
Bioinformatics Experimentation in a New Agent-Based Infrastructure OpenKnowledge

(poster coming soon)

Dietlind L. Gerloff*, Xueping Quan, Adrian Perreau de Pinninck,
Paolo Besana, Siu-wai Leung, Marco Schorlemmer, Dave Robertson

*Dept of Biomolecular Engineering, UC Santa Cruz
School of Informatics, University of Edinburgh
Artificial Intelligence Research Institute, IIA-CSIC

Context of the Work

OpenKnowledge (EU FP6):

A Peer-to-Peer architecture emphasizing
the sharing of knowledge via interaction models

- 3y Targeted Research Project in Computer Science
- 2006-8; currently smaller follow-on projects since
- Testbeds: Bioinformatics + Emergency Response

www.openk.org

OK kernel V1.1 (requires Java 1.5; JRE6)

+ more info, examples

+ demo of “omicslab” scenario

OpenKnowledge - Bioinformatics

Robertson* / Gerloff / Tate / Bundy	University of Edinburgh
Schorlemmer / Abían / Sierra / Augusti	CSIC Barcelona
van Hermelen	Vrije Universiteit Amsterdam
Dasmahaptra / Lewis Shadbolt / Berners Lee	School of Eng & CS Southampton
Motta	Open University
Giunchiglia / Marchese / Bonifacio / Riccardi	University of Trento

*

*

Interest in P2P

- sharing (without depositing)
- ease of joining (“subscribing”) to a task group
- scalability
- no (or less) need for curation and maintenance


P2P in bioinformatics:

- Chinook
- SEED

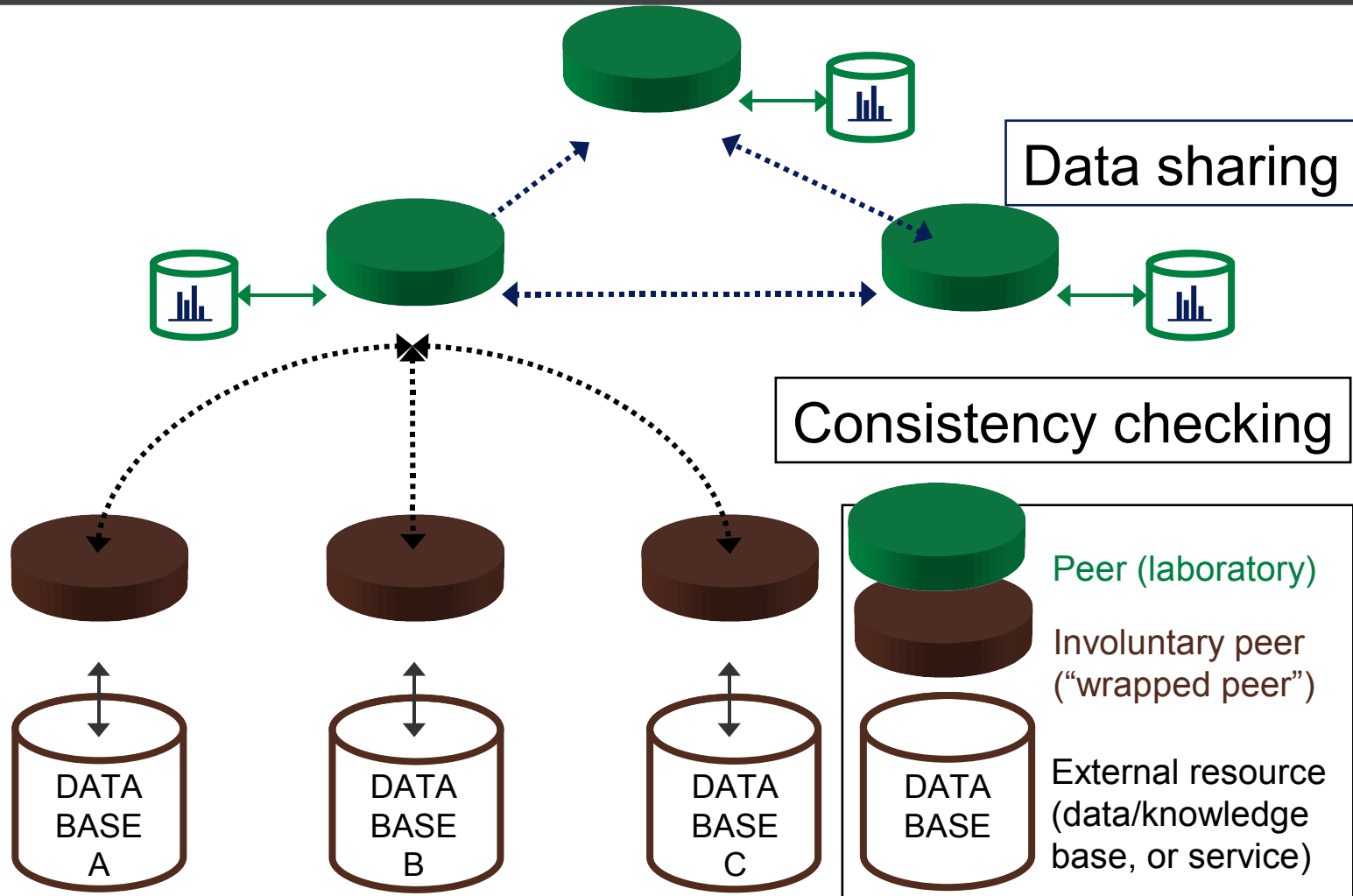
Other related work: Workflow Platforms

- Kepler
- TAVERNA (myExperiment)

OK - Design Philosophy & Goals

- peer-to-peer  as open as possible
- not (as) domain-oriented
- peers groups include non-expert users
(expert / some knowledge / “dummy” users)
- emphasis on sharing knowledge/data of any kind
(less on the distribution of computational load)
- supporting language:
LCC (Lightweight Co-ordination Calculus):
message passing (more details on poster)

Applied Bioinformatics/Biology at High Level: Consistency checking + Data sharing



An Interaction in OK in a Nutshell

1. Peer E (Experimenter) sets a task (asks a question)
2. Discovery Service (DS) looks for peers of potential interest, and suggests interaction partners and interaction models (IM) (if available)
3. The IM specifies roles to be played, and peers subscribe (by agreeing) to interact + in which role(s)
4. One peer is (computational) coordinator
5. Roles are executed via OK-Components (OKC)

Notes:

- many possible IMs for carrying out the same task
- OK is at “infancy” stage! For example, DS in OK currently relies on information provided by peers

Experiment A - yeast protein structure models

Extend structural knowledge through modelling:
Find fragments of 3D-models of *S.cerevisiae* (yeast) proteins that can be trusted

- 6604 yeast protein sequences (some predicted)
- currently only 330 known 3D-structures (in PDB)

Consensus-building is a popular strategy in
Applied Structural Bioinformatics
(typically via meta-WWW-servers)

- Compare modelled structures from three databases: SWISSMODEL (769 models); SAM (2211 models); ModBase (2546 models) in OpenKnowledge
- MaxSub program to identify common substructures

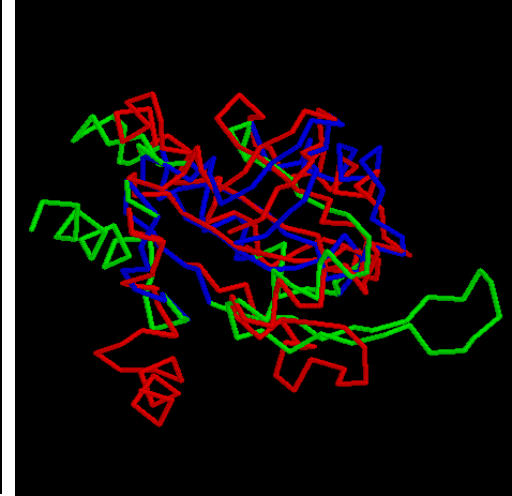
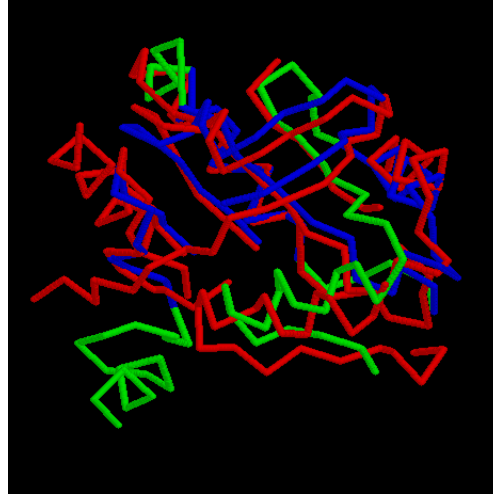
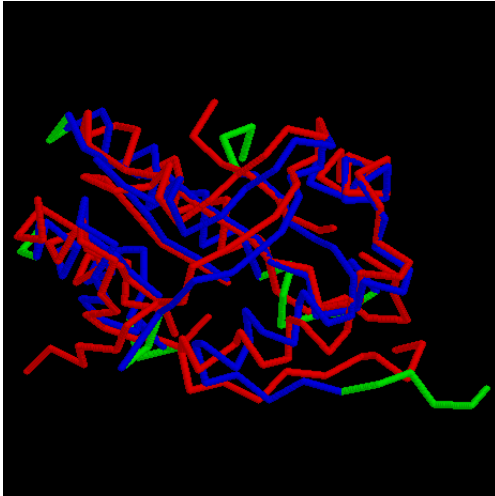
Results

Example: (MaxSub pair-wise common substructure in blue)

YBR024W: SWISS-SAM

ModBase-SAM

SWISS-ModBase



MScore: 0.660

0.459

0.400

CYSP = Comparison of Yeast 3D Structure Predictions

578 three-way supported

MaxSub-substructures > 45 aa

from 545 proteins

(Linked from www.openk.org)

Experiment C - *de novo* peptide sequencing

Look into one another's "trash"

Increase confidence in peptide assignments to MS/MS spectra in proteomics

- one (of several) uses of MS to study protein content
- especially difficult: non-model organisms; mixtures

Data sharing has become a bottleneck in modern biology

- Implement demo example guided by user input
- OK-emulation with 9 peers: 7 proteomics labs + self + NCBI database
- Test data: 38 peptides from ABRF 2006 (contest data; 35% of 78 labs identified >40 or 48 proteins in 2006)

Contact/input from an active community



ProteoRed

instituto nacional de proteómica

www.proteored.org

OK-contact: Joaquín Abián

Jun 11st, 2009

Home

Objectives

Structure

Working Groups

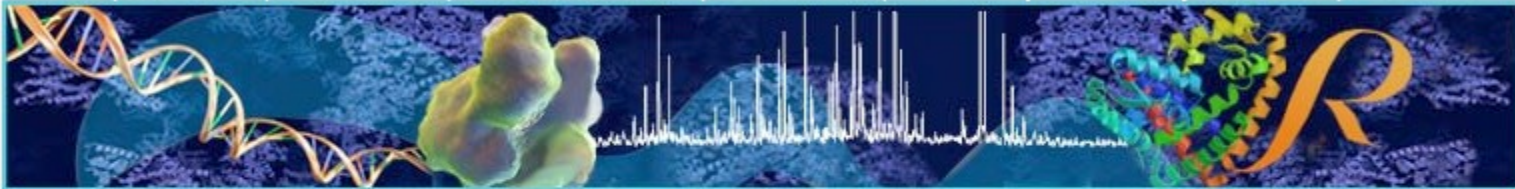
Services

News

Events

Tools

Contact



To ask for a service, or to use the MIAPE generator tool please log on

Login

■ User

guest

■ Password

accept >

To create an account or remember your password or user name, please click [here](#)

Password for guest user: 'guest'

Related Links

HUPO

www.hupo.org

EuPA

www.eupa.org

ABRF

ProteoRed Purpose

The main objective of **ProteoRed** consortium is to coordinate and integrate activities of the Spanish proteomic facilities and services in a common network with a nodal structure, so they will support the whole development of the proteomic research in our country. For this purpose, **ProteoRed** has been constituted as a technological platform composed by six nodes, supporting 17 + 2 proteomic facilities placed all over Spain.

In addition, our purpose include:

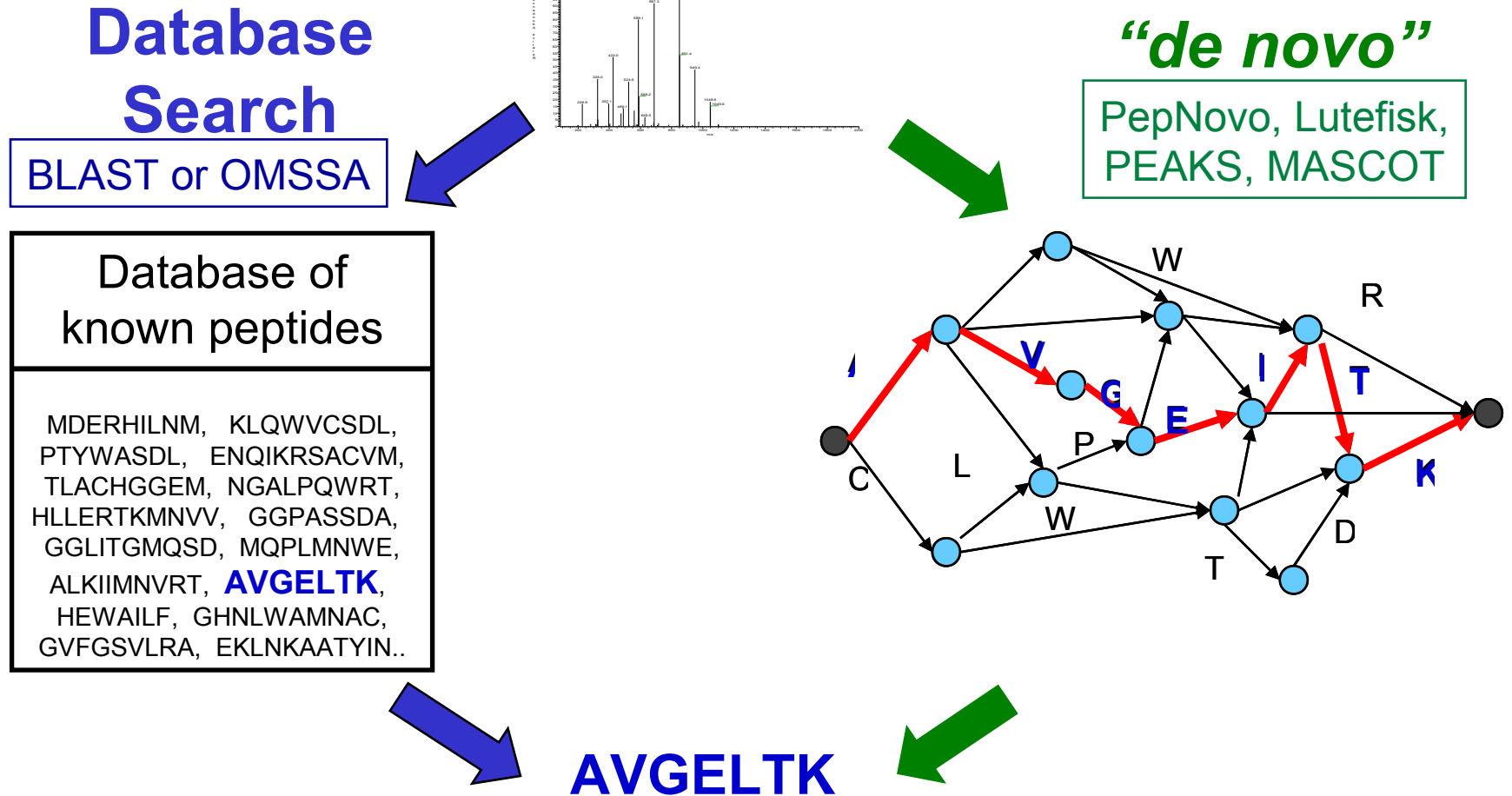
Assure that **ProteoRed** structure will cover the proteomics needs in all Spanish regions considering the differences in research density (more proteomic facilities at Madrid and Barcelona), geographical proximity, economic resources of the customers, and the technological capabilities required by the researchers.

Establish well coordinated and structured protocols for deciding the strategy that will be followed by each node and ProteoRed as a whole for providing proteomics services in Spain. The final objective is to increase the quality and the number of proteomics services provided and the number of projects where those proteomics facilities will participate in next years.

Offer services in all the stages involved in the protein analysis process: protein fractionation, separation and purification of peptides and proteins, Protein and peptide molecular weight analysis, protein identification, characterization, sequencing and diferential proteomics

Improve the capabilities and competitiveness of the Spanish proteomics facilities with the support of the seven working groups created by **ProteoRed** technological platform and coordinated by the

Experiment C: de novo peptide sequencing

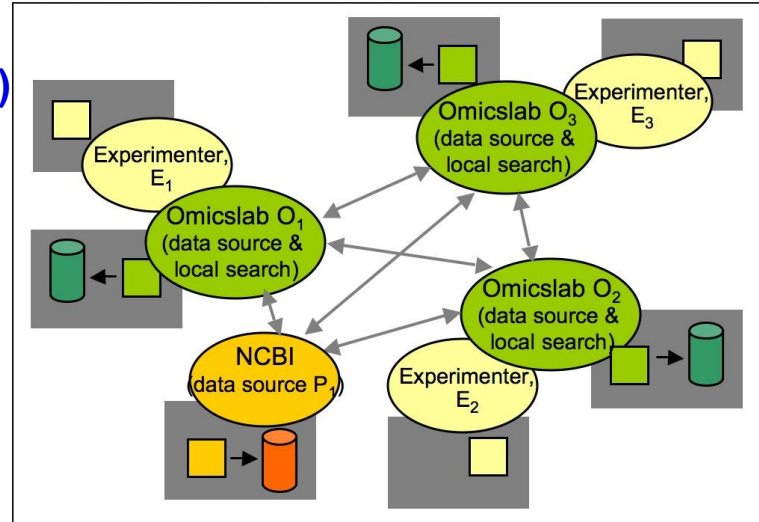


Results

Implementation notes:

- search is executed locally using **BLAST (sequence)** or **OMSSA (spectrum)**
- visualisation: **OK-omics prospector**
- database peer is important for revealing contaminants + holds promise for consistency checks with other data (species, annotation, etc)

3Labs and 1Database:



Successful integration of data sharing
and consistency checking

We found increased correct identification rate compared to
random chance, though not yet significant

Encouragement to run a larger sample

(Demo available at www.openk.org)

Conclusions & Future Work

- Three implementation examples served primarily to:
 - as proof of principle (can be done) & demos
 - to advance OK-development
 - to illustrate complementarity to other infrastructures⇒ benefit from each other's experiences
- Particularly exciting:
 - data sharing example
 - work with data other than sequences (i.e. not strings)
- UCSC group's current/future interests:
 - spur on development and user-base through more integrated examples
 - investigate benefits of raw data sharing (J.Magasin)