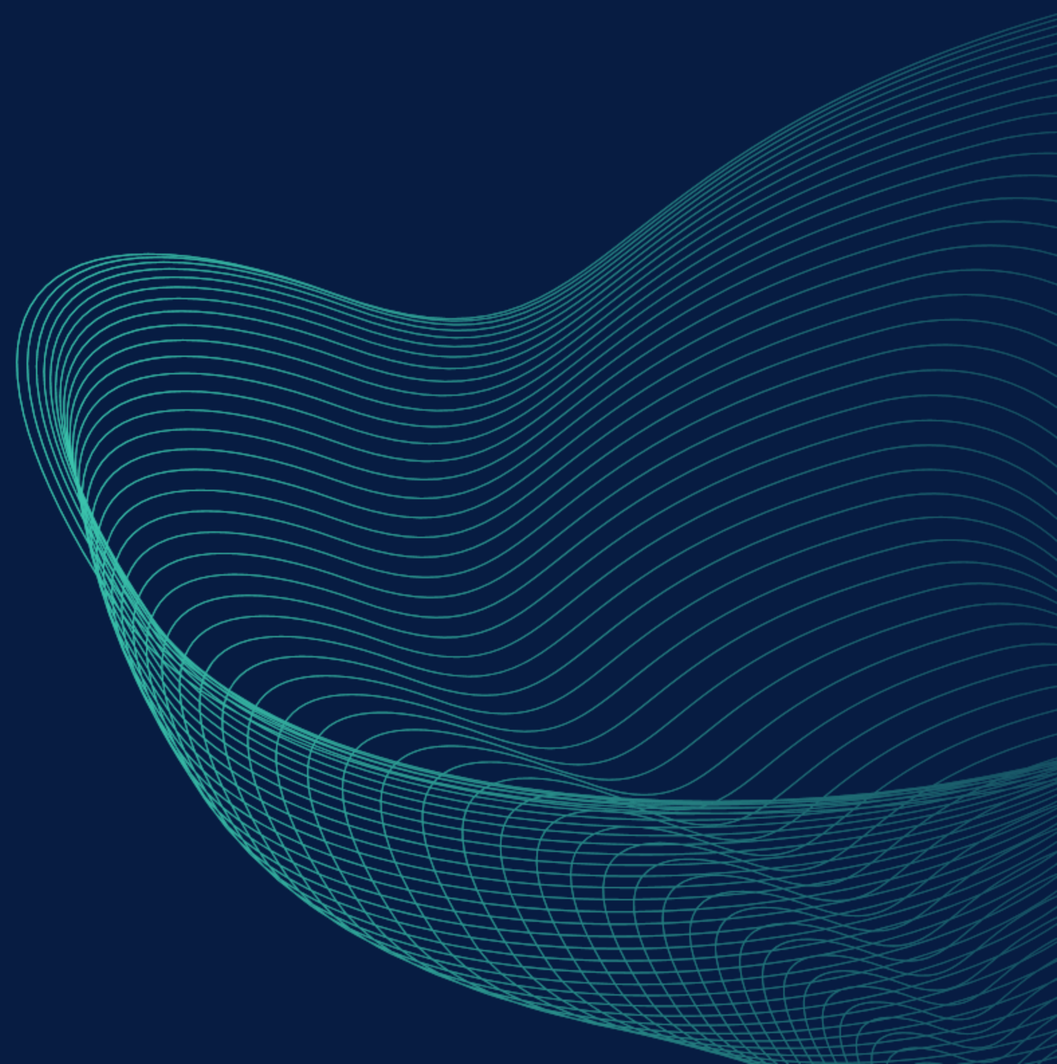
Fixing The Past To Fix The Future

Samuel Munday
sam@data-revival.com
https://linktr.ee/data_revival



Digitally Transforming

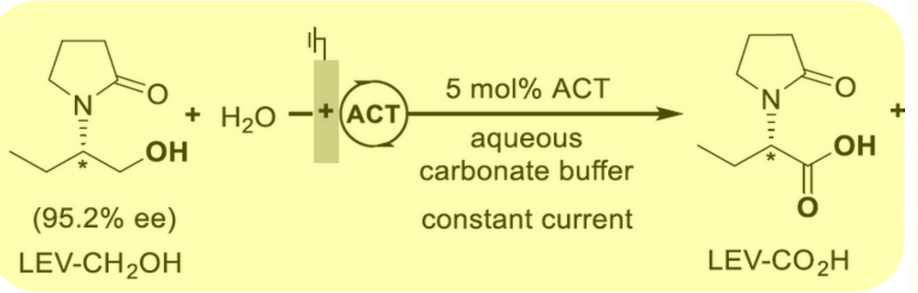
The Past

DATA REVIVAL

How much of our knowledge is unusable with modern digital tools?

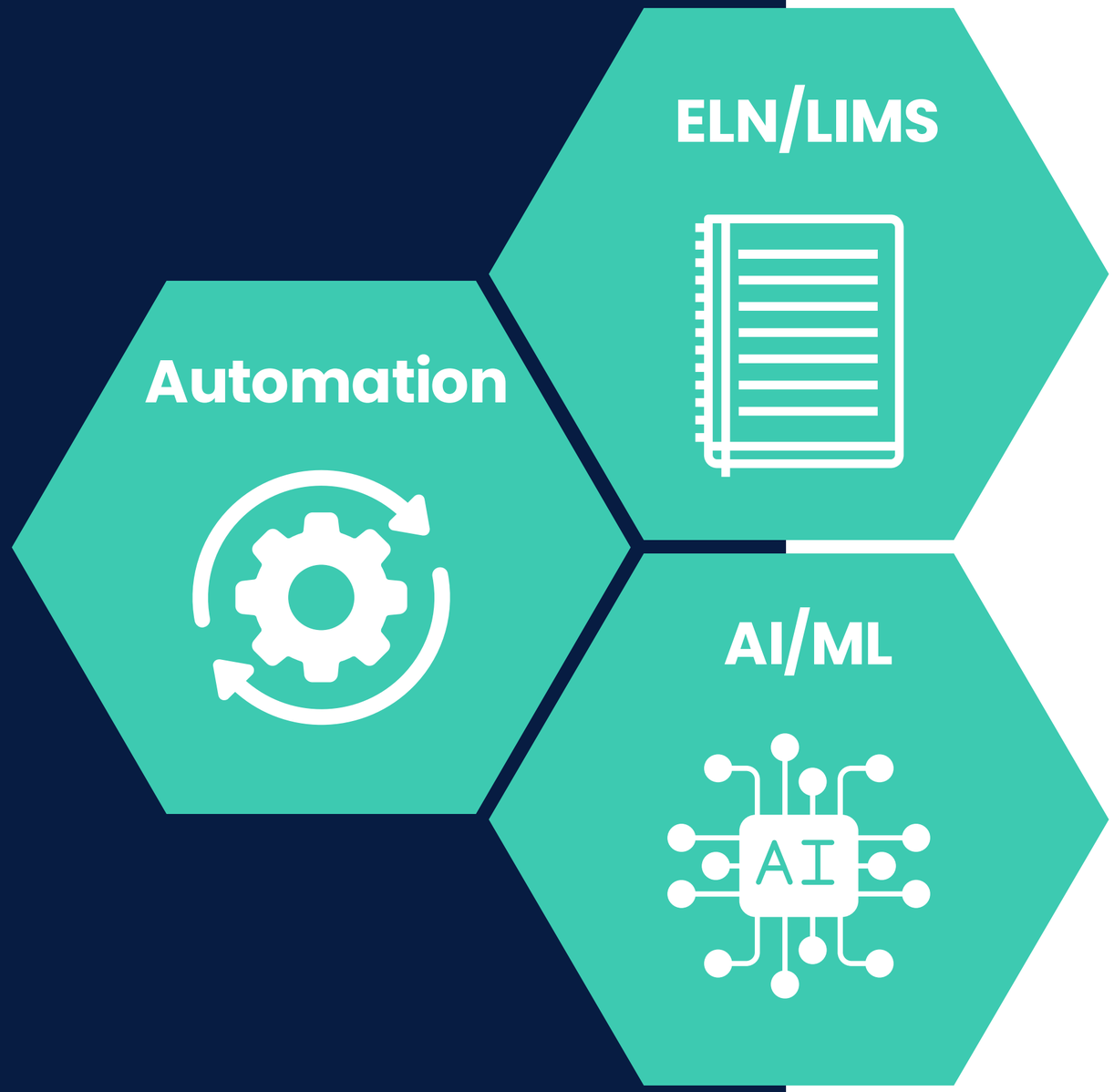
Unstructured data can take many forms

DATA REVIVAL



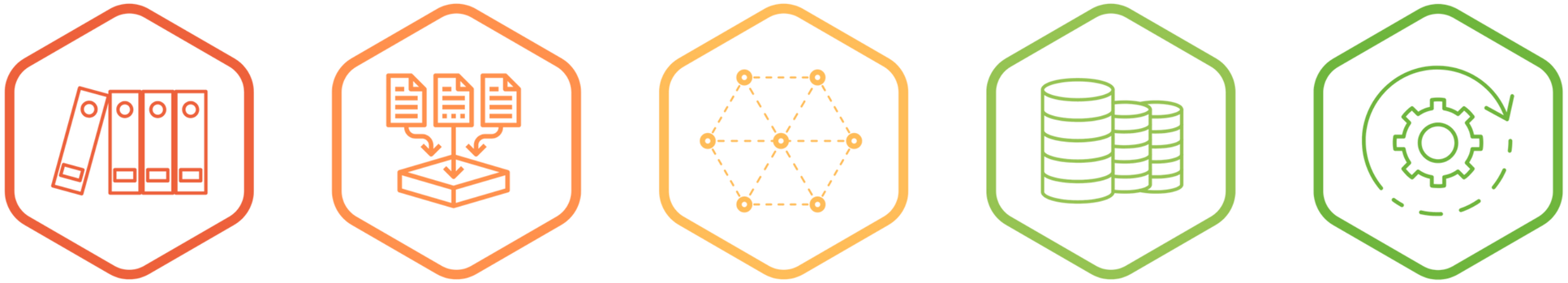
Entry	Format[a]	pH[b]	Time (h)	Yield[c] (%)	FE[d] (%)	ee ret[e] (%)
1	Stirred batch	9.0	11	92	89	92
2	Flow, undivided	9.0	1.2	91	81	97
3	Flow, undivided	8.0	1.2	50	44	98
4	Flow, divided	8.0	0.6	23	N.D.	N.D.
5[f]	Flow, divided	9.0	0.6	92	83	>99

[a]Conditions: 0.1 M (5 mmol) LEV-CH_2OH, 5 mol % ACT, 50 mL of carbonate buffer electrolyte (0.5 M; pH adjusted by varying the ratio of Na_2CO_3/$NaHCO_3$), flow rate = 50 mL min^{-1}, rt. Variable constant currents, set according to the current recorded at 0.7 V vs Ag/AgCl during CV measurement at scan rate of 50 mV/s, corresponded to the following values: batch, I_{app} = 50 mA; undivided flow, I_{app} = 500 mA; divided flow, I_{app} = 1000 mA. electrode dimensions = 10 cm^2 × 0.5 cm. [b]pH controlled by using different ratios of $NaHCO_3$ and Na_2CO_3. [c]^{1}H NMR yields with DMSO as the internal standard. [d]Faradaic Efficiency. [e]% ee ret = (ee of LEV-CO_2H/ee of LEV-CH_2OH)*100; enantiomeric excess (ee) determined by HPLC. [f]The pH of the anode solution maintained by titration with 1.0 M NaOH using a pH controller.

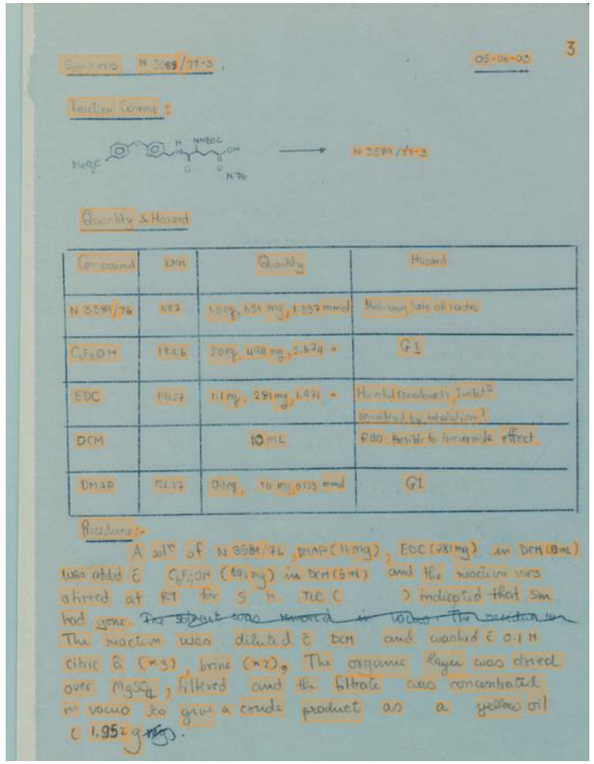


We need to connect this knowledge into new digital systems

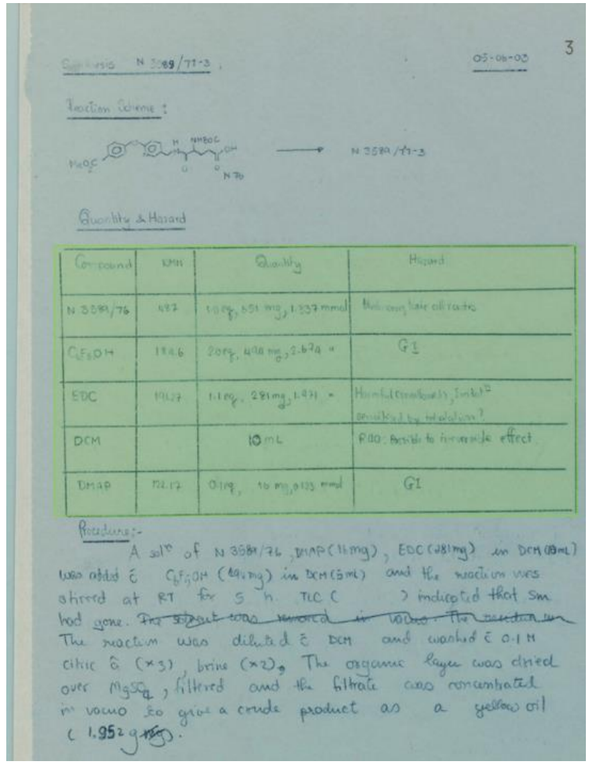
The Data Revival Vision



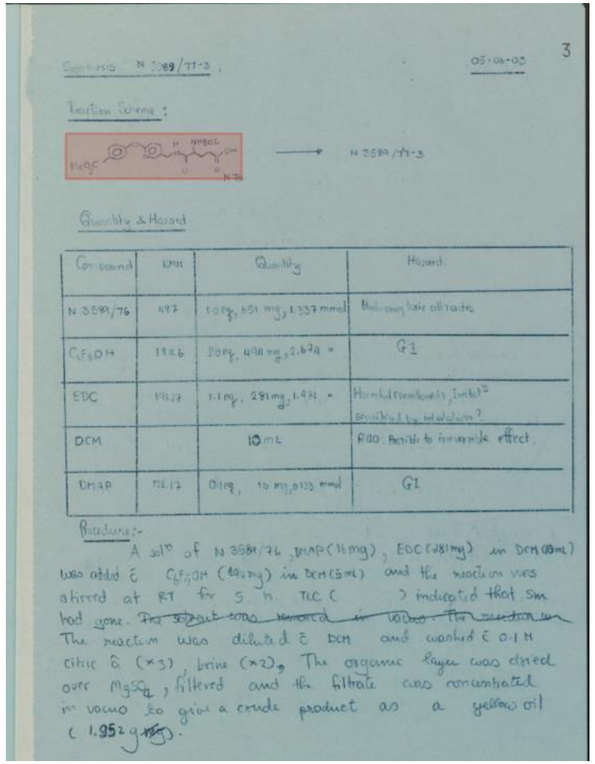
Identification of Data Types



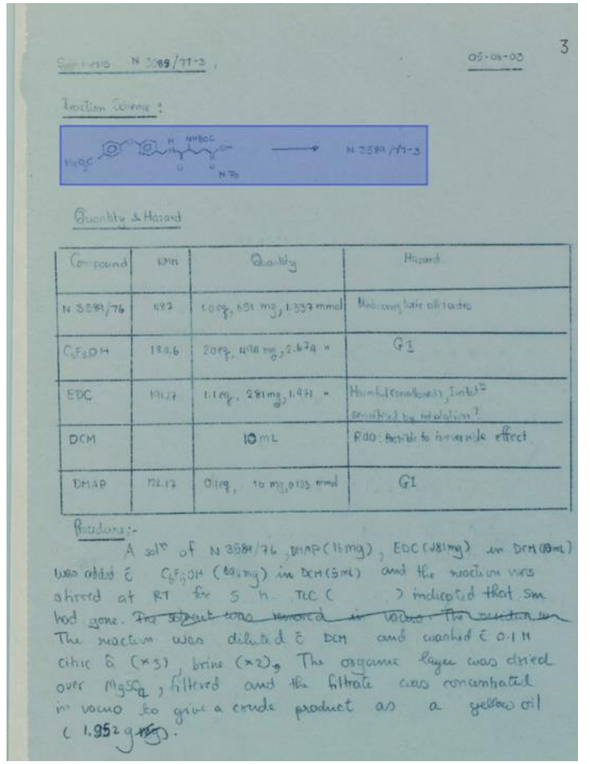
Text Units



Tables

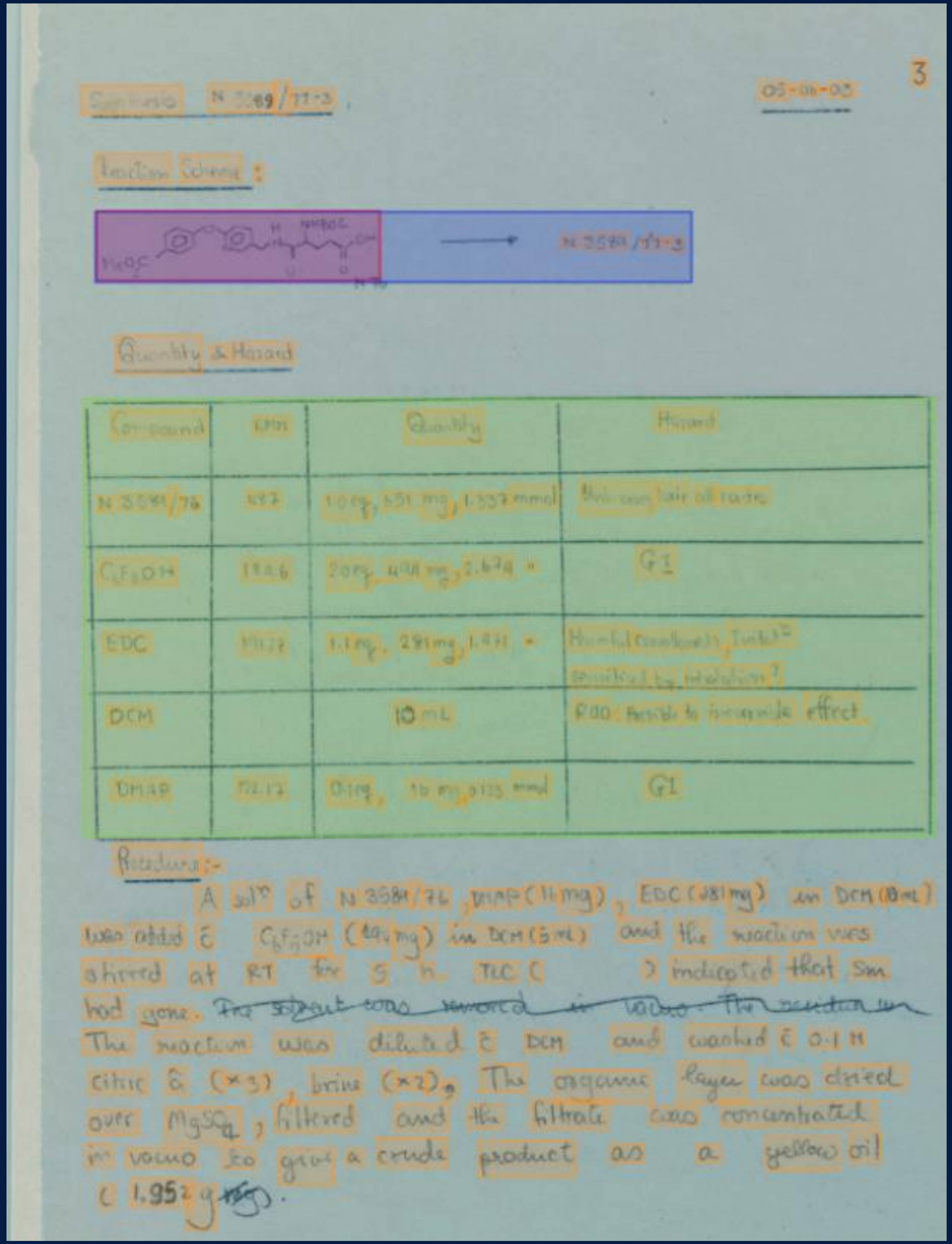


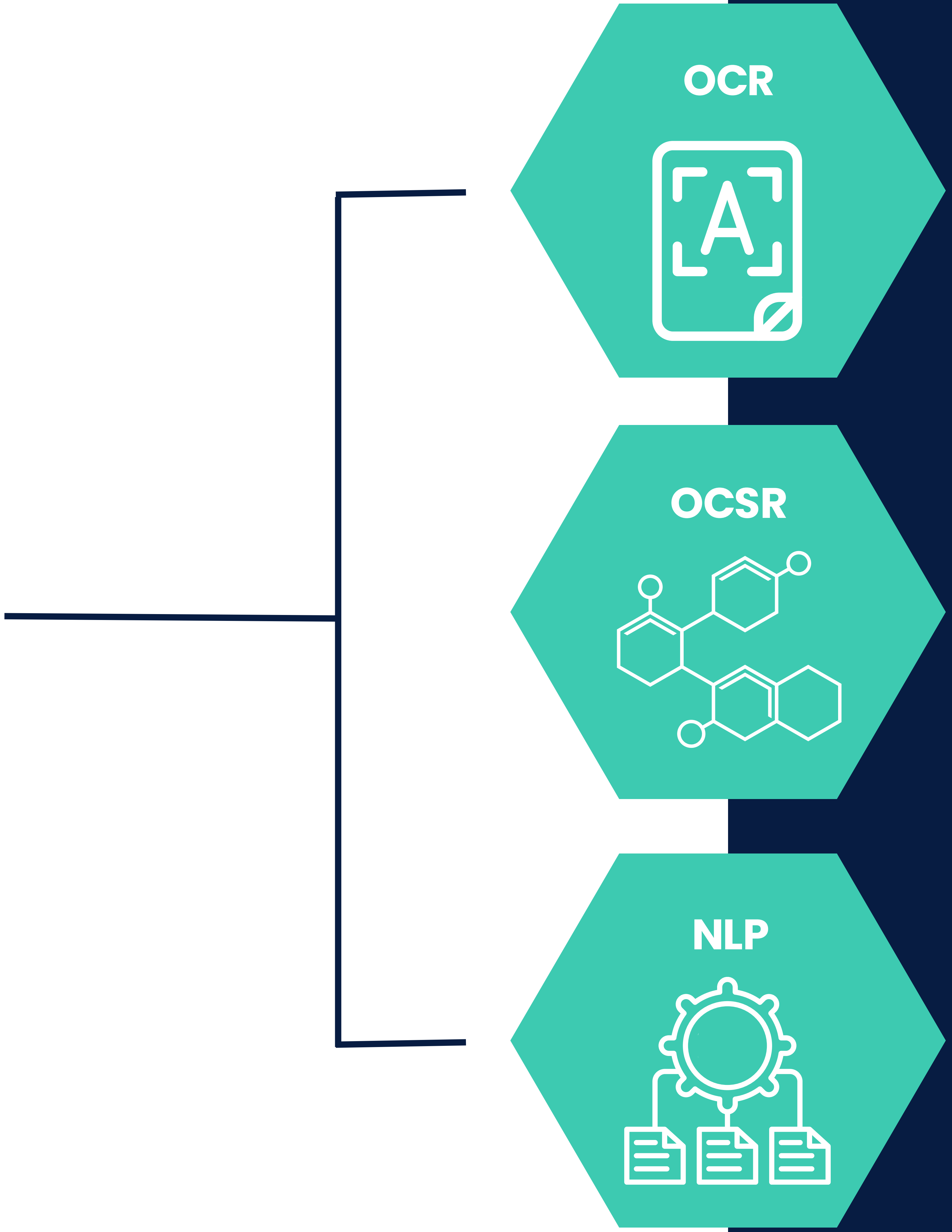
Molecules



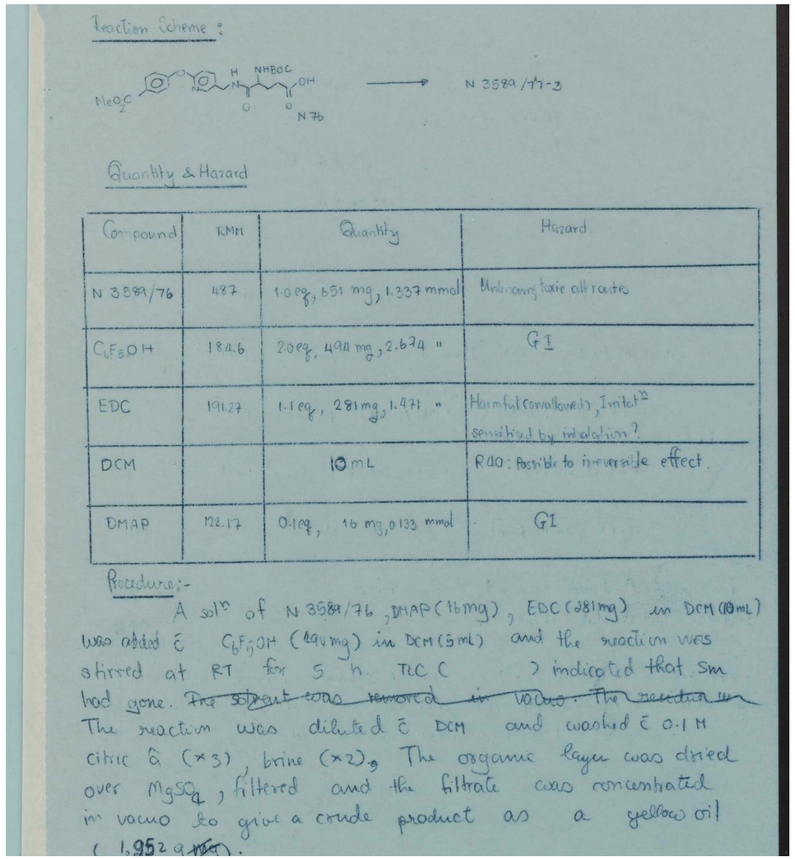
Reaction Schemes

Translation of Data Types





Structuring of Data Types



Reaction Scheme :

N 3589/77-3

Quantity & Hazard

Compound	RMM	Quantity	Hazard
N 3589/76	487	1.0 eq, 651 mg, 1.337 mmol	Unknown, toxic all routes
C_6F_5OH	184.6	2.0 eq, 494 mg, 2.674 "	GI
EDC	191.27	1.1 eq, 281 mg, 1.471 "	Harmful if swallowed, Irritat[12] sensitised by inhalation?
DCM		10 mL	R40: Possible to irreversible effect
DMAP	122.17	0.1 eq, 16 mg, 0.133 mmol	GI

Procedure:-

A soln of N 3589/76, DMAP (16mg), EDC (281mg) in DCM (10mL) was added c C_6F_5OH (494mg) in DCM (5mL) and the reaction was stirred at RT for 5 h. TLC ( ) indicated that SM had gone. ~~The solvent was removed in vacuo. The reaction m~~ The reaction was diluted c DCM and washed c 0.1 M citric a (x3), brine (x2). The organic layer was dried over $MgSO_4$, filtered and the filtrate was concentrated in vacuo to give a crude product as a yellow oil (1.952 g ~~1.95~~).


```
{
    "Document name": "Joe Bloggs - N3389/77-3",
    "Date": "05/06/2003",
    "Molecules": true,
    "Diagrams": false,
    "Text": true,
    "Tables": true,
    "ocr_text": {
    "text1": [

            "A soln of N 3589/76 , DMAP (16mg), EDC (281mg) in DCM (10mL)",
            "was added c C6F5OH (494 mg) in DCM (5 mL) and the reaction was",
            "stirred at RT for 5 in TLC. The reaction was diluted in DCM and washed in 0.1 M",
            "citiric acid (x3), brine (x2). The organic layer was dried",
            "over MgSO4, filtered and the filtrate was concentrated in vaccuo"
    ]
},
    "SMILES": {
        "mol1": {
            "smiles": "OC(CCC(C(NCC(C=C1)=CN=C1OC2=CC=C(C(OC)=O)C=C2)=O)NC(OC(C)(C)C)=O)=O",
            "confidence": 0.94
        }
},
    "Table data": [
        {
            "Compound": "N 3589/76",
            "RMM": "487",
            "Quantity": "1.0 eq, 651 mg, 1.337 mmol",
            "Hazard": "Unknown, toxic all routes",
```

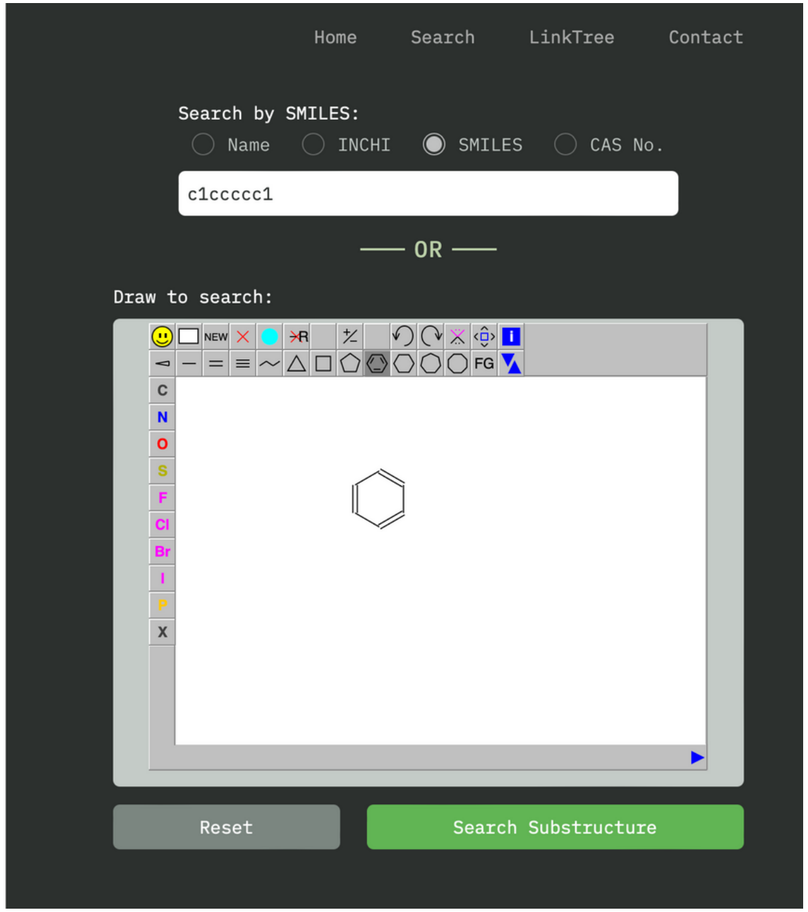
The Data Revival Solution

Search an archive with ease

DATA REVIVAL

Access Lost Chemical Data, Revived By AI Search

Search by identifier, or draw a molecule/fragment to get started.



The Data Revival Solution

We’re building a platform that unlocks the value in your unstrucutred data

DATA REVIVAL

JSON Output

Document Name: 3MG132

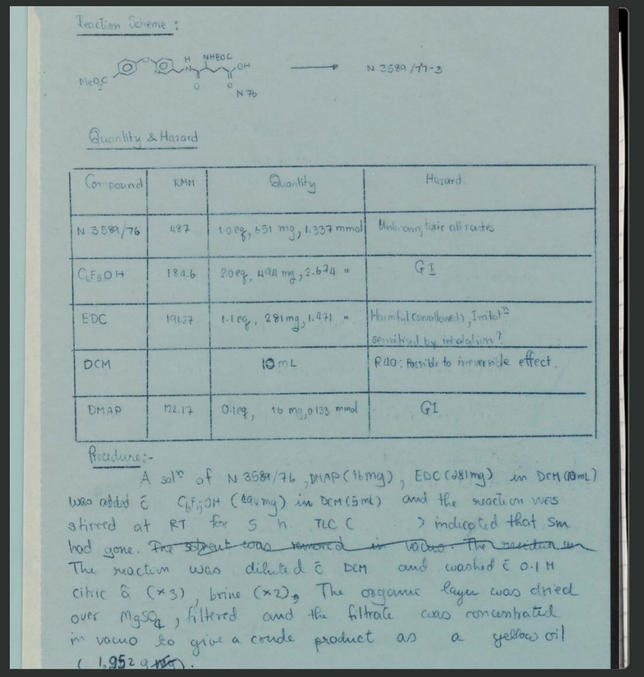
Date: 3/12/92


```
{
    "Document name": "Joe Bloggs – N3389/77–3",
    "Date": "05/06/2003",
    "Molecules": true,
    "Diagrams": false,
    "Text": true,
    "Tables": true,
    "ocr_text": {
    "text1": [

            "A soln of N 3589/76 , DMAP (16mg), EDC (281mg) in DCM (10mL)",
            "was added c C6F5OH (494 mg) in DCM (5 mL) and the reaction was",
            "stirred at RT for 5 in TLC. The reaction was diluted in DCM and washed in 0.1 M",
            "citiric acid (x3), brine (x2). The organic layer was dried",
            "over MgSO4, filtered and the filtrate was concentrated in vaccuo"
    ]
},
    "SMILES": {
        "mol1": {
            "smiles": "OC(CCC(C(NCC(C=C1)=CN=C1OC2=CC=C(C(OC)=O)C=C2)=O)NC(OC(C)(C)C)=O)=O",
            "confidence": 0.94
        }
},
    "Table data": [
        {
            "Compound": "N 3589/76",
            "RMM": "487",
            "Quantity": "1.0 eq, 651 mg, 1.337 mmol",
            "Hazard": "Unknown, toxic all routes",
```


Home Search LinkTree Contact

Original Page



Reaction Scheme :

N 3589/77-3

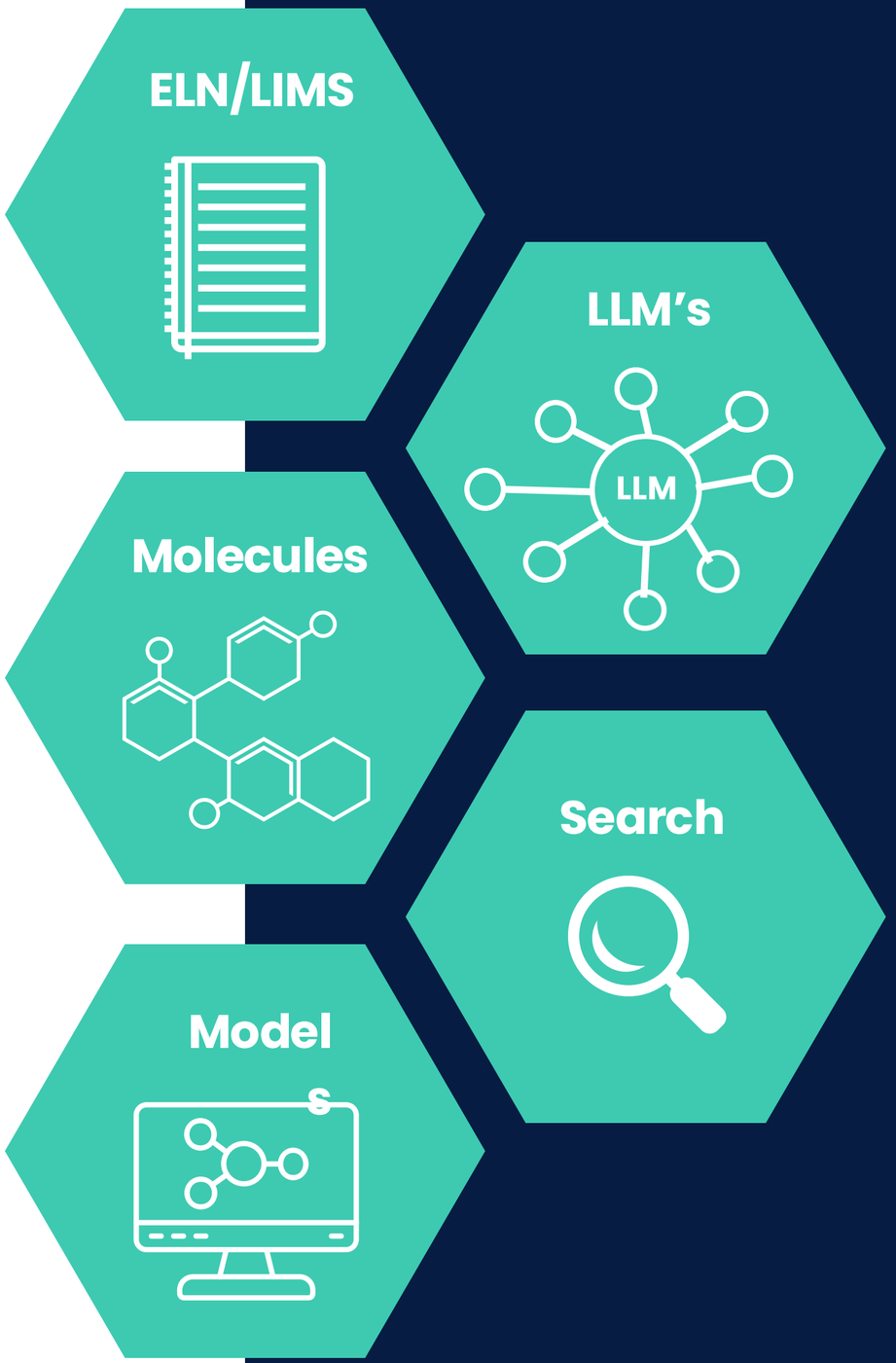
Quantity & Hazard

Compound	RMM	Quantity	Hazard
N 3589/76	487	1.0 eq, 651 mg, 1.337 mmol	Unknown, toxic all routes
C_6F_5OH	184.6	2.0 eq, 494 mg, 2.674 "	GI
EDC	191.27	1.1 eq, 281 mg, 1.471 "	Harmful (small vol), Irritat[2] sensitised by inhalation?
DCM		10 mL	R40: Possible to irreversible effect.
DMAP	122.17	0.1 eq, 16 mg, 0.133 mmol	GI

Procedure:-

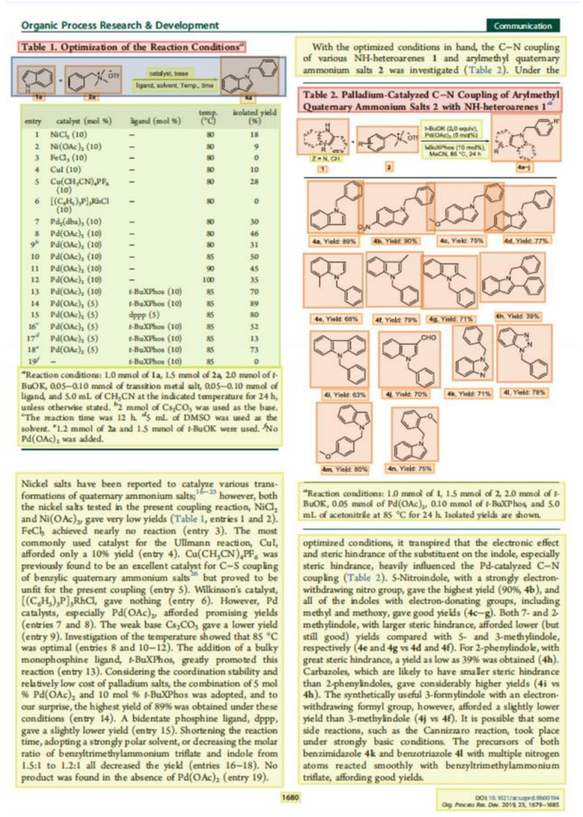
A soln of N 3589/76, DMAP (16mg), EDC (281mg) in DCM (10mL) was added c C_6F_5OH (494mg) in DCM (5mL) and the reaction was stirred at RT for 5 h. TLC ( ) indicated that SM had gone. ~~The solvent was removed in vacuo. The reaction~~ The reaction was diluted c DCM and washed c 0.1 M citric a (×3), brine (×2). The organic layer was dried over $MgSO_4$, filtered and the filtrate was concentrated in vacuo to give a crude product as a yellow oil (1.952 g).

What can be done with this data?



Case Study 1 - The Green Solvents project

Using data to develop new greener solvents



Organic Process Research & Development — Communication

Table 1. Optimization of the Reaction Conditions[a]

entry	catalyst (mol %)	ligand (mol %)	temp. (°C)	isolated yield (%)
1	$NiCl_2$ (10)	–	80	18
2	$Ni(OAc)_2$ (10)	–	80	9
3	$FeCl_3$ (10)	–	80	0
4	CuI (10)	–	80	10
5	$Cu(CH_3CN)_4PF_6$ (10)	–	80	28
6	$[(C_2H_4)_2P]_2RhCl$ (10)	–	80	0
7	$Pd_2(dba)_3$ (10)	–	80	30
8	$Pd(OAc)_2$ (10)	–	80	46
9[b]	$Pd(OAc)_2$ (10)	–	80	31
10	$Pd(OAc)_2$ (10)	–	85	50
11	$Pd(OAc)_2$ (10)	–	90	45
12	$Pd(OAc)_2$ (10)	–	100	35
13	$Pd(OAc)_2$ (10)	*t*-BuXPhos (10)	85	70
14	$Pd(OAc)_2$ (5)	*t*-BuXPhos (10)	85	89
15	$Pd(OAc)_2$ (5)	dppp (5)	85	80
16[c]	$Pd(OAc)_2$ (5)	*t*-BuXPhos (10)	85	52
17[d]	$Pd(OAc)_2$ (5)	*t*-BuXPhos (10)	85	13
18[e]	$Pd(OAc)_2$ (5)	*t*-BuXPhos (10)	85	73
19[f]	–	*t*-BuXPhos (10)	85	0

[a]Reaction conditions: 1.0 mmol of **1a**, 1.5 mmol of **2a**, 2.0 mmol of *t*-BuOK, 0.05–0.10 mmol of transition metal salt, 0.05–0.10 mmol of ligand, and 5.0 mL of CH_3CN at the indicated temperature for 24 h, unless otherwise stated. [b]2 mmol of Cs_2CO_3 was used as the base. [c]The reaction time was 12 h. [d]5 mL of DMSO was used as the solvent. [e]1.2 mmol of **2a** and 1.5 mmol of *t*-BuOK were used. [f]No $Pd(OAc)_2$ was added.

Nickel salts have been reported to catalyze various transformations of quaternary ammonium salts;[16–20] however, both the nickel salts tested in the present coupling reaction, $NiCl_2$ and $Ni(OAc)_2$, gave very low yields (Table 1, entries 1 and 2). $FeCl_3$ achieved nearly no reaction (entry 3). The most commonly used catalyst for the Ullmann reaction, CuI, afforded only a 10% yield (entry 4). $Cu(CH_3CN)_4PF_6$ was previously found to be an excellent catalyst for C–S coupling of benzylic quaternary ammonium salts[20] but proved to be unfit for the present coupling (entry 5). Wilkinson's catalyst, $[(C_2H_4)_2P]_2RhCl$, gave nothing (entry 6). However, Pd catalysts, especially $Pd(OAc)_2$, afforded promising yields (entries 7 and 8). The weak base Cs_2CO_3 gave a lower yield (entry 9). Investigation of the temperature showed that 85 °C was optimal (entries 8 and 10–12). The addition of a bulky monophosphine ligand, *t*-BuXPhos, greatly promoted this reaction (entry 13). Considering the coordination stability and relatively low cost of palladium salts, the combination of 5 mol % $Pd(OAc)_2$ and 10 mol % *t*-BuXPhos was adopted, and to our surprise, the highest yield of 89% was obtained under these conditions (entry 14). A bidentate phosphine ligand, dppp, gave a slightly lower yield (entry 15). Shortening the reaction time, adopting a strongly polar solvent, or decreasing the molar ratio of benzyltrimethylammonium triflate and indole from 1.5:1 to 1.2:1 all decreased the yield (entries 16–18). No product was found in the absence of $Pd(OAc)_2$ (entry 19).

With the optimized conditions in hand, the C–N coupling of various NH-heteroarenes **1** and arylmethyl quaternary ammonium salts **2** was investigated (Table 2). Under the

Table 2. Palladium-Catalyzed C–N Coupling of Arylmethyl Quaternary Ammonium Salts 2 with NH-heteroarenes 1[a]

4a, Yield: 89% · **4b**, Yield: 80% · **4c**, Yield: 75% · **4d**, Yield: 77% · **4e**, Yield: 68% · **4f**, Yield: 79% · **4g**, Yield: 71% · **4h**, Yield: 39% · **4i**, Yield: 63% · **4j**, Yield: 70% · **4k**, Yield: 71% · **4l**, Yield: 78% · **4m**, Yield: 80% · **4n**, Yield: 75%

[a]Reaction conditions: 1.0 mmol of **1**, 1.5 mmol of **2**, 2.0 mmol of *t*-BuOK, 0.05 mmol of $Pd(OAc)_2$, 0.10 mmol of *t*-BuXPhos, and 5.0 mL of acetonitrile at 85 °C for 24 h. Isolated yields are shown.

optimized conditions, it transpired that the electronic effect and steric hindrance of the substituent on the indole, especially steric hindrance, heavily influenced the Pd-catalyzed C–N coupling (Table 2). 5-Nitroindole, with a strongly electron-withdrawing nitro group, gave the highest yield (90%, **4b**), and all of the indoles with electron-donating groups, including methyl and methoxy, gave good yields (**4c**–**g**). Both 7- and 2-methylindole, with larger steric hindrance, afforded lower (but still good) yields compared with 5- and 3-methylindole, respectively (**4e** and **4g** vs **4d** and **4f**). For 2-phenylindole, with great steric hindrance, a yield as low as 39% was obtained (**4h**). Carbazoles, which are likely to have smaller steric hindrance than 2-phenylindoles, gave considerably higher yields (**4i** vs **4h**). The synthetically useful 3-formylindole with an electron-withdrawing formyl group, however, afforded a slightly lower yield than 3-methylindole (**4j** vs **4f**). It is possible that some side reactions, such as the Cannizzaro reaction, took place under strongly basic conditions. The precursors of both benzimidazole **4k** and benzotriazole **4l** with multiple nitrogen atoms reacted smoothly with benzyltrimethylammonium triflate, affording good yields.

1680

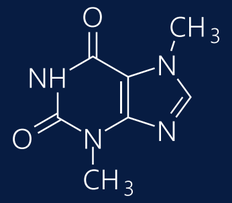
DOI: 10.1021/acs.oprd.9b00194 Org. Process Res. Dev. 2019, 23, 1679–1685

→

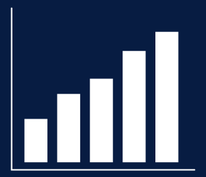
Entry	Catalyst (mol %)	Ligand (mol %)	Temp. (°C)	Isolated Yield (%)	Reactant 1 (Indole 1a)	Specific Reaction Conditions	Reaction Time (h)	Product SMILES
1	NiCl2 (10)	—	80	18	Indole 1a	1.0 mmol of 1a, 1.5 mmol of 2a, 2.0 mmol of t-BuOK, 5.0 mL of CH3CN	24	C1=CC=C(C=C1)CN2C=CC3=CC=CC=C32
2	Ni(OAc)2 (10)	—	80	9	Indole 1a	1.0 mmol of 1a, 1.5 mmol of 2a, 2.0 mmol of t-BuOK, 5.0 mL of CH3CN	24	C1=CC=C(C=C1)CN2C=CC3=CC=CC=C32

400 Lab Notebooks

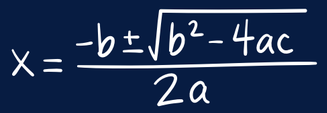
200,000 words of
experimental procedure



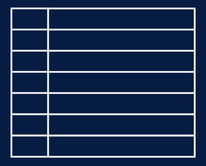
18,000 hand-drawn molecules



8,000 graphs



7,000 equations



5,000 Tables

Case Study 2 – Making the UoS Lab Notebooks FAIR

Thank you!

Samuel Munday
sam@data-revival.com