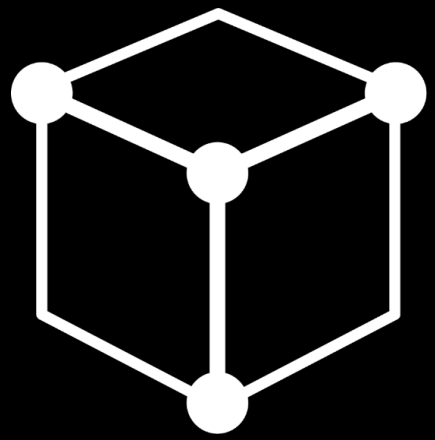


mdbox

A cloud-based repository for molecular dynamics simulations

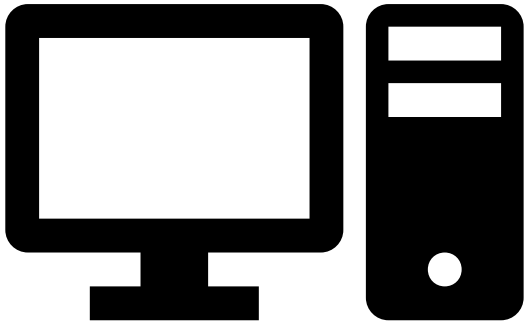
Karmen Čondić-Jurkić

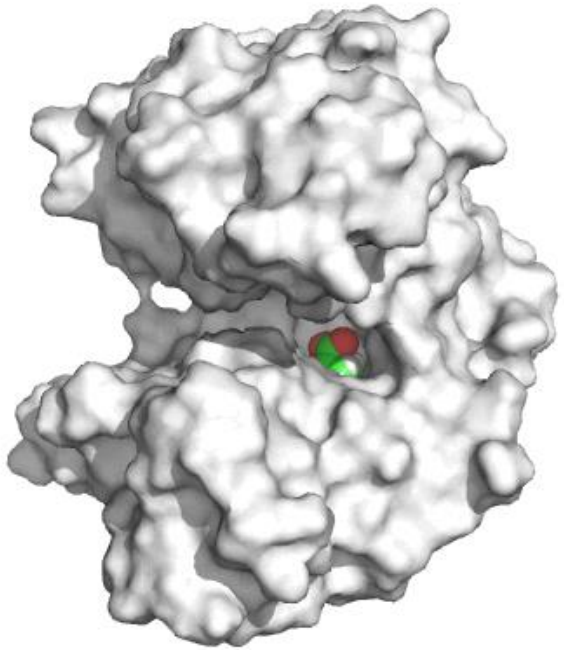
Australian National University

Canberra, Australia

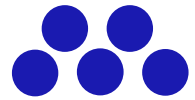
molecular dynamics simulations

- (super)computer simulations of the movement of atoms and molecules





N atoms



3N positions (*x, y, z)*

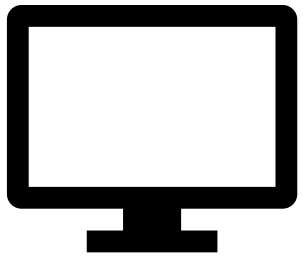
3N velocities (v_x, v_y, v_z)

trajectory

(size in GB)

https://amit1b.wordpress.com/computational-biophysics/

MD simulations workflow

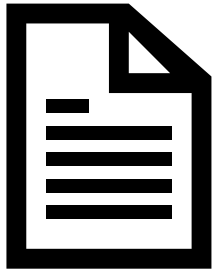


run simulation

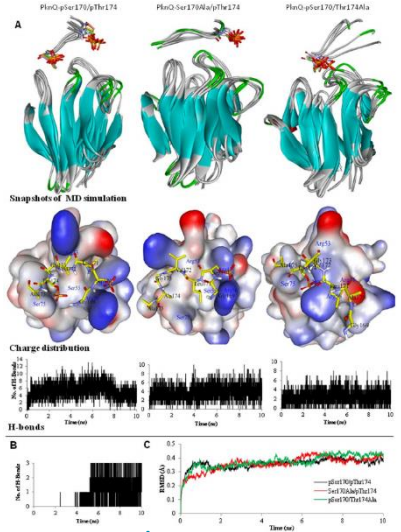
- pick a software
- build your system locally
- select input parameters
- find a powerful (super)computer

analyse trajectories

- retrieve data
- limited number of properties analysed
- custom analysis scripts or software packaged tools (unknown parameters)



publish results



Arora et al, PLoS NTD, 2014

store data

- different file formats for input and output files
- directory tree

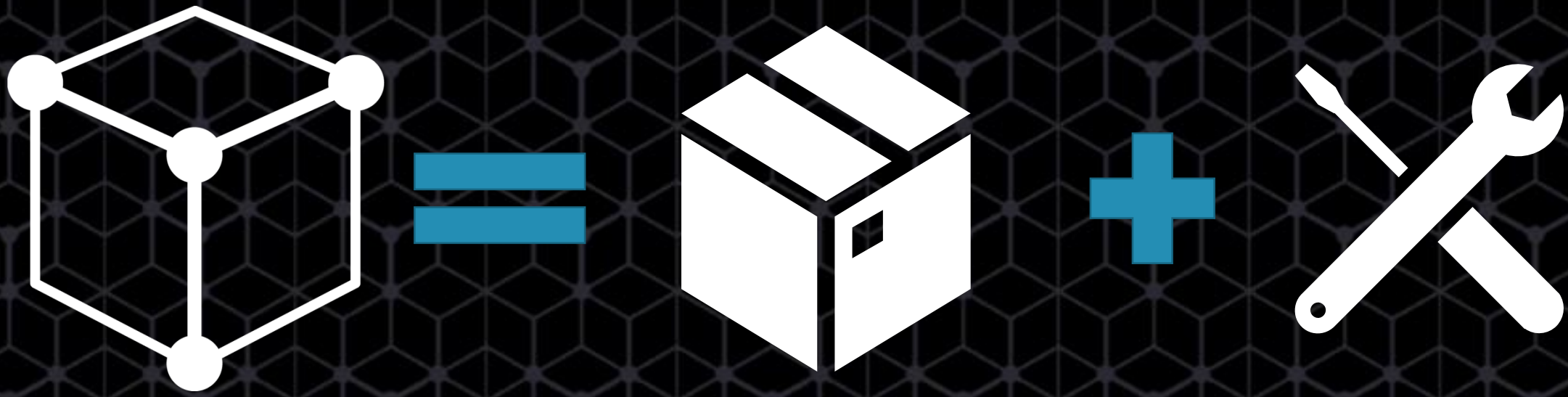
problems?

- poor data management
- poor project management
- difficulties with reproducibility
- lack of transparency
- scientific practice?

not FAIR
at all!

solution

- publicly available repository with a detailed metadata schema and analysis tools



mdbox = storage + toolbox

(general) scientific repositories

institutional repositories

The Dataverse Project

Open Science Framework

do we need one more?

zenodo

figshare

specialised repository

specialised repository

better and more accessible science

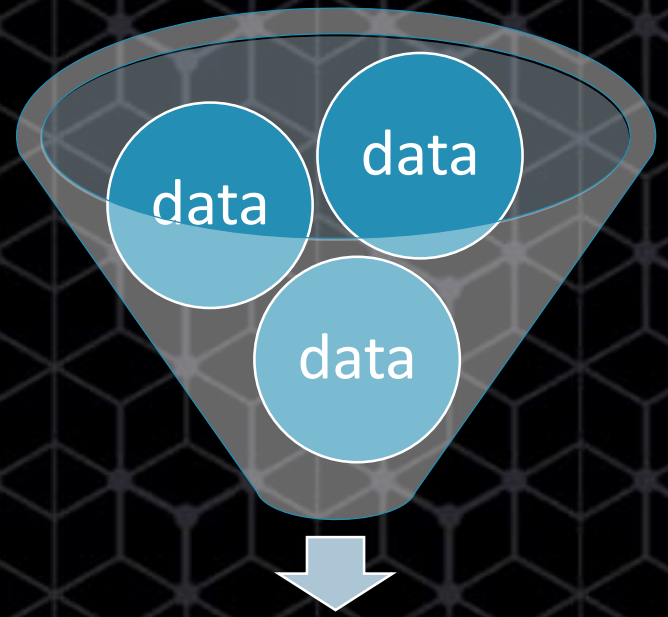
new methods (machine learning)

new analysis & vis. tools / new insights

better documentation for simulation protocols

customised, detailed metadata / curation

general repository



knowledge

prototype: open source components

Analysis: MDTraj [3]

a python library that allows users to manipulate MD trajectories

Metadata: iBIOMES – managing and sharing simulation data in a distributed environment [2]

data model and dictionaries for molecular simulation data indexing

Framework: CKAN - open source data management tool [1]

backend: *Python*, frontend: *Javascript*, web framework: *The Pylons*, ORM: *SQLAlchemy*, database engine: *PostgreSQL*, search: *SOLR*

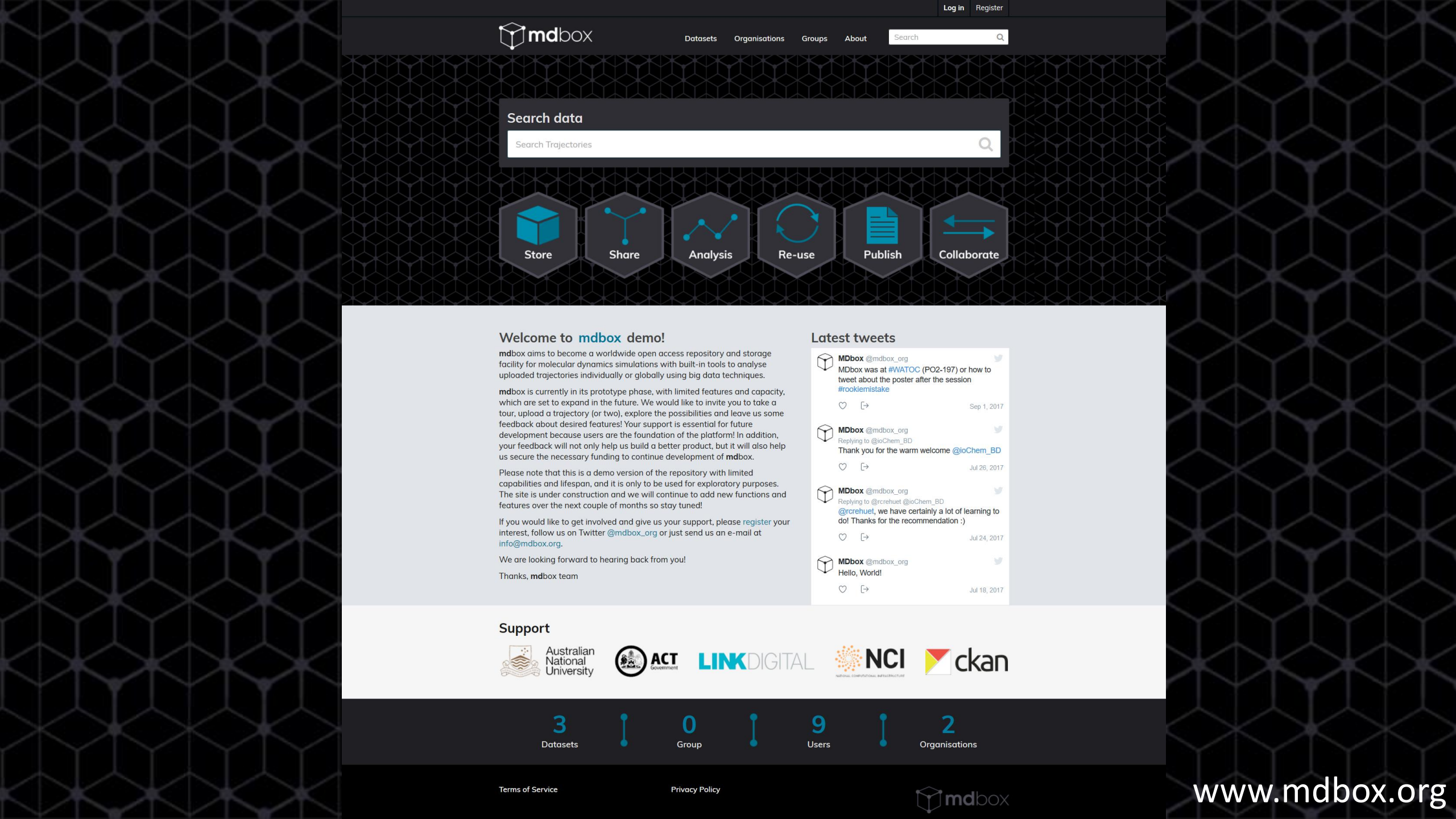
1. www.ckan.org

2.Thibault *et al.* J. Chem. Model. 2013

3. McGibbon *et al*, Biophys. J. 2015

upload/download

- required files for upload:
  - all the input files required to repeat or continue the uploaded simulation (simulation parameters, customised force field files, initial coordinates)
  - all the output files (trajectories and log files) – any trajectory file format is accepted, but will be stored in a single file format (HDF5)
- metadata is created by parsing the uploaded files and via users input (project description, relevant sources, links, etc)
- basic analysis tools implemented (RMSD, RMSF)
- unique identifier assigned to each dataset


Log in
Register
mdbox
Datasets
Organisations
Groups
About
Search
Search data
Search Trajectories
Store
Share
Analysis
Re-use
Publish
Collaborate
Welcome to mdbox demo!
mdbox aims to become a worldwide open access repository and storage facility for molecular dynamics simulations with built-in tools to analyse uploaded trajectories individually or globally using big data techniques.
mdbox is currently in its prototype phase, with limited features and capacity, which are set to expand in the future. We would like to invite you to take a tour, upload a trajectory (or two), explore the possibilities and leave us some feedback about desired features! Your support is essential for future development because users are the foundation of the platform! In addition, your feedback will not only help us build a better product, but it will also help us secure the necessary funding to continue development of mdbox.
Please note that this is a demo version of the repository with limited capabilities and lifespan, and it is only to be used for exploratory purposes. The site is under construction and we will continue to add new functions and features over the next couple of months so stay tuned!
If you would like to get involved and give us your support, please register your interest, follow us on Twitter @mdbox_org or just send us an e-mail at info@mdbox.org.
We are looking forward to hearing back from you!
Thanks, mdbox team
Latest tweets
MDbox @mdbox_org
MDbox was at #WATOC (PO2-197) or how to tweet about the poster after the session #rookiemistake
Sep 1, 2017
MDbox @mdbox_org
Replying to @ioChem_BD
Thank you for the warm welcome @ioChem_BD
Jul 26, 2017
MDbox @mdbox_org
Replying to @rcrehuet @ioChem_BD
@rcrehuet, we have certainly a lot of learning to do! Thanks for the recommendation :)
Jul 24, 2017
MDbox @mdbox_org
Hello, World!
Jul 18, 2017
Support
Australian National University
ACT Government
LINKDIGITAL
NCI
ckan
3
Datasets
0
Group
9
Users
2
Organisations
Terms of Service
Privacy Policy
mdbox
www.mdbox.org

Organisations

Australian National... (2)

Groups

There are no Groups that match this search

Tags

tubulin (1)
test (1)
gromos54a7 (1)
gromacs (1)
equilibrium (1)
NAMD (1)
GlyT2 (1)
GDP state (1)
FimH (1)
CHARMM36 (1)

Show More Tags

Formats

PDB (3)
mdp (2)
xtc (1)
tar.gz (1)
PSF (1)
NAMD configuration (1)
GROMACS trajectory (1)
GROMACS system topo... (1)
GROMACS include top... (1)
GROMACS gro (1)

Show More Formats

Licenses

Creative Commons No... (2)
GNU Free Documentat... (1)

Molecule

protein (3)

Search trajectories...

3 datasets found

Order by: Relevance

GlyT2 S1-gly S2-gly equilbrium

Equilibrium simulation of homology model of GlyT2 in POPC bilayer with NaCl water box. Homology model from: https://doi.org/10.1371/journal.pone.0157583. Zwitterion glycine in...

GROMACS gro | mdp | GROMACS system topology | GROMACS trajectory | PDB

FimH 4xoe DsG CHARMM36 NAMD equilibrium

Equilibrium simulation of mannose bound, associated state of FimH/DsG from H2 E. coli. Water box of 130x130x130 A contains 69497 water molecules that was charged neutralised...

PDB | PSF | NAMD configuration | DCD trajectory

Test dataset

This is a test data set.

tar.gz | GROMACS include topology | PDB | mdp | xtc

You can also access this registry using the API (see API Docs).

GlyT2 S1-gly S2-gly e quilbrium

Followers
0

Organisation

Australian National University

Australian National U niversity

Canberra, Australia read more

Social

Google+

Twitter

Facebook

License

Creative Commons Non-Commercial (Any)

Dataset | Groups | Activity Stream

GlyT2 S1-gly S2-gly equilbrium

Equilibrium simulation of homology model of GlyT2 in POPC bilayer with NaCl water box. Homology model from: https://doi.org/10.1371/journal.pone.0157583. Zwitterion glycine in S1 and S2. Gly in S2 not stable.

Data and Resources

Initial coordinates - gro	Explore
GROMACS configuration file	Explore
Topology file	Explore
GROMACS trajectory file - xtc	Explore
Initial coordinates - pdb	Explore

GlyT2 gromacs gromos54a7

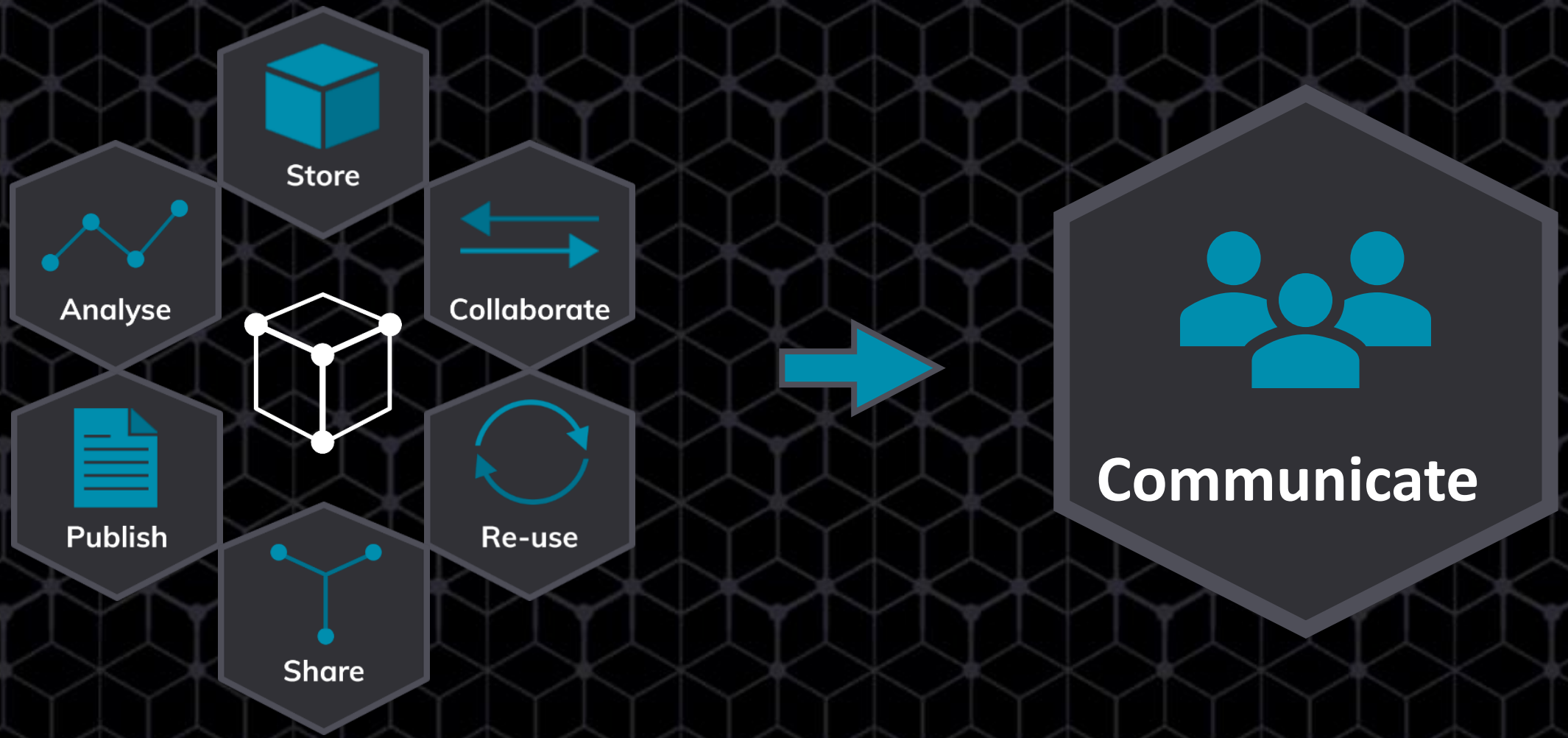
Additional Info

Field	Value
Does your system contain any non-standard residues?	No
Source	
Version	
Are you the author of this dataset?	No
Author	Michael Thomas
Author Email	explodingdinosaurs@gmail.com
Molecule	Protein
Software	GROMACS
Software version	5.1.2
Ion Type	
Ion count	
Solvent Molecule	
Atom Count	
Number of tasks	0
Method name	md
Simulated time(s)	1e-07
Barostats	Berendsen
Thermostats	v-rescale
Electrostatics	pme

community feedback

- **EASY to use:**
  - command line data transfer (background process)
  - automated metadata generation
- browser supported visualisation tools
- integration with computational facilities
- preferably free! (sustainable)

long road ahead…



acknowledgments

www.mdbox.org

- Link Digital
  - Mark Gregson (NCI)
  - Greg von Nessi
  - Steven de Costa
- National Computational Infrastructure (NCI)
  - Rika Kobayashi
  - Claire Trenham (CSIRO)
- MD group @ RSC ANU:
  - Megan O'Mara
  - Michael Thomas

thank you!

@mdbox_org
@karmecon

Australian National University

ACT Government

LINKDIGITAL

NCI

NATIONAL COMPUTATIONAL INFRASTRUCTURE

ckan