



DELSA Endorsed Projects & Working Groups

Working Groups

Social Networking Platform for Tool Brokering/Community Building and Matchmaking

Lead: Brian Glanz (Open Science Federation, Seattle, Washington) and Courtney MacNealy-Koch

Goal: Open a dialog and an organizational effort to build a social networking platform to broker bioinformatics tools. This project would encourage community engagement by crowdsourcing, accelerate discovery by making tools more accessible, and through community ranking, more trustworthy. It would also build community and connect people and resources. Possible examples are VIVO and LinkedIn. Could develop matchmaking 20 questions to determine individuals' skills, interests, tools, review favorites. Could point to publications, tools or resources.

Data Set Accessibility Project

Lead: Corinna Gries (University of Wisconsin-Madison, Madison, Wisconsin)

Goal: Make high quality life sciences data broadly available, traceable and usable. Work to define issues such as: Curation, Sustainability, Rapid growth in data volume, Data provider incentives, Non-trivial processing on the data in the repository, Limited bandwidth from the open Internet to clouds, and Security.

Training Data Scientists

Lead: Geoffrey Fox (Indiana University, Bloomington, Indiana)

Goal: Train new and established scientists to enable more effective use of big data and its cyberinfrastructure. in data enabled science culminating in a certification similar to Microsoft or Cisco certification or existing scientific computing or computational science certificates/curricula. Need to evaluate existing resources such as: UW eScience classes, OGF Grid Computing certificate, and XSEDE HPC University... Possible approach is to focus on particular life science subdomains. Possible solutions are MOOCs.

DELSA International Outreach

Lead: Vural Ozdemir (McGill University, Canada)

Goal: To increase DELSA's international outreach and knowledge translation activities to developing countries specifically, and resource-limited settings more broadly, be they in developing or developed countries. The outreach activities will be explored and achieved through a 'push' and 'pull' hybrid model of innovation diffusion: by engaging with both upstream knowledge generators and expert communities and downstream knowledge-end users and citizens, in the spirit of distributed data intensive life sciences R&D and collective innovation that DELSA endorses.

Projects

Environmental and Medical Sciences Workshop

Lead: Neil Davies (University of California, Berkeley, Berkeley, California)

Goal: To organize a workshop to bring researchers in the medical and environmental sciences together as a way to advance transdisciplinary research.

Deliverable: A one day workshop bringing together medical and environmental scientists

Global Protein Atlas

Lead: Jack Gilbert & Folker Meyer (Argonne National Laboratory, Argonne, Illinois)

Goal: For all the meta-genomes and genomes that are available cluster at the protein level and annotate... MG-RAST, CAMERA, MOPED, PSU, etc... The goal is to characterize all the proteins and answer on the question: what protein is expressed in what organism, what disease, what tissue, what condition, what environment, and in what concentration?

Deliverable: Based on current large scale projects such as Earth Microbiome Project and Human Microbiome Project, we will analyze samples from diverse communities using meta-genomics and meta-proteomics to produce a Global Protein Atlas.

Internet2 Application

Lead: Michael Sullivan (Internet2, Washington, D.C.)

Background: Internet2 is an advanced not-for-profit networking consortium developing revolutionary Internet technologies and leveraging a high-performance network (<http://www.internet2.edu/>). It is currently being adopted by NLM. It has three components: 1) connect pilot place to Internet2; 2) Deploy Science DMZ at the pilot place; and 3) Perform routine exchange of BigData. It is a dedicated data transfer mode to enable fast data transfer mode.

Goal: Create scalable process to connect entities (Research institutes, Universities, Healthcare Providers, and Global Governments) to Internet2.

Conceptualization and Analysis of the 3D Virtual Cell

Lead: Phil Bourne (University of California, San Diego, San Diego, California)

Goal: Biology occurs at multiple scales, from molecules to cells to organisms to environments, which have traditionally been studied separately. Translational biology is now seeking to bring these disparate scales together to achieve a new level of understanding. This understanding requires new algorithm development and analysis tools that can operate on an increasingly large, diverse, complex and widely distributed body of digital biological data. Hence, translational biology is faced with cyberinfrastructure challenges related to data accessibility, software development, software reuse, and sustainability on a scale not seen before. The prior *modus operandi* of software development – let 1000 independent flowers bloom – will no longer scale in studying translational biology. In short, the current software development and maintenance cycle costs too much, takes too long, operates in too many silos, and is therefore unsustainable and cannot meet the increasing challenges of 21st century biological sciences.

Deliverable: We propose an S2I2 Translational Biology Center that focuses on a better understanding of complex biological structures from molecule to cell. The Protein Data Bank (PDB) can be expected to broaden its focus over the next few years, expanding to contain supramolecular complexes – already happening – and later subcellular and cellular structures derived from a variety of experimental methods. With these raw data at hand, new software tools and analysis protocols will be needed to understand and visualize the emergent *3D virtual cell*; tools for structural systems biology.

Pregnancy Atlas Use Case

Lead: Joseph Kemnitz (University of Wisconsin-Madison, Madison, Wisconsin)

Goal: Utilize DELSA and its members and connections for resources that would help the Pregnancy Atlas....The Pregnancy Atlas Consortium has an Integrative Discovery Platform that could be expanded. Help with metrics to assess Platform success.

Deliverable: Additional information for the Pregnancy Atlas such as potential collaborators, CI tools, data formats and funding opportunities. Provide files in a format that could be integrated by the Platform.

ParaMEDIC Use Case

Lead: Wu Feng (Virginia Tech, Blacksburg, Virginia)

Background: Frequent Pain Points experienced by DELSA members include ease of use issues with analysis tools and compute resources, as well as performance issues which may be due to compute problems, data management problems or data representation problems.

Goal: Use the automated, easy to use and integrated high-performance Biocomputing system (including ParaMEDIC: Parallel Metadata Environment for Distributed I/O & Computing) on a Suggested BigData challenge to show what can be done if the system was widely available.

Deliverable: BigData life sciences challenging project successfully accomplished.