NMR Data Management

UKMRM Meeting 2016 - Oxford
Online survey – Academia vs Industry

Academia – 27 Responses (87 %)

Industry – 4 Responses (13 %)

**Does your NMR data store provide a search facility?**

- Yes. SQL database: 50%
- Yes. Other: 50%
- No: 0%

**Does your institution use ELN (Electronic Lab Notebook)?**

- Yes. Compulsory: 25%
- Yes. Some Groups: 25%
- No: 50%

There has been no driving force to improve scientific data management in academia
Funder’s Open Access Data Policies

- All RCUK (EPSRC, STFC, BBSRC, etc.), EC (H2020), the Royal Society, the Wellcome Trust
- Mainly apply to published results

**Data Management Plan (DMP):**
- Describes the files that will be generated (formats, volume), how they will be stored, how compliance will be achieved, what will be shared; where data will be archived

**Data:**
- should be made publicly available
- should be retained - at least <10 years> from last date of access
- include a statement on how to access supporting data (where? On which conditions?)

**Costs:**
- in most cases they can be included in grant applications

**FASTR** (Fair Access to Science and Technology Research Act of 2015)
- Bill introduced to US senate last year
- In summary - every federal agency that spends more than $100 million on grants for research would be required to adopt an open access policy.
How do you deal with the Open Access Data Policies?

Summary

- We know about Open Access Data Policies.
- We mostly see it as another admin chore sent on our heads from above.
- We mostly think that it’s not our business and we let other people to deal with that.

Let’s try to see it from another perspective.
Golden Age of NMR Spectroscopy Workflow

Synthetic Lab

sample

NMR Lab

NMR spectrum

Productivity – 10 to 20 experiments per day
Online survey – Current NMR dataflow

What type of chemistry does your lab support?

**Academia – 27 Responses**

- **SYNTHETIC CHEMISTRY**: 26, 96%
- **STRUCTURAL BIOCHEMISTRY**: 12, 44%
- **METABOLOMICS**: 9, 33%
- **SOLID STATE**: 9, 33%
- **OTHER**: 9, 33%

**Data volume per month**

- Less than 20 GB: 30%, 48%
- 20 – 30 GB: 22%
- More than 30 GB: (only 23 responses)

**Average NMR lab**

- Number of instruments: 6
- Number of sample changers: 4
- Number of users: ~150
- Number of experiments per month: ~4400
Online survey – Current NMR dataflow

Where do you archive NMR data in your lab?

- Obvious trend towards Central File Storage systems
- Very likely impact of Open Access Data policies

More important is how we store and access the data
Online survey – Current NMR dataflow

How do you archive NMR data in your lab?

- Archiving script can be seen unsecure by people maintaining central file store.
- It seems that no one in UK academia uses commercial database for NMR spectra like ACD/Labs Spectrus DB, Mnova DB plugin or Magic Angle LOGS.

Does your NMR data store provide a search facility?

- Yes. SQL database. 1
- Yes. Other. 5
- No. 19
NOMAD – NMR Online Management And Datastore

- **Web based “cloud computing” system** that provides a resilient searchable data-store and centralised control of automation for Bruker NMR instruments.

- **Current version 1.2 (prototype)** – 6 Bruker instruments with sample changers, 260 research users, 350 users in teaching labs

- **Dataflow** – 6 000 to 10 000 Experiments (20 -30 GB) per month
# NOMAD – Submission Portal

**Always REFRESH** this page before you decide/act; the list shown here is **00:00:34** (hh:mm:ss) old and thus might be outdated.

<table>
<thead>
<tr>
<th>Holder</th>
<th>NMR Machine</th>
<th>Username</th>
<th>PIGroup</th>
<th>Solvent</th>
<th>Sample Identifier</th>
<th>ExpNo</th>
<th>Experiment</th>
<th>Parameters</th>
<th>Title</th>
<th>Print</th>
<th>Email</th>
<th>Date</th>
<th>Time</th>
<th>Status</th>
<th>Select</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Alec</td>
<td>tilt2</td>
<td>til</td>
<td>CDC3</td>
<td>09142015-8-tltil2-A</td>
<td>11</td>
<td>c13_deptq.c.and</td>
<td>ns,256</td>
<td>Demo</td>
<td>Y</td>
<td></td>
<td>2015-09-14</td>
<td>12:42</td>
<td>BOOKED</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Alec</td>
<td>tilt2</td>
<td>til</td>
<td>CDC3</td>
<td>09142015-8-tltil2-A</td>
<td>10</td>
<td>proton.c.and</td>
<td></td>
<td>Demo</td>
<td>Y</td>
<td></td>
<td>2015-09-14</td>
<td>12:42</td>
<td>BOOKED</td>
<td></td>
</tr>
</tbody>
</table>

**Automation - Running - Busy until: Mon 13:05 - Day Experiments: 00:12 - Night Experiments: 00:00 - MAX ALLOWANCE: 00:15**

<table>
<thead>
<tr>
<th>Holder</th>
<th>NMR Machine</th>
<th>Username</th>
<th>PIGroup</th>
<th>Solvent</th>
<th>Sample Identifier</th>
<th>ExpNo</th>
<th>Experiment</th>
<th>Parameters</th>
<th>Title</th>
<th>Print</th>
<th>Email</th>
<th>Date</th>
<th>Time</th>
<th>Status</th>
<th>Select</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Marcus</td>
<td>na10</td>
<td>dnh</td>
<td>CDC3</td>
<td>09142015-15-dh-na10-M</td>
<td>10</td>
<td>proton_wide.a.and</td>
<td>Naphynethanol</td>
<td>Y</td>
<td></td>
<td></td>
<td>2015-09-14</td>
<td>12:35</td>
<td>FINISHED</td>
<td>15</td>
</tr>
<tr>
<td>14</td>
<td>Marcus</td>
<td>rmp</td>
<td>ads</td>
<td>CDC3</td>
<td>09142015-14-ads-rmp-M</td>
<td>10</td>
<td>proton_quant.a.and</td>
<td>104 Aldehyde</td>
<td>Y</td>
<td></td>
<td></td>
<td>2015-09-14</td>
<td>12:08</td>
<td>FINISHED</td>
<td>14</td>
</tr>
</tbody>
</table>

**Username**: tilt2

**Password**: **********

**Name**: Tomas Lebl

**PI Group**: tilt

**Available Machines**: Alec

**No of Samples**: 2

<table>
<thead>
<tr>
<th>Holder</th>
<th>Sample Identifier</th>
<th>Solvent</th>
<th>Experiment</th>
<th>Parameters</th>
<th>Print</th>
<th>Email</th>
<th>Title</th>
<th>Duration</th>
<th>Total (D)</th>
<th>More</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>09142015-21-dh21-A</td>
<td>CDC3</td>
<td>proton.c.and</td>
<td>-</td>
<td></td>
<td></td>
<td>Demo 1</td>
<td>00:00:34</td>
<td>00:02:47</td>
<td>Add</td>
</tr>
<tr>
<td>22</td>
<td>09142015-22-dh22-A</td>
<td>DM50</td>
<td>proton.c.and</td>
<td>-</td>
<td></td>
<td></td>
<td>Demo 1</td>
<td>00:00:48</td>
<td>00:04:33</td>
<td>Add</td>
</tr>
</tbody>
</table>

Holders exceeding (00:30:00) will be submitted as NIGHT experiments.
### NOMAD – Admin Portal - user & experiment tables

#### User Configuration
- **ID**: [Input Field]
- **Username**: [Input Field]
- **Real Name**: [Input Field]
- **Special User**: (Only if user is not listed in the university login)
- **Manual User**: (Only if user is allowed to do manual experiments)
- **Inactive User**: (User will not show up in drop-down boxes)
- **Category**
- **PI Group**: [Input Field]

#### User Stats
- **Total Users (Research + Teaching)**: 1,071
- **Total Active Users (Research + Teaching)**: 673
- **Total Active Groups**: 22
- **Very Active Users (Research + Teaching)**: 123
- **Total Research Users**: 883
- **Active Research Users**: 304
- **Inactive Research Users**: 200
- **Special Research Users**: 63
- **Manual Research Users**: 12
- **Total Teaching Users**: 277
- **Active Teaching Users**: 201

#### Quick Filter
- **Include Inactive**: [Checkbox]
- **Category**

#### User Table

<table>
<thead>
<tr>
<th>User ID</th>
<th>Username</th>
<th>Real Name</th>
<th>Special</th>
<th>Manual</th>
<th>Inactive</th>
<th>Category</th>
<th>Group</th>
<th>Actions</th>
<th>Last Activity</th>
<th>Days of Inactivity (UID)</th>
</tr>
</thead>
<tbody>
<tr>
<td>305</td>
<td>lex</td>
<td>Brian Caimbers</td>
<td>YES</td>
<td></td>
<td></td>
<td>Active</td>
<td>pk</td>
<td>Load</td>
<td>2016-06-09 09:29:07.0</td>
<td>22</td>
</tr>
<tr>
<td>1304</td>
<td>bjml</td>
<td>Bethany Laxion</td>
<td>--</td>
<td></td>
<td></td>
<td>Active</td>
<td>COT</td>
<td>Load</td>
<td>2016-04-25 14:21:55.0</td>
<td>60</td>
</tr>
<tr>
<td>776</td>
<td>cm03</td>
<td>Claire Broda</td>
<td>--</td>
<td></td>
<td></td>
<td>Active</td>
<td>MRO</td>
<td>Load</td>
<td>2016-04-05 12:07:22.0</td>
<td>77</td>
</tr>
<tr>
<td>105</td>
<td>ltl0</td>
<td>Laurenza Taylor</td>
<td>--</td>
<td></td>
<td></td>
<td>Active</td>
<td>PK</td>
<td>Load</td>
<td>2016-09-12 16:29:05.0</td>
<td>31</td>
</tr>
<tr>
<td>113</td>
<td>ak7</td>
<td>Pete Kilian</td>
<td>--</td>
<td></td>
<td></td>
<td>Active</td>
<td>PI</td>
<td>Load</td>
<td>2016-05-28 15:45:25.0</td>
<td>65</td>
</tr>
<tr>
<td>51</td>
<td>pm44</td>
<td>Philip Napman</td>
<td>--</td>
<td></td>
<td></td>
<td>Active</td>
<td>MRO</td>
<td>Load</td>
<td>2016-04-03 15:00:41.0</td>
<td>18</td>
</tr>
</tbody>
</table>

#### Experiment Table

<table>
<thead>
<tr>
<th>Exp-ID</th>
<th>Experiment Name</th>
<th>Description</th>
<th>Parameters</th>
<th>Hidden</th>
<th>Composite</th>
<th>Machines</th>
<th>Portals</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>proton.a.and</td>
<td>1H Observe</td>
<td>ns = 0; d1 = 1; expq = 1m 96; ds = 2; td1 = 1</td>
<td>--</td>
<td>--</td>
<td>N M R F</td>
<td>R &amp; T</td>
<td>Load</td>
</tr>
<tr>
<td>2</td>
<td>proton.WIDE.a.and</td>
<td>1H Observe with Expanded Sweep Width (from 15 to -25 ppm)</td>
<td>ns = 0; d1 = 1; expq = 51s; ds = 2; td1 = 1</td>
<td>--</td>
<td>--</td>
<td>N M R F</td>
<td>R-only</td>
<td>Load</td>
</tr>
<tr>
<td>3</td>
<td>proton_Quan.a.and</td>
<td>1H Observe - Quantitative</td>
<td>ns = 0; d1 = 30; expq = 5m 59s; ds = 2; td1 = 1</td>
<td>--</td>
<td>--</td>
<td>N M R F</td>
<td>R-only</td>
<td>Load</td>
</tr>
<tr>
<td>4</td>
<td>solventsup.a.and</td>
<td>1H Observe with Solvent Suppression</td>
<td>ns = 0; d1 = 4; expq = 1m 49s; ds = 4; td1 = 1</td>
<td>--</td>
<td>--</td>
<td>N M R F</td>
<td>R-only</td>
<td>Load</td>
</tr>
</tbody>
</table>
## NOMAD – Admin Portal - Accounting

### Reports Search

- **PI Group:** dp
- **Inactive:**
- **Date From:** 01/09/2015
- **Date To:** 14/09/2015
- **View Duration:**
- **Complete Clear:**

### Storage Data (GB) Loading Status:

**Complete**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto</td>
<td>00:40:46</td>
<td>0.009</td>
<td>£6.00</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>£180.45</td>
</tr>
<tr>
<td>Manual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>£111.05</td>
</tr>
<tr>
<td>jrv74 - Josh Richards</td>
<td>00:58:23</td>
<td>0.015</td>
<td>£8.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>00:31:52</td>
<td>0.006</td>
<td>£12.65</td>
<td>£220.08</td>
</tr>
<tr>
<td>Auto</td>
<td>00:33:02</td>
<td>0.004</td>
<td>£4.78</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>£5.48</td>
</tr>
<tr>
<td>Manual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>£5.48</td>
</tr>
<tr>
<td>tk235 - Tamara Kosikova</td>
<td>02:12:05</td>
<td>0.028</td>
<td>£19.42</td>
<td>00:30:58</td>
<td>0.006</td>
<td>£3.78</td>
<td>00:30:58</td>
<td>0.006</td>
<td>£3.78</td>
<td>00:30:58</td>
<td>0.006</td>
<td>£3.78</td>
<td>00:30:58</td>
<td>0.006</td>
<td>£3.78</td>
<td>00:30:58</td>
<td>0.006</td>
<td>£3.78</td>
<td>£204.04</td>
</tr>
<tr>
<td>Auto</td>
<td>00:33:02</td>
<td>0.004</td>
<td>£4.78</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>00:07:06</td>
<td>0.002</td>
<td>£0.70</td>
<td>£5.48</td>
</tr>
<tr>
<td>Manual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>£5.48</td>
</tr>
<tr>
<td>TOTAL</td>
<td>N/A</td>
<td>0.028</td>
<td>£19.42</td>
<td>N/A</td>
<td>0.074</td>
<td>£16.43</td>
<td>N/A</td>
<td>0.135</td>
<td>£391.84</td>
<td>N/A</td>
<td>0.068</td>
<td>£47.71</td>
<td>N/A</td>
<td>0.065</td>
<td>£40.48</td>
<td>N/A</td>
<td>0.065</td>
<td>£40.48</td>
<td>£518.37</td>
</tr>
</tbody>
</table>

### Total Cost Summary

- **Alec:** £180.45
- **Felix:** £220.08
- **Hector:** £204.04
- **Marcus:** £204.04
- **Noah:** £204.04
- **Robin:** £204.04
- **Total:** £518.37
## NOMAD – User Search Portal

### NOMAD v1.2 - NMR Search Portal

![NOMAD Search Portal](image)

### Search Results

<table>
<thead>
<tr>
<th>ID</th>
<th>Username</th>
<th>PI Group</th>
<th>Holder</th>
<th>Solvent</th>
<th>Sample Identifier</th>
<th>Exp No</th>
<th>Experiment</th>
<th>Parameters</th>
<th>Title</th>
<th>Date</th>
<th>Bucket</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>402569</td>
<td>bn22</td>
<td>jhn</td>
<td>4</td>
<td>CDC23</td>
<td>06/2019-6-jhn-bn22-H</td>
<td>12</td>
<td>ghoop-eG-c and</td>
<td>r=2</td>
<td>Pateaamide D</td>
<td>2016-06-20</td>
<td>K-20160221-105222</td>
<td>Download, View</td>
</tr>
<tr>
<td>402568</td>
<td>bn22</td>
<td>jhn</td>
<td>4</td>
<td>CDC23</td>
<td>06/2019-6-jhn-bn22-H</td>
<td>11</td>
<td>ghory, a and</td>
<td>r=2</td>
<td>Pateaamide D</td>
<td>2016-06-20</td>
<td>K-20160221-105222</td>
<td>Download, View</td>
</tr>
<tr>
<td>402557</td>
<td>bn22</td>
<td>jhn</td>
<td>4</td>
<td>CDC23</td>
<td>06/2019-6-jhn-bn22-H</td>
<td>10</td>
<td>proton c and</td>
<td>r=8</td>
<td>Pateaamide D</td>
<td>2016-06-20</td>
<td>K-20160221-105222</td>
<td>Download, View</td>
</tr>
</tbody>
</table>

- **Always REFRESH** this page before you decide to act, as the list shown here is 06:25:23 (hh:mm:ss) old and thus might be outdated.
- Search Returned 3 Results

---

**Instructions:**

- Use the filter options to narrow down your search.
- Click on the download icon to save the data.
- Click on the view icon to view the details.

---

**Notes:**

- Ensure you refresh the page before taking any action.
- The search results are outdated as of 06:25:23.

---

**Contact:**

- Support: support@nomad.org
- Helpdesk: helpdesk@nomad.org
- Feedback: feedback@nomad.org
NOMAD – PURE link & Open-Data

- NOMAD stores the data securely and provides instant search facility but the data are not publicly available.
- Users have to download selected data and upload them to PURE (Research Information Management System).
- It’s not clear what should be uploaded to PURE and capacity is limited.

![Diagram showing the flow of data between R-Search, PURE-link (private), Open-Data (world readable), PURE, and ELN.]

- Additional Info:
  - Molecular Structure
  - Verification
  - Multiplet reports
  - IUPAC-names
  - Versioning

- URLs:
  - NOMAD – PURE link & Open-Data
  - DOI-number

- Other Information:
  - Project 1
  - Project 1 DOI-number

- Notation:
  - Experiment 1 ✓
  - Experiment 2 ✓
  - Experiment 3 □
  - Experiment 4 ✓
  - Experiment 5 ✓
  - Experiment 6 ✓
  - Experiment 7 □
  - Experiment X □
Nowadays NMR spectroscopy produces vast quantity of data.

- Is that data sufficiently utilised and shared?

Pictures of $^1$H and $^{13}$C spectra in Supporting Information

- Why academia does not use commercial NMR databases?
- Why NMR spectroscopy does not have an equivalent of Protein Data Bank or Cambridge Crystallographic Data Centre?

Likely there has been no driving force.

Could Open Access Data Policies become that driving force?

Maybe
**NOMAD - Current Status and Future Plan**

**Current status**
- Team of 6 people working part time
- We have setup infrastructure and processes for software development and production
- IAA funding available until March 2017

**Short term future plan**
- “PURE link” add-on function and “Open Data Portal”
- NOMAD 2.0 – modular architecture, separate database and control function

**Long term future plan**

![Diagram of NOMAD architecture]
Discussion

• Do you like the concept of NOMAD system? Any comments?

• If yes how to propel further development and distribution?
  
  a) University of St Andrews spin-off
  b) EPSRC or other funding