ESI: Electronic Supplementary Information

A FAIR Research Data Infrastructure for High-Throughput Digital Chemistry

Alice Gauthier,[a‡] Laure Vancauwenberghe,[b‡] Jean-Charles Cousty,[*a] Cyril Matthey-Doret,[b] Robin Franken,[b] Sabine Maennel,[b] Pascal Miéville,[a] and Oksana Riba Grognuz[b]

[a] Swiss Cat+ West Hub, Ecole Polytechnique Fédérale de Lausanne EPFL, 1015 Lausanne, Switzerland.
[b] Swiss Data Science Center - Open Research Data Engagement & Services, EPFL, INN Building, Station 14, 1015 Lausanne, Switzerland.
[‡] Authors contributed equally.
[*] Communication address: jean-charles.cousty@epfl.ch

S1: Simple workflow example for sample tracking across the experimental pipeline:

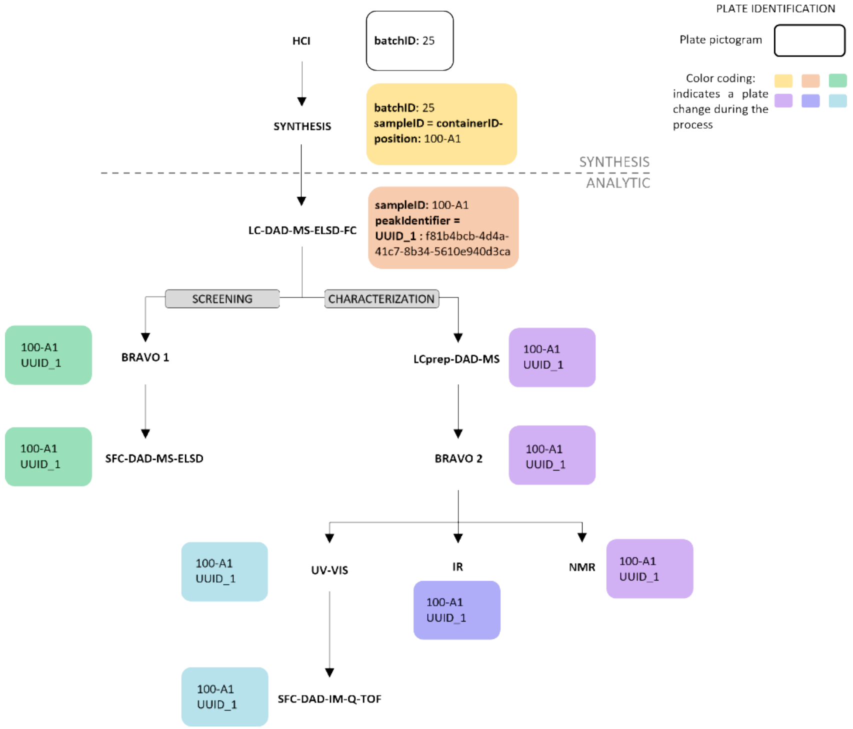


Fig. S1 Color-coded representation of a simple workflow example for sample tracking across experimental stages - Blocks represent successive stages of the experimental pipeline—starting from Human-Computer Interaction (HCI), through synthesis, to analytical characterization. The color of each square encodes the plate identity; the same color, wherever it appears, refers to the same physical plate. Robotic handling ensures that the sample changes container and position at each stage while maintaining identity through persistent tags: *batchID*, *sampleID (containerID-position)*, and *peakIdentifier* (a UUID generated by Agilent software at peak detection).
In this example, a single batch contains one reaction well (one *sampleID=containerID-position* entries corresponding to one reaction well), identified by a unique *sampleID*. The unique reaction is associated with one *peakIdentifier* entrie, corresponding to specific molecular entity synthesised and separated during the process.

S2: Complex workflow example for sample tracking across the experimental pipeline:

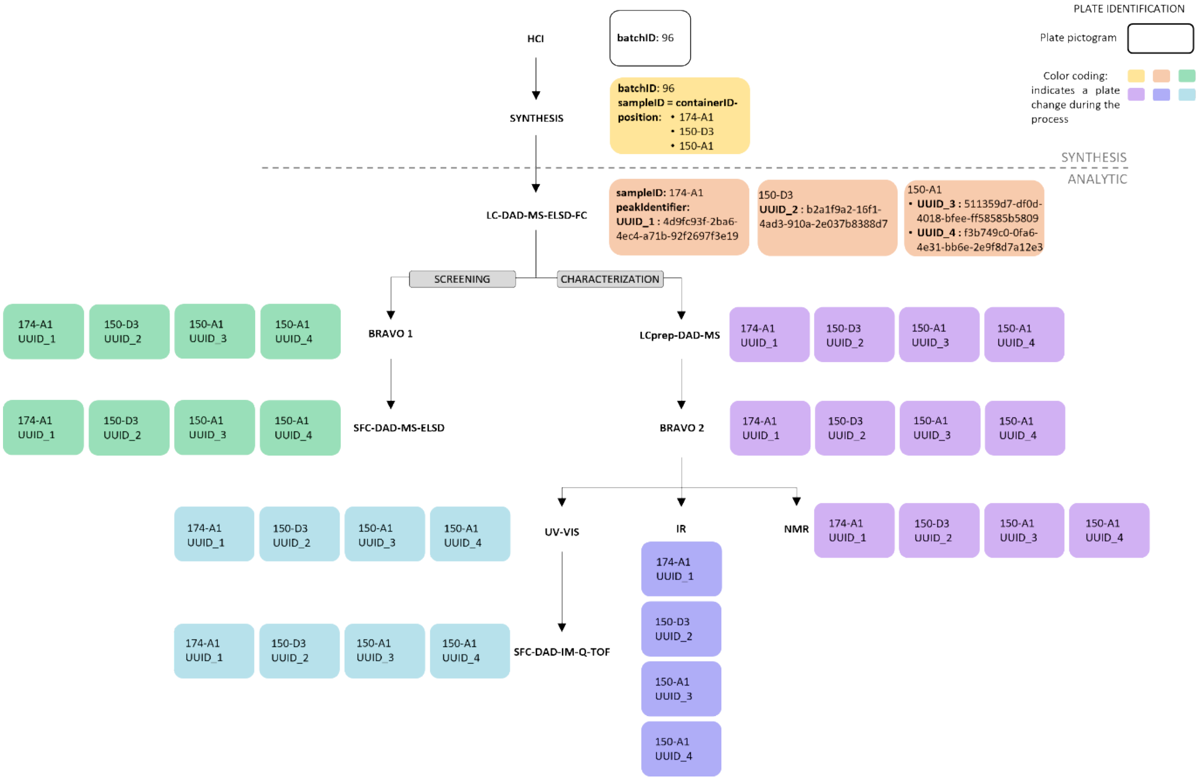
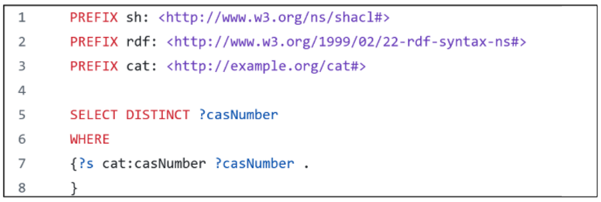


Fig. S2 Color-coded representation of a complex workflow example for sample tracking across experimental stages - Blocks represent successive stages of the experimental pipeline—starting from Human-Computer Interaction (HCI), through synthesis, to analytical characterization. The color of each square encodes the plate identity; the same color, wherever it appears, refers to the same physical plate. Robotic handling ensures that the sample changes container and position at each stage while maintaining identity through persistent tags: *batchID*, *sampleID (containerID-position)*, and *peakIdentifier* (a UUID generated by Agilent software at peak detection).

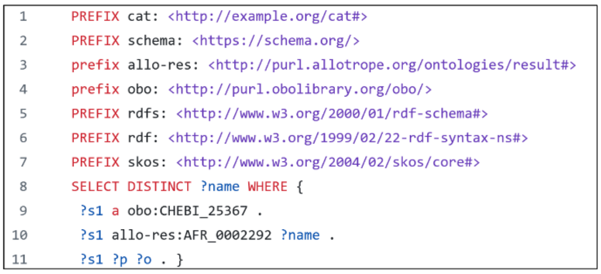
In this example, a single batch contains three distinct reaction wells (three *sampleID=containerID-position* entries corresponding to three reaction wells), each identified by a unique *sampleID*. Each reaction is associated with one or two *peakIdentifier* entries, corresponding to specific molecular entities synthesised and separated during the process. These identifiers enable unambiguous tracking of each compound along a unique path through the workflow, ensuring accurate mapping from synthesis to final analytical characterization.

S3: Example SPARQL Queries for CHEMBORD Metadata Retrieval:



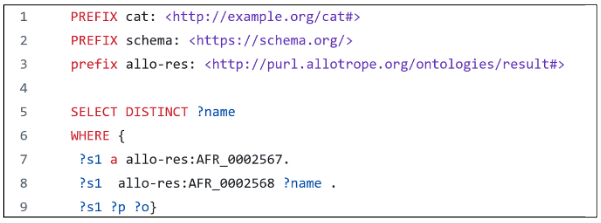
```
PREFIX sh: <http://www.w3.org/ns/shacl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX cat: <http://example.org/cat#>

SELECT DISTINCT ?casNumber
WHERE
{?s cat:casNumber ?casNumber .
}
```

a) CAS numbers Queries



```
PREFIX cat: <http://example.org/cat#>
PREFIX schema: <https://schema.org/>
prefix allo-res: <http://purl.allotrope.org/ontologies/result#>
prefix obo: <http://purl.obolibrary.org/obo/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT DISTINCT ?name WHERE {
 ?s1 a obo:CHEBI_25367 .
 ?s1 allo-res:AFR_0002292 ?name .
 ?s1 ?p ?o . }
```

b) Chemicals Queries

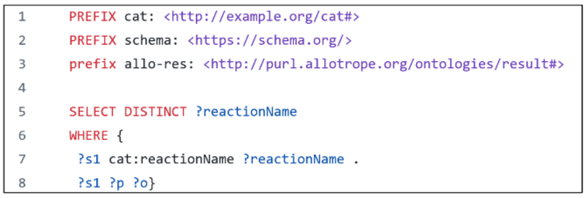


```
PREFIX cat: <http://example.org/cat#>
PREFIX schema: <https://schema.org/>
prefix allo-res: <http://purl.allotrope.org/ontologies/result#>

SELECT DISTINCT ?name
WHERE {
 ?s1 a allo-res:AFR_0002567.
 ?s1  allo-res:AFR_0002568 ?name .
 ?s1 ?p ?o}
```

c) Device documents Queries

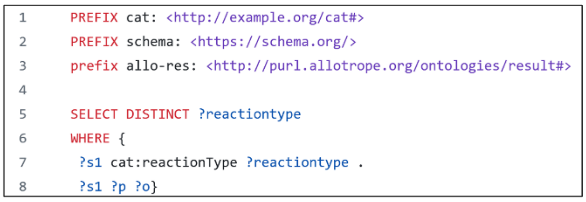
```
PREFIX cat: <http://example.org/cat#>
PREFIX schema: <https://schema.org/>
prefix allo-res: <http://purl.allotrope.org/ontologies/result#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT ?name WHERE {
 ?s1 a ?something .
 ?s1 allo-res:AFR_0001723 ?name .}
```

d) Equipment names Queries



```
PREFIX cat: <http://example.org/cat#>
PREFIX schema: <https://schema.org/>
prefix allo-res: <http://purl.allotrope.org/ontologies/result#>

SELECT DISTINCT ?reactionName
WHERE {
 ?s1 cat:reactionName ?reactionName .
 ?s1 ?p ?o}
```

e) Reaction names Queries



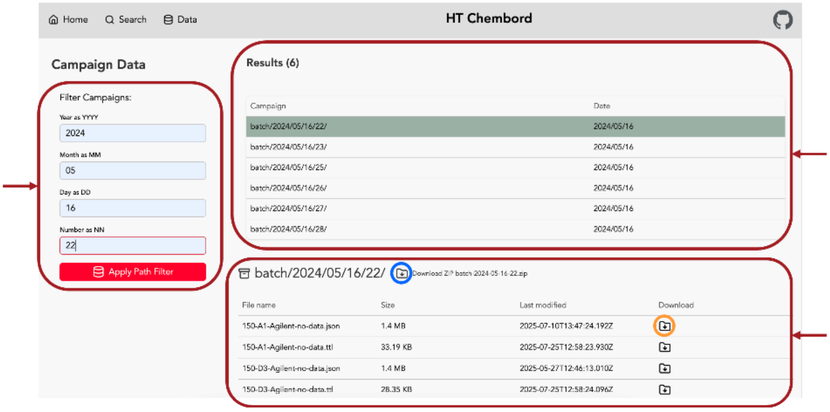
```
PREFIX cat: <http://example.org/cat#>
PREFIX schema: <https://schema.org/>
prefix allo-res: <http://purl.allotrope.org/ontologies/result#>

SELECT DISTINCT ?reactiontype
WHERE {
 ?s1 cat:reactionType ?reactiontype .
 ?s1 ?p ?o}
```

f) Reaction types Queries

Fig. S3 SPARQL query-type exemplars for CHEMBORD metadata retrieval – parametrized *'SELECT DISTINCT'* templates powering the UI search/autocomplete, covering (a) CAS numbers, (b) chemical names, (c) device documents, (d) equipment names, (e) reaction names, and (f) reaction types; shown with declared prefixes (e.g., schema.org, skos, Allotrope Results), final-line parameter insertion, and execution on the QLever endpoint to return paginated results.

S4: Step-by-Step Procedure to Access and Export Data from CHEMBORD UI:

Data Page: Search by Campaign



Campaign Data

Filter Campaigns:

Year as YYYY: 2024

Month as MM: 05

Day as DD: 16

Number as NN: 22

Results (6)

Campaign	Date
batch/2024/05/16/22/	2024/05/16
batch/2024/05/16/23/	2024/05/16
batch/2024/05/16/25/	2024/05/16
batch/2024/05/16/26/	2024/05/16
batch/2024/05/16/27/	2024/05/16
batch/2024/05/16/28/	2024/05/16

batch/2024/05/16/22/ Download ZIP batch-2024-05-16-22.zip

File name	Size	Last modified	Download
150-A1-Agilent-no-data.json	1.4 MB	2025-07-10T13:47:24.192Z	
150-A1-Agilent-no-data.ttl	33.19 KB	2025-07-25T12:58:23.930Z	
150-D3-Agilent-no-data.json	1.4 MB	2025-05-27T12:46:13.010Z	
150-D3-Agilent-no-data.ttl	28.35 KB	2025-07-25T12:58:24.096Z	

1st step: Filtering
The main Data page displays all results stored in S3. Use the search panel to filter by year, month, day, or campaign sequence number.

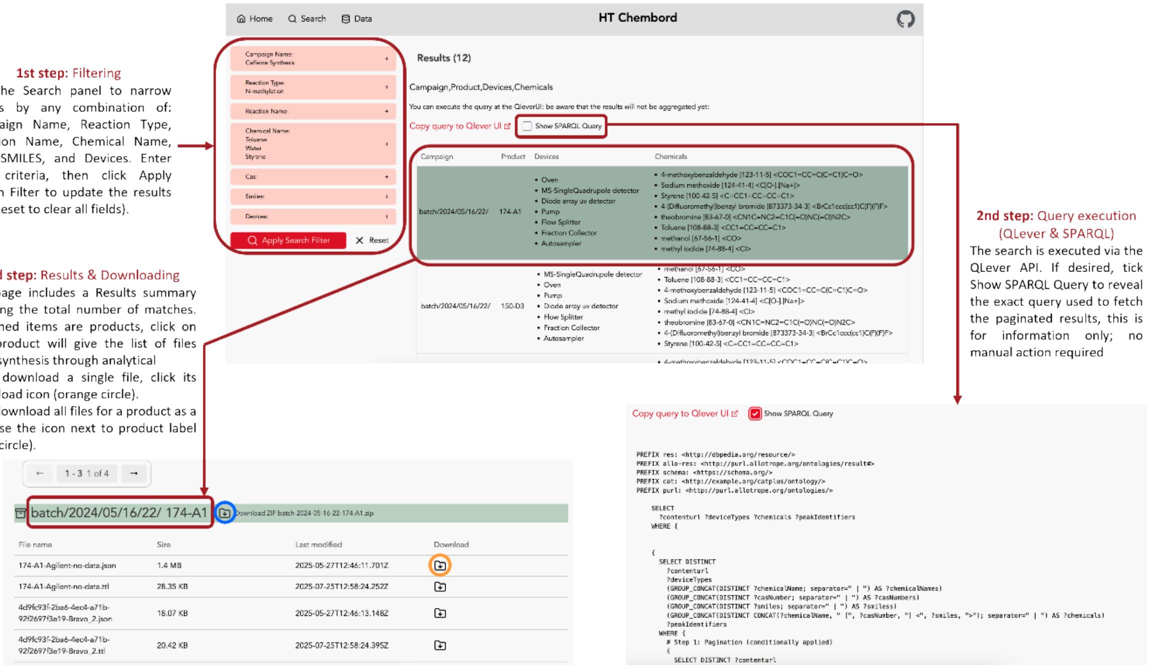
2nd step: Selection
From the filtered results, click the batch, that corresponds to your campaign of interest.

3rd step: Downloading
The full campaign view opens, listing all documents from synthesis through the analytical section.

- To download a single file, click its download icon (orange circle).
- To download the entire campaign as a ZIP, click the icon next to the batch number (blue circle).

Fig. S4.1 Data Page (CHEMBORD UI): Path-filtered campaign selection and downloads - Shows the path filter (Year–Month–Day–Number) on the left, the campaign results list in the center, and the files table for the selected campaign below. The files table lists file name, size, last modified, and download actions. Use the *blue-circled button* to download the entire batch folder as a ZIP; use the *orange-circled icon* to download individual files.

Search Page: By Reaction or Product or Synthetic/Analytical instrument



Campaign Name: Caffeine Synthesis

Reaction Type: N-methylation

Reaction Name:

Chemical Name: Toluene, Water, Styrene

Cas:

Smiles:

Devices:

Results (12)

Campaign,Product,Devices,Chemicals

You can execute the query at the QleverUI: be aware that the results will not be aggregated yet:

Copy query to Qlever UI ☐ Show SPARQL Query

Campaign	Product	Devices	Chemicals
batch/2024/05/16/22/	174-A1	Oven MS-SingleQuadrupole detector Diode array uv detector Pump Flow Splitter Fraction Collector Autosampler	4-methoxybenzaldehyde [123-11-5] <COC1=CC=C(C=C1)C=O> Sodium methoxide [124-41-4] <C[O-].[Na+]> Styrene [100-42-5] <C=CC1=CC=CC=C1> 4-(Difluoromethyl)benzyl bromide [873373-34-3] <BrCc1ccc(cc1)C(F)(F)F> theobromine [83-67-0] <CN1C=NC2=C1C(=O)NC(=O)N2C> Toluene [108-88-3] <CC1=CC=CC=C1> methanol [67-56-1] <CO> methyl iodide [74-88-4] <CI>
batch/2024/05/16/22/	150-D3	MS-SingleQuadrupole detector Oven Pump Diode array uv detector Flow Splitter Fraction Collector Autosampler	methanol [67-56-1] <CO> Toluene [108-88-3] <CC1=CC=CC=C1> 4-methoxybenzaldehyde [123-11-5] <COC1=CC=C(C=C1)C=O> Sodium methoxide [124-41-4] <C[O-].[Na+]> methyl iodide [74-88-4] <CI> theobromine [83-67-0] <CN1C=NC2=C1C(=O)NC(=O)N2C> 4-(Difluoromethyl)benzyl bromide [873373-34-3] <BrCc1ccc(cc1)C(F)(F)F> Styrene [100-42-5] <C=CC1=CC=CC=C1>
			4-methoxybenzaldehyde [123-11-5] <COC1=CC=C(C=C1)C=O>

1st step: Filtering
Use the Search panel to narrow results by any combination of: Campaign Name, Reaction Type, Reaction Name, Chemical Name, CAS, SMILES, and Devices. Enter your criteria, then click Apply Search Filter to update the results (use Reset to clear all fields).

2nd step: Query execution (QLever & SPARQL)
The search is executed via the QLever API. If desired, tick Show SPARQL Query to reveal the exact query used to fetch the paginated results, this is for information only; no manual action required

3rd step: Results & Downloading
The page includes a Results summary showing the total number of matches. Returned items are products, click on one product will give the list of files from synthesis through analytical

- To download a single file, click its download icon (orange circle).
- To download all files for a product as a ZIP, use the icon next to product label (blue circle).

batch/2024/05/16/22/ 174-A1 Download ZIP batch-2024-05-16-22-174-A1.zip

File name	Size	Last modified	Download
174-A1-Agilent-no-data.json	1.4 MB	2025-05-27T12:46:11.701Z	
174-A1-Agilent-no-data.ttl	28.35 KB	2025-07-25T12:58:24.252Z	
4d9fc93f-2ba6-4ec4-a71b-92f2697f3e19-Bravo_2.json	18.07 KB	2025-05-27T12:46:13.148Z	
4d9fc93f-2ba6-4ec4-a71b-92f2697f3e19-Bravo_2.ttl	20.42 KB	2025-07-25T12:58:24.395Z	

Copy query to Qlever UI ☑ Show SPARQL Query

```
PREFIX res: <http://dbpedia.org/resource/>
PREFIX allo-res: <http://purl.allotrope.org/ontologies/result#>
PREFIX schema: <https://schema.org/>
PREFIX cat: <http://example.org/catplus/ontology/>
PREFIX purl: <http://purl.allotrope.org/ontologies/>

    SELECT
      ?contenturl ?deviceTypes ?chemicals ?peakIdentifiers
    WHERE {


    {
      SELECT DISTINCT
        ?contenturl
        ?deviceTypes
        (GROUP_CONCAT(DISTINCT ?chemicalName; separator=" | ") AS ?chemicalNames)
        (GROUP_CONCAT(DISTINCT ?casNumber; separator=" | ") AS ?casNumbers)
        (GROUP_CONCAT(DISTINCT ?smiles; separator=" | ") AS ?smiless)
        (GROUP_CONCAT(DISTINCT CONCAT(?chemicalName, " [", ?casNumber, "] <", ?smiles, ">"); separator=" | ") AS ?chemicals)
        ?peakIdentifiers
      WHERE {
        # Step 1: Pagination (conditionally applied)
        {
          SELECT DISTINCT ?contenturl
```

Fig. S4.2 Search Page (CHEMBORD UI): Query-based product retrieval via QLever with SPARQL script and downloads - Depicts the search panel (filters for campaign name, reaction type, reaction name, chemical name, CAS, SMILES, devices) and the results area with summary count and a table of products including campaign, product, devices, and chemicals. Searches are executed through the QLever API; ticking “*Show SPARQL Query*” reveals the exact query used to obtain the paginated results (for information only). From the product view, download individual files via the *orange-circled icon* or the product folder as a ZIP via the *blue-circled button*.