R in the cloud

Karim Chine

karim.chine@cloudera.co.uk

Cloud Era Ltd

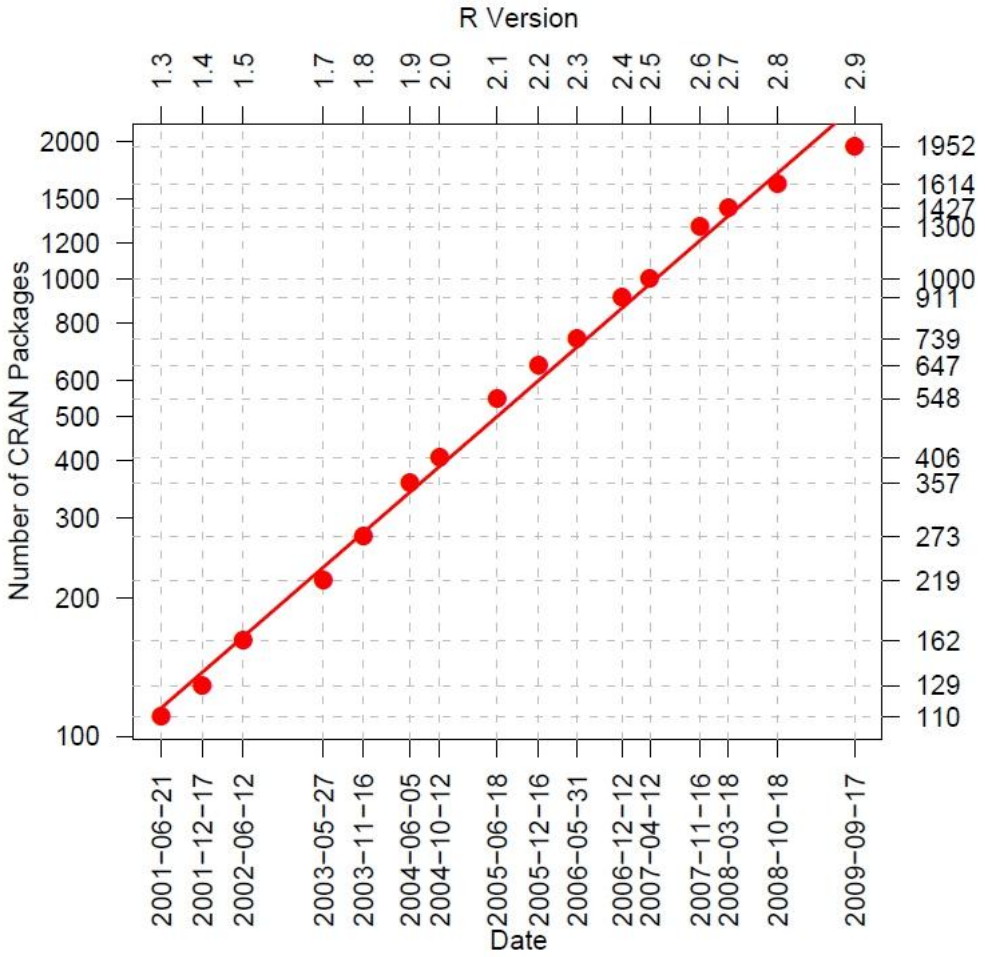
Cambridge – UK

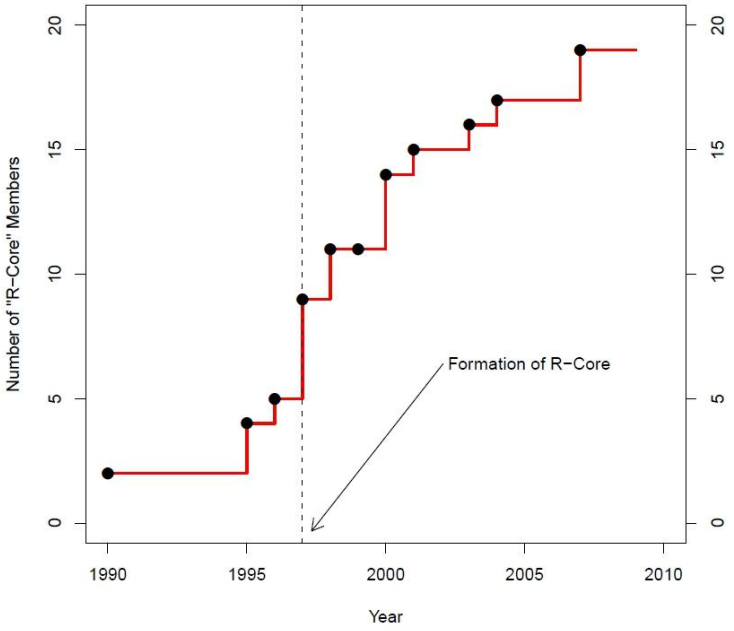
Department of Statistics and Mathematics Seminar

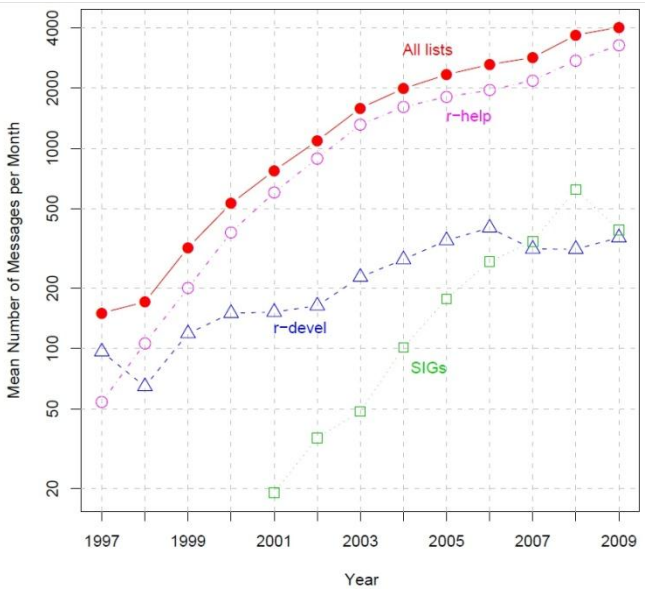
Wirtschaftsuniversitaet Wien

29 April 2010

,*lingua franca* of data analysis







From: John Fox, Aspects of the Social Organization and Trajectory of the R Project, R Journal-Feb 2009

R + Elastic-R : « plug & play » computing environment

Computational Components

R packages : CRAN, Bioconductor, Wrapped C,C++,Fortran code

Scilab modules, Matlab Toolkits, *etc.*

Open source or commercial

Computational User Interfaces

Workbench within the browser

Built-in views / Plugins / Spreadsheets

Collaborative views

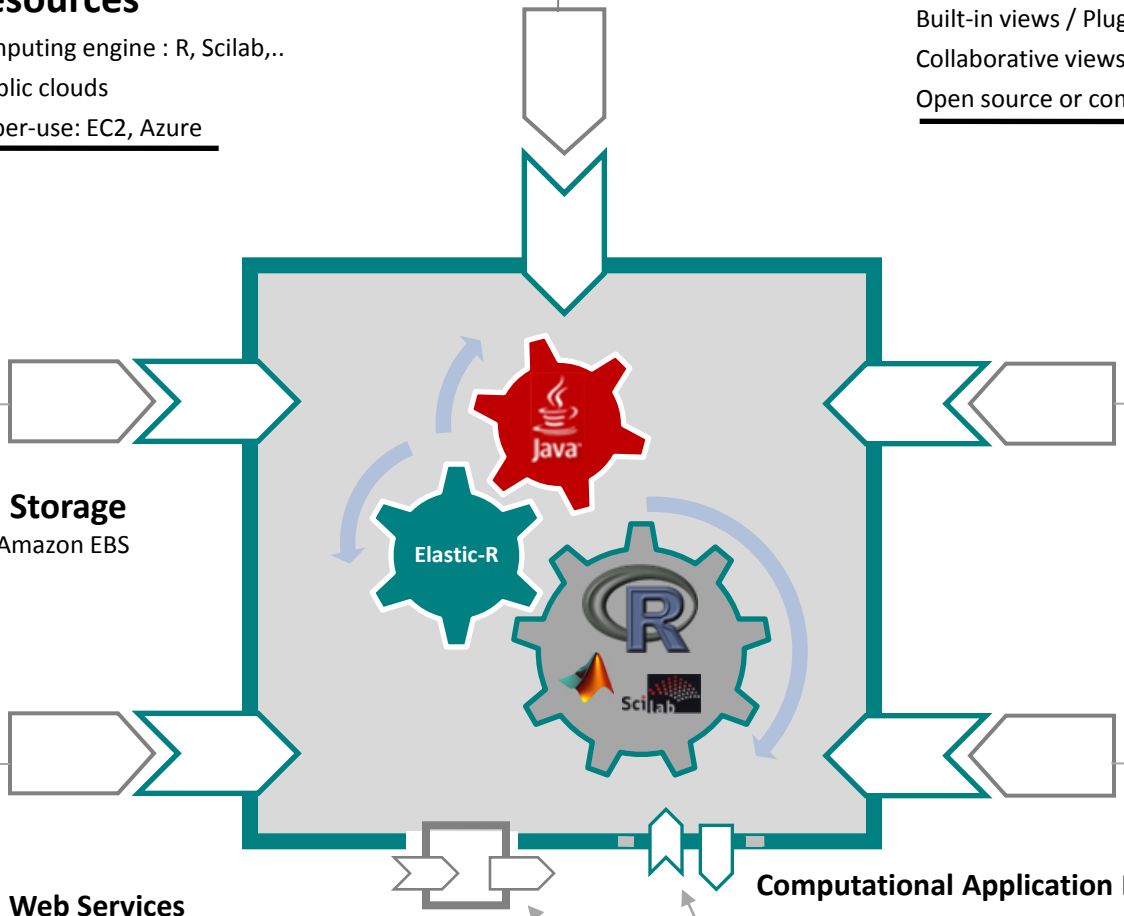
Open source or commercial

Computational Resources

Hardware & OS agnostic computing engine : R, Scilab,..

Clusters, grids, private or public clouds

free: academic grids or pay-per-use: EC2, Azure



Computational Data Storage

Local, NFS, FTP, Amazon S3, Amazon EBS

free or commercial

Computational Scripts

R / Python / Groovy

On client side: interactivity..

On server side: data transfer ..

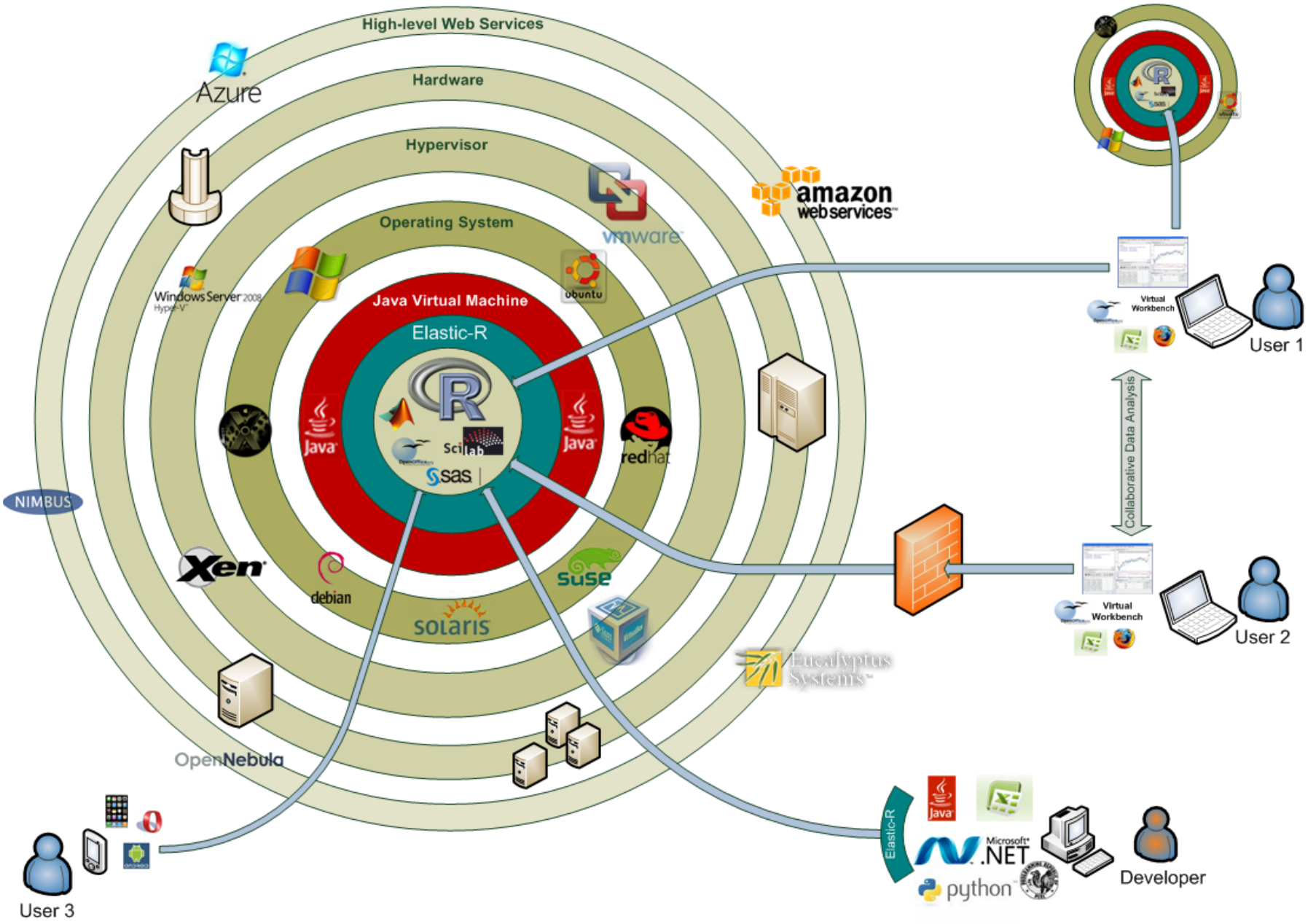
Generated Computational Web Services

Stateful or stateless, automatic mapping of R data objects and functions

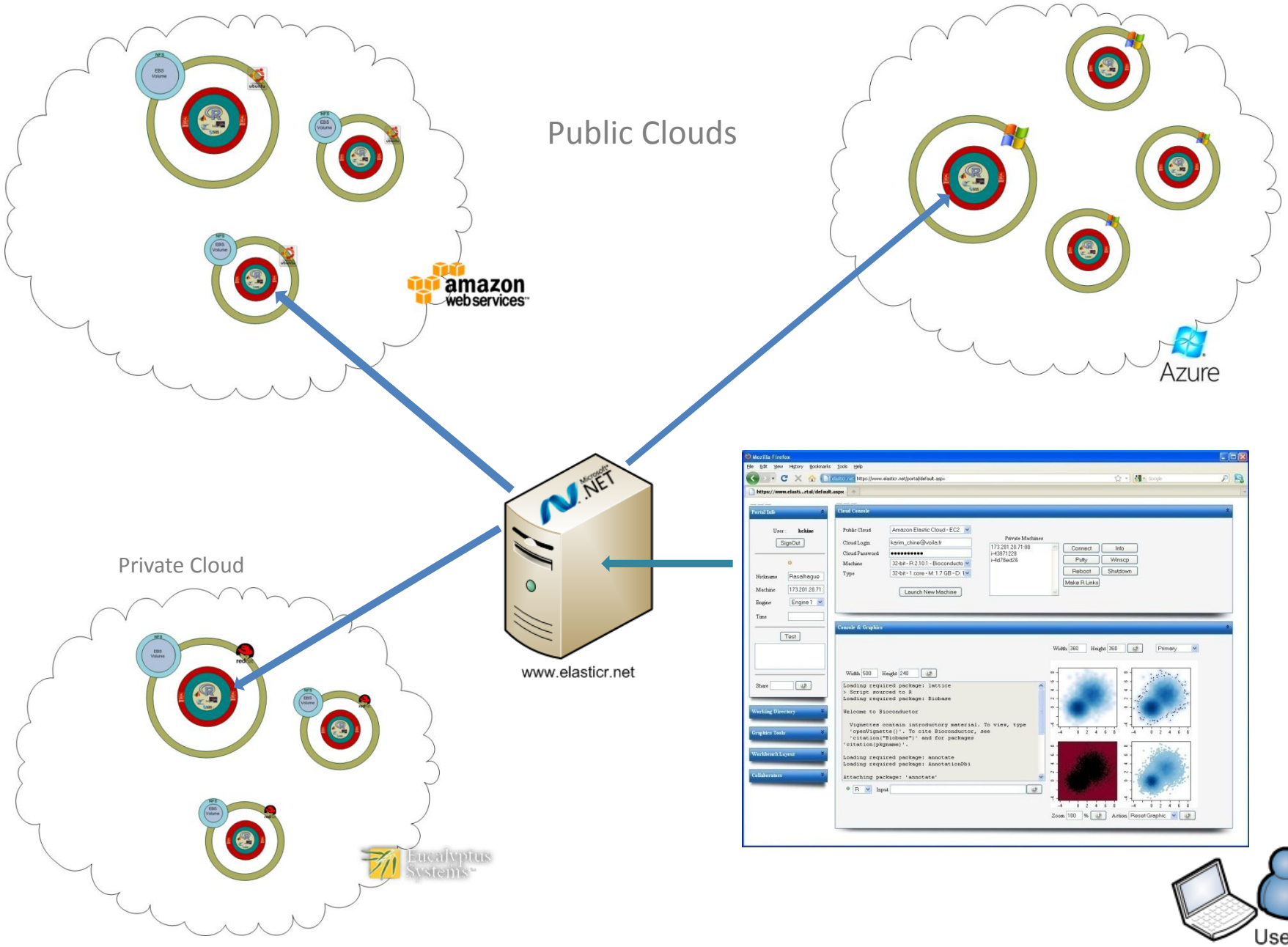
Computational Application Programming Interfaces

Java / SOAP / REST, Stateless and stateful

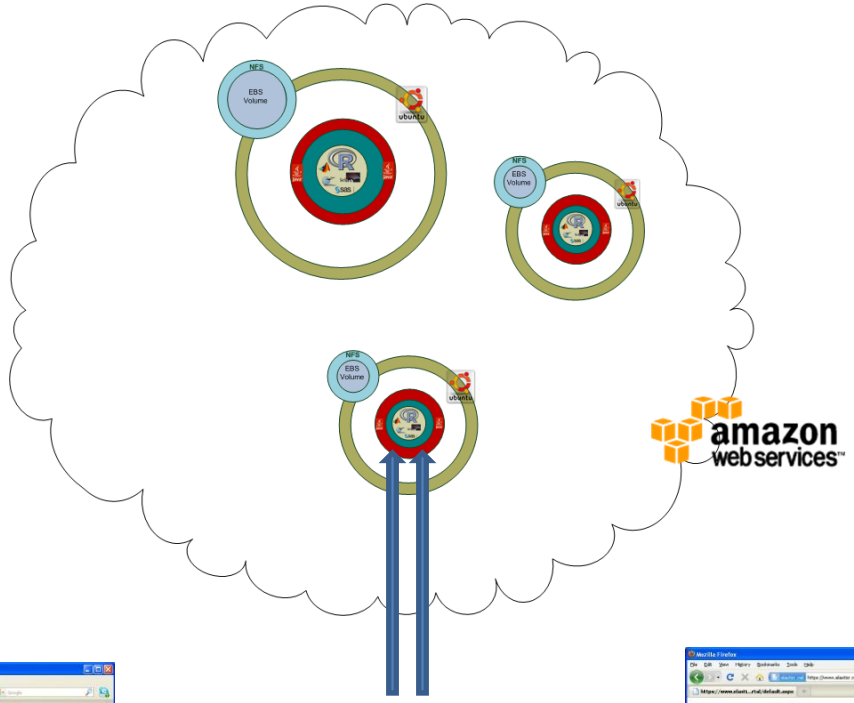
Elastic-R on IaaS-style clouds

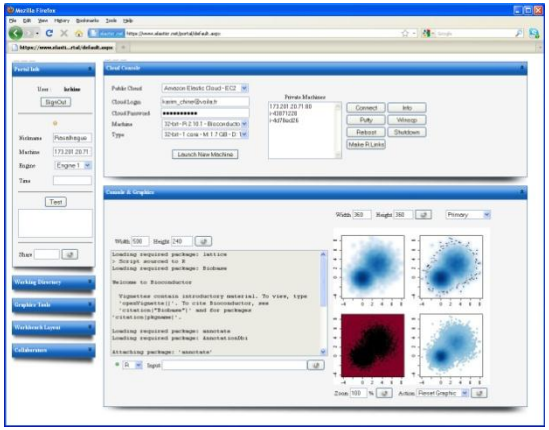


Elastic-R portal: single facade to public/private clouds

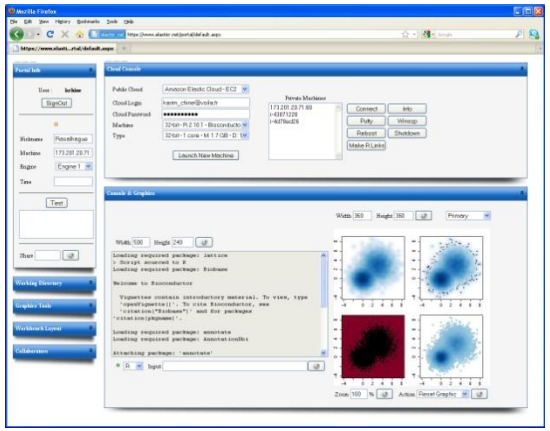


Elastic-R portal : collaborative Virtual Research Environment

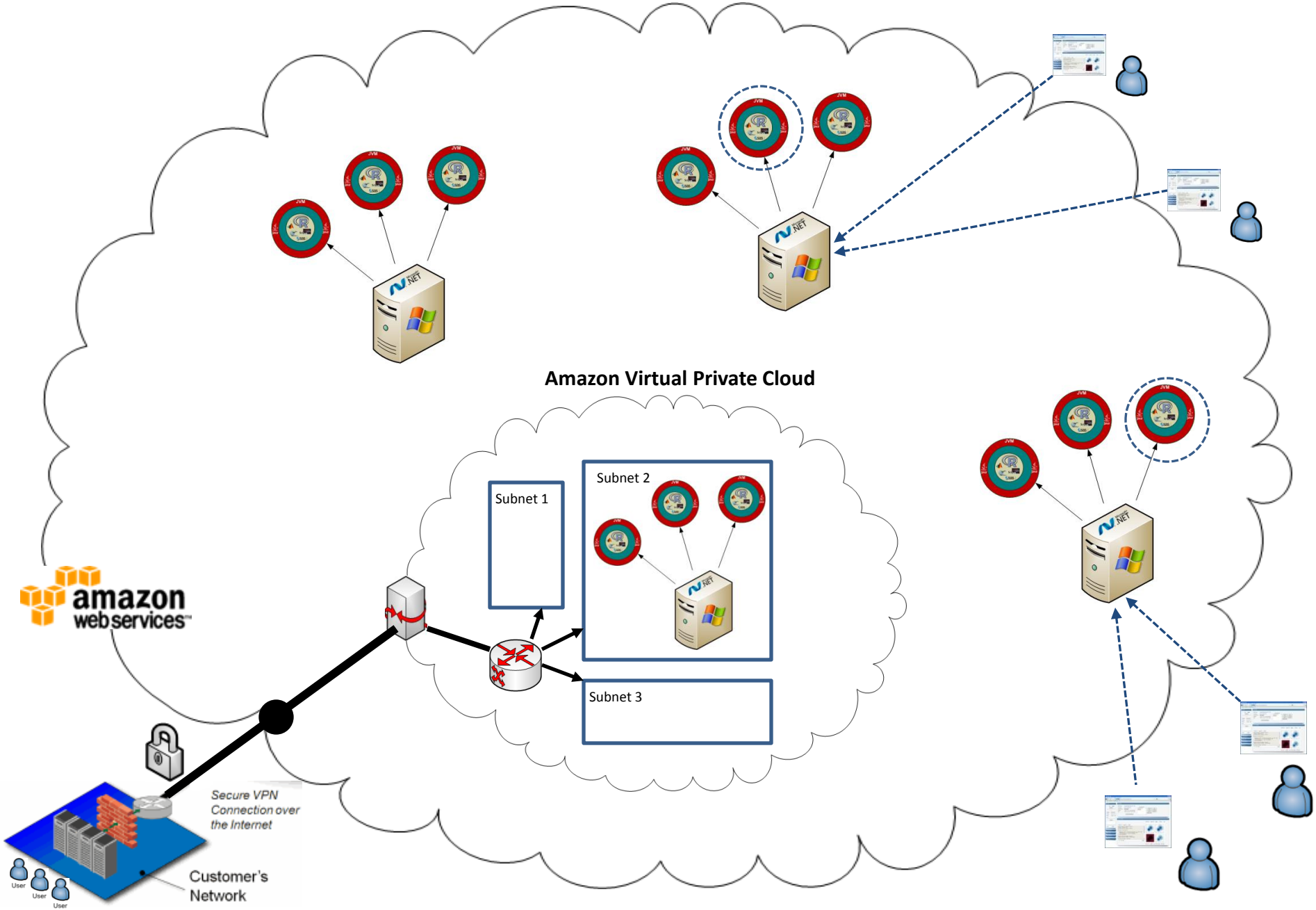




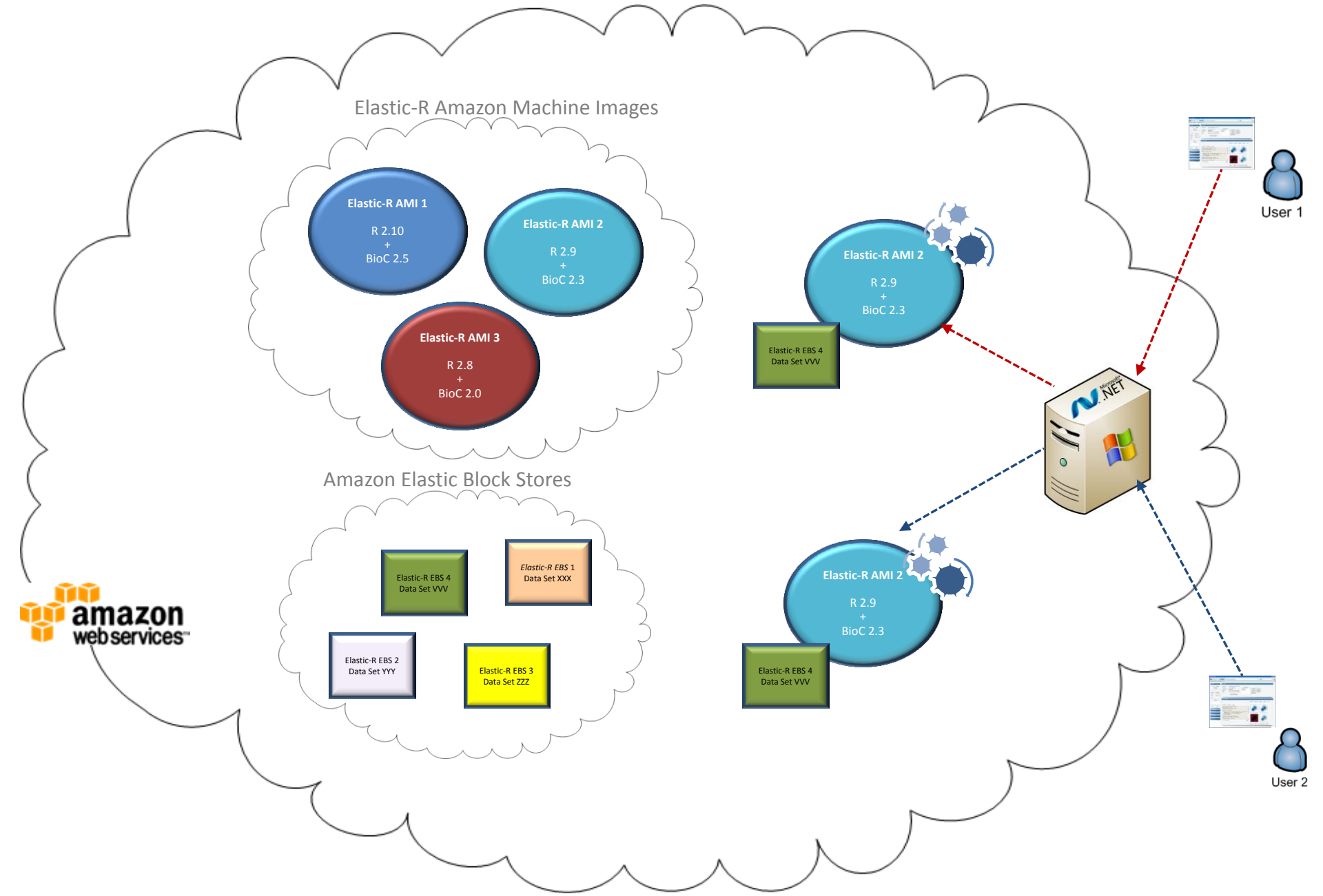
www.elasticr.net



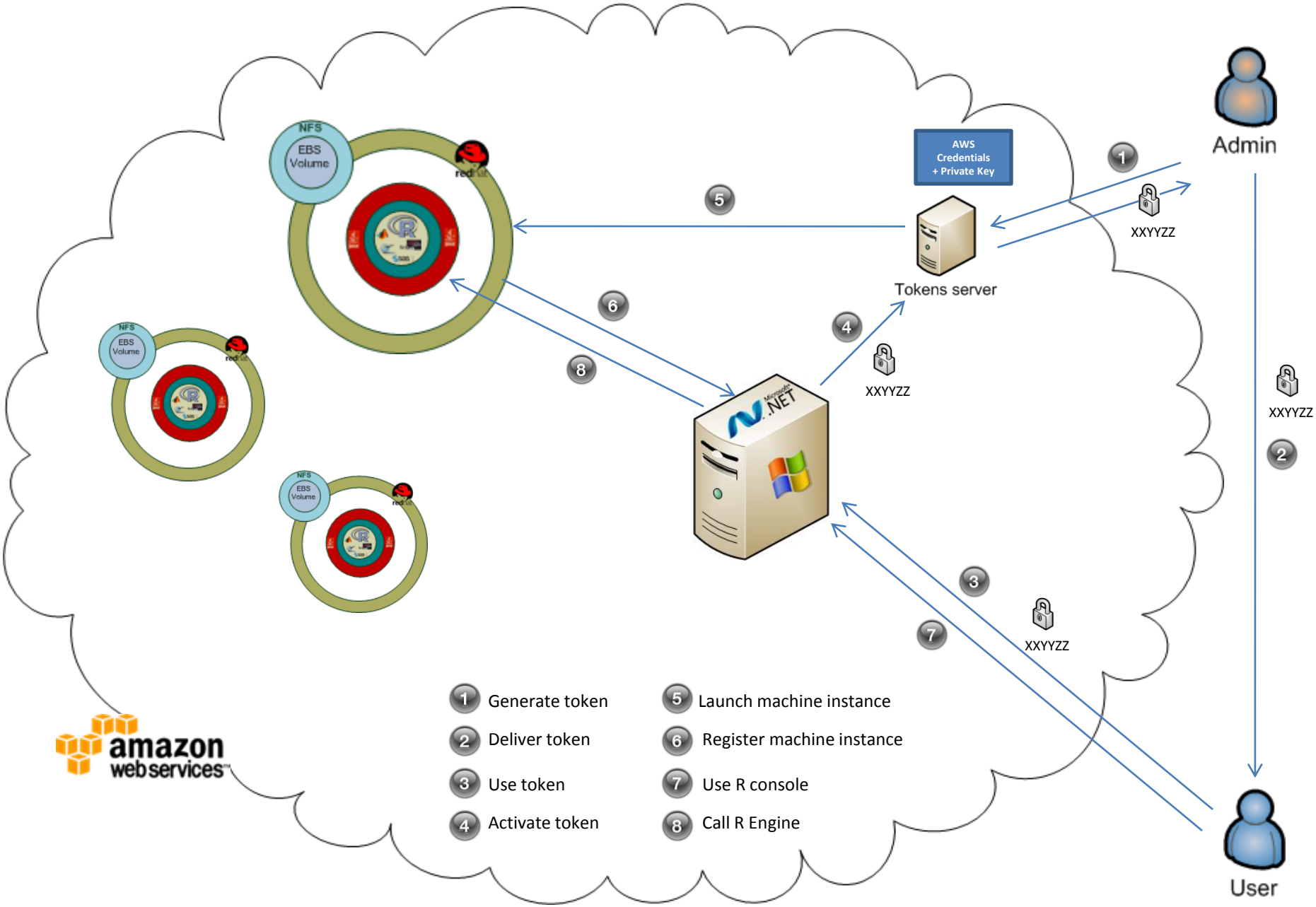
Decentralized collaboration : Elastic-R portal as an EC2 AMI



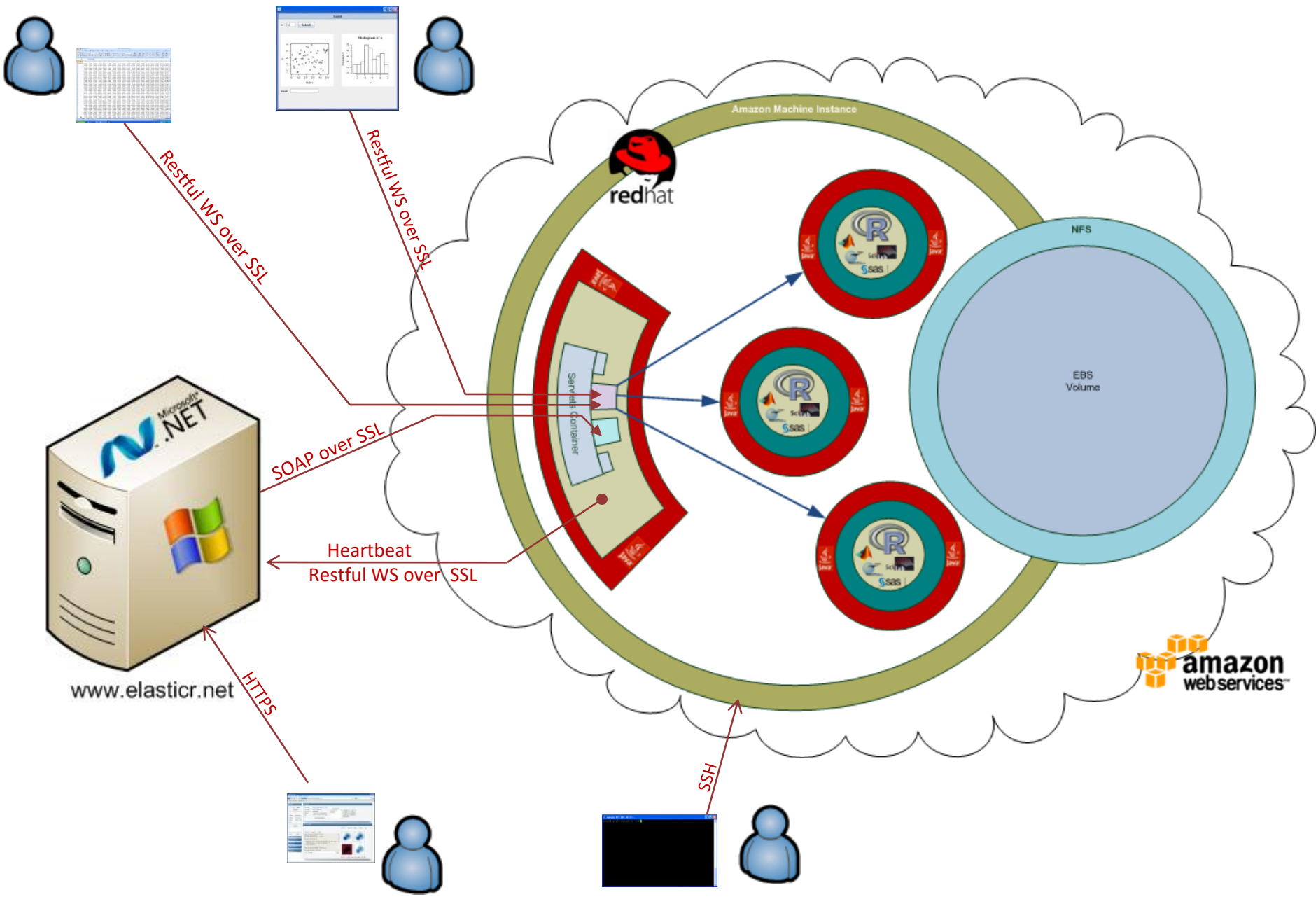
The IaaS-style cloud as a reproducible research platform



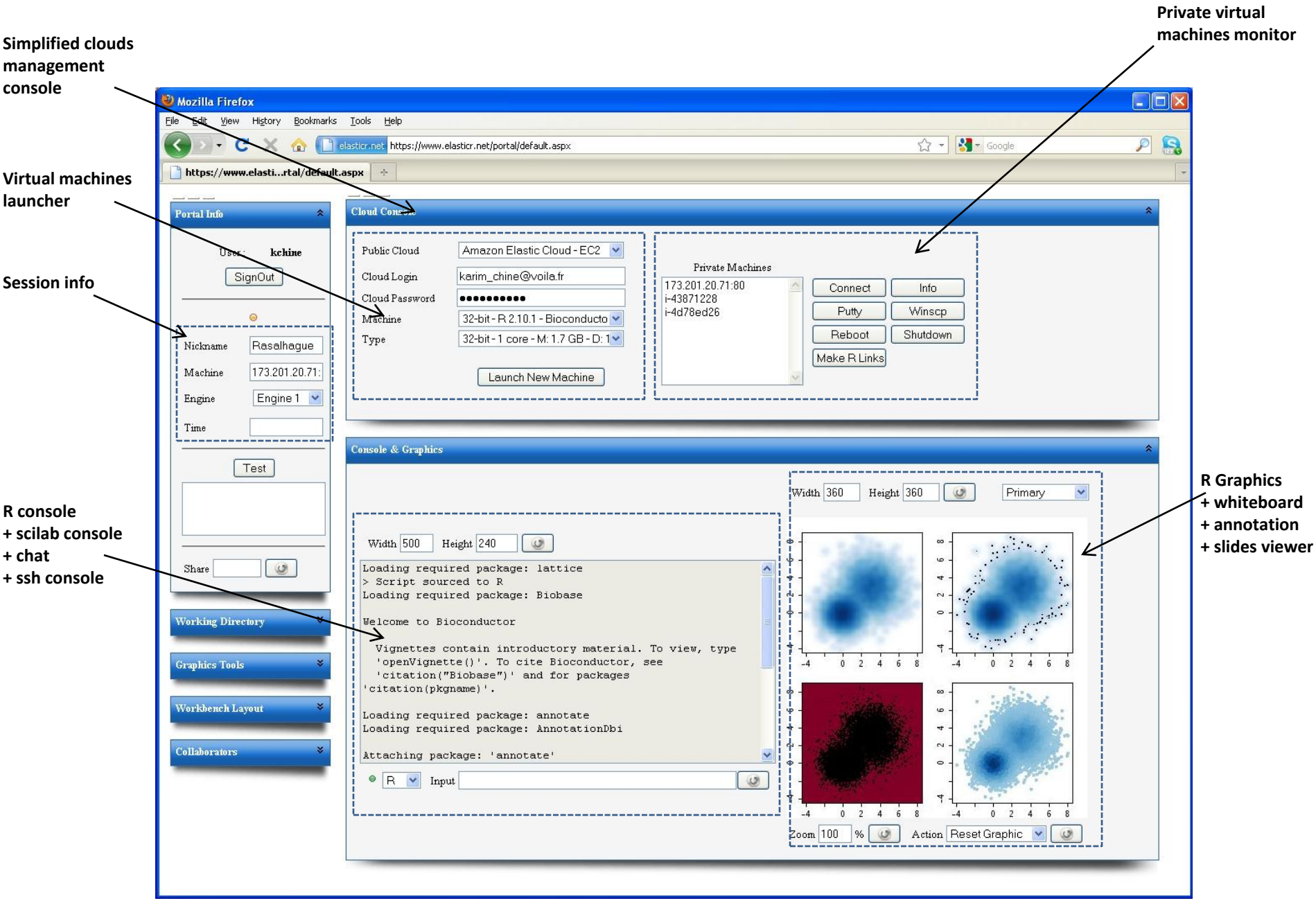
One Amazon account and many users : Elastic-R signed tokens



Elastic-R security



The Elastic-R AJAX Workbench I



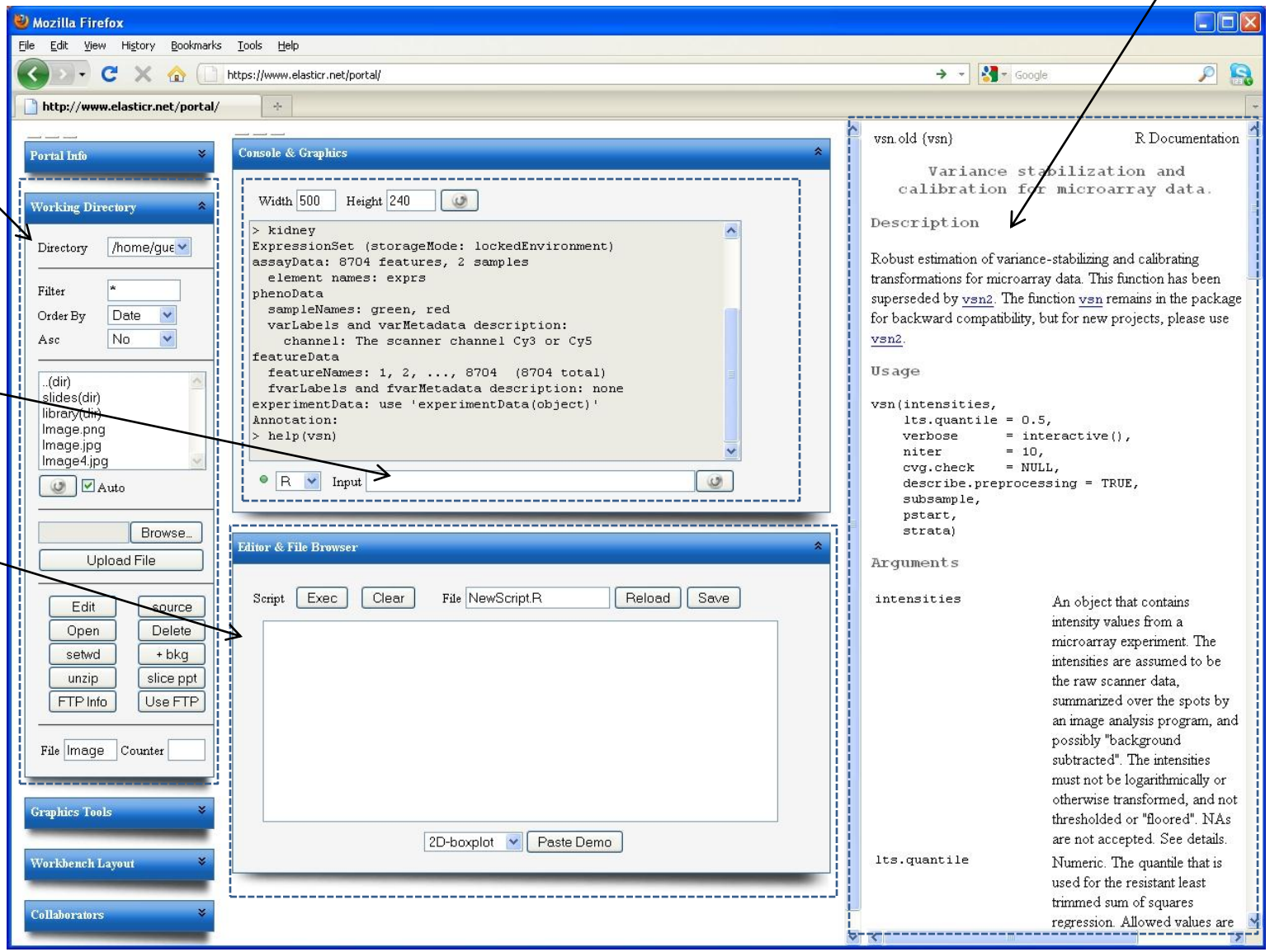
The Elastic-R AJAX Workbench II

Browsable contextual R help

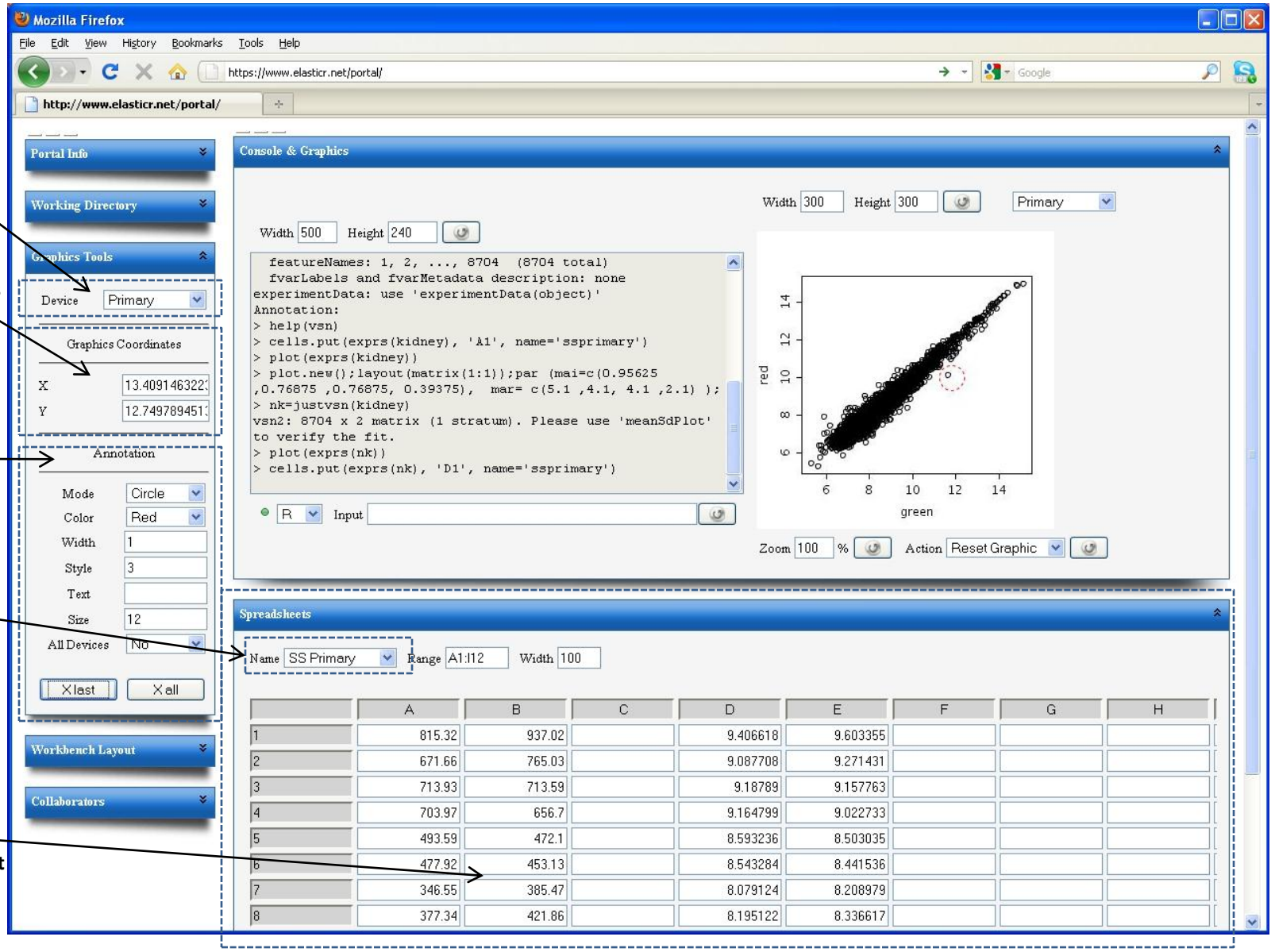
Working directory browser
Files upload/download
to/from cloud machine instance

Collaborative console

Collaborative script editor



The Elastic-R AJAX Workbench III



Graphic device selector

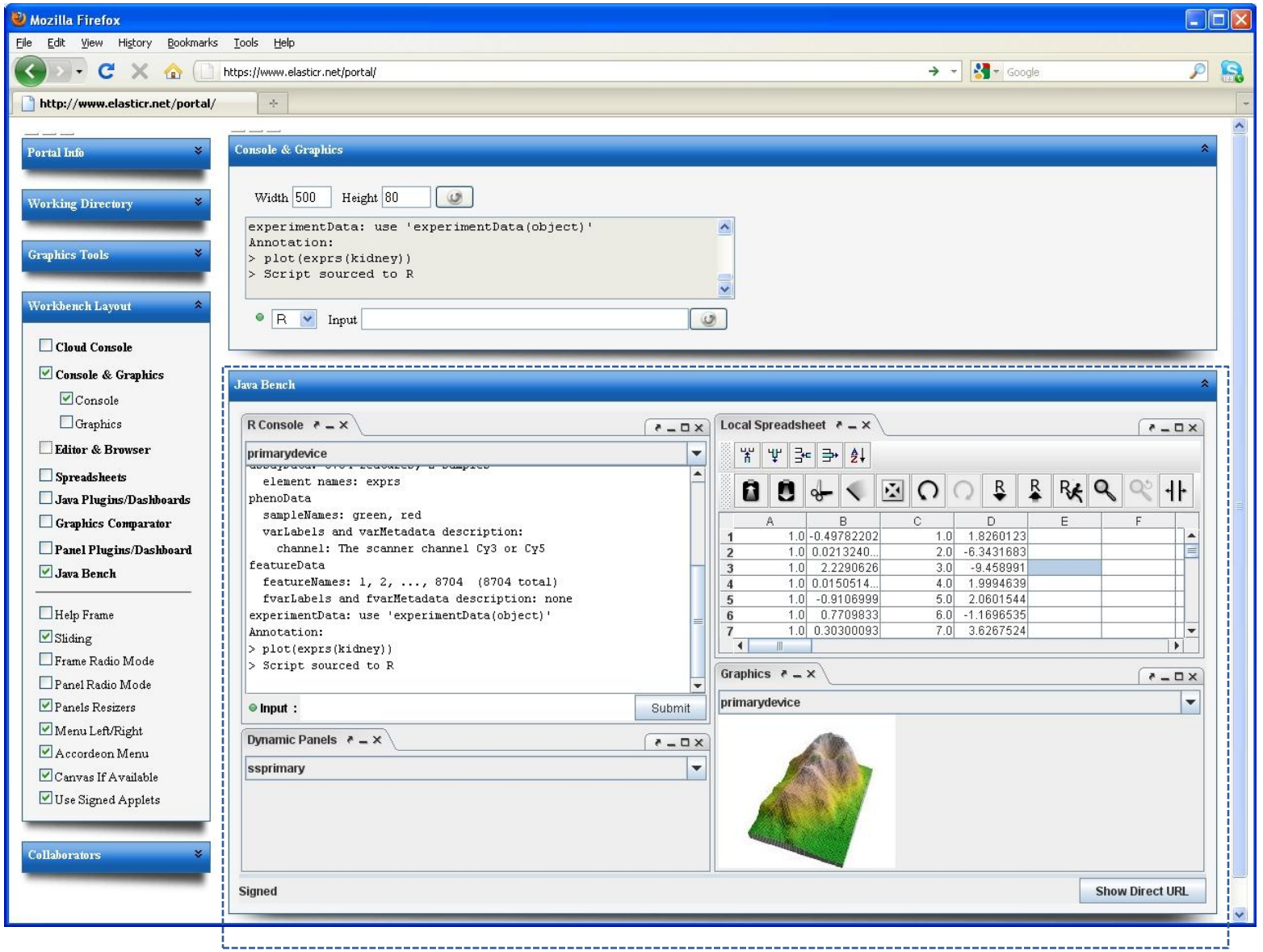
Graphics real coordinates

Graphic tools
Persistent collaborative annotators + virtual laser pointer + whiteboard + ..

Spreadsheet selector

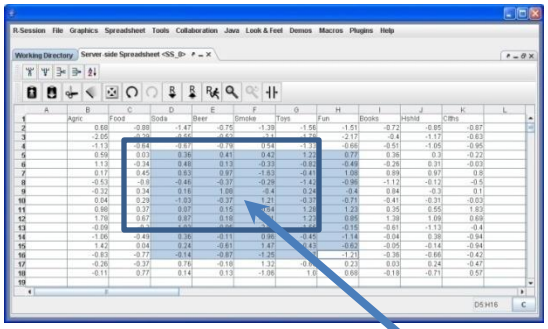
Server-side, R-enabled collaborative spreadsheet

The Elastic-R Java Workbench

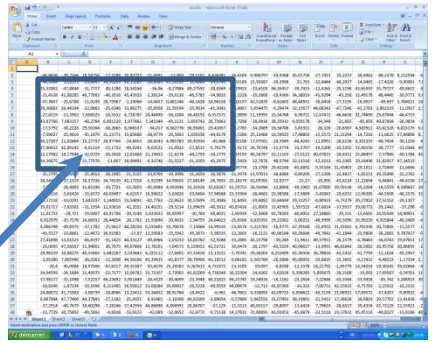


Software + services =applications convergence + collaboration

Elastic-R Java Workbench

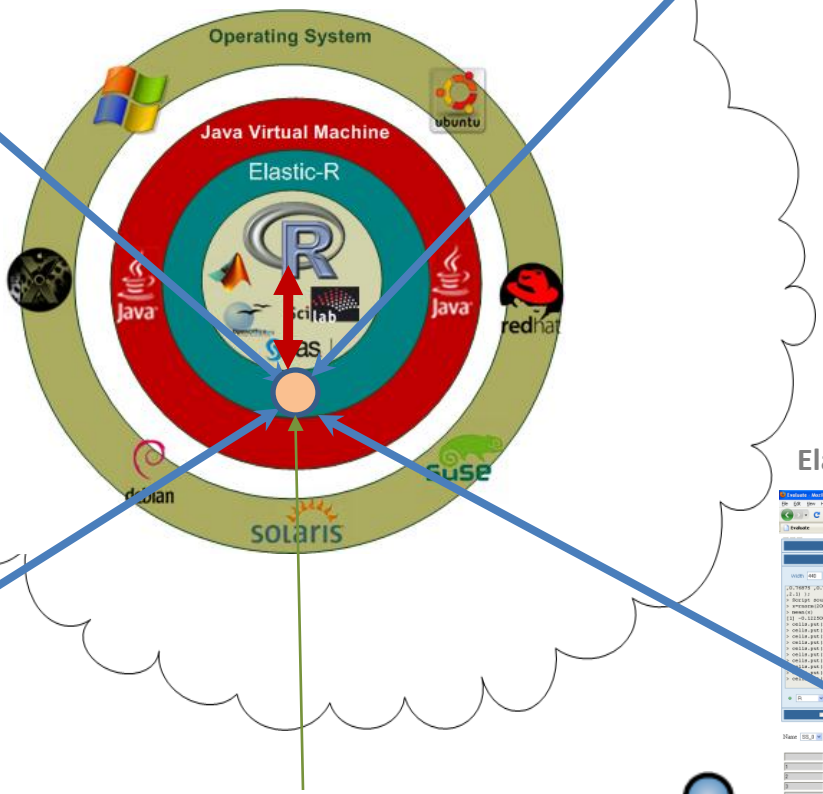


Microsoft Excel

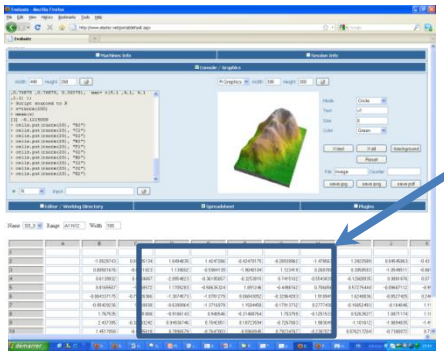


User 1

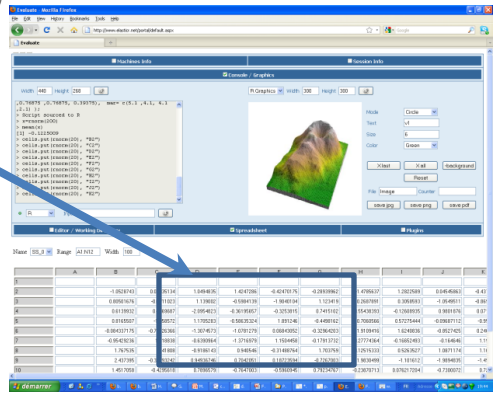
User 2



Elastic-R AJAX Workbench



Elastic-R AJAX Workbench

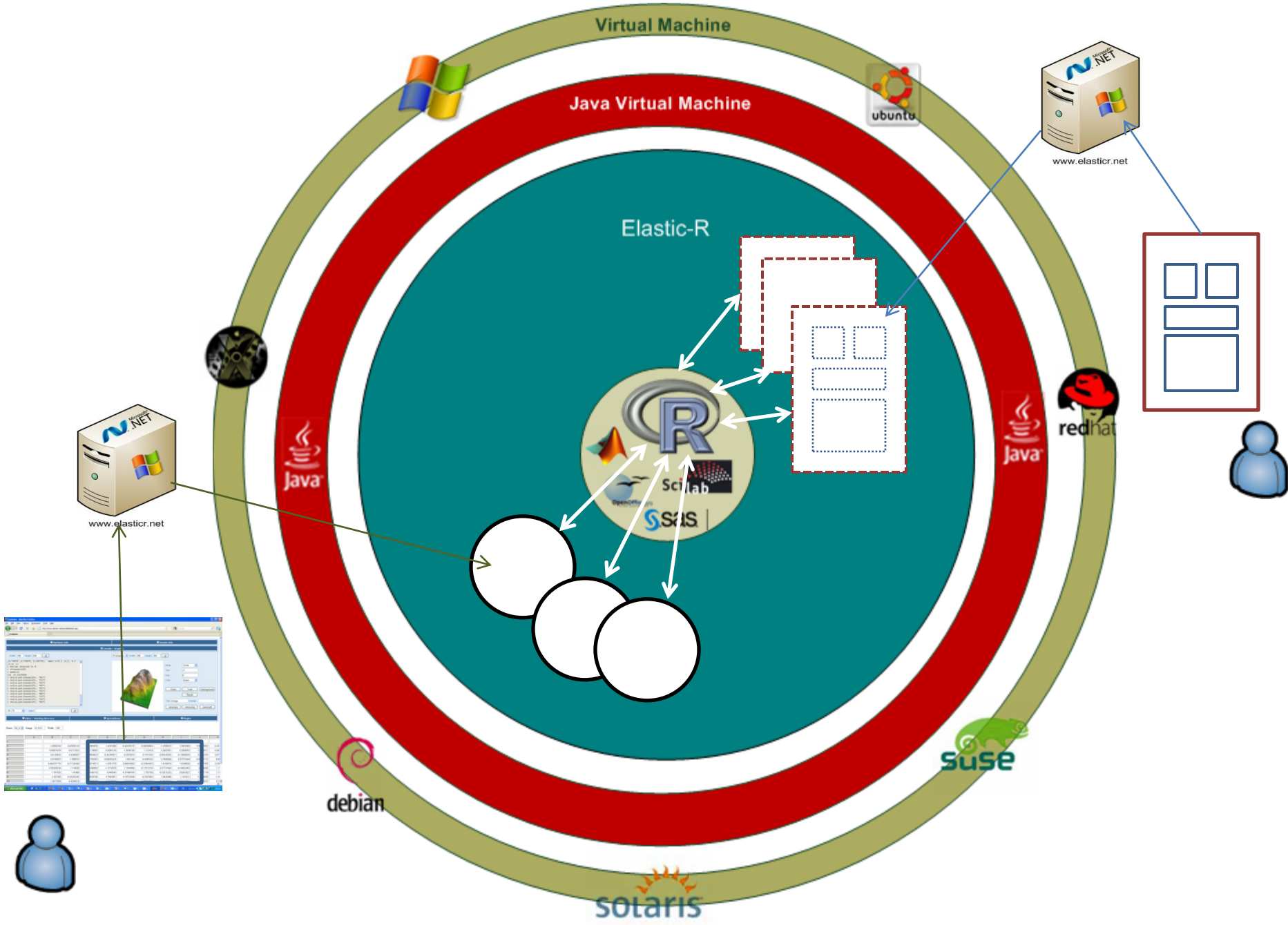


Elastic-R Spreadsheet model

User 3

User 4

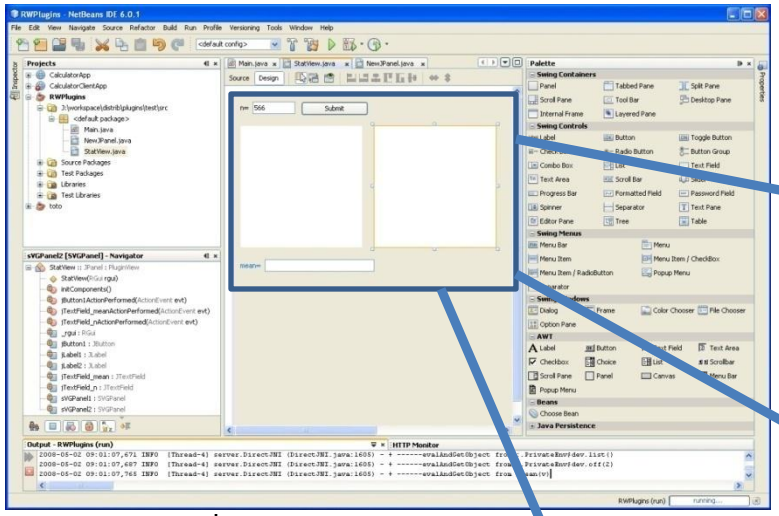
The Elastic-R server-side spreadsheet models / GUI widgets

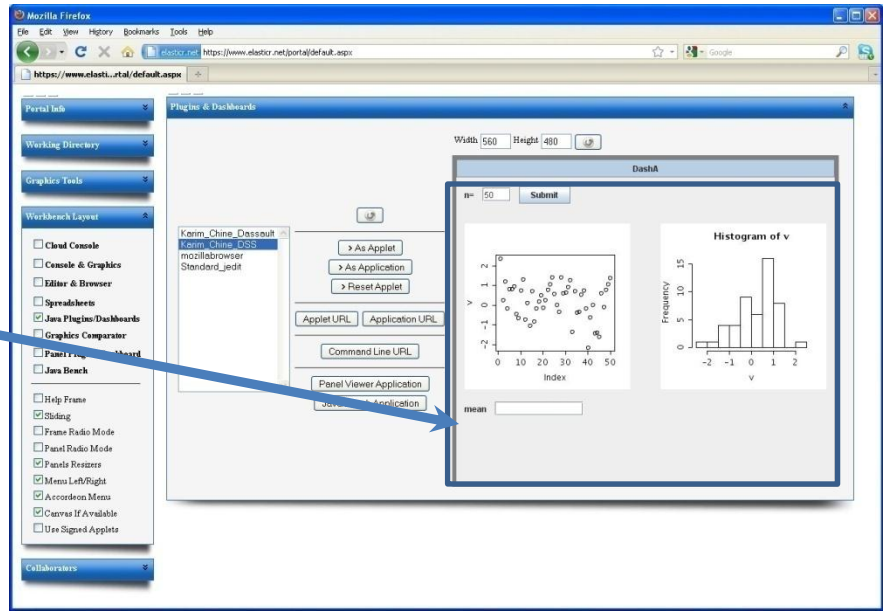


The cloud applications factory

Elastic-R AJAX Workbench

Visual Graphic User Interface Builder



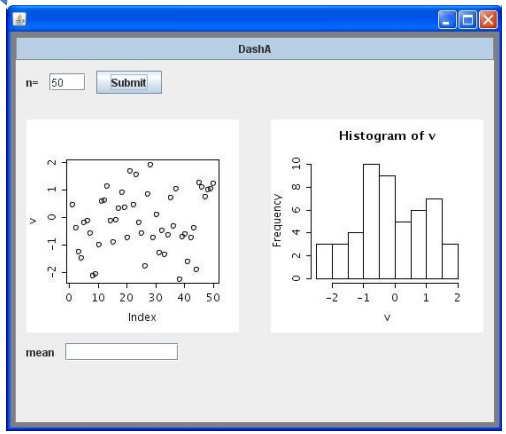


Upload plugin

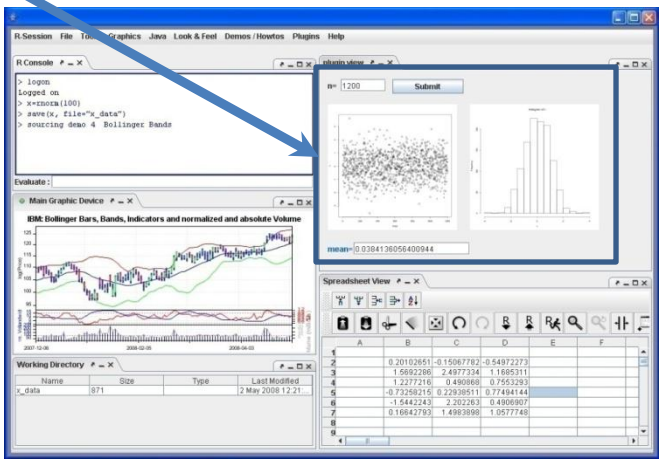
Plugins Repository

- myPlugin
- myDashboard

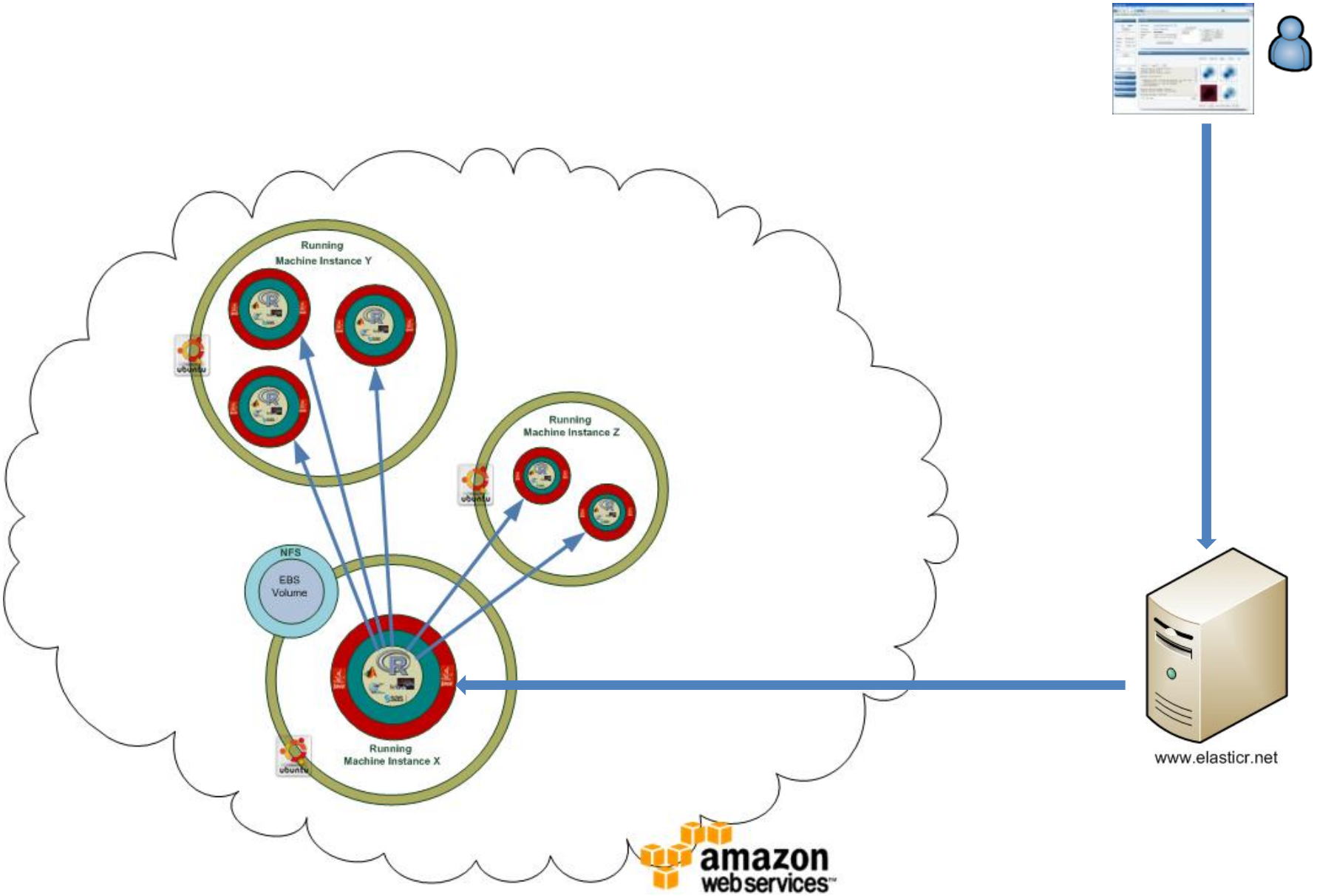
Standalone Application Accessible From a URL



Elastic-R Java Workbench



Elastic-R : user-friendly distributed computing platform



Demo

Useful links

Elastic-R Portal :

www.elasticr.net

Platform Web Site:

www.elasticr.net/platform

Articles :

Karim Chine, "Open Science in the Cloud: Towards a Universal Platform for Scientific and Statistical Computing", Chapter 19 in “Handbook of Cloud Computing”, Springer, 2010 (in press)

Karim Chine, "Scientific Computing Environments in the age of virtualization, toward a universal platform for the Cloud" pp. 44-48, 2009 IEEE International Workshop on Opensource Software for Scientific Computation (OSSC), 2009

Karim Chine, "Biocep, Towards a Federative, Collaborative, User-Centric, Grid-Enabled and Cloud-Ready Computational Open Platform,“ escience,pp.321-322, 2008 Fourth IEEE International Conference on eScience, 2008

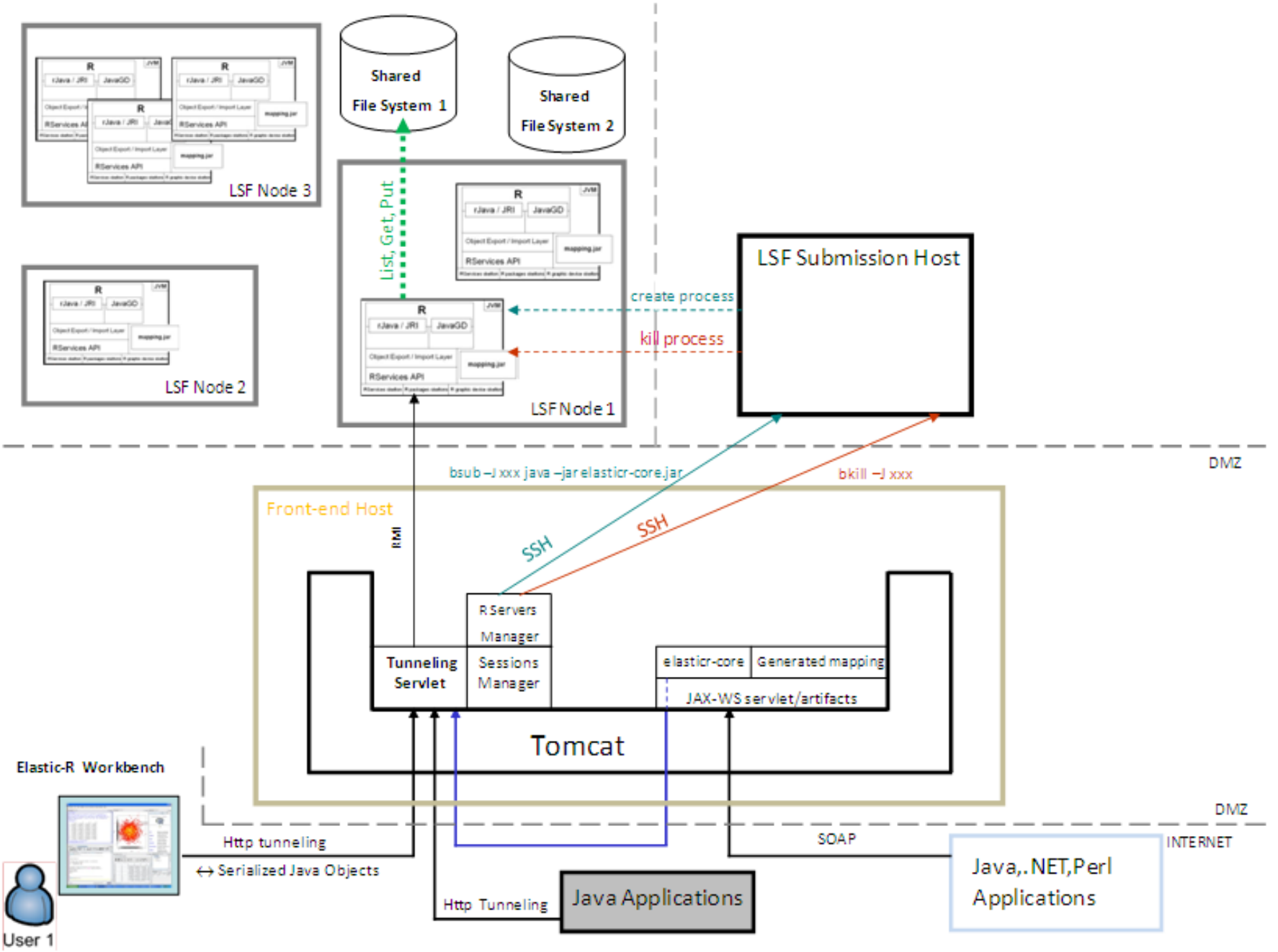
Linkedin Group:

http://www.linkedin.com/groups?home=&gid=2345405

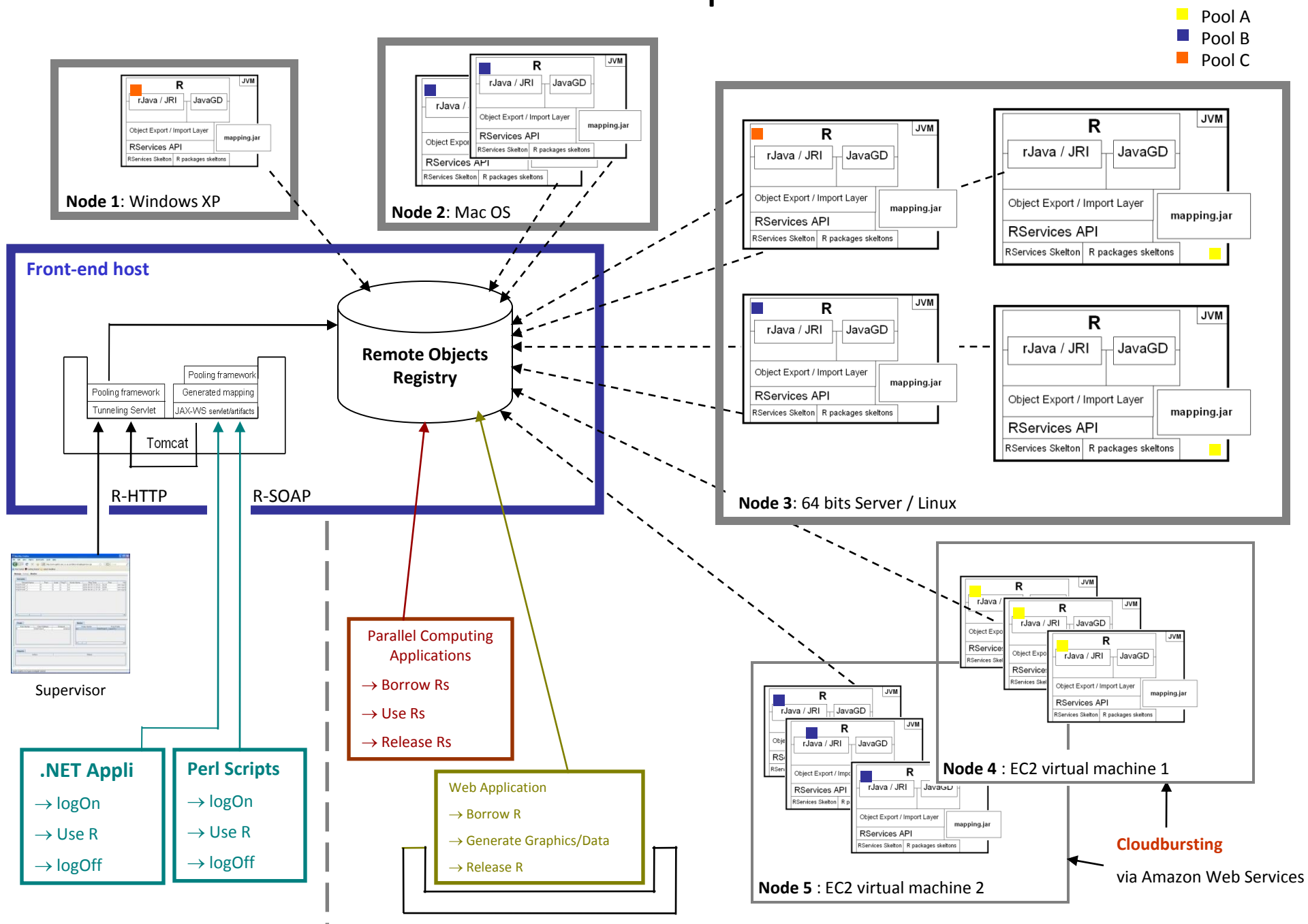
Acknowledgments

ACS: Madi Nassiri **Amazon**: Simone Brunozzi, Deepak Singh **AT&T Research Labs**: Simon Urbanek **ATUGE**: Imen Essafi, Béchir Tourki, Ilyes Gouja, HatemHachicha, Amine Elleuch **Auckland Centre for eResearch**: Nick Jones **Banca d'Italia**: Giuseppe Bruno **Bio-IT World**: Kevin Davies **BNP Paribas**: Ousseynou Nakoulima **Cambridge Healthtech Institute**: Cindy Crowninshield **City University of New York**: Mario Morales, Makram Talih **Columbia University**: Omar Besbes **Dassault Systèmes**: Omri Ben Ayoun, Patrick Johnson **Dataspora**: Michael E. Driscoll **EDF**: Alejandro Ribes **EBI**: Alvis Brazma, Wolfgang Huber, Kimmo Kallio, Misha Kapushesky, Michael Kleen, Alberto Labarga, Philippe Rocca-Serra, Ugis Sarkans, Kirsten Williams, Eamonn Maguire **EPFL**: Darlene Goldstein **ESPRIT**: Farouk Kammoun, Tahar. Benlakhdar **e-Taalim**: Nadhir Douma **ETH Zürich**: Yohan Chalabi, Diethelm Würtz, Martin Mächler **European Commission**: Konstantinos Glinos, Enric Mitjana, Monika Kacik, Ioannis Sagias **FHCRC**: Martin Morgan, Nianhua Li, Seth Falcon **Google**: Olivier Bosquet **FVG LLC**: Lisa Wood **Harvard University**: Tim Clark, Sudeshna Das, Douglas Burke,Paolo Ciccarese **IBM**: Jean-Louis Bernaudin, Pascal Sempe, Loic Simon, Lea A Deleris, Alex Fleischer, Alain Chabrier **Imperial College London**: Asif Akram, Vasa Curcin, John Darlington, Brian Fuchs **Indiana University**:Michael Grobe **INRIA**: David Monteau, Christian Saguez, Claude Gomez, Sylvestre Ledru **JISC**: John Wood, David Flanders **Johnson & Johnson - Janssen Pharmaceutica**: Patrick Marichal **KXEN**: Eric Marcade **Lancaster University**: Robert Crouchley, Daniel Grose **Leibniz Universität Hannover**: Kornelius Rohmeier **LIAMA:** Baogang Hue, Kang Cai **Limagrain**: Zivan Karaman **Mekentosj**: Alexander Griekspoor, Matt Wood **Microsoft**: Eric Le Marois, Tony Hey **Mubadala**: Ghazi Ben Amor **Nature Publishing Group**: Ian Mulvany, Steve Scott **NCeSS**: Peter Halfpenny, Rob Procter, Marzieh Asgari-Targhi, Alex Voss, YuWei Lin, Mercedes Argüello Casteleiro, Wei Jie, Meik Poschen, Katy Middlebrough, Pascal Ekin, June Finch, Farzana Latif, Elisa Pieri, Frank O'Donnell **New York Java User Group**: Frank D Greco **OeRC**: Dimitrina Spencer, Matteo Turilli, David Wallom, Steven Young **OMII-UK**: Neil Chue Hong, Steve Brewer **OpenAnalytics**: Tobias Verbeke **Oracle**: Dominique van Deth, Andrew Bond **OSS Watch**: Ross Gardler **Platform Computing**: Christopher Smith **Royal Society**: James Wilsdon **San Diego Supercomputer Center**: Nancy R. Wilkins-Diehr **Sanger Institute**: Lars Jorgensen, Phil Butcher **Shell**: Wayne.W.Jones, Nigel Smith **Société Générale**: Anis Maktouf **Stanford University**: John Chambers, Balasubramanian Narasimhan, Gunter Walther **SYSTEM@TIC**: Karim Azoum **Technische Universität Dortmund**: Uwe Ligges, Bernd Bischl **Technoforge**: Pierre-Antoine Durgeat **Tekiano**: Samy Ben Naceur **Télécom-ParisTech**: Isabelle Demeure, Georges Hebrail, Nesrine Gabsi **The Generations Network**: Jim Porzak **Total**: Yannick Perigois **Tunisian Ministry of Communication Technologies**: Naceur Ammar, Lamia Chaffai-Sghaier, Mohamed Saïd Ouerghi, Syrine Tlili **Tunisian Ecole Polytechnique**: Riadh Robbana **UC Berkeley**: Noureddine El Karoui, Terry Speed **UC Davis**: Rudy Beran, Debashis Paul, Duncan Temple Lang **UCL**: Daniel Jeffares **UCLA**: Ivo Dinov, Jeroen Ooms **UC San Diego**: Anthony Gamst **UCSF**: Tena Sakai **Université Catholique de Louvain**: Christian Ritter **University of Cambridge**: Ian Roberts, Robert MacInnis Peter Murray-Rust, Jim Downing **University of Manchester**: Carole Goble, Len Gill, Simon Peters, Richard D Pearson, Iain Buchan, John Ainsworth **University of Plymouth**: Paul Hewson **University of Split**: Ivica Puljak **UTK**: Ajay Ohri **World Bank Group-IFC**: Oualid Ammar **Yahoo**: Laurent Mirguet, Rob Weltman **Independant**:Charles Dallas, Romain François

Elastic-R for clusters/grids



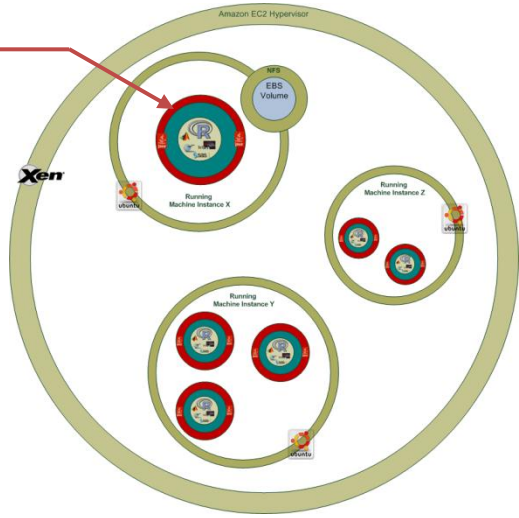
Elastic-R SOA platform



Stateful generated Web Services : Elastic-R for workflow workbenches

Login
Pwd
Options
LogOn

SessionID associated with a reserved Elastic-R Engine



ES
T1
ESon1
T2
ESon2
T3
ESon3
getData
f (ES)

Retrieve Data
logOff
- remove ESonx
- « Clean » Elastic-R Engine
- Put Elastic-R Engine back in the Pool
- kill Elastic-R Engine

T1,**T2**,**T3** : Generated Stateful Web Services for R functions T1,T2 & T3
LogOn, **getData** : R-SOAP methods

ES : ExpressionSet
ESon1, **ES**on2, **ES**on3 : ExpressionSet Object Names

f = T3 o T2 o T1