

The Kymata Atlas: Data Sharing and Data Access Policy

Version 1.00 (Author: AT, 17/08/16) Initial drafting of document.

Version 1.01 (Authors: AT, SL, MT 08/11/16) Edits following amendments from Somaya Langley and Marta Teperek.

1. AIMS	1
2. BACKGROUND	1
3. WHAT KYMATA AIMS TO SHARE BY DEFAULT	2
4. HOW KYMATA AIMS TO SHARE BY DEFAULT	3
4.1 FILE FORMATS	3
4.2 DOCUMENTATION	4
4.3 LICENSES	4
5. INFORMATION KYMATA DOESN'T SHARE BY DEFAULT, AND HOW TO REQUEST IT.	5
6. CITING KYMATA DATA	6
REFERENCES	6

1. Aims

The aim of this document is to set out how the developers of Kymata intend the data associated with the atlas to be shared with others.

2. Background

Kymata is an instantiation of a procedure that estimates ‘information processing pathways’ from neuroimaging data, namely electro- and magnetoencephalography data. The resulting pathways graph is known as a map or atlas. ‘Kymata’ may refer to the front-end and back-end of the procedure used to generate this atlas.

There are four types of raw measurement data that are associated with Kymata, one type of hypothesis data and one form of result data.

Raw measurement data:

- **Experimental stimuli** The stimuli experienced by the participants being recorded. (.wav, .tiff, etc.)

- **Raw EMEG measurements of individuals (plus individual structural meshes)** The electro- and magnetoencephalography recordings of the individual participants experiencing the stimuli, together with the structural T1 scans of their brains. (.fif, format)
- **Average EMEG measurements (uses average structurals)** The above recordings, averaged together into a single, average recording. (.fif format)
- **Average source current estimates (uses average structurals)** Each individual raw electro- and magnetoencephalography recording can be used to estimate individual source current estimates. When these source current estimates are averaged, these result in the average source current estimates. (.stc format)

Hypothesis data:

- **Hypothesized functions.** The mathematical transforms that the brain may (or may not) be engaged in, in code format. (.py, .mat, .gcc, .java, etc.)

Not sure what a function is? See <https://kymata.org/documentation> for more information.

Result data:

- **Processing pathway map** The processing map generated when Kymata is fed the average source current estimates, the experimental stimuli and hypothesized functions. The format of this map is a directed, acyclic graph.

3. What Kymata aims to share by default

By default, Kymata aims to allow public access to all data that is not linked to individuals, using licenses that are as close to free, non-commercial use as possible.

Data that is linked to individuals may be shared, but only if certain caveats are adhered to, including strict anonymization, participant consent, and third-party agreements.

Kymata data only shares anonymized or averaged data, and as such this data is not bound by the Data Protection Act 1998.

4. How Kymata aims to share by default

As a guide to what the developers intend to share, and how, we use the FAIR Guiding Principles for Scientific Data Management and Stewardship (Wilkinson et al., 2016). The main principles are Findability, Accessibility, Interoperability, and Reusability:

Findable	F1. (meta)data are assigned a globally unique and persistent identifier F2. data are described with rich metadata (defined by R1 below) F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource
Accessible	A1. (meta)data are retrievable by their identifier using a standardized communications protocol A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available
Interoperable	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data
Reusable	R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

From Wilkinson et al (2016)

4.1 File formats

In general, Kymata will aim to share data in formats that are open and free, although in some cases this will not be possible because the underlying data is created in a proprietary format.

Here are the current file formats for the data

Data type	File formats
Experimental stimuli	.wav, .tiff (open, free)
Raw brain-activity measurements of individuals (plus individual	.fif format (proprietary Elekta, free

structural meshes)	to open)
Average EMEG measurements (uses average structurals)	.fif format (proprietary Elekta, free to open)
Average source current estimates (uses average structurals)	.stc (open, free)
Hypothesized functions	.py (open, free) .mat (proprietary Mathworks) .c (open, free) .c++ (open, free) .java (open, free) etc. <i>This list is not exhaustive, functions can be written in any programming language.</i>
Processing pathway data	.JSON (open, free)

4.2 Documentation

The documentation for the experimental stimuli, raw brain-activity measurements of individuals, average brain-activity measurements, average source current estimates can be found at <https://kymata.org/datasets>.

The documentation for the hypothesized functions can be found at <https://github.com/kymata-atlas>.

The documentation for the pathways graph can be found at <https://kymata.org/documentation>.

4.3 Licenses

Average EMEG measurements, average source current density data. These are made available under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Hypothesized functions. Each has a different author, and therefore a specific license. Those that are written by the Kymata Hypothesis group (KHG) are made available under a [Creative Commons Attribution 4.0 International License](#).

The processing maps. These are made available under licensing terms drawn up by the University of Cambridge. <https://kymata.org/licenses>. This license stipulates that this data is for research and non-commercial use only.

Individual's raw measurement recordings Permission to access is only given with written consent (see section 5). There is no license, however, users may only use data for research agreed with the PI; users may not share this data; and must acknowledge the funders of the measurements in any publications.

5. Information Kymata doesn't share by default, and how to request it.

Individual data. Individual's measurements (plus meshes) are not publically available. Researchers requesting this data must send the request in writing (or via email) to the Kymata Principal Investigator (support@kymata.org). This data will only be shared if

- a) the subject has given signed consent for the data to be used for research by third parties
- b) this individual's data is anonymized (including the accompanying structural scans)
- c) the requester is a member of a *bona fide* research group which has an official ethics statement authorized by a third-party ethics or medical board of similar standard to that of the proposed study
- d) the requester undertakes, in writing, not to share the data beyond their immediate research group, and to secure this information while it is in their position.
- e) the requester undertakes to cite the data correctly if published.

Final decision lies with the Kymata Principal Investigator.

If shared, the individual's structurals will be face-scrambled to avoid identification. The data is also anonymized (identified only by random ID), and the researcher will not be able to link the meshes to any personal details, aside from their age, gender and population group.

6. Citing Kymata data

Raw measurement data (Experimental stimuli, Raw EMEG measurements of individuals, Average EMEG measurements, Average source current estimates)

- Raw measurement datasets should be cited using the reference given in the dataset documentation accompanying the download.

Hypothesis data (Function code)

- Hypothesis data files should be cited using the Github repository and commit. eg.

Thwaites. A. (2016) The flutter-range vertical displacement function.
<https://github.com/kymata-atlas/KHG-functions/>, commit 6760389f

This reference format will be replaced with Zenodo DOI citable references in the future (<https://guides.github.com/activities/citable-code/>)

Result data (Processing pathway maps)

- Please see <https://kymata.org/documentation> for full details.

References

The Data Protection Act (1998) The Data Protection Act 1998, Chapter 29, UK Parliament, <http://www.legislation.gov.uk/>

Wilkinson MD et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Nature Scientific Data* 3:160018 doi: 10.1038/sdata.2016.18

Thwaites A, Wieser E, Soltan A, Nimmo-Smith I, Zulfiqar I (*in prep.*) Kymata, a graph of information processing in the human cortex. *Nature Scientific Data*