

Establishing and Evaluating Access to Scientific
Data

EUDAT Human Brain Project
Big Databases and Cloud Services
Project Report

Tom Wiesing

May 27, 2016

1

EdN:1

¹EDNOTE: Make an abstract

Contents

1	Introduction	3
1.1	Overview	3
1.2	Collaborators	3
1.3	Project Components	3
2	Human Brain Scans: The Dataset	4
3	Data Ingestion	4
4	Querying the Dataset	4
5	Outlook & Conclusion	4

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

1 Introduction

1.1 Overview

Multi-dimensional arrays arise naturally in multiple occasions and thus the paradigm of Array Databases has become more and more important. They can be used for a variety of scientific applications, ranging from satellite imagery, over medical imaging techniques to mathematically interesting objects.

One such array database is the Rasdaman system [Bau+98]. It promises to allow users to “storing and querying massive multi-dimensional arrays, such as sensor, image, simulation, and statistics data appearing in domains like earth, space, and life science” [Gmb16].

In this project we want to establish and evaluate access to scientific data. In particular we want to insert this scientific data into the Rasdaman database and evaluate how well the database can handle this kind of data. In this case we want to work together with EUDAT and the Human Brain Project.

1.2 Collaborators

The Human Brain Project [Pro16] is a European Commission Future and Emerging Technologies Flagship Project with the aim of “providing research infrastructure in the fields of Neuroscience, computing and brain-related medicine”. In particular we collaborated with Huanxiang Lu and Dr. Sean Hill from the École polytechnique fédérale de Lausanne to gain access to the Human Brain Database. This archive of multiple sources is a database of (not only) human brain scans.

We also collaborated with Peter Wittenburg and Daan Broede from EUDAT [EUD16]. EUDAT is a “collaborative Pan-European infrastructure for research data services, training and consultancy” and will be used to host the data and resulting interfaces created in this project.

1.3 Project Components

Concretely this project consists of (1) gaining access to the scientific data provided by the Human Brain Database, (2) determining how (and if) this data can be represented inside the Rasdaman system, (3) developing a method to properly ingest the data into Rasdaman, (4) asking the collaborators about useful queries that can be performed on the data, (5) running the queries and gaining new insight into the data and finally (6) evaluating how well Rasdaman was able to deal with the provided data and developed queries.

The section of this report is as follows: In Section 2 we introduce in detail the provided dataset. We then continue in Section 3 by describing how we ingested the data into Rasdaman. Next we describe the queries that we developed over the course of this project in Section 4 before concluding with an evaluation and outlook for future work in Section 5.

2 Human Brain Scans: The Dataset

2 EdN:2

3 Data Ingestion

3 EdN:3

4 Querying the Dataset

4 EdN:4

5 Outlook & Conclusion

5 EdN:5

References

- [Bau+98] Peter Baumann et al. “The Multidimensional Database System Ras-DaMan”. In: *Proceedings ACM SIGMOD’98*. Seattle, Washington, USA, June 1998.
- [EUD16] EUDAT. *EUDAT - Research Data Services, Expertise & Technology Solutions*. 2016. URL: <https://www.eudat.eu/> (visited on 05/27/2016).
- [Gmb16] rasdaman GmbH. *the rasdaman raster array database – rasdaman*. 2016. URL: <http://www.rasdaman.org/> (visited on 05/27/2016).
- [Pro16] Human Brain Project. *The Human Brain Project*. 2016. URL: <https://www.humanbrainproject.eu/> (visited on 05/27/2016).

²EdNOTE: Write this

³EdNOTE: Write this

⁴EdNOTE: Write this

⁵EdNOTE: Write this