

## Data Description

Data Number (will be filled by the Commons Center)	DPPSCdbp_2025-06en
Title	PRISM Clinical Information Annotation for Medical Event Summaries from Patient Blogs (Tobyoki)
Creator	Shuntaro Yada (Faculty of Library, Information and Media Science, University of Tsukuba)
Contact information	Shuntaro Yada, <a href="mailto:yada@slis.tsukuba.ac.jp">yada@slis.tsukuba.ac.jp</a>
Summary (Subject and/or Methodology for generation)	This corpus provides clinical information annotations on medical event summaries extracted from 100 breast cancer patient blogs (TobyokiSummary). Annotations follow the PRISM guideline, covering 13 entity types (diseases, anatomical parts, tests, medications, temporal expressions, etc.) and 10 relation types (temporal and spatial relations). All annotations were automatically generated using a large language model (LLM) via the prism-annotator tool, which was developed and released as part of this project.
Update History (Versions, File names, Dates)	v1.0 / tobyoki-prism-dist.zip / March 2026
Data formats	XML (individual document files and merged corpus file), JSON (entity and relation data), HTML (interactive viewer)
Data sizes	Approximately 11 MB (1.5 MB compressed). 100 documents, 14,797 entities, 7,035 relations.
Things to be aware of when using the data (e.g., required metadata)	This dataset contains annotations on medical event summaries derived from patient blog texts ( <a href="https://github.com/sociocom/TobyokiSummary">https://github.com/sociocom/TobyokiSummary</a> ), which does not contain original blog body text. Annotations were automatically generated by an LLM and have not been strictly verified by human.
Related reports and/or papers	(Yada, Sugawara, Kizaki, & Nishiyama (2025). Attempt to extract emotion-cause events from patient narratives (Tobyoki). YANS 2025. [in Japanese]) Shimizu, Hisada, Uno, Yada, Wakamiya, & Aramaki (2025). Exploring LLM Annotation for Adaptation of Clinical Information Extraction Models under Data-sharing Restrictions. Findings of ACL 2025. Yada et al. (2020). Towards a Versatile Medical-Annotation Guideline. LREC 2020.
Other remarks	Annotation tool: <a href="https://github.com/sociocom/prism-annotator">https://github.com/sociocom/prism-annotator</a> (PyPI: prism-annotator)