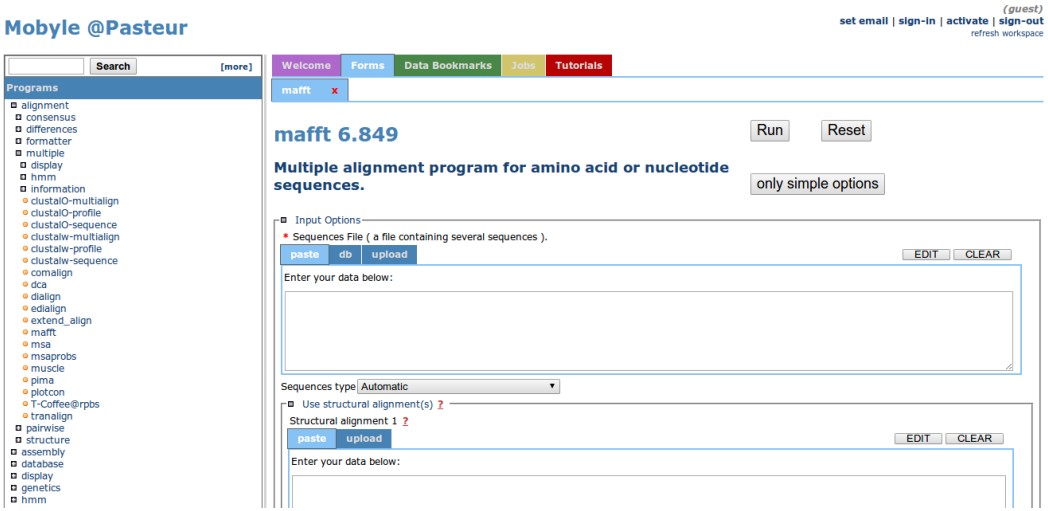
Use of Semantically Annotated Resources in the Mobyle2 Web Framework

Hervé Ménager, Bertrand Néron, Olivia Doppelt-Azeroual,
Olivier Sallou
Mobyle development team

BOSC 2014

Introduction - Mobyle

- A web-based workbench for bioinformatics (v1.5.3)



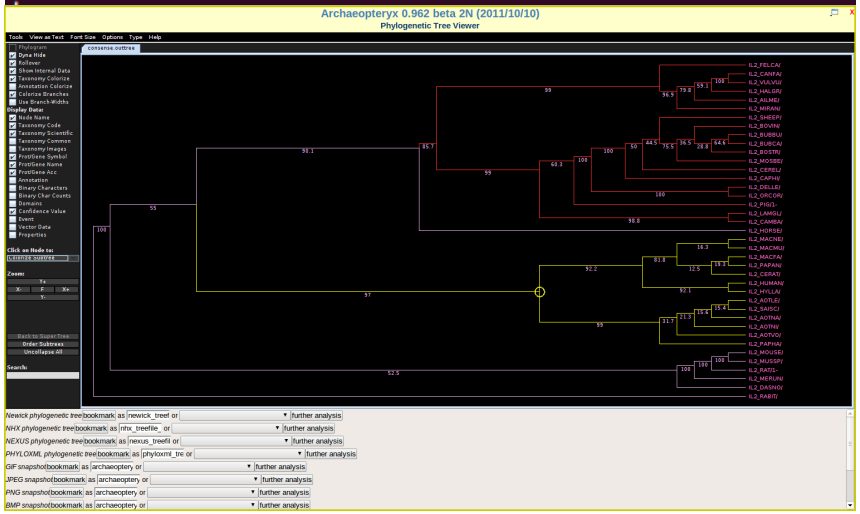
- adapted to the needs of biologists, **sometimes very punctual**.
- also provides a **persistent workspace if needed**.

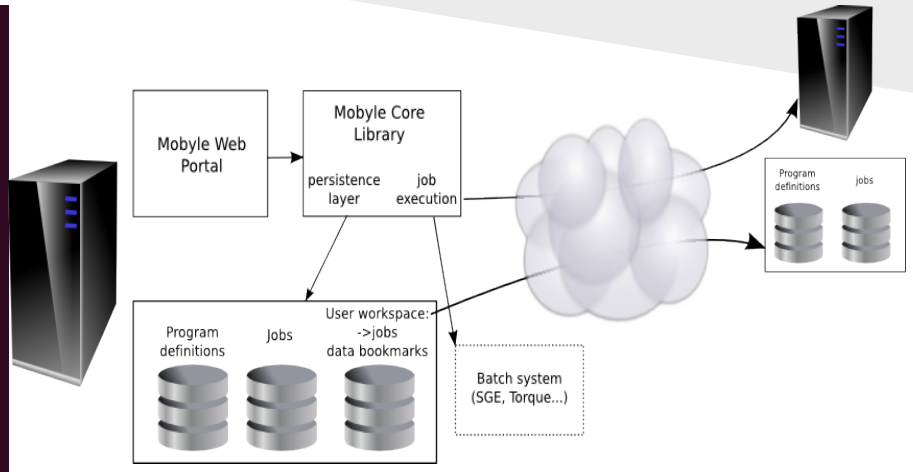
Mobyle 2

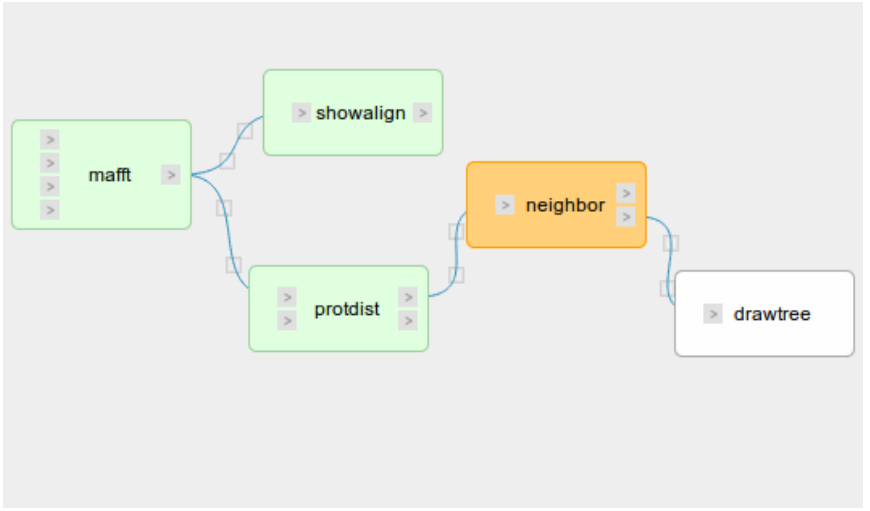
- Major rewrite of Mobyle
- Currently in development
- New features:
  - groupware functionalities
  - secure data sharing
  - REST api
  - ***ontology based service descriptions***

Multiple resources / services

```
hmenager@coop:~$ mafft ~/mobyle_example_data/CBM20.fasta

nseq =  6
distance =  ktuples
iterate =  0
cycle =  2
nthread = 0
done.
scoremtx = 1
Gap Penalty = -1.53, +0.00, +0.00

tuplesize = 6, dorp = p


Making a distance matrix ..
    1 / 6
```







Resource description

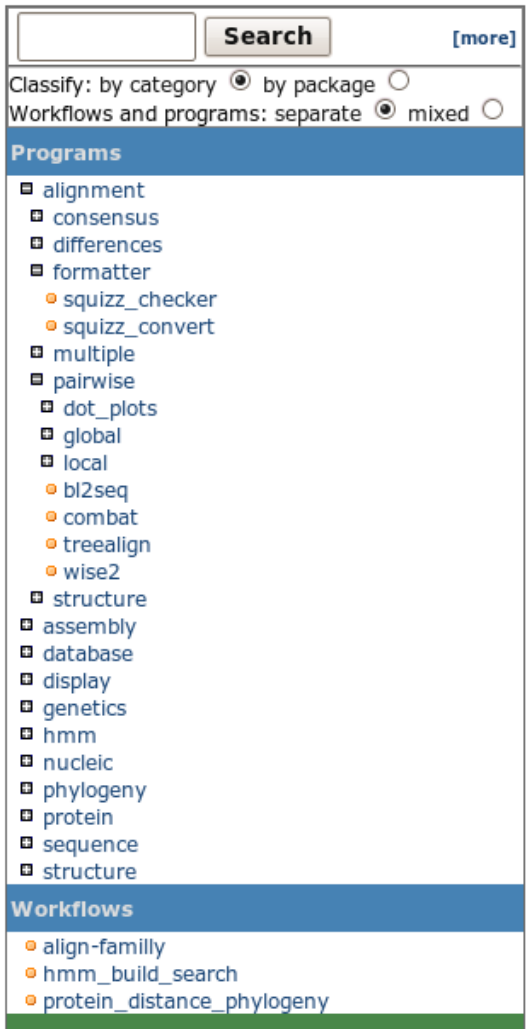
```
<program>
  <head>
    <name>mafft</name>
    <version>6.849</version>
    <doc>
      <title>mafft</title>
      <description>
        <text lang="en">Multiple alignment program for amino acid or nu
      </description>
[...]
      <doclink>http://mafft.cbrc.jp/alignment/software/about.html</docl
    </doc>
    <category>alignment:multiple</category>
  </head>
  <parameters>
    <paragraph>
      <name>input_opt</name>
      <prompt lang="en">Input Options</prompt>
      <parameters>

        <parameter ismandatory="1" issimple="1">
          <name>sequences</name>
          <prompt lang="en">Sequences File ( a file containing several
          <type>
            <datatype>
              <class>Sequence</class>
            </datatype>
            <dataFormat>FASTA</dataFormat>
          </type>
          <format>
            <code proglang="perl">" $sequences"</code>
            <code proglang="python">" " + str( sequences )</code>
```

Classification

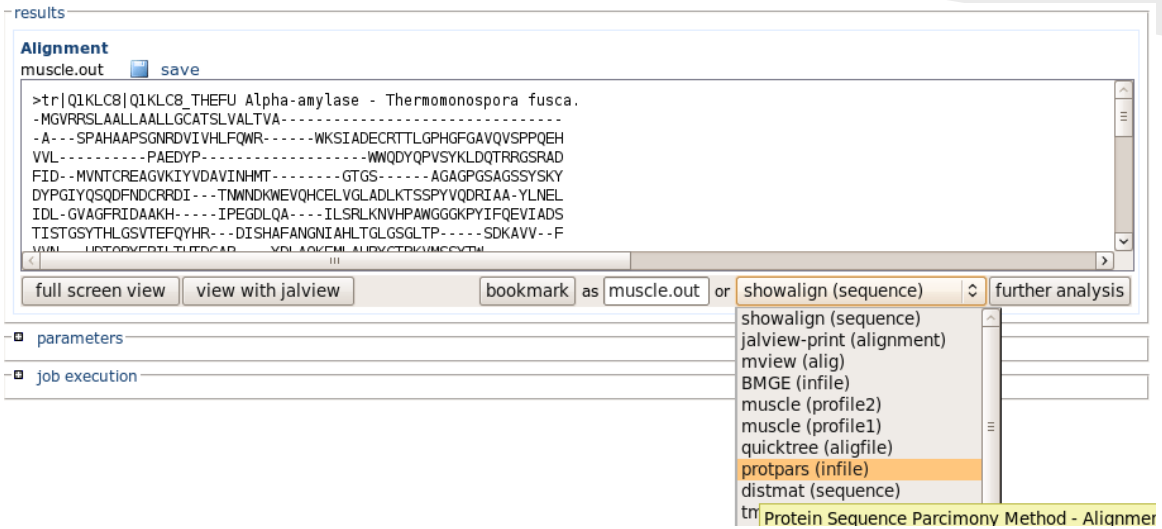
Parameters typing

Discovery and integration

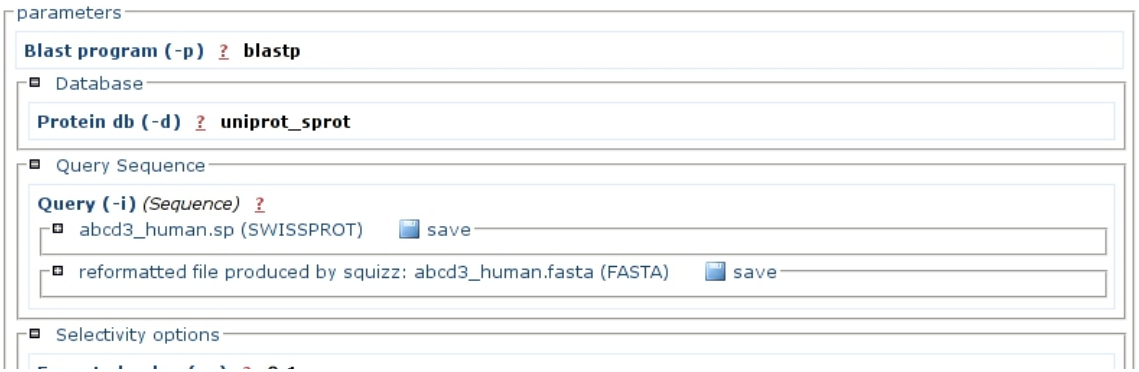
Services classification



Guided chainings



Format detection/conversion



Typing system

- datatype: the “kind” of information stored in the data
  - a datatype is either a reference to a python class
    - e.g.: Sequence -> “`sequence.py`”
  - or a “dynamic” subclass of one of these python classes
    - e.g.: BlastReport, subclass of Report -> “`report.py`”

- data format: the format used to store the data
  - e.g.: FASTA, SWISSPROT, etc.

Shortcomings of Mobyle1 classification and typing system

- freely extendable
  - highly sensitive to annotation errors (PDB or Pdb?),
  - potentially inconsistent between multiple MobyleNet sites,
- confusions between datatypes / data formats
- Mobyle-specific mechanism
  - hard to share/reuse these annotations

Solution: use of an ontology

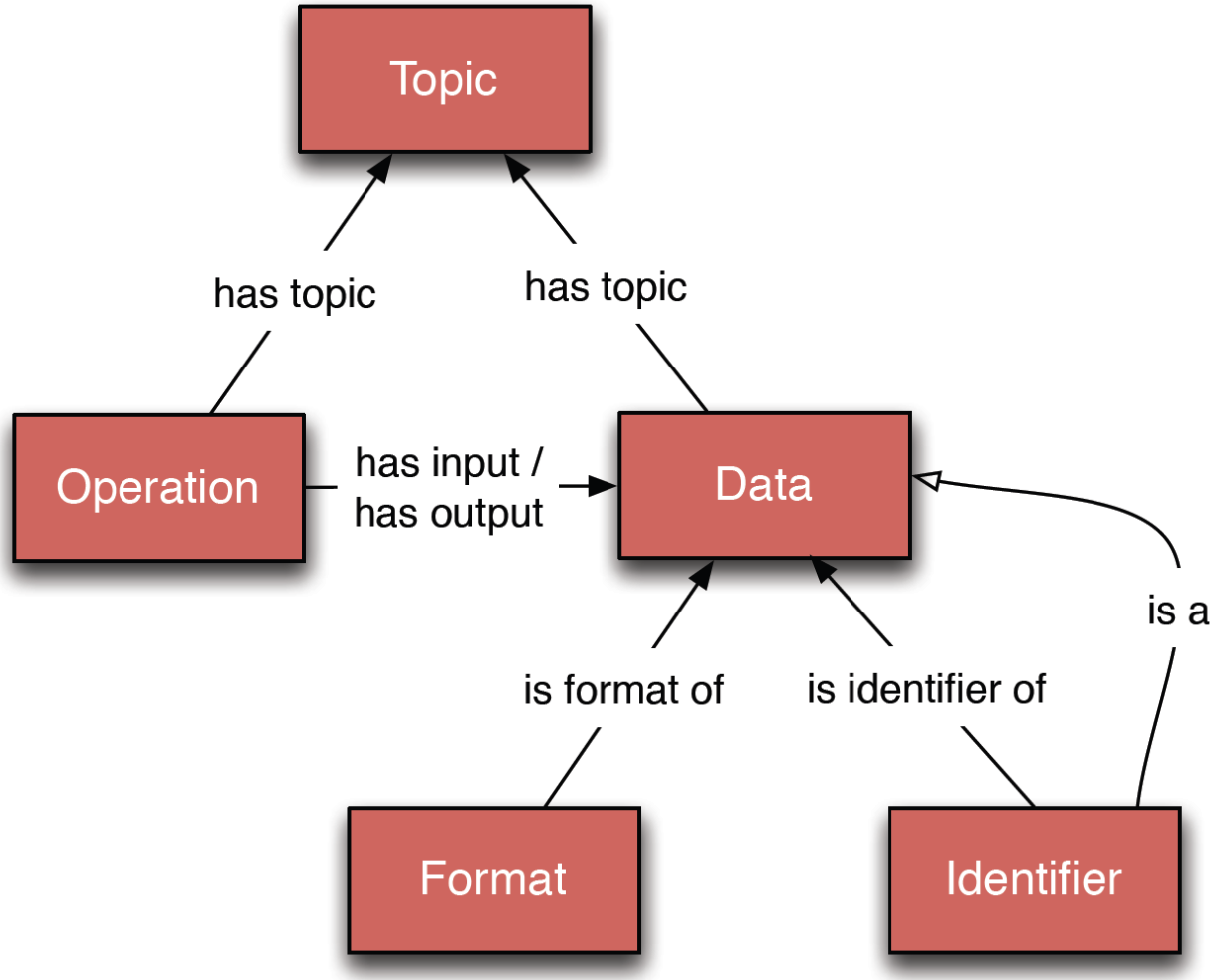
General Properties

- controlled vocabulary,
- describing the relationships between the different terms.

Scope

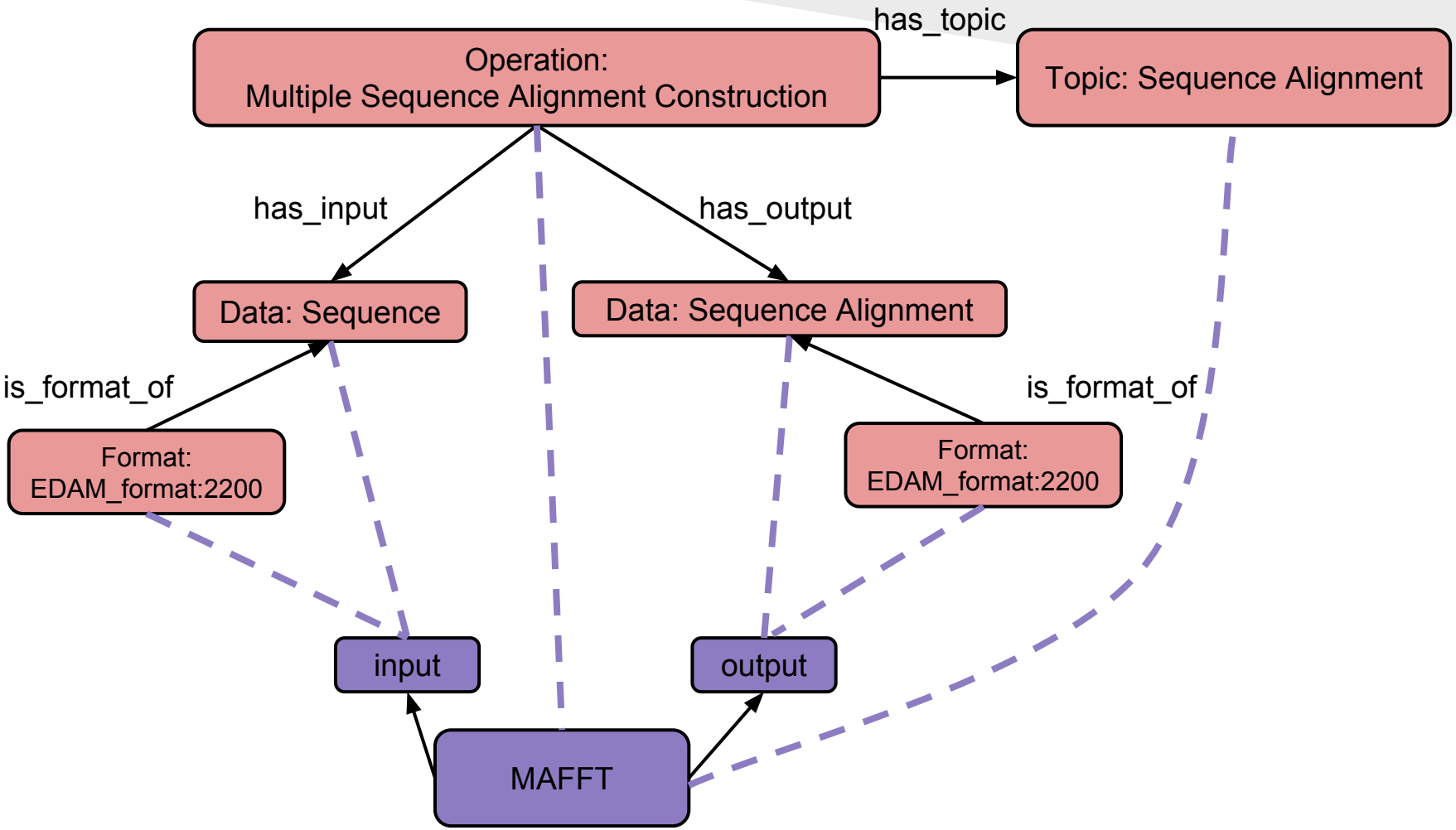
- for the data and parameters:
  - the type of information described,
  - the format used to store it.
- for the services classification:
  - the hierarchy of categories to store the services.

EMBRACE Data And Methods ontology



http://edamontology.org/page

Service annotation: a simple example



Services classification

- Structure analysis
  - Structure prediction
    - Protein secondary structure prediction
      - garnier
      - helixturnhelix
      - pepcoil
    - + Nucleic acid structure prediction
  - Nucleic acid structure analysis
    - + Nucleic acid structure prediction
    - banana
    - btwisted
  - RNA structure and alignment
    - rnaalifold
    - rnaplot 1.8.4
  - + Protein structure analysis
  - + Structure databases
- Phylogenetics
  - kitsch
  - neighbor
  - clique
  - distmat
  - weighbor 1.2.1
- Data handling

- **Usage of EDAM topics hierarchy**

Guided interactive chainings

>FR775944|FR775944.1-Lactobacillus-crispatus-CIP-102990T
cttacggtaagtttgttgttgaaccacttgagcgaggctttggtact
actttaggtaattctttaagaagagtattacttacttctattccaggtactggacttgtt
tatgtgcaaattgatggcgttttacacgagttctcaacagttcctggcgtcagagaagat
gtaaccaagatcatcttgaacttgaagaaacttgaattaaagtctctttcagatgaacaaa
aggttattgaattggacattgaaggtccagccacagtaactgctgatgatcttaaag >>

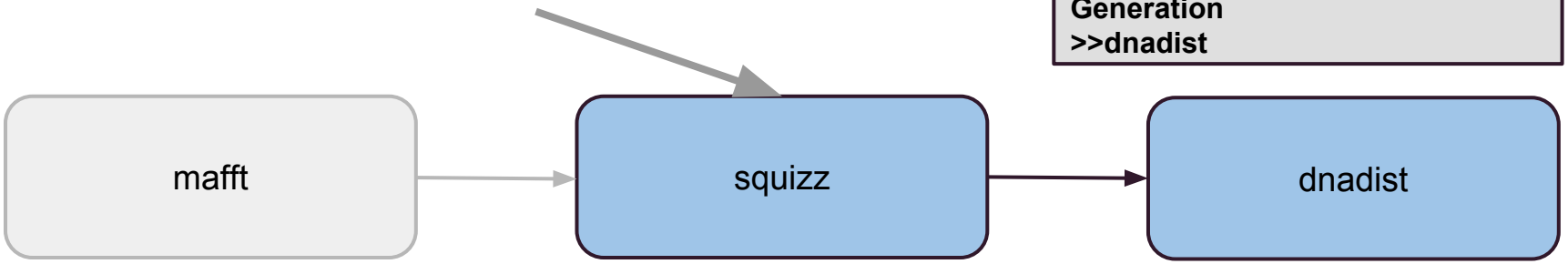
Sequence Profile Generation
>> hmmbuild
Sequence Alignment Rendering
>> boxshade
Sequence Alignment Reformatting
>> mview
>> squizz

Automatic conversions

>FR775944|FR775944.1-Lactobacillus-crispatus-CIP-102990T
----------cttacggtaagtttgttgttgaaccacttgagcgaggctttggtact
actttaggtaattctttaagaagagtattacttacttctattccaggtactggacttgtt
tatgtgcaaattgatggcgttttacacgagttctcaacagttcctggcgtcagagaagat
gtaaccaagatcatcttgaacttgaagaaacttgaattaaagtctctttcagatgaacaa
aaggttattgaattggacattgaaggtccagccacagtaactgctgatgatcttaaa >>

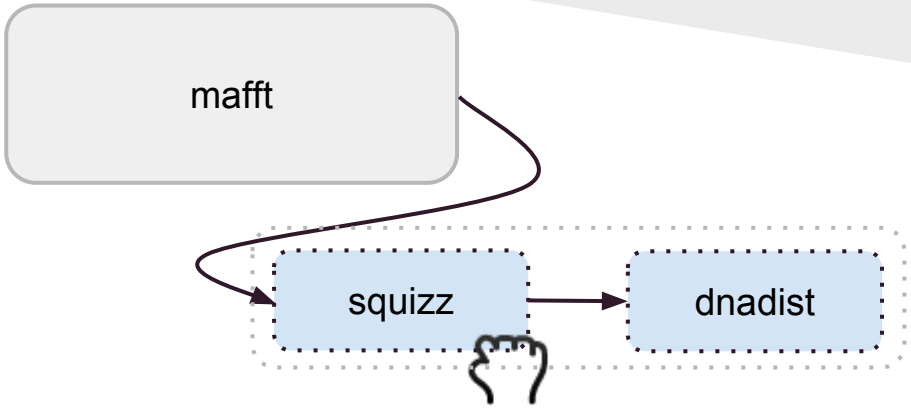
Sequence Profile Generation
>> hmmbuild
Sequence Alignment Rendering
>> boxshade
Sequence Alignment Reformatting
>> mview
>> squizz
Sequence Distance Matrix Generation
>>dnadist

EDAM_operation:0335 (File reformatting)

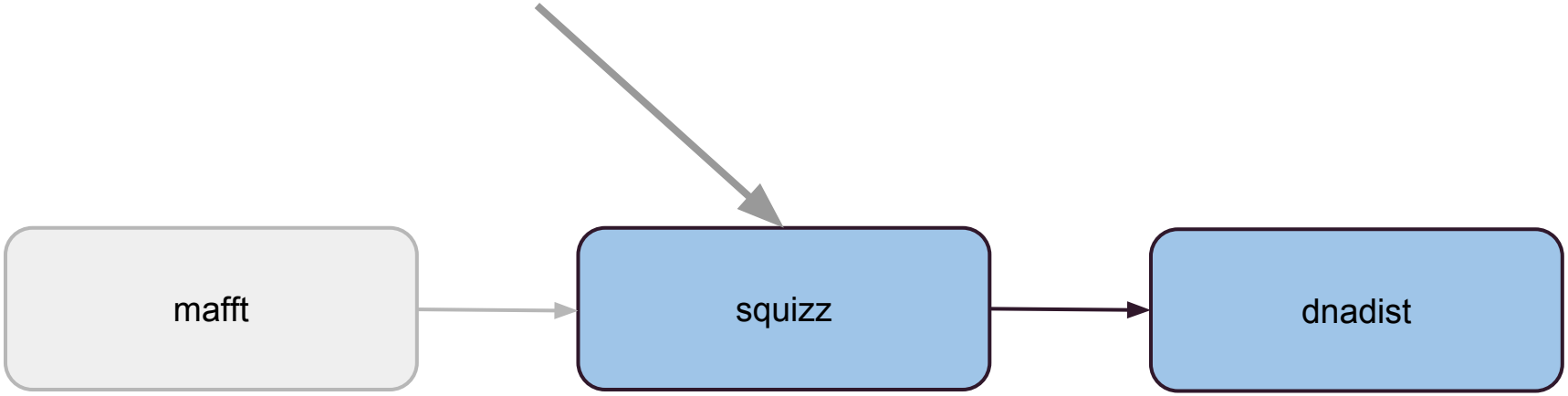


Automatic conversions in workflows

Sequence Profile Generation
>> hmmbuild
Sequence Alignment Rendering
>> boxshade
Sequence Alignment Reformatting
>> mview
>> squizz
Sequence Distance Matrix Generation
>>dnadist

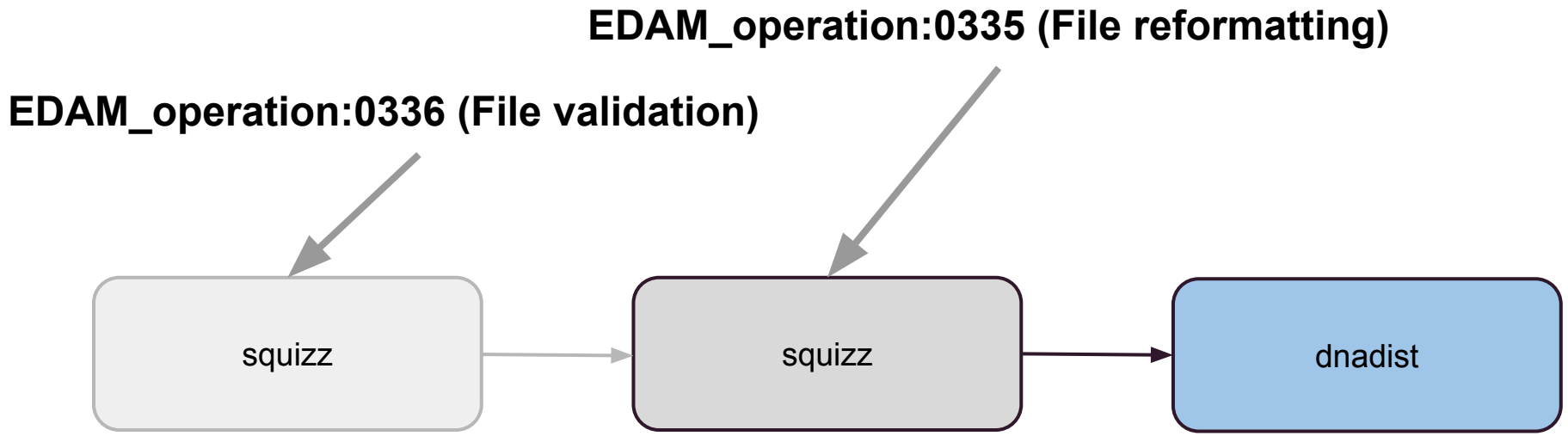


EDAM_operation:0335 (File reformatting)

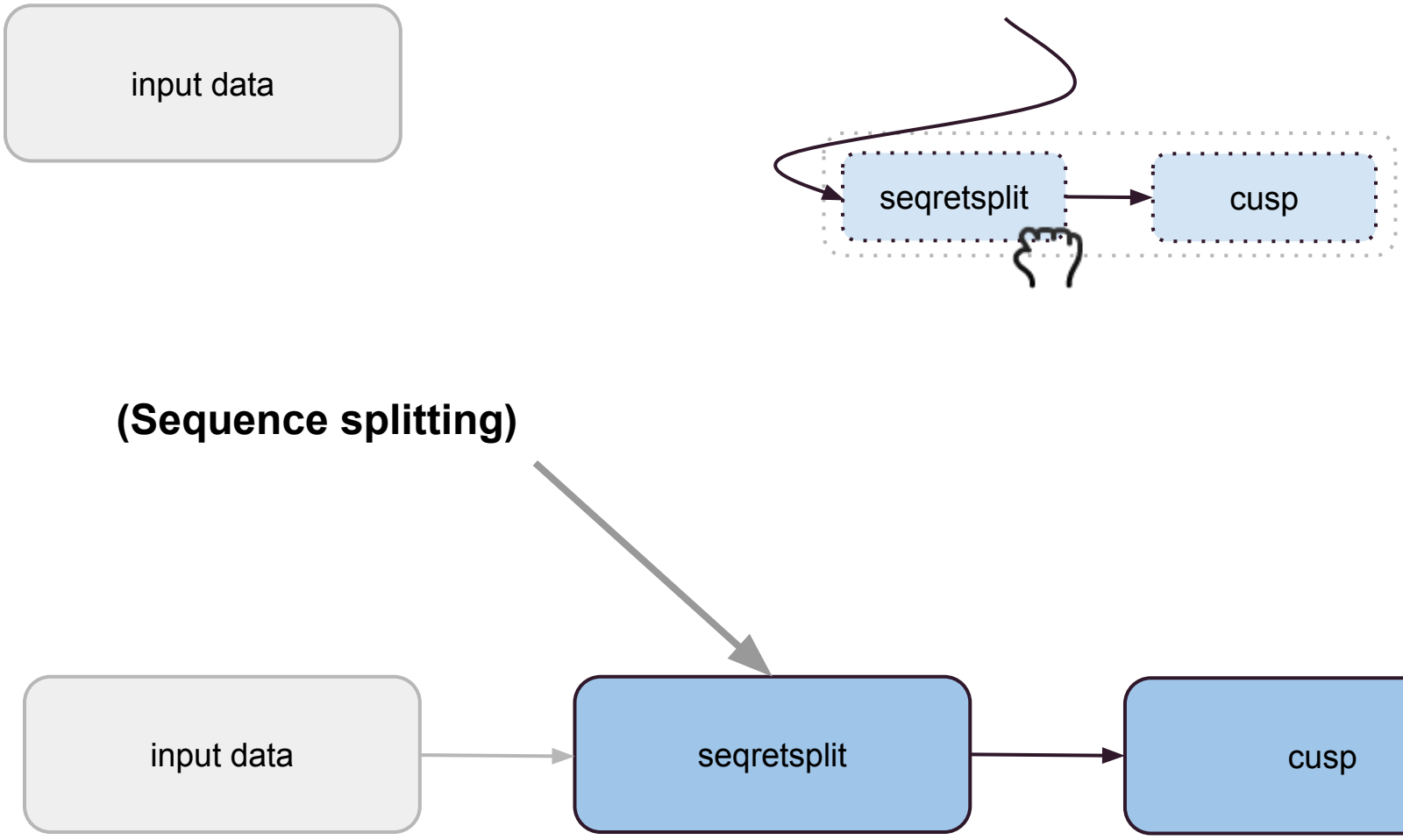


Automatic format detection

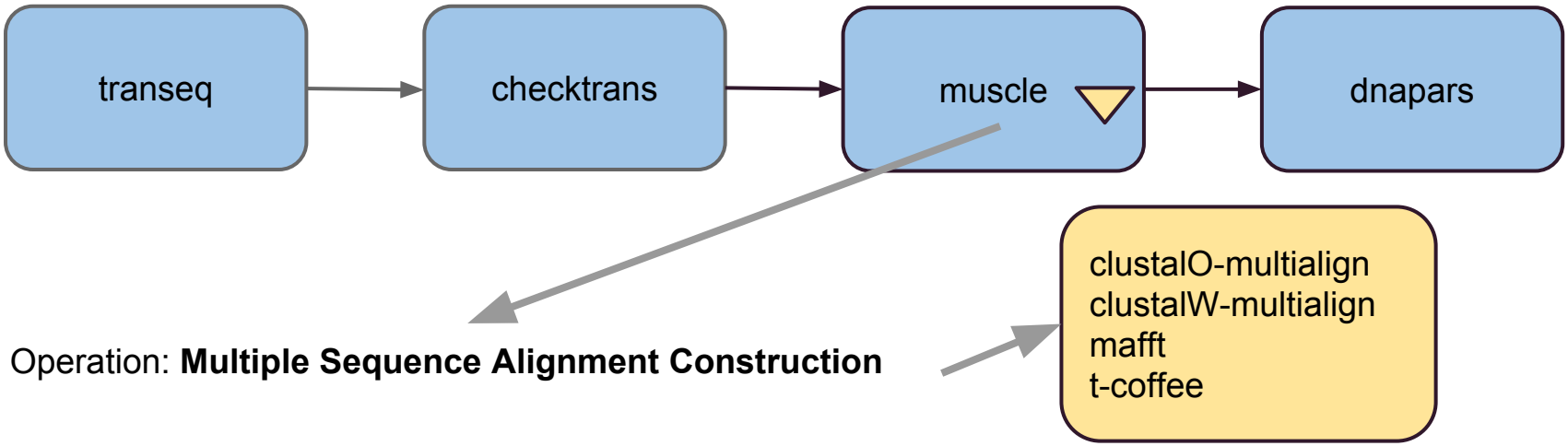
- **if the format of a data item is unknown (e.g., directly uploaded by the user)**
- **detection of the data format, conversion if necessary**



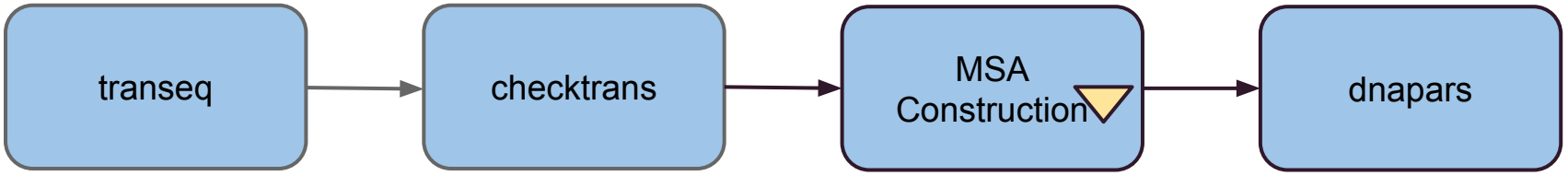
Automatic split and iteration



Equivalent methods



Abstract methods



Operation: **Multiple Sequence Alignment Construction**

MSA Construction Method: clustalw-multialign

clustalO-multialign
mafft
muscle
t-coffee

Conclusion

- Mobyle 2 in development stage

- Ontologies are a useful resource for the integration of bioinformatics resources:
  - “service discovery”
  - implementation of high-level services
  - sharing of resource descriptions with other projects

- Remaining challenges:
  - integration in “user-friendly” interfaces
  - quality of the “annotations” in the service descriptions

Acknowledgements

- Mobyle and Mobyle2 development teams,
- our users,
- EDAM and Tools Registry teams: Jon Ison, Matus Kalas, Kristoffer Rapacki, etc.
- **MASTODONS** funding