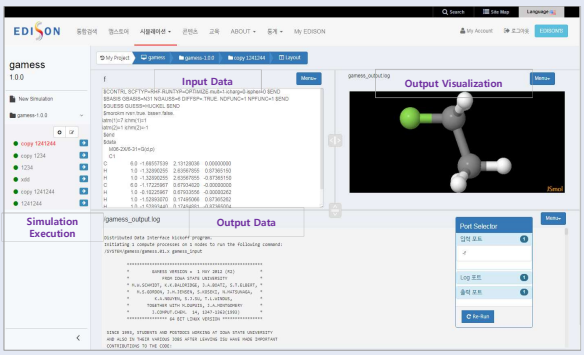
Computational simulation data preprocessing in the EDISON-DATA platform

Sunil Ahn
Division of National Supercomputing,
Korea Institute of Science and Technology Information, South Korea
siahn@kisti.re.kr

EDISON

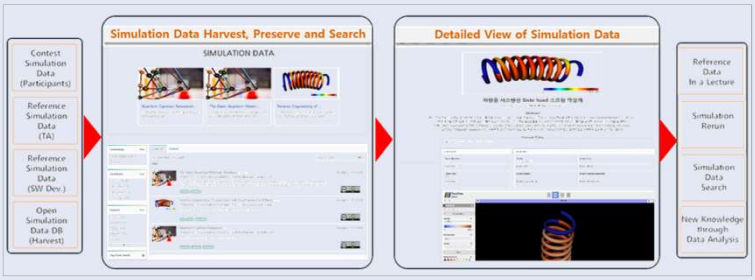
- Education-research-industry Integration through Simulation On the Net : Help computational science & engineering researchers and students access the EDISON portal on the web to learn and virtually experiment on theory and application in the multiple disciplines whenever and wherever
- Objectives

- Develop a web-based simulation SWs/Contests and a fusion environment based on most recent R&D achievements in the science-engineering application area



EDISON-DATA

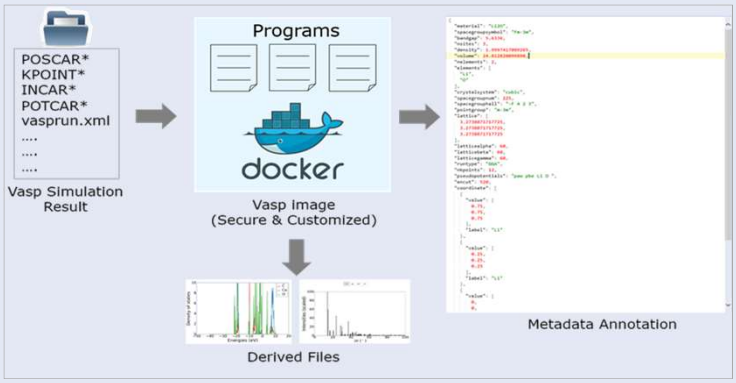
- Simulation Data Repository
  - Solve diversity problem of managing and processing simulation data
  - Access control and embargo for data sharing and protection
  - Version control, checksum and replication for data preservation
  - Restful interface to support data access and analysis



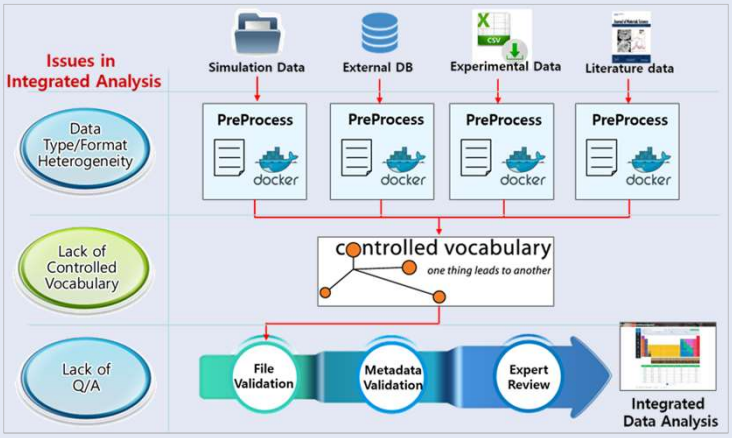
- (pilot application) Material Research Environment
  - Improve the usability by providing easy and convenient environment for Quantum Espresso Simulation
  - Avoid redundant calculations by searching and analyzing simulation data

Preprocessing in EDISON-DATA

- Issue : complexity and diversity of computational simulation data
- Ecosystem is required in which data preprocessing modules are developed, contributed and shared by domain users
- Considerations for having an ecosystem
  - require to use various languages and development tools
  - require specific execution environments depending on simulation SW
  - necessary to provide ways that can be easily used even by users who are not familiar with programming
  - security concerns when preprocessing are developed by 3rd parties
- Techniques to deal with the issue
  - Container technology, Simple dedicated commands set, Python script
  - and combination



- Key to Solve Data heterogeneity for integrated analysis, and to Apply to various fields



Acknowledgements

- This research was supported by KISTI Program (No. K-19-L12-C05-S01), the EDISON Program through the National Research Foundation of Korea (NRF) (No. NRF-2011-0020576) and the National Research Foundation of Korea (NRF)