Karlsruher Institut für Technologie

Institute of Biological and Chemical Systems (IBCS)
Functional Molecular Systems
Institute of Functional Interfaces (IFG)

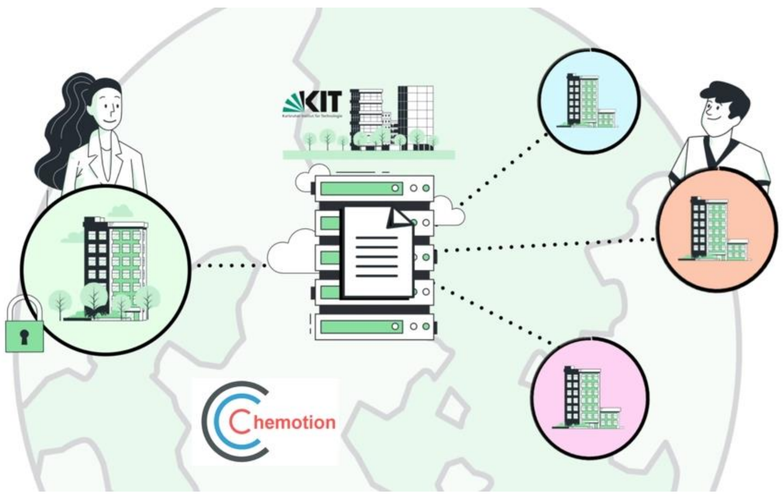
Chemotion Repository

Research Data Management and data publication with Chemotion repository

Nicole Jung, Alexander Welle, Stefan Bräse

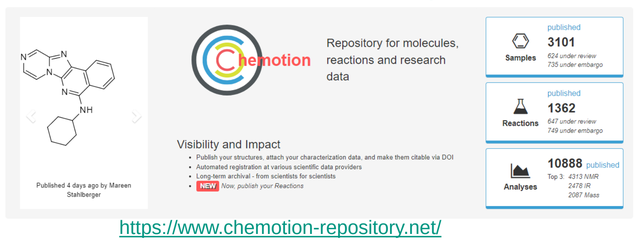
Decentralized ELN-instances and the centrally hosted Chemotion Repository

Data deposition in the repository chemotion complements the traditional publication of research results in scientific journals. The provision of data is possible directly via the upload of data or the transfer of data from Chemotion ELN. The repository is hosted at KIT in Karlsruhe and strongly supported by NFDI4Chem. The repository Chemotion is dedicated to chemistry data and offers discipline specific tools for data deposition and data re-use. The repository is in particular useful for data that describes reactions, samples and analytical measurements.



Chemotion-Repository at one glance

An open access repository
and database for chemistry data

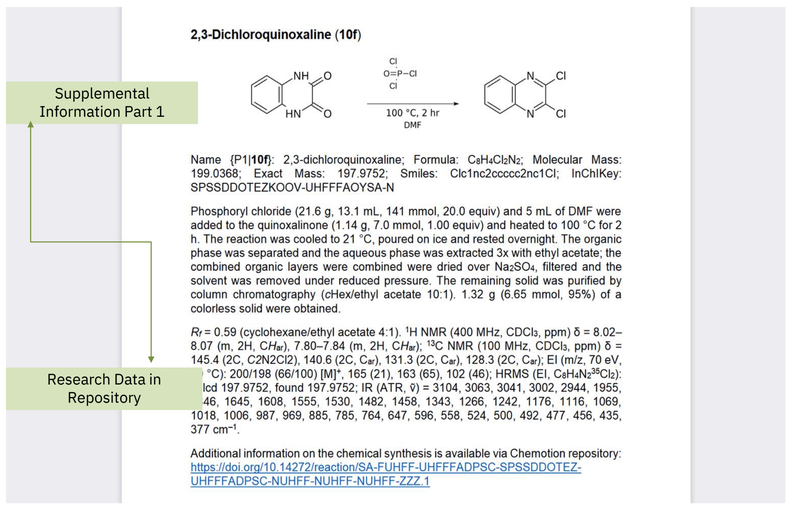


https://www.chemotion-repository.net/

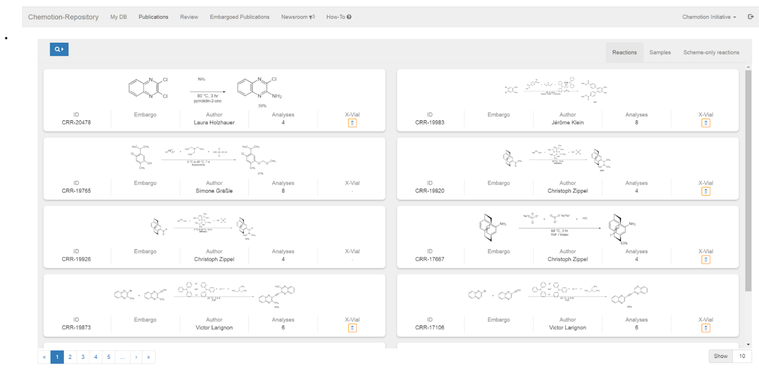
Findable Accessible Interoperable Reusable

Link publication data with research data

We established an automatic process to link the data in the publication with research data in the SI. This is done by embedding the DOIs for samples, reactions or analytical data into the SI. The SI can be generated from the data that is deposited in the repository. The downloaded documents contain the DOIs at the end of each experimental description to that they belong.



Typical examples from chemotion repo



Publications

The data publications initiated by Chemotion Repository are registered to the databases PubChem and DataCite.

The first data publication of Chemotion Repository started in 2014 and included molecules and their analytical datasets. In 2018, the publication of reactions, summarizing also all previously available content such as molecules, has been introduced.

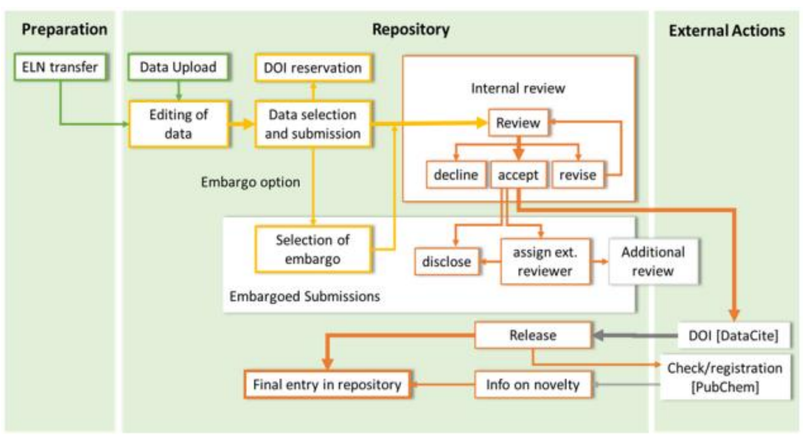
Currently, the available publication types include reactions, samples and analysis; we plan to extend the application from chemistry disciplines to related disciplines such as materials science, biochemistry, biology and many more.

Publication Process / Workflow

All submissions enter an internal reviewing process that focuses on plausibility and formal issues of the submission. The internal review can result in a request to revise the submission. After improving and resubmitting the data, the reviewers can either decline, accept or send back the submission to the author for another review.

For accepted submissions, the repository mints a Digital Object Identifier (DOI) automatically, so the dataset can be cited and referenced.

The released molecule structures are checked for their presence in the database PubChem to give information on the novelty of the compounds.

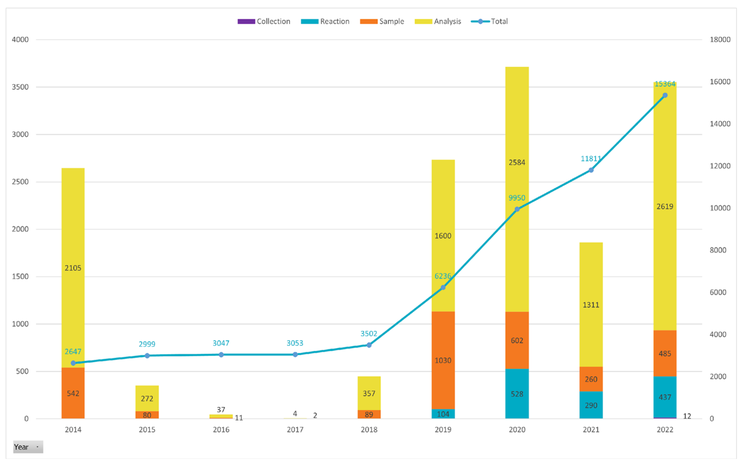


Adding data collections and embargos

Chemotion supports to add an embargo on the data that is deposited. This allows to collect research data prior to its publication. The embargo can be released to make the research data visible to the public whenever the scientists are ready.

While the data is under an embargo, additional external reviewers with read level can be assigned, allowing the review of the data while the publication process is on hold.

Once the embargo collection is released, a DOI for the whole collection is gained and can be added to the data availability statement



until 25.09.2022
Total: cumulative number from 2014

Type	Ontologies	published	under review	under embargo
Collection	-	13	132	132
Reaction	-	1,362	739	753
Sample	-	3,101	724	739
Analysis	NMR	4,313	2,813	1426
Analysis	IR	2,478	1,021	488
Analysis	Mass	2,087	1,380	706
Analysis	DEPT	633	661	320

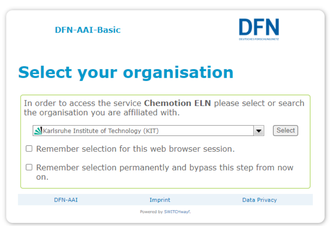
until 02.10.2022

Versioning

The version support for data published in chemotion repository is in the development stage. The planned features enables users to update/ upgrade published publications, keep all records and allow the researchers to easily cite either specific versions or all versions of a record.

Support federated single sign-on

Federated login to Chemotion Repository is supported for the three dominant open web standards for identity: OAuth, SAML and OpenID Connect. Users can login to Chemotion Repository by using, for example, ORCiD and GitHub accounts. Furthermore, Chemotion Repository uses Shibboleth which allows users to login with their home institution credentials when their institutions participate in the DFN-AAI Federation.



APIs / OAI-PMH Harvesting

Chemotion repository provides a public API for users to access the published data. This s currently reworked to facilitate the search, grouping and download of the repository content.

The repository enables the exposure and harvesting of the published data via OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).

Chemotion

NFDI4Chem
ENHANCE YOUR DATA.

Funded by DFG Deutsche Forschungsgemeinschaft
German Research Foundation
NFDI4Chem is supported by DFG under project number 441958208