

Caleydo Web: An Integrated Visual Analysis Platform for Biomedical Data

Samuel Gratzl, Nils Gehlenborg, Alexander Lex, Hendrik Strobel, Christian Partl, and Marc Streit

In many scientific domains data analysis has replaced data acquisition, generation, and storage as the main challenge. This challenge stems not only from volume but also from complexity and heterogeneity of the data. Molecular biology is a prime example for this development trend. We identified six key requirements that a visual analysis platform

for biological data needs to support. Here, we present Caleydo Web, an integrated visual analysis that is designed to address these needs. In this poster we describe its architecture and give an example of how it can be used to create StratomeX.js, a sophisticated visualization technique for cancer subtype characterization.



<http://demo.caleydo.org>

Data Scale and Heterogeneity

The size, complexity, and heterogeneity of datasets like clinical, gene expression, and whole genome sequencing data are challenging with regards to access, processing and interactive visualization.

Identifier Management

Different datasets are often using annotations with different identifiers (e.g., Entrez, Ensembl) for related entities. These identifiers have non-trivial relationships (e.g., 1:1, 1:n, n:m) and granularities (e.g., chromosome, gene), leading to complex mappings.

Multiple Coordinated Views

Integrated analysis of multiple interconnected datasets can lead to new insights. Visually linking entities across different visualizations, annotation systems and granularity levels is essential.

Provenance and Coordination

A study showed that it is not possible to reproduce 90% of the findings in more than 50 cancer genomics studies [1]. Integrated provenance tracking will support the reproduction, interpretation, sharing, presentation, and communication of findings.

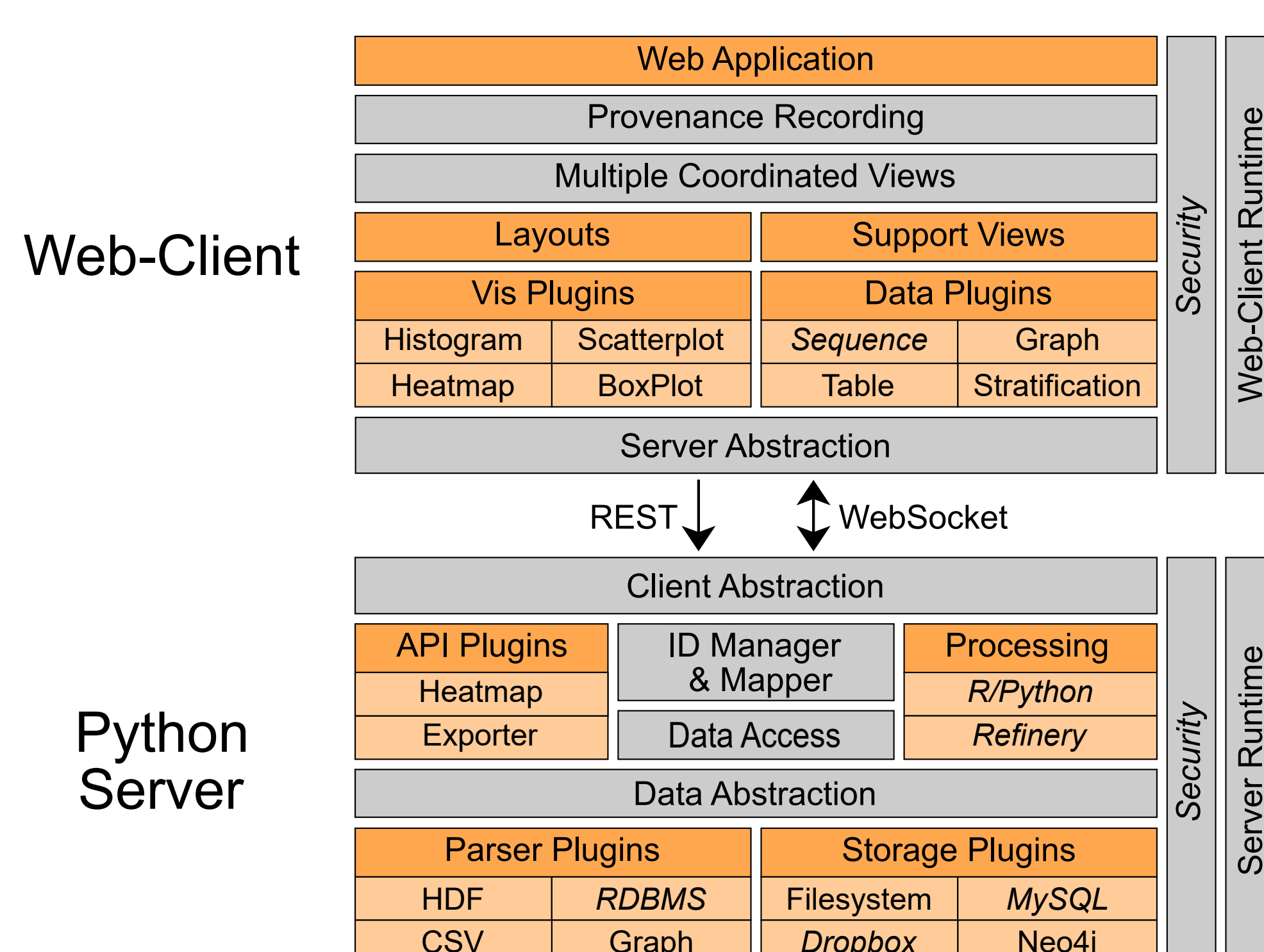
Adaptability

Environments change. New data types, storage backends, visualization techniques emerge and needs to be integrated. The creation of customized setups tailored to specific applications ensures efficiency and effectivity.

Integrated Data Analysis

Integration of algorithms, statistics, and machine learning is crucial. The interplay between algorithmic and visual analysis has to be efficient [3].

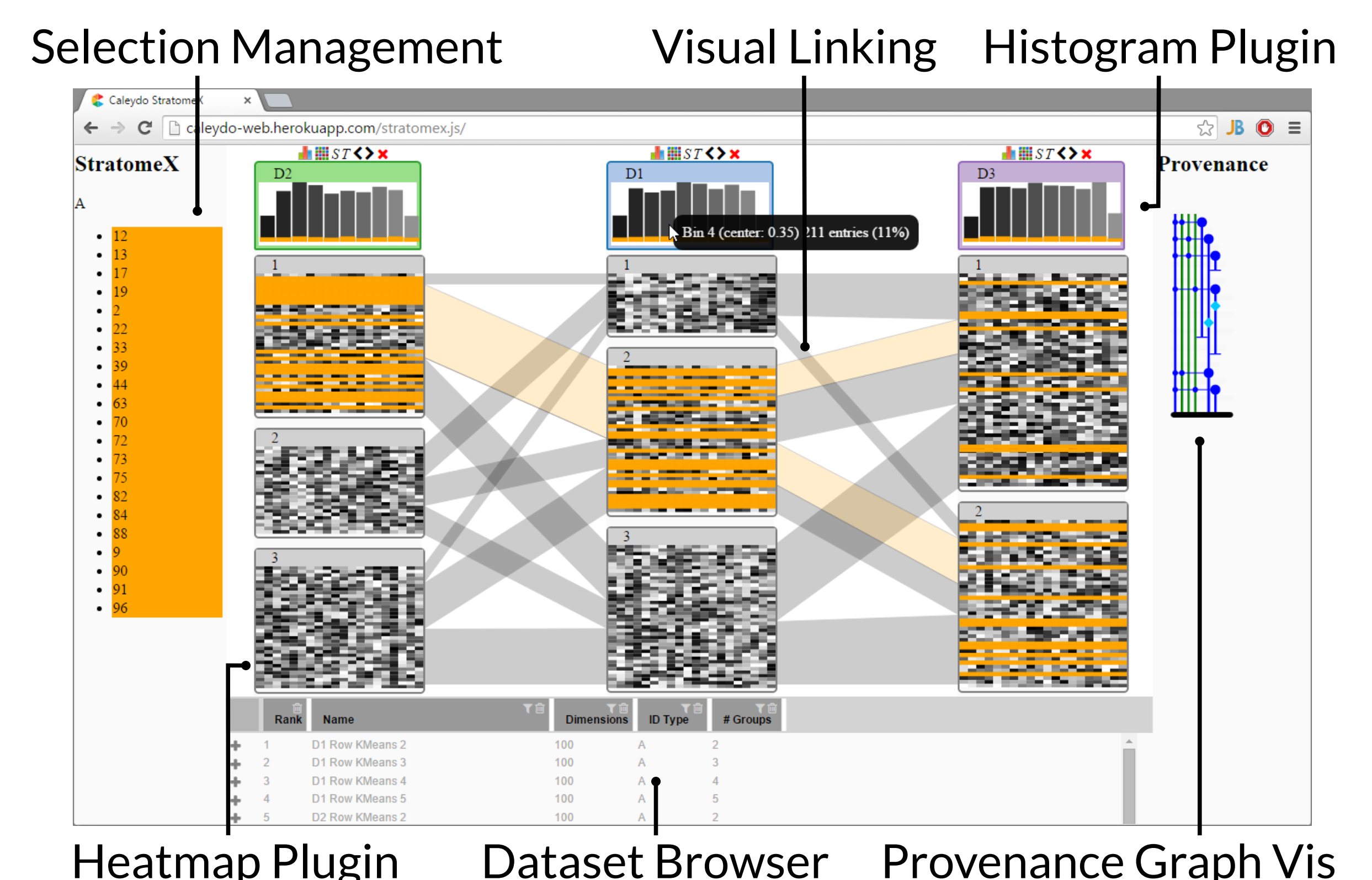
Architecture



Caleydo Web is an open-source framework under the BSD license and hosted on Github (<https://github.com/Caleydo>). The client runtime is written in TypeScript and JavaScript using HTML5. The server runtime is written in Python. Italic labels indicate work in progress. Grey boxes represent Caleydo Web's core, dark orange boxes major plugin types, and light orange boxes custom plugins.

contact@caleydo.org
<https://github.com/Caleydo>
https://twitter.com/caleydo_org
<https://www.youtube.com/user/caleydotugraz>

Example: StratomeX.js



StratomeX.js is a Caleydo Web-based reimplementation of Caleydo StratomeX [2], a cancer subtype visualization technique. Annotations indicate individual components provided by Caleydo Web.

- C. G. Begley and L. M. Ellis. *Drug development: Raise standards for preclinical cancer research*. Nature, 483(7391):531–533, Mar. 2012.
- A. Lex, M. Streit, H.-J. Schulz, C. Partl, D. Schmalstieg, P. J. Park, and N. Gehlenborg. *StratomeX: Visual Analysis of Large-Scale Heterogeneous Genomics Data for Cancer Subtype Characterization*. Computer Graphics Forum (EuroVis '12), 31(3):1175–1184, 2012.
- T. Muhlbacher, H. Piringir, S. Gratzl, M. Sedlmair, and M. Streit. *Opening the Black Box: Strategies for Increased User Involvement in Existing Algorithm Implementations*. IEEE Transactions on Visualization and Computer Graphics (VAST '14), 20(12):1643–1652, 2014.