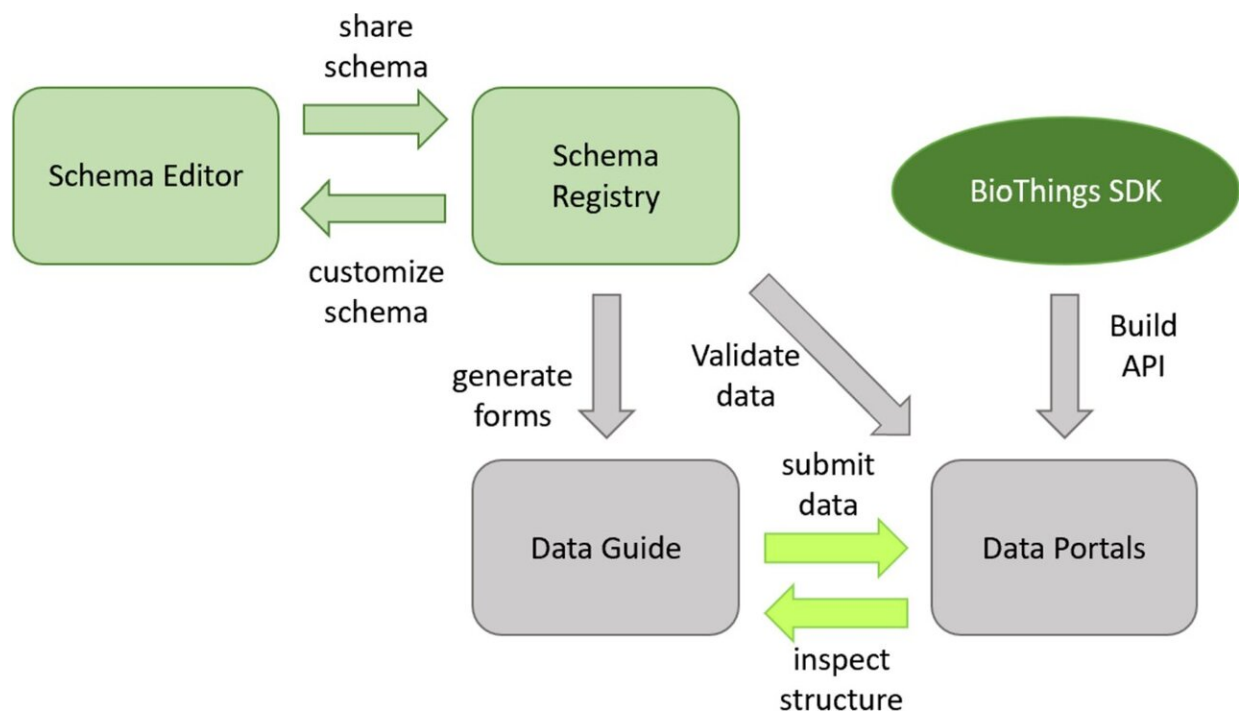


Research team develops behind-the-scenes tool for better biomedical data discovery

May 24 2023



Components of the DDE Schema Playground and how they work together.
 Credit: *BMC Bioinformatics* (2023). DOI: 10.1186/s12859-023-05258-4

Scientists at Scripps Research have developed a new tool that will make datasets, scientific resources and training materials more discoverable online to help more quickly and efficiently facilitate scientific discoveries.

This new tool, called the Data Discovery Engine (DDE) Schema Playground, is described in a paper that published in *BMC Bioinformatics* on April 20, 2023. The DDE Schema Playground is a browser-based resource that empowers scientists to make their data more findable and accessible on the web, which has been a significant barrier in the past.

The resource is an integral part of the Data Discovery Engine—a user-friendly site that helps providers to connect their scientific [datasets](#) to potential target users more efficiently. Researchers can use the Schema Playground to structure information about their datasets in a more interoperable fashion, and portal members can also register their datasets to make the datasets more discoverable and reusable.

"Searching for and finding things efficiently online is hard in general, especially at the complexity level of research assets," says senior author Chunlei Wu, Ph.D., associate professor in the Department of Integrative Structural and Computational Biology at Scripps Research. "Well-structured metadata, often behind the scenes on search engines, is the key for successful online discovery. The DDE provides a suite of behind-the-scenes metadata tools, like the Schema Playground, to bridge between biomedical data providers and researchers as the data consumers."

One of the authors, Scripps Research staff scientist Ginger Tsueng, likens the DDE's utility to making a recipe findable online. Your search results can be broken down by helpful criteria like ratings, prep time, ingredients and so on. These specific, accurate search results are possible because of the metadata (descriptors about the data itself) incorporated into each of those online entries.

But while the metadata for information like recipes is already well standardized and therefore makes them easier to find, this is not the case

for biological datasets, largely because of their complexity. For example, the dataset from an infectious disease preclinical study could differ greatly from a dataset on an immunology clinical trial. Further, every branch of research has its own unique types of metadata, making it very difficult to search among them.

"What Google, Yahoo, Microsoft and others have done for standardizing different types of information, we want to do for biomedical research data and other resources," Tsueng says.

Wu, Tsueng and other team members built the DDE Schema Playground to improve the findability, accessibility, interoperability and reusability of these complex biomedical resources—attributes they collectively refer to as "FAIRness." The DDE leverages the standards of Schema.org, which is an initiative founded by the major search engine companies like Google and Yahoo years ago to help standardize metadata vocabulary.

Schema.org standards help search engines find and make sense of information online, and webpages using these standards allow for more customized search, filter and displays in the [search results](#). For researchers sharing their biological datasets or other biomedically relevant resources online, however, it's been difficult to apply Schema.org standards consistently, as many types of biological information have yet to be standardized in Schema.org.

"Many have created additional guidelines for using these standards and have made them more relevant for biomedical research, but there have been significant technical barriers for finding and using the standards," says Tsueng. "With the DDE, we're helping to create a user-friendly interface so that scientists can more easily upload their information with the right metadata vocabulary, and as a result, others can then search for and find it."

The DDE first started as a project of the National Center for Data to Health and it is currently in use by the NIAID Systems Biology Data Dissemination Working Group (a collective that aims to make research outputs more FAIR), the National COVID Cohort Collaborative (a secure platform for harmonizing clinical data), as well as the Bioschemas community (a grassroots scientific effort that aims to improve the findability of life science resources). Wu and Tsueng are both a part of the Bioschemas steering council, where they encourage scientists to adopt Schema.org standards for their metadata, while also improve public use of this resource.

"We're not trying to make a one-size-fits-all system due to the complexity of the biomedical landscape. Working hand-in-hand with Bioschemas, we instead tried to identify the common component among these datasets and scientific [resources](#), while remaining extensible to fit the diversified use cases—all to help people across different research areas disseminate or access this information," says Wu.

While the DDE Schema Playground represents an important step forward, Wu notes that their goal of making scientific data more findable is an involved and multi-step process. Next, their team will continue to work with the Bioschemas community on standardizing biomedical [metadata](#) and help make it more accessible to people in the life sciences.

More information: Marco A. Cano et al, Schema Playground: a tool for authoring, extending, and using metadata schemas to improve FAIRness of biomedical data, *BMC Bioinformatics* (2023). [DOI: 10.1186/s12859-023-05258-4](https://doi.org/10.1186/s12859-023-05258-4)

Provided by The Scripps Research Institute

Citation: Research team develops behind-the-scenes tool for better biomedical data discovery (2023, May 24) retrieved 20 April 2024 from <https://medicalxpress.com/news/2023-05-team-behind-the-scenes-tool-biomedical-discovery.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.