

Abstraction based cooperation for the design of bioinformatics workflows

Frédéric Cadier, Philippe Picouet

Telecom Bretagne / Institut Télécom – Brest – France

Lab STICC UMR CNRS 3254

European University of Brittany – France

Capitalization of Experimental Knowledge in Bioinformatics (CapExBio)

Station Biologique de Roscoff, OUEST-genopole®, funded by Brittany Council

Roadmap

1. Introduction
2. Communities and boundary object : principles
3. In silico experiment design
4. Towards a workflow based groupware
5. Abstraction to design boundary objects
6. Service abstraction
7. Data format abstraction
8. Conclusion and perspectives

Community and boundary objects - principles

- Activity theory basis [Engestrom 1987] and Sociology approach
 - activity's artefact as a boundary object [Star & Griesemer, 1989]
 - different communities of practices [Wenger, 1998]
 - one activity, one community of interest [Fischer & Ostwald, 2005]

<i>Dimensions</i>	<i>CoPs</i>	<i>CoIs</i>
Nature of problems	Different tasks in the same domain	Common task across multiple domains
Knowledge development	Refinement of one knowledge system; new ideas coming from within the practice	Synthesis and mutual learning through the integration of multiple knowledge systems
Major objectives	Codified knowledge, domain coverage	Shared understanding, making all voices heard
Weaknesses	Group-think	Lack of a shared understanding
Strengths	Shared ontologies	Social creativity; diversity; making all voices heard
People	Beginners and experts; apprentices and masters	Stakeholders (owners of problems) from different domains
Learning	Legitimate peripheral participation	Informed participation

In silico experiment design

- Qualitative study of practices of in-silico experiment design
 - Observing small communities (5 projects - 8 people):
 - SB Roscoff : Researchers + Genomics and informatics service
 - Resources : OUEST-genopole - GenOUEST platform + external contributors (program providers)
 - Observations of work sessions (context and objectives, way to proceed, results analysis)
 - Analysis of exchange means (emails, face to face, data (screen, paper))
 - Individual or group interviews
- Results :
 - Workflow-based activity (shared biological background)
 - the activity is the design of the workflow
 - the workflow is the main activity's artefact
 - 4 communities of practices (know-how & knowledge)
 - 1 community of Interest

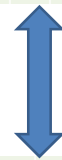
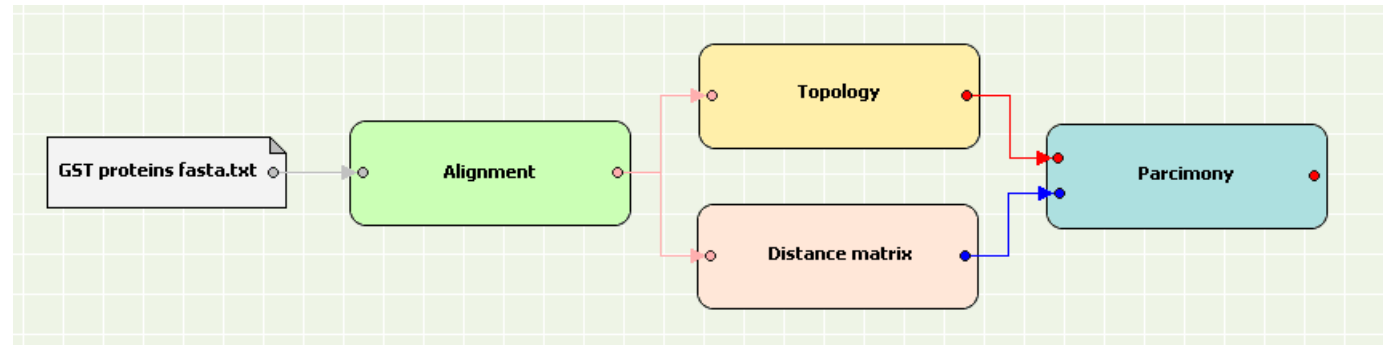
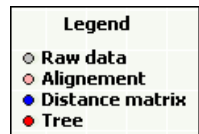
Towards a workflow based groupware

- Distributed community of interest
 - the actors join in the activity to solve the design problem
 - spatial and temporal distributions
 - social dimension
- Need for a groupware
 - informed participation [Fischer & Ostwald, 2005]
 - workspace awareness [Gutwin & Greenberg, 2002]
 - activity reporting : *who, what, where, how, when*
- How to make workflow a valid boundary object ? [Star & Griesemet, 1989]
 - **plasticity** to provide different viewpoints
 - **robustness** to maintain identity across viewpoints

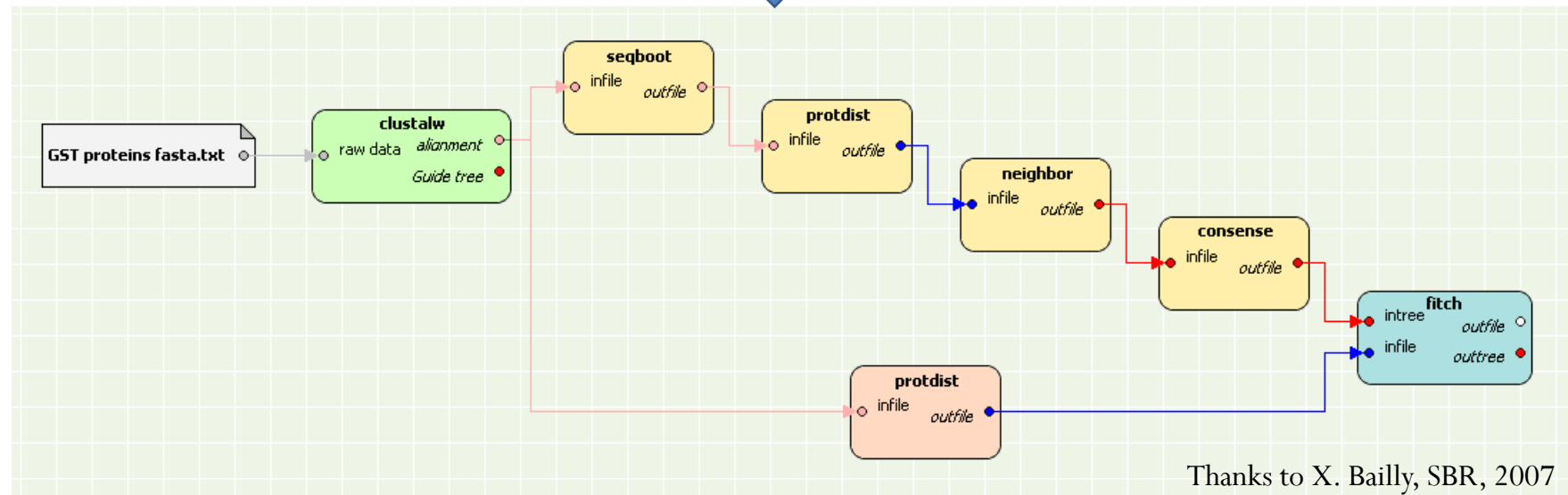
Abstraction to design boundary objects!

- Abstraction provides ability to provide multiple views of a single object, thus providing :
 - Plasticity
 - hiding or showing some information according to user preferences (data format, program versions, availability)
 - Robustness
 - abstraction allows to always point a unique object regardless of the multiple points of view
- We claim that we need richer resource description models to provide workflow abstraction !

User interface : service abstraction



This is the same object, with two representations

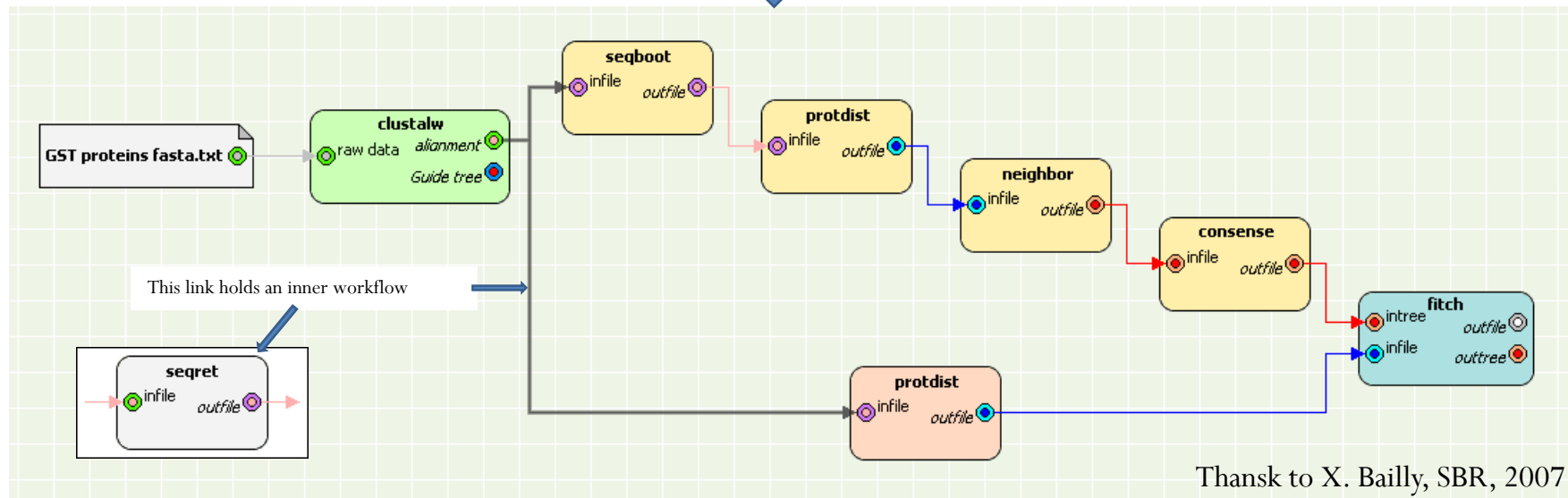
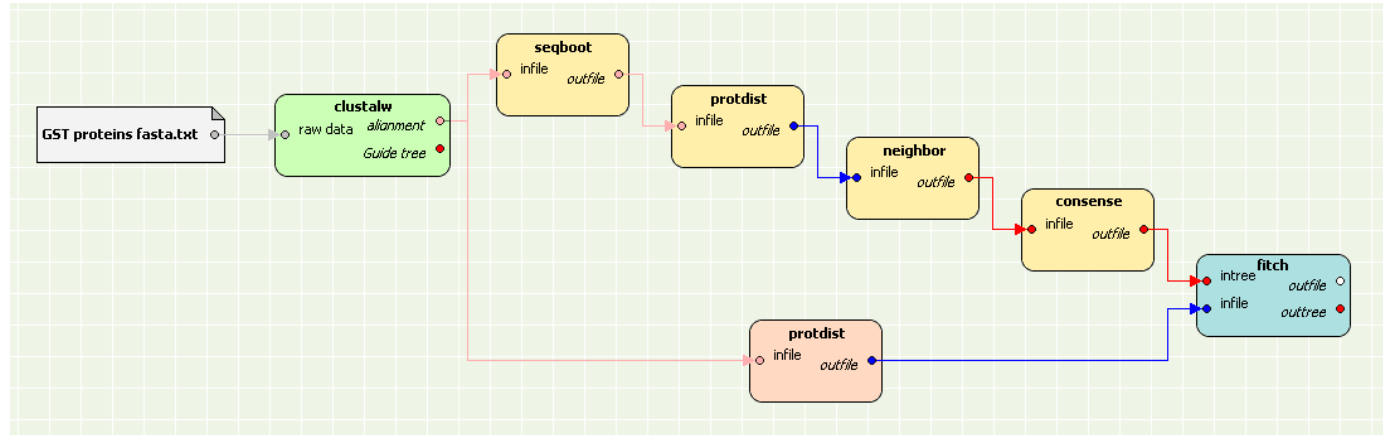
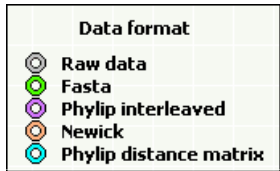
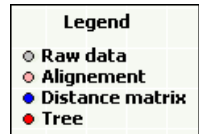


Thanks to X. Bailly, SBR, 2007

Service abstraction

- Services Ontologies « only » provide way to find a program
- We propose to define hierarchies of abstract services
 - Useful to hide program versions (clustal 1.8, clustal 2.0)
 - Useful to group different alternative service with the same function (clustal, muscle)
 - Useful to identify a service-based workflow
- Unifying services and workflows
 - Connecting inputs/outputs of the service to programs I/O
 - Default or unknown values otherwise

User interface : Format abstraction



Thank to X. Bailly, SBR, 2007

Data format abstraction

- Providing format independence is a hard issue
 - Biomoby provides its own XML translation of any format and ensures translation through its type ontology.
 - Some programs offer alternative formats; others offer cumulative formats
- We propose a triple data representation (data, model, format)
 - The couple (data, model) represent the « business » data (i.e. the data viewed by the biologist : ADN alignment).
 - Please note that both data and model may be part of hierarchies
 - Several formats may be associated to a « business » data

Conclusion and perspectives

- Workflow is a valid boundary object for in-silico experiment design activity
- To support workflow-based groupware, we propose to enrich the resource description model:
 - Service ontology can be changed in a service abstraction hierarchy
 - Types ontology have to be split in 3 tuples ontologies (data, model, format) to provide different compatibility views of the workflow
- Capitalization of Experimental Knowledge in Bioinformatics (CapExBio)
 - Prototype is due to end of October 2009 (probably free licence 2010)
- Another boundary object : resource description !