Institut Català d'Investigació Química

Taming Computational Chemistry Data: ioChem-BD and Beyond

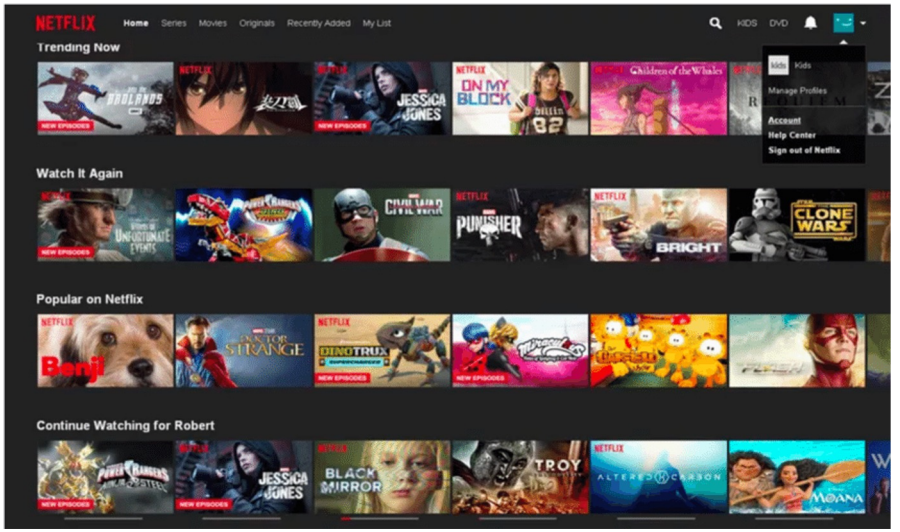
Carles Bo
Institute of Chemical Research of Catalonia (ICIQ)
Tarragona

Research Data Management in HPC
#HPC.NRW
14/5/2024

The NETFLIX disruption

HOW IT STARTED

HOW IT'S GOING



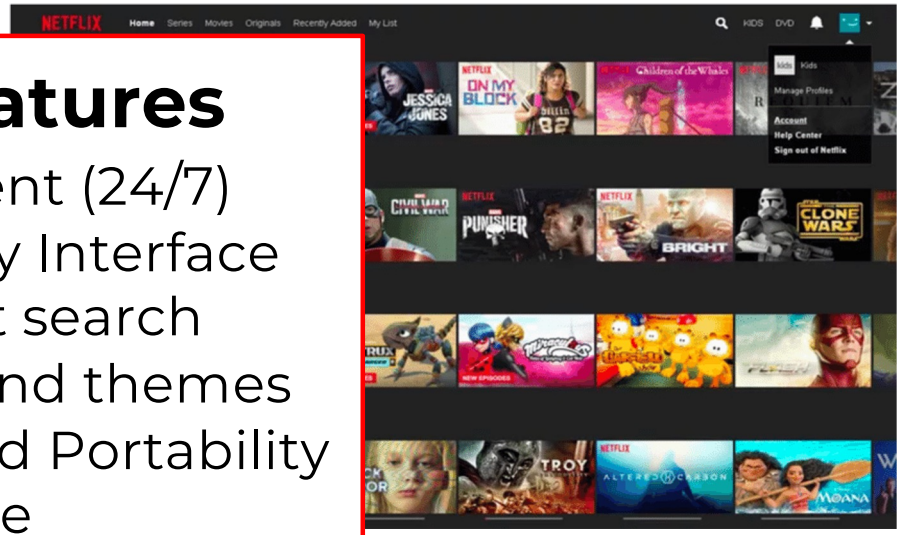
The NETFLIX disruption

HOW IT STARTED

Killer features

- Online content (24/7)
- User-Friendly Interface
- Easy content search
- Categories and themes
- Flexibility and Portability
- Cost-Effective

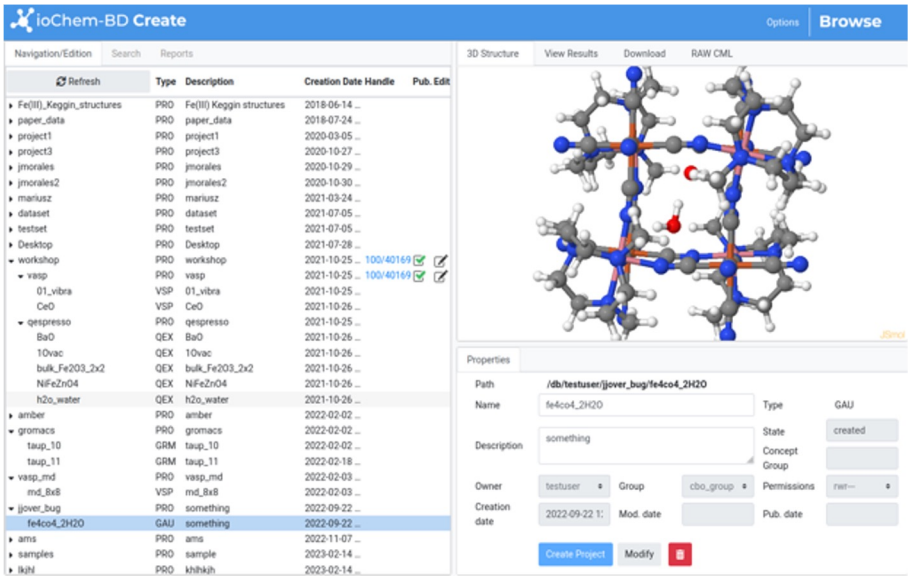
HOW IT'S GOING



The ioChem-BD disruption

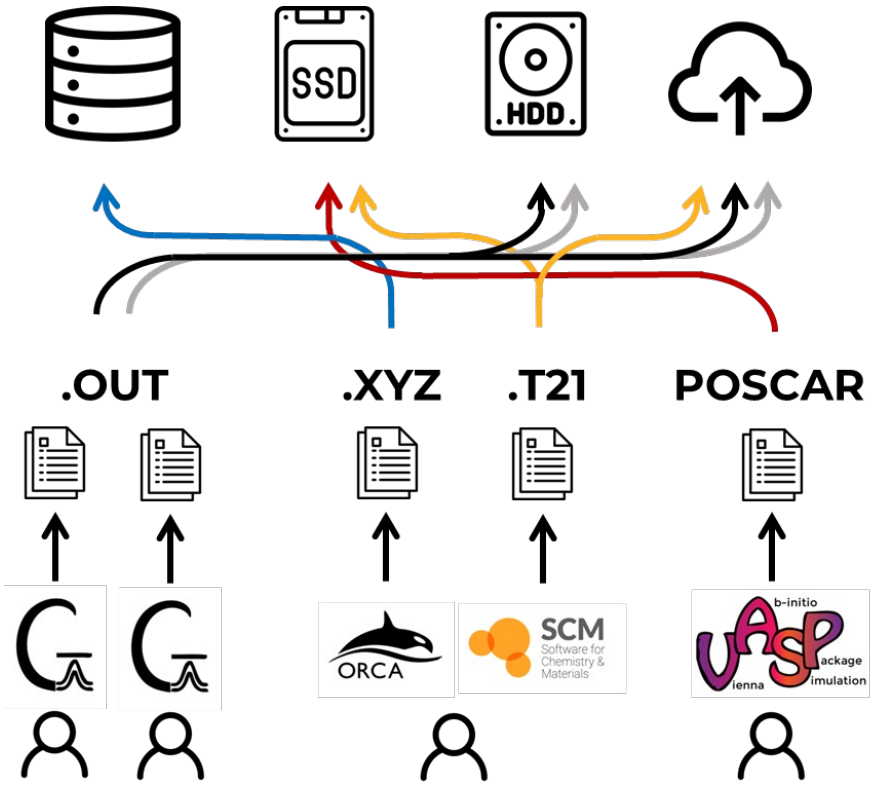
HOW IT STARTED

HOW IT'S GOING

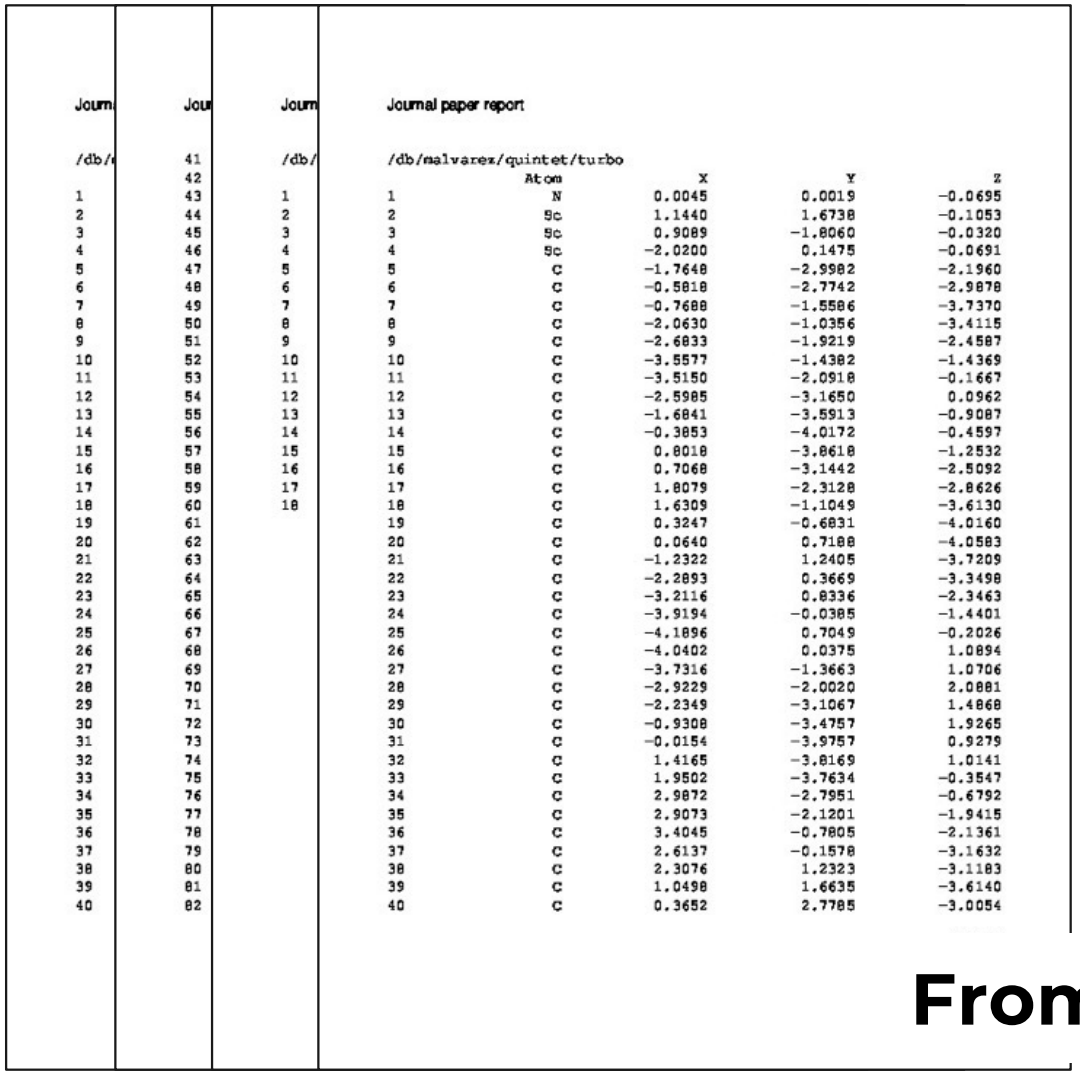


The ioChem-BD disruption

HOW IT STARTED



HOW IT'S GOING



Journal paper report

/db/malvarez/quintet/turbo

	Atom	X	Y	Z
1	N	0.0045	0.0019	−0.0695
2	Sc	1.1440	1.6738	−0.1053
3	Sc	0.9089	−1.8060	−0.0320
4	Sc	−2.0200	0.1475	−0.0691
5	C	−1.7648	−2.9982	−2.1960
6	C	−0.5818	−2.7742	−2.9878
7	C	−0.7688	−1.5586	−3.7370
8	C	−2.0630	−1.0356	−3.4115
9	C	−2.6833	−1.9219	−2.4587
10	C	−3.5577	−1.4382	−1.4369
11	C	−3.5150	−2.0918	−0.1667
12	C	−2.5985	−3.1650	0.0962
13	C	−1.6841	−3.5913	−0.9087
14	C	−0.3853	−4.0172	−0.4597
15	C	0.8018	−3.8618	−1.2532
16	C	0.7068	−3.1442	−2.5092
17	C	1.8079	−2.3128	−2.8626
18	C	1.6309	−1.1049	−3.6130
19	C	0.3247	−0.6831	−4.0160
20	C	0.0640	0.7188	−4.0583
21	C	−1.2322	1.2405	−3.7209
22	C	−2.2893	0.3669	−3.3498
23	C	−3.2116	0.8336	−2.3463
24	C	−3.9194	−0.0385	−1.4401
25	C	−4.1896	0.7049	−0.2026
26	C	−4.0402	0.0375	1.0894
27	C	−3.7316	−1.3663	1.0706
28	C	−2.9229	−2.0020	2.0881
29	C	−2.2349	−3.1067	1.4868
30	C	−0.9308	−3.4757	1.9265
31	C	−0.0154	−3.9757	0.9279
32	C	1.4165	−3.8169	1.0141
33	C	1.9502	−3.7634	−0.3547
34	C	2.9872	−2.7951	−0.6792
35	C	2.9073	−2.1201	−1.9415
36	C	3.4045	−0.7805	−2.1361
37	C	2.6137	−0.1578	−3.1632
38	C	2.3076	1.2323	−3.1183
39	C	1.0498	1.6635	−3.6140
40	C	0.3652	2.7785	−3.0054

From hundreds of pages...

Journal paper supporting information in traditional format

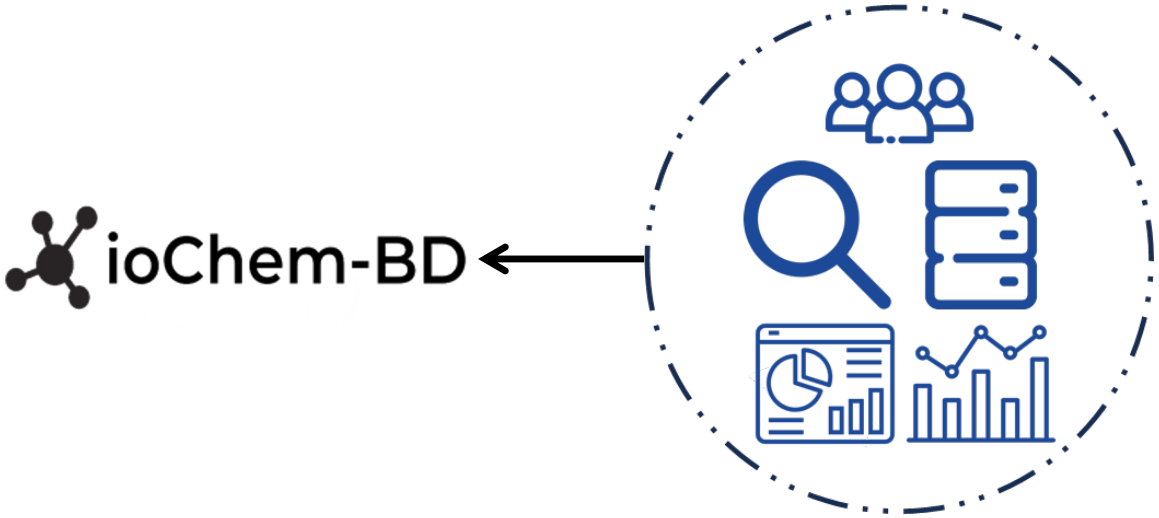
… to a single link:

https://doi.org/10.19061/iochem-bd-1-129


Open research data
Open data
Open science
Data management plans
Research data
Data management
FAIR
DOI
DataCite
DataDOI
OpenAIRE
Open Access
Repository
Knowledge
Publishing
Archiving
Preservation
Openess
Reuse
FAIRRest
Storage
Researchers
University of Tartu Library

R&D Challenges

- Share
- Search
- Store
- Visualize
- Analyse

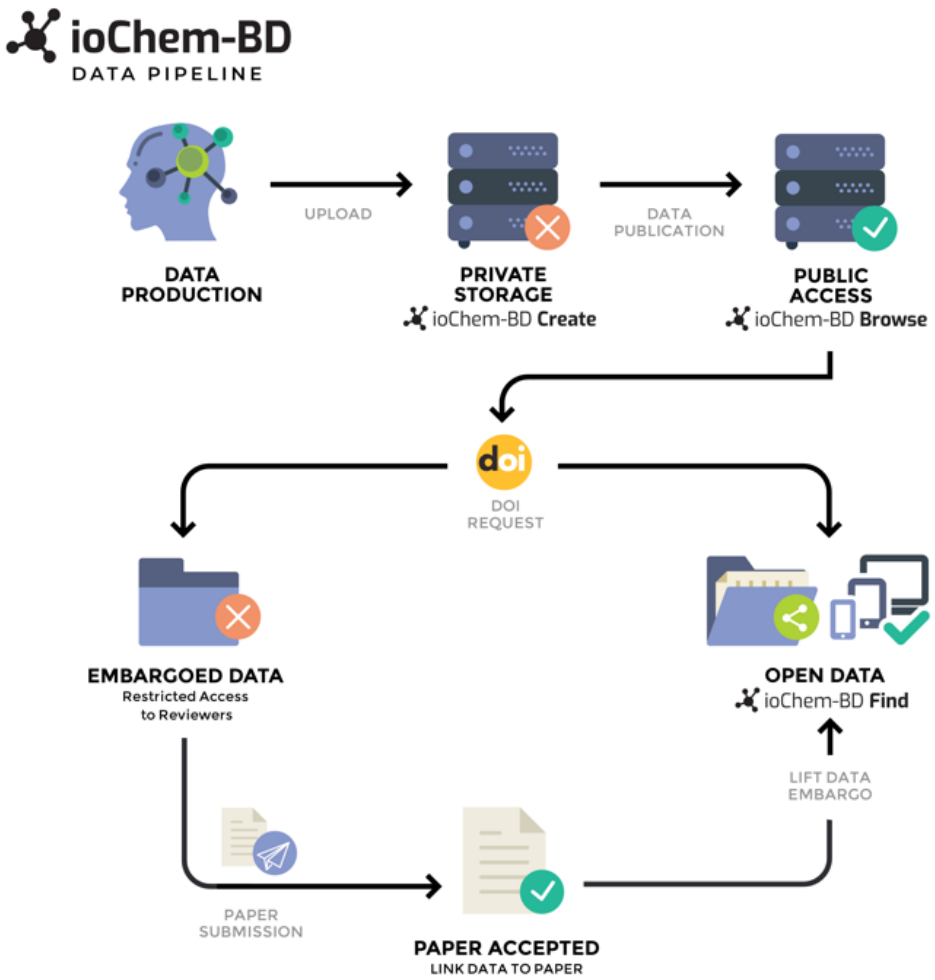


Findable, **A**ccessible, **I**nteroperable and **R**eusable data

FAIR data for new collaborative Data-Driven research

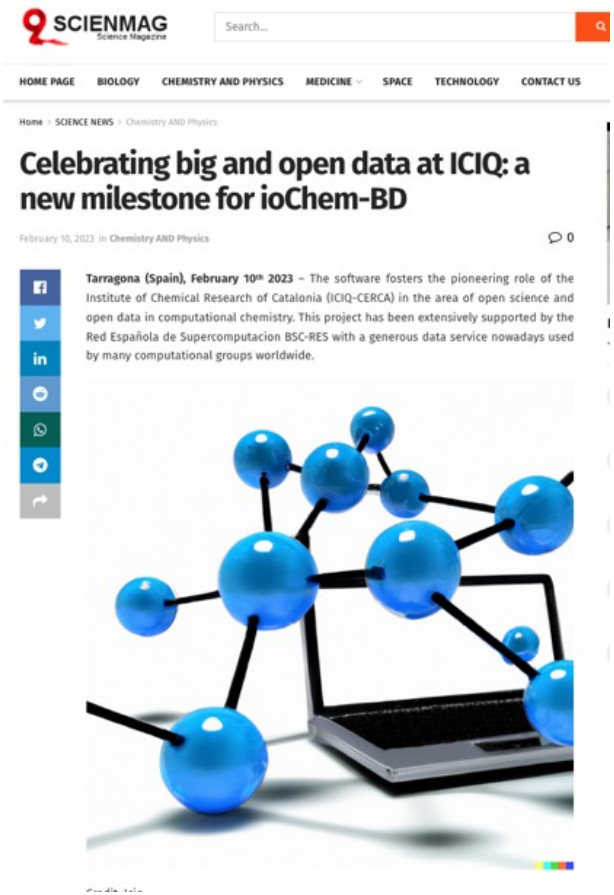
R&D Solutions

- Storage reduction
- Data Organization
- Easy retrieval
- Profitable data
- Adaption to standards



About the platform

- Manage private/public data
- 10+ years' experience
- International Acceptance
- Agreements with companies



About the platform

2011

It all started as a distributed digital repository

of CompChem and material science results.

Initial set of web services

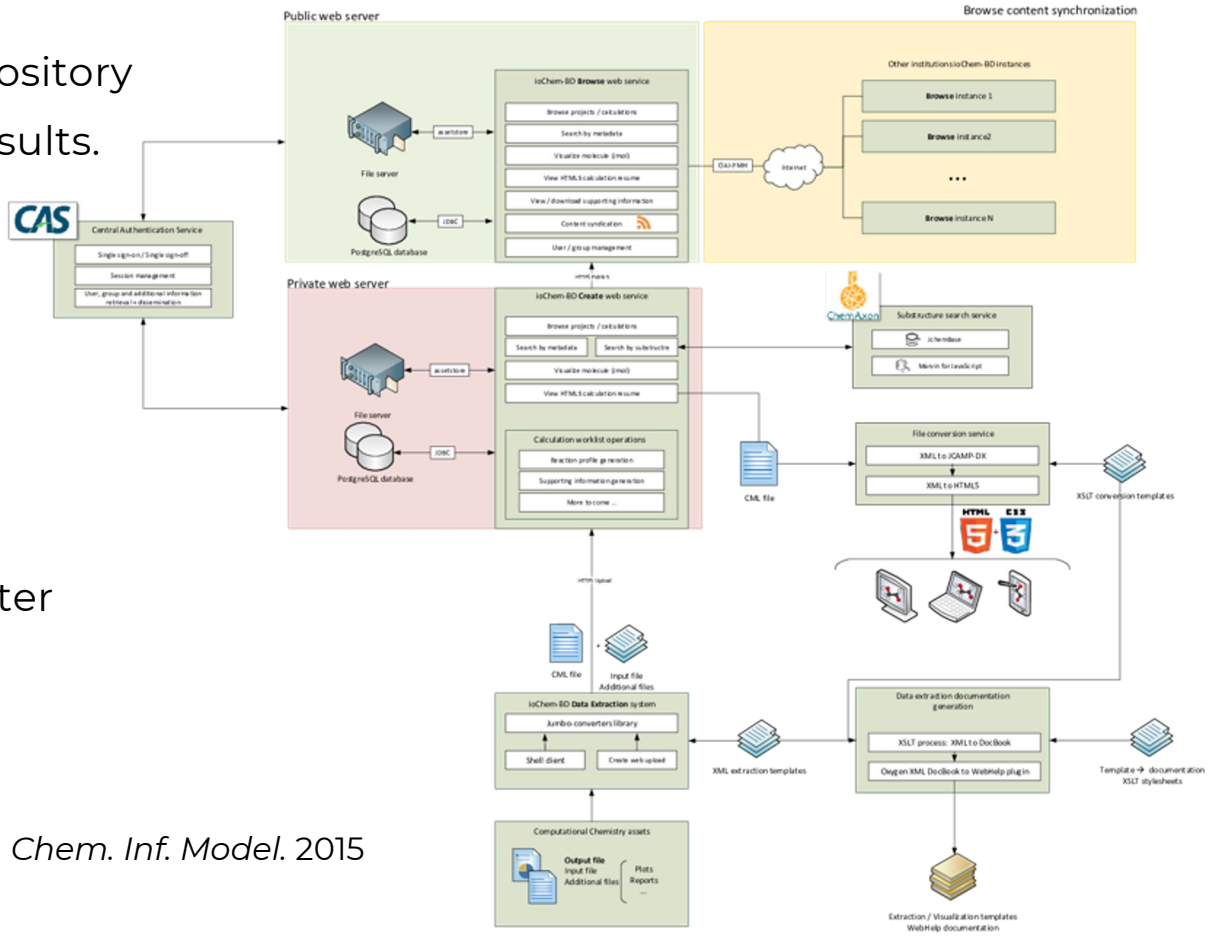
- Private space (Create module)
- Public portal (Browse module)
- Identity portal (CAS)

2017

Central node with all public assets:

- Main service Find module
- Barcelona Supercomputer Center

www.iochem-bd.org

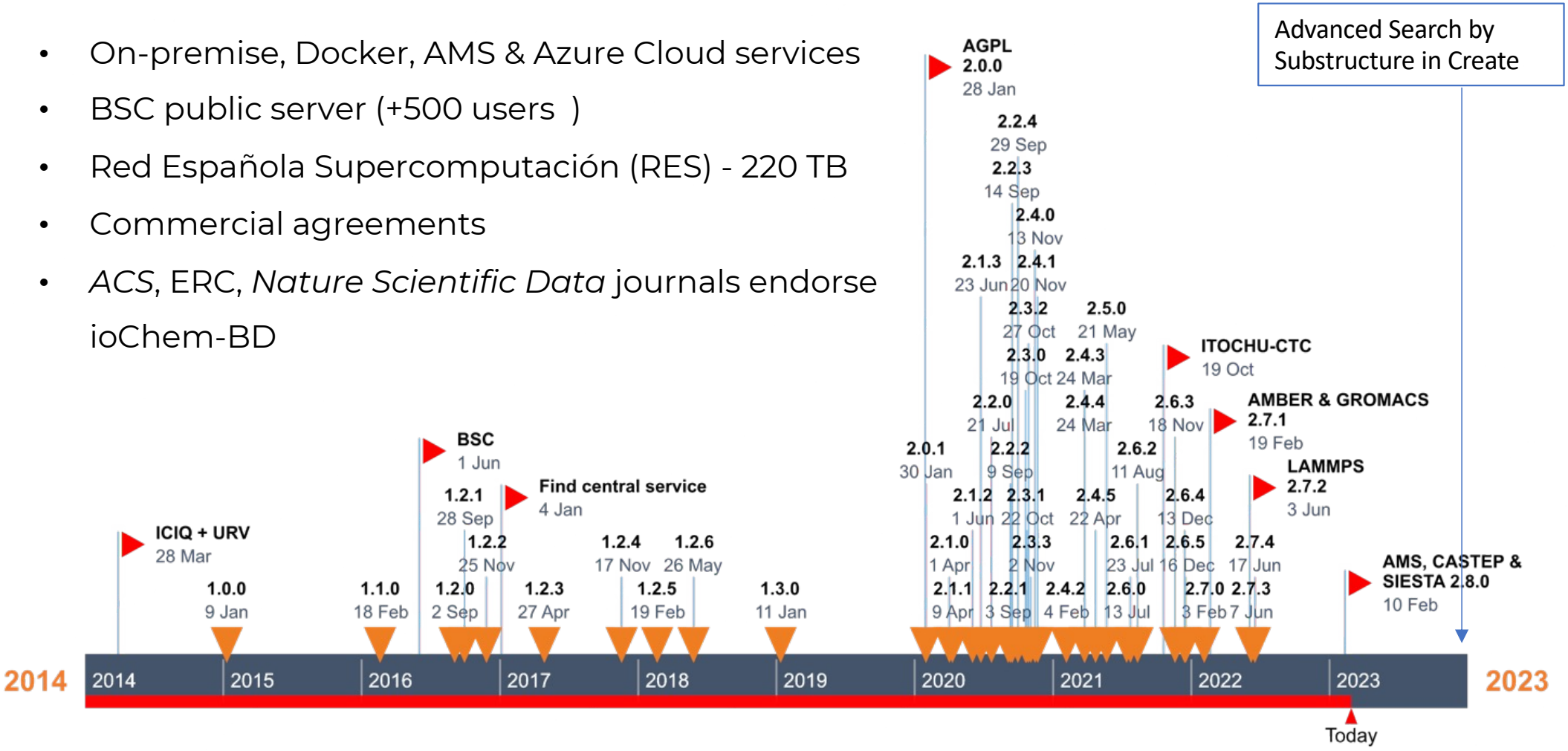


Álvarez-Moreno, de Graaf, López, Maseras, Poblet, Bo, *J. Chem. Inf. Model.* 2015

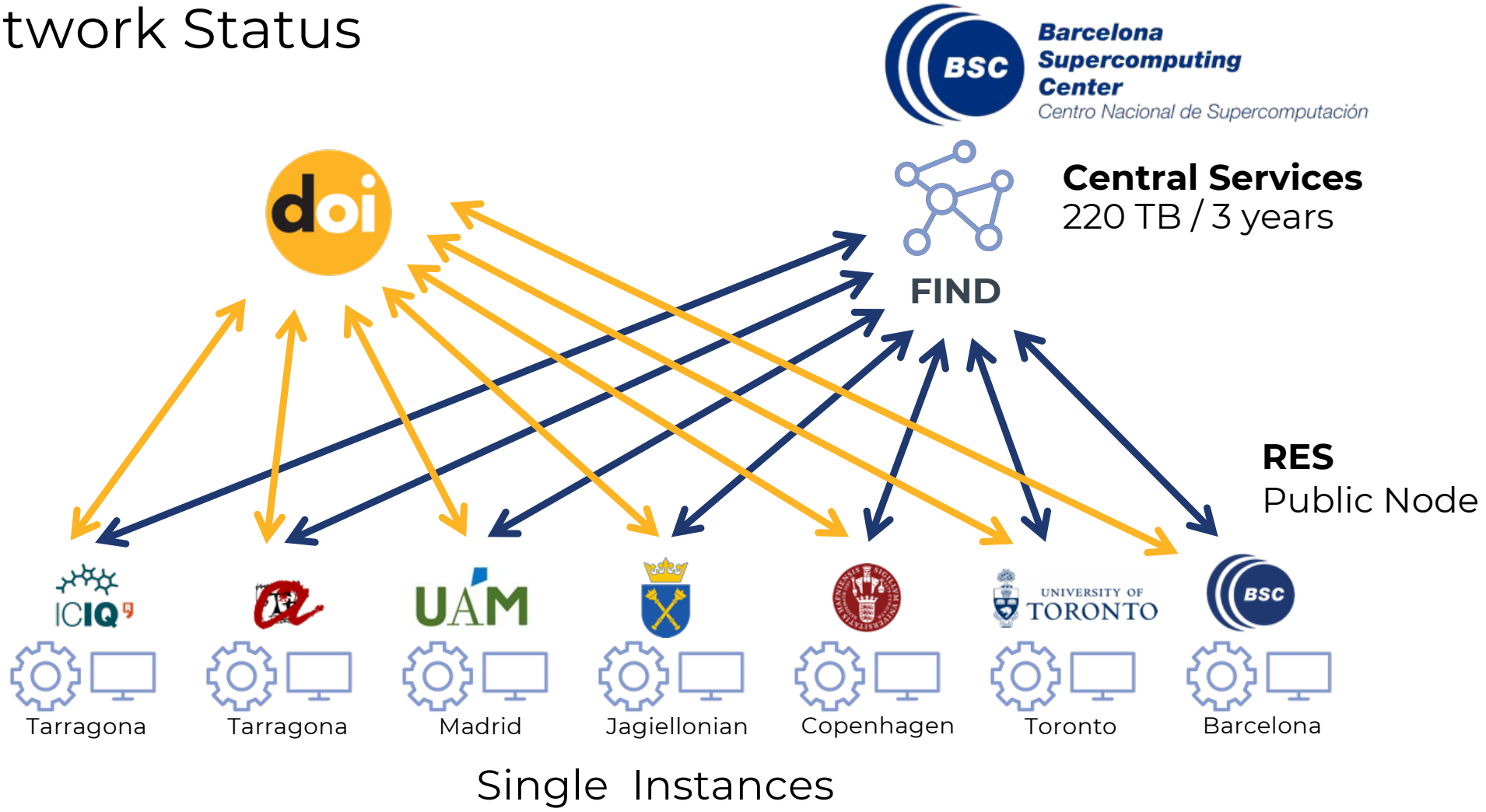
Bo, Maseras, López, *Nat. Catal.* 2018

Today

- On-premise, Docker, AMS & Azure Cloud services
- BSC public server (+500 users)
- Red Española Supercomputación (RES) - 220 TB
- Commercial agreements
- *ACS*, ERC, *Nature Scientific Data* journals endorse ioChem-BD

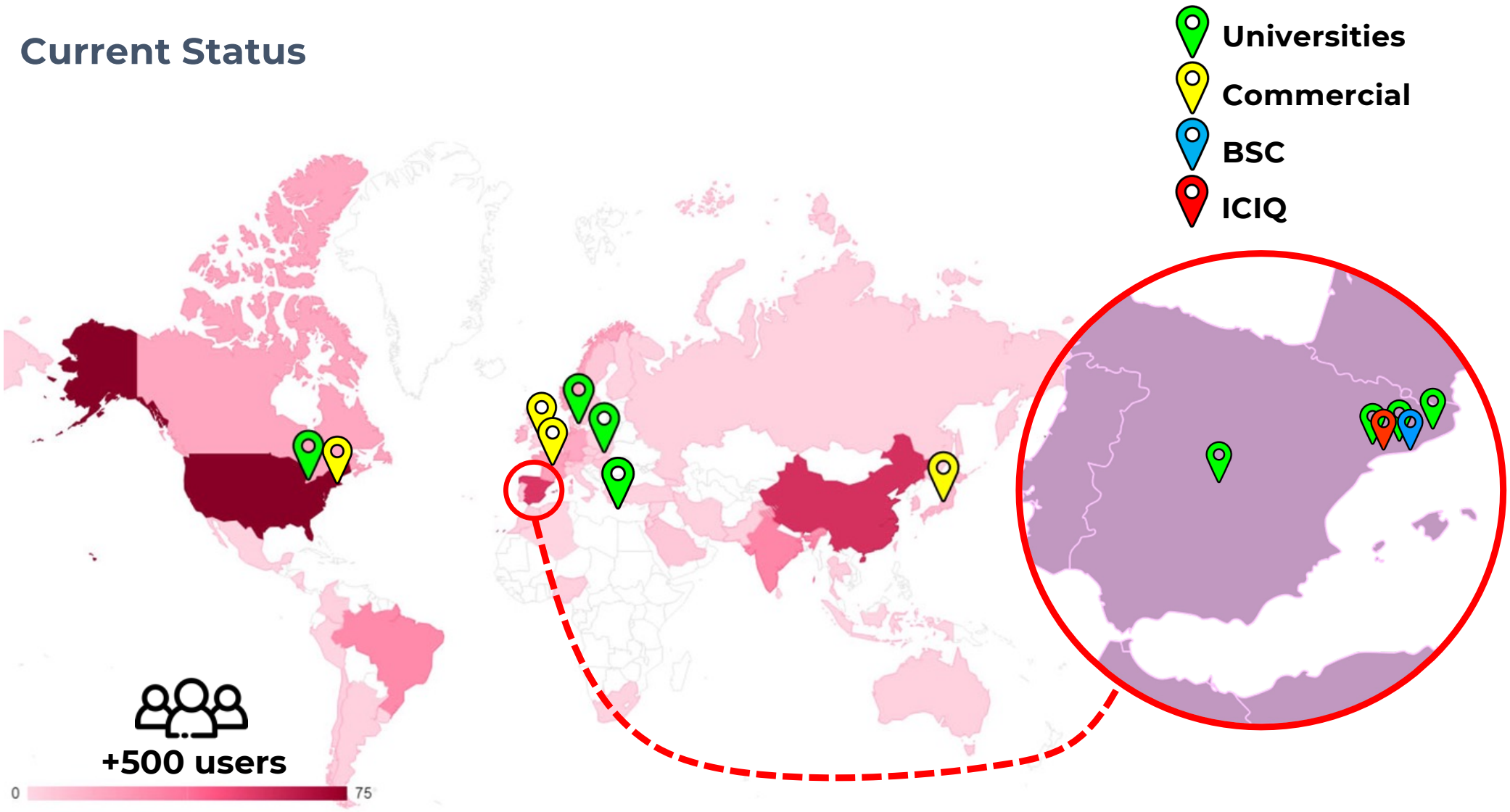


Network Status



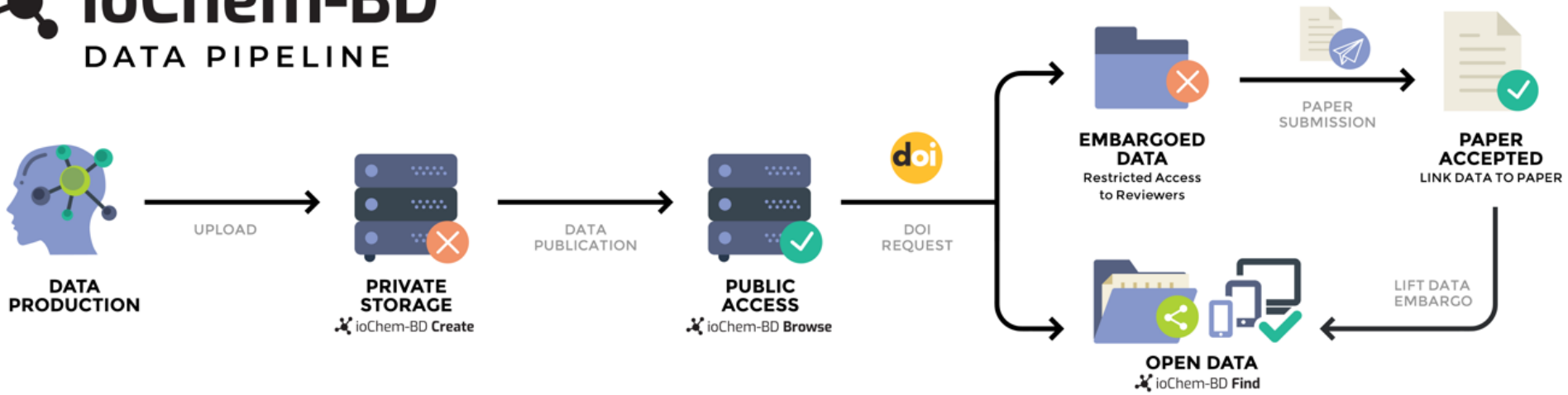
Single Instances

Current Status



ioChem-BD

DATA PIPELINE



CREATE

BROWSE

FIND

ioChem-BD **Create**

... the entry point

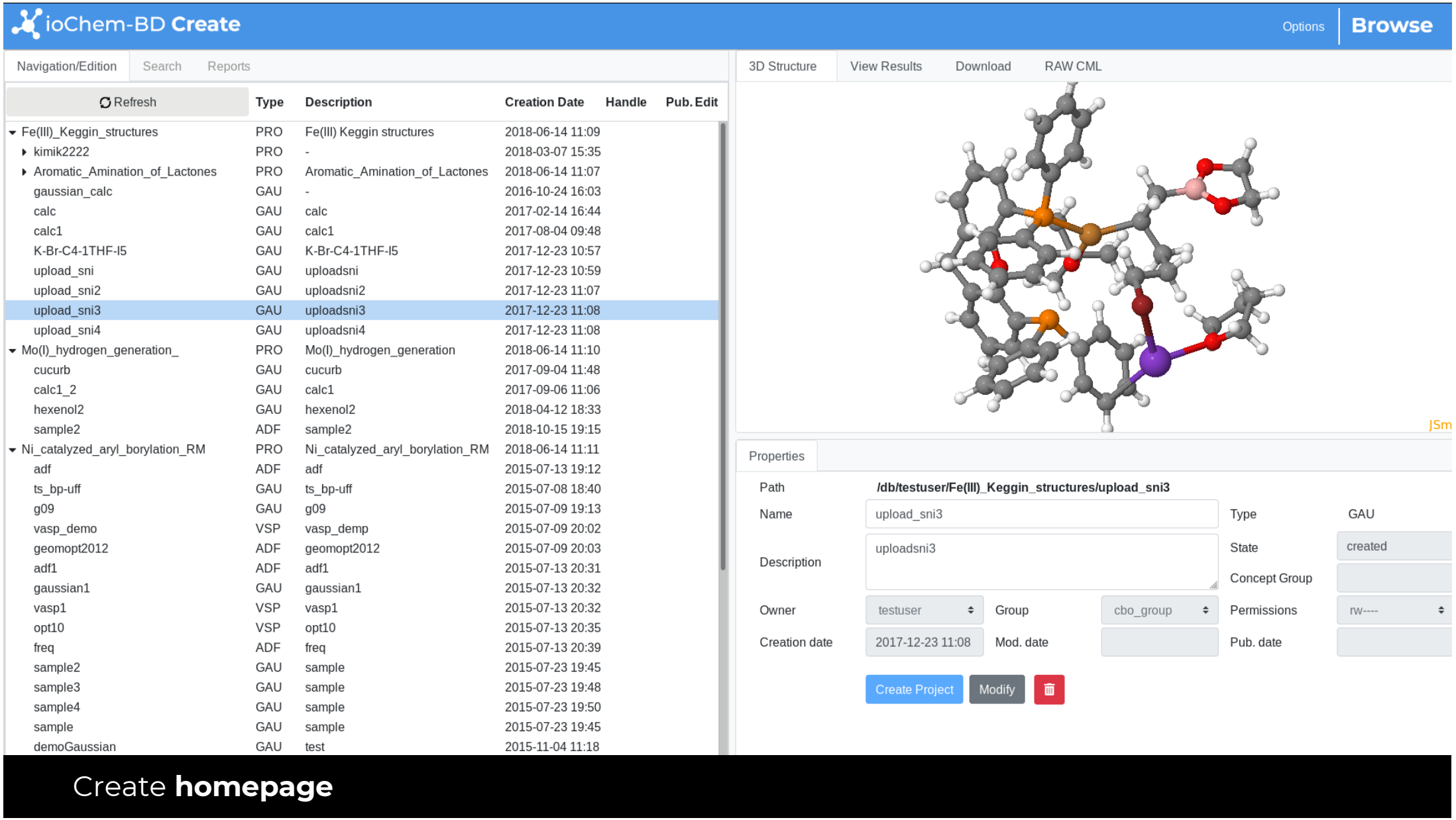
Web service / Desktop App, user-oriented, aimed at handling and managing your files on a daily basis

Features:

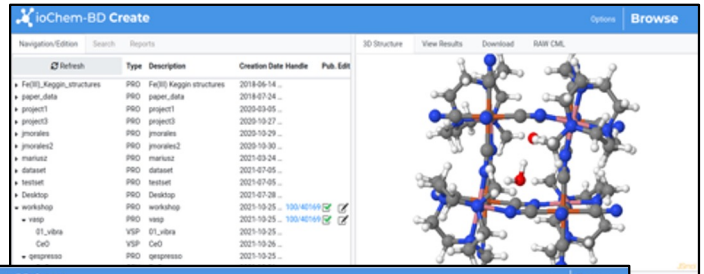
- Private storage area (private server, logged users only)
- Compact and intuitive interface design: Single page application (SPA)
- Multiple visualization tools: JSmol, JCamp-DX and HighCharts
- Search by metadata / Users / Groups
- Reports-generation tools

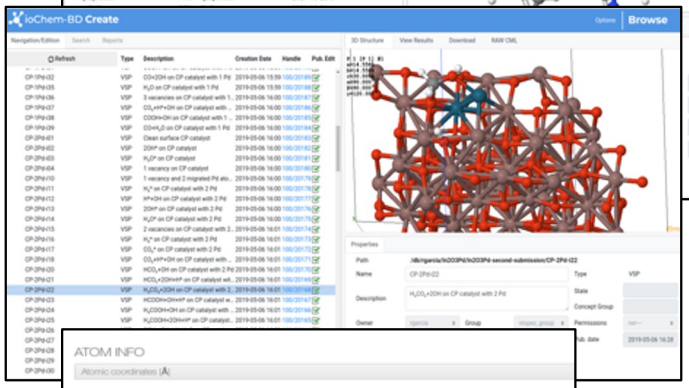
External services:

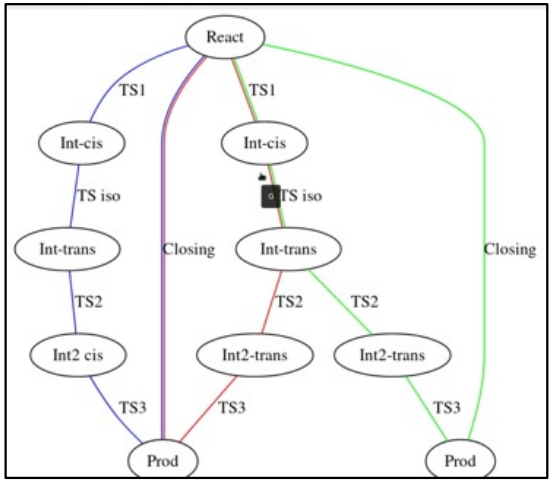
- Shell client interface for task automation
- REST API interface enables remote interaction and process automation (experimental)

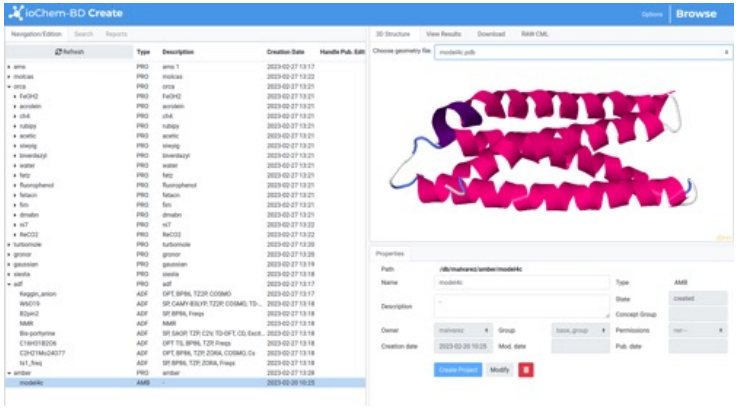


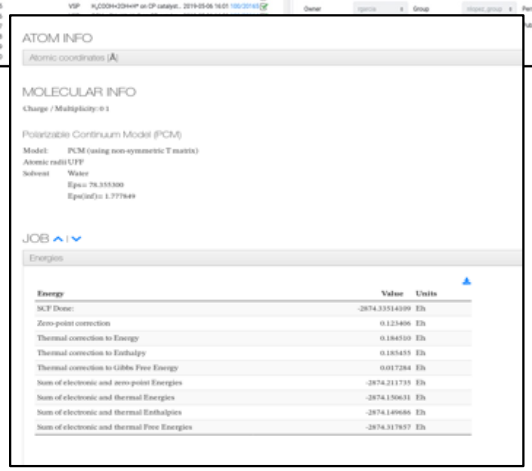
Create **homepage**

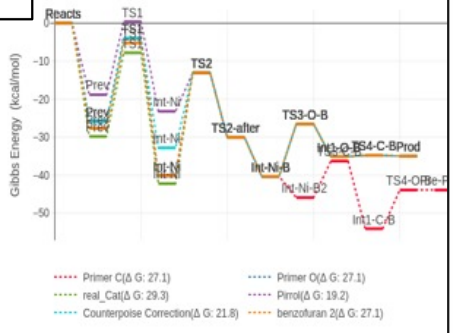

ioChem-BD Create
Browse

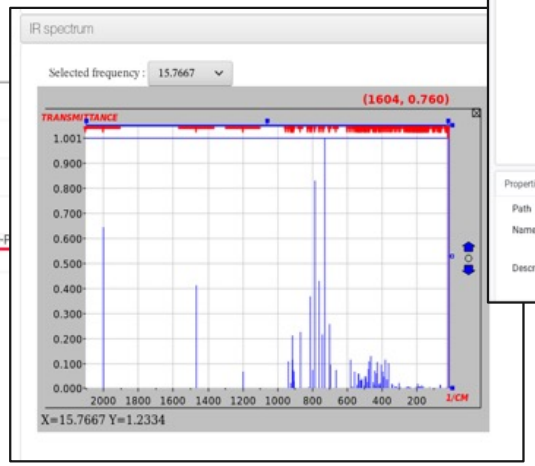

ioChem-BD Create
Browse

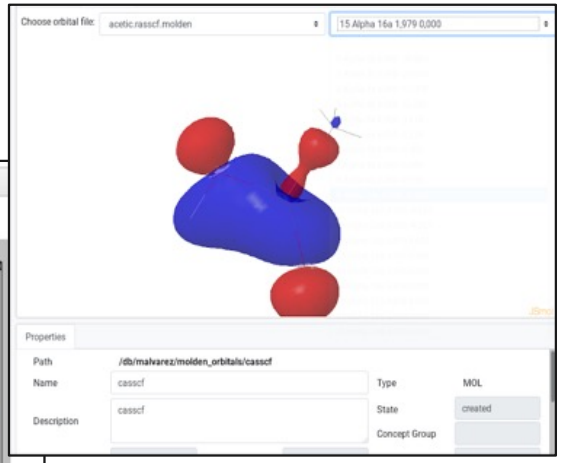

React
TS1
TS1
Int-cis
Int-cis
TS iso
TS iso
Int-trans
Closing
Int-trans
Closing
TS2
TS2
TS2
Int2 cis
Int2-trans
Int2-trans
TS3
TS3
TS3
Prod
Prod


ioChem-BD Create
Browse


ATOM INFO
MOLECULAR INFO
Polarizable Continuum Model (PCM)
JOB
Energies

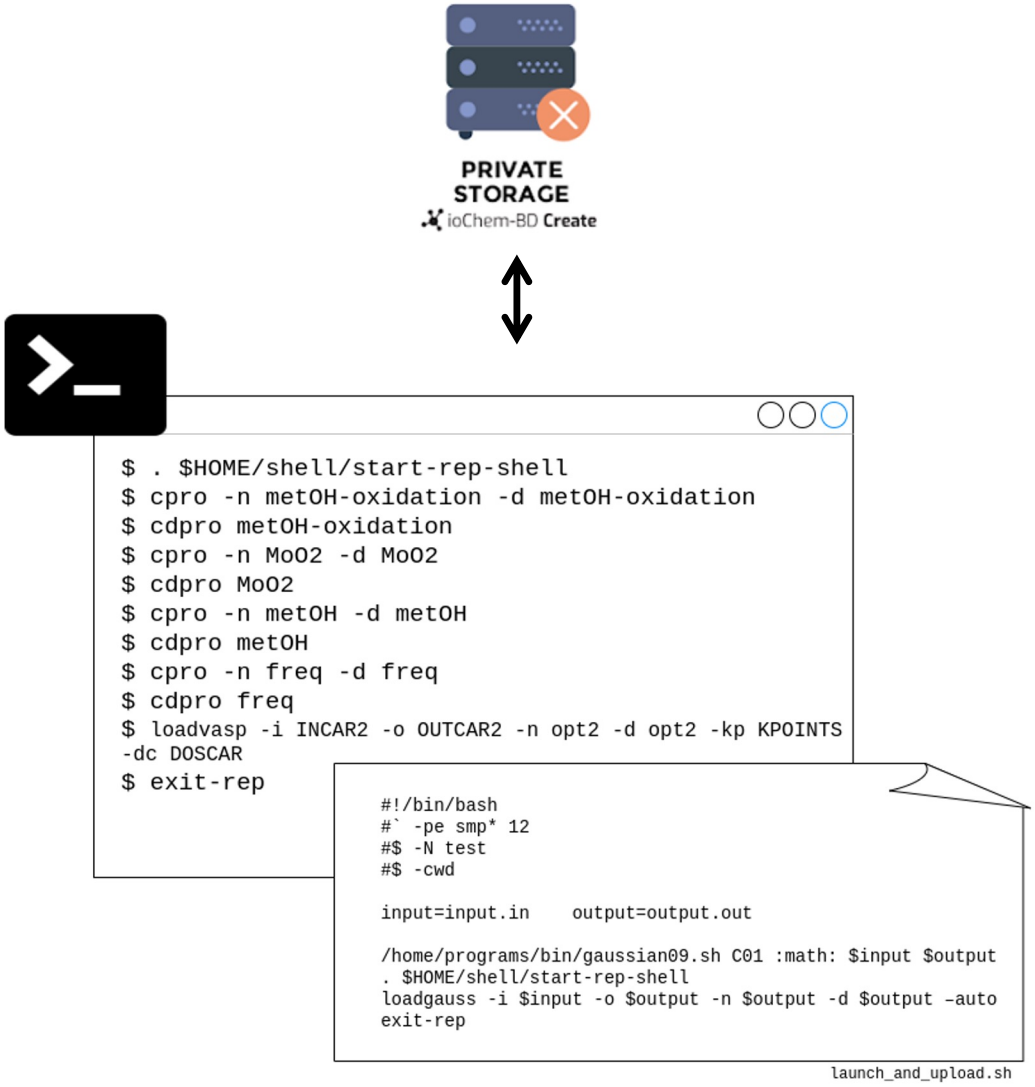

Reacts
TS1
TS2
TS2-after
TS3-O-B
Int-Ni-B
Int-Ni-B2
Int1-C-B
Gibbs Energy (kcal/mol)
Primer C(Δ G: 27.1)
real_Cat(Δ G: 29.3)
Counterpoise Correction(Δ G: 21.8)
Primer O(Δ G: 27.1)
Pirrol(Δ G: 19.2)
benzofuran 2(Δ G: 27.1)


IR spectrum
Selected frequency :
15.7667
(1604, 0.760)
TRANSMITTANCE
1/CM
X=15.7667 Y=1.2334


Choose orbital file:
acetic.rasscf.molden
15 Alpha 16a 1,979 0,000
Properties
Path
/db/malvarez/molden_orbitals/casscf
Name
casscf
Type
MOL
Description
casscf
State
created
Concept Group

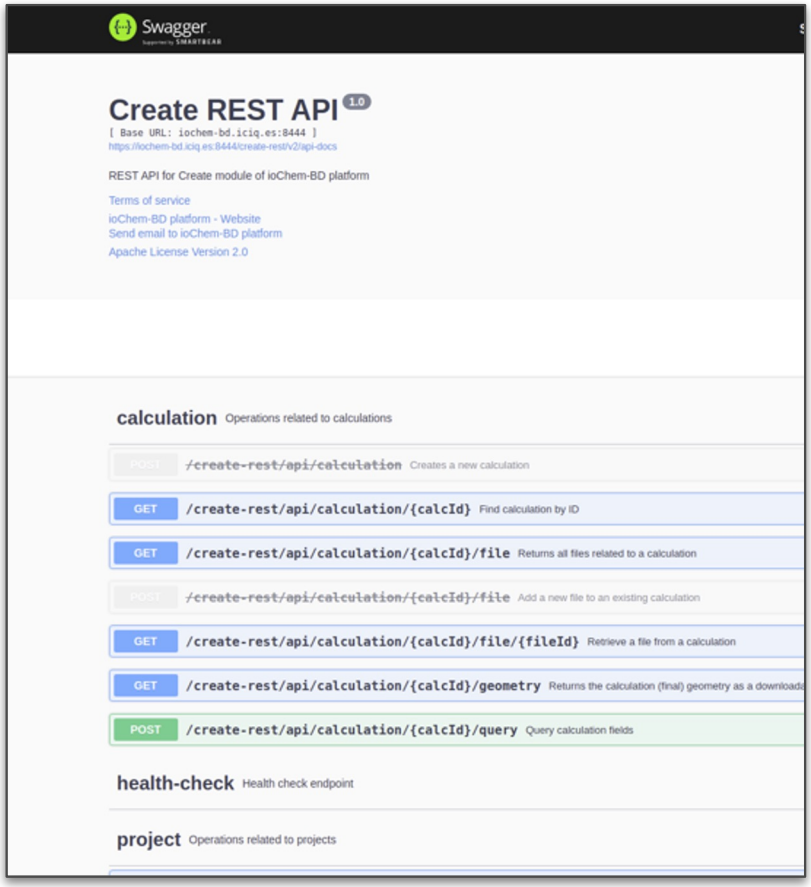
Create Shell client

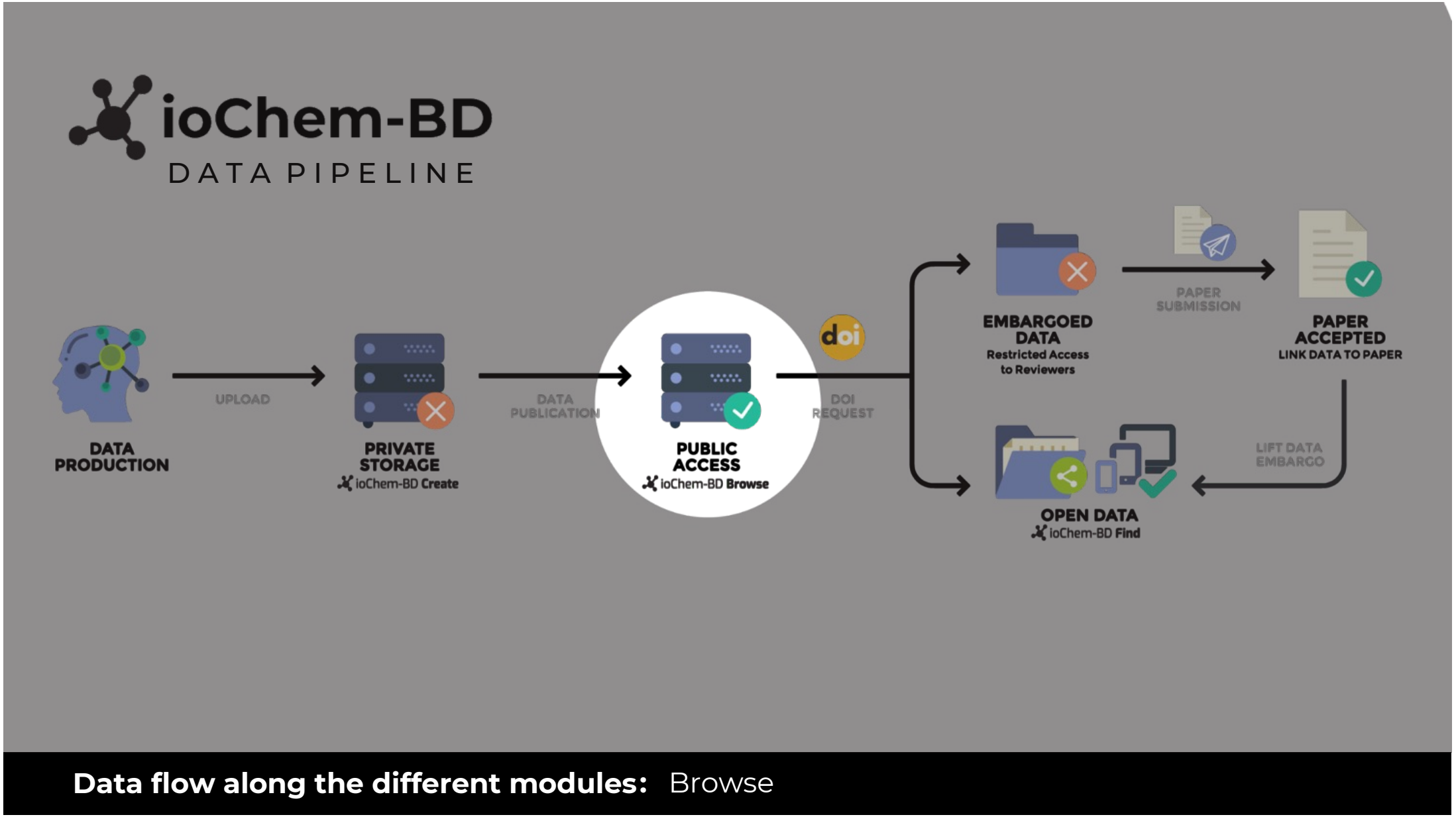
- Command line tool (Unix-like OSes)
- Operates ioChem-BD Create module
- Provides repository specific commands
- Integrable inside scripts for:
  - Large batch uploads or
  - Automatic upload after content creation



Create REST API

- Representational state transfer API (REST)
- Service is:
  - Stateless
  - Extensible
  - Service-oriented
  - Bundled with a uniform interface
- Enables querying and modifying data from a command line or automated service.
- Machine oriented
- Can be easily extended with new endpoints





Data flow along the different modules: Browse

ioChem-BD **Browse**

... the open repository

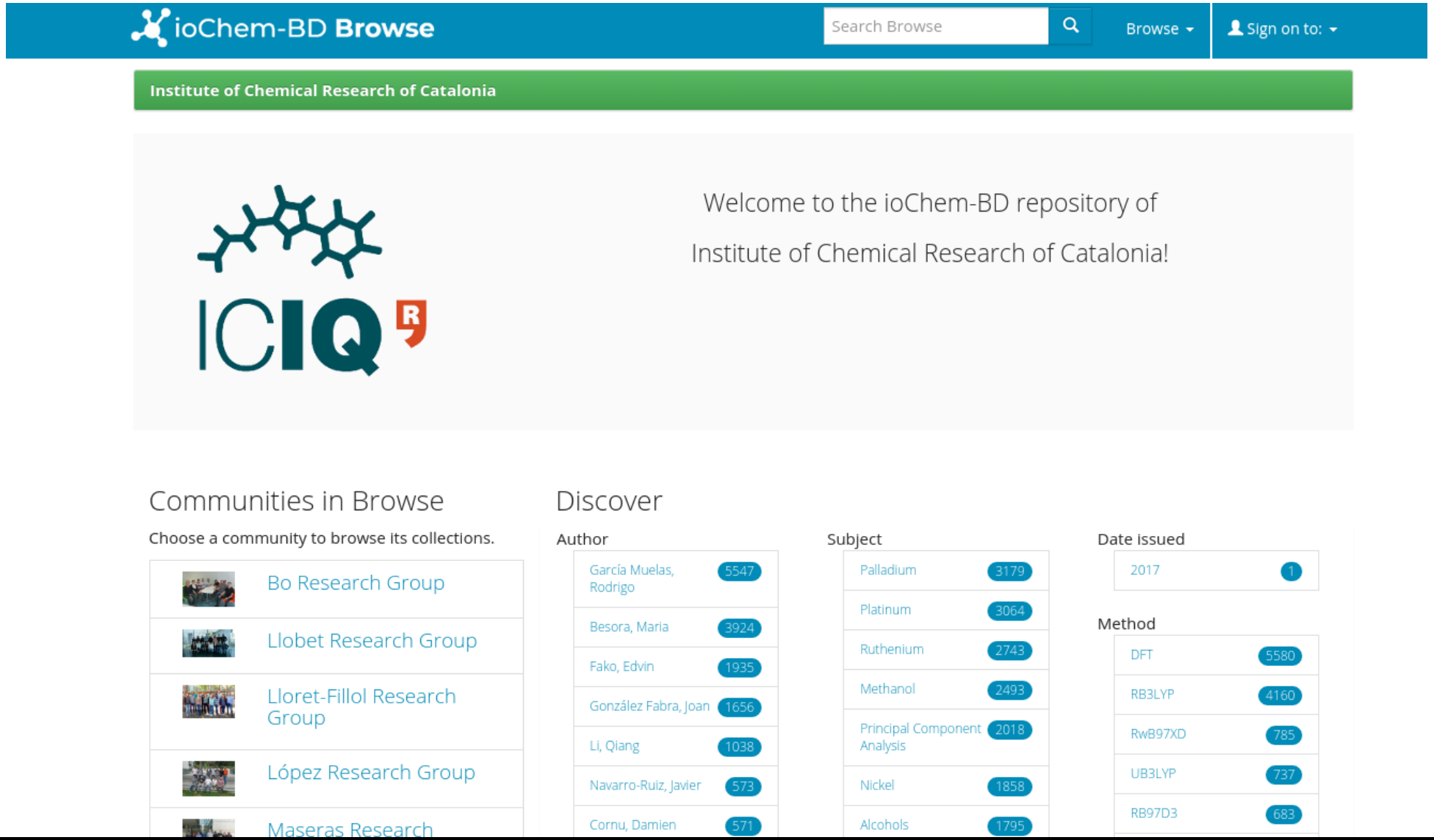
Open web portal where content published from the Create module is displayed.

Features:

- DOIs and permanent storage for users' published datasets
- Easy navigation across dataset collections
- Allows browsing by term and facet navigation
- Same visualization tools as Create: JSmol, JCamp-DX and HighCharts
- Search by multiple administrative and descriptive metadata fields
- Dspace allows easy content syndication, data replication and data harvesting

External services:

- Data services to offer public data to third party services and social networks.
- Multiple services and connectors: REST API, RDF, OAI-PMH



Browse homepage

Collections in this community

Alkali-Driven Disassembly and Reassembly of Molecular Niobium Oxide in Water DOI: 10.19061/iochem-bd-1-88

Alkali counterions for polyoxometalate (POM) clusters control solubility with distinct periodic trends, but evidence for alkali control over speciation is ambiguous. Here we show that a simple Nb-POM, [Nb10O28]6– ({Nb10}), converts to oligomers of (HxNb24O72)(24–x)– ({Nb24}) upon adding only alkali chloride salts, even in buffered neutral solutions.

This dataset derived results are published in:
Manuscript title: Alkali-Driven Disassembly and Reassembly of Molecular Niobium Oxide in Water
Journal: J. Am. Chem. Soc.
DOI: 10.1021/jacs.8b05015

Al(x)Cr(y)_Keggin_oxos DOI: 10.19061/iochem-bd-1-23

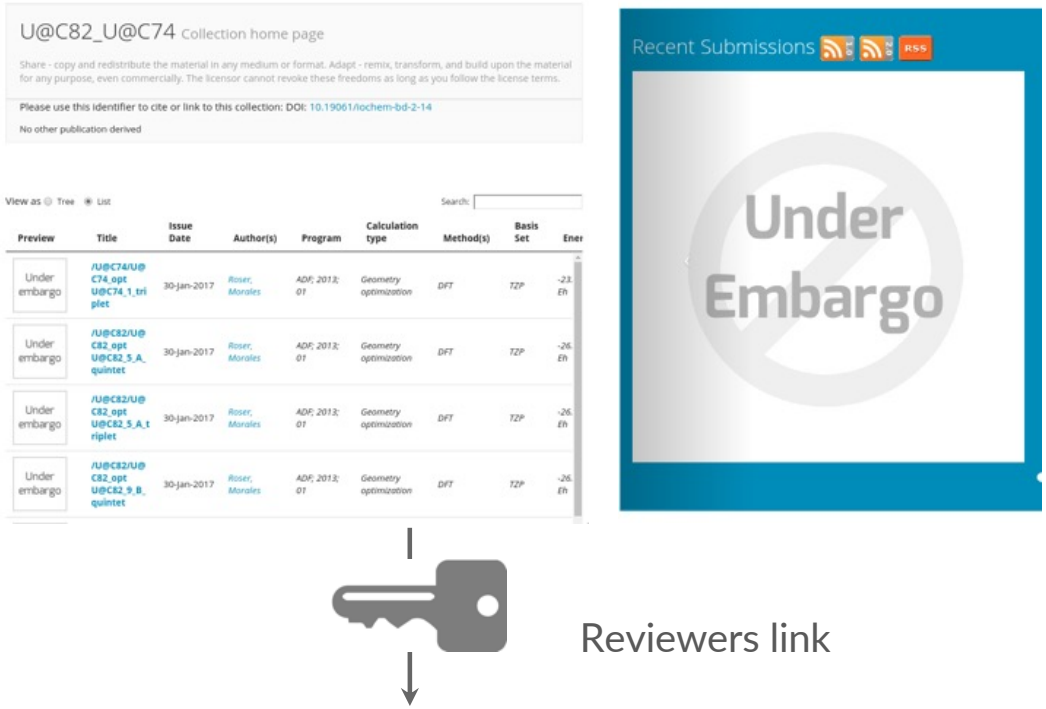
This dataset derived results are published in:
Manuscript title: Crystallizing Elusive Chromium Polycations.
Journal: Chem
DOI: 10.1016/j.chempr.2016.11.006

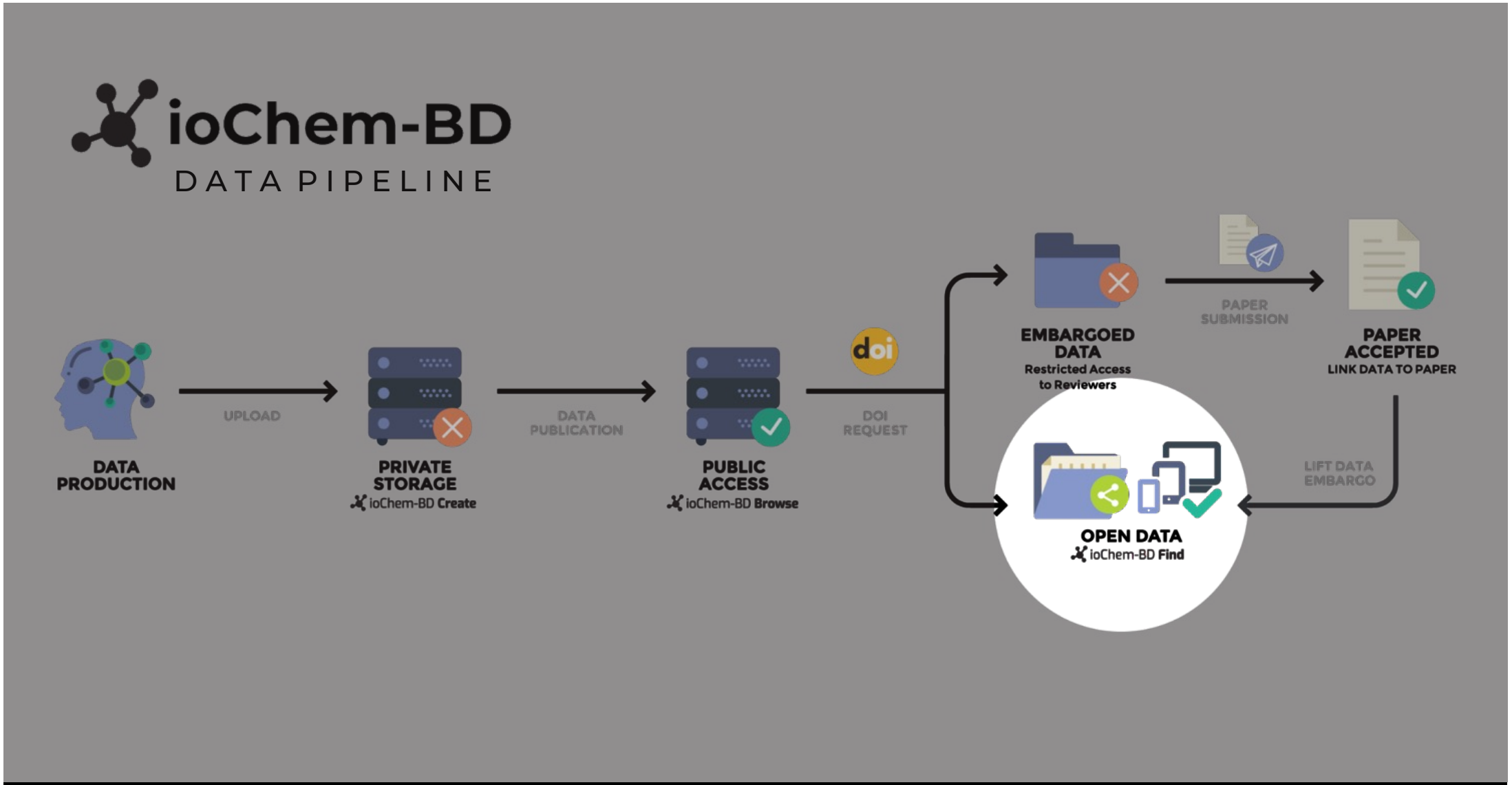
A Metal-Free Synthesis of N-Aryl Carbamates under Ambient Conditions DOI: 10.19061/iochem-bd-1-3

This dataset derived results are published in:
Manuscript title: A Metal-Free Synthesis of N-Aryl Carbamates under Ambient Conditions
Journal: Angew. Chem. Int. Ed.
DOI: 10.1002/anie.201504956

Analysis of the interaction between Keplerate nanocapsules and their anionic ligands DOI: 10.19061/iochem-bd-1-33

This dataset derived results are published in:
Manuscript title: Anions coordinating anions: Analysis of the interaction between anionic Keplerate nanocapsules and their anionic ligands
Journal: PCCP
DOI: 10.1039/C6CP08511C


U@C82_U@C74 collection home page
Share - copy and redistribute the material in any medium or format. Adapt - remix, transform, and build upon the material for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the license terms.
Please use this identifier to cite or link to this collection: DOI: 10.19061/iochem-bd-2-14
No other publication derived
View as Tree List
Search:
Preview
Title
Issue Date
Author(s)
Program
Calculation type
Method(s)
Basis Set
Under embargo
/U@C74/U@C74_opt U@C74_1_triplet
30-Jan-2017
Roser, Morales
ADF; 2013; 01
Geometry optimization
DFT
TZP
/U@C82/U@C82_opt U@C82_5_A_quintet
/U@C82/U@C82_opt U@C82_5_A_triplet
/U@C82/U@C82_opt U@C82_9_B_quintet
Recent Submissions
RSS
Under Embargo
Reviewers link
Institute of Chemical Research of Catalonia / Bo Research Group
/ Alkali-Driven Disassembly and Reassembly of Molecular Niobium Oxide in Water
Geometry:
Actions:
View data
Download geometry
Files in This Item:
File
Size
Format
NB10_Cs_C1_EHFH.in
3.11 kB
Gaussian Input format
Download
output.cml
588.98 kB
Chemical Markup Language
Download
Referenced by:
Manuscript title: Alkali-Driven Disassembly and Reassembly of Molecular Niobium Oxide in Water
Journal: J. Am. Chem. Soc.
DOI: 10.1021/jacs.8b05015



Data flow along the different modules: Find

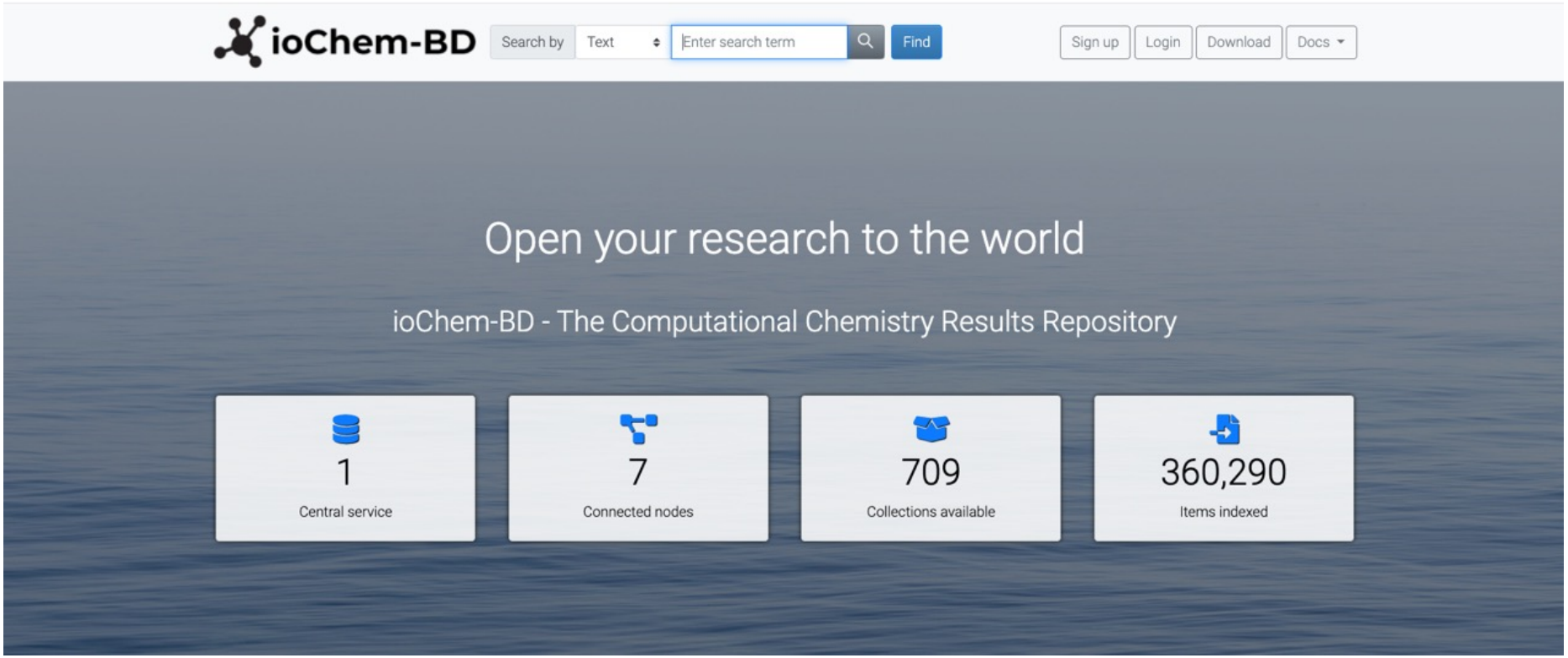
ioChem-BD **Find**

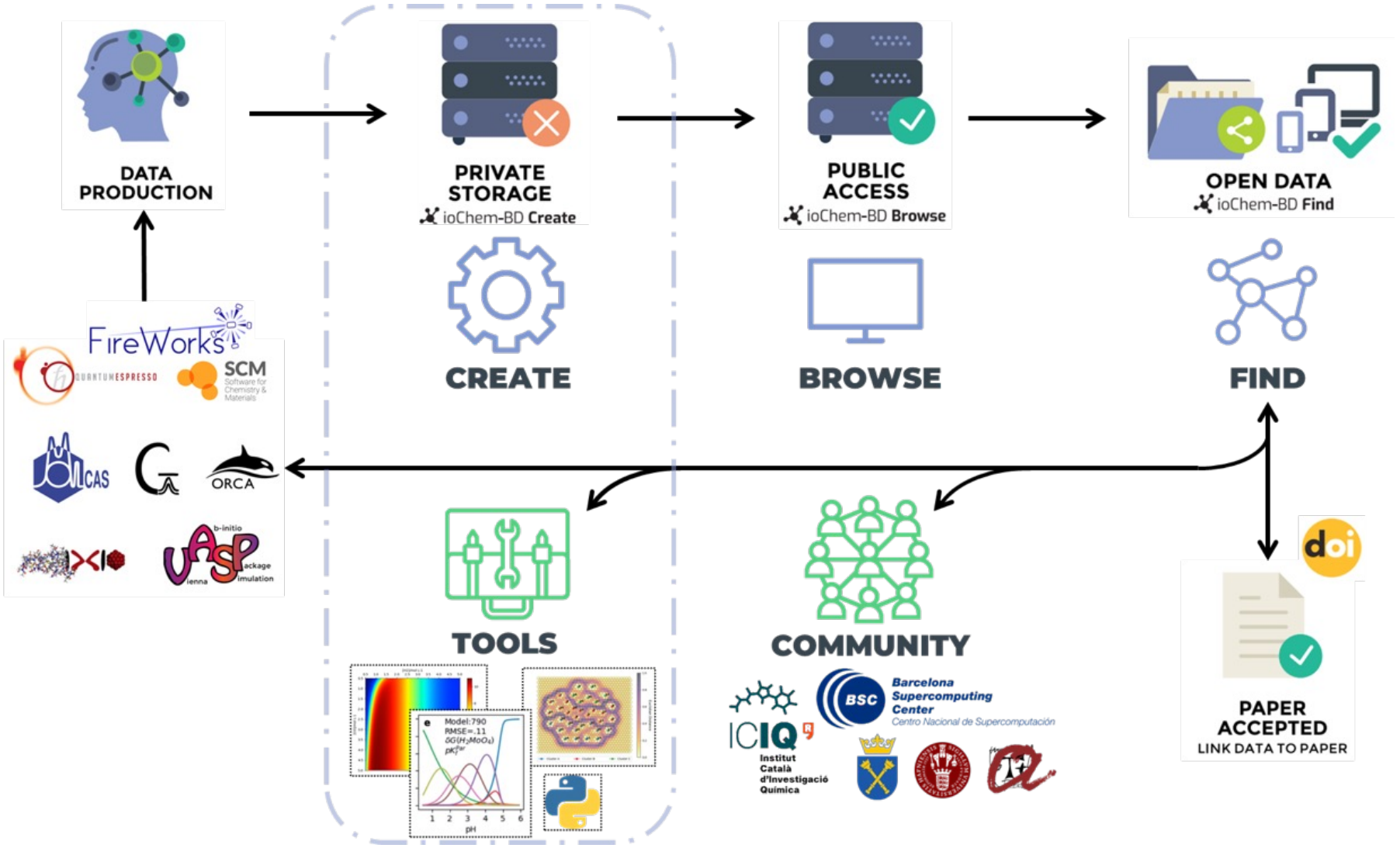
… or the distributed database

Central point and search engine

Features:

- Fed by ioChem-BD distributed nodes whenever a new item/collection is published
- Advanced indexing of chemical data and metadata (InChi, smiles)
- Fast search engine implemented over Apache Solr and RDKit frameworks
- Search by chemical substructure similarity, chemical elements or any text term


ioChem-BD
Search by
Text
Enter search term
Find
Sign up
Login
Download
Docs
Open your research to the world
ioChem-BD - The Computational Chemistry Results Repository
1
Central service
7
Connected nodes
709
Collections available
360,290
Items indexed


DATA PRODUCTION
PRIVATE STORAGE
ioChem-BD Create
PUBLIC ACCESS
ioChem-BD Browse
OPEN DATA
ioChem-BD Find
CREATE
BROWSE
FIND
FireWorks
QUANTUM ESPRESSO
SCM
Software for Chemistry & Materials
CP2K
ORCA
XTB
VASP
b-initio
ienna
ackage
imulation
TOOLS
COMMUNITY
BSC
Barcelona Supercomputing Center
Centro Nacional de Supercomputación
ICIQ
Institut Català d'Investigació Química
doi
PAPER ACCEPTED
LINK DATA TO PAPER

Projects: Data to “papers”

Storage

ioChem-BD **Browse**

+

Editing Webapp

AUTHOREA

ICIQ

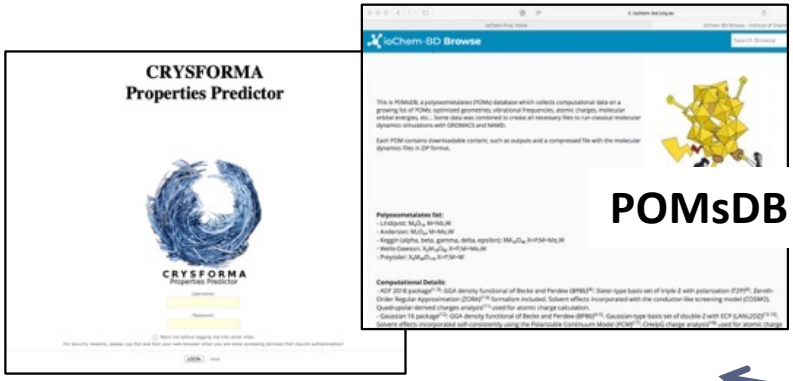
Wiley

QUARXiv -- Preprints for Quantum Chemistry

IJQC's Executable & Interactive Article

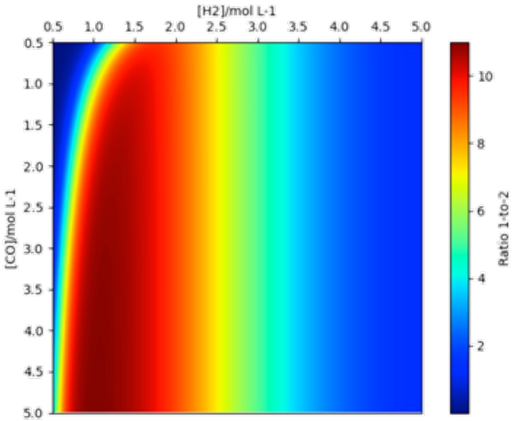
Preprints for Quantum Chemistry, by the Int. Journal of Quantum Chemistry

https://www.authorea.com/338280/_SjnKeMOgx_Dw5IHUyE7iw

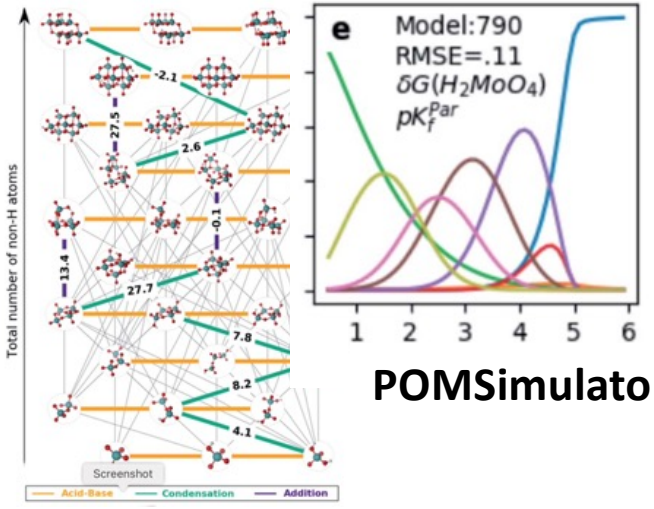


POMsDB

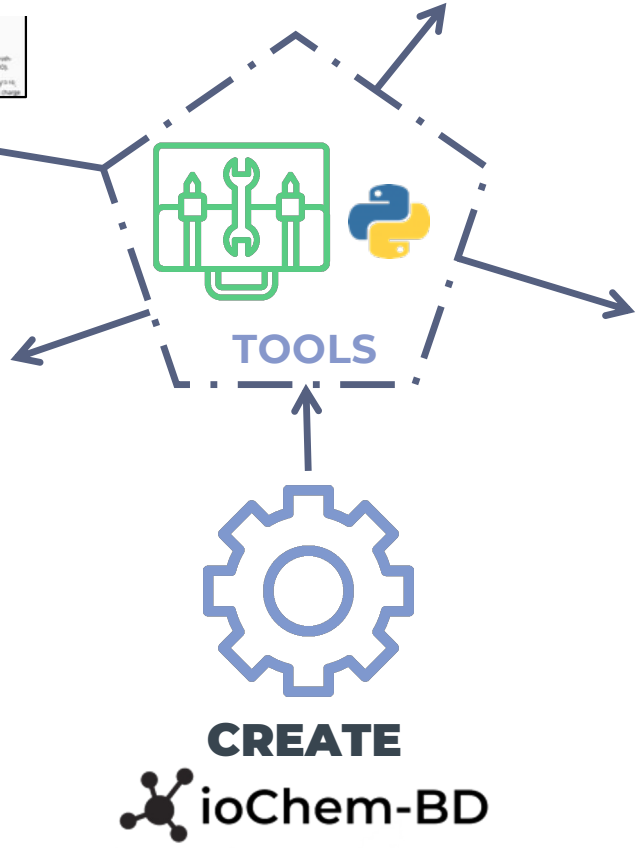
Catalytic Cycles
gTOFfee
amk-tools
OntoRXN / CRN-KG



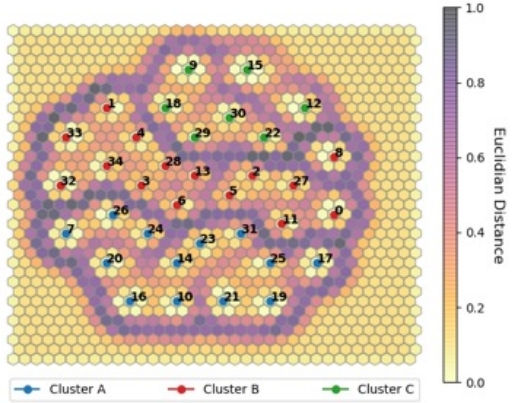
ACS Catal. **10,** 12627–12635 (2020)
ACS Phys. Chem. Au 2, 3, 225–236 (2022)
J. Cheminformatics 14:29 (2022)



POMSimulator

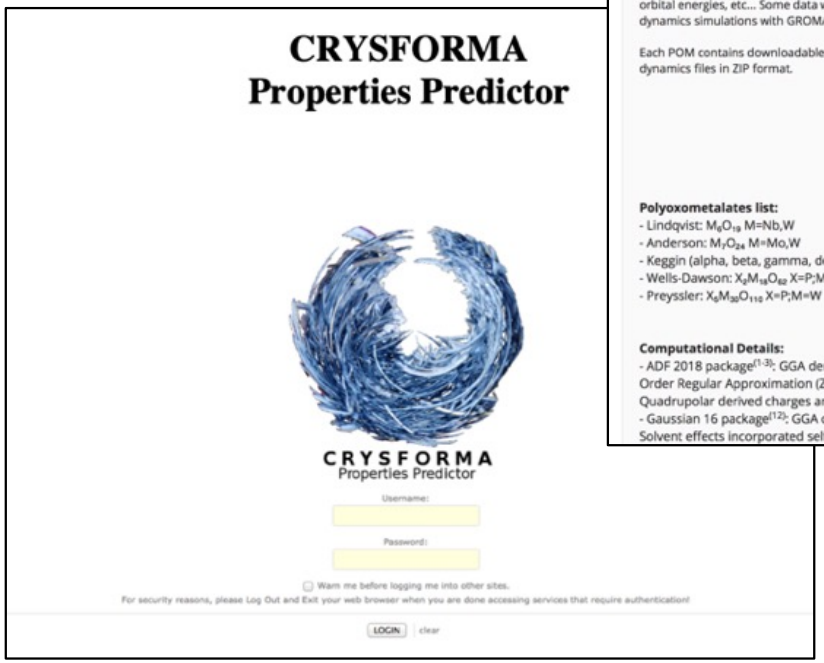


AI – Machine Learning

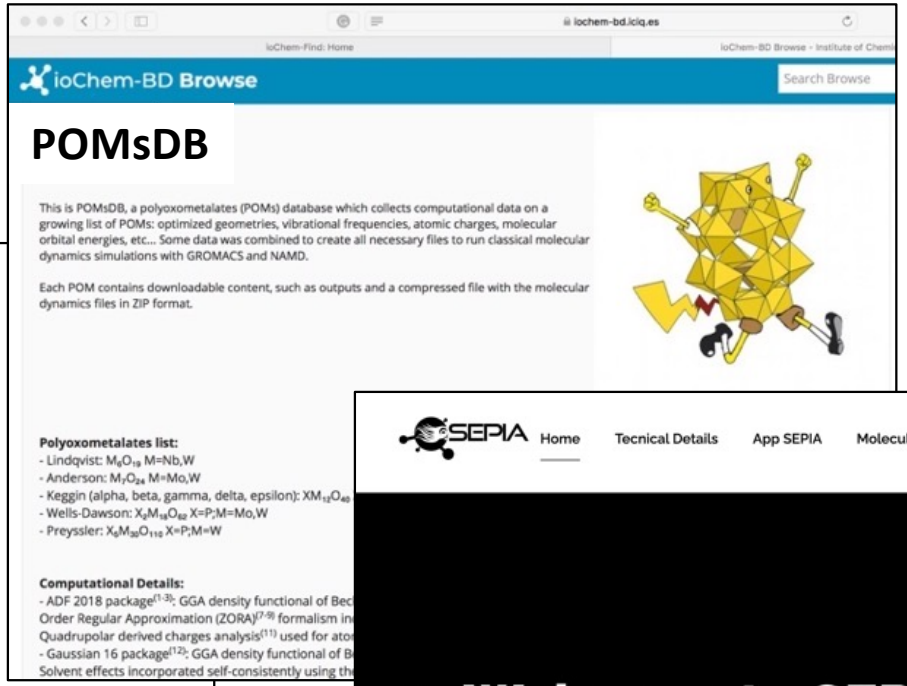


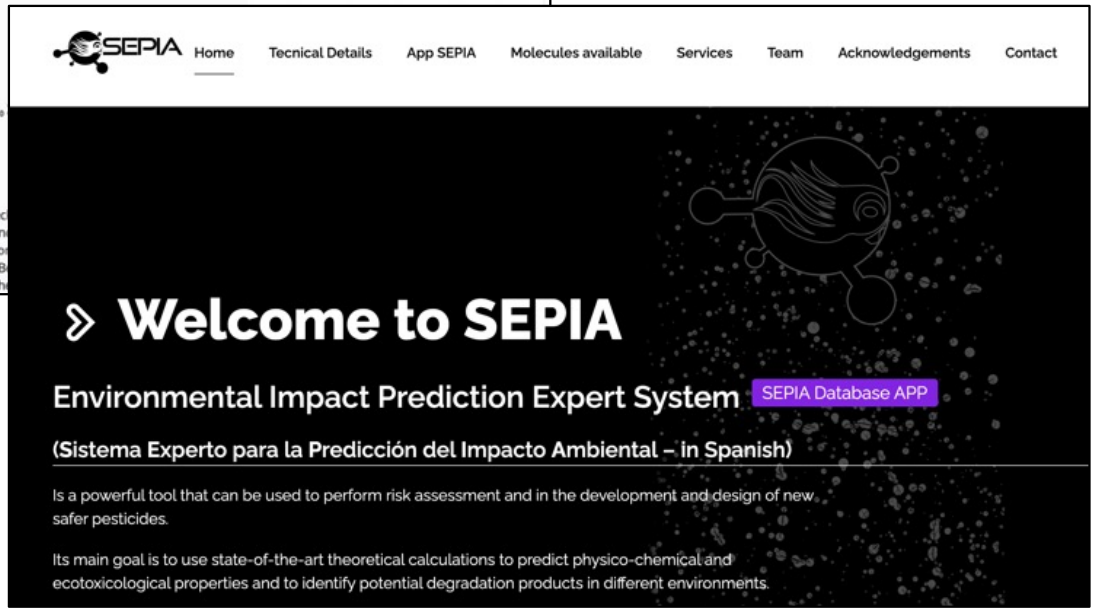
Angew. Chem. Int. Ed. 2022, e202117839

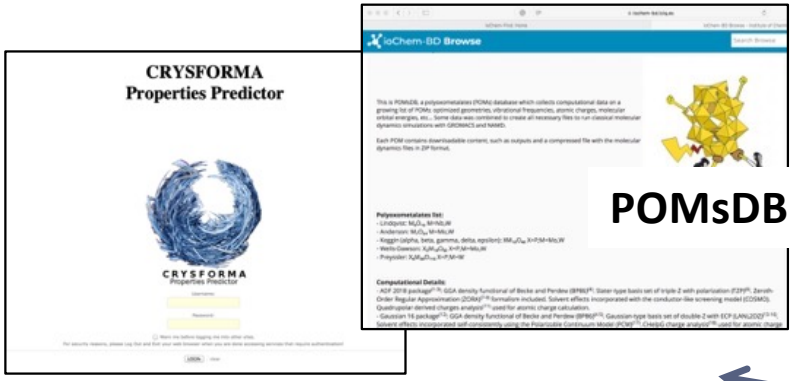
Chem. Sci. **11,** 8448–8456 (2020)
J. Phys. Chem. A, **125**, 5212-5219 (2021)
Inorg.Chem., **61**, 13508–13518 (2022)
J. Am. Chem. Soc., **145**, 18920–18930 (2023)



API co-crystals

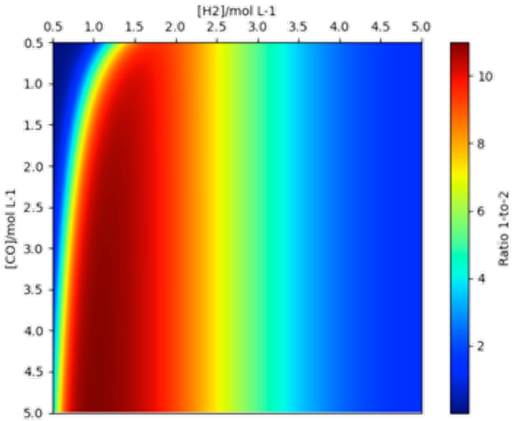




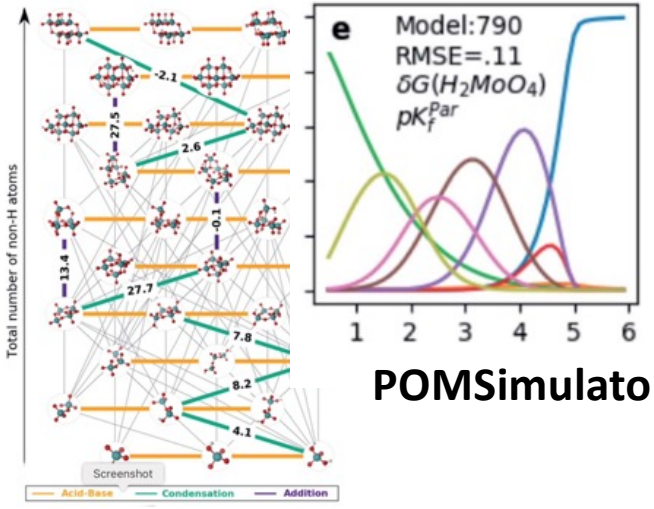


POMsDB

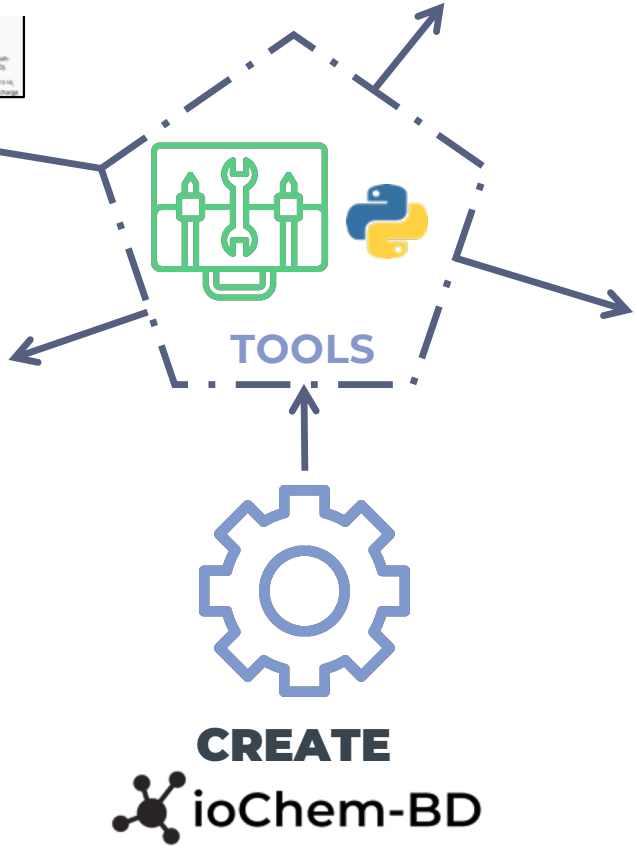
Catalytic Cycles
gTOFfee
amk-tools
OntoRXN / CRN-KG



ACS Catal. **10,** 12627–12635 (2020)
ACS Phys. Chem. Au 2, 3, 225–236 (2022)
J. Cheminformatics 14:29 (2022)

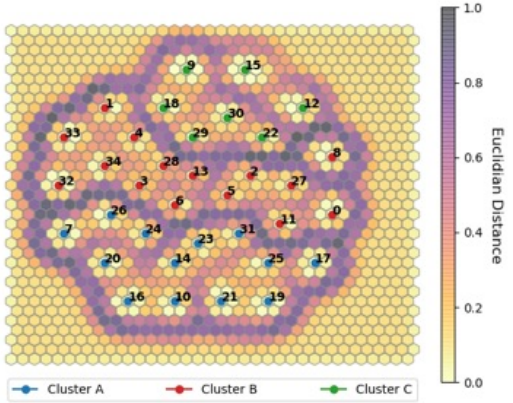


POMSimulator



Chem. Sci. **11,** 8448–8456 (2020)
J. Phys. Chem. A, **125**, 5212-5219 (2021)
Inorg.Chem., 61, 13508–13518 (2022)
J. Am. Chem. Soc., 145, 18920–18930 (2023)

AI – Machine Learning



Angew. Chem. Int. Ed. 2022, e202117839

Node configuration

Server running Find module

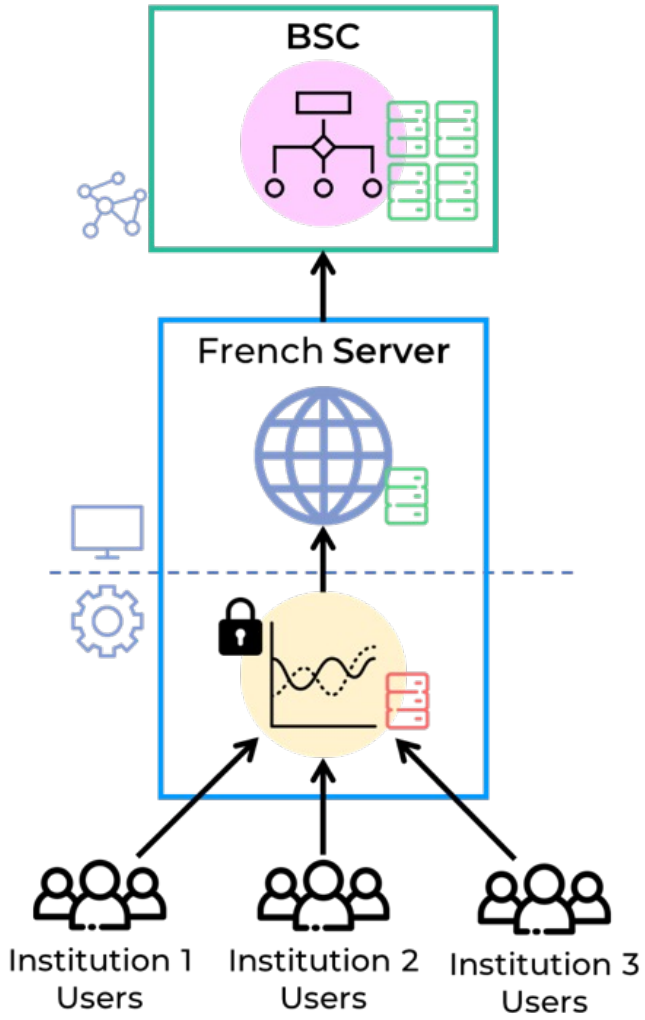
Server running Browse module

Public Database

Private Database

Credentials Required

Public/Private Boundary



ioChem-BD @BSC

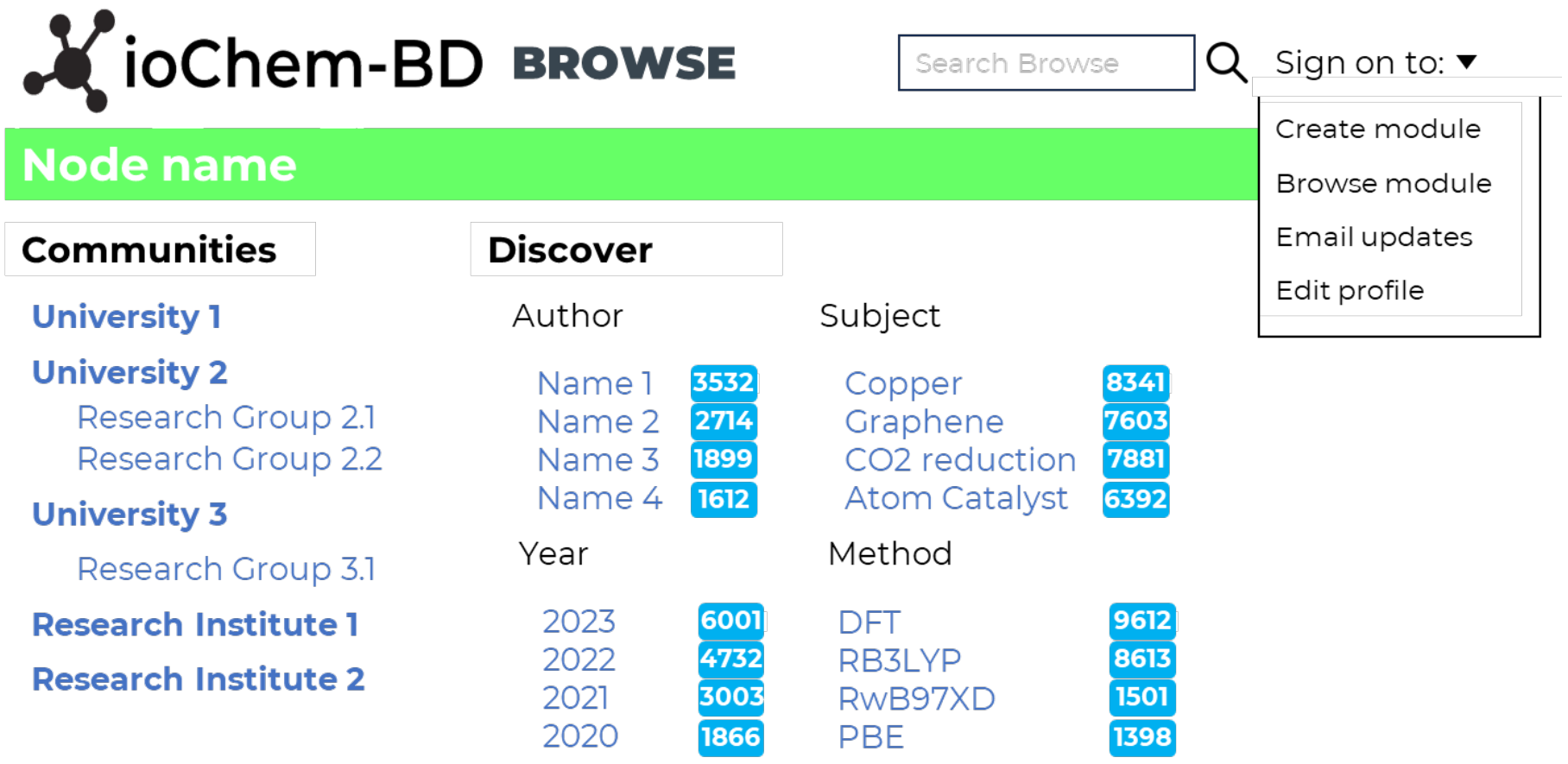
Disk : Totals 20 Tb / Occ. 11 TB

Mem: 16Gb RAM

CPU: 8 cores Intel 2Ghz

Number of Users:	505
Number of Groups:	323
Datasets published:	331
Ítems published:	306619

Node configuration



Concluding remarks:

ioChem-BD ...

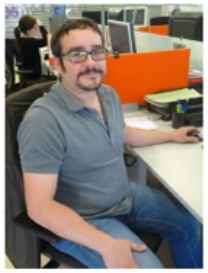
..... is a stable and mature platform .....

.... and there is a lot beyond !!

Team & Institutions

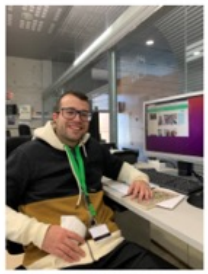
Núria Vendrell
Group Administrative Coordinator

Martin Gumbau
System Manager



Moisés Àlvarez
Programmer

Jair Wells Llanos
Technician



Aleix Mariné Tena
Technician

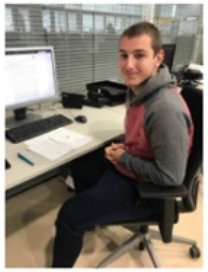
Marc Gruber López
Technician

Dr. Mireia Segado Centellas
Postdoctoral Researcher

Dr. Diego Garay Ruiz
Postdoctoral Researcher

Alba Villar Yanez
PhD Student

Marco Nicaso
PhD Student



Jordi Builes Casasnovas
PhD Student

ICIQ Institut Català d'Investigació Química

UNIVERSITAT ROVIRA i VIRGILI

CERCA Centres de Recerca de Catalunya

GOBIERNO DE ESPAÑA

Generalitat de Catalunya
Departament d'Economia i Coneixement
Secretaria d'Universitats i Recerca

BSC

RES RED ESPAÑOLA DE SUPERCOMPUTACIÓN

CTC Challenging Tomorrow's Changes

TotalEnergies

Pfizer

AstraZeneca

Institut Català d'Investigació Química

@ioChem_BD

Taming Computational Chemistry Data: ioChem-BD and Beyond