Letter from the Coordinator

One Year in: From Seeds to Buds

In the first newsletter I wrote about how the seeds of the BioVeL project were sown. Well, we’re one year into the project now and those seeds have well and truly germinated.

We started by asking our scientists what they felt would be useful to work on – within the constraints of our chosen scientific themes around invasive species, carbon dioxide sequestration and ecosystem functioning and valuation. They quickly pinpointed an old biodiversity and climate change modelling scenario that they felt could have more widespread interest if re-implemented as a workflow. (continued on page 2)

A Workflow for Data Refinement

by Cherian Mathew and Anton Güntsch

Current trends in biodiversity research indicate an increasing dependency on synthesized species data as well as stable occurrence data. Consequently, the availability and quality of such type of data takes on an important role in research experiments. In addition to the traditional small database applications and spreadsheets, a variety of information systems (e.g. GBIF, Catalogue of Life, BioCASE) have emerged as biodiversity data service providers. In this environment of disparate data resources, a considerable amount of time and energy is spent in the mobilization, cleaning and selection of data. This has a direct impact on connecting the resulting data with analytical tools which also have their own specific criteria.

As part of the BioVeL project, we have built a “Data Refinement Workflow” which streamlines the data preparation process by integrating the above mentioned working steps into a Taverna workflow. (continued on page 3-5)

A Workflow for Matrix Population Modelling

by Maria Paula Balcazar and Gerard Oostermeijer

The aim of the Matrix Population Modelling Workflows is to provide an efficient and fast method to analyze demographic data sets or databases. The workflows have been designed to accept input data in a text format, but in the near future they will also accept various other formats (e.g. data retrieved from databases). The output is easily and rapidly exportable to other software programs. (continued on page 6-7)
Letter from the Coordinator, continued

That spawned not only a general purpose ecological niche modelling workflow, but also a workflow for visualizing and selecting spatial data and a data refinement workflow. We will soon be releasing the last of these as we think it’s of high interest to many scientists working with “dirty data.”

Our scientists also pinpointed a simulation model, Biome-BGC which when surrounded by supporting workflows can meet needs in both sequestration and ecosystem functioning studies. Families of workflows are being developed to support stage-based matrix projection modelling of plant and animal populations, analysis of traits in marine microbial ecosystems based on metagenomics sampling, and a phylogenetics pipeline for the taxonomic characterization of bacterial and fungal communities. This scientifically focussed development, which is already leading to draft papers, is underpinned by a wide range of ‘backroom’ technical development. We have to ensure we have the right robust services in place. We are making additions to the Taverna Workflow Management system to more easily support user interaction with workflows and, ultimately to make it easier to use workflows generally. We have begun running familiarisation and awareness raising workshops and we begin to think about establishing our helpdesk to support the first workflows when they are publicly released around the turn of the year.

Generally speaking, we perceive that our work in our first year is well received in the community, and that it can fill a gap many researchers have. We receive numerous unsolicited enquiries from people interested in what we have demonstrated, and we will be responding to each of those positively in the coming weeks as we start to make workflows more publicly available through the specially branded BioVeL pages within myExperiment.

My role in all of this is to arrange the buds and tie them to the trellis, to liberally sprinkle fertiliser and to pick out the weeds. So far, we are weed free.

Alex Hardisty
29th September 2012

Become a «Friend of BioVeL»

«Friends of BioVeL» are members of projects and organizations that support the objectives of the BioVeL project to deploy robust (Web) services for biodiversity analysis and workflow, and to encourage adoption of the workflow approach towards in-silico processing and analysis of biodiversity data. «Friends of BioVeL» are those individuals, projects and organizations that are not formal partners in the BioVeL project consortium but with whom BioVeL wishes to establish strong two-way mutually supportive relations.

To become a «Friend of BioVeL», please write contact@biovel.eu.

Beta Version:

When a new product, such as a software program or a web site, is created or upgraded, there are several stages in development before the manufacturer sells it commercially. The first stage is an “alpha” version, in which software testers within the company test and debug a program as necessary. A “beta” version is the name for the second stage, in which it is often released for free or at a reduced price to a group of external users for additional testing.

The beta version of a software release is considered to be a preview; though it may include many standard features, it is not yet ready for wide release. During this phase, the developers collect feedback from users about the product’s functionality, including what they like and what should be changed before its wide release. A beta version of a program can be either “closed,” which is limited to a specific group of users, or “open,” which is available to the general public. During this testing, developers might release numerous versions of a program, including improvements and bug fixes with each iteration.

Become a «Beta Version» tester

write contact@biovel.eu
Situation

Various forms of information systems which store, aggregate and provide access to taxonomic data are constantly under development and ever increasing in number. In reaction to this development, software applications which consume and analyze taxonomic data are being built to facilitate the processing of large amounts of data produced by the above mentioned data providers. Traditionally, scientists have been constructing ‘conceptual’ workflows, consisting of multiple individual components, which are processed separately. This method requires the user to understand the various technologies involved and manually stitch together the various parts of the workflow. In dealing with taxonomic data in particular, there is an urgent need to streamline the development cycle of such workflows to ensure that technical details (like access to various data sources, data format transformations, etc.) are handled in an automated manner to allow the user to focus on the work of processing, cleaning and selecting data.

Workflow Description

The aim of the Taxonomic Data Refinement Workflow is to provide a streamlined workflow environment for preparing observational and specimen data sets for use in scientific analysis on the Taverna platform.

The workflow has been designed in a way that,
• accepts input data in a recognized format, but originating from various sources (e.g. services, local user data sets),
• includes a number of graphical user interfaces to view and interact with the data,
• the output of each part of the workflow is compatible with the input of each part, implying that the user is free to choose a specific sequence of actions,
• allows for the use of custom-built as well as third-party tools applications and tools.

Currently, the data refinement workflow is made up of three distinct parts:

1. Taxonomic Name Resolution / Occurrence retrieval: is the phase when users can resolve a list of scientific names using taxonomic checklists. This process results in the retrieval of taxonomic information related to each scientific name, including related synonyms as well as other concept information like rank, classification, etc. The resulting information can then be used to retrieve occurrence data or saved as is. The synonym expansion and occurrence retrieval are built on generic frameworks which allow for the inclusion of multiple sources.
Research

A Workflow for Data Refinement, continued

2. **Geo-temporal data selection:** is the phase when users select (or exclude) data records according to different (and unpredictable) geographical or temporal criteria. Geographical selection could be done by drawing polygons, circles, rectangles, etc., on a map as well as filtering data based on geo-markers (e.g. country, latitude/longitude). Records relating to specific time periods can also be isolated using time-based filtering. The web-based ‘BioSTIFF’ client provides these functionalities. See page 9 for a BioSTIFF map.

3. **Data quality checks / filtering:** is the phase when a set of data quality and data integrity rules are applied on the selected data. This allows users to perform data-specific cleaning and filtering. This phase is currently being developed with ‘Google Refine’ as the central interface for accessing the various local as well as external functionalities.
Research

A Workflow for Data Refinement, continued

USE CASES

The Data Refinement Workflow has been used in the following studies:
- Mapping Potential Distributions of Invasive Species in the Baltic Sea using Data from GBIF, Ecological Niche Modeling and Automated Workflows
- A Taxonomic Data Refinement Workflow for comparison of recent species observations with historical records - a case study of the Swedish benthic fauna (see study below)

Meta-genomics Data Refinement

A benchmark exercise was done on the Data Refinement Workflow (Google Refine component) using two data files comprising 11,500 species observation records.

The data sets were:
1. Historical data set with 7,400 species observation data from the Swedish West coast, collected between 1921–1938
2. Recent data set with 4,100 species observation data from the Swedish West coast, collected between 2003–2009

<table>
<thead>
<tr>
<th>TASK DESCRIPTION:</th>
<th>EXCEL data refinement</th>
<th>Google Refine data refinement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RECENT DATA SET (4,100 SPECIES RECORDS)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removed locations according to depth</td>
<td>4 h</td>
<td>20 min</td>
</tr>
<tr>
<td>Removed difficult groups</td>
<td>2 h</td>
<td>5 min</td>
</tr>
<tr>
<td>Removed hydroids from Cnidarians</td>
<td>5 h</td>
<td>30 min</td>
</tr>
<tr>
<td>Fixed Annelides to only include Polychaetes</td>
<td>6 h</td>
<td>40 min</td>
</tr>
<tr>
<td>Clustered replicates</td>
<td>-</td>
<td>30 min</td>
</tr>
<tr>
<td><strong>HISTORICAL DATA SET (7,400 SPECIES RECORDS)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removed locations according to depth</td>
<td>2.5 h</td>
<td>10 min</td>
</tr>
<tr>
<td>Removed difficult groups</td>
<td>3 h</td>
<td>10 min</td>
</tr>
<tr>
<td>Fixed Annelids to only include Polychaeta</td>
<td>6 h</td>
<td>40 min</td>
</tr>
<tr>
<td>Removed all hydroids from Cnidaria</td>
<td>5 h</td>
<td>30 min</td>
</tr>
<tr>
<td>Fixed Arthropods to only include Crustacea</td>
<td>5 h</td>
<td>40 min</td>
</tr>
<tr>
<td><strong>SUM</strong></td>
<td>38.5 h</td>
<td>4.25 h</td>
</tr>
</tbody>
</table>

photos © Matthias Obst
The aim of the Matrix Population Modelling Workflows is to provide an efficient and fast method to analyze demographic data sets or databases (see below). The workflows are based on published R-scripts and executed on the Taverna platform using an Rserve service deployed by BioVeL. They have been designed to accept input data in a text format but, in the near future, they will also accept various other formats (e.g. data retrieved from databases). The output is easily and rapidly exportable to other software programs.

The projection matrix construction and analysis (mega) workflow comprises seven distinct components:

1. **Construction of a density independent stage matrix:**
   In this first step, the user constructs a stage-based projection matrix. The input data is a text file that specifies for each individual plant or animal, and for each study/census year, its life stage and reproductive output (the number of offspring). It also gives the recruitment, to identify the individuals new to the population. The output of this step is a stage-based projection matrix, that can be used as input for the next step. The Taverna workflow is shown in the diagram below.

2. **Analysis of the matrix (eigen analysis):**
   The results of a matrix population analysis (a.k.a. eigen analysis) is a set of statistics that describe the demographic features of the population during the studied time period:
   a) **Lambda**, or dominant eigen value, gives the population growth rate during the time period that is studied. If \( \lambda = 1 \) the population was stable, if \( \lambda > 1 \) the population increased, if \( \lambda < 1 \) it decreased.
   b) The **stable stage distribution**, which describes the relative proportions of individuals in each stage in the population when it grows at the dominant lambda.
   c) The **sensitivity matrix** gives a matrix of values that describe the relative response of lambda to small changes in each of the matrix elements. A high sensitivity value of a certain transition between two life stages means that changes in this transition will have a strong effect on the population growth rate.
   d) The **elasticity matrix** is the same as the sensitivity matrix, but the values are scaled to sum up to 1, allowing for easier comparison between different elements.
   e) The **reproductive value** describes to what extent a plant or animal of a specific life stage is expected to contribute to the next generation.
   f) The **damping ratio** can be considered as a measure of the intrinsic resilience of the population, describing how quickly transient dynamics decay follows a disturbance.

   The demographic statistics described above are a function of the vital rates (i.e. survival, growth and fecundity), which are affected by the genetic composition of the population and environmental variables.

3. **Fundamental matrix:**
   Calculations of age-specific survival rates from stage-based projection matrices. The output includes the mean, variance, and coefficient of variation (CV) of the time a hypothetical individual spends in each stage class, and the mean and variance of the time to death.
   1. \( \mathbf{N} \): fundamental matrix or mean of the time spent in each stage class
   2. \( \mathbf{var} \): variance is the amount of time spent in each stage class
   3. \( \mathbf{CV} \): coefficient of variation of the time spent in each class
   4. \( \mathbf{meaneta} \): mean time to death, or the life expectancy, of each stage
   5. \( \mathbf{vareta} \): variance of time to death

4. **Stage vector projections:**
   Creates a line diagram of the short-term population dynamics showing the convergence to the stable stage distribution using iterative multiplication of the projection matrix with the stage vector.
Research

A Workflow for Population Modelling, continued

5. Generation time ($T$):
The time $T$ required for the population to increase by a factor of $Ro$ (net reproductive rate).

6. Net reproductive rate ($Ro$):
Calculates the generation time, the mean number of offspring by which a newborn individual will be replaced by the end of its life and thus the rate by which the population increases from one generation to the next.

7. Confidence interval of lambda:
Calculates the confidence interval around lambda, to establish whether or not it differs significantly from 1. This is done by creating a large number of projection matrices and corresponding lambda’s by resampling, using a multinomial distribution for transitions and a log normal distribution for fecundities.

The next series of workflows will continue with more statistical analyses of matrix projection models as well as analyses of extinction risks, causes of variations in lambda’s. In collaboration with the Max Planck Institute, we will also integrate IPMPack, a toolkit for the creation and analysis of Integral Projection Models, which are a relatively recent novel development in population modelling.

In the near future, BioVeL workflows will also be able to retrieve matrices for analysis from the largest database in the world containing matrix projection models, integral projection models and life tables for 800 plant and 400 animal species. This database, COMPADRE (COMParative Demographic REsearch) is compiled by the Max Planck Institute for Demographic Research.

For demographic research, it is important to look at all life stages, such as these tiny plant seedlings. From L to R: Hypochaeris radicata, Cirsium palustre, Caltha palustris, Drosera rotundifolia, Ranunculus flammula, Lotus pedunculatus, Salix repens and Isolepis setacea. © Gerard Oostermeijer

This graph shows the relationship between lambda (population growth rate), calculated from projection matrices, and the time since disturbance of the vegetation in the habitat of Gentiana pneumonanthe (Marsh gentian). In the early stages after a disturbance that opens up the vegetation structure, there was ample opportunity for seedling recruitment and the populations grew rapidly, indicated by mean values of lambda higher than 1. After 7 years, the vegetation structure has closed so much by natural succession that light competition prevented most of the seeds from germinating and seedlings could not establish. Because of this, population growth rates (lambda’s) dropped to below one, and the population started to decline. The decline is slow, because adult plants are relatively long-lived. Population extinction would occur after 40-45 years without new disturbance. Based on these results, it was clear what managers needed to do to get this rare and protected species to recover: open up the vegetation by sod removal, mowing or sheep or cattle grazing.
Featured researcher: Gerard Oostermeijer
Institute of Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam (UvA), The Netherlands

My passion for nature began as a little boy, when our school class went on an excursion through a coastal dune reserve. On that day, I decided to go and explore nature by myself, and have done so ever since. It only seemed natural to pursue a career in biology, so I started in 1982 as a biology student at the University of Amsterdam. Although I had a broad interest in nature, I gradually developed into more of a botanist with an initial focus on vegetation ecology. During my MSc research projects, I discovered plant population biology through the lectures given by Dr. Adri Sterk, one of the first plant demographers in The Netherlands. After reading John Harper’s book “Population biology of Plants,” I was hooked, and knew this was going to be my specialisation. Fully inspired, I wrote a proposal for a PhD-project on the population biology of threatened plants. My enthusiasm for the subject got me my PhD degree cum laude in 1996. After my PhD, I did two years of applied conservation research on butterflies for the foundation ‘Dutch Butterfly Conservation,’ but returned to botany when I was offered a position as Assistant Professor at the UvA. I am the chairman of our new foundation Science4Nature, that aims to apply population biology and genetics in practical conservation of threatened species (see www.science4nature.nl). When time permits, I am also an avid nature photographer (see www.oostermeijerphotography.nl).

Population viability models
Following the call of pioneering conservation biologists such as Michael Soulé, I was among the first cohort of researchers to study the effects of reduced population size and increased isolation on different components of population viability. These components were demography (population dynamics), reproductive biology (pollination ecology) and genetics (inbreeding depression, breeding system and genetic diversity). It was the complexity of all interactions between these components that eventually set me on the path of population modelling. This was scary at first, because mathematics had never been my strongest discipline, and I had never had any matrix algebra at school. After struggling for a while to understand Caswell’s rather mathematical textbook on the subject, I finally managed to turn my complex demographic monitoring data into comprehensible matrix projection models, realizing I now had an extremely powerful tool, with which I could not only understand the environmental drivers behind the dynamics of my study populations, but also simulate population dynamics under a variety of environmental scenarios. The matrix modelling toolkit can be used to find the weak links in a species’ life cycle. This knowledge can be used for species conservation management, but also to hit invasive non-native species where it hurts most.

After doing a lot of research on the population ecology and viability of plant species in The Netherlands, Costa Rica, Canary islands, United States, Australia and Taiwan, my interests have expanded to also include metapopulation processes at the landscape scale.

BioVeL Population Modelling
The basic idea of our contribution to BioVeL is to turn the conversion of demographic monitoring data on individual plants and animals into matrix projection models or integral projection models into a breeze. Being created in collaboration with active researchers in the field, the BioVeL Population Modelling service set will standardize the construction and analysis of these models. This will facilitate the peer-review process of papers on this type of models, as well as support collaborative projects on comparative demography. My experience is that streamlining datasets for analyses during such projects is always a time-consuming and tedious process. BioVeL’s services will greatly improve this process.

For more information on BioVeL population modelling workflows, please see pages 6–7.
Introducing the Consortium

Featured researcher: Vera Hernández Ernst
Fraunhofer Gesellschaft, Institute for Intelligent Analysis and Information Systems (IAIS), Sankt Augustin, Germany

I’m working as a researcher in the Knowledge Discovery Department of the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS). In our working group we have a long expertise in the visual exploration and analysis of spatial and temporal data. On one hand we work on intuitive user interfaces for different user groups, from experts to novices, to provide them with powerful tools for visual interactive analysis of data, making information accessible and consumable. On the other hand we use server and business intelligence technology to elaborate models, extract hidden knowledge from the data and perform computations of large and multidimensional data sets. I contribute my expertise to both fields. I have a master degree (Dipl. Inf.) in computer science with a specialization on the development of Computer-Aided-Design (CAD) systems and Geographic Information Systems (GIS). Before joining Fraunhofer in 2004, I had been working in the conception and development of database applications and geospatial systems for five years at two different consulting agencies. One of my major interests and knowledge in this field is the application of standards in the development of GIS tools, and in particular in the Standards from the Open Geospatial Consortium (OGC). I’m also a strategic member of OGC, on behalf of Fraunhofer IAIS.

Although my group has participated in different European and National projects in biodiversity, like the European Forest Information System (EFIS) and Networking activities around it (NEFIS, 2003–2005), my first contact with biodiversity scientists was during my participation in the LifeWatch project (2008–2010) where I worked on the conception of a Reference Model for the ICT infrastructure. Since then, the visualization and analysis possibilities embodied in the different use cases and the relevance of the results for the life of our planet have fascinated me. I’m responsible for the coordination and implementation of the Visualisation and Geospatial Service Sets and for facilitating the seamless flow of data through so-called shim services or transformers. With BioVeL, we contribute to the Biodiversity Spatio-Temporal Interactive Interface (BioSTIF) tool. Our spatial web services provide a mapping tool that allows the representation, computation, and selection of biodiversity data which can be included in any workflow (see page 4).

Visualisation of occurrences of the Gamarus tigrinus and the projected data used during a workflow for computation of ecological niche modeling in BioSTIF
Project Progresses

The First 2012 Deliverables

A Major Report:
With D2.1, BioVeL delivered a major review of three main areas of biodiversity research. The entire report is available by clicking here.


This report reviews the state-of-the-art and research trends for each of the three science areas selected for the BioVeL project. The three science areas are:
- a) biospheric carbon sequestration;
- b) ecosystem functioning and valuation; and
- c) invasive species management.

For each science area the report addresses:
- i) the science-policy context;
- ii) the main scientific questions being addressed in the area;
- iii) the most relevant research groups, networks and projects;
- iv) the existing models and methods;
- v) data source requirements;
- vi) the typical and most challenging workflows; and
- vii) web service requirements.

The report concludes that the needs of biodiversity and ecology sciences for new leading-edge integrative IT tools are growing rapidly, while existing, reliable and open web services and workflow management services to fulfill these needs are still scarce. The planned and on-going developments of the BioVeL project, demonstrating the success of the virtual laboratory approach, will go a long way towards meeting this need, and will open new perspectives for biodiversity analysis. However, several gaps will remain unbridged and there are some constraints as well.

List of Delivered Deliverables (downloads)

D2.2: Use cases, workflows, and sprints in BioVeL September 2011–August 2012 (Sept. 3, 2012)
D3.1: Prioritised list of service sets (April 30, 2012)
D4.1: Communication Plan (Feb. 24, 2012)
D4.2: Specific Support Centre Plan (May 31, 2012)
D7.1: Report of Results of Sprints: 1 and 2 (Feb. 29, 2012)
D8.1: Service Component Management (Aug. 31, 2012)
Project Progresses

The Specific Support Centre

As workflows are burgeoning, we are actively preparing the Specific Support Centre (SSC). We strongly feel that “technical aspects” should not deter from using valuable research tools, and that users should have a positive experience with using BioVeL’s workflows. With this philosophy in mind, we want our workflows and overall services to be user friendly. To this end, the SSC works as an interface between users and the R&D of the project, and is designed to provide effective, timely and friendly users’ support.

The SSC plan of action is progressive and led by the release of the workflows as well as by the needs of our target audience. We are addressing two communities, biodiversity researchers and informatics engineers, and they need to communicate with each other. This is why our first action has been to create a glossary, currently posted on the web site. Frequently Asked Questions will follow soon.

One of the SSC’s goals is to respond to direct inquiries, and we already have a tracking issue system in place (using JIRA, a project tracker software.) “Behind the scene,” the consortium experts will respond to the inquiries, discuss between themselves, share information, and the SSC will process and deliver the response.

Along with the beta versions of the workflows (see pages 3–7), we will also develop tutorials and manuals that too will be tested and improved. These packages might include videos of demonstration and other user-friendly media commonly used in e-learning.

The SSC is also collaborating with LifeWatch’s Service Centre for compatibility and easy transfer. Already we have agreed to use Moodle, an e-learning platform. LifeWatch will be responsible for maintaining this platform in the future while BioVeL’s experts will keep developing the support contents.

Collaboration accelerating

LifeWatch and BioVeL are working together in the creation of their respective SSCs. BioVeL’s SSC is planned and designed so it will easily transfer onto LifeWatch’s platform. Collaborative tools are being put in place and by mid-2013 we should see the first fruits of this with integrated e-learning capabilities.

With the number of workflows available growing, so will training workshops. These will be organised by the SSC at conferences or on demand to present BioVeL services capabilities. The training workshops will be hands-on sessions, teaching how to select a workflow, use it or alter it if necessary, and run it.

We expect that these activities will prove helpful in contributing to the uptake of BioVeL services. At all times we welcome your suggestions and feedback by writing contact@biovel.eu.

MILESTONES

MS1: Setting up requirement review process
MS2: Initial Joint Workshop
MS3: Workshop with Service Providers
MS4: Website published (www.biovel.eu)
MS5: Requirements for WP6 tasks
MS6: Specific small workshop with selected Service Providers
MS7: Workshop with Service Providers (see page 11)

The Milestones reports are available at www.biovel.eu
Events: Consortium Meetings

**MS6 Workshop**, Gothenburg, Sweden, May 10–11, 2012

Developing inter-disciplinary methods for conservation biology and ecosystem research — taxonomic and phylogenetic workflows in biodiversity research

This workshop organised by WP2 leader Matthias Obst, was an important milestone of the project (MS6). Reflecting upon the “agile nature” of the BioVeL project and the considerable positive momentum of the project’s first few months, the consortium decided to hold a large and publicly visible workshop. It was the first meeting where we invited scientists from both the biodiversity fields and informatics backgrounds. The workshop attracted 53 participants, members of the BioVeL consortium (19), BioVeL liaison partner programs and prospective friends (25), and a number of high-profile external scientists (9).

The main, revised objectives of the MS6 workshop were:

• To foster cooperation in the biodiversity science community, principally between project partners and close friends of the project as originally indicated by Letters of Support at the proposal stage; and,

• To identify workflows and use-cases that integrate analytical components from the field of taxonomy and phylogenetics.

The program contained a general introduction to BioVeL and specific presentations of its work on workflows and showcases from the taxonomic, phylogenetic and ecological niche modelling service sets. Friends of BioVeL and external scientists gave presentations on their perspectives, needs and current work. These were followed by breakout discussions and feedback collection from the participants. Important issues, such as services and applications relating to the field of species distribution analysis, synonym resolution, historical biogeography, data cleaning and refinement issues, biodiversity richness analysis, phylogenetic diversity analysis, comparative phylogeography, vicariance analysis, and molecular taxonomy were all discussed. The information collected during the workshop will be used to inform agile planning for the next few months of the project.

It was also a great opportunity to strengthen our relationships with Swedish and Belgium LifeWatch, GBIF, ViBRANT and agINFRA.
Events: Consortium Meetings

MS7, BioVeL Technical Summer Workshop

Schloss Birlinghoven, near Bonn, Germany, July 18–19, 2012

The BioVeL Technical Summer Workshop 2012 was hosted by Fraunhofer IAIS, at Schloss Birlinghoven, near Bonn, Germany, July 18–19, 2012.

The workshop brought together Service Provider partners from the project to discuss technical issues surrounding the provision of robust Web services, to find technical solutions that ultimately can be adopted by all Service Providers offering services to BioVeL. In particular, the workshop concluded on the principles for

i) controlling access to services and ii) handling storage of and access to data needed by services.

Controlling access to services will be aligned with LifeWatch principles, based on the "RM-OA" approach to authentication and authorisation. This means BioVeL will deploy specific Web services for authentication and authorization that will help to protect the services of all Service Providers.

Data management in service-based workflows implies that users are able to gain access to that data. Interference between data of multiple instances of the same workflow or between multiple users has to be taken into account. To meet the needs of BioVeL, the Workspace Access service will be specified and deployed. This will give a common way for users, workflows, and services to interact with the relevant data.

BioVeL Second Annual Meeting

Aix-en-Provence, France, September 3–7, 2012

Over 25 consortium members and four LifeWatch guests participated in BioVeL’s second annual meeting hosted by the French Foundation for Research on Biodiversity (WP4 leader) at its CEntre for the Synthesis and Analysis of Biodiversity (CESAB), near Aix-en-Provence, France, September 3–7, 2012.

This week-long meeting allowed to celebrate the great achievements of the consortium’s first year of activity: the first workflows will be released this Autumn! (see pages 1 and 3–7.)
Events: Consortium Meetings

BioVeL Second Annual Meeting, continued
Aix-en-Provence, September 3–7, 2012

Much of this workshop was dedicated to the release of these workflows that require repeated tests as well as a thorough readiness in implementation to include a user-friendly interface, packaging and annotations, and a specific support activity (see page 11).

The Second Annual Meeting was also a great opportunity for more outreach, dissemination, and collaborations. This meeting allowed for further collaborations with LifeWatch, and we acquired a better understanding of each others’ activities. We also discussed practical aspects of the implementation of BioVeL’s Specific Support Centre so it will easily transfer into LifeWatch’s own SSC.

While in Aix, the consortium also met with researchers from the Tour du Valat, located in Camargue, a fragile ecosystem known for its flamingos' breeding grounds. The Tour du Valat is a wetlands conservation research centre with an emphasis on birds and Mediterranean wet lands ecosystems preservation. These researchers expressed an interest in BioVeL workflows and further conversations are open.

Other collaborations will also be considered with FRB CESAB’s researchers.

Finally, the consortium also prepared for its annual report to the European Commission (November 8, 2012) as such large inter-disciplinary project requires a fair amount of orchestration, information reporting, and planning.
A call to action: Consultation on a 10-year vision for biodiversity informatics

We need funding to pursue biodiversity research, but funding is under huge pressure. One area for which there is potential is in biodiversity informatics because we are witnessing a paradigm shift in the way we handle data. Now it is expected that data will be openly accessible, presenting huge opportunities if only the data can be found when needed.

There is a strong need for a community voice to present the funders with a set of priorities. We have successfully built a number of infrastructural components, such as INSDC, GBIF, Catalogue of Life, but we lack the bridges to allow users to integrate elements from these resources to address a simple question.

With the encouragement of the European Commission, the FP7 funded projects BioVeL, ViBRANT, and agINFRA have carried out a public consultation on the future for biodiversity informatics. More than 60 people have contributed to a White Paper on “A decadal view of biodiversity informatics: challenges and priorities.” This is about to be published, October 2012 and will be available on our website (www.biovel.eu).

If you would like more information on this action, please write contact@biovel.eu.

Events: Upcoming

1st EUDAT Conference
Barcelona, SP, October 22–24, 2012
This international event will bring together data infrastructure providers and practitioners from across the globe to discuss current data infrastructure challenges and solutions, with a focus on interoperability, cross-disciplinarity and cross-border collaboration.

MS9, BioVeL Technical Workshop
Copenhagen, DK, November 2012
The objective of this workshop is to discuss further development and hardening of the web services of important data providers and portals. This workshop will be particularly relevant to the data refinement workflow. The consortium will also invite participants from data providers’ institutions.

Report to the European Commission
Brussels, BE, November 8, 2012
BioVeL consortium will report to the European Commission on its first year of activity.

A Year of Events

Over the last year, consortium members have participated in or presented at a total of 35 events. These include 5 events that where organised for the consortium itself and to which selected guests were occasionally invited.

The consortium participated in 30 conferences, workshops, and other meetings relating to biodiversity research and/or IT development. These were dissemination activities sometimes associated with actual demonstrations of workflows.

The consortium was also involved with 11 events to coordinate its activity with other e-infrastructures, including the planning of BioVeL’s SSC collaboration with LifeWatch. A community White Paper public meeting was also held in July 2012 (see blue box below).

Looking ahead, the calendar is full of upcoming events in the field of biodiversity research and in the computer sciences area. BioVeL partners are planning to participate in these events and offer demonstrations of our workflows.

If you are interested in meeting us at a conference, please write contact@biovel.eu.

See you there!