

Evolutionary Data Leaping to Web 3.0: Some Highlights From NESCent's Third Hackathon

Hilmar Lapp^{1*}, Arlin Stoltzfus^{2,3}, Todd J. Vision^{1,4}, Rutger A. Vos⁵

(1) US National Evolutionary Synthesis Center (NESCent), Durham, NC, USA (2) Center for Advanced Research in Biotechnology, University of Maryland, Rockville, MD, USA (3) National Institute for Standards and Technology (NIST), Gaithersburg, MD, USA (4) Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA (5) Departments of Zoology and Botany, University of British Columbia, Vancouver, BC, Canada

<http://evoinfo.nescent.org/>
Database_Interop_Hackathon

Motivation

There is a rich and meticulously curated variety of evolutionary data available through online databases, yet most of these do not support standard formats or explicit semantics, and can only be queried through interfaces intended for humans. This presents barriers to effective reuse of the data, such as combining, connecting, and visualizing data from different sources in ways often not imagined by the original database creators. Removing these obstacles to data interoperability can facilitate a much broader utilization of the already existing wealth of data, which in turn may further spur efforts to bring an increasing amount of the evolutionary data isolated in silos or still hidden offline into an online web of interoperable data.

To promote the adoption and refinement of standards towards this vision, NESCent sponsored the Evolutionary Database Interoperability Hackathon.

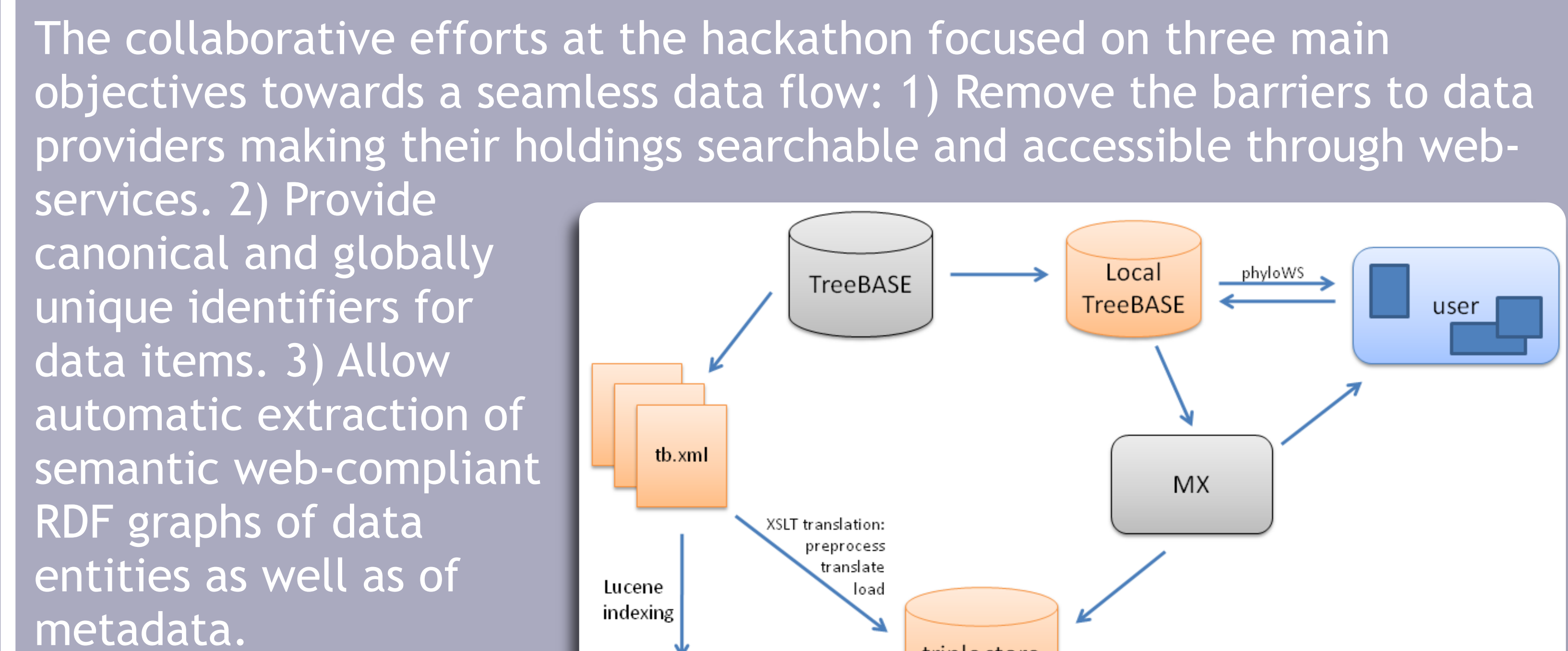
Phyloinformatics Standards

To address the interoperability obstacles, the hackathon aimed to bring to bear three emerging standards for phylogenetics data and databases:

- NeXML (<http://nexml.org>) as a formal, unified syntax that can be validated and embed arbitrary metadata
- Comparative Data Analysis Ontology (CDAO) for explicit, computable semantics of evolutionary data and analyses (<http://www.evolutionaryontology.org>)
- PhyloWS as a standard programmable interface for phylogenetic data providers (<http://evoinfo.nescent.org/PhyloWS>)

All three standards emerged from the Evolutionary Informatics Working Group sponsored by NESCent.

Data Flow and Use Cases



The combination of web-services and semantic web is often referred to as Web 3.0.

Web-Services for Data Access

One subgroup launched a generic publishing toolkit that allows data providers to make their data holdings searchable and accessible through the PhyloWS programming interface standard. Participants also created a web-service prototype for TreeBASE that demonstrates the resolution of canonical PhyloWS-based URLs for trees to NeXML documents.

Data Visualization and Mash-Up

Hackathon Participants

The event brought together 25 data and metadata experts as well as developers from a variety of online data resources, data aggregators, mash-up and visualization projects for phylogenetic data, and the developers of three emerging interoperability standards in phyloinformatics (NeXML, CDAO, and PhyloWS). The participants represented projects from areas as diverse as molecular biology, phylogenetics, systematics, paleontology, and biomedical research. The event took place March 9-13, 2009, on-site at NESCent.



Acknowledgments. The US National Evolutionary Synthesis Center (NESCent, <http://nescent.org>) is a collaborative effort of Duke, NC State University and UNC Chapel Hill, and is supported by the National Science Foundation (EF-0423641). We are indebted to the enthusiasm and energy of the hackathon participants that made the event a success.



Data Annotation With Rich Semantics

In an intensive collaboration, a subgroup of participants drafted a RDFa-compliant specification for how to embed annotation in NeXML with ontology-anchored semantics. They also created a parser and writer toolkit in the Java programming language to implement the specification. The