Improving statistics in macromolecular interactions: ABM and atomistic simulations.

Giovanni La Penna, CNR FI
Gaetano Salina INFN-TOV
Silvia Morante UNI-TOV

September 5th 2019

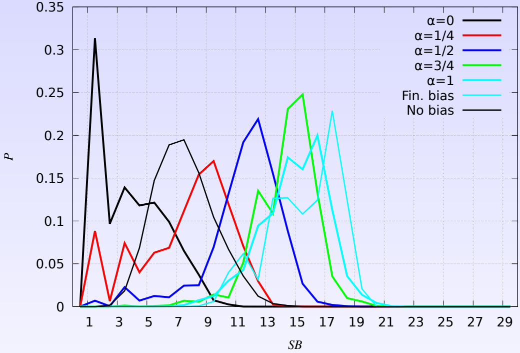
Methods: generalized statistical ensembles

How to combine N independent simulations in one unique statistical ensemble
1) choose a collective variable, CV (CV=energy, as in MC and MD, is not useful for macromolecules);
2) obtain a uniform density in CV: $n(CV) = const$
→ let one trajectory to explore a domain different from the other trajectories
→ "le traiettorie non devono pestarsi i piedi l'una con l'altra"
But trajectories must communicate → detailed balance
→ unique statistical ensemble
→ metadynamics and metastatistics
3) if the probability generating the uniform $n(CV)$ is known:
→ reweighting
→ from metastatistics to thermodynamic ensemble.

Methods: generalized statistical ensembles

To perform 1-2, altruistic metadynamics:

$$F = -V_{bias} = -[(2-\alpha)\sum_{t'\leq t} hg_{j,t'} + \frac{\alpha}{N-1}\sum_{i=1,i\neq j}^{N}\sum_{t'\leq t} hg_{i,t'}]$$

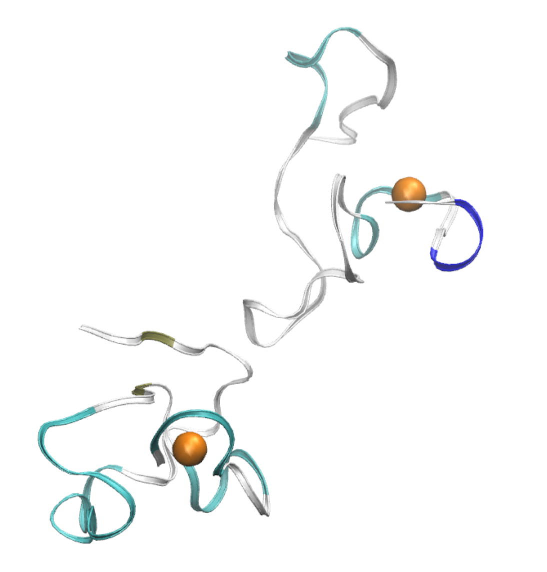


P. Hošek et al., JPCB 2016

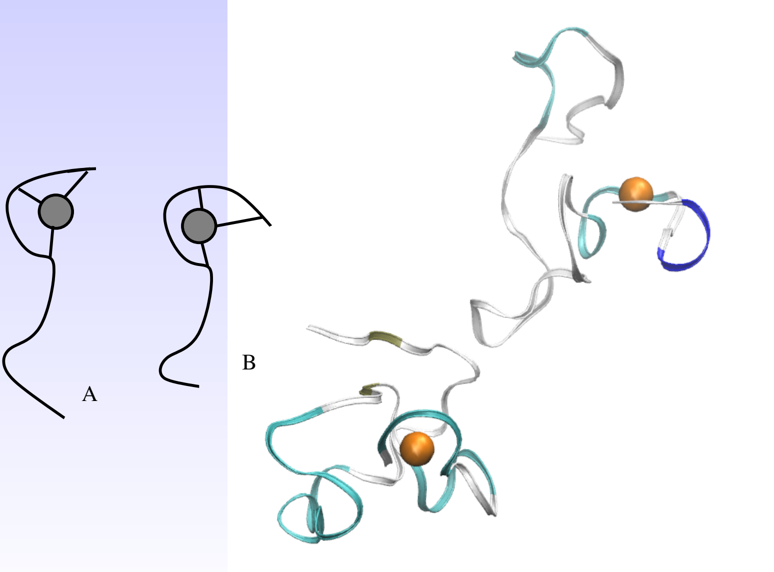
G. La Penna, M.S. Li , Chem. Eur. J. 2018

G. La Penna, M.S. Li , PCCP 2019

Example: Aβ oligomers with Cu 1:1

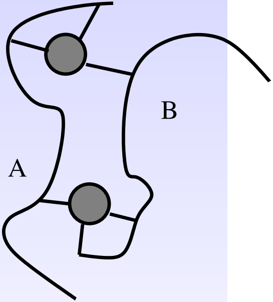


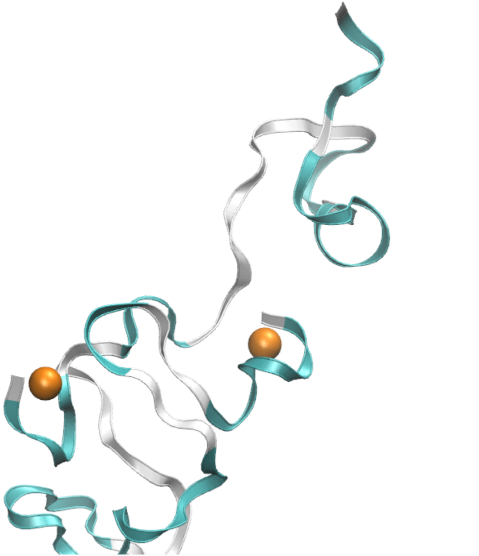
Example: Aβ oligomers with Cu 1:1



Example: Aβ oligomers with Cu 1:1

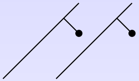
Example: Aβ oligomers with Cu 1:1

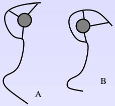




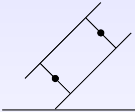
Example: Aβ oligomers with Cu 1:1

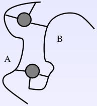
$2\times$ Cu-Aβ_{42}



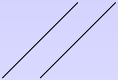
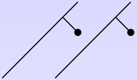
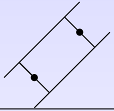
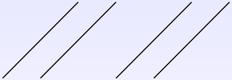
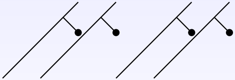
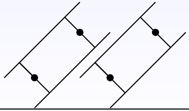


[Cu-Aβ_{42}]$_2$





Example: Aβ oligomers with Cu 1:1, dimers & tetramers

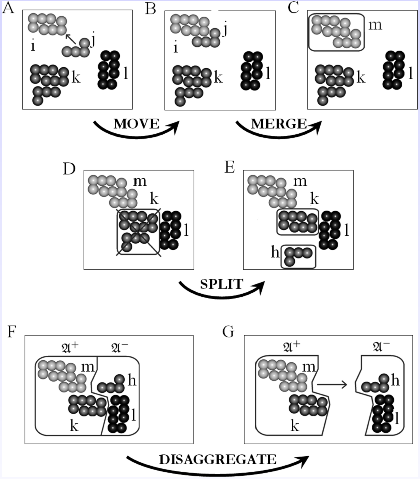
2× Aβ_{42}		128
2× Cu-Aβ_{42}		128
[Cu-Aβ_{42}]$_2$		55
	A-B C-D	
4× Aβ_{42}		125
4× Cu-Aβ_{42}		126
2× [Cu-Aβ_{42}]$_2$		127

Beyond altruism?

Agent based models:
agents=molecular simulations
exchanged information=histograms in CV
...
Direction: from molecular simulations to agents

Direction: from macromolecules to agents

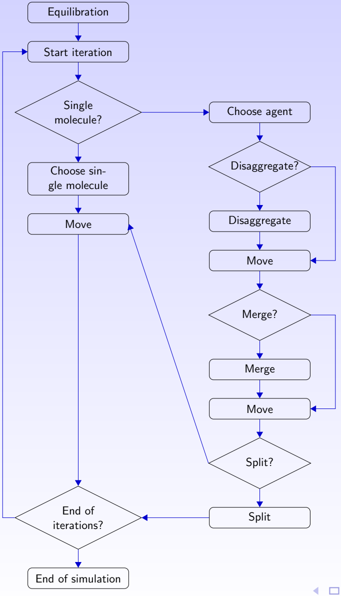
Simulations of N macromolecules provide emerging properties:



Fortuna and Troisi, JPCB, 2009

Thesis: Nicole Corretta, 20 set 2018

Direction: from macromolecules to agents

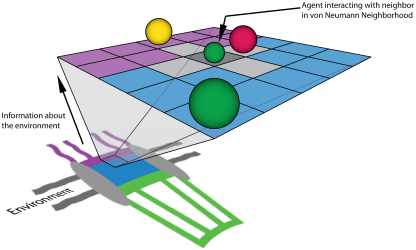


The other direction: from networks to agents

Networks provide linkages between macromolecules
→ transform linkages in rules for ABM
One application:

$$k_{on}/k_{off} \rightarrow P_{on}/P_{off}$$

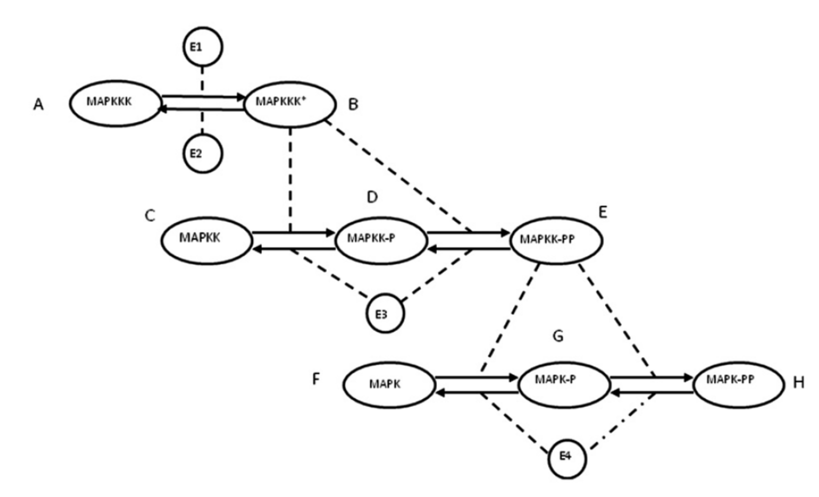
mapping binding/unbinding to swarming of different molecules over a rigid lattice of cells, each containing one or two molecules.
Azimi and Mofrad, PLOS ONE, 2013


Agent interacting with neighbor in von Neumann Neighborhood
Information about the environment
Environment

Reviewed in: Soheilypour and Mofrad, BioEssays, 2018

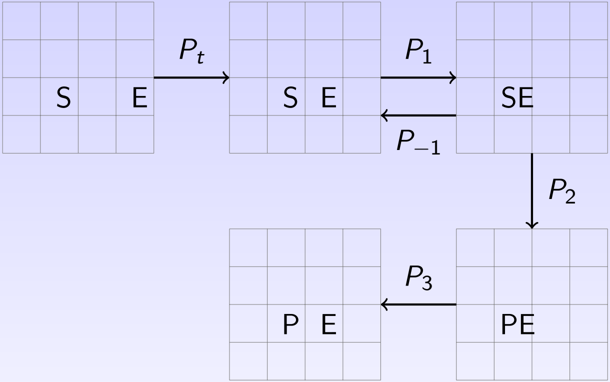
The other direction: from networks to agents

Bonchev et al., SAR QSAR Env. Res. , 2010



→ from networks to simulations (via ABM)

The other direction: from networks to agents



$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \xrightarrow{k_2} PE \xrightarrow{k_3} P + E$$

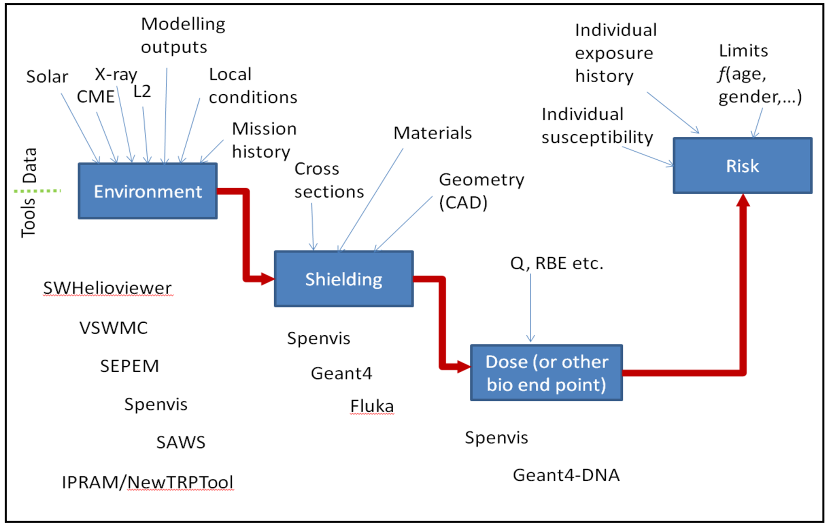
ERFNet Knowledge Base (EKB)

Gaetano Salina
INFN Sezione di Tor Vergata

Possibili sinergie con il Prin,
ovvero la nobile arte del riciclo …

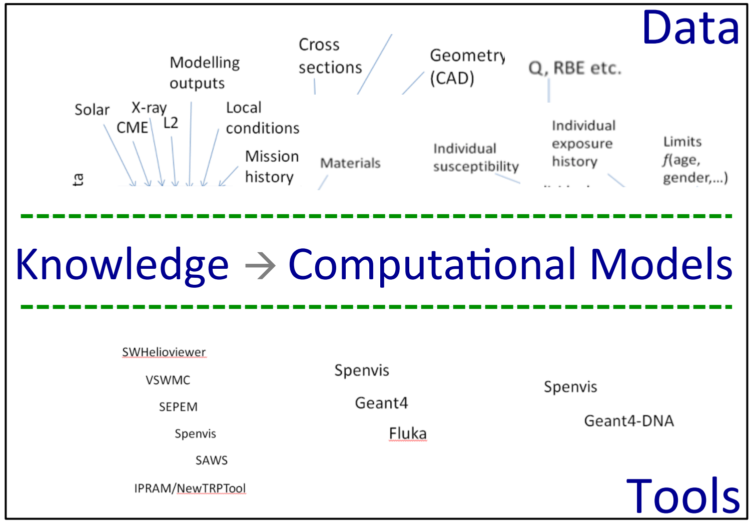
Why EKB ?

Why is not enough a sets of DataBases ?



From Samantha mail …

implies



I am an expert, then I know everything !

Unfortunately this is almost never true!

I do not know, but a friend of mine knows !

It may be a bit ineffective !

I do not have data, but I can use a DataBase to collect them !

You can spend a lot of time to develop the Interface Code!

I can find a solution in some hours !

The problem requires real time !

Why EKB ?

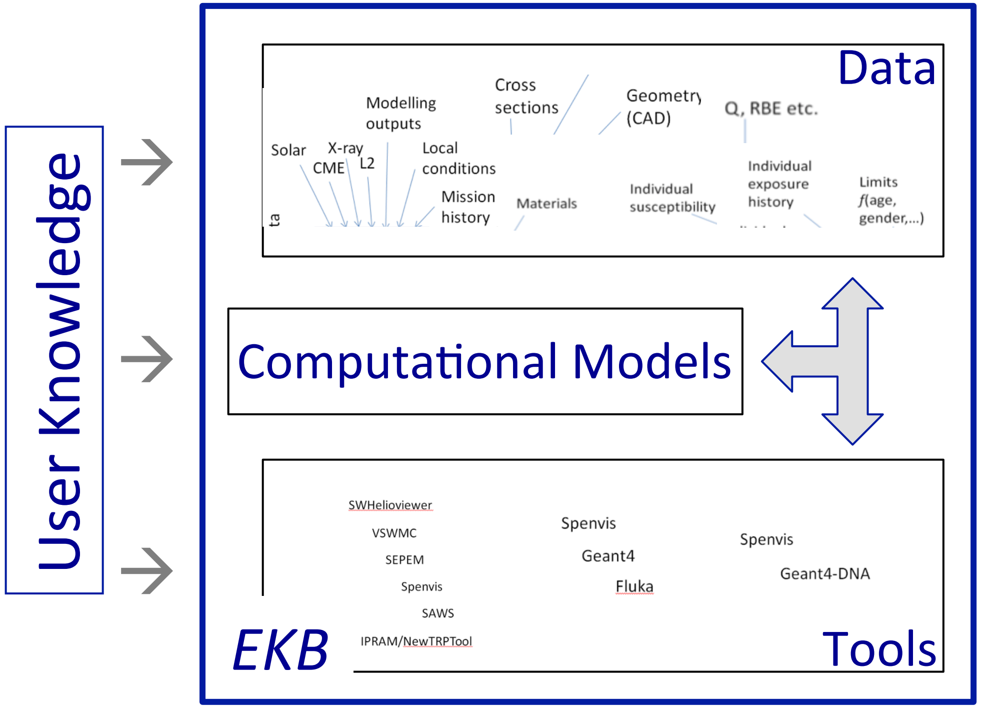
Why is not enough a sets of DataBases ?

ERFNet Knowledge Base is a evolutive, modular and expandible Analysis Enviriorment.

EKB helps the user to performe complex analysis.

The user adds his/her knowledge to EKB in terms of tools, data and Computational Models.

I am supposing an Open Data/Open Source Philosofy



User Knowledge

Data

Solar
CME
X-ray
L2
Modelling outputs
Local conditions
Mission history
Cross sections
Materials
Geometry (CAD)
Individual susceptibility
Q, RBE etc.
Individual exposure history
Limits *f*(age, gender,...)
ta

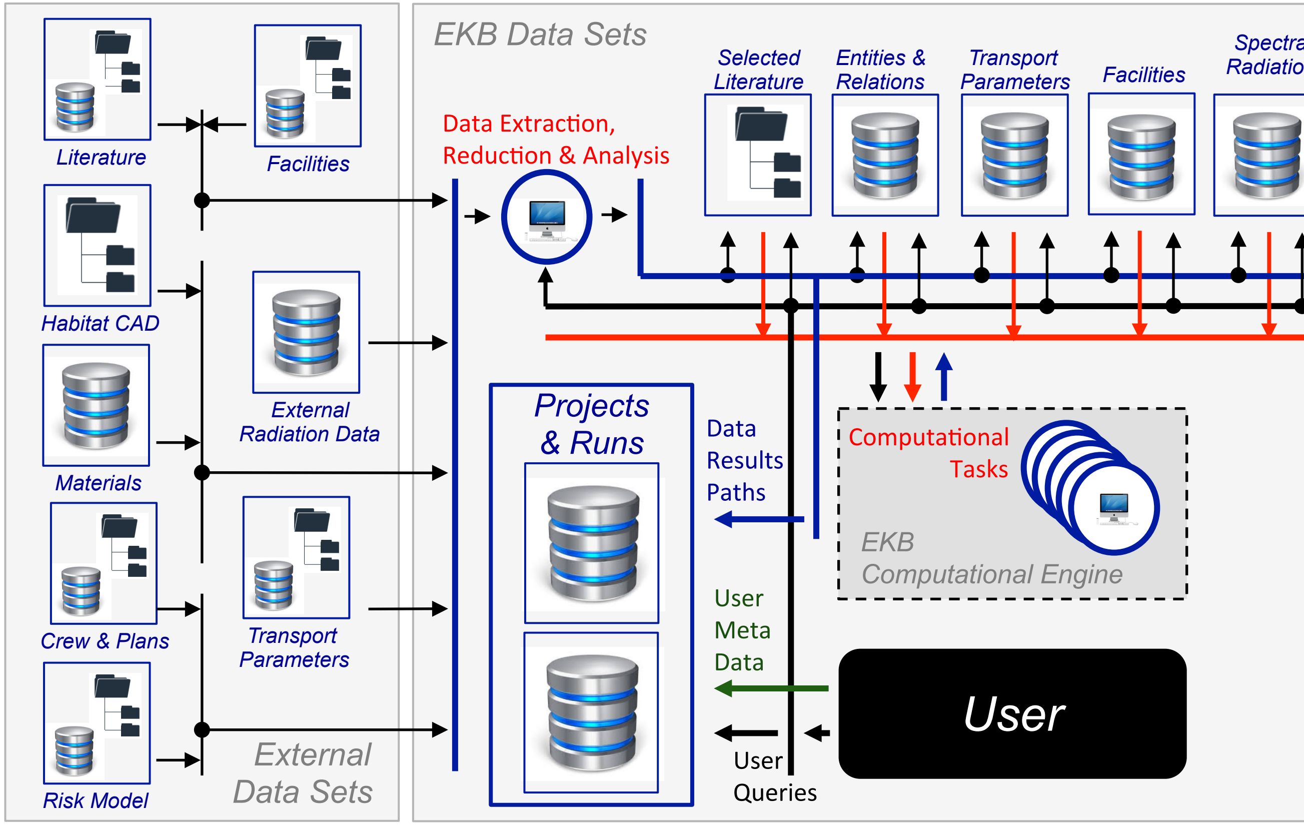
Computational Models

SWHelioviewer
VSWMC
SEPEM
Spenvis
SAWS
IPRAM/NewTRPTool
Spenvis
Geant4
Fluka
Spenvis
Geant4-DNA

EKB

Tools

EKB first Key Point: Projects & Runs DataBase (PRDb)



PRDB DATA STRUCTURE

Projects and Runs (PRDB) are the foundation of EKB. They represent any user operation (**User Knowledge**).

Project: All data, metadata, computational tasks that are involved in a user work. A Project is a collection of Run.

Run: Run is defined as the set of i) input data; ii) a list of Computational Tasks; iii) the output Data, i.e. and iv) some job status parameters.

From Our Proposal:

WP1: Development of an integrated protein variation database (months 1:36)

All the units will collaborate to the database implementation, starting from public resources, enriching the database with information derived from new experiments and computational annotations. …

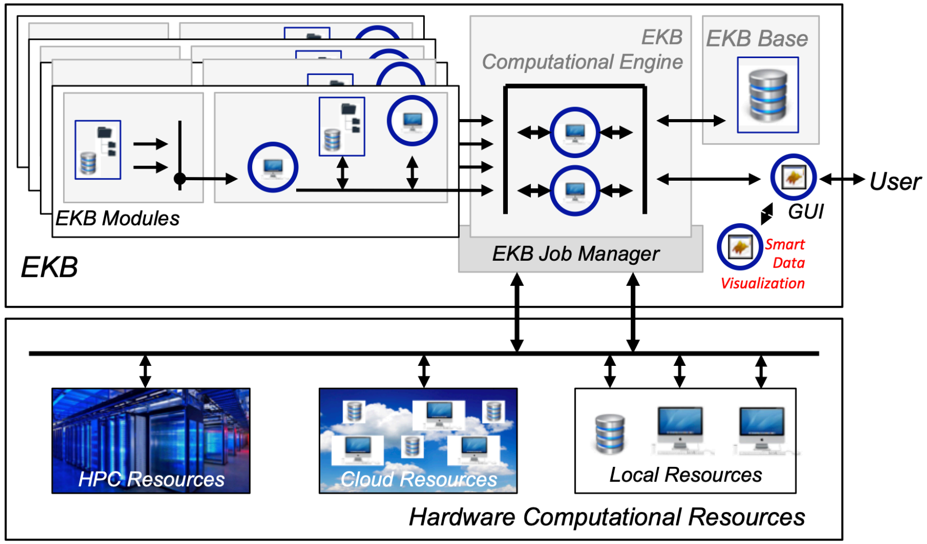
WP2: Development of the Predictors (months 5-30)

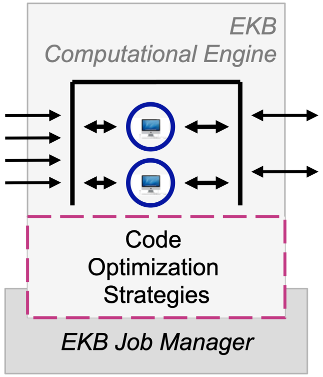
We will develop a new computational tool for predicting the impact of protein variants integrating our state-of-the-art predictors. We will set up a specific Information Technology (IT) infrastructure that will host the software and the data for the project. The main novelty of the new method will consist in the integration of sources of information from different variation databases allowing to capture complementary aspects of the relationship between protein stability, function and disease

EKB second Key Point: EKB Computational Engine

There will be a continuous effort

- to optimize data retrieval and use,
- to develop new algorithms
- to improve the old algorithms
- to optimize the use of hardware resources, in order to reduce the computational time
- to develop new algorithms & tools

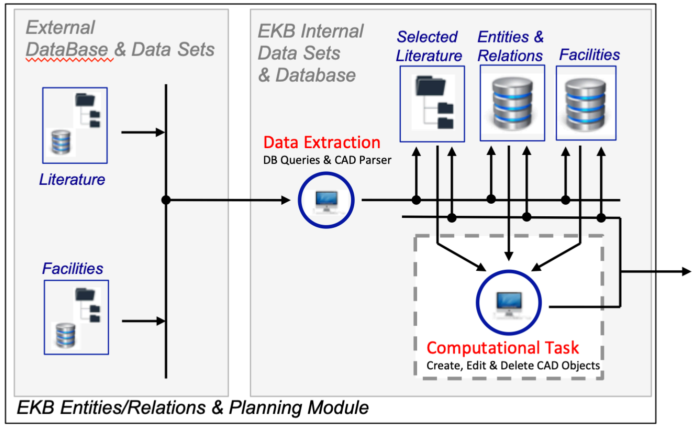




CPU Time is certainly a key parameter when optimizing codes.

EKB Modula Hardware/Software implementation

Entities/Relations & Planner Module


External DataBase & Data Sets
Literature
Facilities
EKB Internal Data Sets & Database
Selected Literature
Entities & Relations
Facilities
Data Extraction
DB Queries & CAD Parser
Computational Task
Create, Edit & Delete CAD Objects
EKB Entities/Relations & Planning Module

The external datasets

& databases structure

Any external repository of documents (research papers, internal documents, expert notes, data tables, etc.). Commonly, these documents contain not structured data or cross-related at all, neither organized in a relational fashion. We focus on:

- PDF Documents: It is the most diffused exchange format for research papers.
- WEB Pages: gathering and structure any data present in it (es. Facilities institutional WEB Pages)
- Raw Textual Data: Any not structured data provided by users in forms of memo, notes, etc.

Facilities web pages providing and requesting information and services.

From Our Proposal:

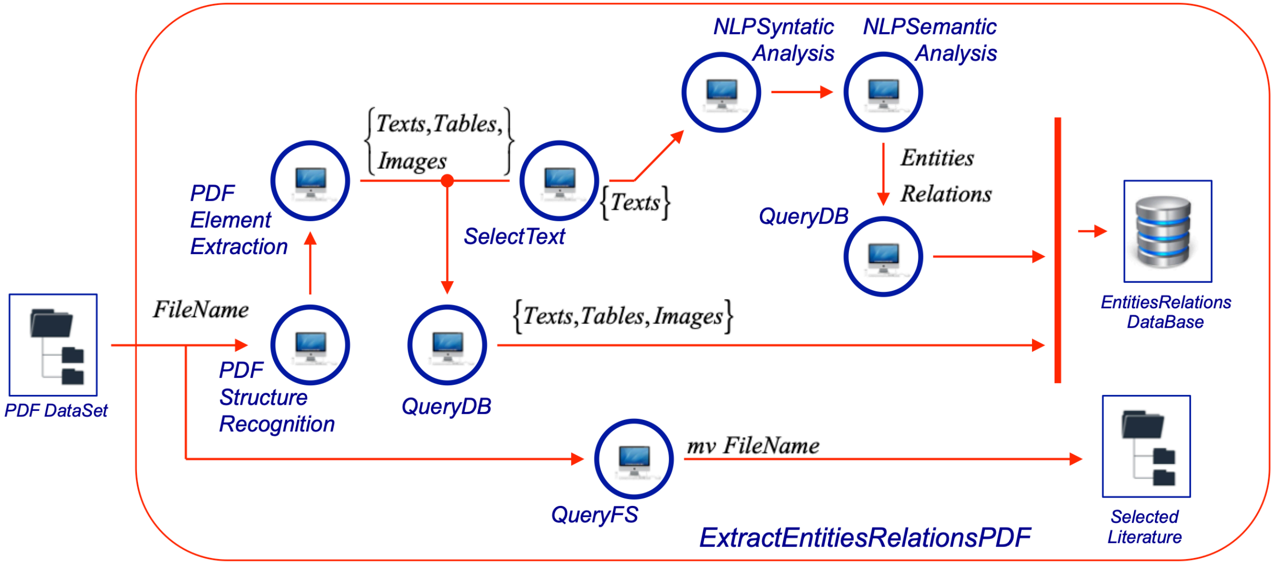
This WP (WP1) includes the development of a semi-automatic tool for extracting experimental data from the literature and update the database. To perform this task we will implement a semi-automatic method based on Natural Language Processing (NLP) techniques. The algorithm will scan each manuscript abstract in PubMed and will return a score related to the probability of finding experimental measures of protein stability in the full-text version of the paper.

EKB Modula Hardware/Software implementation

ENTITIES/RELATIONS & PLANNER MODULE

Natural Language Processing, NLP, plus a bidirectional Long Short- Term Memory (bi-LSTM) plus a CNN layer to perform Entities/Relations extraction.

It extracts Entities and Relations from the PDF file *FileName.* Store *FileName* in *Selected Literature* dataset.



ExtractEntitiesRelationsPDF(*FileName*)