

# VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean

International conference “e-Infrastructures for Excellent Science in Southeast Europe and Eastern Mediterranean”  
15-16 May 2018

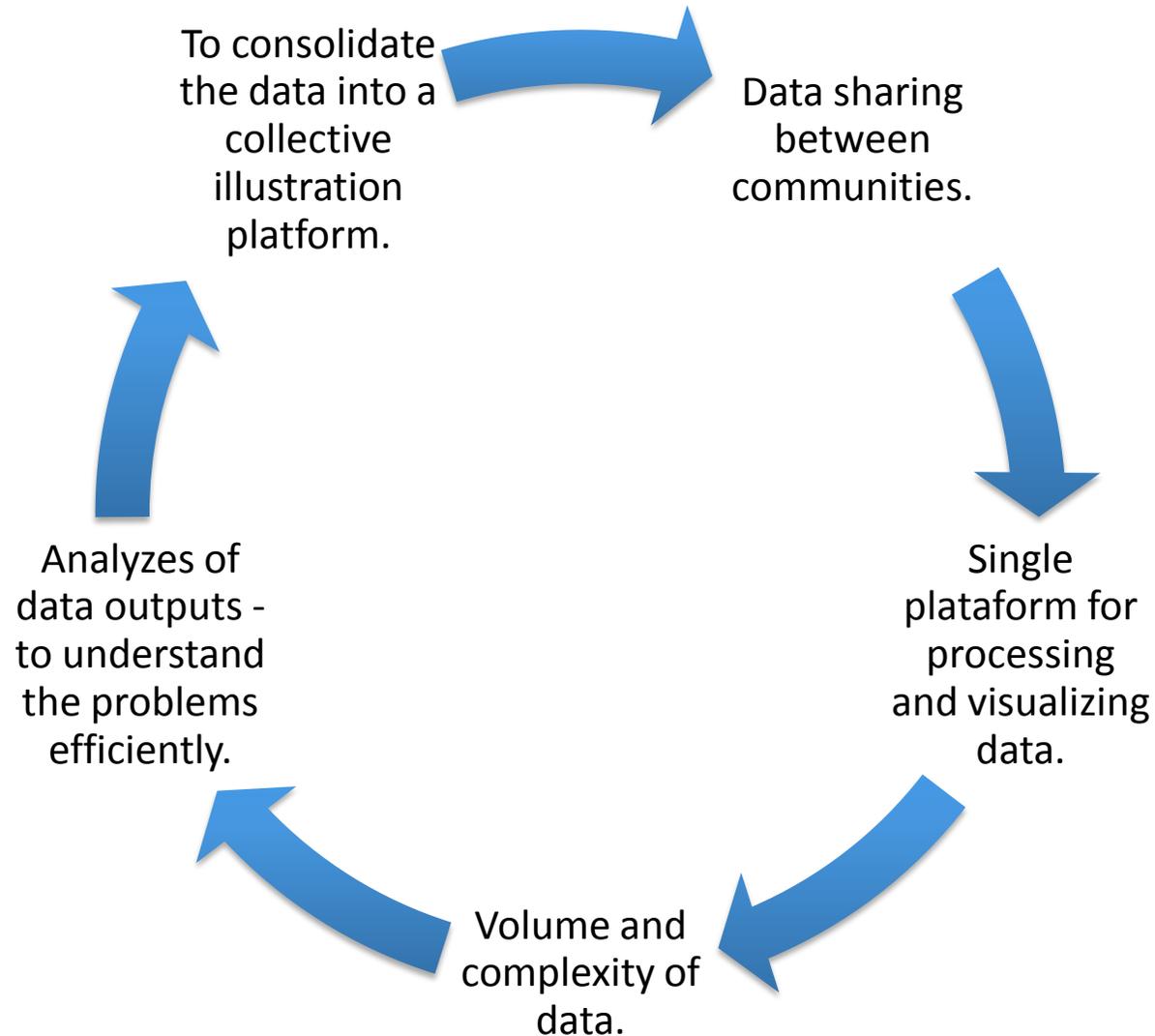
## MDSMS: An integrated Web-based Interactive Data Platform for Molecular Dynamics Simulations

Armen H. Poghosyan  
Bioinformatics Group @ ISEC NAS of Armenia



- ❑ A molecular dynamics (MD) simulations are widely used in the domain of life science to evaluate the equilibrium nature of classical many-body systems. The study of complex systems with a large number of atoms in long trajectory intervals (up to milliseconds) is required to explore many phenomena, which can be realized only with using appropriate HPC resources and storage facilities to manage and visualize these data.
- ❑ Various life science communities from Armenia use HPC resources and generate a large number of research outputs by storing them in local repositories.
  - ❑ Local datasets are usually incomplete
  - ❑ Need to manage data using appropriate metadata and identifiers
  - ❑ No centralized repository to hold all these data

# Challenges



# Scientific Communities



Institution	Scientific Domain	More information
Bioinformatics Group of the International Scientific and Educational Center of the National Academy of Sciences of the Republic of Armenia <b>(Bioinformatics group)</b>	MD simulation of complex systems, including surfactants, polymers, and proteins	<a href="http://bioinformatics.am">http://bioinformatics.am</a>
Molecular Physics department of the Yerevan State University <b>(MolPhys)</b>	Nucleic acids, e.g. double-stranded DNA and single-stranded RNA molecules	<a href="http://www.y-su.am/faculties/hy/Physics">http://www.y-su.am/faculties/hy/Physics</a>

# Bioinformatics Group at ISEC NAS RA



The major fields of the scientific activities of the Group:

- ❑ Modeling and Molecular Dynamics (MD) study of biological membranes.
- ❑ Modeling and MD study of complex systems, including surfactant and polymer solutions.
- ❑ Development of specialized software packages



- direct links of traj./coord. files are available to download.

Repositories :

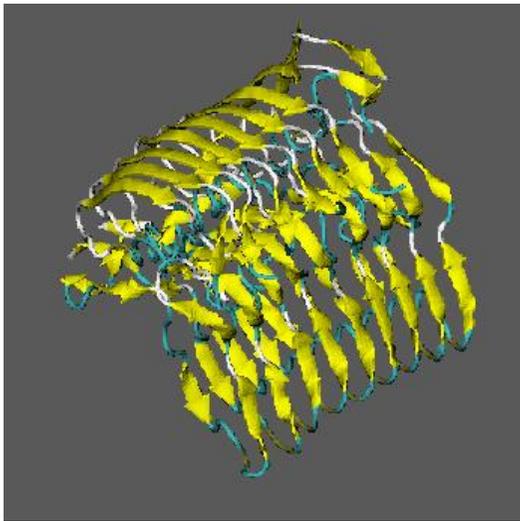
bioinformatics.sci.am, local machines

The screenshot shows the website [bioinformatics.am/downloads/](http://bioinformatics.am/downloads/). The page features a navigation menu with 'Downloads' highlighted. Below the menu, there is a 'Downloads' section with a table of files. An arrow points from the text 'Repositories : bioinformatics.sci.am, local machines' to the 'Downloads' section.

Colloidal systems & Polymers	
<a href="#">sds128_w29.gro</a>	United atom starting .gro configuration of sodium dodecyl sulfate(SDS)water system consisting of 128 SDS and 3750 water molecules (hydration is ~29 water per surfactant).
<a href="#">sds512_w29.gro</a>	United atom starting .gro configuration of SDS/water system consisting of 512 SDS and 15000 water molecules (hydration is ~29 water per surfactant).
<a href="#">sds512_w29_all_atom.pdb</a>	All atom .pdb initial configuration of SDS/water system consisting of 512 SDS and ~15000 water molecules (~ 29 water per surfactant) developed for NAMD package.
<a href="#">sds.itp</a>	Topology file for SDS molecule.

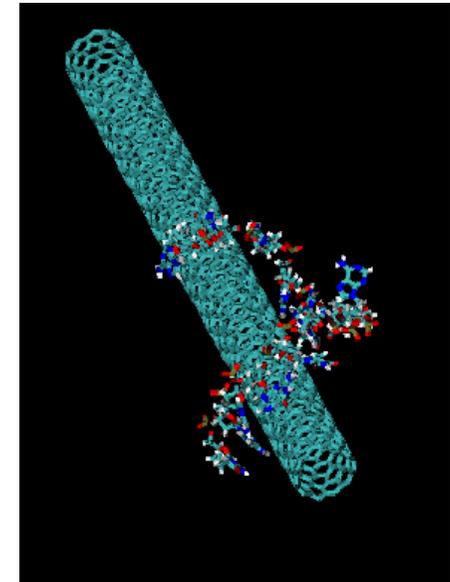
The major fields of scientific activities of the Molecular Physics department :

- Study of the nucleic acids interaction with low-molecular compounds and nano-particles using Molecular Dynamics (MD) simulations, toy models and experimental techniques.



FUS-LC protein fibril fragment

**Repositories:** local machines



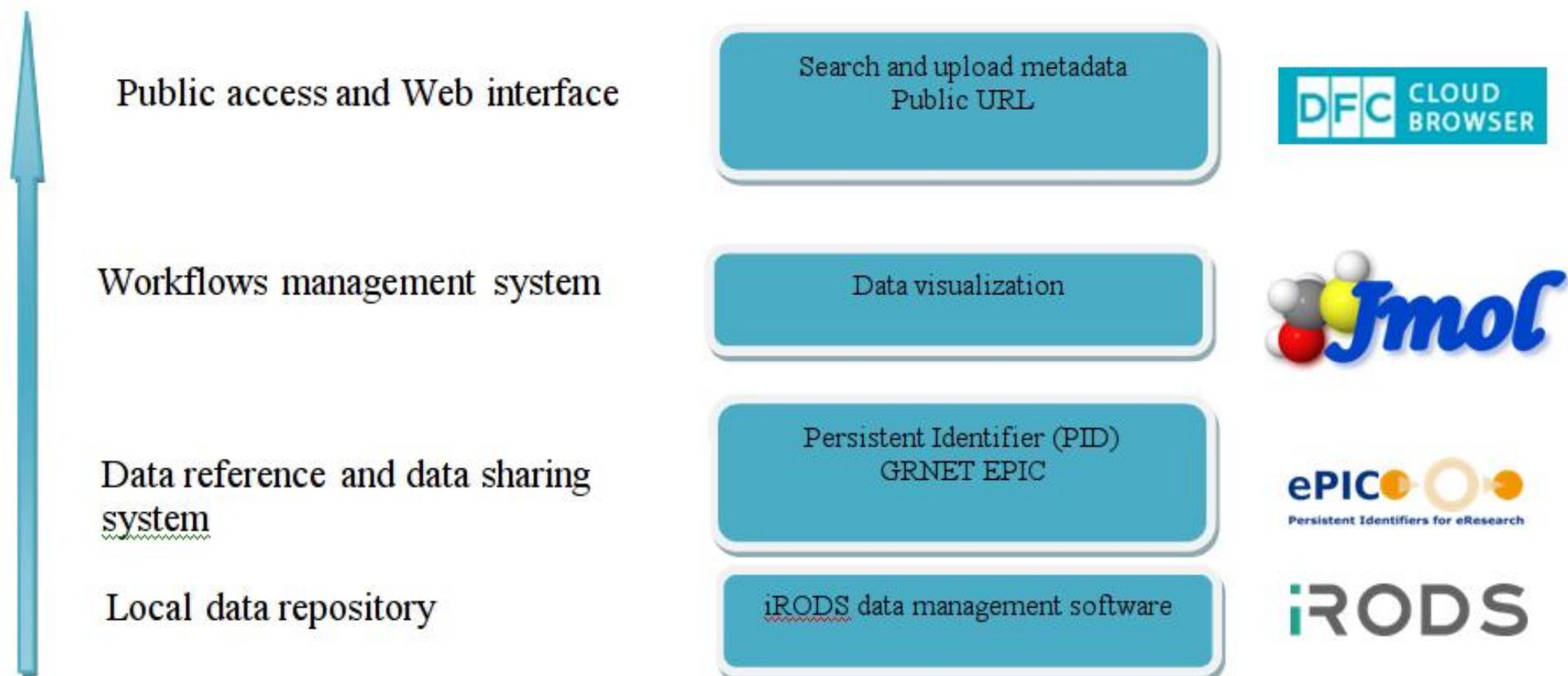
SWCNT wrapped by ssDNA

- Modeling and MD studies of complex systems, including protein fibrils, polymer solutions, etc.

- ❑ An integrated web-based interactive data platform (data repository and workflow management services) for molecular dynamics simulations using the datasets generated by several Armenian life science communities.
- ❑ Data visualization workflow service to perform various analyzes of data and to consolidate the data into one common platform.
- ❑ The mentioned platform is presented as an advanced integrated environment to capture, analyze, process and visualize the research data.

# Web-based interactive data platform

- Diagram of the integrated web-based interactive data platform



# Web-based interactive data platform



## ❑ Local Data repository

- ❑ To preserve the data for future work
- ❑ To assign metadata and persistent identifiers for each data
- ❑ To increase the data discovery over the net
- ❑ To prevent the users from maintaining the data by themselves
- ❑ To enable data sharing between different communities

## ❑ The types of datasets

- *trr/xtc/dcd* file formats - the trajectories of the simulations
- *pdb/gro/itp/psf* file formats – simulation coordinate/topology files

Note that the metadata are stored in Mysql database

**<http://irods.asnet.am:8080/irods-cloud-backend>**

# Web-based interactive data platform



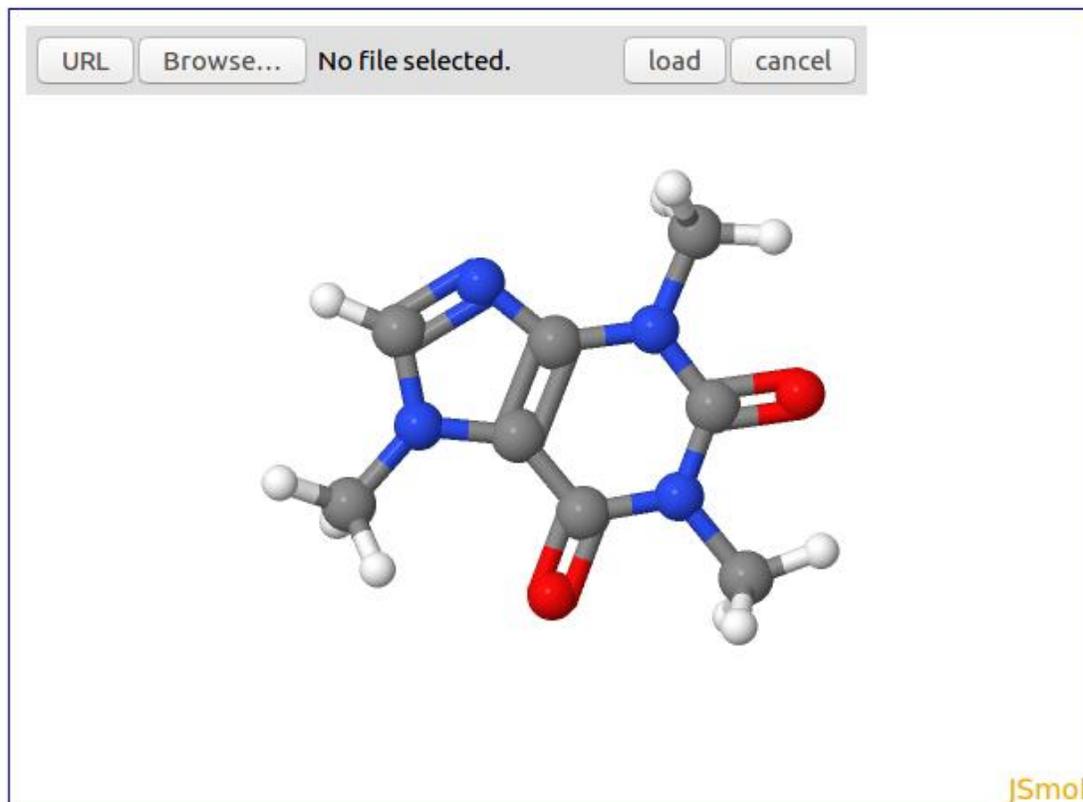
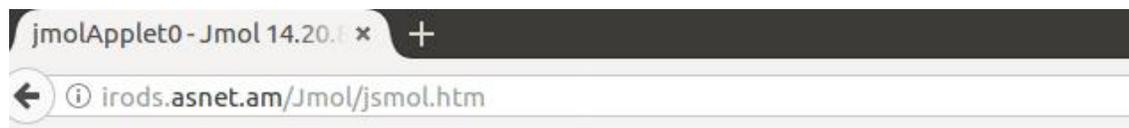
- ❑ Data reference and data sharing
  - ❑ Persistent Identifier (PID)\* is used instead of commonly used Uniform Resource Locator (URL). The PIDs are generated and registered by data centers enabled through European Persistent Identifier Consortium (EPIC).
- ❑ Workflow management system
  - ❑ To run complex workflows that integrate programs, methods, and data
  - ❑ To run different simulations in a single consolidated platform
- ❑ Public access and Web interface
  - ❑ Jmol based interface
  - ❑ browser-based *HTML5* viewer
  - ❑ Java 3D viewer
  - ❑ <http://irods.asnet.am/Jmol/jsmol.htm>

**<http://irods.asnet.am:8080/irods-cloud-backend>**

\* W. Cockshot, et al, *Persistent object management system*, *Softw: Pract. Exper.*, **14** pp. 49–71. (1984)

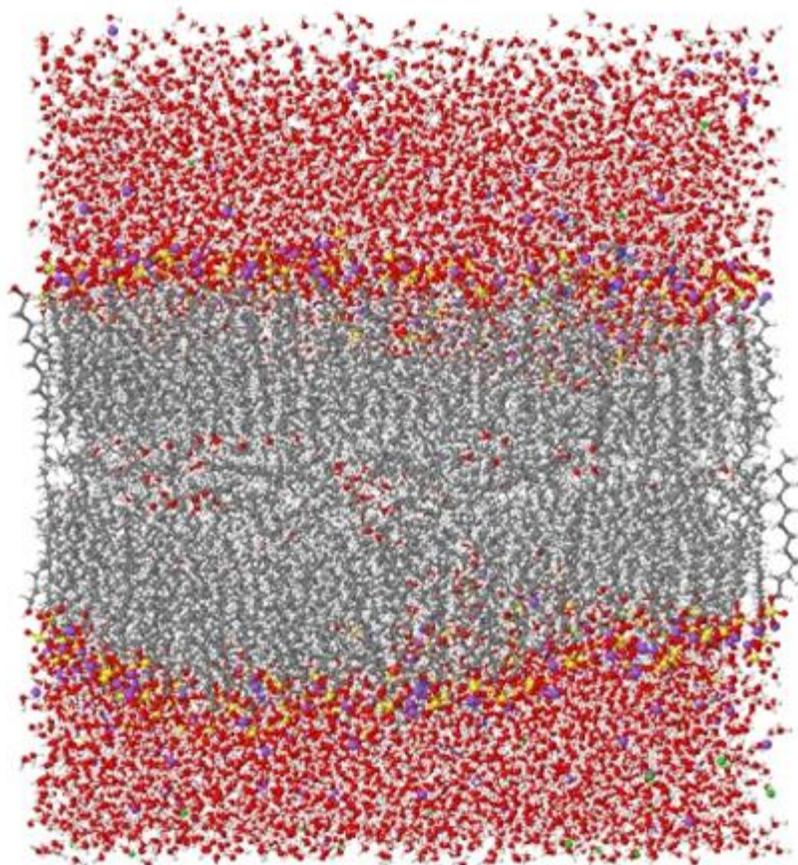
# Web-based interactive data platform

## □ Data visualization interface

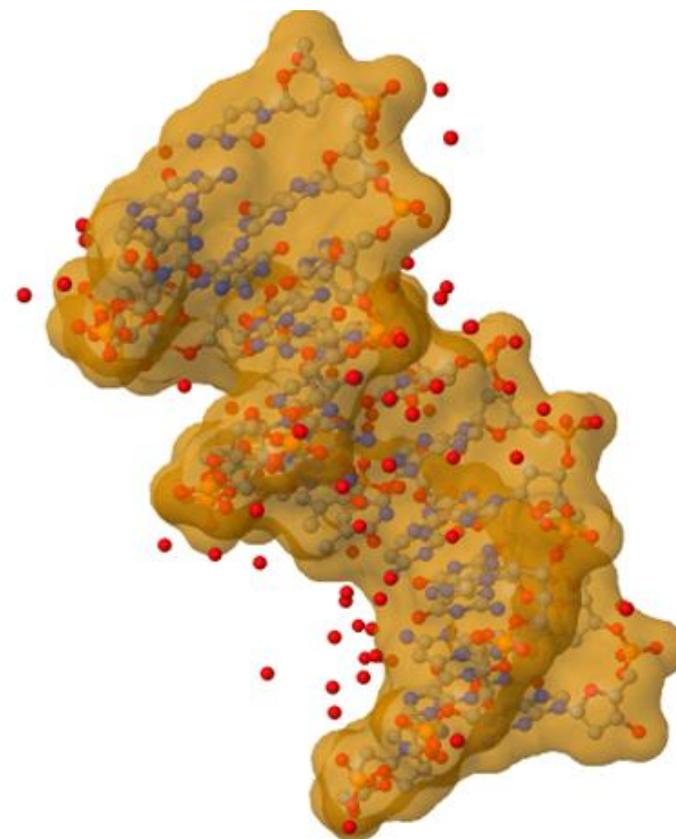


# Web-based interactive data platform

## □ Database example



PDADMAC/SDS/Decanol in water bulk



**d(CGCAAATTTTCGC)2 in presence Na<sup>+</sup> and Cl<sup>-</sup> ions, Water molecules are hidden.**

# Acknowledgment



- The research leading to these results has been co-funded by the European Commission under the H2020 Research Infrastructures contract no. 675121 (project VI-SEEM) and the State Project entitled "Towards the Green e-Infrastructures" funded by the Government of the Republic of Armenia.

## THANKS FOR YOUR ATTENTION