Additional issues, however, remain to be addressed. First, when community-endorsed vocabularies or other (meta)data standards do not include the attributes necessary to achieve rich annotation, there are two possible solutions: either publish an extension of an existing, closely related vocabulary, or—in the extreme case—create and explicitly publish a new vocabulary resource, following FAIR principles ('I2'). Second, to explicitly identify the standard chosen when more than one vocabulary or other (meta)data standard is available, and given that for instance in the life sciences there are over 600 content standards, the BioSharing registry (https://biosharing.org/) can be of use as it describes the standards in detail, including versions where applicable.

The Principles precede implementation

These high-level FAIR Guiding Principles precede implementation choices, and do not suggest any specific technology, standard, or implementation-solution; moreover, the Principles are not, themselves, a standard or a specification. They act as a guide to data publishers and stewards to assist them in evaluating whether their particular implementation choices are rendering their digital research artefacts Findable, Accessible, Interoperable, and Reusable. We anticipate that these high level principles will enable a broad range of integrative and exploratory behaviours, based on a wide range of technology choices and implementations. Indeed, many repositories are already implementing various aspects of FAIR using a variety of technology choices and several examples are detailed in the next section; examples include *Scientific Data* itself and how narrative data articles are anchored to a progressively FAIR structured metadata.

Examples of FAIRness, and the resulting value-added

Dataverse[7]: Dataverse is an open-source data repository software installed in dozens of institutions globally to support public community repositories or institutional research data repositories. Harvard Dataverse, with more than 60,000 datasets, is the largest of the current Dataverse repositories, and is open to all researchers from all research fields. Dataverse generates a formal citation for each deposit, following the standard defined by Altman and King[13]. Dataverse makes the Digital Object Identifier (DOI), or other persistent identifiers (Handles), public when the dataset is published ('F'). This resolves to a landing page, providing access to metadata, data files, dataset terms, waivers or licenses, and version information, all of which is indexed and searchable ('F', 'A', and 'R'). Deposits include metadata, data files, and any complementary files (such as documentation or code) needed to understand the data and analysis ('R'). Metadata is always public, even if the data are restricted or removed for privacy issues ('F', 'A'). This metadata is offered at three levels, extensively supporting the 'I' and 'R' FAIR principles: 1) data citation metadata, which maps to DataCite schema or Dublin Core Terms, 2) domain-specific metadata, which when possible maps to metadata standards used within a scientific domain, and 3) file-level metadata, which can be deep and extensive for tabular data files (including column-level metadata). Finally, Dataverse provides public machine-accessible interfaces to search the data, access the metadata and download the data files, using a token to grant access when data files are restricted ('A').

FAIRDOM (http://fair-dom.org/about): integrates the SEEK[14] and openBIS[15] platforms to produce a FAIR data and model management facility for Systems Biology. Individual research assets (or aggregates of data and models) are identified with unique and persistent HTTP URLs, which can be registered with DOIs for publication ('F'). Assets can be accessed over the Web in a variety of formats appropriate for individuals and/or their computers (RDF, XML) ('I'). Research assets are annotated with rich metadata, using community standards, formats and ontologies ('I'). The metadata is stored as RDF to enable interoperability and assets can be downloaded for reuse ('R').

ISA[16]: is a community-driven metadata tracking framework to facilitate standards-compliant collection, curation, management and reuse of life science datasets. ISA provides progressively FAIR structured metadata to Nature Scientific Data's Data Descriptor articles, and many GigaScience data papers, and underpins the EBI MetaboLights database among other data resources. At the heart is a general-purpose, extensible ISA model, originally only available as a tabular representation but subsequently enhanced as an RDF-based representation[17], and JSON serializations to enable the 'I' and 'R', becoming 'FAIR' when published as linked data (http://elixir-uk.org/node-events/201cisa-as-a-fair-research-object201d-hack-the-spec-event-1) and complementing other research objects[18].

Open PHACTS[19]: Open PHACTS is a data integration platform for information pertaining to drug discovery. Access to the platform is mediated through a machine-accessible interface[20] which provides multiple representations that are both human (HTML) and machine readable (RDF, JSON,

Box 3 | Emergent community/collaborative initiatives with FAIR as a core focus or activity

bioCADDIE (https://biocaddie.org): The NIH BD2K biomedical and healthCAre Data Discovery Index Ecosystem (bioCADDIE) consortium works to develop a Data Discovery Index (DDI) prototype, which is set to be as transformative and impactful for data as PubMed for the biomedical literature[30]. The DDI focuses on finding ('F') and accessing ('A') the datasets stored across different sources, and progressively works to identify relevant metadata[31] ('I') and maps them to community standards ('R'), linking to BioSharing.

CEDAR[32]: The Center for Expanded Data Annotation and Retrieval (CEDAR) is an NIH BD2K funded center of excellence to develop tools and technologies that reduce the burden of authoring and enhancing metadata that meet community-based standards. CEDAR will enable the creation of metadata templates that implement community based standards for experimental metadata, from BioSharing (https://biosharing.org), and that will be uniquely identifiable and retrievable with HTTP URIs, and annotated with vocabularies and ontologies drawn from BioPortal (http://bioportal.bioontology.org) ('F','A','I','R'). These templates will guide users to create rich metadata with unique and stable HTTP identifiers ('F') that can be retrieved using HTTP ('A') and accessible in a variety of formats (JSON-LD, TURTLE, RDF/XML, CSV, etc) ('I'). These metadata will use community standards, as defined by the template, and include provenance and data usage ('R').

These two projects, among others, provide tools and or collaborative opportunities for those who wish to improve the FAIRness of their data.

XML, CSV, etc), providing the 'A' facet of FAIRness. The interface allows multiple URLs to be used to access information about a particular entity through a mappings service ('F' and 'A'). Thus, a user can provide a ChEMBL URL to retrieve information sourced from, for example, Chemspider or DrugBank. Each call provides a canonical URL in its response ('A' and 'I'). All data sources used are described using standardized dataset descriptions, following the global VoID standard, with rich provenance ('R' and 'I'). All interface features are described using RDF following the Linked Data API specification ('A'). Finally, a majority of the datasets are described using community agreed upon ontologies ('I').

wwPDB[4,21]: wwPDB is a special-purpose, intensively-curated data archive that hosts information about experimentally-determined 3D structures of proteins and nucleic acids. All wwPDB entries are stably hosted on an FTP server ('A') and represented in machine-readable formats (text and XML); the latter are machine-actionable using the metadata provided by the wwPDB conforming to the Macromolecular Information Framework (mmCIF[22]), a data standard of the International Union of Crystallography (IUCr) ('F','I' for humans, 'F','I' for IUCr-aware machines). The wwPDB metadata contains cross-references to common identifiers such as PubMed and NCBI Taxonomy, and their wwPDB metadata are described in data dictionaries and schema documents (http://mmcif.wwpdb.org and http://pdbml.wwpdb.org) which conform to the IUCr data standard for the chemical and structural biology domains ('R'). A variety of software tools are available to interpret both wwPDB data and meta-data ('I','R' for humans, 'I','R' for machines with this software). Each entry is represented by a DOI ('F', 'A' for humans and machines). The DOI resolves to a zipped file which requires special software for further interrogation/interpretation. Other wwPDB access points[23–25] provide access to wwPDB records through URLs that are likely to be stable in the long-term ('F'), and all data and metadata is searchable through one or more of the wwPDB-affiliated websites ('F')

UniProt[26]: UniProt is a comprehensive resource for protein sequence and annotation data. All entries are uniquely identified by a stable URL, that provides access to the record in a variety of formats including a web page, plain-text, and RDF ('F' and 'A'). The record contains rich metadata ('F') that is both human-readable (HTML) and machine-readable (text and RDF), where the RDF formatted response utilizes shared vocabularies and ontologies such as UniProt Core, FALDO, and ECO ('I'). Interlinking with more than 150 different databases, every UniProt record has extensive links into, for example, PubMed, enabling rich citation. These links are machine-actionable in the RDF representation ('R'). Finally, in the RDF representation, the UniProt Core Ontology explicitly types all records, leaving no ambiguity—neither for humans nor machines—about what the data represents ('R'), enabling fully-automated retrieval of records and cross-referencing information.

In addition to, and in support of, communities and resources that are already pursuing FAIR objectives, the Data Citation Implementation Group of Force11 has published specific technical recommendations for how to implement many of the principles[27], with a particular focus on identifiers and their resolution, persistence, and metadata accessibility especially related to citation. In addition, the 'Skunkworks' group that emerged from the Lorentz Workshop has been creating software supporting infrastructures[28] that are, end-to-end, compatible with FAIR principles, and can be implemented over existing repositories. These code modules have a particular focus on metadata publication and searchability, compatibility in cases of strict privacy considerations, and the extremely difficult problem of data and metadata interoperability (manuscript in preparation). Finally, there are several emergent projects, some listed in Box 3, for which FAIR is a key objective. These projects may provide valuable advice and guidance for those wishing to become more FAIR.

FAIRness is a prerequisite for proper data management and data stewardship

The ideas within the FAIR Guiding Principles reflect, combine, build upon and extend previous work by both the Concept Web Alliance (https://conceptweblog.wordpress.com/) partners, who focused on machine-actionability and harmonization of data structures and semantics, and by the scientific and scholarly organizations that developed the Joint Declaration of Data Citation Principles (JDDCP[29]),

who focused on primary scholarly data being made citable, discoverable and available for reuse, so as to be capable of supporting more rigorous scholarship. An attempt to define the similarities and overlaps between the FAIR Principles and the JDDCP is provided at (https://www.force11.org/node/6062). The FAIR Principles are also complementary to the 'Data Seal of Approval' (DSA) (http://datasealofapproval.org/media/filer_public/2013/09/27/guidelines_2014-2015.pdf) in that they share the general aim to render data re-usable for users other than those who originally generated them. While the DSA focuses primarily on the responsibilities and conduct of data producers and repositories, FAIR focuses primarily on the data itself. Clearly, the broader community of stakeholders is coalescing around a set of common, dovetailed visions spanning all facets of the scholarly data publishing ecosystem.

The end result, when implemented, will be more rigorous management and stewardship of these valuable digital resources, to the benefit of the entire academic community. As stated at the outset, good data management and stewardship is not a goal in itself, but rather a pre-condition supporting knowledge discovery and innovation. Contemporary e-Science requires data to be Findable, Accessible, Interoperable, and Reusable in the long-term, and these objectives are rapidly becoming expectations of agencies and publishers. We demonstrate, therefore, that the FAIR Data Principles provide a set of mileposts for data producers and publishers. They guide the implementation of the most basic levels of good Data Management and Stewardship practice, thus helping researchers adhere to the expectations and requirements of their funding agencies. We call on all data producers and publishers to examine and implement these principles, and actively participate with the FAIR initiative by joining the Force11 working group. By working together towards shared, common goals, the valuable data produced by our community will gradually achieve the critical goals of FAIRness.

References

1. Roche, D. G., Kruuk, L. E. B., Lanfear, R. & Binning, S. A. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLOS Biol.* **13,** e1002295 (2015).
2. Bechhofer, S. *et al.* Research Objects: Towards Exchange and Reuse of Digital Knowledge. *Nat. Preced.* doi:10.1038/npre.2010.4626.1 (2010).
3. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41,** D36–D42 (2013).
4. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10,** 980–980 (2003).
5. The Uniprot Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204–D212 (2015).
6. Wenger, M. *et al.* The SIMBAD astronomical database-The CDS reference database for astronomical objects. *Astron. Astrophys. Suppl. Ser.* **143,** 9–22 (2000).
7. Crosas, M. "The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data". *D-Lib Mag* **17** (1), p2 (2011).
8. White, H. C., Carrier, S., Thompson, A., Greenberg, J. & Scherle, R. The Dryad data repository: A Singapore framework metadata architecture in a DSpace environment. *Univ. Göttingen*, p157 (2008).
9. Lecarpentier, D. *et al.* EUDAT: A New Cross-Disciplinary Data Infrastructure for Science. *Int. J. Digit. Curation* **8,** 279–287 (2013).
10. Martone, M. E. FORCE11: Building the Future for Research Communications and e-Scholarship. *Bioscience* **65,** 635 (2015).
11. White, E. *et al.* Nine simple ways to make it easier to (re)use your data. *Ideas Ecol. Evol.* **6** (2013).
12. Sandve, G. K., Nekrutenko, A., Taylor, J. & Hovig, E. Ten Simple Rules for Reproducible Computational Research. *PLoS Comput. Biol.* **9,** e1003285 (2013).
13. Altman, M. & King, G. in *D-Lib Magazine* **13,** no. 3/4 (2007).
14. Wolstencroft, K. *et al.* SEEK: a systems biology data and model management platform. *BMC Syst. Biol.* **9,** 33 (2015).
15. Bauch, A. *et al.* openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics* **12,** 468 (2011).
16. Sansone, S.-A. *et al.* Toward interoperable bioscience data. *Nat. Genet.* **44,** 121–126 (2012).
17. González-Beltrán, A., Maguire, E., Sansone, S.-A. & Rocca-Serra, P. linkedISA: semantic representation of ISA-Tab experimental metadata. *BMC Bioinformatics* **15,** S4 (2014).
18. González-Beltrán, A. *et al.* From Peer-Reviewed to Peer-Reproduced in Scholarly Publishing: The Complementary Roles of Data Models and Workflows in Bioinformatics. *PLoS ONE* **10,** e0127612 (2015).
19. Harland, L. Open PHACTS: A Semantic Knowledge Infrastructure for Public and Commercial Drug Discovery Research. *Knowl. Eng. Knowl. Manag. Lect. Notes Comput. Sci.* **7603/2012,** 1–7 (2012).
20. Groth, P. *et al.* API-centric Linked Data integration: The Open PHACTS Discovery Platform case study. *Web Semant. Sci. Serv. Agents World Wide Web* **29,** 12–18 (2014).
21. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
22. Bourne, P. E., Berman, H. M., Watenpaugh, K., Westbrook, J. D. & Fitzgerald, P. M. D. The macromolecular crystallographic information file (mmCIF). *Meth. Enzym* **277,** 571–590 (1997).
23. Rose, P. W. *et al.* The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.* **43,** D345–D356 (2015).
24. Kinjo, A. R. *et al.* Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Res.* **40,** D453–D460 (2012).
25. Gutmanas, A. *et al.* PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.* **42,** D285–D291 (2014).
26. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204–D212 (2015).
27. Starr, J. *et al.* Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Comput. Sci.* **1,** e1 (2015).
28. Wilkinson, M., Dumontier, M. & Durbin, P. DataFairPort: The Perl libraries version 0.231 doi:10.5281/zenodo.33584 (2015).
29. Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. San Diego CA: FORCE11. https://www.force11.org/datacitation (2014).
30. Ohno-machado, L. *et al.* NIH BD2K bioCADDIE white paper—Data Discovery Index. http://dx.doi.org/10.6084/m9.figshare.1362572 (2015).

31. NIH BD2K bioCADDIE WG3 Members. WG3-MetadataSpecifications: NIH BD2K bioCADDIE Data Discovery Index WG3 Metadata Specification v1 doi:10.5281/zenodo.28019 (2015).
32. Musen, M. A. *et al.* The center for expanded data annotation and retrieval. *J. Am. Med. Informatics Assoc.* **22,** 1148–1152 (2015).

Acknowledgements

The original Lorentz Workshop 'Jointly Designing a Data FAIRport' was organized by Barend Mons in collaboration with and co-sponsored by the Lorentz center, The Dutch Techcenter for the Life Sciences and the Netherlands eScience Center. The principles and themes described in this manuscript represent the significant voluntary contributions and participation of the authors at, and/or subsequent to, this workshop and from the wider Force11, BD2K and ELIXIR communities. We also acknowledge and thank the organizers and backers of the NBDC/DBCLS BioHackathon 2015, where several of the authors made significant revisions to the FAIR Principles.

Author Contributions

M.W. was the primary author of the manuscript, and participated extensively in the drafting and editing of the FAIR Principles. M.D. was significantly involved in the drafting of the FAIR Principles. B.M. conceived of the FAIR Data Initiative, contributed extensively to the drafting of the principles, and to this manuscript text. All other authors are listed alphabetically, and contributed to the manuscript either by their participation in the initial workshop and/or by editing or commenting on the manuscript text.

Additional Information

Competing financial interests: M.A. is the *Nature Genetics*' Editor in Chief; S.A.S. is *Scientific Data*'s Honorary Academic Editor and consultant.

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

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0

Mark D. Wilkinson[1], Michel Dumontier[2], IJsbrand Jan Aalbersberg[3], Gabrielle Appleton[3], Myles Axton[4], Arie Baak[5], Niklas Blomberg[6], Jan-Willem Boiten[7], Luiz Bonino da Silva Santos[8], Philip E. Bourne[9], Jildau Bouwman[10], Anthony J. Brookes[11], Tim Clark[12], Mercè Crosas[13], Ingrid Dillo[14], Olivier Dumon[3], Scott Edmunds[15], Chris T. Evelo[16], Richard Finkers[17], Alejandra Gonzalez-Beltran[18], Alasdair J.G. Gray[19], Paul Groth[3], Carole Goble[20], Jeffrey S. Grethe[21], Jaap Heringa[22], Peter A.C. 't Hoen[23], Rob Hooft[24], Tobias Kuhn[25], Ruben Kok[22], Joost Kok[26], Scott J. Lusher[27], Maryann E. Martone[28], Albert Mons[29], Abel L. Packer[30], Bengt Persson[31], Philippe Rocca-Serra[18], Marco Roos[32], Rene van Schaik[33], Susanna-Assunta Sansone[18], Erik Schultes[34], Thierry Sengstag[35], Ted Slater[36], George Strawn[37], Morris A. Swertz[38], Mark Thompson[32], Johan van der Lei[39], Erik van Mulligen[39], Jan Velterop[40], Andra Waagmeester[41], Peter Wittenburg[42], Katherine Wolstencroft[43], Jun Zhao[44] & Barend Mons[45,46,47]

[1]Center for Plant Biotechnology and Genomics, Universidad Politécnica de Madrid, Madrid 28223, Spain. [2]Stanford University, Stanford 94305-5411, USA. [3]Elsevier, Amsterdam 1043 NX, The Netherlands. [4]Nature Genetics, New York 10004-1562, USA. [5]Euretos and Phortos Consultants, Rotterdam 2741 CA, The Netherlands. [6]ELIXIR, Wellcome Genome Campus, Hinxton CB10 1SA, UK. [7]Lygature, Eindhoven 5656 AG, The Netherlands. [8]Vrije Universiteit Amsterdam, Dutch Techcenter for Life Sciences, Amsterdam 1081 HV, The Netherlands. [9]Office of the Director, National Institutes of Health, Rockville 20892, USA. [10]TNO, Zeist 3700 AJ, The Netherlands. [11]Department of Genetics, University of Leicester, Leicester LE1 7RH, UK. [12]Harvard Medical School, Boston, Massachusetts MA 02115, USA. [13]Harvard University, Cambridge, Massachusetts MA 02138, USA. [14]Data Archiving and Networked Services (DANS), The Hague 2593 HW, The Netherlands. [15]GigaScience, Beijing Genomics Institute, Shenzhen 518083, China. [16]Department of Bioinformatics, Maastricht University, Maastricht 6200 MD, The Netherlands. [17]Wageningen UR Plant Breeding, Wageningen 6708 PB, The Netherlands. [18]Oxford e-Research Center, University of Oxford, Oxford OX1 3QG, UK. [19]Heriot-Watt University, Edinburgh EH14 4AS, UK. [20]School of Computer Science, University of Manchester, Manchester M13 9PL, UK. [21]Center for Research in Biological Systems, School of Medicine, University of California San Diego, La Jolla, California 92093-0446, USA. [22]Dutch Techcenter for the Life Sciences, Utrecht 3501 DE, The Netherlands. [23]Department of Human Genetics, Leiden University Medical Center, Dutch Techcenter for the Life Sciences, Leiden 2300 RC, The Netherlands. [24]Dutch TechCenter for Life Sciences and ELIXIR-NL, Utrecht 3501 DE, The

Netherlands. [25]VU University Amsterdam, Amsterdam 1081 HV, The Netherlands. [26]Leiden Center of Data Science, Leiden University, Leiden 2300 RA, The Netherlands. [27]Netherlands eScience Center, Amsterdam 1098 XG, The Netherlands. [28]National Center for Microscopy and Imaging Research, UCSD, San Diego 92103, USA. [29]Phortos Consultants, San Diego 92011, USA. [30]SciELO/FAPESP Program, UNIFESP Foundation, São Paulo 05468-901, Brazil. [31]Bioinformatics Infrastructure for Life Sciences (BILS), Science for Life Laboratory, Dept of Cell and Molecular Biology, Uppsala University, S-751 24, Uppsala, Sweden. [32]Leiden University Medical Center, Leiden 2333 ZA, The Netherlands. [33]Bayer CropScience, Gent Area 1831, Belgium. [34]Leiden Institute for Advanced Computer Science, Leiden University Medical Center, Leiden 2300 RA, The Netherlands. [35]Swiss Institute of Bioinformatics and University of Basel, Basel 4056, Switzerland. [36]Cray, Inc., Seattle 98164, USA. [37]Unaffiliated. [38]University Medical Center Groningen (UMCG), University of Groningen, Groningen 9713 GZ, The Netherlands. [39]Erasmus MC, Rotterdam 3015 CE, The Netherlands. [40]Independent Open Access and Open Science Advocate, Guildford GU1 3PW, UK. [41]Micelio, Antwerp 2180, Belgium. [42]Max Planck Compute and Data Facility, MPS, Garching 85748, Germany. [43]Leiden Institute of Advanced Computer Science, Leiden University, Leiden 2333 CA, The Netherlands. [44]Department of Computer Science, Oxford University, Oxford OX1 3QD, UK. [45]Leiden University Medical Center, Leiden and Dutch TechCenter for Life Sciences, Utrecht 2333 ZA, The Netherlands. [46]Netherlands eScience Center, Amsterdam 1098 XG, The Netherlands. [47]Erasmus MC, Rotterdam 3015 CE, The Netherlands.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

COMMITTEE ON GEOLOGICAL AND GEOTECHNICAL ENGINEERING

Statement of Task

The Committee on Geological and Geotechnical Engineering (COGGE) is the focal point within the Board on Earth Sciences and Resources for scientific, technical, and public-policy issues pertaining to the engineering applications of Earth Sciences. The committee's scope encompasses Earth processes and materials, including the mechanics of rock and soil, and focuses on safe and responsible human development, risk assessment, and mitigation of natural anthropogenic hazards. The committee organizes and oversees studies:

1) to identify, investigate, and report on questions relating to geological and geotechnical engineering to government, industry, academia, and the public;
2) to provide scientific and technical information to inform public policy on geological and geotechnical engineering issues;
3) to identify new technologies and potential applications; and
4) to promote the acquisition and dissemination of knowledge.

In addition, the committee provides a forum for discussion among academic and professional groups, government agencies, and private industry to enhance national and international cooperation and exchange of information.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

COMMITTEE ON GEOLOGICAL AND GEOTECHNICAL ENGINEERING

2023 Committee Member Biographies

W. ALLEN MARR, JR. (Chair) (NAE) is the founder and chief executive officer of Geocomp Corporation, one of the United States' foremost providers of real-time, web-based performance monitoring of civil engineering structures. Among his technical contributions during his 45-year professional career are the development of techniques for monitoring the stability, movement, and pressure in earthwork projects using sensors, wireless communications, automated analysis, and visualization of data. By applying these techniques, Dr. Marr enabled full-scale construction projects to be built more safely and efficiently and at a lower cost. Dr. Marr and his Geocomp colleagues also developed and use the concept of Active Risk Management to help clients identify and proactively manage risks associated with construction and operation of infrastructure. Over the past 30 years, he has consulted on a number of major projects in the United States and abroad including Boston's Central Artery Tunnel, Dulles International Airport, the new World Trade Center, and projects in The Netherlands, Japan, Venezuela, and Korea. He was elected to the National Academy of Engineering for his innovative applications of numerical methods, risk analysis, advanced laboratory techniques, and field instrumentation to geotechnical engineering and construction. In 2018 he serves as president of the ASCE's Academy of GeoProfessionals. Dr. Marr received a B.S. degree in civil engineering from the University of California at Davis and M.S. and Ph.D. degrees in civil engineering from the Massachusetts Institute of Technology.

SCOTT A. ANDERSON is a principal geotechnical engineer at BGC Engineering in Golden Colorado. He has wide ranging geotechnical design and construction experience in the transportation, water resources, mining, and pipeline industries. His experience includes earthwork—retaining structures, foundation design and construction, and ground modification techniques. He is experienced in remote sensing, geophysical and drilling site characterization, and the determination of soil and rock properties and design parameters. He has provided oversight and review of design and construction as well as contributed to research and deployment of training and new technology in many areas of practice. Prior to joining BGC Engineering, Dr. Anderson was the Geotechnical Services Team Leader for the Federal Highway Administration (FHWA) Resource Center from 2008 to 2017; and prior to that, he held geotechnical leadership roles for the Federal Lands Highway Division of FHWA for 6 years. He was awarded the FHWA Engineer of the Year in 2014, he was the recipient of the K.B. Woods Award in 2016 from the Transportation Research Board for contribution to the design and construction of transportation facilities, and he served as the 2017 Jahns Distinguished Lecturer for the Association of Engineering Geologists and the Geological Society of America. Dr. Anderson holds a B.A. and M.S. in engineering geology from the University of Colorado, Boulder and Colorado State University, respectively. He received an M.S. and Ph.D. in civil engineering from the University of California, Berkeley.

PEDRO ARDUINO joined the geotechnical group in the University of Washington's Department of Civil and Environmental Engineering (UW DCEE) in 1997. His primary research interests are in computational geomechanics with emphasis in constitutive modeling of soils, finite element analysis, meshless techniques, soil structure interaction, and hazard analysis. Much of his current research is in the area of landslide and debris flow simulation, soil-structure interaction, and performance-based earthquake engineering. He has conducted research for the National Science Foundation, the Pacific Earthquake Engineering Research (PEER) Center, and the Washington State Department of

Transportation (WSDOT). Dr. Arduino held the Ray Bowen Professorship for Innovation in Engineering Education from 2003 - 2007 and received the Outstanding Teaching Award from the UW DCEE in 2009. Dr. Arduino was a visiting professor at the Universidad Nacional de Córdoba, Argentina in 2004 and 2008 and at the Universidad de los Andes in Colombia also in 2008. He is a member of the ASCE EM Inelasticity and ER Earth and Retaining Structures committees and served on the editorial board of the *Journal of Geotechnical and Geo-environmental Engineering*. Dr. Arduino is a member of Geotechnical Extreme Events Reconnaissance Association and was part of the reconnaissance teams that visited Chile after the 2010 Maule earthquake, Japan after the 2011 Tohoku earthquake, and Mexico after 2017 Morelos-Puebla Mexico earthquake. He has also served as a consultant to private firms and government agencies in the U.S. and abroad. He earned his B.S.C.E. from the Universidad Nacional de Cordoba, his M.S.C.E. from the University of Puerto Rico, and his Ph.D. from the Georgia Institute of Technology.

ROSALYN W. BERNE is the Anne Shirley Carter Olsson Professor of Applied Ethics, and Director of the Online Ethics Center for Engineering and Science (OEC), in the Department of Engineering and Society, School of Engineering and Applied Sciences, where she has been a faculty member since 1999. She earned B.A. and M.A. degrees in Communication Studies and a PhD in Religious Studies-Bioethics, all from the University of Virginia. As a scholar, Berne explores the intersecting realms of emerging technologies, science, fiction and myth, and the links between the human and non-human worlds. Her research and writing span considerations of ethics in engineering practice, biotechnology and nanotechnology, and ethics in engineering education. Published under her name are two academic books -- Nanotalk: Conversations with Scientists and Engineers about Ethics, Meaning, and Belief in the Development of Nanotechnology (2006) and, Creating Life from Life: Biotechnology and Science Fiction (2014); numerous conference papers and journal articles; Waiting in the Silence: A SciFi novel with an ethics focus (2012); and two award winning books in the genre of body-mind-spirit: When the Horses Whisper (2013) and Waking to Beauty (2016). On leave from UVA from 2009-2011, she served as Vice President for Academic Affairs for the Institute for Shipboard Education. On leave from UVA from 2018-2020 Dr. Berne served as Director of the Center for Engineering Ethics and Society at the National Academy of Engineering, and as PI of the Online Ethics Center (OEC). In September 2020 the OEC was transferred from the NAE to UVA and Dr. Berne continued as PI and director of the OEC. She is an emerita advisor to the "Engineering One Planet" project of the Lemelson Foundation; advisor to the Kern Family Foundation KEEN (Kern Engineering Entrepreneurship Network) project, "Educating the Whole Engineer: Engineering Fundamentals, Character Education, and Entrepreneurial Mindset; a Standing Council Member, Engineering Research Visioning Alliance (ERVA); Co-Chair of the Education Committee, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems ("The IEEE Global Initiative"); and serves on the Ethics Committee of the American Society for Engineering Education.

CRAIG A. DAVIS is a professional consultant on geotechnical, earthquake, and lifeline infrastructure system resilience engineering. During his 31.5 year career at the Los Angeles Department of Water and Power (LADWP) he worked as the departmental chief resilience officer, resilience program manager, seismic manager, geotechnical engineering manager and trunk line design manager. Dr. Davis developed a comprehensive LA water system resilience program and is involved in creating policy for improving infrastructure systems to threats and hazards. He has investigated and evaluated numerous dams and tunnels, managed several multimillion dollar projects, and implemented unique and innovative designs while aiding the development of new technologies and their applications. Dr. Davis served on the National Earthquake Hazards Reduction Program (NEHRP) Advisory Committee on Earthquake Hazards Reduction (ACEHR) for 6 years. He is the founding executive committee chairperson for the ASCE Infrastructure Resilience Division. Dr. Davis was honored with the ASCE's 2016 Le Val Lund Practice Award for Lifeline Risk Reduction. Dr. Davis has published over 125 technical papers and also organized and coordinated numerous international workshops and symposiums on geotechnical engineering and lifeline system resilience. Dr. Davis is a California licensed civil and geotechnical engineer and received a B.S. in Civil Engineering from the California Polytechnic State University in San Louis Obispo, an M.S. in civil engineering with emphasis in structural earthquake engineering from the University of Southern California, and a Ph.D. in civil engineering with emphasis in geotechnical earthquake engineering from the University of Southern California.

RULA DEEB is a Senior Principal at Geosyntec Consultants (Oakland, CA). She has over 25 years of experience in private practice and academia addressing the cross-media fate and transport of contaminants and the remediation of complex soil and groundwater sites. Focusing on emerging contaminants including PFAS, Rula leads strategic planning, business

development, and project execution activities for a variety of industrial, municipal, and federal clients. She served as a member of U.S. Environmental Protection Agency's Science Advisory Board Environmental Engineering Committee between 2016 and 2019. She is the recipient of the 2008 Berkeley Engineering Innovation Young Outstanding Leader Award, the 2010 Industry Recognition Award for outstanding contribution as a member of the Interstate Technology and Regulatory Council's Remediation Risk Management team, and the 2019 Association for Women Geoscientists President's Award. Rula received her doctorate in Civil and Environmental Engineering from UC Berkeley and was recently inducted into Berkeley's Civil and Environmental Engineering Academy of Distinguished Alumni in 2019. She is heavily engaged in the National Academy of Engineering's Frontiers of Engineering program. In 2019, she served as a member of the planning committee for a 2019 National Academies of Sciences, Engineering, and Medicine workshop entitled "Identifying Opportunities to Understand, Control, and Prevent Exposure to PFAS.

JAMI G. DWYER is a licensed professional engineer with 27 years of experience in the mining industry specializing in rock mechanics, blasting, operational efficiency, health and safety, maintenance strategies, mine design, and mine planning. Most recently, she was recruited by Barr Engineering to assist with business development in the mining sector for their Engineering and Design Business Unit. Previous to that, Mrs. Dwyer worked for Barrick Gold Corporation for nearly 11 years where she served in a variety of roles including management of engineering, maintenance, and mine operations departments. While with Barrick, she was also selected to lead a special interdisciplinary project team to develop software applications leveraging big data, machine learning, advanced analytics, and predictive analytics to predict failures of mining equipment components. Mrs. Dwyer spent 15 years employed by the National Institute for Occupational Safety and Health Office of Mine Safety and Health Research in Spokane, Washington, where she led and developed several rock mechanics research projects related to innovative geotechnical monitoring technologies, blast damage assessments, and evaluation of ground support. She was also instrumental in developing early versions of software to locate and analyze mine seismicity and rock bursts in deep underground hard rock mines. Mrs. Dwyer has served on the board of directors for the American Rock Mechanics Association and is a past chair of the Society of Mining, Metallurgy, and Exploration's Mining & Exploration Division's Executive Committee. She holds B.S. degrees in applied computer science and mining Engineering from Montana Technological University, and an M.S. degree in mining engineering from the University of Missouri-Rolla.

YOUSSEF HASASH is the Hall Professor of Civil and Environmental Engineering at the University of Illinois at Urbana-Champaign. Professor Hashash's expertise includes underground structures, deep excavations, numerical modeling, earthquake engineering, and static and dynamic soil-structure interaction analysis, visualization and application of information technology, deep learning and artificial intelligence to geotechnical engineering. Professor Hashash's experience encompasses geotechnical design for tunnels and excavations, seismic soil-structure interaction analysis, and construction monitoring, and the use of advanced computational tools for geotechnical practice. Prof. Hashash is the recipient of the 2014 Peck Medal, and the 2000 Presidential Early Career Award for Scientists and Engineers amongst many honors and recognitions. Prof. Hashash received his Ph.D. (1992), M.S. (1998) and B.S. (1997) in Civil Engineering from the Massachusetts Institute of Technology.

OLADOYIN KOLAWOLE is an Assistant Professor in the Department of Civil and Environmental Engineering Department, New Jersey Institute of Technology (NJIT). He is the director of the Biogeomechanics, Sustainability, and Geotechnics (BSG) Lab at NJIT. He integrates geomechanics, energy engineering, and biotechnology to address problems related to geomaterials, sustainability, energy, and the environment. He earned his Ph.D. degree in Petroleum Engineering (Geomechanics concentration) from Texas Tech University. Prior to joining NJIT, He worked as a Postdoctoral Research Associate at the Edward E. Whitacre Jr. College of Engineering, Texas Tech University, and later a Visiting Faculty Fellow at Hope College, Michigan. Oladoyin developed the "biogeomechanics" concept, which investigates mechanical responses of microbial-rock interactions. His multidisciplinary research lab is broadly focused on geomechanics, biogeomechanics, carbon management, energy resources, and geo-hazard mitigation. He previously participated in the Research Experience in Carbon Sequestration (RECS) program of the U.S. Department of Energy (DOE). He is a recipient of the Distinguished Service Award from the American Rock Mechanics Association (ARMA), and He Co-Convened technical sessions on geomechanical topics. He has published more than 35 peer-reviewed journal articles and conference papers, and He volunteers as a peer reviewer for Nature Scientific Reports, Rock Mechanics and Rock Engineering, and other reputable scientific journals, in addition to mentoring

students. He is a professional member of American Society of Civil Engineers (ASCE), American Rock Mechanics Association (ARMA), and other organizations.

JOHN STAMATAKOS is a geologist and geophysicist with extensive domestic and international research experience. His areas of expertise include structural geology, earthquake seismology, tectonics, paleomagnetism, and exploration geophysics. He is currently an institute scientist at Southwest Research Institute (SwRI). During his 25-year tenure at SwRI, he has provided technical support for seismic hazard assessments of critical nuclear facilities, principally in support of U.S. Nuclear Regulatory Commission (NRC) programs. He also supported technical and research activities on a variety of other natural hazard assessments including fault displacements, tsunamis, volcanoes, tornadoes, and other severe storms, floods, and landslides. Dr. Stamatakos has also served as an NRC expert witness in the Atomic Safety Licensing Board's adjudicatory process hearings on volcanic and seismic contentions for several NRC licensing actions. He is currently a member of the participatory peer review panel for the seismic hazard re-evaluation of Spanish nuclear power plants. Dr. Stamatakos is past associate editor of the Geological Society of America Bulletin and EOS and has served as a regular reviewer of papers for many leading scientific journals. Dr. Stamatakos earned his B.A. in geology from Franklin and Marshall College, and his M.S. and Ph.D. from Lehigh University in geology and geophysics. He also completed a two-year post-doctorate study at the Eidgenössische Technische Hochschule, Institut für Geophysik in Zürich, Switzerland and a three-year research and faculty position at the University of Michigan.

For those eligible for reimbursement:

Due to the new reimbursement system still having several technical issues, Emily Bermudez will be personally handling the filing and creation of all your individual expense reports.

To do this, **we need only three things from you:**

- An itemized list of expenses for which you'll be seeking reimbursement (Including dates and price totals **unless you have a receipt)**
- Receipts for **ALL** expenses over $75.
- Let me know if you paid for the hotel and/or air travel.

Please send everything to Ebermudez@nas.edu as soon as you can.

NATIONAL ACADEMIES Sciences Engineering Medicine

PREVENTING DISCRIMINATION, HARASSMENT, AND BULLYING:
POLICY FOR PARTICIPANTS IN NASEM ACTIVITIES

The National Academies of Sciences, Engineering, and Medicine (NASEM) are committed to the principles of diversity, inclusion, integrity, civility, and respect in all of our activities. We look to you to be a partner in this commitment by helping us to maintain a professional and cordial environment. **All forms of discrimination, harassment, and bullying are prohibited in any NASEM activity.** This policy applies to all participants in all settings and locations in which NASEM work and activities are conducted, including committee meetings, workshops, conferences, and other work and social functions where employees, volunteers, sponsors, vendors, or guests are present.

Discrimination is prejudicial treatment of individuals or groups of people based on their race, ethnicity, color, national origin, sex, sexual orientation, gender identity, age, religion, disability, veteran status, or any other characteristic protected by applicable laws.

Sexual harassment is unwelcome sexual advances, requests for sexual favors, and other verbal or physical conduct of a sexual nature that creates an intimidating, hostile, or offensive environment.

Other types of harassment include any verbal or physical conduct directed at individuals or groups of people because of their race, ethnicity, color, national origin, sex, sexual orientation, gender identity, age, religion, disability, veteran status, or any other characteristic protected by applicable laws, that creates an intimidating, hostile, or offensive environment.

Bullying is unwelcome, aggressive behavior involving the use of influence, threat, intimidation, or coercion to dominate others in the professional environment.

REPORTING AND RESOLUTION

Any violation of this policy should be reported. If you experience or witness discrimination, harassment, or bullying, you are encouraged to make your unease or disapproval known to the individual at the time the incident occurs, if you are comfortable doing so. You are also urged to report any incident by:

- Filing a complaint with the Office of Human Resources at 202-334-3400 or hrservicecenter@nas.edu, or
- Reporting the incident to an employee involved in the activity in which the member or volunteer is participating, who will then file a complaint with the Office of Human Resources.

Complaints should be filed as soon as possible after an incident. To ensure the prompt and thorough investigation of the complaint, the complainant should provide as much information as is possible, such as names, dates, locations, and steps taken. The Office of Human Resources will investigate the alleged violation in consultation with the Office of the General Counsel.

If an investigation results in a finding that an individual has committed a violation, NASEM will take the actions necessary to protect those involved in its activities from any future discrimination, harassment, or bullying, including in appropriate circumstances **the removal of an individual from current NASEM activities and a ban on participation in future activities**.

CONFIDENTIALITY

Information contained in a complaint is kept confidential, and information is revealed only on a need-to-know basis. NASEM will not retaliate or tolerate retaliation against anyone who makes a good faith report of discrimination, harassment, or bullying.

Updated December 2, 2021