



# Bioinformatics Resource Portal

Nigel Alfino

Supervisor: Mr. Joseph Bonello, Co-Supervisor: Prof. Ernest Cachia  
Department of Computer Information Systems, Faculty of ICT

## Introduction

In recent years there has been an increase in research performed in the field of bioinformatics, generating a large amount of tools to solve a wide array of problems in biology [1]. Due to this increase in tools, it has become difficult for researchers to track which tools, datasets and methods are available and useful for their area of study [1].

The FAIR Guiding Principles [2] provides four foundational principles that act as guidelines for researchers to ensure that their research is properly managed so that it can be reused as well as made findable [2]. The four principles provided by FAIR are Findability, Accessibility, Interoperability and Reusability. This study uses web crawling techniques to automate the FAIR assessment while providing a searchable archive of resources.

## Aims and Objectives

1. To obtain a searchable archive of tools and datasets.
2. To enhance the archive with a FAIR score indicating the Findability, Accessibility, Interoperability and Reusability.
3. To build a tool that automates the assessment of a resource’s Findability, Accessibility, Interoperability and Reusability and produces a FAIR score.

## Implementation

- Automated assessment tool.
  - Implemented using web crawling techniques.
- Portal
  - Web Application, acts as a searchable repository of tools and datasets.
  - Allows user to refine FAIR scores.
  - Automatically calculates FAIR scores for user defined pipelines.
- User defined pipelines
  - Can be created through the portal.
  - Can contain multiple different tools and datasets.
  - FAIR score calculated based on the scores of the individual tools and datasets.

$$\frac{Findability_{T1} + Findability_{D1} + ... + Findability_{Tn} + Findability_{Dn}}{N_T + N_D}$$

Calculation of pipeline Findability

## Results

Tool	FAIRshake Assessment				Automated Tool			
	F	A	I	R	F	A	I	R
BLAST	75%	100%	N/A	75%	65%	100%	50%	50%
HAMMOCK	100%	100%	N/A	40%	75%	50%	50%	75%
PyroHMMvar	100%	0%	N/A	40%	75%	50%	50%	100%
CytoSPADE	75%	0%	N/A	20%	100%	0%	50%	75%
MEANS	100%	0%	N/A	60%	75%	50%	100%	75%

Dataset	FAIRshake Assessment				Automated Tool			
	F	A	I	R	F	A	I	R
A443654 KINOMEscan	50%	50%	100%	67%	60%	80%	100%	66.67 %
(R)-Roscovitine KINOMEscan	50%	50%	100%	67%	60%	80%	100%	66.67 %
GTEx Portal Datasets	50%	50%	100%	100%	60%	80%	100%	100%
WormBase Datasets	100%	100%	100%	100%	100%	100%	100%	100%

Comparison of results between FAIRshake assessment and automated assessment.

	BLAST		MEANS	
	FAIRshake Assessment	Automated Assessment	FAIRshake Assessment	Automated Assessment
Findability	75%	65%	100%	75%
Tool is freely downloadable.	25%	40%	25%	40%
Tool has a proper description.	25%	25%	25%	25%
Previous versions of tool are available.	0%	0%	25%	10%
Tool has a unique identifier.	25%	0%	25%	0%

Comparison of Findability metrics for BLAST and MEANS

## Evaluation and Limitations

### Evaluation:

- Automated assessment results do not differ too much from FAIRshake results.
  - Differences in results arise from difference in weightings
- Tool FAIR assessment can be mostly automated.
- Dataset FAIR assessment can be fully automated.

### Limitations:

- Inability to discern whether a tool uses ontologies or not or whether the ontologies used were simply omitted from the metadata.
  - Rectified through user input
- Inherent difficulty in obtaining unique identifier (DOI) of tools.

## Future Work

- Further research to improve metrics to calculate a more accurate FAIR score.
- Implementing multi-threading techniques to improve efficiency of FAIR assessment.
- Improving web crawling techniques to overcome Selenium limitations.

## References

[1] Cannata, N., Merelli, E., & Altman, R. B. (2005, Dec). Time to organize the bioinformatics resourceome. PLoS Computational Biology, 1(7). doi: 10.1371/journal.pcbi.0010076

[2] Wilkinson, M. D., Sansone, S.-A., Schultes, E., Doorn, P., da Silva Santos, L. O. B., & Dumontier, M. (2018, June). A design framework and exemplar metrics for FAIRness. Scientific Data, 5, 180118. doi: 10.1038/sdata.2018.118