The Common Fund Data Ecosystem

A report on the state of the ecosystem

Common Fund Data Ecosystem Evaluation Core Team

December 11, 2024

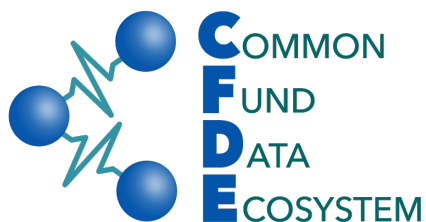A report presented at the annual\meeting on 10 August 2025\

Table of contents

Preface

The National Institutes of Health (NIH) Common Fund Data Ecosystem (CFDE) represents a transformative effort to enhance data sharing, integration, and usability across diverse biomedical research domains. As the volume and complexity of biomedical data continue to grow, the CFDE aims to break down silos, enabling researchers to efficiently find, access, and analyze data critical to advancing science and medicine.

This report focuses on metrics and data as foundational elements of evaluating and guiding the CFDE's impact. Metrics provide a quantitative lens to assess progress, identify challenges, and inform strategic decision-making, while robust data support comprehensive analyses and foster transparency. By examining the ecosystem's usage patterns, user engagement, and research outputs, we aim to provide actionable insights that support both the ongoing development of the CFDE and its alignment with broader NIH goals.

Our approach integrates quantitative metrics—such as data access, downloads, and tool utilization—with qualitative observations that capture user experience and engagement. Together, these elements form a cohesive picture of how the CFDE supports researchers and fulfills its mission of maximizing the impact of NIH-funded data.

This report is intended for stakeholders across the biomedical and data science communities, including NIH program officers, researchers, and policymakers. We hope it serves as a resource to catalyze discussion, innovation, and collaboration as we collectively strive to harness the full potential of biomedical data.

1. Introduction

The NIH Common Fund Data Ecosystem (CFDE) serves as a vital infrastructure supporting the sharing, integration, and usability of biomedical data across a diverse range of research areas. By fostering interoperability and collaboration, the CFDE addresses the growing need for efficient data access and reuse, enabling scientists to tackle pressing questions in biomedicine.

The CFDE comprises five core centers, each focused on a specific aspect of the ecosystem's development and operation, and a network of data coordinating centers (DCCs) representing specific research domains. These DCCs bring together rich datasets, tools, and expertise, creating a foundation for interdisciplinary collaboration. To expand its reach and capabilities, the CFDE has also funded a series of strategic partnerships with external organizations and projects, enhancing the ecosystem's capacity to meet evolving research needs.

In addition to these structural components, CFDE supports a variety of cross-cutting activities. These include meetings, workshops, and webinars designed to facilitate knowledge exchange, foster community engagement, and drive the ecosystem's growth. Collectively, these elements form a dynamic and adaptable framework for advancing biomedical research.

This report provides an analysis of the CFDE through the lens of metrics and data, structured into four sections:

1. **Centers**: This section examines the contributions of the five CFDE centers, focusing on their individual roles and collective impact on the ecosystem's goals.

2. **Data Coordinating Centers (DCCs)**: Here, we analyze the role of DCCs in curating and providing access to high-value datasets and their interactions with the broader CFDE framework.

3. **Partnerships**: This section highlights the impact of CFDE-funded partnerships, emphasizing their role in expanding the ecosystem's scope and capacity.

4. **Cross-Cutting Activities**: Finally, we evaluate the effectiveness of meetings, workshops, and webinars in promoting community engagement and supporting ecosystem objectives.

This structure ensures a comprehensive evaluation of the CFDE's multifaceted components, capturing their distinct contributions while addressing their interdependencies. By organizing the report in this way, we aim to provide stakeholders with a clear understanding of the ecosystem's achievements and opportunities for future growth.

Part I.

Centers

2. Integration and Coordination Center

3. Knowledge Center

4. Data Resource Center

5. Cloud Workspace Implementation Center

6. Training Center

Part II.

Data Coordinating Centers

7. Kids First

- Overview of the center/DCC and its mission.

- Key areas of focus or expertise.

- Contributions to the broader CFDE goals.

7.1. Events

- Summary of significant events (meetings, workshops, webinars).

- Objectives and outcomes of these events.

- Community engagement metrics (e.g., attendance, feedback).

7.2. Milestones

- Key achievements over the reporting period.

- Notable projects, publications, or releases.

- Alignment with CFDE objectives and future milestones.

7.3. Metrics

7.3.1. Google Analytics

- **Website Traffic**: Total visitors, unique visitors, geographic distribution.

- **Engagement**: Average time on page, bounce rate, popular pages.

- **Referrals**: Sources driving traffic (e.g., search engines, social media).

- **User Flow**: Navigation paths, exit points.

7.3.2. GitHub Metrics

- **Repository Activity**: Number of commits, issues opened/closed, pull requests.

- **Contributors**: Active contributors, new contributors.

- **Engagement**: Stars, forks, watchers, and downloads of releases.

- **Usage Insights**: Popular repositories or features, trends in usage over time.

7.3.3. Tool Usage

- Adoption and usage metrics for tools or software developed by the center/DCC.

- User feedback and satisfaction.

7.3.4. Data Downloads

- Volume of data accessed/downloaded.

- Breakdown by dataset or resource.

- Trends over time and geographic distribution.

7.3.5. Other Engagement Metrics

- Newsletter or email subscription statistics.

- Social media activity (e.g., Twitter mentions, followers).

- Training materials access (e.g., tutorials, videos, documentation).

7.4. Scientific impact

7.4.1. Publications

7.4.2. Mentions

Papers that mention data or tools but don't directly cite (full text search based on keywords)

7.4.3. Citation overview

- What are the themes associated with citations and mentions

7.5. Discussion and Recommendations

- Analysis of the metrics and their implications.

- Recommendations for improving engagement, tool adoption, or data accessibility.

- Future plans or initiatives.

Part III.

Partnerships

Part IV.

Cross-cutting Activities

8. Summary

In summary, this book is not yet even close to complete.

References