Demystifying data management: example from physics and chemistry

Petr Štěpánek

University of Oulu Oulu, Finland

16 June 2021

- NMR Research Unit
- Molecular spectroscopy and materials research
- Faculty of Science



What is good data?



Towards reproducible data management

3 / 11 16 June 2021

What is good data?

Laser dode type 406720 Reference wavefull nm 405 Sample 0 DET025A Injection current / mA 0.0068734472 0.5 0.0068734472 2.5 0.012248211 1.5 0.0098213749 3.5 0.039662570347 3.5 0.039662570347 4.5 0.0259643723 5.5 0.0024951377 5.5 0.0024951377 7.5 0.0368383704 6.5 0.0028938970 4.5 0.0259643723 0.6 0.0059946776 7.5 0.0368837047 3.5 0.036862570347 3.5 0.036862570347 3.5 0.036862570347 3.5 0.036862570347 3.5 0.0368838704 4.5 0.0259843723 0.6 0.0559946725 1.1 0.06889389877 7.5 0.0368837047 1.1 0.06889389877 7.5 0.0368837047 1.1 0.06889389877 1.1 0.06889389877 1.1 0.06889389877 1.1 0.0688938977 1.1 0.06889398977 1.1 0.0688939877 1.1 0.0688939877 1.1 0.0688938977 1.1 0.0688938977 1.1 0.0688938977 1.1 0.0688938977 1.1 0.0688938977 1.1 0.0688938977 1.1 0.0688938977 1.1 0.0688938977 1.1 0.068938977 1.1 0.068938977 1.1 0.0689389877 1.1 0.0689389877 1.1 0.0688938977 1.1 0.0689389877 1.1 0.0689398877 1.1 0.0689398877	Purpose	Test of the output laser power											
Reference varielength // eng C 25 Sample Water In 1 cm cell Detector DET025A Injection current /mA Detector voltage // 0.5 0.000674472 0.5 0.000674472 1 0.000674472 2.5 0.00087105471 3 0.006825703547 4.5 0.0256437123 5.5 0.004125746 6.5 0.0026643723 5.5 0.0024851317 6.5 0.0024851317 7.5 0.0036337437 7.5 0.0036337437 8 0.0256437123 9 0.07553444252 10 0.1078326292 11 0.0086837437 12 0.07553444257 13 0.0179524865 11 0.0086807172 11 0.0086807172 12 0.07553444252 13 0.0179524865 14 0.1309651722 13 0.0179524466 14 0.1309651722 15 0.0646407764 14	Laser diode type	L405P20											
Temperature / deg Č 25 Sample Water h 1 cm cell Detector DET025A injection current / mA Detector voltage/ V 0.00069744472 0 1.00192248511 0 2.5 0.01233237 3 0.0366257043 4.5 0.0266347142 5 0.0092182569 3 0.0366257043 4.5 0.026637123 5 0.012849237 7.5 0.036837437 6.5 0.0049435746 6.5 0.004943937 7.5 0.036837437 8 0.0560487746 6.5 0.004943937 7.5 0.036837437 9.5 0.075343425 9.5 0.075343425 10 0.1078305292 11 0.068871712 11.5 0.113744548 12 0.107930246 12.5 0.084404764 13 0.01374246 14.5 0.13282977 <t< td=""><td>Reference wavelength / nm</td><td>405</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	Reference wavelength / nm	405											
Sample Detector Water in 1 cm cell Detector voltage V 0 Detector voltage V 0 Detector voltage V 0 Detector voltage as a function of injection current for 405 nm laser diode 1 0.0066744472 0.0006874472 Detector voltage as a function of injection current for 405 nm laser diode 1 0.0092165462 1 0.005667447 2 0.00921654712 Detector voltage as a function of injection current for 405 nm laser diode 3 0.0056675473 1 4 0.012845117 1 5 0.026437123 0 45 0.0256437123 0 5 0.0044125746 0 7 0.012842077 0 7 0.012842077 0 7 0.012842077 0 9 0.075534547 0 10 0.107832622 10 15 20 25 30 11 0.0086871427 10 15 20 25 30 11 0.017832629 10 15 20 25 30	Temperature / deg C	25											
Detector DETO25A Injection current / mA Detector voltage / 0 0 0.0066/14472 0.5 0.0062/19269 1 0.0192248511 1.5 0.0098831704 2.5 0.012328271 3 0.0366705471 4 0.012329311 5.5 0.0092195269 5 0.012328371 4 0.0123951123 5 0.0123951123 5 0.00961857764 6 0.065944874 6.5 0.0049939977 7 0.0112849237 7.5 0.0368374376 9 0.0755348247 10 0.10783385247 11 0.068948674 12 0.1078328292 13 0.010781246265 11 0.0688871712 12 0.107824865 13 0.0107910246 14 0.13282977 13 0.0107910246 14 0.13282977 15 </td <td>Sample</td> <td>Water in 1 cm cell</td> <td></td>	Sample	Water in 1 cm cell											
Injection current / ma Detector voltage V 0 0.0086744472 0.5 0.0082193269 1 0.0129248511 1.5 0.00982193269 2 0.00982193269 2 0.00982193269 2 0.00982193269 2 0.00982193269 2 0.00982193269 3 0.0366257033 3 0.0366257033 4 0.0129951371 5 0.0094125766 6 0.06569468771 7 0.0112849217 7 0.012849217 7 0.012849217 7 0.012849217 7 0.012849217 7 0.012849217 10 0.052042462 9 0.0753415247 10 0.1078326292 10.5 0.02782058753 11 0.068687172 12 0.07953415247 13 0.017912246 14 0.1392629665 13	Detector	DET025A											
Injection current / mA Detector voltage v 0 0.008674472 1 0.0192248511 1.5 0.009813704 2.5 0.01232377 3 0.0366570347 4 0.0127495112 5 0.0026495112 5 0.01232377 4 0.0127607377 5 0.00496137746 6 0.065946874 6.5 0.00499337747 7 0.011284227 7 0.012842827 7 0.012842827 8 0.0520424262 9 0.05753815247 10 0.1078305262 11 0.0689306573 12 0 13 0.0107910246 12 0.0107910246 13 0.0107910246 14 0.13282977 13 0.0107910246 14 0.13282977 15 0.064404764 16 0.094494977 17 0.010793024													
0 0.0008744472 0.05 Detector voltage as a function of injection current for 405 nm laser diode 1 0.0192248511 2 0.0192248511 2 Detector voltage as a function of injection current for 405 nm laser diode 2 0.0092105462 1 1 3 0.0366257043 1 1 4 0.0137607337 1 1 1 5 0.025647123 0 0 0 5 0.0092165462 0 0 0 5 0.025647123 0 0 0 0 5 0.0094185766 0 0 0 0 0 5 0.0094185766 0 0 0 0 0 0 0 0 0 0 0 0 0 5 0	Injection current / mA	Detector voltage/ V											
0.5 0.008218269 Detector Voltage as a function on rejection current for 405 min aser dicket 1 0.009248511 1 1 2 0.0092165422 1 1 3 0.03662570343 1 1 4 0.0127657043 1 1 1 5 0.0124951317 0 0 1 1 5 0.0124951317 0 0 5 10 15 20 25 30 6 0.0699486974 0 5 10 15 20 25 30 7 0.0124951317 0 5 10 15 20 25 30 7 0.012495137 0 5 10 15 20 25 30 7 0.012495147 0 5 10 15 20 25 30 8 0.06593468673 0 5 10 15 20 25 30 11	C	0.0086744472		Datas	tor voltage	oo o functiv	on of inicati	on ourront	for 40E pm	loc or diodo			
1 0.0192248511 2 0.009815402 2.5 0.0123237 3 0.0366257043 4 0.035670337123 5 0.0224513716 6 0.0569648744 6.5 0.0049153746 6.5 0.00494951371 7.5 0.036837437 8 0.056946874 9 0.0755348247 9 0.0755348247 10 0.1078305292 11 0.068871712 11 0.06887737 12 0.075348425 13 0.113744548 14 0.1352926461 12.5 0.094469573 13 0.1137441548 14 0.1352926491 15 0.084406764 14 0.1352926491 15 0.084406764	0.5	0.0082193269		Jelei	tor voltage	as a function	on or injecti	oncurrent	101 405 1111	laser uluue			
1.5 0.0098312704 2 0.0092165482 3.5 0.035670547 4.5 0.025603712 5 0.00442576 5.5 0.00442576 6.5 0.0124651317 6.5 0.0124691317 7 0.012469297 7.5 0.0058032647 8 0.0550437123 9 0.55 10 15 20 5 11 0.012469297 0 5 12 0 13 0.0755344425 11 0.068047122 11 0.01848297 12 0 13 0.0755344425 11 0.068047122 12 0.0755344425 13 0.017952466 12 0.075534445 13 0.017952466 12 0.0775524865 13 0.017952466 14 0.15926296491 15 0.064640764 14 0.15926296491 15	1	0.0192248511		2							_		
2 0.0092165482 3 0.0366257043 3 0.0366257043 4 0.0187607337 4 0.0187607337 5 0.0124951317 5 0.0024951317 5 0.0049493977 7 0.011244929 7,5 0.0363837437 8 0.052042422 0 5 10 15 20 25 30 10 15 20 25 30 10 15 20 25 30 10 15 20 25 30 10 15 0.025 30 10 15 20 25 30 10 15 0.025 30 10 15 0.025 30 10 15 20 25 30 10 15 20 25 30 10 15 0.05204262 10 0.05204262 10 0.05204265 11 0.068893065 11 0.0688937172 12 0.1078345247 13 0.0179345247 14 0.0688937172 15 0.0984886753 11 0.068887437 15 0.098488753 11 0.06888753 13 0.017910246 14 0.13282929 15 0.098488753 15 0.098489750 15 0.098489750 15 0.098489750 16 Lung 2021 3 / 1	1.5	0.0098831704		1.8 1.6						•			
2.5 0.01233237 3.5 0.035670347 4. 0.01260527043 4. 0.035670347 4. 0.025607123 5. 0.0024951317 5. 0.00249412574 6. 5. 0.004939977 7. 5. 0.0650944257 7. 0.0112842927 0. 5. 0.00124862 8. 5. 0.00509412574 7. 0.0112842927 0. 5. 0.0012462 8. 0.052042452 9. 0.55 0.0059412574 1. 0.052042452 0. 5. 0.0012462 1. 0.0753415247 1. 0.052042452 0. 5. 0.0012462 1. 0.0753415247 1. 0.0686939055 1. 0.0753415247 1. 0.0686939055 1. 0.0753415247 1. 0.0686939055 1. 0.0753415247 1. 0.0686939055 1. 0.0113741548 1. 0.0984586753 1. 0.00945847764 1. 0.0094586753 1. 0.00945847764 1. 0.0094586753 1. 0.0094586755 1. 0.0094586755 1. 0.0094586755 1. 0.00945	2	0.0092165482											
3 0.03662570347 3.5 1.4 0.0187607337 3.5 1.4 0.0187607337 3.5 4 0.0187607337 5.5 0.12 5 0.0124951317 0.6 0.8 0.6 5 0.0094125746 0.6 0.6 0.4 6 0.06594687147 0.7 0.1 7 0.012492927 0.5 0 5 10 15 20 25 30 8 0.0520424262 0.5 0 5 10 15 20 25 30 9 0.0755348227 0 5 10 15 20 25 30 10 0.1078305247 0 5 10 15 20 25 30 11 0.0689390659 0 5 10 15 20 25 30 10.5 0.0278206357 0 5 10 15 20 25 30 11 0.007820486733 0 14 0.107910246 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	2.5	0.01233237							-	,	_		
3.5 0.0365070347/ 4 0 12 4 0.0136070337 0 12 5 0.0266437123 0 0 5 0.0094125746 0 0 6 0.0669946874 0 0 6 0.0669946874 0 0 7 0.012649297 0 5 10 15 20 25 30 7 0.012649297 0 5 10 15 20 25 30 8 0.0520424262 0 5 10 15 20 25 30 9 0.0753915247 0 5 10 15 20 25 30 10 0.107336262 0 5 10 15 20 25 30 11 0.06868737 0 5 10 15 20 25 30 12.5 0.0984586753 1 1 10 10 10 10 10 10 10 113 113 113 113 1137441548<	3	0.0366257043	>	1.4									
4 0.0187607337 0 1 5 0.012695437123 0 0 0 5 0.0094125746 0 0 0 0 6 0.065984687147 0 5 10 15 20 25 30 7 0.012849297 0 5 10 15 20 25 30 7 0.012849297 0 5 10 15 20 25 30 8 0.055348425 0 5 10 15 20 25 30 9 0.0755348425 0 5 10 15 20 25 30 10 0.1078362627 0 5 10 15 20 25 30 11 0.00688371712 10	3.5	0.0356703547	6	1.2							_		
4.5 0.0266437123 5 0.0266437123 5 0.0094125746 5.5 0.0094125746 0 0.6 0.6 0.6 0.6 6.5 0.0084839897 0.6 0.2 0 0.1 15 20 25 30 7.5 0.0084938977 0 5 10 15 20 25 30 8.5 0.0520424262 0 5 10 15 20 25 30 9.5 0.0755345247 0 5 10 15 20 25 30 10.5 0.0278206357 0 5 10 15 20 25 30 11.5 0.11597507 11 0.068871712 11 14 14 14 14 14 14 15 15 15 16 16 16 16 16 16 16 16 16 16 16 16 16 16 15 16 16 16 16 16 16 16 16 16 16 16 <td>4</td> <td>0.0187607337</td> <td>tag</td> <td>1</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	4	0.0187607337	tag	1									
5 0.0124951317 8 0.8 6 0.06909418746 0.4 0.4 6.5 0.0049398977 0.4 0.4 7 0.0112849297 0.5 10 15 20 25 30 8 0.055044252 0 5 10 15 20 25 30 8.5 0.0689390569 injection current /mA 11 11 15 21 12 10 15 20 25 30 11 10.5 0.0753815247 0 5 10 15 20 25 30 11 11.0 0.068939772 10 12 10 12 10 12 11 <td< td=""><td>4.5</td><td>0.0256437123</td><td>2</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>	4.5	0.0256437123	2	1									
5.5 0.0094125746 6 0.6 6 0.069948874 0.2 0.2 7 0.0112849297 0.2 0.2 7.5 0.0363837437 0.5 10 15 20 25 30 8 0.0520424262 0 5 10 15 20 25 30 9 0.0755345247 0 5 10 15 20 25 30 10 0.107832629 0 5 10 15 20 25 30 11 0.0753345247 0 5 10 15 20 25 30 11 0.075345247 0 5 10 15 20 25 30 11 0.075345247 0 5 10		0.0124951317	ctor	0.8					•*		_		
6 0.0669948874 G 0.4 6.5 0.044939877 0.2 0.4 0.2 0.5 0.0 0.2	5.5	0.0094125746	ete	0.6									
6.5 0.0049393977 0.2 7 0.0112849297 0 8 0.055042425 0 9 0.0755345247 0 9 0.0755345247 1 10 0.0755345247 1 11 0.0688390659 1 12 0.1073835247 1 13 0.1073835247 1 14 0.1078365247 1 15 0.0686386733 1 11 0.068686753 1 12 0.1073524665 1 13.5 0.113741548 1 14.5 0.1528296491 1 15 0.0684040764 1 14.5 0.1528296491 1 15 0.068400764 1 14.5 0.1528296491 1 15 0.06400764 1 14.5 0.1528296491 1 15 0.06400764 1		0.0659846874	Ō	0.4									
7 0.011284297 0.2 7.5 0.065042482 0 5 10 15 20 25 30 8 0.052042482 0 5 10 15 20 25 30 9 0.0755345247 0 5 10 15 20 25 30 10 0.1078326292 0 11 0.06887172 0 10 11 0.06887172 0 10 11 11 11 11 0.06886713 0 0 11	6.5	0.0048938977		0.2									
7.5 0.0363337437 0 0 1		0.0112849297		0.2			****						
8 0.05204/2452 0 5 10 15 20 25 30 9 0.0755345247 injection current /mA injection cur	7.5	0.0363837437		0	******			Ť.	-	1	-		
8.5 0.088930059 Injection current /mA 9 0.0755348425 9.5 0.075348425 10 0.1078326292 10.5 0.027820357 11 0.0688971712 12 0.1075326465 13 0.007910246 14 0.13268573 15 0.1137341548 14.5 0.1528296491 15 0.064604764	8	0.0520424262			0	5	10	15	20	25	30		
9 0.0755345247 9.5 0.0755345247 10 0.1078305242 10 5. 0.0278203577 11 0.0688817112 11 0.068881712 12 0.1078524865 12.5 0.084458655 13 0.0107910246 13.5 0.1137441548 14 0.138925469 15 0.0644604764 15 0.0646404764 16 June 2021 3 / 1	8.5	0.0889390659		Injection current / mA									
9.5 0.0753815247 10 0.1078336292 10.5 0.0278206357 11 0.0686871712 11.5 0.1115197507 12 0.1073524865 13 0.0107910246 13.5 0.113741548 14 0.1309511722 15 0.064604764 15 0.05546404 16 0.13292007 17 Dett Ščaćanak 16 Lup 2021 3.4 1	9	0.0755348425											
10 0.107836282 10.5 0.0278206357 11 0.0688871712 11.5 0.1115197507 12 0.1073524865 13.5 0.013730246 13.5 0.1137441548 14 0.130951172 14.5 0.1582956491 15 0.0584640764 15 0.0584900764 15 0.0584900764	9.5	0.0753815247											
10.5 0.027820357 11 0.0688971712 11.5 0.1115197507 12 0.1073524865 13 0.007910246 13.5 0.1137341548 14 0.1309511722 14.5 0.1528296491 15 0.046404764 15 0.046404764	10	0.1078326292											
11 0.0688871712 11.5 0.1115197507 12 0.1073524865 12.5 0.0984586753 13 0.0107910246 13.5 0.1137341548 14 0.1309511722 14.5 0.1528296491 15 0.064604764	10.5	0.0278206357											
11.5 0.1115197607 12 0.1073524865 12.5 0.0884586753 13.5 0.113741548 14.5 0.132920464 14.5 0.1582920491 15 0.064640764 15 0.064640764 16 Lung 2021 3.6 110000	11	0.0688871712											
12 0.1073524865 12.5 0.094586753 13 0.0107010246 13.5 0.113741548 14 0.1306511722 14.5 0.1558296491 15 0.064604764 15 0.0139007 Date Šešánánk 16 Lung 2021 3 / 1	11.5	0.1115197507											
12.5 0.0844586753 13 0.0107910246 13.5 0.113741548 14 0.1309511722 14.5 0.1529296491 15 0.064640766 15 0.064640766 15 0.064640766	12	0.1073524865											
13 0.0107910246 13.5 0.1137341548 14 0.1368511722 14.5 0.1552296491 15 0.064404764 15 0.04404764 16 Une 2021 3.6 1312920102	12.5	0.0984586753											
113 0.113741548 14 0.1309511722 14.5 0.1528296491 15 0.064604764 Date Šešánánek 11000000	13	0.0107910246											
14 0.1308511722 14.5 0.1528296491 15 0.0646404764 Detc Šenának Towards reproducible deta management 16 June 2021 3 / 1	13.5	0.1137341548											
14.5 0.1528296491 15 0.0646404764 Dete Šenának 11000000 Dete Šenának 11000000 3 / 1	14	0.1308511722									_		
15 0.0646404764 Det Štánának 171020102 Det Štánának 16 june 2021 3 / 1	14.5	0.1528296491											
Petr Štopánek Towards reproducible data management 16 lune 2021 3 / 1	15	0.0646404764											
Part Statement In line /1/1 5 / 1	Deta Šta	0 1712020107			nya duaible	data ma			16	June 201	1	2/	1

Mindset

- 1 What was the last time you read an article with a description of an experiment and found it difficult to replicate because it was not well described?
- 2 Try to avoid that

The basic principle

- **1** Record everything
- 2 What, why, when, by whom, what were the conditions
- 3 It is better to have too much than too little
- 4 Store it in a findable way already during the production

Type of data we produce:

- 1 Instrumental measurements
 - e.g., ordered sets of values [f, t, V, I, ...]
 - Experiment description
 - Experiment settings (parameters)
- 2 Results of the theoretical modelling
- Case study: Measurement of spectroscopic effects using a custom-built instrument

Main challenge: a custom-built instrument with many variables



Main challenge: a custom-built instrument with many variables



Description of the experimental parameters

- Digital (can be automatized)
 - Magnetic fields and its shape
 - Detection frequencies
 - Laser intensity
 - Laser temperature
 - ...
- Analog (depend on user)
 - Concentrations
 - Flow rates
 - Solvents
 - Physical properties (tubing diameters)
 - Optical setup
 - Instrument models
 - ...

Data stored in a permanent hierarchy /measurements/YYYMMDD/ Each directory contains

- $1\,$ A text file (.odt) with the experiment description
 - What and why is being measured
 - Hypotheses tested
 - Methodology
 - Description of the experimental parameters
- 2 Actual data measurement file(s). Each file contains
 - What and why is being measured
 - A header with description of the experimental parameters
 - Actual data ([f,t, V, ...] in binary format)
- \Rightarrow Redundancy in parameters description
 - 3 Processed data (.ods)
 - Check of data completeness

Additional documentation: Photography

Handling the data: Instrumental measurements

Data stored in a file structure (TDMS)





- Proprietary format (National Instruments); plugings available (Excel, MATLAB, Python, ...)
- Mainly optimized for recording of continuous data, but customizable
- Each channel automatically has a property field

- Production of data
 - Record data itself
 - Describe the data and parameters (in the file and externally)
 - Backups
- Publication
 - (Format conversion + validity check)
 - Complete metadata (authors, keywords, what kind of data)
 - Record metadata (15 minutes in QVAIN)
 - Cite in the article
 - Deposit (in repository) (15 minutes, depending on the level of data reorganization)

- Try imagining what would someone need to replicate your experiment
- Try to set up your experiment assuming no prior knowledge
- Show your results to someone not familiar with your project