

Perspective/Review

Embedding FAIR Data Practices in Shared Core Resource Facilities

Mousumi Ghosh, Ph.D.¹, Ying Huang, M.D.¹, Emily S. Boja, Ph.D.¹, Jaime M. Guidry Auvil, Ph.D.¹, Subhashini Jagu, Ph.D.¹

¹ NCI Office of Data Sharing, National Institutes of Health

Keywords: Shared Resources, Cancer Centers, NCI-designated Cancer Centers, Data Management and Sharing, FAIR data practices, core-generated datasets, data producers, metadata, data stewardship

<https://doi.org/10.7171/001c.167764>

Data fragmentation, lack of standardization, and incomplete clinical context limit the utility and reuse of datasets in biomedical research. Core facilities (shared resources) generate a large portion of research data and are well positioned to implement FAIR (Findable, Accessible, Interoperable, and Reusable) data practices to align with the NIH Data Management and Sharing Policy. This article highlights strategies for incorporating FAIR data practices into core facility workflows.

Core facilities generate diverse data types across multiple platforms. However, variability in data formats and metadata practices can pose challenges for interoperability and cross-study integration. A critical first step in addressing this challenge is to connect core-generated datasets derived from patient biospecimens with structured clinical metadata. This involves using standardized identifiers, integration with biobanks and electronic health records. Structured metadata capture during experimental workflows to ensure data provenance will enable meaningful downstream analysis.

Equally important is the adoption of standardized data formats and metadata frameworks across technologies. Adoption of community standards for various data types, along with standard metadata schemas, improves data discoverability and reuse. Robust institutional infrastructure (e.g. laboratory information management systems, integrated data platforms) and coordinated governance frameworks can help to implement these efforts. Partnerships with institutional stakeholders and targeted workforce training further support the consistent implementation of effective data stewardship practices.

These approaches position core facilities as central operational hubs for advancing FAIR data principles, enhancing reproducibility, and maximizing the long-term value of biomedical research data.

Introduction

Shared instrumentation and core facilities are foundational components of the modern biomedical research enterprise as they centralize advanced technologies, specialized expertise, and high-value instrumentation. These facilities enable investigators across disciplines to access capabilities that would otherwise be beyond the reach of individual laboratories. As hubs within the institutional research ecosystem, core facilities foster collaboration and team science by connecting researchers, technologies, and data across laboratories and departments. They also play a critical educational role by introducing investigators and trainees to emerging technologies and providing training in experimental design, data generation, and analysis. By adopting standardized protocols and best practices, core facilities contribute significantly to research rigor, reproducibility, and transparency.¹⁻⁷

The increasing scale and complexity of data generated by modern biomedical technologies further elevate the importance of shared core resource facilities in shaping institutional data management practices.⁶ Because cores generate a substantial portion of institutional research data across

platforms, such as genomics, proteomics, and imaging, they are uniquely positioned to embed best practices for data stewardship directly into experimental workflows. Core facilities can ensure clear data provenance and readiness for downstream sharing and reuse by implementing standardized operating procedures, structured metadata capture, and consistent documentation of analytical parameters.

Importantly, shared resource facilities represent a strategic leverage point for advancing FAIR data practices across institutions. Operating as centralized environments that serve diverse research communities, cores can implement consistent workflows for metadata capture, data quality control, and documentation across multiple projects. Through their consultant role in experimental design and user training, core staff can introduce responsible data management practices at the earliest stages of research. By embedding FAIR-aligned data practices⁸ into routine service operations, core facilities can help operationalize emerging policy expectations — such as the National Institutes of Health (NIH) Data Management and Sharing Policy⁹ — while fostering a broader cultural shift toward more transparent, interoperable, and reusable research data.¹⁰⁻¹⁴ Three important steps in achieving this goal include: (1)

Table 1. Minimum set of clinical metadata elements relevant to clinical research

Clinical Data Category	Representative Data Elements	Relevance for Downstream Analysis
Demographics	Age at diagnosis, sex, race/ethnicity	Enables population stratification, epidemiologic analysis, and identification of disparities in disease incidence and outcomes
Disease Characteristics	Cancer type, primary tumor site, histology/morphology, stage, grade, basis of the diagnosis (i.e., pathology, molecular characterization)	Provides essential clinical context for interpreting molecular or imaging data and defining disease subtypes
Molecular and Diagnostic Features	Tumor biomarkers, genomic alterations, molecular subtype, diagnostic assay results	Allows integration of clinical and molecular data and supports precision oncology research
Treatment Information	Treatment modality (surgery, radiation, systemic therapy), treatment regimens, dates of therapy, line of therapy	Enables evaluation of treatment response and association of molecular features with therapeutic outcomes
Clinical Outcomes	Response to therapy, disease progression, recurrence, survival status, progression-free survival	Critical for linking molecular findings to clinical endpoints and evaluating prognostic or predictive biomarkers
Medical History and Comorbidities	Prior cancer history, comorbid conditions, family history	Supports interpretation of disease risk factors and patient heterogeneity
Social and Environmental Factors	Socioeconomic indicators, lifestyle factors (diet, smoking, physical activity), environmental exposures	Important for understanding cancer risk, treatment access, and outcomes
Healthcare Delivery Context	Care setting (inpatient/outpatient), institution type, treatment facility characteristics	Enables implementation science and health systems research
Longitudinal Clinical Data	Follow-up visits, disease course over time, updated outcomes	Enables studies of disease progression, survivorship, and long-term treatment effects

connecting clinical information and patient data to biosample-derived data generated at core facilities, (2) encouraging the use of standardized formats for datatypes and minimum metadata, and (3) institutional support and infrastructure that enables these efforts.

Connecting Shared Core Resource-Generated Data with Clinical Information

A critical first step in operationalizing FAIR data practices in shared resource facilities is ensuring that experimental data generated from patient-derived biospecimens are appropriately linked to relevant clinical information.¹⁵ Many technologies housed in core facilities — including genomics, proteomics, imaging, and other high-throughput platforms — analyze samples derived from tissues, blood, or other biospecimens obtained from patients during clinical care or clinical studies. While the resulting molecular or imaging datasets are often deposited in repositories or shared through publications, the absence of accompanying clinical context substantially limits their interpretability and downstream reuse. To enable meaningful secondary analysis and integration across studies, experimental data must be connected to patient-derived clinical metadata describing the individuals from whom the biospecimens were obtained.^{16,17}

Recent workshops convened by the National Cancer Institute (NCI) Office of Data Sharing identified a core set of clinical features that are broadly required to support translational and clinical cancer research using human-derived data.¹⁸ These features span several major categories, including demographics, disease characteristics, molecular and diagnostic profiles, treatment exposures, and clinical outcomes (Table 1). Establishing systematic mechanisms for connecting experimental outputs generated in core facilities to structured clinical metadata is a key operational challenge. One practical approach may be to establish metadata pipelines that connect biospecimen records and clinical data sources to experimental datasets, using laboratory information management systems (LIMS).

Standardizing Data Formats and Metadata to Enable Interoperability

In addition to connecting experimental data with clinical metadata, a critical requirement for making core-generated datasets FAIR is to adopt standardized data formats. A recent survey of shared instrumentation cores at cancer centers by the NCI Office of Data Sharing demonstrated that data produced by shared resource facilities, including proteomics, metabolomics, genomics, and imaging data, are typically generated using vendor-specific instruments and proprietary software (Figure 1). This data are then mainly shared in heterogeneous formats, making it difficult to in-

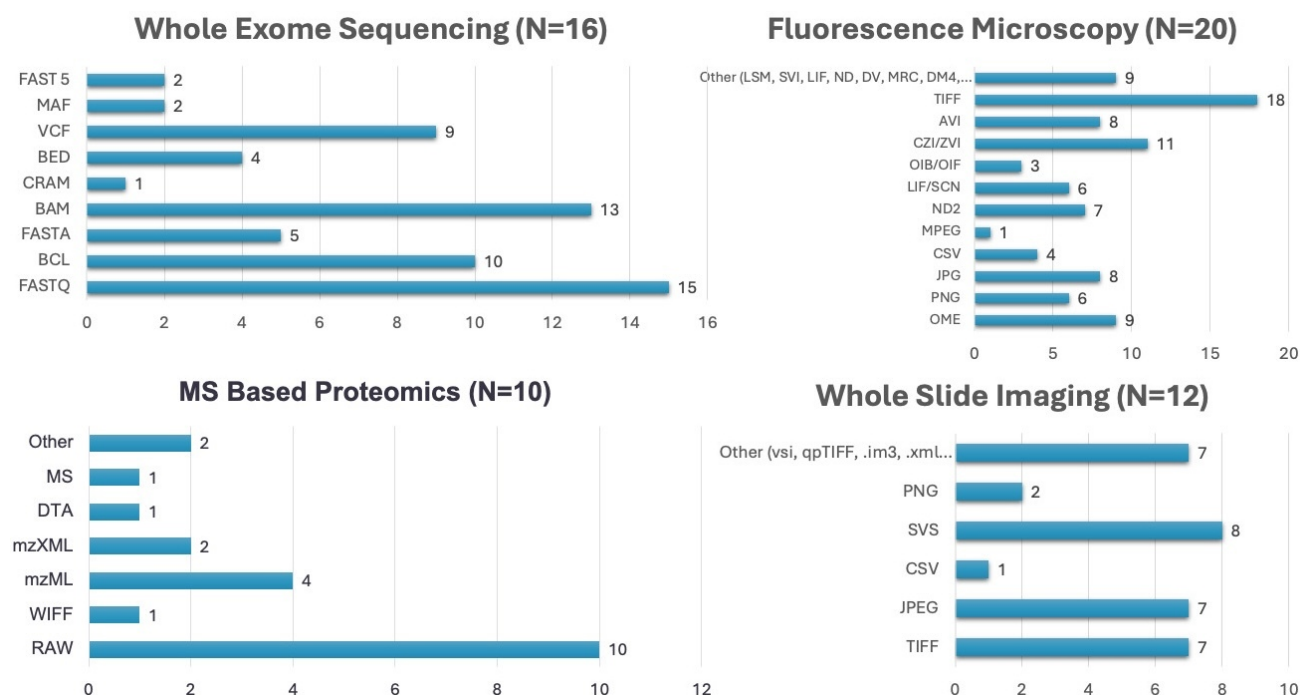


Figure 1. Data format heterogeneity across technologies in shared resource facilities across genomic, proteomic, metabolomic, microscopy and histology data.

tegrate datasets across platforms, institutions, and studies. As a result, even when datasets are publicly shared, the lack of standardized formats can significantly limit their discoverability, interoperability, and reuse.

Equally important is the structured capture of metadata during the experimental workflow. Core facilities routinely collect instrument settings, assay conditions, and analytical parameters as part of standard operating procedures. Extending these workflows to include structured fields for sample preparation details, and study-level metadata ensures that datasets can be accurately contextualized and linked to laboratory data maintained elsewhere in institutional data systems such as electronic laboratory notebooks (ELNs). These practices also support provenance tracking, which is critical for reproducibility and downstream data integration.¹⁹

Adopting community-driven standards is essential for addressing these challenges. In several technology domains commonly supported by shared resource facilities, well-established standards have already been developed. For example, the Human Proteome Organization Proteomics Standards Initiative (HUPO-PSI) has introduced formats such as mzML,²⁰ mzIdentML,²¹ and mzTab²² to standardize mass spectrometry data and other associated metadata.^{23, 24} Similarly, the metabolomics community has promoted standardized reporting and data exchange frameworks through initiatives such as the Metabolomics Standards Initiative (MSI)^{25,26} and repositories including MetaboLights. In imaging, the Open Microscopy Environment (OME) has developed interoperable file formats such as OME-TIFF and OME-NGFF,²⁷ along with metadata guidelines such as REMBI (Recommended Metadata for Biological Images)²⁸ to facilitate consistent data annotation and sharing.

Despite the availability of these frameworks, adopting them across laboratories and instrument vendors remains uneven.

Shared resource facilities can play a pivotal role in advancing data interoperability by embedding standards-compliant workflows into routine service pipelines. For example, core facilities can implement automated export pipelines that convert proprietary instrument outputs into community-supported formats, incorporate standardized metadata templates into sample submission workflows, and provide guidance to investigators on selecting appropriate repositories and data standards for their datasets. By integrating these practices directly into the data generation process, core facilities can ensure that datasets produced across institutional research programs are structured in ways that enable seamless discovery, integration, and reuse.

Building an Institutional Infrastructure to Support FAIR Data Generation in Shared Resource Facilities

Operationalizing FAIR data practices within shared resource facilities requires more than just adopting standards and capturing appropriate metadata; it also depends on the availability of institutional infrastructure that supports consistent data management across the research lifecycle. Because core facilities generate large volumes of highly structured experimental data across multiple technologies, they are uniquely positioned to serve as integration points between laboratory workflows and institutional data

ecosystems. Establishing the appropriate technical and organizational infrastructure (such as LIMS systems) that are integrated with ELNs and biobanks can enable these facilities to efficiently capture, manage, and disseminate research data in ways that align with FAIR principles and funder expectations.^{17,29,30}

Furthermore, partnerships with institutional libraries, research informatics groups, and information technology departments³¹ can support the development of data repositories, metadata catalogs, and data discovery tools that improve the findability and accessibility of core-generated datasets. Institutional data commons³² initiatives and federated data platforms³³ also provide mechanisms for harmonizing datasets across multiple research programs, which enable investigators to identify and reuse data generated across different laboratories and technologies.

Finally, core facility staff routinely provide technical training to investigators and trainees on experimental methods and instrument operation; these interactions also represent valuable opportunities to introduce best practices for data management, metadata documentation, and data sharing. By integrating data stewardship into routine research workflows and training activities, shared resource facilities can help cultivate a research culture in which FAIR data practices are viewed as a standard component of high-quality scientific work.

Together, these institutional infrastructure elements — data management systems, coordinated governance structures, and workforce training — create the foundation needed for shared resource facilities to generate datasets that are well-documented, interoperable, and ready for responsible sharing. Investments in these capabilities will be essential for enabling core facilities to fulfill their role as operational engines for FAIR data generation.

Conclusion

In conclusion, shared resource facilities — such as genomics, proteomics, imaging, and other biophysical cores — play a vital role in operationalizing federal data sharing policy. These facilities generate large-scale datasets whose downstream value depends on consistent linkage to a well-defined clinical context and other metadata. Furthermore, standardization of data across core facilities is critical for the data to be used together. Aligning shared resource practices with NIH data sharing expectations can improve compliance, enhance interoperability, and increase the long-term impact of publicly funded cancer research.

Financial Support/Conflict of Interest

This manuscript is the result of funding in whole or in part by the National Institutes of Health (NIH). It is subject to the NIH Public Access Policy. Through acceptance of this federal funding, NIH has been given a right to make this manuscript publicly available in PubMed Central upon the official date of publication as defined by NIH. The contributions of the NIH author(s) are considered works of the United States government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. The authors have no conflicts of interest to disclose.

Submitted: July 22, 2026 EDT. Accepted: July 22, 2026 EDT.

Published: September 01, 2026 EDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- De Paoli P. Institutional shared resources and translational cancer research. *J Transl Med*. 2009;7:54. doi:[10.1186/1479-5876-7-54](https://doi.org/10.1186/1479-5876-7-54)
- Mische SM, Fisher NC, Meyn SM, et al. A review of the scientific rigor, reproducibility, and transparency studies conducted by the ABRF research groups. *J Biomol Tech*. 2020;31(1):11-26. doi:[10.7171/jbt.20-3101-003](https://doi.org/10.7171/jbt.20-3101-003)
- Charalambakis NE, Ambulos NP Jr, Hockberger P, et al. Establishing a national strategy for shared research resources in biomedical sciences. *FASEB J*. 2021;35(11):e21973. doi:[10.1096/fj.202101393](https://doi.org/10.1096/fj.202101393)
- Mische SM. Maximizing shared research resources: next steps. *J Biomol Tech*. 2022;33(4). doi:[10.7171/3fc1f5fe.8b8abaaf](https://doi.org/10.7171/3fc1f5fe.8b8abaaf)
- Ambulos NP Jr, Mische SM, Fox JW. Greater reliance on shared resources can enrich the future success of academic research. *J Biomol Tech*. 2025;36(2):3fc1f5fe.e2548884. doi:[10.7171/3fc1f5fe.e2548884](https://doi.org/10.7171/3fc1f5fe.e2548884)
- Strom TA, Haugstad G, Shu J, Seshadri R. Shared instrumentation facilities: benefiting researchers and universities, and sustaining research excellence. *MRS Bull*. 2020;45(5):331-335. doi:[10.1557/mrs.2020.130](https://doi.org/10.1557/mrs.2020.130)
- Smith DB, Ferdous A, Oxford JT. Shared core facilities serve as hubs for biomedical research network at institutions of emerging excellence. *J Biomol Tech*. 2023;34(1). doi:[10.7171/3fc1f5fe.2f22458d](https://doi.org/10.7171/3fc1f5fe.2f22458d)
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3(1):160018. doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)
- Rahaman T. Open data and the 2023 NIH data management and sharing policy. *Med Ref Serv Q*. 2023;42(1):71-78. doi:[10.1080/02763869.2023.2168103](https://doi.org/10.1080/02763869.2023.2168103)
- Schmidt C, Hanne J, Moore J, et al. Research data management for bioimaging: the 2021 NFDI4BIOIMAGE community survey. *F1000Res*. 2022;11. doi:[10.12688/f1000research.121714.1](https://doi.org/10.12688/f1000research.121714.1). PMID:36405555
- DeVries M, Fenchel M, Fogarty RE, Kim BD, Timmons D, White AN. Name it! Store It! Protect It!: a Systems Approach to Managing Data in Research Core Facilities. *J Biomol Tech*. 2017;28(4):137-141. doi:[10.7171/jbt.17-2804-003](https://doi.org/10.7171/jbt.17-2804-003)
- Donaldson DR, Koepke JW. A focus groups study on data sharing and research data management. *Sci Data*. 2022;9(1):345. doi:[10.1038/s41597-022-01428-w](https://doi.org/10.1038/s41597-022-01428-w)
- Dragon JA, Gates C, Sui SH, et al. Bioinformatics core survey highlights the challenges facing data analysis facilities. *J Biomol Tech*. 2020;31(2):66-73. doi:[10.7171/jbt.20-3102-005](https://doi.org/10.7171/jbt.20-3102-005)
- Syn SY, Kim S. Characterizing the research data management practices of NIH biomedical researchers indicates the need for better support at laboratory level. *Health Info Libr J*. 2022;39(4):347-356. doi:[10.1111/hir.12433](https://doi.org/10.1111/hir.12433)
- Fouad K, Vavrek R, Surles-Zeigler MC, et al. A practical guide to data management and sharing for biomedical laboratory researchers. *Exp Neurol*. 2024;378:114815. doi:[10.1016/j.expneurol.2024.114815](https://doi.org/10.1016/j.expneurol.2024.114815)
- Marsolek W, Wright SJ, Luong H, Braxton SM, Carlson J, Lafferty-Hess S. Understanding the value of curation: A survey of researcher perspectives of data curation services from six US institutions. *PLOS One*. 2023;18(11):e0293534. doi:[10.1371/journal.pone.0293534](https://doi.org/10.1371/journal.pone.0293534)
- Oxford JT, Cornell KA, Romero JJ, et al. Center of Biomedical Research Excellence in matrix biology: building research infrastructure, supporting young researchers, and fostering collaboration. *Int J Mol Sci*. 2020;21(6):2141. doi:[10.3390/ijms21062141](https://doi.org/10.3390/ijms21062141)
- Ghosh M, Huang Y, Basehore HK, et al. Making human derived data FAIR: feedback from NCI office of data sharing workshop. *J Natl Cancer Inst*. Published online 2025:djaf379. doi:[10.1093/jnci/djaf379](https://doi.org/10.1093/jnci/djaf379)
- Johnston LR, Curty R, Braxton SM, et al. Understanding the value of curation: A survey of US data repository curation practices and perceptions. *PLOS One*. 2024;19(6):e0301171. doi:[10.1371/journal.pone.0301171](https://doi.org/10.1371/journal.pone.0301171)
- Martens L, Chambers M, Sturm M, et al. mzML—a community standard for mass spectrometry data. *Mol Cell Proteomics*. 2011;10(1):R110.000133. doi:[10.1074/mcp.R110.000133](https://doi.org/10.1074/mcp.R110.000133)
- Jones AR, Eisenacher M, Mayer G, et al. The mzIdentML data standard for mass spectrometry-based proteomics results. *Mol Cell Proteomics*. 2012;11(7):M111.014381. doi:[10.1074/mcp.M111.014381](https://doi.org/10.1074/mcp.M111.014381)

22. Griss J, Jones AR, Sachsenberg T, et al. The mzTab data exchange format: communicating mass-spectrometry-based proteomics and metabolomics experimental results to a wider audience. *Mol Cell Proteomics*. 2014;13(10):2765-2775. doi:[10.1074/mcp.O113.036681](https://doi.org/10.1074/mcp.O113.036681)
23. Snyder M, Mias G, Stanberry L, Kolker E. Metadata checklist for the integrated personal OMICS study: proteomics and metabolomics experiments. *OMICS J Integr Biol*. 2014;18(1):81-85. doi:[10.1089/omi.2013.0148](https://doi.org/10.1089/omi.2013.0148)
24. Taylor CF, Paton NW, Lilley KS, et al. The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol*. 2007;25(8):887-893. doi:[10.1038/nbt1329](https://doi.org/10.1038/nbt1329)
25. Alseekh S, Aharoni A, Brotman Y, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods*. 2021;18(7):747-756. doi:[10.1038/s41592-021-01197-1](https://doi.org/10.1038/s41592-021-01197-1)
26. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007;3(3):211-221. doi:[10.1007/s11306-007-0082-2](https://doi.org/10.1007/s11306-007-0082-2)
27. Moore J, Allan C, Besson S, et al. OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies. *Nat Methods*. 2021;18(12):1496-1498. doi:[10.1038/s41592-021-01326-w](https://doi.org/10.1038/s41592-021-01326-w)
28. Sarkans U, Chiu W, Collinson L, et al. REMBI: recommended Metadata for Biological Images-enabling reuse of microscopy data in biology. *Nat Methods*. 2021;18(12):1418-1422. doi:[10.1038/s41592-021-01166-8](https://doi.org/10.1038/s41592-021-01166-8)
29. Rehnert M, Takors R. FAIR research data management as community approach in bioengineering. *Eng Life Sci*. 2023;23(1):e2200005. doi:[10.1002/elsc.202200005](https://doi.org/10.1002/elsc.202200005)
30. Schmidt C, Boissonnet T, Dohle J, et al. A practical guide to bioimaging research data management in core facilities. *J Microsc*. 2024;294(3):350-371. doi:[10.1111/jmi.13317](https://doi.org/10.1111/jmi.13317)
31. Samuel SM, Sevryugina YV, MacEachern M, Saylor K, Woodbrook R. Stepping up to the moment: collaborating on a data management and sharing workshop series. *J Med Libr Assoc*. 2025;113(3):252-258. doi:[10.5195/jmla.2025.2070](https://doi.org/10.5195/jmla.2025.2070)
32. Sansone SA, Rocca-Serra P, Field D, et al. Toward interoperable bioscience data. *Nat Genet*. 2012;44(2):121-126. doi:[10.1038/ng.1054](https://doi.org/10.1038/ng.1054)
33. Beyvers S, Schlegel J, Brehm L, Hansen M, Goesmann A, Förster F. Towards FAIR and federated data ecosystems for interdisciplinary research. *PLOS Comput Biol*. 2026;22(2):e1013806. doi:[10.1371/journal.pcbi.1013806](https://doi.org/10.1371/journal.pcbi.1013806)