

# SMART LABS, AUTOMATION & TECHNOLOGY EUROPE

## Technical and Data Requirements of Digitalising Scientific Research

Dr Samantha Kanza  
University of Southampton



# OUTLINE

01

- Introduction

02

- Landscape of Digitisation

03

- Barriers to Digital Research

04

- Technical & Data Requirements



# 01 INTRODUCTION

- About Me
- My PhD
- Motivation
- PSDI
- Digital Research



Dr Samantha Kanza

## ABOUT ME

- Senior Enterprise Fellow at the University of Southampton
- Coordinates AI4SD & Future Blood Testing Network
- Digital & Semantic Web Researcher for PSDI Project
- Research Interests: Semantic Web Technologies, IoT, Research Data Management, Digitisation, Lab of the Future, Paperless Labs, Re-use of Technology



[www.southampton.ac.uk/people/5xm57p/doctor-Samantha-kanza](http://www.southampton.ac.uk/people/5xm57p/doctor-Samantha-kanza)



[s.kanza@soton.ac.uk](mailto:s.kanza@soton.ac.uk)



[@SamiKanza](https://twitter.com/SamiKanza)



[linkedin.com/in/skanza](https://linkedin.com/in/skanza)



# PhD & 1<sup>st</sup> PUBLICATION

What Influence would a Cloud Based Semantic Laboratory Notebook have on the Digitisation and Management of Scientific Research?

**BMC** Part of Springer Nature

Search  Menu

## Journal of Cheminformatics

Home About [Articles](#) Submission Guidelines About The Editors Calls For Papers

Download PDF

Research article | [Open Access](#) | Published: 24 May 2017

### Electronic lab notebooks: can they replace paper?

[Samantha Kanza](#) , [Cerys Willoughby](#), [Nicholas Gibbins](#), [Richard Whitby](#), [Jeremy Graham Frey](#), [Jana Eriavec](#), [Klemen Zupančič](#), [Matjaž Hren](#) & [Katarina Kovač](#)

[Journal of Cheminformatics](#) **9**, Article number: 31 (2017) | [Cite this article](#)

14k Accesses | 20 Citations | 78 Altmetric | [Metrics](#)

#### Abstract

Despite the increasingly digital nature of society there are some areas of research that remain firmly rooted in the past; in this case the laboratory notebook, the last remaining paper component of an experiment. Countless electronic laboratory notebooks (ELNs) have been created in an attempt to digitise record keeping processes in the lab, but none of them have become a 'key player' in the ELN market, due to the many adoption barriers that have been identified in previous research and further explored in the user studies presented here. The main issues identified are the cost of the current available ELNs, their ease of use (or lack of it) and their accessibility issues across different devices and operating systems. Evidence suggests that whilst scientists willingly make use of generic notebooking software, spreadsheets and other general office and scientific tools to aid their work, current ELNs are lacking in the required functionality to meet the needs of the researchers. In this paper we present our extensive research and user study results to propose an ELN built upon a pre-existing

nature

Explore content  About the journal  Publish with us  Subscribe  Sign up for alerts  RSS feed

TOOLBOX | 06 August 2018

### How to pick an electronic laboratory notebook

Choosing wisely from a burgeoning array of digital tools can help researchers to record experiments with ease.

[Roberta Kwak](#)

[Twitter](#) [Facebook](#) [Email](#)



Illustration by The Project Twins

Since at least the 1990s, articles on technology have predicted the imminent, widespread adoption of electronic laboratory notebooks (ELNs) by researchers. It has yet to happen – but more and more scientists are talking the plunge.

One barrier to uptake is the wide range of products available. ELNs comprise software that helps researchers to document experiments, and that often has features such as protocol templates, collaboration tools, support for electronic signatures and the ability to manage the lab inventory. But the ELN market encompasses considerable variety; a study conducted in 2016 by the University of Southampton, UK, identified 72 active products ([S. Kanza et al. J. Cheminformatics 9, 31; 2017](#)). "It's just insane," says Sian Jones, a petroleum engineer at the Delft University of Technology in the Netherlands. "It does become very confusing." And many researchers simply lack the time or motivation to make the move to ELNs.

Access this article via [University of Southampton](#)

Download PDF

#### Subjects

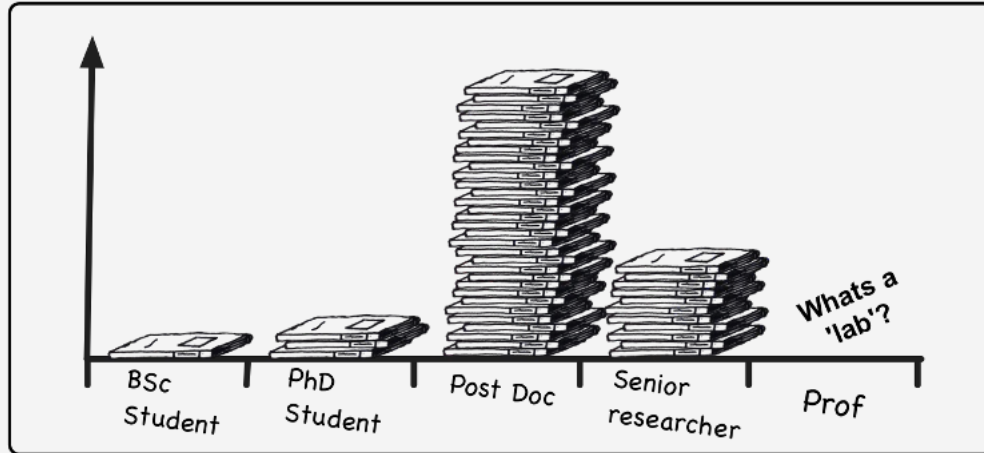
[Lab life](#) [Research management](#) [Research data](#)

[Technology](#)



# Motivation: PEOPLE ARE STILL USING PAPER?!

Lab book use at various levels of academia



"Lab book use at various levels of academia" by ErrantScience.com is licensed under [CC-BY-NC](https://creativecommons.org/licenses/by-nc/4.0/)

## Advantages of Paper

- Portable
- Robust
- Can be securely stored
- Don't need a power supply
- Ease / Flexibility of data entry

## Disadvantages of Paper

- Easy to lose/destroy
- Harder to search
- Harder to backup
- Harder to share
- Harder to readily access



"Electronic Lab Notebooks are great, but not on vacation" Cartoon by Phil Johnson for MIT

## Advantages of ELNs

- Accessible
- Searchable
- Easy to backup
- Easy to share data
- Data is quickly retrievable

## Disadvantages of ELNs

- Slow / inflexible data entry
- Requires power supply
- May require internet
- Concerns about tech in lab
- Harder to create diagrams

# PHYSICAL SCIENCES DATA INFRASTRUCTURE

## Case Study 6: Process Recording & Digital Research Notebooks

Aim: Assess process recording requirements and the associated digital landscape. Investigate Digital Research Platforms (DRP) and evaluate their suitability as generic recording systems to support diverse workflows



**PSDI**  
PHYSICAL SCIENCES  
DATA INFRASTRUCTURE



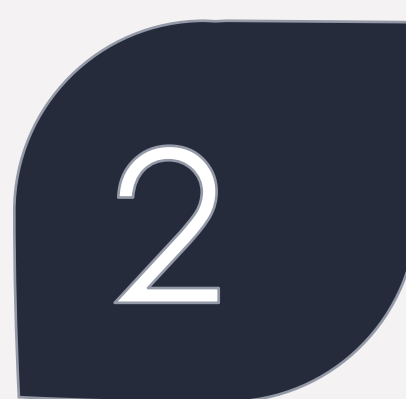
### **The Vision of PSDI**

PSDI aims to accelerate research in the physical sciences by providing a data infrastructure that brings together and builds upon the various data systems researchers currently use.

# DIGITAL RESEARCH

## LITERATURE REVIEW

- Systematic Literature Review of different aspects of digital technologies in the Physical Sciences



## QUALITATIVE RESEARCH

- Research to understand the current landscape of digitisation and use of technology
- Focus Groups with different science groups (Physics, Chemistry, Biology)
- Ethnography in different chemistry laboratories
- Surveys of the Physical Sciences Community



## SOFTWARE INVESTIGATIONS

- Investigate usage of non ELN software for Chemists
- ELN Landscape assessment

## TECHNICAL PROTOTYPES

- Semantic tagging and annotation of scientific documents
- Smart Lab prototyping



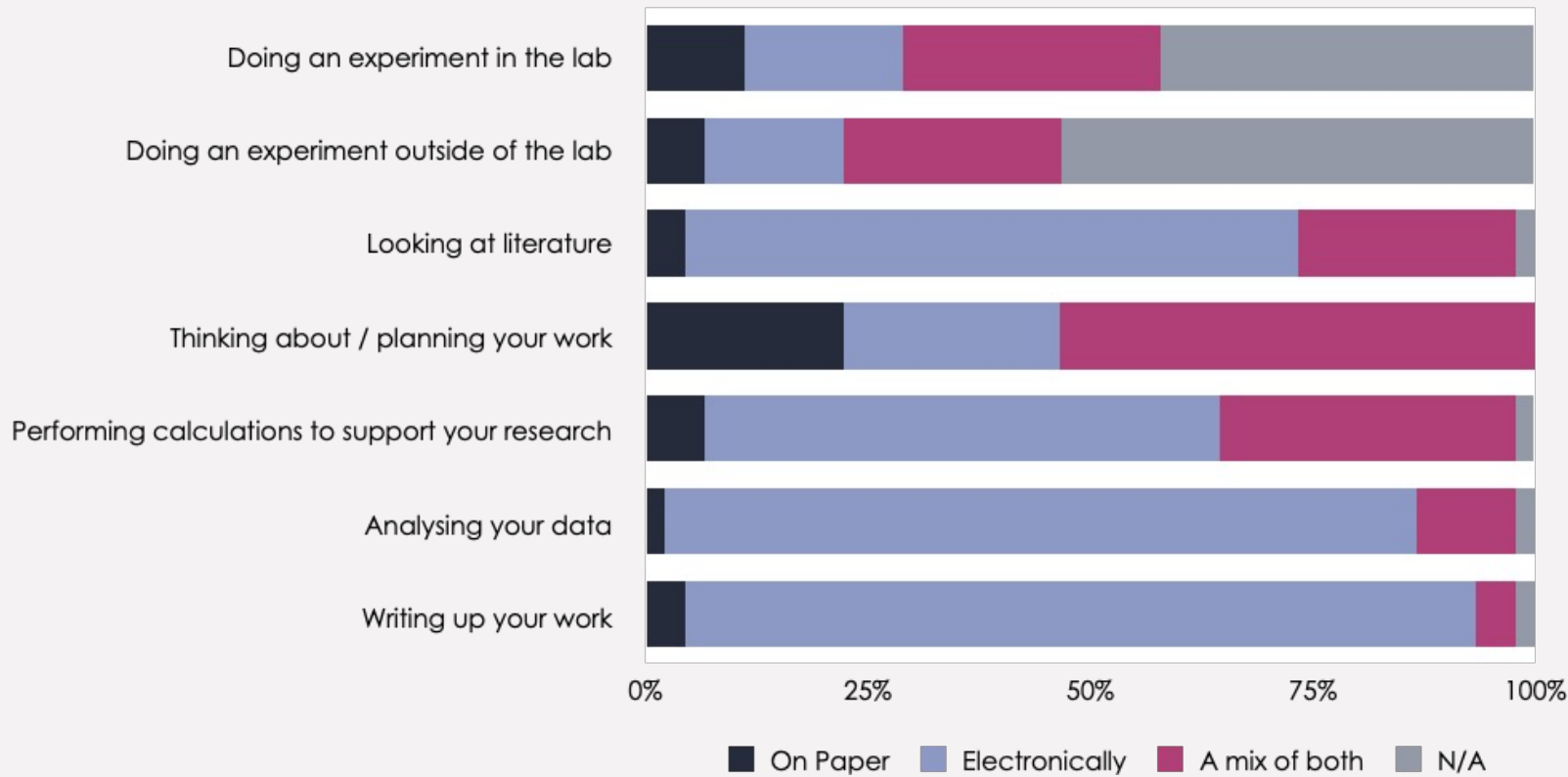


# 02 CURRENT LANDSCAPE

- Use of Paper & Electronic
- Organising and Linking
- ELN Landscape Assessment
- Use of Notebooking Software
- Use of Data Sharing Software
- Use of Other Software
- Smart Lab Prototype
- Changes since COVID-19

# USE OF PAPER & ELECTRONIC

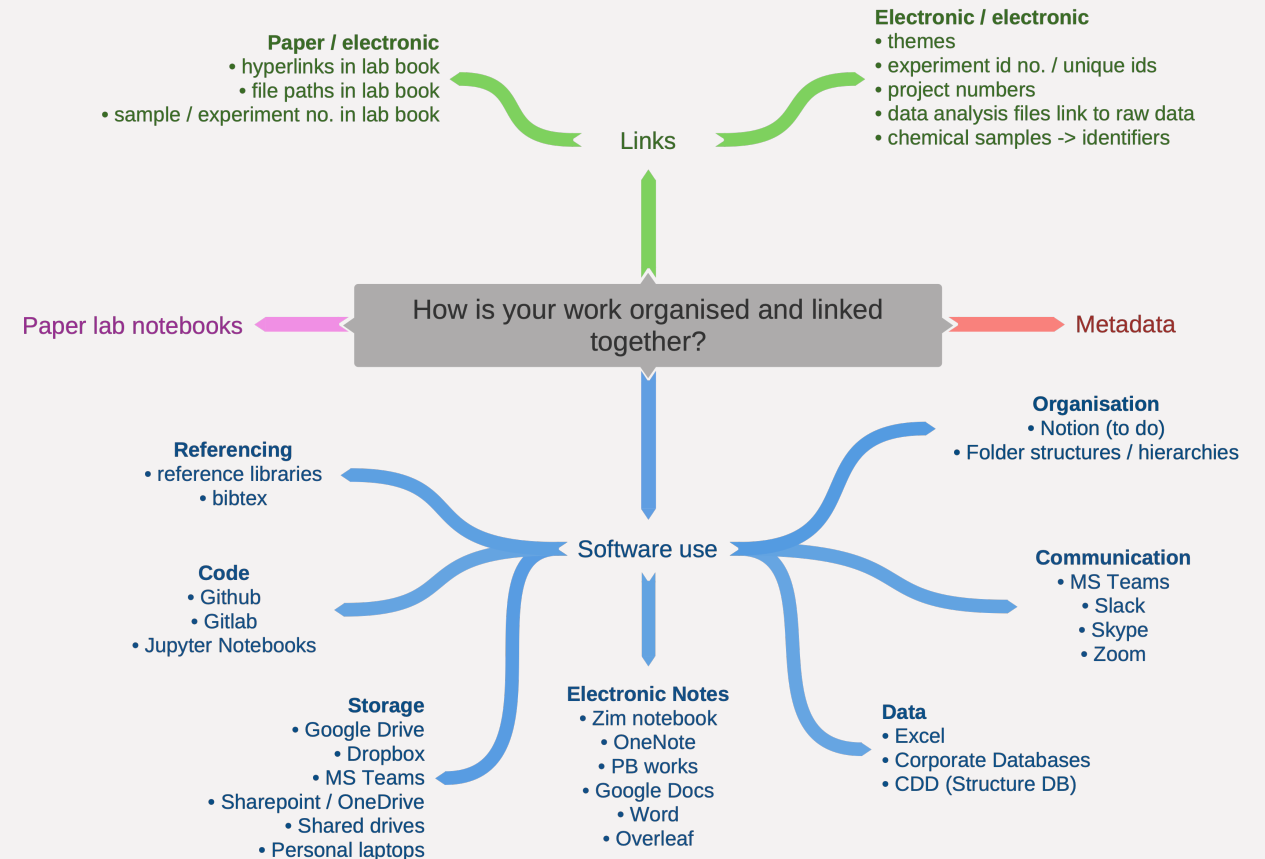
For each of the following types of work, how do you currently record it?



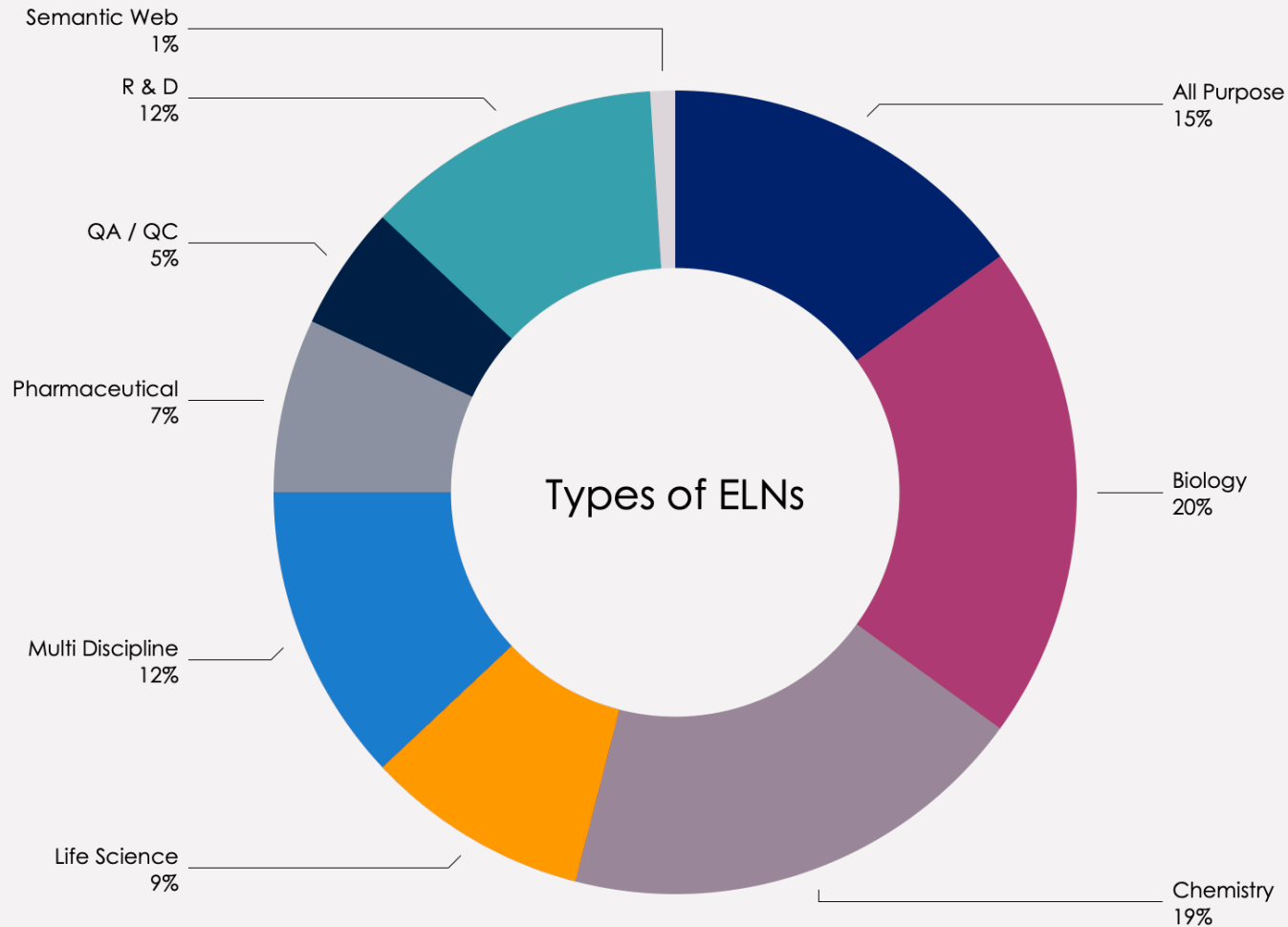
- Researchers work in different ways using a mix of paper/electronic methods
- Paper still more used for planning, with a heavy reliance on computational methods for analysis and writeup
- Reduction in only using paper

# ORGANISING & LINKING

- Necessity to link between paper and electronic demonstrates a use of paper
- Clear pattern of using codes/links to bridge this gap
- High level of software usage to organize and link work



# ELN LANDSCAPE ASSESSMENT



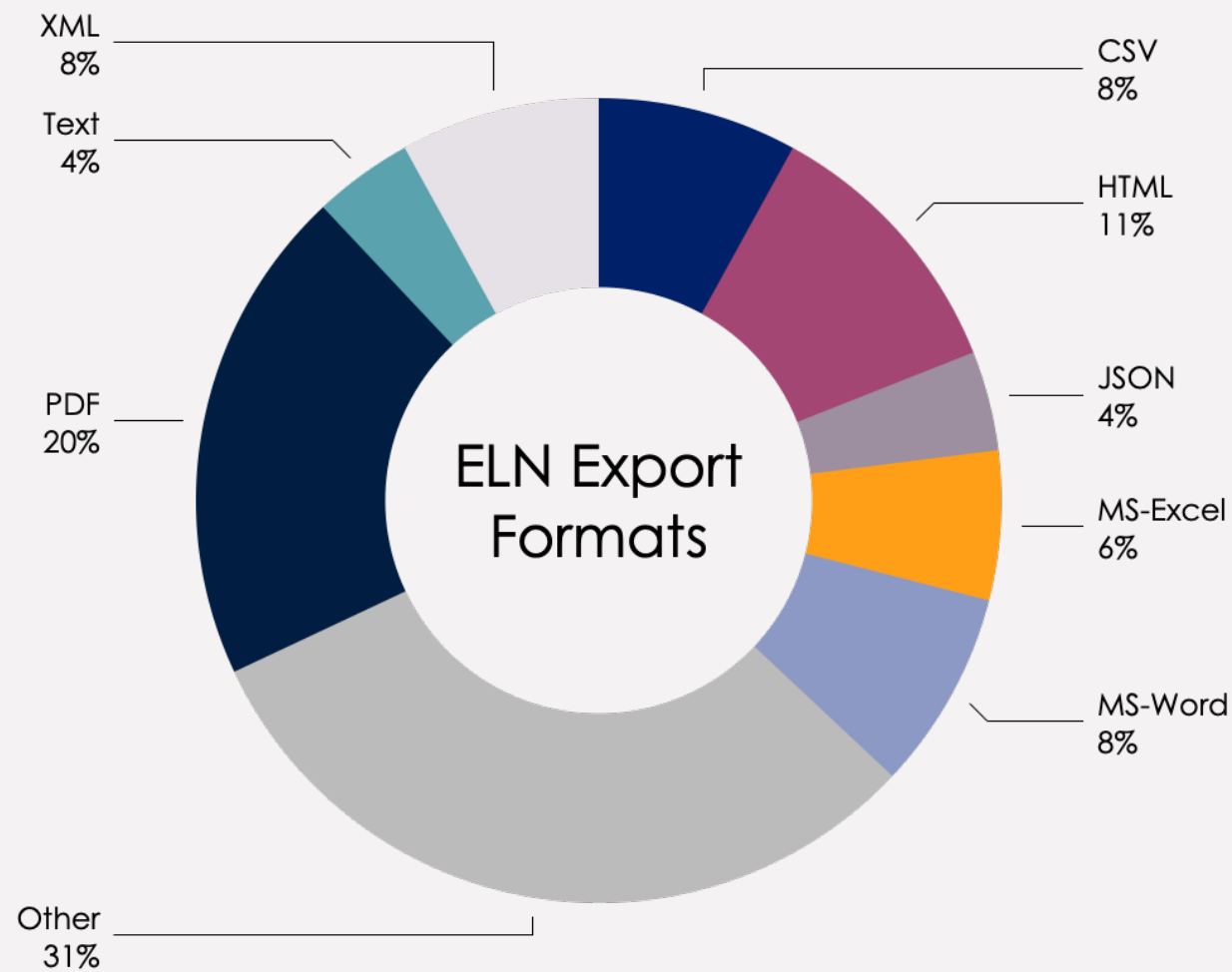
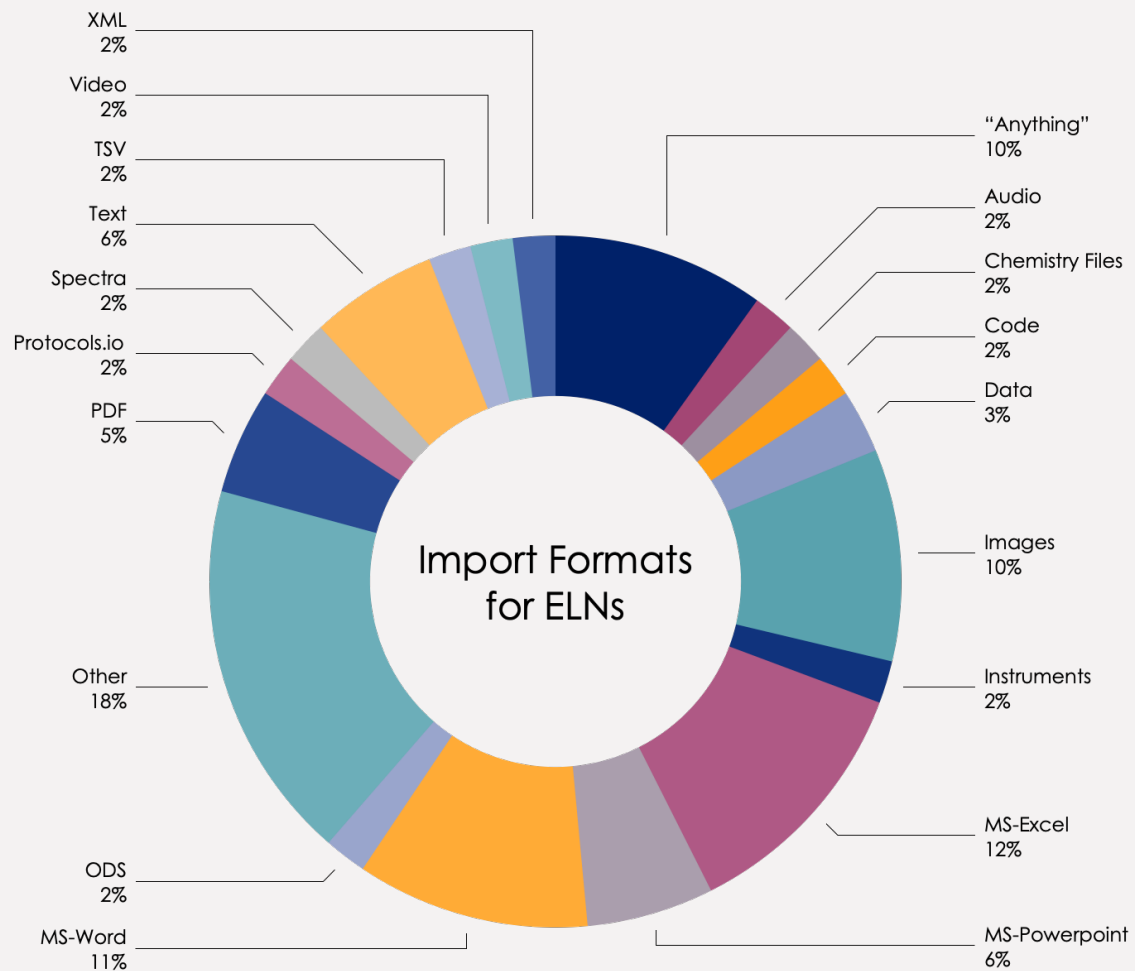
● > 120 ELNs have been created in the last 2 decades

● Over 30 have become inactive

● Over 80 on the market



# ELN IMPORT/EXPORT FORMATS



# USE OF NOTEBOOKING SOFTWARE

Do you use a Digital Research Notebook?

- 55% - Yes
- 45% - No



# DATA SHARING SOFTWARE

<b><u>Communication</u></b> MS Teams Jitsi Discord Slack Zoom Skype Mattermost	<b><u>Organisation / Decision making</u></b> Notion Mural MS Teams Asana	<b><u>Domain software (structures)</u></b> Avogadro Chemoffice Diamond Mercury
<b><u>Notetaking</u></b> Google Docs Lyx Word OneNote Overleaf Notion	<b><u>ELNs</u></b> CDD Science Cloud Lab Archives	<b><u>Data Analysis</u></b> Spreadsheets Origin
<b><u>Code</u></b> Bash scripts Python routines Mercurial Git(hub) Gitlab Bitbucket*	<b><u>Cloud Storage</u></b> Sharepoint/ OneDrive MS Teams Dedicated Cloud Dropbox Google Drive	<b><u>Bespoke software</u></b>

Increase in use of communication and organizational software! ●

Switch to MS Teams/Office 365 (not necessarily COVID!) ●

Higher use of software/scripts to share code ●

# USE OF OTHER SOFTWARE

Category	Totals (/206)	Percentage
Crystallographic Software	26	12.44%
Coding Software	22	10.53%
Molecular Modelling & Simulation Software	22	10.53%
Quantum Chemistry and Solid State Physics Software	21	10.05%
Data Visualisation & Analysis	19	9.09%
General document processing	18	8.61%
Other	13	6.22%
Spectroscopic Software	10	4.78%
Image processing Software	9	4.31%
Chemical Database & Informatics Software	8	3.83%
Organisational Software	7	3.35%
Chemistry Bibliographic Databases	5	2.39%
Database Software	5	2.39%
Instrument Control	5	2.39%
Simulation (non-chemical)	5	2.39%
Communication Software	4	1.91%
Molecular Editor Software	3	1.44%
Nanostructures Modelling Software	2	0.96%
Machine Learning	2	0.96%
CAD Software	2	0.96%
Workflow Software	1	0.48%

## #SOFTWARE PACKAGES

>200 different software packages identified

## CATEGORIES

Categorised using categories from PhD research and identified additional categories

## DIVERSITY

Demonstrates the wide ranging need for generic and specialist software in the physical sciences



# DIGITAL SMART LAB

## VOICE

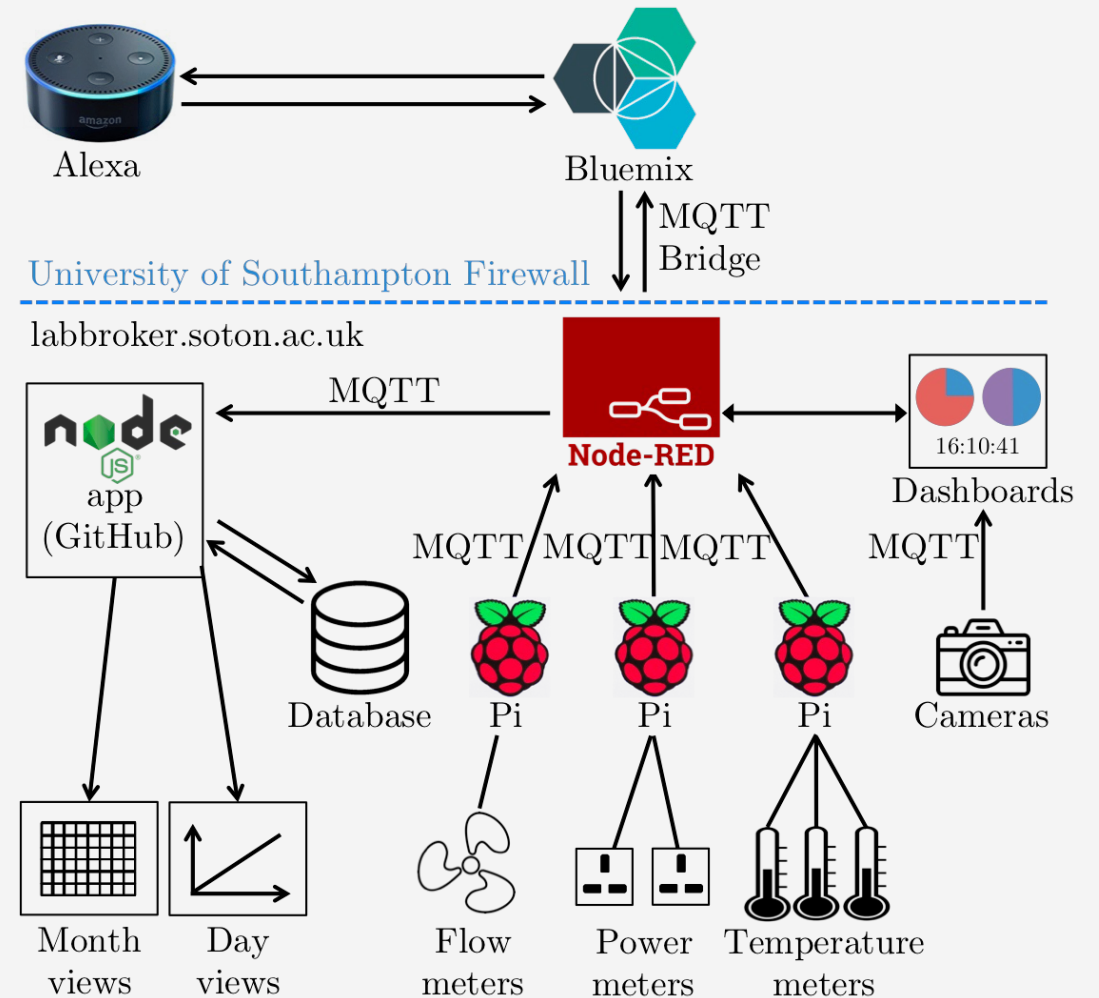
- Useful in the laboratory, but only if it functions correctly
- Users want voice control as well as access
- Noise/Chemical Names is an issue

## WARNINGS

- Users want warnings from the lab that will save them time. E.g. your experiment is failing

## VISUAL EFFECTS

- Dashboards and camera feeds are useful
- Dashboards provide lots of info in one go
- Cameras enable multiple views of the lab and can help with equipment setup



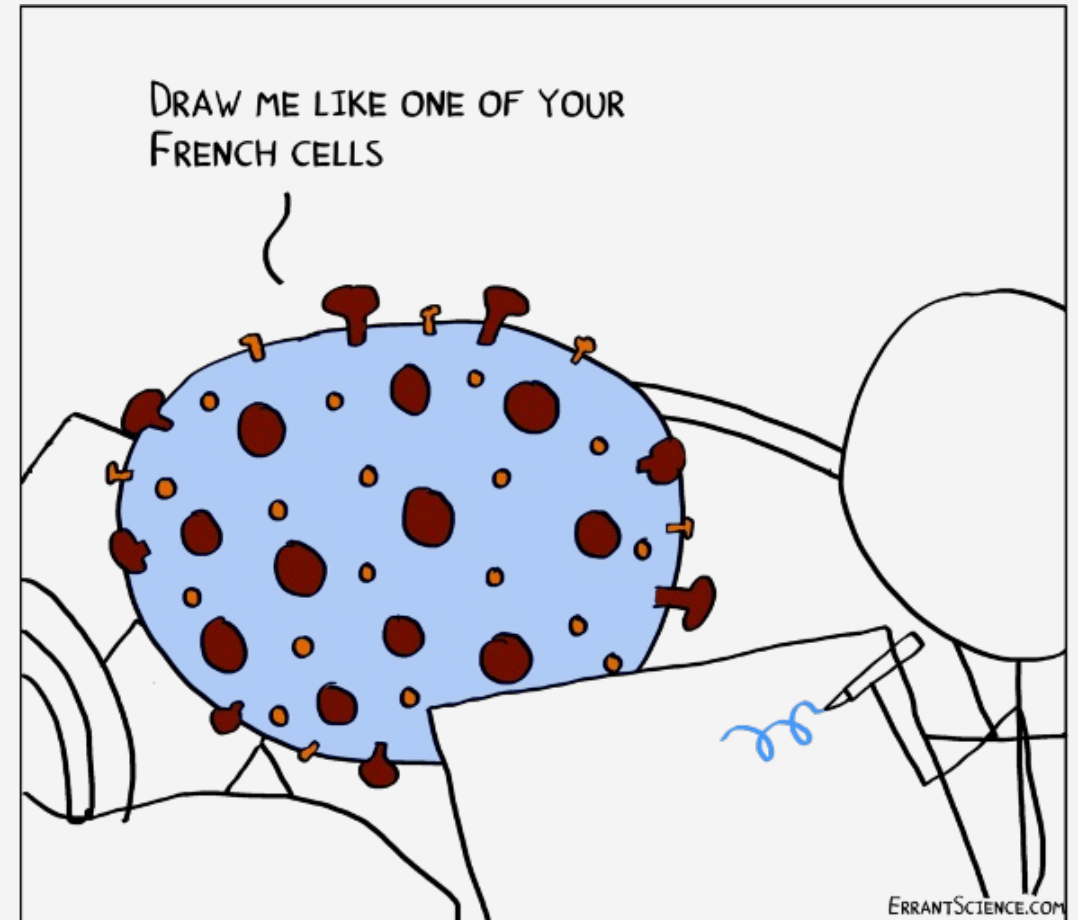
# COVID-19 CHANGES

COVID-19 has led to an increased use of digital tools to communicate, digitize work and share data in an electronic form. Noticed a higher prevalence of companies working on lab voice assistants.

The uptake of software usage reported via our survey is mainly centered around:

- Notetaking software
- Task Management software
- Literature Linking
- Software/Code support
- Domain based solutions

We still don't have a viable solution to increasing digitization!

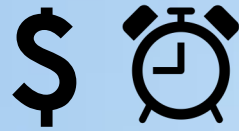




# 03

## BARRIERS TO DIGITAL RESEARCH

- Logistical Barriers
- People Barriers
- Data Barriers
- Standards Barriers
- ELN Barriers
- Hardware & Software Barriers



# LOGISTIAL BARRIERS

## Cost

- Lack of funding
- Cost of Research
- Cost of software /OA Fees

## Time

- Lack of time for projects
- Time to learn and use systems
- Current systems in place making digitising everything very time consuming



# PEOPLE BARRIERS



## Attitude

- People are afraid of change
- Unwilling to adopt new systems
- Knowledge gets lost with people

## Training

- Researchers need to be trained to use new systems

This is a socio-technical issue. People are just as important as technology



# DATA BARRIERS

## Un-FAIR Data

- Frequently people are unable to locate, access or re-use data

## Metadata/Provenance

- Data often lacks context
- Time consuming to capture metadata
- Results in no provenance

## Data Size

- Large datasets are harder to store/share

# STANDARDS BARRIERS

## Range of Data Standards

- Lots of different standards/formats
- Not always possible to convert between
- Much software offers a wider range of import than export formats
- Concern about being locked into proprietary formats

## Lack of Adoption of Standards

- Still too many cases of data not adhering to standards

# ELN Barriers

Category	Barriers	% of 169
Accessibility: Use in and out of the lab	You'd need to enter data in both the lab and write-up area	74
	No easy access to appropriate hardware in the lab	12.5
Ease of use: Not as easy as paper	ELN was too difficult to use	22
	Does not capture the right information for me	7
	Difficult to capture some kinds of information in an ELN	80
ELN attitude: Requires change of attitude from higher ups / from the onset	Only makes sense if the whole department adopts it	20
	Belief that students/post docs would resist adoption	11
Cost: People don't want to pay	Up front costs and licensing fees	74
	Additional infrastructure costs (e.g. computers)	27
	Future development and costs of applications	90
	On-going costs of the system	93
Data Portability: Data can't always be moved between notebooks or machines	Data will be tied into a commercial package	84
Other	Other	11



# HARDWARE & SOFTWARE BARRIERS

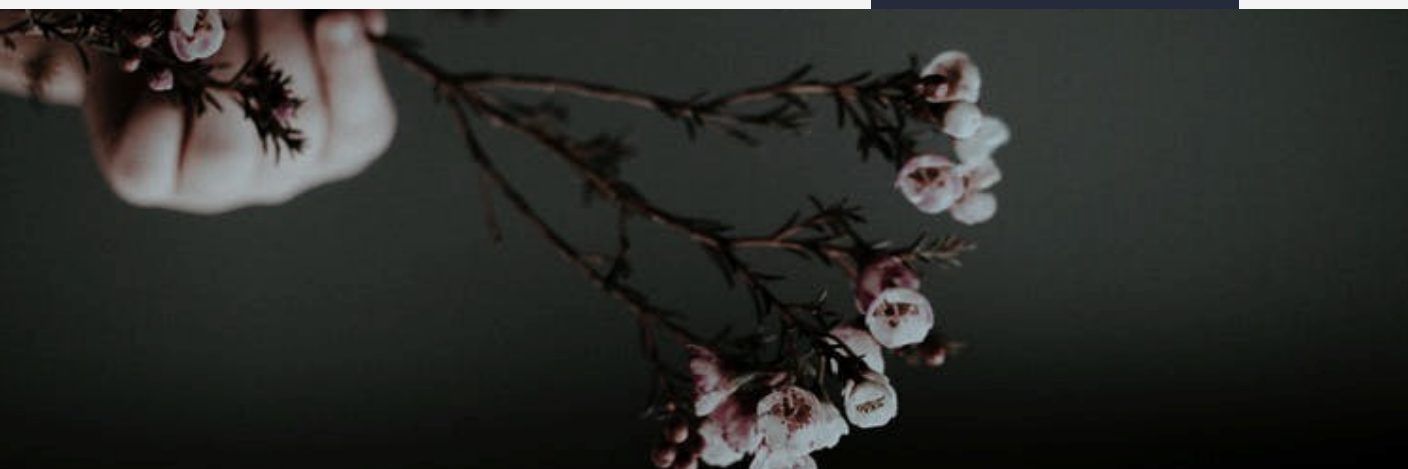
Insufficient data storage capabilities

Software

- We need more modern compatible software
- Concerns about security of different software

Hardware

- Legacy equipment requires legacy software / outdated data formats



# 04

## TECHNICAL & DATA REQUIREMENTS

- What do users want?
- Software Integration
- Data Standards and Management
- Semantic Enrichment
- The Lab of the Future
- Our Vision of a PSDI Infrastructure



# WHAT DO USERS WANT?

Feature Category	Description
Generic	API Access, Automation, GUI, Localisation, Remote Access, Synchronisation
Notebooking	Content Support, Interaction/Access, File Links, Organisation/Reconfiguration, Paper Integration, Referencing/ Literature, Word Processing
Data	Access, Conversion, Exchange, Integration, Management, Quality, Retention, Security, Standards, Support, FAIR, Identifiers, Provenance
Publishing & Sharing	Documentation & Instructions, DOIs, Export, Licensing, Open Access, Publishing, Sharing, Social Media, Researcher Attribution, Repositories
Collaboration & Management	Auditing, Comments, Notifications, Subscribe, Team Management
Domain Based Features	Chemical/Molecules, Default Lists, Equipment Interface, Experiment Planning/Recording, Health & Safety, LIMS/ELN, Link to Domain based databases & software
Coding Support	Coding, Versioning
Metadata, Semantics & AI	AI Tools/Integration, Metadata, Semantics
Searching	Search By: Domain, Characteristics Search, Keyword/Concept via Content Types, Literature & Notebook, Indexing
Customisation & Extension	Personalisable, Templates
Training & User Support	Training, User Documentation

# SOFTWARE INTEGRATION

- The scientific community are using a wide range of different types of software
- Identify which software is most popular with the scientific community and which meets their needs
- Create middleware/methods of integrating domain-based software with generic notebooking software where ELNs are not suitable or desired
- Build confidence and trust in security of software, and data needs to be stored in an acceptable location

# DATA STANDARDS & MANAGEMENT

- There is both a saturation and lack of data standards
- Identify common data and metadata standards (domain and generic) and investigate which ones are widely adopted / successful and where there are gaps
- Investigate conversion methods for different data formats! (Or stop using questionable proprietary formats...)
- We need better tools to enable FAIR data

# SEMANTIC ENRICHMENT

- Users want semantic enrichment of their documents but this requires extensive further work
- Investigate and evaluate relevant ontologies/schemas for use with notebooking and domain based software and identify gaps
- Quick wins: Are there viable taxonomies that can be converted into ontologies?
- Semantic annotation/markup also needs to be considered

# THE LAB OF THE FUTURE

- The lab of the future doesn't involve a keyboard
- Consider how to incorporate voice into the lab
- Hybrid devices / Smart Notebooks should look to replace regular paper notebooks
- We have smart homes, now is the time for smart labs!

# Our Vision of a PSDI Infrastructure





# CONCLUSIONS – WE'RE ALL IN THIS TOGETHER



## DIGITISE MORE!

We need to solve two problems: Digitising more and managing data better. Both require improved methods and infrastructures.



## WE NEED PEOPLE!

Most “desired features” link to managing data better. The barriers to digitising seem more linked to people, and the barriers to managing data better are linked to people, data, hardware and software.



## ENABLE THE FUTURE!

We need better methods for scientists to record their research in a laboratory environment that doesn't necessitate the use of a keyboard! Hybrid and Voice Technologies are the future!

# RELEVANT PUBLICATIONS

- Kanza, S., Willoughby, C., Gibbins, N., Whitby, R., Frey, J.G., Erjavec, J., Zupančič, K., Hren, M. and Kovač, K., 2017. Electronic lab notebooks: can they replace paper?. *Journal of cheminformatics*, 9(1), p.31. <https://doi.org/10.1186/s13321-017-0221-3>
- Kanza, S., 2018. What influence would a cloud based semantic laboratory notebook have on the digitisation and management of scientific research? (Doctoral dissertation, University of Southampton). <https://eprints.soton.ac.uk/421045/>
- Kanza, S., Gibbins, N. and Frey, J.G., 2019. Too many tags spoil the metadata: investigating the knowledge management of scientific research with semantic web technologies. *Journal of cheminformatics*, 11(1), p.23. <https://doi.org/10.1186/s13321-019-0345-8>
- Knight, N.J., Kanza, S., Cruickshank, D., Brocklesby, W.S. and Frey, J.G., 2020. Talk2Lab: The Smart Lab of the Future. *IEEE Internet of Things Journal*, 7(9), pp.8631-8640. <https://doi.org/10.1109/JIOT.2020.2995323>
- Kanza, S., Willoughby, C., Bird, C.L. and Frey, J.G., 2021. eScience Infrastructures in Physical Chemistry. *Annual review of physical chemistry*, 73. <https://doi.org/10.1146/annurev-physchem-082120-041521>
- Kanza, S., 2021. Guidelines for Chemistry Labs Looking to Go Digital. *Digital Transformation of the Laboratory: A Practical Guide to the Connected Lab*, pp.191-197. <https://doi.org/10.1002/9783527825042.ch13>
- Kanza, S., 2021. Understanding and Defining the Academic Chemical Laboratory's Requirements: Approach and Scope of Digitalization Needed. *Digital Transformation of the Laboratory: A Practical Guide to the Connected Lab*, pp.179-189. <https://doi.org/10.1002/9783527825042.ch12>
- Kanza, S., 2021. Academic's Perspective on the Vision About the Technology Trends in the Next 5–10 Years. *Digital Transformation of the Laboratory: A Practical Guide to the Connected Lab*, pp.297-301. <https://doi.org/10.1002/9783527825042.ch22>
- Kanza, S. and Knight, N.J., 2022. Behind every great research project is great data management. *BMC Research Notes*, 15(1), pp.1-5. <https://doi.org/10.1186/s13104-022-05908-5>

# ACKNOWLEDGEMENTS & QUESTIONS

- **AI4SD Team:** Jeremy Frey & Mahesan Niranjan & Nicola Knight & Victoria Hooper (University of Southampton)
- **PhD Supervisors:** Jeremy Frey & Nicholas Gibbins (University of Southampton)
- **Electronic Lab Notebook Research:** Cerys Willoughby & Nicholas Gibbins & Richard Whitby & Jeremy Frey (University of Southampton), Jana Erjavec, Klemen Zupančič, Matjaz Hren & Kristina Kovač (SciNote)
- **Talk2Lab Team:** Nicola Knight & Don Cruickshank & Bill Brocklesby & Jeremy Frey (University of Southampton)
- **PSDI Team:** Simon Coles, Jeremy Frey, Nicola Knight, Cerys Willoughby & Colin Bird (University of Southampton)





THANK YOU FOR  
YOUR ATTENTION