



Author Once, Publish Everywhere: Portable Metadata Authoring with the CEDAR Embeddable Editor

RESEARCH PAPER

MARTIN J. O'CONNOR

MARCOS MARTÍNEZ-ROMERO

ATTILA L. EGYEDI

METE U. AKDOGAN

MICHAEL V. DORF

MARK A. MUSEN

**Author affiliations can be found in the back matter of this article*

]u[ubiquity press

ABSTRACT

High-quality, 'rich' metadata are essential for making research data findable, accessible, interoperable, and reusable. The Center for Expanded Data Annotation and Retrieval (CEDAR) has long addressed this need by providing tools to design machine-actionable metadata templates that encode community standards in a computable form. To make these capabilities more accessible within real-world research workflows, we have developed the CEDAR Embeddable Editor (CEE)—a lightweight, interoperable Web Component that brings structured, standards-based metadata authoring directly into third-party platforms. The CEE dynamically renders metadata forms from machine-actionable templates and produces semantically rich metadata in JSON-LD format. It supports ontology-based value selection via the BioPortal ontology repository, and it includes external authority resolution for persistent identifiers such as ORCIDs for individuals and RORs for research organizations. Crucially, the CEE requires no custom user-interface development, and it supports deployment across diverse platforms. The CEE has been successfully integrated into generalist scientific data repositories such as Dryad and the Open Science Framework, demonstrating its ability to support creation of descriptive, discipline-specific metadata. By supporting the embedding of metadata authoring within existing research environments, the CEE can facilitate the adoption of community standards and help improve metadata quality across scientific disciplines.

CORRESPONDING AUTHOR:

Martin J. O'Connor

Division of Computational
Medicine, Stanford, CA 94305,
USA

martin.oconnor@stanford.edu

KEYWORDS:

Metadata; Metadata and
systems interoperability;
Metadata management;
Linked Data; FAIR

TO CITE THIS ARTICLE:

O'Connor, M.J., Martínez-Romero, M., Egyedi, A.L., Akdogan, M.U., Dorf, M.V. and Musen, M.A. 2026 Author Once, Publish Everywhere: Portable Metadata Authoring with the CEDAR Embeddable Editor. *Data Science Journal*, 25: 2, pp. 1–18. DOI: <https://doi.org/10.5334/dsj-2026-002>

The FAIR Guiding Principles have become a cornerstone of modern data stewardship, emphasizing that data should be findable, accessible, interoperable, and reusable—not just for humans, but also for machines (Wilkinson et al., 2016). While these principles are broadly endorsed, operationalizing them remains a challenge, particularly for the descriptive metadata that must accompany scientific datasets to make the datasets FAIR in practice. The difficulty lies in the abstract nature of the FAIR principles themselves, which call for ‘rich’ and ‘community-relevant’ metadata without specifying how those qualities should be formally represented or assessed.

The CEDAR Workbench (Musen et al., 2015) has addressed this challenge by offering a system for creating machine-actionable metadata templates that encode community standards in a structured and computable form. These templates define not only the required metadata fields but also the value types and controlled vocabularies appropriate for each field. Originally, users completed these templates using a centralized Web interface—the CEDAR Metadata Editor—hosted within the CEDAR Workbench. This environment ensured that metadata conformed to both the structural and semantic conventions of the relevant scientific community (Musen et al., 2022). However, the centralized model required researchers to leave their native platforms and engage with an external tool, limiting its integration into routine workflows.

To address this limitation, we developed the CEDAR Embeddable Editor (CEE): a reusable, standards-compliant Web Component that brings CEDAR’s metadata authoring capabilities directly into third-party applications. Although CEDAR already provides a strong infrastructure for template design and management, researchers still need tools that fit smoothly into their everyday workflows. The CEE meets this need by allowing external systems to leverage CEDAR’s structured templates without requiring users to leave their existing environments—embedding metadata creation within the flow of research and data submission.

Critically, CEE eliminates the need for platform developers to build or maintain custom user interfaces for metadata entry and display. Because the editor dynamically renders forms based on CEDAR templates and returns structured metadata in JSON-LD, the presentation and data-creation logic are entirely encapsulated within the component. This design ensures that any changes to metadata templates—whether structural updates, controlled vocabulary adjustments, or community-standard revisions—are automatically reflected in the rendered forms without further development effort. As a result, the CEE enables deployment of standards-compliant metadata interfaces across diverse platforms, while reducing the engineering overhead typically associated with metadata user interface development and maintenance.

Building on these motivations, this paper introduces the design and implementation of the CEE. It explains how the CEE fits into the broader CEDAR architecture, how it supports domain-specific metadata creation through template-driven rendering, and how it enables consistent, ontology-linked metadata output in JSON-LD format. In addition to ontology-based value selection, the CEE also supports external authority fields, allowing users to incorporate validated identifiers from registries such as ORCID (Haak et al., 2012), ROR (Giglia et al., 2020), and the EPA CompTox API (Williams et al., 2017) system directly into metadata records. We present real-world deployments of the CEE in platforms such as the Open Science Framework (Olson and Corker, 2024) and Dryad (Lippincott, 2024) and show how embedding standards-aligned metadata tools directly into data publication workflows can promote wider adoption of community practices and improve metadata quality across disciplines.

RELATED WORK

A wide array of tools and frameworks have been developed to support the creation of metadata for research datasets, particularly in response to the increasing emphasis on the FAIR principles.

Several general-purpose tools, such as the Dublin Core Metadata Generator (Steffel, 2011) and the DataCite Metadata Generator (DataCite, 2024), offer lightweight interfaces for generating metadata compliant with high-level standards. While useful for basic dataset documentation, these tools are constrained to a single, highly abstract metadata model and lack support for the detailed, context-specific attributes required in many research domains. As a result, the metadata they produce are often too shallow and generic to support meaningful data discovery, interoperability, or machine-actionable reuse.

To address the limitations of generic tools, various domain-specific frameworks have emerged. Examples include ISA-Tab ([Rocca-Serra et al., 2010](#)) for life sciences, and the Ecological Metadata Language (EML) ([Fegraus et al., 2005](#)) for environmental science. These systems embed richer semantics and structured fields aligned with disciplinary standards. However, they often presume substantial familiarity with community-specific ontologies and metadata practices. This learning curve can be a significant barrier, particularly for interdisciplinary users or researchers new to metadata curation. They also are tailored to their particular scientific disciplines and therefore do not represent general solutions.

Many technically capable metadata systems are hampered by usability issues. Tools such as Metatool ([Berkley, Jones and Boettiger, 2001](#)) and Morpho ([Higgins, Berkley and Jones, 2002](#)) illustrate this tension; while powerful, their complex interfaces and steep learning curves have limited broader adoption. Tools that do not prioritize intuitive interaction risk producing inconsistent or incomplete metadata, even when supported by sound metadata models.

Frameworks such as FAIRsharing ([Sansone et al., 2019](#)) and RO-Crate ([Soiland-Reyes et al., 2022](#)) have emerged to bridge the FAIR principles and implementation. FAIRsharing catalogs community standards and encourages reuse, while RO-Crate provides a packaging format for datasets and associated metadata. However, their adoption often requires substantial customization and technical overhead, limiting accessibility for many research groups.

Another critical limitation of current tools is their poor integration with external research platforms. Although systems such as OpenRefine ([Verborgh, R and Wilde, 2013](#)) facilitate metadata enrichment and linking, they are often external to the platforms where data are managed and shared. Embedding them typically demands additional development effort and can introduce friction into data workflows. This integration gap limits the routine use of metadata tools at the point of data publication or sharing.

The CEDAR Workbench aims to address many of the longstanding challenges in metadata authoring by introducing machine-actionable templates that encode community standards and domain-specific requirements in a reusable, computable form. However, early CEDAR integrations with third-party systems required researchers to leave their native environments and use a centralized Web interface—the CEDAR Metadata Editor—for metadata entry. While this interface enabled form-based metadata authoring with ontology-based value selection and semantic validation, integration into external systems was cumbersome. Metadata had to be authored on the CEDAR Workbench itself, after which downstream submission workflows needed to extract, transform, and relay the records manually or via batch processes. This approach also required platform developers to maintain custom routing code and to synchronize template changes by hand, creating brittle integrations that were difficult to scale or update. These early integrations also raised authentication and single sign-on (SSO) challenges, as external platforms had to broker separate CEDAR logins or implement custom token-passing logic, adding friction for users and complicating platform security models. For example, an early integration with the LINCS ([Egyedi et al., 2018](#)) demonstrated the feasibility of an end-to-end CEDAR-facilitated metadata pipeline but also revealed significant friction—researchers were forced to shift between platforms, and repository operators needed to implement ad hoc transformations, authentication handoffs, and validation logic to reconcile CEDAR-hosted metadata with their internal systems. These challenges underscored the need for a mechanism to bring metadata authoring directly into external workflows rather than requiring external workflows to adapt to CEDAR.

The CEE was developed to overcome these limitations by enabling CEDAR templates to be rendered directly within third-party applications. Delivered as a lightweight, standards-compliant Web Component, the CEE eliminates the need for external metadata portals or custom interface development. It dynamically adapts to template changes, supports ontology-based value recommendations, and outputs structured metadata enriched with persistent identifiers from external authorities such as ORCID and ROR. Crucially, it requires no custom user interface development, allowing deployment across diverse platforms without the need for additional frontend engineering. This approach combines domain specificity, usability, and seamless integration—directly addressing the technical and adoption barriers encountered in earlier systems.

In summary, while existing metadata tools have advanced standardization and semantic expressivity, few offer the combination of rich domain modeling, intuitive design, and direct integration that the CEE enables.

To extend and improve the capabilities of the CEDAR Workbench, the team developed the CEE. Building on the template-driven infrastructure of the CEDAR Workbench, the CEE extends CEDAR's metadata creation capabilities into external platforms through a lightweight, self-contained Web Component. It enables structured, ontology-based metadata, authoring directly within host applications, combining flexibility, semantic precision, and ease of deployment to support a wide range of scientific workflows.

BACKGROUND: OVERVIEW OF THE CEDAR WORKBENCH

The CEDAR (Center for Expanded Data Annotation and Retrieval) Workbench (Gonçalves et al., 2017; Musen et al., 2015) is developed and maintained within Stanford University's Division of Computational Medicine (formerly the Stanford Center for Biomedical Informatics Research), which oversees its ongoing development and maintenance. CEDAR was created to address a central challenge in scientific data stewardship: While the FAIR principles urge that metadata should be 'rich' and compliant with community standards, few tools provide a practical means for researchers to achieve these goals (Musen et al., 2022). The burden of interpreting abstract guidelines, selecting appropriate standards, and ensuring consistency across records typically falls on individual researchers or data curators—often without sufficient training, resources, or tooling. CEDAR was conceived to lower these barriers through automation, standardization, and intuitive user interfaces focused on community-driven design.

At the heart of the CEDAR Workbench is the concept of machine-actionable metadata templates (O'Connor et al., 2016). These templates encode the structure, semantics, and possible expected values of metadata records using a formal, JSON Schema-based model. Unlike free-text reporting guidelines, which are often ambiguous and difficult to enforce, CEDAR templates make metadata specifications explicit and computable. Each template defines a collection of fields, their data types, required constraints, and associated controlled vocabularies drawn from community ontologies accessible through the BioPortal ontology repository (Vendetti et al., 2025). CEDAR supports a rich array of field types, including text, number, date/time, Boolean, checkboxes, image, and video fields. Fields can also be defined as single-valued or multi-valued. Templates may be assembled from reusable components—known as template elements—that encapsulate commonly used metadata patterns (e.g., specimen descriptors, instrumentation settings, study design attributes).

The typical CEDAR metadata development workflow begins with template creation. Domain experts use CEDAR's Template Designer (Figure 1) to build templates that reflect the information needs of a specific scientific community, experimental method, or data-sharing requirement. Through a drag-and-drop interface, users define the structure of a metadata record without writing code.

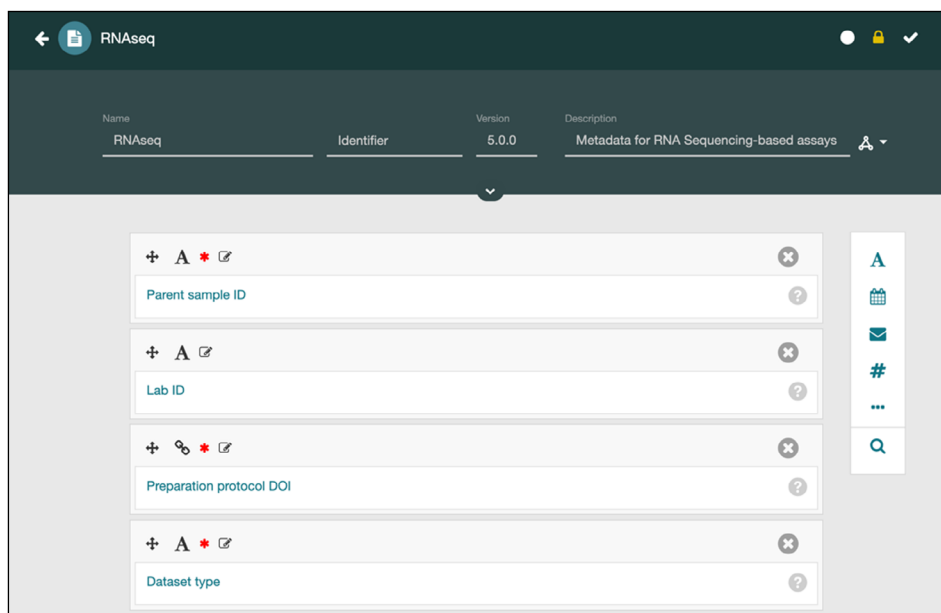


Figure 1 A CEDAR template in the Template Designer tool. Template authors can use the Template Designer to interactively build templates to describe metadata reporting guidelines. Here, the template describes metadata for an RNAseq-based assay.

Once a template is published, it becomes the blueprint for generating a user-facing metadata authoring interface. In the CEDAR Workbench, a tool called the Metadata Editor (Figure 2) can be used to automatically generate a Web-based form from a template's schema. Form fields are rendered according to their type (e.g., text, number, date, controlled term) and, if the fields are ontology-linked, users are guided to select appropriate ontology terms through drop-down autocompletion menus backed by BioPortal's ontology services. Constraints such as required fields, allowed values, or term hierarchies are automatically enforced by the editor, ensuring consistency and adherence to the underlying specification.

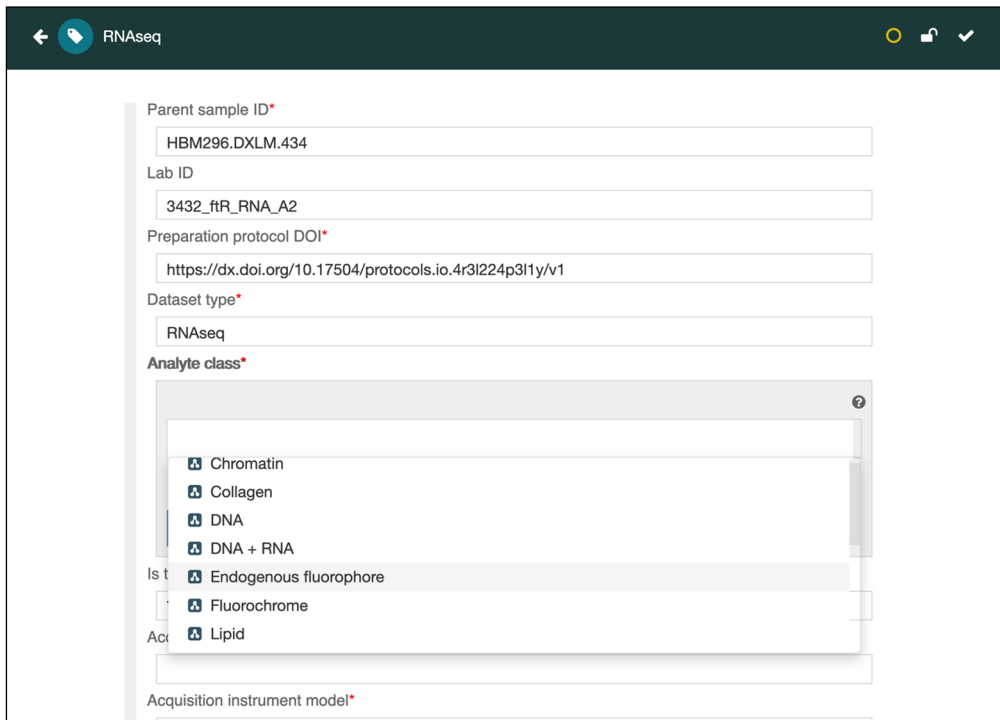


Figure 2 CEDAR-generated web-based form illustrating how metadata can be acquired from users in the Metadata Editor tool. Here, the form is generated automatically from a template that specifies metadata for a RNAseq-based biomedical assay. In this case, the metadata author has completed four initial fields and is being presented with a set of choices for a field called Analyte class. Once completed, the user can save the completed form, whereupon it is stored in CEDAR.

As a user completes the form, CEDAR incrementally builds a structured metadata instance that conforms to the selected template. The resulting record is exported in JSON-LD (Sporny, Kellogg and Lanthaler, 2020), a linked data format that preserves the semantic structure of the metadata and allows for easy integration with downstream systems and repositories. This output format is key to ensuring the metadata's interoperability and machine-readability, as each data element is unambiguously linked to the vocabulary or ontology from which it was drawn.

Beyond template creation and instance authoring, the CEDAR Workbench also supports metadata sharing, collaboration, and validation. Templates and metadata instances can be versioned, published to community folders, or shared with specific collaborators. Additionally, CEDAR offers tools for metadata validation and data quality assessment, which help curators evaluate completeness and compliance before metadata are submitted to external repositories.

Major biomedical consortia, such as the Human BioMolecular Atlas Program (HuBMAP) (HuBMAP Consortium, 2019) and the Cellular Senescence Network (SenNet) (SenNet, 2025), are using CEDAR as the foundation of their entire metadata management workflow. Others, for example HEAL (Baker, Koroshetz and Volkow, 2021), have started integrating CEDAR as a central component of their metadata management processes. CEDAR is also incorporated in data portals, such as Health-RI, a national initiative in the Netherlands to improve the reuse of health data for policy, research, and innovation (ZonMw, 2024). The RADx Data Hub (Martínez-Romero et al., 2025), an initiative to create a national resource for managing COVID-19 data, uses CEDAR to support its metadata management capabilities.

The CEDAR Workbench offers a comprehensive solution for defining, generating, and validating high-quality metadata. By transforming metadata standards into machine-actionable templates and automatically generating intuitive, standards-compliant interfaces, CEDAR empowers scientific communities to formalize and scale their metadata practices.

While the CEDAR Workbench provides a robust environment for designing and managing community-aligned metadata templates, its original deployment model required users to work within a dedicated platform. This separation often limited adoption, as researchers prefer to interact with metadata tools inside the systems that they already use to manage data. Platform developers, in turn, faced significant barriers in integrating external metadata tools, especially those requiring custom interface work or ongoing maintenance. The CEE addresses this gap by bringing CEDAR's full template-driven metadata capabilities directly into third-party environments—enabling structured, ontology-linked metadata to be authored and displayed within existing research platforms, without the need for bespoke user interface development.

COMPONENT ARCHITECTURE AND EMBEDDING

The CEE is implemented as a lightweight, standards-compliant Web Component designed for integration into a wide range of Web applications. It is distributed as a framework-agnostic package via 'npm', with no external dependencies. Once installed in an environment, the component can be invoked using the custom HTML tag 'cedar-embeddable-editor'.

Because it conforms to the Web Components standard, the CEE works identically across modern frontend frameworks—including React, Angular, and Vue—as well as in static HTML contexts. Developers can instantiate it either directly in the DOM or within an HTML iframe and supply the necessary inputs via custom parameters. Required inputs include a template (a CEDAR metadata template in JSON Schema format) and a configuration object, which controls language settings, display mode, and interface features. Optionally, an existing metadata instance may also be passed for editing or display.

This architecture allows the editor to be embedded in diverse platforms—from data repositories to mobile application interfaces—with minimal integration effort and without custom UI development.

TEMPLATE-DRIVEN UI GENERATION

At the core of the CEE is its ability to interpret and render CEDAR templates. As mentioned, these templates may include complex nested elements and single- or multi-valued fields. Templates also accommodate display specifications such as field visibility conditions and language-specific labels via language maps. When a template is passed to the CEE, it is interpreted at runtime and dynamically rendered into a fully functional metadata form. Any changes to the template—such as updated ontology bindings or newly required fields—are immediately reflected in the rendered form, eliminating the need for redeployment or manual schema-to-UI synchronization.

METADATA ACQUISITION AND OUTPUT

The CEE guides users through metadata entry in real time, using template-driven validation to ensure conformity with structural and semantic requirements. Required fields are flagged, value types are enforced, and inconsistent or incomplete records are surfaced during authoring via contextual feedback. As users complete a form, the CEE incrementally constructs a metadata instance that conforms precisely to the supplied template.

As discussed, all output is serialized as JSON-LD ([Sporny, Kellogg and Lanthaler, 2020](#)), a linked data format that preserves the structure of the metadata and encodes controlled term values with their corresponding vocabulary term or ontology reference. Completed metadata instances can be retrieved using the component's API, stored locally, sent to a backend, submitted to the CEDAR repository, or exported for downstream use. The use of JSON-LD ensures that resulting metadata adhere to a standard representation, are machine-readable, and can be easily processed by downstream systems, including analytic pipelines and data catalogs ([Figure 3](#)).

Beyond supporting structured metadata authoring, the CEDAR ecosystem also provides a pathway for assigning DOIs to metadata created with CEDAR templates. Through its integration with DataCite, CEDAR enables users to supply the descriptive information required for DOI registration using a DataCite-aligned metadata form, after which the DOI and its associated

metadata are registered and made publicly accessible. When a host platform does not offer its own landing pages, these metadata can be displayed through CEDAR OpenView, which provides a stable, human-readable view that aligns with DataCite’s expectations for DOI landing pages. This approach ensures that metadata produced through the CEE can be incorporated into persistent identifier workflows while supporting both machine-readable structure and durable public accessibility.

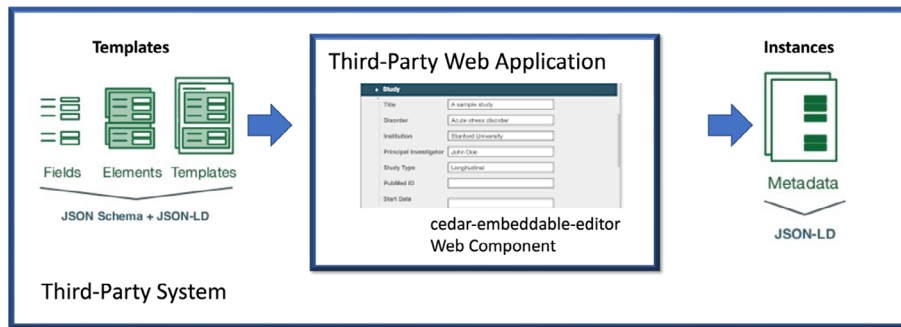


Figure 3 Integration workflow of the CEE within a third-party Web application. On the left, domain-specific metadata templates—composed of fields, elements, and templates defined using JSON Schema and JSON-LD—are created in a third-party system. These templates are rendered in the central component using the `cedar-embeddable-editor` Web Component, which allows users to input structured metadata directly within the host application. On the right, completed metadata instances are generated in JSON-LD format, ready for storage, reuse, or publication.

SEMANTIC INTEGRATION

A key feature of the CEE is its built-in support for ontology-based value selection. Fields can be linked to ontologies, ontology branches, and value sets hosted on the BioPortal ontology repository (Vendetti et al., 2025). During metadata entry, the editor suggests relevant terms—displayed with labels and synonyms—sourced from the configured ontologies and value sets. When a user selects a term, the editor records both its human-readable label and its IRI, ensuring semantic precision and enabling downstream reuse.

INTERACTION MODES AND ACCESSIBILITY

The CEE also supports integration with external authority systems for persistent identifier fields (Figure 4). Currently supported authorities include ORCID for identifying individual researchers, the Research Organization Registry (ROR) for research organizations, Research Resource Identifiers (RRIDs) for standardized biological resources, and PubMed IDs for linking to indexed publications. We have recently added the US Environmental Protection Agency CompTox API as an authority for selecting per- and polyfluoroalkyl substances (PFAS)—a class of synthetic chemicals widely studied in environmental science and toxicology. This integration enables researchers to link PFAS-related metadata fields to precise chemical identifiers maintained by the EPA, supporting more accurate reporting and downstream reuse in environmental datasets. Templates may also include fields for machine-interpretable licenses—such as SPDX identifiers or other controlled license vocabularies—allowing platforms to capture standardized licensing information in a form that is both human- and machine-readable. This improves the consistency of rights metadata and supports downstream automation in repository workflows and DOI-registration pipelines.

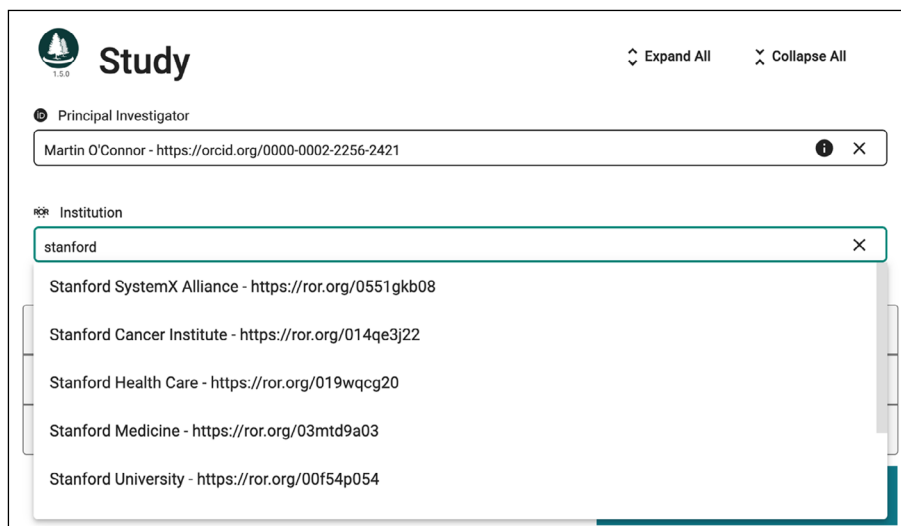


Figure 4 This image shows a metadata entry form rendered by the CEE, with the Principal Investigator field completed using a valid ORCID identifier. Below, the Institution Research Organization Registry (ROR) field shows the CEE’s autocomplete capability, allowing users to search and select standardized institutional identifiers. This external authority integration ensures precise, interoperable metadata by linking to authoritative external registries.

For these fields, the CEE performs REST-based searches against external authority services to autocomplete user input and to validate selections. These requests are routed through an intermediating microservice provided by the CEDAR system, which facilitates communication with multiple external APIs by handling authentication, standardizing request formats, and normalizing responses. When users select from the returned options, the CEE records globally resolvable, authoritative identifiers. Additional external authority fields are planned for future releases of the CEE. The CEE supports multiple operational modes to accommodate different usage scenarios:

1. **Form Entry Mode:** The default mode, used for creating new metadata instances.
2. **Edit Mode:** Allows modification of a previously authored metadata instance.
3. **View-Only Mode:** Displays a template or instance without permitting edits—useful for audit trails, data reviews, and read-only contexts such as metadata visualization platforms.

The CEE integrates visually with its host environment by inheriting local styles, maintaining a consistent user interface. Accessibility is a core feature: the CEE supports complete keyboard operability and screen reader compatibility via standardized WAI-ARIA role annotations ([W3C, 2021](#)).

The CEE ships with built-in support for multiple languages. Language selection is configured via the initialization object passed to the editor. Developers can add new translations by mirroring the structure of the provided English language map and updating the configuration. The editor gracefully falls back through available language maps when keys are missing, with diagnostics surfaced through a debug trace.

To support accurate metadata entry, the editor provides real-time validation. Color-coded prompts highlight missing required fields, datatype mismatches, and ontology term issues as the user interacts with the form. A collapsible validation ribbon summarizes all outstanding problems in a centralized view, enabling users to resolve issues before submission. For more comprehensive assessments, the CEE can generate structured data quality reports based on the template specification—flagging missing values, unselected controlled terms, or inconsistencies across the metadata record.

DEPLOYMENT AND CONFIGURATION

The CEE is designed for low-friction deployment. Because the UI is dynamically rendered based on centrally managed templates, updates to metadata standards—such as revised field definitions or updated value sets—are automatically reflected in the editor without any code changes to the host application.

The CEE's configuration object—passed as a JavaScript property or bound to the HTML tag—exposes fine-grained control over appearance and behavior, including:

1. Template and instance injection
2. Form mode (edit/view-only)
3. Language and localization paths
4. Display of headers, footers, and validation ribbons
5. Styling integration via scoped CSS classes and slot-based layout hooks

Templates used by the CEE are initially authored within the CEDAR Workbench by the third-party deployer—typically a developer, data manager, or community standards lead responsible for metadata integration within a given platform. The deployer uses the Workbench's visual editor to construct machine-actionable metadata templates, specifying field structure, cardinality, validation constraints, controlled vocabulary bindings, and language-localized labels. Once finalized, each template can be exported from CEDAR as a standalone JSON Schema document. Rather than relying on CEDAR to serve these templates at runtime, most third-party systems store and manage them locally—either embedding the JSON directly in their application logic or maintaining a versioned template registry within their infrastructure.

At runtime, the CEE receives the JSON Schema template as a property or API-delivered object and renders the corresponding metadata form on the client side. This design decouples form generation from CEDAR's live infrastructure, allowing the third-party platform to retain full control over which template versions are in use and when updates are deployed. The deployer is thus responsible not only for designing the template but also for maintaining its lifecycle—tracking changes, managing backward compatibility, and aligning with evolving community standards or platform requirements.

Similarly, all metadata instances authored using the CEE are retained and managed entirely by the host platform. As users complete and submit forms, the resulting metadata—structured as JSON-LD and semantically enriched with persistent identifiers and ontology-linked terms—are returned to the host application via the CEE's API. From there, the third-party system is responsible for storing, indexing, validating, and linking these metadata records with associated digital objects (e.g., datasets, software packages, protocols). CEDAR does not store or archive metadata created through the CEE unless explicitly configured to do so. This model helps ensure that integrations are lightweight and that metadata ownership, persistence, and discoverability remain under the control of the embedding platform.

The CEE supports styling integration through a set of scoped class names applied to UI elements, allowing developers to override fonts, spacing, or color schemes using host-defined CSS. Additionally, the component exposes elements that enable custom theming or layout control within the host application's design system. These features ensure that the editor blends visually with surrounding components and remains consistent across embedding environments.

The component has been deployed in environments built using React, Django, and custom JavaScript frameworks, and tested across all major browsers and a selection of mobile devices.

RESULTS

The CEE's lightweight deployment model and framework-agnostic architecture have made it suitable for integration across a variety of research platforms. Its ability to render rich, standards-based metadata forms without requiring platform-specific interface development has enabled adoption in diverse environments. These deployments demonstrate how the editor supports community-specific metadata practices and enables semantic interoperability without disrupting existing workflows.

Within the Open Science Framework (OSF) ([Foster and Deardorff, 2017](#)), the CEE is embedded into the metadata authoring workflow during project registration and dataset submission ([Olson and Corker, 2024](#)). As part of an initiative by the Center for Open Science to improve metadata quality and field-specific coverage, OSF first prompts researchers to complete general metadata describing the overall study—such as title, contributors, funding, and study purpose—before selecting a discipline-specific metadata template ([Figure 5](#)). Upon this selection, the CEE is invoked to render the corresponding structured form within the OSF interface ([Figure 6](#)). These templates, designed in collaboration with domain experts and curated in the CEDAR Workbench, include discipline-specific schemas for cognitive neuroscience, clinical trials, ecology, and other areas of active research. Each template encodes not only the fields required for rich documentation of study context and methods but also includes bindings to relevant ontologies through BioPortal—for example, recommending terms from the Cognitive Atlas or MeSH depending on the selected domain.

By embedding these templates directly through the CEE, OSF provides users with structured guidance that encourages the entry of consistent, semantically valid metadata. Users do not need to leave the OSF environment or install additional tools. The editor enforces constraints and offers intelligent autocompletion for ontology-linked fields in real time. The resulting JSON-LD output ensures that the metadata can be programmatically analyzed, reused by other platforms, or federated with disciplinary metadata repositories. This integration also lays the foundation for more granular, scalable metadata capture across OSF-hosted projects. By decoupling template logic from the platform's user interface code, OSF can support ongoing expansion of template coverage across fields without requiring frontend redevelopment—enabling a long-term strategy for increasing metadata richness and FAIR compliance across diverse scientific disciplines.

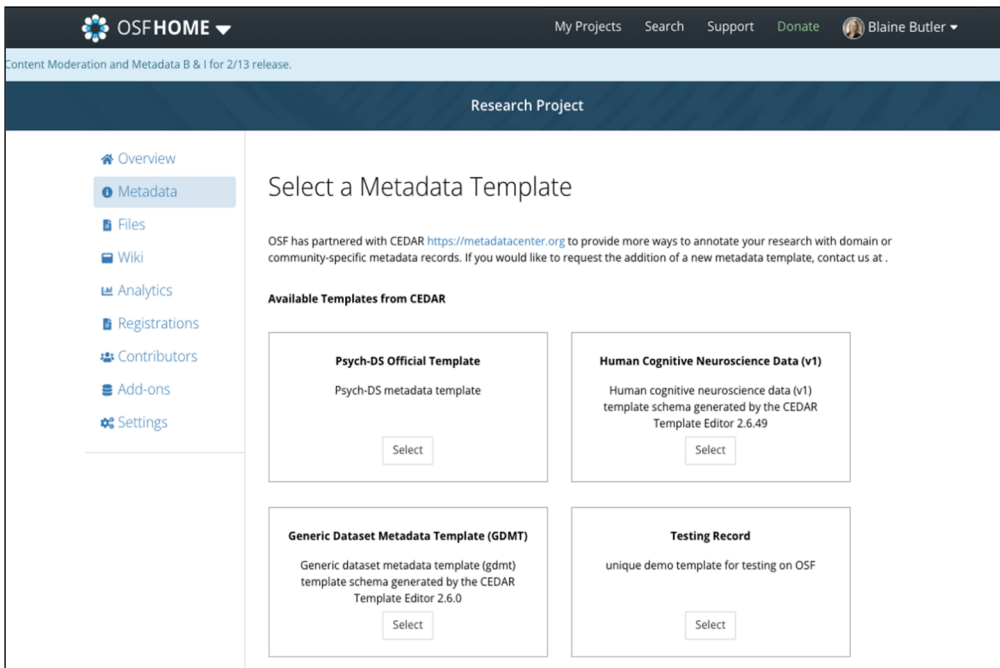


Figure 5 This OSF interface presents users with a selection of metadata templates defined using CEDAR. Each template—such as those for Psych-DS, human cognitive neuroscience data, or generic datasets—encodes a community- or domain-specific metadata standard. Upon selecting a template, users are presented with a structured metadata entry form rendered by the CEE, enabling guided, standards-compliant metadata creation directly within the OSF platform.

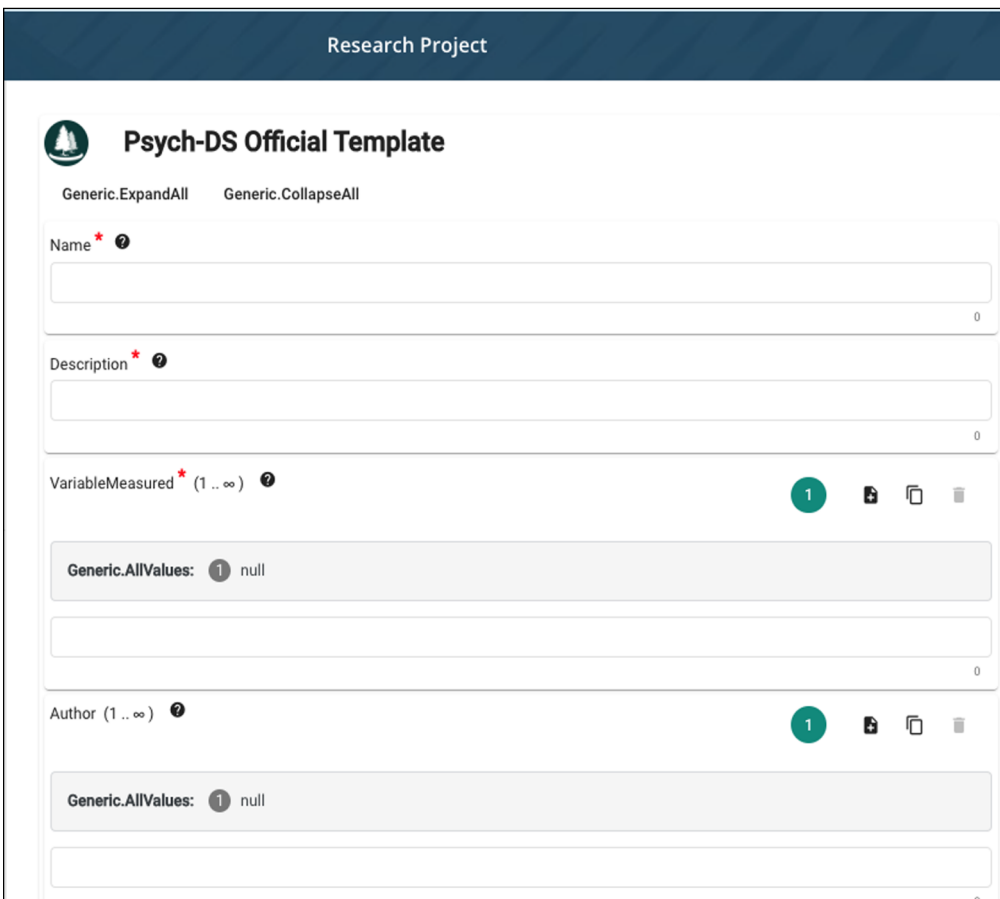


Figure 6 Psych-DS template rendered within the Open Science Framework (OSF) using the CEE. The form provides a structured interface for entering metadata fields defined by the Psych-DS metadata standard. By embedding the CEE, OSF enables users to create semantically rich, standards-compliant metadata records directly within their research workflows.

Dryad's (White et al., 2008) deployment of the CEE focuses on improving metadata quality across multiple domains, including neuroscience and high-throughput biology. In partnership with domain experts and CEDAR developers, Dryad first introduced a neuroscience-specific metadata template to capture detailed contextual information essential for reuse (Figure 7). This template includes fields for experimental paradigm, participant demographics, imaging modality, cognitive domain, and data acquisition protocols, many of which are linked to controlled vocabularies such as the Cognitive Atlas and OpenNeuro Vocabulary via BioPortal. More recently, Dryad also integrated a set of metadata templates developed by the HuBMAP Consortium (HuBMAP Consortium, 2025; Jain et al., 2023), which describe assay workflows, sample characteristics, and tissue metadata for cellular-resolution datasets. These templates

allow Dryad to support metadata capture for a broader range of biological submissions. During dataset submission, users are prompted with CEDAR-rendered forms that reflect the selected template. The editor enforces constraints, provides ontology-based term recommendations in real time, and exports completed records as JSON-LD. The resulting metadata are stored by Dryad alongside the associated dataset, making them available for indexing, discovery, and downstream reuse. Dryad staff noted that these integrations lowered the barrier to submitting specialized domain-specific data, helping to align metadata practices with the FAIR principles and community standards (Lippincott, 2024).

The image shows a screenshot of a web-based metadata form. The main content area is titled 'Preprocessing' and contains several sections: 'Preprocessing status' with checkboxes for 'Preprocessed' and 'Raw'; a text field for 'Information about the preprocess used to produce the dataset'; a 'Preprocessing Pipeline' section with a text input and a '1' icon; a text field for 'Provide a link to the location where the preprocessing code is hosted'; and a 'Preprocessing Script' section with a text input and a '1' icon. Below these are dropdown menus for 'Standard', 'Source dataset', 'Experiment', and 'Analysis'. On the right side, there is a sidebar titled 'related to this Data' with a list of items, each with a 'remove' button. At the bottom of the sidebar, it says 'All progress saved' and 'Proceed to README'. The left sidebar contains 'Standardized metadata' and 'Related works' sections. The footer includes 'Privacy', 'Accessibility', 'Terms', and 'Copyright (c) 2024 Dryad'.

Figure 7 Neuroscience metadata template rendered within the Dryad data repository using the CEE. The displayed section captures structured metadata related to data preprocessing, including status, pipeline references, and script locations. By embedding CEE, Dryad enables researchers to input detailed, standards-based metadata directly within the dataset submission workflow, supporting reproducibility and alignment with open data best practices.

The RADx and HuBMAP programs use the CEE in a read-only mode to display templates and metadata authored externally but rendered for user-friendly inspection. In these contexts, the CEE provides a consistent interface for reviewing metadata associated with complex biological datasets. The RADx Data Hub (Martinez-Romero et al., 2025), for instance, uses the CEE to present metadata records for COVID-19 diagnostic projects. The visual rendering of nested, structured information improves readability. In HuBMAP, the CEE serves as the authoritative public interface for viewing all metadata templates used across the consortium (Martínez-Romero et al., 2025). Each template is rendered in a readable, structured format directly from its machine-actionable specification, enabling researchers, tool developers, and reviewers to inspect the exact metadata requirements in a standardized, accessible form.

A unifying feature across all CEE deployments is the use of JSON-LD as the output format for metadata instances. Regardless of the domain or platform, metadata authored through the CEE are encoded in a semantically structured, machine-readable format that aligns with linked data principles. This common representation enables metadata to be validated, indexed, and transformed using standard Web-based tools, while preserving field-level semantics and ontology references. Platform developers across OSF, Dryad, and HuBMAP emphasized the value of this consistency—not only for immediate use in submission pipelines and UI display, but also for long-term goals such as metadata federation, reuse across repositories, and automated schema evolution. JSON-LD also facilitates integration with persistent identifier systems (e.g., ORCID, ROR) and supports downstream applications including DOI registration, search indexing, and metadata enrichment via external services.

Across these deployments, a consistent theme has been the need for controlled, predictable adoption of new CEE releases. Embedding platforms maintain their own UI logic, workflow constraints, and metadata pipelines, and therefore cannot absorb upstream code changes without first verifying compatibility. To support this requirement, the CEE now follows a formal release process with stable npm-distributed versions, a public changelog, and advance notice for major or breaking changes. As part of integration testing, partner platforms evaluated new versions of the CEE across common environments—including React-, Django-, and custom JavaScript-based systems—and validated behavior in all major browsers (Chrome, Firefox, Safari, and Edge) as well as selected mobile devices. This coordinated process allows platforms to adopt updates intentionally, assess interface stability under their own configurations, and manage transitions between template versions without disrupting ongoing metadata submission workflows. These shared deployment patterns have shaped the current governance model for the CEE's versioning and support lifecycle.

Deployments also surfaced practical concerns around performance, accessibility, and page-load behavior, particularly for platforms that must adhere to strict indexing or accessibility criteria. Performance benchmarking using tools such as Google PageSpeed Insights and Lighthouse was used across sites to evaluate load times, bundle size, and render behavior, revealing opportunities to optimize resource loading, reduce dependencies, and refine metadata tagging. In response, the CEE team implemented targeted improvements—including local font embedding, dependency reductions, enhanced meta tags, and removal of heavier libraries—which yielded measurable gains in desktop and mobile performance scores. Accessibility assessments using automated frameworks such as axe-core provided additional validation, prompting refinements to WAI-ARIA annotations, keyboard navigation behavior, and screen-reader cues to better align with WCAG AA expectations. These iterative refinements, informed by feedback and testing practices shared across multiple deployment partners, have strengthened the robustness, accessibility, and production readiness of the CEE.

Together, these deployment experiences demonstrate that the CEE can lower the barrier to structured metadata authoring and provides a practical path for extending FAIR practices across a wide range of research infrastructures.

DISCUSSION

The CEE brings structured, standards-based metadata authoring directly into third-party platforms through a lightweight, framework-agnostic Web Component. It dynamically renders forms from machine-actionable templates, enforces community-defined validation rules, supports ontology-based value selection and external identifier resolution, and outputs metadata in semantically rich JSON-LD format. By encapsulating all metadata logic within the component itself, the CEE eliminates the need for custom interface development while ensuring consistency with evolving metadata standards.

The adoption and integration of the CEE across multiple platforms underscores the growing value placed on metadata solutions that combine domain-specific adherence to standards with flexible deployment options. The CEE demonstrates that it is possible to deliver structured, standards-aligned interfaces for authoring metadata that are both technically robust and easy to embed within existing research workflows. By leveraging CEDAR's machine-actionable templates and its ontology infrastructure, the editor makes high-quality metadata entry accessible to researchers without requiring them to leave the environments in which they already manage and publish their work.

Compared to traditional metadata tools—whether general-purpose form builders or discipline-specific systems such as ISA-Tab ([Rocca-Serra et al., 2010](#)) or EML ([Fegraus et al., 2005](#))—the CEE strikes a unique balance between expressiveness, usability, and integration. Generic tools often lack the schema flexibility or ontology linkage required to support rich metadata with appropriate persistent identifiers, whereas more powerful domain-specific tools tend to be standalone applications with steep learning curves and limited adaptability. The CEE's lightweight, component-based architecture offers a middle path, enabling host platforms to support multiple community-defined schemas with minimal development burden. Its 'no UI coding' paradigm lowers the barrier for third-party integration, allowing developers to rely on centrally managed templates that evolve independently of platform-specific user interfaces.

The architecture of the CEE highlights a broader design principle applicable well beyond metadata: decoupling form logic from platform code enables dynamic, standards-driven interfaces that can evolve independently of the systems embedding them. In any domain where users complete structured forms—whether for scientific metadata, grant applications, clinical records, or regulatory submissions—this separation allows user interfaces to reflect updated specifications, validation rules, and vocabularies without requiring software redeployment. The CEE's use of centrally defined, machine-actionable templates and its runtime rendering model illustrate how form-filling systems can become both more maintainable and more responsive to changing community needs. By outputting structured, semantically rich data—in this case, JSON-LD—the editor ensures compatibility with downstream processing, indexing, and integration tasks. This approach points to a more generalized model for building intelligent, standards-aware form interfaces across a wide range of applications.

The deployment of CEDAR templates developed by the HuBMAP Consortium across both Dryad and the Open Science Framework (OSF) further illustrates the portability and ecosystem value of the CEE model. In collaboration with the HuBMAP metadata team, a set of ten templates—primarily covering assay descriptions, sample metadata, and tissue characterization workflows—were authored in the CEDAR Workbench using domain-specific schema elements and controlled vocabulary bindings ([HuBMAP Consortium, 2025](#)). These templates, originally designed to capture rich metadata for high-throughput biology experiments, have been integrated into both Dryad and OSF, where they are used to guide metadata entry during dataset submission and project registration. Notably, the same set of exported JSON Schema templates is being reused in both platforms without modification, with each system embedding the CEE and applying its own local configuration and styling.

These deployments demonstrate a core strength of the CEE architecture: templates are fully portable and decoupled from any specific implementation or backend. Because the editor renders forms dynamically from JSON Schema templates at runtime and does not rely on server-side logic or templating engines, a single template can be reused across multiple platforms with no duplication of development effort. For research communities, this means that a well-defined metadata standard can be expressed once—using CEDAR's authoring tools—and then embedded consistently across the tools their researchers already use ([Musen et al., 2022](#); [Musen et al., 2025](#)). For platform developers, it enables schema alignment across systems while preserving autonomy over data flow, storage, and integration. As the HuBMAP deployment shows, shared metadata standards can now be operationalized across generalist and domain-specific infrastructure alike, promoting both metadata harmonization and localized control. This portability also creates new opportunities for maintaining authoritative, version-controlled template libraries that span multiple services and disciplines.

The results across deployments highlight that embedding structured metadata authoring into research workflows leads to more complete, consistent, and semantically valid metadata. The OSF and Dryad integrations show how the CEE can adapt to domain-specific needs while maintaining a unified implementation model. Meanwhile, the use of the CEE in read-only mode by RADx and HuBMAP illustrates the value of a consistent rendering layer for metadata review and quality control. Across these deployments, the feedback from researchers and platform administrators consistently emphasizes improved metadata quality, enhanced user experience, and reduced maintenance costs.

The CEE architecture enables a clean separation between metadata governance and platform implementation. Templates are defined and versioned within the CEDAR Workbench by community curators, standards developers, or platform-specific leads, while the embedding platforms remain responsible for how and when those templates are integrated into production workflows. This separation allows domain experts to evolve metadata standards over time—adding fields, adjusting validation logic, or updating ontology bindings—without requiring downstream systems to redevelop their user interfaces. In effect, the CEE makes it possible to delegate stewardship of semantic and structural metadata rules to the appropriate knowledge holders, while preserving implementation independence and stability for platform developers. This distributed governance model aligns with real-world research infrastructure, where metadata policies are negotiated across institutions, consortia, and sponsors, but must be supported by a diverse array of technical systems.

Long-term sustainability has been a central design consideration for the CEDAR ecosystem. The CEE is part of a broader suite of open-source tools developed within Stanford University's Division of Computational Medicine, which also maintains long-running community resources such as Protégé (initiated in 1990) and BioPortal (operational since 2005). CEDAR itself has been continuously supported since 2015 through a combination of federal grants and institutional investment, and its codebase is openly maintained on GitHub with active contributions from both the core team and external collaborators. This track record reflects a mature governance model in which tools are versioned, publicly documented, and sustained over decades, giving adopters confidence that the CEE and associated services will continue to be maintained, improved, and supported by an established development and stewardship community.

Because the CEE renders metadata forms from machine-actionable templates, the JSON-LD outputs produced by different communities share a common structural foundation even when their schemas diverge. Core elements such as titles, creators, dates, identifiers, and descriptive fields often appear across templates, allowing downstream systems to identify these commonalities and apply cross-template mappings for aggregation or federation workflows. While CEDAR does not impose a universal schema, the explicit definition of each field—together with its ontology bindings and persistent identifiers—makes such mappings transparent and easier to maintain than in systems where forms and semantics are embedded directly in platform code. This approach supports a balance between expressiveness and interoperability: communities can define richly detailed, domain-specific templates, while still enabling consistent extraction of shared metadata elements for indexing, search, or cross-repository integration.

Although DOI creation within the CEDAR ecosystem is handled through a DataCite-aligned metadata form, as described earlier, template designers still play a central role in determining which template fields support those DOI workflows. In practice, communities that intend their templates to feed DOI registration simply include the required DataCite descriptive elements as fields or assemble them from reusable components that reflect common DataCite metadata elements. CEDAR does not currently provide an automated way to declare field-to-DataCite mappings, but the explicit, machine-actionable structure of templates makes these relationships transparent and straightforward to apply when a DOI is later registered through the CEDAR-DataCite integration. Designers who need richer disciplinary detail can extend these core components with additional fields and ontology bindings tailored to their domain. Planned enhancements to template inheritance and shared component libraries are expected to make it even easier to build new domain-specific templates on top of a DataCite-aligned foundation.

While the FAIR principles have become a cornerstone of research data policy, their implementation often remains aspirational, especially in domains lacking dedicated tooling. The CEE offers a concrete mechanism for operationalizing core aspects of FAIR by embedding machine-actionable metadata logic directly into submission workflows. Through real-time validation, persistent identifier resolution (e.g., ORCID, ROR), and ontology-based field constraints, the CEE can help ensure that the metadata researchers submit are not only structured and complete, but semantically meaningful and machine-interpretable. Because these features are encoded in reusable templates and automatically enforced by the editor, platforms can raise the metadata quality floor without burdening researchers with additional training or manual standards compliance. In this sense, the CEE shifts FAIR from a set of external recommendations into an embedded capability—making adherence to best practices a natural outcome of the submission process.

The CEE also serves as an effective bridge between general-purpose data platforms and the more specialized needs of disciplinary repositories. Many open data environments face a tension between broad accessibility and domain specificity: generalist platforms must support heterogeneous users and content types, while domain repositories demand high-quality, semantically rich metadata. By embedding the CEE and adopting discipline-specific templates, generalist platforms such as Dryad and OSF can support structured metadata capture without fragmenting their infrastructure or duplicating interface logic. This capability allows researchers to submit community-aligned metadata within familiar environments, while ensuring that those records remain compatible with downstream curation, indexing, or federation workflows in specialized ecosystems. As more research platforms converge on shared template libraries,

the CEE provides a scalable mechanism for harmonizing metadata practices across the broader research data landscape. Although our deployment examples focus on datasets, this same template-driven approach applies equally to other digital research objects—including instruments, protocols, software, and project-level records—where structured, standards-aligned metadata are needed.

Nevertheless, there are limitations. While the editor supports complex schemas, some scientific domains may still require capabilities not yet expressible in CEDAR's current metadata model. For example, CEE does not currently support advanced conditional logic such as dynamic field branching or question-skipping behavior based on prior responses—functionality that is often critical in clinical studies, environmental assessments, and longitudinal data collection instruments. Similarly, cross-field validation rules that involve dependencies across multiple metadata fields are not yet supported. These features are planned for future development, along with more expressive constraint languages and enhanced form logic.

Additionally, while the editor provides robust form-based entry, some users prefer working in tabular environments—particularly for bulk metadata entry or for records that naturally align with spreadsheet structures. Integration with spreadsheet-based metadata-entry tools, along with support for round-tripping between structured templates and spreadsheet formats, is an area of ongoing work (O'Connor et al., 2025). Another promising direction is the automated extraction of metadata from primary data files (e.g., imaging headers, instrument logs), which could prepopulate template fields and reduce manual burden, while preserving structured output. Addressing these limitations will help extend the applicability of the CEE to an even broader set of metadata use cases and scientific workflows.

By aligning a flexible technical architecture with a governance model for community-developed templates, the CEE offers a practical approach for how structured, standards-based metadata creation can be embedded across the research ecosystem. As adoption grows, we anticipate that the CEE will contribute not only to the quality and consistency of metadata, but also to the normalization of FAIR practices in everyday research workflows.

CONCLUSIONS

The CEE is a lightweight, interoperable Web Component that brings standards-based, ontology-linked metadata authoring and visualization directly into third-party platforms. Developed within the CEDAR Workbench ecosystem, the CEE leverages machine-actionable metadata templates to render structured, semantically rich metadata forms without requiring custom user interface code. The editor has been integrated into a variety of platforms such as Dryad and the Open Science Framework, supporting both Web and mobile environments. A key distinction of the CEE is its ability to generate forms directly from template specifications, incorporate ontology-based value selection, and output structured JSON-LD—all while sparing platforms the need to build or maintain custom metadata interfaces. This combination of template-driven UI generation, semantic grounding, and zero-UI-coding deployment is not typical of existing metadata tools and enables consistent, standards-aligned metadata capture across heterogeneous environments. By adhering to the same JSON-LD-based metadata model as CEDAR, the CEE ensures consistency across systems and automatically reflects updates to metadata specifications. Underpinned by the broader CEDAR architecture—which includes tools for template design, metadata validation, ontology integration via BioPortal, and spreadsheet interoperability—the CEE enables scientific communities to encode and enforce metadata reporting guidelines in a reusable format. In doing so, it allows communities to deliver high-quality metadata tools with reduced engineering effort and provides a practical path toward making FAIR metadata the default rather than the exception.

AI TOOL USE DISCLOSURE

Generative AI tools were used to assist with grammar refinement and to provide editorial feedback on structure and clarity. No content was generated autonomously by AI tools; all substantive contributions were made by the authors.

The CEDAR Workbench and the CEE source code are publicly available under the BSD 2-Clause License and maintained by the Division of Computational Medicine in Stanford University. The CEE software is available for execution at <https://github.com/metadatacenter/cedar-embeddable-editor>. All CEDAR Workbench source code is available from the Metadata Center landing page at <https://github.com/metadatacenter>. Metadata templates referenced in the Dryad and OSF deployments are available via the respective platform repositories or by request. All deployments use publicly available ontologies and value sets accessible through the BioPortal ontology repository at <https://bioportal.bioontology.org>.

ETHICS AND CONSENT

This research did not involve human participants, personal data, or animal subjects and therefore did not require institutional ethical review.

FUNDING INFORMATION

This work was supported in part by Grant 2134956 from the National Science Foundation, by a grant from the Templeton World Charitable Foundation, and by Award OT2 DB000009 and Grant U24 GM143402 from the National Institutes of Health.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

MJO'C and MMR led the conceptualization and development of the CEE. ALE, MUA, and MVD contributed to deployment, implementation, and architecture. MAM supervised the project. All authors reviewed and approved the final manuscript.

AUTHOR AFFILIATIONS

Martin J. O'Connor  orcid.org/0000-0002-2256-2421

Division of Computational Medicine, Stanford, CA 94305, USA

Marcos Martínez-Romero  orcid.org/0000-0002-9814-3258

Division of Computational Medicine, Stanford, CA 94305, USA

Attila L. Egyedi  orcid.org/0000-0003-0730-5053

Division of Computational Medicine, Stanford, CA 94305, USA

Mete U. Akdogan  orcid.org/0000-0003-1520-1481

Division of Computational Medicine, Stanford, CA 94305, USA

Michael V. Dorf  orcid.org/0009-0005-2374-6344

Division of Computational Medicine, Stanford, CA 94305, USA

Mark A. Musen  orcid.org/0000-0003-3325-793X

Division of Computational Medicine, Stanford, CA 94305, USA

REFERENCES

- Baker, R.G., Koroshetz, W. and Volkow, N.D.** (2021) 'The helping to end addiction long-term (HEAL) initiative of the National Institutes of Health', *JAMA*, 326(11), pp. 1005–1006. Available at: <https://doi.org/10.1001/jama.2021.13300>
- Berkley, C., Jones, M.B. and Boettiger, C.** (2001) *Metatool: A metadata creation tool for ecological data*. National Center for Ecological Analysis and Synthesis (NCEAS), University of California, Santa Barbara.
- Cellular Senescence Network (SenNet).** (2025) *Metadata*. Available at: <https://docs.sennetconsortium.org/libraries/ingest-validation-tools/schemas/source-murine/> (Accessed: 17 November 2025).
- DataCite Metadata Working Group.** (2024) *DataCite Metadata Schema Documentation for the publication and citation of research data and other research outputs (Version 4.6)*. DataCite e.V. Available at: <https://doi.org/10.14454/mzv1-5b55> (Accessed: 17 November 2025).

- Egyedi, A.L. et al.** (2018) 'Using semantic technologies to enhance metadata submissions to public repositories in biomedicine', *Semantic Web Applications and Tools for Health Care and Life Sciences (SWAT4 LS)*.
- Fegraus, E.H. et al.** (2005) 'Maximizing the value of ecological data with structured metadata: An introduction to Ecological Metadata Language (EML) and principles for metadata creation', *Bulletin of the Ecological Society of America*, 86(3), pp. 158–168. Available at: [https://doi.org/10.1890/0012-9623\(2005\)86\[158:MTVOED\]2.0.CO;2](https://doi.org/10.1890/0012-9623(2005)86[158:MTVOED]2.0.CO;2)
- Foster, E.D. and Deardorff, A.** (2017) 'Open Science Framework (OSF)', *Journal of the Medical Library Association*, 105(2), pp. 203–206. Available at: <https://doi.org/10.5195/jmla.2017.88>
- Giglia, E. et al.** (2020) 'Introducing the Research Organization Registry (ROR)', *Insights*, 33(1), p. 9.
- Gonçalves, R.S. et al.** (2017) 'The CEDAR Workbench: An ontology-assisted environment for authoring Metadata that describe scientific experiments', *Semantic Web ISWC*, 10588, pp. 103–110. Available at: https://doi.org/10.1007/978-3-319-68204-4_10
- Haak, L.L. et al.** (2012) 'ORCID: A system to uniquely identify researchers', *Learned Publishing*, 25(4), pp. 259–264. Available at: <https://doi.org/10.1087/20120404>
- Higgins, D., Berkley, C. and Jones, M.B.** (2002) 'Managing heterogeneous ecological data using Morpho', in *Proceedings of the 14th International Conference on Scientific and Statistical Database Management*.
- HuBMAP Consortium.** (2019) 'The human body at cellular resolution: the NIH Human BioMolecular Atlas Program', *Nature*, 574(7777), pp. 187–192. Available at: <https://doi.org/10.1038/s41586-019-1629-x>
- HuBMAP Consortium.** (2025) *HuBMAP Metadata Specifications*. Available at: <https://docs.hubmapconsortium.org/metadata> (Accessed: 17 November 2025).
- Jain, S. et al.** (2023) 'Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP)', *Nature Cell Biology*, 25(8), 1089–1100. Available at: <https://doi.org/10.1038/s41556-023-01194-w>
- Lippincott, S.** (2024) 'New at Dryad: Empowering reuse of cognitive neuroscience data', *Dryad News*, 9 April. Available at: <https://blog.datadryad.org/2024/04/09/new-at-dryad-empowering-reuse-of-cognitive-neuroscience-data/> (Accessed: 17 November 2025).
- Martínez-Romero, M. et al.** (2025) 'A Cloud-Based Platform for Harmonized COVID-19 Data: Design and Implementation of the RADx Data Hub', *JMIR Public Health Surveillance*, 11. Available at: <https://doi.org/10.2196/72677>
- Musen, M.A. et al.** (2015) 'The Center for Expanded Data Annotation and Retrieval', *Journal of the American Medical Informatics Association*, 22(6). Available at: <https://doi.org/10.1093/jamia/ocv048>
- Musen, M.A. et al.** (2022) 'Modeling community standards for metadata as templates makes data FAIR', *Scientific Data*, 9, p. 696. Available at: <https://doi.org/10.1038/s41597-022-01815-3>
- Musen, M.A., O'Connor, M.J., Hardi, J. and Martínez-Romero, M.** (2025) 'Knowledge engineering for open science: building and deploying knowledge bases for metadata standards', *AI Magazine*, 47(1). Available at: <https://doi.org/10.1002/aaai.70048>
- O'Connor, M.J. et al.** (2016) 'An Open Repository Model for acquiring knowledge about scientific experiments', *20th International Conference on Knowledge Engineering and Knowledge Management*. 4 November. Springer, Cham, pp. 762–777. Available at: https://doi.org/10.1007/978-3-319-49004-5_49
- O'Connor, M.J. et al.** (2025) 'Ensuring adherence to standards in experiment-related Metadata entered via spreadsheets', *Scientific Data*, 12, p. 265. Available at: <https://doi.org/10.1038/s41597-025-04589-6>
- Olson, E. and Corker, K.** (2024) 'Making it easy to make specialized research FAIR: Introducing field-specific metadata templates for OSF', *Center for Open Science*, 18 April. Available at: <https://www.cos.io/blog/cedar-embeddable-editor> (Accessed: 17 November 2025).
- Rocca-Serra, P. et al.** (2010) 'ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level', *Bioinformatics*, 26(18). Available at: <https://doi.org/10.1093/bioinformatics/btq415>
- Sansone, S.A. et al.** (2019) 'FAIRsharing as a community approach to standards, repositories and policies', *Nature Biotechnology*, 37(4), pp. 358–367. Available at: <https://doi.org/10.1038/s41587-019-0080-8>
- Soiland-Reyes, S. et al.** (2022) 'Packaging research artefacts with RO-Crate', *Data Science*, 5(2), pp. 97–138. Available at: <https://doi.org/10.3233/DS-210053>
- Sporny, M., Kellogg, G. and Lanthaler, M.** (2020) *JSON-LD 1.1: A JSON-based Serialization for Linked Data* (W3C Recommendation). Available at: <https://www.w3.org/TR/json-ld11/> (Accessed: 17 November 2025).
- Steffel, N.** (2011) 'Better, faster, stronger: Building a better Dublin Core Generator', *International Conference on Dublin Core and Metadata Applications*.
- Vendetti, J. et al.** (2025) 'BioPortal: An open community resource for sharing, searching, and utilizing biomedical ontologies', *Nucleic Acids Research*, 53(W1). Available at: <https://doi.org/10.1093/nar/gkaf402>
- Verborgh, R. and De Wilde, M.** (2013) *Using OpenRefine*. Packt Publishing.

- White, H.C. et al.** (2008) 'The Dryad Data Repository: A Singapore Framework Metadata Architecture in a DSpace Environment', *International Conference on Dublin Core and Metadata Applications*.
- Wilkinson, M. et al.** (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3, p. 160018.
- Williams, A.J. et al.** (2017) 'The CompTox Chemistry Dashboard: A community data resource for environmental chemistry', *Journal of Cheminformatics*, 9, 61. Available at: <https://doi.org/10.1186/s13321-017-0247-6>
- W3C.** (2021) *Accessible Rich Internet Applications (WAI-ARIA) 1.2* (W3C Recommendation). Available at: <https://www.w3.org/TR/wai-aria-1.2/> (Accessed: 17 November 2025).
- ZonMw.** (2024) *First steps towards data FAIRification in COVID-19 research*. Available at: <https://www.zonmw.nl/en/article/first-steps-towards-data-fairification-covid-19-research> (Accessed: 17 November 2025).

TO CITE THIS ARTICLE:

O'Connor, M.J., Martínez-Romero, M., Egyedi, A.L., Akdogan, M.U., Dorf, M.V. and Musen, M.A. 2026 Author Once, Publish Everywhere: Portable Metadata Authoring with the CEDAR Embeddable Editor. *Data Science Journal*, 25: 2, pp. 1–18. DOI: <https://doi.org/10.5334/dsj-2026-002>

Submitted: 16 July 2025

Accepted: 17 December 2025

Published: 27 January 2026

COPYRIGHT:

© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.