



Project C

Sage Infrastructure Tools Project

- Carole Goble, University of Manchester, UK
- Ted Liefeld, Broad Institute
- Alex Pico, Gladstone Institutes

- Marc Hadfield



Acknowledgments

Manchester

- Carole Goble
- Peter Li
- Paul Fisher
- Charlotte Hooson-Sykes

Broad Institute

- Ted Liefeld

Gladstone Institutes

- Alex Pico

Alitara

- Marc Hadfield
- Derek Kobylarz

Sage Bionetworks

- Andrew Kasarskis
- Justin Guinney
- Xavier Schildwachter
- Bin Zhang
- Kaitlin Thaney

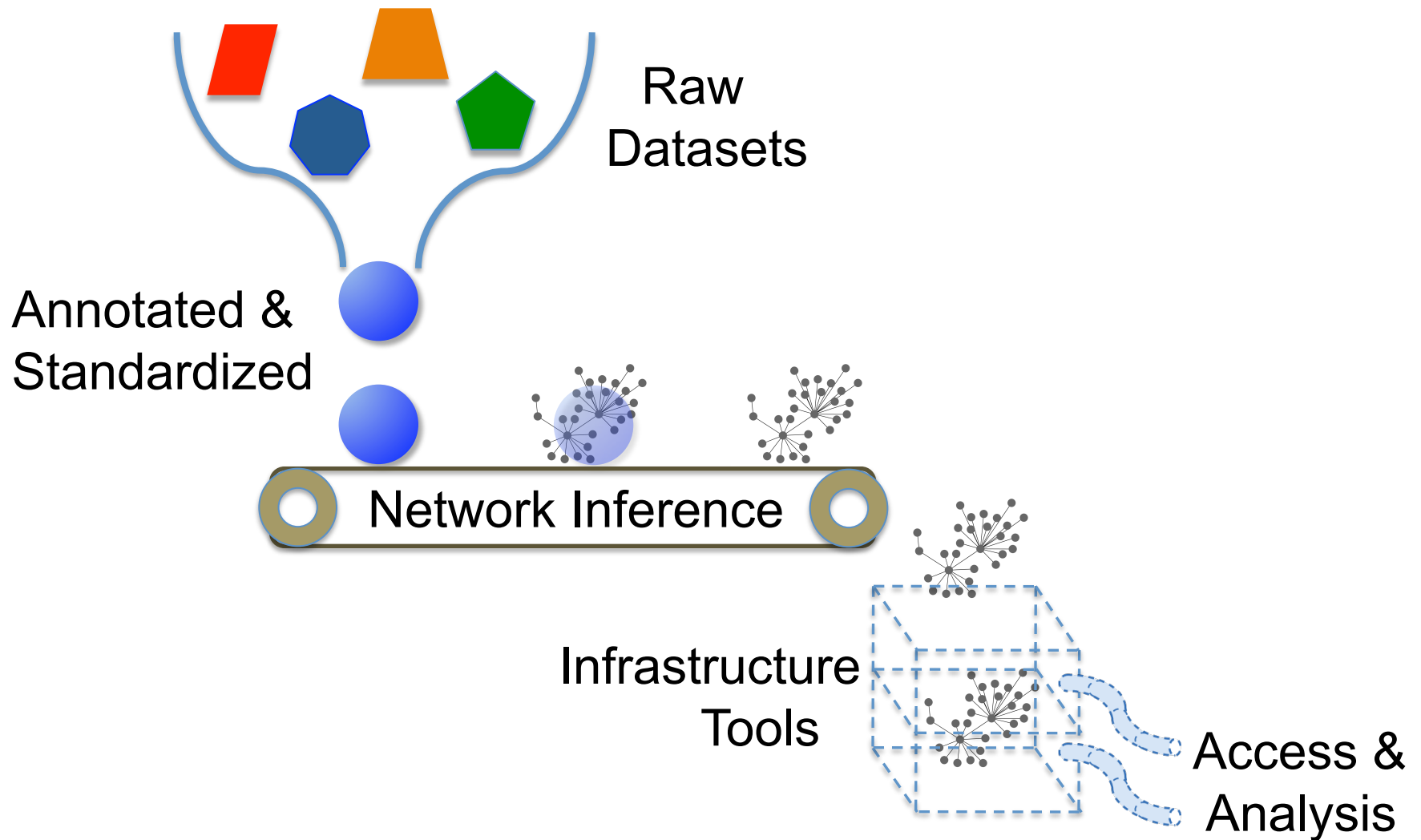
Congress Workstreams

- Jessie Tenenbaum
- Ilya Kuperschmidt
- Jim Davies

And Many More...

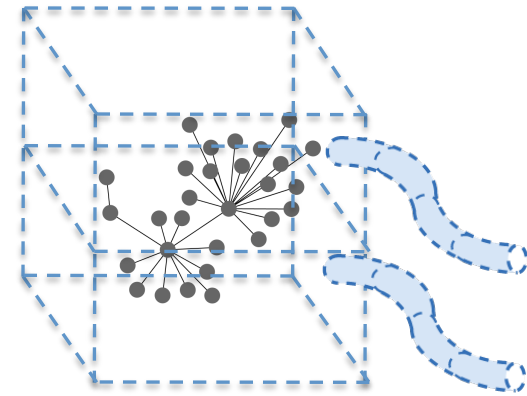
- Michel Dumontier
- Tim Clark

Project Workstream C: Tools



Project Workstream C: Tools

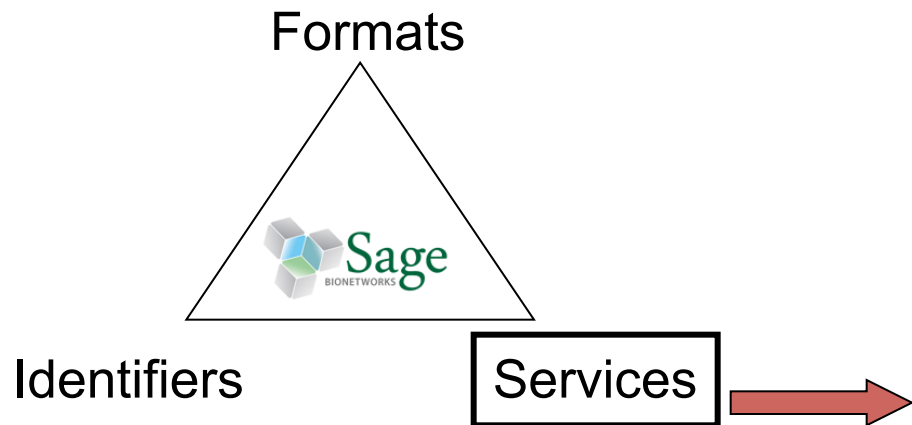
- Formatted network data
- Standard interfaces
- Extensible tools
- Easy to access, easy to use



- Next 10 minutes:
 - example use cases and tools
 - core concepts
 - challenges ahead

Overview

Sage Commons



Work Group C - Tools

- Alitora
 1. Query Networks
 2. Store Annotations
 3. Web/API Access
- Gene Pattern
 1. Key Driver Analysis
 2. Integration with Cytoscape
- Taverna
 1. Service / Tool integration
 2. Workflow re-use
 3. Large-scale and systematic data analyses



Use Case: Find Networks of Interest

<http://saas.alitora.com/sagedemo/>

About Sage Commons Demo

Visualize Sage Networks

John Visio [preferences](#) [help](#) [logout](#)

DEMO

Sage Commons

Meme Search Results:

- BioChem** Com.Memomics.AlitoraSystem BCAR3

Breast tumors are initially dependent on resistant. Breast cancer anti-estrogen resistant signal transduction that causes estrogen-kinase signaling molecules, and is partly

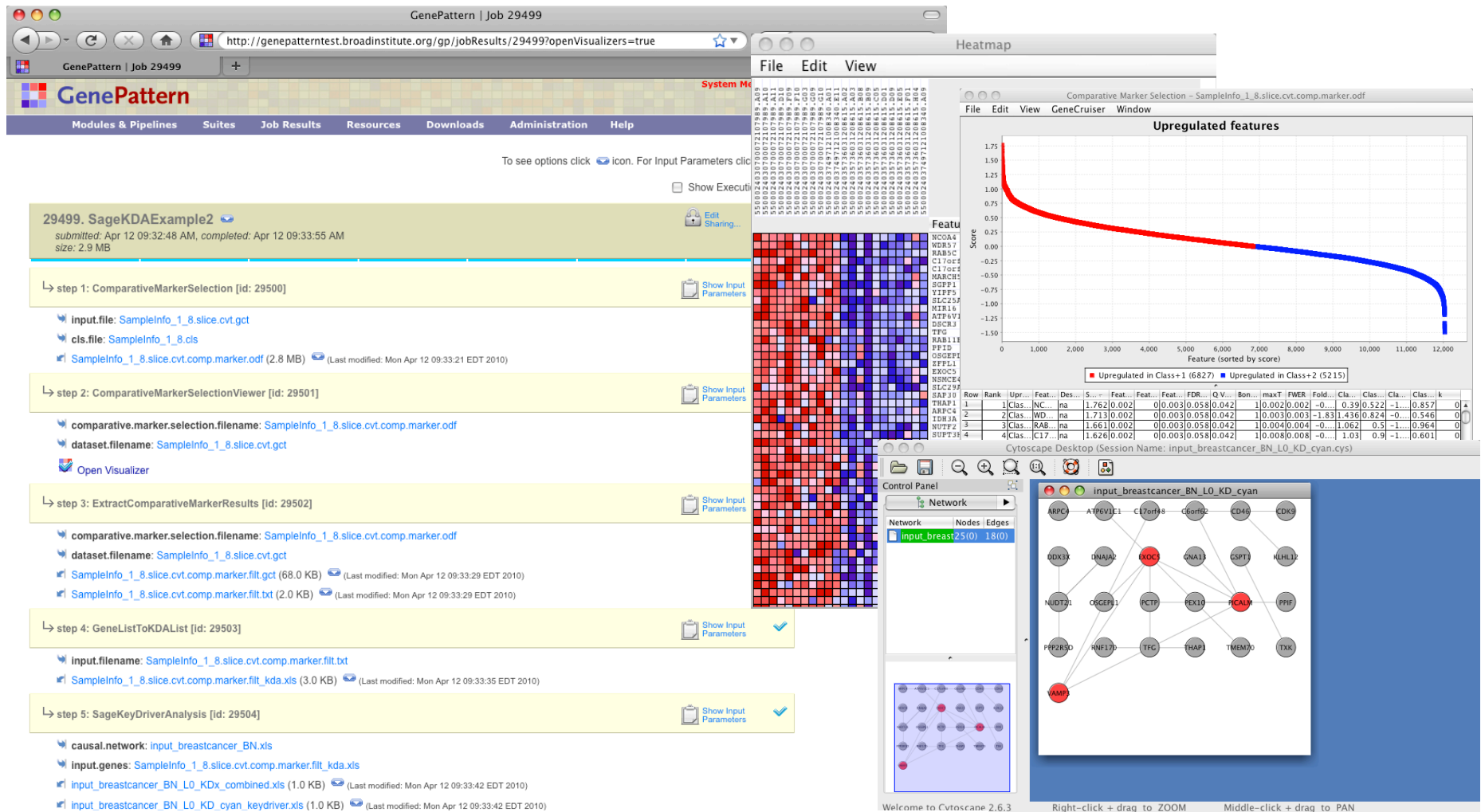
Associated networks

 - adipose male SCN6 sage network
 - muscle male SCN20 sage network
 - muscle female SCN19 sage network
 - adipose female SCN5 sage network
- OMIM** Com.Memomics.AlitoraSystems.c BREAST CANCER
- OMIM** Com.Memomics.AlitoraSystems.c BREAST CANCER-RELATED RE
- BioChem** Com.Memomics.AlitoraSystem BRMS1

The screenshot shows the Cytoscape Desktop interface. The main window displays a network graph titled 'MemomicsNetwork2' with various nodes and edges. The control panel on the left is open to the 'Expansion Criteria' tab, showing filters for 'Nodes & Edges' and 'Time & Confidence'. The 'Nodes & Edges' section includes checkboxes for 'Include / Exclude' and lists of 'edge types' (SystemSegment, omim, sage, geneOntology, clinicalTrialsToCompanies, diseaseOntology, articleOntology, drugBank2entrez) and 'relation types' (relation types). The 'Node types' section includes checkboxes for 'Include / Exclude' and lists of 'node types' (node types, SystemSegment, semanticHacker, opencalais, sage, articleOntology, companies). The 'Data Panel' at the bottom shows a table with columns for 'ID' and 'typedId'. The status bar at the bottom indicates 'Welcome to Cytoscape 2.6.3' and provides instructions for zooming and panning.

This gene reduces the metastatic potential, but not the tumorigenicity, of human breast cancer and melanoma cell lines. The protein encoded by this gene localizes primarily to the nucleus and is a

Use Case: Key Driver Analysis



GenePattern | Job 29499

http://genepatterntest.broadinstitute.org/gp/jobResults/29499/openVisualizers=true

GenePattern | Job 29499

Modules & Pipelines Suites Job Results Resources Downloads Administration Help

To see options click icon. For Input Parameters click icon.

29499. SageKDAExample2
submitted: Apr 12 09:32:48 AM, completed: Apr 12 09:33:55 AM
size: 2.9 MB

step 1: ComparativeMarkerSelection [id: 29500]

- input.file: SampleInfo_1_8.slice.cvt.gct
- cis.file: SampleInfo_1_8.cis
- SampleInfo_1_8.slice.cvt.comp.marker.odf (2.8 MB) (Last modified: Mon Apr 12 09:33:21 EDT 2010)

step 2: ComparativeMarkerSelectionViewer [id: 29501]

- comparative.marker.selection.filename: SampleInfo_1_8.slice.cvt.comp.marker.odf
- dataset.filename: SampleInfo_1_8.slice.cvt.gct

Open Visualizer

step 3: ExtractComparativeMarkerResults [id: 29502]

- comparative.marker.selection.filename: SampleInfo_1_8.slice.cvt.comp.marker.odf
- dataset.filename: SampleInfo_1_8.slice.cvt.gct
- SampleInfo_1_8.slice.cvt.comp.marker.filter.gct (68.0 KB) (Last modified: Mon Apr 12 09:33:29 EDT 2010)
- SampleInfo_1_8.slice.cvt.comp.marker.filter.txt (2.0 KB) (Last modified: Mon Apr 12 09:33:29 EDT 2010)

step 4: GeneListToKDAList [id: 29503]

- input.filename: SampleInfo_1_8.slice.cvt.comp.marker.filter.txt
- SampleInfo_1_8.slice.cvt.comp.marker.filter_kda.xls (3.0 KB) (Last modified: Mon Apr 12 09:33:35 EDT 2010)

step 5: SageKeyDriverAnalysis [id: 29504]

- causal.network: input_breastcancer_BN.xls
- input.genes: SampleInfo_1_8.slice.cvt.comp.marker.filter_kda.xls
- input_breastcancer_BN_L0_KDx_combined.xls (1.0 KB) (Last modified: Mon Apr 12 09:33:42 EDT 2010)
- input_breastcancer_BN_L0_KD_cyan_keydriver.xls (1.0 KB) (Last modified: Mon Apr 12 09:33:42 EDT 2010)

Heatmap

Comparative Marker Selection - SampleInfo_1_8.slice.cvt.comp.marker.odf

Upregulated features

Score

Feature (sorted by score)

■ Upregulated in Class+1 (6827) ■ Upregulated in Class+2 (5215)

Feat	Row	Rank	Up	Des	S	Feat	Feat	FDR	Q V	Bon	maxT	FWER	Fold	Clas	Clas	Clas	Clas	k	
WDR4	1	1	Clas	NC	na	1.762	0.002	0.003	0.058	0.042	1	0.002	0.002	-0.1	0.39	0.522	-1.1	0.857	0
IDR1A	2	2	Clas	WD	na	1.713	0.002	0.003	0.058	0.042	1	0.003	0.003	-1.83	1.436	0.824	-0.1	0.546	0
NOTIF2	3	3	Clas	RAB	na	1.661	0.002	0.003	0.058	0.042	1	0.004	0.004	-0.1	1.062	0.5	-1.1	0.964	0
NOTIF3	4	4	Clas	C17	na	1.626	0.002	0.003	0.058	0.042	1	0.008	0.008	-0.1	1.03	0.9	-1.1	0.601	0

Cytoscape Desktop (Session Name: input_breastcancer_BN_L0_KD_cyan.cys)

Control Panel

Network

Nodes Edges

input_breast25(0) 18(0)

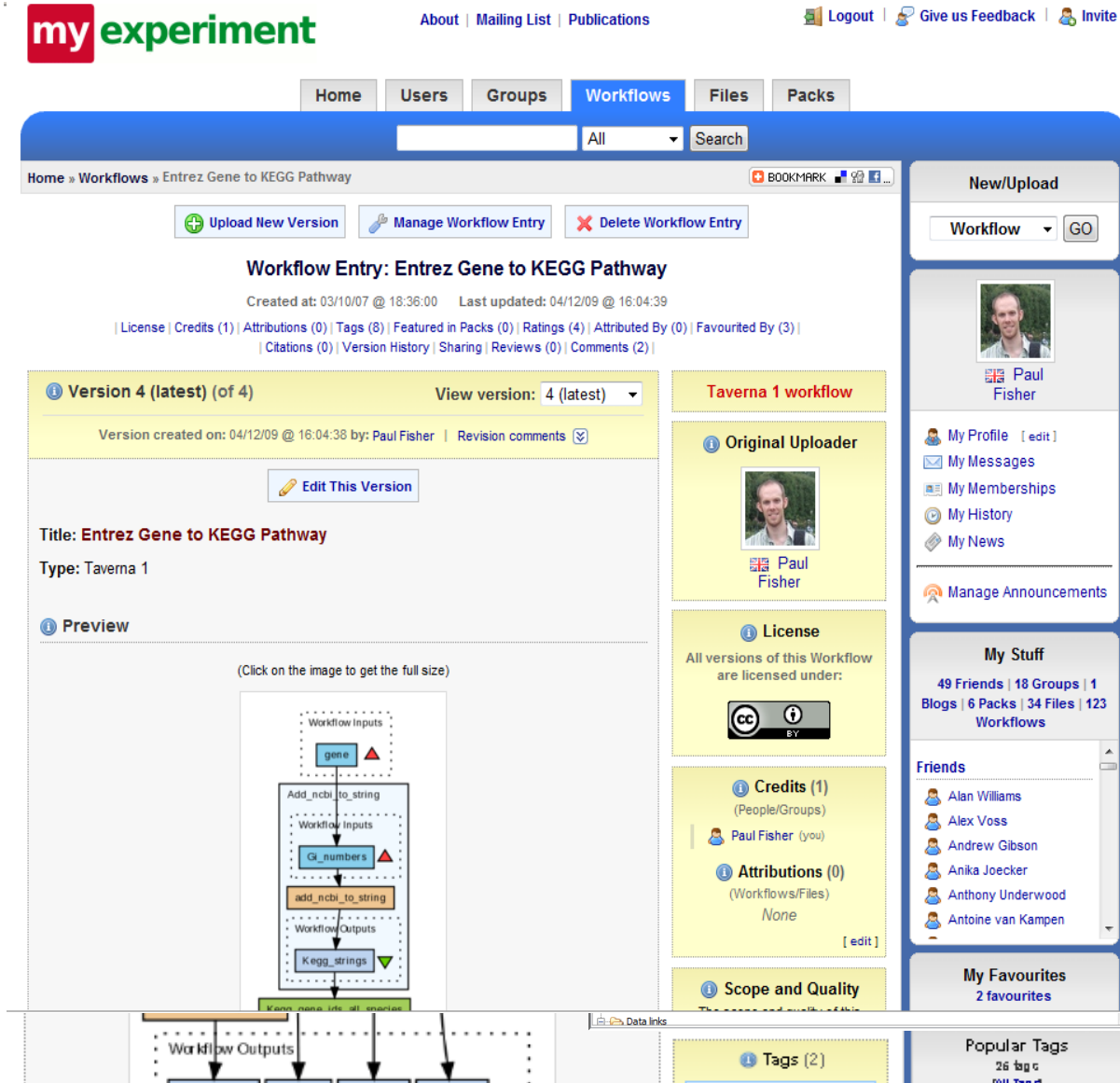
Network diagram showing nodes and edges.

Welcome to Cytoscape 2.6.3

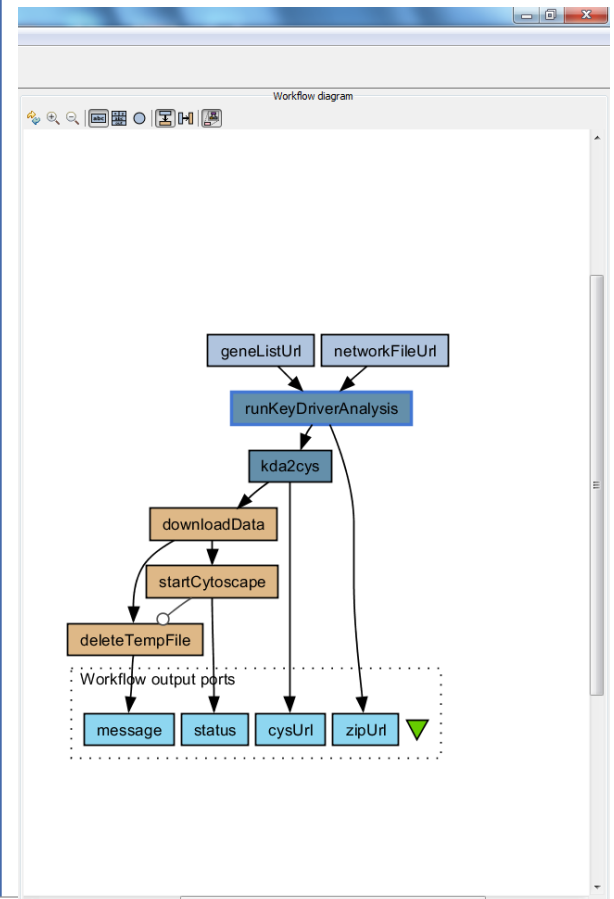
Right-click + drag to ZOOM Middle-click + drag to PAN

Use Case: Taverna

<http://www.myexperiment.org>

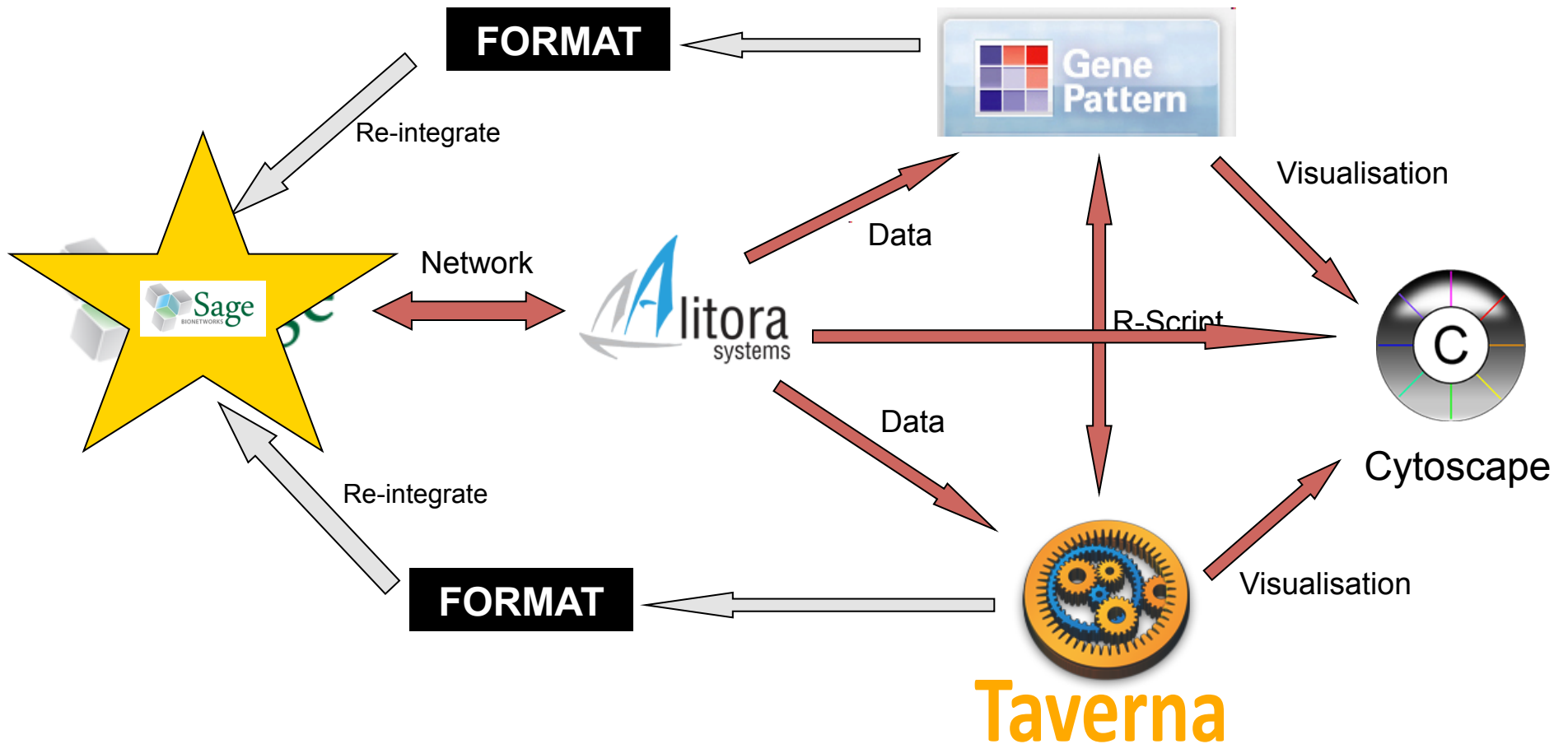


The screenshot shows the 'myexperiment.org' website interface. The main content area displays a workflow entry titled 'Entrez Gene to KEGG Pathway'. It includes a search bar, navigation tabs (Home, Users, Groups, Workflows, Files, Packs), and a sidebar with user information for Paul Fisher. The workflow details section shows 'Version 4 (latest) (of 4)' with a 'Preview' section containing a workflow diagram. The diagram shows a sequence of steps: 'gene' (input), 'Add_ncbi_to_string', 'Gt_numbers', 'add_ncbi_to_string', and 'Kegg_strings' (output). The sidebar also lists 'Original Uploader', 'License', 'Credits', and 'Attributions'.



<http://www.taverna.org.uk>

The SAGE Pipeline



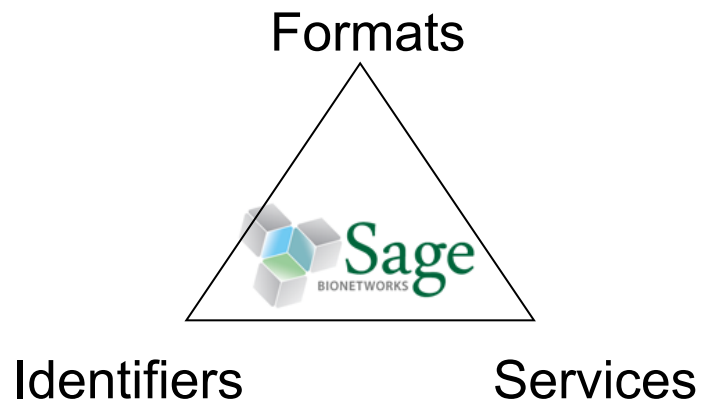


Core principles

1. Maximize ***access***
2. Maximize ***use***
3. Maximize ***reuse***

- ✓ Distribute multiple file formats
- ✓ Make use of existing standards and tools
- ✓ Design for flexible, extensible solutions
- ✓ Support collaboration and community annotation

Implications for Sage infrastructure



Map to
standards



Appropriate
interfaces

Lessons Learned:

1. Standard network & gene list file formats are critical to the success of infrastructure tools.
2. Current dataset and network repositories fall short of providing a community resource with adequate standards and extensible tools.

Challenges Ahead:

1. Preparing for increasing *scale* and *scope* of data
2. Preparing for future data types and analyses



This afternoon

- Drill down into demos and experiences
- Guests
 - Tim Clark – SWAN, Web 3.0, neurodegeneration
 - Michel Dumontier – Bio2RDF
- Audience participation!
 - Opportunities, Barriers and Incentives
 - Platforms, datasets, services and tools
 - Technologies and Standards
 - Gaps
 - Directions for Sage Bionetworks
 - Planning Saturday