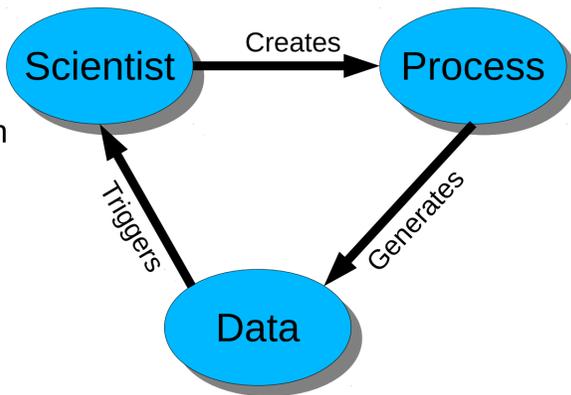


Motivation

Science is increasingly becoming data-centric. Now more than ever huge datasets are available at the scientists' fingertips. Adding to this, many scientific processes which act on such data are being developed but go unnoticed without any system to exploit their full potential.

Science is often locked into a perpetual cycle of generating more data. Open Linked Data exposes more data through semantics. Our idea is to apply the same notion to Processes thus having a system of **Networked Open Processes (NOP)**.



Framework

TReQL

```

WANT TO spec:SequenceAlign AS sq THEN spec:
  SequenceDisplay AS sd
WHERE { sq.hasInputParameter.objectType IS spec:
  Fasta }
INVOKE WITH {
  SPARQL { sparql_query } FOR sq.hasInputParameter.
  sequence1,
  URL fasta_url FOR sq.hasInputParameter.sequence2
}
  
```

Higher-level User Interfacing

user

OWL Reasoners / SPARQL

Graph Embelishment and Execution Engine

core public components

Semantic Processes Tripple Stores

Linked Open Provenance Data

Migratable Process Stores

POI Registry

Message Passing Exchanges

Optimized Scientific Process Containers

Traditional Service Containers

resources

Cloud

Grid

Desktop

Browser

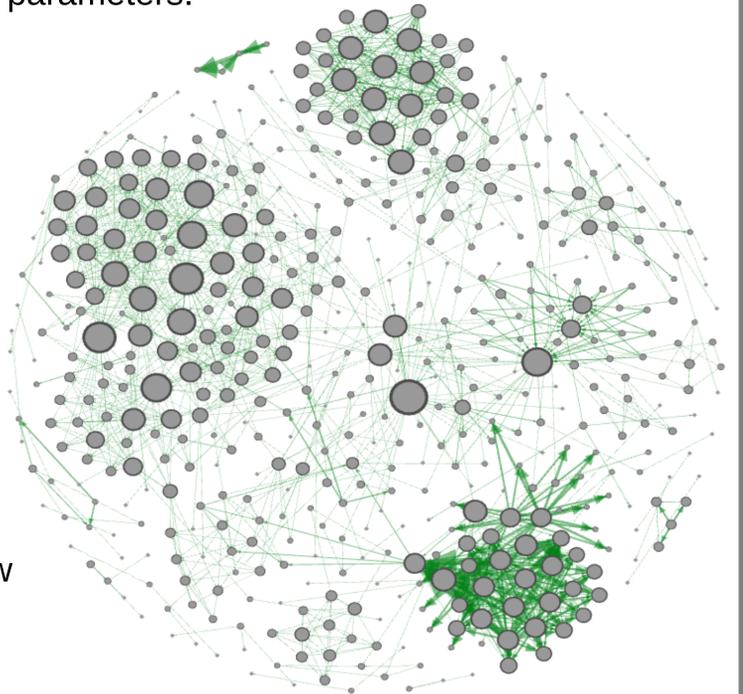
Since we do not intend to re-invent the wheel, the system should also work with traditional services as long as some annotation is given. Though, these services have limitations when dealing large datasets. Such web services are **anchored** and therefore can not move closer to the data. They can not easily **replicate** themselves to work on partitioned data and they can not achieve p2p **communication** thus a 3rd party always has to relay data or references to it.

Objective

The main objective is to create a framework for scientific distributed computing through sharing and exposing processes. The graph below depicts potential Interoperability between services described in SADI framework. Vertices represent operations while edges represent parameters.

Clusters form around common edges (common data types).

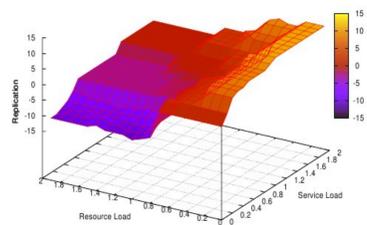
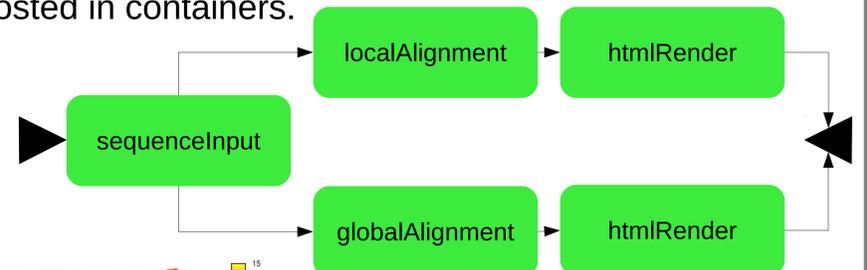
Degree distribution follows power-law



As with Open Linked Data, **Semantic Repositories** expose endpoints to process descriptions. The semantic repositories are the source of the linked processes.

We believe that at a certain point a process/service should become immutable. The preservation of such a process allows for better data provenance. Also, as opposed to web services, processes should be addressable on non changeable attributes such as location and ownership. The **POI(Process Object Identifier) Registry** serves that purpose This approach allows a process uniquely identifiable and not anchored to a URL thus can migrate/replicated freely on purposely deployed containers.

Ultimately processes can be hosted **dormant** or **active**. Dormant processes are kept in **Process Stores** which allows processes to be archived. Active processes are hosted in containers.



- **Dynamic** deep-network web service deployment
- **Direct** service to service communication
- **Fuzzy Elastic** prediction-based service scaling