

SVAI - NF Hackathon

September 2019



Databiology Company Overview

- Global software and services company
- Providing specialist solutions for global life sciences sector
- Business hubs in Oxford, San Francisco, Hong Kong and Mexico
- Databiology's mission is to be the enterprise platform for globally distributed biomedical data, applications and infrastructure



Our Solution: Biomedical Information Management & Process Orchestration



ANY Biomedical Data



ANY Application



databiology

Configure | Command | Collaborate



ANY Compute Platform

Objectives

- Demonstrate data and metadata import and management
- Demonstrate FAIR principles applied in Data and Analysis (APPs)


Main participant


- Juan Caballero, CSO (remotely from Mexico)



Data import

- After registration in Synapse, data was imported into our public instance
<https://www.lab.databiology.net/>
into a private project
- Access is restricted and data is secured all the time

 Projects Subjects Episodes Samples Extracts Workunits Datasets Admin

Project 255  running NF Hackathon 2019 - Cutaneous Neurofibroma Dataset

Project 255 - NF Hackathon 2019 - Cutaneous Neurofibroma Dataset

Details

Members 2

Roles 2

Comments 0

Subjects 16

Episodes 0

Samples 86

Extracts 86

Workunits 13

Resources 142

Datasets 0

Charges 0

Bookings 0

Reviews 2

Mails

Log

Tree

Terms and Conditions of Usage

Id 255

Name NF Hackathon 2019 - Cutaneous Neurofibroma Dataset

Status running

Storage Size 102.829GB / 38.530KB / 0

Organization Databiology

Department/Faculty Databiology

Institute/Group Databiology

Approver Databiology Support

Leader Juan Caballero

Contact Juan Caballero

Applicant Juan Caballero

Technologies General, Genomics, Proteomics, Sequencing, Transcriptomics

Support Mode Userlab

Summary This project includes high-throughput data from 11 different patient tumors and manuscript describing this resource can be found here
<https://www.nature.com/articles/sdata201745>

Description sdata201745.pdf

Start Date 2019-09-13

End Date 2022-09-13

Disease neurofibroma

First Extension Report

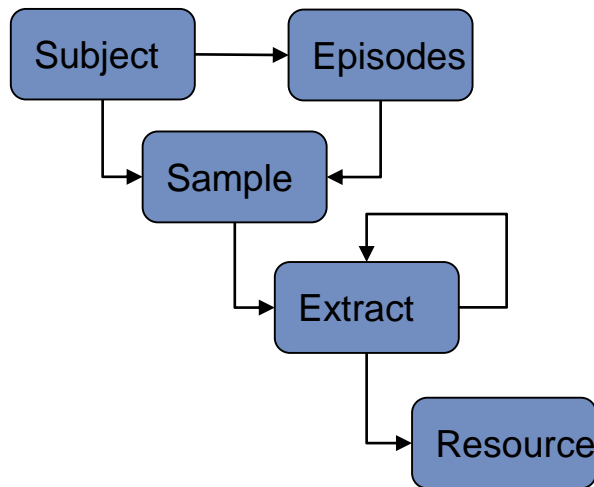
Second Extension Report

Third Final Report

Created By Juan Caballero (jcaballero) 2019-09-13 18:49

Data and metadata

- Data was modeled in our data schema



- The data model is flexible and FAIR-complete
- Data is indexed after import, enabling search immediately

databiology Projects Subjects Episodes Samples Extracts Workunit

Project 255 running NF Hackathon 2019 - Cutaneous Neurofibroma Dataset

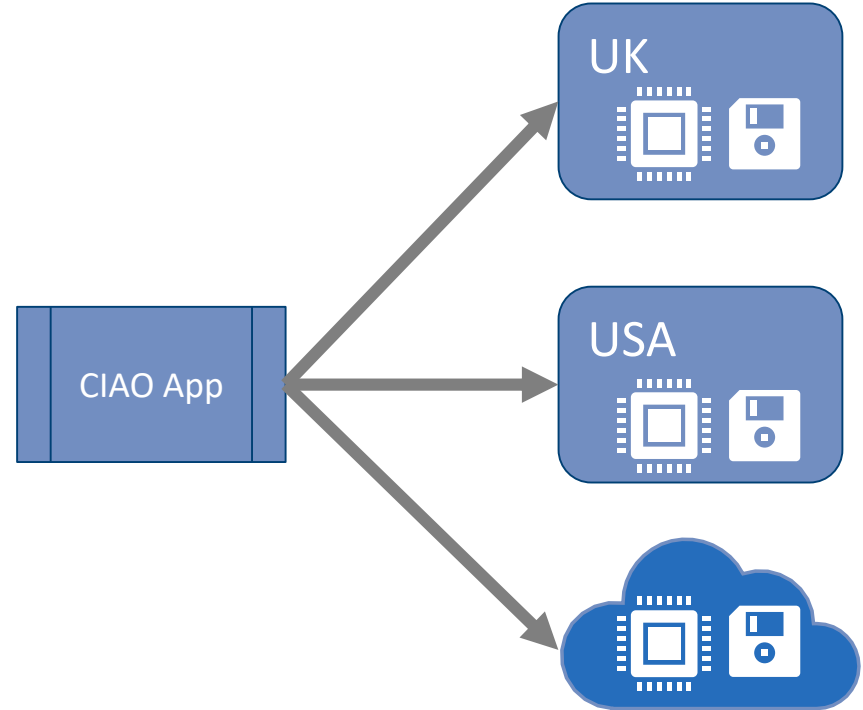
Subject : 216 - CT000001

Details		Id 216
Episodes	0	Individual Id CT000001
Samples	7	Type cnf_patient
Extracts	7	Project 255
Resources	6	Age of NF1 Diagnosis Older than 20 years old
		Age of Neurofibromas Older than 20 years old
		Armpit Freckles TRUE
		Birthdate 1965-07-03
		Bone Fractures FALSE
		Breast Cancer FALSE
		Cafe au Lait Spots Yes - more than 6, less than 20
		Concomitant Medications paroxetine
		Date First Period 1980
		Date Last Period 2013-07-16
		Date of Surgery 2013-07-16
		DateTime of Informed Consent 2013-07-15
		During Pregnancy FALSE
		Ethnicity Not Hispanic or Latino
		Family NF1 TRUE
		Female Childbearing Age Yes, female of child bearing age with NF1
		Gender Female
		Genetic Test Confirmation NF1 TRUE

databiology

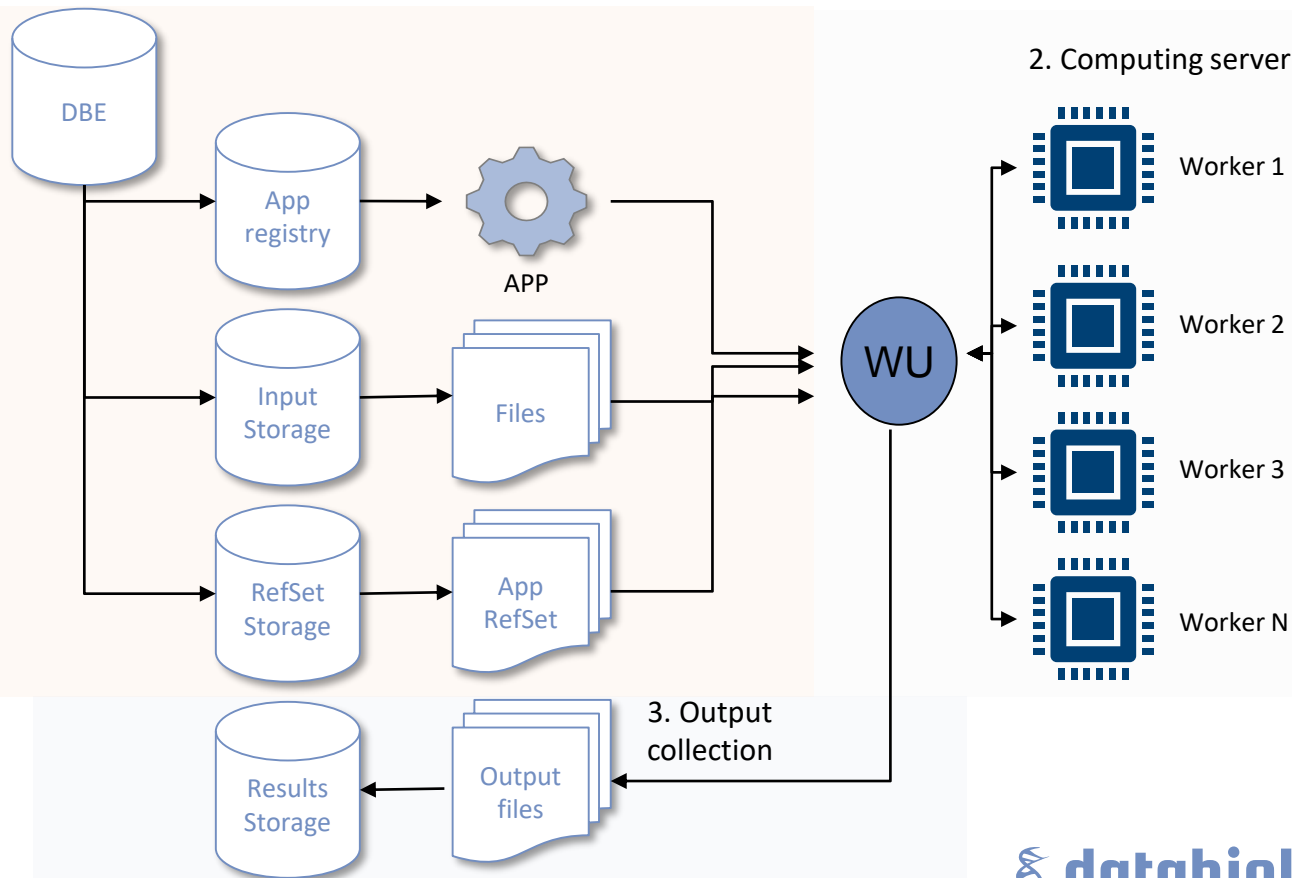
Real portable apps

- Analysis are performed with our CIAO apps (based on Docker)
- Any developer can create Apps with minimal requirements
- Apps are portable (can be run in diverse infrastructures)
- Apps are also FAIR-complete

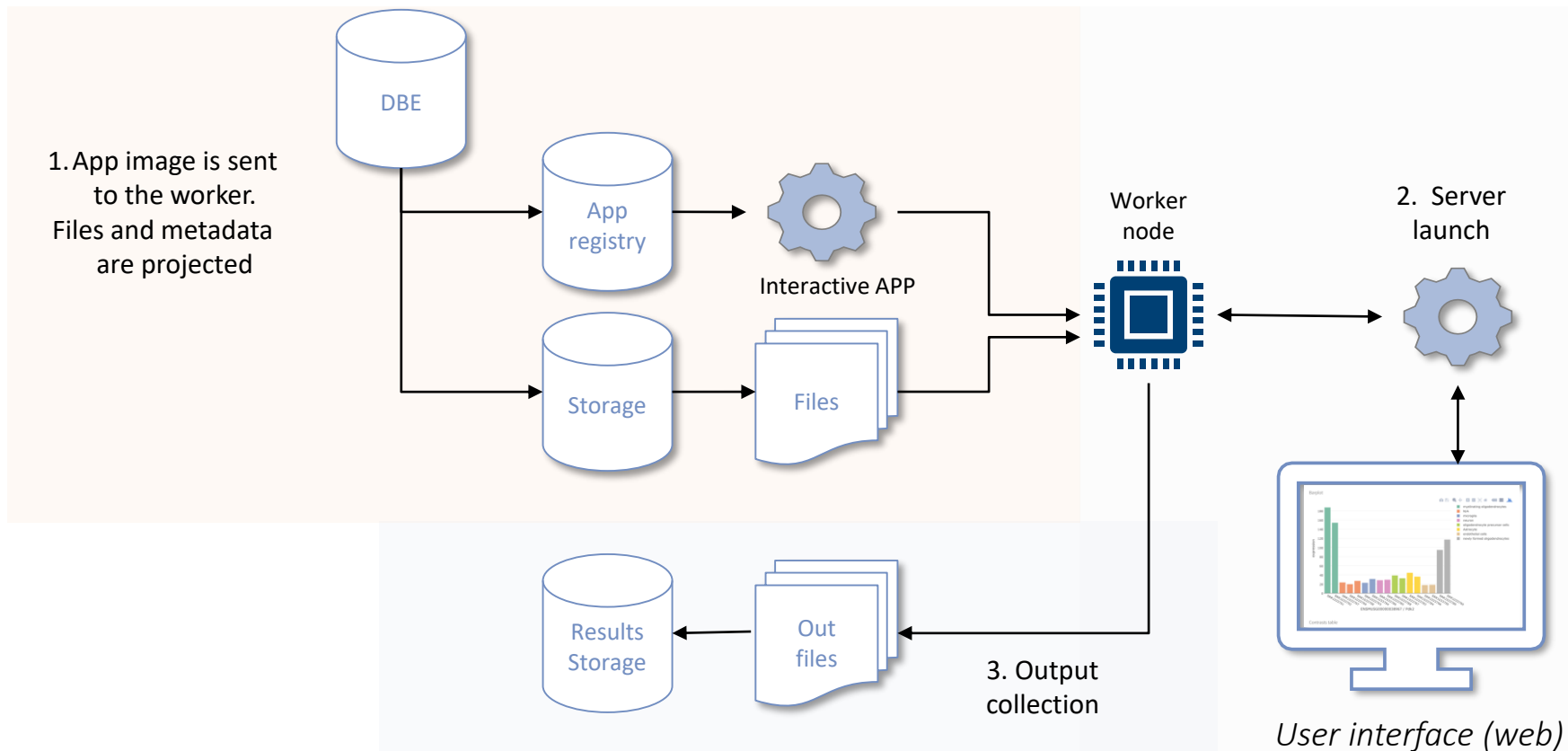


Workunits

1. App are run as Workunits, the container image is sent to the workers. Files and metadata are projected configuring the app automatically



Interactive App



WGS data analysis

- VCFs were analyzed with Exomiser, prioritizing genes related to HP:0001067 (Neurofibromas)
- All analysis were performed in a Workunit, keeping record of data inputs, parameters, app version, infrastructure, etc., making the analysis FAIR and reproducible.

The screenshot displays the databiology web application interface. At the top, a navigation bar includes links for Projects, Subjects, Episodes, Samples, Extracts, Workunits, Datasets, and Admin. Below this, a header section shows 'Project 255' and a status indicator 'running' for the 'NF Hackathon 2019 - Cutaneous Neurofibroma Dataset'. The main content area is titled 'Workunit : 7419 - Exomiser'. On the left, a sidebar lists various details: Parameters (41), Comments (0), Resources (3), Input Resources (1), Executables, Log, and Data provenance. The right side of the page provides detailed information about the workunit, including its ID (7419), Name (Exomiser), Status (available), Size (516.655MB), Storage (DBIO Result Storage), End Date (2019-09-15 21:04), Application (Exomiser), Submitter (DBIO Static Submitter), and Input Resource (SL107656.vcf attached). It also shows the Project (255), Created By (Juan Caballero), and Modified By (Juan Caballero). At the bottom, there are buttons for 'Rerun Analysis', 'Edit', 'Delete', 'Mark Deleted', and 'Export archive'.

Details	
Parameters	41
Comments	0
Resources	3
Input Resources	1
Executables	
Log	
Data provenance	

Id	7419
Name	Exomiser
Status	available
Size	516.655MB
Storage	DBIO Result Storage
End Date	2019-09-15 21:04
Application	Exomiser
Submitter	DBIO Static Submitter
Input Resource	SL107656.vcf attached
Project	255
Created By	Juan Caballero (jcaballero) 2019-09-15 20:54
Modified By	Juan Caballero (jcaballero) 2019-09-15 21:04

[Rerun Analysis](#) [Edit](#) [Delete](#) [Mark Deleted](#) [Export archive](#)

Example output

Prioritised Genes

NF1

Exomiser Score: **0.998**

Phenotype Score: **1.000**

Variant Score: **1.000**

Phenotype matches:

Phenotypic similarity 0.890 to [Watson syndrome](#) associated with NF1.

Best Phenotype Matches:

HP:0001067, Neurofibromas - HP:0001067, Neurofibromas

Phenotypic similarity 1.000 to mouse mutant involving [NF1](#).

Best Phenotype Matches:

HP:0001067, Neurofibromas - MP:0010314, increased neurofibroma incidence

Proximity score 0.508 in [interactome](#) to [NF2](#) and phenotypic similarity 0.883 to [Neurofibromatosis type 2](#) associated with NF2.

Best Phenotype Matches:

HP:0001067, Neurofibromas - HP:0009595, Occasional neurofibromas

Proximity score 0.508 in [interactome](#) to [NF2](#) and phenotypic similarity 0.810 to mouse mutant of [NF2](#).

Best Phenotype Matches:

HP:0001067, Neurofibromas - MP:0010314, increased neurofibroma incidence

Known diseases:

OMIM:162200 Neurofibromatosis, type 1 - autosomal dominant

OMIM:162210 Neurofibromatosis, familial spinal - autosomal dominant

OMIM:193520 Watson syndrome - autosomal dominant

OMIM:601321 Neurofibromatosis-Noonan syndrome - autosomal dominant

OMIM:607785 Leukemia, juvenile myelomonocytic - autosomal dominant

ORPHA:139474 17q11.2 microduplication syndrome

ORPHA:638 Neurofibromatosis-Noonan syndrome

Gene scores under compatible inheritance modes:

AUTOSOMAL_DOMINANT

Exomiser Score: **0.998**

Phenotype Score: **1.000**

Variant Score: **1.000**

Variants contributing to score:

STOPGAIN chr17:g.29588751C>T [0/1] rs760703505 (variation viewer)

Variant score: 1.000 **CONTRIBUTING VARIANT**

Transcripts:

Pathogenicity Data:

No pathogenicity data

Frequency Data:

No frequency data

Metadata is projected

- Besides data transfer, metadata is also projected into the running container, facilitating complex analysis

```
File Edit View Language
1 [
2   {
3     "status": "available",
4     "junk": false,
5     "name": "SL102344.vcf",
6     "created": "2019-09-14 01:09:29.000310",
7     "relativepath": "p255/General/SynapseImporter/wu7408/j21617/SL102344.vcf",
8     "description": null,
9     "size": 646842638,
10    "storage": "8",
11    "modified": "2019-09-14 20:01:54.000918",
12    "id": 1610332,
13    "workunit": "7408",
14    "project": "255",
15    "succeedingworkunit": [
16      "7421",
17      "7420",
18      "7418"
19    ],
20    "modifiedby": "jcaballero",
21    "createdby": "jcaballero",
22    "remoteiri": null,
23    "extract": {
24      "project": "255",
25      "arraydesignname": "",
26      "resource": [
27        "1610332"
28      ],
29      "name": "SL102344",
30      "created": "2019-09-14 20:00:04.000565",
31      "modified": "2019-09-14 20:00:04.000673",
32      "sample": {
33        "origin": "",
34        "initialtimepoint": "",
35        "sex": "",
36        "strain": "",
37        "modifiedby": "jcaballero",
38        "id": 79268,
39        "extract": [
40          "817808"
41        ],
42        "diseasestate": "",
43        "subject": {
44          "description": null,
45          "created": "2019-09-14 19:43:41.000569",
46          "modified": "2019-09-14 19:43:41.000569",
47          "individualid": "CT000005",
```



Contact us or follow us online!

Databiology Ltd.
Magdalen Centre
The Oxford Science Park
Oxford, OX4 4GA
United Kingdom

+44-1865-784426
contactus@databiology.com

Databiology Inc.
201 Spear Street, Suite 1100
San Francisco, CA 94105
USA

+1-415-426-3592
contactus@databiology.com

Databiology Hong Kong Ltd.
Unit E, 6/F Golden Sun Centre
59-67 Bonham Street West
Sheung Wan, Hong Kong
Hong Kong (SAR)

+852-8193-4005
contactus@databiology.com

