# Bioinformatics Resource Portal

Nigel Alfino, BSc. IT (Hons) Software Development
Supervisor: Mr Joseph Bonello, Co-Supervisor: Prof. Ernest Cachia
Keywords: FAIR, FAIR Assessment, Reproducibility

Bioinformatics is the field of study that applies computational techniques to improve the understanding and organisation of biological data. A major difficulty in performing research in bioinformatics is finding the right tool or dataset to use. This obstacle arises from a lack of effort made in the bioinformatics community to make tools or datasets reusable. This issue is compounded by the lack of an indexable resource of tools and datasets, as well as a lack of assessment of their reusability. Moreover, reproducible research in biology and medicine is also an issue. A study produced by ATCC, 2019, shows that over 70% of researchers were unable to reproduce the findings of other scientists [3].

The FAIR guiding principles [2] provide four principles to measure a tool or dataset's Findability, Accessibility, Interoperability and Reusability, as well as guidelines on how to score these measurements. This study aims to create a searchable portal of tools together with a semi-automated assessment tool that calculates a FAIR score for tools. This will allow a researcher to make an informed decision on which tool is appropriate and to enable other researchers to determine how easy it is to reproduce a study. The FAIR score provides researchers with a level of trust in the resources they use since the FAIR scores indicate how usable a tool is from a scientific and applicable point-of-view and the degree of interoperability it has with respect to different contexts. The results as can be seen from the portal are shown in figure 1.

## BLAST

| Findability | | 90.0% |
|---|---|---|
| Download Link | Link provided on website. | 40.0% |
| DOI | https://doi.org/10.1002/cpet.8 | 25.0% |
| Description | The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. | 25.0% |
| Older Versions | Previous versions of tool unavailable. Findability could be improved if previous versions of tool were made available. | 0.0% |

Refine Findability

| Accessibility | | 100.0% |
|---|---|---|
| API availability | Tool supports API usage. | 50.0% |
| Command Line | Tool can be accessed through command line. | 50.0% |

Refine Accessibilitiy

Figure 1: Summary Information of Tool

The proposed semi-automated assessment tool uses web crawling techniques to obtain information based on a set of pre-defined criteria. Assessment results are accessible through a portal were additional information can be provided to refine the FAIR score. Researchers can also calculate scores for Bioinformatics pipelines, i.e. a series of tools and datasets used sequentially in a study, based on the individual FAIR scores of tools and datasets.

Our results show that the majority of the FAIR assessment criteria of tools, datasets and pipelines can be automated. However, some additional information, such as unique identifiers for tools and determining whether a tool uses ontologies or not, may be required as additional input from the user as this information may not always be available online. This can be seen in figure 2.

## BLAST - Findability

**Download Link**
`https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=[`

**DOI Link**
`https://doi.org/10.1002/cpet.8`

**Description**
The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

**Previous Versions**
○ Previous Versions available on website/public repository.
● Previous Versions unavailable.

[ Update ]

Figure 2: User Refinement

**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

[3] ATCC. (2019). Six factors affecting reproducibility in life science research and how to handle them. Nature News. Retrieved from https://www.nature.com/articles/d42473-019-00004-y