BLAST, Pipelines, and FAIR

Statement of Work

Create a workflow for BLAST Reciprocal Best Hit (RBH) that is more FAIR than present methods

How to prove FAIR improvements?

F A I R - what will we improve?

FA: data will not be better findable or accessible by this workflow (out of scope)

I: Interoperability improves ….

- Machine AND human readabilty (CWO viewer / cross platform script)
- i/o ontolgies contained in files (EDAM_… standard)

R: Reusabilty improves ....

- Backward tracking capability as all in i/o files
- Each step in the pipeline gets also more reusable by this workflow (modular approach - modular reuse of components).

```
                    ┌─────────────────┐
                    │   BLAST all     │
                    │  Protein Seqs   │
                    │      from       │
                    │   Organism A    │
                    │    against      │
                    │   Organism B    │
                    └─────────────────┘

┌─────────────────┐                        ┌─────────────┐      ┌─────────────┐
│    Extract      │                        │    Parse    │      │  Check for  │
│    Protein      │                        │   Output    │ ───→ │   Overlap   │
│   Sequences     │                        └─────────────┘      └─────────────┘
│      from       │
│   Organisms     │
│    A and B      │
└─────────────────┘
                    ┌─────────────────┐
                    │   BLAST all     │
                    │    Protein      │
                    │     Seqs        │
                    │   Organism B    │
                    │    against      │
                    │       A         │
                    └─────────────────┘
```