AHRQ CEDAR

Option Period Final Project Report

Agency for Healthcare Research and Quality
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 75FCMC18D0047

Prepared by:

CMS Alliance to Modernize Healthcare (The Health FFRDC)

A Federally Funded Research and Development Center

AHRQ Publication 23-0080-3-EF
September 2023

Disclaimer of Conflict of Interest

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Funding Statement

This project is presently funded under contract/grant number 75FCMC18D0047 from the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (HHS). The opinions expressed in this document are those of the authors and do not reflect the official position of AHRQ or HHS.

Suggested Citation

CMS Alliance to Modernize Healthcare (The Health FFRDC). AHRQ CEDAR: Option Period Final Project Report. Prepared under Contract No. 75FCMC18D0047. AHRQ Publication No. 23-0080-3-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2023.

Acknowledgments

The MITRE team thanks and recognizes members of AHRQ who provided information, collaboration, insight, and guidance throughout the project, including:

- Dr. Lionel Banez
- Steve Bernstein
- Dr. Christine Chang
- Dr. Roland Gamache
- Dr. Edwin Lomotan
- Dr. Justin Mills
- Mary Nix
- James Swiger
- Dr. Mario Terán

MITRE and AHRQ also extend thanks to the many end users and subject matter experts who contributed to CEDAR's outreach and engagement activities.

Executive Summary

In 2020, the Center for Evidence and Practice Improvement (CEPI) at the Agency for Healthcare Research and Quality (AHRQ) established the CEPI Evidence Discovery and Retrieval (CEDAR) project in conjunction with the Centers for Medicare & Medicaid Services (CMS) Alliance to Modernize Healthcare (Health FFRDC), operated by MITRE. CEDAR's purpose is to assist AHRQ in disseminating patient-centered outcomes research (PCOR) evidence and make the CEPI repositories of research evidence more findable, accessible, interoperable, and reusable (FAIR). After a successful Base Period (2020 – 2022) in which the CEDAR tool was developed, an Option Period (2022 – 2023) included the public launch of CEDAR and activities to understand its utility to future users of its content.

Project Activities

During the Option Period, MITRE undertook the following activities to expand upon the Base Period's accomplishments.

- The team updated the Environmental Scan surveying the landscape of information retrieval, identifying two major themes:
    1. CEDAR end users (clinicians, researchers, developers, and knowledge managers) access clinical research and evidence in different ways and for different purposes, and they desire a varying quantity and quality of the search results.
    2. All target end users expressed a need for healthcare research and evidence that they can trust to be accurate and impartial.
- The team expanded upon the functionality of the CEDAR prototype developed in the Base Period by enhancing the importer functionality, administration application, application programming interface (API), and demonstration user interface following an Agile development process. This tool was publicly launched for open access in November 2022.
- In partnership with Computable Publishing, LLC, the team:
    1. Conducted a multicomponent pilot to use the CEDAR API in an applied setting; this included the expansion of CEDAR functionality in the partnering effort to demonstrate a bidirectional data exchange capability that increased the number of accessible resources.
    2. Hosted focus groups with varied target end-users.
    3. Assessed the FAIRness of the CEDAR API as well as Computable Publishing, LLC's application.
- The Pilot activities assessed CEDAR's acceptability and maturity to inform CEDAR adoption and use.
- Outreach efforts uncovered several considerations, including the need to better understand CEDAR's problem statement, as well as CEDAR's usefulness for target end users. Should AHRQ decide to advance CEDAR, then it should apply these lessons to future undertakings.

- A coordinated team executed activities in accordance with specific timelines in the management of the project.

The project team successfully created CEDAR, publicly released the tool, and modified it based on feedback from target end users. The first 3 years of CEDAR development and testing serve as a demonstration project. This work will further AHRQ's mission and impact the healthcare community either by continuing to evolve CEDAR with expanded functionality for target end-user needs, or by applying lessons learned from this experience in future initiatives.

Lessons Learned and Recommendations

The CEDAR demonstration project aimed to improve user access to trusted clinical research and evidence from AHRQ CEPI resources by making them available through a single search, which is consistent with AHRQ's legislative mandate from the Affordable Care Act and the PCOR Trust Fund to disseminate evidence. To further encourage the use of CEDAR beyond this demonstration, AHRQ should develop a clear and forward-looking value proposition statement for CEDAR that defines the specific problems addressed by CEDAR. The value proposition should also detail how CEDAR offers benefits that are superior to other available information-retrieval options. The value proposition may need to be tailored to address different end-user types, and their specific interface needs. Input from stakeholders suggests that this may be achieved by expanding on the trust inherent in AHRQ (and therefore in CEDAR) by growing the quantity of indexed information, while preserving its high quality and timeliness, integrating CEDAR with emerging large-language models, and designing the user interface to meet target end users in their existing workflow.

Contents

Figures

Tables

Introduction

The Agency for Healthcare Research and Quality (AHRQ) disseminates findings from patient-centered outcomes research (PCOR), along with other types of research evidence into clinical practice through clinical decision support (CDS). The AHRQ Center for Evidence and Practice Improvement (CEPI) maintains public repositories of research evidence and PCOR findings. Programs that support these repositories encompass the Systematic Review Data Repository (SRDR), the Effective Health Care (EHC) Program, Evidence-based Practice Centers (EPC) Reports, CDS Connect, and the U.S. Preventive Services Task Force (USPSTF).

In 2020, AHRQ began work with the CMS Alliance to Modernize Healthcare (the Health FFRDC), operated by The MITRE Corporation (MITRE), to establish the CEPI Evidence Discovery And Retrieval (CEDAR) project to make the CEPI repositories more findable, accessible, interoperable, and reusable (FAIR).[1]

During the first 2 years of the project (i.e., the 2020–2022 Base Period) the CEDAR project team developed a standards-based application programming interface (API) capable of indexing resources—inclusive of repositories and artifacts (henceforth described as "resources")—from the CEPI PCOR repositories through a single software-accessible endpoint, as well as a Reference Implementation (RI). In partnership with the American Academy of Family Physicians (AAFP), the project team piloted a prototype to demonstrate the feasibility for a third-party developer to create their own CEDAR API application.

This report provides an overview of the following accomplishments for the CEDAR Project in the year following the Base Period (i.e., the 2023 Option Period).

- Updating the Environmental Scan to reflect the latest industry standards and best practices (e.g., those defined by Health Level Seven [HL7]).
- Enhancing the RI to reflect FAIR guiding principles.
- Conducting a pilot.
- Engaging in outreach activities (including the public launch of CEDAR).

All publicly available reports related to CEDAR can be found on the AHRQ CEDAR Project page.

Objectives and Outcomes

The desired outcome of CEDAR was to provide clinicians, researchers, implementers, patients, and others with timely and efficient access to PCOR findings so that they can incorporate up-to-date evidence in healthcare decision making.

From the development of CEDAR in the Base Period to the public launch of CEDAR in the Option Period, the project objectives evolved from focusing on dissemination of the API to assessing CEDAR's ability to make CEPI resources more FAIR.

Project Tasks

The Option Period continued the following five core tasks:

- Update Environmental Scan and Gap Analysis
- Further develop the RI
- Conduct Pilot
- Conduct Outreach
- Project Management

Notable Accomplishments

During the Option Period, the project team achieved several significant milestones and notable accomplishments.

- **Public Launch:** CEDAR was publicly launched at the American Medical Informatics Association (AMIA) Annual Symposium in November 2022. The project team provided a technical demonstration of its capabilities and met with potential users and implementers of CEDAR. These interactions with the AMIA community not only marked the beginning of the dissemination of the CEDAR API and RI, but also began conversations with interested stakeholders to identify potential collaborations and future next steps.
- **Integrated Stakeholder Feedback:** The project team obtained and integrated feedback from four target end-user types (clinicians, researchers, developers, and knowledge managers) to understand the perceived utility of CEDAR. The team used these findings to revise the problem statement and value proposition; this activity also highlighted AHRQ's potential to impact the healthcare community.
- **Expanded Search Feature:** The project team partnered with Computable Publishing, LLC, the author of the Fast Evidence Interoperability Resources (FEvIR) citation platform,[2] to test the RI in an applied setting. This collaboration expanded the quantity of AHRQ sources indexed by CEDAR. A new search capability permitted access to PubMed and Medline content, thereby enhancing the tool's scope and potential impact.

Update Environmental Scan and Gap Analysis

In the Option Period, the CEDAR project team updated the Environmental Scan initially developed during the Base Period.

Purpose, Objectives, and Methods

The purpose of the Option Period Environmental Scan and Gap Analysis was to provide updates to Base Period Environmental Scan findings, as well as to document any evolution or changes in the landscape of healthcare information retrieval. The objective of updating the scan was to inform other project activities, including refinements to outreach, pilot, API, and RI initiatives.

The project team conducted informal interviews with stakeholders and industry partners, performed a literature review, and examined industry leading practices relevant to CEDAR's operating requirements and potential use—including a review of successful practices for API adoption.

Findings and Observations

The project team identified key themes from the scan.

- Different users have distinct preferences for search methods and functionalities. For example, clinicians envisioned CEDAR as a product that can incorporate streamlined evidence into their workflows. By contrast, researchers prioritized extended search capabilities to refine search results.
- Access to trustworthy information resources remained a top priority in evidence retrieval, reflected in the observed confidence in AHRQ's reputation. Furthermore, this trusted information needs to be affordable so that it can be incorporated into current processes, especially by providers who may not be able to obtain subscriptions to commercial products.
- Two CEPI repositories relevant to CEDAR underwent significant upgrades during this time: the EHC added new artifacts, and SRDR+ released an updated version with improved access and features.

The project team's analysis identified opportunities for the future development of CEDAR or similar evidence-retrieval demonstrations.

- Develop specific use cases and target end users (i.e., users of CEPI resources) by defining user needs, expectations, and scenarios for use of the tool. Those results can then be leveraged to prioritize development activities.
- Improve the FAIR and CEDAR FAIR (C-FAIR) Tool[3] assessments by incorporating Transparency, Responsibility, User Focus, Sustainability, Technology (TRUST)[4] principles.
- Promote new tools by developing a strong brand associated with AHRQ—an organization the community views as a trustworthy source of information—that emphasizes unique features (i.e., that CEDAR is reliable, available at no cost through a trusted source, and worthwhile for users accessing AHRQ CEPI data).
- Create an "implementation kit" to simplify the process of integrating tools into other applications and to increase platform adoption.
- Develop alternate implementations of CEDAR on non-desktop platforms so that users can test and leverage CEDAR capabilities in diverse settings (e.g., smart phone apps or widgets).
- Identify new features that can be incorporated into new and existing tools, based on common issues found when indexing other repositories, that will enhance information retrieval (e.g., explore use of CEDAR RI or API in lieu of existing repository searches).

- Ensure that the tool incorporates the appropriate breadth and depth of information desired by target end-user groups (e.g., include information in CEDAR beyond selected CEPI resources and PCOR data).

Further Develop Reference Implementation

In the Base Period, the project team defined CEDAR as a concept, built the API using HL7 Fast Healthcare Interoperability Resources (FHIR),[5] and developed the first version of the RI.

In the Option Period, the project team enhanced the RI in response to ongoing feedback from the user community by:

- Refining the specificity of search results and numeric content rankings.
- Improving backend stability.
- Allowing for error alerts.

Appendix B provides screenshots of the CEDAR website, available as of August 2023, to provide a visual guide of the updated content. Appendix C details CEDAR development updates and their release schedule.

Purpose, Objectives, and Methods

The purpose of the Option Period was to further develop and enhance CEDAR based on feedback from the Base Period and the Option Period. The objective of these refinements was to encourage greater use of the tool and to promote access for target end users. The Option Period RI development employed an Agile process, which allows for ongoing regular feedback from the AHRQ team and responsiveness to the user community. The iterative approach is especially effective for testing and pilot programs; it facilitates the incorporation of feedback in rapid cycles across short 2-week sprint cycles.

Security Review

Throughout the course of the project, the project team examined and assessed its code for cybersecurity purposes, in keeping with Department of Health and Human Services (HHS) and AHRQ Federal security standards. Security reviews are essential to reduce the risk of cyberattacks and protect sensitive information.

Technical Documentation

Over the course of the entire 3-year period of performance, in conjunction with the development of the CEDAR RI, the project team maintained technical documentation, including webpages explaining how to get started with the API, an interactive API exploration tool, and an installation guide.

Findings and Observations

The project team identified several areas for RI improvement through direct feedback from target end users, as well as input from the other tasks during the Base Period and Option Period. Updates resulting from this information included the following.

- Importer Functionality
    - Detecting large numbers of changed artifacts and flagging imports with significant changes for verification by an administrator before making the changes visible via the CEDAR API.
    - Using the National Institutes of Health (NIH) bookshelf API instead of HyperText Markup Language (HTML) page scraping for NIH-hosted artifacts improved the robustness of the importer and incorporated additional metadata elements.
- Administration Application
    - Adding a page that summarizes and graphs search statistics over a selectable time period, allowing administrators to visualize CEDAR API usage.
    - Adding a feature to display the top 10 click-thru and top 10 returned artifacts for all repositories and each repository separately, to allow administrators to identify popular artifacts.
    - Adding a feature to allow an administrator to review flagged imports and to approve or reject the associated updates.
    - Adding lookup of internet protocol (IP) addresses to allow administrative users to better determine the source of CEDAR API search requests.
    - Adding support for viewing optional client identification numbers for CEDAR API searches when captured in the search log to allow voluntary tracking of CEDAR client usage.
    - Using Bidirectional Encoder Representations from Transformers[6] (BERT) language models to calculate artifact similarity and agglomerative hierarchical clustering to group similar artifacts to reveal common themes in artifact history and content across distinct AHRQ repositories. BERT language models allow text to be converted to vectors; the vectors for two pieces of text can then be compared for semantic similarity, assigning a similarity score. Each pair of artifact titles in CEDAR was compared and assigned a similarity score. Once title similarities were calculated, artifacts were grouped into clusters based on their title similarity scores across resources.
- API
    - Adding the ability to track returned artifacts for each search, to provide additional information and insights to CEDAR administrators.
    - Adding copyright information on source repository results, to allow client applications to display that information to users.
    - Adding organization description to API results, to allow client applications to display those descriptions to users.
    - Improving synonym support in search terms, to improve performance of searches.
    - Demonstrating User Interface (UI).
    - Fixing styling issues on mobile devices to allow the demonstration UI to function correctly on phones and tablets.

- Adding related searches with search results to provide additional search options to users.
- Flagging external links with an icon and a pop-up to warn users that the linked content is not managed by HHS when they are visiting externally managed sites.
- Adding validation for user-supplied dates to ensure searches behave as expected when date ranges are provided.
- Improving the user search experience (by displaying current search terms in the sidebar and in the search box) to enable users to track their search terms.

The project team made many additional smaller enhancements, fixed bugs, and tweaked code for additional efficiency and improvements to all CEDAR components. Ongoing development ideas are tracked in a log as part of the Agile development process that AHRQ will retain for use by future development teams as needed. See Appendix C for additional details on enhancements made to the RI throughout the Option Period.

Conduct Pilot

The Option Period Pilot included conducting focus groups, developing and implementing a bidirectional data exchange capability, and creating an application with a third-party partner.

Purpose, Objectives, and Methods

The Base Period Pilot was designed to demonstrate the RI in an applied setting with target end users. The project team received feedback that, although target end users need more information than the contents of a CEDAR search, users still preferred the limited quantity and highly trustworthy quality of results returned by CEDAR. To address these seemingly conflicting demands, the project team established a second pilot effort in the Option Period that could explore the underlying interests driving the Base Period findings.

The following were objectives of the Option Period Pilot:

- Further engage target end users to determine the feasibility of integrating CEDAR into an external application.
- Discover opportunities to improve CEDAR.
- Gain data-driven insights on the future capacities and usefulness of the tool.

The project team undertook activities to identify and recruit target end users to participate in focus groups, revise CEDAR use cases, partner with a third-party organization to expand CEDAR capabilities and develop a CEDAR application, apply the C-FAIR tool scoring (a tool developed during the Base Period, for more information see the CEDAR Base Period Final Report and Appendix D) to the CEDAR API and the third-party application alike, and evaluate the acceptance and maturity of CEDAR.

Focus Groups

The project team evaluated the efforts from the Base Period to identify key target end-users to incorporate in the Option Period Pilot. These included clinicians, researchers, developers, and knowledge managers. To further understand these user categories, the team organized several focus groups to obtain qualitative input from these target end-users on the CEDAR RI. The data collected from the focus group sessions were analyzed using Machine Learning (ML) and Natural Language Processing (NLP) techniques performed using Python.

Use Case Revisions

Upon finishing the focus group discussions, the project team revised the list of CEDAR use cases based on input from focus group discussions, outreach activities, and input and feedback from subject matter experts (SMEs).

CEDAR RI Usage Statistics

The project team released a publicly available CEDAR demonstration user interface during the Option Period. The team actively promoted knowledge of this tool in three ways: 1) including a link to the publicly available AHRQ CEDAR RI in the email originally circulated to 120 individuals when recruiting focus group participants, 2) encouraging focus group participants to use the RI in their work even after the discussion and share it with possibly interested colleagues, and 3) promoting the RI at the Clinical Decision Support Innovation Collaborative (CDSiC) 2023 Annual Meeting. In addition, the CEDAR RI is available and easily discoverable on the AHRQ website.

Throughout the piloting period, the project team monitored the frequency of use of the CEDAR RI.

Third-Party Partnership with Computable Publishing, LLC (FEvIR)

AHRQ and MITRE partnered with Computable Publishing, LLC, to expand the capabilities of the CEDAR API. Computable Publishing, LLC, is responsible for the creation and maintenance of the FEvIR Platform, an open-source platform that indexes information from Medline, ClinicalTrials.gov, and other resources using HL7 FHIR citation standards. The goal of this partnership was to create a FEvIR/CEDAR application that demonstrates a bidirectional data-exchange capability, in which each platform pulls data from the other's and expands the number of CEDAR and FEvIR search results alike.

C-FAIR Tool Scoring

The team evaluated the CEDAR API and the FEvIR/CEDAR application using the C-FAIR Tool, an evaluation metric tool created by MITRE to objectively measure the FAIRness of the CEPI resources during the Option Period.

The project team first evaluated the CEDAR API using the C-FAIR Tool after making enhancements and updates during the Option Period to assess any impact to FAIRness. After creating the FEvIR/CEDAR application, the team also assessed the new tool's adherence to FAIR principles using the C-FAIR Tool.

Acceptance and Maturity

The project team evaluated API and RI adoption and maturity[7] using the Technology Acceptance Model (TAM)[8] and the Capability Maturity Model Integration (CMMI).[9] These evaluations helped create recommendations to advance the adoption and use of CEDAR's technology.

Findings and Observations

Of all the pilot activities, partnering with Computable Publishing, LLC's FEvIR platform resulted in the greatest impact to CEDAR. The collaboration demonstrated bidirectional data exchange between CEDAR and FEvIR: FEvIR imported CEPI artifacts from CEDAR using CEDAR's FHIR API and CEDAR imported Medline artifacts from FEvIR using FEvIR's public API. This collaboration resulted in improvements to both the CEDAR data model and to the FEvIR API. The functionality allowing CEDAR artifacts to be imported into FEvIR is available on the FEvIR website as the "FEvIR: CEDAR Search Importer."

The pilot also identified 14 major themes that relate to target end users' needs, along with minor themes from conversations that may further define the problem statement CEDAR is intended to address. This information helped to not only characterize the utility of CEDAR, but also describe how it could be used in potential use cases.

Although the C-FAIR tool's scoring of CEDAR had not changed since the Base Period (39 of 42 total possible points), the CEDAR API improved in a C-FAIR scoring criterion by exposing copyright data (this criterion already met full marks in other related aspects). FEvIR: CEDAR Search Importer also scored 39 out of 42 total possible points. FEvIR's application failed to achieve full scores for one aspect of a single criterion: The FEvIR: CEDAR Search Importer does not make metadata available to the public in real time when a resource or artifact is archived or unavailable. Otherwise, FEvIR: CEDAR Search Importer scored full points for all other criteria in the C-FAIR Tool. Its direct and intentional alignment to the FHIR Citation Resource (and the FHIR standard in general) supports its high overall score.

The pilot determined that CEDAR was not broadly or widely adopted during the Option Period. Nevertheless, the focus-group discussions and technical pilot with FEvIR demonstrated the CDS community's positive acceptance and reception of CEDAR overall, as well as positive attitudes about using AHRQ-developed tools. Despite this, many users and participants expressed skepticism about CEDAR's perceived usefulness relative to alternative search tools.

Finally, the team conducted an informal review of the development of CEDAR's software that revealed a mix of processes following CMMI Maturity Levels 2 ("Managed") and 3 ("Defined")[10] as part of a standards-based approach to indexing and presenting Evidence-Based Medicine (EBM) and PCOR resources within CEPI. The project team applied a consistent Agile development approach in 2-week sprint cycles across the project, as well as in weekly development and integration sessions with the FEvIR team. CEDAR's longevity has been reliant on a small team, rather than a consistent organizational-wide approach; therefore, the structure lacks a longer-term strategy for maintaining and updating the API, RI, and supporting repositories.

Key insights from the pilot activities included the following.

- The healthcare community noted that trust is an important feature; the community has repeatedly stated that AHRQ's primary value is the trustworthiness of its resources. This characteristic is a major driver for why end users seek information provided by AHRQ.
- Trust in AHRQ's evidence needs to be balanced with the ability to keep information current, especially in terms of balancing conflicting evidence.
- CEDAR is a niche tool, requiring specialized attention and a careful approach to product management. Users have alternate options to access similar data or ways to obtain the AHRQ data (e.g., Google, Chat Generative Pre-Trained Transformer [Chat-GPT], PubMed, and UpToDate). It is possible that a small group of early adopters could help increase the tool's visibility.
- The partnership between CEDAR and FEvIR demonstrates the capability of expanding CEDAR to encompass other standards-based PCOR information.
- CEDAR applications (including the developed CEDAR demonstration UI) represent an opportunity to provide information directly to doctors or to patients, as well as to mediate information through the patient/doctor relationship.
- Users expressed a preference for tools that are integrated into standardized, existing workflows that do not create additional burden to use (e.g., as a link in the electronic health record [EHR] that clinicians can access directly from the patient's record). To be successfully adopted by potential users (and to lower any barriers to such adoption), the content, capabilities, and UI of CEDAR need to be intuitive to navigate.

- Participants in the pilots expressed a willingness to use an improved version of CEDAR. They also emphasized the need for its enhancement to either supplement or expand its content in order for it to evolve into a more beneficial solution.

Conduct Outreach

In the Option Period, outreach efforts evolved with the maturity of CEDAR to yield deeper insights to guide the tool's potential future trajectory.

Purpose, Objectives, and Methods

The purpose of the Option Period Outreach was to shift focus from creating a framework for *how* users could adopt CEDAR to obtaining pointed insights into *whether* CEDAR would be adopted. The objective of stakeholder engagement focused on capturing feedback on CEDAR's relevance to users' current and desired practices, as well as soliciting suggestions on how CEDAR could be modified to make it more suitable for its intended audience. This objective was achieved by collecting input from thought leaders and SMEs, and by promoting CEDAR as a resource available to multiple user groups. The team followed this method in multiple forums: presentations and exhibits at conferences, group conversations (e.g., focus groups), and individual calls with specific SMEs.

Findings and Observations

The project team identified several findings that can inform future demonstration projects or the evolution of CEDAR, including the following key updates.

- Target end users and community members consulted in the Option Period confirmed ongoing trust in AHRQ and its evidence; from the outset, these individuals had a high regard for the impartiality in how AHRQ generates evidence, which is consistent with sentiments from the Base Period.
- Meeting with smaller audiences provided more insights for future improvements than were obtained through large-scale conferences, which may be most beneficial for demonstration efforts that are newer and less entrenched.
- Stakeholders appreciated AHRQ's commitment to make CEPI information more accessible, and they were open to considering the value of CEDAR to their professional work; however, most did not express a specific use for an API indexing CEPI repositories. Future demonstration efforts may benefit from initial market analysis to identify the problem before pursuing a specific development solution.

- Most target end users did not have a third-party user interface in place to perform CEDAR API integration and to incorporate the tool into their existing workflow, which resulted in an unintended deterrent to adoption. This issue might be remedied in the future by further defining use cases (i.e., how users interact with the tool) and target user groups (i.e., the individuals intended to act out the use cases) at the outset of development.
- Target end users and commercial vendors wanted more high-quality content from AHRQ included in CEDAR, and to have it presented in more intuitive ways (e.g., tagging artifacts to link related content), including possibly incorporating it in commercial health products. AHRQ might consider partnering with groups such as the National Science Foundation or HHS agencies for future demonstration efforts.
- Stakeholders suggested including patients as target end users through the creation of a patient-centered application or interface that enables the tailoring of content. Increased patient access to information may promote the improvement of an individual's own care (whether self-directed or in collaboration with their clinical teams) by leveraging PCOR evidence to understand their diagnoses or gain insight into treatment options. This direction may be an option for future demonstration efforts.

Project Management

The project structure and processes established in the Base Period proved to be an effective means to manage the tasking; they were maintained throughout the Option Period. The CEDAR project tasking was managed by a leadership team comprised of a Project Leader, a Deputy Project Leader, a Technical Lead, and Task Leads for the pilot and outreach tasks. The leadership team met every other week throughout the Option Period to ensure collaboration and communication between MITRE and AHRQ, and to collaboratively address any issues or questions that arose. Additional MITRE leadership (including a Program Manager, Chief Engineer, and Department Manager) provided oversight and management through a regular touch-base, as well as more extensive program check-ins and ad hoc outreach.

Throughout the project, the project team used an Agile management approach to manage tasking and to track project progress. The team planned each sprint at the beginning of the 2-week period, then held review meetings with AHRQ leadership from key CEPI programs at the end of each period.

Lessons Learned and Recommendations

The Option Period produced key lessons learned and recommendations to evolve CEDAR's value proposition and/or enhance future demonstration projects.

The following four concepts are broadly relevant to the future of CEDAR and information retrieval.

Trust

Trust remains a key factor in healthcare and health research,[11] as indicated through the proposed integration of TRUST principles into the FAIR principles. AHRQ's reputation for issuing valuable, high-quality, and trustworthy content can be leveraged in evolving CEDAR and future demonstration efforts. Trust is a differentiator for AHRQ tools; it can be an asset when defining the brand and conducting outreach.

Recommendation

- Add TRUST principles to the C-FAIR Tool evaluation criteria to augment the original FAIR principles and add further value to C-FAIR assessments.

Emerging Large-Language Models

During the Option Period, large-language models (e.g., ChatGPT) began to influence online search practices, with the expectation that they will revolutionize the paradigm of data accessibility and establish new standards for user experiences in data retrieval. Their functionality augments the traditional web search capabilities by providing an intuitive conversational approach that fosters engagement and ideation. Integrating this functionality with CEDAR may address many of the comments from end users, such as a desire for aggregated information or information that is easily distilled.

Recommendation:

- Consider selective use of new technology, technology enablers, and capabilities in AHRQ's overall strategy for disseminating evidence-based medicine and PCOR resources. Specifically, where these solutions can enable more accurate processing of natural language, they may provide more context-aware responses that improves the relevance and richness of search results.

Best Practices for Product Development

Demonstration projects need a value proposition, including a problem statement, to guide iterative development and ongoing engagement with target end users. This information can inform proactive planning across the required developmental phases (e.g., prototype, test, mature). Concurrently with each active development phase, project leadership should consider the phase that will follow, assessing what is feasible and viable. For example, does the project warrant continued testing, is it ready to be matured, or is it ready to be ended to allow the application of lessons learned to other efforts?

Recommendations:

- Develop the tool with user-centric design methods that provide information that is tailored to the target end user (e.g., providing aggregate information to patients).
- Establish messaging when searches yield low volume or no results to assure users of the result.
- Continue to innovate on delivering FAIR-based access to CEPI and other AHRQ PCOR evidence.
- Add TRUST principles to the C-FAIR Tool evaluation criteria. This augments the original FAIR principles and adds further value to C-FAIR assessments.
- Balance "quality curation" and "speed-to-publication" using Artificial Intelligence (AI) or Robotic Process Automation (RPA) tools. For example, checking for missing AHRQ funding references can be accomplished using pattern matching; the document then can be returned via email or workflow to the author for resolution before curation commences. Significant savings of time and costs could be realized by having automatic treatment of "simple" issues, allowing the curator to focus on tasks that require more specialized skills.
- Consider partnerships with other aligned agencies and organizations, including the Office of the Assistant Secretary for Planning and Evaluation (ASPE) and the Patient-Centered Outcomes Research Institute (PCORI), to facilitate dissemination of PCOR material to relevant end users.

CEDAR's Value Proposition

CEDAR provides a trustworthy option for users to retrieve information efficiently and effectively from five AHRQ CEPI repositories. CEDAR provides a clear utility by indexing trusted data in a publicly accessible, user-friendly API and demonstration interface. Evolving CEDAR into a mature tool will require ongoing maintenance and management to ensure smooth operation of the existing functionality. Fundamentally, CEDAR needs to maintain the high quality and trust of the information it indexes; it must also grow its content to add breadth to the available evidence and meet users within their existing workflows.

Recommendations:

- Solicit and verify the CEDAR's value proposition from target end users sufficiently early in its development to incorporate that feedback during phases of planning, requirements, design, development, and release.
- Describe this value proposition and its differentiators to maximize development of a brand that increases visibility and promotes its impactful use.
- Establish Communities of Interest (COI) to serve as project advisors when determining future direction and as champions of the product. For example, COIs can provide valuable feedback (positive and negative) that are not subject to restrictions imposed by the Paperwork Reduction Act (PRA); further, they can help communicate the value of new AHRQ resources to wider audiences.
- Establish an ongoing development cycle roadmap that functions alongside the project work plan to proactively guide decision making concurrently with preparing for the next development phase.
- Complement the product's Agile development approach by prioritizing long-range development options, a strategic approach that accounts for the tool's problem statement, value proposition, and target users.

Appendix A: Abbreviations and Acronyms

Table 1: Acronyms, Abbreviations, and Definitions

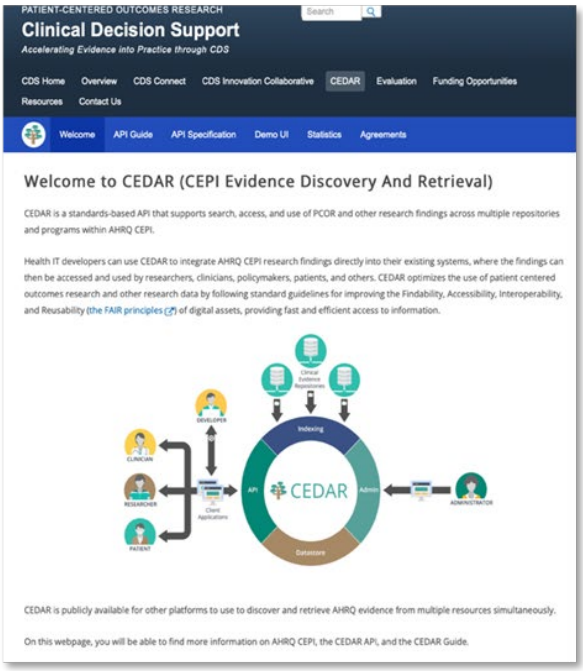
Acronym or Abbreviation	Definition
AAFP	American Academy of Family Physicians
AHRQ	Agency for Healthcare Research and Quality
AI	Artificial Intelligence
AMIA	American Medical Informatics Association
API	Application Programming Interface
ASPE	Office of the Assistant Secretary for Planning and Evaluation
BERT	Bidirectional Encoder Representations from Transformers
C-FAIR	CEDAR FAIR Tool
CDS	Clinical Decision Support
CDSiC	CDS Innovation Collaborative
CEDAR	CEPI Evidence Discovery And Retrieval
CEPI	Center for Evidence and Practice Improvement
Chat-GPT	Chat Generative Pre-Trained Transformer
CMMI	Capability Maturity Model Integration
COI	Community of Interest
DOI	Digital Object Identifier
EBM	Evidence-Based Medicine (on FHIR)
EHC	Effective Health Care Program
EHR	Electronic Health Record
EPC	Evidence-based Practice Centers
FAIR	Findable, Accessible, Interoperable, and Reusable
FEvIR	Fast Evidence Interoperability Resources
FFRDC	Federally Funded Research and Development Center
FHIR	Fast Healthcare Interoperability Resources
FORCE11	The Future of Research Communications and e-Scholarship
GUID	Globally Unique Identifier
HHS	Department of Health and Human Services

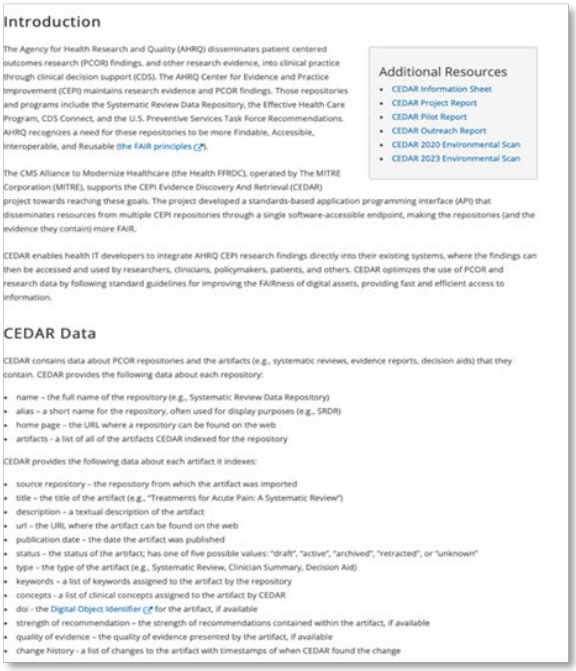
Acronym or Abbreviation	Definition
HL7	Health Level Seven
HTML	HyperText Markup Language
IHE	Integrating the Healthcare Enterprise
IP	Internet Protocol
IT	Information Technology
JAMA	Journal of the American Medical Association
MeSH	Medical Subject Headings
MITRE	The MITRE Corporation
ML	Machine Learning
NIH	National Institutes of Health
NLP	Natural Language Processing
OP	Option Period
PCOR	Patient-Centered Outcomes Research
PCORI	Patient Centered Outcomes Research Institute
PRA	Paperwork Reduction Act
RI	Reference Implementation
RPA	Robotic Process Automation
SEI	Software Engineering Institute
SME	Subject Matter Expert
SQL	Structured Query Language
SRDR	Systematic Review Data Repository
TAM	Technology Acceptance Model
TRUST	Transparency, Responsibility, User Focus, Sustainability, Technology
UI	User Interface
URL	Uniform Resource Locator
USPSTF	U.S. Preventive Services Task Force
XML	Extensible Markup Language

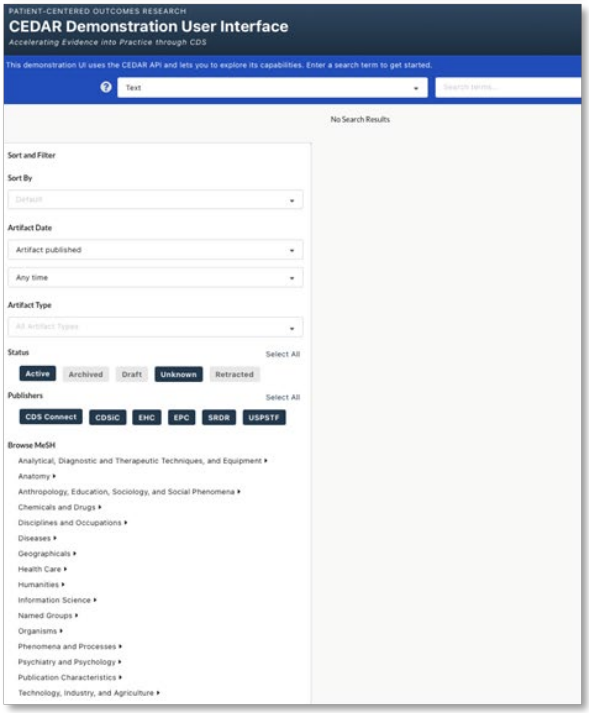
Appendix B. The CEDAR Experience

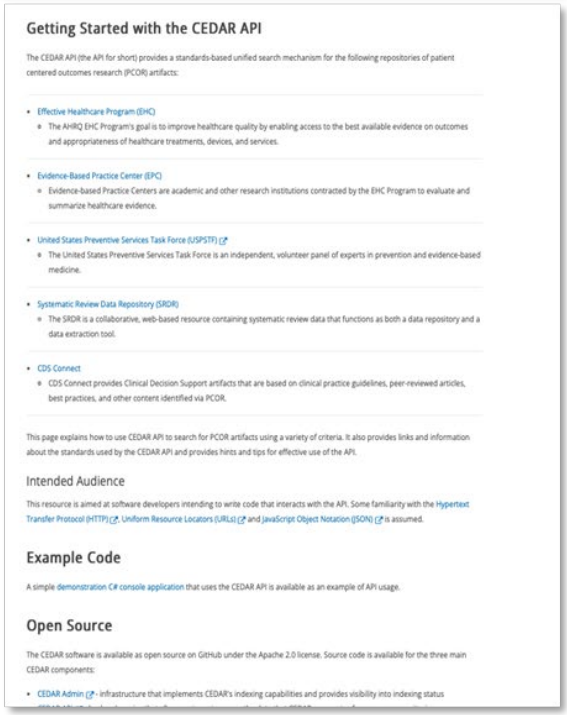
The following screen shots provide a visual display of the CEDAR experience for reference. More can be seen using the internet archive (i.e., the Wayback Machine) or on digital.ahrq.gov.

Figure 1: Selected CEDAR Screenshots









Appendix C. CEDAR Development Release Notes

CEDAR Admin

v0.8.0 - 2023-03-14

- Adds a page that summarizes search statistics
  - Summarizes searches that have taken place over a selectable time period
  - Graphs the number of searches by day over that time period
  - Lists the top IP addresses performing searches over that time period
  - Allows exclusion of one or more IP addresses from the total tally of results
  - Allows clicking on an IP address to see the search logs for that Internet Protocol (IP) address
- Adds feature to display top 10 click thru and top 10 returned artifacts
- Improves visibility of flagged imports and adds page showing complete import history
- Improves parsing of dates when indexing EPC artifacts
- Improves handling of Evidence-based Practice Center (EPC) technology assessment artifacts
- Fixes a Uniform Resource Locator (URL) issue with email notifications
- Fixes an efficiency issue with displaying search logs
- Fixes an Extensible Markup Language (XML) namespace issue when handling National Institutes of Health (NIH) bookshelf imports
- Updates U.S. Preventive Services Task Force (USPSTF) indexer to stop indexing blocked Journal of the American Medical Association (JAMA) Network pages
- Refactors complex Structured Query Language (SQL) queries into database views

v0.7.2 – 2023-02-22

- Uses NIH bookshelf metadata service for NIH imports instead of scraping HTML
- Fixes an exception handling issue with importer
- Updates to Ruby 3.0.3 and Rails 6.1.7

v0.7.1 – 2023-01-18

- Temporarily works around importer issue of descriptions being removed from source repository
- Adds more configurability to email settings

v0.7.0 – 2023-01-05

- Adds feature to detect a large number of changes on import and require approval
- Adds repository descriptions
- Flags external links with icon and adds a popup
- Fixes issue with importing concepts that are empty strings
- Fixes issue where description fields were not always correctly synched
- Adds appropriate user agent on importer requests
- Adds introductory text to site and updates links
- Updates dependencies

v0.6.0 – 2022-10-03

- Addresses error when displaying logs for searches that do not complete
- Removes redundant synonym expansions

- Addresses styling issue with menu display
- Updates the EPC importer
    - Identifies archived artifacts using a HyperText Markup Language (HTML) meta tag
    - Improves support for EPC artifact dates
    - Removes duplicate warnings for missing artifact dates
    - Re-orders import log view so latest import is at the top of the page
- USPSTF importer updates
    - Uses general recommendation pubDate instead of topicYear
    - Adds general recommendation topicType to CEDAR keywords
    - Uses tool keywords to supplement those on the associated general recommendation
- Marks artifacts as retracted after two weeks of failed import attempts
- Supports pruning of older database backups
- Fixes an issue where concepts may only have a Spanish Medical Subject Headings (MeSH) code
- Updates README with additional information

CEDAR API

v0.8.0 – 2023-03-14
- Tracks returned artifacts for each search

v0.7.2 – 2023-02-22
- Adds support for returning numerical content rankings with search results
- Updates to Ruby 3.0.3, Sinatra 3.0.5, and Rails 6.1.7

v0.7.1 – 2023-01-18
- Removes flagged and suppressed artifact versions from search results
- Fixes the base URL used in structure definitions and search results
- Adds copyright information on source repositories to results

v0.7.0 – 2023-01-05
- Adds related search links to search results
- Fixes issues with handling punctuation and hyphens in free text searches
- Adds organization description to API results
- Updates C# example code
- Updates swagger documentation
- Completes move of documentation from API repository to static content repository
- Updates dependencies

v0.6.0 – 2022-10-03
- Supports hyphens in title search
- Extends the CEDAR logger to support debug logging level
- Fixes issue where text search for exact title did not return matching item
- Updates synonym expansion to use database stemming
- Updates synonym handling to include phrase synonyms for simple queries
- Updates Swagger documentation

CEDAR UI

v0.7.1 – 2023-02-22

- Fixes issue with related search behavior where search concepts were not correctly cleared
- Fixes styling of CDS footer on mobile browsers

v0.7.0 – 2023-01-05

- Adds support for displaying related searches that a user can click on
- Styles interface to display nicely on mobile devices
- Displays descriptions of source repositories
- Flags external links with icon and adds a popup
- Updates dependencies

v0.6.0 – 2022-10-03

- Adds validation for user-supplied dates
- Updates URLs to use relative references

Appendix D: C-FAIR Tool

This appendix captures the content provided in all tabs of the C-FAIR Tool. The first tab of the C-FAIR Tool, titled "Instructions," is included here.

Purpose: The C-FAIR Tool provides a tangible and quantifiable method to manually assess a health information resource's adherence to FAIR (Findable, Accessible, Interoperable, Reusable) principles[12] A "resource," as defined by the Agency for Healthcare Research and Quality, is a "practical, research-based tool ... to help a variety of healthcare organizations, providers and others make care safer in all healthcare settings." [13]

Description: The C-FAIR Tool is a multi-tab Excel workbook. The "Criteria" tab of the workbook contains four tables with four columns. Column A (the first, left-most column) is titled "Original Principle" and contains the original FAIR Data Principles. The second column (B), titled "#," includes unique IDs for each criterion of the C-FAIR Tool. The third column (C), titled "C-FAIR Tool Criteria," lists each criterion to factor into the evaluation of the C-FAIR Tool. Column D (the fourth, right-most column) is titled "Description" and describes each of the C-FAIR Tool criteria.

The "Scoring" tab of the C-FAIR Tool contains four tables with six columns. The first, second, and third columns (A – C) are identical to the second, third, and fourth columns in the "Criteria" tab (B – D, respectively). The fourth column of the "Scoring" tab, also titled "Scoring," is used to describe each factor in the C-FAIR Tool's scoring system. There are multiple criteria for each component of FAIR, and each criterion can have a maximum of three points, representing full adherence; a minimum of zero points, representing no adherence; or one or two points, representing partial adherence to a FAIR criterion. The C-FAIR Tool includes scoring definitions for each criterion. The fifth column, titled "Score," is the area where a user should provide each numeric value. Once each criterion is assigned a score, the total score is calculated for each component and, at the bottom of the column, the component scores are totaled to provide an overall resource score. The sixth, right-most column, titled "Notes," is the area where a user can include any notes, comments, or justification as to why a certain score was given for a specific criterion.

Instructions: To manually assess a health information resource's adherence to FAIR, enter a numeric value in the "Score" column on the "Scoring" tab for each criterion. As able, include all notes, comments, or justification as to why a certain score was given for each criterion in the "Notes" column.

The following table captures content from the C-FAIR Tool, titled "C-FAIR Tool Criteria" and "C-FAIR Tool Scoring."

Table 2: C-FAIR Tool Principles, Criteria, and Scoring

Original FAIR Principle	#	C-FAIR Tool Criteria	Description	C-FAIR Tool Scoring
F1.1 (meta)data are assigned a globally unique and externally persistent identifier	**F1.1.1**	**Digital objects have a persistent identifier as a metadata element**	Ensures that data and metadata are findable over time. Persistence is a guarantee that an identifier will remain; the responsibility of the resource owner, this criterion measures the presence of the identifier and/or the intent to maintain persistence.	0 = no 2 = unclear 3 = yes
F1.1 (meta)data are assigned a globally unique and externally persistent identifier	**F1.1.2**	**Digital objects have a globally unique identifier (GUID) as a metadata element**	A single identifier for the same set of metadata. A GUID is an identifier formatted according to special conventions to support uniqueness within an organization and across all organizations. Examples of GUIDs are digital object identifiers (DOI) #s for digital resources or Open Researcher and Contributor Identifiers (OrcID) #s.	0 = no 3 = yes
F2: Data are described with rich metadata	**F2.1**	**Data are described with rich, defined metadata**	Ensures data discovery through the representation of data in a formal, structured way. Rich metadata include URL, identifier, version, name, title, status, experimental, date, publisher, contact, description, useContext, jurisdiction, purpose, copyright, approvalDate, lastReviewDate, and effectivePeriod.	0 = no metadata 1 = 1 – 6 of defined metadata elements 2 = 7 – 12 of defined metadata elements 3 = 13 – 18 of defined metadata elements

Original FAIR Principle	#	C-FAIR Tool Criteria	Description	C-FAIR Tool Scoring
F2: Data are described with rich metadata	**F2.2**	**Documentation, such as descriptions and definitions of the data, metadata, and resource, is available**	Documentation increases the understanding of the data, metadata, and entire resource. Score is determined by the presence of these materials.	0 = none 1 = documentation describes one of the three elements (data, metadata, or resource) 2 = documentation describes two of the three 3 = documentation describes all three
F3: (Meta)data are registered or indexed in a searchable resource	**F3.1**	**The resource makes metadata available to be searchable by another system**	Examines the method of data extraction, which impacts the ease in which metadata are pulled correctly into CEDAR.	0 = metadata are not available 1 = metadata are available through web searches 2 = metadata are available through static database extract 3 = metadata are available programmatically via API
A1: (Meta)data are retrievable by their identifier using a standard communication protocol	**A1.1**	**The process supporting (meta)data retrieval should be open and not proprietary**	Users should have open access to the metadata; an API is the optimal mechanism.	0 = retrieval process is proprietary; metadata unavailable to any user 1 = retrieval process is proprietary; metadata available to some users 2 = retrieval process is not proprietary; access to full metadata is available for users 3 = retrieval process is available using an API; access to full metadata is available to all users

Original FAIR Principle	#	C-FAIR Tool Criteria	Description	C-FAIR Tool Scoring
A1: (Meta)data are retrievable by their identifier using a standard communication protocol	**A1.2**	**Resources have instructions for how to access the metadata**	Documentation that describes how the data can be accessed is available. To be considered accessible, (meta)data should have instructions (e.g., user license, copyright, documentation of restrictions) that support accessing the data.	0 = no documentation/ instructions 3 = documentation/ instructions
A2: Metadata are accessible, even when the data are no longer available	**A2.1**	**Metadata remain available if the data become unavailable**	Metadata should persist even after the underlying data can no longer be sustained. This will improve accessibility; the metadata can still provide users with clarifying information about the current state of the data.	0 = metadata become unavailable when the data become unavailable 3 = metadata remain available when the data become unavailable
I1: (Meta)data use formal, accessible, shared, and broadly applicable language for knowledge representation	**I1.1**	**Metadata include community accepted keywords and a defined taxonomy to describe the underlying data**	Metadata should be represented and described using keywords, phrases, and/or standards that are prevalent and accepted in the health information technology (IT) community. This will allow for the exchange of information once systems and users can interact using a universally standard knowledge representation.	0 = no defined taxonomy/ community accepted keywords 3 = defined taxonomy/community accepted keywords are included
I2: (Meta)data use vocabularies that follow FAIR Principles	**I2.1**	**Metadata use a comprehensive, controlled vocabulary that is referenced and linked**	Metadata will need to be referenceable (having its own identifier) as well as linked (there should be a way to navigate to a standard's appropriate digital resource location).	0 = no common vocabulary is included in the metadata 3 = a common vocabulary is included in the metadata

Original FAIR Principle	#	C-FAIR Tool Criteria	Description	C-FAIR Tool Scoring
I3: (Meta)data include qualified references to other metadata	**I3.1**	**Metadata include qualified references to associated metadata**	Identifies a link between artifacts digital objects across resources; for a digital object in a given resource, the metadata should include an identifier of the external/linked data.	0 = no metadata of any associated external resource included 3 = metadata of any associated external resource included
R1.1: (Meta)data are released with clear and accessible data usage license	**R1.1.1**	**Resources should provide documented data usage license**	A data usage agreement/license is needed to provide guidance on the reuse and sharing of the data in a resource. This is measured by the presence of a data usage agreement that can be understood by a user.	0 = no data usage license provided 3 = full data usage license documentation
R1.2: (Meta)data are associated with their provenance	**R1.2.1**	**Metadata are associated with detailed provenance**	Ensure that metadata are current and relevant. Provenance ensures that the data are relevant and can be accessed. Six key elements: origin, creator, version, timestamp, Persistent ID, latest version.	0 = no provenance elements provided 1 = 1 – 2 of defined provenance elements are provided 2 = 3 – 4 of defined provenance elements are provided 3 = 5 – 6 of defined provenance elements are provided
R1.3: (Meta)data meet domain-relevant community standards	**R1.3.1**	**Metadata are expressed in a standardized data format**	Standardized data formats (e.g., "fully structured data formats") that are most prevalent in this domain.	0 = no standardized, domain-supported data format 3 = standardized, domain-supported data format is used

Appendix E. References / Citations

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. "The FAIR Guiding Principles for scientific data management and stewardship." Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18.

[2] Fast Evidence Interoperability Resources (FEvIR®) Platform [Project]. Contributors: Brian S. Alper [Authors/Creators]. In: Fast Evidence Interoperability Resources (FEvIR) Platform, FOI 29394. Revised 2023-06-23. Available at: https://fevir.net/resources/Project/29394.

[3] MITRE created the C-FAIR Tool specifically to quantify FAIR access to CEPI information. This tool was adapted from the FAIR Guiding Principles to assess the total information within the CEPI repositories instead of the individual evidence artifacts within the repositories. See the CEDAR Base Period Final Report for more information.

[4] TRUST is a set of guiding principles (Transparency, Responsibility, User focus, Sustainability, and Technology) established through community consensus that demonstrate a digital repository's trustworthiness. They provide a common framework to facilitate discussion and implementation of best practice in digital preservation by all stakeholders.

[5] FHIR® is a registered trademark of Health Level Seven (HL7).

[6] Devlin, Jacob, Chang, Ming-Wei. "Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing" (Google Research Blog). Available at: https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html.

[7] The "maturity" of the API and RI represent the use of industry best practices to assure quality. More "mature" processes used to develop software focus on continuous process improvement, including (but not limited to) structured, documented, repeatable processes; quality assurance reviews to identify and address bugs before deployment; and other governance and independent verification and validation processes for objectivity in reviews. In this context, "maturity" does not, for instance, refer to how close the software is to implementing a defined development roadmap or completely implementing all items in a feature backlog.

[8] Davis F D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manage Inf Syst Q. 1989; 13:319–340.

[9] Software Engineering Institute (SEI) at Carnegie Mellon University. Pittsburgh, PA. Available at: https://www.sei.cmu.edu/about/index.cfm.

[10] Process Areas may be assessed together in groups of processes divided into five maturity levels. A "Staged Model" is a formal assessment that follows this approach, addressing a specific goal (e.g., improvement in control and planning of tasks associated with the process). An alternate practice followed for many years is the "Continuous Model," where the user can select the order of improvement that best fits the organization's business needs; thus focusing attention on certain Process Areas, even if the Staged Model would group them into different maturity levels. The project team conducted an informal assessment of maturity.

[11] "Trust In Health Care: Insights From Ongoing Research," Health Affairs Forefront, January 11, 2022.DOI: 10.1377/forefront.20220110.928032

[12] FORCE11: The Future of Research Communications and e-Scholarship: Past, Present, and Inviting Input for the Future. Presentation, the National Center for Biomedical Ontology. October 10, 2012.

[13] AHRQ Data Resources. https://www.ahrq.gov/data/resources/index.html/.