



Project no. 043338

Project acronym: EMERGENCE

Project title: A foundation for Synthetic Biology in Europe

Instrument: NEST Pathfinder

Thematic Priority: Synthetic Biology

**D1.2: Report on the first workshop on development of the European IT infrastructure for synthetic biology**

Due date of deliverable: February 2007

Actual submission date: July 2007

Start date of project: 1.12.2006

Duration: 36 months

**Alfonso Valencia**

Structural and Computational Biology Programme Spanish National Cancer Research Centre (CNIO) Melchor Fernandez Almagro, 3. E-28029 Madrid

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
<b>PU</b>	Public	x
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

## **GROUP WORK of THE integration of Bioinformatic /GENOMICS information with MIT Repository of Biological Parts.**

Initial phase dedicated to understand the organization and structure of the MIT 'Registry of Standardized Biological Parts' (<http://partsregistry.org/>). Basic activity around the methods and protocols necessary to make possible the navigation from the MIT repository into the available genomic and biochemical information and symmetrically to represent the experimental information available in the repository in the main genomics representation tools and systems.

Important monitoring the activity of the active group of scientist working in related issues (<http://partsregistry.org/>), foresee changes in the repository, apparition of other likely federated repositories associated to specific labs, and the emergence of protocols for the collaboration and information exchange between those federated repositories (the EMERGENCE action is also participating in this discussion with the presence of the group of Luis Serrano (see the BioBrick Standards Mailing List [http://biobricks.org/pipermail/standards\\_biobricks.org/](http://biobricks.org/pipermail/standards_biobricks.org/)))

The work has proceeded in collaboration between the CNIO and MIT groups, particularly Randy Rettberg (MIT, main developer of the MIT registry) and the CNIO group, including the three days hands-on workshop in Madrid (5 and 6/9/2007) and the participation of Ildelfonso Cases (CNIO) in the MIT workshop for MIT registry developers workshop 5-6/11/2007.

### ***1. Analysis of the limitations and future most needed developments in the repository and by extension to be incorporated in any other associated repository of parts.***

Quality Checks: Confirm user provided annotation (CDS, promoters, etc), Check redundancy and subparts, Presence of non-allowed restriction sites, Predict expression/cloning/amplification problems

Cross-Linking with other databases: Match part sequence to external databases such as Genebank/EMBLBank, Uniprot, Genome Databases, etc.

Use of Ontologies and Control Vocabularies: Normalize vocabulary for elements, as promoters, etc, using standard ontologies (i.e. Sequence Ontology)

Add/Link to Additional Data: Variations, Homologues, Properties (other substrates, inhibitors, Km, etc), Associated literature.

The Repository is addressing some of these limitations and is currently under redesigning and improvement, increasing the number of available tools, making some of the information computer readable and reorganizing parts categories ([http://partsregistry.org/BioBrick Part Program](http://partsregistry.org/BioBrick_Part_Program) ) In parallel, the Registry is encouraging part contributors to better characterize and document their part through a Part Promotion Process based on the well establish system of peer-review (see [http://partsregistry.org/Part Promotion Process](http://partsregistry.org/Part_Promotion_Process))

***Prototyping the connectivity of the MIT repository with the Biological databases, including comparative analysis fo the available technologies.***

Several common methodologies used on the genomics database area were considered:

***Flat-File Distribution: SQL direct queries: DAS:*** stands for *Distributed Annotation System*. ***Webservices/SOAP/BioMoby: API:***

During the technical workshop in 2007 we selected DAS for the prototype mainly due to the easier nature of its implementation and the possibility of collaborating with the BioSapiens NoE largely based on this technology and currently the most active collaborative project in bioinformatics in Europe.

Three experimental DAS servers, with minimal features, were implemented on the MIT

site:

- A DAS reference server that basically provide the sequence and IDs of the MIT Repository.
- A DAS annotation server providing annotation about the parts, including subparts, and many other features like, coding sequences, promoters, TF binding sites, terminators, ribosomal binding sites, mutations, etc.
- A DAS annotation server that uses Uniprot as Reference Server which when queried with a Uniprot ID returns if the protein is included in any available part in the repository.

For the demonstration purposes the DAS servers was registered in the “DAS registration service” (<http://www.dasregistry.org/>), what makes this prototype openly available to the community.

For example the Dasty, a DAS Client (Prlic et al. BMC Bioinformatics, 2007) can show the availability of a Biological Part containing the queried Protein.

***Initial implementation of tools and methods to visualize / access the information repository from genomics tools.***

We have developed a system (CARGO; <http://cargo2.bioinfo.cnio.es>, Cases *et al.*, NAR 2007,) for the representation of biological information extracted from different databases and methods. Cargo uses widgets similar to the ones implemented in Google and Macintosh desktop, and currently collects information from more than 20 sources.

In the context of this project CARGO has been extended to incorporate and represent information obtained from the repository using the DAS servers described above. This prototype shows the feasibility of this implementation and set the basis for the integration with biological/ genomic information of the MIT or any other repository of parts accessible via DAS servers.

The pilot experience for the visualization of Parts in the CARGO environment is the “IGEM parts viewer”, which takes advantage of the experimental DAS server described above to display a sketch of the part along with the sequence and provide links to Uniprot when an ID is included in the part annotation.