



University of
St Andrews

FOUNDED
1413



<https://nomad.wp.st-andrews.ac.uk/>

NOMAD - NMR Online Management And Datastore

Challenges of NMR data mining

*NMReDATA Symposium
SMASH 2019 - Porto*

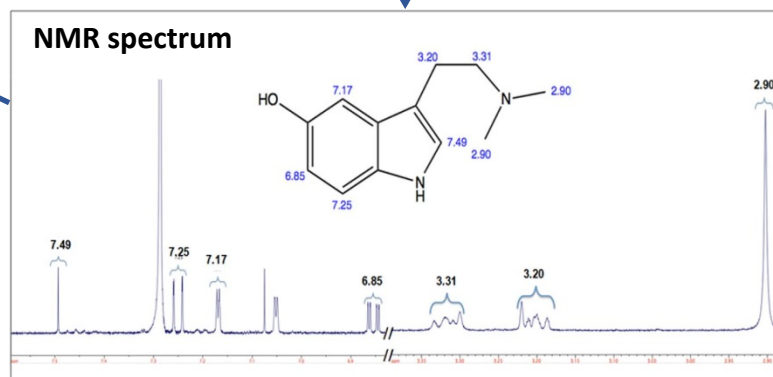
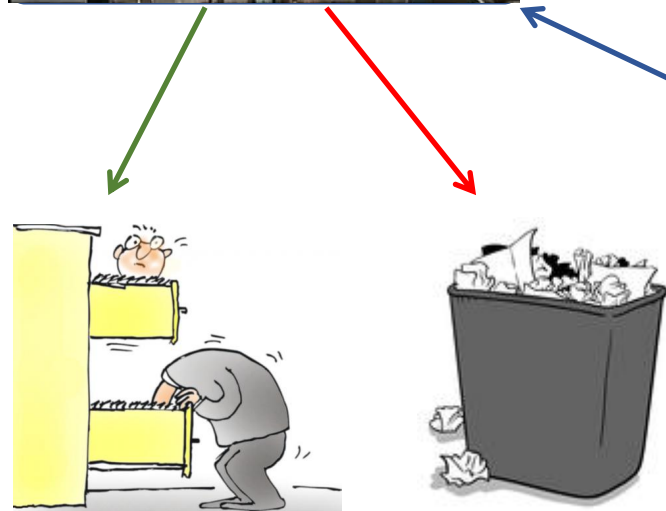
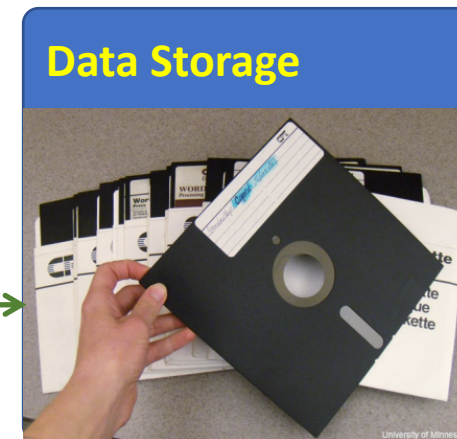
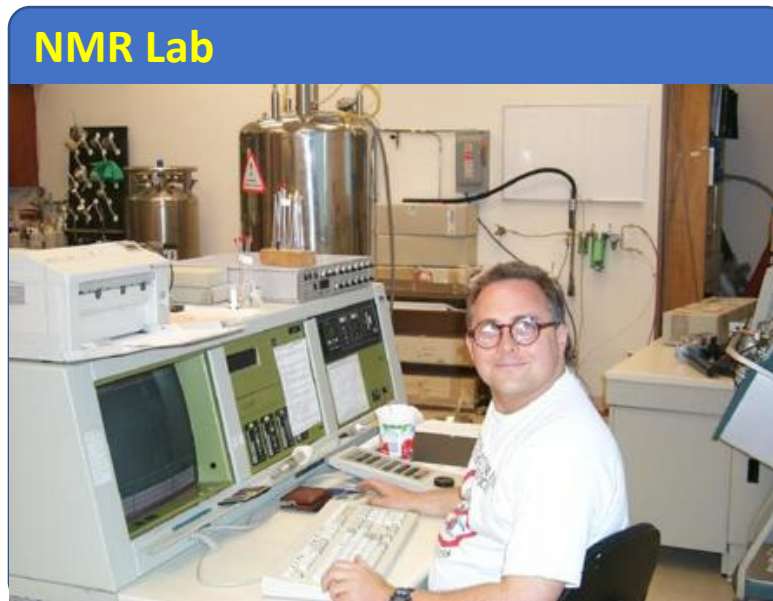
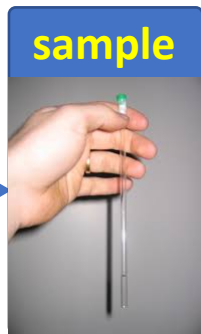
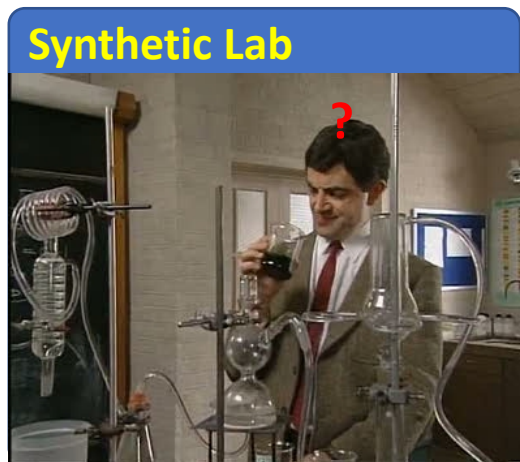
Why there isn't an equivalent of CCDC for NMR spectroscopy? [Google for NMR data]

Data Format

X-ray crystallography	NMR Spectroscopy
<p>Well defined format for data deposition</p> <p>Crystallographic Information File (CIF)</p>	<p>It is unclear what NMR data consist of</p> <ul style="list-style-type: none">• Wide variety of experimental methods ^1H, ^{13}C, COSY, HSQC, HMBC ... <i>much more</i>• Vendor formats• Many people still believe that NMR data is image or PDF file of 1D spectrum

- Is NMReDATA going to solve the problem?
- What are the other challenges of NMR data mining?

NMR lab workflow (90's - Golden Age)

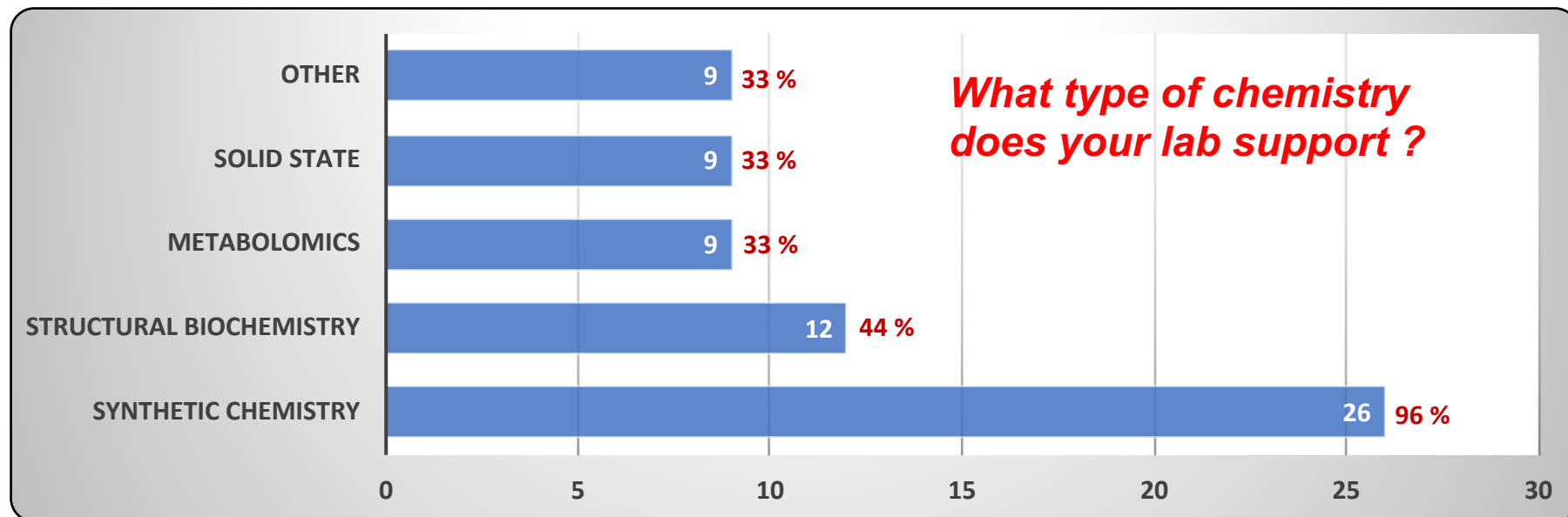


Productivity
10 to 20 experiments
per day

There was a good chance to define a uniform format for NMR data deposition

Present NMR lab workflow

Analysis of 27 responses from UK academia online survey (2016)



Average NMR lab in UK academia

Number of instruments : 6

Number of sample changers : 4

Number of users : ~ 150

Number of experiments per month : ~ 4400

NMR lab in St Andrews

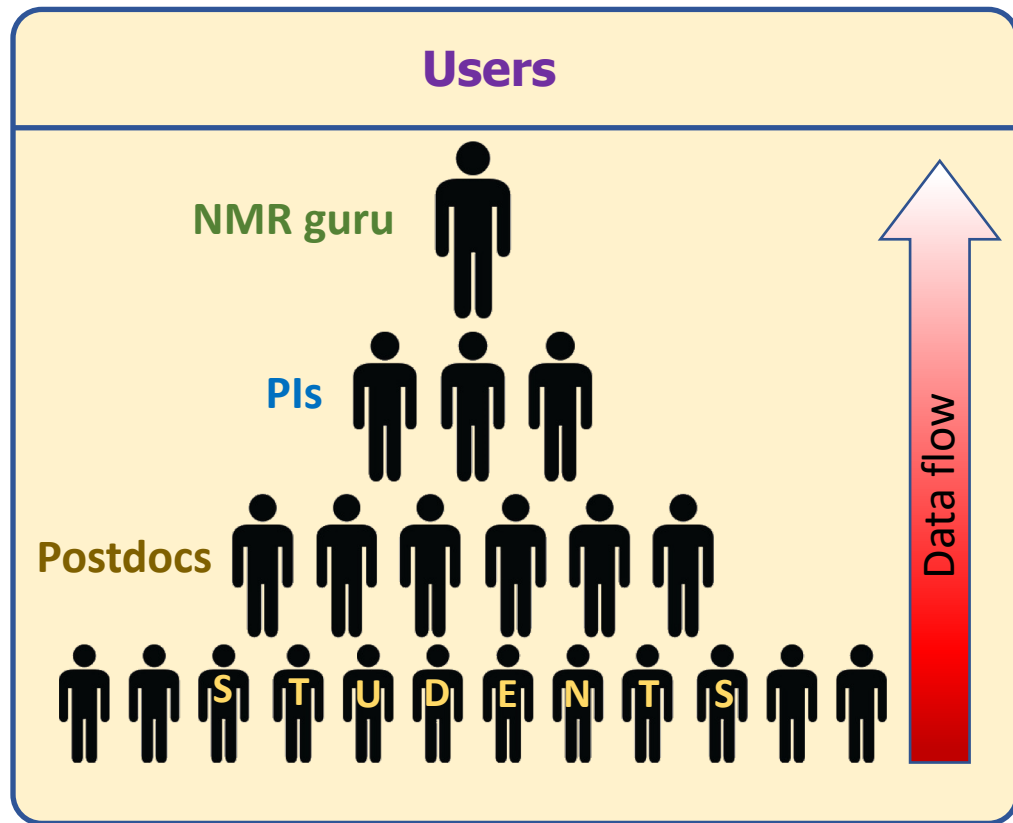
Number of instruments : 6

Number of sample changers : 6


Number of users : ~ 260 research users + 350 students in teaching labs

Number of experiments per month : 6000 – 10 000 (20 – 30GB)

Present NMR lab workflow & Challenges of NMR data management/mining

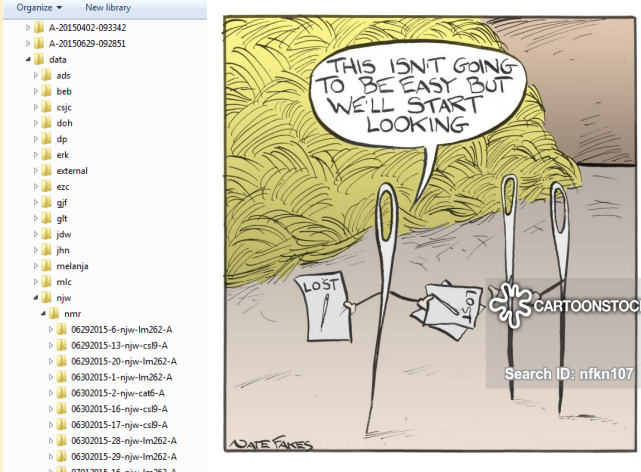


NMR Lab



- Sample changers
- 24/7 operation
- High throughput

Data Storage



Network shared drive without any search facility

Who is going to do NMR data deposition?

NMR guru
gets to see only small fraction of data

PIs
can't be bothered to use NMR software

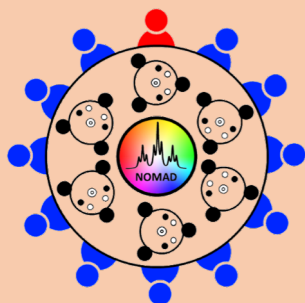
Students & Postdocs
are unreliable and elusive



NOMAD - NMR Online Management And Datastore

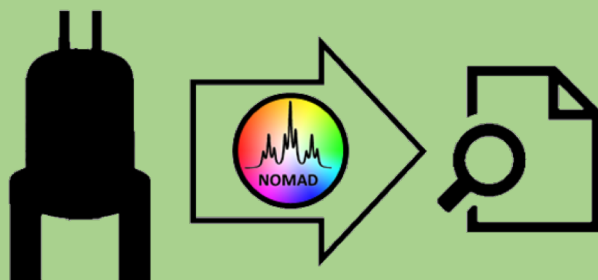
Smart and complete solution for NMR laboratory data management

Control



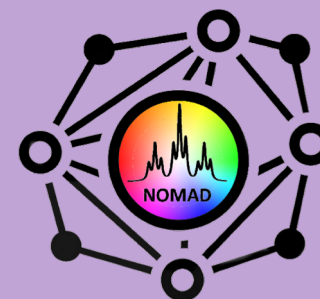
centralised dashboards for admins
traffic control & usage statistics
remote control & monitoring

Database



secure automatic data storing
raw data + provenance metadata
easy annotating, editing sharing

Open Access Data



streamline publishing of raw data
decentralised repository
driven by community

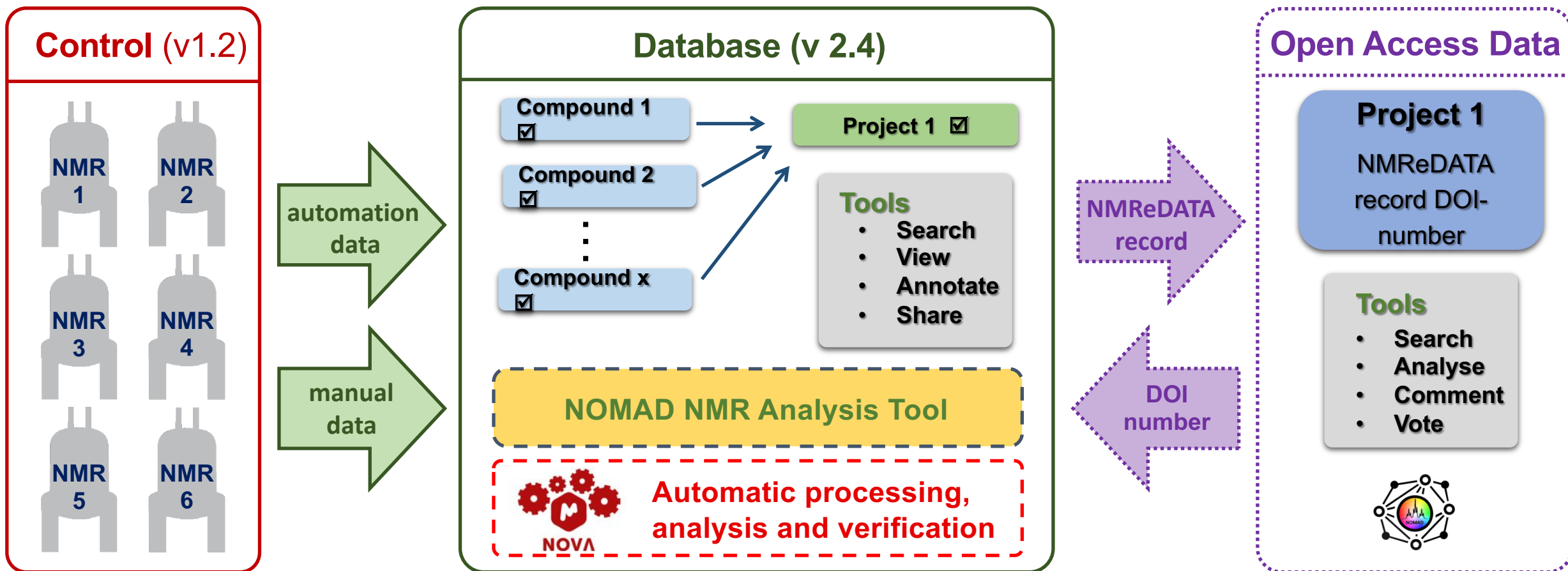
Version 1 (2012 – 2016)

Version 2 (2016 – 2019)

Vision

NOMAD - NMR Online Management And Datastore

Smart and complete solution for NMR laboratory data management



— in production

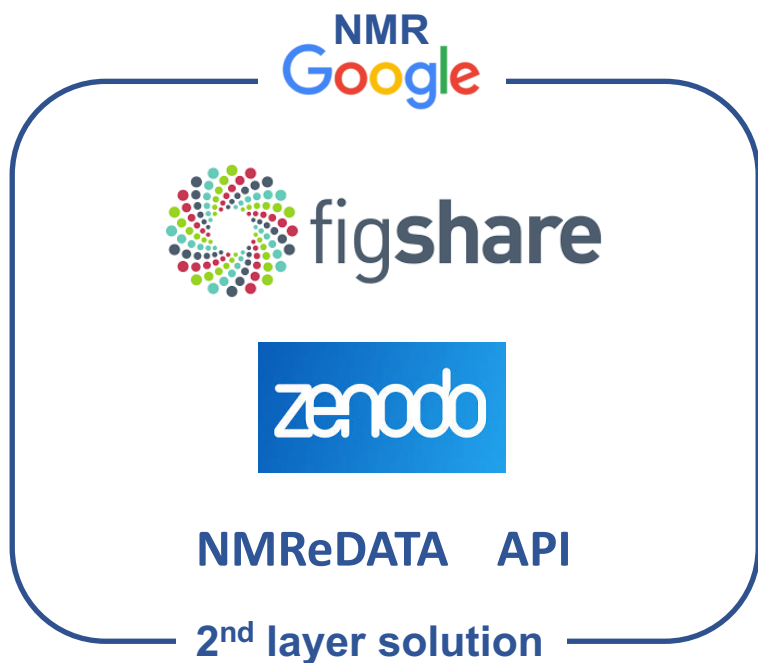
- - - prototype

..... vision



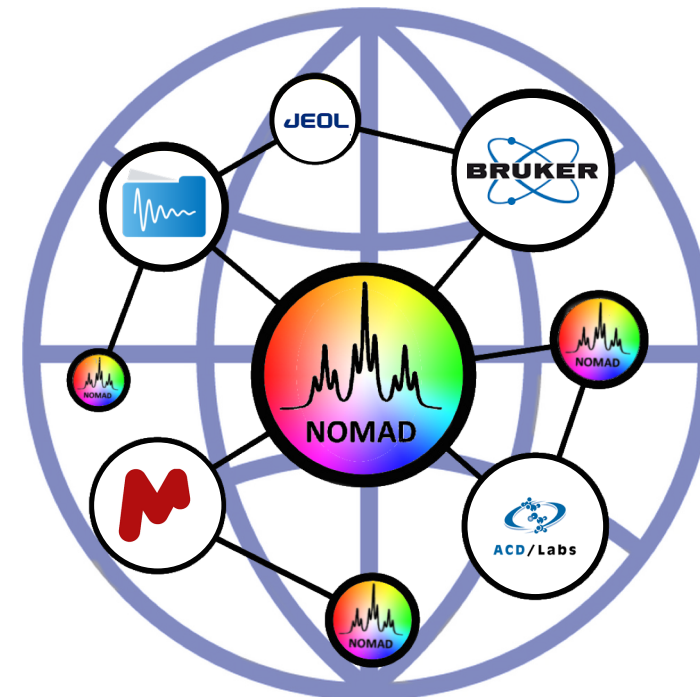
Futuristic Vision of NMR Depository Architecture

Centralised Database



- Implementing of NMR tools requires 2nd layer solution
- Central points of failure
- It's complicated to control ownership of data

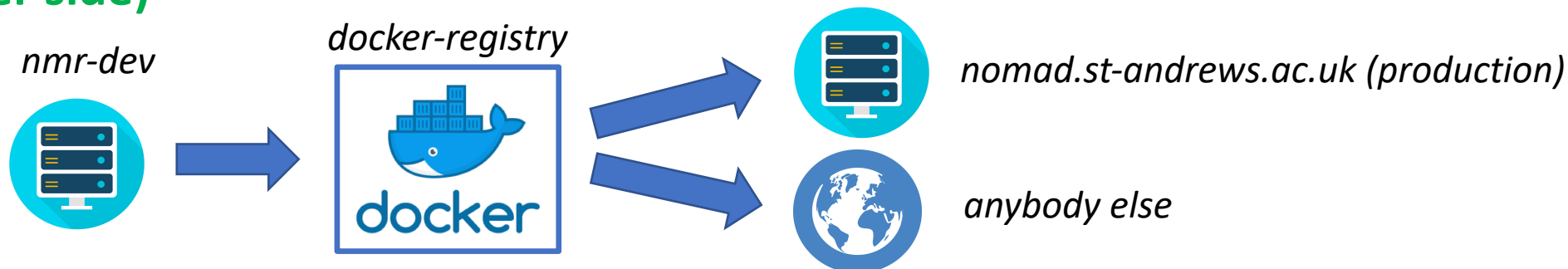
Distributed Database



- Local NMR data managements system share resources to create P2P network of nodes
- Published data is distributed and replicated on the nodes
- Ownership of data is controlled by private keys

Can I have NOMAD running in my lab?

1) Docker (server side)



2) Local networking & interfacing with instruments (challenge)

- *Sharing folders*
- *Synchronising & bucketing scripts*
- *Mailhost configuration*
- *IconNMR configuration (optional)*
- *Connection to AD servers for authentication (optional)*



Generic installation instructions

Work in progress

Get in touch if you want to give it a try and collaborate!



Funding

EPSRC

Engineering and Physical Sciences
Research Council



University of
St Andrews

FOUNDED
1413

Present



Future



People behind NOMAD



Shyam Reyal



Simone Conte



Michalis Psalios



Daria Savanovich



Ignacy Debicki



Juan Karsten



Khawar Shehzad



Iain Carson



Nattasan Inchaiya



Jake Rivett