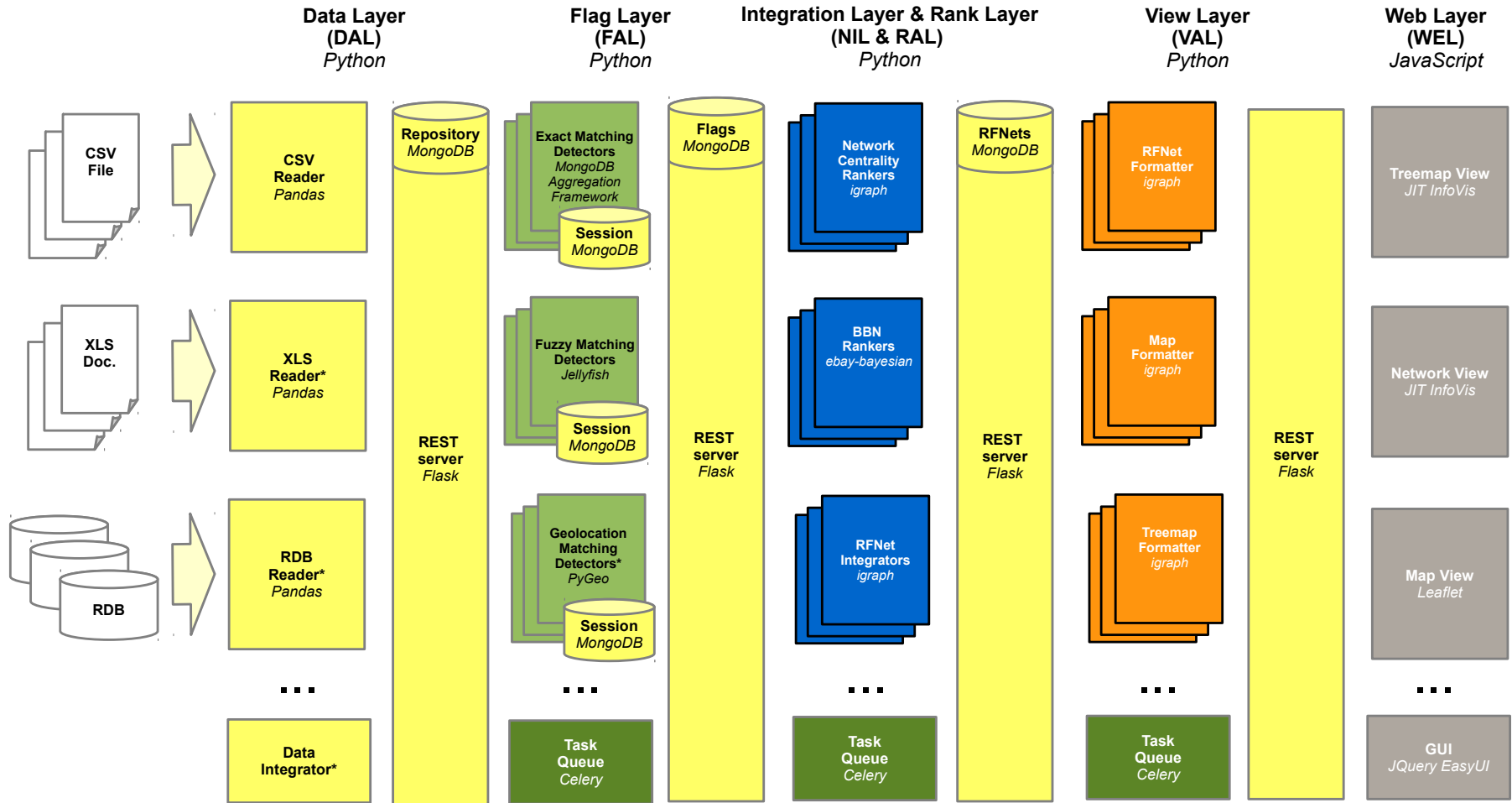# SAFARI
Architecture and Software Stack

Alberto Garcia-Robledo, Abel Sanchez, Rongsha Li,
Juan-Carlos Murillo-Torres, John Williams and Sascha Boheme
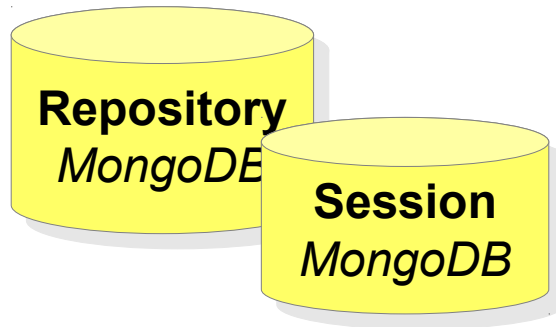
Massachusetts Institute of Technology
MIT Geospatial Data Center

SAFARI Software Architecture

Data Layer (DAL) *Python*

Flag Layer (FAL) *Python*

Integration Layer & Rank Layer (NIL & RAL) *Python*

View Layer (VAL) *Python*

Web Layer (WEL) *JavaScript*

CSV File → **CSV Reader** *Pandas*

XLS Doc. → **XLS Reader*** *Pandas*

RDB → **RDB Reader*** *Pandas*

Data Integrator*

Repository *MongoDB*

REST server *Flask*

Exact Matching Detectors *MongoDB Aggregation Framework*

Session *MongoDB*

Fuzzy Matching Detectors *Jellyfish*

Session *MongoDB*

Geolocation Matching Detectors* *PyGeo*

Session *MongoDB*

Task Queue *Celery*

Flags *MongoDB*

REST server *Flask*

Network Centrality Rankers *igraph*

BBN Rankers *ebay-bayesian*

RFNet Integrators *igraph*

Task Queue *Celery*

RFNets *MongoDB*

REST server *Flask*

RFNet Formatter *igraph*

Map Formatter *igraph*

Treemap Formatter *igraph*

Task Queue *Celery*

REST server *Flask*

Treemap View *JIT InfoVis*

Network View *JIT InfoVis*

Map View *Leaflet*

GUI *JQuery EasyUI*

* = to be developed

- MongoDB is an open-source NoSQL document database.

- JSON-style documents with dynamic schemas.

- Rich, document-based queries.

- Flexible aggregation and MapReduce data processing.

Who's using it:

REST server *Flask*

Flask

web development, one drop at a time

- Lightweight Web application framework for Python

- Microframework: it keeps the core simple but extensible.

- RESTful request dispatching.
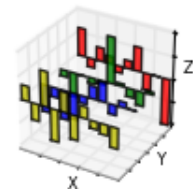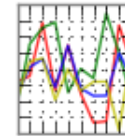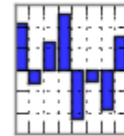
- Extensions available to enhance features as desired.

Who's using it:

Media Queries

Comiker

BATTLEFIELD 3
Development News Aggregator

CSV
Reader
*Pandas*

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

- Easy-to-use data structures and data analysis tools.

- Efficient DataFrame object for data manipulation.

- R/W between in-memory data and text files, CSV, Microsoft Excel, SQL DBs, and HDF5.

- High performance merging and joining of data sets.

Celery

**Task
Queue
*Celery***

- Asynchronous task queue based on distributed message passing.

- Tasks are executed concurrently on a single or more worker servers.

- Support for RabitMQ, Redis, Beanstalk and MongoDB brokers.

Who's using it:

igraph
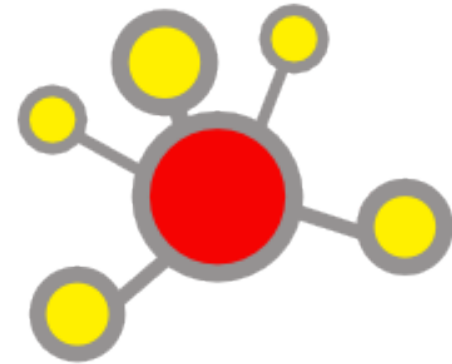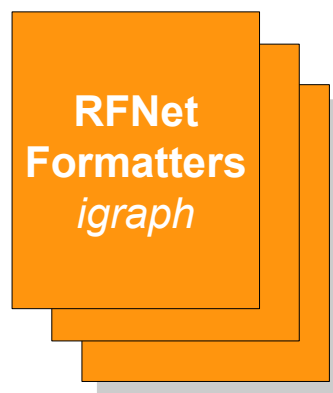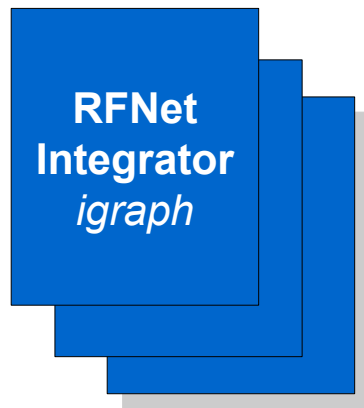
■ High performance library for complex network research and SNA.

■ Algorithms for measuring structural properties, node centrality, K-decomposition and community detection.

■ Algorithms for generating 2D/3D layouts: Fruchterman-Reingold, Kamada-Kawai, Reingold-Tilford, Distributed Recursive Layout, etc.

BBN
Rankers
igraph

eBay / **bayesian-belief-networks**

- Pythonic Bayesian Belief Network package.

- Exact inference on BBNs specified as pure python functions.

- Discrete and continuous variables.

- Different inference engines: junction tree, sum product, etc.
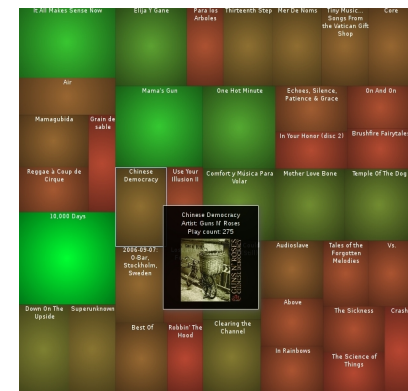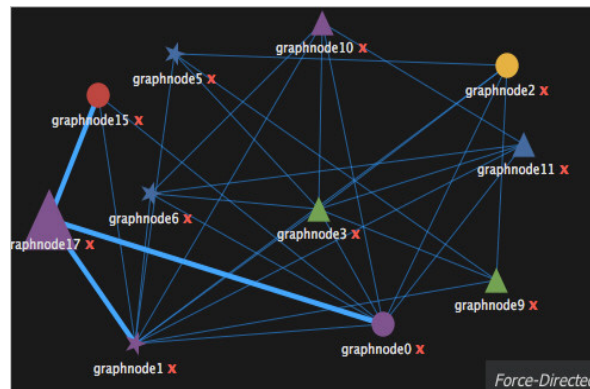
Who's using it: ebay™

JS InfoVis Toolkit
http://philogb.github.io/jit/

Treemap View
JS I...

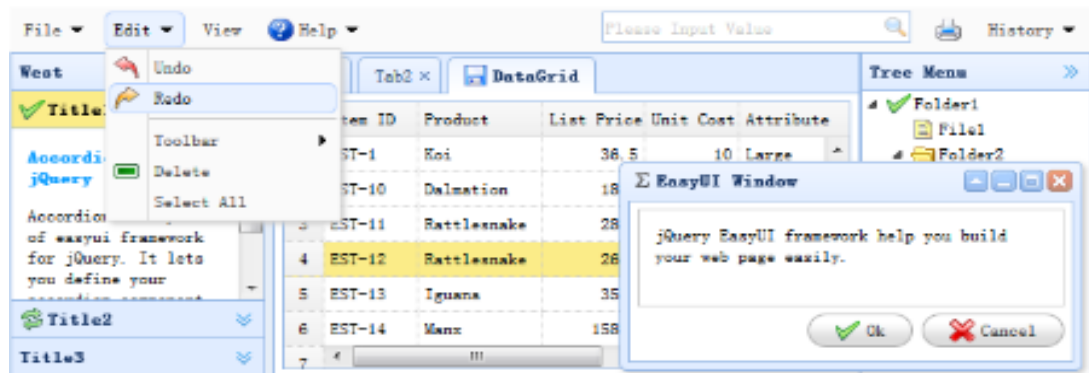**Network View**
JS InfoVis

- Tools for creating Interactive Data Visualizations for the Web.

- Based on the HTML5 canvas.

- Graph, radial and hierarchical network visualizations.

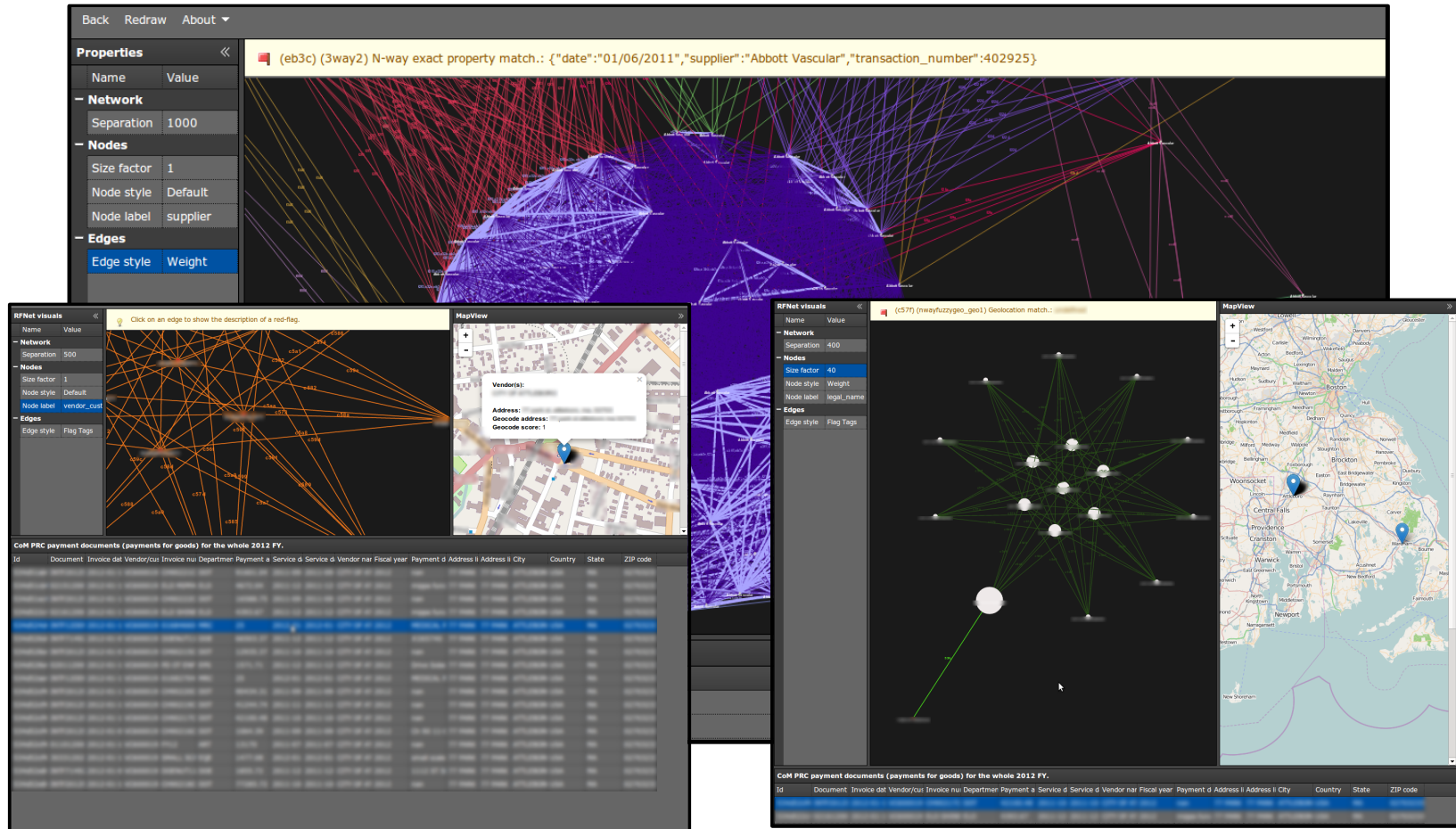- Treemap, stacked sunburst, area, bar and pie charts.

GUI
jQuery EasyUI

- Collection of user-interface plugin based on HTML5 and jQuery.

- Essential functionality for building modern, interactive, javascript applications.

- Datagrid, treegrid, panel, combo and more for building cross-browser web page.

SAFARI: Web-Based Visual Analytics

Conclusions

- **Analysis integration**. Enable SMEs to integrate different analysis techniques for processing large amounts of payment documents.

- **Big data analysis**. Help SMEs to make sense of a large amount of RFs spread across data.

- **Focus**. Help SMEs to focus on the most suspicious payments by exploiting modern high-performance multi-core computers and visualization techniques.
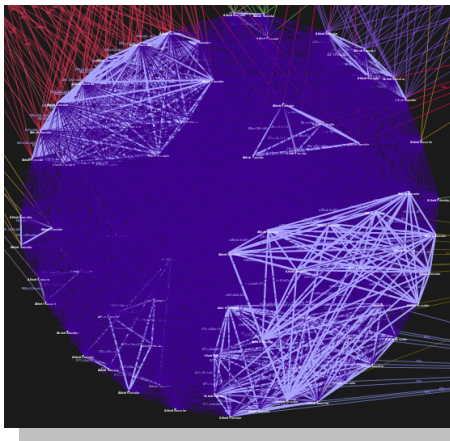
Integration Ranking Visualization

False positive minimization