Open Data MICCAI 2026

I. Author Guidelines for Dataset Papers

Accepted papers will be presented at the Open Data MICCAI 2026 session and will be invited for submission to the MELBA Resource Track Special Issue on Open Data MICCAI 2026.

All submissions must follow the **MELBA Resource manuscript structure** and the additional Open Data MICCAI requirements described below.

II. General Submission Requirements

- **Page limit:** Maximum 8 pages (text, figures, tables) + up to 2 pages for references.
- **Template:** MELBA journal LaTeX template for the Resource Track [GitHub]
- **Manuscript type:** Resource Track.
- **Cover letter:** Not required at submission stage.
- **FAIR principles:** the resource must adhere to FAIR principles (Findable, Accessible, Interoperable, Reusable).
- **Reproducibility emphasis:** Submissions must prioritize transparency, documentation quality, and usability.
- **Repository Link**: to improve our review process, we now require the authors to provide the link to their repository upon submission. If the dataset is going to be shared with restricted access, the authors are invited to provide a private link or create a reviewer account for the Open Data 2026 reviewers to be able to access and review the repository.

III. Manuscript Structure

Each paper must follow the structure required by the MELBA template. Below we provide guidelines for each section. The aim is to help you improve your manuscript by enhancing its scientific soundness and style.

Note that papers that fail to adhere to instructions marked as Must Have may be desk-rejected.

1. Title

The title should clearly and concisely identify the resource and its scope.

Must Have - Clear, concise dataset/tool name

- Indication of modality/domain
- No overly broad claims

2. Abstract

The abstract should provide a compact overview of the resource, its motivation, and its intended use.

Must Have
- Purpose of the dataset/tool
- Imaging modality/domain
- Key characteristics and scale

Recommended
- Intended ML or clinical applications (e.g. segmentation of a structure, classification of disease)
- Key quantitative statistics
- Repository link
- Summary of validation experiments

3. Background

This section contextualizes the resource within the existing literature and clearly defines the gap it addresses.

Must Have
- Scientific or clinical motivation for building the resource
- Identified gap in existing resources
- Discussion of similar datasets/tools
- Clear statement of novelty

4. Summary

The summary should describe the vision, scope, audience, and readiness of the resource.

Must Have
- Vision and objectives
- AI-readiness statement
- Resources needed to use the dataset/tool

5. Discussion

The discussion should critically reflect on the strengths, limitations, and broader implications of the resource.

Must Have - Strengths - Known limitations - Biases or representation concerns - Recommendations for responsible use
Recommended - Planned extensions - Comparison with parallel initiatives

6. Resource Availability

This section must clearly explain where and how the resource can be accessed and under which conditions.

Must Have

- Data/code location:
  - Permanent public repository link
  - DOI (preferred)
  - Clear access procedure
- Potential use cases
  - Target ML tasks
  - Clinical domain
- Licensing
  - An explicit standard license statement compatible with open data licenses
    - If unsure, consult our licensing guide and your institution's legal or data protection office before submission.
  - License file present in the repository (e.g. `LICENSE.txt`)
  - License name and URL cited in the paper itself
  - Confirmation that the license is compatible with any source/upstream data licenses
  - License version explicitly noted (e.g. 4.0, not just "CC BY")
- Ethical considerations
  - Description of data composition
    - If data was collected across multiple sites, ethics approval from each contributing institution must be declared.
    - Inclusion of vulnerable populations (e.g., children)
    - Data provider location
  - Ethical approval statement/number
    - If not required, an explanation of exemption or a formal statement
    - Description of what data is approved for open release and under which conditions including who takes the decision of final release

- Informed consent description
  - Describe if explicit secondary use consent, broad consent, waiver of consent, or legal basis other than consent
- Define data sensitivity level
- Anonymization/pseudonymization procedures
- Compliance with original licenses (if derived data)
- A statement of intended use and known misuse risks
- Official contact for data and ethics questions or requests

Recommended

- Data/code location:
  - Version history (e.g. previous dataset version)
- Licensing
  - A plain-language summary of what users can and cannot do, beyond what the license name already communicates.
  - If a restrictive clause is used (e.g. NC, SA), a brief explanation of why, so reviewers and users understand the constraint.
  - A separate license declared for any accompanying code (since CC is not designed for software)
  - A license badge in the repository README
- Ethical considerations
  - Data protection impact assessment (provide full version as supplement)
  - Ethical risk impact assessment
  - Declare whether an ethical use clause accompanies the dataset and, if not, why the chosen license is considered sufficient.
    - An ethical use clause attached to the repository, referencing at minimum: prohibition on re-identification attempts, requirement to report suspected data breaches, and commitment to use data in accordance with applicable research ethics standards. The Responsible AI Licenses (RAIL) framework offers ready-made templates designed for exactly this purpose.
  - A versioning policy stating how future changes, record removals, or participant withdrawals will be handled and communicated.

7. Methods

This section must provide sufficient methodological detail to ensure transparency and reproducibility.

Must Have

- **Data details:** To describe the cohort and dataset composition in sufficient detail for reproducibility.
  - Inclusion and exclusion criteria
  - Data collection timeline (retrospective vs prospective)
  - Demographic description
    - Must have: age, sex
  - Clinical/molecular variables (if applicable)

- **Methods Used for Data Creation:** To describe acquisition, preprocessing, and preparation steps from raw data to released dataset.
    - Acquisition equipment (manufacturer, model)
    - Acquisition parameters and Imaging protocol
    - Data format (e.g. DICOM, PNG, CSV)
    - Preprocessing pipeline (preferably with open-source code)
    - Data anonymization pipeline (preferably with open-source code)
    - Known artifacts or limitations
    - Modality-specific information (*see section IV*)
    - Annotation procedure and ground truth definition
    - Annotator expertise

Recommended

- Data details
    - Recruitment flow diagram
    - Diversity indicators (e.g. ethnicity)
- Methods used for data creation
    - Site harmonization details if the dataset is multi-center
    - Inter-observer variability
    - Clinical standard followed for generating the labels (e.g. Bi-RADS)

8. Validation

This section must demonstrate the quality, reliability, and usability of the resource.

Must Have

- Information on quality control procedures (preferably with open-source scripts)
- Quality control metrics
- Subsections should correspond to the “Methods Used for Data Creation in section III.7)

Recommended

- Example ML/AI use case

9. Conflicts of Interest

Must Have

- Financial disclosures

10. Acknowledgments

Must Have - Funding sources and grant numbers

11. References

Must Have - Adhere to the style provided in the MELBA template - Provide DOI for the referenced works when possible

IV. Recommended modality-specific metadata

We encourage you to provide as much metadata as possible in a standardized format. The tables below serve as a recommended reporting checklist. Not every variable will apply to every dataset; authors should report what is relevant and explain any omissions.
The modalities listed are not exhaustive; if your dataset involves a modality not covered below, please report analogous acquisition and protocol parameters relevant to your domain.

A. Imaging Modalities

Modality	Recommended information to be reported in the paper
MRI	- Sequence types - Magnetic field strength - Scanner manufacturer and model - Slice thickness and spacing - Voxel size / spatial resolution - Acquisition plane and parameters (TR, TE, TI if applicable) - Flip angle - Usage of contrast agent
CT	- Tube voltage (kVp) - Tube current (mA or mAs) - Slice thickness - Pixel spacing - Reconstruction algorithm - Contrast use (yes/no, type, timing phase)
PET	- Radiotracer type (e.g., FDG, amyloid tracer) - Injected dose (if quantitative analysis is intended) - Uptake time

	- Acquisition duration - Reconstruction method
X-Ray	- Projection type (AP, PA, lateral) - Tube voltage (kVp) - Tube current (mA or mAs) - Pixel spacing
Ultrasound	- Probe type (linear, convex, phased-array) - Imaging mode (B-mode, Doppler, elastography) - Frequency range - Depth settings - Resolution (axial/lateral if known)
OCT	- Scan pattern (e.g., macular cube) - Axial resolution - B-scan spacing - Scan area dimensions
Fundus	- Field of view (e.g., 45°) - Resolution - Color vs grayscale
Digital pathology	- Magnification level (e.g., 20x, 40x) - Pixel resolution (µm/pixel) - Staining protocol (e.g., H&E, IHC marker type) - Tissue preparation protocol
Microscopy	- Objective magnification and NA - Imaging modality (brightfield, fluorescence, confocal) - Channel description (wavelengths) - Resolution - Sample preparation protocol
Endoscopic videos	- Endoscope type (rigid, flexible) - Procedure type (e.g., laparoscopy, colonoscopy, bronchoscopy) - Imaging modality (white light, NBI, fluorescence, ICG) - Resolution (e.g., 1920×1080) - Frame rate (fps) - File format and encoding
Mammography / Digital Breast Tomosynthesis	- Compression type - Pixel spacing - Detector type - Number of projection angles and reconstruction method (for DBT)

B. Non-Imaging Modalities

Datasets submitted to Open Data MICCAI are expected to center on medical imaging. Non-imaging data modalities are welcome when provided as part of a multimodal resource alongside imaging data. Standalone non-imaging datasets are not excluded but should demonstrate clear relevance to the medical image computing community. The following metadata is recommended for commonly paired non-imaging modalities.

Modality	**Recommended information to be reported in the paper**
EEG	- Amplifier model - Sampling rate - Montage (e.g., 10-20 system) - channel count - Filter settings (High/Low-pass) - Recording condition (e.g., Sleep, Task).
DNA Sequencing	- Sequencing platform & model - Library type (WGS, WES, Panel) - Average read depth & length - Reference genome version.
Electronic Health Records	- Data model (e.g., OMOP, FHIR) - Timeframe & setting (Inpatient/Outpatient) - Included variables (ICD codes, Vitals)
Clinical Reports (Unstructured Text)	- Report type (Radiology, Pathology) - Language - Processing status (Raw vs. Cleaned)