Implementing RNA_DENOVO in

R.O.S.I.E

Rosetta Online Server that Includes Everyone

Sergey Lyskov, GrayLab
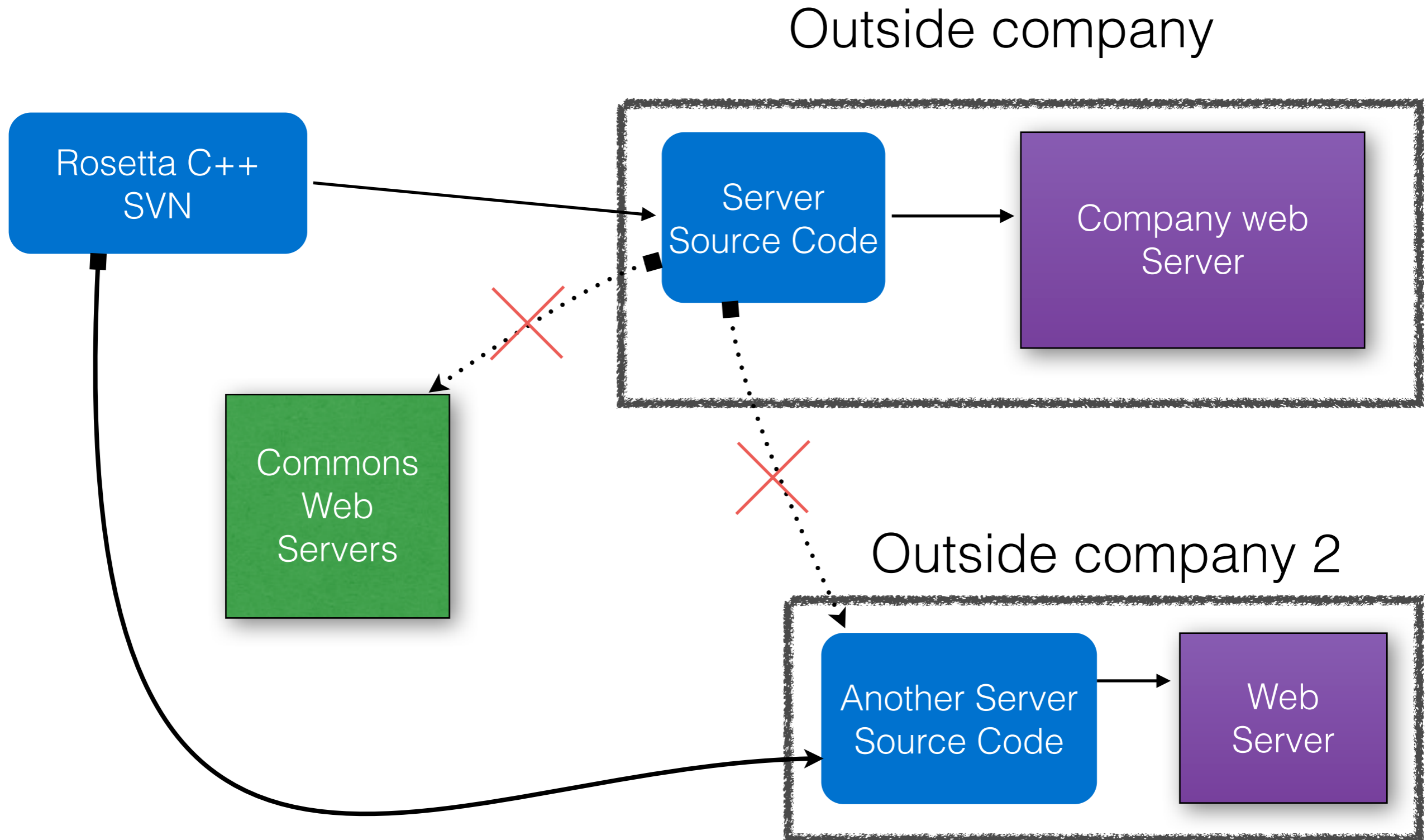
http://rosie.rosettacommons.org

Talk overview

- Importance and benefits of having a central hub for Rosetta servers.

- How much it took to implement RNA_Denovo (our test case)

- New licensing approach when dealing with startups that want to develop commercial servers

- Discussion

Current state of server affairs and problems that we encounter

Typical lifecycle of a Rosetta web server:

- New student joins the lab

- Implements a beautiful web server in very advanced but obscure framework that no one but him could understand

- Graduates!

- Things work until next major OS update/Database/disk crash... and then no one knows how to fix it => server is quietly retired

- New student joins the lab... AND STARTS FROM SCRATCH

- <u>Bottom line</u>: 3-5 years of work get lost from humanity knowledge pool forever...

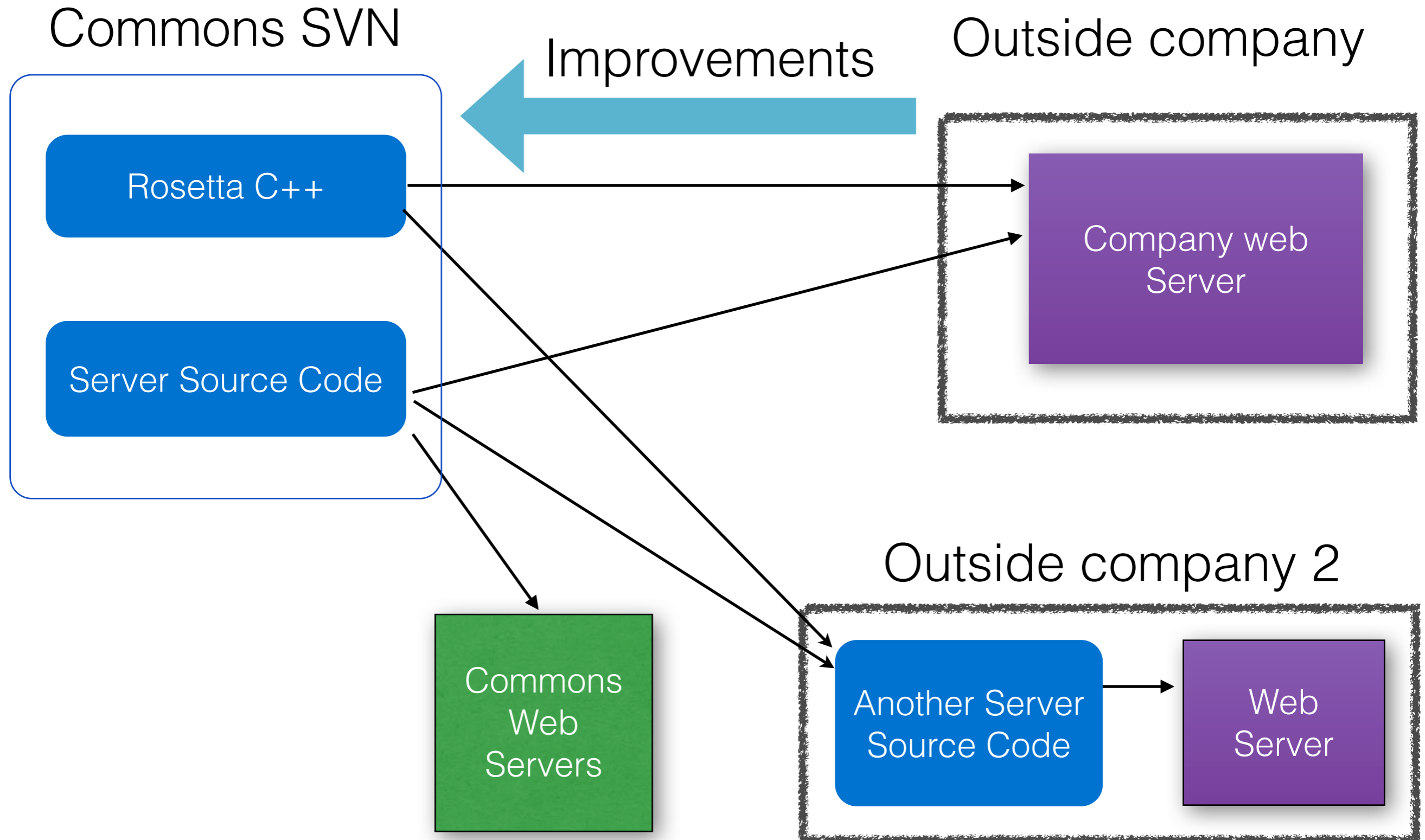# Our previous server licensing strategy created conflict of interest situation:

Outside company

Rosetta C++ SVN

Server Source Code

Company web Server

Commons Web Servers

Outside company 2

Another Server Source Code

Web Server

So what we need is to create
a **central pool** where code/
knowledge will be **accumulating** and
which will allow us to build on
previous work

This will allow up to keep source code well-
maintained and up to date with the latest
web frameworks

And new strategy allows us to avoid conflict of interest

Commons SVN

Outside company

Improvements

Rosetta C++

Server Source Code

Company web Server

Outside company 2

Commons Web Servers

Another Server Source Code

Web Server

R.O.S.I.E

Rosetta Online Server that Includes Everyone

- So it's already in SVN `trunk/RosettaServer`

- All basic functionality like job queue, user management and etc. is implemented and protocols developers do not need to reinvent the wheel

- BackEnd is written in a way that allows it to be de-coupled from particular HPC system and allow later to be adapted to any particular job manager or cloud computing.

- And we implemented two protocols Docking and RNA_Denovo to test the grounds

RNA Denovo implementation

http://rosettaserver.graylab.jhu.edu/rna_denovo/viewjob/274

- BackEnd (computational part): 2 files less then 10kB of code in Python

- FrontEnd (WebServer part): 5 files, 10kB of code and 20kB of html like templates.

- As we develop our codebase time for simple protocols implementation could be weeks or even less!

Advantages

- Central pool for all server projects that will share code and improvements.

- Since Commons owns the source code there will be no legal obstacles to deploy one or more academic servers and rebrand them as we see fit, and also sell it.

- Because academic servers and commercial ones are running the same code there will be no question of additional validity of results for com. servers.

- No conflict of interest between Commons and startups!

- Clear separation of responsibilities. With this approach, Commons could concentrate on implementing scientific parts of the servers right and worry less of implementation of payment systems, security and other things with which we have no expertise.

- Less code to write for outside companies. It will take much less effort to create startup and offer some of Rosetta web servers, so most of our grad students will be able to do it!
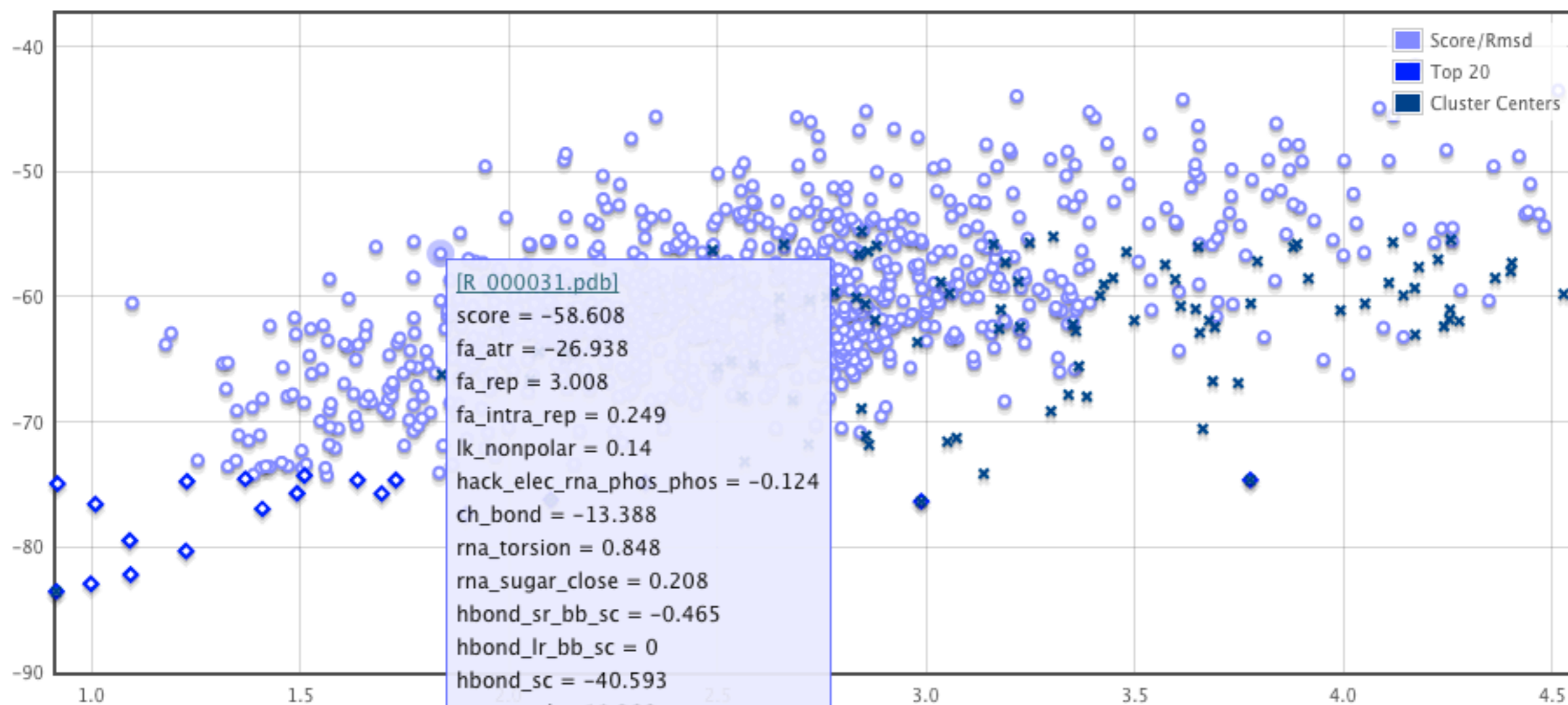
Wrapping up:

- Docking and RNA_Denovo modes now implemented.

- Add Sequence Tolerance protocol (collaboration with Tanja Kortemme and Shane O'Connor, in progress),

- Erraser protocols (collaboration with Rhiju Das and Fang-Chieh Chou, in progress)

- Antibody protocol as soon as it is ported to Rosetta3 code base.

- We are beta testing Server VirtualBox image that has all env. installed (ask me for URL, user name and password if you want to check it out).

Our future goals:

- More protocols by next RosettaCon (10?)

- Get more developers from different labs involved (streamline the VM approach to simplify things)

<Rhiju part>

Discussion

Score/RMSD

-40
-50
-60
-70
-80
-90

1.0   1.5   2.0   2.5   3.0   3.5   4.0   4.5

Score/Rmsd
Top 20
Cluster Centers

[R_000031.pdb]
score = -58.608
fa_atr = -26.938
fa_rep = 3.008
fa_intra_rep = 0.249
lk_nonpolar = 0.14
hack_elec_rna_phos_phos = -0.124
ch_bond = -13.388
rna_torsion = 0.848
rna_sugar_close = 0.208
hbond_sr_bb_sc = -0.465
hbond_lr_bb_sc = 0
hbond_sc = -40.593
geom_sol = 18.366
atom_pair_constraint = 0.082
rms = 2.64
rms_stem = 1.268
N_WC = 2
N_NWC = 3
N_BS = 10
f_natWC = 1
f_natNWC = 0.2
f_natBP = 0.429

The only 5000 lowest ener...   ...ph points to see exact score and download particular result file.

Score data [Download original score fil...

decoy	score	fa_atr	fa_rep	fa_intra_rep	lk_nonpolar	hack_elec_rn;	ch_bond	rna_torsion	rna_sugar_clc	hbond_sr_b
R_000567	-83.543	-30.501		0.259	0.878	0.039	-16.406	0.599	0.474	-0.056
R_000631	-82.924	-30.455		0.25	0.954	0.107	-16.277	0.423	0.679	-0.175
R_000395	-82.2	-31.388		0.258	0.974	-0.053	-15.742	0.544	0.666	0
R_000824	-80.313	-31.005	2.971	0.256	1.123	0.2	-17.297	0.346	0.887	-0.077
R_000780	-79.472	-31.134	3.526	0.261	0.994	0.128	-15.928	0.637	0.887	-0.112
R_000179	-77.367	-28.665	3.159	0.28	0.774	-0.047	-15.215	0.627	0.579	-0.227
R_000571	-76.926	-29.306	2.696	0.247	0.871	0.016	-16.087	0.422	0.723	-0.1
R_000164	-76.559	-31.046	4.81	0.255	0.911	0.139	-16.568	0.536	0.519	-0.155