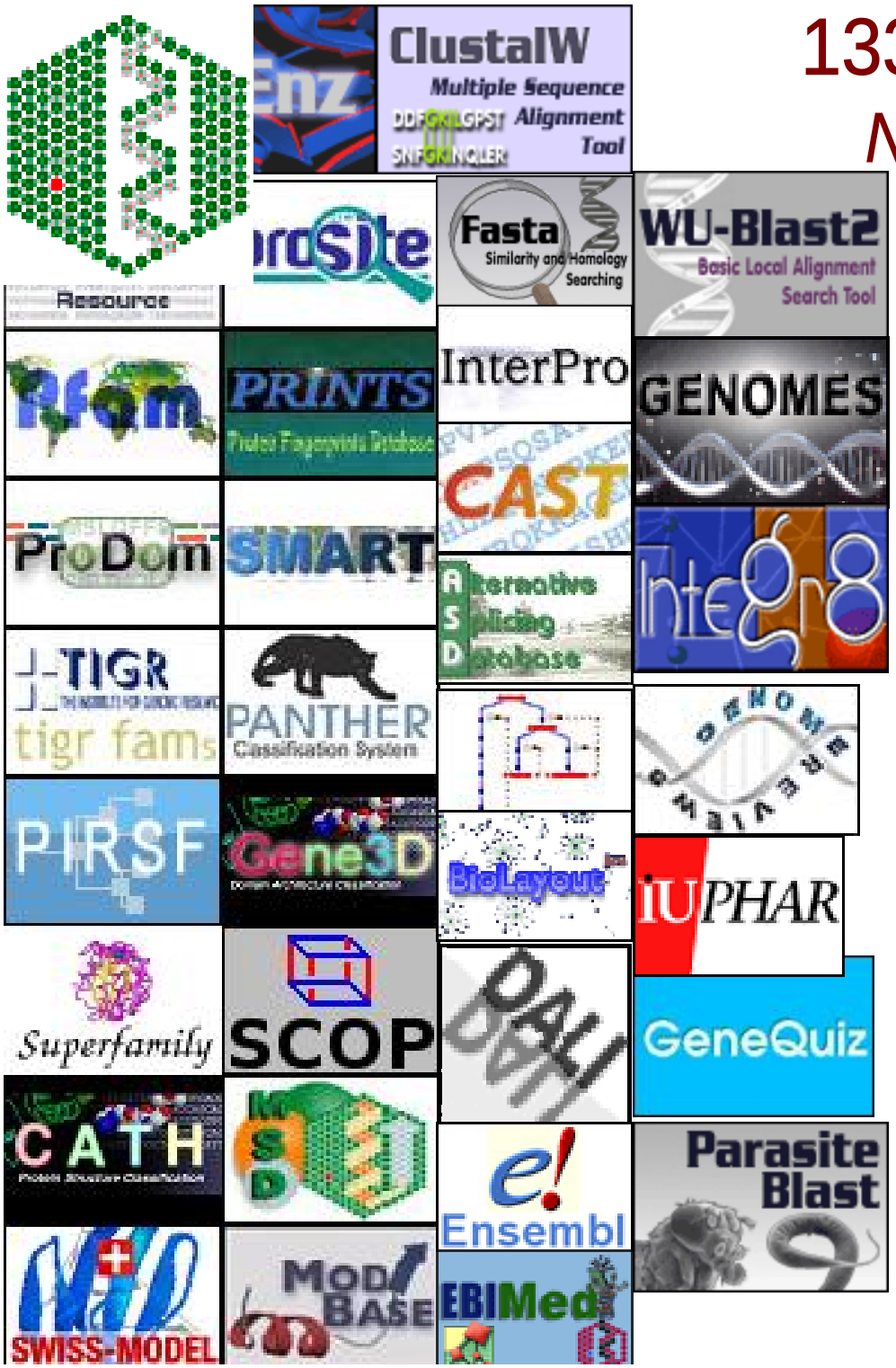
Computation Institute

Data is knowledge, knowledge data …

Ian Foster

Argonne NATIONAL LABORATORY | THE UNIVERSITY OF CHICAGO

www.ci.anl.gov
www.ci.uchicago.edu



1330 molecular bio databases

Nucleic Acids Research Jan 2011 (96 in 2001, 1070 in 2010)

- Proteomics
- Genomics
- Transcriptomics
- Protein sequence prediction
- Phenotypic studies
- Phylogeny
- Sequence analysis
- Protein structure prediction
- Protein-protein interaction
- Metabolomics
- Model organism collections
- Systems biology
- Health epidemiology
- Organisms
- Disease ....

Slide: Carole Goble

The BioCatalogue: providing a curated catalogue of Life Science Web Services

t Activity

joined the
e

n added a tag
Service:
ice

n added a tag
Service:
ice

n added a tag
Service:
ice

n added a tag
Service:
ice

n added a tag
Service:
ice

n added a tag
Service:
ice

n added a tag
Service:
ice

n added a tag

The BioCatalogue currently has **2053 services**, **148 service providers** and **509 members**

"Web Services are hard to find"

DISCOVER

- Find the right Web Service
- Powerful search and filtering
- Information from providers and community

More info

"My Web Services are not visible"

REGISTER

- Easily register Web Services
- Instantly available to everyone
- Providers can advertise, describe and monitor their Services

More info

"Web Services are poorly described"

ANNOTATE

- Anyone can describe and annotate
- Ongoing expert curation
- Social curation by the community

More info

"Web Services are volatile"

MONITOR

- Services change and get outdated
- BioCatalogue monitors Services
- Monitors availability and reliability

More info

Site An

BioCatalogue
hour on Wed
09:30 (BST)
By Jiten Bhaga

Sandbox dow
21st April 201
By Jiten Bhaga

Sandbox not
weekend
By Jiten Bhaga

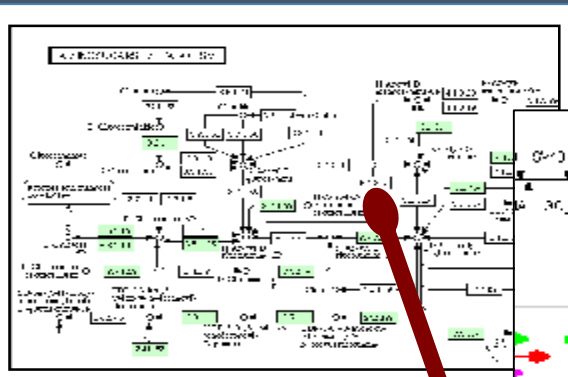
Webinar on S
Scientists: Bu
e-Laboratorie
By Jiten Bhaga

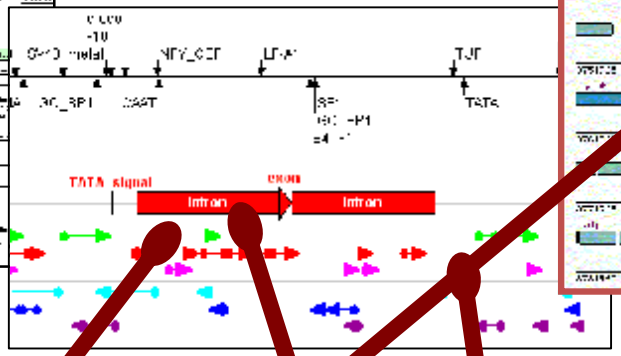
BioCatalogue
Resources
By Jiten Bhaga

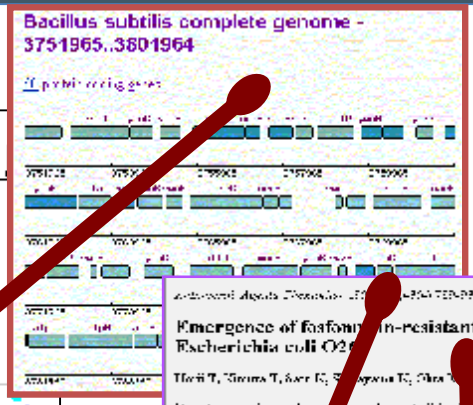
Late

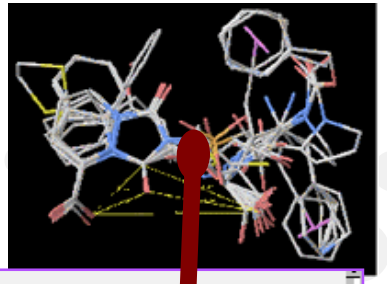
Psic

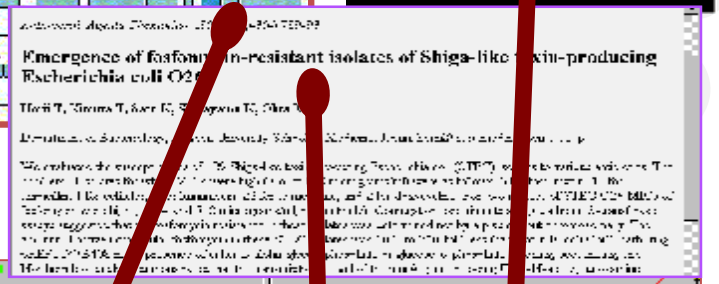
Knowledge results from data integration

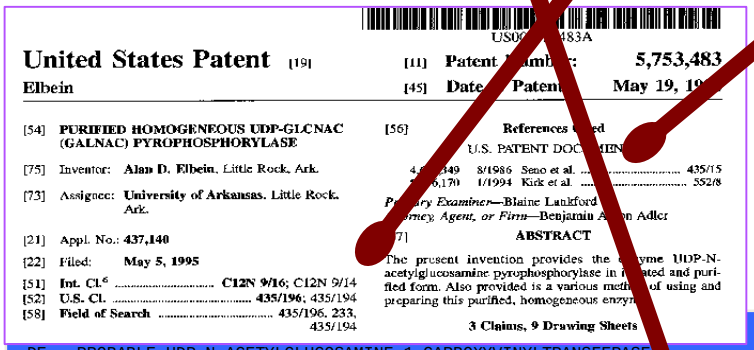












United States Patent [19]
Elbein

[11] Patent Number: 5,753,483
[45] Date of Patent: May 19, 19

[54] PURIFIED HOMOGENEOUS UDP-GLCNAC (GALNAC) PYROPHOSPHORYLASE

[75] Inventor: Alan D. Elbein, Little Rock, Ark.

[73] Assignee: University of Arkansas, Little Rock, Ark.

[21] Appl. No.: 437,140

[22] Filed: May 5, 1995

[51] Int. Cl.[6] C12N 9/16; C12N 9/14
[52] U.S. Cl. 435/196; 435/194
[58] Field of Search 435/196, 233, 435/194

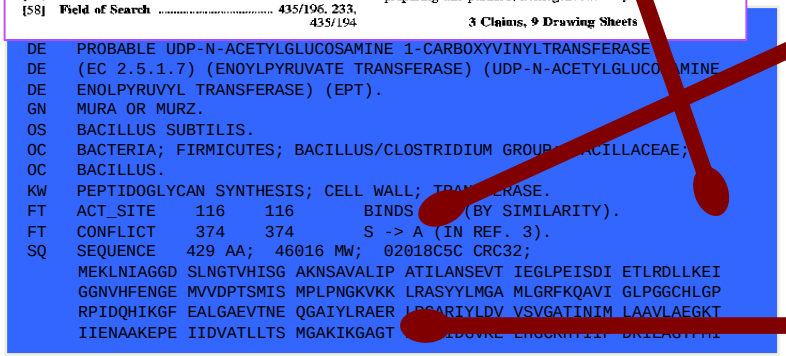
[56] References Cited
U.S. PATENT DOCUMENTS
4,[illegible]349 8/1986 Seno et al. 435/15
5,[illegible]170 1/1994 Kick et al. 552/8

Primary Examiner—Blaine Lankford
Attorney, Agent, or Firm—Benjamin [illegible] Adler

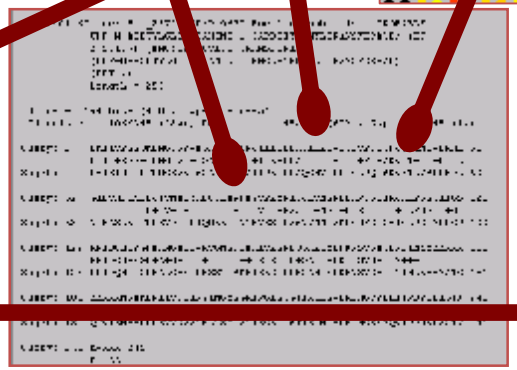
[57] ABSTRACT

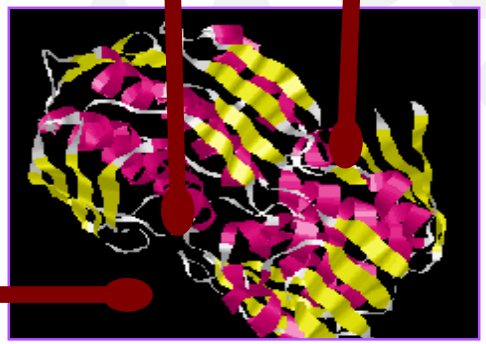
The present invention provides the enzyme UDP-N-acetylglucosamine pyrophosphorylase in [illegible]ated and purified form. Also provided is a various methods of using and preparing this purified, homogeneous enzyme.

3 Claims, 9 Drawing Sheets



```
DE   PROBABLE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE
DE   (EC 2.5.1.7) (ENOYLPYRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE
DE   ENOLPYRUVYL TRANSFERASE) (EPT).
GN   MURA OR MURZ.
OS   BACILLUS SUBTILIS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
KW   PEPTIDOGLYCAN SYNTHESIS; CELL WALL; TRANSFERASE.
FT   ACT_SITE    116    116       BINDS [illegible] (BY SIMILARITY).
FT   CONFLICT    374    374       S -> A (IN REF. 3).
SQ   SEQUENCE   429 AA;  46016 MW;  02018C5C CRC32;
     MEKLNIAGGD SLNGTVHISG AKNSAVALIP ATILANSEVT IEGLPEISDI ETLRDLLKEI
     GGNVHFENGE MVVDPTSMIS MPLPNGKVKK LRASYYLMGA MLGRFKQAVI GLPGGCHLGP
     RPIDQHIKGF EALGAEVTNE QGAIYLRAER LRGARIYLDV VSVGATINIM LAAVLAEGKT
     IIENAAKEPE IIDVATLLTS MGAKIKGAGT [illegible]
```





[source: GlaxoSmithKline]

Didn't find what you need? Click here to search external workflow repositories.

Search Results

14 Workflows found for "cabig".

Note: some items may not be visible to you, due to viewing permissions.

« previous 1 2 next »

Taverna 2

Lymphoma type prediction based on microarray data (v7)

View

Download (v7)

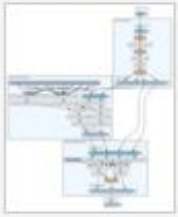
Created: 11/05/10 @ 19:04:30 | **Last updated:** 11/05/10 @ 19:04:32

Credits: Wei Tan Madduri Stian Soiland-Reyes

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Wei Tan



Scientific value Using gene-expression patterns associated with DLBCL and FL to predict the lymphoma type of an unknown sample. Using SVM (Support Vector Machine) to classify data, and predicting the tumor types of unknown examples. Steps Querying training data from experiments stored in caArray. Preprocessing, or normalize the microarray data. Adding training and testing data into SVM service to get classification result.

Rating: 0.0 / 5 (0 ratings) | **Versions:** 7 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 2

Viewed: 298 times | **Downloaded:** 121 times

Tags (11):

cabig | cagrid | cancer research | classification | gene expression | globus | grid service | lymphoma | lymphoma workflow | microarray | support vector machine

Taverna 2

Lymphoma type prediction based on microarray data (v7)

View
Download (v7)

Created: 11/05/10 @ 19:04:30 | **Last updated:** 11/05/10 @ 19:04:32

Credits: Wei Tan, Madduri, Stian Soiland-Reyes

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Wei Tan

Scientific value Using gene-expression patterns associated with DLBCL and FL to predict the lymphoma type of an unknown sample. Using SVM (Support Vector Machine) to classify data, and predicting the tumer types of unknown examples. Steps Querying training data from experiments stored in caArray. Preprocessing, or normalize the microarray data. Adding training and testing data into SVM service to get classification result.

Rating: 0.0 / 5 (0 ratings) | **Versions:** 7 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 2

Viewed: 295 times | **Downloaded:** 121 times

Tags (11):

cabig | cagrid | cancer research | classification | gene expression | globus | grid service | lymphoma | lymphoma workflow | microarray | support vector machine

Taverna 2

Support-Vector-Machine (SVM) based data classification (v2)

View
Download (v2)

Created: 24/08/10 @ 22:34:31 | **Last updated:** 24/08/10 @ 22:34:32

Credits: Wei Tan

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Wei Tan

Support-Vector-Machine based data classificationSupport-Vector-Machine based data classification using genePattern SVM service, the input should be in genePattern STATML format.

Rating: 0.0 / 5 (0 ratings) | **Versions:** 2 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 0

Viewed: 59 times | **Downloaded:** 18 times

Tags (5):

cabig | cagrid | classification | functional workflow | support vector machine

Taverna 2

Using CQL to query protein sequence data (v1)

View
Download (v1)

Created: 07/05/09 @ 14:26:59

Credits: Stian Soiland-Reyes

Attributions: Using CQL to query protein sequence data

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Stian Soiland-Reyes

To query protein sequence infomation out of 3 caGrid data services: caBIO, CPAS and GridPIR Adapted from http://www.myexperiment.org/workflows/600

Rating: 0.0 / 5 (0 ratings) | **Versions:** 1 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 0

Viewed: 106 times | **Downloaded:** 49 times

Tags (10):

cabig | cabio | cagrid | cancer research | cpas | globus | grid service | gridpir | protein | sequence

Taverna 2

caDSR Data service query in caGrid (v3)

View
Download (v3)

Created: 25/05/10 @ 19:16:58 | **Last updated:** 11/06/10 @ 04:50:15

Credits: Wei Tan

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Wei Tan

This workflow is used as an example in this wiki article: https://wiki.cagrid.org/display/knowledgebase/How+to+Create+caGrid+Workflow+Using+Taverna+2 Tested with Taverna 2.1.2 as of 6/10/2010

Taverna 1

Using CQL to query protein sequence data (v1)

View
Download (v1)

Created: 05/12/08 @ 22:03:03 | **Last updated:** 14/07/09 @ 19:32:17

Credits: Wei Tan

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Wei Tan

To query protein sequence infomation out of 3 caGrid data services: caBIO, CPAS and GridPIR. Scientific value To query protein sequence information out of 3 caGrid data services: caBIO, CPAS and GridPIR. To analyze a protein sequence from different data sources. Steps Querying CPAS and get the id, name, value of the sequence. Querying caBIO and GridPIR using the id or name obtained from CPAS.

Rating: 0.0 / 5 (0 ratings) | **Versions:** 1 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 0

Viewed: 159 times | **Downloaded:** 42 times

Tags (9):

cabig | cabio | cagrid | cancer research | cpas | globus | grid service | gridpir | protein

Taverna 1

hierarchical microarray clustering (v1)

View
Download (v1)

Created: 05/12/08 @ 18:31:53 | **Last updated:** 05/12/08 @ 20:33:37

Credits: Wei Tan

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Wei Tan

To illustrate our caGrid plug-in's application, we tested it with a microarray hierarchical clustering workflow that involves services hosted at multiple institutions. Microarrays are a high-throughput technology used to measure the expression of tens of thousands of genes in different tissues or cells. Scientists represent the data from each microarray via a vector (profile) in which each element represents a gene's expression level. They use clustering analysis to identify sim...

Rating: 0.0 / 5 (0 ratings) | **Versions:** 1 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 0

Viewed: 133 times | **Downloaded:** 54 times

Tags (8):

cabig | cagrid | cancer research | clustering | gene expression | globus | grid service | microarray

Taverna 2

An example workflow with WSRF service (v1)

View
Download (v1)

Created: 28/07/09 @ 17:24:05

Credits: Wei Tan, Stian Soiland-Reyes

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Wei Tan

An example workflow with WSRF service The service used is http://sdgrid.cs.uchicago.edu:8106/wsrf/services/CounterService The service is included in Globus toolkit installation (ws-core). If the it is down you can host your own and modify the workflow definition file to point to your own. The workflow first creates a counter instance and adds the value 10 for two times to the same counter. Therefore the result should be 20.

Rating: 0.0 / 5 (0 ratings) | **Versions:** 1 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 0

Viewed: 116 times | **Downloaded:** 25 times

Tags (5):

cabig | example | globus | grid service | wsrf

Taverna 1

caDSR metadata query in caGrid (v1)

View
Download (v1)

Created: 05/12/08 @ 17:45:17 | **Last updated:** 05/12/08 @ 18:20:25

Credits: Wei Tan

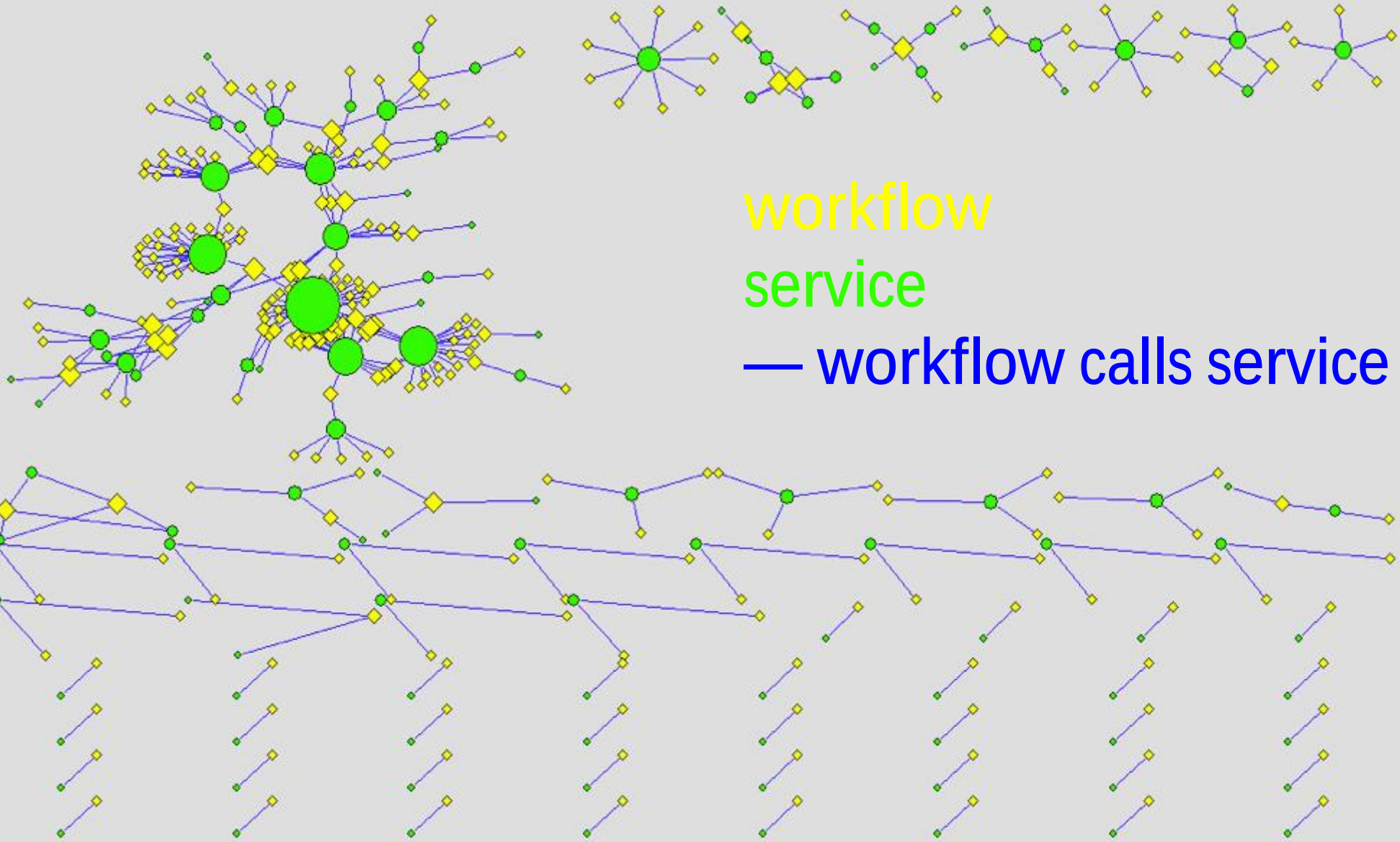
License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

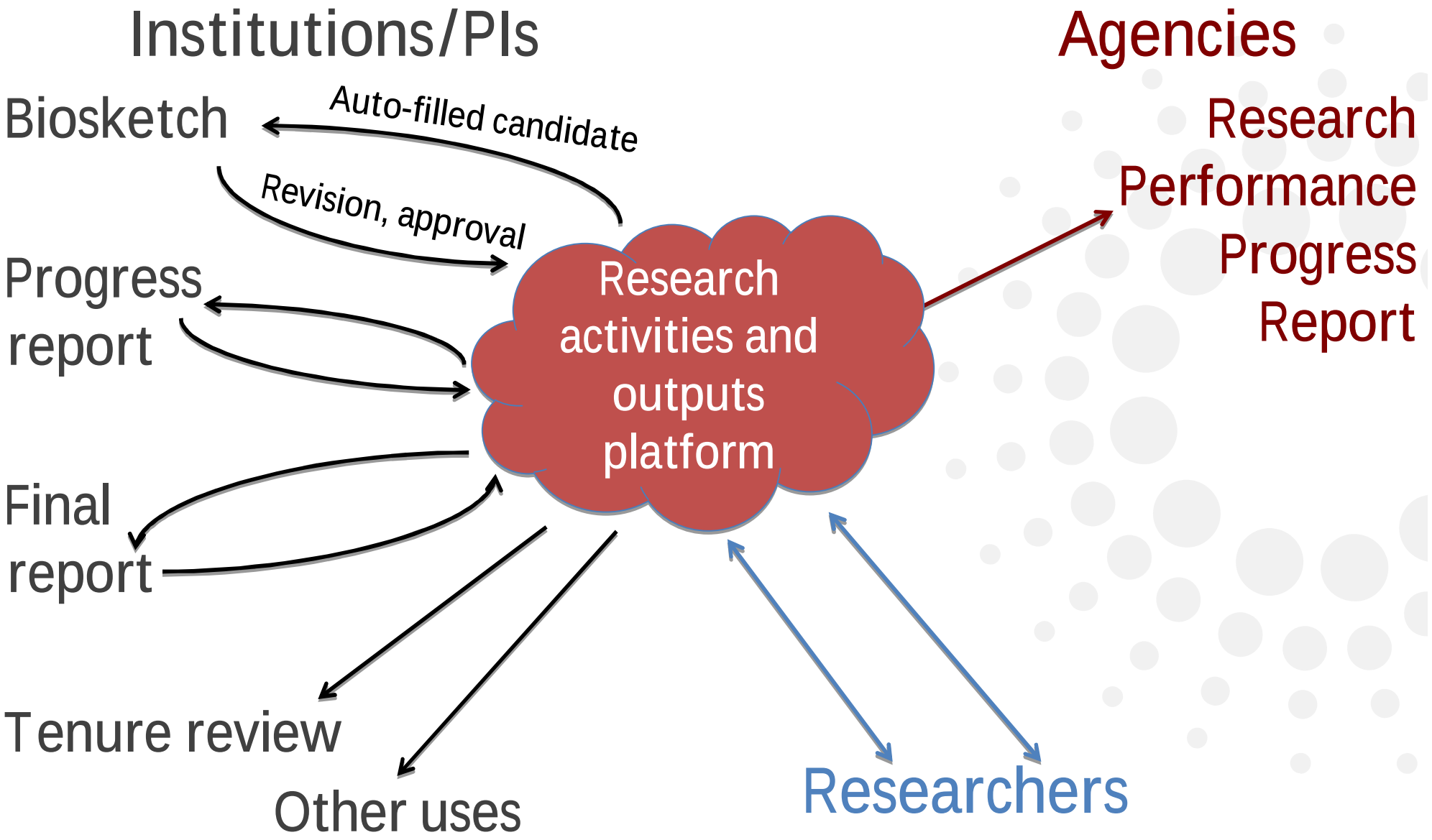
Wei Tan

This workflow shows the coordinated use of two services in caGrid, i.e., the caDSR (Cancer Data Standards Repository) and EVS (Enterprise Vocabulary Services) services. caDSR is to define a comprehensive set of standardized metadata descriptors for cancer research terminology used in information collection and analysis. EVS provides resources and services to meet NCI needs for controlled terminology, and to facilitate the standardization of terminology and information

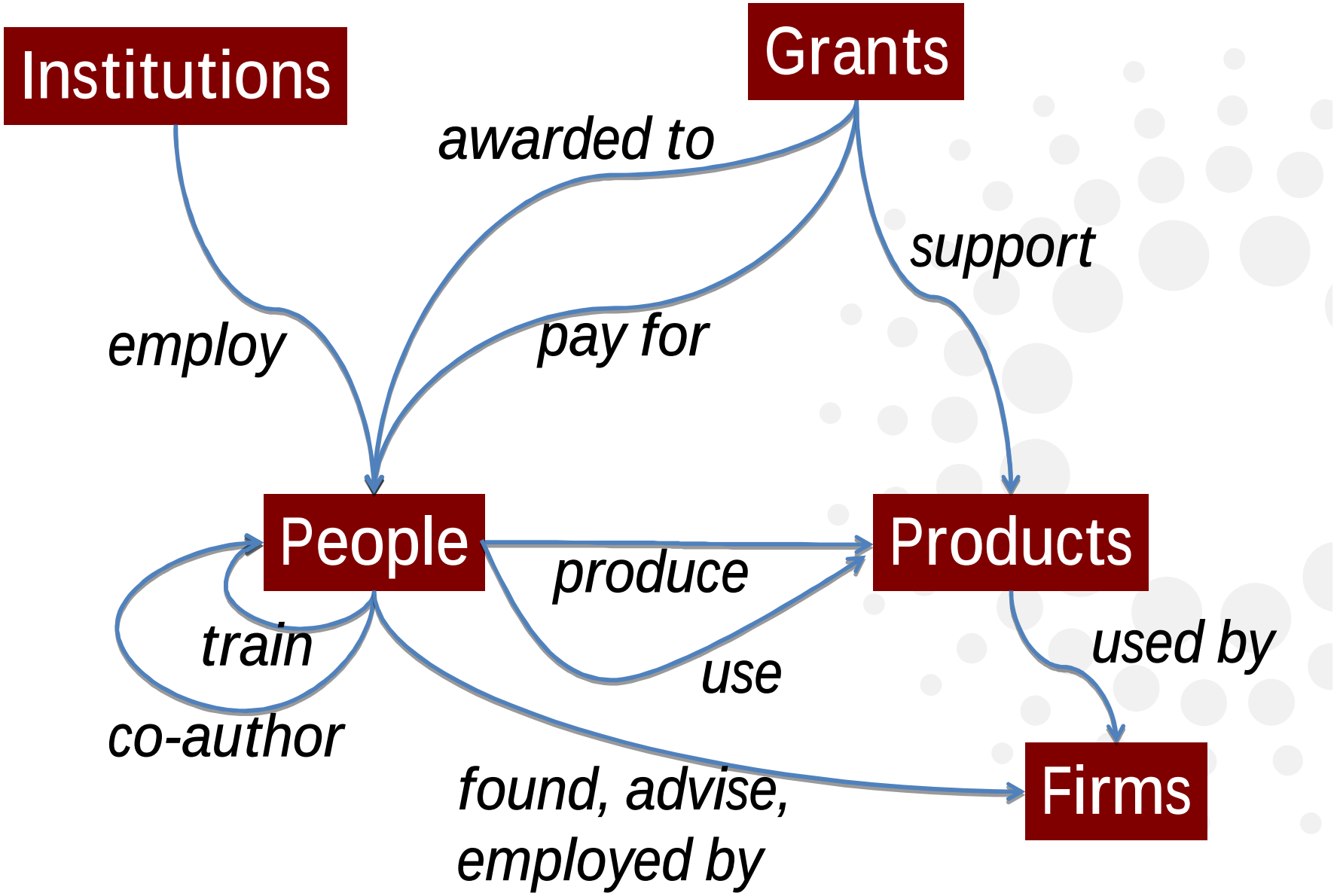
The workflow-service relation [Wei Tan]



Schematic overview of possibilities



Aim for dual use in outcomes data



Knowledge is data ("metaknowledge")

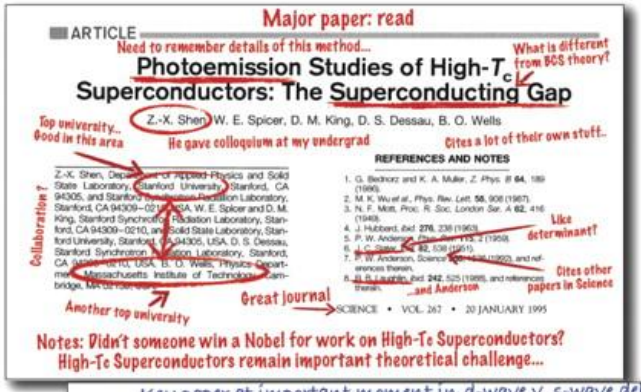
New student

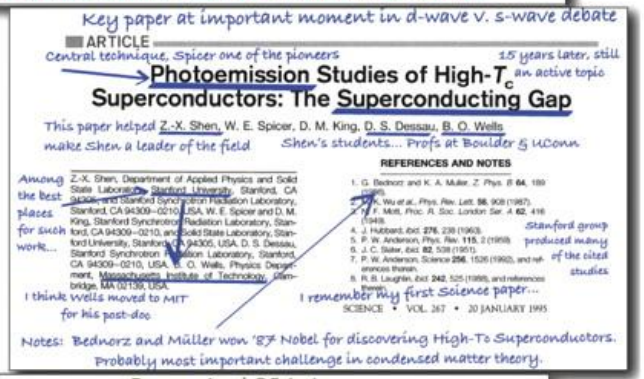
Leading scientist

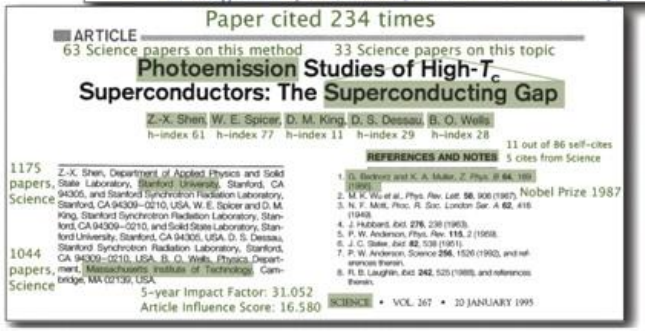
Computer system

Reader

Perception







Evaluation (e.g., authors)

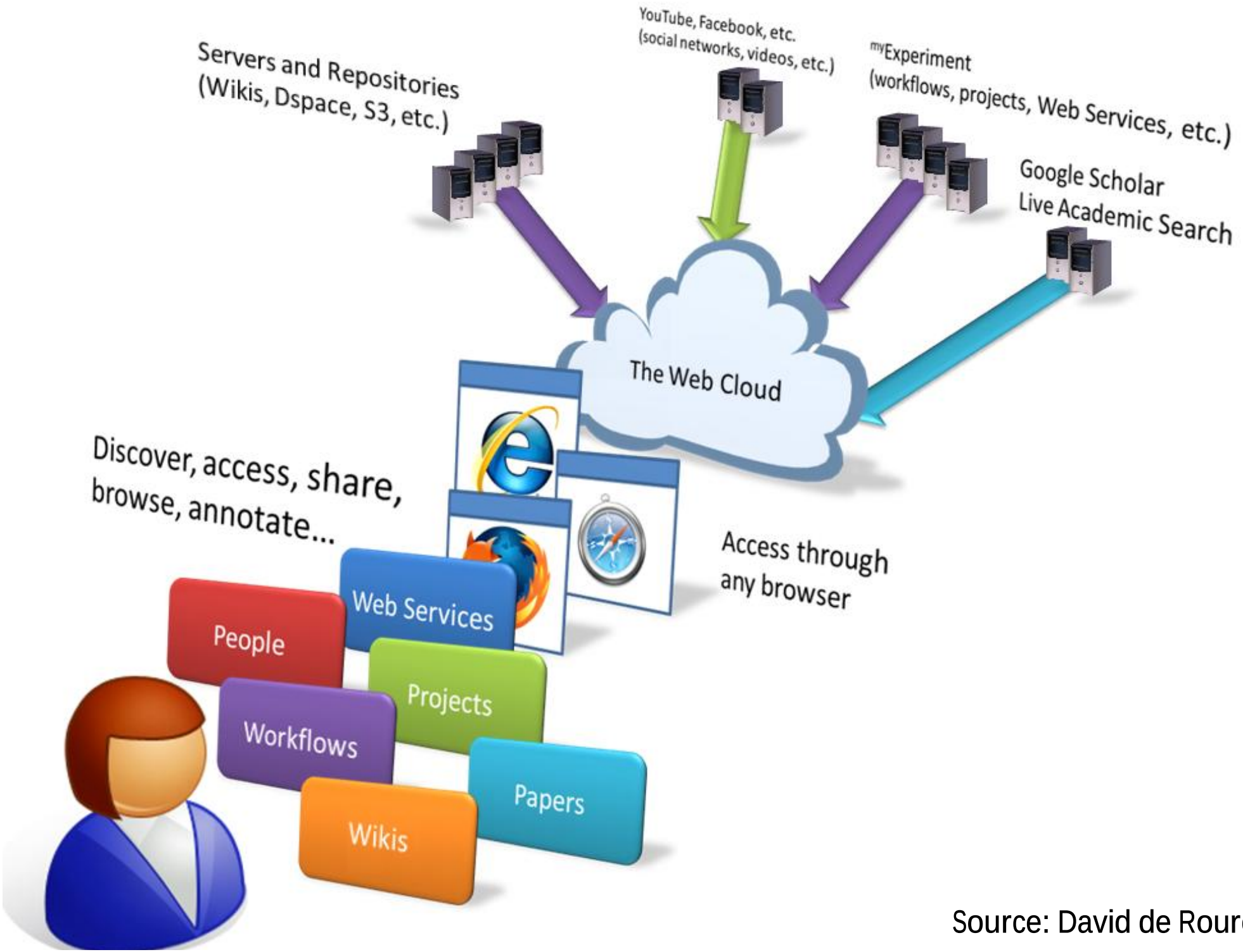
Name	Bin	
W.E. Spicer Z.-X. Shen	Best	Vague, unordered impression of field
D.S. Dessau B.O. Wells D.M. King	Good	

Name	Rank	
W.E. Spicer	1	Detailed, ordered ranking of field
Z.X. Shen	2	
D.S. Dessau	3	
B.O. Wells	4	
D.M. King	5	

Name	Citations	h-index	# papers	Avg. Cites
W.E. Spicer	26,859	77	726	37.0
Z.X. Shen	15,115	61	273	55.4
D.S. Dessau	5,108	29	80	63.9
B.O. Wells	4,002	28	68	58.9
D.M. King	2,347	11	14	167.6

Comparison

J A Evans, J G Foster Science 2011;331:721-725



Source: David de Roure

Computation Institute

Thank you!

foster@uchicago.edu

foster@anl.gov

@ianfoster twitter

Argonne NATIONAL LABORATORY | THE UNIVERSITY OF CHICAGO

www.ci.anl.gov
www.ci.uchicago.edu