EarthCube Program Bibliography (2012-2023)

Abduallah, Y., J. T. L. Wang, Y. Nie, C. Liu, and H. Wang, 2021: DeepSun: machine-learning-as-a-service for solar flare prediction. *Res. Astron. Astrophys.*, **21**, 160, https://doi.org/10.1088/1674-4527/21/7/160.

——, V. K. Jordanova, H. Liu, Q. Li, J. T. L. Wang, and H. Wang, 2022: Predicting Solar Energetic Particles Using SDO/HMI Vector Magnetic Data Products and a Bidirectional LSTM Network. *ApJS*, **260**, 16, https://doi.org/10.3847/1538-4365/ac5f56.

Aberger, C. R., A. Lamb, S. Tu, A. Nötzli, K. Olukotun, and C. Ré, 2017: EmptyHeaded. *ACM Trans. Database Syst.*, **42**, 1–44, https://doi.org/10.1145/3129246.

Abernathey, R. P., and Coauthors, 2021: Cloud-Native Repositories for Big Scientific Data. *Comput. Sci. Eng.*, **23**, 26–35, https://doi.org/10.1109/mcse.2021.3059437.

Al-Saadi, A., I. Paraskevakos, B. C. Gonçalves, H. J. Lynch, S. Jha, and M. Turilli, 2021: Comparing workflow application designs for high resolution satellite image analysis. *Future Generation Computer Systems*, **124**, 315–329, https://doi.org/10.1016/j.future.2021.04.023.

Balasubramanian, V., and Coauthors, 2018: Harnessing the Power of Many: Extensible Toolkit for Scalable Ensemble Applications. *Harnessing the Power of Many: Extensible Toolkit for Scalable Ensemble Applications*, https://doi.org/10.1109/ipdps.2018.00063.

Bhatt, A., T. Valentic, A. Reimer, L. Lamarche, P. Reyes, and R. Cosgrove, 2020: Reproducible Software Environment: a tool enabling computational reproducibility in geospace sciences and facilitating collaboration. *J. Space Weather Space Clim.*, **10**, 12, https://doi.org/10.1051/swsc/2020011.

Blumberg, K. L., A. J. Ponsero, M. Bomhoff, E. M. Wood-Charlson, E. F. DeLong, and B. L. Hurwitz, 2021: Ontology-Enriched Specifications Enabling Findable, Accessible, Interoperable, and Reusable Marine Metagenomic Datasets in Cyberinfrastructure Systems. *Front. Microbiol.*, **12**, https://doi.org/10.3389/fmicb.2021.765268.

Bolton, D. C., P. Shokouhi, B. Rouet-Leduc, C. Hulbert, J. Rivière, C. Marone, and P. A. Johnson, 2019: Characterizing Acoustic Signals and Searching for Precursors during the Laboratory Seismic Cycle Using Unsupervised Machine Learning. *Characterizing Acoustic Signals and Searching for Precursors during the Laboratory Seismic Cycle Using Unsupervised Machine Learning*, **90**, 1088–1098, https://doi.org/10.1785/0220180367.

——, S. Shreedharan, J. Rivière, and C. Marone, 2020: Acoustic Energy Release During the Laboratory Seismic Cycle: Insights on Laboratory Earthquake Precursors and Prediction. *JGR Solid Earth*, **125**, https://doi.org/10.1029/2019jb018975.

Bolukbasi, B., and Coauthors, 2013a: Open Data: Crediting a Culture of Cooperation. *Science*, **342**, 1041–4042, https://doi.org/10.1126/science.342.6162.1041-b.

——, and Coauthors, 2013b: Open Data: Crediting a Culture of Cooperation. *Science*, **342**, 1041–4042, https://doi.org/10.1126/science.342.6162.1041-b.

Brewer, T. E., and Coauthors, 2019: Ecological and Genomic Attributes of Novel Bacterial Taxa That Thrive in Subsurface Soil Horizons. *mBio*, **10**, https://doi.org/10.1128/mbio.01318-19.

van den Brink, L., and Coauthors, 2018: Best practices for publishing, retrieving, and using spatial data on the web. *SW*, **10**, 95–114, https://doi.org/10.3233/sw-180305.

Bromwich, D. H., and Coauthors, 2018: The Arctic System Reanalysis, Version 2. *The Arctic System Reanalysis, Version 2*, **99**, 805–828, https://doi.org/10.1175/bams-d-16-0215.1.

Calyam, P., and Coauthors, 2020: Measuring success for a future vision: Defining impact in science gateways/virtual research environments. *Concurrency Computat Pract Exper*, **33**, https://doi.org/10.1002/cpe.6099.

Cantrall, C., and T. Matsuo, 2021: Deriving column-integrated thermospheric temperature with the N₂ Lyman–Birge–Hopfield (2,0) band. *Atmos. Meas. Tech.*, **14**, 6917–6928, https://doi.org/10.5194/amt-14-6917-2021.

Cash, B. A., and N. J. Burls, 2019: Predictable and Unpredictable Aspects of U.S. West Coast Rainfall and El Niño: Understanding the 2015/16 Event. *Predictable and Unpredictable Aspects of U.S. West Coast Rainfall and El Niño: Understanding the 2015/16 Event*, **32**, 2843–2868, https://doi.org/10.1175/jcli-d-18-0181.1.

Cervone, G., L. Clemente-Harding, S. Alessandrini, and L. Delle Monache, 2017: Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. *Renewable Energy*, **108**, 274–286, https://doi.org/10.1016/j.renene.2017.02.052.

Chan, M. A., S. E. Peters, and B. Tikoff, 2016: The Future of Field Geology, Open Data Sharing and CyberTechnology in Earth Science. *TSR*, **14**, 4–10, https://doi.org/10.2110/sedred.2016.1.4.

Chen, B., and Coauthors, 2020: Measurement of magnetic field and relativistic electrons along a solar flare current sheet. *Nat Astron*, **4**, 1140–1147, https://doi.org/10.1038/s41550-020-1147-7.

Choi, I., A. J. Ponsero, M. Bomhoff, K. Youens-Clark, J. H. Hartman, and B. L. Hurwitz, 2018: Libra: scalable*k*-mer–based tool for massive all-vs-all metagenome comparisons. *Libra: scalablek-mer–based tool for massive all-vs-all metagenome comparisons*, **8**, https://doi.org/10.1093/gigascience/giy165.

Choi, Y.-D., and Coauthors, 2021: Toward open and reproducible environmental modeling by integrating online data repositories, computational environments, and model Application Programming Interfaces. *Environmental Modelling & Software*, **135**, 104888, https://doi.org/10.1016/j.envsoft.2020.104888.

Cholia, S., and Coauthors, 2020: Towards Interactive, Reproducible Analytics at Scale on HPC Systems. *Towards Interactive, Reproducible Analytics at Scale on HPC Systems*, https://doi.org/10.1109/urgenthpc51945.2020.00011.

Chuah, J., M. Deeds, T. Malik, Y. Choi, and J. L. Goodall, 2020: Documenting Computing Environments for Reproducible Experiments. *Documenting Computing Environments for Reproducible Experiments*, https://doi.org/10.3233/apc200106.

Cox, S., J. Klump, and K. Lehnert, 2018: Connecting Scientific Data and Real-World Samples. *Eos*, **99**, https://doi.org/10.1029/2018eo090337.

Crawford, O., D. Al-Attar, J. Tromp, J. X. Mitrovica, J. Austermann, and H. C. P. Lau, 2018: Quantifying the sensitivity of post-glacial sea level change to laterally varying viscosity. *Quantifying the sensitivity of post-glacial sea level change to laterally varying viscosity*, **214**, 1324–1363, https://doi.org/10.1093/gji/ggy184.

David, C. H., Y. Gil, C. J. Duffy, S. D. Peckham, and S. K. Venayagamoorthy, 2016: An introduction to the special issue on Geoscience Papers of the Future. *Earth and Space Science*, **3**, 441–444, https://doi.org/10.1002/2016ea000201.

Deng, J., W. Song, D. Liu, Q. Li, G. Lin, and H. Wang, 2021: Improving the Spatial Resolution of Solar Images Using Generative Adversarial Network and Self-attention Mechanism*. *ApJ*, **923**, 76, https://doi.org/10.3847/1538-4357/ac2aa2.

Dere, A., 2019: iSamples Sample Management Training Module for Soil Cores. *iSamples Sample Management Training Module for Soil Cores*, https://doi.org/10.1594/IEDA/100709.

Di, L., Z. Sun, E. Yu, J. Song, D. Tong, H. Huang, X. Wu, and B. Domenico, 2016: Coupling of Earth science models and earth observations through OGC interoperability specifications. *Coupling of Earth science models and earth observations through OGC interoperability specifications*, https://doi.org/10.1109/igarss.2016.7729933.

Doan, K., A. O. Oloso, K.-S. Kuo, T. L. Clune, H. Yu, B. Nelson, and J. Zhang, 2016: Evaluating the impact of data placement to spark and SciDB with an Earth Science use case. *Evaluating the impact of data placement to spark and SciDB with an Earth Science use case*, https://doi.org/10.1109/bigdata.2016.7840621.

Dove, N. C., and Coauthors, 2020: Continental-scale patterns of extracellular enzyme activity in the subsoil: an overlooked reservoir of microbial activity. *Environ. Res. Lett.*, **15**, 1040a1, https://doi.org/10.1088/1748-9326/abb0b3.

Duncan, C. J., M. A. Chan, E. Hajek, D. Kamola, N. M. Roberts, B. Tikoff, and J. D. Walker, 2021: Bringing sedimentology and stratigraphy into the StraboSpot data management system. *Bringing sedimentology and stratigraphy into the StraboSpot data management system*, **17**, 1914–1927, https://doi.org/10.1130/ges02364.1.

Dye, M., D. S. Stamps, M. Mason, and E. Saria, 2022: Toward Autonomous Detection of Anomalous GNSS Data Via Applied Unsupervised Artificial Intelligence. *Int. J. Semantic Computing*, **16**, 29–45, https://doi.org/10.1142/s1793351x22400025.

Elag, M. M., P. Kumar, L. Marini, J. D. Myers, M. Hedstrom, and B. A. Plale, 2017: Identification and characterization of information-networks in long-tail data collections. *Environmental Modelling & Software*, **94**, 100–111, https://doi.org/10.1016/j.envsoft.2017.03.032.

Emile-Geay, J., D. Khider, N. McKay, Y. Gil, D. Garijo, and V. Ratnakar, 2018: LinkedEarth: supporting paleoclimate data standards and crowd curation. *PAGES Mag*, **26**, 62–63, https://doi.org/10.22498/pages.26.2.62.

Essawy, B. T., J. L. Goodall, H. Xu, and Y. Gil, 2017: Evaluation of the OntoSoft Ontology for describing metadata for legacy hydrologic modeling software. *Environmental Modelling & Software*, **92**, 317–329, https://doi.org/10.1016/j.envsoft.2017.01.024.

——, ——, W. Zell, D. Voce, M. M. Morsy, J. Sadler, Z. Yuan, and T. Malik, 2018: Integrating scientific cyberinfrastructures to improve reproducibility in computational hydrology: Example for HydroShare and GeoTrust. *Environmental Modelling & Software*, **105**, 217–229, https://doi.org/10.1016/j.envsoft.2018.03.025.

——, ——, D. Voce, M. M. Morsy, J. M. Sadler, Y. D. Choi, D. G. Tarboton, and T. Malik, 2020: A taxonomy for reproducible and replicable research in environmental modelling. *Environmental Modelling & Software*, **134**, 104753, https://doi.org/10.1016/j.envsoft.2020.104753.

Fan, Y., and Coauthors, 2014: DigitalCrust - a 4D data system of material properties for transforming research on crustal fluid flow. *Geofluids*, **15**, 372–379, https://doi.org/10.1111/gfl.12114.

Farley, S. S., A. Dawson, S. J. Goring, and J. W. Williams, 2018: Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions*, **68**, 563–576, https://doi.org/10.1093/biosci/biy068.

Ferdowsi, B., J. D. Gartner, K. N. Johnson, A. Kasprak, K. L. Miller, W. Nardin, A. C. Ortiz, and A. Tejedor, 2021a: Earthcasting: Geomorphic Forecasts for Society. *Earth's Future*, **9**, https://doi.org/10.1029/2021ef002088.

——, ——, ——, ——, ——, ——, ——, and ——, 2021b: Earthcasting: Geomorphic Forecasts for Society. *Earth's Future*, **9**, https://doi.org/10.1029/2021ef002088.

Fleishman, G. D., G. M. Nita, N. Kuroda, S. Jia, K. Tong, R. R. Wen, and Z. Zhizhuo, 2018: Revealing the Evolution of Non-thermal Electrons in Solar Flares Using 3D Modeling. *ApJ*, **859**, 17, https://doi.org/10.3847/1538-4357/aabae9.

——, S. A. Anfinogentov, A. G. Stupishin, A. A. Kuznetsov, and G. M. Nita, 2021a: Coronal Heating Law Constrained by Microwave Gyroresonant Emission. *ApJ*, **909**, 89, https://doi.org/10.3847/1538-4357/abdab1.

——, L. Kleint, G. G. Motorina, G. M. Nita, and E. P. Kontar, 2021b: Energy Budget of Plasma Motions, Heating, and Electron Acceleration in a Three-loop Solar Flare. *ApJ*, **913**, 97, https://doi.org/10.3847/1538-4357/abf495.

Fredericks, J., and M. Botts, 2018: Promoting the capture of sensor data provenance: a role-based approach to enable data quality assessment, sensor management and interoperability. *Open geospatial data, softw. stand.*, **3**, https://doi.org/10.1186/s40965-018-0048-5.

Fuka, D. R., A. S. Collick, P. J. A. Kleinman, D. A. Auerbach, R. D. Harmel, and Z. M. Easton, 2016: Improving the spatial representation of soil properties and hydrology using topographically derived initialization processes in the SWAT model. *Hydrol. Process.*, **30**, 4633–4643, https://doi.org/10.1002/hyp.10899.

Gaigalas, ?, ? Di, and ? Sun, 2019: Advanced Cyberinfrastructure to Enable Search of Big Climate Datasets in THREDDS. *IJGI*, **8**, 494, https://doi.org/10.3390/ijgi8110494.

Garijo, D., and Coauthors, 2014a: Workflow Reuse in Practice: A Study of Neuroimaging Pipeline Users. *Workflow Reuse in Practice: A Study of Neuroimaging Pipeline Users*, https://doi.org/10.1109/escience.2014.33.

——, O. Corcho, Y. Gil, B. A. Gutman, I. D. Dinov, P. Thompson, and A. W. Toga, 2014b: FragFlow Automated Fragment Detection in Scientific Workflows. *FragFlow Automated Fragment Detection in Scientific Workflows*, https://doi.org/10.1109/escience.2014.32.

——, Y. Gil, and O. Corcho, 2017: Abstract, link, publish, exploit: An end to end framework for workflow sharing. *Future Generation Computer Systems*, **75**, 271–283, https://doi.org/10.1016/j.future.2017.01.008.

——, M. Osorio, D. Khider, V. Ratnakar, and Y. Gil, 2019: OKG-Soft: An Open Knowledge Graph with Machine Readable Scientific Software Metadata. *OKG-Soft: An Open Knowledge Graph with Machine Readable Scientific Software Metadata*, https://doi.org/10.1109/escience.2019.00046.

Gil, Y., and V. Ratnakar, 2016: Dynamically Generated Metadata and Replanning by Interleaving Workflow Generation and Execution. *Dynamically Generated Metadata and Replanning by Interleaving Workflow Generation and Execution*, https://doi.org/10.1109/icsc.2016.89.

——, and D. Garijo, 2017: Towards Automating Data Narratives. *Towards Automating Data Narratives*, https://doi.org/10.1145/3025171.3025193.

——, V. Ratnakar, and D. Garijo, 2015: OntoSoft. *OntoSoft*, https://doi.org/10.1145/2815833.2816955.

——, and Coauthors, 2016a: Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science*, **3**, 388–415, https://doi.org/10.1002/2015ea000136.

——, D. Garijo, S. Mishra, and V. Ratnakar, 2016b: OntoSoft: A distributed semantic registry for scientific software. *OntoSoft: A distributed semantic registry for scientific software*, https://doi.org/10.1109/escience.2016.7870916.

——, ——, V. Ratnakar, D. Khider, J. Emile-Geay, and N. McKay, 2017: A Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata Annotations. *A Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata Annotations*, 231–246, https://doi.org/10.1007/978-3-319-68204-4_24.

——, and Coauthors, 2018: Intelligent systems for geosciences. *Commun. ACM*, **62**, 76–84, https://doi.org/10.1145/3192335.

Glazner, A., and J. D. Walker, 2020: StraboTools: A Mobile App for Quantifying Fabric in Geology. *GSAT*, https://doi.org/10.1130/gsatg454a.1.

Goble, C., S. Cohen-Boulakia, S. Soiland-Reyes, D. Garijo, Y. Gil, M. R. Crusoe, K. Peters, and D. Schober, 2020: FAIR Computational Workflows. *Data Intellegence*, **2**, 108–121, https://doi.org/10.1162/dint_a_00033.

Gonçalves, B. C., and H. J. Lynch, 2021: Fine-Scale Sea Ice Segmentation for High-Resolution Satellite Imagery with Weakly-Supervised CNNs. *Remote Sensing*, **13**, 3562, https://doi.org/10.3390/rs13183562.

Gonçalves, B. C., B. Spitzbart, and H. J. Lynch, 2020: SealNet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery. *Remote Sensing of Environment*, **239**, 111617, https://doi.org/10.1016/j.rse.2019.111617.

Granger, B. E., and F. Perez, 2021: Jupyter: Thinking and Storytelling With Code and Data. *Comput. Sci. Eng.*, **23**, 7–14, https://doi.org/10.1109/mcse.2021.3059263.

Griffin, J. S., and Coauthors, 2017: Microbial diversity in an intensively managed landscape is structured by landscape connectivity. *Microbial diversity in an intensively managed landscape is structured by landscape connectivity*, **93**, https://doi.org/10.1093/femsec/fix120.

Grimm, E. C., J. Blois, T. Giesecke, R. Graham, A. Smith, and J. Williams, 2018: Constituent databases and data stewards in the Neotoma Paleoecology Database: History, growth, and new directions. *PAGES Mag*, **26**, 64–65, https://doi.org/10.22498/pages.26.2.64.

Gui, Z., and Coauthors, 2016: Developing Subdomain Allocation Algorithms Based on Spatial and Communicational Constraints to Accelerate Dust Storm Simulation. *PLoS ONE*, **11**, e0152250, https://doi.org/10.1371/journal.pone.0152250.

Gundersen, O. E., Y. Gil, and D. W. Aha, 2018: On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications. *AIMag*, **39**, 56–68, https://doi.org/10.1609/aimag.v39i3.2816.

Hallett, B., 2019: iSamples Sample Management Training Module for Rock Outcrop Samples. *iSamples Sample Management Training Module for Rock Outcrop Samples*, https://doi.org/10.1594/IEDA/100691.

He, Y., Y. Zhou, T. Wen, S. Zhang, F. Huang, X. Zou, X. Ma, and Y. Zhu, 2022: A review of machine learning in geochemistry and cosmochemistry: Method improvements and applications. *Applied Geochemistry*, **140**, 105273, https://doi.org/10.1016/j.apgeochem.2022.105273.

Held, N., J. Saunders, J. Futrelle, and M. Saito, 2018: Harnessing the Power of Scientific Python to Investigate Biogeochemistry and Metaproteomes of the Central Pacific Ocean. *Harnessing the Power of Scientific Python to Investigate Biogeochemistry and Metaproteomes of the Central Pacific Ocean*, https://doi.org/10.25080/majora-4af1f417-010.

Held, N. A., and Coauthors, 2020: Co-occurrence of Fe and P stress in natural populations of the marine diazotroph <i>Trichodesmium</i> *Biogeosciences*, **17**, 2537–2551, https://doi.org/10.5194/bg-17-2537-2020.

Hoffman, L., M. R. Mazloff, S. T. Gille, D. Giglio, and A. Varadarajan, 2022: Ocean Surface Salinity Response to Atmospheric River Precipitation in the California Current System. *Ocean Surface Salinity Response to Atmospheric River Precipitation in the California Current System*, **52**, 1867–1885, https://doi.org/10.1175/jpo-d-21-0272.1.

Hogan, A., P. Hitzler, and K. Janowicz, 2016: Linked Dataset description papers at the Semantic Web journal: A critical assessment. *SW*, **7**, 105–116, https://doi.org/10.3233/sw-160216.

Hsu, C. -T., T. Matsuo, A. Maute, R. Stoneback, and C. -P. Lien, 2021: Data-Driven Ensemble Modeling of Equatorial Ionospheric Electrodynamics: A Case Study During a Minor Storm Period Under Solar Minimum Conditions. *JGR Space Physics*, **126**, https://doi.org/10.1029/2020ja028539.

Hsu, L., B. McElroy, R. L. Martin, and W. Kim, 2013a: Building a Sediment Experimentalist Network (SEN): sharing best practices for experimental methods and data management. *TSR*, **11**, 9–12, https://doi.org/10.2110/sedred.2013.4.9.

——, ——, ——, and ——, 2013b: Building a Sediment Experimentalist Network (SEN): sharing best practices for experimental methods and data management. *TSR*, **11**, 9–12, https://doi.org/10.2110/sedred.2013.4.9.

——, R. L. Martin, B. McElroy, K. Litwin-Miller, and W. Kim, 2015a: Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities. *Geomorphology*, **244**, 180–189, https://doi.org/10.1016/j.geomorph.2015.03.039.

——, ——, ——, ——, and ——, 2015b: Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities. *Geomorphology*, **244**, 180–189, https://doi.org/10.1016/j.geomorph.2015.03.039.

Hu, W., and G. Cervone, 2019: Dynamically Optimized Unstructured Grid (DOUG) for Analog Ensemble of numerical weather predictions using evolutionary algorithms. *Computers & Geosciences*, **133**, 104299, https://doi.org/10.1016/j.cageo.2019.07.003.

——, D. Del Vento, and S. Su, 2017: Proceedings of the 2020 Improving Scientific Software Conference. *Proceedings of the 2020 Improving Scientific Software Conference*, https://doi.org/10.5065/P2JJ-9878.

——, G. Cervone, A. Merzky, M. Turilli, and S. Jha, 2022: A new hourly dataset for photovoltaic energy production for the continental USA. *Data in Brief*, **40**, 107824, https://doi.org/10.1016/j.dib.2022.107824.

Hulbert, C., B. Rouet-Leduc, P. A. Johnson, C. X. Ren, J. Rivière, D. C. Bolton, and C. Marone, 2018: Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geosci*, **12**, 69–74, https://doi.org/10.1038/s41561-018-0272-8.

Husson, J. M., and S. E. Peters, 2017: Atmospheric oxygenation driven by unsteady growth of the continental sedimentary reservoir. *Earth and Planetary Science Letters*, **460**, 68–75, https://doi.org/10.1016/j.epsl.2016.12.012.

Illarionov, E., A. Kosovichev, and A. Tlatov, 2020: Machine-learning Approach to Identification of Coronal Holes in Solar Disk Images and Synoptic Maps. *ApJ*, **903**, 115, https://doi.org/10.3847/1538-4357/abb94d.

Janowicz, K., and Coauthors, 2016: Moon Landing or Safari? A Study of Systematic Errors and Their Causes in Geographic Linked Data. *Moon Landing or Safari? A Study of Systematic Errors and Their Causes in Geographic Linked Data*, 275–290, https://doi.org/10.1007/978-3-319-45738-3_18.

Jenkins, C., 2018: Sediment Accumulation Rates For the Mississippi Delta Region: a Time-interval Synthesis. *Sediment Accumulation Rates For the Mississippi Delta Region: a Time-interval Synthesis*, **88**, 301–309, https://doi.org/10.2110/jsr.2018.15.

Jia, X., and Coauthors, 2021: Physics-Guided Machine Learning from Simulation Data: An Application in Modeling Lake and River Systems. *Physics-Guided Machine Learning from Simulation Data: An Application in Modeling Lake and River Systems*, https://doi.org/10.1109/icdm51629.2021.00037.

Jiang, H., J. Wang, C. Liu, J. Jing, H. Liu, J. T. L. Wang, and H. Wang, 2020: Identifying and Tracking Solar Magnetic Flux Elements with Deep Learning. *ApJS*, **250**, 5, https://doi.org/10.3847/1538-4365/aba4aa.

Jiang, P., M. Elag, P. Kumar, S. D. Peckham, L. Marini, and L. Rui, 2017: A service-oriented architecture for coupling web service models using the Basic Model Interface (BMI). *Environmental Modelling & Software*, **92**, 107–118, https://doi.org/10.1016/j.envsoft.2017.01.021.

Jiang, Y., and Coauthors, 2018: A Smart Web-Based Geospatial Data Discovery System with Oceanographic Data as an Example. *IJGI*, **7**, 62, https://doi.org/10.3390/ijgi7020062.

Kadlec, J., B. StClair, D. P. Ames, and R. A. Gill, 2015: WaterML R package for managing ecological experiment data on a CUAHSI HydroServer. *Ecological Informatics*, **28**, 19–28, https://doi.org/10.1016/j.ecoinf.2015.05.002.

——, A. W. Miller, and D. P. Ames, 2016: Extracting Snow Cover Time Series Data from Open Access Web Mapping Tile Services. *J Am Water Resour Assoc*, **52**, 916–932, https://doi.org/10.1111/1752-1688.12387.

Karnauskas, K. B., and D. Giglio, 2022: Argo Reveals the Scales and Provenance of Equatorial Island Upwelling Systems. *Geophysical Research Letters*, **49**, https://doi.org/10.1029/2022gl098744.

Kelbert, A., 2014: Science and Cyberinfrastructure: The Chicken and Egg Problem. *Eos Trans. AGU*, **95**, 458–459, https://doi.org/10.1002/2014eo490006.

Kenigsberg, A. R., J. Rivière, C. Marone, and D. M. Saffer, 2020: Evolution of Elastic and Mechanical Properties During Fault Shear: The Roles of Clay Content, Fabric Development, and Porosity. *J. Geophys. Res. Solid Earth*, **125**, https://doi.org/10.1029/2019jb018612.

Kerkez, B., and Coauthors, 2016: Cloud Hosted Real-time Data Services for the Geosciences (CHORDS). *Geosci. Data J.*, **3**, 4–8, https://doi.org/10.1002/gdj3.36.

Khalsa, S. J. S., C. A. Mattmann, and R. Duerr, 2017: Deep web crawling for insights from polar data. *Deep web crawling for insights from polar data*, https://doi.org/10.1109/igarss.2017.8126974.

Khider, D., and Coauthors, 2019: PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data. *Paleoceanography and Paleoclimatology*, **34**, 1570–1596, https://doi.org/10.1029/2019pa003632.

Kosovichev, A., 2021: Intelligent Databases and Machine-Learning Analysis Tools for Heliophysics. *Intelligent Databases and Machine-Learning Analysis Tools for Heliophysics*, https://doi.org/10.6084/M9.FIGSHARE.14848713.V1.

Krisnadhi, A., and Coauthors, 2015: The GeoLink Modular Oceanography Ontology. *The GeoLink Modular Oceanography Ontology*, 301–309, https://doi.org/10.1007/978-3-319-25010-6_19.

Kumar, P., 2015: Hydrocomplexity: Addressing water security and emergent environmental risks. *Water Resour. Res.*, **51**, 5827–5838, https://doi.org/10.1002/2015wr017342.

Kuo, K.-S., A. Oloso, K. Doan, T. L. Clune, and H. Yu, 2016: Implications of data placement strategy to Big Data technologies based on shared-nothing architecture for geosciences. *Implications of data placement strategy to Big Data technologies based on shared-nothing architecture for geosciences*, https://doi.org/10.1109/igarss.2016.7730983.

——, Y. Pan, F. Zhu, J. Wang, M. L. Rilee, and H. Yu, 2018: A Big Earth Data Platform Exploiting Transparent Multimodal Parallelization. *A Big Earth Data Platform Exploiting Transparent Multimodal Parallelization*, https://doi.org/10.1109/igarss.2018.8518304.

——, H. Yu, Y. Pan, and M. L. Rilee, 2019: Leveraging STARE for Co-aligned Data Locality with netCDF and Python MPI. *Leveraging STARE for Co-aligned Data Locality with netCDF and Python MPI*, https://doi.org/10.1109/igarss.2019.8900423.

Kuroda, N., G. D. Fleishman, D. E. Gary, G. M. Nita, B. Chen, and S. Yu, 2020: Evolution of Flare-Accelerated Electrons Quantified by Spatially Resolved Analysis. *Front. Astron. Space Sci.*, **7**, https://doi.org/10.3389/fspas.2020.00022.

Le, H., D. Samaras, and H. J. Lynch, 2021: A convolutional neural network architecture designed for the automated survey of seabird colonies. *Remote Sens Ecol Conserv*, **8**, 251–262, https://doi.org/10.1002/rse2.240.

Lehnert, K., J. Klump, L. Wyborn, and S. Ramdeen, 2019: Persistent, Global, Unique: The three key requirements for a trusted identifier system for physical samples. *BISS*, **3**, https://doi.org/10.3897/biss.3.37334.

Leonard, L., K. Madduri, and C. J. Duffy, 2016: Tuning Heterogeneous Computing Platforms for Large-Scale Hydrology Data Management. *IEEE Trans. Parallel Distrib. Syst.*, **27**, 2753–2765, https://doi.org/10.1109/tpds.2015.2499741.

Lepore, C., R. Abernathey, N. Henderson, J. T. Allen, and M. K. Tippett, 2021: Future Global Convective Environments in CMIP6 Models. *Earth's Future*, **9**, https://doi.org/10.1029/2021ef002277.

Li, J., T. Matsuo, and L. M. Kilcommons, 2022: Assimilative Mapping of Auroral Electron Energy Flux Using SSUSI Lyman-Birge-Hopfield (LBH) Emissions. *JGR Space Physics*, **127**, https://doi.org/10.1029/2021ja029739.

Li, Z., C. Yang, K. Liu, F. Hu, and B. Jin, 2016: Automatic Scaling Hadoop in the Cloud for Efficient Process of Big Geospatial Data. *IJGI*, **5**, 173, https://doi.org/10.3390/ijgi5100173.

Liu, H., C. Liu, J. T. L. Wang, and H. Wang, 2020a: Predicting Coronal Mass Ejections Using *SDO*/HMI Vector Magnetic Data Products and Recurrent Neural Networks. *ApJ*, **890**, 12, https://doi.org/10.3847/1538-4357/ab6850.

——, Y. Xu, J. Wang, J. Jing, C. Liu, J. T. L. Wang, and H. Wang, 2020b: Inferring Vector Magnetic Fields from Stokes Profiles of GST/NIRIS Using a Convolutional Neural Network. *ApJ*, **894**, 70, https://doi.org/10.3847/1538-4357/ab8818.

Liu, N., J. Jing, Y. Xu, and H. Wang, 2022: Multi-instrument Comparative Study of Temperature, Number Density, and Emission Measure during the Precursor Phase of a Solar Flare. *ApJ*, **930**, 154, https://doi.org/10.3847/1538-4357/ac6425.

Lopez, L. A., R. Duerr, and S. J. S. Khalsa, 2015: Optimizing apache nutch for domain specific crawling at large scale. *Optimizing apache nutch for domain specific crawling at large scale*, https://doi.org/10.1109/bigdata.2015.7363976.

M. C. Carvalho, L. A., D. Garijo, C. Bauzer Medeiros, and Y. Gil, 2018: Semantic Software Metadata for Workflow Exploration and Evolution. *Semantic Software Metadata for Workflow Exploration and Evolution*, https://doi.org/10.1109/escience.2018.00132.

Ma, X., 2022: Knowledge graph construction and application in geosciences: A review. *Computers & Geosciences*, **161**, 105082, https://doi.org/10.1016/j.cageo.2022.105082.

Mahabal, A. A., D. Crichton, S. G. Djorgovski, E. Law, and J. S. Hughes, 2016: From Sky to Earth: Data Science Methodology Transfer. *Proc. IAU*, **12**, 17–26, https://doi.org/10.1017/s1743921317000060.

Maidment, D. R., 2016a: Open Water Data in Space and Time. *J Am Water Resour Assoc*, **52**, 816–824, https://doi.org/10.1111/1752-1688.12436.

——, 2016b: Conceptual Framework for the National Flood Interoperability Experiment. *J Am Water Resour Assoc*, **53**, 245–257, https://doi.org/10.1111/1752-1688.12474.

Malik, T., 2020: Artifact Description/Artifact Evaluation. *Artifact Description/Artifact Evaluation*, https://doi.org/10.1145/3456287.3465479.

Manne, N. N., S. Satpati, T. Malik, A. Bagchi, A. Gehani, and A. Chaudhary, 2022: CHEX. *Proc. VLDB Endow.*, **15**, 1297–1310, https://doi.org/10.14778/3514061.3514075.

Marsicek, J., S. Goring, S. Marcott, S. Meyers, S. Peters, I. Ross, B. Singer, and J. Williams, 2018: Automated extraction of spatiotemporal geoscientific data from the literature using GeoDeepDive. *PAGES Mag*, **26**, 70–70, https://doi.org/10.22498/pages.26.2.70.

Matheny, A. M., P. Marchetto, J. Powell, A. Rechner, J. Chuah, E. McCormick, and S. A. Pierce, 2019: LEAF: Logger for ecological and atmospheric factors. *HardwareX*, **6**, e00079, https://doi.org/10.1016/j.ohx.2019.e00079.

Matsuo, T., 2019: Recent Progress on Inverse and Data Assimilation Procedure for High-Latitude Ionospheric Electrodynamics. *Recent Progress on Inverse and Data Assimilation Procedure for High-Latitude Ionospheric Electrodynamics*, 219–232, https://doi.org/10.1007/978-3-030-26732-2_10.

Mayernik, M. S., 2018: Scholarly resource linking: Building out a "relationship life cycle." *Proc. Assoc. Info. Sci. Tech.*, **55**, 337–346, https://doi.org/10.1002/pra2.2018.14505501037.

——, 2021: Credibility via Coupling. *Engaging STS*, **7**, 10–32, https://doi.org/10.17351/ests2021.769.

——, and Coauthors, 2016: Building Geoscience Semantic Web Applications Using Established Ontologies. *Building Geoscience Semantic Web Applications Using Established Ontologies*, **15**, https://doi.org/10.5334/dsj-2016-011.

McGranaghan, R. M., A. Bhatt, T. Matsuo, A. J. Mannucci, J. L. Semeter, and S. Datta-Barua, 2017: Ushering in a New Frontier in Geospace Through Data Science. *JGR Space Physics*, **122**, https://doi.org/10.1002/2017ja024835.

McNutt, M., K. Lehnert, B. Hanson, B. A. Nosek, A. M. Ellison, and J. L. King, 2016: Liberating field science samples and data. *Science*, **351**, 1024–1026, https://doi.org/10.1126/science.aad7048.

Miller, K. L., W. Kim, and B. McElroy, 2019a: Laboratory Investigation on Effects of Flood Intermittency on Fan Delta Dynamics. *J. Geophys. Res. Earth Surf.*, **124**, 383–399, https://doi.org/10.1029/2017jf004576.

——, ——, and ——, 2019b: Laboratory Investigation on Effects of Flood Intermittency on Fan Delta Dynamics. *J. Geophys. Res. Earth Surf.*, **124**, 383–399, https://doi.org/10.1029/2017jf004576.

Modrak, R. T., D. Borisov, M. Lefebvre, and J. Tromp, 2018: SeisFlows—Flexible waveform inversion software. *Computers & Geosciences*, **115**, 88–95, https://doi.org/10.1016/j.cageo.2018.02.004.

Moges, E., B. L. Ruddell, L. Zhang, J. M. Driscoll, and L. G. Larsen, 2022: Strength and Memory of Precipitation's Control Over Streamflow Across the Conterminous United States. *Water Resources Research*, **58**, https://doi.org/10.1029/2021wr030186.

Mookerjee, M., D. Vieira, M. A. Chan, Y. Gil, C. Goodwin, T. F. Shipley, and B. Tikoff, 2015a: We need to talk: Facilitating communication between field-based geoscience and cyberinfrastructure communities. *GSAT*, 34–35, https://doi.org/10.1130/gsatg248gw.1.

——, ——, ——, ——, ——, ——, and ——, 2015b: We need to talk: Facilitating communication between field-based geoscience and cyberinfrastructure communities. *GSAT*, 34–35, https://doi.org/10.1130/gsatg248gw.1.

——, ——, M. Chan, Y. Gil, T. Pavlis, F. Spear, and B. Tikoff, 2015c: Field Data Management: Integrating Cyberscience and Geoscience. *Eos*, **96**, https://doi.org/10.1029/2015eo036703.

——, and Coauthors, 2023: Cyberinfrastructure for collecting and integrating geology field data: Community priorities and research agenda. *Cyberinfrastructure for collecting and integrating geology field data: Community priorities and research agenda*, https://doi.org/10.1130/2022.2558(01).

Moon, T., M. Fisher, H. Simonoko, and T. Stafford, 2017: QGreenland. *QGreenland*, https://doi.org/10.5281/ZENODO.6369184.

——, A. Ahlstrøm, H. Goelzer, W. Lipscomb, and S. Nowicki, 2018: Rising Oceans Guaranteed: Arctic Land Ice Loss and Sea Level Rise. *Curr Clim Change Rep*, **4**, 211–222, https://doi.org/10.1007/s40641-018-0107-0.

Morlighem, M., and Coauthors, 2017: BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation. *Geophysical Research Letters*, **44**, https://doi.org/10.1002/2017gl074954.

Mullendore, G. L., M. S. Mayernik, and D. C. Schuster, 2021: Open Science Expectations for Simulation-Based Research. *Front. Clim.*, **3**, https://doi.org/10.3389/fclim.2021.763420.

Narock, T., and Coauthors, 2014: The OceanLink project. *The OceanLink project*, https://doi.org/10.1109/bigdata.2014.7004347.

Nita, G. M., N. M. Viall, J. A. Klimchuk, M. A. Loukitcheva, D. E. Gary, A. A. Kuznetsov, and G. D. Fleishman, 2018: Dressing the Coronal Magnetic Extrapolations of Active Regions with a Parameterized Thermal Structure. *ApJ*, **853**, 66, https://doi.org/10.3847/1538-4357/aaa4bf.

Njinju, E. A., and Coauthors, 2019: Lithospheric Structure of the Malawi Rift: Implications for Magma-Poor Rifting Processes. *Tectonics*, **38**, 3835–3853, https://doi.org/10.1029/2019tc005549.

——, D. S. Stamps, K. Neumiller, and J. Gallager, 2021: Lithospheric Control of Melt Generation Beneath the Rungwe Volcanic Province, East Africa: Implications for a Plume Source. *JGR Solid Earth*, **126**, https://doi.org/10.1029/2020jb020728.

Oloso, A., K.-S. Kuo, T. Clune, P. Brown, A. Poliakov, and H. Yu, 2016: Implementing connected component labeling as a user defined operator for SciDB. *Implementing connected component labeling as a user defined operator for SciDB*, https://doi.org/10.1109/bigdata.2016.7840945.

PAGES2k Consortium, ?, 2017: A global multiproxy database for temperature reconstructions of the Common Era. *Sci Data*, **4**, https://doi.org/10.1038/sdata.2017.88.

Peckham, S. D., A. Kelbert, M. C. Hill, and E. W. H. Hutton, 2016: Towards uncertainty quantification and parameter estimation for Earth system models in a component-based modeling framework. *Computers & Geosciences*, **90**, 152–161, https://doi.org/10.1016/j.cageo.2016.03.005.

——, M. Stoica, E. Jafarov, A. Endalamaw, and W. R. Bolton, 2017: Reproducible, component-based modeling with TopoFlow, a spatial hydrologic modeling toolkit. *Earth and Space Science*, **4**, 377–394, https://doi.org/10.1002/2016ea000237.

Pennington, D., I. Ebert-Uphoff, N. Freed, J. Martin, and S. A. Pierce, 2019: Bridging sustainability science, earth science, and data science through interdisciplinary education. *Sustain Sci*, **15**, 647–661, https://doi.org/10.1007/s11625-019-00735-3.

Peters, S., I. Ross, J. Czaplewski, A. Glassel, J. Husson, V. Syverson, A. Zaffos, and M. Livny, 2017a: A New Tool for Deep-Down Data Mining. *Eos*, https://doi.org/10.1029/2017eo082377.

Peters, S. E., and J. M. Husson, 2017: Sediment cycling on continental and oceanic crust. *Geology*, **45**, 323–326, https://doi.org/10.1130/g38861.1.

——, C. Zhang, M. Livny, and C. Ré, 2014: A Machine Reading System for Assembling Synthetic Paleontological Databases. *PLoS ONE*, **9**, e113523, https://doi.org/10.1371/journal.pone.0113523.

——, J. M. Husson, and J. Wilcots, 2017b: The rise and fall of stromatolites in shallow marine environments. *Geology*, **45**, 487–490, https://doi.org/10.1130/g38931.1.

Peucker-Ehrenbrink, B., 2020: Land2Sea database, Version 2.0. *Land2Sea database, Version 2.0*, https://doi.org/10.1594/PANGAEA.892680.

——, and G. J. Fiske, 2019: A continental perspective of the seawater 87Sr/86Sr record: A review. *Chemical Geology*, **510**, 140–165, https://doi.org/10.1016/j.chemgeo.2019.01.017.

Pham, Q., T. Malik, D. H. T. That, and A. Youngdahl, 2018: Improving Reproducibility of Distributed Computational Experiments. *Improving Reproducibility of Distributed Computational Experiments*, https://doi.org/10.1145/3214239.3214241.

Ponsero, A. J., and B. L. Hurwitz, 2019: The Promises and Pitfalls of Machine Learning for Detecting Viruses in Aquatic Metagenomes. *Front. Microbiol.*, **10**, https://doi.org/10.3389/fmicb.2019.00806.

Ponsero, A. J., M. Bomhoff, K. Blumberg, K. Youens-Clark, N. M. Herz, E. M. Wood-Charlson, E. F. Delong, and B. L. Hurwitz, 2020: Planet Microbe: a platform for marine microbiology to discover and analyze interconnected ‘omics and environmental data. *Planet Microbe: a platform for marine microbiology to discover and analyze interconnected ‘omics and environmental data*, **49**, D792–D802, https://doi.org/10.1093/nar/gkaa637.

Radcliffe, D. E., and Coauthors, 2015: Applicability of Models to Predict Phosphorus Losses in Drained Fields: A Review. *J. Environ. Qual.*, **44**, 614–628, https://doi.org/10.2134/jeq2014.05.0220.

Raheem, A., H. Cavus, G. C. Coban, A. C. Kinaci, H. Wang, and J. T. L. Wang, 2021: An investigation of the causal relationship between sunspot groups and coronal mass ejections by determining source active regions. *An investigation of the causal relationship between sunspot groups and coronal mass ejections by determining source active regions*, **506**, 1916–1926, https://doi.org/10.1093/mnras/stab1816.

Regalia, B., G. McKenzie, S. Gao, and K. Janowicz, 2016: Crowdsensing smart ambient environments and services. *Trans. in GIS*, **20**, 382–398, https://doi.org/10.1111/tgis.12233.

Reiss, M. A., and Coauthors, 2021: The Observational Uncertainty of Coronal Hole Boundaries in Automated Detection Schemes. *ApJ*, **913**, 28, https://doi.org/10.3847/1538-4357/abf2c8.

Richard, S. M., G. Pearthree, A. K. Aufdenkampe, J. Cutcher-Gershenfeld, M. Daniels, B. Gomez, D. Kinkade, and G. Percivall, 2014a: Community-Developed Geoscience Cyberinfrastructure. *Eos Trans. AGU*, **95**, 165–166, https://doi.org/10.1002/2014eo200001.

——, ——, ——, ——, ——, ——, ——, and ——, 2014b: Community-Developed Geoscience Cyberinfrastructure. *Eos Trans. AGU*, **95**, 165–166, https://doi.org/10.1002/2014eo200001.

Rilee, M. L., K.-S. Kuo, T. Clune, A. Oloso, P. G. Brown, and H. Yu, 2016: Addressing the big-earth-data variety challenge with the hierarchical triangular mesh. *Addressing the big-earth-data variety challenge with the hierarchical triangular mesh*, https://doi.org/10.1109/bigdata.2016.7840700.

Rosenberg, D. E., and Coauthors, 2020: The Next Frontier: Making Research More Reproducible. *J. Water Resour. Plann. Manage.*, **146**, https://doi.org/10.1061/(asce)wr.1943-5452.0001215.

Roux, S., and Coauthors, 2018: Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat Biotechnol*, **37**, 29–37, https://doi.org/10.1038/nbt.4306.

Sadler, J. M., D. P. Ames, and S. J. Livingston, 2015: Extending HydroShare to enable hydrologic time series data as social media. *Extending HydroShare to enable hydrologic time series data as social media*, **18**, 198–209, https://doi.org/10.2166/hydro.2015.331.

Sadykov, V. M., and A. G. Kosovichev, 2017: Relationships between Characteristics of the Line-of-sight Magnetic Field and Solar Flare Forecasts. *ApJ*, **849**, 148, https://doi.org/10.3847/1538-4357/aa9119.

Sadykov, V. M., A. G. Kosovichev, I. N. Kitiashvili, and A. Frolov, 2019a: Statistical Properties of Soft X-Ray Emission of Solar Flares. *ApJ*, **874**, 19, https://doi.org/10.3847/1538-4357/ab06c3.

——, ——, I. N. Sharykin, and G. S. Kerr, 2019b: Statistical Study of Chromospheric Evaporation in Impulsive Phase of Solar Flares. *ApJ*, **871**, 2, https://doi.org/10.3847/1538-4357/aaf6b0.

Saito, M. A., and Coauthors, 2019: Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. *J. Proteome Res.*, **18**, 1461–1476, https://doi.org/10.1021/acs.jproteome.8b00761.

——, and Coauthors, 2020: Development of an Ocean Protein Portal for Interactive Discovery and Education. *J. Proteome Res.*, **20**, 326–336, https://doi.org/10.1021/acs.jproteome.0c00382.

Saunders, J. K., D. A. Gaylord, N. A. Held, N. Symmonds, C. L. Dupont, A. Shepherd, D. B. Kinkade, and M. A. Saito, 2020: METATRYP v 2.0: Metaproteomic Least Common Ancestor Analysis for Taxonomic Inference Using Specialized Sequence Assemblies—Standalone Software and Web Servers for Marine Microorganisms and Coronaviruses. *J. Proteome Res.*, **19**, 4718–4729, https://doi.org/10.1021/acs.jproteome.0c00385.

——, and Coauthors, 2022: Microbial functional diversity across biogeochemical provinces in the central Pacific Ocean. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, https://doi.org/10.1073/pnas.2200014119.

Schaen, A. J., and Coauthors, 2020: Interpreting and reporting 40Ar/39Ar geochronologic data. *Interpreting and reporting 40Ar/39Ar geochronologic data*, **133**, 461–487, https://doi.org/10.1130/b35560.1.

Sermet, Y., and I. Demir, 2019: Towards an information centric flood ontology for information management and communication. *Earth Sci Inform*, **12**, 541–551, https://doi.org/10.1007/s12145-019-00398-9.

Sessa, J., S. Butts, T. Karim, G. Nelson, C. Norris, D. Serratos, and M. Uhen, 2018: The ePANDDA project: linking the Paleobiology Database, iDigBio, and iDigPaleo for biological and paleontological research, collections management, and outreach. *BISS*, **2**, e26644, https://doi.org/10.3897/biss.2.26644.

Sessa, J. A., and Coauthors, 2017: THE EPANDDA PROJECT: LINKING THE PALEOBIOLOGY DATABASE, IDIGBIO, AND IDIGPALEO FOR BIOLOGICAL AND PALEONTOLOGICAL RESEARCH,

COLLECTIONS MANAGEMENT, AND OUTREACH. *THE EPANDDA PROJECT: LINKING THE PALEOBIOLOGY DATABASE, IDIGBIO, AND IDIGPALEO FOR BIOLOGICAL AND PALEONTOLOGICAL RESEARCH, COLLECTIONS MANAGEMENT, AND OUTREACH*, https://doi.org/10.1130/abs/2017am-298208.

Sethi, R. J., and Y. Gil, 2016: Reproducibility in computer vision: Towards open publication of image analysis experiments as semantic workflows. *Reproducibility in computer vision: Towards open publication of image analysis experiments as semantic workflows*, https://doi.org/10.1109/escience.2016.7870918.

——, and ——, 2017: Scientific workflows in data analysis: Bridging expertise across multiple domains. *Future Generation Computer Systems*, **75**, 256–270, https://doi.org/10.1016/j.future.2017.01.001.

Shi, X., and Coauthors, 2018: Long-Lasting Poloidal ULF Waves Observed by Multiple Satellites and High-Latitude SuperDARN Radars. *JGR Space Physics*, **123**, 8422–8438, https://doi.org/10.1029/2018ja026003.

Shi, Y., D. M. Oliveira, D. J. Knipp, E. Zesta, T. Matsuo, and B. Anderson, 2019: Effects of Nearly Frontal and Highly Inclined Interplanetary Shocks on High-Latitude Field-Aligned Currents (FACs). *Space Weather*, **17**, 1659–1673, https://doi.org/10.1029/2019sw002367.

——, D. J. Knipp, T. Matsuo, L. Kilcommons, and B. Anderson, 2020a: Modes of (FACs) Variability and Their Hemispheric Asymmetry Revealed by Inverse and Assimilative Analysis of Iridium Magnetometer Data. *J. Geophys. Res. Space Physics*, **125**, https://doi.org/10.1029/2019ja027265.

——, ——, ——, ——, and ——, 2020b: Event Studies of High-Latitude FACs With Inverse and Assimilative Analysis of AMPERE Magnetometer Data. *J. Geophys. Res. Space Physics*, **125**, https://doi.org/10.1029/2019ja027266.

Shimizu, C., P. Hitzler, and M. Horridge, 2017: Rendering OWL in Description Logic Syntax. *Rendering OWL in Description Logic Syntax*, 109–113, https://doi.org/10.1007/978-3-319-70407-4_21.

Shreedharan, S., D. C. Bolton, J. Rivière, and C. Marone, 2020: Preseismic Fault Creep and Elastic Wave Amplitude Precursors Scale With Lab Earthquake Magnitude for the Continuum of Tectonic Failure Modes. *Geophys. Res. Lett.*, **47**, https://doi.org/10.1029/2020gl086986.

Song, J., and L. Di, 2017: Near-Real-Time OGC Catalogue Service for Geoscience Big Data. *IJGI*, **6**, 337, https://doi.org/10.3390/ijgi6110337.

Sperhac, J. M., and Coauthors, 2020: *GHub* : Building a glaciology gateway to unify a community. *Concurrency Computat Pract Exper*, **33**, https://doi.org/10.1002/cpe.6130.

Stafford, T., M. Fisher, and Twilamoon Science, 2017: nsidc/qgreenland: v1.0.1. *nsidc/qgreenland: v1.0.1*, https://doi.org/10.5281/ZENODO.4558266.

Stern, C., R. Abernathey, J. Hamman, R. Wegener, C. Lepore, S. Harkins, and A. Merose, 2022: Pangeo Forge: Crowdsourcing Analysis-Ready, Cloud Optimized Data Production. *Front. Clim.*, **3**, https://doi.org/10.3389/fclim.2021.782909.

Stocks, K., S. Diggs, C. Olson, A. Pham, R. Arko, A. Shepherd, and D. Kinkade, 2018: SeaView: Bringing Together an Ocean of Data. *Oceanog*, **31**, 71–71, https://doi.org/10.5670/oceanog.2018.111.

Straneo, F., and Coauthors, 2019: The Case for a Sustained Greenland Ice Sheet-Ocean Observing System (GrIOOS). *Front. Mar. Sci.*, **6**, https://doi.org/10.3389/fmars.2019.00138.

Sun, Z., H. Fang, M. Deng, A. Chen, P. Yue, and L. Di, 2015: Regular Shape Similarity Index: A Novel Index for Accurate Extraction of Regular Objects From Remote Sensing Images. *IEEE Trans. Geosci. Remote Sensing*, **53**, 3737–3748, https://doi.org/10.1109/tgrs.2014.2382566.

——, L. Di, H. Fang, C. Zhang, E. Yu, L. Lin, X. Tan, and P. Yue, 2016a: Embedding Pub/Sub mechanism into OGC web services to augment agricultural crop monitoring. *Embedding Pub/Sub mechanism into OGC web services to augment agricultural crop monitoring*, https://doi.org/10.1109/agro-geoinformatics.2016.7577653.

——, ——, C. Zhang, L. Lin, H. Fang, X. Tan, and P. Yue, 2016b: Combining OGC WCS with SOAP to faciliate the retrieval of remote sensing imagery about agricultural fields. *Combining OGC WCS with SOAP to faciliate the retrieval of remote sensing imagery about agricultural fields*, https://doi.org/10.1109/agro-geoinformatics.2016.7577652.

——, H. Fang, L. Di, P. Yue, X. Tan, and Y. Bai, 2016c: Developing a web-based system for supervised classification of remote sensing images. *Geoinformatica*, **20**, 629–649, https://doi.org/10.1007/s10707-016-0252-3.

——, and Coauthors, 2017a: GeoFairy: Towards a one-stop and location based Service for Geospatial Information Retrieval. *Computers, Environment and Urban Systems*, **62**, 156–167, https://doi.org/10.1016/j.compenvurbsys.2016.11.007.

——, and Coauthors, 2017b: Establish cyberinfrastructure to facilitate agricultural drought monitoring. *Establish cyberinfrastructure to facilitate agricultural drought monitoring*, https://doi.org/10.1109/agro-geoinformatics.2017.8047054.

——, and Coauthors, 2017c: CyberConnector: a service-oriented system for automatically tailoring multisource Earth observation data to feed Earth science models. *Earth Sci Inform*, **11**, 1–17, https://doi.org/10.1007/s12145-017-0308-4.

——, L. Di, and H. Fang, 2018: Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *International Journal of Remote Sensing*, **40**, 593–614, https://doi.org/10.1080/01431161.2018.1516313.

——, ——, and J. Gaigalas, 2019: SUIS: Simplify the use of geospatial web services in environmental modelling. *Environmental Modelling & Software*, **119**, 228–241, https://doi.org/10.1016/j.envsoft.2019.06.005.

——, ——, A. Burgess, J. A. Tullis, and A. B. Magill, 2020a: Geoweaver: Advanced Cyberinfrastructure for Managing Hybrid Geoscientific AI Workflows. *IJGI*, **9**, 119, https://doi.org/10.3390/ijgi9020119.

——, ——, B. Cash, and J. Gaigalas, 2020b: Advanced cyberinfrastructure for intercomparison and validation of climate models. *Environmental Modelling & Software*, **123**, 104559, https://doi.org/10.1016/j.envsoft.2019.104559.

——, ——, S. Cvetojevic, and Z. Yu, 2020c: GeoFairy2: A Cross-Institution Mobile Gateway to Location-Linked Data for In-Situ Decision Making. *IJGI*, **10**, 1, https://doi.org/10.3390/ijgi10010001.

——, and Coauthors, 2022: A review of Earth Artificial Intelligence. *Computers & Geosciences*, **159**, 105034, https://doi.org/10.1016/j.cageo.2022.105034.

Tamma, V., M. Dragoni, R. Gonçalves, and A. Ławrynowicz, 2016: Ontology Engineering. *Ontology Engineering*, https://doi.org/10.1007/978-3-319-33245-1.

Tan, X., and Coauthors, 2015: Building an Elastic Parallel OGC Web Processing Service on a Cloud-Based Cluster: A Case Study of Remote Sensing Data Processing Service. *Sustainability*, **7**, 14245–14258, https://doi.org/10.3390/su71014245.

——, and Coauthors, 2017: Parallel Agent-as-a-Service (P-AaaS) Based Geospatial Service in the Cloud. *Remote Sensing*, **9**, 382, https://doi.org/10.3390/rs9040382.

Theurich, G., and Coauthors, 2016: The Earth System Prediction Suite: Toward a Coordinated U.S. Modeling Capability. *The Earth System Prediction Suite: Toward a Coordinated U.S. Modeling Capability*, **97**, 1229–1247, https://doi.org/10.1175/bams-d-14-00164.1.

Thomer, A., ??, ??, and ??, 2018: iSamples user stories: common themes and areas for future work. *iSamples user stories: common themes and areas for future work*, https://doi.org/10.6084/M9.FIGSHARE.4272164.V1.

Tikoff, B., V. Chatzaras, J. Newman, and N. M. Roberts, 2019: Big data in microstructure analysis: Building a universal orientation system for thin sections. *Journal of Structural Geology*, **125**, 226–234, https://doi.org/10.1016/j.jsg.2018.09.019.

Ton That, D. H., G. Fils, Z. Yuan, and T. Malik, 2017: Sciunits: Reusable Research Objects. *Sciunits: Reusable Research Objects*, https://doi.org/10.1109/escience.2017.51.

Tucker, G. E., and Coauthors, 2022: CSDMS: a community platform for numerical modeling of Earth surface processes. *Geosci. Model Dev.*, **15**, 1413–1439, https://doi.org/10.5194/gmd-15-1413-2022.

Tucker, T., D. Giglio, M. Scanderbeg, and S. S. P. Shen, 2020: Argovis: A Web Application for Fast Delivery, Visualization, and Analysis of Argo Data. *Argovis: A Web Application for Fast Delivery, Visualization, and Analysis of Argo Data*, **37**, 401–416, https://doi.org/10.1175/jtech-d-19-0041.1.

Uhen, M. D., S. Goring, J. Jenkins, and J. Williams, 2018: EarthLife Consortium: Supporting digital paleobiology. *PAGES Mag*, **26**, 78–79, https://doi.org/10.22498/pages.26.2.78.

Uhen, M. D., P. I. Buckland, S. J. Goring, J. P. Jenkins, and J. W. Williams, 2021: The EarthLife Consortium API: an extensible, open-source service for accessing fossil data and taxonomies from multiple community paleodata resources. *Frontiers of Biogeography*, **13**, https://doi.org/10.21425/f5fbg50711.

University of Oregon, 2017: Estimating the Freshwater Flux from the Greenland Ice Sheet Workshop Report, American Geophysical Union, 2018. *Estimating the Freshwater Flux from the Greenland Ice Sheet Workshop Report, American Geophysical Union, 2018*, https://doi.org/10.18739/A24M9198B.

Valcke, S., A. Craig, R. Dunlap, and G. D. Riley, 2016: Sharing Experiences and Outlook on Coupling Technologies for Earth System Models. *Sharing Experiences and Outlook on Coupling Technologies for Earth System Models*, **97**, ES53–ES56, https://doi.org/10.1175/bams-d-15-00239.1.

Valentine, D., I. Zaslavsky, S. Richard, O. Meier, G. Hudman, B. Peucker-Ehrenbrink, and K. Stocks, 2020: EarthCube Data Discovery Studio: A gateway into geoscience data discovery and exploration with Jupyter notebooks. *Concurrency Computat Pract Exper*, **33**, https://doi.org/10.1002/cpe.6086.

Walker, J., 2021: Geology in an Online World. *GSAT*, **31**, 4–7, https://doi.org/10.1130/gsatprsadrs20.1.

Walker, J. D., and Coauthors, 2019: StraboSpot data system for structural geology. *StraboSpot data system for structural geology*, **15**, 533–547, https://doi.org/10.1130/ges02039.1.

Werthmüller, D., R. Rochlitz, O. Castillo-Reyes, and L. Heagy, 2021: Towards an open-source landscape for 3-D CSEM modelling. *Towards an open-source landscape for 3-D CSEM modelling*, **227**, 644–659, https://doi.org/10.1093/gji/ggab238.

Williams, J. W., and Coauthors, 2018a: The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. *Quat. res.*, **89**, 156–177, https://doi.org/10.1017/qua.2017.105.

Williams, J. W., D. Kaufman, A. Newton, and L. von Gunten, 2018b: Building open data: Data stewards and community-curated data resources. *PAGES Mag*, **26**, 50–51, https://doi.org/10.22498/pages.26.2.50.

——, ——, ——, and ——, 2018c: Building and harnessing open paleodata. *PAGES Mag*, **26**, 49–49, https://doi.org/10.22498/pages.26.2.49.

Woodbury, D. H., D. P. Ames, J. Kadlec, S. Duncan, and G. Gault, 2016: A New Open-Access HUC-8 Based Downscaled CMIP-5 Climate Model Forecast Dataset for the Conterminous United States. *J Am Water Resour Assoc*, **52**, 906–915, https://doi.org/10.1111/1752-1688.12437.

Wood-Charlson, E. M., and E. F. DeLong, 2021: EarthCube Oceanography and Geobiology Environmental 'Omics Research Coordination Network Workshop 1 Report. *EarthCube Oceanography and Geobiology Environmental 'Omics Research Coordination Network Workshop 1 Report*, https://doi.org/10.13140/RG.2.1.4908.4561.

Wu, J., L. Yu, and H. Yu, 2015: Texture-based edge bundling: A web-based approach for interactively visualizing large graphs. *Texture-based edge bundling: A web-based approach for interactively visualizing large graphs*, https://doi.org/10.1109/bigdata.2015.7364046.

Wyngaard, J., H. Lynch, J. Nabrzyski, A. Pope, and S. Jha, 2017: Hacking at the Divide Between Polar Science and HPC: Using Hackathons as Training Tools. *Hacking at the Divide Between Polar Science and*

HPC: Using Hackathons as Training Tools, https://doi.org/10.1109/ipdpsw.2017.177.

Xie, Y., E. He, X. Jia, H. Bao, X. Zhou, R. Ghosh, and P. Ravirathinam, 2021a: A Statistically-Guided Deep Network Transformation and Moderation Framework for Data with Spatial Heterogeneity. *A Statistically-Guided Deep Network Transformation and Moderation Framework for Data with Spatial Heterogeneity*, https://doi.org/10.1109/icdm51629.2021.00088.

——, X. Jia, H. Bao, X. Zhou, J. Yu, R. Ghosh, and P. Ravirathinam, 2021b: Spatial-Net. *Spatial-Net*, https://doi.org/10.1145/3474717.3483970.

——, ——, S. Shekhar, H. Bao, and X. Zhou, 2021c: Significant DBSCAN+: Statistically Robust Density-based Clustering. *ACM Trans. Intell. Syst. Technol.*, **12**, 1–26, https://doi.org/10.1145/3474842.

——, S. Shekhar, and Y. Li, 2023: Statistically-Robust Clustering Techniques for Mapping Spatial Hotspots: A Survey. *ACM Comput. Surv.*, **55**, 1–38, https://doi.org/10.1145/3487893.

Yang, C., Q. Huang, Z. Li, K. Liu, and F. Hu, 2016: Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, **10**, 13–53, https://doi.org/10.1080/17538947.2016.1239771.

——, M. Yu, F. Hu, Y. Jiang, and Y. Li, 2017: Utilizing Cloud Computing to address big geospatial data challenges. *Computers, Environment and Urban Systems*, **61**, 120–128, https://doi.org/10.1016/j.compenvurbsys.2016.10.010.

Youens-Clark, K., M. Bomhoff, A. J. Ponsero, E. M. Wood-Charlson, J. Lynch, I. Choi, J. H. Hartman, and B. L. Hurwitz, 2019: iMicrobe: Tools and data-driven discovery platform for the microbiome sciences. *iMicrobe: Tools and data-driven discovery platform for the microbiome sciences*, **8**, https://doi.org/10.1093/gigascience/giz083.

Youngdahl, A., D.-H. Ton-That, and T. Malik, 2019: SciInc: A Container Runtime for Incremental Recomputation. *SciInc: A Container Runtime for Incremental Recomputation*, https://doi.org/10.1109/escience.2019.00040.

Yu, L., and H. Yu, 2016: A study of scientific visualization on heterogeneous processors using Legion. *A study of scientific visualization on heterogeneous processors using Legion*, https://doi.org/10.1109/ldav.2016.7874341.

——, F. Zhu, H. Yu, J. Wang, and K.-S. Kuo, 2016a: Feature extraction and tracking for large-scale geospatial data. *Feature extraction and tracking for large-scale geospatial data*, https://doi.org/10.1109/igarss.2016.7729384.

——, M. L. Rilee, Y. Pan, F. Zhu, K.-S. Kuo, and H. Yu, 2017: Visual analytics with unparalleled variety scaling for big earth data. *Visual analytics with unparalleled variety scaling for big earth data*, https://doi.org/10.1109/bigdata.2017.8257966.

Yu, S., B. Chen, K. K. Reeves, D. E. Gary, S. Musset, G. D. Fleishman, G. M. Nita, and L. Glesener, 2020: Magnetic Reconnection during the Post-impulsive Phase of a Long-duration Solar Flare: Bidirectional Outflows as a Cause of Microwave and X-Ray Bursts. *ApJ*, **900**, 17, https://doi.org/10.3847/1538-4357/aba8a6.

Yu, X., C. Duffy, Y. Gil, L. Leonard, G. Bhatt, and E. Thomas, 2016b: Cyber-Innovated Watershed Research at the Shale Hills Critical Zone Observatory. *IEEE Systems Journal*, **10**, 1239–1250, https://doi.org/10.1109/jsyst.2015.2484219.

Yuan, Z., D. Ton That, S. Kothari, G. Fils, and T. Malik, 2018: Utilizing Provenance in Reusable Research Objects. *Informatics*, **5**, 14, https://doi.org/10.3390/informatics5010014.

Yurchyshyn, V., X. Yang, G. Nita, G. Fleishman, V. Abramenko, S. Inoue, E.-K. Lim, and W. Cao, 2022: Magnetic Field Re-configuration Associated With a Slow Rise Eruptive X1.2 Flare in NOAA Active Region 11944. *Front. Astron. Space Sci.*, **9**, https://doi.org/10.3389/fspas.2022.816523.

Zhang, C., Z. Sun, G. Heo, L. Di, and L. Lin, 2016: A GeoPackage implementation of common map API on Google Maps and OpenLayers to manipulate agricultural data on mobile devices. *A GeoPackage* *implementation of common map API on Google Maps and OpenLayers to manipulate agricultural data on mobile devices*, https://doi.org/10.1109/agro-geoinformatics.2016.7577654.

——, L. Di, Z. Sun, E. G. Yu, L. Hu, L. Lin, J. Tang, and Md. S. Rahman, 2017a: Integrating OGC Web Processing Service with cloud computing environment for Earth Observation data. *Integrating OGC Web Processing Service with cloud computing environment for Earth Observation data*, https://doi.org/10.1109/agro-geoinformatics.2017.8047065.

Zhang, D., P. K. Dera, P. J. Eng, J. E. Stubbs, J. S. Zhang, V. B. Prakapenka, and M. L. Rivers, 2017b: High Pressure Single Crystal Diffraction at PXˆ2. *JoVE*, https://doi.org/10.3791/54660.

Zhang, L., E. Moges, J. W. Kirchner, E. Coda, T. Liu, A. S. Wymore, Z. Xu, and L. G. Larsen, 2021: <scp>CHOSEN</scp> : A synthesis of hydrometeorological data from intensively monitored catchments and comparative analysis of hydrologic extremes. *Hydrological Processes*, **35**, https://doi.org/10.1002/hyp.14429.

Zheng, C. L., V. Ratnakar, Y. Gil, and S. K. McWeeney, 2015: Use of semantic workflows to enhance transparency and reproducibility in clinical omics. *Genome Med*, **7**, https://doi.org/10.1186/s13073-015-0202-y.

Zheng, W., 2022: Glacier geometry and flow speed determine how Arctic marine-terminating glaciers respond to lubricated beds. *The Cryosphere*, **16**, 1431–1445, https://doi.org/10.5194/tc-16-1431-2022.

——, S. Grigsby, F. Sapienza, J. Taylor, T. Snow, F. Pérez, and M. Siegfried, 2021: Mapping ice flow velocity using an easy and interactive feature tracking workflow. *Mapping ice flow velocity using an easy and interactive feature tracking workflow*, https://doi.org/10.5281/ZENODO.5496306.

Zhou, Y., J. Wu, L. Yu, H. Yu, and Z. Tang, 2016: A geohydrologie data visualization framework with an extendable user interface design. *A geohydrologie data visualization framework with an extendable user interface design*, https://doi.org/10.1109/bigdata.2016.7840865.

Zhu, R., Y. Hu, K. Janowicz, and G. McKenzie, 2016: Spatial signatures for geographic feature types: examining gazetteer ontologies using spatial statistics. *Trans. in GIS*, **20**, 333–355, https://doi.org/10.1111/tgis.12232.