Managing data with key-value stores

Chuck Ha, Holly Miller, Catherine N. Norton
MBLWHOI Library, MBL

Introduction
As more data becomes readily available, data managers and scientists need tools to deal with the different types of data. Some data doesn’t easily fit into traditional relational databases without complex database design. We saw this complication while developing MouseDB and decided to use a key-value store.

Our Usage
The first step of the MouseDB project was to gather data. But where were we going to store this data? We didn’t even know exactly what data we wanted. We didn’t want to spend time engineering a database because we knew the requirements were going to change and thus the schema would change as well. We wanted something simple that would allow us to easily access the data we needed. We settled on an Apache project, CouchDB that fit our needs perfectly. Instead of querying relational tables, we simply ask for a document and get the data in JSON (JavaScript Object Notation) format.

Key-value stores are:
• Not a silver bullet
• Useful when the data is flat and has few relations
• Easy to access programmatically
• Popular enough to have libraries in most languages
• Good when the schema of the data you need is unknown

What worked for us
• Changing requirements meant not having a schema saved us lots of time
• Open source plugins extended CouchDB to allow us to use it as a database for our web application
• CouchDB was Easy to setup
• Simple database replication

MouseDB
The MouseDB project is all about taking the vast amount of mouse strain data available, making a subset of the data based on its research applications and making the data searchable.

Funded through a gift from the George F Jewett Foundation