

my **experiment** and the
Rise of Social Machines

David De Roure

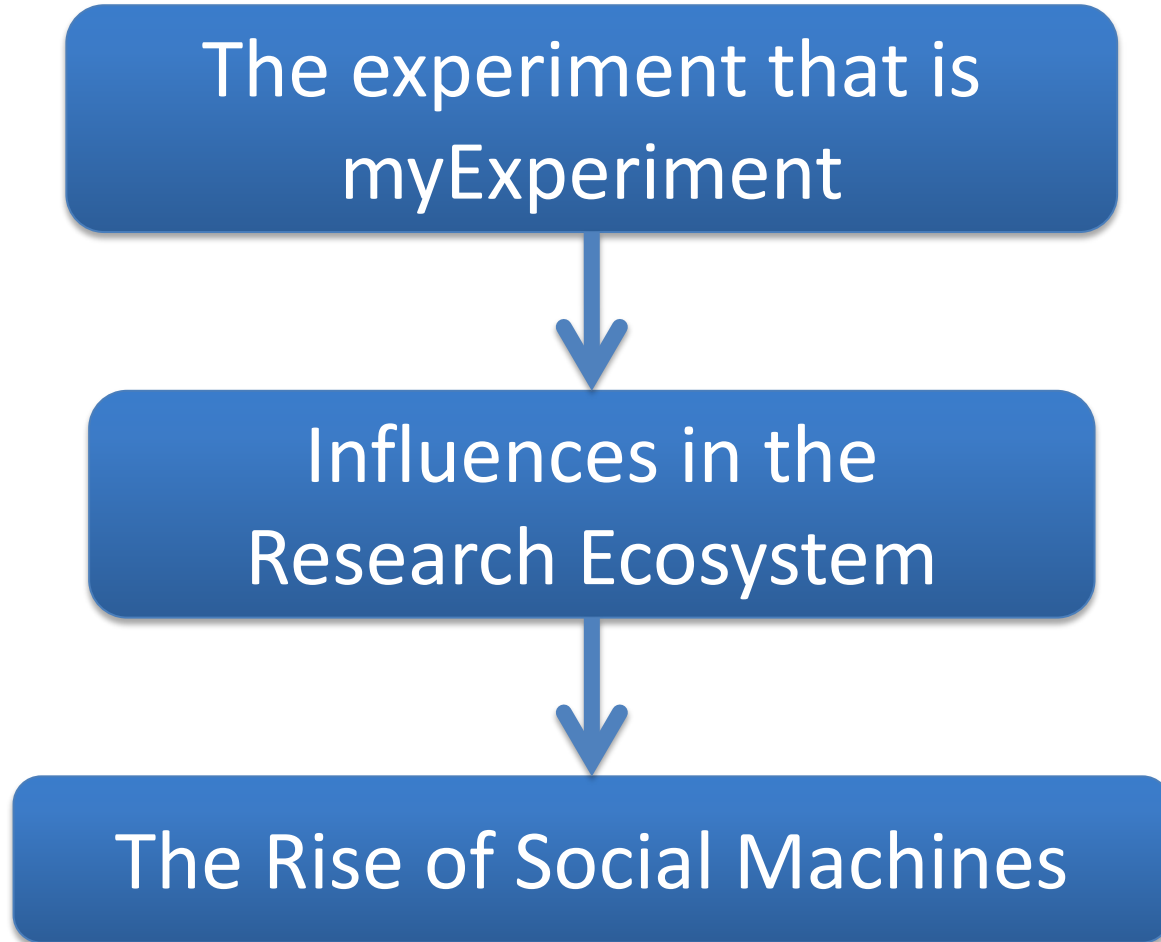


Workflow of Talk

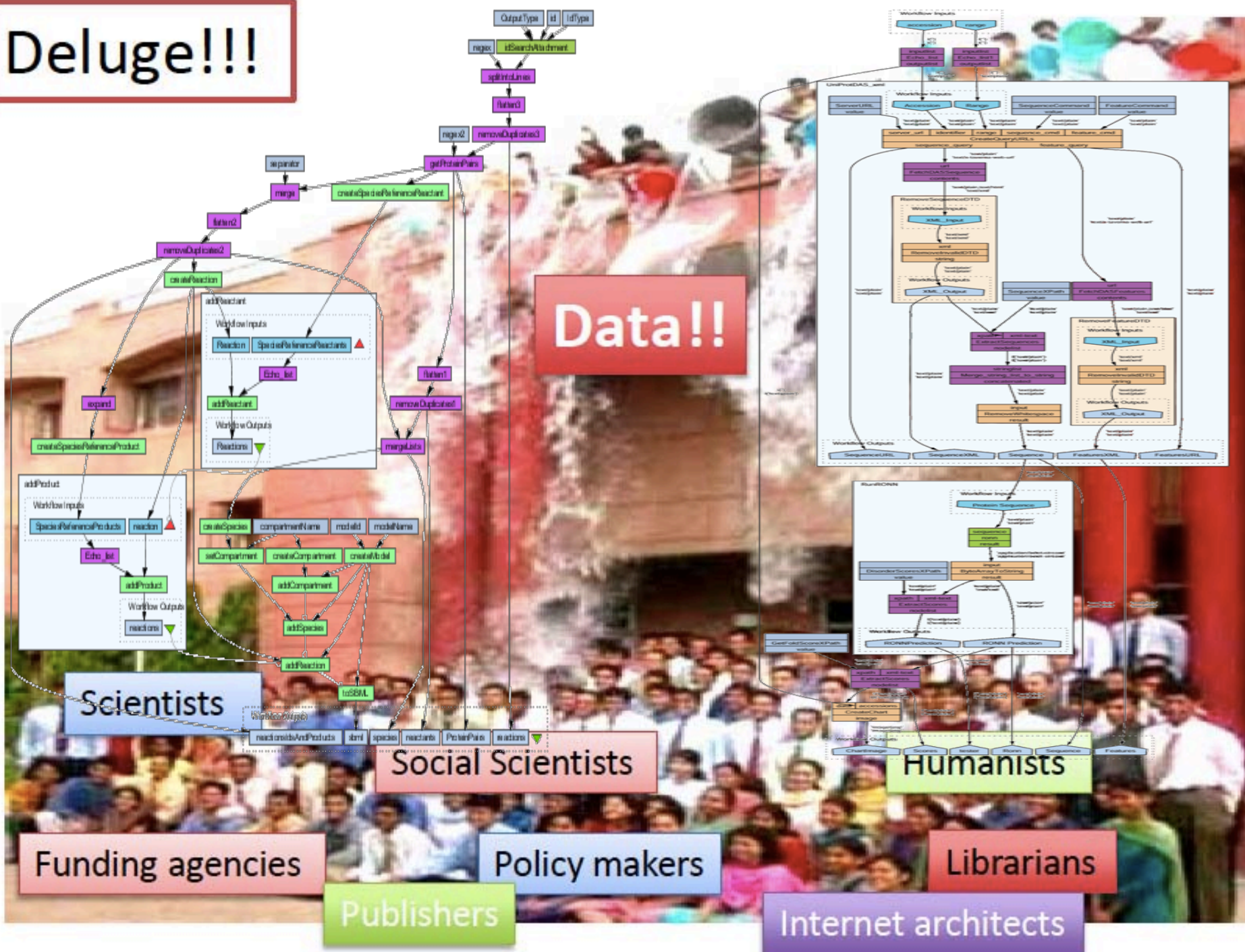
The experiment that is
myExperiment

Influences in the
Research Ecosystem

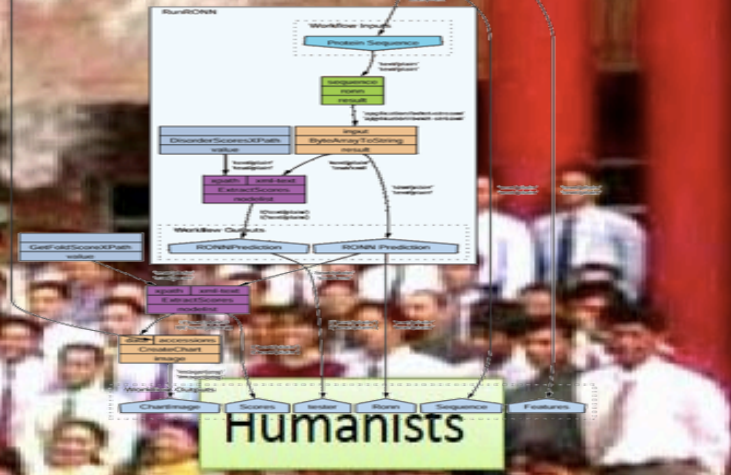
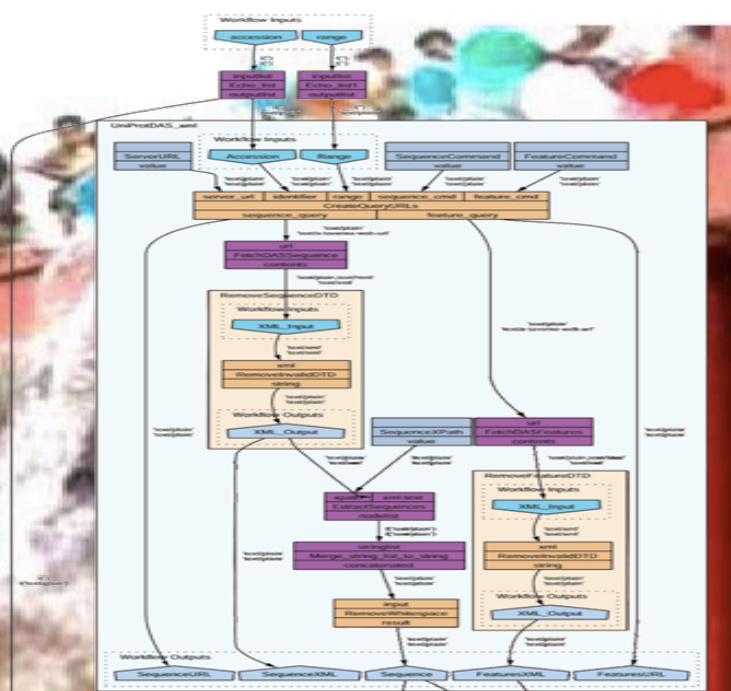
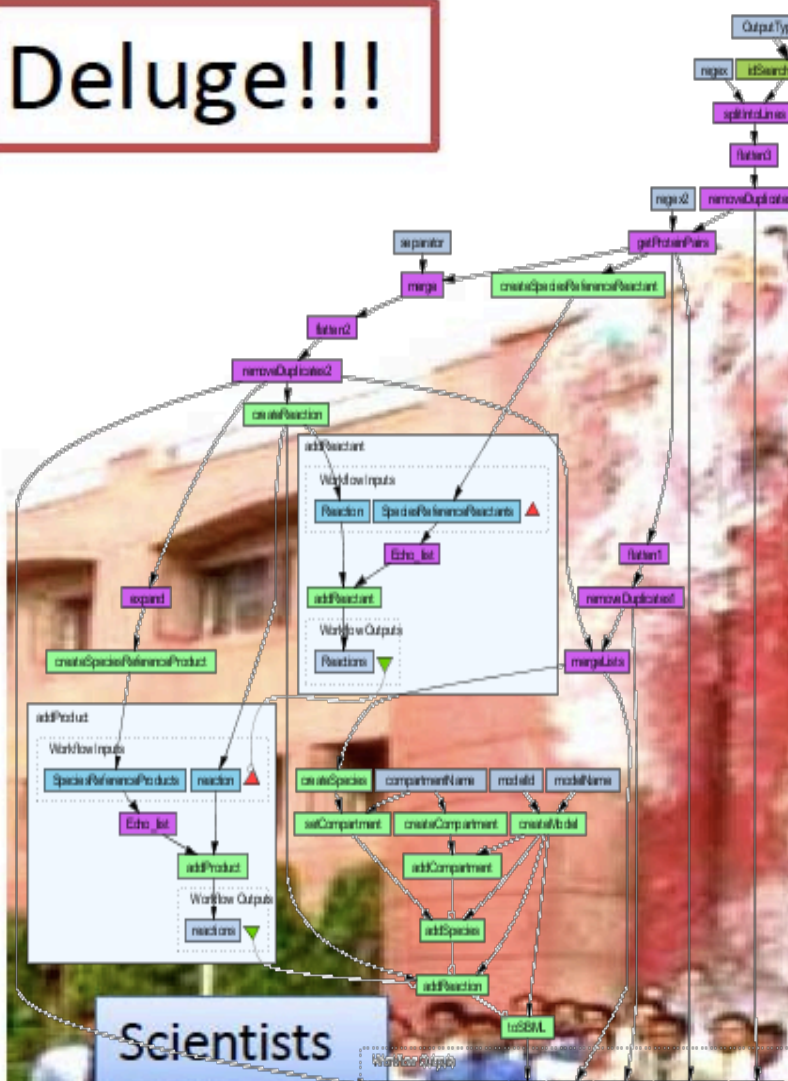
The Rise of Social Machines



Deluge!!!



Data!!



Scientists

Social Scientists

Humanists

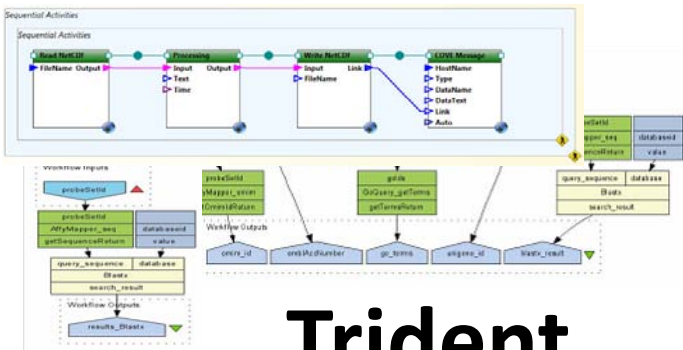
Funding agencies

Policy makers

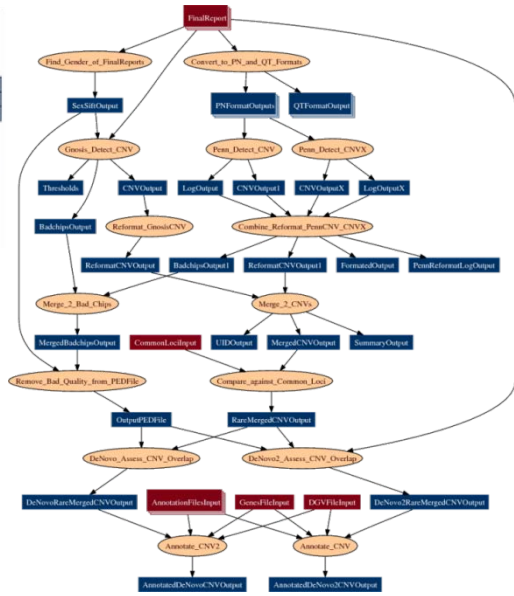
Librarians

Publishers

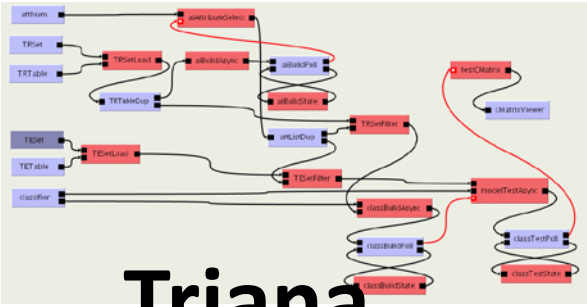
Internet architects



Trident



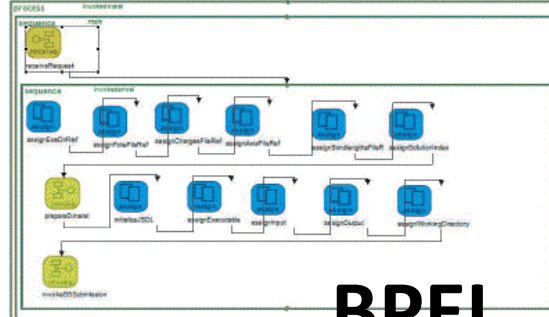
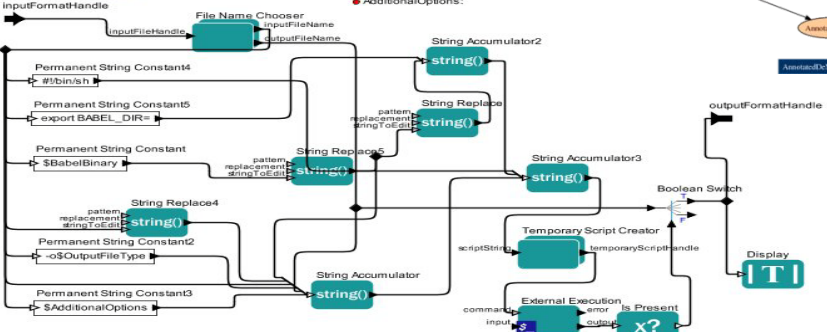
Pegasus



Triana

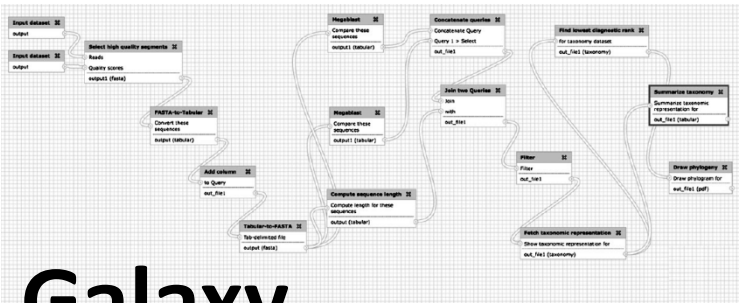
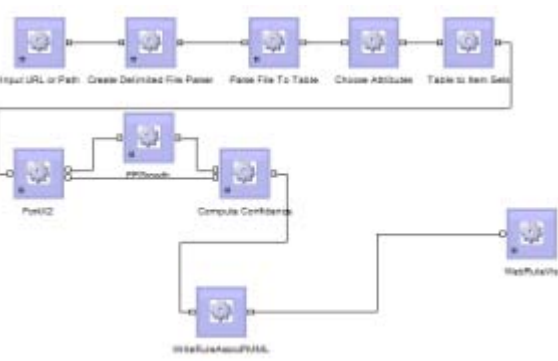
Kepler

- BabelBinary: fopt/babel-1.6/babel
- OutputFileType:
- OutputFileDirectory:
- AdditionalOptions:



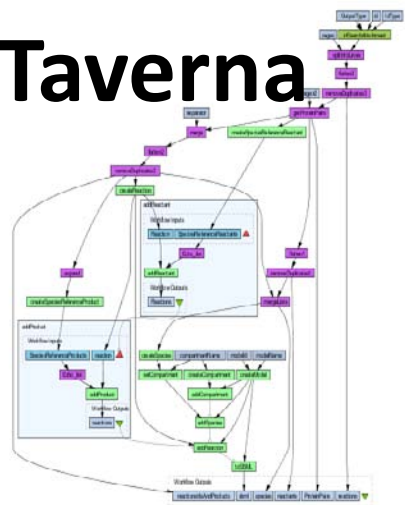
BPEL

Meandre



Galaxy

Taverna



“A biologist would rather share their toothbrush than their gene name”



flickrTM



Broadcast YourselfTM

amazon.com[®]

The Amazon smile logo is a curved orange arrow pointing from left to right, positioned below the text.

slideshare

Present Yourself

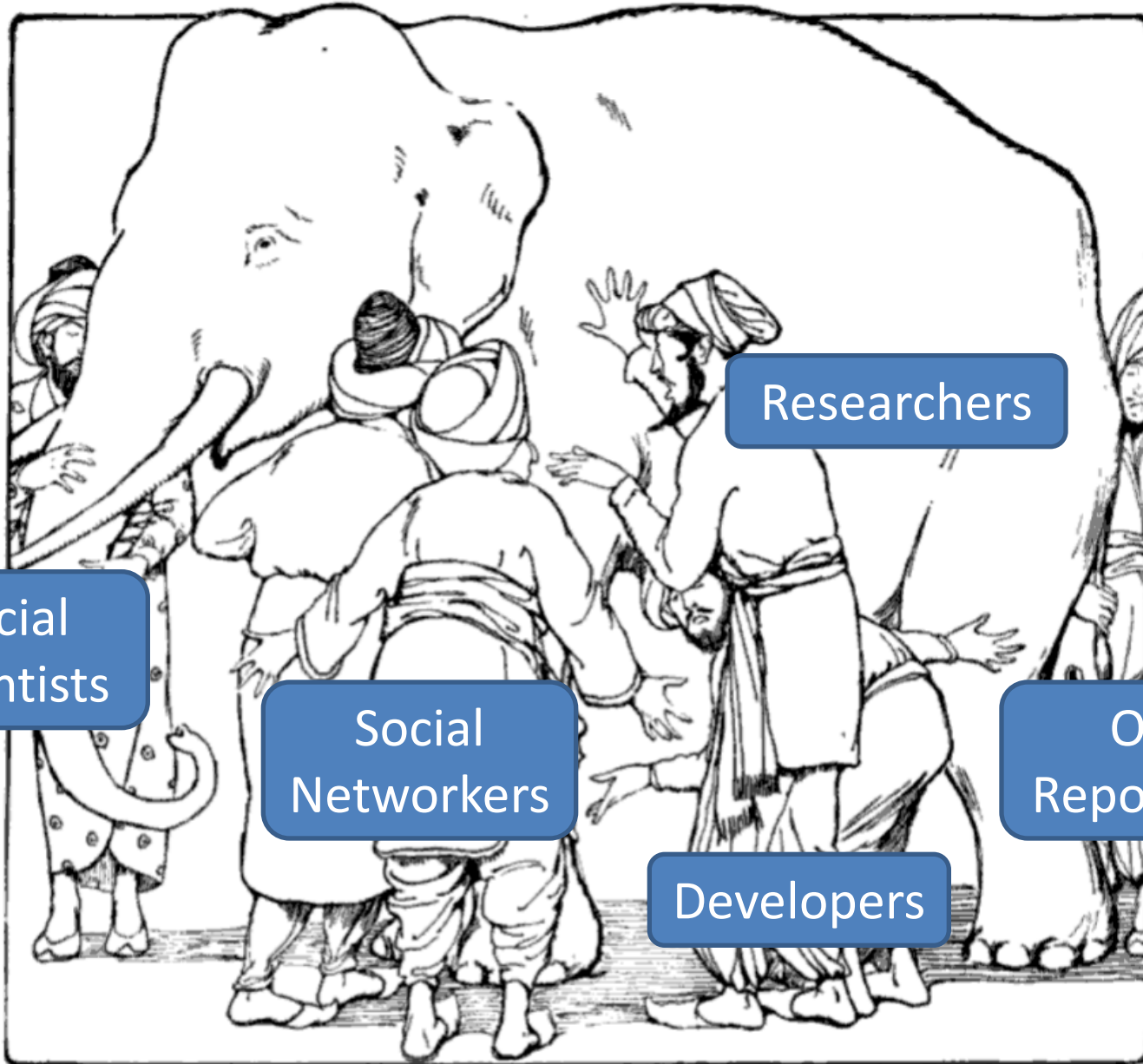
my **experiment**



Not Facebook for scientists!



The experiment that is **my** experiment



Researchers

Social
Scientists

Social
Networkers

Open
Repositories

Developers

my experiment



- “Facebook for Scientists”
...but different to Facebook!
- A repository of research methods
- A community social network of people and things
- A probe into researcher behaviour
- A provocation in scholarly communication
- Open source (BSD) Ruby on Rails app
- REST and SPARQL interfaces, supports Linked Data
- Influenced BioCatalogue, MethodBox and SysMO-SEEK

myExperiment has 300 groups, 2600 workflows and 260 packs



Workflow Entry: Success-Abandonment-Classification

Created at: 06/02/08 @ 14:35:41 Last updated: 02/07/08 @ 17:15:25

[License](#) | [Credits \(2\)](#) | [Attributions \(0\)](#) | [Tags \(6\)](#) | [Featured in Packs \(1\)](#) | [Ratings \(1\)](#) | [Attributed By \(0\)](#) | [Favourited By \(2\)](#) | [Citations \(0\)](#) | [Version History](#) | [Reviews \(0\)](#) | [Comments \(0\)](#)

Version 3 (latest) (of 3)

View version: 3 (latest) [dropdown]

Version created on: 06/02/08 @ 14:35:41 by: [Andrea Wiggins](#) | [Revision comments](#) [dropdown]

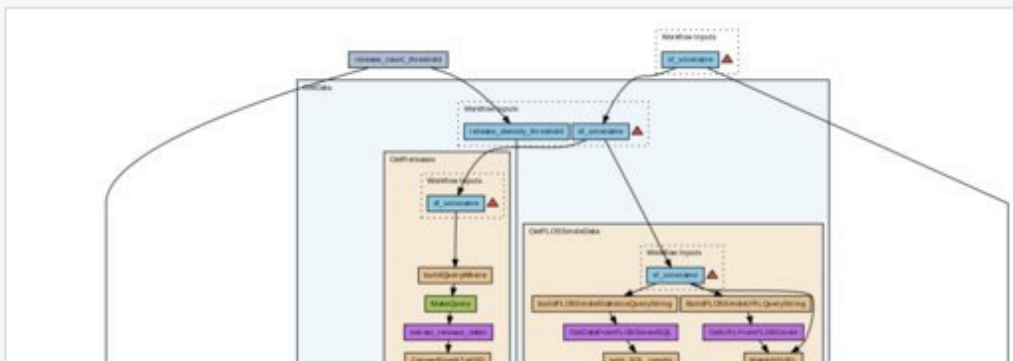
Last edited on: 02/07/08 @ 17:15:25 by: [Andrea Wiggins](#)

Title: Success-Abandonment-Classification

Type: Taverna 1

Preview

(Click on the image to get the full size)



Workflow Type
Taverna 1

Original Uploader



[Andrea Wiggins](#)

License

All versions of this Workflow are licensed under:



Credits (2)

(People/Groups)

[Andrea Wiggins](#)

[James Howison](#)

Attributions (0)

New/Upload

Workflow [dropdown] GO



[David De Roure](#)

- [My Profile](#) [edit]
- [My Messages \(3\)](#)
- [My Memberships \(1\)](#)
- [My History](#)
- [My News](#)

[Manage Announcements](#)

3 new messages

- [Hi David](#)
- [Sean Bechhofer is n...](#)
- [Invitation to 'Wf4E...](#)

2 new friendship requests

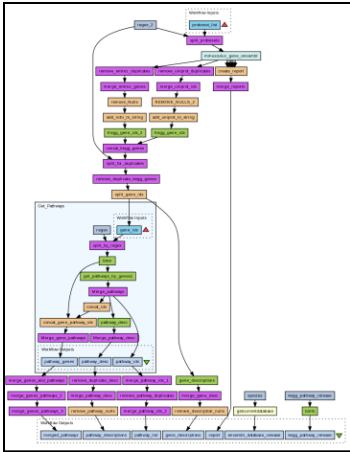
- [Yehia El-khatib](#)
- [mihailionita_me](#)

1 new group request

- [From Pique](#)
(for Group: Wf4Ever)

Paul's Research Object

Workflow 16



QTL

Results

```

path:mmu04060 Cytokine-cytokine receptor interaction - Mus musculus (mouse)
path:mmu00970 Aminoacyl-tRNA biosynthesis - Mus musculus (mouse)
path:mmu00240 Pyrimidine metabolism - Mus musculus (mouse)
path:mmu03010 Ribosome - Mus musculus (mouse)
path:mmu04080 Neuroactive ligand-receptor interaction - Mus musculus (mouse)
path:mmu04210 Apoptosis - Mus musculus (mouse)
path:mmu05220 Chronic myeloid leukemia - Mus musculus (mouse)
path:mmu04612 Antigen processing and presentation - Mus musculus (mouse)
path:mmu00271 Methionine metabolism - Mus musculus (mouse)
path:mmu04912 GPCR signaling pathway - Mus musculus (mouse)
path:mmu04330 Notch signaling pathway - Mus musculus (mouse)
path:mmu04640 Hematopoietic cell lineage - Mus musculus (mouse)
path:mmu00361 Lipid metabolism - Mus musculus (mouse)
path:mmu04110 Cell cycle - Mus musculus (mouse)
path:mmu04530 Tight junction - Mus musculus (mouse)
path:mmu02010 ABC transporters - General - Mus musculus (mouse)
    
```

produces

Analysis Protocol for Candidate Genes and Pathways

This protocol is aimed at providing a guide to the interpretation of the results obtained from both the QTL and microarray workflows. Each workflow provides a series of text files, which are to be used as a means of obtaining the pathways which relate to differentially expressed genes in the necessary study and genes located within the chosen QTL region.

The output from each workflow consists of the following files:

- ensembl_database_release.txt
- pathway_descriptions.txt
- gene_descriptions.txt
- merged_pathways.txt
- merged_pathways.txt
- map_external_gene_reference.txt
- report.txt
- pathway_list.txt

ensembl_database_release.txt

The current release of the Ensembl database for the chosen species, e.g. *Mus musculus*. Although this uses the programmatic interface of Ensembl, it can be used to identify which release was used to generate the list of genes in the QTL region or mapping of Affymetrix probeids identifiers.

Included in

Feeds into

Included in

Published in

Logs

produces



complete.ppt



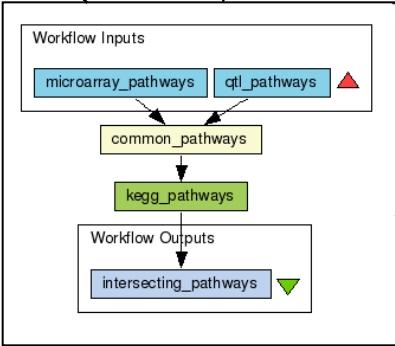
Included in

Metadata

Included in

Slides

Paper



Common pathways Workflow 13

produces

Published in

Primer Name	Left Flank Nucleotide sequence	Right Flank Nucleotide sequence
DAXX_1274_1812	CAGGAGGAATGGCGAGTG	AGCTTAGTCTTCCCAAGCC
DAXX_140754_456_1070	CTTGATAGGATTGGGACTGGG	TCTCTCTCTCTCTCTCTC
DAXX_2270_2720	TGGCAGAGAGATGGTTC	ATGGTCAAGGGAAGGGAAA
DAXX_2644_3187	TGTGTGATTGGCTGTTGTT	GCAAAATACGAGGAGTCTGGG
DAXX_exon5	TCTCTCTCTACCAATCAAA	AGCAGAACTAACACCACAAGG
Daxx_Upst_479_1104	CAGGCTTCTCATCAACACC	TGTCTATGCTGTGCGAGG

Results



Tags (19)

Creator tags

affymetrix | african trypanosomiasis | cattle | data-driven | disease | entrez | genotype | Kegg Pathways | KeggID | link-integration | **microarray** | mouse | **pathway** | pathway-driven | phenotype | sleeping sickness | swissprot | uniprot | web services

[edit]

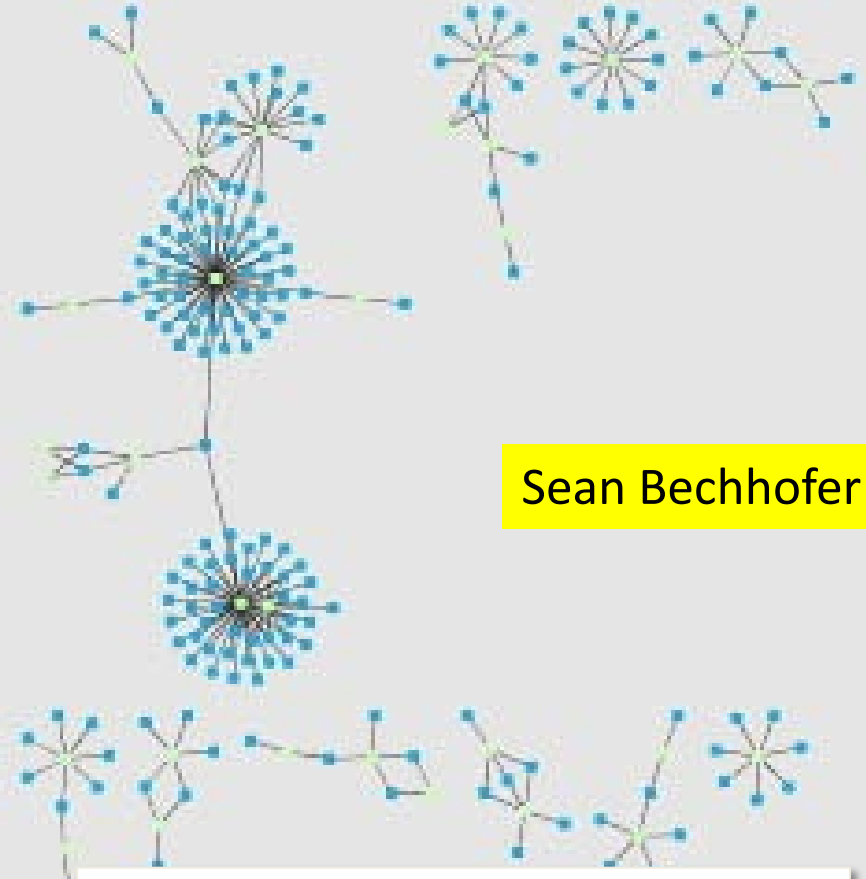
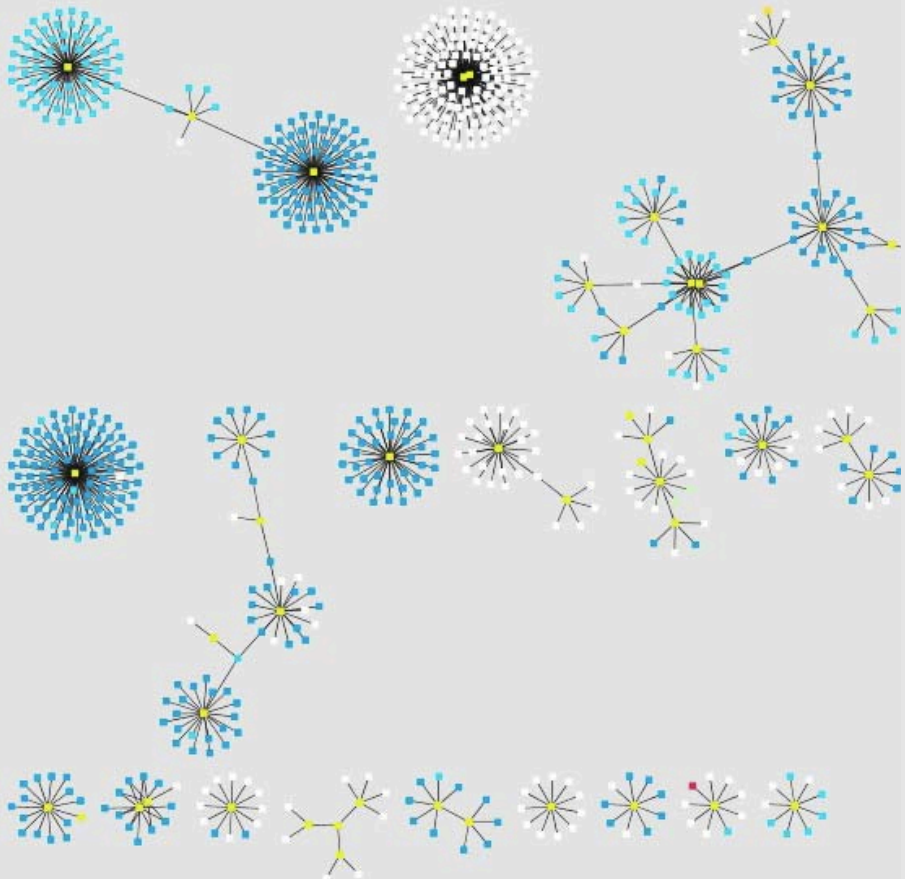
Add Tags

Linked Data support

rdf.myexperiment.org

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs so that they can discover more things





Sean Bechhofer

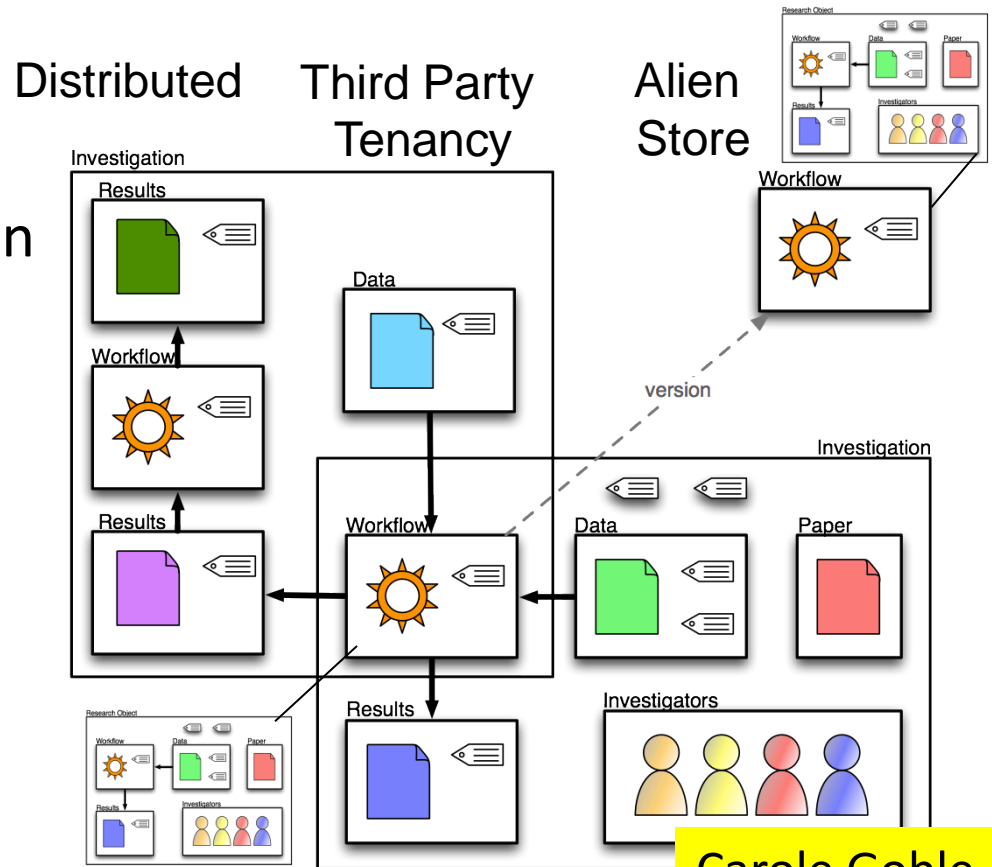
```
SELECT?pack ?contrib
WHERE {
  ?pack rdf:type mepack:Pack.
  ?pack ore:aggregates ?contrib.
}
```

```
SELECT?wf ?uri
WHERE {
  ?wf mebase:has-current-version ?v.
  ?v mecomp:executes-dataflow ?d.
  ?d mecomp:has-component ?c.
  ?c rdf:type mecomp:WSDLProcessor.
  ?c mecomp:processor-uri ?uri.
}
```

Research Objects

Reproducibility, Integrated Publishing

- Workflow
 - Provenance
 - Conservation & Preservation
 - Executable Publication
- Human
 - Credit Tracking
 - Unit of Scholarship
 - Crowd management
- Semantics
 - Acquisition & Publishing
 - Encoding, Encapsulation & Annotation: OAI-ORE, AO...



Carole Goble



Technical Objects Social Objects

Product Family

www.mygrid.org.uk

- The BioCatalogue
Life Science Web Service Registry
www.biocatalogue.org
- MethodBox
Browse datasets, shop for variables and share knowledge
www.methodbox.org
- SysMO-Seek
Find, share and exchange Data, Models and Processes
www.sysmo-db.org

BioCatalogue 

 MethodBox



nature
biotechnology

 Galaxy

#preserving #linking #using #sharing

JCDL 2012

ACM/IEEE - CS Joint Conference
on Digital Libraries

HUB BUB 2012
the HUBzero conference 

SocialObjects
2011
October 9, 2011
Boston, MA, USA

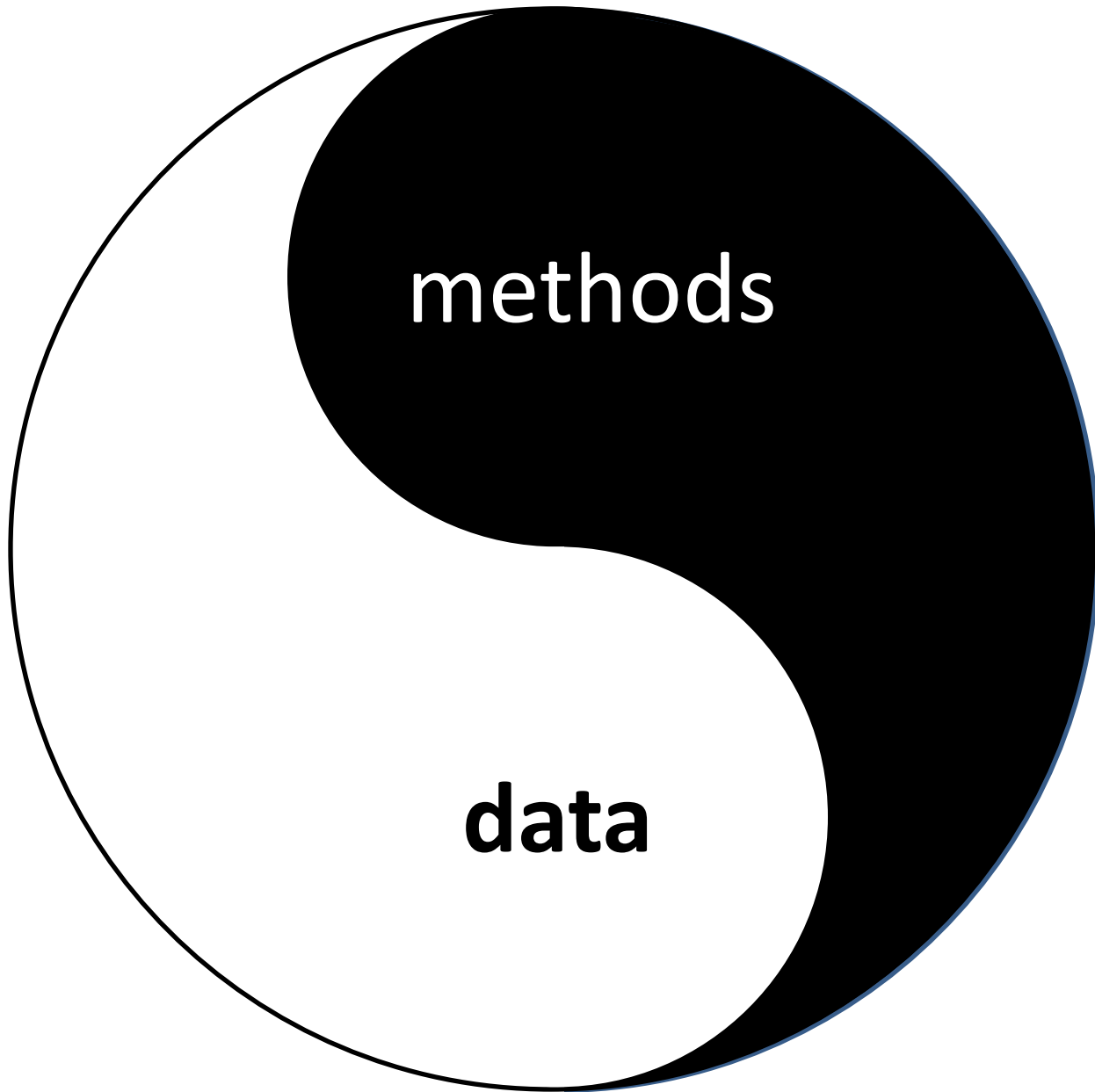
FORCE 11
the Future of Research Communications and e-Scholarship

 IEEE SERVICES 2011
The 7th World Congress on Services
July 4-9, 2011, Washington DC, USA

OR 2010

Bio·IT World
CONFERENCE & EXPO '10

WEBIST
2012



methods

data

Beyond the PDF

Workshop: January 19-21, 2011
University of California San Diego

- Program
- Discussion
- Notes from the breakout session
- Workshop Papers
- Webcasts
- Blogs and reviews
- Demos
- Attendees
- Files
- SMA Prototype Files
- Topics
- Links
- Publications
- The Team
- Photos
- Calendar of Upcoming Events
- Sponsors
- Sitemap

Template tips

[Learn more about working with templates.](#)

[How to change this sidebar.](#)

Navigation

Workshop: January 19-21, 2011 University of California San Diego



ELSEVIER

executable paper grand challenge

knowledge e

Home News About the Challenge

The Winners of the Executable Paper Challenge



Pictured left to right: Ann Gabriel, Elsevier; Piotr Nowakowski and Eryk Ciepiela, ACC CYF

Elsevier, a leading publisher of scientific journals, is the publisher of the Executable Paper Challenge. The winners of the challenge are the authors of the paper "Reproducing computer science research" presented at the Conference on Computational Science and e-Scholarship.

The winners, selected from a pool of 70...

FORCE11

the Future of Research Communication

About Force 11 Publications Editors' Pict

Force11 is a virtual community working to improve research communications through advanced tools and practices. We invite you to join us.

Force 11 Publications

[Force11 White Paper: Improving Future Research Communication and e-Scholarship](#)
[Read More](#)

Microsoft Research

Home Our Research Connections Careers
About Us News Media Resources Events Community

2011 eScience Workshop: Transforming Scholarly Communication

October 23-25, 2011 | Cambridge, Massachusetts

Home [Agenda](#) [Videos](#)



View videos of the presentations from the 2011 eScience Workshop, Transforming Scholarly Communication. Microsoft Research and Harvard University sponsored this invitation-only event from October 23 to 25, 2011, at the Microsoft New England Research and Development Center in Cambridge, Massachusetts. The workshop—co-sponsored in close collaboration with the Alfred P. Sloan Foundation, and the Gordon and Betty Moore Foundation—convened experts in scholarly communication from academia, the publishing industry, scholarly and learned societies, and funding agencies.



Harvard University

The workshop identified and promoted the adoption of new platforms and approaches to facilitate research in the eSciences. In response to growing interest in new forms of communication, the workshop helped participants transition to these new paradigms effectively and sustainably.



Microsoft Research New England Research and Development Center

Related Website

[tumblr Site](#)

Related Links

- [Jim Gray eScience Award](#)
- [eScience at Microsoft](#)
- [eScience Group](#)
- [The Fourth Paradigm](#)

Related Events

- [IEEE International Conference on e-Science](#)
- [2011 eScience in Action Workshop](#)
- [eScience Workshop 2010](#)

Contact Us

For more information, contact esci@microsoft.com

The R dimensions



Reusable. The key tenet of Research Objects is to support the sharing and reuse of data, methods and processes.

Repurposeable. Reuse may also involve the reuse of constituent parts of the Research Object.

Repeatable. There should be sufficient information in a Research Object to be able to repeat the study, perhaps years later.

Reproducible. A third party can start with the same inputs and methods and see if a prior result can be confirmed.

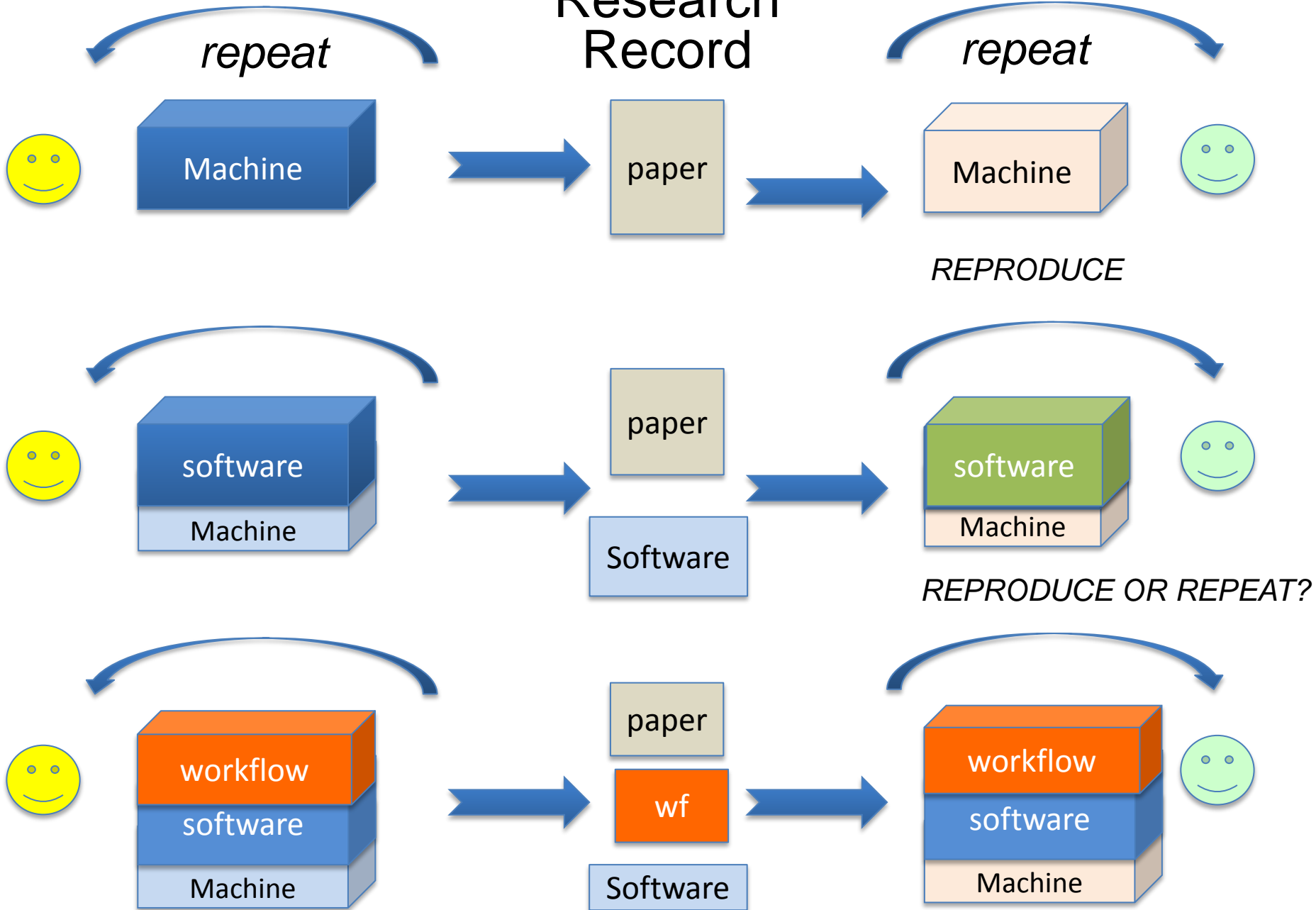
Replayable. Studies might involve single investigations that happen in milliseconds or protracted processes that take years.

Referenceable. If research objects are to augment or replace traditional publication methods, then they must be referenceable or citeable.

Revealable. Third parties must be able to audit the steps performed in the research in order to be convinced of the validity of results.

Respectful. Explicit representations of the provenance, lineage and flow of intellectual property.

Research Record



The Executable Thesis



PhD Student



new data

executable
thesis

new results

A futuristic cityscape with a flying vehicle and industrial structures. The scene is set in a high-tech environment with a large, multi-tiered industrial building on the left, featuring green glowing panels and complex piping. In the center, a large, metallic, disc-shaped flying vehicle with a bright light on its underside is hovering in the sky. The background shows a dense city with various buildings and a body of water. The sky is a mix of blue and white, suggesting a bright, sunny day.

Notifications and automatic re-runs

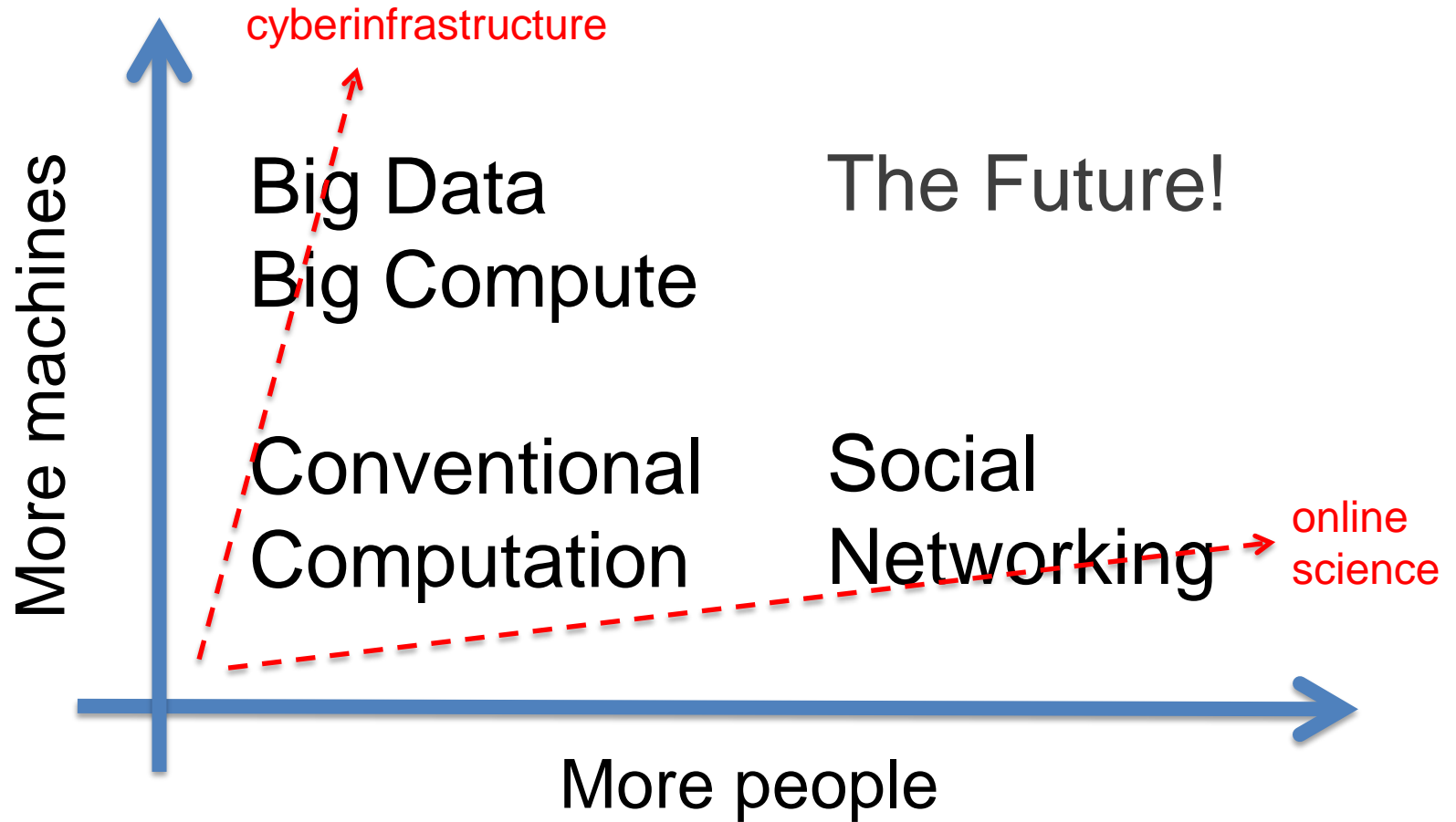
Autonomic
Curation

Self-repair

New research?

Machines are users too

A Big Picture



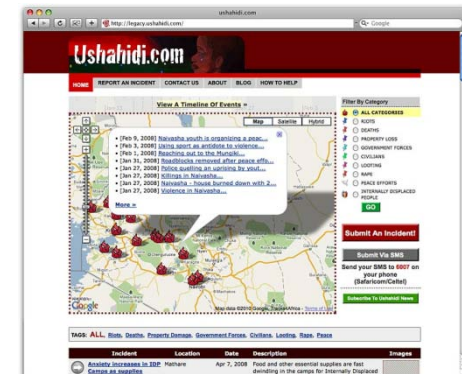
The Order of Social Machines

Real life is and must be full of all kinds of social constraint – the very processes from which society arises. Computers can help if we use them to create abstract social machines on the Web: processes in which the people do the creative work and the machine does the administration... The stage is set for an evolutionary growth of new social engines.

Berners-Lee, *Weaving the Web*, 1999

An Example Social Machine

- The Kenyan election on the 27th December 2007
- wave of riots, killings and turmoil
- African blogger Erik Hersman read a post by another blogger Ory Okolloh
- Resulted in Ushahidi...
- “Nobody Knows Everything, but Everyone Knows Something.”
- Local observers to submit reports using the Web or SMS messages from mobile phones



Ushahidi Variants

Washington Snowmageddon

The interface features a blue header with the title and 'PICnet' logo. Below is a navigation bar with 'HOME', 'REQUESTS/REPORTS', 'FIND/OFFER HELP', and 'WHAT'S ALL THIS ABOUT?'. A 'FILTERS' section includes 'REPORTS', 'NEWS', 'PICTURES', 'VIDEO', 'ALL', 'VIEWS', 'CLUSTERS', and 'TIME'. The main area shows a map of Washington with numerous black dots representing reports. Below the map, there are three numbered steps: 1. Report a problem or solution (with an icon of a person shouting), 2. Look for reports near you (with an icon of a person looking at a phone), and 3. Use the report's comment section to connect with others (with an icon of two people talking).

Japan Fukushima



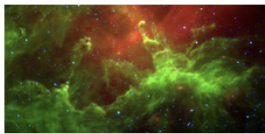
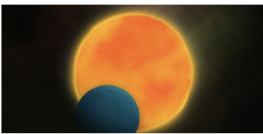
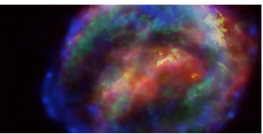
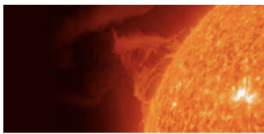
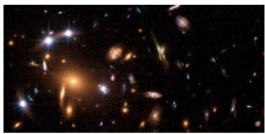
Port au Prince Haiti



Middle East Gaza



The homepage features the Ushahidi logo and navigation links for 'About Us', 'Products', 'Get Involved', and 'Downloads'. A central banner reads: 'We're a non-profit tech company that changes the way information flows in the world.' Below the banner are three sections: 'What is the Ushahidi Platform' (with a 'Watch The Video' button), 'Manage the River of Information' (featuring 'GETTING STARTED WITH SWIFT RIVER' and 'Watch The Video' buttons), and 'Crowdsourcing in "the cloud"' (featuring the 'CROWDMAP' logo).



How do galaxies form?
 NASA's Hubble Space Telescope archive provides hundreds of thousands of galaxy images.

GALAXY ZOO

Explore the surface of the Moon
 We hope to study the lunar surface in unprecedented detail.

MOON ZOO

Study explosions on the Sun
 Explore interactive diagrams to learn out about the Sun and the spacecraft monitoring it.

SOLAR STORMWATCH

How do galaxies merge?
 One important area of research in astronomy studies the role of interacting galaxies.

GALAXY ZOO

Search for exploding stars
 Help to find Supernovae, astronomers are ready to follow up.

GALAXY ZOO

Find planets around stars
 Lightcurve changes from the Kepler spacecraft can indicate transiting planets.

planethunters.org

How do stars form?
 We're asking you to help us find and draw circles on infrared image data from the Spitzer Space Telescope.

THE MILKY WAY PROJECT

ZOO NIVERSE

REAL SCIENCE ONLINE

Climate



Model Earth's climate using wartime ship logs

Help scientists recover worldwide weather observations made by Royal Navy ships.

oldWeather

Humanities



Study the lives of ancient Greeks

The data gathered by Ancient Lives helps scholars study the Oxyrhynchus collection.

ANCIENT LIVES

Nature



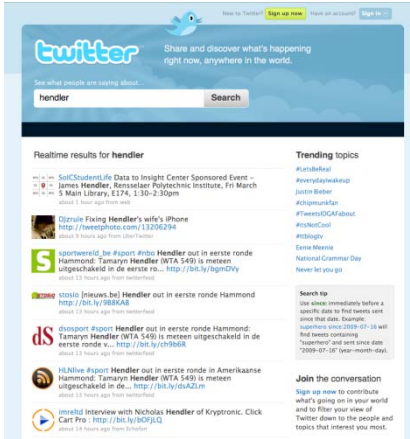
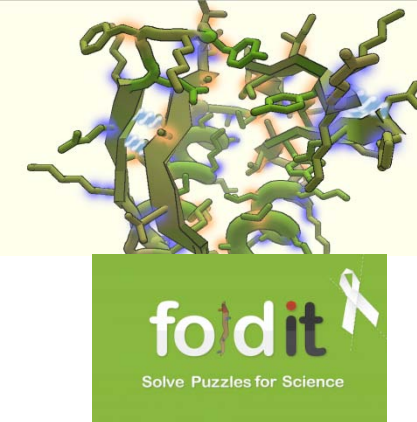
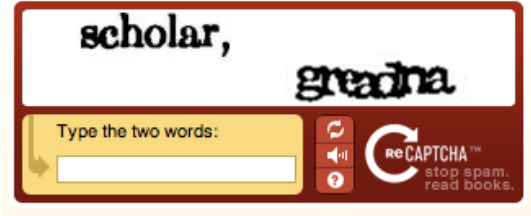
Hear Whales communicate

You can help marine researchers understand what whales are saying

WHALEFM



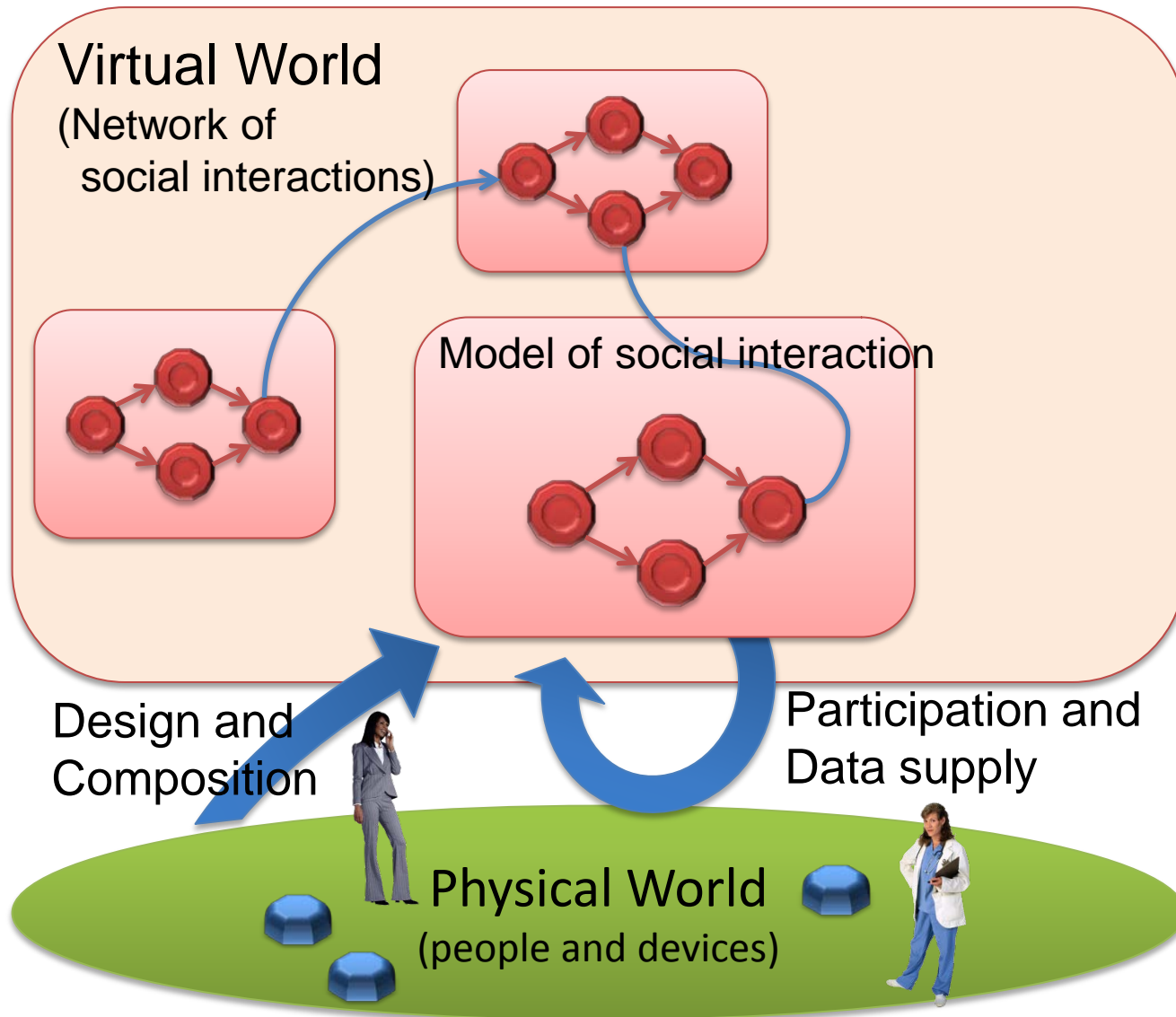
Some other machines?



Dimensions

- Number of people
- Number of machines
- Scale of data
- Varieties of data
- Type of machine problem solving
- Type of human problem solving
- Empowering of individuals, groups, crowds
- Time criticality
- Extent of wide area communication
- Need for urgent mobilization
- Specification of goal state

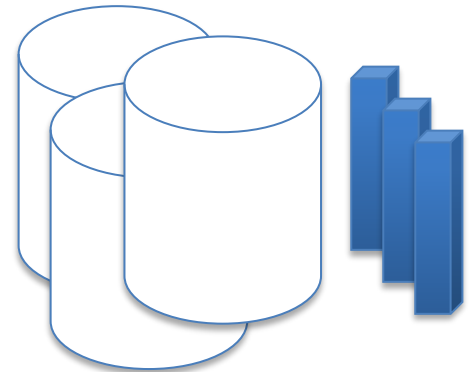
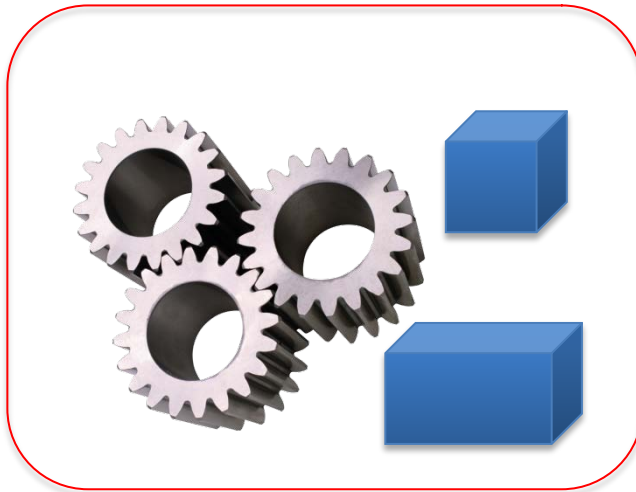
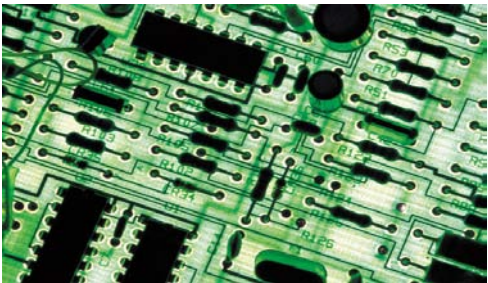
Building a Social Machine



hubzero

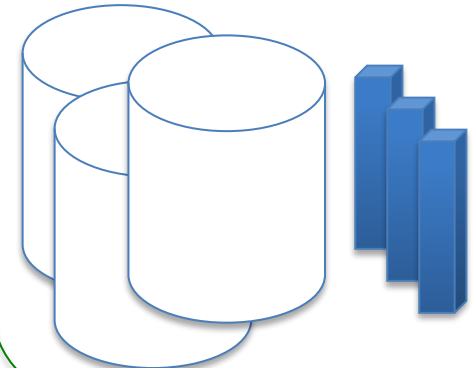
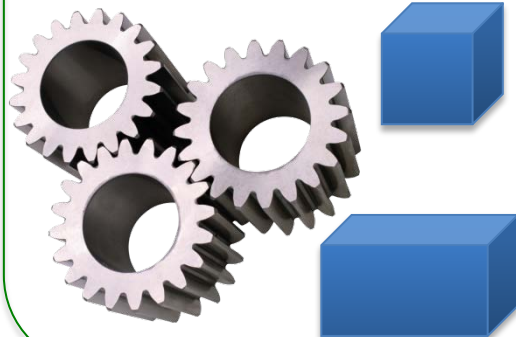
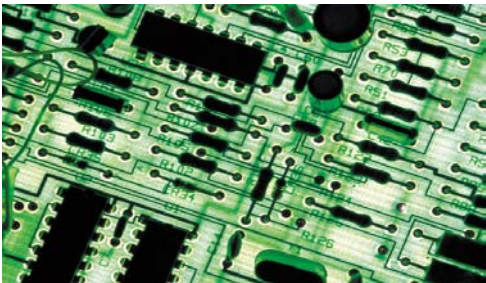
- A platform for creating social machines
- A community of builders and users
- Machine assisted human problem solving
- Future quadrant and moving up





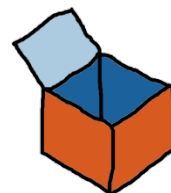
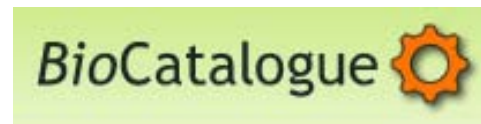
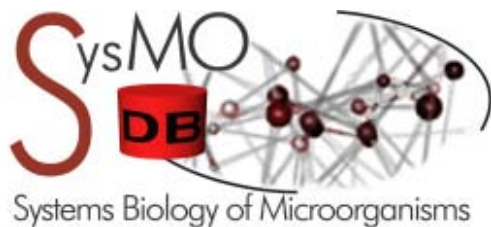
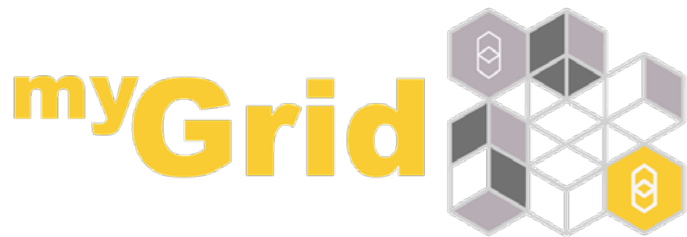


The users of a website, the website, and the interactions between them, together form our fundamental notion of a “machine”



Discussion

- myExperiment
 - Workflows as social objects
- Research Objects are emerging
 - Reproduce, Reconstruct
 - Curate, Preserve
- HUBzero is a Social Machine platform
 - Noshir Contractor has commenced analysis!



MethodBox



Digital Social Research
Harnessing technology for social science research

Links

- myExperiment project wiki
<http://wiki.myexperiment.org/>
- Workflow Forever project (Wf4Ever)
<http://www.wf4ever-project.org/>
- Future of Research Communication (FORCE11)
<http://force11.org/>
- Web Science
<http://webscience.org/>
- hubzero
<http://hubzero.org/>

- D. De Roure, C. Goble and R. Stevens. **The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflows** *Future Generation Computer Systems* 25, pp. 561-567.
- S. Bechhofer, I. Buchan, D De Roure et al. **Why linked data is not enough for scientists,** *Future Generation Computer Systems*
- D. De Roure, David and C. Goble, **Anchors in Shifting Sand: the Primacy of Method in the Web of Data.** *WebSci10*, April 26-27th, 2010, Raleigh, NC, US.
- D. De Roure, S. Bechhofer, C. Goble and D. Newman, **Scientific Social Objects,** *1st International Workshop on Social Object Networks (SocialObjects 2011).*
- D. De Roure, K. Belhajjame, P. Missier, P. et al **Towards the preservation of scientific workflows.** *8th International Conference on Preservation of Digital Objects (iPRES 2011).*
- Carole A. Goble, David De Roure and Sean Bechhofer **Accelerating scientists' knowledge turns.** Will be available at www.springerlink.com
- Khalid Belhajjame, Oscar Corcho, Daniel Garijo et al **Workflow-Centric Research Objects: First Class Citizens in Scholarly Discourse,** SePublica2012 at ESWC2012, Greece, May 2012
- Kevin R. Page, Ben Fields, David De Roure et al **Reuse, Remix, Repeat: The Workflows of MIR,** 13th International Society for Music Information Retrieval Conference (ISMIR 2012) Porto, Portugal, October 8th-12th, 2012
- Jun Zhao, Jose Manuel Gomez-Perezy, Khalid Belhajjame et al, **Why Workflows Break - Understanding and Combating Decay in Taverna Workflows,** eScience 2012, Chicago, October 2012



david.deroure@oerc.ox.ac.uk

www.oerc.ox.ac.uk/people/dder

www.scilogs.com/eresearch

@dder

www.myexperiment.org/packs/335

www.slideshare.net/davidderoure/myexperiment-and-the-rise-of-social-machines