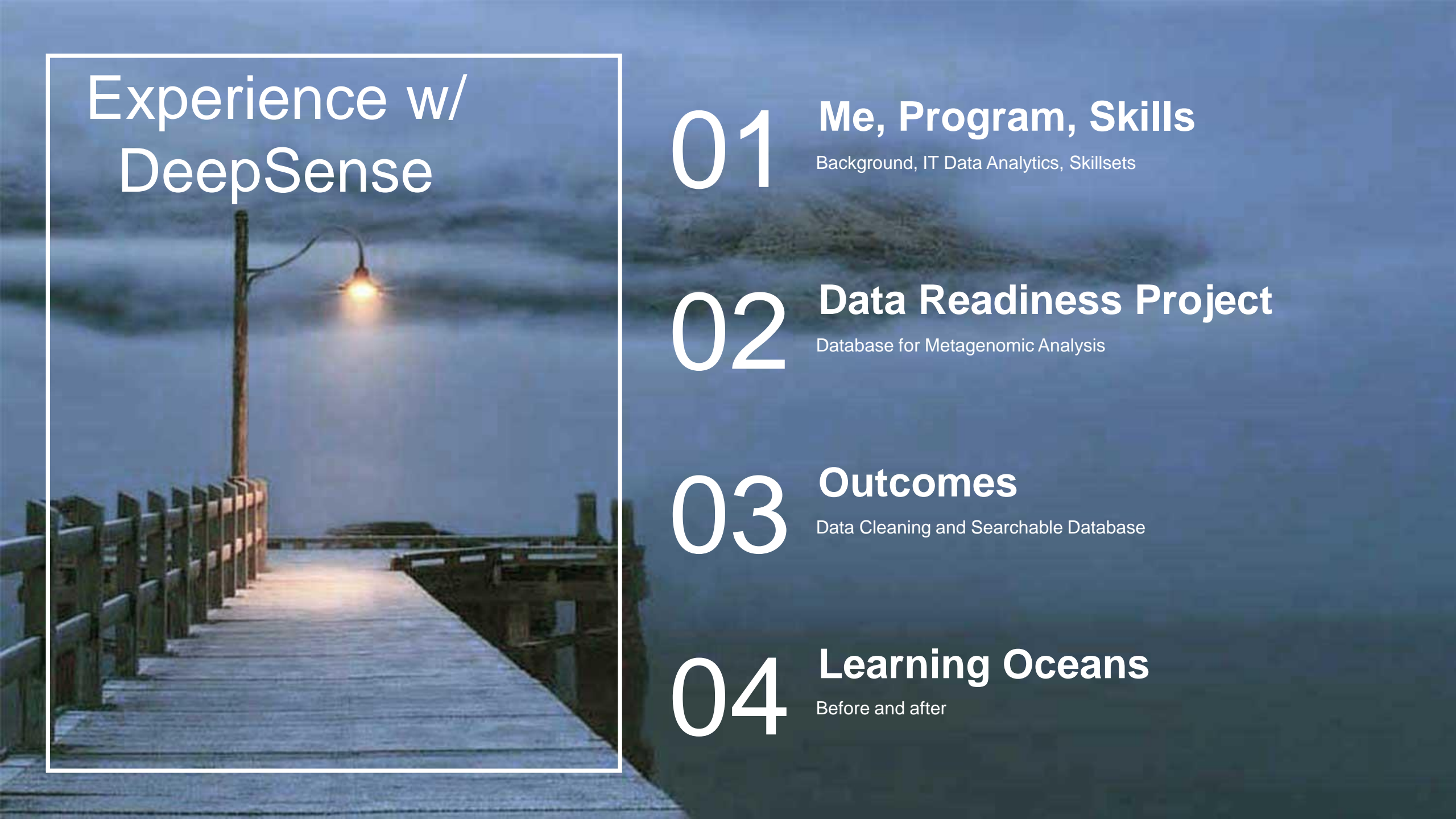


# Data Readiness Project

Cheolsoo Im,  
DeepSense Alumni,  
NSCC Data Analytics



# Experience w/ DeepSense

01

## Me, Program, Skills

Background, IT Data Analytics, Skillsets

02

## Data Readiness Project

Database for Metagenomic Analysis

03

## Outcomes

Data Cleaning and Searchable Database

04

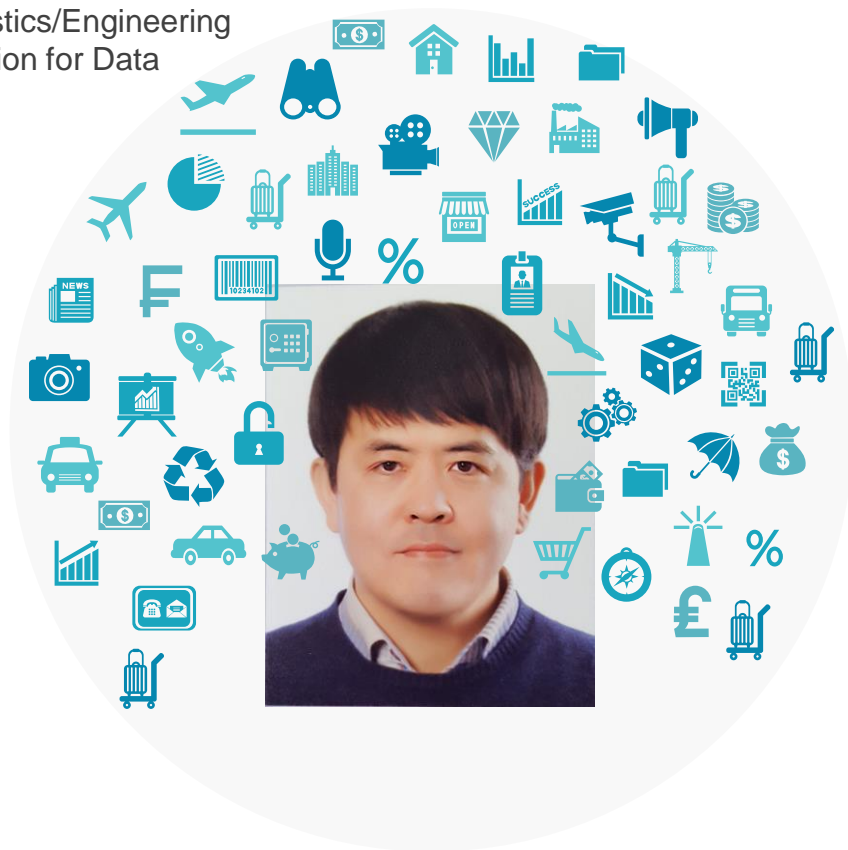
## Learning Oceans

Before and after

# Me, Program, Skills

## Background

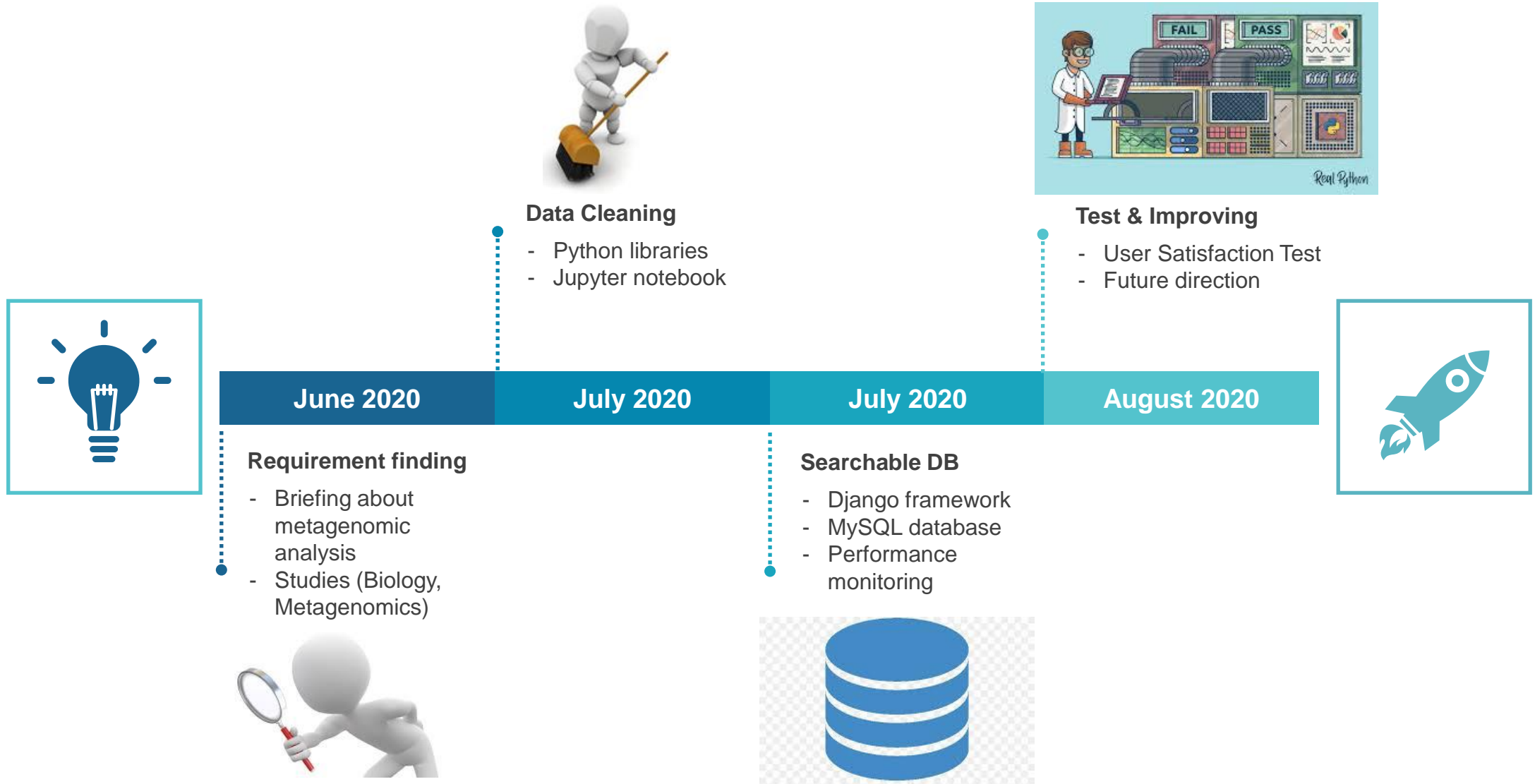
Statistics/Engineering  
Passion for Data



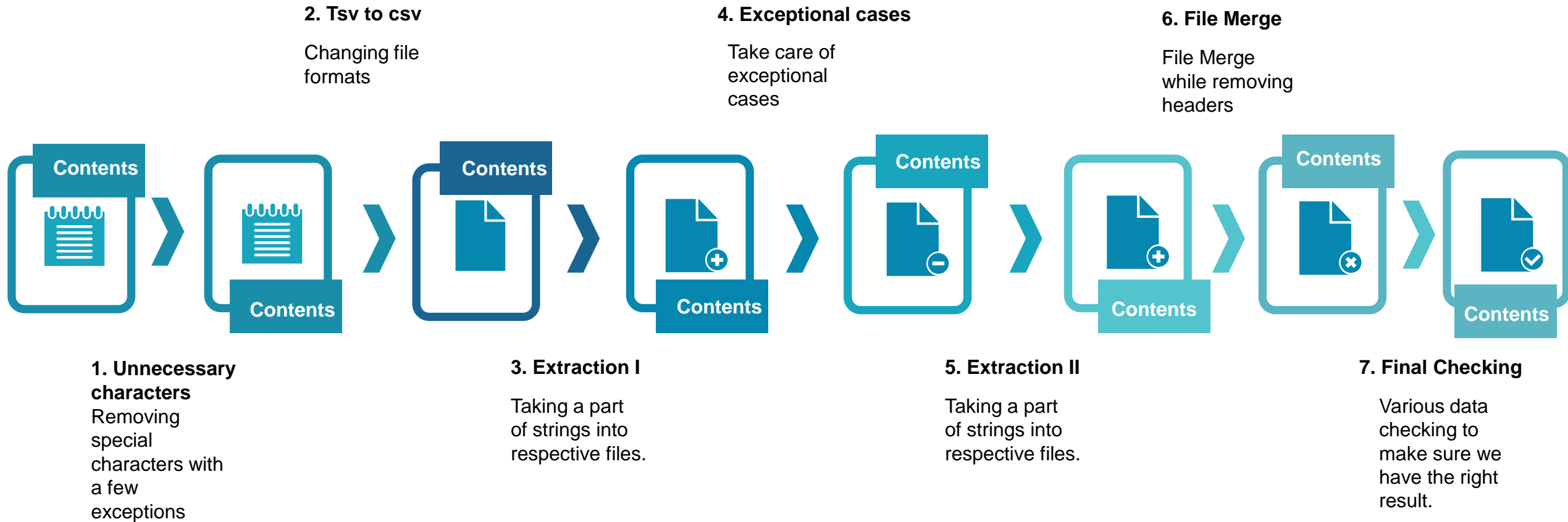
NSCC IT Data Analytics



# Project milestones



# Data Cleaning







# Final Outcome

Searchable Database

# Metagenomics Dashboard

Select sense full4 to change | Dja x +

127.0.0.1:8000/main/Sense/sensefull4/

Apps Search - IBM Traini... IBM Security Learni... Institute of Technol... Master the Mainfra... IBM Z - Group hom... My Enrollments Preferences Search - IBM Traini... Other bookmarks

DeepSense  
Metagenomics  
Dashboard

Welcome, coopstudent1. [View site](#) / [Change password](#) / [Log out](#)

ADMIN

Authentication and AuthorizationSense

Sense full4s

Search 1755792 results (Show all)

Import Export Add sense full4

	Gene callers id	Co assembly	Contig	COG FUNCTION
<input type="checkbox"/>	<a href="#">235311</a>	BB16	c_000000172261	DNA polymerase III alpha subunit!!!DNA polymerase III epsilon subunit or related 3 5 exonuclease
<input type="checkbox"/>	<a href="#">173021</a>	BB16	c_000000124941	PKD repeat!!!Sugar lactone lactonase YvrE
<input type="checkbox"/>	<a href="#">111831</a>	BB16	c_000000079224	Uncharacterized N terminal domain of the transcription elongation factor GreA
<input type="checkbox"/>	<a href="#">84474</a>	BB16	c_000000059410	DNA binding response regulator OmpR family contains REC and winged helix wHTH domain!!!K sensing histidine kinase KdpD!!!AraC type DNA binding domain and AraC containing proteins!!!Periplasmic ligand binding sensor domain!!!Periplasmic ligand binding sensor domain
<input type="checkbox"/>	<a href="#">298642</a>	BB16	c_000000220382	Glutamate 1 semialdehyde aminotransferase
<input type="checkbox"/>	<a href="#">298641</a>	BB16	c_000000220382	Predicted O methyltransferase YrrM
<input type="checkbox"/>	<a href="#">298640</a>	BB16	c_000000220382	
<input type="checkbox"/>	<a href="#">298639</a>	BB16	c_000000220382	
<input type="checkbox"/>	<a href="#">298638</a>	BB16	c_000000220382	Lysyl tRNA synthetase class II !!!Uncharacterized membrane protein
<input type="checkbox"/>	<a href="#">297760</a>	BB16	c_000000219717	Na proline symporter!!!Na proline symporter
<input type="checkbox"/>	<a href="#">297759</a>	BB16	c_000000219717	TPR repeat
<input type="checkbox"/>	<a href="#">297758</a>	BB16	c_000000219717	Ligand binding SRPBCC domain

By co assembly

All  
[AZCF](#)  
[AZCS](#)  
[AZOF](#)  
[AZOS](#)  
[BB16](#)

By direction

All  
[f](#)  
[r](#)

By gene\_callers\_id

By contig

Type here to search

12:04 PM  
2020-09-01

# Deeper look on individual records

Change sense full4 | Django site x +

127.0.0.1:8000/main/Sense/sensefull4/1755792/change/

Apps Search - IBM Traini... IBM Security Learni... Institute of Technol... Master the Mainfra... IBM Z - Group hom... My Enrollments Preferences Search - IBM Traini... Other bookmarks

Authentication and Authorization Sense

Sense full4s

Home > Sense > Sense full4s > c\_000000172261

Gene callers id: \* 235311

Contig: \* c\_000000172261

Co assembly: \* BB16

Start: \* 485

Stop: \* 4817

Direction: \* r

COG CATEGORY: \* L

COG CATEGORY ACCESSION: \* L

COG FUNCTION: \* DNA polymerase III alpha subunit!!!DNA polymerase III epsilon subunit or related 3 5 exonuclease

COG FUNCTION ACCESSION: \* COG0587!!!COG0847

Dna sequence: \* ATGTTTTAGTTTTGATACCGAAACCACTGGCTTGCCAAAACGGGGGATGCGCCTATTTTCAGA

Save

Save and add another

Save and continue editing

Delete

History

12:12 PM 2020-09-01



# User management

The screenshot displays the DjangoSuit web application's user management interface. At the top, there's a navigation bar with the DjangoSuit logo and a user profile section. Below this, a sidebar contains tabs for 'Authentication and Authorization' and 'Sense'. The main content area is titled 'Users' and features a search bar and a table of users. The table has columns for Username, Email address, First name, Last name, and Staff status. Two users are listed: 'coopstudent1' and 'testuser', both with a green checkmark in the Staff status column. To the right of the table, there are three filter panels: 'By staff status', 'By superuser status', and 'By active', each with 'All', 'Yes', and 'No' options. At the bottom, there's an 'Action:' dropdown menu and a 'Go' button. The footer includes the DjangoSuit logo, copyright information, and a search bar.

Username	Email address	First name	Last name	Staff status
<input type="checkbox"/> coopstudent1				Yes
<input type="checkbox"/> testuser				Yes

# Performance management

Select sense full4 to change | Django 3.0

127.0.0.1:8000/main/Sense/sensefull4/

Apps Search - IBM Traini... IBM Security Learni... Institute of Technol... Master the Mainfra... IBM Z - Group hom... My Enrollments Preferences Search - IBM Traini... Other bookmarks

## DeepSense Metagenomics Dashboard

Welcome, coopstudent1. [View site](#) / [Change password](#) / [Log out](#)

ADMIN

Authentication and Authorization **Sense**

Sense full4s

Search 1755792 results (Show all) Import Export

	Gene callers id	Co assembly	Contig	COG FUNCTION
<input type="checkbox"/>	235311	BB16	c_000000172261	DNA polymerase III alpha subunit!!!DNA polymerase III epsilon subunit or related 3 5 exonuclease
<input type="checkbox"/>	173021	BB16	c_000000124941	PKD repeat!!!Sugar lactone lactonase YvrE
<input type="checkbox"/>	111831	BB16	c_000000079224	Uncharacterized N terminal domain of the transcription elongation factor GreA
<input type="checkbox"/>	84474	BB16	c_000000059410	DNA binding response regulator OmpR family contains REC and winged helix wHTH domain!!!K sensing histidine kinase KdpD!!!AraC type DNA binding domain and AraC containing proteins!!!Periplasmic ligand binding sensor domain!!!Periplasmic ligand binding sensor domain
<input type="checkbox"/>	298642	BB16	c_000000220382	Glutamate 1 semialdehyde aminotransferase
<input type="checkbox"/>	298641	BB16	c_000000220382	Predicted O methyltransferase YrrM
<input type="checkbox"/>	298640	BB16	c_000000220382	
<input type="checkbox"/>	298639	BB16	c_000000220382	
<input type="checkbox"/>	298638	BB16	c_000000220382	Lysyl tRNA synthetase class II !!!Uncharacterized membrane protein
<input type="checkbox"/>	297760	BB16	c_000000219717	Na proline symporter!!!Na proline symporter
<input type="checkbox"/>	297759	BB16	c_000000219717	TPR repeat
<input type="checkbox"/>	297758	BB16	c_000000219717	Ligand binding SRPBCC domain

Hide »

Versions ☒ Log out

DJANGO 3.0

Time ☒

CPU: 2811.59ms  
(9363.07ms)

Settings ☒

Headers ☒

Request ☒

CHANGELIST\_VIEW

SQL ☒

7 QUERIES IN 7963.02ms

Static files ☒

13 FILES USED

Templates ☒

ADMIN/IMPORT\_EXPORT/CHANG

Cache ☒

0 CALLS IN 0.00ms

Signals ☒

7 RECEIVERS OF 12  
SIGNALS

Logging ☒

Type here to search

12:09 PM 2020-09-01

# Learning the potential

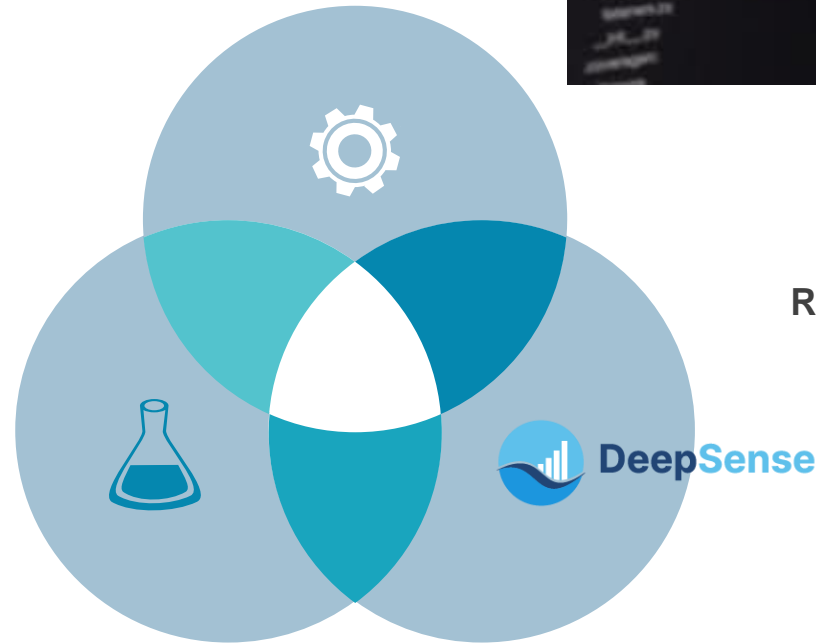
## Power of analytical tool

Python/ MySQL and more



## Big Data in Metagenomics Research

Scale and availability

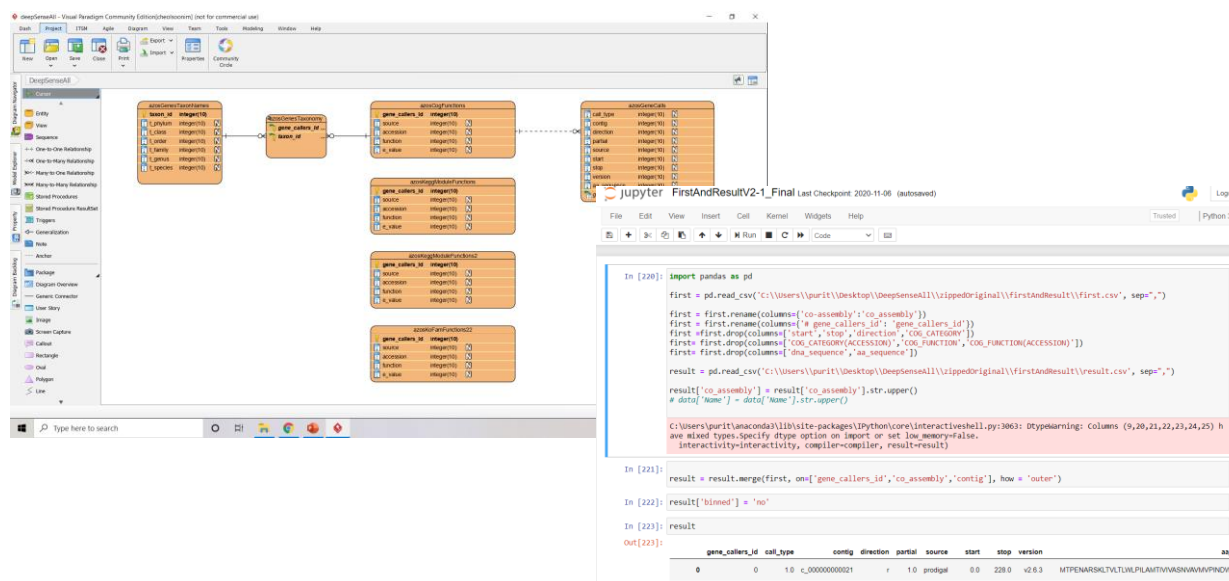


## Research

Dalhousie Univ, IBM  
COVE, Ocean Frontier Institute



# After Summer





THANK YOU