Data Catalog Task Force

Chair: Sean Hill, Krembil Centre for Neuroinformatics, Canada
Co-Chair: Tom Johnstone, National Imaging Facility, Australia

Mathew Abrams, INCF
Nargis Akter, Japan Brain/MINDS
Amy Bernard, Allen Institute
Jan Bjaalie, EBRAINS/HBP
Andrew Davison, EBRAINS/HBP
Kenji Doya, Japan Brain/MINDS
Mohameth François Sy, Blue Brain Project
Satrajit Ghosh, US Brain Initiative - DANDI
Tom Gillespie, US Brain Initiative, SPARC, EBRAINS
Jeffrey Grethe, US Brain Initiative, SPARC
Megumi Maruyama, Japan Brain/MINDS
Aggie McMahon, Kavli Foundation
Visakh Muraleedharan, INCF
JB Poline, Canadian Open Neuroscience Platform
Alexander Woodward, Japan Brain/MINDS
Yi Zeng, China Brain Project

Data Catalog Task Force Objective:

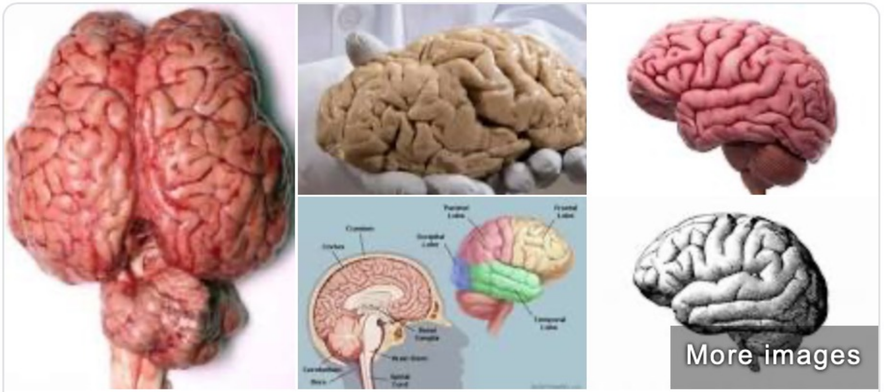
Establish a sustainable data search capability for the output of the international brain initiatives

The Google Knowledge Graph

Knowledge Graph

The Knowledge Graph is a knowledge base used by Google and its services to enhance its search engine's results with information gathered from a variety of sources. The information is presented to users in an infobox next to the search results. Wikipedia

The Google Knowledge Graph



Human brain

The **brain** is one of the largest and most complex organs in the human body. It is made up of more than 100 billion nerves that communicate in trillions of connections called synapses. ... The **brain** stem is between the spinal cord and the rest of the **brain**. Basic functions like breathing and sleep are controlled here. May 18, 2019

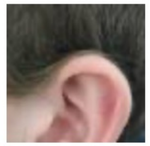
Brain (Human Anatomy): Picture, Function, Parts, Conditions ...

https://www.webmd.com › brain › picture-of-the-brain

People also search for

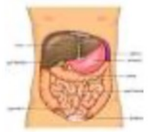
View 5+ more


Eye


Ear

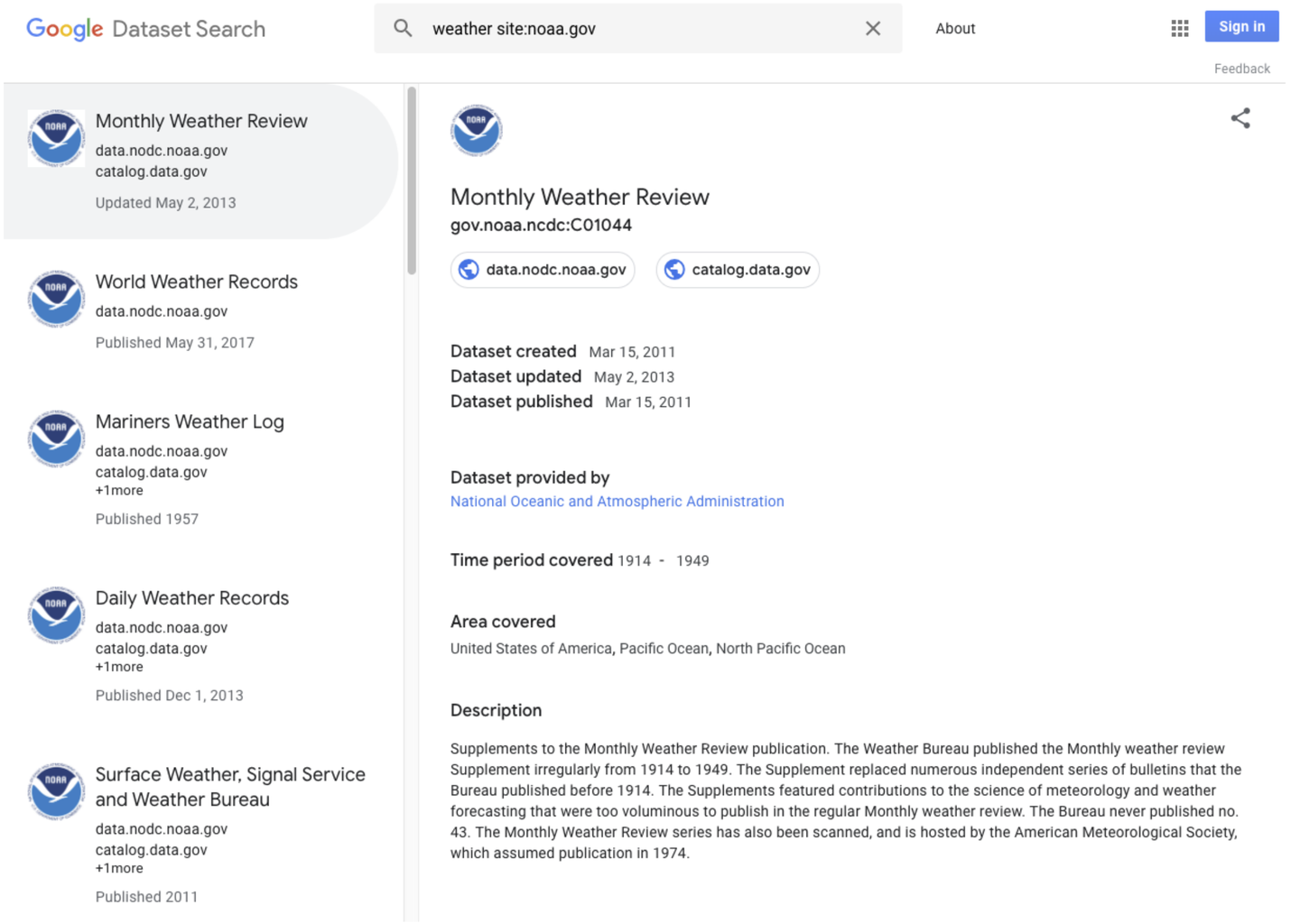

Spinal cord


Hypothal…


Liver

Feedback

Google Dataset Search



https://developers.google.com/search/docs/data-types/dataset

Currently only supports whole datasets, not granular individual data entities.

Structured data

Google Search

Search

HOME | GUIDES | REFERENCE | CASE STUDIES | APIS | TOOLS | SUPPORT

Introduction to Structured Data

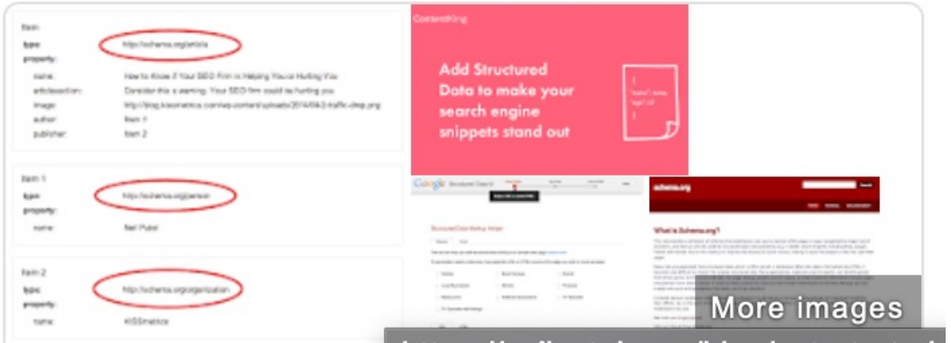
☆ ☆ ☆ ☆ ☆

Google Search works hard to understand the content of a page. You can help us by providing explicit clues about the meaning of a page to Google by including structured data on the page. Structured data is a standardized format for providing information about a page and classifying the page content; for example, on a recipe page, what are the ingredients, the cooking time and temperature, the calories, and so on.

Google uses structured data that it finds on the web to understand the content of the page, as well as to gather information about the web and the world in general. For example, here is a JSON-LD structured data snippet that might appear on the contact page of the Unlimited Ball Bearings corporation, describing their contact information:

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "Organization",
  "url": "http://www.example.com",
  "name": "Unlimited Ball Bearings Corp.",
  "contactPoint": {
    "@type": "ContactPoint",
    "telephone": "+1-401-555-1212",
    "contactType": "Customer service"
  }
}
</script>
```

schema.org



Schema.org

Schema.org is a collaborative community activity with a mission to "create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond." Wikipedia

Year started: 2011

Base standards: URI, HTML5, RDF, Microdata, ISO 8601

Related standards: RDFa, Microformat, RDFS, OWL, N-Triples, Turtle, JSON, JSON-LD, CSV

Organization: Google, Yahoo, Microsoft, Yandex

Domain: Semantic Web

License: CC-BY-SA 3.0

The schemas are a set of 'types', each associated with a set of properties. The types are arranged in a hierarchy.

The core vocabulary currently consists of 614 Types, 902 Properties, and 114 Enumeration values.

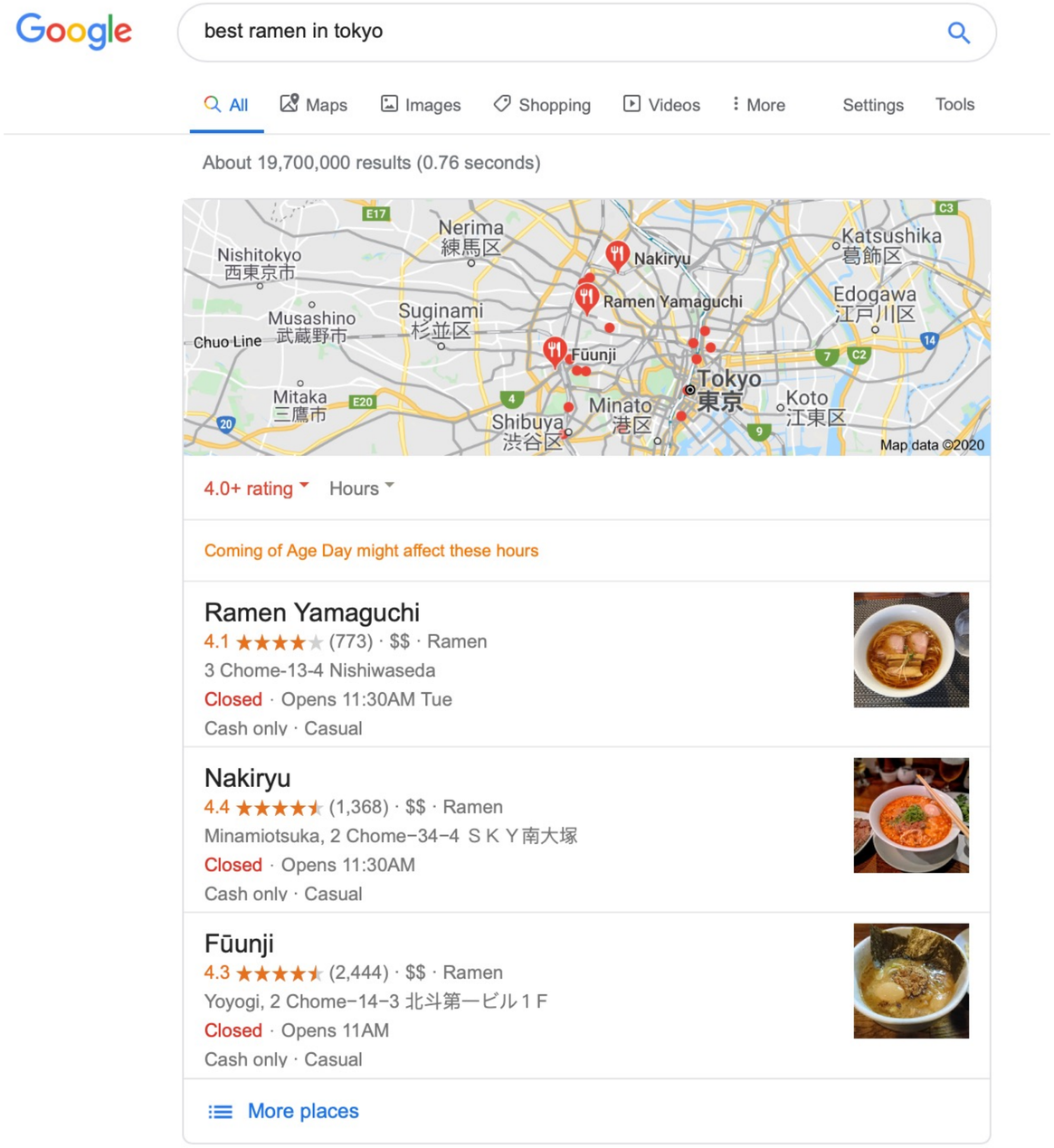
- Creative works: CreativeWork, Book, Movie, MusicRecording, Recipe, TVSeries ...
- Embedded non-text objects: AudioObject, ImageObject, VideoObject
- Event
- Health and medical types: notes on the health and medical types under MedicalEntity.
- Organization
- Person
- Place, LocalBusiness, Restaurant ...
- Product, Offer, AggregateOffer
- Review, AggregateRating
- Action

Why do businesses implement structured data?

Structured data helps sites appear in Google's Knowledge Graph

For sites that appear in highly competitive verticals, **getting the edge over your competition is critical**, and one way to do this is by establishing your site presence with Google and **appear**ing **in the Knowledge Graph**.

To enable your business Knowledge Graph card, earlier you would have needed to **add the necessary Corporate Contact markup on the homepage of your website.**

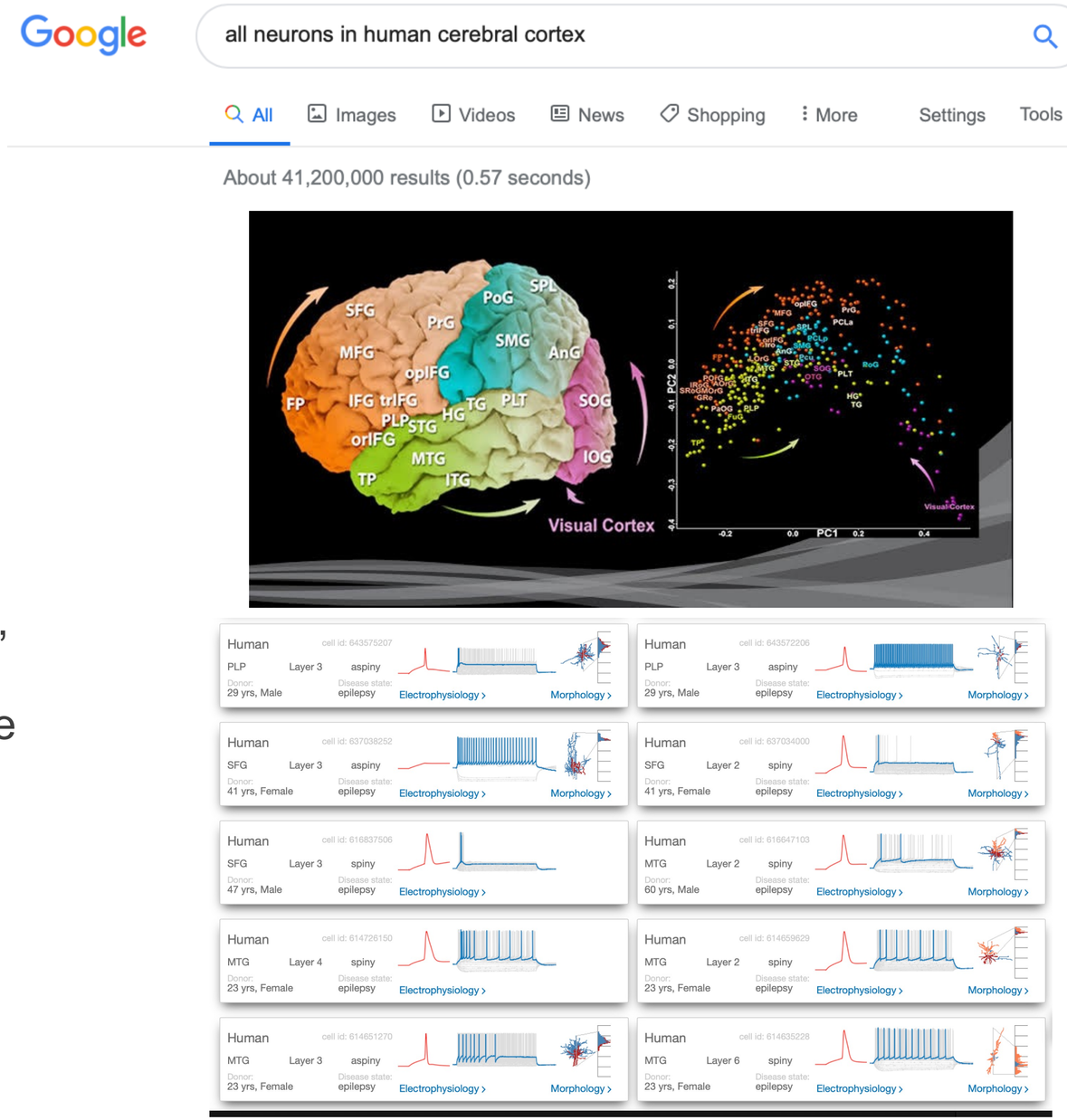


Towards an International Brain Initiative Knowledge Graph…

Why should neuroscientists implement structured data descriptions in their websites?

What are the **incentives** in neuroscience that will make it so motivating to establish the presence of your data in an **International Brain Initiative Knowledge Graph**?

To enable your neuroscience data Knowledge Graph card, earlier you would have needed to **add the necessary Brain Research Data markup on your website.** Then the a data card could be automatically added to the IBI Knowledge Graph.



First review all current brain initiative approaches to structured metadata:

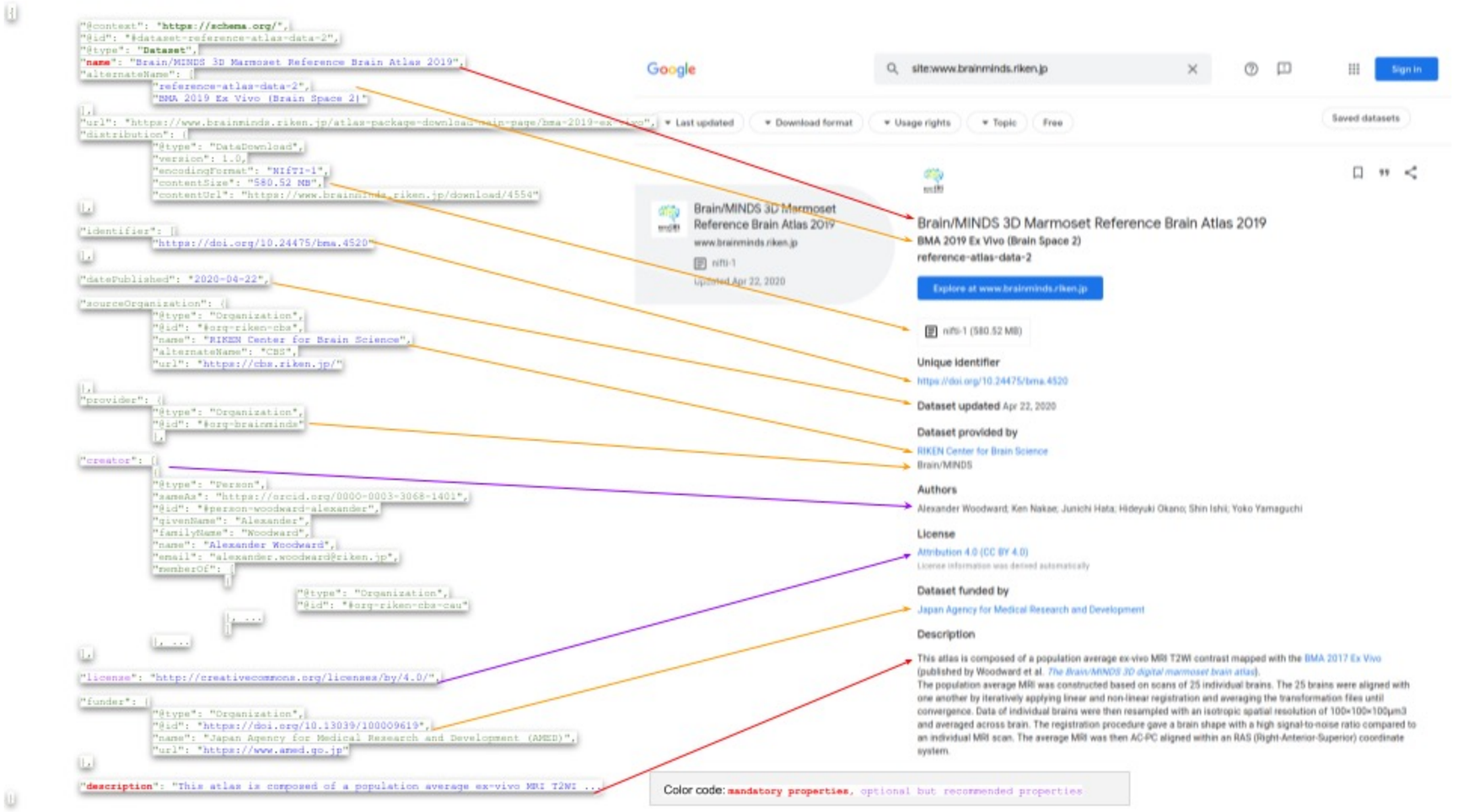
- Human Brain Project - EBRAINS Knowledge Graph
- Blue Brain Project - Neuroshapes
- Japan Brain/MINDS Data Portal
- Canadian Open Neuroscience Platform
- INCF KnowledgeSpace
- DANDI: Distributed Archives for Neurophysiology Data Integration
- SPARC Data Structure

Example: Brain/MINDS DATA PORTAL

- Launched in 2014, Brain/MINDS is Japan’s Brain Mapping Project.
- The core institute for the project is RIKEN and it was sponsored by MEXT in 2014 and now AMED since 2015.
- The unique appeal of the Brain/MINDS project is its focus on mapping the brain of a small new world monkey, the common marmoset (Callithrix jacchus).

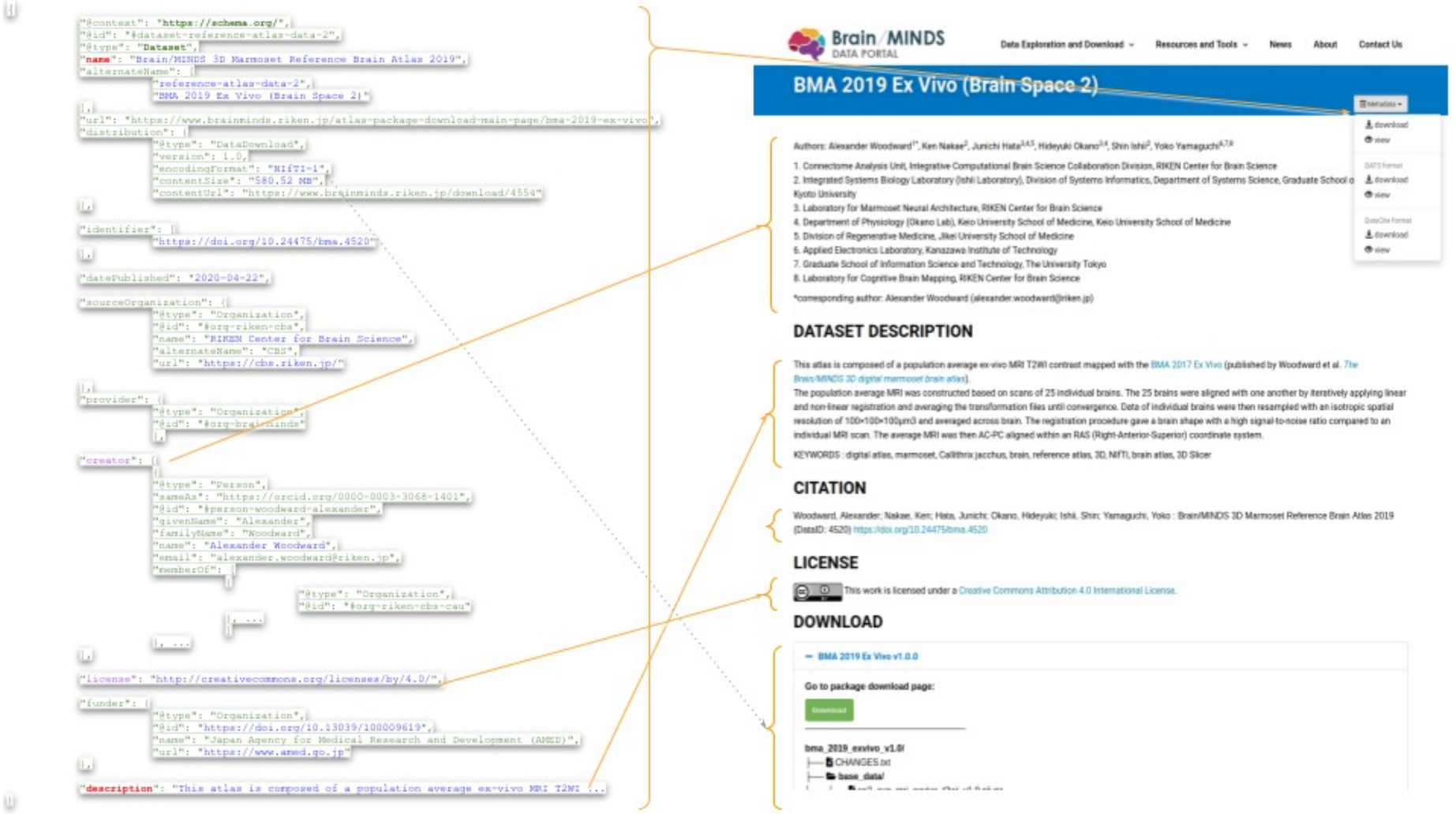
Brain/MINDS DATA PORTAL

schema.org Dataset metadata for Brain/MINDS 3D Marmoset Reference Brain Atlas

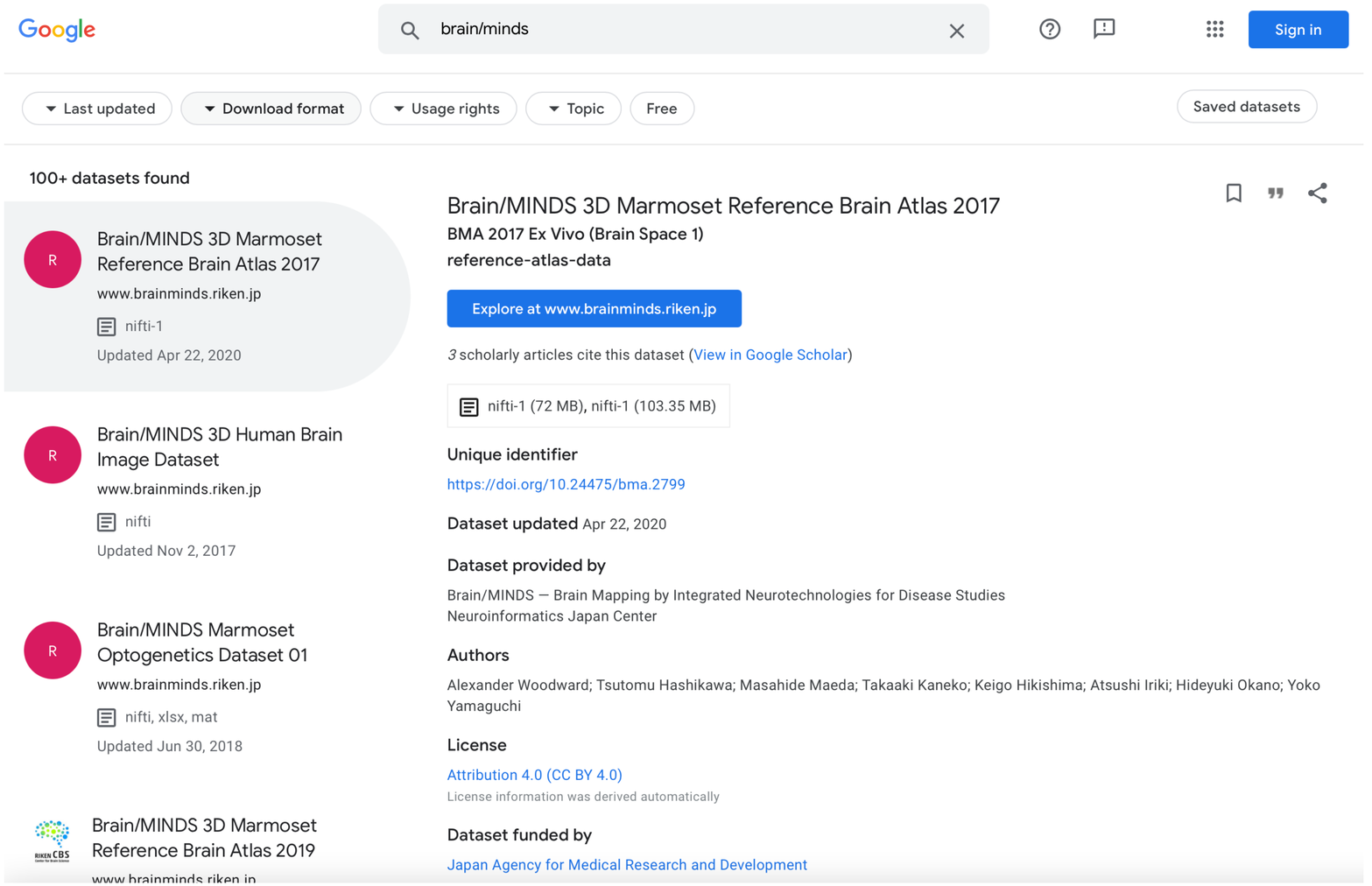


Brain/MINDS DATA PORTAL

Generated dataset landing page for Brain/MINDS 3D Marmoset Reference Brain Atlas



Google Dataset Search



https://developers.google.com/search/docs/data-types/dataset

Currently only supports whole datasets, not granular individual data entities.

Limitations of Google Dataset Search:

- No guarantee that your datasets will be indexed or searchable
- No application programming interface to build other search engines
- No support for neuroscience specific metadata
- Limited to entire datasets
  - not clear when it might support neuroscience record level data

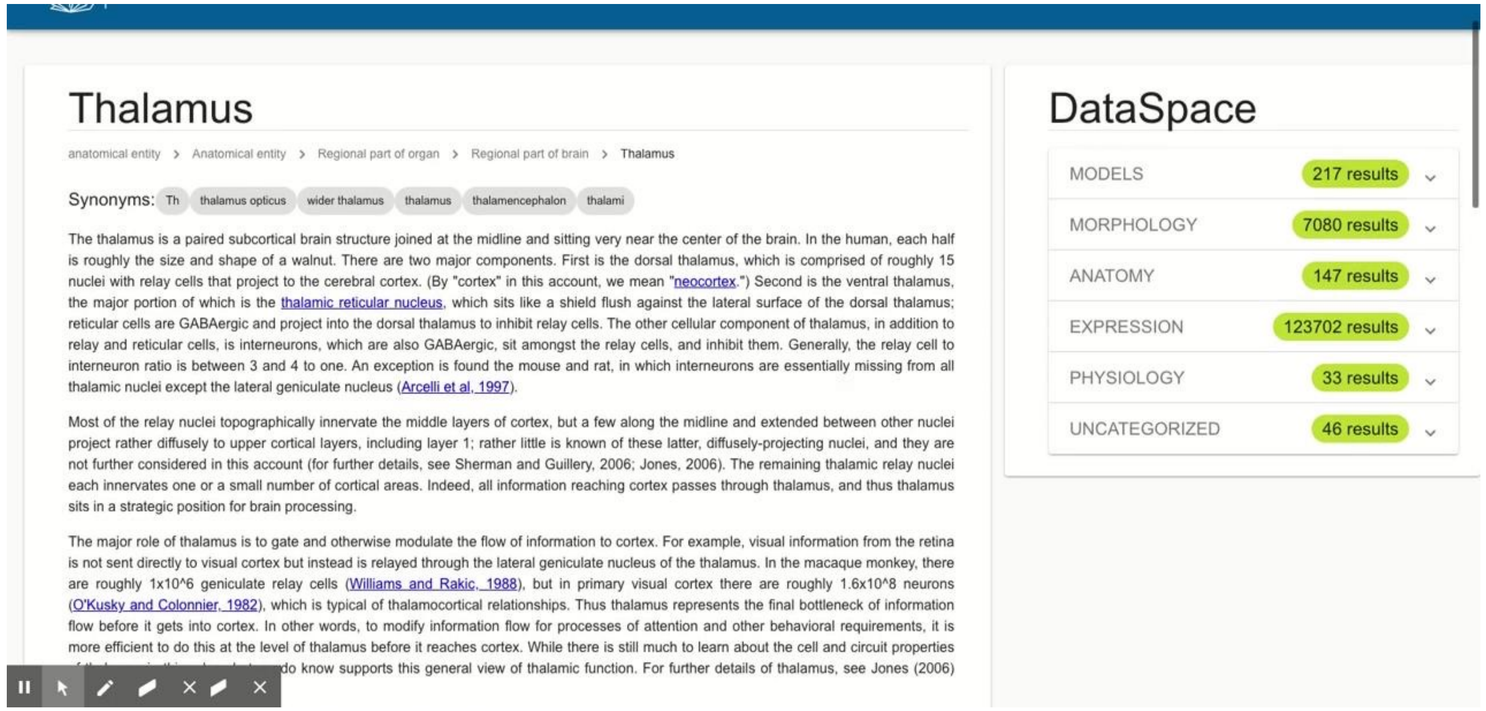
INCF KnowledgeSpace - knowledge-space.org

A community-based encyclopedia/data discoverability portal for neuroscience that links brain research concepts to the data, models, and literature that support them.

Core concepts:

- Provides a search engine for neuroscience resources including data, computational models and literature
- Supports encyclopedia articles with links to related resources
- Provides an API to support FAIR data

Features:



- Descriptions of neuroscience research concepts
- Links to ontologies that define concepts
- Links to PubMed entries associated with the concepts
- Links to data/models related to the concepts

Supported by:

Proof of Concept IBI Data Search:

KnowledgeSpace

Search…

RESOURCES ABOUT

brain MINDS

SEARCH

FILTERS

KEYWORDS

- [ ] Brain (7)
- [ ] Callithrix Jacchus (6)
- [ ] NIfTI (4)
- [ ] Marmoset (4)
- [ ] MRI (3)
- [ ] 3D (2)
- [ ] 3D Slicer (2)
- [ ] ECoG (2)
- [ ] Electrocorticography (2)
- [ ] Brain Atlas (2)

Brain/MINDS Results

7 records found

Title	Description
Brain/MINDS 3D Marmoset Reference Brain Atlas 2019	This atlas is composed of a population average ex-vivo MRI T2WI contrast mapped with the [BMA 2017 Ex Vivo] (https://www.brainminds.riken.jp/reference-atlas-data/) (published by Woodward et al. _[The Brain/MINDS 3D digital marmoset brain atlas](https://www.nature.com/articles/sdata20189)_). The population average MRI was constructed based on scans of 25 individual brains. The 25 brains were aligned with one another by iteratively applying linear and non-linear registration and averaging the transformation files until convergence. Data of individual brains were then resampled with an isotropic spatial resolution of 100×100×100μm3 and averaged across brain. The registration procedure gave a brain shape with a high signal-to-noise ratio compared to an individual MRI scan. The average MRI was then AC-PC aligned within an RAS (Right-Anterior-Superior) coordinate system.
Brain/MINDS 3D Marmoset Reference Brain Atlas 2017	The dataset includes NIfTI files of MRI T2 ex-vivo data; reconstructed Nissl stained images of the same brain, registered to the shape of the MRI; brain region segmentation (with separate color lookup table); and gray, mid-cortical and white matter boundary segmentation. In addition, a 3D Slicer scene file is provided that can be used for testing the dataset within the freely downloadable 3D Slicer software (https://www.slicer.org/). The scene file can be dragged directly into 3D Slicer and the atlas can be used immediately. Files can be downloaded individually or as one zip file. </br>The atlas can be viewed online via the Zooming Atlas Viewer (ZAV) by clicking [here] (https://www.bminds.brain.riken.jp/ZAViewer_Reference_Brain/).
Developmental Age-Specific Brain Templates Dataset	The dataset includes NifTI files of MRI T1-weighted images data and T2-weighted images at the age of 1 month, 3 months, 6 months, 12 months, 18 months, and 24 months. The templates at the age of 18 and 24 months were registered to the stereotaxic coordinates defined in Paxinos[^1] and Woodward et al[^2]. For the template at the age of 1, 3, 6, 12 months, we newly defined the stereotaxic coordinates. For details regarding the procedure, please download [Additional_information_about_the_dataset.pdf](https://www.brainminds.riken.jp/download/3815) [^1]: Paxinos G. The marmoset brain in stereotaxic coordinates. Academic Press. 2012. [^2]: Woodward A, Hashikawa T, Maeda M, Kaneko T, Hikishima K, Iriki A, Okano H, Yamaguchi Y. The Brain / MINDS 3D digital marmoset brain atlas. DOI: https://doi.org/10.24475/bma.2799
Brain/MINDS 3D Human Brain Image Dataset	Human brain images obtained with 3T MRI. The dataset includes T1-weighted images of patients with schizophrenia, those with major depressive disorder, and those with bipolar disorder, as well as of healthy controls. This dataset will contribute to promoting research on brain mapping in human psychiatric disorders. See detailed dataset description at https://www.brainminds.riken.jp/human-brain-images-about
Brain/MINDS Marmoset Optogenetics Dataset 01	Eight weeks after the virus injections, we applied 3.5 V to each LED for 200 ms to monitor the development of neural responses to photostimuli. In each session, eight LEDs were pseudo-randomly illuminated and 50 stimulation trials were performed for every LED. Inter stimulus intervals were fixed at 2 s. ECoG data were sampled at 1KHz. Further information can be found on [TychoWiki](http://wiki.neurotycho.org/Marmoset_Optogenetics_Details)
Brain/MINDS	We used auditory stimuli of different durations (AD) and frequencies (AF). In AD, 10 types of pure sinusoidal tones (1 ms rise/fall) with different durations (10, 25, 50, 75, 100, 125, 150, 175, 200, and 225 ms; 1000 Hz; 2000 stimuli in total)

KnowledgeSpace integrates IBI Data Sources:

Next steps:

- **Prepare an inventory** of available data repositories in the brain initiatives, metadata standards, and governance models for data access;
- **Establish a sustainable governance process** for developing, approving, and revising a standard metadata model(s) for describing neuroscience datasets;
- **Develop tools, best practices, and guidelines** for publishing datasets using the standard metadata model;
- **Configure and operate a core infrastructure and portal** for ingesting, indexing, and searching dataset metadata;
- **Define and implement interoperability** with existing data portals and infrastructures, and interoperability guidelines for new repositories;
- **Develop and disseminate tutorials**, training materials, and educational activities for dataset publishing;
- **Develop and enact a global communications strategy** to internal stakeholders, funding organizations, and neuroscience researcher communities regarding the activities and products of this project.