



2 Tools Interoperability and Service Registry

Leads: Bengt Persson (SE), Søren Brunak / Peter Løngreen (DK)

Team of relevant node staff/experts contributing to the work:

Kidane Tekle (NO), Victor de la Torre (ES), Joaquin Dopazo (ES), Federico Zambelli (IT), UNIMI, Mattia D'Antonio (IT), CINECA, Helen Parkinson, EMBL-EBI, Kristoffer Rapacki (DK)

2.1 Remit

Identification of tools and resources was highlighted as a key challenge in the report on industry needs and expectations from ELIXIR; to meet this need ELIXIR will develop and provide a Service Registry (led by elixir.dk). In addition, biological tools and resources need to meet requirements of increasingly complex analysis pipelines and the interoperability of software, services and underlying resources is often a bottleneck in applied bioinformatics. This Work stream will address and deliver concrete objectives and proposals to address:

1. Overview of the Service Registry and how this will interface and exchange with workflow based systems
2. A plan to address Tools Interoperability within ELIXIR
3. Address the standard, processes and tools to drive and increase efficiency in data-resource integration and development (e.g. through efforts such as PSI, DAS, BioJS and IMEX).

2.2 Introduction and Specific challenges

During the Preparatory Phase, four essential components of tool integration were identified – tool discovery, user friendliness, benchmarking, and interoperability. Furthermore, it was concluded that multiple access routes to tools are necessary – both human-friendly interfaces and programmatic access. The latter can be achieved using Web Services for easy access via multiple routes.

The creation of a comprehensive, searchable tools registry is prerequisite to the other objectives mentioned above: tool discovery, interoperability and benchmarking. Therefore, the initial work has focused on this task. The basic registry structure has been settled, consisting of a tool description model, the data itself and an upkeep strategy intended to secure a sustainable high quality service. The tool description model contains administrative, technical and scientific information, sufficient for the user to locate, understand, compare and select suitable tools. Whenever possible the information will be provided using defined ontologies. The data will obey the model. The upkeep strategy envisions a distributed, collaborative effort involving a community of tool providers, large institutions as well as individual scientists. The interfaces to the registry will make possible the interactions needed: convenient browsing by the users and updates by the tool providers, using a variety of suitable mechanisms. Both adding and removing of data will be possible.

Since the Preparatory Phase, new protocols have appeared (BioXSD is currently the recommended standard for data interchange in workflow construction) and additional will appear in the future. In addition, workflow tools have become increasingly popular. One central task for PoW2 is to keep ELIXIR tools and services up-to-date with standardised and widely used access routes. Furthermore, PoW2 will maintain a comprehensive Tools Registry with adequate descriptions of each tool using defined ontologies, making it easy for users to find relevant tools. New tools should be added fast, and obsolete tools should be removed or tagged appropriately. A regular and automated benchmarking procedure will add quality assessments to each tool and help the users in selecting the optimal tools for their analyses.

Life science users will benefit from tools packages. PoW2 will keep regular interactions with the biomedical community at large in order to continuously be updated with their needs. Furthermore,



we expect an increased need for tools acting in a secure compute environment for analysis of sensitive data.

ELIXIR will give priority to key tools of central importance for the life science community. ELIXIR will also give priority to sustainability of tools, adding kite-mark to a limited number of core tools upon which additional services can be built.

2.3 Collaborations and services beyond Europe, GAP analysis

ELIXIR will collaborate in relevant areas with bioinformatics infrastructures world-wide, e.g. NCBI. Identification of major gaps in the ELIXIR service catalogue for this area will be done in Jan/Feb when all node contributions are compiled, including description of on-going efforts or options for addressing the gaps.

2.4 ELIXIR Objectives up to 2018

The order of the list below reflects the order in which the work will proceed. Thus, the work on tool discovery will precede the use considerations, the enabling of multiple interfaces will open up for the work on tool interoperability etc. The dates are approximate as parts of the effort will be distributed and undertaken in parallel.

- Comprehensive Tools Registry (**Dec 2014**)
- Enabling Tools with multiple interfaces – human, machine, downloadable code, ... (**Dec 2015**)
- Assuring tools compatibility using standardised interfaces (**Dec 2016**)
- Automated system for regular tool benchmarking (**Dec 2017**)
- Organisation of tools competitions analogous to the CASP meetings
- A core set of tools with ELIXIR kite marking

We expect PoW2 to have a large and fundamental impact to bioinformatics analyses of biological data from all domains of life. Furthermore the offered services will be of large importance to users in academia, health care, agricultural and pharmaceutical companies, and the society at large.

Specific and timed deliverables and milestones for node services will be done in the Jan/Feb version. Based on the service gap analysis, we will describe key objectives for building up new services.

2.5 Relevant Node services

The purpose is to collate the relevant Node services into a coherent offering from the distributed infrastructure. The planned work will be described in detail and the role of each partner specified. This will be done in the longer version later, including graphic illustration of the coherent offering.

PoW2 is mainly dependent upon PoW1 (data), PoW3 (tech services) and PoW5 (training). (Furthermore, PoW7 (management) is of course of central importance for all PoWs.) Finally, PoW2 will use ontologies and vocabulary from PoW4, and the domain specific services in PoW6 will depend on tools from PoW2.

Format:

2.5.1 Node Service 1

Refer to the Node application:

- Each Node with mapped services to this PoW will need to provide a short summary statement of services in line with the Node application and the SAB feedback.
- Include collaborations between Nodes as well as organisations outside ELIXIR.



Provide an indicative dependency graph:

- Is the service dependent on other ELIXIR offerings?
- Are there major dependencies on the service?

It is expected that this section will match Annex "Service Delivery Plan" in Collaboration Agreements and should include ancillary services such as training.

2.5.2 Node Service 2

2.6 Resources

- This will be detailed in accordance with the Service Delivery Plans
- Provide details of in kind contributions
- Identify gaps in current resource needs
- Provide options for addressing gaps in resources (e.g. additional grants needed, long term sustainability, ...)