AMERICAN ACADEMY OF OPHTHALMOLOGY®

Editorial

# Barriers to Extracting and Harmonizing Glaucoma Testing Data: Gaps, Shortcomings, and the Pursuit of FAIRness

Niloofar Radgoudarzi, BS - *La Jolla, California*
Shahin Hallaj, MD - *La Jolla, California*
Michael V. Boland, MD, PhD - *Boston, Massachusetts*
Brian Stagg, MD - *Salt Lake City, Utah*
Sophia Y. Wang, MD, MS - *Palo Alto, California*
Benjamin Xu, MD, PhD - *Los Angeles, California*
Swarup S. Swaminathan, MD - *Miami, Florida*
Eric N. Brown, MD, PhD - *Nashville, Tennessee*
Aiyin Chen, MD - *Portland, Oregon*
Catherine Q. Sun, MD - *La Jolla, California*
Dilru C. Amarasekera, MD - *Philadelphia, Pennsylvania*
Jonathan S. Myers, MD - *Philadelphia, Pennsylvania*
Murtaza Saifee, MD - *West Hills, California*
William Halfpenny, MB BChir, MEng - *La Jolla, California*
Keri Dirkes, MPH - *La Jolla, California*
Linda Zangwill, PhD - *La Jolla, California*
Kerry E. Goetz, PhD, MS - *Bethesda, Maryland*
Michelle Hribar, PhD, MS - *Portland, Oregon; Bethesda, Maryland*
Sally L. Baxter, MD, MSc - *La Jolla, California*

***We describe potential improvements and barriers to data FAIRness in glaucoma testing and imaging, aiming to inform future efforts to standardize and enhance data availability for research.***

Glaucoma is a progressive optic neuropathy and is the leading cause of irreversible blindness globally with nearly 95 million people affected by the disease worldwide.[1,2] As the global population ages, the prevalence of glaucoma is anticipated to increase in the coming decades.[3,4] The significant prevalence and elevated incidence of blindness due to glaucoma raises the condition as a public health concern.[5] Leveraging big data sources to investigate novel therapeutic approaches, understand disease processes, identify risk factors, and more, presents an avenue for exploring comprehensive ways to tackle this condition.

Ophthalmology has been pioneering research in big data and artificial intelligence (AI), due to the widespread availability of noninvasive, rapid, and cost-effective ophthalmic imaging and accumulation of data through electronic health records.[6,7] Additionally, various ophthalmic big data sources, such as the American Academy of Ophthalmology IRIS® (Intelligent Research in Sight) Registry—the world's largest eye disease clinical registry managed by the American Academy of Ophthalmology—[8,9]along with administrative and health insurance databases, national biobanks, crowd-sourced data, and international epidemiology consortia, further contribute to advancing this field.[8,10,11]

Diagnosing and monitoring glaucoma requires a combination of data sources aside from the clinical examination of the optic nerve head, such as fundus photographs and/or OCT images, intraocular pressure, visual fields (VFs), and relevant factors such as family history and medical history.[12] Historically, lack of standards adoption and interoperability has impeded imaging and testing data from being included in big data sources, which are limited to diagnosis codes, procedure codes, and medications. Lack of large datasets with OCT and VF data, nonstandardized imaging or postprocessing protocols between devices, and inconsistency in the reporting metrics are major hurdles in using these data for analyses.[12] Analyzing testing and imaging data at scale is critical for advancing glaucoma research. In particular, there is currently a gap in linking systemic or primary care health data with ophthalmic phenotypes due to the lack of ophthalmic data in large databases; for example, the National Institutes of Health (NIH) *All of Us* Research Program includes extensive systemic data but limited ophthalmic data, whereas the



https://doi.org/10.1016/j.xops.2024.100621
ISSN 2666-9145/24

IRIS Registry has abundant ophthalmic data but limited systemic data.[8,13–15] This is important in glaucoma given that there is strong interest in identifying modifiable risk factors besides intraocular pressure that may become additional therapeutic targets.

The FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, introduced in a 2016 article in *Scientific Data*,[16] were developed by the international research community to enhance machine-actionability, enabling computational systems to autonomously find, access, interoperate, and reuse data, and have since been widely accepted by research institutions globally. To achieve *findability* for both humans and machines, datasets require clear descriptions, identification, and registration or indexing. This includes having sufficiently detailed descriptive metadata as well as a unique and persistent identifier such as a digital object identifier. *Accessibility* demands that datasets be available through defined access procedures, preferably automated, with consistently accessible metadata, even if the actual data are restricted. *Interoperability* is facilitated by adhering to common, published standards, when available, for the conceptualization, expression, and structure of both data and metadata, ensuring that they can integrate seamlessly with other datasets. *Reusability* entails detailed descriptions of data characteristics and their provenance according to relevant community standards, accompanied by transparent and accessible usage conditions.[16,17]

Adoption of the FAIR principles is an area of emphasis for the NIH and was included as part of the NIH Strategic Plan for Data Science, which includes incentivizing FAIR digital data deposition in NIH-funded repositories.[18] Additionally, NIH supports the NIH Common Data Elements Repository that aligns with FAIR data principles, enabling interoperability across datasets and facilitating efficient research practices. Common data elements reduce the workload for planning data collection often undertaken by clinical trial coordinating centers by minimizing the need for data collection design and consensus building and post hoc harmonization efforts.[19]

Given the growing focus on aligning data with FAIR principles, we evaluated current workflows for extracting glaucoma testing data. Here, we describe potential improvements and barriers to data FAIRness in glaucoma testing and imaging, aiming to inform future efforts to standardize and enhance data availability for research.

We represent a group of glaucoma researchers and informatics experts from 10 United States academic institutions who convened an online half-day workshop in February 2024 to share practices for large-scale extraction of VF and OCT data for glaucoma research (Fig 1). We included industry representatives to discuss current and future data extraction methods, share cross-institutional insights, and engage with vendors on tools to enhance data collection and sharing strategies. Details provided included devices used, use of Picture Archiving and Communication Systems, file types used, need for additional software licenses for access, and batch export functionality. Each institution shared challenges and socio-technical considerations, including collaborations with device manufacturers and information technology teams. Industry representatives were queried about standards conformance and future tools development.

The following barriers in extracting and harmonizing glaucoma imaging and testing data were identified based on consensus review. There was a wide range of approaches to extracting VF and OCT data for research, including manual transcription, optical character recognition of PDF documents, exporting structured data in various file formats (e.g., PDF, XML, JSON, DCM, VOL, and IMG), or exporting raw test or pixel data (Table 1). Many institutions employed multiple approaches simultaneously. Key challenges identified included the lack of standardized Digital Imaging and Communications in Medicine (DICOM) conformance for some modalities, dependence on proprietary Picture Archiving and Communication systems and additional licenses to access or export data, limited functionality in some cases to batch export data for a large number of patients, lack of familiarity among researchers with manufacturers' export options, and dependence on enterprise information technology support. The interoperability of data across organizations is also limited due to nonstandardized methods of representing VF pointwise data for varying grids and the lack of coverage of glaucoma imaging data elements among standardized medical terminologies. These issues represent gaps in alignment with FAIR principles (Table 2).

The challenges in extracting and harmonizing glaucoma testing and imaging data have significant implications for both clinical practice and research. In the clinical context, difficulty in transferring VF and OCT data from devices or Picture Archiving and Communication systems to electronic health records and other practices hampers the seamless coordination of care between health care providers.[20,21] From a research perspective, these barriers complicate the inclusion of detailed ophthalmic data in large-scale analyses that could otherwise enhance our understanding of glaucoma progression and treatment responses. Studies that attempt to correlate systemic conditions and medications with glaucoma outcomes often rely on proxies like using procedure codes from insurance claims,[13,14,22] but this is an imperfect approximation, and it would be ideal to correlate these associations with structural or functional progression using imaging data. The overarching move towards enhanced clinical interoperability, as emphasized by the United States Department of Health and Human Services, Office of the National Coordinator for Health Information Technology,[20] further highlights the necessity for overcoming these data barriers to enhance clinical interoperability and the availability of comprehensive, standardized datasets for advancing biomedical research.

Achieving accurate and reliable communication of image-based data proves challenging in environments employing multiple proprietary devices and data storage methods. The DICOM standard, recognized globally as a preferred medical imaging standard, offers a detailed framework for formatting and exchanging images along with associated information.[23] This specification encompasses various data, including image descriptions, patient demographics, capture protocols, and study

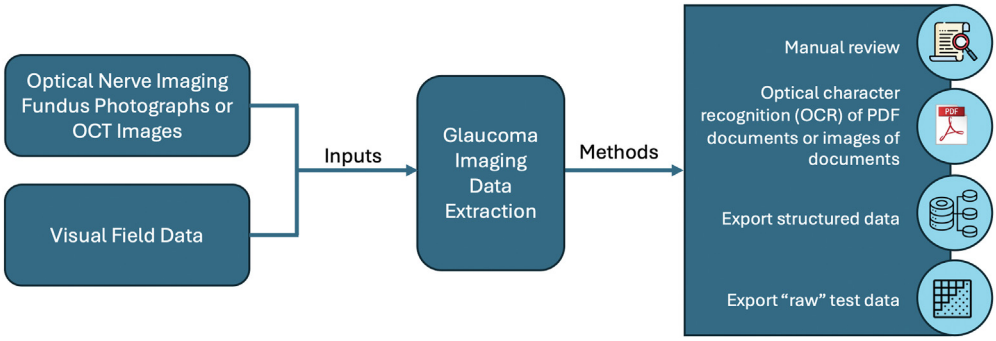


**Figure 1.** Overview of glaucoma imaging data extraction processes used by the institutions represented in this workshop.

outcomes. The American Academy of Ophthalmology has urged ophthalmic imaging device manufacturers to adopt standardized image formats conformant with DICOM, emphasizing the inclusion of raw images and metadata in standardized DICOM format for enhanced interoperability and data sharing.[24] Although DICOM has seen widespread adoption in radiology, cardiology, and radiotherapy, conformance remains inconsistent in ophthalmic imaging technologies.[24,25]

To advance the efficacy of clinical care and research and fully harness the potential of big data and AI in ophthalmology, several advancements are needed. First, there must be widespread full implementation of the DICOM standards across all imaging device vendors to ensure data from any

Table 1. Methods of Glaucoma Imaging Data Extraction

| Method of Data Extraction | Description | Challenges |
|---|---|---|
| Manual review | A team of researchers manually reviews the values in PDF documents or printouts and transcribes them | Highly labor-intensive and time-consuming<br>Substantial risk for transcription errors |
| OCR of PDF documents or images of documents | Several institutions use computational scripts leveraging regular expressions to secondarily parse data from PDF documents or images of structured reports/printouts and output them in formats more amenable to downstream analysis (i.e., CSV files) | Faster and more accurate than manual review but does not directly utilize the underlying data<br>Essentially a "workaround"<br>Risk for OCR errors and sensitivity to formatting/font changes |
| Export structured data (i.e., XML, FDA, vol, proprietary DICOM files, or standard DICOM files) | This approach directly exports structured data associated with the VF or OCT and avoids tedious manual review or an external workaround process<br>Can be done on an individual patient level or batches of multiple patients<br>Can be done using proprietary export tools or by locating and copying the DICOM database of the devices using command line. | Typically requires vendor-specific PACS and additional licensing of manufacturer-developed export tools<br>Batch export functionality not available for all imaging modalities (particularly raw images)<br>Even when batch export functionality is available, size of batches (e.g., number of patients whose data can be exported at once) can be limited<br>Potential impact on PACS performance when trying to export from the same database that is being used for daily clinical care |
| Export "raw" data | This entails exporting the underlying pointwise or pixel data in the image<br>Useful for developing artificial intelligence models/deep learning applications | Similar to the above, often needs vendor-specific PACS and additional licensing to allow export<br>Batch export functionality not available for all imaging modalities<br>Even when batch export functionality is available, size of batches (e.g. number of patients whose data can be exported at once) can be limited<br>Potential impact to PACS performance when also used for daily clinical care |

DICOM = Digital Imaging and Communications in Medicine; OCR = optical character recognition; PACS = picture archiving and communication systems; VF = visual field.

Table 2. Gaps in FAIRness for Glaucoma Imaging

| Principle | Current Gaps | Strategies for the Future |
|---|---|---|
| Findable | Glaucoma imaging data are not typically included in many large "big data" resources; often siloed in individual institutions<br>Some institutions also have multiple PACS (each focusing on 1 manufacturer's data), causing patient data to be spread across multiple systems and less findable | Improve pathways for data sharing<br>Assign persistent identifiers for imaging metadata and index into searchable resources |
| Accessible | Access is limited by requirements for proprietary PACS and need for additional licenses to export data<br>Some batch export options are not currently available or have limited capabilities<br>Access to these data also often depends on enterprise IT support, which can be variable | Encourage open access and removal of licensing requirements<br>Data should be accessible and exportable from any PACS<br>Develop data access workflows that are less dependent on enterprise IT |
| Interoperable | Current device manufacturers are at varying levels of DICOM conformance with most lacking adoption of existing standards that include data needed for research<br>Lack of representation of data elements in standardized terminologies<br>Workflows within institutions are not built to accept DICOM files, and modifying them could be resource-intensive | Advocate for DICOM conformance and data export functionality to be continued priorities for manufacturers<br>Develop standard concepts for data elements relevant for glaucoma imaging |
| Reusable | Multiple, competing schema for representing VF pointwise data (e.g., sequential numbering vs. x-y coordinates) make data difficult for reuse and limits ability to conduct cross-institutional analyses | Develop standard schema for pointwise representation that are applicable to multiple grids<br>Develop scripts to convert legacy schema to the standard schema |

DICOM = Digital Imaging and Communications in Medicine; FAIR = Findable, Accessible, Interoperable, and Reusable; IT = information technology; PACS = picture archiving and communication systems; VF = visual field.

device are viewable across different image viewers, and not limited by proprietary constraints. Despite this need, many researchers continue to use workaround methods such as optical character recognition or XML due to a lack of familiarity with the benefits that DICOM tools have to offer. If researchers become aware of and utilize DICOM's advantages, it could significantly reduce errors associated with these alternative methods. Another critical aspect is ensuring that data accessibility for physicians and patients is readily available without restrictions from proprietary licenses, facilitating both individual care and bulk data analysis. Additionally, there is a pressing need for additional standardization and sharing of common labels and definitions, data variables, as well as data sharing procedures. Embracing standardization would serve the interests of ophthalmologists, patients, and clinical care quality by promoting interoperability, facilitating comprehensive dataset creation for research, and advancing machine learning and AI algorithms.

The National Eye Institute, the Food and Drug Administration, and the Office of the National Coordinator for Health Information Technology are key organizations collaborating to improve the adoption of standardized imaging practices in ophthalmology.[20] The National Eye Institute is advancing this effort by integrating DICOM compliance into funding criteria and developing an open-access GitHub repository for shared coding scripts. The Food and Drug Administration supports these initiatives by formally recognizing the DICOM standard for ocular imaging devices, which allows manufacturers to streamline premarket submissions by declaring conformance with this standard. The Office of the National Coordinator for Health Information Technology is leveraging its regulatory and coordination authority to drive the integration of ocular imaging data into electronic health records, thereby facilitating easier and more consistent access to imaging data across health care systems.[20] In parallel, as AI models grow more prevalent, market pressure for standardization and interoperability may rise, driven by patient demands for transparency and data access, similar to how patient portals led to widespread note-sharing adoption.[26] Continued multi-institutional collaborative efforts and ongoing engagement with manufacturers are needed in the pursuit of making these data more FAIR.

## Footnotes and Disclosures

Footnotes and Disclosures

Michelle Hribar, PhD, MS, an Editor of this journal, was recused from the peer-review process of this article and had no access to information regarding its peer-review.

Disclosures: All authors have completed and submitted the ICMJE disclosures form.

The authors have made the following disclosures:

A.C.: Grants — NIH grant: GCAEI0492E.

M.V.B.: — Consulting — Carl Zeiss Meditec, Topcon Healthcare, Abbvie, Janssen; Payment or honoraria — Carl Zeiss Meditec.

B.S.: Grants — National Eye Institute (K23EY032577), Research to Prevent Blindness.

J.M.: Grants − Abbvie, Avisi, Elios, Equinox, Glaukos, Guardian, Laboratories Thea, Nicox, Olleyes, Santen; Consulting fees − Avisi, Glaukos, Olleyes.

N.R.: − Grants − Research to Prevent Blindness (Medical Student Eye Research Fellowship).

S.L.B.: Grants − National Institutes of Health, Research to Prevent Blindness; Consulting − Topcon; Payment or honoraria − Topcon; Support for attending meetings and/or travel − Topcon; Receipt of equipment − Optomed.

C.S.: Grants − National Eye Institute, All May See Foundation, American Glaucoma Society.

S.H.: Grants − OT2OD032644.

S.S.: Grants − National Eye Institute (K23-EY033831); Consulting − Sight Sciences, Abbvie/Allergan, Lumata Health; Payment or honoraria − Heidelberg Engineering; Participation in a data safety monitoring or advisory board − Topcon, Abbvie/ Allergan; Stock or stock options − Lumata Health; Other services − Alcon, Bausch & Lomb.

L.M.Z.: Grants − The Glaucoma Foundation, DRCR Retina Network/ JAEB, Center for Health Research, National Institutes of Health, National Eye Institute, The Krupp Foundation, Heidelberg Engineering; Consulting − AbbVie, Topcon Medical Systems; Patents − AISight Health; Stock − AISight Health Inc.; Receipt of equipment, materials, or other services −Carl Zeiss Meditec, Icare, Optovue/Visionix, Optomed, Topcon, Heidelberg Engineering.

The other authors have no proprietary or commercial interest in any materials discussed in this article.

This work was supported by National Institutes of Health (Bethesda, MD, USA) grants DP5OD029610, P30EY022589, R01EY034146, OT2OD032644, and an unrestricted departmental grant from Research to Prevent Blindness (New York, NY, USA). The sponsor or funding organizations had no role in the design or conduct of this research.

All authors have completed and submitted the ICMJE disclosures form.

An abstract version of this manuscript has been presented as a poster at the Association for Research in Vision and Ophthalmology (ARVO) Annual Meeting, May 5−9, 2024, Seattle, Washington, United States.

Abbreviations and Acronyms:
AI = artificial intelligence; DICOM = Digital Imaging and Communications in Medicine; FAIR = Findable Accessible, Interoperable, and Reusable; IRIS = Intelligent Research in Sight; NIH = National Institutes of Health; VF = visual field.

Correspondence:
Sally L. Baxter, MD, MSc, University of California San Diego, 9415 Campus Point Drive, MC 0946, La Jolla, CA 92093-0946. E-mail: s1baxter@health.ucsd.edu.

## References

1. Jayaram H, Kolko M, Friedman DS, Gazzard G. Glaucoma: now and beyond. *Lancet*. 2023;402:1788−1801.
2. Kang JM, Tanna AP. Glaucoma. *Med Clin North Am*. 2021;105:493−510.
3. Tham Y-C, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081−2090.
4. Zhang N, Wang J, Li Y, Jiang B. Prevalence of primary open angle glaucoma in the last 20 years: a meta-analysis and systematic review. *Sci Rep*. 2021;11:13762. https://doi.org/10.1038/s41598-021-92971-w.
5. Allison K, Patel D, Alabi O. Epidemiology of glaucoma: the past, present, and predictions for the future. *Cureus*. 2020;12:e11686. https://doi.org/10.7759/cureus.11686.
6. Lee CS, Brandt JD, Lee AY. Big data and artificial intelligence in ophthalmology: where are we now? *Ophthalmol Sci*. 2021;1:100036. https://doi.org/10.1016/j.xops.2021.100036.
7. Wang SY, Pershing S, Lee AY. AAO Taskforce on AI and AAO Medical Information Technology Committee. Big data requirements for artificial intelligence. *Curr Opin Ophthalmol*. 2020;31:318−323.
8. American Academy of Ophthalmology. IRIS Registry. https://www.aao.org/iris-registry; 2024. Accessed February 9, 2024.
9. Pershing S, Lum F. The American Academy of Ophthalmology IRIS Registry (Intelligent Research In Sight): current and future state of big data analytics. *Curr Opin Ophthalmol*. 2022;33(5):394−398.
10. Cheng C-Y, Soh ZD, Majithia S, et al. Big data in ophthalmology. *Asia Pac J Ophthalmol (Phila)*. 2020;9:291−298.
11. Allen N, Sudlow C, Downey P, et al. UK Biobank: current status and what it means for epidemiology. *Health Policy Technol*. 2012;1:123−126.
12. Huang X, Islam MR, Akter S, et al. Artificial intelligence in glaucoma: opportunities, challenges, and future directions. *Biomed Eng OnLine*. 2023;22:126. https://doi.org/10.1186/s12938-023-01187-8.
13. Zheng W, Dryja TP, Wei Z, et al. Systemic medication associations with presumed advanced or uncontrolled primary open-angle glaucoma. *Ophthalmology*. 2018;125:984−993.
14. Christopher M, Gonzalez R, Huynh J, et al. Proactive decision support for glaucoma treatment: predicting surgical interventions with clinically available data. *Bioengineering (Basel)*. 2024;11. https://doi.org/10.3390/bioengineering11020140.
15. National Institutes of Health. All of Us Research Program. https://allofus.nih.gov; 2024. Accessed June 15, 2024.
16. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.
17. Boeckhout M, Zielhuis GA, Bredenoord AL. The FAIR guiding principles for data stewardship: fair enough? *Eur J Hum Genet*. 2018;26:931−936.
18. National Institutes of Health. Data Management. https://sharing.nih.gov/data-management-and-sharing-policy/data-management. Accessed February 9, 2024.
19. National Institutes of Health. Common Data Elements: Standardizing Data Collection. https://www.nlm.nih.gov/oet/ed/cde/tutorial/index.html. Accessed February 9, 2024.
20. Goetz KE, Reed AA, Chiang MF, et al. Accelerating care: a roadmap to interoperable ophthalmic imaging standards in the United States. *Ophthalmology*. 2024;131:12−15.
21. Cai CX, Halfpenny W, Boland MV, et al. Advancing toward a common data model in ophthalmology: gap analysis of general eye examination concepts to standard Observational Medical Outcomes Partnership (OMOP) concepts. *Ophthalmol Sci*. 2023;3:100391. https://doi.org/10.1016/j.xops.2023.100391.
22. Baxter SL, Marks C, Kuo T-T, et al. Machine learning-based predictive modeling of surgical intervention in glaucoma using systemic data from electronic health records. *Am J Ophthalmol*. 2019;208:30−40.

23. Bidgood WD Jr, Horii SC, Prior FW, Van Syckle DE. Understanding and using DICOM, the data interchange standard for biomedical imaging. *J Am Med Inform Assoc*. 1997;4:199–212.
24. Lee AY, Campbell JP, Hwang TS, et al. Recommendations for standardization of images in ophthalmology. *Ophthalmology*. 2021;128:969–970.
25. Larobina M. Thirty years of the DICOM standard. *Tomography*. 2023;9:1829–1838.
26. Peterson JL, Houser SH. Unlocking patient portals: health information professionals navigating challenges and shaping the future. *Perspect Health Inf Manag*. 2024;21. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11102060/.