BioVeL: Biodiversity Virtual e-Laboratory

Saverio Vicario[1], Alex Hardisty[2], Niobe Haitas[3]

[1]CNR - Institute for Biomedical Technologies, Italy
[2]Cardiff University, United Kingdom
[3]HealthGrid, France

E-solutions for the management of biodiversity in the 21st century

Scientists are being pressured to provide convincing evidence of changes to contemporary biodiversity, to identify factors causing decline in biodiversity and to predict the impact of, and to suggest ways of combating, biodiversity loss. Altered species distributions, the changing nature of ecosystems and increased risks of extinction all have impacts in important areas of societal concern. Biologists and environmental scientists are asked to provide decision support for managing the biodiversity component of our environment at multiple scales (genomic, organism, habitat, ecosystem, landscape) to prevent and mitigate such losses. Generating the evidence and providing decision support relies, increasingly, on large collections of data held in digital formats, and the application of substantial computational capability and capacity to analyse and model such data, and to run simulations.

The BioVeL approach

BioVeL, a 3-year FP7 project, aims to catalyse the energy and knowledge present in the research community, helping to address the challenge of understanding and managing biodiversity. More precisely, the goal of the BioVeL project is to provide a seamlessly connected informatics
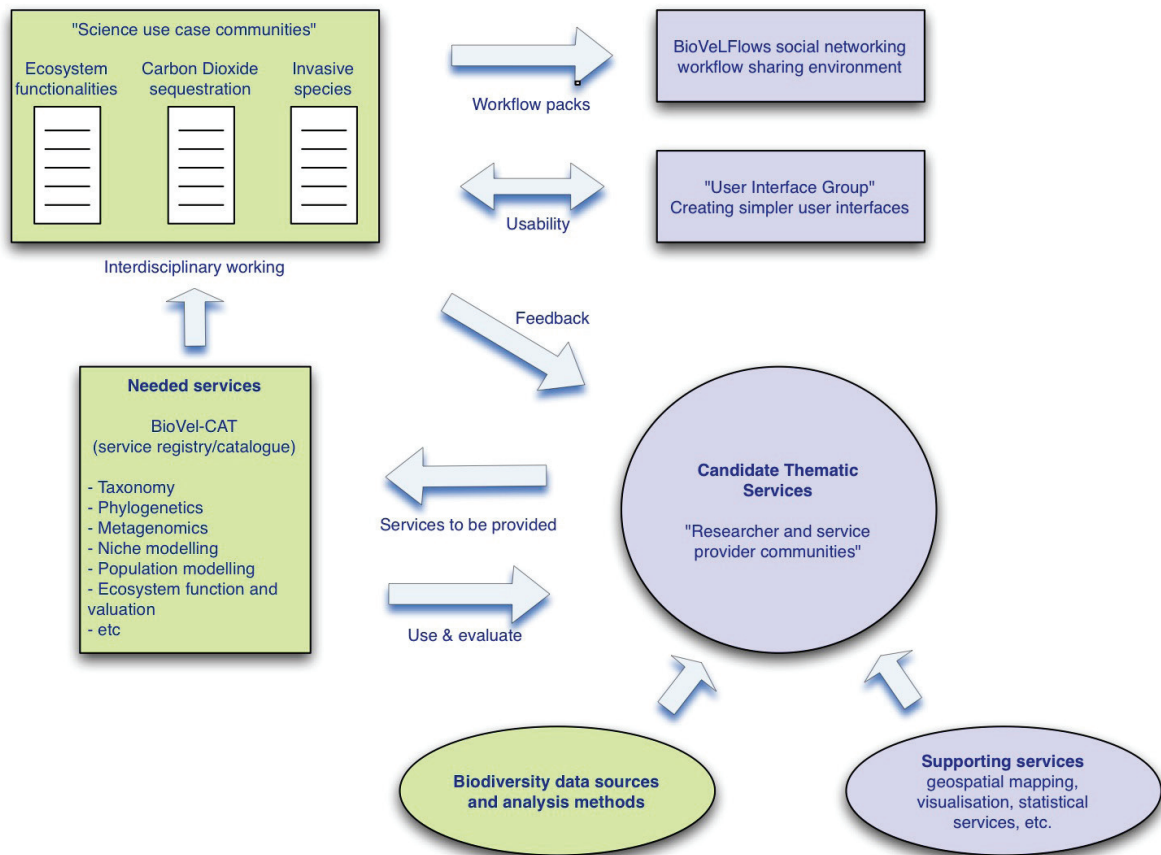

Figure 1. Algorithm selection module of the Superclusteroid tool.

environment that makes it easier for biodiversity scientists both to carry out *in silico* analysis of relevant biodiversity data, and to pursue *in silico* experiments based on the composition and execution of complex digital data-manipulation and -modelling tasks. In September, the Biodiversity Virtual e-Laboratory (BioVeL) project had its kick-off. In BioVeL scientists and technologists will work together to meet the needs and demands for 'e-Science', and to create a production-quality informatics infrastructure to enable pipelining of data and analysis into efficient, integrated workflows. Workflows represent a way of speeding up scientific advance when that advance is based on the manipulation of digital data (Gil *et al.*, 2007).

BioVeL will not produce new applications or software, but will help developers to expose useful software programs as Web services. It will allow users to access and compose them in workflows, and to share and comment the composed workflows, using an adaptation of the myGrid[1] suite. Furthermore, a large section of the project is dedicated to engaging the community along three lines of action: 1) following the myGrid paradigm, all services and workflows will be inserted within a social network framework that will ensure feedback and quality control of best practice; 2) the project will designate "pals", who are persons knowledgeable in a specific scientific field, who will make the connection between the ICT part of the project and the user community; 3) a series of workshops will ensure more formal occasions to engage with, and collect feedback from, the community of users and developers. To allow a fast reaction cycle between input from the community and ICT developers, the project will unfold using an agile-process paradigm. To seed the infrastructure with the first workflows, the project will focus on three science use cases from the biodiversity community (ecosystems services, CO_2 sequestration, invasive species management), and the workflows will be built taking applications from the following areas of knowledge: taxonomy, phylogenetics, metagenomics, ecological niche modelling, ecological population modelling, ecosystem functioning and valuation. For each of these areas, the project will nominate a "pal". The "pal" will help to compose the first workflows, together with the experts of the three focal questions, and ensure its initial correct use.

On the top of these six areas of knowledge there will be a geospatial set, to allow integration of biological applications in a spatial context (*i.e.*, biogeography or phylogeography). Furthermore, to handle all the problems linked to format and congruity of newly-formed workflows, a shim service set taken from Taverna[2] and the EDAM ontology[3] will be used. Figure 1 shows a conceptual scheme of BioVel.

We invite all interested developers to become friends of BioVeL, to participate in the workshops and to follow the progress of the project from now onwards. We think that a large social component of the project will both facilitate interaction and feedback among developers and user scientists, and increase the impact of research on society.

References

1. Gil Y, Deelman E, Ellisman M, Fahringer T, Fox G, Gannon D, Goble C, Livny M, Moreau L, Myers J (2007) Examining the Challenges of Scientific Workflows, *Computer* **40**, 24-32.