

genSpace: Community Driven Knowledge Sharing in geWorkbench

Gail Kaiser's Programming Systems Lab

enable (v.t.) : *to make possible, practical, or easy*



PROGRAMMING SYSTEMS LAB
COLUMBIA UNIVERSITY

<http://www.psl.cs.columbia.edu/>

Computer Science Department, Columbia University

June 1, 2012

Knowledge Sharing

- Scientists collaborating together in the same lab on the same project share:
 - Data: specimens, samples, materials, analyses
 - Tools: instruments, software, hardware
 - Knowledge: open discussion, whiteboard
- However, there are temporal (time) and physical (space) constraints
- This model does not scale to communities of scientists working on different projects but who could possibly learn from each other's expertise, experience, etc.

CSCW Approaches

- Most current generation Computer-Supported Cooperative Work systems enable data sharing and/or tool sharing (e.g., EMSL Collaboratories, UIUC BioCoRE, Monod)
- However, these systems support relatively limited knowledge sharing - how/when/where/why to use tools and data
- Knowledge sharing is partially enabled through labor intensive approaches: pubs, email lists, wikis, chat, shared display, etc. — may be outdated, requires active participation
- We seek to enable automatic knowledge sharing — without requiring "extra work" by scientists

genSpace and geWorkbench

- As geWorkbench continually expands with new analysis and visualization modules, choosing the right tools to use and learning how to chain these tools in sequence (workflows) can be very daunting, especially to novice users
- genSpace aims to alleviate this problem by providing recommendations about what to do first and what to do next
- Records, aggregates and data mines geWorkbench users' activities with tools and tool sequences
- Users can ask questions such as what are the most frequently used tools and which workflows start with or contain a given tool, with the answers derived (and updated periodically) from the collective experience of geWorkbench users

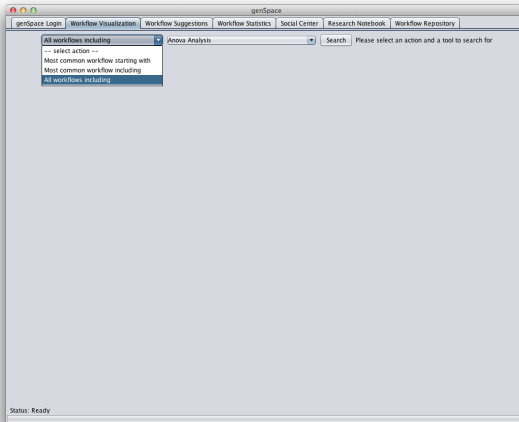
Features

Our overall goal is to deliver community driven knowledge sharing, and to that end, we have implemented many features including:

- Tool and workflow recommendations
- Workflow sharing and visualization
- Community portal (website for tool and workflow comments)
- Social networking features (friends, networks, chat, profiles)
- Research notebook with automatic updates

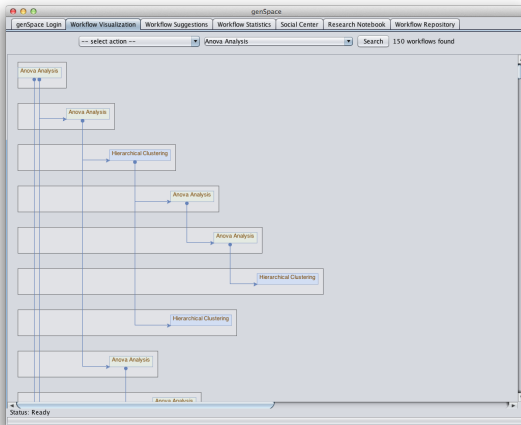
Workflow Visualization

Viewing all workflows with a given tool



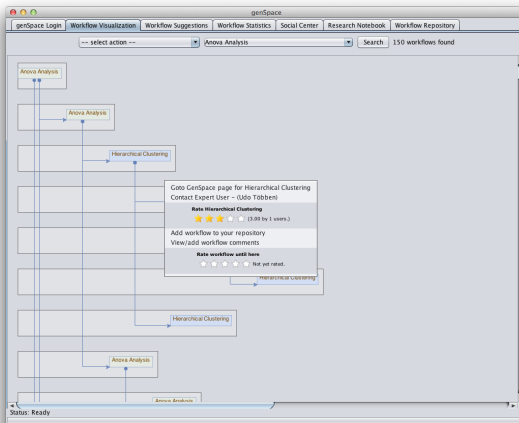
Workflow Visualization

Results



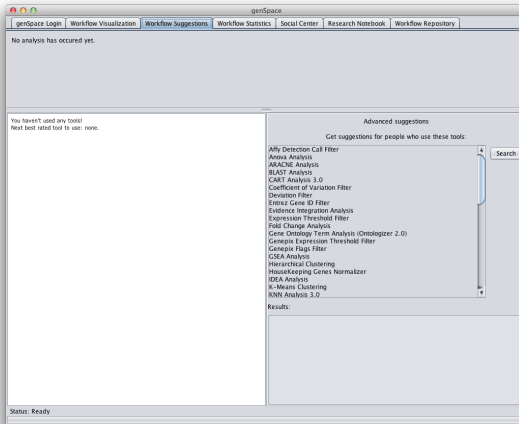
Workflow Visualization

Rating tools and identifying expert users



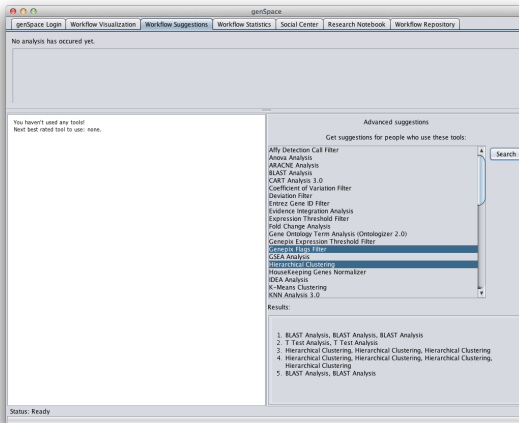
Workflow Suggestions

Default view



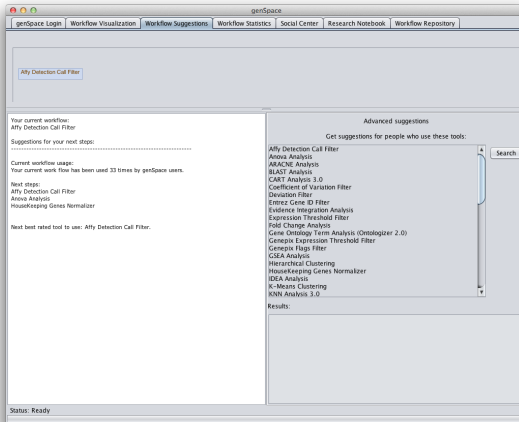
Workflow Suggestions

Can be focused by tool



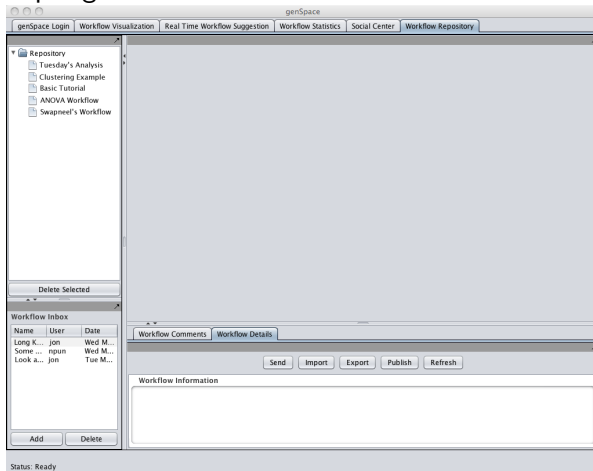
Workflow Suggestions

Updates as user performs analyses



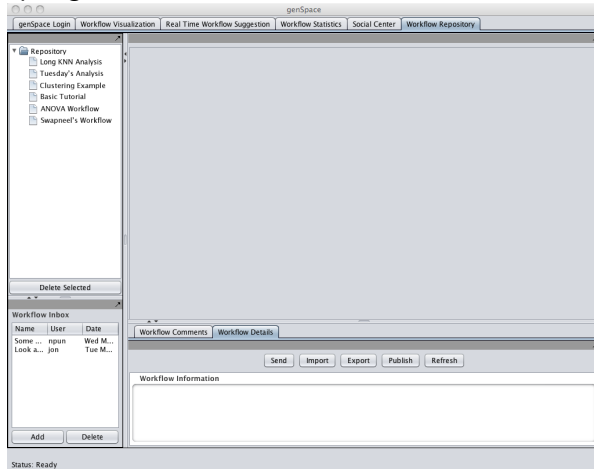
Workflow Repository

Before accepting a shared workflow



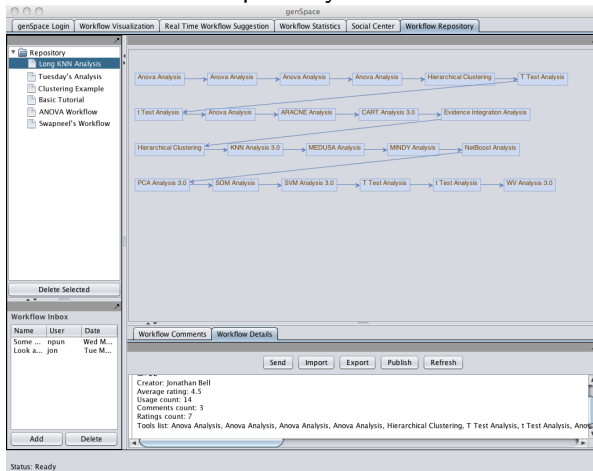
Workflow Repository

After accepting a shared workflow

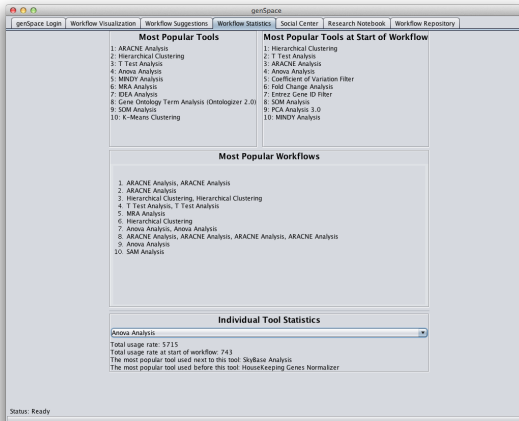


Workflow Repository

Viewing a workflow in the repository

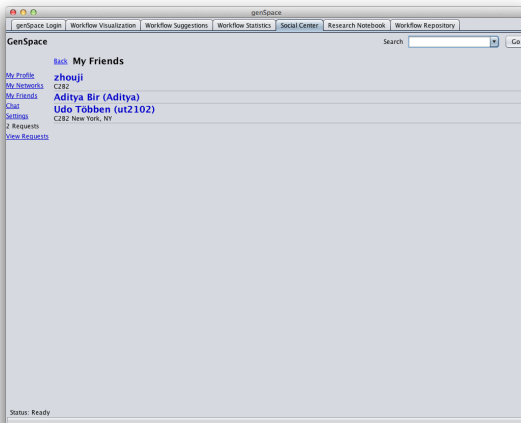


Workflow Statistics



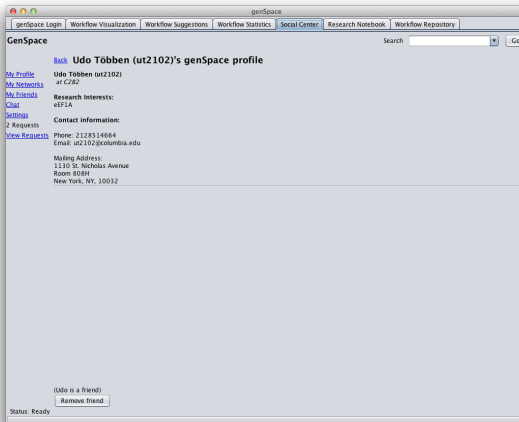
Social Networking

Viewing friends



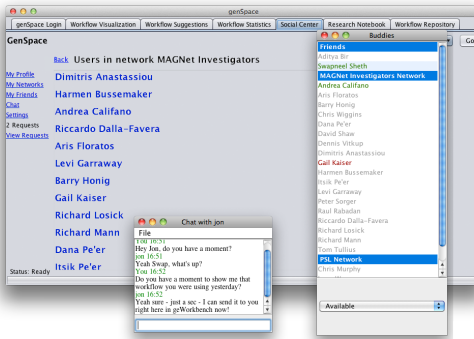
Social Networking

Viewing profiles



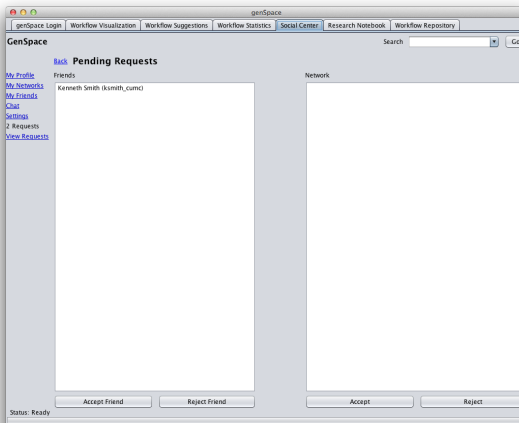
Social Networking

Users by network and chat



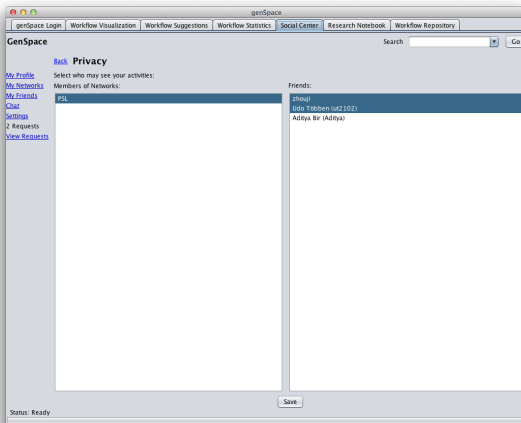
Social Networking

Friend and network requests



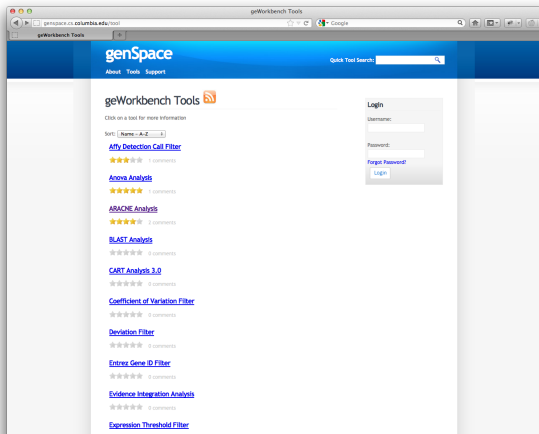
Social Networking

Privacy settings



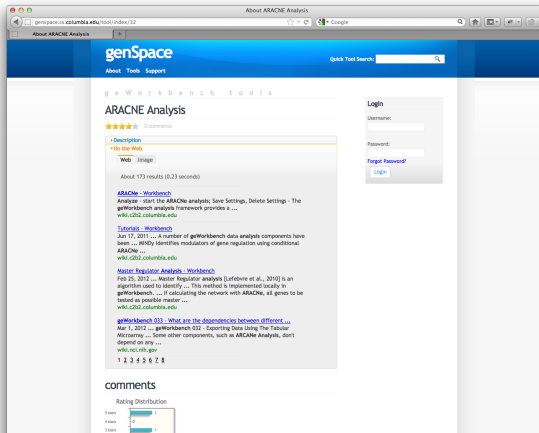
Community portal

Tool list



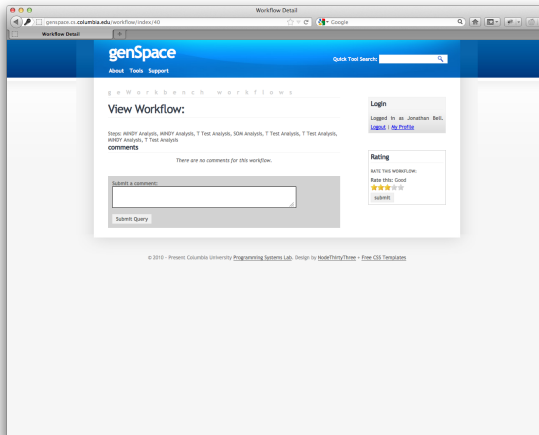
Community portal

Tool detail



Community portal

Workflow detail



New this year

This year's focus was on socially-oriented recommendations and collaboration tools.

- “People like you”
- Collaborative filtering and diversity in recommendations
- Research notebook

Additional notable tasks completed:

- Re-engineered genSpace recommendation platform to use Apache Mahout (scalable open source machine learning library)
- Developed prototype web-based genSpace user interface to integrate with the in-development web-based geWorkbench

“People like you”

Based on social connections and geWorkbench usage

The screenshot shows the genSpace web application interface. At the top, there is a navigation bar with tabs: genSpace Login, Workflow Visualization, Workflow Suggestions, Workflow Statistics, Social Center, Research Notebook, and Workflow Repository. Below the navigation bar, there are two tabs: Data Visibility and General Profile. The General Profile tab is active, showing a form for user information. The form fields are: First Name (Jonathan), Last Name (Bell), Lab Affiliation * (PSL), Email Address, Phone, Address 1, Address 2, City, State, and ZIP Code. There are Save and Logout buttons. Below the form, there are two sections: Your Recommended Workflows and People Like You. The Your Recommended Workflows section lists: 1. ARACNE Analysis, ARACNE Analysis; 2. Hierarchical Clustering, Hierarchical Clustering; 3. Hierarchical Clustering, T Test Analysis; 4. MRA Analysis; 5. Anova Analysis. The People Like You section lists: 1. swap - Swapneeh Sheth; 2. leonw - Leon Wu; 3. chris - Chris Murphy; 4. kaiser - Gail Kaiser; 5. Anonymous. There is a checkbox labeled Filter to My Networks. The status bar at the bottom says Status: Ready.

genSpace	
genSpace Login Workflow Visualization Workflow Suggestions Workflow Statistics Social Center Research Notebook Workflow Repository	
Data Visibility General Profile	
First Name	Jonathan
Last Name	Bell
Lab Affiliation *	PSL
Email Address	
Phone	
Address 1	
Address 2	
City	
State	
ZIP Code	
Save	
Logout	
Filter to My Networks ☐	
Your Recommended Workflows	People Like You
1. ARACNE Analysis, ARACNE Analysis	1. swap - Swapneeh Sheth
2. Hierarchical Clustering, Hierarchical Clustering	2. leonw - Leon Wu
3. Hierarchical Clustering, T Test Analysis	3. chris - Chris Murphy
4. MRA Analysis	4. kaiser - Gail Kaiser
5. Anova Analysis	5. Anonymous
Status: Ready	

Recommendations based on collaboration networks

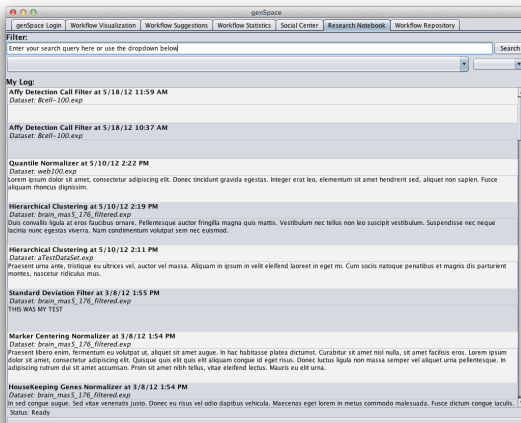
- genSpace tool recommendations now take into account a user's collaboration networks (including implicit “people like you” networks)
- This creates an issue of diversity:
 - Personalized recommendations can lead to a “bubble”
 - For example, if all members of a lab's recommendations are based heavily on their lab mates, there will be no exposure to new ideas
 - genSpace solution: mix in results from other (visible) networks automatically, and allow users to specifically request more diverse results

Research Notebook

- Permanent archive of geWorkbench experiments
- Automatically logs every analysis, including dataset and parameters
- Annotations can optionally be added to any analysis
- History is fully searchable

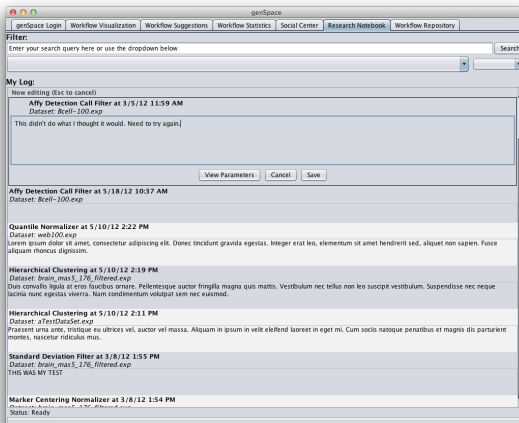
Research Notebook

Viewing the notebook



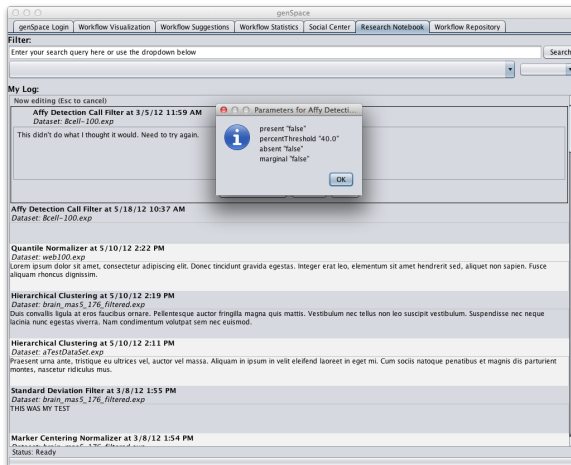
Research Notebook

Editing a notebook entry



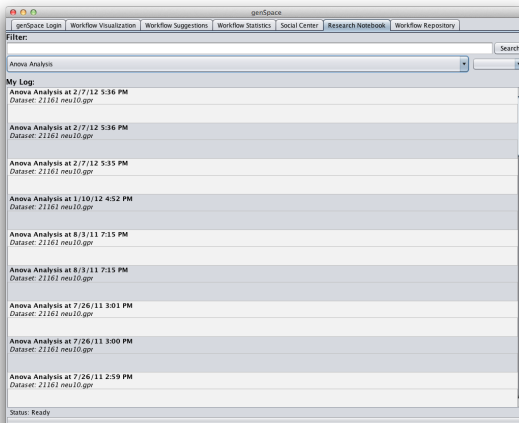
Research Notebook

Viewing parameters from an analysis



Research Notebook

Filtering view by tool



Future Work

- Expand research notebook feature to allow groups of researchers to share notes
- Fully develop a web-based genSpace to complement the web-based geWorkbench
- Full support for workflows that include aborted and long-duration tools in geWorkbench

genSpace: Community Driven Knowledge Sharing in geWorkbench

Gail Kaiser's Programming Systems Lab

enable (v.t.) : *to make possible, practical, or easy*



PROGRAMMING SYSTEMS LAB
COLUMBIA UNIVERSITY

<http://www.psl.cs.columbia.edu/>

Computer Science Department, Columbia University

June 1, 2012