

應用於精準醫療的 大數據平台研究

資工數位4A 102013025 江泓昇

資工數位4A 102013042 吳嘉恩

指導老師：朱學亭



什麼是Big Data（大數據、巨量資料）？

Can't find your data? [Click here for more information.](#) Dismiss ✕

NIH NATIONAL CANCER INSTITUTE GDC Data Portal Home Projects Data Analysis Quick Search Login Cart 0 GDC Apps

Harmonized Cancer Datasets Genomic Data Commons Data Portal

Get Started by Exploring

Projects

Data

Perform Advanced Search Queries, such as:

Cases of kidney cancer diagnosed at the age of 20 and below	182 Cases	1,514 Files
CNV data of female brain cancer cases	459 Cases	1,788 Files
Gene expression quantification data in TCGA-GBM project	166 Cases	522 Files

Cases by Primary Site

Primary Site	Approximate Number of Cases
Kidney	1,500
Brain	1,100
Nervous System	1,100
Breast	1,100
Lung	1,100
Blood	1,000
Colorectal	900
Uterus	600
Ovary	600
Head and Neck	500
Thyroid	500
Prostate	500
Stomach	500
Skin	400
Bladder	400
Bone	400
Liver	300
Cervix	300
Adrenal Gland	200
Soft Tissue	200
Bone Marrow	200
Pancreas	200
Esophagus	200
Testis	100
Thymus	100
Pleura	100
Eye	100
Lymph Node	100
Bile Duct	100

DATA PORTAL SUMMARY
Data Release 4.0 - October 31, 2016

PROJECTS
 39

PRIMARY SITE
 29

CASES
 14,531

FILES
 274,821

Infrastructure

Data is continuously being processed and harmonized by the GDC. View GDC system statistics.

Compute Infrastructure	12,800 Cores	87.96 TB RAM
Storage Infrastructure	4.98 PB Used	5.42 PB Total

[View Data Download Statistics Report »](#)

Documentation

Learn how to use the GDC Data Portal to its full potential with common topics such as:

- [Browse Data using Facet Search](#)
- [Search Data with Advanced Search Technology](#)
- [Project Based Data Availability](#)
- [Controlled Access Data](#)
- [Visit the Documentation Website »](#)

GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

Data Portal

Website

Data Transfer Tool

API

Data Submission Portal

Documentation

Legacy Archive

GDC cBio Portal

[Site Home](#) | [Policies](#) | [Accessibility](#) | [FOIA](#)
 U.S. Department of Health and Human Services | National Institutes of Health | National Cancer Institute | USA.gov
 NIH... Turning Discovery Into Health ®
 UI 1.4.1 @ 1416c89, API 1.5.0 @ 9f568ba

參考網站：<https://gdc-portal.nci.nih.gov/>

系統環境：



(1) MongoDB

(2) Express

(3) AngularJS

(4) NodeJS

資料庫間的比較：

Relation DataBase	MongoDB
資料庫 (Database)	DataBase
資料表 (Table)	Collection
資料 (Record / Row)	Document
欄位 (Column)	Field
主鍵 (Primary Key)	_id

資料庫指令示意圖：

Collection

↓

```
db.users.insert(
```

Document

↓

```
{
  name: "sue",
  age: 26,
  status: "A",
  groups: [ "news", "sports" ]
})
```

Document

```
{
  name: "sue",
  age: 26,
  status: "A",
  groups: [ "news", "sports" ]
}
```

insert →

Collection

{ name: "al", age: 18, ... }
{ name: "lee", age: 28, ... }
{ name: "jan", age: 21, ... }
{ name: "kai", age: 38, ... }
{ name: "sam", age: 18, ... }
{ name: "mel", age: 38, ... }
{ name: "ryan", age: 31, ... }
{ name: "sue", age: 26, ... }

users

系統前端：

HTML



CSS



JS

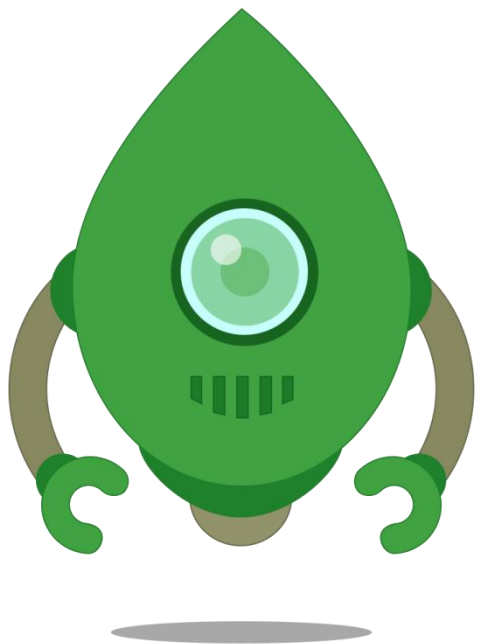


Data-Driven Documents

程式語言：



管理工具：



Robomongo

伺服器管理



Atom

網頁撰寫

Robomongo 0.9.0

File View Options Window Help

New Connection (2)

- System
- TCGA
 - Collections (1)
 - Dali
 - Functions
 - Users

```
db.getCollection('Dali').find({})
```

New Connection localhost:27017 TCGA

```
db.getCollection('Dali').find({})
```

Dali 0.077 sec.

Key	Value	Type
(1) ObjectId("583c07aedfa94afa9672c0ff")	{ 9 fields }	Object
_id	ObjectId("583c07aedfa94afa9672c0ff")	ObjectId
SiteId	15	Int32
SiteName	大里	String
ItemId	36	Int32
ItemName	二氧化碳	String
ItemEngName	CO2	String
ItemUnit	ppm	String
MonitorDate	2016-07-07-12	String
Concentration	434.33	Double
(2) ObjectId("583c07aedfa94afa9672c100")	{ 9 fields }	Object
(3) ObjectId("583c07aedfa94afa9672c101")	{ 9 fields }	Object
(4) ObjectId("583c07aedfa94afa9672c102")	{ 9 fields }	Object
(5) ObjectId("583c07aedfa94afa9672c103")	{ 9 fields }	Object

Robomongo – MongoDB的Schema

開始 Demo

謝謝聆聽