General introduction to www.c6h6.org

18.1.2021

Philosophy
- Store
- Retrieve original data
- Process
- Extract knowledge
- Export / publish

STORE

DB

RETRIEVE

EXTRACT KNOWLEDGE

Comparison
Similarity
Machine learning

SHARE

Reporting (PDF)
Patent
Export data
Zenodo
...

PROCESS

Chromatogram
Mass analysis
Image analysis

External links

- https://www.c6h6.org

Reusable building blocks


SINGAPORE AIRLINES

Open-source scientific data project

- Started in 2014
- 3 GitHub organisations:
  - https://github.com/cheminfo
  - https://github.com/mljs
  - https://github.com/image-js
- Over 60 developers have contributed
- Over 170 open source MIT projects !
- Over 100’000 download per week !

ml-matrix

Matrix manipulation and computation library.

Installation

```
$ npm install ml-matrix
```

Usage

As an ES module

```
import { Matrix } from 'ml-matrix';

const matrix = Matrix.ones(5, 5);
```

As a CommonJS module

```
const { Matrix } = require('ml-matrix');

const matrix = Matrix.ones(5, 5);
```

API Documentation

Examples

Standard operations

```
const { Matrix } = require('ml-matrix');
```

Install

```
> npm i ml-matrix
```

Weekly Downloads

35,427

Version	License
6.5.3	MIT

Unpacked Size	Total Files
382 kB	40

Issues	Pull Requests
8	1

Homepage

github.com/mljs/matrix

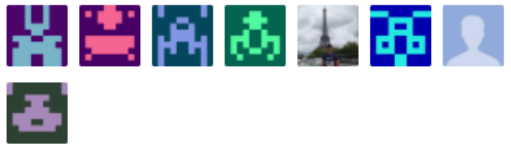
Repository

github.com/mljs/matrix

Last publish

2 months ago

Collaborators



Open-source data processing in the browser

STORE

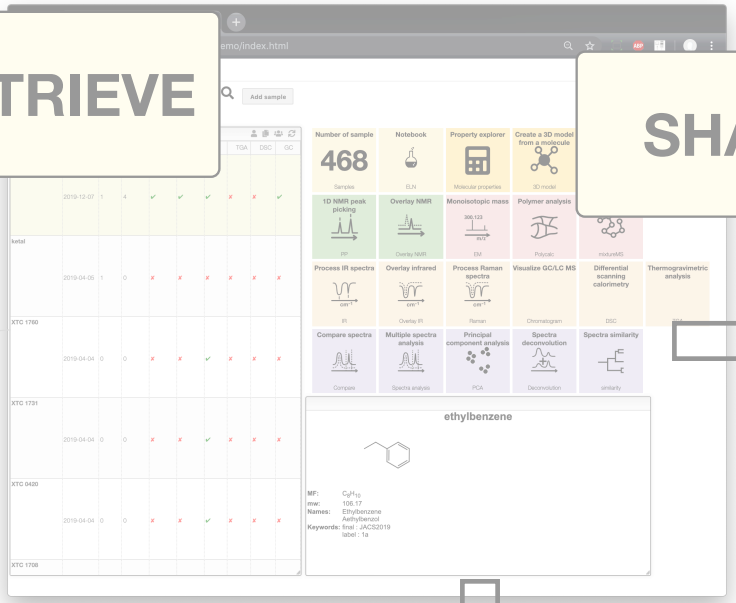
DB

EXTRACT KNOWLEDGE

Comparison
Similarity
Machine learning

RETRIEVE

SHARE

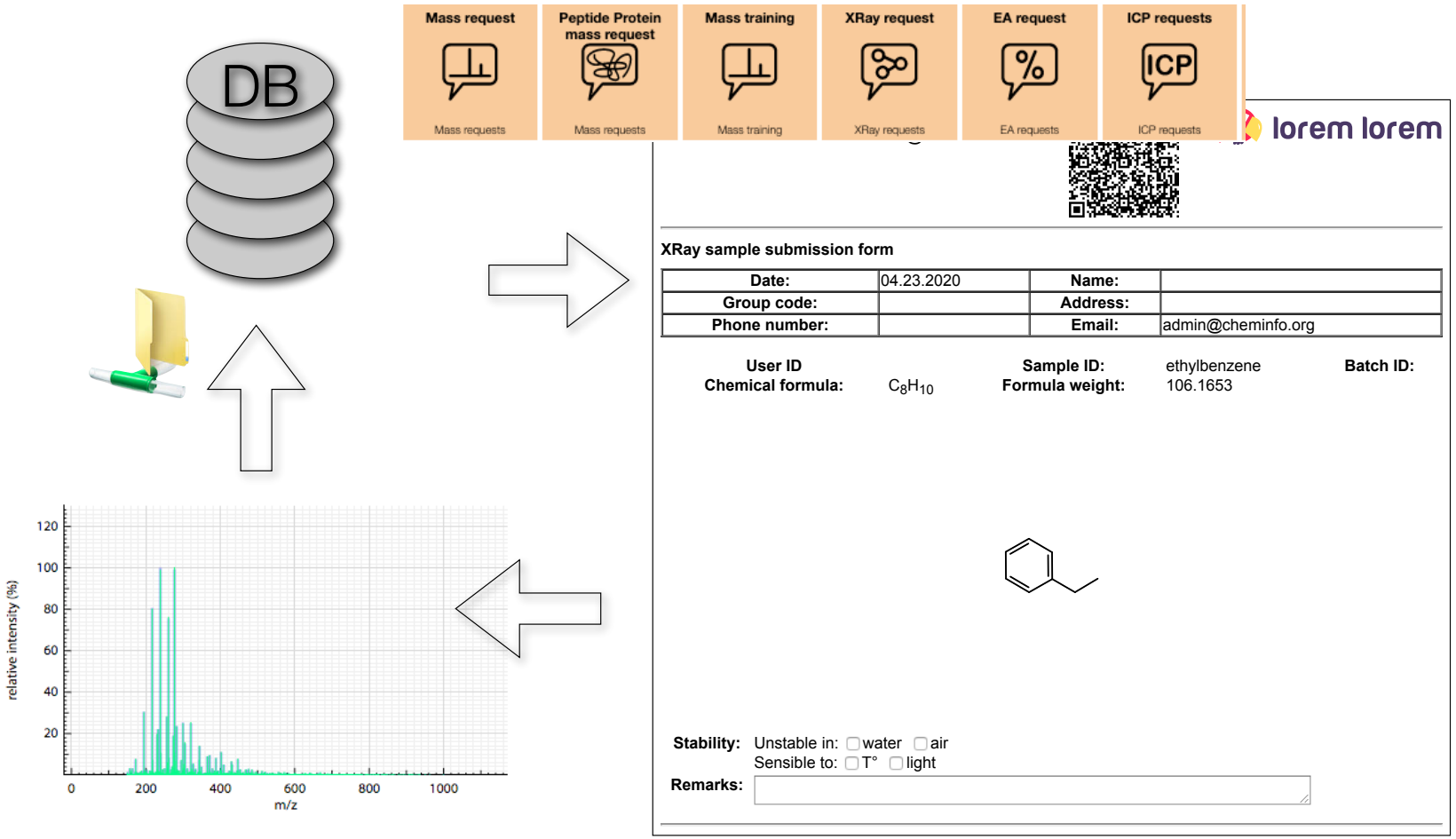


Reporting (PDF)
Patent
Export data
Zenodo
...

PROCESS

Chromatogram
Mass analysis
Image analysis

Traceability and analytical results

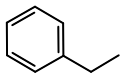


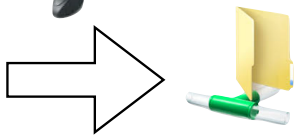
Traceability and analytical results

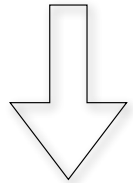
ethylbenzene

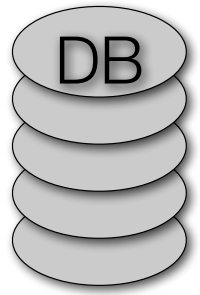
lpatiny@gmail.com_ethylbenzene_

2017-06-19
2020-04-23

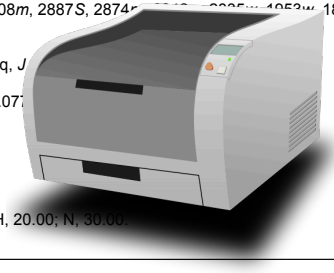








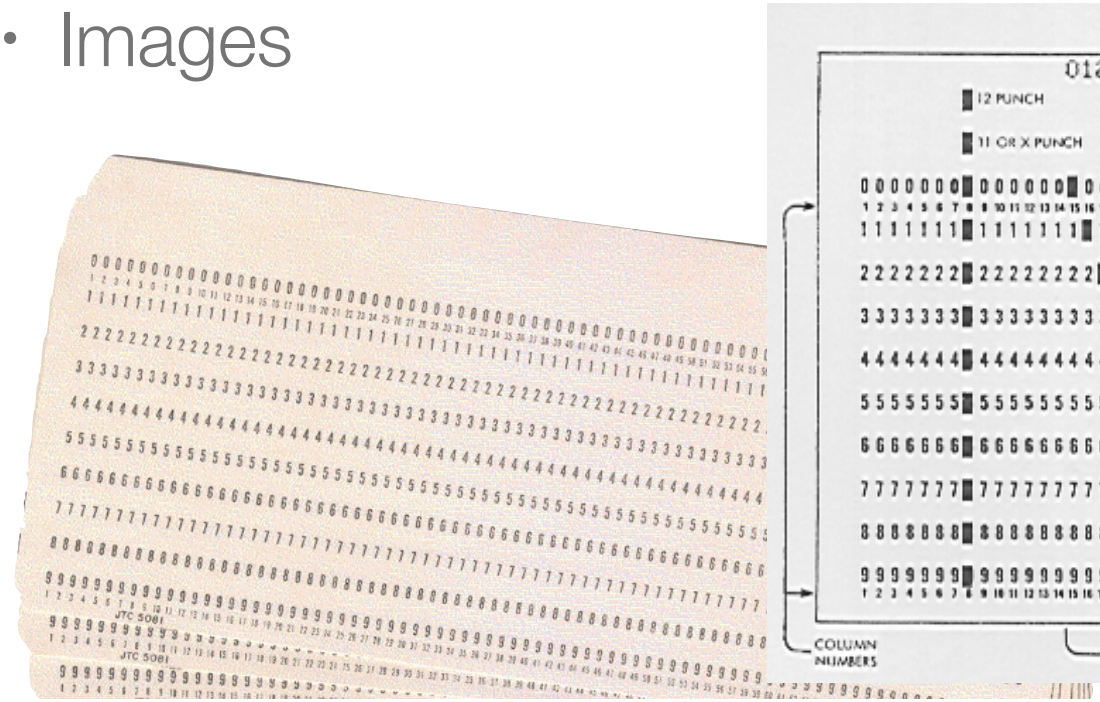
mf	C_8H_{10}
mw	106.1653
em	106.0783
bp (°C)	136 - 137
mp (°C)	-95
density	0.866 - 0.867
IR	IR (cm^{-1}): 3118*w*, 3097*m*, 3077*S*, 3040*S*, 3015*m*, 2978*S*, 2945*S*, 2908*m*, 2887*S*, 2874*m*, 1875*w*, 1812*w*, 1614*m*, 1594*w*, 1502*S*
NMR	^1H NMR (DMSO, 400 MHz): δ 7.28 (2H, m), 7.20 (3H, m), 2.60 (2H, q, *J*
MASS	HRMS (nanochip-ESI/LTQ-Orbitrap) m/z: $[M]^+$ Calcd for $C_8H_{10}^+$ 106.07
	MS m/z: 106 (28), 91 (100), 77 (10), 65 (11), 51 (11), 39 (7)
	HRMS m/z: $[M]^+$ Calcd for $C_8H_{10}^+$ 106.0777; Found 106.0000.
EA	Anal. Calcd for C_8H_{10}: C, 90.51; H, 9.49. Found: C, 52.00; H, 4.30.
	Anal. Calcd for C_8H_{10}: C, 90.51; H, 9.49; N, 0.00. Found: C, 10.00; H, 20.00; N, 30.00.

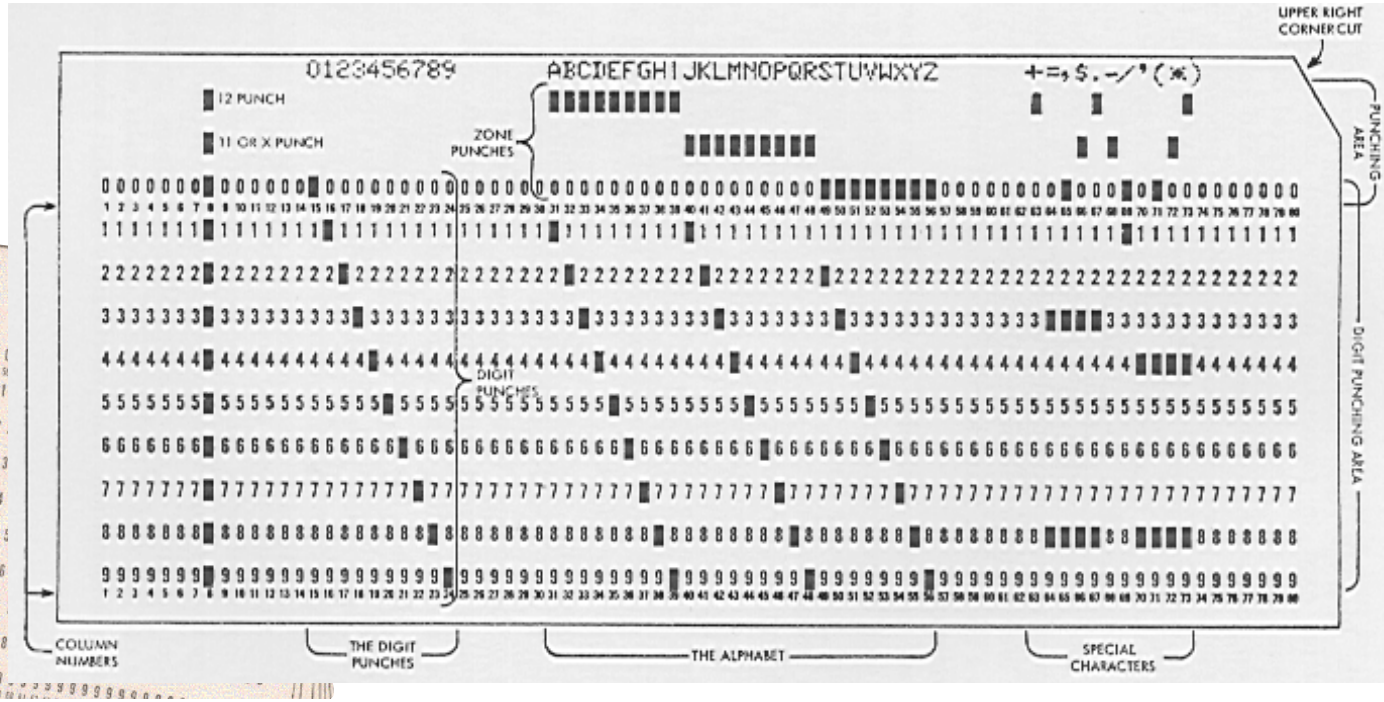


Traceability and chemical library

Perennity: File format

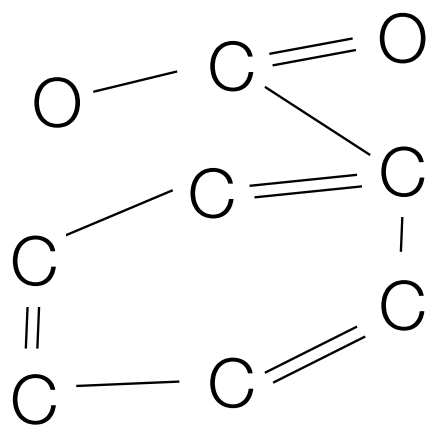
- Chemical structures and reactions
- Spectra and chromatograms
- Images





Connection Table

```
  9  9  0  0  0  0  0  0  0  0999 V2000
   -2.4700    1.3000    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -1.7200   -0.0025    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.7800   -1.0625    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -0.2200   -0.0025    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.5300   -1.3000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.0300   -1.3000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.7825   -0.0025    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.0300    1.2975    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.5300    1.2975    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  2  3  2  0  0  0  0
  2  4  1  0  0  0  0
  4  5  1  0  0  0  0
  5  6  2  0  0  0  0
  6  7  1  0  0  0  0
  7  8  2  0  0  0  0
  8  9  1  0  0  0  0
  4  9  2  0  0  0  0
M  END
```



Molfile

```
Benzoic Acid
  ChemDraw02260010222D

  9  9  0  0  0  0  0  0  0  0999 V2000
   -2.4700    1.3000    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -1.7200   -0.0025    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.7800   -1.0625    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -0.2200   -0.0025    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.5300   -1.3000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.0300   -1.3000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.7825   -0.0025    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.0300    1.2975    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.5300    1.2975    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  2  3  2  0  0  0  0
  2  4  1  0  0  0  0
  4  5  1  0  0  0  0
  5  6  2  0  0  0  0
  6  7  1  0  0  0  0
  7  8  2  0  0  0  0
  8  9  1  0  0  0  0
  4  9  2  0  0  0  0
M  END
```

} 3 lines

1 molecule

Can be read / written by:

- IsisDraw
- ChemDraw
- MarwinSketch
- ACD/ChemSketch
- ...

https://github.com/cheminfo/openchemlib-js

Spectra and chromatograms

Text files

- CSV (comma separated values)
- TSV (tab separated values)

https://github.com/cheminfo/xy-parser

JCAMP-DX

- Jcamp-DX : Joint Committee on Atomic and Molecular Physical Data - Data Exchange
- Text file allowing to store IR, Mass, UV, NMR (1D, 2D, nD), ...
- http://www.jcamp-dx.org

https://github.com/cheminfo/convert-to-jcamp

https://github.com/cheminfo/jcampconverter

Description of a JCAMP-DX file: PEAK TABLE

```
##TITLE=Pd(PPh3)(PPh2C6H4SO3)(CD2Cl2)
##JCAMP-DX=4.24
##DATA TYPE=MASS SPECTRUM
##ORIGIN=
##OWNER=lpatiny
##$USER=lpatiny
##$CONDITIONS=
##XUNITS=M/Z
##YUNITS=relative abundance
##FIRSTX=50
##LASTX=2000
##FIRSTY=0
##LASTY=153300
##NPOINTS=139710
##PEAK TABLE=(XY..XY)
50    10
60    20
100   15
110   100
267   70
...
##END=
```

- LDR : Labeled Data Records
- 80 characters (or less) on one line
- Starts with Data Labels
- unlimited number of lines

Mass format: MzXML MzData MzML

- Widely used exportation format for mass spectrometer (GC / LC / MS / MS
- Some examples on: https://github.com/cheminfo/mzData

https://github.com/cheminfo/mzData

NetCDF (Analytical Instrument Association [AIA])

- Store array-oriented scientific data
- Open standard
- Binary format
- Wide usage, from atmospheric research as well as for mass electra
- Some examples: https://github.com/cheminfo/netcdfjs

https://github.com/cheminfo/netcdfjs

Images

Images formats

	TIFF	GIF	JPEG	PNG	SVG
Vector graphics					X
Bits per channel	8, 16, 32	256 indexed colors	8	8, 16	
Alpha		1 bit		8 or 16 bits	
Compression	yes / no may be destructive	X	X (destructive)	X	

https://github.com/image-js/fast-png

https://github.com/image-js/tiff