

Development of Integrative Bioinformatics Applications using Cloud Computing resources and Knowledge Organization Systems (KOS).

Jonas S Almeida

Department of Bioinformatics and Computational Biology, The University of Texas M D Anderson Cancer Center, 1515 Holcombe Blvd Houston, TX 77030, USA.

Helena F Deus

Institute of Chemical and Biological Technology, Universidade Nova de Lisboa, Oeiras, Portugal.

Wolfgang Maass

Research Center for Intelligent Media, Furtwangen University, Furtwangen, Germany

Almeida JS, Deus HF, Maass W. (2010) S3DB core: a framework for RDF generation and management in bioinformatics infrastructures. BMC Bioinformatics. 2010 Jul 20;11(1):387. [[PMID 20646315](#)].



Semantic Web Applications and Tools for Life Sciences
December 10th, 2010, Berlin, Germany

• Wang X, R Gorlitsky, and JS Almeida (2005) *From XML to RDF: How Semantic Web Technologies Will Change the Design of 'Omic' Standards*. *Nature Biotechnology*, Sep;23(9):1099-103 [[PMID:16151403](#)].

• Almeida JS, C Chen, R Gorlitsky, R Stanislaus, M Aires-de-Sousa, P Eleutério, JA Carriço, A Marezek, A Bohn, A Chang, F Zhang, R Mitra, GB Mills, X Wang, HF Deus (2006) *Data integration gets 'Sloppy'*. *Nature Biotechnology* 24(9):1070-1071. [[PMID:16964209](#)].

• Deus FH, R Stanislaus¹, DF Veiga, C Behrens, II Wistuba, JD Minna, HR Garner, SG Swisher, JA Roth, AM Correa, B Broom, K Coombes, A Chang, LH Vogel, JS Almeida (2008) *A Semantic Web management model for integrative biomedical informatics*. *PLoS ONE*. Aug 13;3(8):e2946 [[PMID: 18698353](#)].

• Almeida JS, Deus HF, Maass W. (2010) *S3DB core: a framework for RDF generation and management in bioinformatics infrastructures*. *BMC Bioinformatics*. 2010 Jul 20;11(1):387. [[PMID 20646315](#)].

• Deus HF, DF Veiga, PR Freire, JN Weinstein, GB Mills, JS Almeida (2010) *Exposing The Cancer Genome Atlas as a SPARQL endpoint*. *Journal Biomedical Informatics* [[PMID 20851208](#)].

• Correa MC, HF Deus, AT Vasconcelos, Y Hayashi, JA Ajani, SV Patnana, JS

Almeida JS, R Stanislau, E Krug, J Arthur (2005) Normalization and Analysis of residual variation in 2D Gel Electrophoresis. *Proteomics* 5(5):1242-9 [[PMID:15732138](#)].

Mitas M, JS Almeida, K Mikhitarian, WE Gillanders, DN Lewin, DD Spyropoulos, L Hoover, A Graham, T Glenn, P King, DJ Cole, R Hawes, CE Reed, BJ Hoffman (2005) Accurate discrimination of Barrett's esophagus and esophageal adenocarcinoma using a quantitative algorithm and multi-marker real-time RT-PCR. *Clin Cancer Res.* 2005 Mar 15;11(6):2205-14 [[PMID:15788668](#)].

Nunes S, R Sá-Leão, J Carriço, CR Alves, R Mato, A Brito Avó, J Saldanha, JS Almeida, I Santos Sanches, and H de Lencastre (2005) Trends in drug resistance, serotypes and molecular types of *Streptococcus pneumoniae* colonizing pre-school age children attending in Lisbon, Portugal – a summary of four years of annual surveillance. *J Clin Microbiol.* 2005 Mar;43(3):1285-93 [[PMID:15750097](#)].

Stanislau R, C Chen, J Franklin, J Arthur, JS Almeida (2005) AGML Central: AGML Compatible Proteomic Database. *Bioinformatics*, 21(9):1754-7 [[PMID:15647304](#)].

Garcia, S., J.S. Almeida (2005) Nearest neighbor embedding with different time delays. *Physical Review E* 71, 037204 [[PMID: 15903641](#)]; also selected for reprinting in *Vol 9, Issue 7 of Biological Physics Research*.

Mikhitarian, K., Gillanders, W.E., Almeida, J.S., Hebert Martin R., Varela J.C., Metcalf, J.S., Cole, D.J., and Mitas, M. (2005) An innovative microarray strategy identifies informative molecular markers for the detection of micrometastatic breast cancer. *Clinical Cancer Research* 11(10):3697-704. [[PMID:15897566](#)]

Almeida JS, Nowotny H. (2005) The emergence of the ERC. *Science.* 307(5713):1200 [[PMID:15731424](#)].

McKillen DJ, YA Chen, C Chen, MJ Jenny, HF Trent, J Robalino, DC McLean, PS Gross, RW Chapman, GW Warr, JS Almeida (2005) Marine Genomics: A clearing-house for genomic and transcriptomic data of marine organisms. *BMC Genomics* 2005, 6:34 [[doi:10.1186/1471-2164-6-34](#)].

Frazao N, Brito-Avo A, Simas C, Saldanha J, Mato R, Nunes S, Sousa NG, Carrico, JA, Almeida JS, Santos-Sanches I, de Lencastre H. (2005) Effect of the Seven-Valent Conjugate Pneumococcal Vaccine on Carriage and Drug Resistance of *Streptococcus pneumoniae* in Children Attending Day-Care Centers in Lisbon. *Pediatr Infect Dis J.* 2005 Mar;24(3):243-252. [[PMID:15750461](#)].

Almeida JS, DJ McKillen, YA Chen, PS Gross, RW Chapman, G Warr (2005) Design and Calibration of Microarrays as Universal Transcriptomic Environmental Biosensors. *Comparative and Functional Genomics*, 6(3):132-137(6). [[doi:10.1002/cfg.466](#)].

Nunes S, Sa-Leao R, Carrico J, Alves CR, Mato R, Avo AB, Saldanha J, Almeida JS, Sanches IS, de Lencastre H. (2005) Trends in Drug Resistance, Serotypes, and Molecular Types of *Streptococcus pneumoniae* Colonizing Preschool-Age Children Attending Day Care Centers in Lisbon, Portugal: a Summary of 4 Years of Annual Surveillance. *J Clin Microbiol.* 2005 Mar;43(3):1285-93 [[PMID:15750097](#)].

Wolf G, JS Almeida, MAM Reis and JG Crespo (2005). Modelling of the extractive membrane bioreactor process based on natural fluorescence fingerprints and process operation history. *Water Science and Technology*, 51 (6-7): 51-58. [[PMID:16003961](#)]

Wolf G, JS Almeida, MAM Reis and JG Crespo (2005) Non-mechanistic modelling of complex biotfilm reactors and the role of process operation history. *Journal of Biotechnology*, 117 (4): 367-383. [[PMID:15925719](#)].

Wang X, R Gorlitsky, and JS Almeida (2005) From XML to RDF: How Semantic Web Technologies Will Change the Design of 'Omic' Standards. *Nature Biotechnology*, Sep;23(9):1099-103 [[PMID:16151403](#)].

Garcia S.P., Jonas S. Almeida, JS (2005) Multivariate phase space reconstruction by nearest neighbor embedding with different time delays. *Physical Review E* 72, 027205. [[PMID:16196759](#)].

Ces JC, Varghese S, Bland AM, Taylor TP, Self SE, Stanislau R, Almeida JS, Arthur JM (2005) Prediction of urinary protein markers in lupus nephritis. *Kidney Int. Dec*;68(6):2588-92 [[PMID:16316334](#)].

Carrico JA, Pinto FR, Simas C, Nunes S, Sousa NG, Frazao N, de Lencastre H, Almeida JS (2005) Assessment of band-based similarity coefficients for automatic type and subtype classification of microbial isolates analyzed by pulsed-field gel electrophoresis. *J Clin Microbiol.* 2005 Nov;43(11):5483-90. [[PMID:16272474](#)].

Mato R, Sanches IS, Simas C, Nunes S, Carrico JA, Sousa NG, Frazao N, Saldanha J, Brito-Avo A, Almeida JS, Lencastre HD. (2005) Natural History of Drug-Resistant Clones of *Streptococcus pneumoniae* Colonizing Healthy Children in Portugal. *Microb Drug Resistance* 11(4):309-22. [[PMID:16359190](#)].

Muller LN, de Brouwer JF, Almeida JS, Stalder B, Voigt E, B. W. Analysis of a marine phototrophic biofilm by confocal laser scanning microscopy using the new image quantification software PHLIP. *BMC Ecology and Evolution* 6(1):1 [[PMID:16412253](#)].

Chen YA, Chou CC, Lu X, Slate EH, Peck KA, Hill EG, Smolka AJ, Morris JS, Coombes KR, Baggerly KA, Almeida JS (2007) PrepMS: TOF MS data graphical preprocessing tool. *Bioinformatics*. 15;23(2):264-5 [[PMID:17121773](#)].

Muller M, Wagner CL, Annibale DJ, Knapik J, Hill EG, Smolka AJ, Morris JS, Coombes KR, Baggerly KA, Almeida JS (2007) Parameter selection for semi-supervised classification: a decision tree-based approach to the problem of feature selection. *BMC Medical Informatics and Decision Making* 6:1 [[PMID:17148689](#)].

Kropievitch YV, Almeida JS (2006) mGrid: A parallel Matlab library for user-scale distribution. *BMC Bioinformatics* 7:149 [[PMID:17148689](#)].

Bland, A.M., L.R. D'Eugenio, M.A. Dugan, M.G. Janech, J.S. Almeida, M. Zilean J.M. Arthur. Comparison of Variability Associated with Sample Preparation in Two-Dimensional Gel Electrophoresis of Cardiac Tissue. *J. Biomol. Tech.* In Press: 2006. [[PMID:16870710](#)].

Cupit P, P Rolghamre, JS Almeida, K Ekdahl (2006) Modeling Pneumococcal Resistance to Penicillin in Southern Sweden Using Artificial Neural Networks. *Microbial Drug Resistance* 12(3):149-157. [[PMID:17002540](#)]

Almeida JS, Oates JC, Arthur JM. (2006) The need for concurrent calibration and discrimination statistics in predictive models. *Kidney Int.* 70(1):231-2. [[doi:10.1038/sj.ki.5001519](#)].

Almeida EO, Almeida JS, Marino S, Lall R, Goel G, Neves AR, Santos H (2006) Regulation of glycolysis in *Lactococcus lactis*: an unfinished systems biological case study. *Syst Biol (Stevenage)* Jul;153(4):286-98 [[PMID:16986630](#)].

Carrico JA, Silva-Costa C, Melo-Cristino J, Pinto FR, de Lencastre H, Almeida JS, Ramirez M. (2006) Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *J Clin Microbiol.* 44(7):2524-32 [[PMID:16986630](#)].

Almeida JS, C Chen, R Gorlitsky, R Stanislau, M Aires-de-Sousa, P Eleutério, JA Carriço, A Mareczek, A Bohn, A Chang, F Zhang, R Mitra, GB Mills, X Wang, HF Deus (2006) Data integration gets 'Sloppy'. *Nature Biotechnology* 24(9):1070-1071. [[PMID:16964209](#)].

Almeida, J.S., S. Vinga (2006) Computing distribution of scale independent motifs in biological sequences. *Algorithms for Molecular Biology*. 1:18. [[PMID:17049089](#)].

Mancia A, Lundqvist ML, Romano TA, Peden-Adams MM, Fair PA, Kindy MS, Ellis BC, Gattoni-Celli S, McKillen DJ, Trent HF, Ann Chen Y, Almeida JS, Gross PS, Chapman RW, Warr GW. (2007) A dolphin peripheral blood leukocyte cDNA microarray for studies of in vivo stress reactions. *Dev Comp Immunol.* 31(5):520-9 [[PMID:17084893](#)].

Kropievitch YV, Hill EG, Smolka AJ, Morris JS, Coombes KR, Baggerly KA, Almeida JS. (2007) PrepMS: TOF MS data graphical preprocessing tool. *Bioinformatics*. 15;23(2):264-5 [[PMID:17121773](#)].

Frazaio J, Almeida JS, McKillen D, Colglazier J, Trent Iii HF, Chen YA, Peck ME, Browdy CL, Chapman RW, Warr GW, Gross PS (2007) Physiol Genomics. 14;29(1):44-56 [[PMID:17148689](#)].

Wolf G, JS Almeida, JG Crespo, MA Reis (2007) An improved method for two-dimensional fluorescence monitoring of complex bioreactors. *J Biotechnol.* 128(4):801-12. [[PMID:17291616](#)].

Pinto FR, Carrico JA, Ramirez M, Almeida JS. (2007) Ranked Adjusted Rand: integrating distance and partition information in a measure of clustering agreement. *BMC Bioinformatics* 8(1):44. [[PMID:17286861](#)].

Varghese SA, Powell TB, Budisavljevic MN, Oates JC, Raymond JR, Almeida JS, Arthur JM (2007) Urine Biomarkers Predict the Cause of Glomerular Disease. *J Am Soc Nephrol. J Am Soc Nephrol.* 18(3):913-22 [[PMID: 17301191](#)]

Bohn A, Zippel B, Almeida JS, Xavier JB (2007) Stochastic modeling for characterisation of biofilm development with discrete detachment events (sloughing). *Water Sci Technol.* 2007;55(8-9):257-64. [[PMID: 17546994](#)]

Garcia SP, DeLancey LB, Almeida JS, Chapman RW (2007) Ecoforecasting in real time for commercial Wsheries: the Atlantic white shrimp as a case study. *Mar Biol* (2007) 152:15–24. [[DOI 10.1007/s00227-007-0622-3](#)]

Jenny MJ, Chapman RW, Mancia A, Chen YA, McKillen DJ, Trent H, Lang P, Escoubas JM, Bachere E, Boulo V, Liu ZJ, Gross PS, Cunningham C, Cupit PM, Tanguy A, Guo X, Moraga D, Boutet I, Huvet A, De Guise S, Almeida JS, Warr GW (2007) A cDNA Microarray of *Carinaria gigas*. *Mar Biotechnol (NY)*. 2007 Aug 1; [[PMID: 17688268](#)]

Vilela M, Borges CC, Vinga S, Vanconcelos AT, Santos H, Voit EO, Almeida JS. (2007) Automated smoother for the numerical decoupling of dynamics models. *BMC Bioinformatics* 8(1):305. [[PMID: 17711581](#)]

Vinga S, Almeida JS. (2007) Local Renyi entropic profiles of DNA sequences. *BMC Bioinformatics*. 2007 Oct 16;8(1):393. [[PMID: 17939871](#)]

Sá-Leão R, Nunes S, Brito-Avó A, Alves CR, Carriço JA, Saldanha J, Almeida JS, Santos-Sanches I, de Lencastre H. (2008) High rates of transmission of and colonization by *Streptococcus pneumoniae* and *Haemophilus influenzae* within a day care center revealed in a study. *J Clin Microbiol.* Jan;46(1):225-34. [[PMID: 18003797](#)]

Stanislau R, Jm Arthur, B Rajagopalan, R Moerschell, B McGoethlen, JS Almeida (2008). An open-source representation for 2-DE-centric proteomics and support infrastructure for data storage and analysis, *BMC Bioinformatics*. Jan 7;9:4. [[PMID: 18179696](#)]

Arthur JM, Janech MG, Varghese SA, Almeida JS, Powell TB (2008) Diagnostic and prognostic biomarkers in acute renal failure. *Contrib Nephrol* 160:53-64. [[PMID: 18179696](#)]

Vilela M, I-Chun Chou , S Vinga , ATR Vasconcelos , EO Voit, JS Almeida (2008) Parameter optimization in S-system models. *BMC Systems Biology* 2:35. [[PMID: 18416837](#)]

Robinson CJ, Swift S, Johnson DD, Almeida JS (2008) Prediction of pelvic organ prolapse using an artificial neural network. *American Journal of Obstetrics and Gynecology Jun 2.* [[PMID: 18533119](#)]

Deus FH, R Stanislau, DE Veiga, C Behrens, I Wistuba, JD Minna, HR Garner, SG Swisher, JA Roth, AM Correa, B Broom, K Coombes, A Chang, LH Vogel, JS Almeida (2008) A Semantic Web management model for integrative biomedical informatics. *PLoS ONE* 3(3):e2946 [[PMID: 18698353](#)].

Hennessy BT, M Murph, M Nanjundan, M Carey, N Auersperg, JS Almeida, Coombes KR, Liu J, Lu Y, Gray JW, Mills GB. Ovarian cancer: linking genomics to new target discovery and molecular markers--the way ahead. (2008) *Adv Exp Med Biol.* 617:23-40. [[PMID: 18722723](#)]

Stanislau R, M Carey, HF Deus, K Coombes, BT Hennessy, GB Mills, JS Almeida (2008) RPPAML/RIMS: A meta data format and an information management system for Reverse Phase Protein Arrays. *BMC Bioinformatics* 9(1):555. [[PMID 19102723](#)].

Freire P, M Vilela, HF Deus, YW Kim, D Koul, H Colman, KD Aldape, O Bogler, WKA Yung, K Coombes, GB Mills, AT Vasconcelos, JS Almeida (2008) Exploratory Analysis of the Copy Number Alterations in Glioblastoma Multiforme. *PLoS ONE* 3(12): e4076 [[PMID: 19102723](#)]

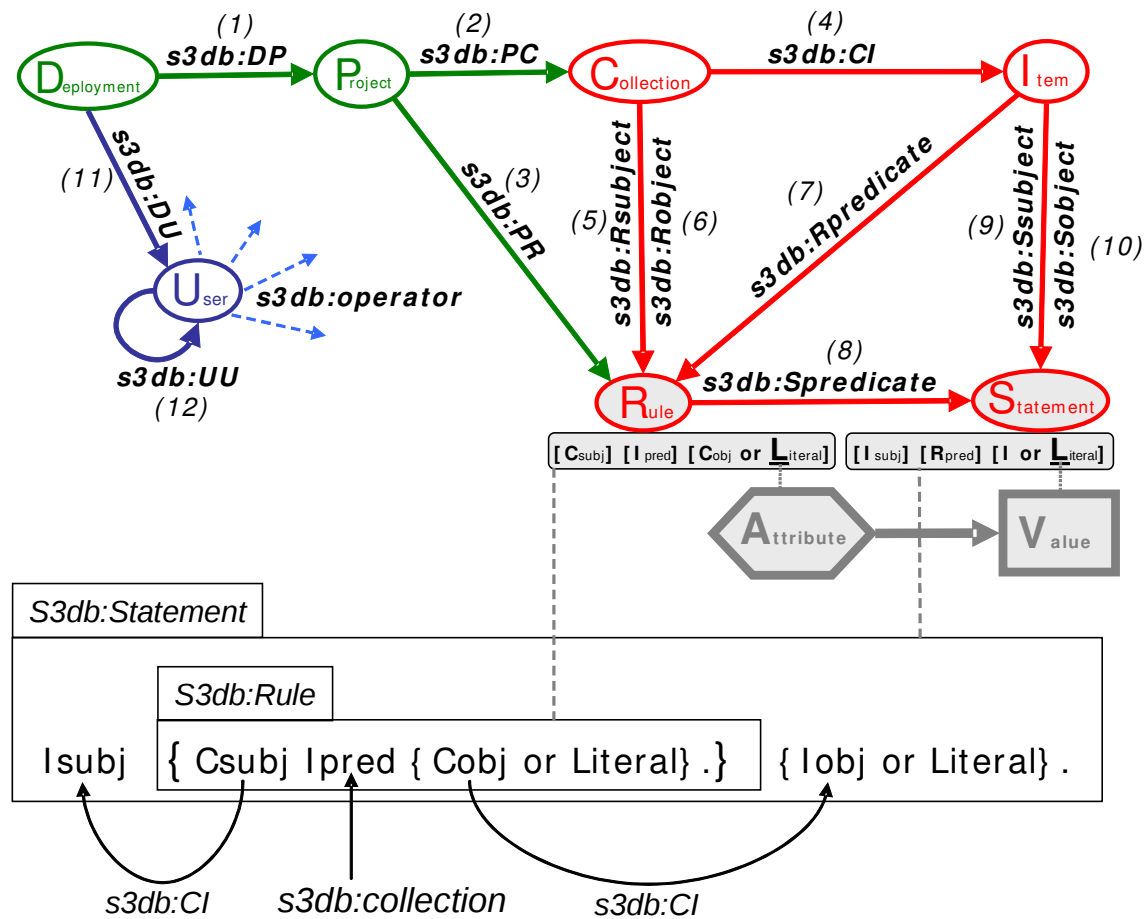
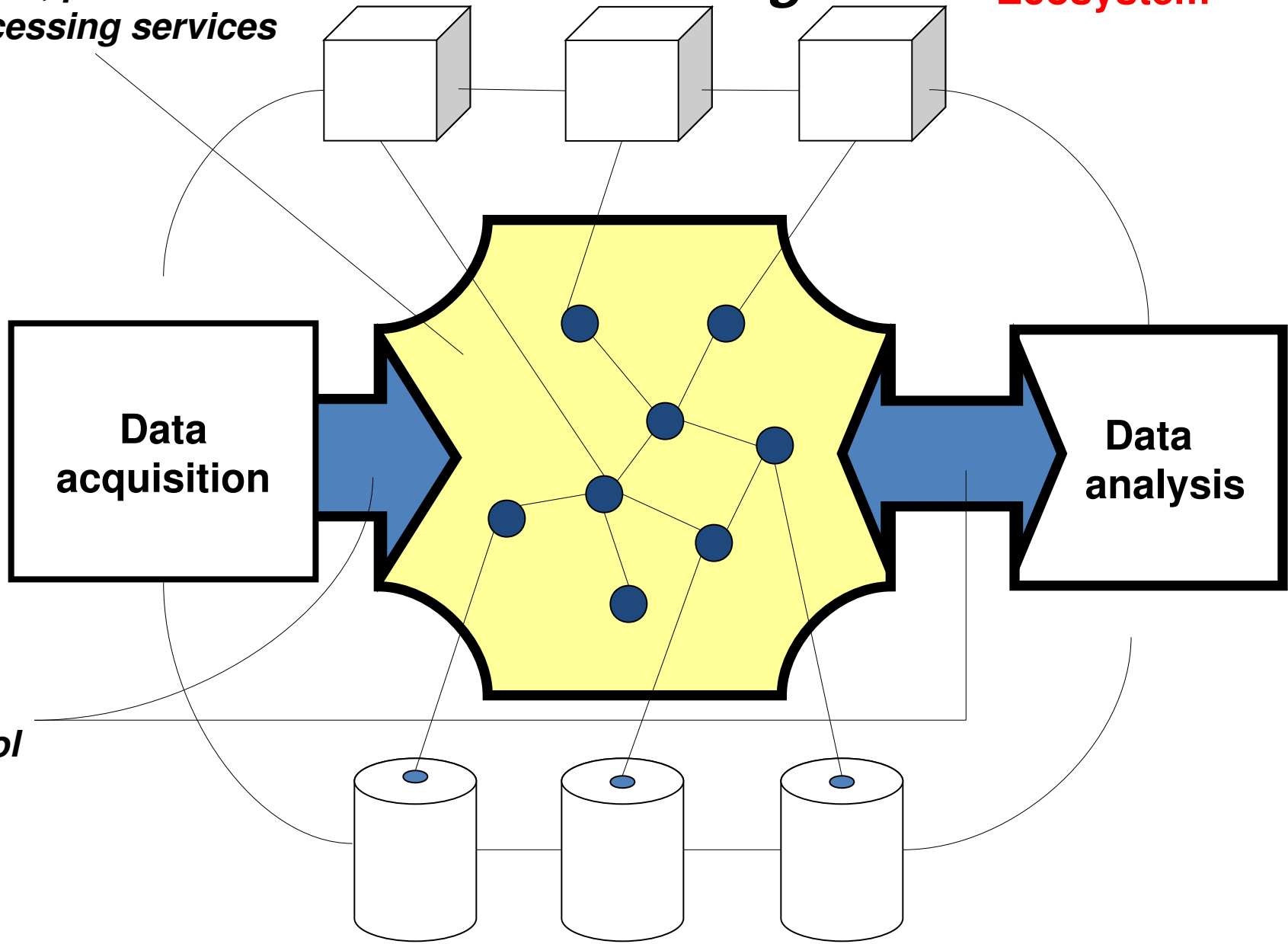


Figure 1 – Two views of the S3db core model. Top diagram - solid arrows describe relationship between the seven core entities; Dashed arrows (*s3db:operatorState*) indicate operators which have states that describe the relationship between users and each of the core entities. This core model encapsulates the key relationship between *s3db:rule* and *s3db:statement*, detailed in the lower part of the figure using N3 notation - the *s3db:rule* is a dyadic predicate and it is also, as a whole, the predicate of the *s3db:statement*. If the object of the *s3db:rule* triple is a literal attribute, then the object of the statement that rule predicates will be the attribute's literal value. Otherwise the statement object is the item of the collection indicated as object of the rule. The statement subject is invariable an item from the collection indicated as subject of the predicate rule. See text for nomenclature and definitions.

**RDF metadata linking URIs
of raw data, processed data
and processing services**

Processing

**Computational
Ecosystem**



Nature Precedings : doi:10.1038/npre.2011.5537.1 : Posted 11 Jan 2011

**REST
protocol**

Stores

Jonas Almeida @ Univ Texas MDAnderson Cancer Ctr

**RDF metadata linking URIs
of raw data, processed data
and processing services**

Syntactic interoperability

**REST, S(W)OA
and Cloud computing**

- *Organic development of analytical software applications integrated with other initiatives/resources.*

- *Programmatic interoperability by exposing API through REST.*

- *Interoperability with legacy systems because they are special realizations of more generic RDF based abstractions.*

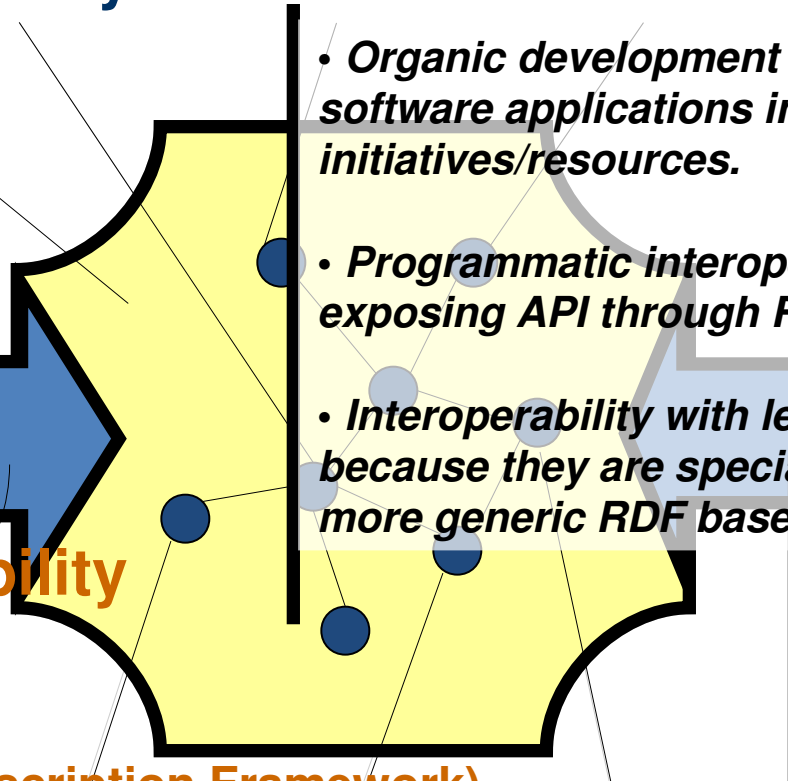


Semantic Interoperability

**REST
protocol**

RDF bus (Resource Description Framework)

Merged representation of data structures and workflows



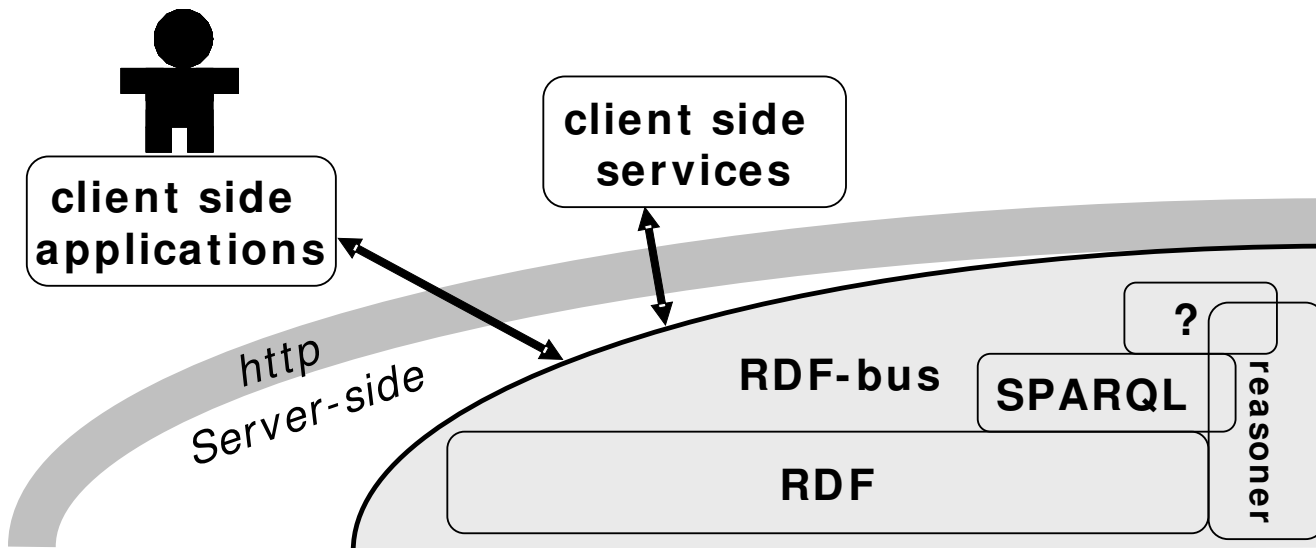


Figure 1 - Web-based infrastructure architecture composed of server side representation and client side presentation + data analysis computational services. This disposition moves to the client side both the assembly of interfaces as well as the computational intensive data analysis services – such as computational statistics modules. As a consequence, all server side components are standardized and can therefore benefit from cloud computing scaling.



Entrez Nucleotide

My NCBI
[Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

- Search Nucleotide for [Save Search](#)
- Jan 2014
- About Entrez
- Entrez Nucleotide
- Help | FAQ
- Entrez Tools
- Check sequence revision history
- LinkOut
- MyNCBI (Cubby)
- Related resources
- BAST
- Reference sequence project
- Nature Precedings
- Search for Genes
- Submit to GenBank
- Search for full length cDNAs

CGCTCAGGATACGACTTCGGCCGCTAGAGGATCGGATCCCCGGG...ATTATAATAGCTCGATCGATCT
TTCTCTATATCCGCGGAATGGG...FATATACACACACAGCGCCGCATAGCATGACTGATCT#
CCCACTGATACAGTTGAGTTCGCTTCGCATACGTCTGAG...ATAGCATGACTGATCT#
CACAGACTACGCGTTCAGTACTTAC TAAC CAATTCGGGAGAGGGGCGGAA TGGCGAGAG

NCBI Nucleotide

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for [] Go Clear

Limits Preview/Index History Clipboard Details

Display GenBank Send all to file

Range: from begin to end Reverse complemented strand Features: SNP CDD MGC HPRD STS

1: [CX769772](#). Reports est_1_vannamei579...[gi:58180125] Links

CX769772 446 bp mRNA linear EST 25-JAN-2005
 est_1_vannamei579 Luc_gil127t32d Litopenaeus vannamei cDNA, mRNA
 Sequence: [REDACTED]

ACCESSION CX769772
 VERSION CX769772.1 GI:58180125

EST [REDACTED]

SOURCE Litopenaeus vannamei (Pacific white shrimp)
 ORGANISM [Litopenaeus vannamei](#)
 Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
 Penaeidae; Litopenaeinae

REFERENCE 1 (bases 1-446)
 AUTHORS Bartlett, F.C., Cuthbertson, B.J., Shepard, E.F., Chapman, R.W., Gross,
 P.S., Robalino, J., Almeida, J.S., Chen, Y., McKillen, D.J., Wu, S. and
 Warr, G.W.
 Litopenaeus vannamei EST Library (L99-05) at marinegenomics.org
 Unpublished (2004)

COMMENT Contact: Gross, P.S.
 Department of Biochemistry and Molecular Biology
 Medical University of South Carolina
 Box 959509, 188 Ashley Avenue, Charleston, SC 29425, USA
 Tel: 843 792 8388
 Fax: 843 792 4850
 Email: grossp@musc.edu

source [REDACTED]
 [REDACTED]
 [REDACTED]
 /organism="Litopenaeus vannamei"
 /mol_type="mRNA"
 /db_xref="taxon:6689"

Posted 11 Jan 2004
 10.1038/nature01537
 Precedings: doi:10.1038/nature01537



PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for [] Go Clear

Limits Preview/Index History Clipboard Details

Display GenBank Send all to file

Range: from begin to end Reverse complemented strand Features: SNP CDD MGC HPRD STS

1: [CX769772](#). Reports est_1_vannamei579...[gi:58180125] Links

LOCUS CX769772 446 bp mRNA linear EST 25-JAN-2005

DEFINITION est_1_vannamei5798 LvG-gill27t32d Litopenaeus vannamei cDNA, mRNA sequence.

ACCESSION CX769772

VERSION CX769772.1 GI:58180125

KEYWORDS EST.

SOURCE Litopenaeus vannamei (Pacific white shrimp)

ORGANISM [Litopenaeus vannamei](#)
 Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
 Penaeidae; Litopenaeus.

REFERENCE 1 (bases 1 to 446)

AUTHORS Bartlett,T.C., Cuthbertson,B.J., Shepard,E.F., Chapman,R.W., Gross ,P.S., Robalino,J., Almeida,J.S., Chen,Y., McKillen,D.J., Wu,S. and Warr,G.W.

TITLE Litopenaeus vannamei EST Library (L99-05) at marinegenomics.org

JOURNAL Unpublished (2004)

COMMENT Contact: Gross PS
 Department of Biochemistry and Molecular Biology
 Medical University of South Carolina
 Box 250509, 173 Ashley Avenue, Charleston, SC 29425, USA
 Tel: 846 792 8503
 Fax: 843 792 4850
 Email: grossp@musc.edu.

FEATURES Location/Qualifiers
 source 1..446
 /organism="Litopenaeus vannamei"
 /mol_type="mRNA"
 /db_xref="taxon:6689"

11 Jan 2004
 Posted
 11:58
 Apr 2011
 10:1038
 Precedings
 Nature



PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for [] Go Clear

Limits Preview/Index History Clipboard Details

Display GenBank Send all to file

Change: from end Reverse complemented strand Features: [] SNP [] CDD [x] MGC [] HPRD [] STS

1: CX769 vannamei579...[gi:58180125]

Locus 446 bp mRNA linear EST 25-JAN-2005
 DEFINITION 5798 LvG-gill127t32d Litopenaeus vannamei cDNA, mRNA

FASTA
 EST/GSS
 XML :58180125
 GBSeq XML
 TinySeq XML
 SOURCE vannamei (Pacific white shrimp)
 ORGANISM vannamei
 Graph
 Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
 Penaeidae; Litopenaeus.

REFERENCE
 1 (bases 1 to 446)
 AUTHORS Bartlett,T.C., Cuthbertson,B.J., Shepard,E.F., Chapman,R.W., Gross ,P.S., Robalino,J., Almeida,J.S., Chen,Y., McKillen,D.J., Wu,S. and Warr,G.W.

TITLE Litopenaeus vannamei EST Library (L99-05) at marinegenomics.org
 JOURNAL Unpublished (2004)
 COMMENT Contact: Gross PS
 Department of Biochemistry and Molecular Biology
 Medical University of South Carolina
 Box 250509, 173 Ashley Avenue, Charleston, SC 29425, USA
 Tel: 846 792 8503
 Fax: 843 792 4850
 Email: grossp@musc.edu.

FEATURES Location/Qualifiers
 source 1..446
 /organism="Litopenaeus vannamei"
 /mol_type="mRNA"
 /db_xref="taxon:6689"

Jan 2004
 Posted
 11:59
 Apr 2001
 10:38
 Nature Precedings

Links



Search for

Limits Preview/Index History Clipboard Details

Display Features: SNP CDD MGC HPRD STS

1: [CX769](#)

446 bp mRNA linear EST 25-JAN-2005
 5798 LvG-gill127t32d *Litopenaeus vannamei* cDNA, mRNA

FASTA
 EST/GSS
 XML :58180125
GBSeq XML
 TinySeq XML *vannamei* (Pacific white shrimp)
 INSDSeq [vannamei](#)
 Graph *azoa; Arthropoda; Crustacea; Malacostraca;*

Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea; Penaeidae; Litopenaeus.

1 (bases 1 to 446)
REFERENCE
AUTHORS Bartlett,T.C., Cuthbertson,B.J., Shepard,E.F., Chapman,R.W., Gross ,P.S., Robalino,J., Almeida,J.S., Chen,Y., McKillen,D.J., Wu,S. and Warr,G.W.

TITLE *Litopenaeus vannamei* EST Library (L99-05) at marinegenomics.org
JOURNAL Unpublished (2004)

COMMENT
 Contact: Gross PS
 Department of Biochemistry and Molecular Biology
 Medical University of South Carolina
 Box 250509, 173 Ashley Avenue, Charleston, SC 29425, USA
 Tel: 846 792 8503
 Fax: 843 792 4850
 Email: grossp@musc.edu.

FEATURES
 Location/Qualifiers
 source 1..446
 /organism="Litopenaeus vannamei"
 /mol_type="mRNA"
 /db_xref="taxon:6689"

11 Jan 2004
 Posted
 11:58
 Apr 2001
 10:38
 Apr 2001
 Nature Precedings



PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for [] Go Clear

Limits Preview/Index History Clipboard Details

Display GBSeq XML Send all to file

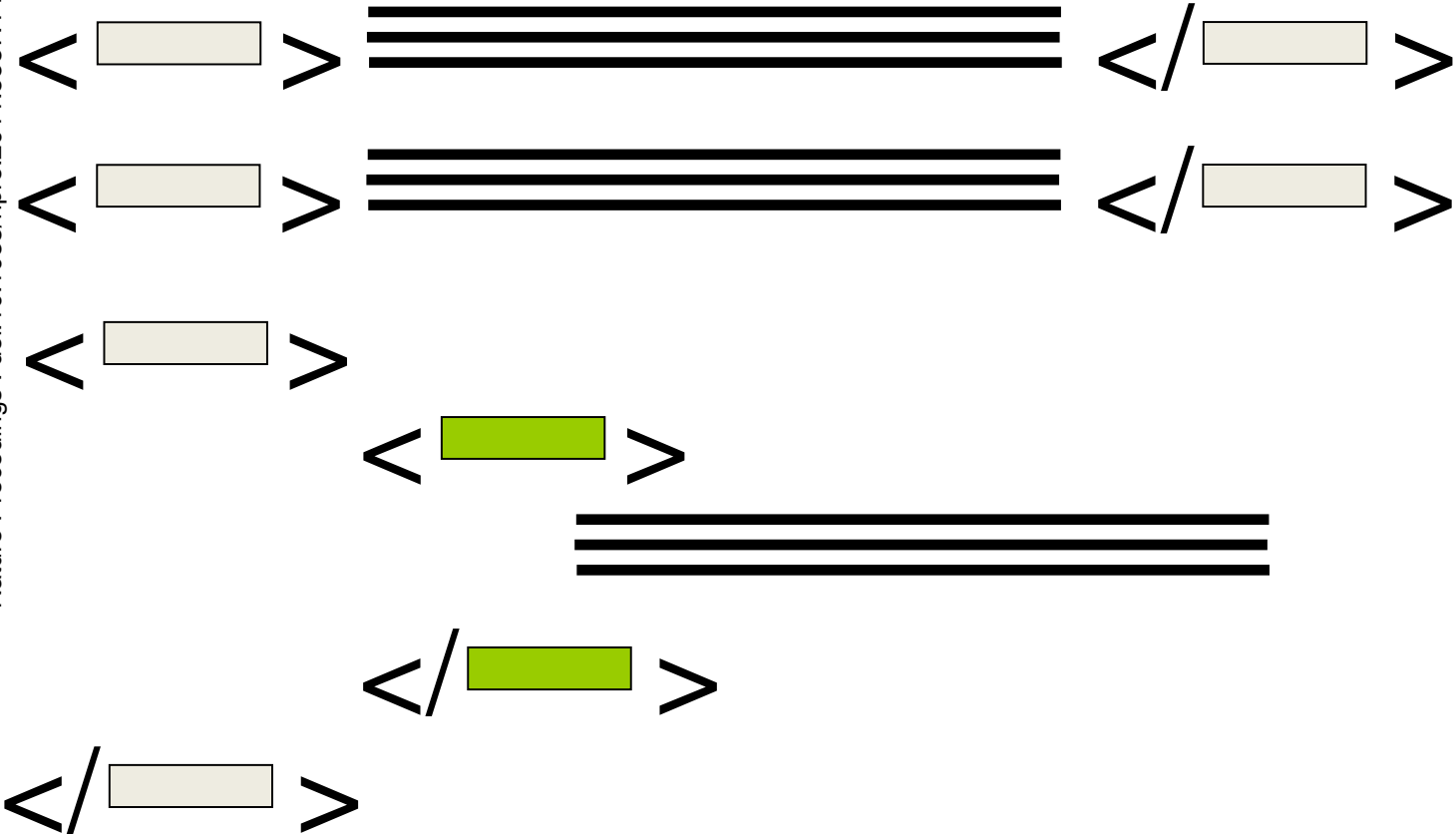
Range: from begin to end Reverse complemented strand Features: SNP CDD MGC HPRD STS

1: CX769772. Reports est_1_vannamei579... [gi:58180125] Links

```

<?xml version="1.0" encoding="UTF-8" type="text/plain">
<!DOCTYPE GBSeq PUBLIC "-//NCBI//NCBI-GBSeq/XML" http://www.ncbi.nlm.nih.gov/dtd/NCBI-GBSeq.dtd">
<GBSeq>
  <GBSeq_locus>CX769772</GBSeq_locus>
  <GBSeq_length>446</GBSeq_length>
  <GBSeq_strandedness value="single stranded">1</GBSeq_strandedness>
  <GBSeq_moltype value="mrna">5</GBSeq_moltype>
  <GBSeq_topology value="linear">1</GBSeq_topology>
  <GBSeq_division>EST</GBSeq_division>
  <GBSeq_update-date>25-JAN-2005</GBSeq_update-date>
  <GBSeq_create-date>25-JAN-2005</GBSeq_create-date>
  <GBSeq_definition>est_1_vannamei5798 LvG-gill127t32d Litopenaeus vannamei cDNA, mRNA sequence</GBSeq_definition>
  <GBSeq_primary-accession>CX769772</GBSeq_primary-accession>
  <GBSeq_accession-version>CX769772.1</GBSeq_accession-version>
  <GBSeq_other-seqids>
    <GBSeqid>gnl|dbEST|27391918</GBSeqid>
    <GBSeqid>gb|CX769772.1|</GBSeqid>
    <GBSeqid>gi|58180125</GBSeqid>
  </GBSeq_other-seqids>
  <GBSeq_keywords>
    <GBKeyword>EST</GBKeyword>
  </GBSeq_keywords>
  <GBSeq_source>Litopenaeus vannamei (Pacific white shrimp)</GBSeq_source>
  <GBSeq_organism>Litopenaeus vannamei</GBSeq_organism>
  <GBSeq_taxonomy>Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca; Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata</GBSeq_taxonomy>
  <GBReference>
    <GBReference_reference>1 (bases 1 to 446)</GBReference_reference>
  </GBReference>

```







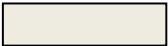



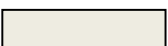











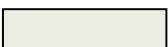






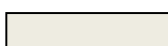
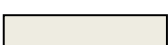

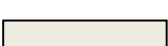













(T-Box)

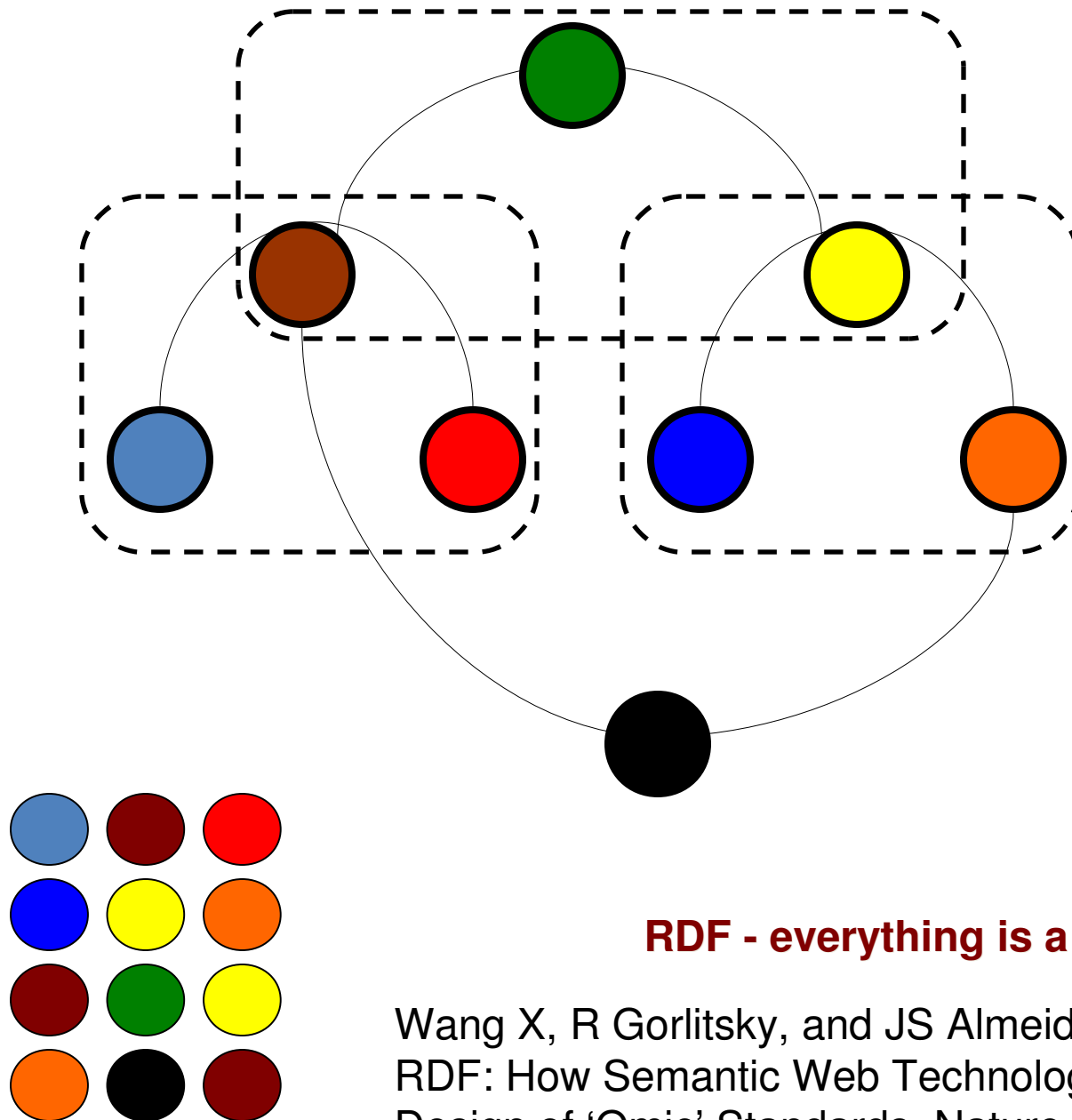
(A-Box)

Rules

Statements

	rel1	
	rel2	
	rel3	
	rel4	
	rel5	
	rel6	
	rel0	

	rel0	
	rel1	
	rel1	
	rel6	
	rel5	
	rel1	
	rel3	
	rel1	
	rel6	
	rel5	
	rel1	
	rel1	
	rel3	
	rel1	
	rel1	

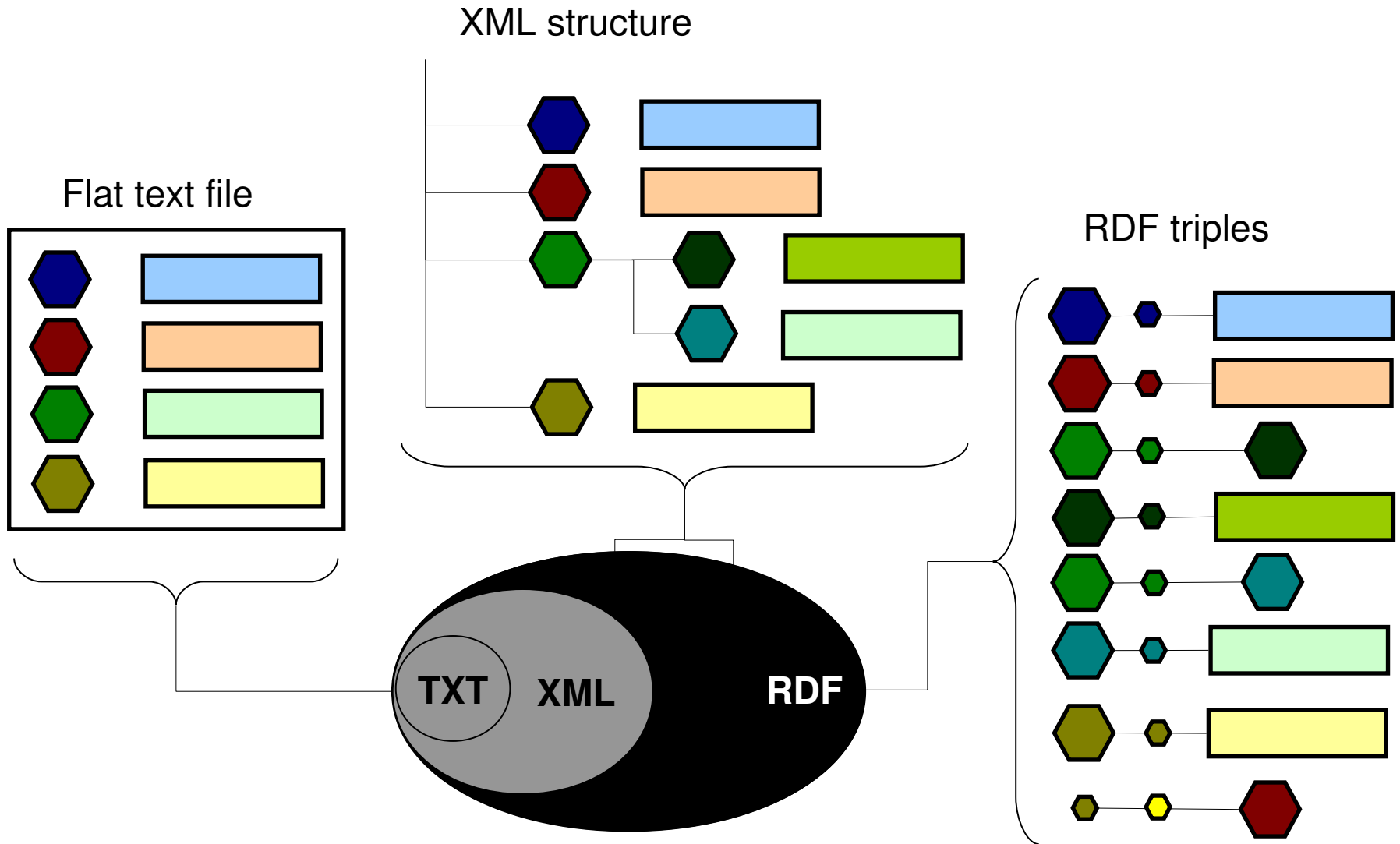


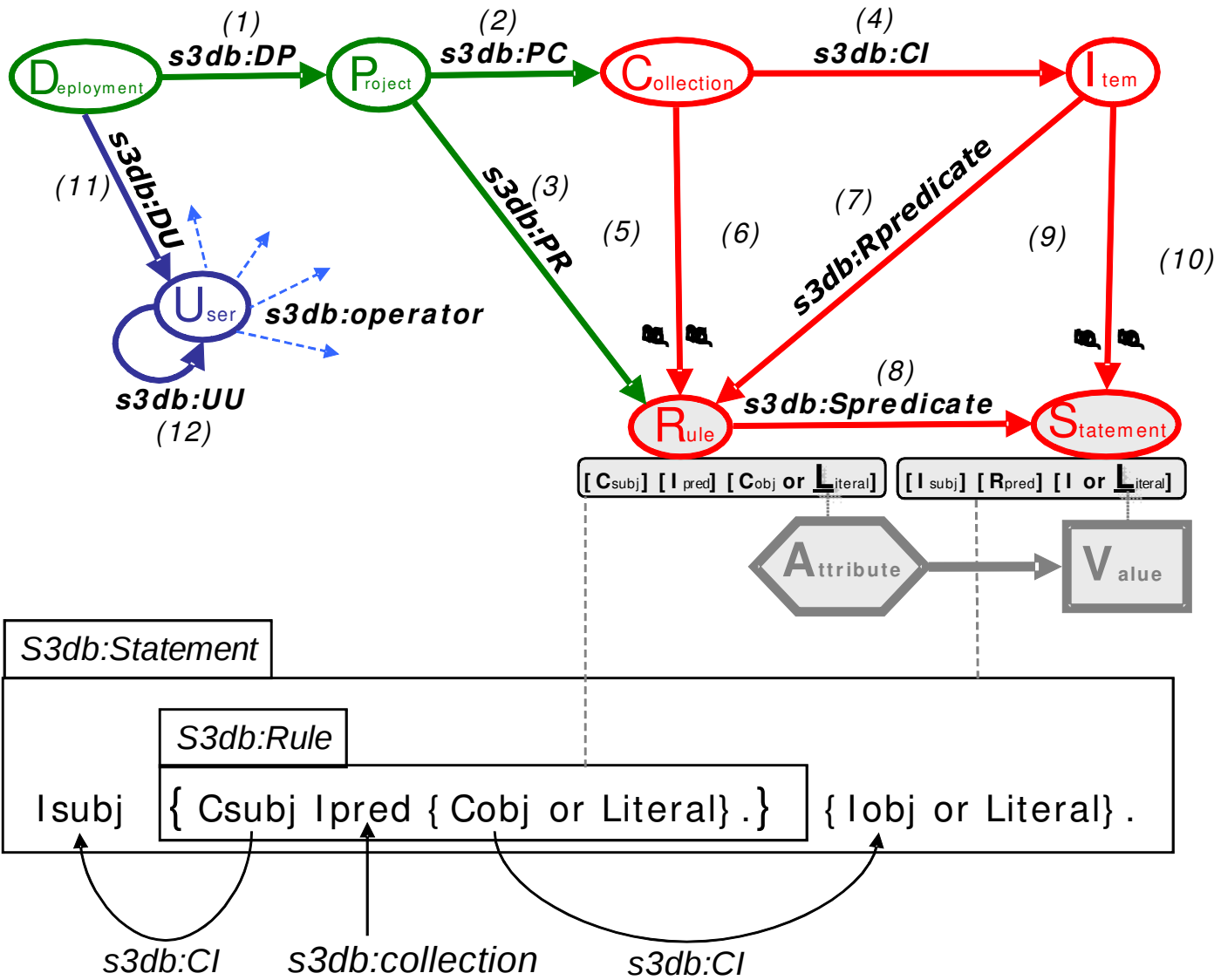
RDF - everything is a resource

Wang X, R Gorlitsky, and JS Almeida (2005) From XML to RDF: How Semantic Web Technologies Will Change the Design of 'Omic' Standards. Nature Biotechnology, Sep;23(9):1099-103 [[PMID:16151403](#)].

A brief history of data

Nature Precedings : doi:10.1038/npre.2011.5537.1 : Posted 11 Jan 2011





Minimal description of the core 12 relationships and 1 operator between the 7 s3db entities, using notation 3 (N3).

(s3db:deployment s3db:project s3db:collection s3db:item s3db:rule s3db:statement s3db:user) rdfs:subClassOf s3db:entity.

(s3db:DP s3db:PC s3db:PR s3db:CI s3db:CI s3db:Rsubject s3db:Robject s3db:Rpredicate s3db:Ssubject s3db:Sobject s3db:Spredicate) rdfs:subClassOf s3db:relationship.

1. *s3db:DP* rdfs:domain *s3db:deployment*; rdfs:range *s3db:project*.

2. *s3db:PC* rdfs:domain *s3db:project*; rdfs:range *s3db:collection*.

3. *s3db:PR* rdfs:domain *s3db:project*; rdfs:range *s3db:rule*.

4. *s3db:CI* rdfs:domain *s3db:collection*; rdfs:range *s3db:item*.

5. *s3db:Rsubject* owl:inverseOf *rdf:subject*; rdfs:domain *s3db:collection*; rdfs:range *s3db:rule*.

6. *s3db:Robject* owl:inverseOf *rdf:object*; rdfs:domain *s3db:collection*; rdfs:range *s3db:rule*.

7. *s3db:Rpredicate* owl:inverseOf *rdf:predicate*; rdfs:domain *s3db:item*; rdfs:range *s3db:rule*.

8. *s3db:Spredicate* owl:inverseOf *rdf:predicate*; rdfs:domain *s3db:rule*; rdfs:range *s3db:statement*.

9. *s3db:Ssubject* owl:inverseOf *rdf:subject*; rdfs:domain *s3db:item*; rdfs:range *s3db:statement*.

10. *s3db:Sobject* owl:inverseOf *rdf:object*; rdfs:domain *s3db:item*; rdfs:range *s3db:statement*.

11. *s3db:DU* rdfs:domain *s3db:deployment*; rdfs:range *s3db:user*.

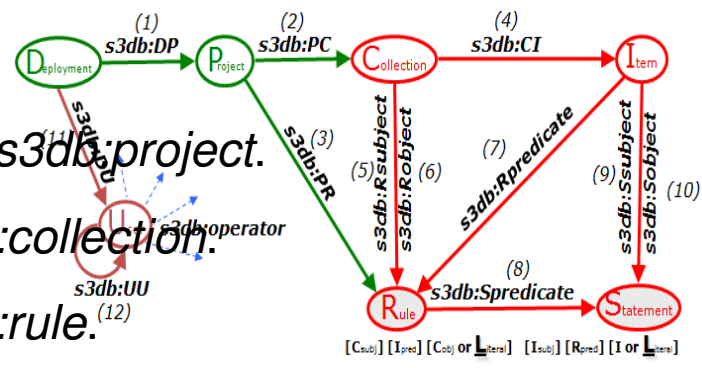
12. *s3db:UU* rdfs:domain *s3db:user*; rdfs:range *s3db:user*.

s3db:user *s3db:operator* *s3db:entity*.

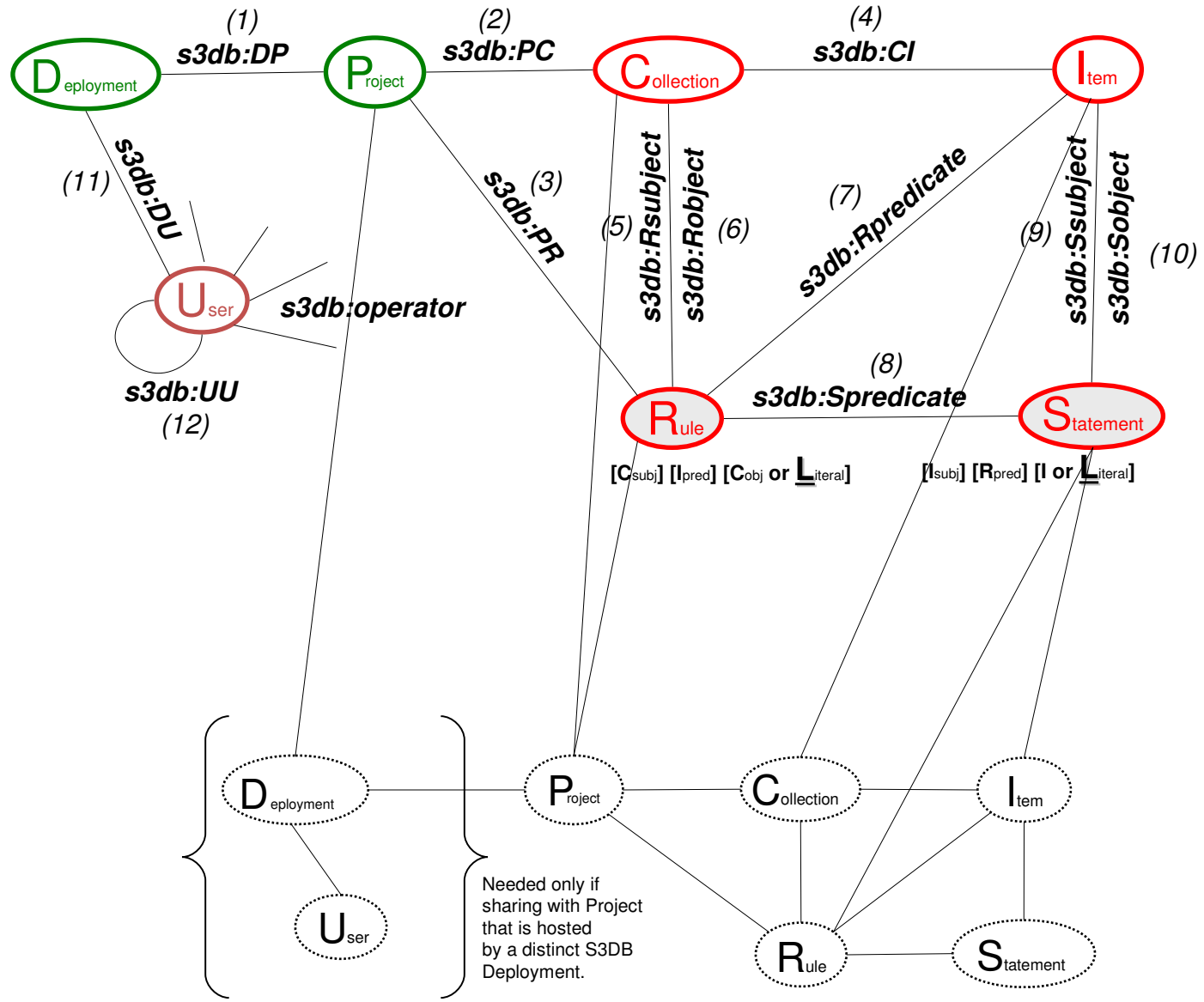
All relationships except for *s3db:operator* (last row) are *s3db:relationship* (first row). The inversion of RDF subject, predicate and object in relations 5-10 may appear capricious at this point but it will simplify the identification of automata for the propagation of *s3db:operator* states in the next section. Specifically, it will allow the definition of Equation 3 such that the direction of the arrows in Figure 2 is the same as the propagation of *s3db:operator* states.

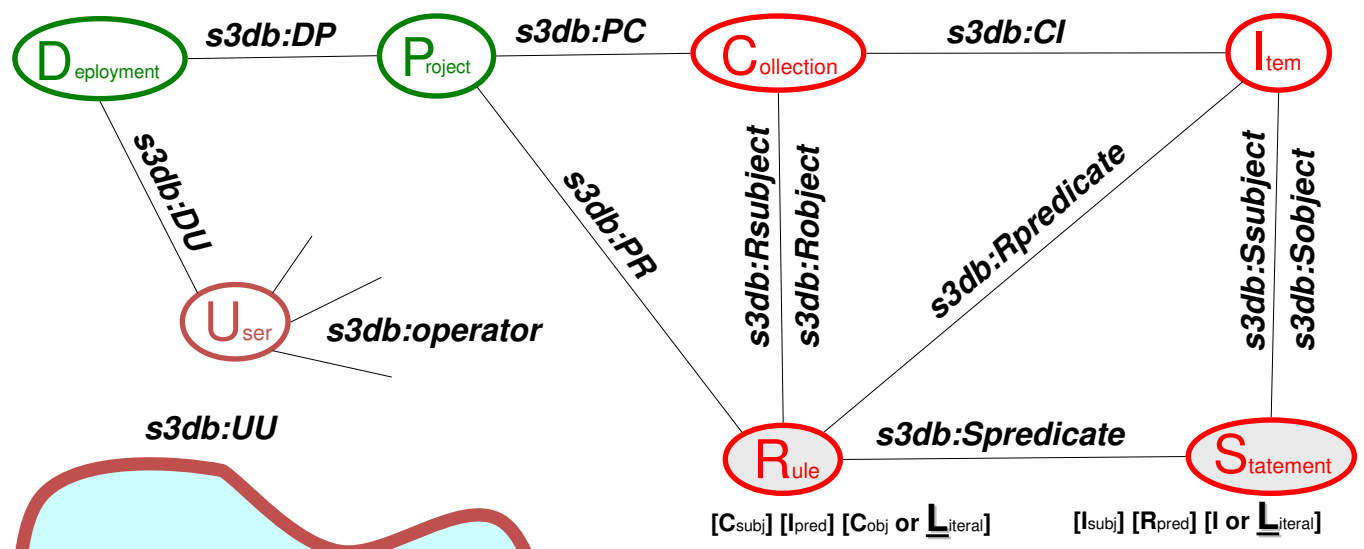
Nature Precedings doi:10.1038/npre.2011.5537.1

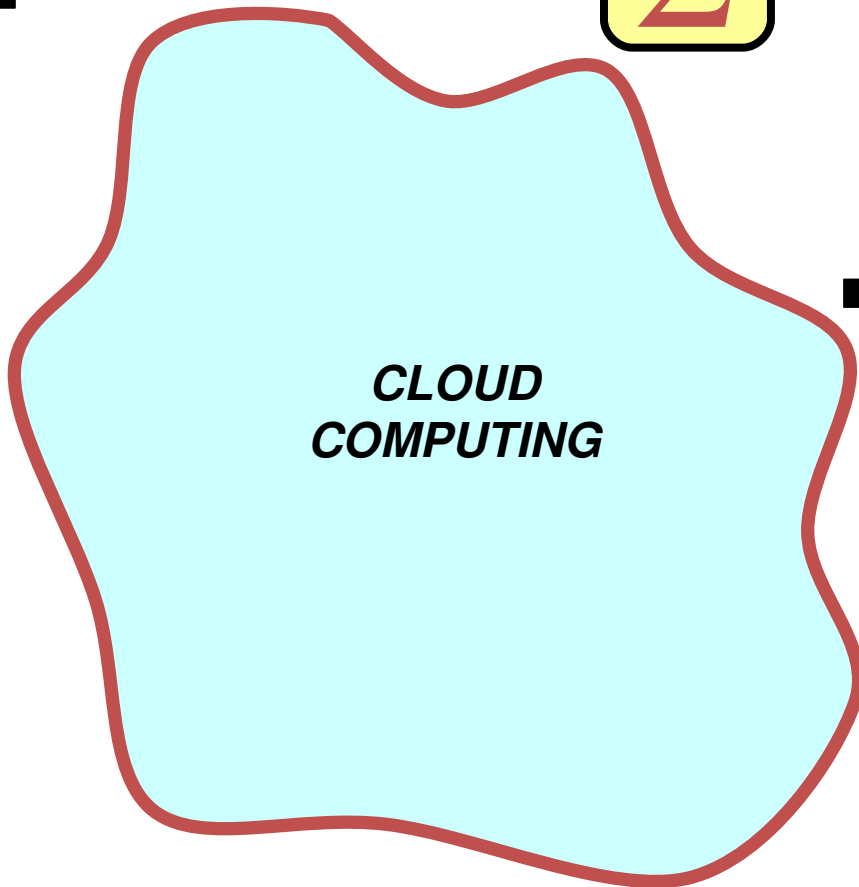
14 Jan 2011

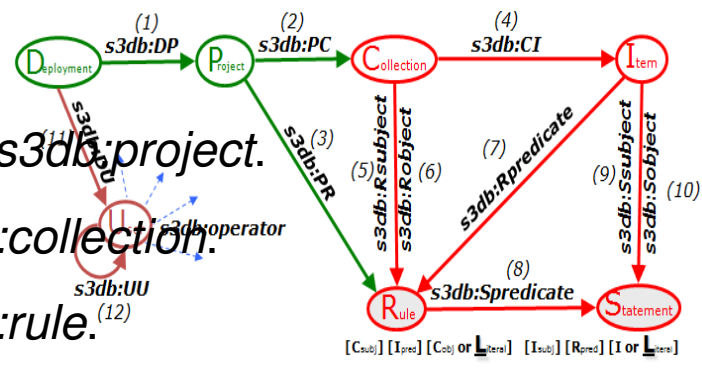


- 1. `s3db:DP` `rdfs:domain` `s3db:deployment`; `rdfs:range` `s3db:project`.
- 2. `s3db:PC` `rdfs:domain` `s3db:project`; `rdfs:range` `s3db:collection`.
- 3. `s3db:PR` `rdfs:domain` `s3db:project`; `rdfs:range` `s3db:rule`.
- 4. `s3db:CI` `rdfs:domain` `s3db:collection`; `rdfs:range` `s3db:item`.
- 5. `s3db:Rsubject` `owl:inverseOf` `rdf:subject`; `rdfs:domain` `s3db:collection`; `rdfs:range` `s3db:rule`.
- 6. `s3db:Robject` `owl:inverseOf` `rdf:object`; `rdfs:domain` `s3db:collection`; `rdfs:range` `s3db:rule`.
- 7. `s3db:Rpredicate` `owl:inverseOf` `rdf:predicate`; `rdfs:domain` `s3db:item`; `rdfs:range` `s3db:rule`.
- 8. `s3db:Spredicate` `owl:inverseOf` `rdf:predicate`; `rdfs:domain` `s3db:rule`; `rdfs:range` `s3db:statement`.
- 9. `s3db:Ssubject` `owl:inverseOf` `rdf:subject`; `rdfs:domain` `s3db:item`; `rdfs:range` `s3db:statement`.
- 10. `s3db:Sobject` `owl:inverseOf` `rdf:object`; `rdfs:domain` `s3db:item`; `rdfs:range` `s3db:statement`.









- 1. *s3db:DP* *rdfs:domain* *s3db:deployment*; *rdfs:range* *s3db:project*.
- 2. *s3db:PC* *rdfs:domain* *s3db:project*; *rdfs:range* *s3db:collection*.
- 3. *s3db:PR* *rdfs:domain* *s3db:project*; *rdfs:range* *s3db:rule*.
- 4. *s3db:CI* *rdfs:domain* *s3db:collection*; *rdfs:range* *s3db:item*.
- 5. *s3db:Rsubject* *owl:inverseOf* *rdf:subject*; *rdfs:domain* *s3db:collection*; *rdfs:range* *s3db:rule*.
- 6. *s3db:Robject* *owl:inverseOf* *rdf:object*; *rdfs:domain* *s3db:collection*; *rdfs:range* *s3db:rule*.
- 7. *s3db:Rpredicate* *owl:inverseOf* *rdf:predicate*; *rdfs:domain* *s3db:item*; *rdfs:range* *s3db:rule*.
- 8. *s3db:Spredicate* *owl:inverseOf* *rdf:predicate*; *rdfs:domain* *s3db:rule*; *rdfs:range* *s3db:statement*.
- 9. *s3db:Ssubject* *owl:inverseOf* *rdf:subject*; *rdfs:domain* *s3db:item*; *rdfs:range* *s3db:statement*.
- 10. *s3db:Sobject* *owl:inverseOf* *rdf:object*; *rdfs:domain* *s3db:item*; *rdfs:range* *s3db:statement*.

f *subClassOf* $s3db : operator.$

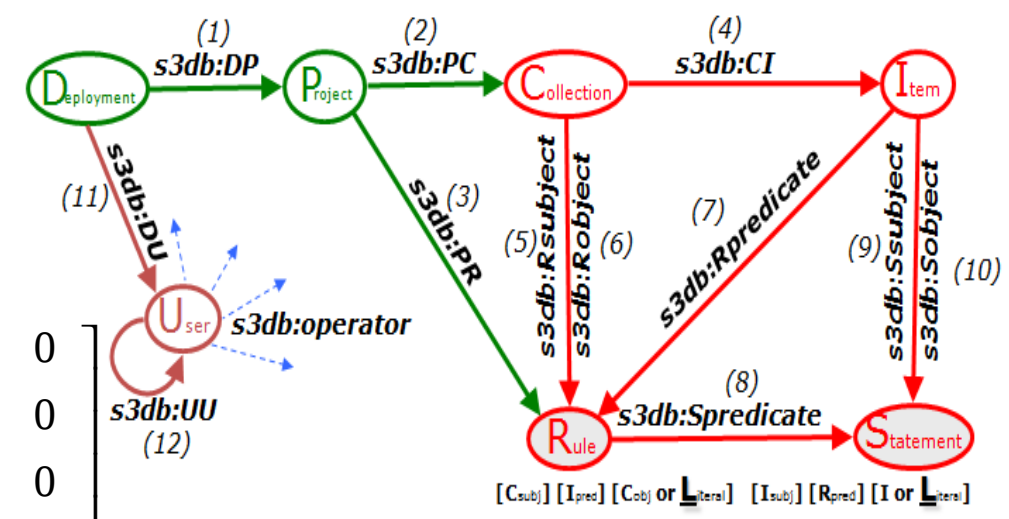
(ϕ, Φ_i) *subClassOf* $f.$

U_some_user (ϕ, Φ_i) $E_some_entity.$

$$i = \text{merge}(\{\Phi_A, \phi_a\}) \rightarrow \begin{cases} i_{|A=null} = \max(a) \\ i_{|A \neq null} = \min(A) \end{cases}$$

$$T_{3DB}^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ (1) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (2) & 0 & 0 & 0 & 0 & 0 \\ 0 & (3) & [(5), (6)] & 0 & (7) & 0 & 0 \\ 0 & 0 & (4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (8) & [(9), (10)] & 0 & 0 \\ (11) & 0 & 0 & 0 & 0 & 0 & (12) \end{bmatrix}$$

$$\begin{bmatrix} D \\ P \\ C \\ R \\ I \\ S \\ U \end{bmatrix}_{k+1} = \text{merge}(T \times \begin{bmatrix} D \\ P \\ C \\ R \\ I \\ S \\ U \end{bmatrix}_k)$$



$$f_{object,k+1} = merge([f_{object,k}, migrate(f_{subject,k})])$$

$$l = length(f)$$

$$l = 1 \rightarrow migrate(f) = f = f[1]$$

$$l > 1 \rightarrow migrate(f) = f[2, \dots, l]$$

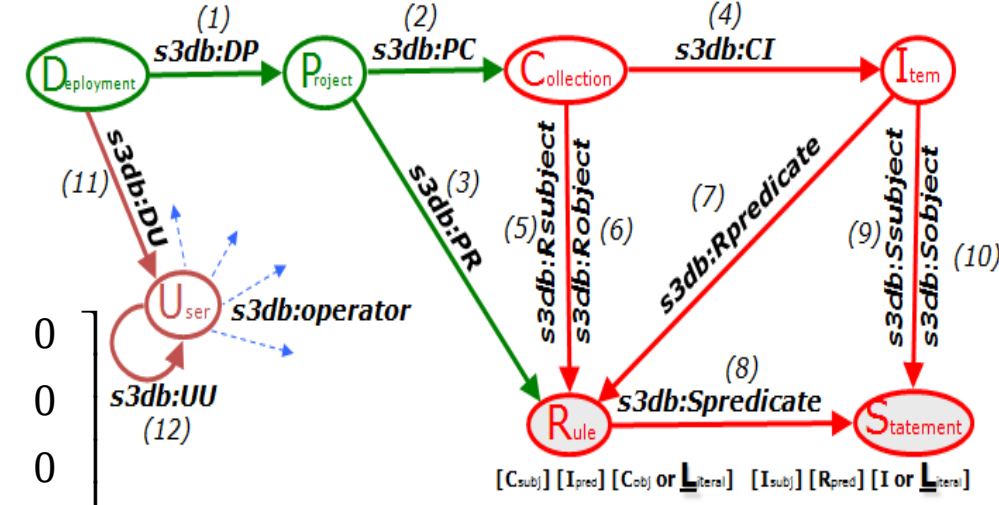
$$i = [1 + m, \dots, 2m]$$

$$i > l, i - m > 0 \rightarrow f[i] = f[i - m]$$

$$i > l, i - m \leq 0 \rightarrow f[i] = f[i - 1]$$

$$migrate(f, m) = f[m + 1, 2m]$$

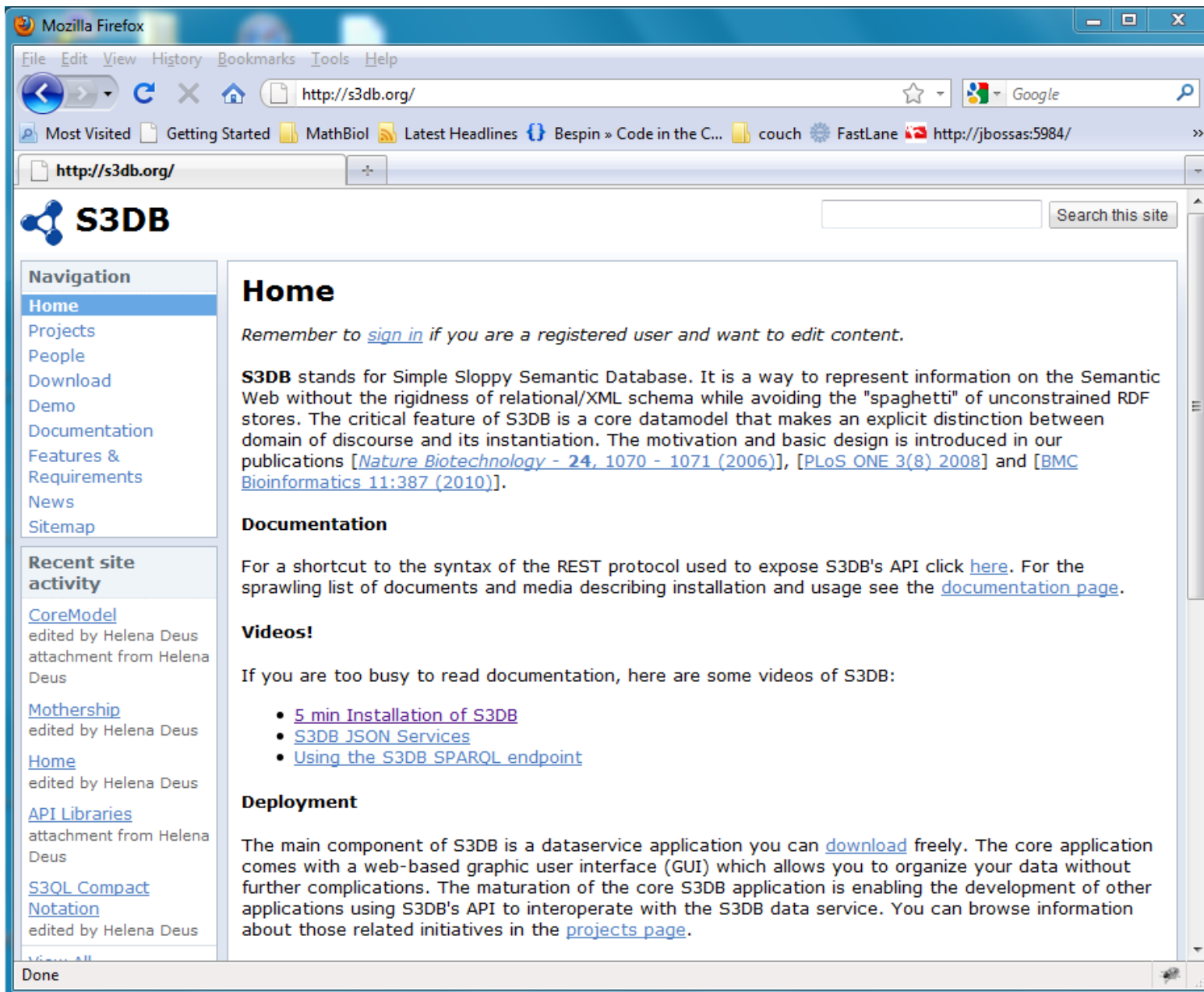
$$T_{SDB}^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ (1) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (2) & 0 & 0 & 0 & 0 & 0 \\ 0 & (3) & [(5), (6)] & 0 & (7) & 0 & 0 \\ 0 & 0 & (4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (8) & [(9), (10)] & 0 & 0 \\ (11) & 0 & 0 & 0 & 0 & 0 & (12) \end{bmatrix}$$



$$E_{k+1} = E_k$$

$$\begin{bmatrix} D \\ P \\ C \\ R \\ I \\ S \\ U \end{bmatrix}_{k+1} = \text{merge}(T \times \begin{bmatrix} D \\ P \\ C \\ R \\ I \\ S \\ U \end{bmatrix}_k)$$

<http://s3db-operator.googlecode.com>



The screenshot shows a Mozilla Firefox browser window displaying the S3DB website. The address bar shows the URL <http://s3db.org/>. The page features a navigation menu on the left, a main content area with a 'Home' section, and a 'Recent site activity' sidebar. The 'Home' section includes a sign-in reminder, a description of S3DB, documentation links, a video list, and deployment information.

Navigation

- Home
- Projects
- People
- Download
- Demo
- Documentation
- Features & Requirements
- News
- Sitemap

Recent site activity

- [CoreModel](#)
edited by Helena Deus
attachment from Helena Deus
- [Mothership](#)
edited by Helena Deus
- [Home](#)
edited by Helena Deus
- [API Libraries](#)
attachment from Helena Deus
- [SQL Compact Notation](#)
edited by Helena Deus

Home

Remember to [sign in](#) if you are a registered user and want to edit content.

S3DB stands for Simple Sloppy Semantic Database. It is a way to represent information on the Semantic Web without the rigidity of relational/XML schema while avoiding the "spaghetti" of unconstrained RDF stores. The critical feature of S3DB is a core datamodel that makes an explicit distinction between domain of discourse and its instantiation. The motivation and basic design is introduced in our publications [[Nature Biotechnology - 24, 1070 - 1071 \(2006\)](#)], [[PLoS ONE 3\(8\) 2008](#)] and [[BMC Bioinformatics 11:387 \(2010\)](#)].

Documentation

For a shortcut to the syntax of the REST protocol used to expose S3DB's API click [here](#). For the sprawling list of documents and media describing installation and usage see the [documentation page](#).

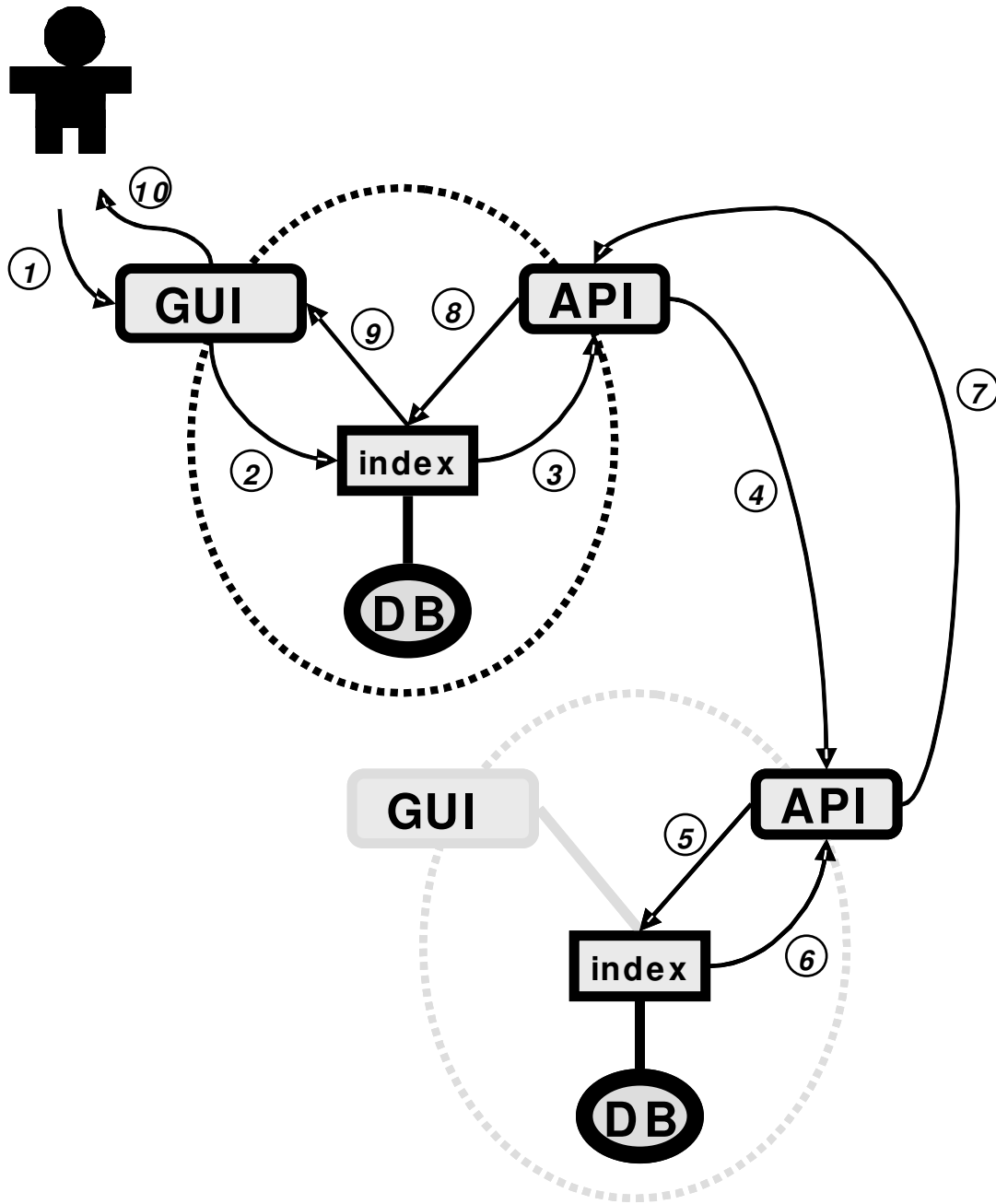
Videos!

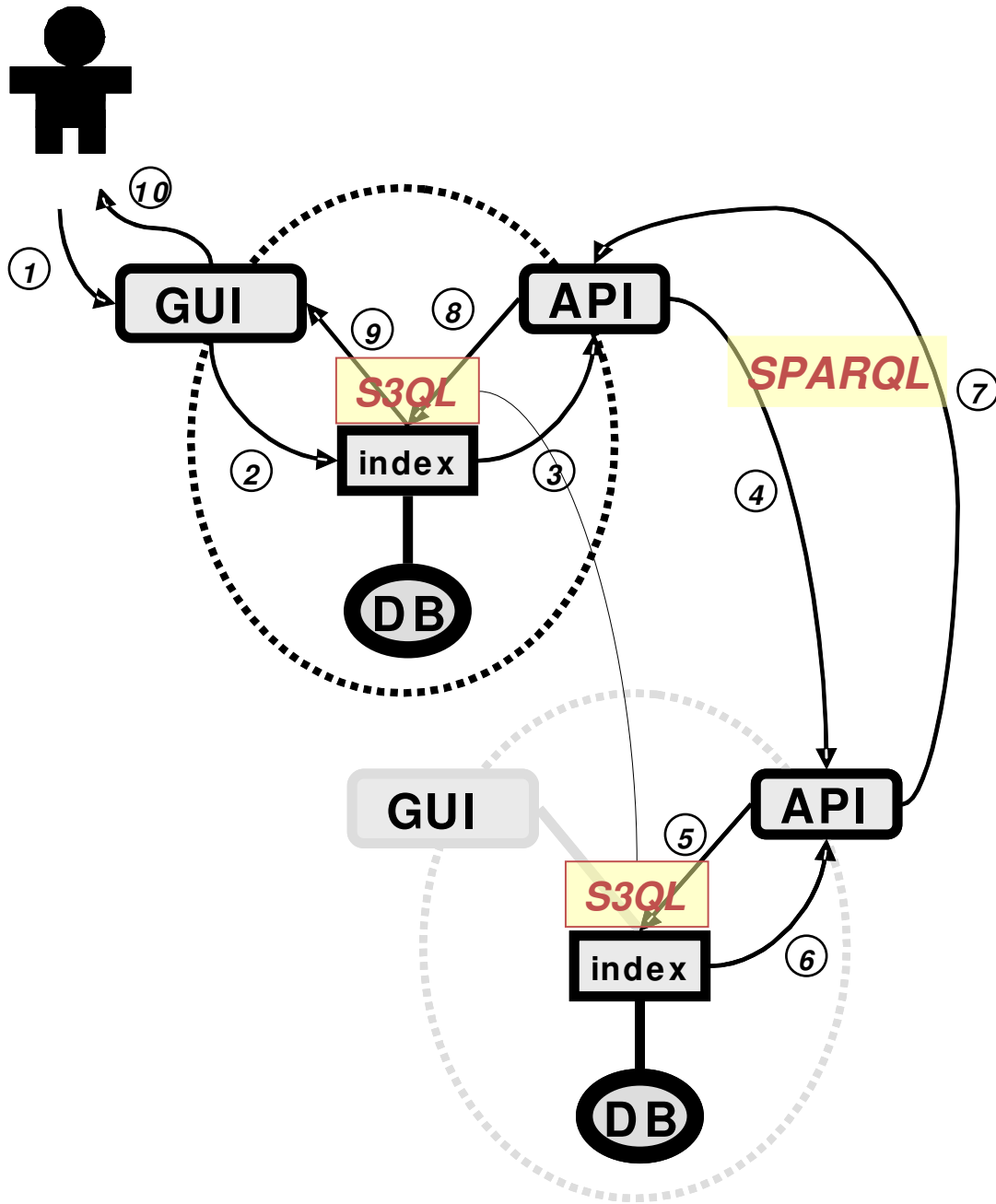
If you are too busy to read documentation, here are some videos of S3DB:

- [5 min Installation of S3DB](#)
- [S3DB JSON Services](#)
- [Using the S3DB SPARQL endpoint](#)

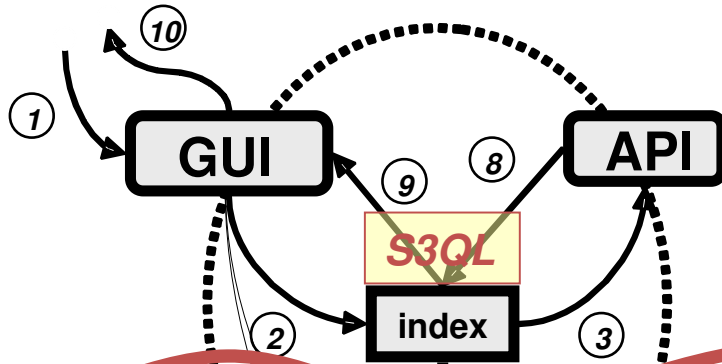
Deployment

The main component of S3DB is a dataservice application you can [download](#) freely. The core application comes with a web-based graphic user interface (GUI) which allows you to organize your data without further complications. The maturation of the core S3DB application is enabling the development of other applications using S3DB's API to interoperate with the S3DB data service. You can browse information about those related initiatives in the [projects page](#).



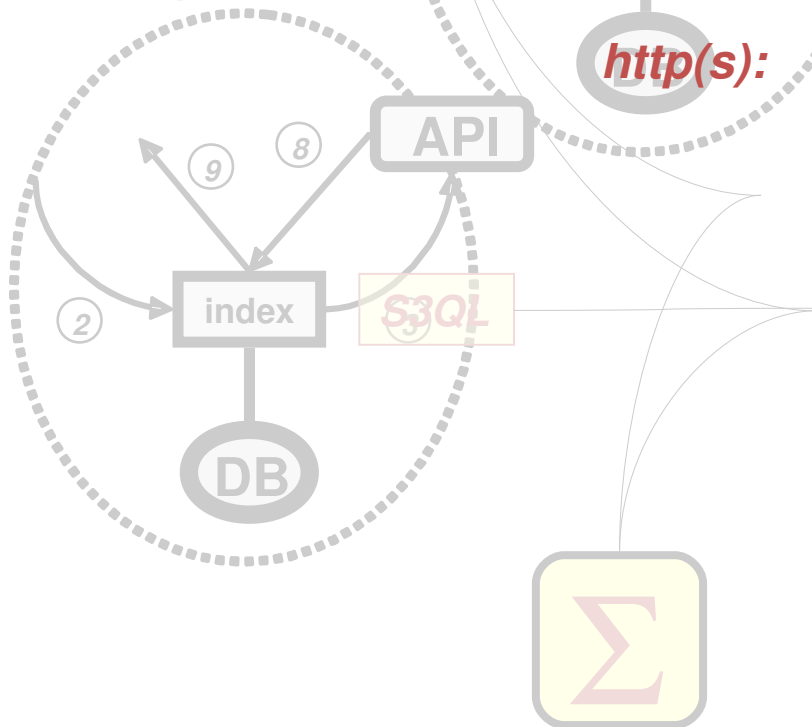


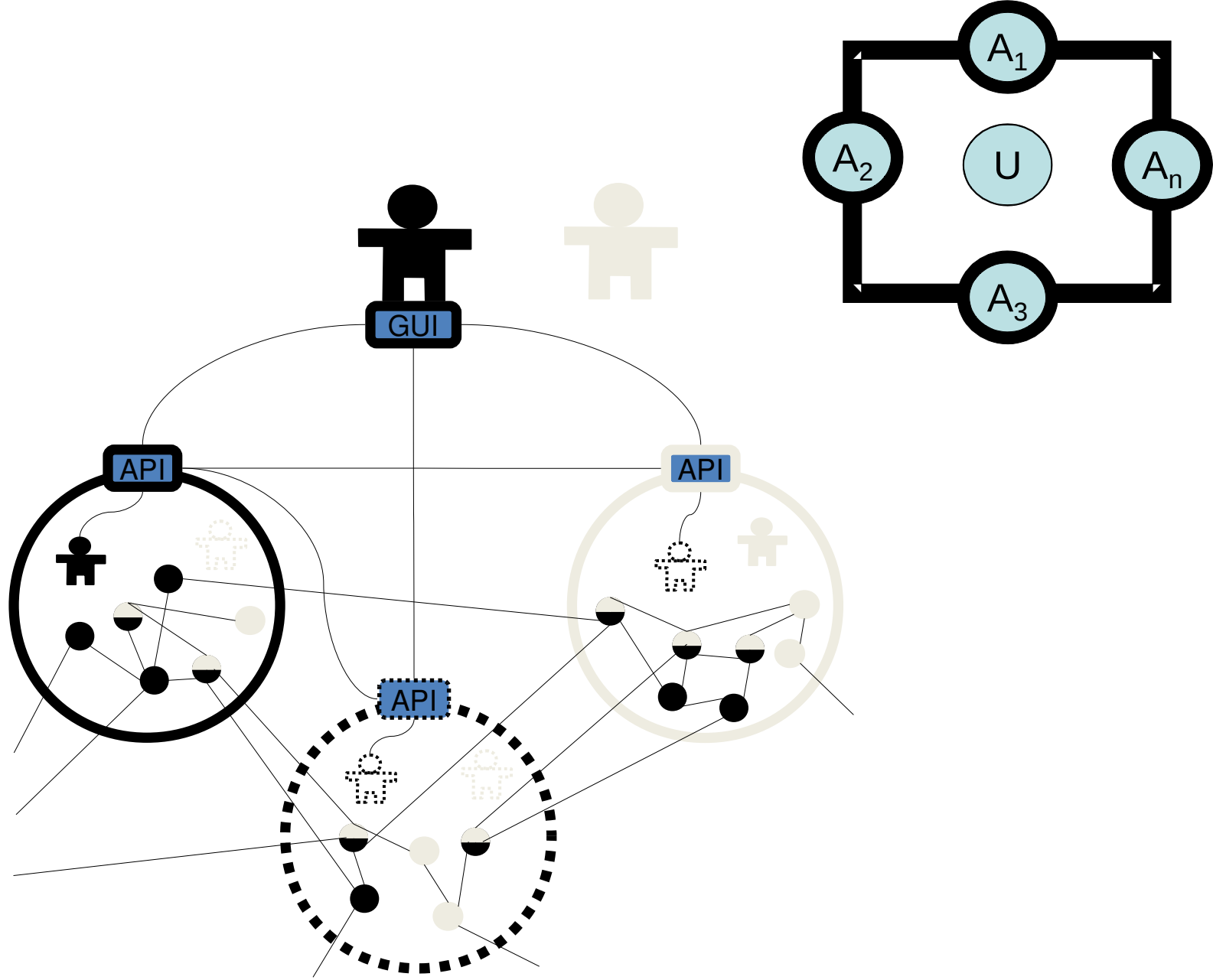
Nature Precedings : doi:10.1038/npre.2011.5537.1 : Posted 11 Jan 2011



SPARQL

http(s):





Data of Sample (Project: UCSF and MDACC ovary cancer cell lines)

Page 1

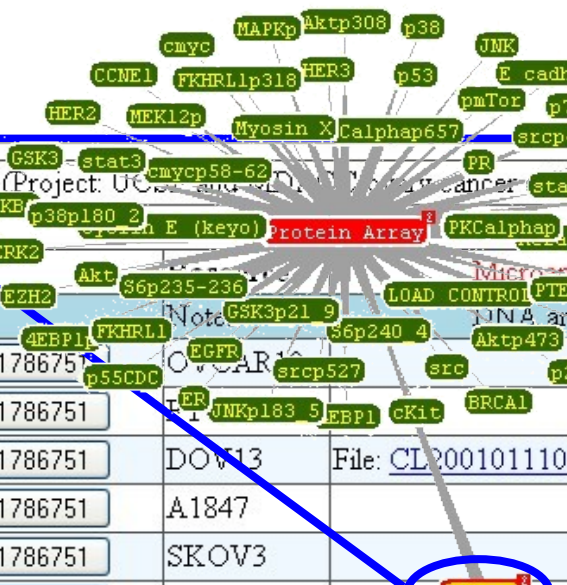
Resource	Notes
006389600116178675	EGFR
0183129001161786751	BRCA1
0308397001161786751	Down 13 File: CL200101110
0427069001161786751	A1847
0573273001161786751	SKOV3
0108892001161786750	OCC1
0239669001161786750	ES2

Data of Patient (Project: UCSF and MDACC ovary cancer cell lines)

Page 1

Resource	Resource	Clinical	Clinical	Trial
UID	Notes	Clinical Diagnostic	Clinical Outcome	Sample
003875	UCSF Patient#1814-6	UCSF Patient#1814-6 (id 003719)	UCSF Patient#1814-6 (id 003758)	UCSF Patient#1814-6 (id 003836)
003874	UCSF Patient#1173-1	UCSF Patient#1173-1 (id 003718)	UCSF Patient#1173-1 (id 003757)	UCSF Patient#1173-1 (id 003835)
003873	UCSF Patient#1161-9	UCSF Patient#1161-9 (id 003717)	UCSF Patient#1161-9 (id 003756)	UCSF Patient#1161-9 (id 003834)
003872	UCSF Patient#1156-18	UCSF Patient#1156-18 (id 003716)	UCSF Patient#1156-18 (id 003755)	UCSF Patient#1156-18 (id 003833)
003871	UCSF Patient#1149-18	UCSF Patient#1149-18 (id 003715)	UCSF Patient#1149-18 (id 003754)	UCSF Patient#1149-18 (id 003832)
003870	UCSF Patient#1123-3	UCSF Patient#1123-3 (id 003714)	UCSF Patient#1123-3 (id 003753)	UCSF Patient#1123-3 (id 003831)
003869	UCSF Patient#1106-17	UCSF Patient#1106-17 (id 003713)	UCSF Patient#1106-17 (id 003752)	UCSF Patient#1106-17 (id 003830)
003868	UCSF Patient#1072-3	UCSF Patient#1072-3 (id 003712)	UCSF Patient#1072-3 (id 003751)	UCSF Patient#1072-3 (id 003829)
003867	UCSF Patient#1069-37	UCSF Patient#1069-37 (id 003711)	UCSF Patient#1069-37 (id 003750)	UCSF Patient#1069-37 (id 003828)
003866	UCSF Patient#1068-43	UCSF Patient#1068-43 (id 003710)	UCSF Patient#1068-43 (id 003749)	UCSF Patient#1068-43 (id 003827)
003865	UCSF Patient#1065-20	UCSF Patient#1065-20 (id 003709)	UCSF Patient#1065-20 (id 003748)	UCSF Patient#1065-20 (id 003826)
003864	UCSF Patient#1032-5	UCSF Patient#1032-5 (id 003708)	UCSF Patient#1032-5 (id 003747)	UCSF Patient#1032-5 (id 003825)
003863	UCSF Patient#1028-21	UCSF Patient#1028-21 (id 003707)	UCSF Patient#1028-21 (id 003746)	UCSF Patient#1028-21 (id 003824)
003862	UCSF Patient#1010-2	UCSF Patient#1010-2 (id 003706)	UCSF Patient#1010-2 (id 003745)	UCSF Patient#1010-2 (id 003823)
003861	UCSF Patient#972-29	UCSF Patient#972-29 (id 003705)	UCSF Patient#972-29 (id 003744)	UCSF Patient#972-29 (id 003822)
003860	UCSF Patient#878-17	UCSF Patient#878-17 (id 003704)	UCSF Patient#878-17 (id 003743)	UCSF Patient#878-17 (id 003821)
003859	UCSF Patient#747-4	UCSF Patient#747-4 (id 003703)	UCSF Patient#747-4 (id 003742)	UCSF Patient#747-4 (id 003820)
003858	UCSF Patient#703-4	UCSF Patient#703-4 (id 003702)	UCSF Patient#703-4 (id 003741)	UCSF Patient#703-4 (id 003819)
003857	UCSF Patient#389-6	UCSF Patient#389-6 (id 003701)	UCSF Patient#389-6 (id 003740)	UCSF Patient#389-6 (id 003818)
003856	UCSF Patient#387-3	UCSF Patient#387-3 (id 003700)	UCSF Patient#387-3 (id 003739)	UCSF Patient#387-3 (id 003817)

OCC1	(id 032356 Cancer 34374)
ES2	(id Patient 16184372)
OAW28	(id 0437638001161784370)
EF021	(id 0306293001161784368)
OC314	(id 0297762001161784366)
UPN251	(id 0351945001161784364)
IGROV1	(id 0448438001161784362)
OVCA432	(id 0583980001161784389)
OVCA429	(id 0766966001161784367)
OVCA420	(id 0870650001161784385)
OVCAR3	(id 0862921001161784383)
OSE80	(id 0035834001161784382)
IOSE29	(id 0160421001161784380)
OSE1162	(id 0226465001161784378)
EF027	(id 0295613001161784376)

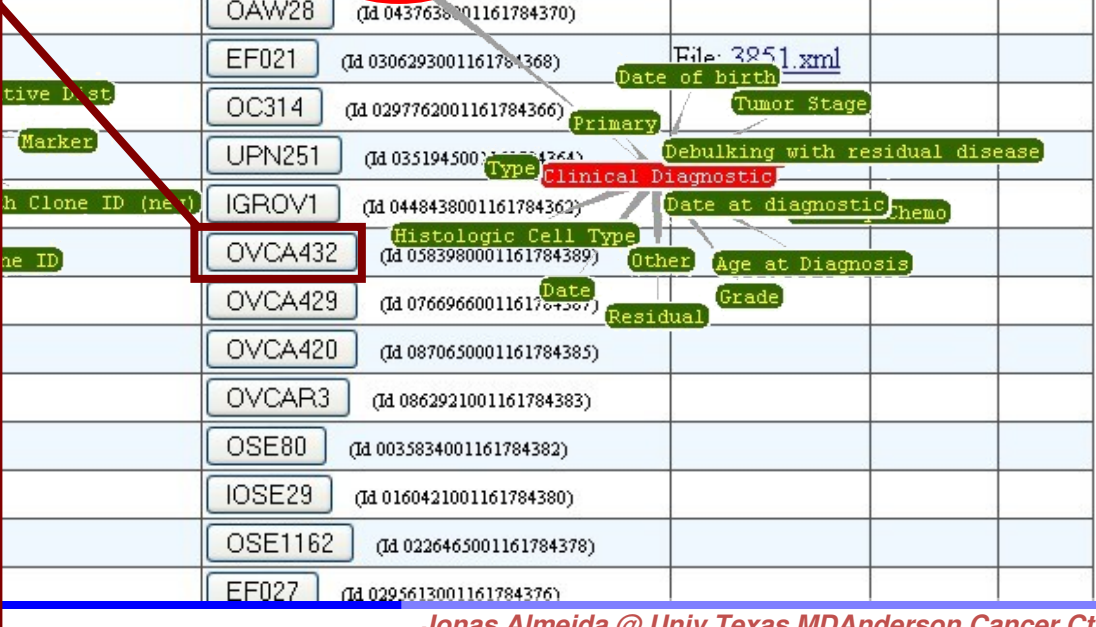


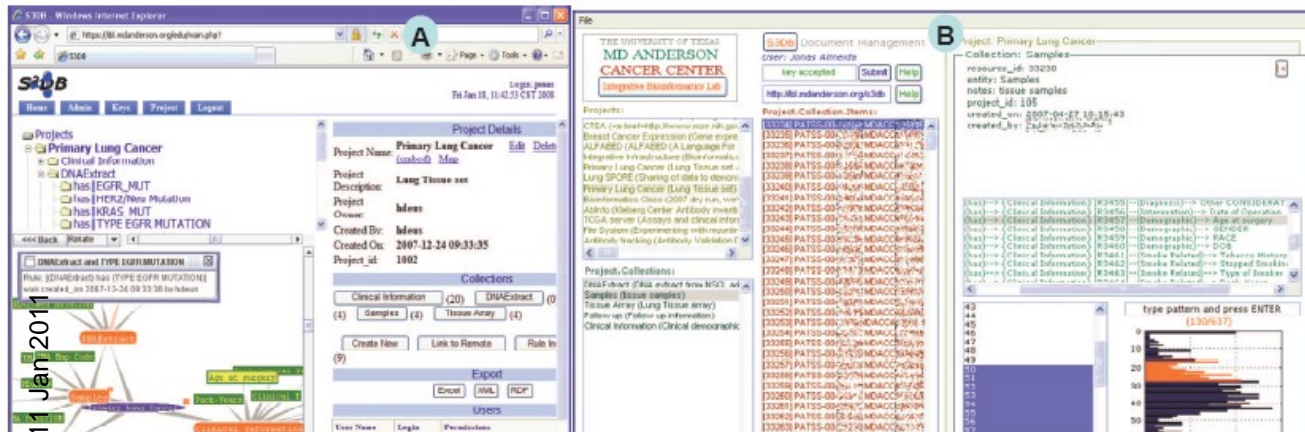
http://localhost - Mozilla Firefox

[Insert All Page] [Refresh]

ID	Resource	Created On	Created By	Notes
0583980001161784389	Protein Array	2006-10-25 08:53:09	oresearcher	OVCA432

- Antibody | 4EBP1
 - 4.49
- Antibody | 4EBP1p
 - 2.9
- Antibody | Akt
 - 3.9





Snapshots of interfaces using S3DB's API (Application Programming Interface). These applications exemplify why the semantic web designs can be particularly effective at enabling generic tools to assist users in exploring data documenting very specific and very complex relationships. Snapshot **A** was taken from S3DB's web interface, which is included in the downloadable package. This interface was developed to assist in managing the database model and, therefore, is centered on the visualization and manipulation of the domain of discourse, its Collections of Items and Rules

Nature Precedings | doi: 10.1038/npre.2011.5537.1 | Posted 11 Jan 2011

OPEN ACCESS Freely available online **PLoS ONE. Aug 13;3(8):e2946**



A Semantic Web Management Model for Integrative Biomedical Informatics

Helena F. Deus^{1,2}, Romesh Stanislaus¹, Diogo F. Veiga¹, Carmen Behrens³, Ignacio I. Wistuba^{3,4}, John D. Minna⁵, Harold R. Garner^{5,6,7,8,9}, Stephen G. Swisher¹⁰, Jack A. Roth¹⁰, Arlene M. Correa¹⁰, Bradley Broom¹, Kevin Coombes¹, Allen Chang¹, Lynn H. Vogel^{1,11}, Jonas S. Almeida^{1*}

1 Department of Bioinformatics and Computational Biology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States of America, 2 Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Lisboa, Portugal, 3 Department of Thoracic/Head and Neck Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States of America, 4 Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States of America, 5 Hamon Center for Therapeutic Oncology Research, Simmons Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, 6 Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, 7 Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, 8 Center for Biomedical Inventions, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, 9 Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, 10 Department of Thoracic and Cardiovascular Surgery, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States of America, 11 Department of Biomedical Informatics, Columbia University, New York, New York, United States of America

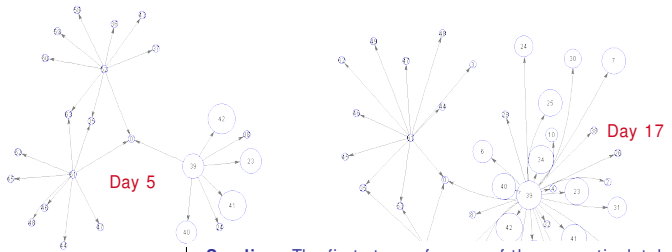
Abstract

Background: Data, data everywhere. The diversity and magnitude of the data generated in the Life Sciences defies automated articulation among complementary efforts. The additional need in this field for managing property and access permissions compounds the difficulty very significantly. This is particularly the case when the integration involves multiple domains and disciplines, even more so when it includes clinical and high throughput molecular data.

Methodology/Principal Findings: The emergence of Semantic Web technologies brings the promise of meaningful

Ontology-centric web client

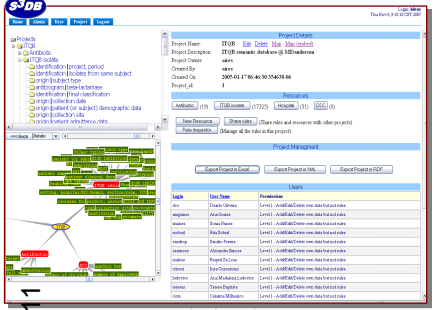
S3DB is equipped with REST application programming interface (API), that is, client applications can be easily weaved by composing URL calls with variable values.



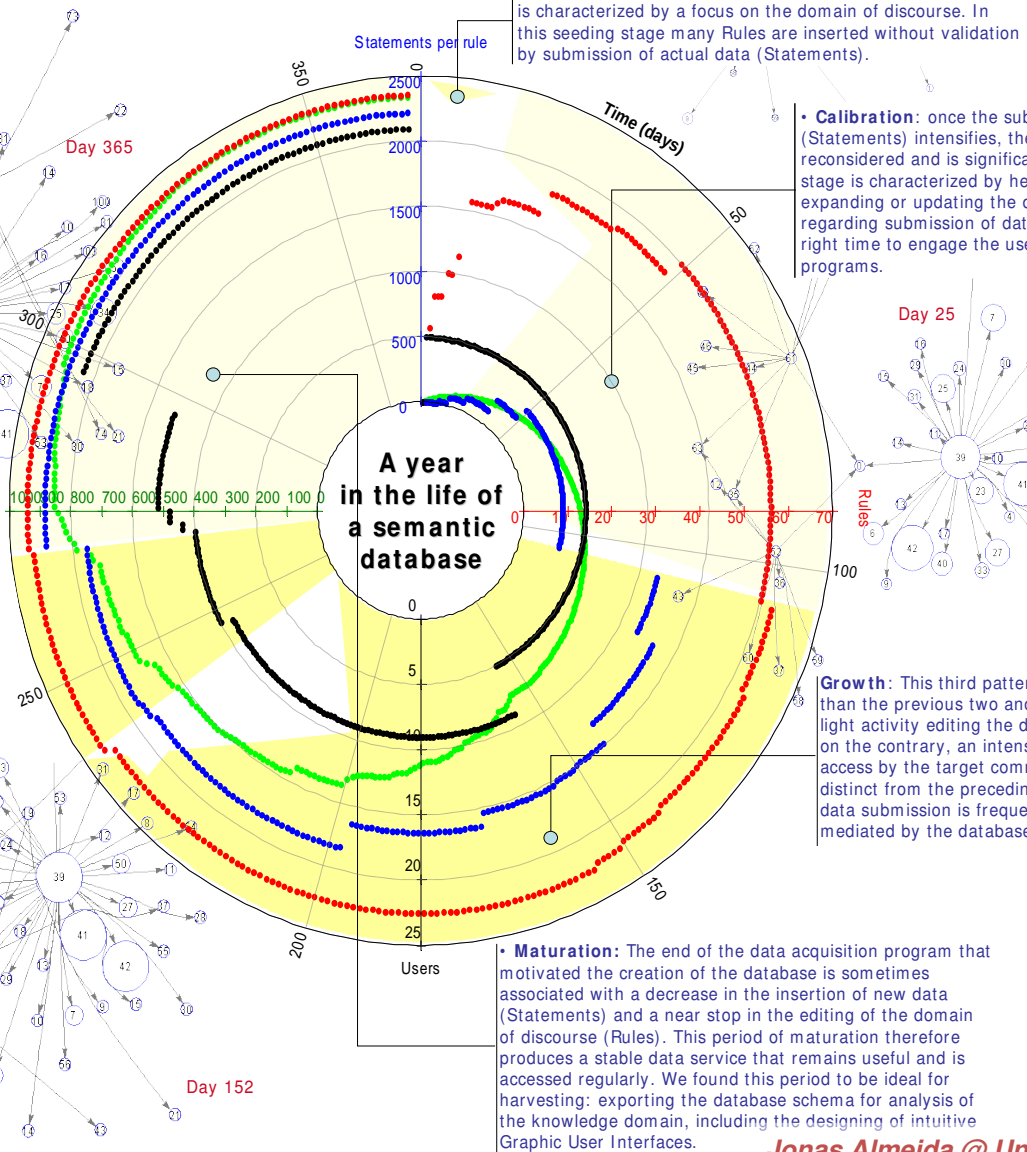
Document-centric clients

...and client side applications can be easily developed, relying only on the REST protocol to interoperate with the S3DB DBMS service.

S3DB is being used for a variety of molecular epidemiology domains, for example, for Cancer Research:



- Entity
- Spa typing
- Doubling time
- monthly fee
- RPO
- collection site
- patient admittance data
- patient (or subject) demographic data
- MLST
- Clal-meca:Tn554
- PEGE
- disk inhibition
- project, station
- leukodisks
- hemolysates
- other
- Ribotyping
- Phage typing
- Small hybridization bands
- Antibiotic
- activation class
- full name
- subject type
- disk inhibition
- collection date
- project, peptid
- setting, hospital/DCC/heard, service/room, ICU
- MIC
- isolates from same subject
- beta-lactamase
- susceptibility
- Agr
- PCR genes amplification
- country, state/province/county, city
- name
- 3-4 letter code
- alternative name
- MIC
- ITQB isolate
- susceptibility
- isolate reference
- 80oC
- country, state/province/county, city
- country, city
- number of rooms
- number of employees
- outdoor area
- indoor area
- code
- species and tests
- final classification
- Hospital
- patient clinical data
- LN2 freezing
- Dot-blot
- Rep-PCR
- 57 SCCmec typing
- 58 category
- 59 speciality
- 60 bed size
- 61 DCC
- 62 number of children
- 63 name
- 63 target
- 67 mechanism and genes
- 81 Plasmid analysis
- 95 MRSA frequency
- 96 MRSE frequency
- 97 antibiotic consumption
- 100 institution
- 101 LN2 viability test
- 102 ClB
- 103 enterotoxins
- 104 exfoliatins

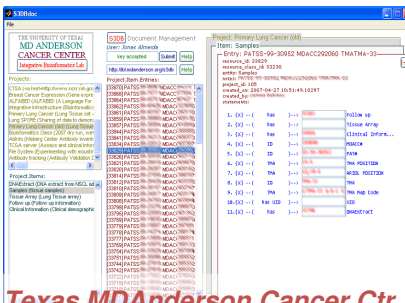
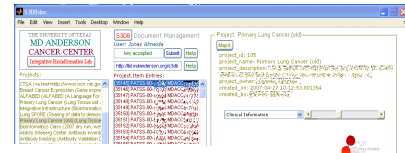


Seeding: The first stage of usage of the semantic database is characterized by a focus on the domain of discourse. In this seeding stage many Rules are inserted without validation by submission of actual data (Statements).

Calibration: once the submission of data triples (Statements) intensifies, the seed data model is reconsidered and is significantly edited. This second stage is characterized by heavy activity both regarding expanding or updating the domain of discourse and also regarding submission of data. We found this to be the right time to engage the user community with training programs.

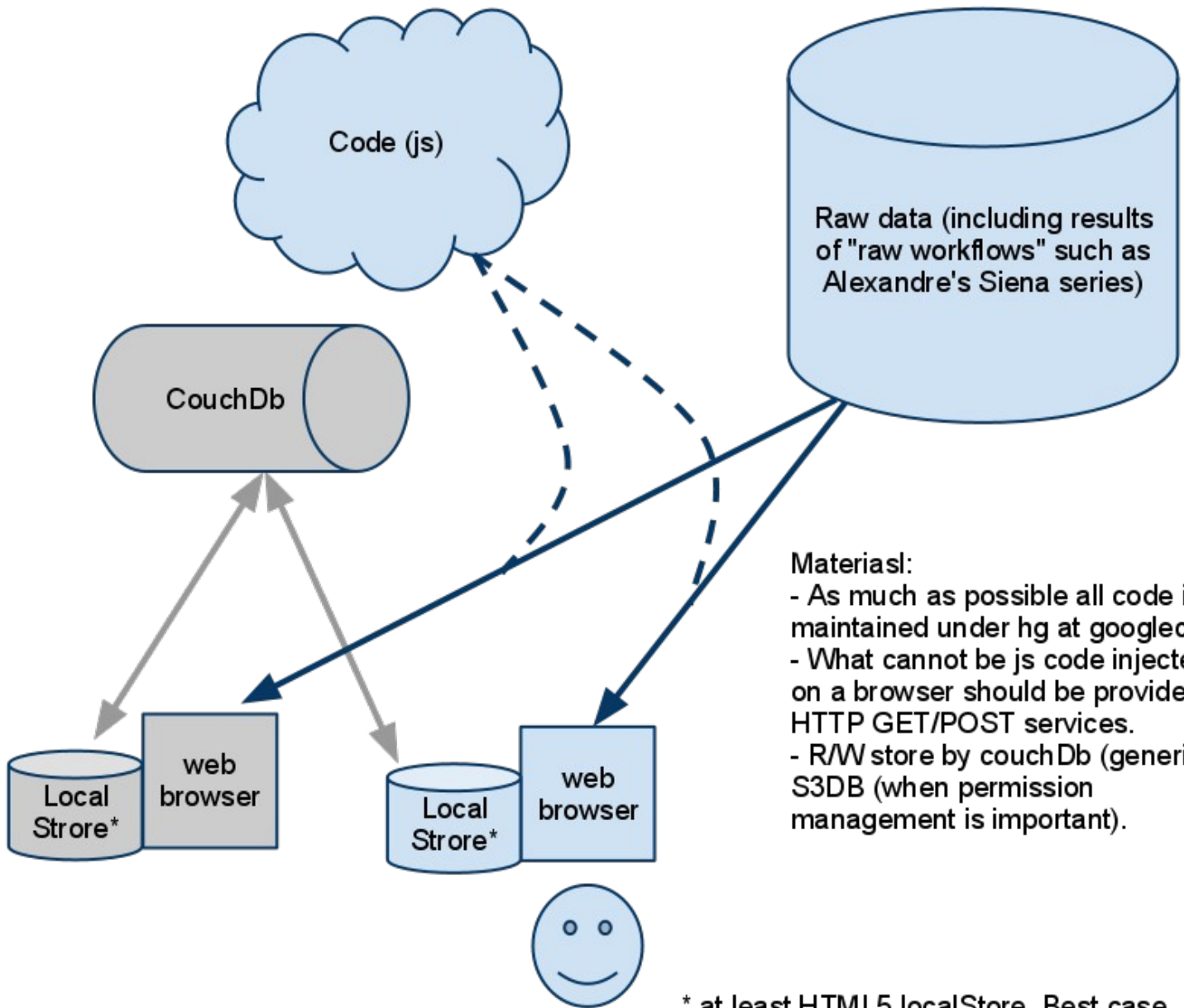
Growth: This third pattern of usage is much longer than the previous two and corresponds to a relative light activity editing the domain of discourse while, on the contrary, an intensification of the database access by the target community of users. This is distinct from the preceding Calibration state where data submission is frequently aided or even mediated by the database developers.

Maturation: The end of the data acquisition program that motivated the creation of the database is sometimes associated with a decrease in the insertion of new data (Statements) and a near stop in the editing of the domain of discourse (Rules). This period of maturation therefore produces a stable data service that remains useful and is accessed regularly. We found this period to be ideal for harvesting: exporting the database schema for analysis of the knowledge domain, including the designing of intuitive Graphic User Interfaces.



<http://cnviewer.googlecode.com>

<http://link.inesc-id.pt/pneumopath>



- Materials:
- As much as possible all code in js, maintained under hg at googlecode.
 - What cannot be js code injected on a browser should be provided as HTTP GET/POST services.
 - R/W store by couchDb (generic) or S3DB (when permission management is important).

* at least HTML5 localStore. Best case scenario would be to have a local deployment of couchDb.

Conclusions

1.KOS: Domain neutral ontologies are particularly conducive to variable discovery.

2.Cloud: If the real-world domain expert is part of the exercise then the OS is the browser and the “command line” is its console.

[MDACC Stat → UAB Path]