A Journal for Human and Machine

James Hendler[1†], Ying Ding[2†] & Barend Mons[3]

[1] Rensselaer Institute for Data Exploration and Applications, Rensselaer Polytechnic Institute, Troy, NY12180, USA

[2] School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, USA

[3] Leiden University Medical Centre, The Netherlands, Poortgebouw N-01, Rijnsburgerweg 10 2333 AA Leiden, The Netherlands

It is with great pride to bring you this new journal of *Data Intelligence*. This journal has at least two major purposes that we hope embrace. First, it will embrace the traditional role of a journal in helping to facilitate the communication of research and best practices in scientific data sharing, especially across disciplines, an area that is continually growing in importance for the modern practice of science. Second, we will be experimenting with new methods of enhancing the sharing of this communication, and examples of the field, by utilizing the increasing power of intelligent computing systems to further facilitate the growth of the field. The journal's title, combining "data," the field we will support, and "intelligence," a means to that end, is meant to connote this growing interaction.

Since the establishment of the first academic journals in the mid 1600's, academic publishing has been a key part of scientific infrastructure, facilitating knowledge sharing and scholarly communication. Journals, at their best, publish high-quality scientific articles so that researchers can be aware of recent advancements in their fields and can have access to archival publications of the "giants" whose shoulders they stand on. The best papers can also inspire researchers to pursue new scientific adventures.

In the past few decades, journals have taken on another, somewhat unexpected role, being used to rank scientists and often impacting their future careers (e.g., hiring, promotion, and future funding). As such, academic publishing can not only help support the communication of science, but increasingly they are taking on a role in defining new subfields where researchers can come together and share information while enhancing their careers. Despite the changing nature of publications, and the search for alt-metrics, we still today see journals as necessary to enhancing scientific communication among researchers and practitioners with common interests, enabling them to forge scientific sub-disciplines and/or work across current disciplines to share their ideas.

Somewhat ironically, as scientists across a number of fields are under pressure to share their scientific data, a growing community of researchers has struggled to find a place to share their ideas about how best

to do this. For example, the US National Academies of Science, Engineering and Medicine, recently held a symposium on "International Coordination for Science Data Infrastructure,"[1] and a number of emerging efforts were discussed among participants who largely were unaware of many of the other ongoing efforts in the area. In this vein, we can see this journal, *Data Intelligence,* as a publication aimed at providing a common communications space for a community of researchers who have not had a place to exchange their ideas as to how best to share data across a wide range. Also, without a reputable journal to publish in, researchers working in this emerging field have been at a real disadvantage academically as papers have been spread over a wide range of publications from different disciplines making it hard to find, and thus to cite, research that builds on common techniques across these domains.

However, as well as this important academic goal, this journal will strive to do more. For these past 350 years of journal publication, the intended readers of the journals have been other scientists. But in the past two decades, with the advent of the Semantic Web with its increasingly powerful knowledge graphs, better metadata standards, and new linked-data tools, there has been a growing interest in the use of machines to enhance scientific information sharing and an increasing capability of artificial intelligence systems to help facilitate the practice of science[2].

Given the speed with which AI technologies are advancing, and the better processing available when data are machine readable, it is clear that we need to start to explore how to build a new generation of journal publication which has the ability to accumulate, disseminate, and create knowledge that is simultaneously contributed to both humans and machines.

Thus, as we hope the name of the journal implies, a key goal of our publication is to go beyond the traditional journal practice and to increasingly help, as it were, to deliver intelligence using data. We admit that it is not yet crystally clear to us how we should differentiate data intelligence from the more general field of artificial intelligence and machine learning technologies. However, our focus is on the sharing of data using these technologies.

Further, we are living in the cusp of an exponentially increasing curve with respect to the data that are becoming available to scientists and researchers (and many others, of course, but our focus as a scientific journal is on the use of data in scientific research and engineering). The advent of the "Internet of Things" will make even more data from sensors and devices available, and scientific instrumentation will produce ever more machine-readable outputs, which will need to be processed for the human scientist to digest. Metaphorically speaking, the only way to control this breaking wave of data (or some might say to tame the data monster) humans need help from machines.

One of the key methods for providing an interface between the machines that are increasingly producing data and the humans who need to process data has been the development of better metadata approaches and the linking of this metadata across applications. The use of metadata is not a new idea, and it has been used to help humans to represent and categorize data even before computerization, for example in the century-long practice in libraries for managing the retrieval of books or periodicals. However, one of the goals of this Journal will be to better understand the needs of scientific metadata and to explore how

---

[1] https://www.nap.edu/catalog/25015/international-coordination-for-science-data-infrastructure-proceedings-of-a-workshop
[2] http://science.sciencemag.org/content/346/6206/171

humans and machines can collaboratively create and reuse metadata to empower knowledge generation and sharing.

In essence, the ultimate goal of this Journal is to help us to explore an emerging ecosystem of scientific data in which human researchers and increasingly capable machines work together to enhance research across diverse fields ranging from the traditional sciences and engineering disciplines, the social sciences and humanities, and to emerging fields that cross the artificial boundaries between these areas. We want to understand how scientific and research data can be timely captured and represented using metadata to add to the "giant global graph" of knowledge proposed by Tim Berners-Lee[③]. This vision is that the linked data of many different kinds form a fusion of linked information where new knowledge can be inferred, and feedback loops can be created to renew and update older data.

In the natural world, an ecological balance is defined as "a state of dynamic equilibrium within a community of organisms in which genetic, species, and ecosystem diversity remain relatively stable, subject to gradual changes through natural succession.[④]" With this journal, we want to explore creating exactly that kind of ecological system within the world of knowledge—where information can be continually changing, but rarely destabilizing.

The tools for the creation and curation of such knowledge are still in their infancy, and we hope as the journal grows over time we will be able to both report on the experiments in data sharing that are being pursued by our contributors, as well as to see how best to create new models that can enhance the sharing of information. We will start as an open access purveyor of papers, but at the same time we will be exploring the development and publication of online information to accompany the articles and/or the publication of human-readable descriptions of metadata, ontologies, and other sources being shared online.

As an example of the kind of work we hope to enable, consider the introductory paper of this issue in which Barend Mons describes how FAIR (Findable, Accessible, Interoperable, and Reusable) data principles can be realized in practice. In particular, he explores an envisioned Internet of FAIR Data and Services (IFDS) that could play a critical role in helping scientists, especially in the findability aspect of FAIR.

Data Intelligence in its role as "the first journal that is also for machines" hopes to explore how we can be an exemplar of creating potential solutions in which all journals, data repositories, and software repositories, in addition to what they already do and publish, also produce a FAIR data point (FDP) with rich metadata to be indexed by multiple search and matching engines, so as to participate in this envisioned IFDS.

In short, even though we are starting out using a traditional publishing model, enhanced by relatively simple article-related metadata, as time goes on we hope to be helping to forge a community of data sharers who can increasingly take advantage of the emerging machine intelligence models that can enhance the practice of research across our many disciplines. We hope you will join us on this journey of exploration by reporting on your experiments, your data sharing technologies, and your shared data resources. We look forward to seeing where we can go together as we experiment with these exciting new *data* models that machine *intelligence* is helping to enable.


[③] see https://en.wikipedia.org/wiki/Giant_Global_Graph

[④] WWF, http://wwf.panda.org

AUTHOR BIOGRAPHY

James Hendler is a Professor and Director of the Rensselaer Institute for Data Exploration and Applications at Rensselaer Polytechnic Institute. He is one of the early Semantic Web developers. Hendler held a longstanding position as professor at the University of Maryland where he was the Director of the Joint Institute for Knowledge Discovery and held joint appointments in the Department of Computer Science, the Institute for Advanced Computer Studies and the Institute for Systems Research. He was the Director for Semantic Web and Agent Technology at the Maryland Information and Network Dynamics Laboratory. Hendler served as an "Internet Web Expert" for the U.S. government, providing guidance to the Data.gov project and in 2016 he was appointed to the National Academies Board on Research Data and Information. He is a Fellow of the American Association for Artificial Intelligence, the British Computer Society, the Institute of Electrical and Electronics Engineers, the American Association for the Advancement of Science and the Association for Computing Machinery.
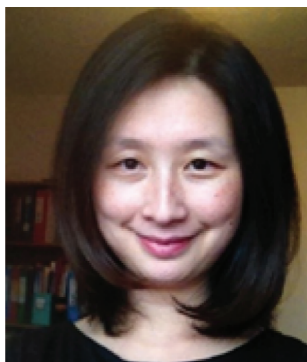
Ying Ding is a Professor at School of Informatics, Computing, and Engineering, Indiana University and currently the associate director for data science online program. She is the Changjiang Guest Professor at Wuhan University and Elsevier Guest Professor at Tongji University, China. She has been involved in various NIH, NSF and European-Union funded projects. She has published more than 200 papers in journals, conferences and workshops, and served as the program committee member for more than 180 international conferences. She is the co-editor of book series called Semantic Web Synthesis by Morgan & Claypool publisher. She is co-author of the book *Intelligent Information Integration in B2B Electronic Commerce* published by Kluwer Academic Publishers, and co-author of the book chapter in *Spinning the Semantic Web* published by Massachusetts Institute of Technology (MIT) Press. She is the co-founder of Data2Discovery Company advancing cutting edge technologies in data science. Her current research interests include data-driven science of science, data-driven discovery, Semantic Web, scientific collaboration and the application of Web Technology.

Barend Mons is Professor of BioSemantics at the Human Genetics Department of Leiden University Medical Center and founder of the BioSemantics group. He was elected CODATA President in 2018. Next to his leading role in the research of the group, Barend plays a leading role in the international development of 'data stewardship' for biomedical data. For instance, he was head-of-node of ELIXIR-NL at the Dutch Techcentre for Life Sciences (until 2015), is Integrator Life Sciences at the Netherlands eScience Center, and board member of the Leiden Centre of Data Science. In 2014, Barend initiated the FAIR data initiative and in 2015, he was appointed Chair of the European Commission's High Level Expert Group for the "European Open Science Cloud", from which he retired by the end of 2016. Presently, Barend is co-leading the GO FAIR initiative, an initiative to kick start developments towards the Internet of FAIR data and services, which will also contribute to the implementation of components of the European Open Science Cloud. The focus of the contribution of the BioSemantics group is on developing an interoperability backbone for biomedical applications in general and rare disease in particular.