Mitigating Barriers and Maximising Opportunities in Chemistry Research Data Management

John Jolliffe, Samantha Pearman-Kanza, Nicola Knight

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

Outline

- Who are NFDI4Chem and PSDI?
- Current landscape of Chemistry Research Data Management (RDM)
- Barriers to and Opportunities in Chemistry RDM
- International Chemistry RDM Co-operations

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

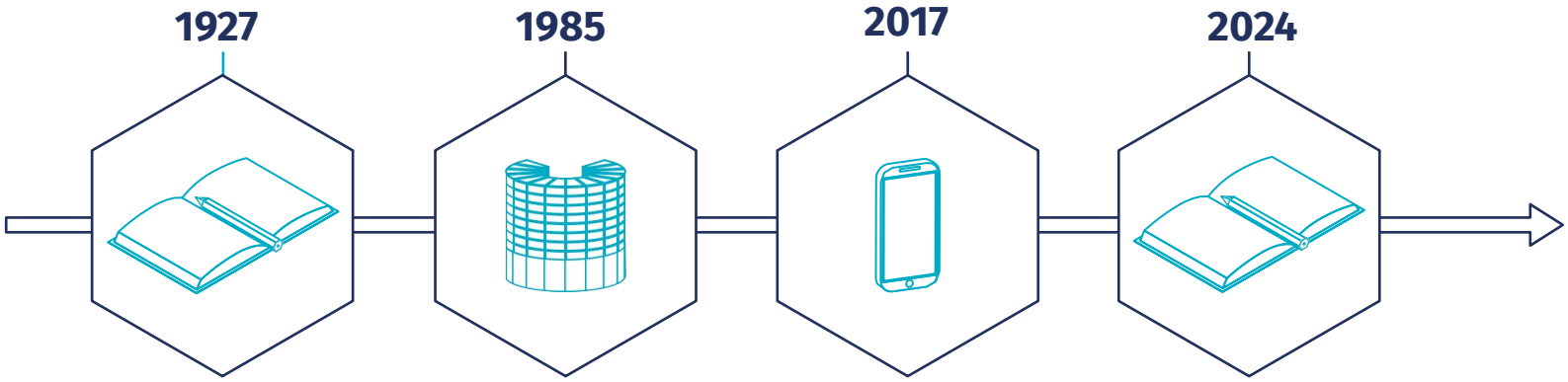
NFDI4Chem
ENHANCE YOUR DATA.

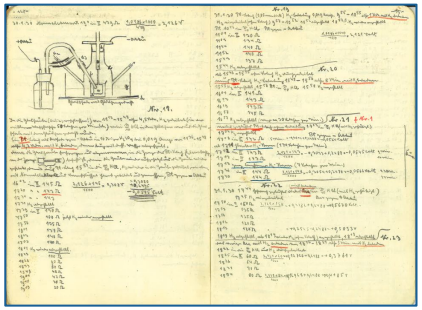
Why are we here?
Who are we?

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE
YOUR
DATA.

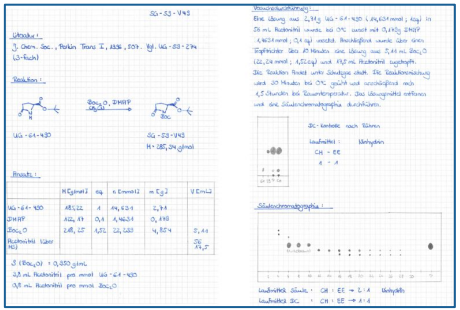
Speed of Evolution





Cray-2, 1985
1.9 GFlops
244 MHz CPU Speed

IPhone 8
325.0 GFlops
2.39 GHz CPU Speed



PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

National Research Data Infrastructure

- German National Research Data Infrastructure
- Founded in 2018 – funded by federal government and states
- **Main goal:** Improvement and creation of Research Data Infrastructure & Management.
- Provision of tools, trainings and initiation of cultural change.
- Part of federal governments digitalisation strategy
- Different consortia representing different scientific disciplines

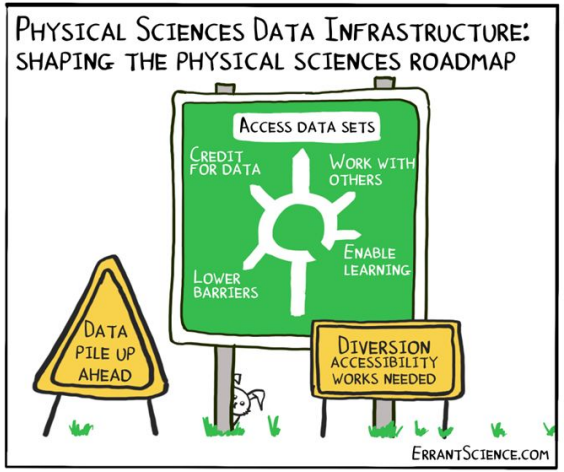
PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

DFG Deutsche Forschungsgemeinschaft

Composition of NFDI4Chem

Physical Sciences Data Infrastructure



CC BY-ND 4.0 Errant Science - https://errantscience.com

- Funded by UKRIs ‘Digital Research Infrastructure’ fund
- Pilot in 2021, development activities started late 2022
- Creating connections between existing experimental and computational facilities
- Not just technology components, but skills, culture and community
- At the core supporting Chemistry, Materials and related disciplines, but interfacing with Life, Medical, Engineering and Environmental Sciences

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

Who is involved in PSDI (so far)

CCDC advancing structural science

Cardiff University / Prifysgol Caerdydd

CatalysisHub

CC BY-ND 4.0 Errant Science - https://errantscience.com

DCC

University of Sheffield | Henry Royce Institute

Loughborough University

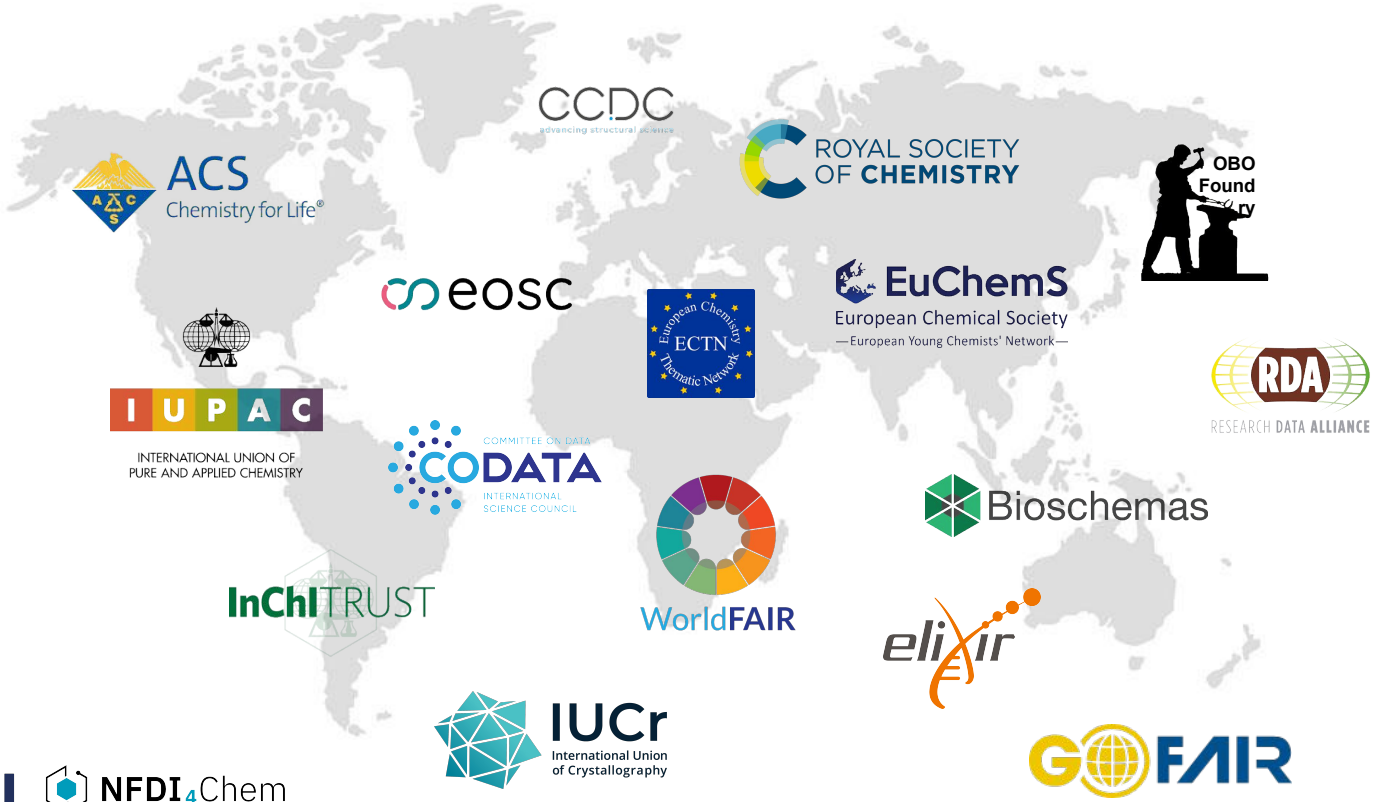
UKRI Science and Technology Facilities Council – Scientific Computing

University of Southampton

PSDI Physical Sciences Data Infrastructure

NFDI4Chem Enhance Your Data.

International Network of PSDI and NFDI4Chem



PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

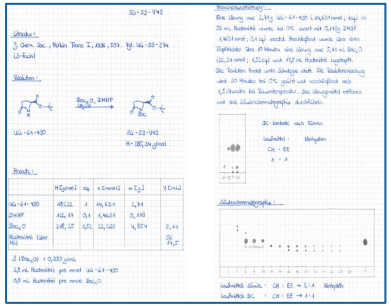
What is the status Quo for Research Data in Chemistry?

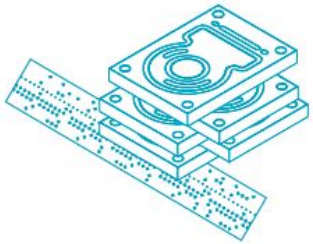
PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

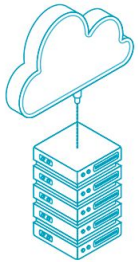
Status Quo

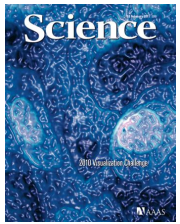
Analogue / Digital

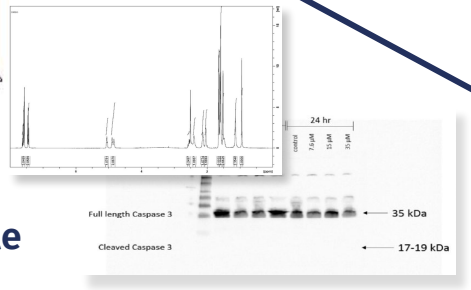




Digital





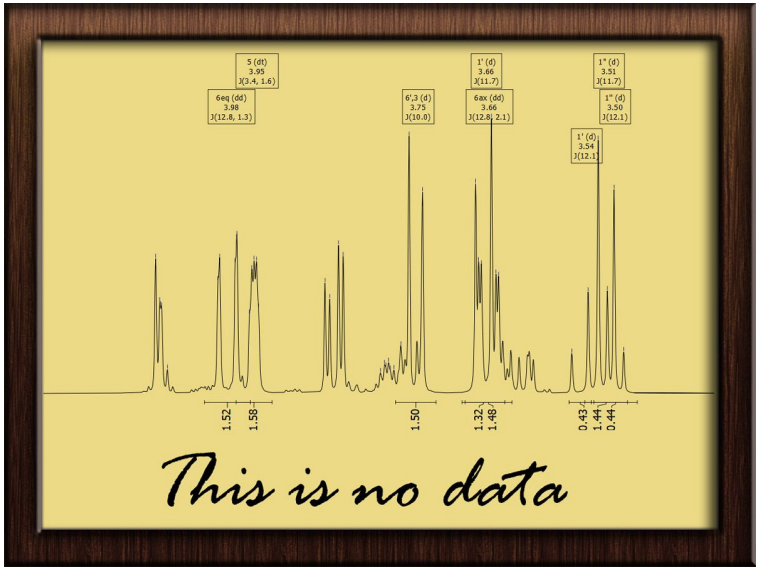


Not machine readable

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

Status Quo



PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

Worst Case

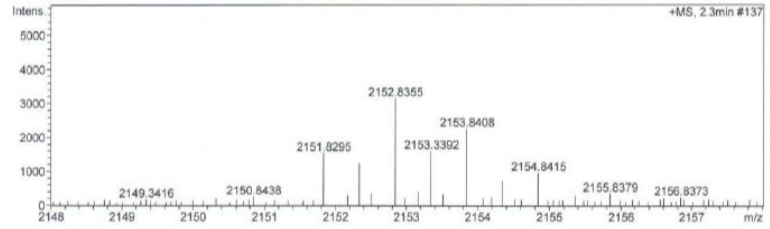
Mass Spectrum SmartFormula Report

Analysis Info

Analysis Name	[redacted]20328.d	Acquisition Date	5/13/2020 10:10:26 PM
Method	20150915.m	Operator	[redacted]
Sample Name	2015110-7-76	Instrument / Ser#	maXis 4G 21240
Comment			

Acquisition Parameter

Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	1.0 Bar
Focus	Active	Set Capillary	4500 V	Set Dry Heater	220 °C
Scan Begin	300 m/z	Set End Plate Offset	-500 V	Set Dry Gas	6.0 l/min
Scan End	2900 m/z	Set Collision Cell RF	2000.0 Vpp	Set Divert Valve	Waste



Meas. m/z # Formula Score m/z err [ppm] Mean err [ppm] mSigma rdb e⁻ Conf N-Rule

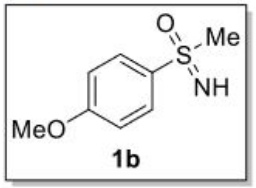
$C_{124}H_{128}O_{32}$ $[M+Na]^+$

2151.828492

2152.831487

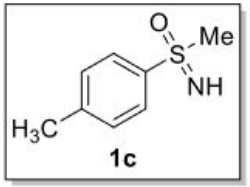
2151.8295

Looks neater but still worst case



1b

1-Methoxy-4-(*S*-methylsulfonimidoyl)benzene (1b). Isolated in 95% yield (907.1 mg, 4.90 mmol) as a faint yellow soild. ^{1}H NMR (600 MHz, DMSO-d_6) δ 7.85 (d, J = 8.8 Hz, 2H), 7.11 (d, J = 8.8 Hz, 2H), 4.05 (s, 1H), 3.84 (s, 3H), 3.02 (s, 3H). ^{13}C NMR (151 MHz, DMSO-d_6) δ 162.3, 135.7, 129.4, 114.1, 55.6, 46.3.



1c

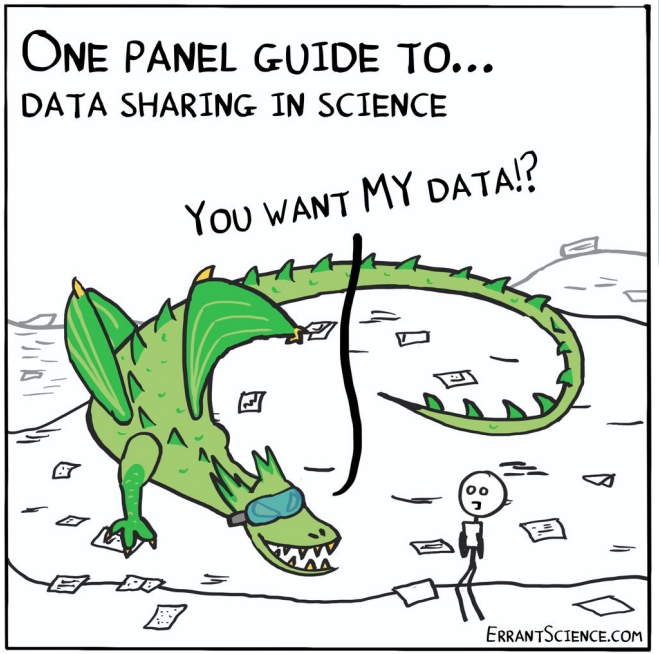
1-Methyl-4-(*S*-methylsulfonimidoyl)benzene (1c). Isolated in 72% yield (608.4 mg, 3.60 mmol) as a faint yellow oil. ^{1}H NMR (600 MHz, DMSO-d_6) δ 7.81 (d, J = 8.0 Hz, 2H), 7.40 (d, J = 7.6 Hz, 2H), 4.12 (s, 1H), 3.03 (s, 3H), 2.39 (s, 3H). ^{13}C NMR (151 MHz, DMSO-d_6) δ 143.2, 141.7, 129.9, 127.8, 46.3, 21.4.

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

More examples of poor data sharing

- Data available on request
- A spreadsheet called results.xlsx with no description
- Method as in Smith et al 2016, where that paper references White et al 2014, which references another paper…
- 1 file with all the information together and no description
- Compound identifiers not included with the data files
- And more…



CC BY-ND 4.0 Errant Science - https://errantscience.com

results.xlsx

Data availability

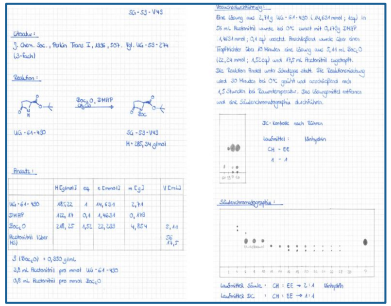
The data that support the findings of this study are available from the corresponding authors upon reasonable request.

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

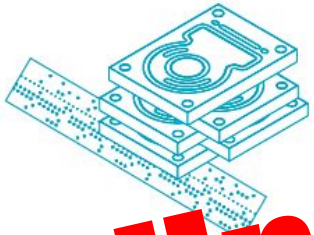
NFDI4Chem ENHANCE YOUR DATA.

Status Quo

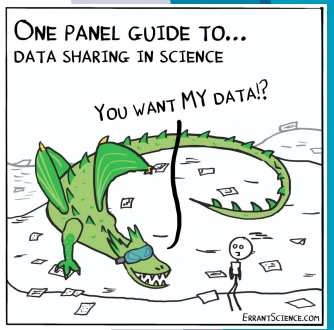
Analogue / Digital



results.xlsx

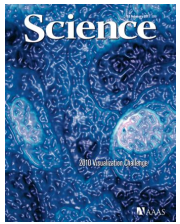


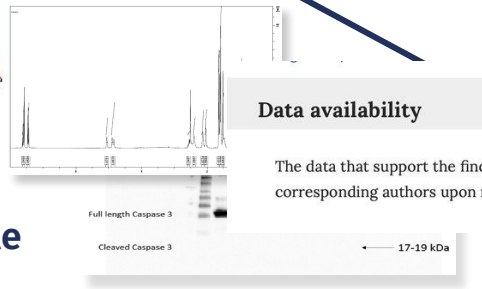
Digital



UnFAIR

Findable
Accessible
Interoperable
Reusable



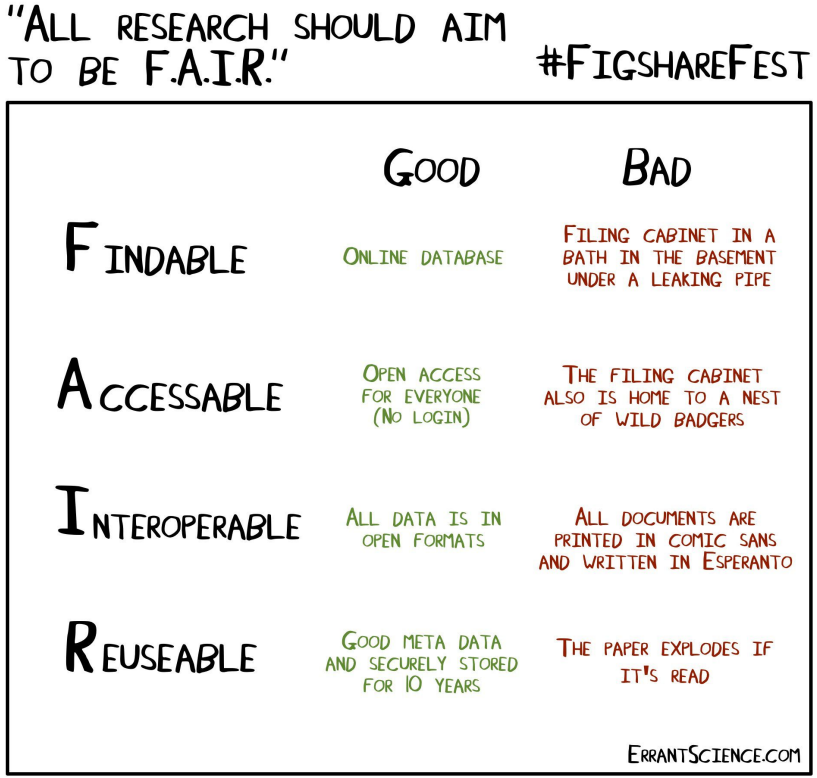


Not machine readable

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

What does FAIR mean?



CC BY-ND 4.0 Errant Science - https://errantscience.com

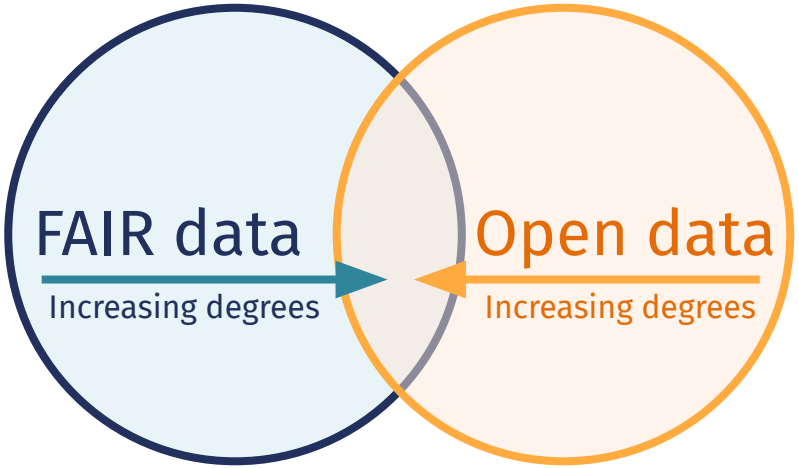
PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

FAIR and Open

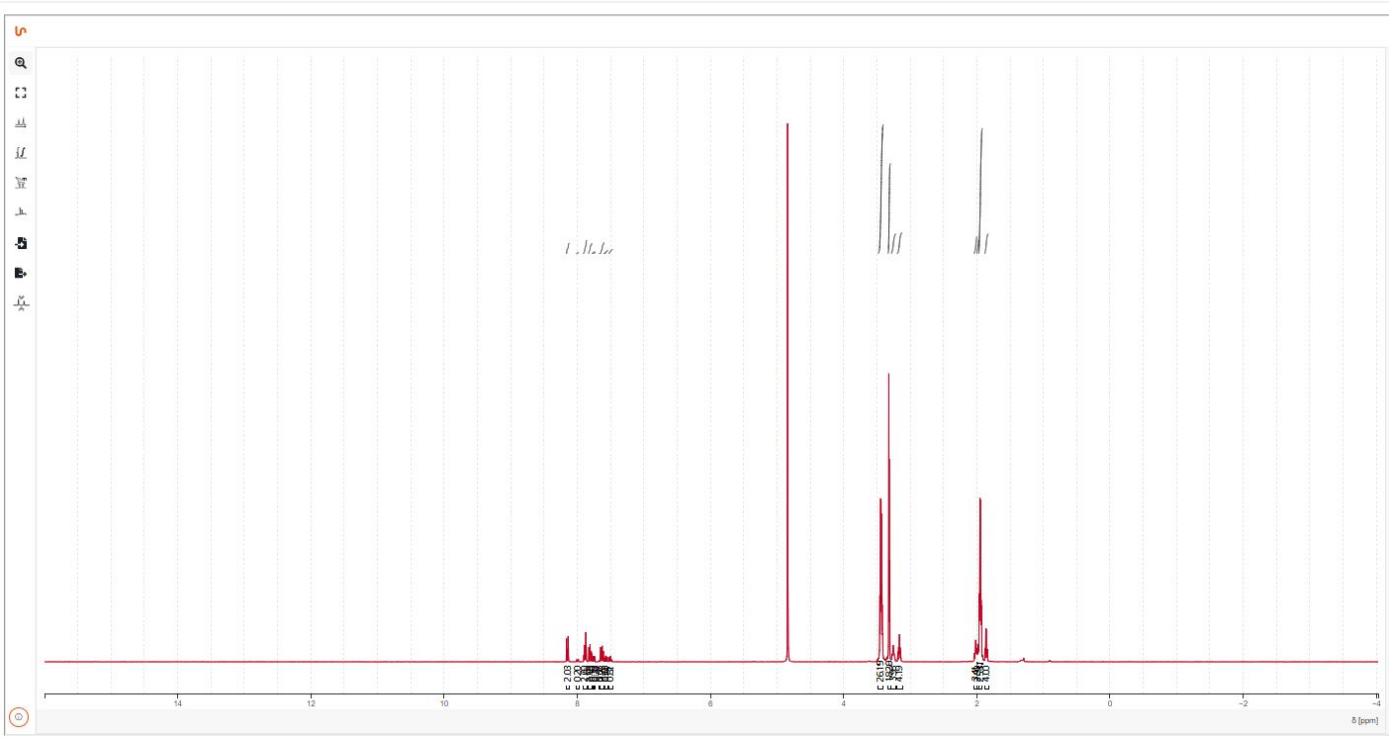
- Not all Open Data are FAIR, not all FAIR data are necessarily open

Findable
Accessible
Interoperable
Reusable



PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE
NFDI4Chem ENHANCE YOUR DATA.

An example of FAIR Data: 1H-NMR Spectrum



PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

F is for Findable

- To be Findable:
  - It needs to exist
  - But existing ≠ findable
  - Provide your users with pointers!
  - Structured metadata according to standardised metadata schemata to make it findable:

    → Metadata answers the “w” questions: who did what, how, when and where?

- Example for NMR: Dataset can be found via persistent identifier such as a doi (digital object identifier)

  https://dx.doi.org/10.14272/KDWZDHRURBJEIN-UHFFFAOYSA-N/CHMO0000593

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

A is for Accessible

- To be accessible:
  - Human and machine accessible
  - What should and shouldn't be accessible?
  - If access is restricted or complex, have you provided relevant information?
  - Technically accessible ≠ Easily accessible
- Example for NMR:
  - Dataset is published online in data repositories
  - Ideally, in data repositories where all NMR data are published in a standardised way

Chemotion Repository

nmrXiv

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

I is for Interoperable

- To be interoperable
  - Consider your data standards
  - Use Common and Shared Vocabularies
    - For Data and Metadata
  - Use Ontologies/Knowledge Graphs to the best of their potential

- Example for NMR:
  - Data is published in non-proprietary, open data format such as jcamp.dx

~~*.mnova~~ *.jdx

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

What is JCAMP-DX

- IUPAC standard format family for spectral data exchange
- Non-proprietary, open data format
- Text-based format for spectroscopic data
- Started for Infrared spectroscopy
- Expanded to include **NMR** and other methods
- In principle, any two-dimensional data can be depicted with JCAMP-DX

```
##TITLE=HMS-CCP3737_10
##JCAMP-DX=5.0
##DATA TYPE=LINK
##BLOCKS=1


$$ === CHEMSPECTRA SPECTRUM ORIG ===
##TITLE=HMS-CCP3737_10
##JCAMP-DX=5.00
##DATA TYPE=NMR SPECTRUM
##DATA CLASS=XYDATA
##$CSCATEGORY=SPECTRUM
##ORIGIN=HMS
##OWNER=HMS
##.OBSERVE FREQUENCY=400.13240078
##.OBSERVE NUCLEUS=^1H
##SPECTROMETER/DATA SYSTEM=
##.SHIFT REFERENCE=INTERNAL, CDC13, 1, 15.92546
##.SOLVENT NAME=CDC13
##.PULSE SEQUENCE=zg30
##XUNITS=PPM
##YUNITS=ARBITRARY
##XFACTOR=0.0001248
##YFACTOR=361.654927
##FIRSTX=15.987510479959688
##LASTX=-4.037606784383083
##MAXX=15.987510479959688
##MAXY=361654927.0
##MINX=-4.037606784383083
##MINY=-47376.795437
##NPOINTS=65536
##XYDATA= (X++(Y..Y))
65536dB0B4bb2a5aa6a6a4b0b7b9b2a4aB4B9A9A3A7A0@ga1gaa2b2b1b8b6b7b3hA3B0A2
65498@a3a3a5b4b4a7a3bA8A8B2A7Aia7a3Accca3a3fa9b1b0cda1ieha9b0dA3A5@a2a8
65457@B0A7IHA7IA6B8B7B1Dihbha2EA2ga7a0a6a0A1B4A5FCba7c6d1b2a2b6c7c9d0d2
65417c2a6giA3B7B6B4Ea2a6a2CE@BD@a7d6c5b0b4a9a0A3Eia2DA4Fca4a9a0a2a5b8
65378d0b3@E@a0dHA4Gfa6a3GB2D1D0B0A4B3C8A4a7b4bA0IFBGCBbea1b4d4e7d8b8b0
65337dCeiaBa7c5d0d0c9e1g1f7d5a5ca7c2b4b7b2a6ia2c1d0b4@A3B6A9A2A7Ibhfa8
65298a7BeeFH@A@a1b8c3a5EB4C4A1FB4A0a2a4a0A6C8B2aa5a4iBD@a1gA7D1D6B4A6
65258A2HA2B4B7A0b0c5b7cca8GB7B4B9B0A8B4A3@A1B6D2B3a1b2b1c3a5B4C9C1B4B2
65223Ab2b6b0b0c9c3d3e8e7e4d5a6cdBA9A7ICA4A8fc1c3b4c1b6a2a6@AGA6@dCDea7
65183b9c5c5b7a7a5a2b1b6b3a4BDdb3c5c3b6h@HCa0a5bA6C3C0hc0b4iBA7C0B6A7A7
65145A7Dfegb1b8fea4b2c5d6e8f5e8c6bI@AAA3Fa3b3c6c6c9c9b4b1b4b2a6DC0C6C1
65106C2D1C4A4Ca1b4c8c4b9c2eGfIB2A9Ibfa7c6d8c5b1gbieIA9B6C4D2D8F6H7G2E4
65067D7C4B6B7A4bBA6Bb4c8d3d1c0c0a9bA6B4Gb1b8dB9C0e@B4B2H@a3a2HC7E2D7C5
65029B6B7HA4C9C8B4aeba1a0c3d8f8i4g6c5e7h7h4h9h4h6f5b4a0cA0A7A0a8c7a5H
64994@b1b1A1B9C9B7B0A6A3A6A9Ea5FC0C6E2F5D8B0ac0d8f0e5d8c5egbaa5iHB4E2
64957F4C9A6A3BFB2B4D1E5C6AfEB1C1H@A4Ica8d6d5b4dBHc@a@GA7C6C1B3Faa6bB3
64915B4ECDia6a5a2i@GFA2A7F@cb8c9iA1aFA1@b0b7ia5b7a4c5h5g7e1d8d6c8d2d3
```

R is for Re-useable

- This isn't JUST about the data! You need to consider:
  - Metadata, Data, Tools, Code, Methods, Context
  - How could/would your work be re-used, replicated, reproduced or repurposed
    - Re-use – re-use the data in the same manner
    - Replicate – repeat entire research from scratch including data collection and analysis
    - Reproduce – reanalyse the existing data in the same manner
    - Repurpose – use existing data for a new purpose
- Example for NMR:
  - Re-use of data is clearly defined via an appropriate CC-license such as CC-BY-4.0

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

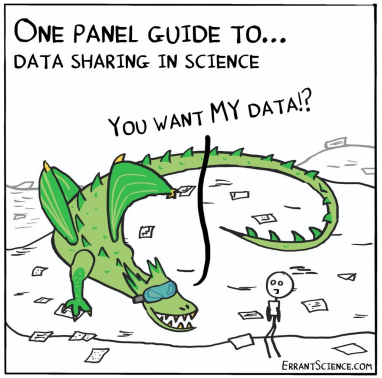
This is only the tip of the "R" Iceberg

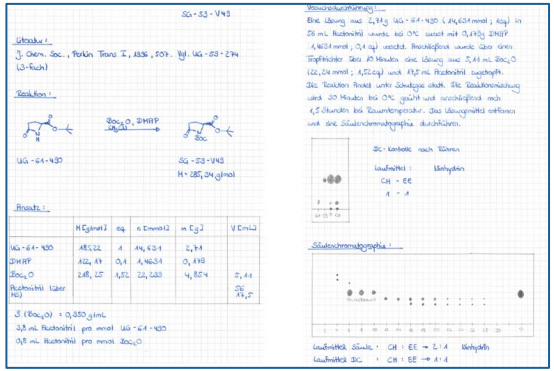
Why FAIR Data and how do we get there?

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

Our Vision

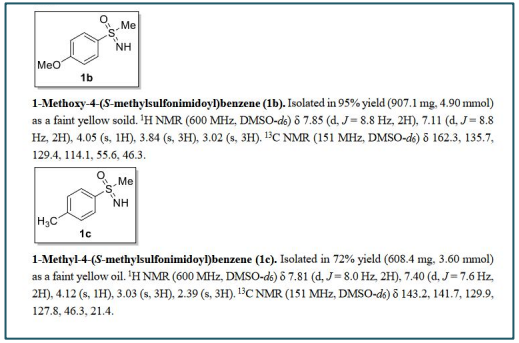




Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

results.xlsx



1-Methoxy-4-(*S*-methylsulfonimidoyl)benzene (1b). Isolated in 95% yield (907.1 mg, 4.90 mmol) as a faint yellow soild. ^{1}H NMR (600 MHz, DMSO-*d*$_6$) δ 7.85 (d, *J* = 8.8 Hz, 2H), 7.11 (d, *J* = 8.8 Hz, 2H), 4.05 (s, 1H), 3.84 (s, 3H), 3.02 (s, 3H). ^{13}C NMR (151 MHz, DMSO-*d*$_6$) δ 162.3, 135.7, 129.4, 114.1, 55.6, 46.3.

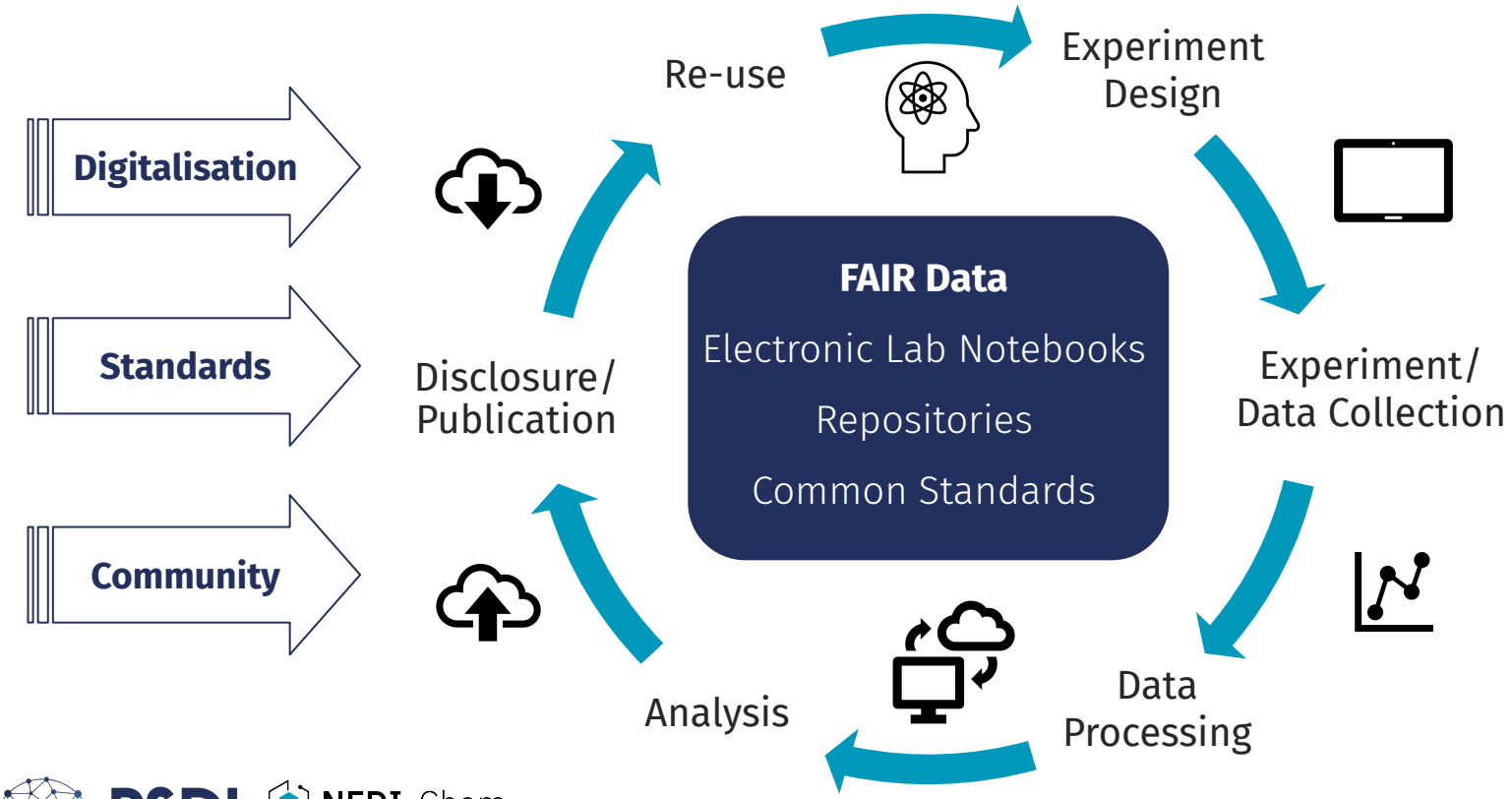
1-Methyl-4-(*S*-methylsulfonimidoyl)benzene (1c). Isolated in 72% yield (608.4 mg, 3.60 mmol) as a faint yellow oil. ^{1}H NMR (600 MHz, DMSO-*d*$_6$) δ 7.81 (d, *J* = 8.0 Hz, 2H), 7.40 (d, *J* = 7.6 Hz, 2H), 4.12 (s, 1H), 3.03 (s, 3H), 2.39 (s, 3H). ^{13}C NMR (151 MHz, DMSO-*d*$_6$) δ 143.2, 141.7, 129.9, 127.8, 46.3, 21.4.

FAIR?

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

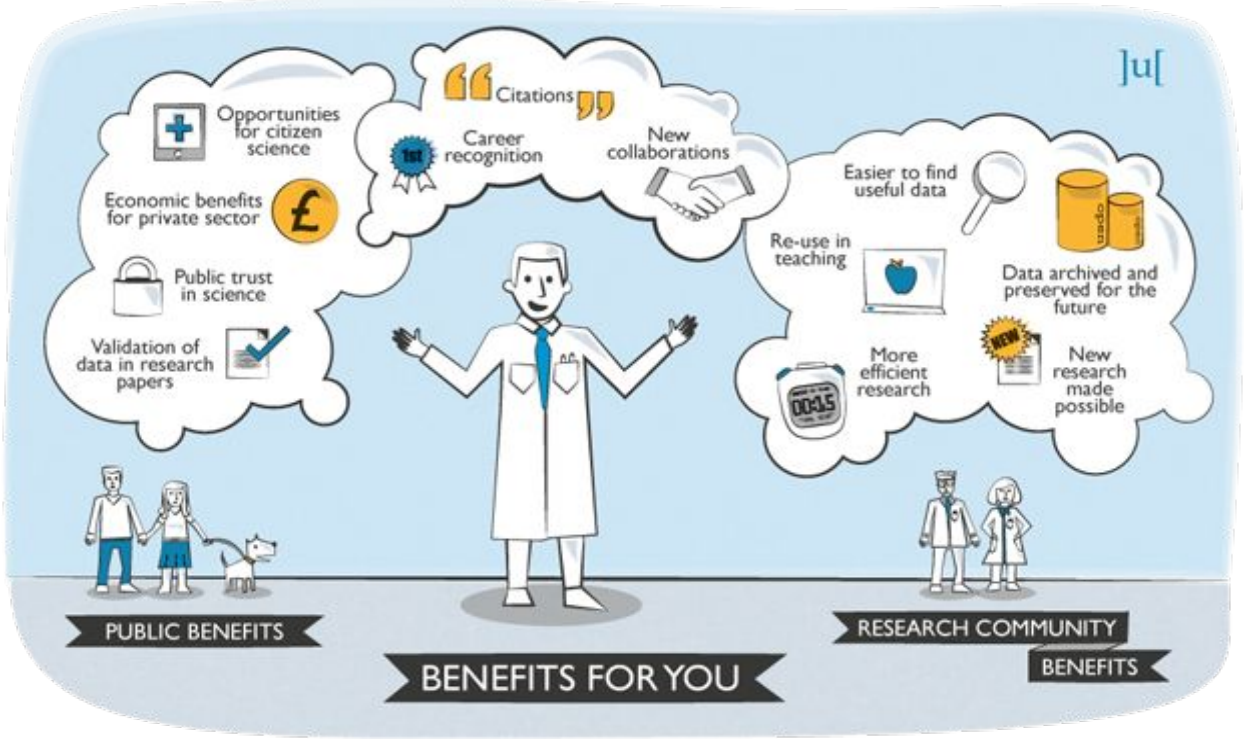
Our Vision: All chemists publish FAIR data



PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

Why FAIR Data?



https://www.slideshare.net/brianhole/the-journal-of-open-archaeology-data-and-prime-incentivising-open-data-archiving

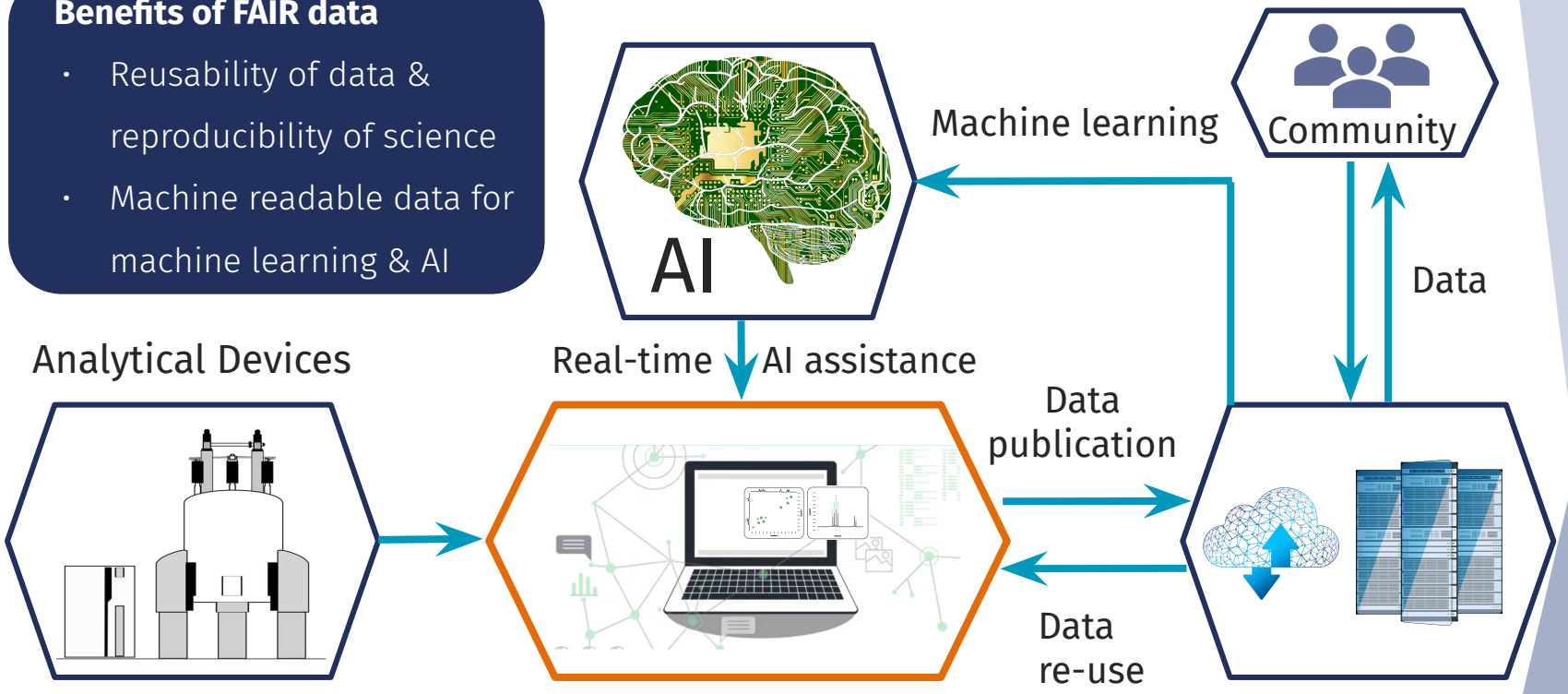
PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

Why FAIR Data?

Benefits of FAIR data

- Reusability of data & reproducibility of science
- Machine readable data for machine learning & AI

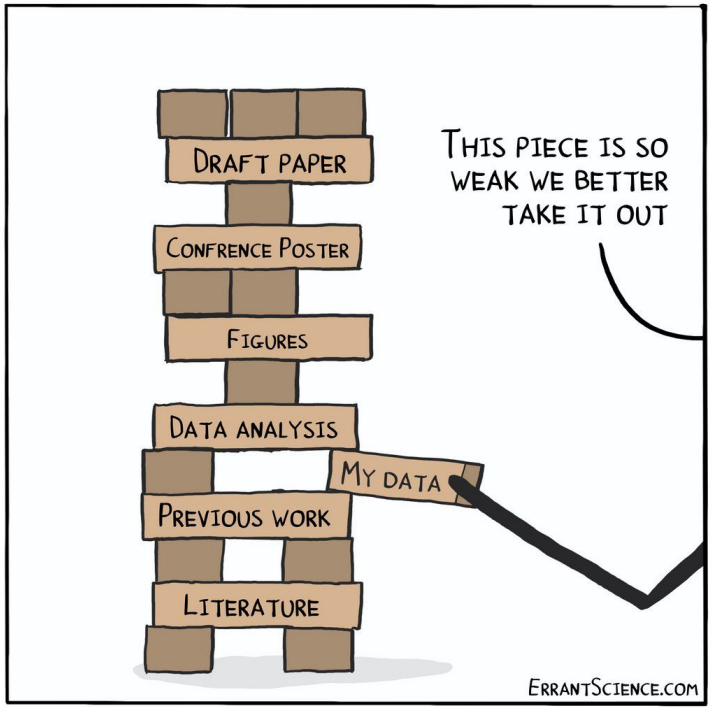


PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

What are the barriers?
How can these be addressed?

Current Barriers & Challenges to Open and Transparent Digital Research



CC BY-ND 4.0 Errant Science - https://errantscience.com

- Data
- Standards
- Software
- Hardware
- Cost
- Time
- Trust
- People / Adoption

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

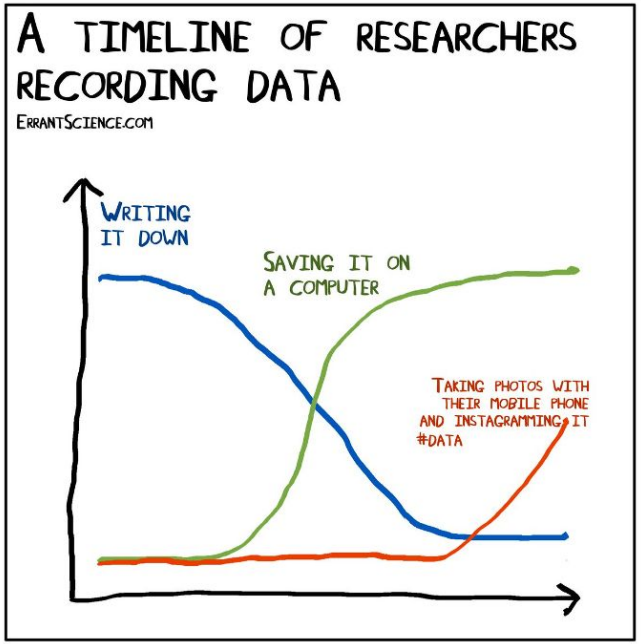
People/Adoption Barriers & Mitigations

Barriers

- **Resistance to change:** Hard to persuade people to change their processes
- **Group mentality:** Generally need to get entire group on board
- **Duplication of work:** Concerns about more work without seeing an obvious reward

Mitigations

- **Management:** A top down approach can make or break this
- **Demonstration:** Need to demonstrate how this will be of benefit
- **Start Afresh:** Implement this with a set of keen early adopters and or new starters for cooperation and to demonstrate success



CC BY-ND 4.0 Errant Science - https://errantscience.com

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

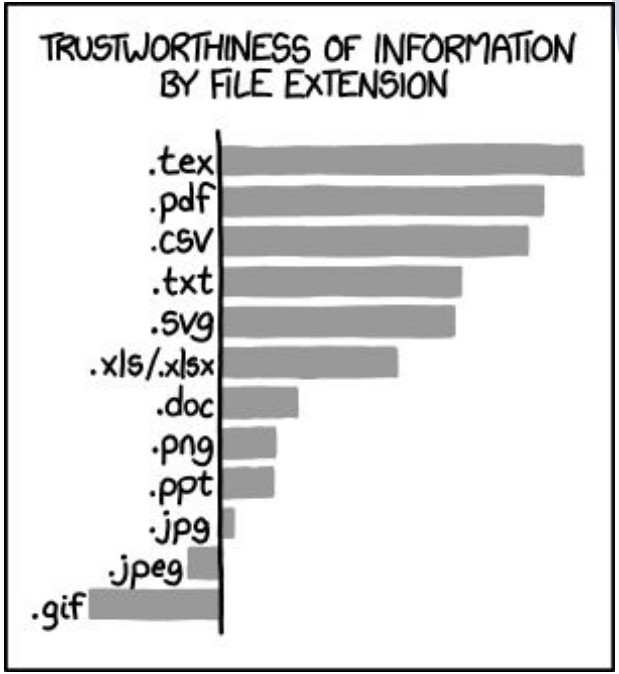
Data Barriers & Mitigations

Barriers

- **Un-FAIR Data:** Data frequently isn't findable, accessible, interoperable or re-useable
- **Metadata/Provenance:** Data often lacks context, its time consuming to capture metadata which leads to lack of provenance
- **Data Size:** Datasets are frequently too large to easily store/share

Mitigations

- **Education:** We need more training on how to how to make our data properly FAIR
- **Tooling & Guidelines:** We need better tools and guidelines to aid with capturing FAIR data and Metadata e.g. ELNs
- **Computing Resources:** We need access to centralised computing resources for storing/sharing data.

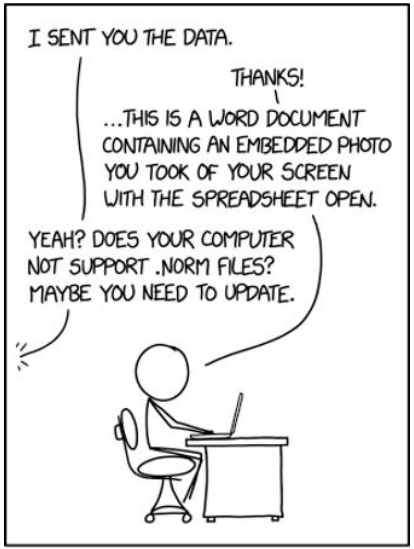


"File Extensions" by XKCD is licensed under CC BY-NC 2.5

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

Standards Barriers & Mitigations



"NORM Normal File Format" by XKCD is licensed under CC BY-NC 2.5

Barriers

- **Too many standards:** There are both too many, and not enough standards
- **Proprietary Formats:** Industry software uses proprietary formats, causing vendor lock in and lack of compatibility
- **Lack of Interoperability:** There are many formats that don't work across multiple pieces of software

Mitigations

- **Open Source Formats:** Software should use open source formats that work across different tools and can be easily converted to/from
- **Format Conversion:** We need tools to convert between different data formats

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

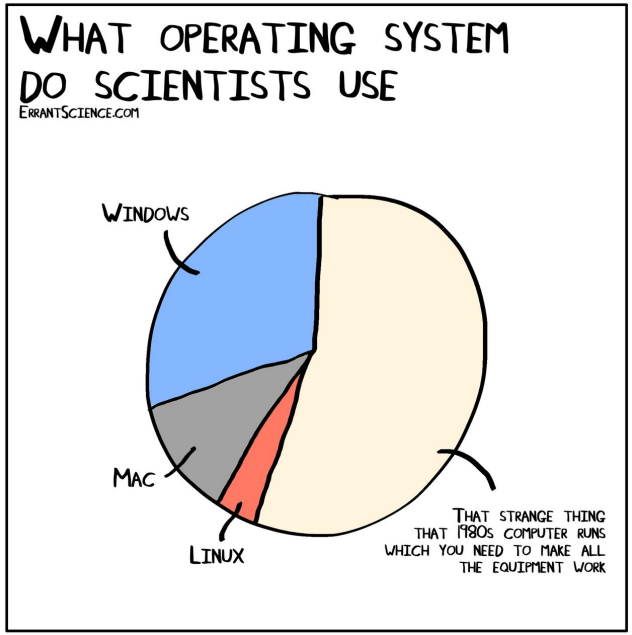
Software Barriers & Mitigations

Barriers

- **Software Choice Overload:** There are too many different pieces of software for digital research and data capture
- **Software Compatibility:** Software is frequently incompatible
- **Software Quality:** Users typically won't engage with software of poor quality

Mitigations

- **Resource Catalogues:** We need comprehensive information on the range of software available to make a suitable choice
- **Format Conversion:** We need tools to convert between different data formats so data can be used with different software
- **Software Design:** Involve actual software engineers, and train the scientists who make software in better practices

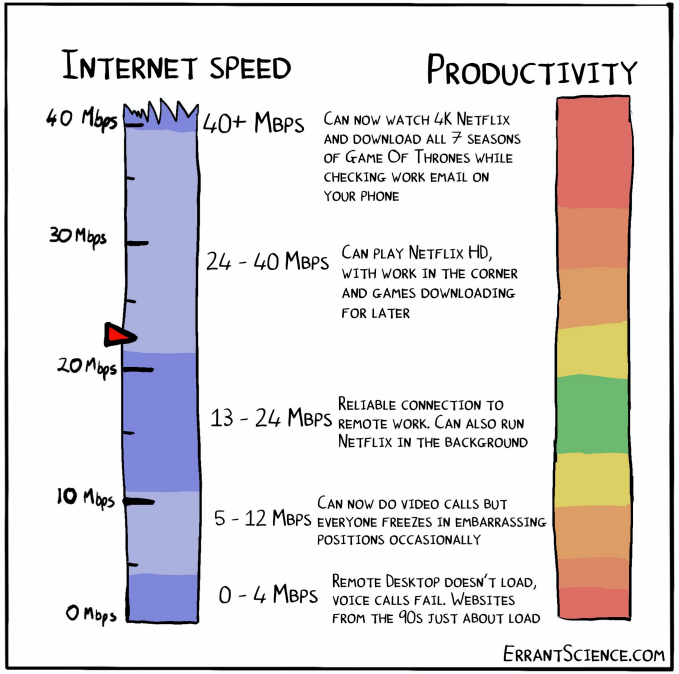


CC BY-ND 4.0 Errant Science - https://errantscience.com

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

Hardware Barriers & Mitigations



CC BY-ND 4.0 Errant Science - https://errantscience.com

Barriers

- **Clunky Hardware:** Frequently available hardware won't run digital tools adequately, disengaging users completely
- **Legacy Equipment:** Too many laboratories use legacy equipment which requires legacy software and leads to producing outdated data formats
- **Hardware Cross Contamination:** Moving hardware in and out of the lab is dangerous and leads to contamination
- **Poor Internet Connection:** If there is lack of connectivity anything cloud/web based won't work

Mitigations

- **Invest in new Equipment:** We need to invest in new equipment to ensure smoother running of software and to reduce the requirement for legacy software/data
- **Dedicated Lab Hardware:** We need specific hardware for inside the lab, and separate hardware for outside the lab to reduce contamination
- **Internet Provisions:** Either have hardwired connections or boost your WIFI

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

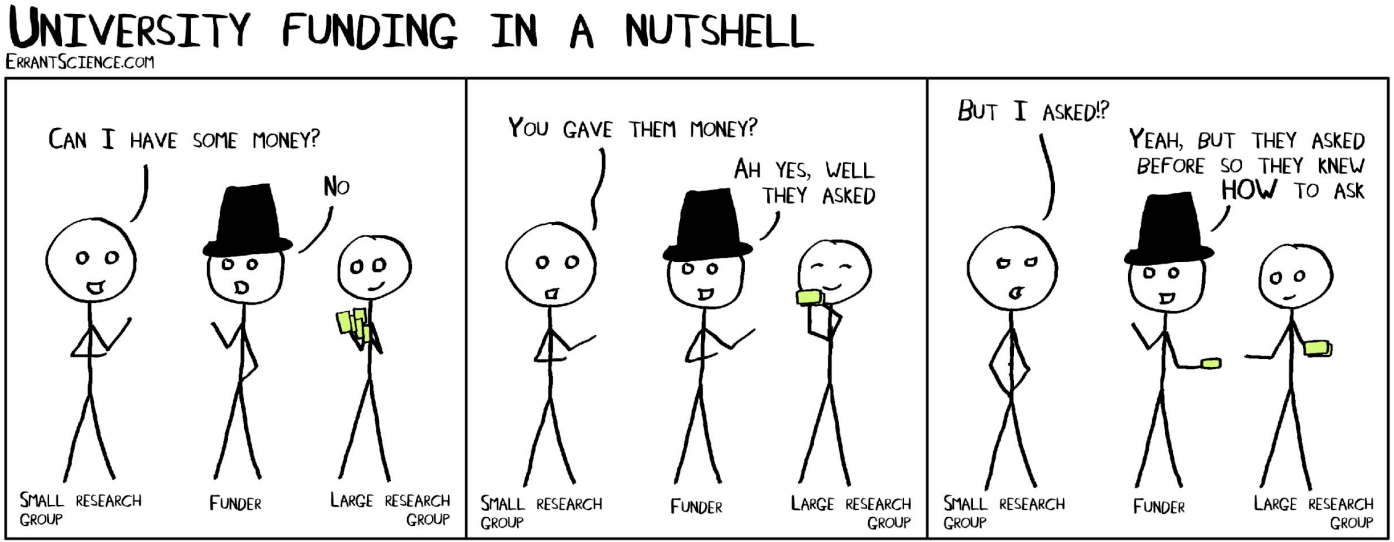
Cost Barriers

Barriers

- **Costs:** Funding, Research, Software, Hardware, Publishing

Mitigations

- **Embed in Funding Applications:** Extra costs to perform research data management properly with the appropriate expertise need to be embedded in applications
- **Investment:** Institutions need to invest in the appropriate hardware, software, education and addition staff e.g. Data Stewards to do this properly



CC BY-ND 4.0 Errant Science - https://errantscience.com

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

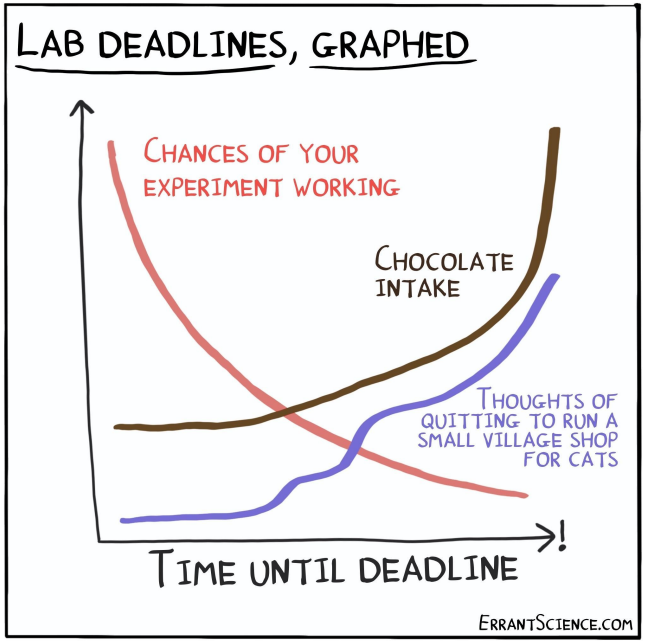
Time Barriers & Mitigations

Barriers

- **Learning:** Time taken to learn and use new systems
- **Do it Properly:** Doing research data management (RDM) properly is very time consuming
- **Current Systems:** without effective tools digitising data and capturing research in a transparent way is very time consuming

Mitigations

- **Start Early:** Make Data Management Plans and embed RDM into your entire project
- **Build in Time:** Projects need to formally build in time for research data management rather than seeing it as a “nice to have” or something that can be tacked on quickly at the end
- **Better Systems:** We need to develop better tools for capturing research and data

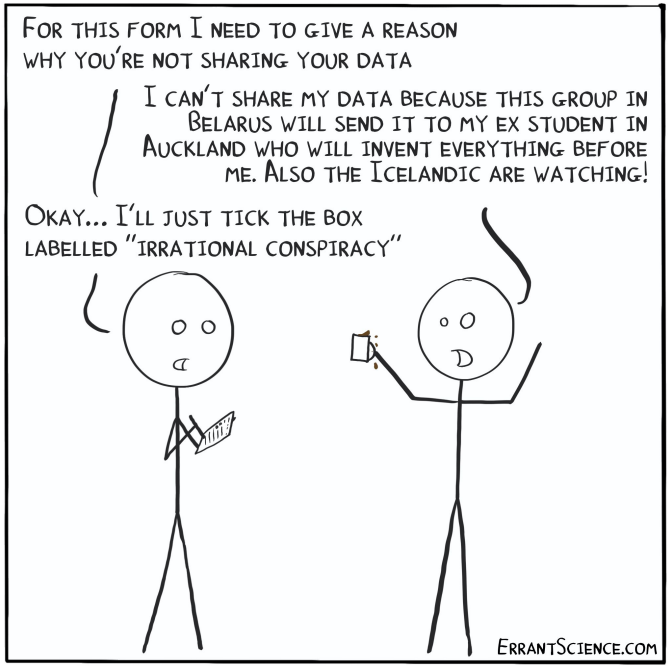


CC BY-ND 4.0 Errant Science - https://errantscience.com

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

Trust Barriers & Mitigations



CC BY-ND 4.0 Errant Science - https://errantscience.com

Barriers

- **Cloud Storage:** Many researchers don't trust storing their data/research online due to concerns about:
  - **Data Privacy:** Sharing/Hacking
  - **Vendor Lockin:** Software using proprietary formats and lack of data exit strategy

Mitigations

- **Education:** We need more education about the genuine dangers of cloud based storage weighed up against paper / local storage
- **Storage Location:** There are options to store data "in the cloud" but with local servers e.g. MS Sharepoint and to have clever backup strategies to subvert hacking vulnerabilities

Planning ahead - key message

- There are still many barriers to overcome but the community is working towards solutions
- We need to remember to:
  - Start early and plan for your data, don't leave data management to the last minute
  - FAIR is a FOUR letter word, but has many nuances
  - Collaboration is key - this is a human AND technical endeavour
    We need to invest more time, money and resource into being FAIR

CC BY-ND 4.0 Errant Science - https://errantscience.com

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

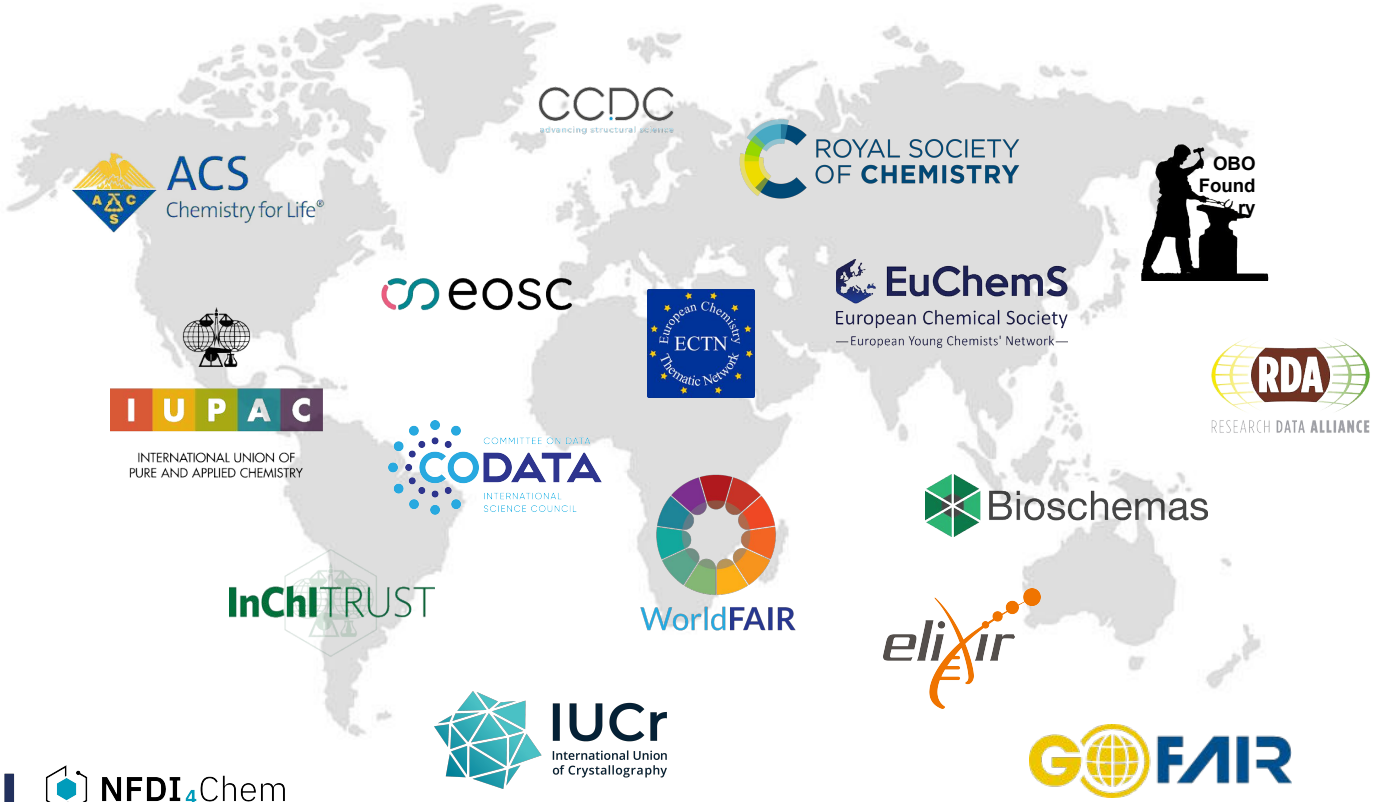
These challenges are international

What international efforts are there?

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

International Network of PSDI and NFDI4Chem



PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.

IUPAC and Standards

- The International Union of Pure and Applied Chemistry (IUPAC) is the world authority on digital standards in chemistry; chemical nomenclature and terminology standardized methods for measurement; and on atomic weights
- IUPAC has been creating the common language of chemistry for more than a century.

IUPAC

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

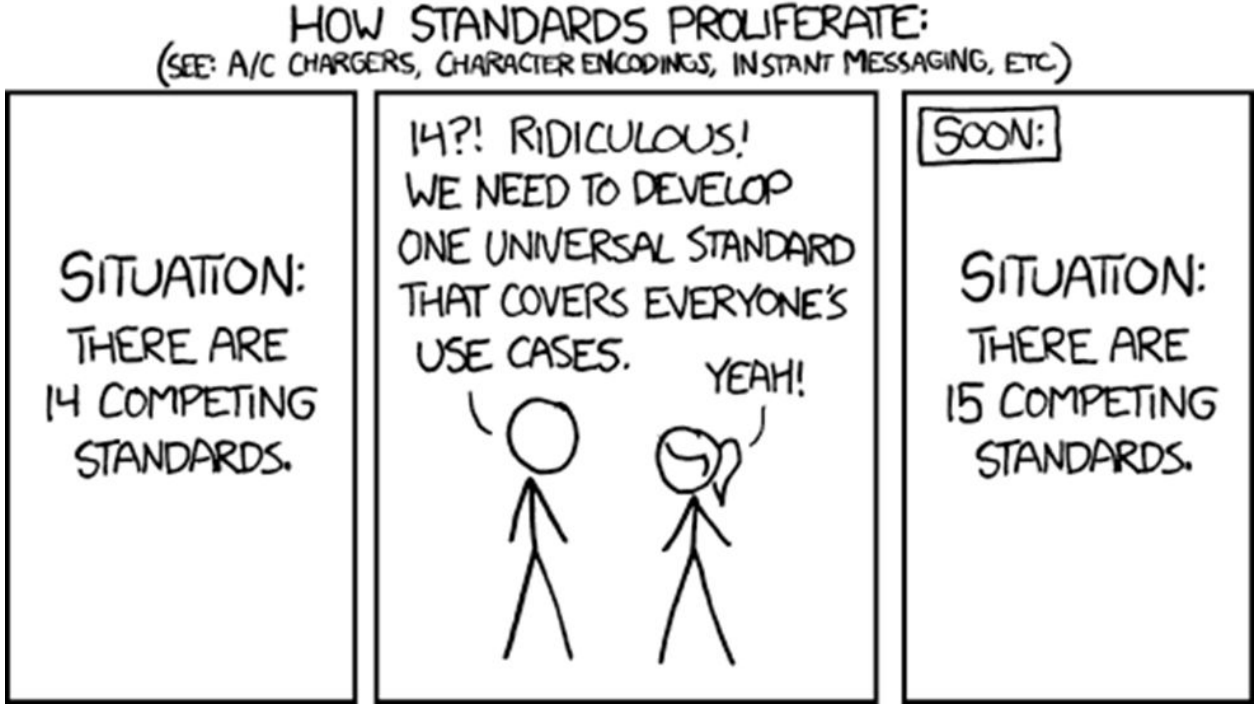
PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

IUPAC and Standards

- Standards are important to advance the goals of FAIR data because:
  - **Uniformity and Clarity:** Ensure consistent chemical nomenclature and terminology.
  - **Seamless Data Exchange:** Facilitate smooth data sharing and collaboration.
  - **Reproducibility:** Enable accurate reproduction of experiments and results.
  - **Interoperability:** Promote data and tool interoperability.
  - **Global Collaboration:** Support effective collaboration worldwide.
  - **Enhanced Data Reusability:** Simplify data reuse, saving time and resources

Standards - the Challenge



"Standards" by XKCD is licensed under CC BY-NC 2.5

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

NFDI4Chem & PSDI Involvement in IUPAC

- Participation in IUPAC's involvement in World FAIR
  - FAIR Chemistry Cookbook
- InChI Trust
  - Development and maintenance of the InChI Standard
  - Promotion of InChI adoption in scientific communities
  - Collaboration with software developers for integration
- Committee on Publications and Cheminformatics Data Standards
  - Standard for FAIR Data Management of Spectroscopic Data (FAIRSpec)
  - Machine-Accessible Periodic Table
  - Development of guidelines for publication standards
- IUPAC Gold Book
  - Incorporation of new scientific terms and concepts, with regular updates

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

IUPAC INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

NFDI4Chem's involvement in the InChI Trust

- Since 2020 involvement in the organometallics subgroup (until now, bonds to metals are disconnected)
- Since 2022: two InChI developers working at RWTH Aachen University on the next version of the InChI
- Since 2023: Sonja Herres-Pawlis was elected into the InChI Trust
- 2024: new InChI version v1.07 out as test version

  → the InChI has moved to GitHub, is now FAIR

  → free webdemo: https://iupac-inchi.github.io/InChI-Web-Demo/

- Organometallics/inorganics are programmed **now** into the InChI
- InChI in Action: technical workshop on 24th and 25th july in Aachen (contact Sonja Herres-Pawlis)

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE
NFDI4Chem ENHANCE YOUR DATA.
InChITRUST
VolkswagenStiftung

IUPAC FAIR Cookbook (PSDI)

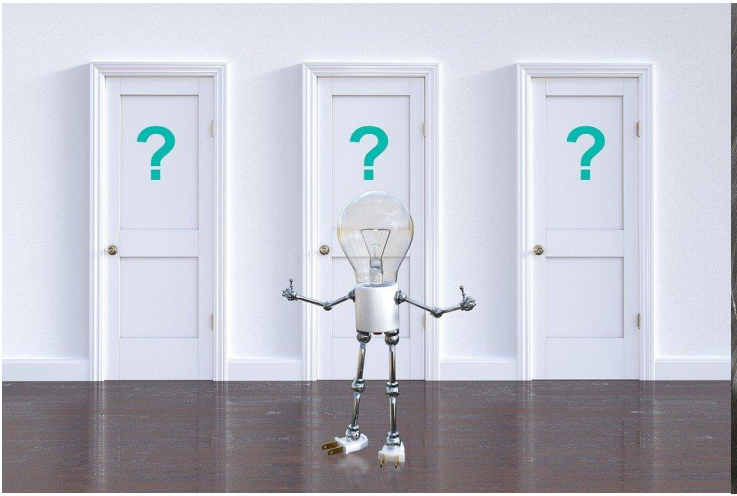
- Living community resource for managing digital chemical data effectively
- Enabling the move toward sharing and reusing data, code, metadata, etc.
- Step by step recipes for working with chemical data such as: using APIs to access chemical data or using chemical libraries in Python

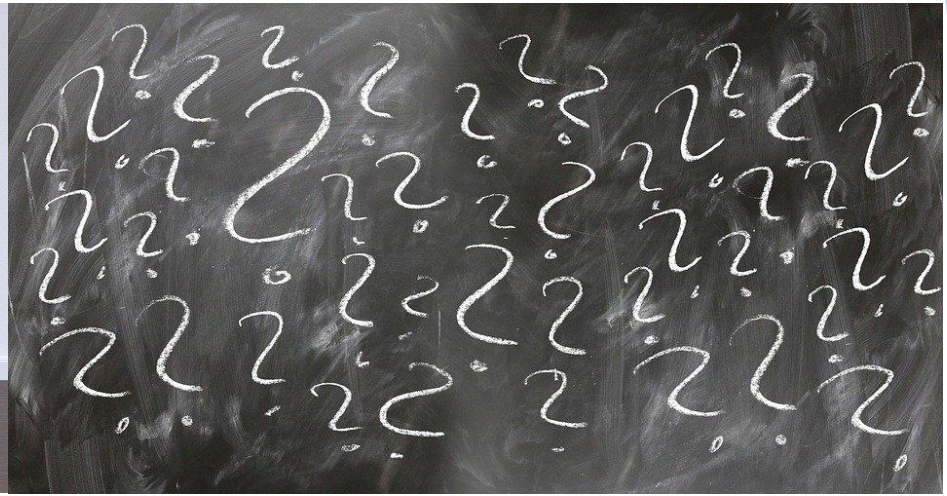
PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem ENHANCE YOUR DATA.

http://bit.ly/CookFAIR or
https://iupac.github.io/WFChemCookbook

Questions?





PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

NFDI4Chem
ENHANCE YOUR DATA.